Efficient inference of clade-specific diversification rates with probabilistic approaches

Bjørn Tore Kopperud

München 2025

Efficient inference of clade-specific diversification rates with probabilistic approaches

Bjørn Tore Kopperud

Dissertation zur Erlangung des Doktorgrades
an der Fakultät für Geowissenschaften
der Ludwig-Maximilians-Universität München

vorgelegt von
Bjørn Tore Kopperud

München, 1 Juli 2025

Erstgutachter: Prof. Dr. Sebastian Höhna
Zweitgutachter: Prof. Dr. Gudrun Kadereit
Tag der mündlichen Prüfung: 15.01.2026

Contents

Acknowledgements **ix**

List of published papers and submitted manuscripts **xi**

Summary **xvii**

1 Introduction **1**

1.1 General motivation 1
1.2 Historical perspective on birth-death models 2
1.3 Modeling species diversification with a stochastic branching process 3
1.4 Probability of observing the reconstructed tree 6
1.5 Inferring speciation and extinction rates 11
1.6 Extensions of the phylogenetic birth-death model 15
1.7 Aims of the thesis 17

2 CRABS: Congruent rate analyses **19**

2.1 Summary 20
2.2 Introduction 21
2.3 Theory and usage 22
2.4 Discussion and conclusions 32
2.5 Acknowledgments 34
2.6 Data Availability 34

3 Rapidly changing diversification rates and non-identifiability **35**

3.1 Significance statement 36
3.2 Abstract 37
3.3 Introduction 38
3.4 Results 40
3.5 Discussion 46
3.6 Concluding remarks 49
3.7 Materials and methods 49

4 Shifts in the tempo of origination **55**
4.1 Abstract . . . 56
4.2 Introduction . . . 57
4.3 Materials and methods . . . 59
4.4 Results . . . 72
4.5 Discussion . . . 85
4.6 Conclusion . . . 89

5 Diversification rate shifts across the Tree of Life **91**
5.1 Abstract . . . 92
5.2 Estimating branch rates and rate shift events across the Tree of Life . . . 94
5.3 Upshifts vs. downshifts in net diversification . . . 97
5.4 Variation in diversification shift rate across the Tree of Life . . . 98
5.5 Shifts in speciation vs. extinction rate . . . 99
5.6 Caveats in diversification rate estimates . . . 101
5.7 Conclusions . . . 102

6 Bayesian branch-specific diversification methods **103**
6.1 Abstract . . . 104
6.2 Lay summary . . . 104
6.3 Introduction . . . 106
6.4 Methods . . . 108
6.5 Results and discussion . . . 112

7 RevGadgets: visualizing RevBayes analyses **123**
7.1 Abstract . . . 124
7.2 Introduction . . . 125
7.3 Phylogenies . . . 127
7.4 Posterior estimates of numerical parameters . . . 129
7.5 Ancestral-state estimates . . . 129
7.6 Diversification rates . . . 133
7.7 Model adequacy . . . 133
7.8 Conclusions . . . 136

A Supplementary Material for Chapter 2 **139**
S2.1 Calculation of the equivalent rate functions within the congruence class . . . 139
S2.2 Accuracy of the method . . . 141

B Supplementary Material for Chapter 3 **145**
S3.1 Extended datasets . . . 145
S3.2 Detailed exploration of the congruent models for three empirical datasets . . . 150
S3.3 Evaluating alternative trend thresholds . . . 154
S3.4 Posterior samples . . . 156
S3.5 Convergence assessment . . . 157
S3.6 Congruent speciation rates are more constrained than extinction rates . . . 161
S3.7 Hypothetical models: alternative extinction rate . . . 163
S3.8 Hypothetical models: alternative speciation rate . . . 174
S3.9 How steep is sufficiently steep? . . . 183

C Supplementary Material for Chapter 4 **185**
S4.1 Statistical inference of ancestral diversification rates . . . 185
S4.2 Notation . . . 191
S4.3 Extended methods . . . 192
S4.4 RevBayes equivalence . . . 197
S4.5 Variance of branch-specific speciation rates . . . 199
S4.6 The likelihood surface . . . 202
S4.7 Estimation error in branch-specific diversification rates . . . 202

D Supplementary Material for Chapter 5 **209**
S5.1 Materials and methods . . . 211
S5.2 Supplementary figures . . . 214
S5.3 Supplementary text . . . 216

E Supplementary Material for Chapter 6 **237**
S6.1 Empirical datatsets . . . 238
S6.2 Convergence . . . 239
S6.3 Methods comparison . . . 240
S6.4 Quantifying uncertainty in rate estimates . . . 245
S6.5 Results for individual chronograms . . . 248

Bibliography **249**

Acknowledgements

First and foremost I want to thank my supervisor Sebastian Höhna. I am grateful for all the guidance and opportunities you have given me, in terms of learning about phylogenetic models, statistical inference, and learning about how to be an academic. If there were two takeaways from what I learned, I would first say that asking "how can you simulate under the model?" is an incredibly powerful pedagogical tool. For the second, I learned better strategies when it comes to problem solving. Often, it is important to internalize and firmly believe that a problem is possible to solve. With this mindset, the means to solve the problem (be it analytical solutions, approximations or computer simulations) become less of a hurdle, and it is easier to not give up.

Furthermore, I wish to thank all past and present members of the Höhna lab. Luiza Fabreti, Killian Smith, Ronja Billenstein and Allison Hsiang, thank you for the warm welcome when I first arrived in Germany. It was a bit unfortunate with the pandemic, but despite that I really enjoyed our time together, and all the interesting discussions we had. Since I stayed in the lab for a long time, I got to see several "generations" of the lab. Thus, I also want to thank Alessio Capobianco, Wenjie Zhu, John T. Clarke, Nicola Heckeberg, Walker Sexton and Haoqing Du, for being such wonderful office mates and always being up for a chat. Basanta Khakurel and Gustavo Darlim—the greenhouse partners—thank you as well for always being good company. Priscilla Lau, I am glad to have been a small part of your project on macroevolutionary adaptation. If there is one thing I will miss the most from the Höhna lab, it might be our daily lunch ritual.

I would also like to thank all the students, postdocs and professors that I got a chance to get to know from outside the lab, particularly the department of Earth and Environmental Sciences, Paleontology & Geobiology and the Bavarian state collection. Joëlle van der Sprong, Luis Porras, Manon Hullot, Ömer Coskun, Dan Mills, Baran Karapunar, Charalampos Kevrekidis, Marit Renken, Alexandra Fernandes, Max Kellermann, Elena Cuesta, Victor Beccari—thank you for all the good moments, and I apologize to those whom I failed to mention. I also want to thank Gert Wörheide, Dirk Erpenbeck and Oliver Rauhut for organizing various cross-lab activities, including joint lab meetings and the student seminars, where I could learn more

about sponge biology and paleontology.

My decision to start the doctoral study did not come without prior influence. Therefore, I want to thank my mentors, colleagues and friends from my time at the University of Oslo. Without you I surely would not have decided to study evolution in the first place. Thomas F. Hansen, thank you for introducing me to the world of academic research, about probabilistic models of evolution, and macroevolution in general. As per your teaching, I have tried my best to report parameter values with an accompanying description of the units. I am also indebted to Olja Toljagić and Kjetil L. Voje for their guidance and support. I also want to thank Masahito Tsuboi for being my informal mentor, and for giving me the opportunity to work with all kinds of deers with strange antlers.

I am also grateful for my time in the bryozoology lab at the Natural History Museum in Oslo. There, I indirectly learned a lot about paleontology, and the importance of fossils has stuck with me ever since. Lee Hsiang Liow, thank you for giving me the opportunity of working with you, your lab and your precious cheilostome bryozoans. Even though I was unduly unqualified as a natural language processing researcher, you supported me and believed in me, and I really appreciate that. I would also like to thank the other bryozoologists (and non-bryozoologists) Scott Lidgard, Franziska Franeck, Russell Orr, Mali Ramsfjell, Emanuela Di Martino, Trond Reitan, Arthur Porto, Marianne Nilsen and others for the interesting times.

I also want to express my gratitude to the numerous scientific collaborators that I have had the opportunity to work with. Andrew F. Magee, thanks for your always analytical comments about diversification rate methods. Mark Grabowski, thanks for the interesting work we did on comparative primate brain-size evolution. Jason Pienaar, it was great to finally meet you at Evolution in Montreal—your ideas about models of multiple traits evolving together and in response to each other are really interesting. Carrie Tribble, thank you for giving me the opportunity to contribute to RevGadgets, and also Jesús Martínez-Gómez and Michael Song for the project about branch-rate methods. Savannah Olroyd, thank you for allowing me to be a small part of your study on the evolution of hearing ability in fossil synapsids.

I also want to mention my study group from the University of Oslo, including Simen Fredriksen, Solveig Haug, Arrian Karbassioon, Andreas Langeland Jenssen and Even Mikkelsen Bjørgesæter. Thank you for all the fun skiing trips and barbecue meetings, both in the past and hopefully in the future.

Finally, I want to thank my parents Astrid and Øystein, as well as my brother Trond Ivar Kopperud for their support.

List of published papers and submitted manuscripts

The following is a list of published papers and submitted manuscripts that are part of the thesis. These are all related to phylogenetic diversification rate analyses.

Höhna S, **BT Kopperud**, AF Magee. 2022. CRABS: Congruent rate analyses in birth–death scenarios.
Published in: Methods in Ecology and Evolution 13:2709–2718. (**Chapter 2**)

Kopperud, BT, AF Magee, S Höhna. 2023. Rapidly changing speciation and extinction rates can be inferred in spite of nonidentifiability.
Published in: Proceedings of the National Academy of Sciences 120. (**Chapter 3**)

Kopperud BT, S Höhna. 2025. Phylogenetic Estimation of Branch-Specific Shifts in the Tempo of Origination.
Published in: Systematic Biology, 74(6), 985-1006. (**Chapter 4**).

Kopperud BT, A Capobianco, JT Clarke, L Palazzesi, S Höhna. 2026. The nature and prevalence of diversification rate shifts across the Tree of Life.
Published in: Evolution Letters (in press). (**Chapter 5**).

Martínez-Gómez J, MJ Song, CM Tribble, **BT Kopperud**, WA Freyman, S Höhna, CD Specht, CJ Rothfels. 2024. Commonly used Bayesian diversification methods lead to biologically meaningful differences in branch-specific rates on empirical phylogenies.
Published in: Evolution Letters 8:189–199. (**Chapter 6**)

Tribble CM, WA Freyman, MJ Landis, JY Lim, J Barido-Sottani, **BT Kopperud**, S Höhna, MR May. 2022. RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes.
Published in: Methods in Ecology and Evolution 13:314–323. (**Chapter 7**)

During my stay as a doctoral student, I also contributed to the following published articles. Overall, they are concerned with investigations about the evolution of morphological characters (continuous and discrete). There is also one article about natural language processing of biodiversity data. Since they diverge from the main topic of diversification rate analyses, they are not part of the thesis.

Tsuboi M, **BT Kopperud**, C Syrowatka, M Grabowski, KL Voje, C Pélabon, TF Hansen. 2020. Measuring complex morphological traits with 3D photogrammetry: a case study with deer antlers.
Published in: Evolutionary Biology 47:175–186.

Kopperud, BT, S Lidgard, LH Liow. 2022. Enhancing georeferenced biodiversity inventories: automated information extraction from literature records reveal the gaps.
Published in: PeerJ 10:e13921.

Grabowski M, **BT Kopperud**, M Tsuboi, TF Hansen. 2023. Both diet and sociality affect primate brain-size evolution.
Published in: Systematic Biology 72:404–418.

Grabowski M, J Pienaar, KL Voje, S Andersson, J Fuentes-González, **BT Kopperud**, DS Moen, M Tsuboi, J Uyeda, TF Hansen. 2023. A cautionary note on "a cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies".
Published in: Systematic Biology 72:955–963.

Tsuboi M, **BT Kopperud**, M Matschiner, M Grabowski, C Syrowatka, C Pélabon, TF Hansen. 2024. Antler Allometry, the Irish Elk and Gould revisited.
Published in: Evolutionary Biology 51(1):149–165.

Olroyd SL, **BT Kopperud**. 2025. Allometry of sound reception structures and evidence for a mandibular middle ear in non-mammalian synapsids.
Published in: Evolution 79(6):905–921.

Author contributions

Chapter 2: SH and AF conceived the study. SH, BTK and AF developed the R package `CRABS`. SH and BTK drafted the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Chapter 3: BTK and SH conceived the study. All authors contributed to methods development. BTK performed the analyses. BTK and SH drafted the manuscript. All authors revised and approved the manuscript.

Chapter 4: BTK and SH contributed to methods development. BTK performed the analyses. BTK and SH wrote the manuscript.

Chapter 5: BTK and SH conceived the study. BTK and SH contributed to methods development. BTK, AC, JTC and LP contributed to data collection and curation. BTK performed the analyses. BTK and SH drafted the manuscript. All authors revised and approved the manuscript.

Chapter 6: JM-G, MJS, and CMT designed and performed the analyses. JM-G, MJS, CMT, SH, BTK, WAF, CDS, and CJR wrote the manuscript.

Chapter 7: CMT and MRM designed the R package. All authors contributed code and examples. CMT and MRM drafted the manuscript. All authors revised and approved the final version of the manuscript.

Summary

Understanding the origin and maintenance of biodiversity is one of the most important topics in the natural sciences. The concept of a species is a fundamental unit in biology, and major lines of research use species richness—or the number of species for a particular group of shared origin—as a measure of the degree of biodiversity. Unlike in modern civilization where there are genealogies and censuses describing recent history, there is limited available information about the history of species and populations. Therefore, the evolutionary history of the species we see today are best seen as hypotheses. Importantly, we are interested in learning about questions such as i) during which time periods did species originate more often, ii) which clades speciated more often than other clades, and iii) how often do species go extinct. To answer these questions, we can model how species originate and go extinct using a stochastic branching process. The most widely used branching process is the birth-death model, and variants thereof. The thesis contains seven chapters that study varying aspects of the birth-death model. This includes theoretical properties as well as empirical findings.

The **introduction** provides an overview of the most important aspects of the birth-death model as used in phylogenetics, an example of how to fit the birth-death model, and a summary of the aims of the thesis.

Chapter 2 is a description of an R-package called `CRABS` (Höhna et al., 2022). In 2020, Louca and Pennell published an article about the identifiability of time-varying phylogenetic diversification models. Under certains assumptions, they demonstrated that speciation and extinction rates (as a function of time) are not jointly identifiable, even if the phylogeny is arbitrarily large. This means that, given a hypothesis for time-varying speciation and extinction rates, it is possible to construct an infinite amount of alternative hypotheses (a so-called congruent class of models) that explain the data equally well. Although the issue was already pointed out by Kubo and Iwasa (1995), the renewed interest in identifiability led to several discussions in the diversification research community (Morlon et al., 2022; Helmstetter et al., 2022). Our impression was that the findings of Louca and Pennell (2020b) were interesting but the implications were not well understood. `CRABS` is a method that aims to investigate the congruence class of time-varying diversification rate models by sampling from the congruence class, and visualizing common patterns of the alternative hypotheses/models. The aim of `CRABS` is to provide tools for gaining better intuition of the properties of a specific congruence

class. Later, Andréoletti and Morlon (2023) extended the functionality of `CRABS` by allowing for more sophisticated sampling of alternative models.

Chapter 3 investigates specific diversification rate patterns in more detail (Kopperud et al., 2023), using the methods of `CRABS`. We also analyze the congruence class for empirical phylogenies (New World/African true parrots, tyrant flycatchers and woodpeckers). The impact of non-identifiability on hypothesis testing depends on the shapes of the time-varying speciation and extinction rates. Specifically, we find that if there is a signal for time-constant diversification rates, then the congruent models can be in disagreement with each other. If the speciation and extinction rates were rapidly changing, however, then the congruent models are in agreement. Thus, we argue that rapidly changing diversification rates can be inferred in spite of non-identifiability. Legried and Terhorst (2022, 2023) used a more theoretical approach for analyzing the non-identifiability issue. Using mathematical proofs, they demonstrated that time-varying speciation and extinction rates are jointly identifiable under specific conditions. Specifically, if the rate functions are piecewise polynomial with finite number of pieces (this includes piecewise-constant and spline functions), then the model is identifiable.

Chapter 4 veers from hypotheses about time-varying diversification rates, and instead focuses on hypotheses that certain clades in a phylogeny diversified more rapidly or slowly than others. Analyses of clade-specific or lineage-specific diversification rates can broadly be separated in two approaches. First, the change in diversification rate can be associated with a specific hypothesis, such as "did the evolution of sexual reproduction lead to more rapid diversification?" This implies we have knowledge of the predictor trait at the tips of the phylogeny, and we can use trait-dependent diversification rate models (e.g., Maddison et al., 2007) to test the hypothesis. Second, clade-varying diversification rates can be investigated without a specific hypothesis in mind. Chapters 4, 5 and 6 are of the latter approach. Kopperud and Höhna (2025) introduce a new method and software implementation (`Pesto`) for inference of branch-specific diversification rates. `Pesto` uses the same generative process (`LSBDS`) as the one presented by Höhna et al. (2019). One of the main challenges with the `LSBDS` model as implemented in `RevBayes` is the time it takes to run an analysis. For exceptionally species-rich phylogenies, obtaining enough samples of the posterior distribution (i.e., simulating using Markov-chain Monte Carlo; MCMC) may take several months or even years. Consequently, this has limited the application of the method to phylogenies of poor or moderate species richness. The `Pesto` method overcomes this limitation by i) fixing the parameters of the model (it is an empirical Bayes approach), and ii) by using a dynamic programming algorithm to estimate branch-specific diversification rates and branch-specific rate shift events. The dynamic programming algorithm is known in Bayesian inference literature (Pearl, 1988; Teo et al., 2025), and has been leveraged in estimating ancestral discrete character states (Revell, 2025), but we are the first to use it for branch-specific diversification rate estimation. The software implementation is also optimized specifically to be able to fit the

model on species-rich phylogenies. The ability of fitting the birth-death-shift model rapidly even to species-rich phylogenies ($> 20,000$ taxa) has two advantages. First, it becomes more feasible to investigate hypotheses of diversification rate shifts in species-rich phylogenies, such as on the scale of flowering plants. Second, it becomes possible to set up elaborate simulation and inference experiments, where we can learn more about the behaviour and potential biases of the method.

In **Chapter 5**, we apply the `Pesto` method to a broad selection of empirical phylogenies, spanning over 300,000 described species. We investigate several hypotheses about the prevalence and nature of diversification rate shifts. Overall, we find that all empirical phylogenies experienced at least one rate shift event, indicating that diversification rates are heterogeneous across the Tree of Life. Moreover, we find that the tempo of diversification rate shifts is higher in younger clades than in older clades. We also infer substantially more upshifts in net-diversification than downshifts. This is in strong contrast to the model assumptions, where we assume that downshifts and upshifts are equally probable. However, simulation studies reveal that downshifts are much more difficult to infer as opposed to upshifts. We are also interested in investigating whether a rate shift event is due to a change in the speciation or extinction rate, although we find that the result is strongly driven by the prior assumptions.

Chapter 6 is comparison of several methods for inferring branch-specific diversification rates (Martínez-Gómez et al., 2024). Following the development of *Bayesian analysis of Macroevolutionary Mixtures* (`BAMM`, Rabosky et al. 2013; Rabosky 2014) and its prolific use, several methods for Bayesian inference of branch-specific diversification rates have emerged. These include the cladogenetic diversification shift model (`CLaDS2` Maliet et al., 2019), the multi-type birth-death model (`MTBD` Barido-Sottani et al., 2020) and the lineage-specific birth-death-shift model (`LSBDS` Höhna et al., 2019). These methods are conceptually similar, yet they have different model assumptions and/or inference behaviour. Since there are a wealth of available methods, it is not trivial for empiricists to decide which method to use. The article aims to review `BAMM`, `CLaDS2`, `MTBD`, `LSBDS` and `Pesto` by testing whether they yield different results when applied to the same empirical phylogenies. Martínez-Gómez et al. (2024) find that the results are meaningfully different. Notably, `BAMM`, `LSBDS` and `Pesto` are most similar. `MTBD` also gives comparable results, but `CLaDS2` is most divergent, as expected due to the assumption of cladogenetic diversification rate shifts.

Chapter 7 is a description of an R-package called `RevGadgets` (Tribble et al., 2022). It is designed to be used together with the software `RevBayes` (Höhna et al., 2016a), which is a framework for fitting phylogenetic and macroevolutionary models. The prefix `Rev` is an homage to Reverend Thomas Bayes (1701–1761), the namesake of Bayesian statistics. Bayesian statistics, coupled with numerical computing techniques such as MCMC, enables researchers to analyze data using complex and parameter-rich evolutionary models. These analyses often result in large MCMC log files of several gigabytes, and the data structures

within can be elaborate. To summarize the results and produce a figure, the log files obtained from a `RevBayes` analysis must be post-processesed and one must use a graphical plotting software for visualization. The `RevGadgets` program is a collection of such post-processing tools and visualization tools. It includes tools to plot posterior distributions of simple continuous parameters, and phylogenetic trees. It can also accommodate various macroevolutionary analyses that condition on the phylogeny, including but not limited to time-varying diversification rate analyses (Palazzesi et al., 2022), state-dependent diversification rate analyses (Tribble et al., 2025), ancestral reconstruction of discrete characters (Peoples et al., 2025) or analyses of continuous characters (Roberts-Hugghis et al., 2025). With these tools, the primary aim of `RevGadgets` is to provide a common standard for visualizing results of typical macroevolutionary analyses in a pedagogical manner. A secondary aim is to increase the productivity of researchers by unifying the tools in a central repository.

Chapter 1

Introduction

1.1 General motivation

Species are better understood in the context of their evolutionary history, in the same manner that people are better understood in the light of genealogical history. The question of "who were my ancestors?" is as relevant on a million-year time scale as it is on a generational time scale. The study of species diversification is the study of learning about successes and failures of species. For instance, flowering plants are comprised of hundreds of thousands of species, while their cousins the gymnosperms number about a thousand. Mammals are at present the most dominant group of large terrestrial animals, but from the Jurassic the dinosaurs were the dominant group. The major failure of non-Avian dinosaurs came with the Cretaceous-Paleogene mass extinction event (Raup and Sepkoski, 1982). Such descriptions of events and features of diversification histories are often available for well-known groups, but not for all groups. Furthermore, while major radiations and mass extinctions are important events, they do not by themselves represent a comprehensive view of the process of diversification. To learn about the fundamental process of diversification, we would also like to compare diversification histories across different groups of organisms. If these groups are distantly related, separated in time, or if some groups are more studied than others, one must consider if these aspects can bias our view of the process of diversification. With these considerations in mind, I will in this thesis focus on assessing the process of diversification using quantitative methods instead of qualitative assessments.

Broadly speaking, there are two approaches for quantitative study of diversification dynamics in macroevolution. The first approach is to make use of occurrence data from the fossil record to study species or genus richness, as is frequently done in paleontology. Richness can be represented naively by plotting the time interval between the first and last occurrences per species, and counting the number of species that were alive per time bin (i.e., range-through curves, Taylor and Waeschenbach, 2015). Alternatively, one can standardize for sampling (Alroy et al., 2001), or adopt richness estimators that were originally developed for ecology

(e.g., capture-mark-recapture, Liow and Finarelli, 2014). Crucially, studies of diversification dynamics that use fossil occurrence data conceptualize their analyses in the context of a group or a clade of species. However, there is no explicit link between a phylogenetic history of the group and the fossil occurrence observations (but see Gupta et al., 2020; Andréoletti et al., 2022). A second approach is used in this thesis. Here, the link is made explicit, meaning that we use a stochastic process that can be used to generate or simulate a phylogeny (Nee et al., 1994b). The most commonly used stochastic process is the birth-death model, which I will introduce in more detail in the following sections. The phylogeny is almost always assumed to take the shape of a bifurcating tree. This means that species are either allowed to go extinct, or to split in two by means of a "speciation event". Reticulated evolution, such as horizontal gene transfer or hybridization events are usually not considered in branching processes (but see Fogg et al., 2023).

While there is a strong separation between approaches used in paleontological and neontological studies, this dichotomy is largely artificial. In recent years, extensive work has been carried out to bridge the gap between paleontological and neontological methods. In particular, birth-death models have been extended to also take into account fossil samples as observations (Heath et al., 2014; Barido-Sottani et al., 2019; Gupta et al., 2020; MacPherson et al., 2022). The main challenge with analyzing fossil phylogenies is that it is much more difficult to estimate fossil timetrees than extant timetrees. Thus, when one is browsing the published literature to see what kinds of phylogenies are available, there are a substantial amount of species-rich extant timetrees, but fewer and less species-rich fossil timetrees. This does not mean, however, that fossil phylogenies are unimportant. In particular, extinction rates are estimated more robustly when fossil observations are included in the tree (Didier et al., 2017; Mitchell et al., 2019; do Rosario Petrucci et al., 2025). Despite this, phylogenies that include fossil samples as observations are not accounted for in this thesis.

1.2 Historical perspective on birth-death models

The first developments of a stochastic branching process can be traced back to Yule (1924), who developed a model for the growth of the number of species and genera. A Yule model, or a "birth-only" model allows for lineages to speciate with rate λ, but does not allow for extinction. Assuming no extinction might be realistic on short time scales, however it is less realistic on longer time scales. If we consider the ancient history of the fossil record, then almost all species that have existed have gone extinct (Raup, 1986). Feller (1939) extended the Yule model to the "birth-and-death" model, which allows for extinction events to occur with rate μ. Later, Kendall (1948) presented a generalized birth-death model, which permits the speciation and extinction rates to vary over time. Birth-death models were not used in a phylogenetic context until methods for reconstructing phylogenetic relationships began to

be developed (Edwards and Cavalli-Sforza, 1964; Felsenstein, 1973). Thereafter, variants of the birth-death model have increasingly been adopted in empirical studies for reconstructing phylogenies and learning about diversification rate dynamics (Raup et al., 1973; Thompson, 1975). With advances in molecular sequencing technology (Sanger et al., 1977), more and more molecular data have become available (Venter et al., 2001) and phylogenetic reconstruction has proliferated (Huelsenbeck et al., 2001) to the point of becoming the cornerstone analysis in macroevolutionary studies (e.g. Rabosky et al., 2018; Machado et al., 2023; Quintero et al., 2024). In turn, the availability of phylogenetic trees has enabled more comprehensive and elaborate quantitative investigations of the process of diversification.

Perhaps the most standard diversification rate analysis involves estimating constant speciation and extinction rates (Hey, 1992; Nee et al., 1994b). In the next section, I will explain more in detail how one can estimate constant speciation and extinction rates from an extant phylogeny. However, numerous extensions of the birth-death model have been developed, where the aim is to learn more about specific aspects of diversification (Morlon et al., 2024). In Section 1.6 I will give an overview of some extensions of the birth-death model that have been developed.

1.3 Modeling species diversification with a stochastic branching process

In this and the following two sections I will give some more detailed explanations of the properties of the phylogenetic birth-death model. This includes how phylogenies are simulated, where waiting times for speciation and extinction events are random variables. Next, I describe an approach for calculating the probability of observing the reconstructed tree. This approach is a tree-pruning algorithm, and readers who are familiar with other tree-pruning algorithms (such as in models of molecular or morphological evolution) may find some similarities. It is not always necessary to use a tree-pruning algorithm for calculating the probability of the tree. For the constant-rate birth-death model in particular, the probability of the tree depends only on the divergence times and not on the topology. However, the tree-pruning method is a flexible approach that can easily be extended to more complex models, including those that allow for branch-specific diversification rates. Finally, I discuss how the parameters of the birth-death model can be estimated.

The aim of these sections is to give the reader an overview of the most basic properties and assumptions of the standard birth-death model, which for the most part also apply to more complex birth-death models. These aspects are often overlooked in journal articles, as page space is devoted more so to communicating novel insights than established concepts. However, in order to interpret empirical diversification analyses, it is crucial to have at least a basic understanding of the birth-death model, and how it is fitted to empirical data.

1.3.1 The generative birth-death process

The birth-death process is a generative model of evolution. Together with some starting conditions and rules for when to stop, it constitutes a random variable. The outcome of this random variable is a phylogenetic tree. The set of all possible outcomes of phylogenetic trees is a probability distribution, and some trees are more probable than others. This probability distribution is determined by the parameters of the birth-death model. Sampling phylogenetic trees using the birth-death process is also often referred to as "simulating phylogenies". Two events form the building blocks of a phylogenetic tree: speciation events and extinction events.

The time until a speciation or extinction event occurs is called a waiting time. Let's represent the waiting time with the symbol T. Consider an example in which events can occur in discrete time: in the 1st million year interval (myr), in the 2nd myr interval, and so on. Suppose that in one myr interval, a speciation event has a probability of p to happen, and $1-p$ not to happen. The probability of a speciation event to occur in the 2nd myr interval, is the probability of it not happening in the 1st myr, times the probability that it occurred in the final myr: $(1-p) \times p$. If $p = 0.5$, then the probability halves each myr, and the probability is $(1-p) \times p = 0.25$ for the event to happen in the second myr. However, if we condition on that the event did not occur, i.e., dividing by $(1-p)$, we return to the original event probability, p. The probability of an event occurring after time $t+1$, conditional on that the event did not yet occur, is the same as the unconditional probability of an event happening after 1 myr:

$$\Pr(T = t+1 | T > t) = \Pr(T = 1) = p. \tag{1}$$

In other words, the remaining waiting time does not depend on how much time has already passed. This distribution is known as the geometric distribution (Fig. 1.1).

In macroevolutionary models, however, time is usually thought of as a continuous measure. The analogue of the geometric distribution in the continuous domain is known as the exponential distribution. Like the geometric distribution, it has only one parameter, the rate λ (or sometimes the scale $1/\lambda$), and the probability density halves every $\ln(2)/\lambda$ units of time. The rate parameter λ is not usually something that we know, but an unknown parameter that will be estimated. As it turns out, the geometric and the exponential are the only probability distributions that can represent memoryless waiting times. The memoryless property (also called the Markov property) is not primarily motivated in biological realism, but rather to make birth-death models mathematically tractable. Although some work has been done on relaxing this assumption (e.g., Alexander et al. 2016, see also Saulsbury et al. 2024), most phylogenetic birth-death models are memoryless, including the ones used in this thesis.

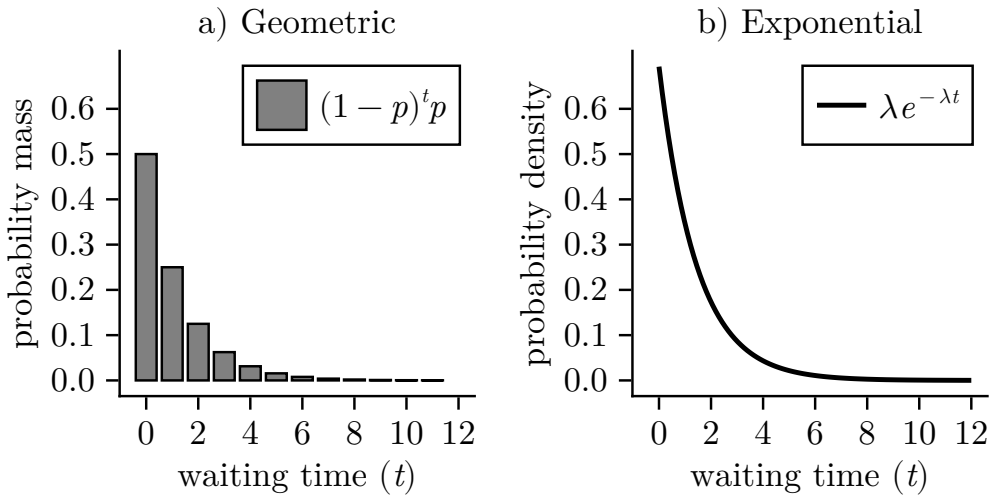


Figure 1.1: Possible waiting times in discrete or continuous time. In discrete time (a), the geometric distribution describes the probability of an event occurring after a waiting time t. With parameter $p = 0.5$, the probability mass is halved for each unit t in time. In continuous time (b), the exponential distribution is analogous. To set up an exponential distribution with a half-life of one in units of t, the rate parameter λ must be equal to $\ln(2)$.

1.3.2 The complete birth-death process

The exponential distribution has a parameter λ, which governs the overall tempo at which the speciation events happen. A high rate (λ) will correspond to short waiting times, i.e., many speciation events in a short time interval. Extinction events are modeled analogously, but with a different rate: μ, the rate of extinction. We assume that speciation and extinction are independent events, and so we can model the time until the next event occurs as a transformed variable $T \sim \text{Exp}(\lambda + \mu)$. After a waiting time randomly drawn from T, the event is either a speciation event (with probability $\lambda/(\lambda + \mu)$) or an extinction event (with probability $\mu/(\lambda + \mu)$). When a branching event happens, we also assume that the two child lineages evolve independently of each other. This assumption can be relaxed, for example by having the process depend on the overall richness or diversity at a given time (Rabosky and Lovette, 2008a; Bokma, 2009), but no such models were used in this thesis.

1.3.3 The reconstructed birth-death process

With phylogenetic trees, one must discriminate between "complete" and "reconstructed" trees (Fig. 1.2). Complete trees are the simple scenario, where all events (branching and extinction) are recorded in the tree. In a simulation study, or perhaps in certain empirical cases on tiny time scales, it is possible to obtain a complete phylogenetic tree. In practice, however, and especially on macroevolutionary timescales, it is almost impossible to obtain a complete phylogenetic tree. More often, we work with a phylogenetic tree that has been reconstructed by

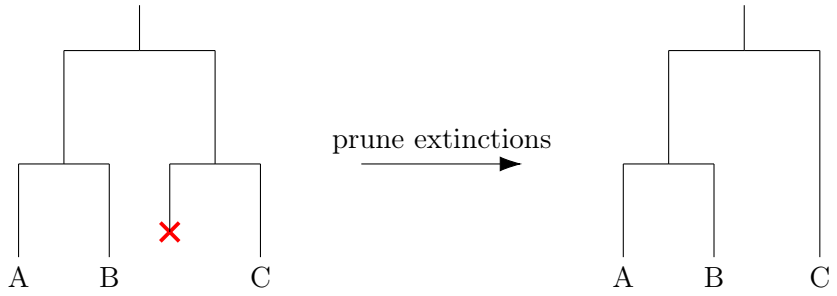


Figure 1.2: An example showing how the complete tree (left) and the reconstructed tree (right) are related. In the complete tree, there were three speciation events, one extinction event (the red cross), and three species survived until the present. In the reconstructed tree, the extinction event is pruned from the phylogenetic history, and it appears as if there were only two speciation events.

some means. For example, a phylogenetic tree can be inferred using a nucleotide alignment, a model of evolution by means of nucleotide substitutions, a clock model, and a framework for probabilistic inference (e.g., Bayesian inference). The trees produced by such procedures are so-called "reconstructed trees". In reconstructed trees, one can not actually see the extinction events. Even if the tree has been reconstructed using fossil data, these are also not extinction events (they are sampling events). Since the extinction events are not visible in reconstructed phylogenies, the branching events leading to the extinct lineages are also not visible in the reconstructed tree. Furthermore, one is also not usually able to include all taxa in the reconstructed tree. Incomplete taxon sampling has a similar effect of obscuring information: since some taxa are missing, some branching events will also be missing.

Reconstructed trees do not have complete information about the times of branching events, nor the extinction events. Therefore, it can be tricky to learn about the underlying but partially obscured processes of speciation and extinction. One powerful approach to study the birth-death process is to consider the various possible events that could have led to the "observed" reconstructed phylogeny.

1.4 Probability of observing the reconstructed tree

Here I will showcase how the probability of the reconstructed phylogeny can be calculated using a pruning algorithm. The overall description and notation follows that of Maddison et al. (2007). First, I derive the extinction probability $E(t)$. The extinction probability does not depend on the topology. Second, I derive the probability of observing the phylogeny, $D(t)$. The probability of observing the phylogeny is calculated by visiting each branch in a particular order. Daughter branches are visited before parent branches. This is called a postorder tree traversal, or a tree pruning algorithm. If the reader is not interested in the details regarding how to calculate the probability of the tree, he or she may skip until Section 1.5.

Extinction probability.— Let us begin with deriving the probability of a species at age

t in the past going extinct before the present ($t = 0$). Consider the possible events that can occur on a lineage, in a small time interval Δt in the past, such that the lineage goes extinct before the present. We will assume that only one extinction or speciation event can happen in the interval, but not both. The possible events leading to the extinction of the lineage are shown in Fig. 1.3. If Δt is sufficiently small, the probability of an extinction event happening

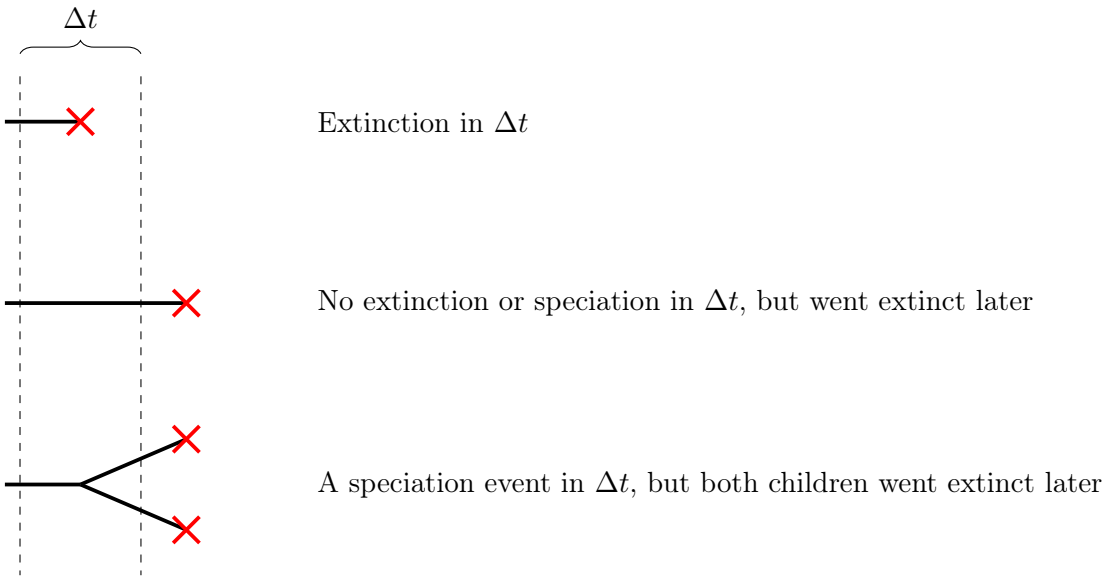


Figure 1.3: Three possible scenarios in a small time interval Δt that result in the lineage not leaving any survivors.

in the time Δt is approximately $\mu\Delta t$, and the probability of there not being an extinction event is approximately $(1 - \mu\Delta t)$. The extinction probability at the present, $E(t = 0)$, is equal to zero if we have complete taxon sampling. Intuitively, this makes sense, since over a zero-length time span there is not enough time to go extinct. If we have uniformly distributed incomplete taxon sampling, $E(t)$ is no longer the "probability of going extinct before the present", but instead more precisely "the probability of a lineage at time t not being sampled in the phylogeny", and hence $E(t = 0) = 1 - \rho$, where ρ is the taxon sampling fraction. For brevity, I will continue to use the term "extinction probability" for $E(t)$. The probabilities of the three possible events can be summed to be

$$\begin{aligned} &E(t + \Delta t) = \\ &\mu\Delta t && \text{extinction in } \Delta t \\ &+ (1 - \mu\Delta t)(1 - \lambda\Delta t)E(t) && \text{no events in } \Delta t, \text{went extinct later} \\ &+ (1 - \mu\Delta t)\lambda\Delta t E(t)^2 && \text{speciation in } \Delta t, \text{both went extinct later.} \end{aligned} \tag{2}$$

We can expand the parentheses, and gather the terms by $E(t)$:

$$E(t+\Delta t) = \mu\Delta t + (1^2 + \Delta t^2\mu\lambda - \Delta t\mu - \Delta t\lambda)E(t) + (\Delta t\lambda - \Delta t^2\lambda\mu)E(t)^2. \tag{3}$$

Notice that some terms include Δt^2. If Δt is small, then Δt^2 is tiny, and thus we can remove the terms including Δt^2:

$$\begin{aligned} E(t+\Delta t) =& \mu\Delta t + (1 - \Delta t\mu - \Delta t\lambda)E(t) + \Delta t\lambda E(t)^2 \\ E(t+\Delta t) =& \mu\Delta t + E(t) - \Delta t\mu E(t) - \Delta t\lambda E(t) + \Delta t\lambda E(t)^2. \end{aligned} \tag{4}$$

If we subtract $E(t)$ from both sides, and divide by Δt, we are left with the difference equation:

$$\begin{aligned} \frac{E(t+\Delta t) - E(t)}{\Delta t} &= \mu - \mu E(t) - \lambda E(t) + \lambda E(t)^2 \\ &= \mu - (\mu+\lambda)E(t) + \lambda E(t)^2, \end{aligned} \tag{5}$$

or if we take the limit as $\Delta t \to 0$, we get the differential equation

$$\frac{dE}{dt} = \mu - (\mu+\lambda)E(t) + \lambda E(t)^2. \tag{6}$$

This differential equation is exactly the same if we replace the constant speciation and extinction rates with functions that are allowed to vary arbitrarily through time (Kendall, 1948):

$$\frac{dE}{dt} = \mu(t) - (\mu(t)+\lambda(t))E(t) + \lambda(t)E(t)^2. \tag{7}$$

If λ and μ are constants, then a relatively simple closed-form solution exists,

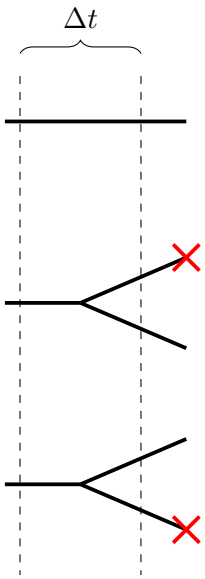
$$E(t) = 1 - \frac{e^{(\lambda-\mu)t}}{\rho^{-1} + \frac{\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)t} - 1\right)}. \tag{8}$$

If $\lambda(t)$ or $\mu(t)$ are time-variable, then a slightly more involved solution also exists (Morlon et al., 2011)

$$E(t) = 1 - \frac{e^{\int_0^t \lambda(u)-\mu(u)du}}{\rho^{-1} + \int_0^t e^{\int_0^s \lambda(u)-\mu(u)du}ds}. \tag{9}$$

Since Eq. 9 involves two nested integrals, however, it is not always practical to evaluate it as-is.

Tree probability.— The probability of observing a branch or a subtree can be derived in a similar manner as we derived the extinction probability. First, we consider which events are possible in a small time interval Δt, for an ancestral lineage to survive and to result in a single descendant lineage (Fig. 1.4). The probability of these three events can be summed as



No speciation and no extinction in Δt

Speciation event in Δt, but the left descendant went extinct later

Speciation event in Δt, but the right descendant went extinct later

Figure 1.4: Three possible scenarios in a small time interval Δt that result in the lineage leaving exactly one descendant lineage in the reconstructed tree.

$$
\begin{aligned}
D(t+\Delta t) = & \\
& (1-\mu\Delta t)\times && \text{in all cases no extinction in } \Delta t \\
& \quad [(1-\lambda\Delta t)D(t) && \text{no speciation in } \Delta t \\
& \quad + \lambda\Delta t D(t)E(t) && \text{speciation in } \Delta t\text{, left subtree went extinct} \\
& \quad + \lambda\Delta t D(t)E(t)] && \text{speciation in } \Delta t\text{, right subtree went extinct,}
\end{aligned}
\tag{10}
$$

and we can refactor it as

$$
\begin{aligned}
D(t+\Delta t) &= D(t) - D(t)(\lambda+\mu)\Delta t + 2\lambda\Delta t D(t)E(t) \\
\frac{D(t+\Delta t) - D(t)}{\Delta t} &= -D(t)(\lambda+\mu) + 2\lambda D(t)E(t).
\end{aligned}
\tag{11}
$$

Thus, the differential equation is

$$
\frac{dD}{dt} = -(\lambda+\mu)D(t) + 2\lambda D(t)E(t). \tag{12}
$$

When a branch is visited, we find the solution to Eq. 12 over the time span of the branch, given some initial value. The initial value for $D(t)$ is set to the sampling fraction ρ at the tips of the phylogeny. When we encounter a speciation event in the phylogeny, we assume that the left and the right subtrees evolved independently of each other. Therefore, the joint probability of the two subtrees is the product of the two subtree probabilities

$$
D_{\text{ancestor}}(t) = D_{\text{left}}(t) \times D_{\text{right}}(t) \times \Delta t\lambda, \tag{13}
$$

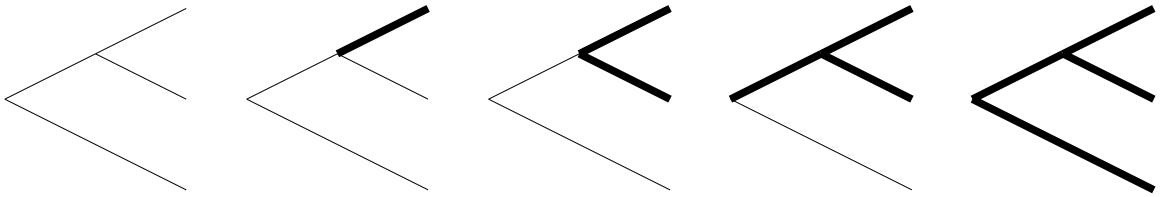


Figure 1.5: A three-taxon phylogeny where the daughter branches are visited (thick lines) before parent branches (read from left to right). This is a postorder tree traversal, also called a tree pruning algorithm.

times the probability that there was a speciation event in the short time interval when the branching event occurred ($\Delta t\lambda$). Since Δt is an arbitrarily small constant, $\Delta t\lambda$ is usually simplified to λ (Maddison et al., 2007). Therefore, the initial value for $D(t)$ at an internal node becomes

$$D_{\text{ancestor}}(t) = D_{\text{left}}(t) \times D_{\text{right}}(t) \times \lambda. \tag{14}$$

These steps are repeated across the branches of the tree, where the daughter branches are visited before their parent branches (Fig. 1.5). We terminate the probability calculation at the "root" node in the phylogeny, i.e., the node that represents the most-recent common ancestor of the group. Thus, the probability of observing the reconstructed phylogeny, given that we know the speciation rate and the extinction rates (λ and μ), can be phrased as

$$\Pr(\Psi|\lambda,\mu) = D_{\text{root}}(t_{\text{root}}), \tag{15}$$

where the symbol Ψ represents the reconstructed phylogeny. Often times we will also impose certain conditions on the probability of observing the tree. For instance, we can condition on that i) there was a speciation event at the root node, and ii) that both the left and the right subtrees survived until the present

$$\Pr(\Psi|\lambda,\mu) = \frac{D_{\text{root}}(t_{\text{root}})}{\underbrace{\lambda}_{\text{condition i}} \times \underbrace{(1 - E(t_{\text{root}}))^2}_{\text{condition ii}}}. \tag{16}$$

Imposing these conditions helps to improve the accuracy of parameter estimation (explained in the next section) by taking into account that the reconstructed trees we study are obtained in a systematically biased manner. For instance, if the clade did not leave any extant descendants (i.e., it went completely extinct), we would not fit the birth-death process in the first place.

1.4.1 Alternative approaches for the probability of the tree

In the previous section, I explained how one could calculate the probability of the reconstructed tree. For a reader who is not well-acquainted with birth-death models, this may seem like a complicated affair. Much of the complexity in the probability calculations comes due to that we are integrating out all possible "complete-tree" histories that are hidden due to extinction

events or incomplete sampling of extant species. However, this is not the only way to reason about the phylogeny in a probabilistic manner.

An alternative approach is to instead consider the probability of the complete tree, including both extinction events and unsampled taxa. This presents a problem, as we only observe the reconstructed phylogeny and not the complete phylogeny. As it turns out, it is possible to simulate these hidden extinction events and unsampled extant species, with an approach called "data augmentation" (Quintero et al., 2024, 2025). Effectively, this is a replacement of the analytical integration of hidden extinction events and unsampled extant species with a kind of numerical integration. With the data augmentation approach, the probability calculations are greatly simplified. However, it introduces more complexity in terms of the computer code that is required to sample the hidden evolutionary events. Data augmentation approaches have great potential. They are especially flexible in terms of how the generative model is specified, enabling researchers to test a wider range of hypotheses. Nevertheless, most phylogenetic birth-death approaches, including the ones presented in this thesis, do not use data augmentation.

1.5 Inferring speciation and extinction rates

Estimating the parameter values (the speciation and extinction rates λ, μ) is often the goal when conducting a diversification rate analysis. There are several ways of estimating the parameters, however likelihood-based methods are perhaps the most principal. The actual "likelihood" is the same as the probability of observing the phylogeny

$$\text{Likelihood}(\lambda, \mu|\Psi) = \Pr(\Psi|\lambda, \mu). \tag{17}$$

If the likelihood and the probability are the same, one may wonder what is the benefit of introducing additional jargon. The difference is that the probability is a function of the phylogeny Ψ, whereas the likelihood is a function of the parameters λ and μ. A probability function must sum or integrate to one over the possible observations (all possible phylogenies) for it to represent a valid probability distribution (Devore and Berk, 2012, p.57). The likelihood function, however, does not integrate to one. As such, the likelihood function has no probabilistic interpretation.

There are two main paradigms in likelihood-based statistical inference. The first and perhaps the simpler method is that of maximum likelihood. The second method is Bayesian inference. I will summarize these two paradigms and compare with how parameter estimation is performed in the various chapters of the thesis.

1.5.1 Maximum likelihood

The maximum likelihood approach involves finding the combination of parameter values λ and μ that maximize the likelihood function. Overall, the method of maximum likelihood has desirable properties. Under certain assumptions, the maximum-likelihood estimator is said to be *consistent* (Wald, 1949). A consistent estimator is an estimator that converges to the true parameter value, given sufficiently many independent observations. In macroevolutionary studies of diversification rates there is typically only one observation of the phylogeny, but species-rich phylogenies are generally more informative than species-poor phylogenies. An important assumption for the method of maximum-likelihood to work, is that the model must be *identifiable*. If a model is not identifiable, then there exists some combination of parameter values $\lambda_1 \neq \lambda_2$ and/or $\mu_1 \neq \mu_2$ such that

$$\Pr(\Psi|\lambda_1, \mu_1) = \Pr(\Psi|\lambda_2, \mu_2), \tag{18}$$

i.e., the probability function is not unique across the parameter values. The constant-rate birth-death model is identifiable, provided that the taxon sampling fraction is fixed (Stadler, 2009). The identifiability of time-variable birth-death models will be discussed in **Chapters 2 and 3**.

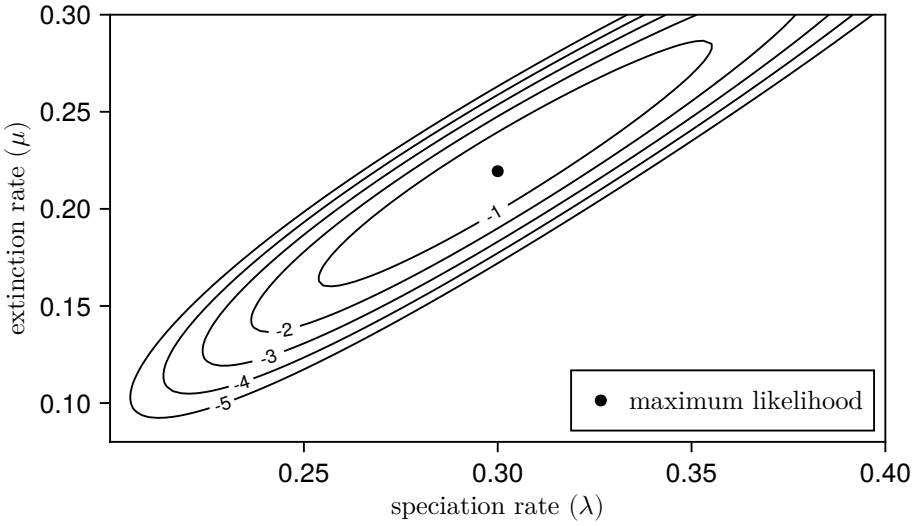


Figure 1.6: The log-likelihood surface of the constant-rate birth-death process, as a function of the speciation rate and the extinction rate. The log likelihood is rescaled such that the maximum log-likelihood value is zero. The contour rings represent the distance to the maximum log-likelihood value.

In the case of a simple model such as the Yule model, it is possible to derive an analytical equation for the maximum-likelihood estimator of the birth rate. For more complex models, however, analytical solutions are usually not available. Instead, numerical techniques can be used to find the maximum-likelihood parameter values. A visually intuitive way of finding

the maximum-likelihood parameter values is to perform a grid search and plot the likelihood surface (Fig. 1.6). Grid searches are however computationally expensive, especially if there are more than a few parameters, and care must be taken to select appropriate limits and spacing of the grid. "Smarter" and more efficient numerical optimization techniques can also be used, for instance gradient ascent or Newton's method. These have the disadvantage, however, of sometimes getting stuck in local optima. The main philosophical challenge with maximum-likelihood estimation is that it is not a good framework for assessing the parameter estimates in a probabilistic manner. In particular, we would like for the statistical inference procedure to inform us about whether a parameter estimate is precise or uncertain.

1.5.2 Bayesian inference

Bayesian statistics is a framework for reasoning about probabilities in the context of subjective knowledge. In Bayesian philosophy, probability is interpreted as a "strength of belief". This is in contrast to frequentist philosophy, where the interpretation is "relative frequency". In order to use Bayesian inference on the parameters of a model, one must first establish the causal dependencies among the parameters of the model. A convenient way to do so, is to draw a

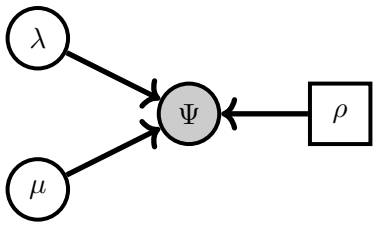


Figure 1.7: A Bayesian belief network representing the causal dependencies (arrows) of random variables (circles) and fixed parameters (square). The node representing the phylogeny (Ψ) is shaded, meaning that we have observed a reconstructed phylogeny. λ is the speciation rate, μ is the extinction rate, and ρ is the taxon sampling probability.

picture representing the Bayesian belief network (Pearl, 1988), also known as a probabilistic graphical model (Höhna et al., 2014). In Fig. 1.7 I drew the Bayesian belief network of the constant-rate birth-death model, where there is a causal influence of the speciation and extinction rates (λ, μ) on the phylogenetic tree (Ψ). Using Bayes theorem, we can pose the question "what is the probability of λ and μ, given that we know Ψ?"

$$\overbrace{\Pr(\lambda, \mu|\Psi)}^{\text{Posterior}} = \frac{\overbrace{\Pr(\Psi|\lambda, \mu)}^{\text{Likelihood}} \times \overbrace{\Pr(\lambda)\Pr(\mu)}^{\text{Priors}}}{\underbrace{\Pr(\Psi)}_{\text{Normalizing factor}}}. \tag{19}$$

Here, $\Pr(\lambda)$ and $\Pr(\mu)$ are the prior probabilities of the speciation and extinction rates, and the likelihood was covered previously. The priors represent the strength of belief in what the speciation and extinction rates are, before having observed the phylogeny. The prior strength

of belief in the parameters is combined with the likelihood, resulting in the posterior strength of belief (Fig. 1.8). This notion of updating one's beliefs when new evidence is presented is the core idea of Bayesian statistics. If one wishes to obtain a point estimate for the speciation and extinction rates, one can for example find the parameter values that maximize the posterior probability (using hill climber methods), although this is not usually done.

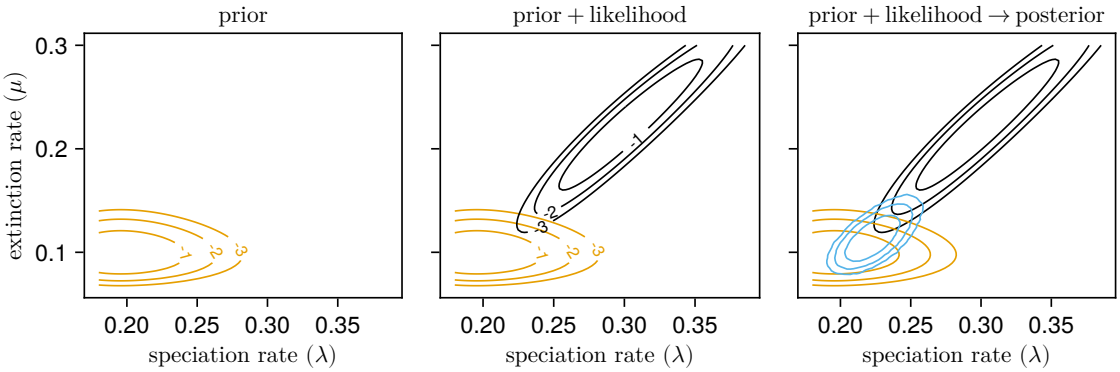


Figure 1.8: An example of how Bayes theorem is used to update the strength of belief in extinction and speciation rates when new evidence is presented. The prior strength of belief (in orange) combines with the probability of observing the data (the likelihood, in black), resulting in an updated, posterior strength of belief (in blue). The contours represent log probabilities, re-scaled such that the maximum log-probability is zero. The model is the constant-rate birth-death process. The data is an empirical phylogeny of primates with 233 species (Vos and Mooers, 2006). The posterior distribution is approximated using the Metropolis-Hastings algorithm.

Notice that I neglected to mention the unconditional probability of the phylogeny, $\Pr(\Psi)$. This is the likelihood times the priors, integrated over all possible values of the parameters:

$$\Pr(\Psi) = \int_0^\infty \int_0^\infty \Pr(\Psi|\lambda,\mu) \times \Pr(\lambda)\Pr(\mu) d\lambda d\mu. \tag{20}$$

Operationally, since the parameters are integrated, or "marginalized" out, it is often referred to as a marginal probability. I think it is more intuitive to interpret its role as a normalizing factor, in that it must be a certain value in order for the posterior probability to be a valid probability function (i.e., the posterior probability function must integrate to one). One challenge is that the normalizing factor can typically be difficult or impossible to calculate, analytically or otherwise. Therefore, one can try to avoid computing the normalizing factor, for instance by setting it to one. Then, the posterior probability function is not valid in the narrow sense. Despite this, it can still be interpreted as an unnormalized probability function.

As it turns out, it is possible to investigate the posterior distribution using simulation approaches, even if the posterior probability is unnormalized. For instance, the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) is a popular method that can be used to generate samples from the posterior distribution. However, I will not explain the Metropolis-Hastings algorithm here, nor any other kinds of Markov-chain Monte Carlo

(MCMC) algorithms. Besides the philosophical aspect of incorporating prior knowledge, Bayesian inference with MCMC has another useful advantage. Namely, MCMC methods are an excellent tool for assessing the extent of uncertainty in the parameter estimates. Since MCMC samples are drawn from the posterior distribution, it is straightforward to assess parameter uncertainty. A typical measure for this is the 95% credible interval. It is also possible to assess parameter uncertainty in the context of maximum-likelihood inference, for example by computing a support interval (Edwards, 1972). Support intervals are however impractical for models with more than one or two parameters, and therefore it is rarely done. Nevertheless, when fitting any statistical model, it is important to consider estimation uncertainty. Ideally, one should also report a quantitative estimate for how large the uncertainty is.

1.5.3 Inference techniques used in this thesis

The chapters in this thesis use several different approaches for parameter estimation. In **Chapters 2 and 3**, we investigate nonidentifiability of the time-varying birth-death process. Whether a model is identifiable is a fundamental property of the model, and does not depend on which technique is used to estimate the parameters. That being said, we motivate the empirical examples using time-varying birth-death models that were fitted using Bayesian inference (Magee et al., 2020). The method presented in **Chapter 4** uses a combination of inference techniques. First, we fit the parameters of the birth-death-shift model using maximum likelihood. Second, we investigate branch-specific diversification rates using techniques motivated by Bayesian inference. This two-step approach is also known as an "empirical Bayes" approach. In **Chapter 5** we use the same statistical inference technique as in **Chapter 4**. In **Chapter 6**, we review several methods for estimating lineage-specific diversification rates, and they all make use of Bayesian inference. In **Chapter 7**, we are not concerned with estimating parameters, but with summarizing and visualizing `RevBayes` analyses. However, almost all `RevBayes` analyses use Bayesian inference with Markov-chain Monte Carlo sampling, and the parameter proposals are typically Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970), Gibbs proposals (Geman and Geman, 1984), or special tree topology proposals (Robinson, 1971; Hein, 1990).

1.6 Extensions of the phylogenetic birth-death model

Beyond the constant-rate birth-death model for extant phylogenies, many variations of the birth-death model have been developed. By relaxing the model assumptions, we can gain more nuanced insights about the process of diversification (reviewed in Morlon et al., 2024). For the purpose of perspective, I discuss some variations and give citations.

Time-heterogeneous models.— Nee et al. (1994b) provided a theoretical framework for phylogenetic investigations of time-variable diversification rates, in particular by deriving

the probability of the reconstructed phylogeny. Rabosky and Lovette (2008b) and Morlon et al. (2011) extended this work by investigating time-varying patterns in empirical phylogenies, including the "explosive early" hypothesis. Time-variable models may also be linked to a specific hypothesis, for example the question of whether grassland diversity is related to atmospheric carbon dioxide (Palazzesi et al., 2022). This kind of model usually assumes that the environmental variable is equal across lineages of the tree, but varying over time. Furthermore, one can test whether mass extinction events are more likely to explain the phylogeny than short periods of high extinction (Höhna, 2015; May et al., 2016).

Lineage-heterogeneous models.— Another class of models is designed for testing whether there is evidence of lineage-specific heterogeneity in diversification rates. If diversification rates change along a phylogeny, the changes may be numerous and small (Maliet and Morlon, 2022; Budd and Mann, 2025), but the more common assumption is for the changes to be few and substantial (Rabosky et al., 2013; Höhna et al., 2019). Several frameworks for Bayesian inference of lineage-heterogeneous diversification rates have been developed (reviewed in **Chapter 6**, Martínez-Gómez et al. 2024). `Pesto` is one such method, where the focus is not only to infer lineage-specific diversification rates, but also to infer where on the phylogeny there was a diversification rate shift event (**Chapter 4**, Kopperud and Höhna 2025). Overall, `Pesto` and similar methods do not test specific causal hypotheses. This means that, even if there is strong statistical support for a diversification rate shift event, we do not learn anything about the underlying cause of the rate shift event.

Trait-dependent models.— If one has a specific causal hypotheses in mind, as for why diversification rates have changed across lineages, it is possible to test the hypothesis more directly. For example, one can test whether diversification rates are being influenced by lineage-specific traits that themselves also evolve on the phylogeny (Maddison et al., 2007; FitzJohn, 2012). One challenge with trait-dependent diversification rate models is that the correlation between trait and diversification rate can often be spurious. A spurious correlation is due to mutual dependence on some unmeasured factors, instead of an inherent causal link between the trait and the diversification process. To account for this, methods for inferring diversification rate changes due to observed and hidden traits have been developed (Beaulieu et al., 2013).

Protracted speciation models.— Most phylogenetic birth-death models assume that speciation events happen instantaneously. According to the model, there is in one moment a single species, and after a snap of the fingers there are two species. From a biological point of view, this is far from realistic. If speciation happens due to geographic isolation of two populations (i.e., allopatric speciation), it can take a substantial amount of time before the two populations are sufficiently separated (e.g., in terms of reproductive compatibility) and subsequently become two different species (Hernández-Hernández et al., 2021). Even with more rapid modes of speciation, such as reproductive isolation due to polyploidization,

evidence suggests that the process of speciation is not instant (Slotte et al., 2008). Protracted speciation processes are one approach for modeling speciation not as an instantaneous event but as a gradual process (Hua et al., 2022; Veron et al., 2025). While no doubt more realistic than standard birth-death models, it can be more challenging to fit protracted speciation models to empirical phylogenies.

Diversity-dependent models.— Standard birth-death models are fundamentally exponential models of evolution. If the net-diversification rate is positive, then this leads in expectation to an ever-increasing number of lineages. Over a long time span, this may be unrealistic. An alternative view might be that there is a self-limiting aspect to species growth. Analogously to how population growth is limited by factors such as land and resources, it may be that species growth is limited by intrinsic factors such as competition for ecological niche space (Etienne et al., 2023). Diversity-dependent birth-death models incorporate this by letting the diversification rates depend on the current species diversity (Rabosky and Lovette, 2008a). Recall that in the standard birth-death model, sister lineages evolve independently of each other. As this is not true for a diversity-dependent diversification model, it becomes more complicated to express the probability of the phylogeny.

General framework for testing diversification rate hypotheses.— Consider the scenario where one wishes to test two separate hypotheses about diversification. For example, whether a) there were lineage-specific rapid radiation events, or b) rates of extinction declined due to a change in global temperature. It is relatively straightforward to test these hypotheses separately. However, it may be the case that both hypotheses are strongly supported in comparison with a constant-rate birth-death model. This gives rise to several questions. Is the former hypothesis confounded by the latter, or the opposite? Does a joint model of lineage-specific shifts and climate-dependent extinction rate explain the data better than either hypothesis separately? To answer these questions, one should ideally implement a joint model where diversification rates a influenced both by lineage-specific events and environmental factors. This is not only a mathematical problem, but also a challenge in terms of designing the software to be flexible enough. Overall, some progress has been made in terms of providing more flexible ways to set up birth-death models (Ronquist et al., 2021). However, the notion of a general framework for joint testing of multiple alternative hypotheses remains an open problem in the field of macroevolutionary diversification dynamics.

1.7 Aims of the thesis

The general motivation of this thesis is to improve our understanding of the underlying process of diversification that has given rise to the species diversity we see today. To do so, we aim to advance the state-of-the-art in the field of phylogenetic diversification rate investigations. This includes model-theoretical approaches, simulation studies, writing of software, empirical

analyses and more. Each manuscript has more specific aims:

- **Chapter 2**: Improve the understanding of the nonidentifiability of time-varying birth-death models, specifically by providing software for sampling the congruence class, and to visualize the overall trends within the congruence class.
- **Chapter 3**: Provide better insights into what types of time-varying diversification rate patterns are robust to the issue of nonidentifiability, with empirical examples.
- **Chapter 4**: Design a method for inference of branch-specific diversification rates, that is feasible even for exceptionally species-rich phylogenies ($>$ 20k taxa).
- **Chapter 5**: Investigate the prevalence of branch-specific shifts in diversification rates in empirical phylogenies, as well as the nature or type of these rate shift events.
- **Chapter 6**: Provide a better understanding of empirical estimates of branch-specific diversification rates, by comparing several contemporaneous inference frameworks.
- **Chapter 7**: Provide better tools for presentation of the results of common phylogenetic Bayesian analyses.

Each chapter was written to stand on their own as independent manuscripts. However, there are some overlaps in topic. In particular, chapters 2 and 3 are about nonidentifiability of the time-varying birth-death model. Chapters 4, 5 and 6 are all concerned with the inference of branch-specific diversification rates and branch-specific diversification rate shift events.

Chapter 2

CRABS: Congruent rate analyses in birth-death scenarios

A **coconut crab** (*Birgus latro*) is capable of producing a pinching force of up to 3300 Newton.

Oka et al. (2016)

2.1 Summary

1. Diversification rates inferred from phylogenies are not identifiable if the rates are allowed to vary freely over time. There are infinitely many combinations of speciation and extinction rate functions that have the exact same likelihood score for a given phylogeny, building a congruence class. The specific shape and characteristics of such congruence classes have not yet been studied. Whether speciation and extinction rate functions within a congruence class share common features is also not known.

2. Prior hypotheses typically render diversification rates identifiable, but the results then depend on these a priori hypotheses and assumptions. In order to test the robustness of diversification results to these a priori hypotheses, we use two different approaches to explore congruence classes: (i) constructing congruent models under alternative hypotheses, and (ii) sampling alternative rate function within the congruence class.

3. Our methods are implemented in the open-source R package `CRABS`. `CRABS` provides a flexible approach to explore the congruence class and provides summaries of rate functions within a congruence class. The summaries can highlight common trends, i.e., increasing, flat or decreasing rates.

4. Although there are infinitely many equally likely diversification rate functions, these can share common features. `CRABS` can be used to assess if diversification rate patterns are robust despite the non-identifiability of the birth-death model.

In our example, we clearly identify three phases of diversification rate changes that are common among all models which we sampled from the congruence class. Thus, congruence classes are not necessarily a problem for studying historical patterns of biodiversity from phylogenies.
Key-words: Birth-death models, macroevolution, diversification rates, identifiability, congruence class.

2.2 Introduction

In macroevolution, one prominent avenue of research is to estimate macroevolutionary rates of diversification from molecular phylogenies (Ricklefs, 2007; Morlon, 2014). Specifically, many studies are interested in inferring time-varying diversification rates. Time-varying diversification rates are used to study monotonic slowdowns/increases using continuous functions (*e.g.,* Rabosky, 2006; Morlon et al., 2011; Höhna, 2014), abrupt shifts in diversification rates (*e.g.,* Stadler, 2011; May et al., 2016; Magee et al., 2020), mass extinction (*e.g.,* Höhna, 2015; May et al., 2016; Culshaw et al., 2019; Magee and Höhna, 2021) and correlations to environmental factors (*e.g.,* Condamine et al., 2013, 2019; Palazzesi et al., 2022). While these models are typically identifiable when estimated from time-calibrated phylogenies, they rely on specific a priori hypotheses on the form of the time-variation of diversification rates (Morlon et al., 2022). Unfortunately, when relaxing these hypotheses by allowing for any continuous diversification rate function, time-varying diversification rates are no longer identifiable (Kubo and Iwasa, 1995; Louca and Pennell, 2020b). That is, infinitely many combinations of speciation and extinction rate functions, summarized within a *congruence class*, result in the same likelihood given a phylogenetic tree (Kubo and Iwasa, 1995; Louca and Pennell, 2020b).

However, the existence of infinitely many equivalently likely rate functions does not imply that one cannot draw any general conclusions (Morlon et al., 2022; O'Meara and Beaulieu, 2021). We do not know yet which diversification rate functions are within a congruence class, and if these diversification rate function share some specific features (e.g., rate changes at the same time). If we obtained estimates of diversification rate functions for our study group, then we could be interested in exploring all or a sample of diversification rate functions included in the congruence class to identify shared features. Furthermore, we could test if different specific diversification rate scenarios are included within a congruence class, for example, if a model with exponentially increasing/decreasing speciation rates is included in the congruence class. If the corresponding extinction rate is negative, which is biologically unrealistic and therefore not allowed, then the diversification scenario is not contained within the congruence class (see also Louca and Pennell, 2021). Similarly, we could assess if models with rate shifts at specific times, corresponding to alternative diversification scenarios, are included in the congruence class. These tests focus on the robustness of inferred diversification rate pattern.

Here we provide the `R` package `CRABS` (Congruent Rate Analyses in Birth-death Scenarios) that (1) converts between models within the same congruence class, (2) samples from the congruence class, and (3) shows common trends among models within the same congruence class. Conversion between congruent models is useful if a researcher wants to explore alternative diversification scenarios, for example, "what if the extinction rate was not constant but instead exponentially increased through time?" Sampling from the congruence class can highlight general features of a congruence class, for example, if a researcher has estimated a given pattern of diversification rates and wants to know if most models within the congruence

class show a certain trend (e.g., a rate shift at time t). Finally, all explored models can be analyzed to show common patterns of diversification rate increases and decreases. With our `R` package `CRABS`, researchers can test which patterns are robust despite the non-identifiability of speciation and extinction rates.

2.3 Theory and usage

In this section we provide the theory behind `CRABS`, as well as a brief overview on how to use it. We start by explaining how `CRABS` processes any diversification rate function to construct the congruence class. Next, we explain how `CRABS` obtains alternative (congruent) matching pairs of speciation and extinction rates functions. Then, we show how to explore the congruence class by constructing specific diversification scenarios and using diversification rate functions generated from a process or a distribution. Finally, we demonstrate how alternative diversification rate models within the congruence class can be summarized to assess for shared trends.

2.3.1 Reference Speciation and Extinction Rates

Our exploration of the congruence class starts with obtaining a *reference* set of speciation and extinction rates. As an empirical example, we estimated speciation and extinction rates for the primates phylogeny from Springer et al. (2012) using a horseshoe Markov random field (HSMRF) prior distribution (Carvalho et al., 2010; Magee et al., 2020) as implemented in `RevBayes` (Höhna et al., 2016a). The specific details about the data set and Markov chain Monte Carlo settings are not important for this study, but can be found at `https://revbayes.github.io/tutorials/divrate/ebd.html`. We include the samples from the posterior distribution in our package for convenience. We will use this example to showcase how to explore the congruence class with `CRABS`.

`CRABS` accepts any type of function for the reference model. As an example, we will use here a first order spline function, *i.e.,* a piecewise linear rate function for the reference model. In `R`, we can set up the piecewise linear rate functions as follows.

```
library(CRABS)
data(primates_ebd)
lambda <- approxfun(primates_ebd$time, primates_ebd$lambda)
mu <- approxfun(primates_ebd$time, primates_ebd$mu)
```

2.3.2 Constructing the Congruence Class

The central idea in `CRABS` is to construct the congruence class given a speciation and extinction rate function. A congruence class is fully specified by either the pulled net-diversification rate

$r_p(t)$ and the speciation rate at the present (λ_0), or the pulled speciation rate $\lambda_p(t)$. That is, any combination of speciation and extinction rate function that result in the same pulled net-diversification rate $r_p(t)$ and speciation rate at the present (λ_0), or pulled speciation rate $\lambda_p(t)$, belong to the same congruence class (Louca and Pennell, 2020b). We can therefore set up the congruence class by constructing the pulled net-diversification rate $r_p(t)$. The pulled net-diversification rate is defined as (Louca et al., 2018; Louca and Pennell, 2020b)

$$r_p(t) = \lambda(t) - \mu(t) + \frac{1}{\lambda(t)} \frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} \quad . \tag{1}$$

The equation for the pulled speciation is (Louca et al., 2018; Louca and Pennell, 2020b):

$$\lambda_p(t) = (1 - E(t))\lambda(t) \quad , \tag{2}$$

where $E(t)$ is the probability that the lineage observed at time t goes extinct before the present, and the equation for the pulled extinction rate is

$$\mu_p(t) = \lambda_0 - r_p(t) \quad . \tag{3}$$

Currently we only use the pulled speciation rate and pulled extinction rate for visualization purposes (e.g., Fig. 2.1). We use the pulled net-diversification rate to construct the congruence class because it simplifies the equations.

Continuing with the primates data example, we can construct the congruence class in `CRABS` as follows.

```
times <- seq(0, max(primates_ebd$time), length.out = 1000 )
my_model <- create.model(lambda, mu, times)
```

The pulled rates are computed for a set of pre-defined time points. Here, we use a fine grid of one thousand time points to achieve a high resolution in our plots. In our experience, one thousand time points are sufficient for plotting purposes because a higher resolution does not improve visibility.

Then, we can plot the diversification rates together with their pulled counterparts in `CRABS` using `plot(my_model)` (Fig. 2.1). Studying the pulled speciation and pulled net-diversification rates itself can highlight aspects of the congruence class (Louca et al., 2018; Helmstetter et al., 2022).

2.3.3 Transforming speciation and extinction rates

A researcher can either provide another speciation rate function (or extinction rate function) and `CRABS` computes the corresponding extinction rate function (or speciation rate function, respectively) so that the new pair also belongs to the same congruence class.

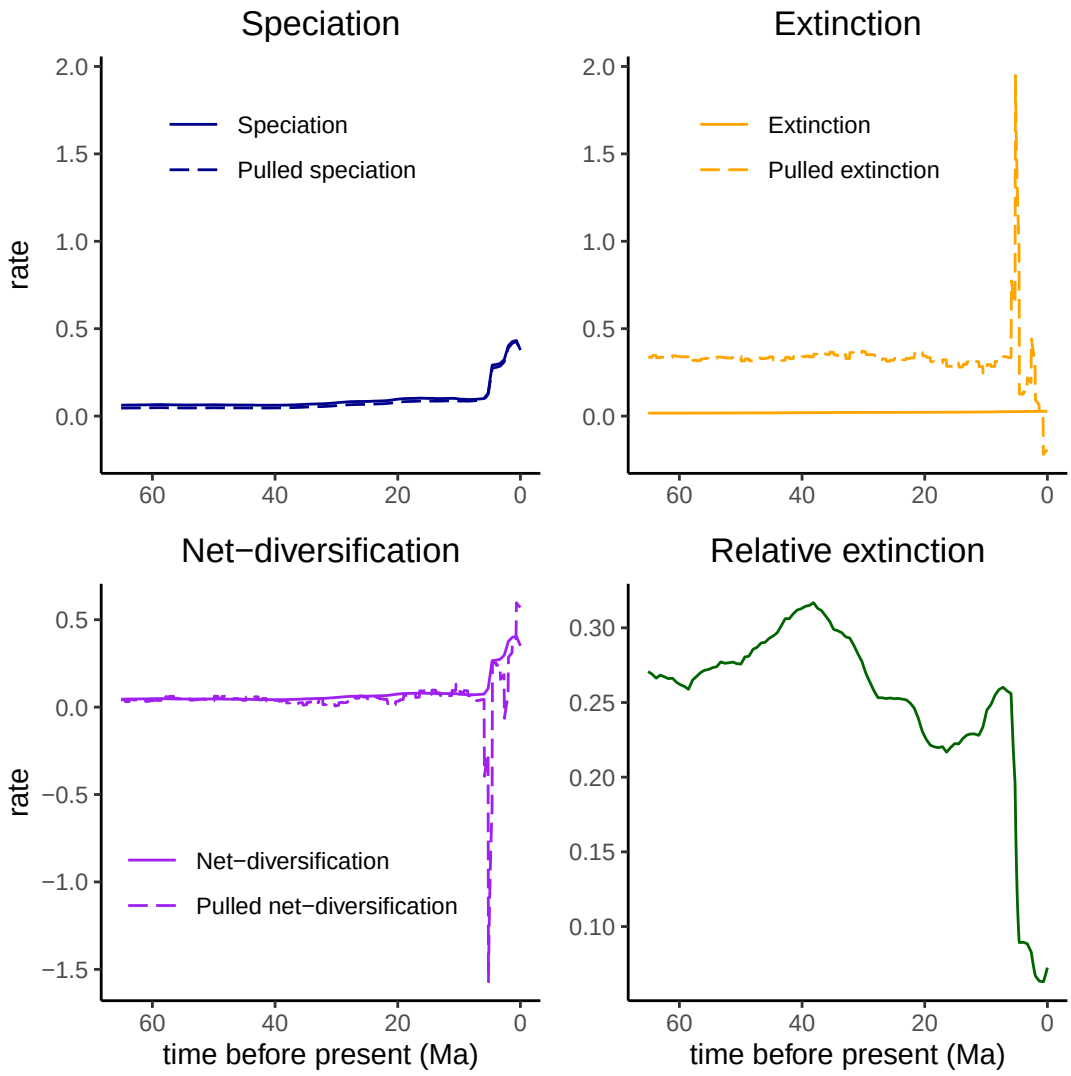


Figure 2.1: Estimated diversification rates for the primates phylogeny by Springer et al. (2012) using an episodic birth-death model with a horseshoe Markov random field prior (Magee et al., 2020) in `RevBayes` (Höhna et al., 2016a). Here we show the speciation rate (λ), extinction rate (μ), net-diversification rate ($\lambda - \mu$) and relative extinction rate (μ/λ). The original rates define the *reference* model in `CRABS` (solid lines). `CRABS` automatically computes the pulled speciation rate, the pulled extinction rate, as well as the pulled net-diversification rate (dashed lines), which characterize the congruence class.

If a user provides a new extinction rate function, $\mu'(t)$, and wishes to know the corresponding speciation rate, $\lambda'(t)$, we can compute it from the following relationship, obtained by solving Eq. 1 for λ:

$$\frac{d\lambda'}{dt} = -(\lambda'(t))^2 + \lambda'(t)(\mu'(t) + r_p(t)) \quad . \tag{4}$$

We discuss in the Supplementary Material how this differential equation can be solved. We note that λ_0 is equal for all models in the congruence class. Conversely, if a user provides a new speciation rate function, then we can instead solve Eq. 1 for μ and get

$$\mu'(t) = \lambda'(t) - r_p(t_i) + \frac{1}{\lambda'(t)}\frac{\lambda'(t)}{dt} \quad . \tag{5}$$

These two transformations allow us to explore the congruence class. We only need to propose an alternative speciation rate function, or an alternative extinction rate function, and then compute their counterpart for the new model to be within the congruence class.

2.3.4 Constructing congruent models under specific hypotheses

A first option to explore the congruence class is to test for specific alternative diversification scenarios. For example, one can explore if a linearly or exponentially decreasing speciation rate function is contained within the congruence class while restricting that extinction rates must be non-negative. In principle, there are no limitations to the choice of specific diversification scenarios and we provide several examples in our vignette (`https://afmagee.github.io/CRABS`). This option is useful when a researcher has a specific idea regarding when diversification rates could have changed and what shape the diversification rates function might have.

In our primate HSMRF analysis, we inferred that the speciation rate changed abruptly in the last few million years, but the extinction rate remained comparably constant (Fig. 2.1). We note that the speciation rate appears to drive the changes in the "observed" net-diversification rate, i.e., in our originally inferred net-diversification rate. As an illustration, we explore here the alternative scenario if it instead was the extinction rate that drove the changes in the net-diversification rate. Because the net-diversification rate is defined as $\delta(t) = \lambda(t) - \mu(t)$, we construct our new diversification scenario for the extinction rate function as $\mu'(t) = \mu_0 - \lambda(t)$ where μ_0 is any arbitrary value with $\mu_0 \geq \sup(\lambda(t))$ to ensure that $\mu'(t) \geq 0$. In `CRABS`, we only need to specify the alternative extinction rate functions $\mu'(t)$ and then call `congruent.models` (which works similar to the function `congruent_hbds_model` in the `R` package `castor` (Louca and Doebeli, 2018)).

```
mu_scaling <- c(1.1, 1.2, 1.5, 2.0)
mu_0s <- max(lambda(my_model$times)) * mu_scaling
```

```
mu_prime <- list()
for (i in seq_along(mu_scaling)){
  mu_prime[[i]] <- local({
    mu_0 <- mu_0s[i]
    function(t) mu_0 - lambda(t)
  })
}

alt_models <- congruent.models(my_model, mus = mu_prime)
plot(alt_models)
```

Even though we constructed the extinction rate functions to be responsible for the changes in the net-diversification rates, we observe that the corresponding speciation rate functions of the same congruence class are never constant (Fig. 2.2). This results in somewhat different net-diversification rate functions although both recent rate increases remain.

2.3.5 Exploring congruent models, i.e., sampling from the congruence class

A second option is to explore the congruence class. However, because there are infinitely many rate functions contained within the congruence class, we cannot explore every single model within the congruence class. Instead, we can sample randomly from a distribution of rate functions within the congruence class. We only need to specify how to create a random sample. For example, we could sample from different exponentially decreasing rate functions, or we could sample from a (discretized) Brownian motion process.

In `CRABS`, we provide ways to randomly sample from several flexible rate distributions. These distributions can be specified using the function `sample.basic.model`, and Table 2.1 provides an overview of the options for this function. A more detailed description is provided in our vignette (`https://afmagee.github.io/CRABS`).

As an example, we assume that an alternative extinction rate function corresponds to a (log-scale) Brownian motion. Note that we cannot sample from Brownian motion as a fully continuous function so we sample instead at specific time points and use a first order spline function to interpolate between these time points. As a starting point for the Brownian motion at the present time $t_0 = 0$, we sample μ_0' from a lognormal distribution. The distribution is centered around the reference estimate μ_0. Each preceding μ_i is distributed as lognormal(μ_{i-1}, σ^2), where σ is drawn from a half-Cauchy distribution with parameters set following Magee et al. (2020). We select the variance for the initial lognormal such that the central 95%-quantile of μ_0' spans 2.5 times two orders of magnitude on the rate scale. We repeat the entire procedure to draw each rate function.

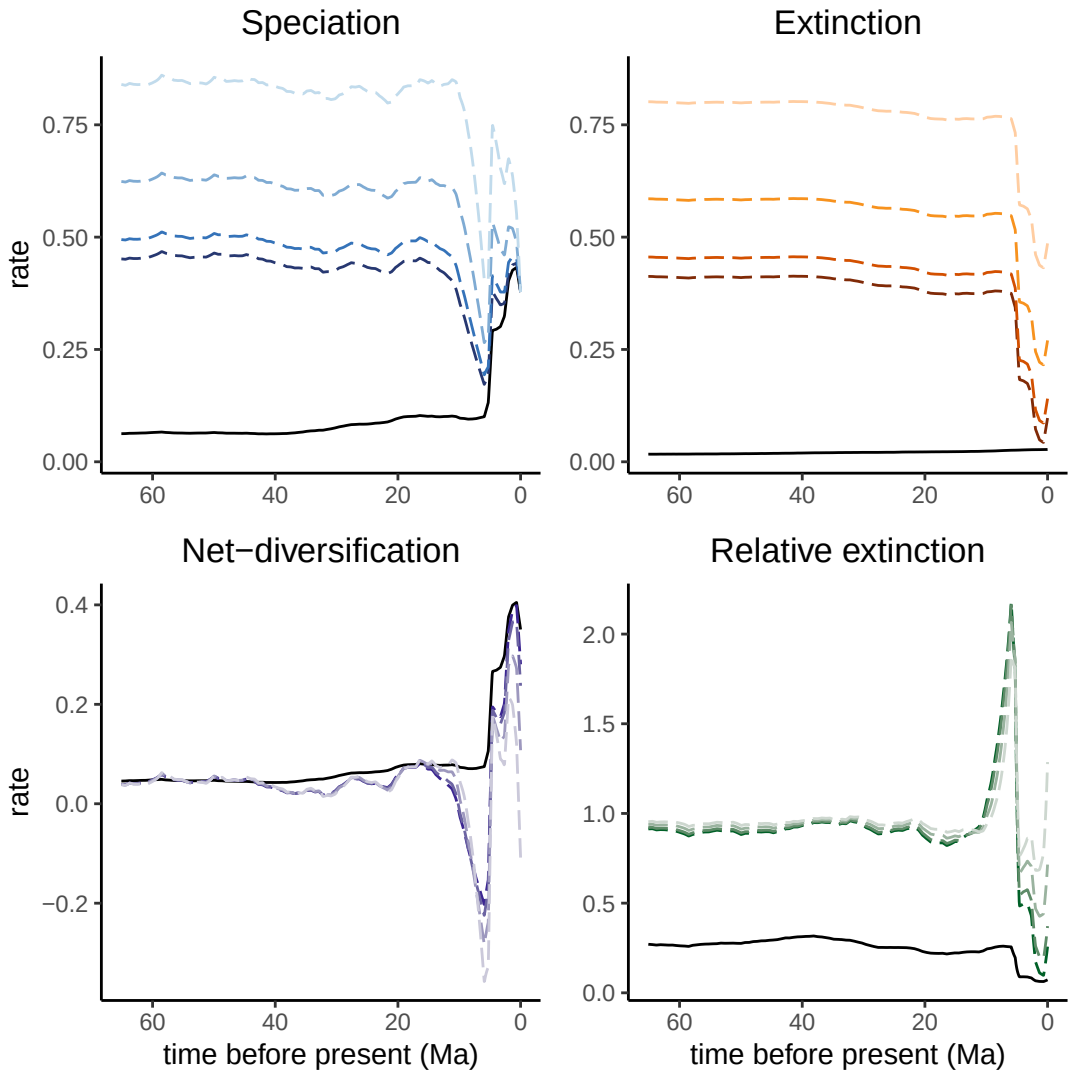


Figure 2.2: Alternative models contained in the congruence class. The solid black line depicts the *reference* model (*i.e.*, the model which we inferred and provided to `CRABS`). The four dashed lines in each plot show four different examples where we assumed that the extinction rate function was driving the net-diversification rates. The speciation rate functions, net-diversification rate functions and relative extinction rate functions are computed (i.e., deduced) given the extinction rate functions to enforce that the models remain in the congruence class.

Table 2.1: Options for sampling basic models using `sample.basic.model`.

Argument	Description
times	the vector of time points that we wish to sample at
rate0	the rate at the present. If not specified, will draw a random log-normally distributed rate at the present
rate0.median	median rate at present
rate0.logsd	(log-scale) standard deviation of the rate at present
model	either "exponential", "linear", "episodic", "MRF" for Markov random field
direction	increase or decrease for the deterministic trend
noisy	whether or not to add stochastic noise (the "MRF" component)
MRF.type	one of "HSMRF", for horseshoe-, or "GMRF" for Gaussian Markov random field
monotonic	whether to enforce that the rate always changes in the same direction
fc.mean	mean fold-change
min.rate	minimum rate value used for rejection sampling
max.rate	maximum rate value used for rejection sampling

```
extinction_rate_samples <- function() {
  sample.basic.models(
    times = primates_ebd[["time"]],
    model = "MRF",
    MRF.type = "GMRF",
    max.rate = 0.5,
    mrf.sd.scale = 2.5,
    rate0.median = mu(0.0))
}
```

Then, we use this function `extinction_rate_samples` —which specifies how to generate new samples for the extinction rate function— to sample from the congruence class, which is done using the function `sample.congruence.class`. Here, we sample 20 new extinction rate functions and automatically compute the corresponding speciation rate functions.

```
samples <- sample.congruence.class(
       my_model,
       num.samples=20,
       rate.type="extinction",
       sample.extinction.rates=extinction_rate_samples)
```

We can plot the BM-samples in `CRABS`, using `plot(samples)` (Fig. 2.3, left and middle columns). Note that we strongly recommend users to apply several different approaches for

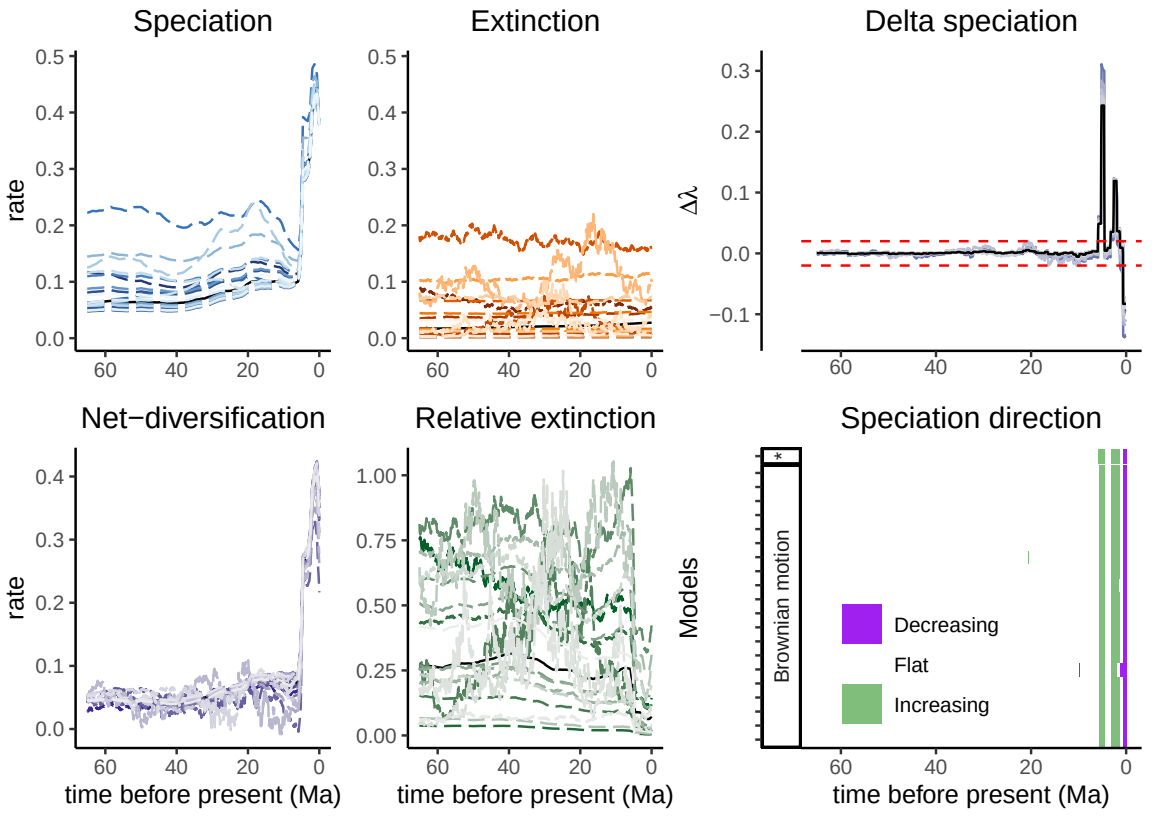


Figure 2.3: Congruent models where the extinction rates were sampled from a Brownian motion. The speciation rates were computed (i.e., deduced) to match the extinction rates so that the models remain within the same congruence class. The net-diversification rates and relative extinction rate functions result from the speciation and extinction rate functions. The right column depicts the change in the speciation rate ($\Delta\lambda = (\lambda_{i-1} - \lambda_i)/\Delta t$), and a summary indicating whether the speciation rate function is decreasing, flat or increasing assuming a threshold for $\Delta\lambda$ of $\pm\epsilon = 0.02$. The asterisk (*) in the trends plot (bottom row, right panel), and the solid black lines represent the *reference model.*

sampling more comprehensively from the congruence class (for examples see our vignette: `https://afmagee.github.io/CRABS`) and to draw more samples than in this demonstrative example.

2.3.6 Summarizing trends over congruent models

Once several models from the congruence class are obtained, we compute summaries of our sampled rate functions. Recall that we are primarily interested in changes in diversification rates over time for these models. Therefore, we compute the amount of rate change within a small interval Δt

$$\Delta\lambda_i = \frac{\lambda_{i-1} - \lambda_i}{\Delta t} \quad , \tag{6}$$

where the index i refers to the i^{th} grid point used for drawing. Next, we extract at which times the change in diversification rate ($\Delta\lambda$) is larger than a pre-defined threshold ϵ: if $\Delta\lambda$ is

larger than ϵ then we paint this time as a rate increase, and if $\Delta\lambda$ is smaller than $-\epsilon$ then we paint this time as a rate decrease. Alternatively, we can compute the normalized rate change of the rate function:

$$\Delta\bar{\lambda}_i = \frac{\Delta\lambda_i}{\lambda_i} \quad . \tag{7}$$

The absolute rate change represents the change in the speciation rate per time, while the normalized rate change represents the fold-change in the speciation rate per time. The absolute rate change has the advantage that we can easily specify a comparable threshold over the entire diversification history, while the normalized rate change has the advantage that we can easily specify a threshold that is comparable between analyses/datasets.

Outliers can signal single rate changes detected by $\Delta\lambda$ but could also be noise. Therefore, we implemented an option to smooth trends (either increases or decreases) by removing singleton outliers. We define an outlier as a time interval where both neighbors share the same trend but the interval itself has a different trend. The outlier is then replaced with the same trend as both its neighbors. This smoothing can clarify the overall signal to detect the total number of directional changes. However, smoothing might delete sharp or instantaneous rate changes.

In `CRABS`, we summarize the directional changes in the congruence class by specifying a threshold ϵ, in units of rate change per time.

```
summarize.trends(samples, threshold = 0.02)
```

The summary provides us with an overview of the trends: how many of the sampled models have speciation rate functions that were increasing or decreasing at a given time. For the example primate dataset and our choice to generate model samples from the congruence class, we observe that sampled models had two speciation rate increases very recently, and one additional speciation rate decrease at the present (Fig. 2.3). A more detailed description of how to summarize and interpret trends is provided in our vignette (`https://afmagee.github.io/CRABS`).

2.3.7 Accommodating uncertainty in rates

In the above example, we explored the congruence class for the posterior median diversification rates. We can also explore congruence classes for samples from posterior distributions. That is, as an example, we compute the congruence class for 20 samples from the posterior distribution and draw 20 alternative rate functions for each posterior sample. First, we load our posterior samples.

```
data(primates_ebd_log)
posterior <- read.RevBayes(
               primates_ebd_log,
               n_times = 1000,
```

```
              max_t = 65,
              n_samples = 20)
```

We plot the speciation rate functions from the posterior sample in the left panel of Fig. 2.4 (see Supplementary Scripts for information about plotting). Next, we generate congruent models for each of the samples from the posterior.

```
samples <- sample.congruence.class.posterior(
              posterior,
              num.samples = 20,
              rate.type = "extinction",
              model="MRF",
              MRF.type = "GMRF",
              max.rate = 3.0,
              rate0.median = mu(0.0))
```

This selection yields 20 samples from the posterior, which are not congruent, but for each 20 samples we generated a subset of 20 additional congruent models. Next, we can summarize and plot the directions of change in the speciation rate function.

```
summarize.posterior(samples, threshold = 0.02)
```

Directional changes are summarized to show the number of models with an increase (or decrease, Fig. 2.4, right panel), rather than to keep results from the same model consistent across rows. In contrast to results based solely on the posterior median model (Fig. 2.3), the posterior samples show more disagreement between trends in direction of rate changes. Nevertheless, we observe general agreement among rate functions for the same three rate changes; two speciation rate increases in the last eight million years followed by one speciation rate decrease very close to the present.

2.3.8 Summary of available functions

Finally, we present an overview of the most important functions available in `CRABS` (Table 2.2). We demonstrated the core functionality in the previous sections, and we also provide a vignette where we explore more detailed features of the package. We designed `CRABS` to have few, but generic functions, that allow for flexibility in exploration of the rate space in the congruence class.

We designed `CRABS` to use standard `ggplot` objects (Wickham, 2016) so that plots can easily be manipulated as any other `ggplot` objects. For example, it is possible to change the axis limits, axis labels, or the time scale. This allows for flexibility in visualizing the congruence class for other data sets than we have exemplified here.

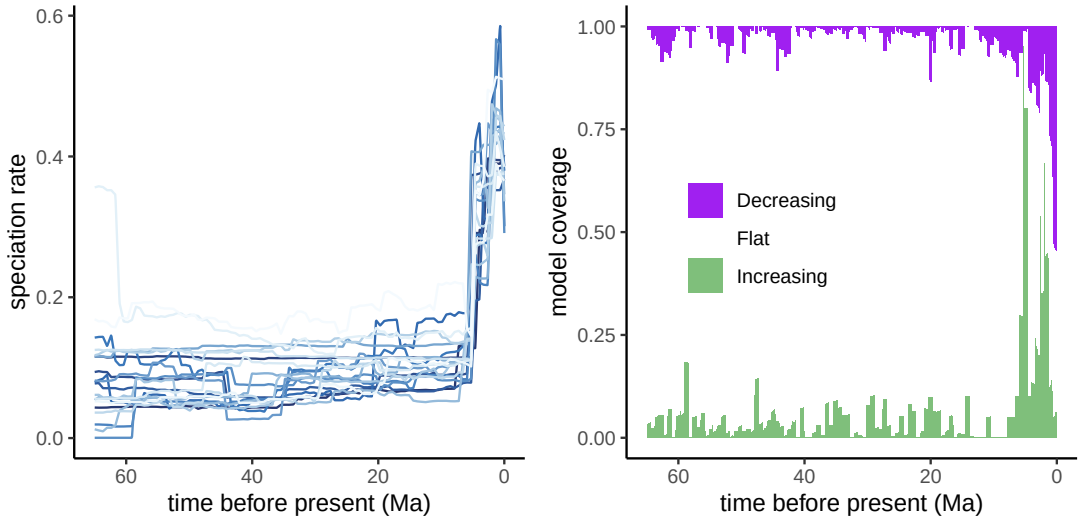


Figure 2.4: Left column: twenty posterior samples from the primates analyses. Right column: summary of trends over the posterior samples and congruence class samples for each posterior sample.

Table 2.2: A summary of the core functions used in CRABS.

Function	Description
congruent.models	constructs models that are congruent with a reference model
create.model	creates a CRABS model object
model2df	converts a CRABS model object to a data frame, e.g., for plotting
plot.CRABS	plots a birth-death model
plot.CRABSset	plots a set of congruent models
sample.basic.models	samples rate functions from various distributions (see Table 2.1)
sample.congruence.class	samples models from the congruence class
sample.congruence.class.posterior	sample from the congruence classes of the posterior samples
sample.rates	samples custom rate functions
summarize.posterior	plots a summary of the directional trends in the posterior
summarize.trends	plots a summary of the directional trends in the congruence class
read.RevBayes	reads a RevBayes log file

2.4 Discussion and conclusions

In this paper we present the `R` package `CRABS`, Congruent Rate Analyses in Birth-death Scenarios. `CRABS` is available on CRAN and the source code is available from GitHub (`https://github.com/afmagee/CRABS`). `CRABS` enables easy testing of the impact of non-identifiable

diversification rates. Specifically, with `CRABS` anyone can explore specific alternative diversification rate scenarios (e.g., explore if a model with constant speciation rates is included within the congruence class), or explore equally probable diversification rate models for shared characteristics. Thus, non-identifiability of diversification rates can be incorporated into conclusions about the process shaping historical biodiversity.

In this paper, we have numerically approximated the congruence class by computing the congruent models at given grid points. Between these grid points we choose to linearly interpolate in our plotting functions. Importantly, our motivation for the exploration of the congruence class is primarily visual (see Fig. 2.3). Alternative models produced by `CRABS` are not —in a strict mathematical sense— in the same congruence class as they do not yield the exact same likelihoods. However, the precision can be arbitrarily increased by selecting a finer time grid. The precision can be increased sufficiently so that the features are visually indistinguishable (see Fig. S2.1).

In our main example of exploring a congruence class, we sampled alternative rate functions from a Brownian motion process. This choice reflects our belief that Brownian motion might be a good approximation of how diversification rates have changed over time, though other approaches should be considered. The biological motivation to apply Brownian motion is that diversification rates rarely change drastically from one time-step to another and the change is instead gradual. We can also model rate changes with rapid shifts using a HSMRF model (Magee et al., 2020). Alternatively, we could assume that rates of speciation and extinction fluctuate around some central tendency, and model the behaviour of the rates using an Ornstein-Uhlenbeck process (see Condamine et al. 2018, Magee et al. 2020 and Palazzesi et al. 2022 for comparisons of diversification rate models through time). To assist with this, `CRABS` provides functions to generate alternative rate functions with stochastic changes, as well as diversification rate functions with deterministic trends (e.g., exponential and linear). The existing functions to explore the congruence class can accept any type of rate functions. Although it remains unclear what the best approach to sample alternative rate functions is, we provide some guidelines here. First, we strongly recommend to use multiple different approaches to draw samples of rate function to suggest alternative rates. These multiple rate functions can be summarized together in the same trend plots. Second, we strongly recommend to explore both alternative speciation and extinction rate functions to not bias rate changes on only either the speciation or the extinction rate.

The primary output of `CRABS` is the congruence class and models belonging to this congruence class. Since changes in diversification rates are generally of interest, rather than the rates themselves, we focus on summaries of the congruence class showing directional trends in diversification rates (increasing, decreasing or flat). However, these summaries strongly depend on the chosen threshold for assessing whether the diversification rate was changing or not. We chose an arbitrary threshold of $\epsilon = 0.02$, which signifies that any rate change of ± 0.02

per million years is a significant trend. We recommend in practice testing summaries using a variety of thresholds. We strongly recommend researchers begin by exploring and visualizing specific models (as in Section 2.3.4), before examining broader sections of the congruence class (as in Sections 2.3.5 and 2.3.6). Additionally, we strongly recommend users to apply different approaches with varying degrees of smoothness for sampling from the congruence class (for examples see our vignette: `https://afmagee.github.io/CRABS`).

Finally, non-identifiability of diversification rates extends to diversification rates inferred from phylogenies with fossil taxa and phylodynamic models (Louca et al., 2021). In `CRABS`, we currently only focus on speciation and extinction rate functions and omit fossilization rates. However, it is possible to analyze a congruence class obtained from a phylogeny with fossil taxa using an approach analogous to what we described here. If a specific new extinction rate function was chosen, then both the speciation rate function and fossilization rate function can be computed given the congruence class. The results with `CRABS` are still valid for speciation and extinction rates, and could be considered as if the fossilization was simply not shown.

Non-identifiability of diversification rates has questioned the reliability and interpretation of diversification rate estimates from current approaches. Non-identifiability is a very real problem and should not be neglected. `CRABS` provides a new tool to assess if the conclusions drawn about patterns of diversification rates are robust despite the existence of infinitely many alternative diversification models.

2.5 Acknowledgments

We thank Luiza Fabreti, Ronja Billenstein and Killian Smith for feedback about visualizing and interpreting congruence classes. We are also grateful to Matthew Pennell, Jérémy Andréoletti and Hélène Morlon for constructive comments that greatly improved this manuscript.

2.6 Data Availability

`CRABS` is available on CRAN and the source code is available from GitHub (`https://github.com/afmagee/CRABS`). A released version is archived and available from Zenodo (`https://doi.org/10.5281/zenodo.7079514`; Kopperud et al., 2022b). A detailed description of example code is provided in our vignette (`https://afmagee.github.io/CRABS`).

Chapter 3

Rapidly changing speciation and extinction rates can be inferred in spite of non-identifiability

Orator Fuller Cook Jr. coined the term **speciation**, meaning the process where one ancestral species splits into two new species.

Cook (1906)

3.1 Significance statement

Evolutionary biologists have long been interested in studying the tempo and mode of emergence and extinction of species. One widely used approach to study rates of speciation and extinction is the stochastic birth-death model. Recent research has shown that the time-varying birth-death model is not identifiable. Specifically, a phylogenetic tree with extant tips does not contain sufficient information to uniquely infer time-varying rates. Non-identifiability can be a major problem if it prohibits drawing a conclusion about the diversification history. However, if the class of equally likely rates leads to the same conclusion, then non-identifiability is a minor problem. We show that periods of rapid change can still be inferred, and therefore qualitative conclusions can still be drawn.

Keywords: macroevolution, diversification, birth-death, identifiability, congruence class

3.2 Abstract

The birth-death model is commonly used to infer speciation and extinction rates by fitting the model to phylogenetic trees with exclusively extant taxa. Recently, it was demonstrated that speciation and extinction rates are not identifiable if the rates are allowed to vary freely over time. The group of birth-death models that have the same likelihood is called a congruence class and there is no statistical evidence to favor one model over the other. This issue has led researchers to question if and what patterns can reliably be inferred from phylogenies of only extant taxa, and whether time-variable birth-death models should be fitted at all. We explore the congruence class in the context of several empirical phylogenies as well as hypothetical scenarios. For these empirical phylogenies we assume that we inferred the true congruence class. Thus, our conclusions apply to any empirical phylogeny for which we robustly inferred the true congruence class. When we summarize shared patterns in the congruence class, we show that strong directional trends in speciation and extinction rates are shared among most models. Therefore, we conclude that inference of strong directional trends is robust. Conversely, estimates of constant rates or gentle slopes are not robust and must be treated with caution. Interestingly, the space of valid speciation rates is narrower and more limited in contrast to extinction rates, which are less constrained. These results provide further evidence and insights that speciation rates can be estimated more reliably than extinction rates.

3.3 Introduction

Macroevolutionary species diversification is often investigated using molecular phylogenetic trees, i.e., phylogenetic trees with only extant taxa (Ricklefs, 2007; Morlon, 2014). Phylogenetic trees are an instrumental resource for learning about diversification phenomena such as rapid radiations (Boschman and Condamine, 2022; Melo et al., 2022), key innovations (Burress and Muñoz, 2021), mass extinctions (May et al., 2016; Magee and Höhna, 2021), biogeographic drivers (Goldberg et al., 2011; Thomson et al., 2021), diversity-dependence (Etienne et al., 2012), state-dependence (Maddison et al., 2007), or environmental drivers (Condamine et al., 2019; Palazzesi et al., 2022). These applications typically use some variation of the reconstructed birth-death model (Nee et al., 1994b) fitted to an extant phylogenetic tree. The birth-death model has two main components: the speciation rate (λ) that controls the tempo of emergence of new species (birth), and the extinction rate (μ) that controls the tempo of elimination of species (death). In this study, we consider birth-death models that do not vary across lineages, but do allow for rates to vary through time (Stadler, 2011; Magee et al., 2020).

Recently, Louca & Pennell (Louca and Pennell, 2020b) demonstrated that diversification rates are not identifiable when using birth-death models fitted to extant phylogenetic trees. Given continuously differentiable rate functions, they presented a class of so-called congruent models that are equally likely. A congruence class is defined by the set of rate functions that yield exactly the same distribution of reconstructed phylogenies. One consequence of this definition of the congruence class is that all models within a congruence class have the same speciation rate at the present, $\lambda(t = 0) = \lambda_0$. Additionally, some extinction rate functions require negative speciation rates to yield the same likelihood (Louca and Pennell, 2021) which renders these models biologically invalid. Both constraints provide grounds for optimism that the space of diversification rate models within a congruence class is (locally) restricted.

Unfortunately, no amount of data can alleviate the problem of non-identifiability, i.e., time-varying diversification rates cannot be inferred even for a phylogeny of millions of extant taxa. One approach for non-identifiable models is to estimate compound parameters, such as θ in population genetics (the product of population size and mutation rate, Watterson, 1975) and branch lengths in phylogenetics (the product of the evolutionary time and the clock rate). Once the compound parameter is estimated, one can disentangle both components by additional assumptions, e.g., the mutation rate in population genetics or fossil calibrations or relaxed clock priors in phylogenetics (Donoghue and Yang, 2016). For time-varying diversification rates, Louca & Pennell (Louca and Pennell, 2020b) proposed to infer diversification using transformed parameters: so-called pulled rates. Pulled diversification rates are identifiable, however they can be challenging to interpret in the framework of the birth-death model (Helmstetter et al., 2022) and additional knowledge to later disentangle the rates is not yet known. Another approach is to focus on space of models itself that are non-identifiable. Specifically, there are infinitely many rate functions that are increasing, as there are infinitely many values

between, say, 0.9 and 1 ($0.9 < \lambda < 1$). Nevertheless we can draw qualitative conclusions that all rate functions are increasing, or that all values of λ are positive and smaller than 1. Thus, time-varying diversification rates may be constrained albeit non-identifiable.

One concern is that congruence classes can contain contradictory temporal trends: for example, increasing, decreasing and flat rate functions in the same time period (Louca and Pennell, 2020b). Conversely, it is possible that some diversification rate patterns (e.g., shifts or directional trends) are shared among all models within a congruence class, even if there are infinitely many models. We wish to know whether certain rate patterns can reliably be inferred, and so we need to analyze the congruence class. Here we explore the congruence class using the recently developed method `CRABS` (Höhna et al., 2022). Specifically, we explore a variety of alternative models within the congruence class, including both specific alternative rate hypotheses and rates sampled randomly from stochastic processes. Finally, we inspect and analyze the produced alternative models for common trends.

In this study, we investigate time-varying diversification rates in a set of phylogenetic clades from Condamine et al. (Condamine et al., 2019). We use the time-calibrated phylogenetic trees to estimate diversification rates under an episodic birth-death model using Bayesian inference (Magee et al., 2020), which gives one model from the congruence class. The episodic birth-death model uses constant speciation and extinction rates within an interval but arbitrary rates between intervals and, with sufficiently many episodes and smoothing priors, can efficiently be fitted as almost any arbitrary rate function (Magee et al., 2020; Palazzesi et al., 2022). From the available datasets, we select three focal datasets that we believe are useful exemplars to illustrate how the congruence class behaves under certain scenarios. These scenarios include (a) constant rates, (b) a single rate-shift, and (c) a more complex history of rate increases and decreases. These three datasets, which we will discuss in detail, are: the New World/African true parrots (Psittacidae, n = 330), the tyrant flycatchers (Tyrannidae, n = 419), and the woodpeckers (Picidae, n = 223). A selection of nine additional empirical datasets can be found in the supplemental material. To capture a more complete range of possible diversification scenarios, we augmented our selection of empirical datasets with hypothetical scenarios. Specifically, we considered rates with exponential increase and decrease, sigmoidal increase and decrease (i.e., an up-shift and a down-shift), a modal burst event (i.e., a shorter period of time with accelerated rates), and linear increase and decrease. Taken together, these empirical and hypothetical models cover a wide range of realistic diversification scenarios. Importantly, here we assume that we inferred the true congruence class and wish to explore the characteristics of this congruence class. Thus, our examples serve as a reference for other studies that inferred similar congruence classes regardless of the size of phylogeny. However, robustly inferring the congruence class in the first place is not the focus of this study.

3.4 Results

3.4.1 Estimating time-varying diversification rates

We estimated time-varying diversification rates using an episodic birth-death model (Stadler, 2011; Höhna, 2015) implemented in `RevBayes` (Höhna et al., 2016a). We used the horseshoe Markov random field (HSMRF) distribution as a prior distribution on the speciation and extinction rates to model autocorrelation among the episodes (Magee et al., 2020). The HSMRF distribution can be compared to a Brownian motion, as both distributions bring about a temporal smoothing effect through time. Unlike a Brownian motion, however, the HSMRF distribution has fat tails, so large jumps between episodes are more probable (Magee et al., 2020). This analysis provides us with a reference model with speciation and extinction rates that we will investigate further. We estimated diversification rates for twelve datasets although we only show the results of three exemplary datasets in the main text (New World/African true parrots, tyrant flycatchers and woodpeckers) and nine more examples are shown in the Supplementary Material (Figs. S3.1 to S3.3).

The extinction rates for all three datasets are approximately constant (Figs. 3.1 and 3.2). The diversification history of the New World/African true parrots is relatively constant; they diversified at the same approximately constant rate throughout the entire span of the phylogeny (Fig. 3.2, Psittacidae). The speciation rate of the tyrant flycatchers shows a single rate shift (Fig. 3.2, Tyrannidae). Thus, their diversification history can be split in three parts: first a period of constant low diversification until around 45 Ma, second a short period of a sharp increase in speciation rate, and third another relatively constant but high period of diversification starting at 40 Ma and continuing towards the present (Fig. 3.2, Tyrannidae). The inferred speciation rate of woodpeckers depicts a more complex diversification rate scenario with increases and decreases. The speciation rate inferred using the woodpeckers dataset exhibit an initial speciation rate increase around the time the wrynecks (Jynginae) and the piculets (Picumninae) split off (25 Ma, Fig. 3.2, Picidae). After that, there is a constant period (25-18 Ma), followed by another increase in the speciation rate corresponding to the radiation of the true woodpeckers (Picinae). After peaking at around 8 Ma, the woodpecker speciation rate decreases towards the present. The three exemplary datasets show three qualitatively distinct patterns: constant vs. single shift vs. complex diversification. In the following analyses, we assume that we have inferred the correct congruence class (i.e., one representative model in the infinite set of equally likely models). We refer the reader to other studies that focus on estimating the correct congruence class (e.g., Louca et al., 2021).

3.4.2 Alternative speciation rates may be strongly constrained

We used the R-package `CRABS` (Höhna et al., 2022) to explore the congruence class, i.e., to compute the pulled diversification rates and to propose alternative congruent models. `CRABS`

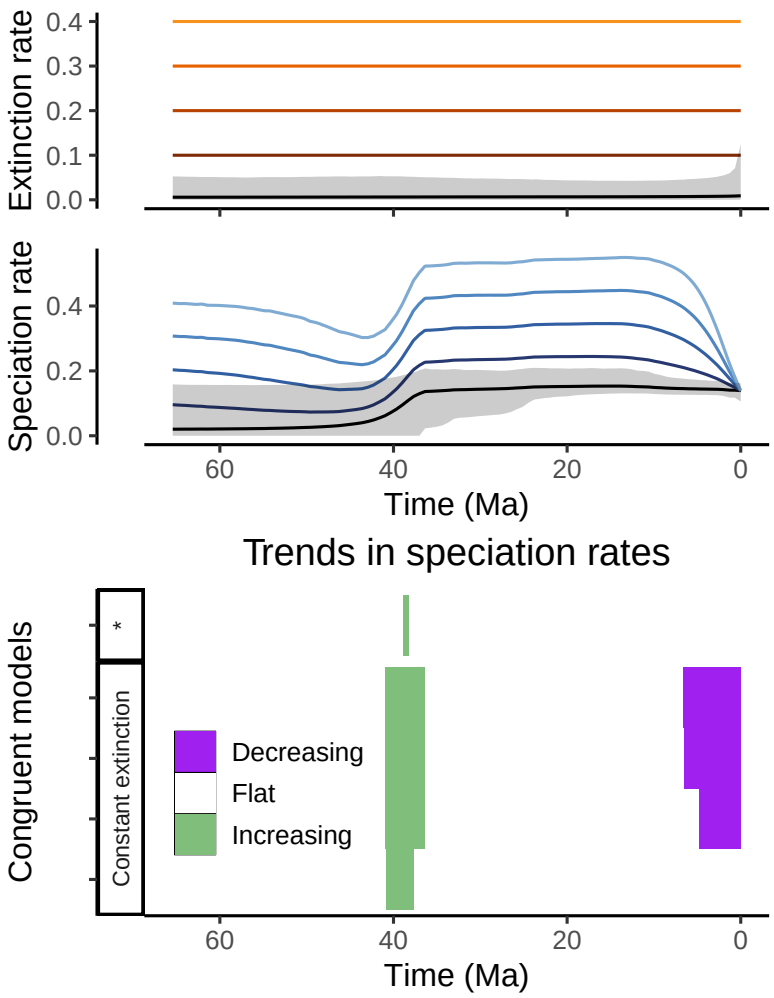


Figure 3.1: Exploring the congruence class by sampling alternative models. Top plot: We estimated speciation and extinction rates for the tyrant flycatchers (Tyrannidae) using an episodic birth-death model implemented in `RevBayes`. The black line represents the posterior median, and the shaded area is the 95% credible interval. The reference model (in black, and denoted *) has a near-zero extinction rate, and a time-varying speciation rate with a period of rapid increase. We proposed four alternative extinction rates ($\mu(t) \in \{0.1, 0.2, 0.3, 0.4\}$, in orange) and used `CRABS` to compute the congruent speciation rates (in blue). Bottom plot: When the slope of the speciation rate ($\Delta\lambda_i = (\lambda_i - \lambda_{i-1})/\Delta t$) is steep and positive ($\Delta\lambda > \epsilon = 0.02$), we consider the trend to be increasing, and analogously for decreasing ($\Delta\lambda < -\epsilon$). This figure corresponds to a subset of Fig. 3.2 (Tyrannidae, congruent models C, *).

takes in a proposed alternative speciation (or extinction) rate, and uses the pulled diversification rate to determine the extinction (or speciation) rate necessary to remain in the congruence class (Fig. 3.1).

We made a detailed examination of alternative models for the New World/African true parrots, tyrant flycatchers and woodpeckers (Fig. 3.2). Using a set of alternative but qualitatively different rate functions (see Figs. 3.1 and 3.2 and Figs. S3.5 to S3.7), we proposed both new extinction rates and new speciation rates. We found that proposing alternative speciation

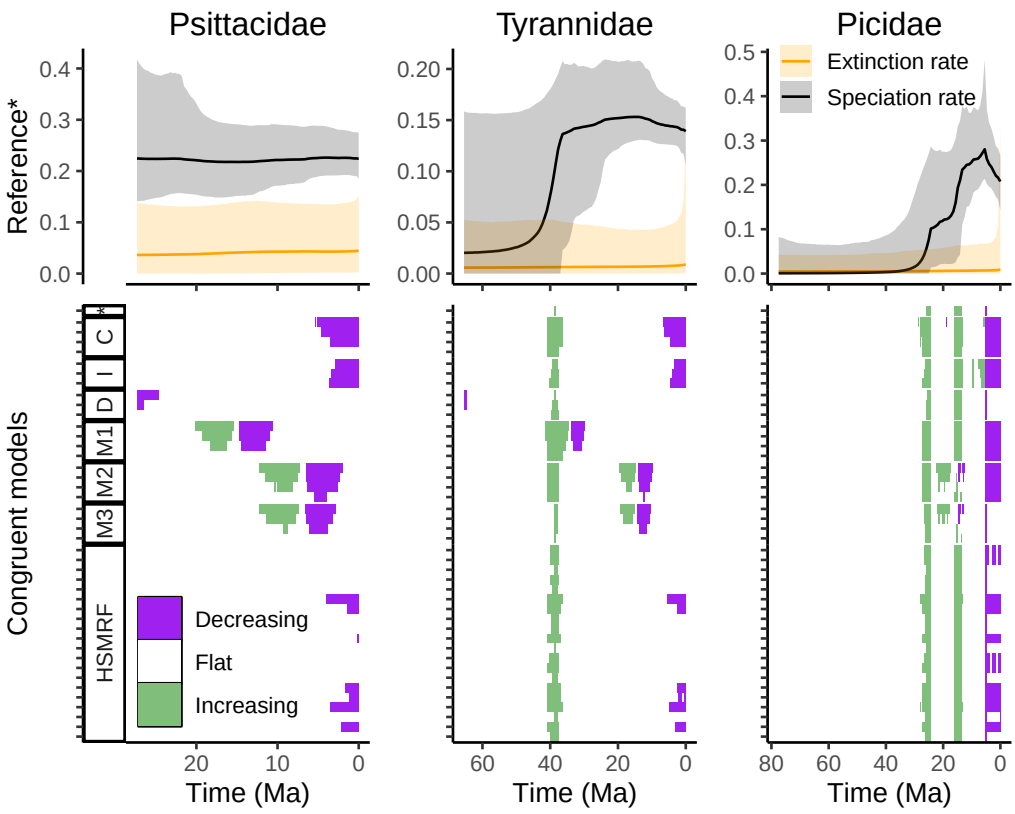


Figure 3.2: A summary of the congruence class for three bird families: New World/African true parrots (Psittacidae), tyrant flycatchers (Tyrannidae), and woodpeckers (Picidae). The top row depicts the posterior median and 95% credible interval speciation and extinction rates, for the episodic birth-death model implemented in `RevBayes`. The bottom row depicts summaries of directional trends in the congruence class, where each row is a model. The models include alternative extinction rates that are constant (C), exponentially increasing (I), exponentially decreasing (D), modal (M1-M3), HSMRF-distributed (horseshoe Markov random field), and the reference model (*). If the slope of the speciation rate function is greater than the threshold $\epsilon = 0.02$ (in speciation rate units per million years), or less than $-\epsilon$, we say that the function is increasing, or decreasing, respectively. The trends in the net-diversification rate can be seen in Fig. S3.1, an additional nine datasets can be found in Figs. S3.2 to S3.4, and the exact rate shapes of the alternative models are in Figs. S3.5 to S3.7.

rates is more difficult than proposing alternative extinction rates. When we proposed alternative speciation rates, we often obtained corresponding extinction rates that were negative. This implies that many of the proposed speciation rate functions are not valid choices. Thus, it appears that all extinction rate functions are contained within the congruence class but not all speciation rate functions. In the next sections, we focus on proposing new extinction rates rather than new speciation rates.

3.4.3 Rapid changes are shared among congruent models

We wish to compare shared features across the models, which raises the question, how do we compare features among birth-death models? Biologists are often interested in whether

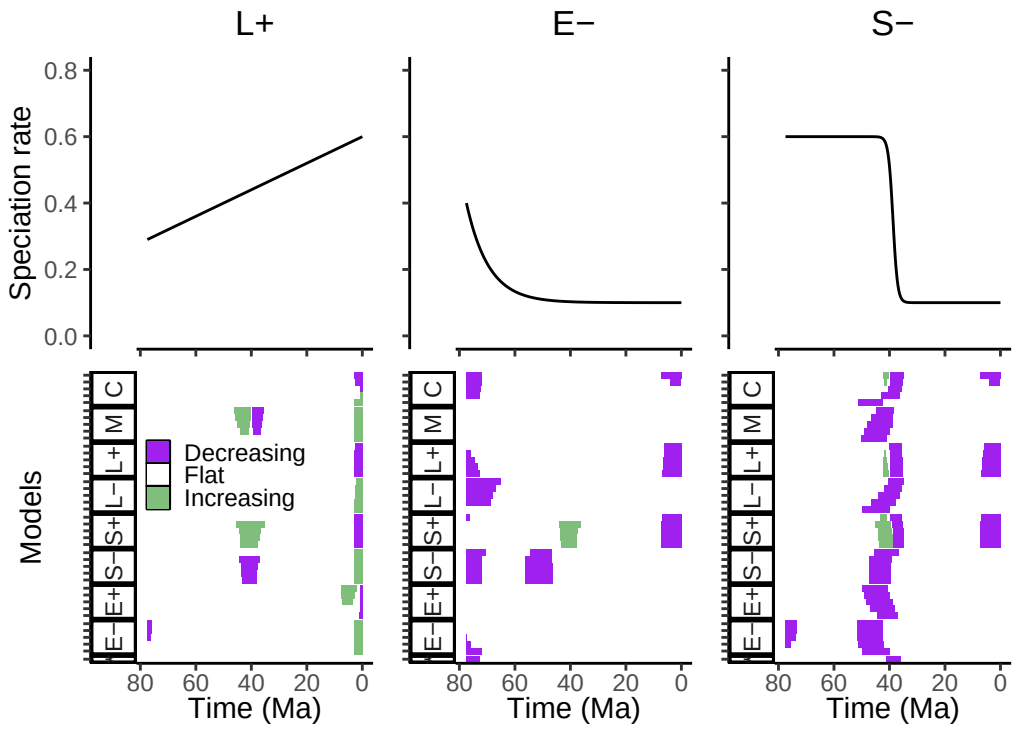


Figure 3.3: Three hypothetical scenarios: a linear increase in speciation rate, an exponential decrease, and a sigmoidal decrease. The reference model is the speciation rate in black, and a constant extinction rate of $\mu = 0.28$. For each congruence class, we proposed various alternative extinction rates: constants (C), modals (M), sigmoidal increase (S+), sigmoidal decrease (S–), linear increase (L+), linear decrease (L–), exponential increase (E+) and exponential decrease (E–), with five rate trajectories for each of these shape categories (Fig. S3.18). Each congruence class results in a set of models with inferred speciation rate curves (Figs. S3.22, S3.23 and S3.26). We used a threshold of $\epsilon = 0.02$ rate units per Ma to compute the significant directional trends in the inferred speciation rates. We selected a time scale similar to the woodpecker phylogeny for interpretability. See Fig. S3.17 for the trends in the net-diversification rates.

the rates increased, decreased, or remained constant. Thus, we investigated if the directional trends were significantly increasing or decreasing within a short time period. First, we show the significant directional trends for the New World/African true parrots (Psittacidae), the tyrant flycatchers (Tyrannidae), and the woodpeckers (Picidae). Then we show the significant directional trends for the hypothetical scenarios.

We explored the congruence class for three empirical datasets (Fig. 3.2). We used a variety of alternative extinction rate trajectories to represent a wide range of possible extinction scenarios (Figs. S3.5 to S3.7). The significant directional trends in the speciation rates are consistent across the congruence class when the slope of the speciation rate is steep (Fig. 3.2, Tyrannidae and Picidae). Specifically, we were able to recover the following events. The single rate shift in the tyrant flycatchers is unanimously represented in our sample of the congruence class (Fig. 3.2, middle column, bottom plot). For the woodpeckers, we were able to recover two events where the speciation rate increased, and another event where the

speciation rate decreased (Fig. 3.2, right column, bottom plot). When the speciation rate is relatively constant or shallow, then there is no consistent directional trend in the speciation rate across the congruence class. In other words, when the rates are constant, it is trivial to propose an alternative congruent model that contradicts the diversification pattern of the reference model. These results are corroborated by an additional nine empirical datasets that we present in the Supplementary material (Figs. S3.2 to S3.4).

Despite analyzing several empirical systems, there were only a limited set of interesting scenarios. For example, we never inferred speciation rates that were monotonously decreasing across the time span of the phylogeny, or that resembled exponential growth or decline. To complement our analysis of the behaviour of the congruence class using empirical datasets, we added hypothetical scenarios with more variable shapes. These scenarios are commonly used in diversification rate analyses (see Höhna, 2014; Morlon et al., 2016; Höhna et al., 2016b; Magee et al., 2020). Specifically, we assessed a shallow linear increase, an exponential decrease, and a sigmoidal decrease (Fig. 3.3). The assessments from the hypothetical scenarios are similar to the empirical scenarios. A shallow change in the speciation rate is not well recovered by the inferred rates in the congruence class (L+ and E–, Fig. 3.3). The steep descent of the sigmoidally decreasing model is recovered by our sample of the congruence class. Unlike the empirical datasets, however, this descent is not precisely overlapping in time. The proposed models exhibit a decrease in speciation rate between 30 and 50 Ma, but with some temporal distortion (Fig. 3.3, S–). In the supplemental material, we also investigated hypothetical models where the extinction rate has a distinct pattern, and we explored the congruence class by proposing alternative speciation rates. Similar to speciation rates, we show that rapidly changing extinction rates are also robust to the congruence class (Figs. S3.27, S3.28, S3.30 and S3.32 to S3.34). See also Fig. S3.35 and the associated paragraph for a discussion of how rapid is sufficiently rapid.

For both the empirical and hypothetical scenarios, we observed the strongest disagreement of trends in the intervals near the present (Figs. 3.2 and 3.3). Speciation rate trends near the present are strongly driven by the constraint that all speciation rate functions must be equal at the present. By choosing higher alternative extinction rates than the reference extinction rate, we obtain higher alternative speciation rates for most of the evolutionary history which then must drastically drop towards the present to λ_0 (Figs. S3.5 to S3.7). Conversely, lower alternative extinction rates than the reference extinction rate leads to lower speciation rates until the present where the speciation rate rapidly increases towards λ_0. Abrupt rate change in the speciation rate near the present may be an artefact of the congruence class. However, if one judges the actual rate values instead of the trends, then the rates are most robust in the near present (Fig. 3.4).

3.4.4 Speciation rates are more constrained than extinction rates

While setting up congruent models by proposing alternative rates, we observed that extinction rates varied more than speciation rates. In other words, the resulting distribution of congruent speciation rates was more constrained than the distribution of congruent extinction rates. In order to investigate this in a quantitative manner, we assessed a hypothetical model with a single steep rate increase (time-varying λ and constant μ, Figs. 3.4 and S3.16). Next, we proposed random alternative rates (drawn from an autocorrelated Brownian motion, an Ornstein-Uhlenbeck process, and uncorrelated samples from a lognormal distribution). The variance across the draws of congruent rates at a given time was in all cases greater for extinction rates than for speciation rates (Figs. 3.4 and S3.16) which confirms that the space of speciation rates is smaller than the space of extinction rates. Interestingly, the variance in the speciation rate is diminished in the period where the rate shift happens. This lends further credibility to the notion that periods of rapid change are more reliably inferred than periods of gentle or no change in rates. Similarly, the uncertainty, represented by the standard deviation of the rates at a time, is smaller closer towards the present. This shows that the actual value of the speciation rate is more robustly inferred towards the present.

3.4.5 The effect of diversification rate uncertainty is substantial

Thus far we have explored the congruence class and investigated the robustness of diversification rate patterns of a single point estimate of the rates for each episode, either the posterior median or a hypothetical rate. However, we must acknowledge there is substantial uncertainty in the estimates of the speciation and extinction rates. There is uncertainty both in the magnitude of the rates but also in the timing of the rate shifts (Fig. 3.2, top row; Fig. S3.11). The uncertainty in the timing of the rate shifts motivated us to test different window sizes to assess whether rates have changed within a larger time span (Fig. S3.12).

Recall that the posterior median trend summaries had some unambiguous trends: the three rapid changes in the woodpeckers dataset and the single speciation rate increase in the flycatcher dataset (Fig. 3.2). When we instead study the congruence class and estimation uncertainty together, then the signal of the three intervals is only partially present (Fig. 3.5). Thus, the diversification rate trends are no longer unambiguous. The assessment of the rate shift is sensitive to the window size, with too small and too large windows not recovering the trend (Fig. S3.12). Regardless of the chosen window, no diversification-rate patterns are preserved across all, or even half, of the posterior samples. In our examples, the variation across models sampled from the posterior is much greater than the variation from within each of those congruence classes.

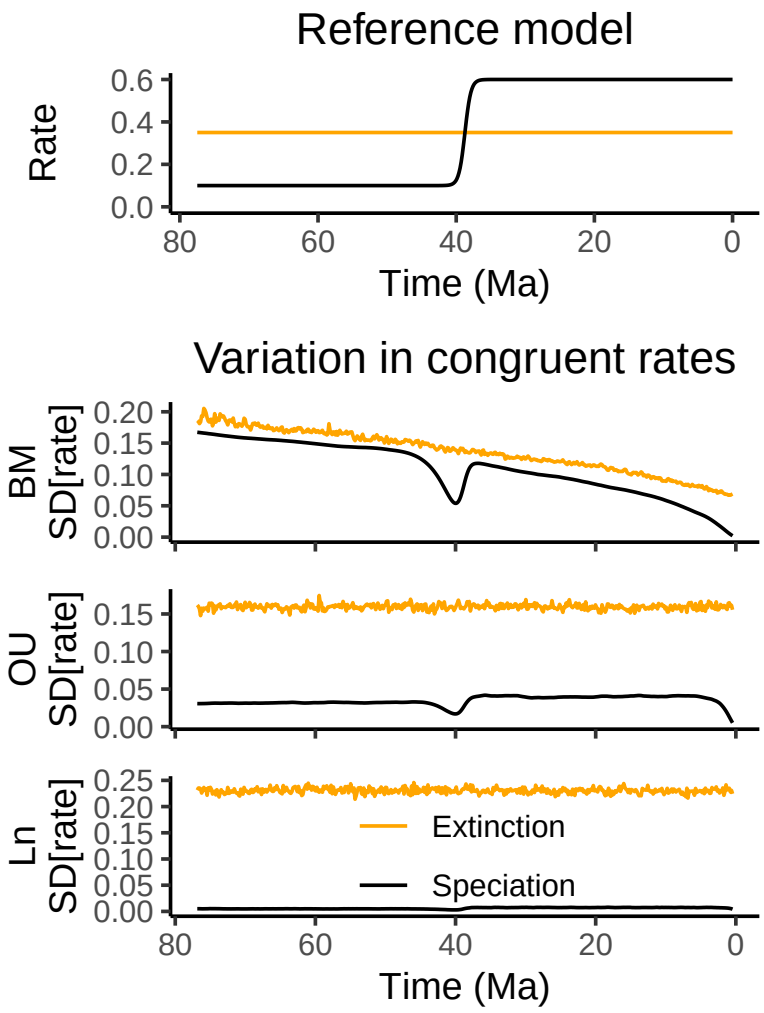


Figure 3.4: We set up a hypothetical reference model with a sigmoidally increasing speciation rate (black), and a constant extinction rate (orange). In the following rows we proposed rates that were drawn from a Brownian motion (with $\sigma = 0.02$), an Ornstein-Uhlenbeck process (with $\sigma = 0.05$, $\alpha = 0.5$, $\theta = 0.6$), and rate values drawn independently from a lognormal distribution (with mean = 0.6, logsd = 0.05). For all proposed speciation and extinction rates we set the initial value at the present to 0.6. The standard deviation (SD) illustrates the amount of variation in the congruent rates, for each time interval. Importantly, the standard deviation of the extinction rate was always higher than the standard deviation of the speciation rates, $\mathrm{SD}_\mu(t) > \mathrm{SD}_\lambda(t)$, which indicates that the space of speciation rate functions is smaller compared to the space of extinction rate functions within the congruence class. See Fig. S3.16 for an extended version.

3.5 Discussion

3.5.1 Accuracy of speciation rate estimates near the present

The common expectation is that diversification rates can be estimated more robustly in the near present compared with the more ancient parts of the evolutionary history of the study group (e.g., May et al., 2016). From the definition of the congruence class, the value of the

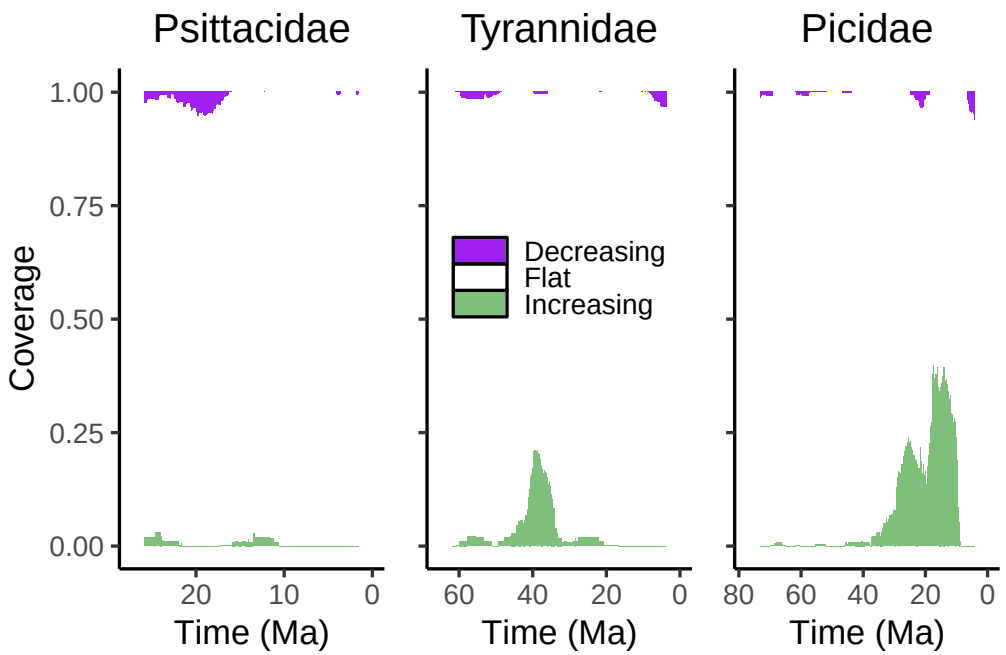


Figure 3.5: The interplay between the variation of rates in the posterior distribution and the congruence class. Each panel represents 110 samples from the posterior distribution. These samples are not congruent. For each of the samples, we constructed an additional 10 congruent models with horseshoe-distributed extinction rates. The coverage represents the proportion of models that had a decreasing, increasing, or constant speciation rate, for a particular time slice. The coverage is sorted, meaning each row does not correspond to a single model. We calculated the slopes using a window size of 55 ($\Delta\lambda_i = (\lambda_i - \lambda_{i-55})/(t_i - t_{i-55})$). For other window sizes see Fig. S3.12.

speciation rate at the present is fixed to λ_0 for all models. This means that if we are able to infer the correct congruence class, then we are also able to infer the correct speciation rate at the present. Indeed we observed that the uncertainty in rate estimates is smallest near the present (Fig. 3.4, Fig. S3.16). However, the trend of the speciation rate towards the present is entirely determined by the choice of the extinction rate in the near present. A greater assumed extinction rate near the present results in greater a speciation rate near the present and thus a decline towards the present. On the other hand, a smaller assumed extinction rate near the present results in a smaller speciation rate near the present and thus an increase towards the present. Paradoxically, our assessment of trends in speciation rate is most volatile in the near-present. This means our ability to estimate the actual rates is most accurate near the present, however our ability to assess the trends in rates is worse near the present.

3.5.2 The space of valid speciation rates versus extinction rates

We demonstrated that the variation in congruent extinction rate is greater than the variation in congruent speciation rates, in a typical model where the speciation rate varies through time. The reason for this discrepancy in distribution is not well understood. However, we

suspect that the speciation rates being constrained at the present (λ_0) contributes to this effect. Indeed, from our albeit limited exploration of congruent models through proposals of alternative rates, it appears as if all extinction rates are included in the space of valid rate functions. On the other hand, when we proposed an alternative speciation rate that was smaller than the reference rate, this sometimes led to negative extinction rates (see Figs. S3.31 to S3.34). This phenomenon has also been observed by Louca & Pennell Louca and Pennell (2021) where the alternative model within the congruence class had a negative extinction rate. Negative extinction rates do not make sense, and if we exclude those rate shapes, then the valid space for speciation rates is shrunk even further.

We conclude that the space of valid alternative speciation rates is narrower than the space of valid alternative extinction rates. This supports previous evidence that speciation rates are more reliably and accurately estimated than extinction rates from molecular phylogenies (e.g., Maddison et al., 2007; Pyron and Burbrink, 2013; Höhna, 2014; Title and Rabosky, 2019). Furthermore, since in all our examples the pure-birth model was contained in the congruence class, inferring the congruence class using a pure-birth model may be a more robust and efficient approach for inference of the congruence class (see also Helmstetter et al., 2022).

3.5.3 Posterior distribution

We have assessed the congruence class issue primarily by using a single diversification rate (the posterior median, since we used Bayesian inference). Other lines of research that use maximum likelihood to fit the model also typically result in a single rate trajectory. Both the posterior median and the maximum likelihood estimate are point predictions of the model fit; neither indicate whether the rate estimates are precise or uncertain. We have demonstrated that the uncertainty induced by credible intervals is substantial in our selection of models. The rate uncertainty can also be substantial for models inferred using maximum likelihood, even if the uncertainty is not explicitly investigated as is often the case.

Our results indicate that the variation of rates induced by the congruence class is, in a sense, less than the variation represented by the inherent uncertainty in the time-varying model. In other words, we argue that the congruence class is a smaller problem compared to the uncertainty in rate estimates of time-varying models. The inferred trends were robust to the congruence class but less so to the posterior uncertainty. If a shift was present in the reference model, then all models in the congruence class agreed on the same shift unambiguously, whereas the posterior samples varied in the timing of the rate shifts.

One of the main challenges for inferring trends when including estimation uncertainty is the timing of the rapid rate shift. Often, there is considerable uncertainty of the exact timing of the rate shift but less uncertainty whether the shift occurred. However, our current trend analyses focus on whether a trend occurred at a given time interval, and therefore did not agree among all diversification rate samples. This problem is potentially more severe when

phylogenetic and divergence time uncertainty is included in the assessment of diversification rate patterns. Instead of our test for a rate change at a given time interval one might instead ask if there was at least one rate change (or how many such rate shifts there were).

3.6 Concluding remarks

Non-identifiability of diversification rates is a real issue and should be addressed when studying temporal patterns in diversification histories. Yet, the congruence class may not be a problem when testing for macroevolutionary diversification hypotheses in a qualitative fashion. Based on our selection of datasets and alternative models, we argue that abrupt changes in speciation and extinction rates are robust to the congruence class. Conversely, periods of constant or flat speciation and extinction rates are not robust, and it is trivial to construct congruent but contradictory evolutionary hypotheses. Furthermore, the restriction of the congruence class that all models share the same speciation rate at the present λ_0, often induces diversification rate changes near the present. Rates are robustly estimated in the near present, however rate shifts near the present are not and should be treated cautiously. As such, we argue that while quantitative rate estimates are difficult to infer, it is still possible to draw qualitative conclusions about the general pattern. Additionally, we conclude that the space of possible speciation rate functions is more restricted than the space of possible extinction rate functions.

3.7 Materials and methods

3.7.1 Empirical datasets used as case studies

The empirical phylogenetic trees were obtained from Condamine et al. (Condamine et al., 2019). Specifically, the three highlighted bird families were derived from the study by Jetz et al. (Jetz et al., 2012). In our study, we used a single phylogeny as the observation, thus treating the phylogenetic trees as known without error. We also knew the proportion of taxa sampled (ρ) as the fraction of tips divided by the number of described taxa. This sampling fraction was directly used in our study to correct for uniform incomplete taxon sampling (Höhna et al., 2011).

3.7.2 Estimation of the reference diversification model

We fitted an episodic birth-death model to the empirical trees in `RevBayes` (Höhna et al., 2016a). Episodic models divide the time period into discrete bins, i.e., intervals, and we assume constant speciation and extinction rates within an interval. By incorporating sufficiently many time bins, an episodic model can approximate a model with continuously time-varying rates. Evidence suggests that it is possible to estimate the correct congruence class, i.e., any model that is congruent to the true model and thus has the same likelihood as the true model (Magee

et al., 2020), if not the true model when using episodic models (Legried and Terhorst, 2022). We treated the rate functions as piecewise constant (100 pieces), and we used a horseshoe Markov random field (HSMRF) process to specify a prior distribution on the speciation and extinction rates. We followed Magee et al. (Magee et al., 2020) when specifying the parameters of the HSMRF distribution.

Specifically, we assumed the following prior distribution for the extinction rate μ_0 at the present and all consecutive extinction rates μ_i:

$$\begin{aligned} \xi &= 0.0021 \\ \gamma &\sim \text{HalfCauchy}(0,1) \\ \ln(\mu_0) &\sim \text{Uniform}(-10,10) \\ \sigma_i &\sim \text{HalfCauchy}(0,1) \\ \ln(\mu_i) &\sim \text{Normal}(\ln(\mu_{i-1}), (\sigma_i \times \xi \times \gamma)^2) \end{aligned} \tag{1}$$

where ξ scales the overall magnitude of the changes in accordance with the number of episodes (here 100), and σ_i is the local scaling factor. We picked $\xi = 0.0021$ such that we expected *a priori* $\ln(2)$ shift events of magnitude two or higher. An exactly matching set of prior distributions were assumed for the speciation rates.

We used `RevBayes` to fit the model in a Bayesian framework, and we used the Metropolis-Hastings algorithm to sample from the posterior distribution. We ran four chains in parallel for 50,000 iterations, discarded the first 5,000 iterations as burnin, and compared the chains using the Kolmogorov-Smirnov test to assess convergence (Fabreti and Höhna, 2022, see Supplementary materials). We used the posterior median estimates for extinction rate and speciation rate for further analyses.

3.7.3 Exploring the congruence class

We imported the fitted time-varying birth-death models into the R-package `CRABS`: Congruent Rate Analyses in Birth-death Scenarios (Höhna et al., 2022). We used the pulled net-diversification rate, r_p, to construct the congruence class (Louca and Pennell, 2020b):

$$r_p(t) = \lambda(t) - \mu(t) + \frac{1}{\lambda(t)} \frac{d\lambda}{dt}, \tag{2}$$

where t is in units of time before present. `CRABS` uses numerical approximations, i.e., piecewise linear rate functions, to solve the derivative and to construct the congruence class. We used a grid of 500 time points for the piecewise approximations, which has been shown to give precise approximations of the underlying continuous functions (Höhna et al., 2022).

In order to construct an alternative model in the congruence class, `CRABS` requires an alternative extinction or speciation rate function. Then, `CRABS` solves Eq. 2 for either λ or

μ, depending on which alternative rate function is proposed, to compute the corresponding alternative extinction or alternative speciation rate function, respectively. We detail how the alternative rate functions are proposed in the next sections.

Specific diversification hypotheses

We employed two strategies to sample alternative models from the congruence class, by generating alternative extinction or speciation rates. First, we selected a set of specific hypotheses that are biologically distinct. For the extinction rates, we constructed constant rate functions, increasing rate functions, decreasing rate functions, and a scenario in which there is a short-lived increased (modal) in extinction rate (Figs. S3.5 to S3.7). Similarly, we constructed a set of alternative speciation rate functions, including linear increasing speciation rates, and an abrupt up-shift or down-shift (Figs. S3.5 to S3.7).

Drawing random diversification rates

Our second approach to exploring the congruence class was to sample rate functions from a random probability distribution. Specifically, we used the horseshoe Markov random field (HSMRF) (Carvalho et al., 2010; Magee et al., 2020) distribution to represent temporally autocorrelated rates. First, the simulation starts by drawing the global variance parameter $\gamma \sim \text{HalfCauchy}(0, 1)$. Then, for the case of simulating extinction rates, we draw $\mu_0 \sim \text{lognormal}(\hat{\mu}_0, 1.175)$, where the index 0 is at the present and $\hat{\mu}_0$ is the estimated extinction at the present for the reference model. In the case of simulating speciation rates, we explicitly set $\lambda_0 = \hat{\lambda}_0$ because the congruence class requires that all alternative models have the same speciation rate at the present.

For each preceding episode i we draw both a new local scaling factor $\sigma_i \sim \text{HalfCauchy}(0, 1)$ and then a new rate $\ln(\mu_i) \sim \text{Normal}(\ln(\mu_{i-1}), (\sigma_i \times \xi \times \gamma)^2)$, where ξ scales the overall magnitude of the changes in accordance with the number of episodes (here 100), such that we expect *a priori* $\ln(2)$ shift events of magnitude two or higher. This results in a distribution of $\ln(\mu_i)$ that is centered around the previous $\ln(\mu_{i-1})$, but with fat tails on each end. The most probable draw is little to no change, but large jumps are also possible.

3.7.4 Sampling congruent models from the posterior

We sampled 110 rate functions from the posterior distribution to represent the uncertainty in the parameter estimates. The samples were equidistant to minimize MCMC autocorrelation. Each of these diversification rate samples have different likelihoods, meaning they are not congruent. The individual posterior samples have more rapidly and sporadically changing rates, as opposed to the relatively smooth median summaries (Fig. S3.11).

For each of these 110 samples, we constructed the congruence class and sampled 10 alternative models using HSMRF-distributed extinction rate functions. We fixed the initial episode (μ_0) of the proposed extinction rates to being equal to the posterior sample initial extinction rate. This avoids extra artificial bias in directional trends near the present, since the speciation rate must always be equal at the present (λ_0). If we had chosen the same value for μ_0 for all posterior samples, then the directional trends in speciation rate in the near present ($\Delta\lambda$) would be completely determined by the first episode in the posterior samples.

For large proposed extinction rates, we used rejection sampling to ensure that the maximum of the proposed rate was not greater than the maximum of $\mu_0 + 0.1$ across the posterior samples (woodpeckers: 0.32, flycatchers: 0.62, parrots: 0.28). For some posterior-congruent models, we encountered numerical issues preventing us from computing the corresponding speciation rate, and we consequently discarded these samples.

We assessed the common trends in the speciation rate functions in the resulting models (Fig. 3.5). Since the number of models are now large, we sorted the model axis such that the "increasing" and "decreasing" bars are stacked. This means that one row no longer corresponds to a single model. However, we can still assess the fraction of significant trend directions for a particular time interval.

3.7.5 Assessing significant rate changes

We assessed common patterns in diversification rates by computing the trends in diversification rate changes. Specifically, we computed the slope of the speciation rate function, as $\Delta\lambda_i = (\lambda_i - \lambda_{i-k})/(t_i - t_{i-k})$, where the window size k is unity or some larger integer (see Fig. S3.12). If the slope is greater than some threshold ϵ, we interpret the rate function as increasing. If the slope is smaller than $-\epsilon$, we interpret the trend as decreasing. If the difference is between ϵ and $-\epsilon$, we interpret it as flat.

The selection of ϵ is to some extent arbitrary, and therefore we explored three values for $\epsilon \in \{0.01, 0.02, 0.05\}$. Choosing a too small ϵ will result in small noisy changes being detected as significant changes (Fig. S3.8). Conversely, a too large ϵ will result in no significant changes being detected at all (Fig. S3.10). In our analyses, we mainly used $\epsilon = 0.02$ rate units per Ma as the threshold.

3.7.6 Specifying and exploring hypothetical scenarios

We constructed a set of specific scenarios that we deemed either biologically realistic or interesting to complete the empirical exploration of the congruence class. Specifically, we assessed constant rates, a modal burst event (i.e., a temporary increased rate over a short period of a few million years), and linearly increasing and decreasing rates. We also considered an up-shift and a down-shift, as well as exponentially increasing and decreasing rates. For each scenario, we selected one reference speciation rate, and used in all cases a constant reference extinction

rate of $\mu = 0.28$. Further, we used all of the proposed rate functions as proposed alternative extinction rate functions to explore the congruence class. Each individual rate function can be found in the Supplementary materials, in Figs. S3.19 to S3.26.

3.7.7 Data and code availability

The data and code necessary for reproducing the analyses and the figures is available on github (`http://github.com/kopperud/cc_exploration`).

3.7.8 Acknowledgements

We thank Matt Pennell, Luke Harmon and one anonymous reviewer for comments that helped improve the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether-Program (Award HO 6201/1-1 to SH) and by the European Union (ERC, MacDrive, GA 101043187). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. AFM was partially supported by National Science Foundation grant DGE-1762114 and National Institutes of Health grant R01 AI153044.

Chapter 4

Phylogenetic estimation of branch-specific shifts in the tempo of origination

The word **algorithm** or algorism comes from the Persian mathematician al-Khwārzimī (c. 780-850). His full name is Abū ‘Abd Allāh Muḥammad ibn Mūsā al-Khwārizmī, meaning “Father of Abdullah, Mohammed, son of Moses, native of Khwārizm.” The Khwārizm region is located south of what was Lake Khwārizm (the Aral Sea).

(Knuth, 1997)

4.1 Abstract

Studying rates of species diversification is one of the key themes in macroevolution. In particular, we are interested in if some clades in a phylogeny diversify more rapidly/slowly than others due to branch-specific diversification rates. A common approach in neontological studies is to use a phylogenetic birth-death process to model species diversification. Specifically, the birth-death-shift process is used to model branch-specific shifts in the tempo of diversification. Here, we present `Pesto`, a new method and software for estimating branch-specific diversification rates that does not rely on Markov chain Monte Carlo simulations for fitting the model and instead deterministically computes the posterior mean branch-specific diversification rates using only two traversals of the tree. This method is blazingly fast: the birth-death-shift model can be fitted to large phylogenies ($>$ 20k taxa) in minutes on a personal computer while also providing branch-specific inference of diversification rate shift events. Thus, we can robustly infer branch-specific diversification rates and the number of diversification rate shift events for large-scale phylogenies as well as exploring the characteristic of the birth-death-shift model through complex and large-scale simulations. Here, we first describe the method and the software implementation `Pesto` and explore its behavior using trees simulated under the birth-death-shift model. Then, we explore the behavior of inferring significant branch-specific diversification rate shifts using both Bayes factors and effect sizes. We find few to none false positive inferences of diversification rate shift events but many false negatives (reduced power). The most difficult parameter to estimate is the rate at which diversification rate shifts occur (shift rate). Despite this, branch-specific diversification rate estimates are precise and nearly unbiased.

Key Words: diversification, speciation, extinction, phylogeny, birth-death, macroevolution, origination, clade-varying, branch-specific

4.2 Introduction

Estimating rates of diversification is one of the key interests in macroevolutionary biology and paleontology. Diversification rates are fundamental to our understanding of how historical biodiversity has changed. Recent developments in the field have focused on birth-death models that describe the tempo of lineage diversification (Ricklefs, 2007; Morlon, 2014). This process is highly variable; diversification rates may vary over time with episodes of mass extinctions (Raup and Sepkoski, 1982) or elevated speciation rates (Sepkoski, 1998) and vary among branches with clades in high or low turnover (Sanderson and Donoghue, 1994, 1996; Jetz et al., 2012). Moreover, we can speculate that diversification rate variation over time is associated with climatic or environmental factors (Hua and Wiens, 2013; Condamine et al., 2018; Palazzesi et al., 2022) whereas diversification rate variation among branches is associated with, amongst others, key innovations or changes in habitat (Hodges and Arnold, 1995; Allen and Gillooly, 2006; Weir and Schluter, 2007; Silvestro et al., 2014). Thus, we are ultimately interested in linking diversification rates to, for example, underlying ecological and environmental factors (Condamine et al., 2018). However, as a first step, we should infer diversification rates independently to learn if there was any diversification rate variation for the specific study group (Budd and Mann, 2025). Here, we will not consider branch-specific shifts in diversification rates in relation to the evolution of novel characters or changes in ecosystems, however we are nonetheless interested in identifying the branch-specific diversification rate shifts themselves. In particular, we are interested in where on the phylogeny the diversification rates shifted and the size of the shifts, in order to test hypotheses about the generation of biodiversity.

In recent years, several statistical approaches were developed to estimate branch-specific diversification rates from phylogenies (`BAMM`, Rabosky 2014; `CLaDS2`, Maliet et al. 2019; `LSBDS`, Höhna et al. 2019; `MTBD`, Barido-Sottani et al. 2020; `MiSSE`, Vasconcelos et al. 2022; `RPANDA` Mazet et al. 2023; see Martínez-Gómez et al. 2024 for a review). One key limitation of existing methods is that they are prohibitively slow. Existing methods rely on simulation techniques (i.e., Markov-chain Monte Carlo simulations together with data augmentation or stochastic mapping) with long run times (i.e., often days to weeks). To complicate matters, these highly complex simulation algorithms are slow to converge and can often fail (Martínez-Gómez et al., 2024). Thus, current methods are (i) impractical for phylogenies that contain many tips (i.e., $> 10,000$) which should be of main interest to study diversification rate variation (Rabosky et al., 2018), and (ii) too computationally demanding to explore model assumptions and behavior via large-scale simulation studies. We solve the computational bottleneck by developing a novel approach that deterministically computes branch-specific diversification rates on a given phylogeny. Thus, our approach does not suffer from convergence issues. Our approach uses the same underlying birth-death-shift process as developed by Barido-Sottani et al. (2020) and Höhna et al. (2019) and inherits its biological assumptions

and statistical properties. Therefore, our new approach and the implementation in `RevBayes` (Höhna et al., 2016a, 2019) give identical estimates of the posterior mean branch-specific diversification rates, if the same parameters (i.e., diversification rate categories and shift rate) and model assumptions are chosen. However, our new approach is orders of magnitude faster.

An outstanding question is whether we can reliably infer the number and location of diversification rate shifts, or if we can only reliably estimate branch-specific diversification rates (Shi and Rabosky, 2015; Moore et al., 2016; Mitchell and Rabosky, 2017; Höhna et al., 2019). A change in diversification rates could either be explained by several small or a few large shifts. Therefore, the number of estimated diversification rate shifts depends strongly on the prior assumption on the shift rate (Moore et al., 2016; Höhna et al., 2019). Nevertheless, hypothesis testing using Bayes factors can be used to assess the support for alternative hypotheses despite sensitivity to the prior (Kass and Raftery, 1995). In this study we develop an efficient approach to infer the number and location of diversification rate shifts on phylogenies using Bayes factors (see also Shi and Rabosky, 2015).

Our approach for inferring branch-specific diversification rates and branch-specific rate shift events is inspired by Bayesian inference of ancestral discrete characters (Yang et al., 1995; Nielsen, 2002). However, we discovered that the probabilistic description of the birth-death-shift model is incompatible with established theory on Bayesian belief networks (Pearl, 1988). Specifically, we are not able to represent branch-specific diversification rates as random variables that are separable from the phylogeny (see Fig. S4.1). Thus, if we assume the phylogeny as "observed" data, then the branch-specific diversification rates are "unobserved" data instead of parameters of the model. This is an issue, as the prior probability and the likelihood of branch-specific diversification rates are conflated (see also May and Rothfels, 2023). We suspect that the inseparability of prior and likelihood is also relevant for previous approaches that used Bayesian inference to fit identical (Höhna et al., 2019; Barido-Sottani et al., 2020) or very similar stochastic processes (Rabosky, 2014; Maliet et al., 2019; Quintero et al., 2024). Nevertheless, we conceptualize our approach using ideas and terminology from Bayesian inference theory (for more discussion see Appendix S4.1).

We provide a `Julia` implementation of our new algorithm called *Phylogenetic Estimation of Shifts in the Tempo of Origination (`Pesto`)*. The algorithm requires only two tree traversals: one postorder (from the tips to the root) and one preorder (from the root to the tips) tree traversal. The first pass over the tree is used to compute the likelihood scores and the second pass is used to compute and map the posterior mean branch-specific diversification rates. We show the time complexity of the algorithm for various tree sizes, demonstrating that `Pesto` infers robust branch-specific diversification rates and diversification rate shifts for phylogenies with $> 10,000$ taxa in a few minutes on a standard personal computer. We validate `Pesto` by showing that branch-specific diversification rate estimates are identical to estimates obtained using `LSBDS` (Höhna et al., 2019), up to some Monte Carlo error, when the identical model is

used. Furthermore, we explore the behavior of `Pesto` using a simulation study and show that parameter estimation is robust given sufficiently large phylogenies. For example, estimates of the shift rate are less reliable when there are few shifts, i.e., if shifts are rare and/or if the phylogeny is small. Branch-specific diversification rates are robustly inferred even for small trees (minimum 100 taxa). Finally, we present a simulation study to explore the ability to estimate the number and location of diversification rate shifts. `Pesto` is conservative in the inference of diversification rate shift events with virtually no false positives, but has low power (i.e., more false negative shift event inferences).

4.3 Materials and methods

4.3.1 The Underlying Theory of `Pesto`

Our new method *Phylogenetic Estimation of Shifts in the Tempo of Origination (`Pesto`)* relies on and extends existing methods for estimating branch-specific diversification rates. The underlying stochastic process in `Pesto` is equivalent to the *birth-death-shift* process (Höhna et al., 2019). The existing implementations of the birth-death-shift process and similar processes use computationally expensive techniques such as Markov chain Monte Carlo simulation with data augmentation (Rabosky, 2014; Maliet et al., 2019; Barido-Sottani et al., 2020) or stochastic mapping (Höhna et al., 2019) to infer branch-specific diversification rates. The fundamental novelty of `Pesto` is our deterministic algorithm for calculating the posterior mean branch-specific diversification rates and the number of diversification rate shifts, conditional on the specific choice of diversification rate categories and shift rate. Unlike with simulation approaches, the calculations are performed only once and there is no randomness or inherent uncertainty in the inference procedure. Hence, we can run analyses in `Pesto` using trees that have tens of thousands of tips in a matter of minutes on a personal computer. We describe the details of the method in the following sections.

The birth-death-shift model.— `Pesto` uses the same underlying birth-death-shift process as described by Höhna et al. (2019), which we recapitulate here. The birth-death-shift process has a speciation rate λ, an extinction rate μ, as well as a shift rate η. Suppose one begins a tree simulation with a starting (root) node that represents the common ancestor of the clade, and that the root node has two descendant branches of unknown lengths. After a waiting time modeled as an exponentially distributed random variable (with rate $\lambda+\mu+\eta$), there are three possible events. First, there can be a speciation event, with probability $\lambda/(\lambda+\mu+\eta)$, where two new branches are created. Second, there can be an extinction event, with probability $\mu/(\lambda+\mu+\eta)$, where the branch terminates. Third, there can be a rate shift event, with probability $\eta/(\lambda+\mu+\eta)$. At a rate-shift event, a new set of diversification rates (i.e. the pair λ', μ') is chosen randomly, which we will explain in the following section. Note that we as-

sume that the new diversification rates are independent of the old diversification rates (unlike Maliet et al., 2019; Quintero et al., 2024). After some time t has passed, the tree has either gone entirely extinct or it has survived. If the process survived, then each species is sampled with probability ρ. Finally, extinction events and unsampled species are pruned, resulting in a so-called "reconstructed tree", meaning that the tree appears as if it was reconstructed from molecular or morphological data at the present.

Base distribution of diversification rates.— When a diversification rate shift event occurs in the birth-death-shift process, new diversification rates are drawn randomly from a base distribution. For example, `BAMM` uses an exponential base distribution (Rabosky, 2014) and Höhna et al. (2019) proposed to use instead a log-normal distribution. However, any non-negative base distribution could in principle be used. Thus, in addition to the exponential and the log-normal base distributions, we also explored the use of a log-uniform and a gamma base distribution. We parameterized the base distributions in terms of their means and variances where applicable. Specifically, we specified the exponential distribution with a mean of $\hat{\lambda}$, the log-normal distribution with a median of $\hat{\lambda}$ and standard deviation of $H \approx 0.587$ (meaning that the 2.5%–97.5% quantile interval of the base distribution spans one order of magnitude), and the log-uniform with a mean of $\hat{\lambda}$ and a span of one order of magnitude. For the gamma distribution, we set the scale and shape parameters such that the mean was $\hat{\lambda}$, and the variance was equal to that of the log-normal distribution. This results in four base distributions whose location is controlled by the $\hat{\lambda}$-parameter, and the variance is either fixed or also determined by the $\hat{\lambda}$ parameter. We will discuss specific choices of $\hat{\lambda}$ and the variance later in section *Estimating the parameters of the birth-death-shift model*.

Discretization of diversification rates.—

Simulating a phylogeny under the birth-death-shift process with continuous base distributions is relatively straight forward. Moreover, modeling the base distributions as being fully continuous might be the preferable approach for inference. However, it is not trivial to calculate the probability of observing the phylogeny when using continuous base distributions. For the probability calculation to be computationally tractable, we follow the birth-death-shift implementation of the `LSBDS` model (Höhna et al., 2019) where an approximation approach is used. Specifically, the continuous base distributions are discretized by splitting them into n bins with equal probability mass (i.e., the bins represent quantiles). This results in a vector of diversification rate *classes*, i.e., $\vec{\lambda}$ for speciation rates, and $\vec{\mu}$ for extinction rates. Next, we select all pairwise combinations of $\vec{\lambda}$ and $\vec{\mu}$ to form the vectors $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_K]$ and $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_K]$, which represent the $n^2 = K$ diversification rate *categories* (see Fig. 4.1). Choosing a higher number of rate classes will result in a better approximation of the base distributions (Höhna et al., 2019). In the limit of infinitely many rate classes, the birth-death-shift process with discrete and continuous base distributions for the diversification rates are equivalent. Therefore, we expect that inferences made using the discrete base distribu-

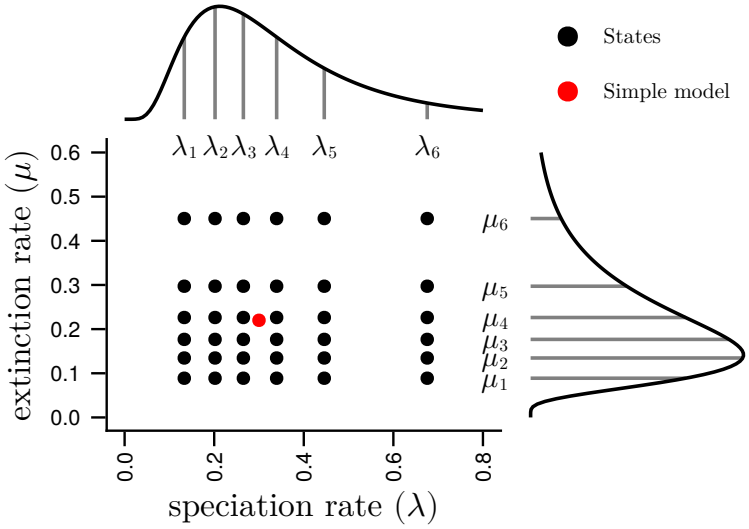


Figure 4.1: Schematic of diversification rate categories. First, `Pesto` estimates the speciation rate ($\hat{\lambda} \approx 0.30$) and the extinction rate ($\hat{\mu} \approx 0.22$) under the constant-rate birth-death process (red dot). These constant-rate estimates are used as the mean of base distributions. Second, we select the number of quantiles ($n = 6$) and the standard deviation $H = 0.587$ chosen such that the 2.5%–97.5% quantile interval spans one order of magnitude. Third, we consider all pairwise combinations of the quantiles (black dots, $K = n^2 = 36$). The log-normal distribution with $n = 6$ quantiles and a variance of H is the default option, however other distributions and other number of quantiles and variances can be specified. The rate category values can also be specified manually. Each black dot is a rate category in the birth-death-shift model, and together they represent the range of rate heterogeneity allowed under the model. Along each branch and across time, we calculate the probability that the process was in each rate category using Eq. 7.

tion should converge and stabilize as more bins are used, however at the cost of increased computational time.

4.3.2 Computing the likelihood of the phylogeny

In this section we will explain how we compute the probability density of observing the phylogeny, as well as how we set up the likelihood function. While there are no known analytical solutions for the probabilities and probability densities, it is possible to find solutions using numerical methods. The idea is to express the change in probability (density) by considering which events (i.e., speciation, extinction or rate shift events) are possible in a small time interval Δt, and taking the limit as $\Delta t \to 0$, in other words using differential equations (see Maddison et al., 2007; FitzJohn, 2012, for more details on how the differential equations are derived). As the data, we assume a rooted, bifurcating phylogenetic tree with branch lengths in units of time. We also assume that the tree is known without error. Note that the likelihood function (Eq. 4) is analogous to the likelihood function of a state-dependent birth-death process without observed states (Maddison et al., 2007; FitzJohn, 2012; Beaulieu and O'Meara, 2016; Freyman and Höhna, 2018) and we repeat it here as we require the likelihood function and the probability density computations to derive our algorithm to deterministically compute the posterior mean branch-specific diversification rates.

Conceptually, our algorithm contains the following steps. First, we calculate the extinction probabilities ($\mathbf{E}(t)$), which are invariant to the topology and divergence times. Second, we calculate the probability density of observing the tree ($\mathbf{D}(t)$). Since the probability density depends on the topology and divergence times, it needs to be computed for each branch in a postorder tree traversal. Third, we use the probability density of observing the tree to set up the likelihood function. Note that we write scalars in plain typeface and vectors and matrices in bold. We treat vectors as matrices with a single column and transposed vectors as matrices with a single row. See table S1 for an overview of the notation and symbols that are used.

Computing extinction probabilities.— In order to calculate the extinction probabilities, we use the following system of differential equations (Maddison et al., 2007):

$$\frac{d\mathbf{E}(t)}{dt} = \boldsymbol{\mu} - (\boldsymbol{\lambda} + \boldsymbol{\mu}) \odot \mathbf{E}(t) + \boldsymbol{\lambda} \odot \mathbf{E}(t) \odot \mathbf{E}(t) + \mathbf{Q}\mathbf{E}(t). \tag{1}$$

More precisely, $E_j(t)$ represents the probability that a lineage that was in rate category j at time t went extinct before the present or was not sampled in the phylogeny. Note that we use the symbol $\odot$ to represent element-wise multiplication. The matrix $\mathbf{Q}$ has entries $-\eta$ on the diagonal, and entries $\eta/(K-1)$ elsewhere. For example in the case of a three-category model, $\mathbf{Q}$ would be

$$\mathbf{Q} = \begin{bmatrix} -\eta & \eta/2 & \eta/2 \\ \eta/2 & -\eta & \eta/2 \\ \eta/2 & \eta/2 & -\eta \end{bmatrix}. \tag{2}$$

We allow for the possibility that the taxa are incompletely sampled, but we assume an equal probability of taxon sampling (FitzJohn, 2010; Höhna et al., 2011). Therefore, the initial values of the extinction probabilities are all set to $E_j(t = 0) = 1 - \rho$, where ρ is the known/fixed taxon sampling fraction.

Computing the probability density of the observed branches.— The probability of observing the branch (or subtree) at time t is likewise represented as a set of differential equations (Maddison et al., 2007; FitzJohn, 2012):

$$\frac{d\mathbf{D}_M(t)}{dt} = (-\boldsymbol{\lambda} - \boldsymbol{\mu} + 2\boldsymbol{\lambda} \odot \mathbf{E}(t)) \odot \mathbf{D}_M(t) + \mathbf{Q}\mathbf{D}_M(t), \tag{3}$$

where M is the branch index. Specifically, if M is a terminal branch, then $D_{M,j}(t)$ is the probability of observing the branch if the process was in category j at time t. If M is an internal branch, then $D_{M,j}(t)$ represents the probability density of observing the subtree descended from branch M. For initial values, we assign $D_{M,j}(t = 0) = \rho$ for all categories j if M is a terminal branch. For internal branches, the initial (youngest) value of a branch M is assigned to $\mathbf{D}_M(t) \coloneqq \mathbf{D}_A(t) \odot \mathbf{D}_B(t) \odot \boldsymbol{\lambda}$, where the branches A and B are the descendants of branch M. We iterate and solve $\mathbf{D}_M(t)$ for all branches M in a postorder traversal of the tree (Fig. 4.2**a**–**c**) until the root node ($\mathbf{D}_R(t)$) is reached.

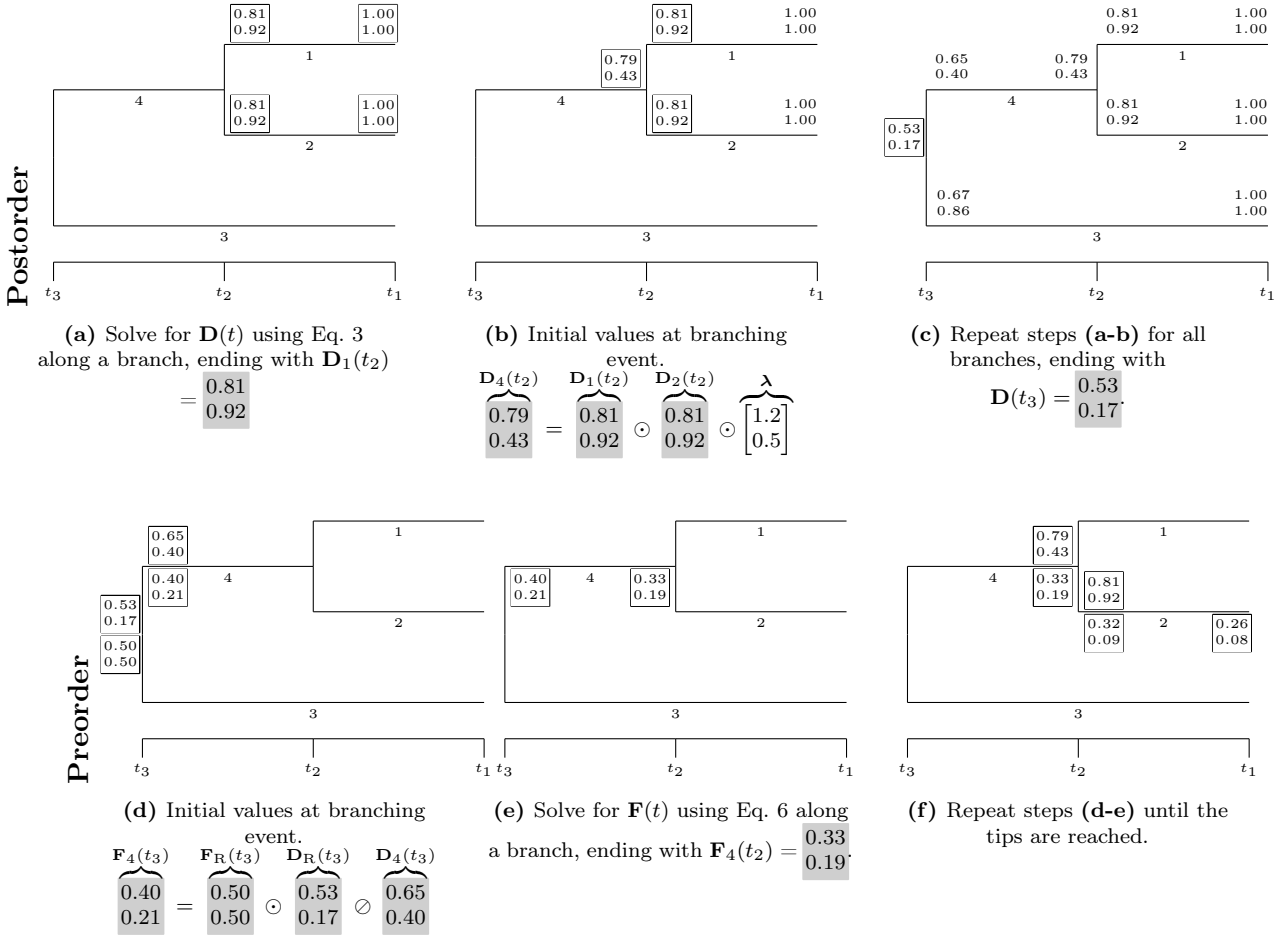


Figure 4.2: The algorithm for the two tree traversals on a miniature tree. The postorder pass **(a–c)** computes the probabilities given the descendants ($\mathbf{D}(t)$, written above the branches). The preorder pass **(d–f)** computes the probabilities given the tree up to but not including the descendants ($\mathbf{F}(t)$, written below the branches). The white boxes on the trees correspond to the white boxes in the caption. We set $\mathbf{D}(t_1) = [1.0, 1.0]$ at the tips, and $\mathbf{F}_R(t_3) = [0.5, 0.5]$ at the root. The densities $\mathbf{D}(t)$ and $\mathbf{F}(t)$ are used to calculate the probability that the process was in a particular rate category at a particular time, see Eq. 7. In this example, we used divergence times $t_1 = 0$, $t_2 = 0.1$, $t_3 = 0.2$, speciation rates $\boldsymbol{\lambda} = [1.2, 0.5]$, extinction rates $\boldsymbol{\mu} = [1.0, 0.3]$, shift rate $\eta = 0.01$, extant sampling fraction $\rho = 1$, we did not condition on survival, and we did not re-scale the initial values.

Computing the probability of observing the tree (likelihood).— The likelihood function for the birth-death-shift process is defined as the weighted average of the probabilities at the root node (Eq. 3) over all diversification rate categories. When calculating the likelihood of observing the tree, we condition on (i) that the two branches subtending from the root survived until the present, and (ii) that there was a speciation event at the root:

$$L(\eta, \hat{\lambda}, \hat{\mu}, \rho|\Psi) = \frac{1}{K}\sum_{j=1}^{K}\frac{D_{R,j}(t_R)}{\lambda_j(1 - E_j(t_R))^2}, \tag{4}$$

where t_R is the age of the root, Ψ represents the phylogeny (i.e., topology and branch lengths),

and we follow Pagel (1994) in using a flat prior probability ($1/K$) of being in any of the diversification rate categories at the root (see also Goldberg and Igić, 2008; FitzJohn et al., 2009; FitzJohn, 2012, for alternative root treatments).

Estimating the parameters of the birth-death-shift model.— Höhna et al. (2019) argued for estimating the shift rate η jointly with the parameters of the base distributions ($\hat{\lambda}$ and $\hat{\mu}$) using Bayesian inference while Rabosky et al. (2014b) advocates using an empirical Bayes approach to estimate the prior means of the base distributions under a constant rate birth-death process and the shift rate η is replaced with a geometric distribution directly on the number of shift events (Mitchell and Rabosky, 2017). We discovered, however, that estimating the base distribution parameters and the shift rate jointly is more difficult than previously thought. The joint likelihood surface includes many local optima and a hill-climber algorithm can easily get stuck in these local optima (Fig. S4.6). As a compromise, we settled for a two-step approach (Fig. 4.1). This involves first finding the maximum-likelihood estimates for $\hat{\lambda}, \hat{\mu}$ using the simpler lineage-homogeneous constant-rate birth-death model (as done in Rabosky et al., 2014b). Second, we infer η by maximizing the likelihood of the birth-death-shift model while keeping the other parameters fixed

$$L(\eta|\Psi) = L(\eta, \hat{\lambda}, \hat{\mu}, \rho|\Psi). \tag{5}$$

This is also called an estimated likelihood function (Pawitan, 2001, p. 274), with which we did not encounter any local optima issues. We note, however, that the estimated base distribution parameters and shift rate using the two-step approach are not necessarily the same as the true maximum likelihood estimates, and the optimal approach for selecting or estimating the diversification rate base distribution and shift rate requires further research.

4.3.3 The posterior distribution of rate shift histories

It is not possible to directly observe the true rate shift history on a phylogeny except in the case of simulated phylogenies. When we wish to make inferences about the birth-death-shift process that led to an “observed" phylogeny, we consider the diversification rate shift history to be an unobserved random variable. As we can not observe it directly, we calculate the likelihood by considering all possible diversification rate shift histories that could have led to the phylogeny in question (Eq. 4). In other words, when calculating the likelihood, we integrate across all possible realizations of the diversification rate shift history. By doing so, however, we are omitting several aspects of interest. For example, we are interested in the branch-specific diversification rates and how many diversification rate shift events occurred, across branches and across time. Therefore, we wish to investigate the posterior distribution of diversification rate shift histories, i.e., the probability distribution of the diversification rate shift histories conditioned on the data. Höhna et al. (2019) showed how one could approximate

the posterior distribution of the diversification rate shift histories using stochastic mappings similar to stochastic character mapping for discrete character evolution (Huelsenbeck et al., 2003). By simulating many samples of the posterior distribution of diversification rate shift histories, one can compute summaries such as the posterior mean branch-specific diversification rate, the posterior mean number of diversification rate shift events, or the posterior probability that there was at least one diversification rate shift. In order for this sampling approach to yield a good approximation of the posterior distribution, one must i) simulate with small step sizes in time, and ii) repeat the stochastic mapping thousands or millions of times. Accordingly, one of the main disadvantages of the stochastic mapping approach is its slow run time and computational burden, although it is more efficient than data augmentation (Höhna et al., 2019).

In the next sections, we will outline a deterministic algorithm that gives equivalent summaries of the posterior distribution. Instead of using stochastic mappings, this algorithm is entirely non-random (i.e., it is deterministic), and thus will yield the same results every time the algorithm is executed. Notably, we will compute i) the posterior probability that the process was in a particular diversification rate category, ii) the posterior mean branch-specific diversification rates, iii) the posterior mean number of branch-specific rate shift events, iv) the posterior probability that there was at least one rate shift event at a given branch, and v) the Bayes factor for the support of the hypothesis that there was at least one rate shift event at a given branch. These calculations are facilitated by simple formulae or differential equations that need to be solved only once per branch. Compared with stochastic mapping, our approach is much more computationally efficient and consequently the inferences can be made in a fraction of the time.

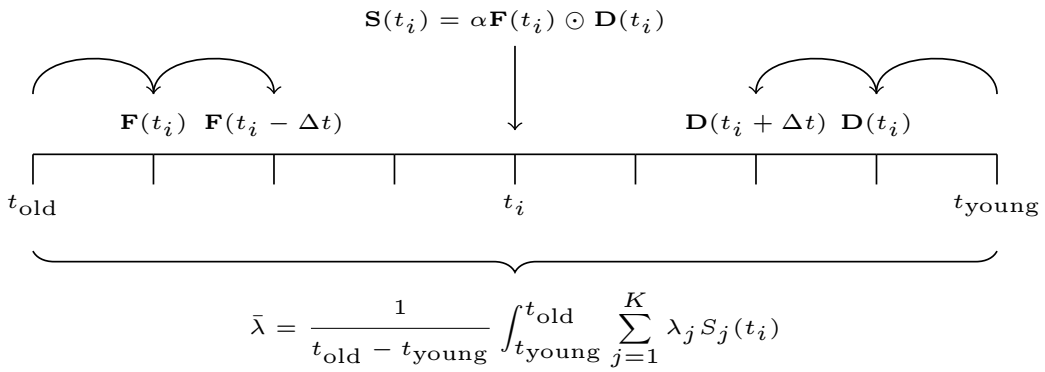


Figure 4.3: A simplified schematic of the backward-forward algorithm along a branch. First, we find the solution for $\mathbf{D}(t)$, by starting from t_{young} and going towards the past. Second, we find the solution for $\mathbf{F}(t)$, by starting from t_{old} and going towards the present. Third, we calculate the ancestral rate category probabilities $\mathbf{S}(t)$, which is $\mathbf{F}(t)$ element-wise times $\mathbf{D}(t)$, normalized by α such that $\mathbf{S}(t)$ sums to one for any particular time t. Fourth, we compute the posterior mean speciation rate, averaged across the time interval of the branch ($\bar{\lambda}$). Note that the fixed time intervals Δt are plotted for illustrative purposes. In the actual implementation, the number of time intervals, and the size of the time intervals can vary.

Preorder traversal to calculate posterior rate category probabilities.— We use a dynamic programming approach to compute the posterior probabilities of the rate categories (Fig. 4.3), specifically the backward-forward algorithm (Rabiner, 1989) extended to accommodate a tree structure (Pearl, 1982). The motivation for applying the backward-forward algorithm is based on the fact that the birth-death-shift process can along a lineage be seen as a hidden Markov model where the diversification rate categories are the hidden states. The backward pass of the algorithm was explained previously, where we calculated the probabilities $\mathbf{D}_M(t)$. In the forward pass, we calculate the forward probabilities $\mathbf{F}_M(t)$. Specifically, $F_{M,j}(t)$ represents the probability that the rate category was j on branch M at time t, given the part of the phylogeny that is not descended from branch M at time t. At the root node R, there is no ancestral node, meaning that $F_{R,j}(t_R)$ represents the prior probability of the rate category being j. We assume that the ancestral diversification rate is a random variable that is distributed according to the (discretized) base distribution. Therefore, we initialize $F_{R,j}(t)$ to $1/K$ at the root node (recall that each diversification rate category represents quantile intervals with equal probability mass). When initializing $F_{R,j}(t)$, we also condition on i) survival and ii) that there was a speciation event at the root, in the same manner as in the likelihood function (Eq. 4). Since the rate category variable is the only random variable ($\hat{\lambda}, \hat{\mu}, \eta, \rho$ are fixed to their estimates), there are no other priors (see Fig. S4.1). In order to solve for $\mathbf{F}_M(t)$ along the branches, we use the following set of differential equations

$$\frac{d\mathbf{F}_M(t)}{dt} = (\boldsymbol{\lambda} + \boldsymbol{\mu} - 2\boldsymbol{\lambda} \odot \mathbf{E}(t)) \odot \mathbf{F}_M(t) - \mathbf{Q}\mathbf{F}_M(t). \tag{6}$$

Note that $d\mathbf{F}/dt$ is equal to $d\mathbf{D}/dt$, except the sign has changed on both the left and the right sides. The reason for the change in sign is that we are solving $d\mathbf{F}/dt$ in forward time (from ancient to recent), and therefore the time differential dt is negative.

Once we have solved $\mathbf{F}_M(t)$, we compute posterior probabilities of the rate categories (see the supplementary material for a discussion on the posterior). They are given as follows

$$\mathbf{S}_M(t) = \alpha \mathbf{F}_M(t) \odot \mathbf{D}_M(t), \tag{7}$$

where α is a normalizing factor such that $\mathbf{S}_M(t)$ sums to one (Fig. 4.3, see Pearl 1982, 1988). In summary, the preorder traversal consists of two steps. First, we initialize $\mathbf{F}_M(t) \coloneqq \mathbf{S}_P(t) \oslash \mathbf{D}_M(t)$, where P is the parental branch, t is the divergence time, and $\oslash$ represents element-wise division. Second, we solve $\mathbf{F}(t)$ along branch M. We repeat these two steps for all descendant branches, in a preorder traversal of the tree (Fig. 4.2**d**–**f**). When the preorder traversal is completed, the backward-forward algorithm is also completed.

Estimating the posterior mean branch-specific diversification rates.— Estimating branch-specific diversification rates is a key result for any method that fits the birth-death-shift process to a phylogeny. We consider the diversification rate shift history and thus also

the branch-specific diversification rates as random variables. For a given branch M and a time t, the moments of the speciation rate $\lambda_M(t)$ are weighted averages of the speciation rate categories λ_j, with the weights given by their posterior probabilities, which we can use to compute the posterior mean and variance

$$\begin{aligned}\mathbb{E}[\lambda_M(t)] &= \sum_j \lambda_j S_{M,j}(t) \\ \mathrm{Var}[\lambda_M(t)] &= \sum_j \lambda_j^2 S_{M,j}(t) - \left(\sum_j \lambda_j S_{M,j}(t)\right)^2 .\end{aligned} \tag{8}$$

This can for example be used to compute tip rates, where the posterior mean speciation rates are computed at the present. To summarize over a branch as a whole, we use the symbol $\bar{\lambda}_M$, and get the expectation by integrating over the time span and dividing by the length of the branch:

$$\mathbb{E}(\bar{\lambda}_M) = \frac{1}{t_{\text{old}} - t_{\text{young}}} \int_{t_{\text{young}}}^{t_{\text{old}}} \sum_j \lambda_j S_{M,j}(t) dt \tag{9}$$

meaning that $\mathbb{E}(\bar{\lambda}_M)$ is the posterior mean speciation rate for branch M, and $t_{\text{young}}, t_{\text{old}}$ represent the youngest and oldest times of branch M (Fig. 4.3). If one is interested in the variance of $\bar{\lambda}_M$, it is possible to approximate the integral by splitting the branch into c discrete bins represented by $\lambda(t_1), \lambda(t_2), \ldots$ etc. Then, we can apply the standard formula for the variance of a linear combination of correlated random variables

$$\mathrm{Var}[\bar{\lambda}_M] = \frac{1}{c^2} \sum_{a=1}^{c} \mathrm{Var}[\lambda_M(t_a)] + \frac{1}{c^2} \sum_{a=1}^{c} \sum_{b>a}^{c} \mathrm{Cov}[\lambda_M(t_a), \lambda_M(t_b)]. \tag{10}$$

Since the time bins are not independent, it is necessary to calculate the covariance, which one can calculate as

$$\begin{aligned}\mathrm{Cov}[\lambda_M(t_a), \lambda_M(t_b)] = &\sum_{j=1}^{K} \sum_{i=1}^{K} S_{M,j}(t_a) \times P_{M,ij}(t_a, t_b) \\ &\times (\lambda_j - \mathbb{E}[\lambda_M(t_a)]) \times (\lambda_i - \mathbb{E}[\lambda_M(t_b)]),\end{aligned} \tag{11}$$

where $S_{M,j}(t_a)$ is the marginal probability of the diversification rate category being j at time t_a, and $P_{M,ij}(t_a, t_b)$ is the transition probability from category j at time t_a to category i at a younger time t_b, normalized such that $P_{M,ij}(t_a, t_b)$ sums to one over the departure categories j. We note that computing the variance of the branch-specific speciation rates is significantly more computationally expensive than computing the mean. For more details on how to calculate the posterior variance of the speciation rate, see the supplementary material.

If one is interested in the posterior mean and variance of the branch-specific extinction rate, net-diversification rate, or relative extinction rate, one can substitute λ_j for μ_j, $\lambda_j - \mu_j$,

or μ_j/λ_j, respectively, and use the same formulae. In order to evaluate the integrals, we use Gauss-Legendre quadrature with $n = 10$ points. We want to emphasize that previous methods (Freyman and Höhna, 2019; Höhna et al., 2019) used Eq. 7 to simulate the branch-specific diversification rates (i.e., stochastic mappings), whereas `Pesto` uses Eq. 9 to compute the posterior mean deterministically. To present the results, we plot the posterior mean branch-specific diversification rates on the tree, however standard visualizations like scatter plots or histograms can also be used.

Estimating the number of rate shifts.— Another key property of the birth-death-shift process that we are interested in is the number of diversification rate shift events that occurred across a phylogeny. As for the branch-specific diversification rates, the number of diversification rate shifts for a particular branch is also a random variable, which we will summarize by computing the posterior mean. The idea is to compute the average number of diversification rate shifts in a small time interval Δt, by computing the probability of a rate shift from rate category j to a different rate category i, and multiplying by the posterior probability that the process was in rate category j at the time. By summing the number of diversification shifts over several small time intervals $\Delta t_1, \Delta t_2, \ldots$ across the branch, we can calculate the total number of rate shifts across the time span of the branch. In the limit of $\Delta t \to 0$, the number of rate shifts converge to zero, however the change in the average number of rate shifts per time can still be expressed as a differential equation:

$$\frac{d\hat{N}_{M,ij}(t)}{dt} = \begin{cases} -S_{M,j}(t)\dfrac{D_{M,i}(t)}{D_{M,j}(t)}\dfrac{\eta}{K-1} & \text{if } j \neq i \\ 0 & \text{if } j = i, \end{cases} \tag{12}$$

where $\hat{N}_{M,ij}(t)$ is the posterior mean number of accumulated rate shifts from a diversification rate category j to another diversification rate category i from the beginning of branch M until a younger time t. The initial value is $\hat{N}_{M,ij}(t_{\text{old}}) = 0$, and the time steps dt are negative. For the posterior mean number of diversification rate shifts to be of any note (i.e., non-zero), two conditions must be met. First, the posterior probability that the process was in the departure category j at time t must be non-zero. Second, the probability density (of observing the tree, given the process was in the rate category) of the arrival category i must be equal or larger than that of the departure category j (i.e., $D_{M,i}(t) \geq D_{M,j}(t)$). For details on how the expression is derived, see the supplementary material. Since we can compute $\hat{N}_{M,ij}$ for all diversification rate shifts from any diversification rate category j to any other category i, we

can build a rate shift matrix:

$$\hat{\mathbf{N}}_M = \begin{bmatrix} 0 & \hat{N}_{M,12} & \dots & \hat{N}_{M,1K} \\ \hat{N}_{M,21} & 0 & & \\ \vdots & & \ddots & \vdots \\ \hat{N}_{M,K1} & & \dots & 0 \end{bmatrix}, \tag{13}$$

and group the elements $\hat{N}_{M,ij}$ depending on which specific question we are interested in. For example, one can group $\hat{N}_{M,ij}$ by whether it was the speciation rate or extinction rate that shifted. Another use case is to estimate the number of diversification rate shifts of a minimum shift size (e.g., $\lambda_i - \lambda_j > 0.2$). In the remainder of the manuscript, we will only focus on the total number of shifts, $\hat{N}_M = \sum_i \sum_j \hat{N}_{M,ij}$, regardless of whether it was a shift in the speciation or extinction rate or both. If only the total number of diversification rate shifts is of interest, it is possible to simplify Eq. 12 to the following

$$\frac{d\hat{N}_M(t)}{dt} = \begin{cases} \frac{-\eta}{K-1}\left[\sum_i D_i(t) \sum_j \frac{S_j(t)}{D_j(t)} - 1\right] & \text{if } j \neq i \\ 0 & \text{if } j = i, \end{cases} \tag{14}$$

yielding the same result. We emphasize that $\hat{N}$ is not the probability that a diversification rate shift occurred. In a realization of the birth-death-shift process, the number of diversification rate shift events are integers, for example 0, 1 or 2. The posterior mean number of diversification rate shifts $\hat{N}$ is a measure of the average or expected number of diversification rate shifts, which can take continuous values, for example 0.01 or 0.95. While this may seem counter-intuitive, it is not unlike how a six-sided dice roll has an expectation of 3.5.

Identification of branches with strong support for a diversification rate shift.— In the previous section we derived an approach to estimate the posterior mean branch-specific number of diversification rate shifts. Next, we focus on estimating whether a branch experienced at least one significant diversification rate shift. More specifically, we calculate the posterior probability of no diversification rate shifts in a small time interval Δt. Over a longer time interval such as a branch, the joint probability of no diversification rate shifts can then be calculated as the product of posterior probabilities over successive small time intervals $\Delta t_1, \Delta t_2$, etc. For computational convenience we replace the product with a sum by using log-transformed probabilities and we express the change in the sum as a differential equation. This allows us to use standard differential equation solvers (Tsitouras, 2011), which give more precise solutions. Specifically, we calculate the posterior probability of no diversification rate shifts originating from rate category j on branch M, conditional on the process beginning in rate category j, represented by the quantity $X_{M,j}(t)$. The following differential equation

describes how $\log X_{M,j}(t)$ changes over time

$$\frac{d \log X_{M,j}(t)}{dt} = \frac{\eta}{K-1} \frac{\sum_{i \neq j} D_{M,i}(t)}{D_{M,j}(t)}, \tag{15}$$

where the initial value is $\log X_j(t_{\text{old}}) = 0$, and $dt < 0$. We provide details of the derivation in the supplementary material. Next, we take the weighted average across the diversification rate categories j, with the weights being the posterior probabilities for the ancestral rate categories (at the beginning of the branch), and take the complement. This gives us the posterior probability that there was at least one rate shift

$$P_M(\geq 1 \text{ shifts}) = 1 - \sum_j S_{M,j}(t_{\text{old}}) e^{\log X(t_{\text{young}})}. \tag{16}$$

If we want a measure of support for the hypothesis of at least one diversification rate shift on a branch, we can calculate the Bayes factor (Shi and Rabosky, 2015)

$$\text{Bayes factor} = \frac{\frac{P_M(\geq 1 \text{ shifts})}{\pi_M(\geq 1 \text{ shifts})}}{\frac{P_M(0 \text{ shifts})}{\pi_M(0 \text{ shifts})}}, \tag{17}$$

where we use $\pi_M(0 \text{ shifts}) = e^{-(t_{\text{old}} - t_{\text{young}})\eta}$ as an approximation for the prior probability of 0 diversification rate shifts —i.e., the probability of 0 events under a Poisson distribution with rate $(t_{\text{old}} - t_{\text{young}})\eta$.

In our exploration of how the model fitted to several test datasets, we noticed that very short branches, which have an extremely low prior probability of a diversification rate shift, can produce strongly supported Bayes factors even though the posterior mean number of diversification rates shifts is very low (e.g., $\hat{N} < 0.05$). Therefore, when considering whether a branch showed strong support for a diversification rate shift event, we employed two criteria. First, we required the rate shift event to have a Bayes factor of more than 10 (Jeffreys, 1961; Kass and Raftery, 1995). Second, we required the branch to deviate substantially from an estimate of 0 rate shifts, thus we picked $\hat{N} > 0.5$ as a conservative criterion.

4.3.4 Birth-death-shift simulations

To assess the performance of `Pesto`, we designed a procedure to simulate trees under the birth-death-shift process. We use a forward-simulation approach, in other words we begin at the root and move towards the tips. This approach generates "complete trees", i.e., trees that also include extinct tips. Our simulator for complete trees rejects the tree if i) the number of lineages exceeds the maximum, or ii) if one or both subtrees descending from the root went extinct. In other words, we only accept trees where both the left and the right descendant branches of the root node have at least one surviving species (i.e., conditioning on survival

of both daughter lineages of the root, see Eq. 4). Next, we prune the extinct lineages from the complete trees to form "reconstructed" trees. The information about diversification rate shifts on the surviving lineages is kept, while the records of diversification rate shifts on the extinct lineages is discarded. These final trees only have tips at the present, and appear as if the tree was reconstructed or inferred based on molecular or morphological characters of extant species.

Simulation specifics.— We simulated phylogenies under different settings to assess the performance of the `Pesto` inferences in different scenarios. Specifically, we used several models with three rate categories, ranging from low to high net-diversification rates (Table 4.1), with constant relative extinction (μ/λ). The initial diversification rate category was always the one with the lowest net-diversification rate, meaning that i) the first rate shift that could happen must be an upwards shift, and that we ii) expected upward shifts to be more common. We varied the range of rate variation (from tiny to large), meaning that the difference among the rate categories λ_j within each model was varying. Moreover, we simulated trees with a time interval ranging from 25 to 125 Ma, resulting in different distributions of tree size. For computational purposes, we rejected simulations with too many tips ($> 50,000$), leading to a truncated distribution. We simulated 100 phylogenetic trees for each setting, meaning that there were in total 2000 simulated trees (5 time intervals and 4 levels of rate variation).

Models	Rate variation	λ_1	λ_2	λ_3	μ_1	μ_2	μ_3
A	Tiny variation	0.12	0.15	0.18	0.08	0.10	0.12
B	Small variation	0.12	0.18	0.24	0.08	0.12	0.16
C	Moderate variation	0.12	0.21	0.30	0.08	0.14	0.20
D	Large variation	0.12	0.24	0.36	0.08	0.16	0.24

Table 4.1: We set up a series of three-category models with net-diversification rates $\boldsymbol{r} = [0.04, 0.04+\beta, 0.04+2\beta]$, and relative extinction rates of $\mu/\lambda = 2/3$. $\beta \in \{0.01, 0.02, 0.03, 0.04\}$ controls the range of allowed rate variation, from tiny to large. We simulated trees of varying height, including 25, 50, 75, 100 and 125 Ma between the most-recent common ancestor and present. All simulations begin with the first category (λ_1, μ_1), which results in up-shifts being common in the simulated trees. The speciation and extinction rates were chosen such that the variation in tree size was not too large (i.e. such that $n_{\text{tips}} > 10^5$ was relatively rare). We selected a shift rate such that rate shift events are much less frequent than the branching events remaining in the reconstructed tree ($\eta = (\lambda_1 - \mu_1)/50 = 0.0008$).

We tested the performance of `Pesto` under three different scenarios. First, we used the true diversification rate categories $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ and the true shift rate η. Second, we used the true diversification rate categories, but we assumed the shift rate to be unknown. Third, we assumed that both the diversification rate categories and the shift rate were were unknown. This allowed us to assess the estimates of the shift rate parameter, the inference of significant branch-specific diversification rate shifts, as well as the estimates of the branch-specific diversification rates, in various scenarios of parameter (un)certainty. To assess the inference of branch-specific diversification rate shift inferences, we used metrics that are standard in

binary classification problems, including accuracy, false positive ratio, and false negative ratio (Powers, 2011).

4.3.5 Implementation

We implemented our inference method in the `Julia` (Bezanson et al., 2017) module `Pesto`. `Pesto` is an acronym for *Phylogenetic Estimation of Shifts in the Tempo of Origination*. The source code and documentation is available on github in the form of vignettes with example code (`https://github.com/kopperud/Pesto.jl`). To solve the differential equations numerically, we use the Runge–Kutta algorithm of Tsitouras (2011) implemented in `DifferentialEquations.jl` (Rackauckas and Nie, 2017). We implemented the simulations in a different module, `BirthDeathSimulation.jl` (`https://github.com/kopperud/BirthDeathSimulation.jl`), where it is possible to export the tree as an extended Newick string with metadata for each node. We plotted the figures using `ggtree` (Yu et al., 2017), `Makie.jl` (Danisch and Krumbiegel, 2021) and Ti*k*Z.

4.4 Results

We implemented an extremely efficient method for inferring branch-specific diversification rates and diversification rate shifts under the birth-death-shift process. `Pesto` uses the same underlying model, the birth-death-shift process, as the `LSBDS` implementation in `RevBayes` (Höhna et al., 2016a, 2019). Note that the `LSBDS` implementation of the birth-death-shift model can estimate the parameters of the base distributions (and the shift rate) jointly using hierarchical models, whereas our approach in `Pesto` can not. Therefore, the results of a `Pesto` and `LSBDS` analysis may in practice be different. If the same model is chosen, however (i.e., the prior distributions and discretization is identical) the two methods will give exactly the same estimates. We show the equivalence in Fig. S4.3 and thus validated our implementation. The efficiency of our method enabled us to set up many tests for exploring various model choices and their impact on parameter estimates, which we explore in the following sections.

The primary use case of `Pesto` is to estimate branch-specific diversification rates and to infer branch-specific diversification rate shift events. As an example, we show the estimated number of accumulated diversification rate shifts and the average speciation rates —plotted with different colors and mapped on the branches of the phylogeny— for primates (Vos and Mooers, 2006, Fig. 4.4). Based on the primates phylogeny, our posterior mean estimate for the total number of diversification shifts is $\hat{N} = 4.9$, summed across all diversification rate shifts and across all branches in the phylogeny. However, only one branch experienced a significant diversification rate shift event. On the branch that led to the Old World Monkeys clade (Cercopithecidae), we estimated that the number of diversification rate shifts is far higher than what is typical on other branches ($\hat{N} = 0.91$ vs close to zero) and this branch had strong

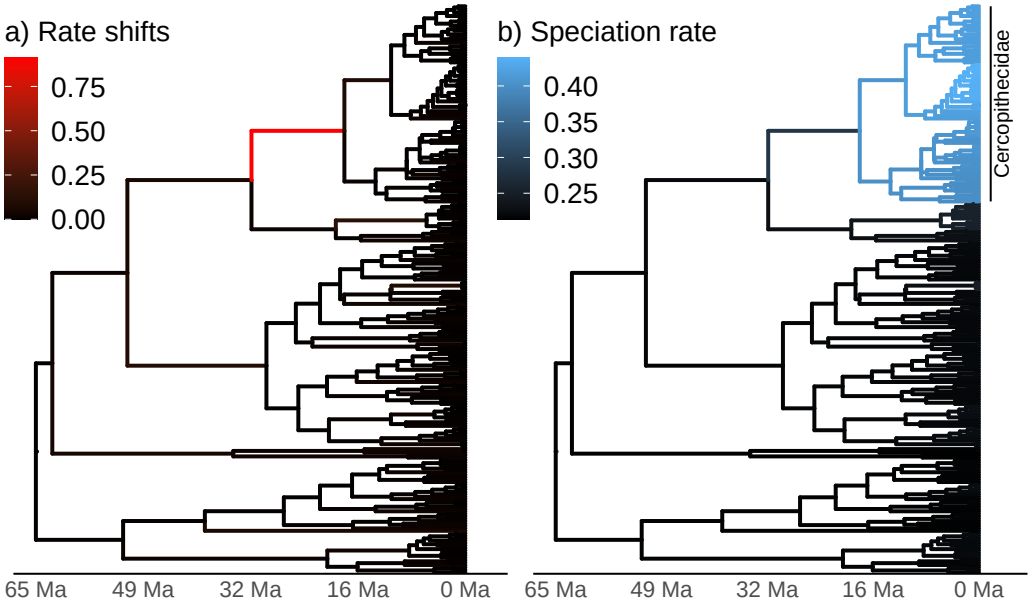


Figure 4.4: Posterior mean number of diversification rate shifts (a) and posterior mean branch-specific speciation rates (b) on the primate phylogeny (Vos and Mooers, 2006). We estimated the highest number of diversification rate shifts on the branch that led to the Old World Monkeys clade (Cercopithecidae). There are many branches with a low (but non-zero) posterior mean number of diversification rate shifts, which are not discernible using a simple color gradient. The estimate for the shift rate is $\eta = 0.0032$ number of diversification rate shift events per time per lineage. The posterior mean number of diversification rate shifts is $\hat{N} = 0.91$ on the red branch, whereas the across the whole tree it is $\hat{N} = 4.9$. There are 233 species in the primate phylogeny, and we assumed an equal taxon sampling probability of $\rho = 0.62$ (i.e., we assumed that there were 376 species in total, Groves 2005). See Fig. 4.1 for the model design.

statistical support to have experienced at least one diversification rate shift event (with a Bayes factor of 173). This corresponds to the visual interpretation of one clear diversification rate shift (Fig. 4.4), whereas the additional diversification rate shifts are inferred with low frequency and low magnitude distributed over the remaining branches.

4.4.1 Evaluation and Exploration of `Pesto` & Methods Choices

Impact of the diversification rate priors.— When a rate shift occurs under the birth-death-shift process, the new diversification rates are determined based on a base distribution (Höhna et al., 2019). This base distribution could have many different shapes, for example, a log-normal distribution, a log-uniform distribution, an exponential distribution (e.g., as used in `BAMM`, Rabosky 2014), or a gamma distribution. So far, little is known about which base distribution is most realistic for diversification rates and what impact the choice of the base distribution has. The base distribution governs the rate heterogeneity allowed under the model. If the base distribution is too narrow or too wide, then the model may not be adequate to explain the data. Importantly, each potential base distribution has a different shape and therefore spreads the values differently.

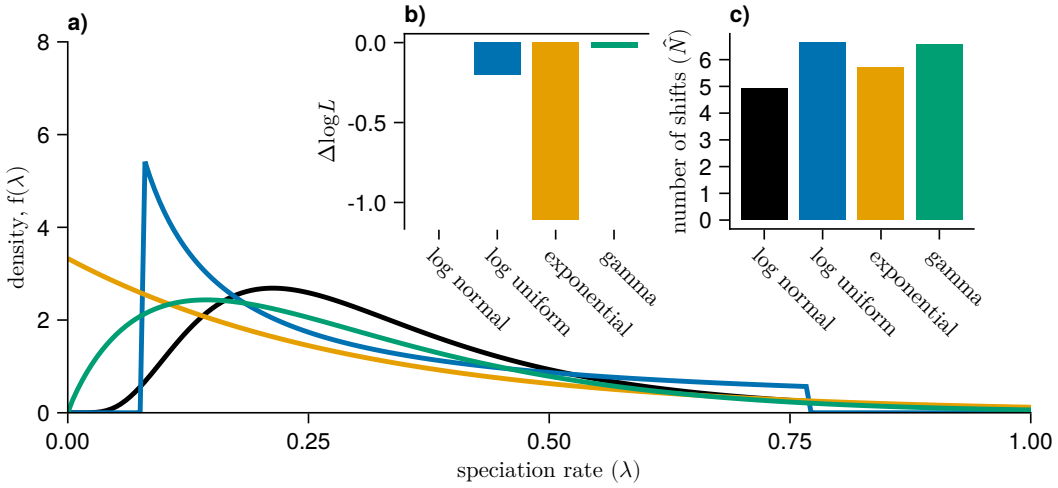


Figure 4.5: Impact of different base distributions on the branch-specific diversification rate analyses. Panel a) depicts the probability density functions of the log-normal (log-mean = $\log(\hat{\lambda})$, $\sigma = 0.587$), log-uniform (mean = $\hat{\lambda}$, one order of magnitude range), exponential (mean = $\hat{\lambda}$) and gamma (mean = $\hat{\lambda}$, variance equal to the variance of the log-normal) distributions for the speciation rate. For all models, we drew 10 quantiles from the distributions, and we estimated η conditional on $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ using maximum likelihood. Panel b) shows the relative support (ΔlogL) for each prior distribution. Panel c) depicts the expected number of shifts across all branches in the primate phylogeny.

For the primates example we see that the choice of base distribution has an impact on the estimated number of rate shifts (Fig. 4.5c). The posterior mean number of rate shifts were 4.9, 6.6, 5.7 and 6.6 for the log-normal, log-uniform, exponential and gamma base distributions, respectively. The pattern of one strongly supported branch was supported for the log-normal,

log-uniform and gamma distributions ($\hat{N} \approx 0.91$ and Bayes factor in the range of 109–173). The exponential distribution also showed strong support for the Old World Monkeys branch ($\hat{N} = 0.92$ and a Bayes factor of 162), but additionally weaker support for a second branch (in the guenon clade, *Cercopithecus*). This second branch had a low number of estimated rate shifts ($\hat{N} = 0.02$) and weaker but still strong statistical support (with a Bayes factor of 11), but the overall magnitude of the rate shift was small. The pattern of many branches with few number of rate shifts ($\hat{N} < 0.2$) was consistent across all four choices of the base distribution. The branch-specific diversification rate estimates were quantitatively different across the four base distributions (at most by 0.05 speciation rate units), but all four choices were consistent in inferring much higher rates for the Old World Monkeys clade.

In principle it could be possible to choose base distributions based on model testing (Fig. 4.5b). For the remainder of this study we choose the log-normal distribution as the base distribution because it was the most supported distribution in this test case and has two main advantages. First, the standard deviation and thus the range of values can be specified independently of the mean (which is not possible for the exponential distribution). Second, the distribution is centered around an *a priori* specified value compared to assigning the highest density to small values, as is done for the log-uniform distribution (Fig. 4.5a). Overall, we expect the log-normal and gamma prior distributions to produce similar results as both distributions can resemble one another, although the gamma distribution is usually more skewed towards smaller values (Fig. 4.5a).

Impact of the number of rate categories.— The base distributions of the birth-death-shift process are, in theory, full continuous distributions. Therefore, the number of possible rate categories is infinite. However, the approach implemented in `Pesto` is limited to a finite number of rate categories. Ideally, we would like to pick as many rate categories as possible, as the model will potentially fit better. In practice, we want to pick as few rate categories as possible, since more categories increase the computational time. In other words, we need to establish at what point adding more rate categories does not change the number of inferred diversification rate shifts, nor the branch-specific diversification rate estimates. We tested whether the number of allowed rate categories had a systematic impact on the estimates of the number of shifts and the branch-specific diversification rates for the primates tree (Fig. 4.6). When the state space is small, for example increasing the number of categories from 4 to 9, or from 9 to 16, there is some impact on the number of diversification rate shifts and the branch-specific diversification rates (Fig. 4.6). However, the estimates appear to stabilize when additional categories are used ($n \geq 6$). This indicates that the number of diversification rate shifts and branch-specific diversification rates are not systematically biased by the decision to use discrete rate categories.

Impact of the shift rate parameter on the number of diversification rate shifts.— Estimating the number of diversification rate shift events is arguably the most exciting result

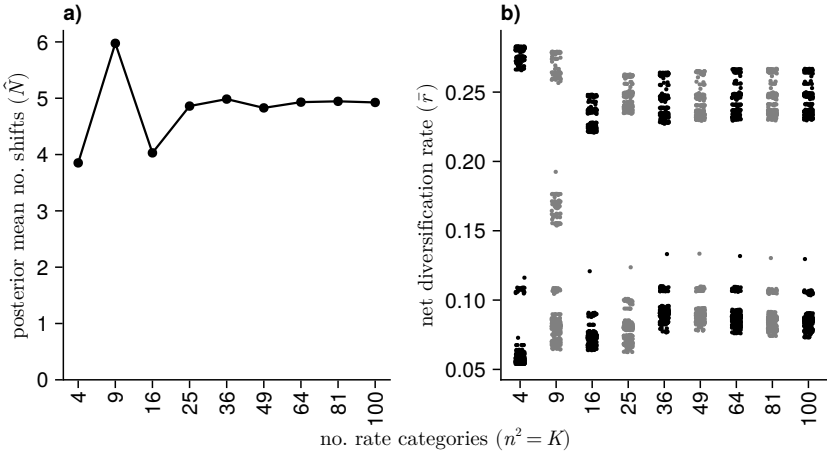


Figure 4.6: The impact of the number of rate categories on the number of rate shifts (a) and branch-specific rate estimates (b). All analyses are made on the primates phylogeny (Vos and Mooers, 2006). The bimodal distribution for average branch rates in (b) reflects the Old World Monkeys clade vs all of the other primates (see Fig. 4.4). The number of rate categories does not systematically bias the posterior mean number of rate shifts (a), nor the branch-specific rate estimates (b).

of branch-specific diversification rate methods (Alfaro et al., 2009; Rabosky, 2014). Previously, Moore et al. (2016) argued that estimates of the number of diversification rate shifts are not reliable and driven entirely by the choice of prior. That is, for a low rate of shifts, the methods infer few diversification rate shift events (or conversely, many events for a high shift rate). Our exploration of the impact of the shift rate parameter η on the posterior mean number of diversification rate shift events (Fig. 4.7) discovered several important aspects. First, the total number of shifts is sensitive to the shift rate η (see also Höhna et al., 2019). Second, the posterior mean number of diversification rate shifts is almost perfectly correlated with the prior number of shifts when the shift rate η is large (Fig. 4.7c). When η is small, however, the posterior number of diversification rate shifts is invariant to the rate prior (Fig. 4.7c). Thus, there is some but weak information in the phylogeny to estimate the number of diversification rate shifts.

A common approach to assess the support for a hypothesis, e.g., how many diversification rate shift events occurred along the phylogeny, is to use Bayes factors (Jeffreys, 1961; Kass and Raftery, 1995). Bayes factors are defined as the ratio of marginal likelihoods of the data between two hypotheses, e.g., if there was at least one diversification rate shift on the branch vs. no diversification rate shifts. The marginal likelihood is the integral over all parameter values, e.g., the branch-specific diversification rates, weighted by the prior of the parameters. Thus, Bayes factors have the advantage that they account for the prior on the number of diversification rate shift events. Nevertheless, Bayes factors are not independent of the prior on the shift rate (Fig. 4.7b). Since the marginal likelihood is often difficult to compute, as for the birth-death-shift process, one can compute instead the Bayes factor as the posterior ratio divided by the prior ratio (Eq. 17). We implemented and explored branch-specific Bayes factors

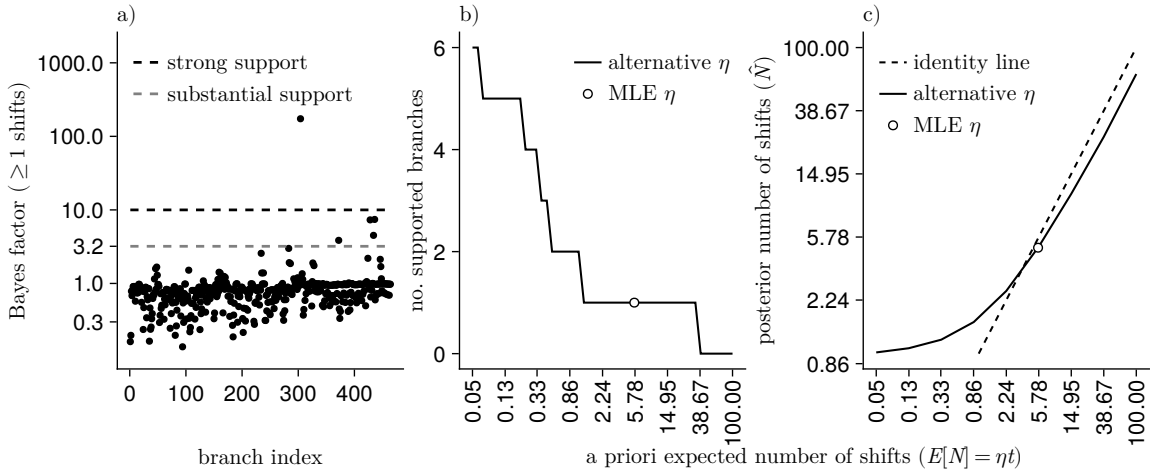


Figure 4.7: Branch-specific support for there being at least one shift, and the impact of the shift rate parameter (η) on the diversification rate shift inferences, for the primates phylogeny. The Bayes factors per branch (see Eq. 17) are computed using the maximum-likelihood estimate of the shift rate ($\eta = 0.0032$). The most supported branch, with a Bayes factor of 173, is the branch that led to the Old World Monkeys clade. If the Bayes factor is above 10, we consider the diversification rate shift to be strongly supported. Panel b) shows the impact of the shift rate (η) on the number of supported branches. With our maximum-likelihood estimate of the shift rate, there was one strongly supported branch. If the shift rate is much smaller, then there is more than one strongly supported branch. The posterior mean number of diversification rate shifts ($\hat{N}$) is highly correlated to the prior expectation ($E[N] = \eta t$, where t is the tree length), except when the prior is small (c). Note that when the prior and posterior mean number of diversification rate shifts diverge (c) the Bayes factors become larger, which in turn results in statistical support for more branches (b).

for detecting diversification rate shifts (Fig. 4.7a,b). Indeed, we found that the Bayes factors are not independent of the shift rate parameter (η), which controls how many diversification rate shifts we expect *a priori*. Nevertheless, for the maximum likelihood estimate of the shift rate ($\eta = 0.0032$) we inferred one branch with a strong support for a diversification rate shift (Bayes factor = 173).

Run time.— The run time of `Pesto` is the main advantage over similar methods that infer branch-specific diversification rates. We only perform two passes of the pruning algorithm to compute the likelihood function (backward pass) and find the posterior mean number of diversification rate shifts and branch-specific diversification rates (forward pass, Fig. 4.2). Furthermore, our approach has the tremendous advantage that the diversification rate category probabilities are computed precisely, without the common Monte Carlo uncertainty. Thus, convergence of branch-specific diversification rate estimates is guaranteed and convergence assessment is not necessary. Every analysis will yield exactly the same results, provided the same model and phylogeny are used.

In Fig. 4.8 we provide an overview of how long one can expect to run the program with various number of taxa in the phylogeny. The entire analysis takes about 1 second for a phylogeny with 100 taxa, whereas it takes less than 20 minutes for a phylogeny with 20,000 taxa. The backwards pass and the forwards pass (calculating $\mathbf{D}(t)$ and $\mathbf{F}(t)$) scales approximately

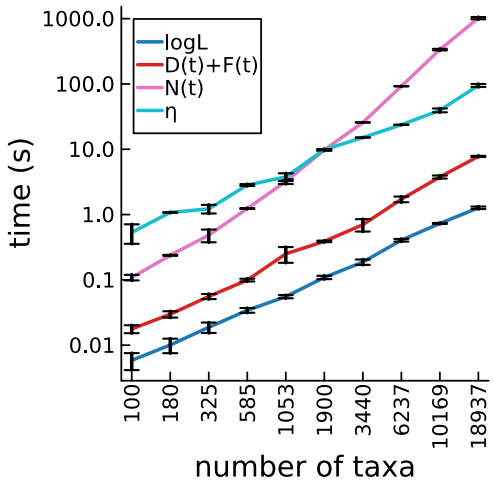


Figure 4.8: Time benchmarks of the inference procedure. The run time of `Pesto` is plotted, partitioned into the log-likelihood calculation (Eq. 4), the branch-rate calculation (postorder and preorder pass, $\mathbf{D}(t)$, Eq. 3 and $\mathbf{F}(t)$, Eq. 6), the maximum-likelihood estimation of the shift rate ($\hat{\eta}$), and the number of shifts calculation ($\hat{\mathbf{N}}(t)$, Eq. 12). We repeated these for a range of simulated phylogenies, all with $n^2 = 36$ rate categories. All time measurements are repeated five times, and we present the median ± the interquartile range. Time spent for installation, just-in-time compilation and garbage collection is not shown.

linearly with the number of tips, and quadratically with the number of rate categories. Computing the posterior mean number of diversification rate shifts $\hat{\mathbf{N}}$ is considerably slower, and scales worse than linear (but better than quadratic) with the number of tips. Fig. 4.8 shows the run times using a single thread, however in the default settings we use multi-threading to calculate both the likelihood and $\hat{\mathbf{N}}$ in parallel, meaning that one can expect the computation to be faster in practice. In our experience, finding the maximum-likelihood estimate for the shift rate parameter η is often the bottleneck of the inference procedure for small trees, requiring around 50-60 likelihood evaluations before a good estimate for η is found. We performed the time benchmarks using a desktop machine with an Intel Core i7-9700 CPU with a nominal clock rate of 3.0 GHz. Note that we did not want to compare the speed of `Pesto` to alternative implementations, e.g., `LSBDS`, as other implementations need to be carefully tailored to set up the length of the MCMC chain to achieve convergence. Nevertheless, `Pesto` runs easily 100 to 1000 times faster than `LSBDS`.

4.4.2 Simulation Study

In order to test the robustness of `Pesto`, we performed parameter estimation on phylogenies for which we knew the true model parameters, i.e., simulations. Note that we did not simulate under the exact same model as our inference settings, as we predefined different scenarios of rate variation (Table 4.1). Our prior settings for the inference on the rate variation is conservative, i.e., wider than realistic, which renders simulations computationally or practically infeasible.

Furthermore, our simulations were not conditioned on specific outcomes, for example by conditioning on exactly one rate shift event, as has been done in some previous simulation studies (Rabosky, 2014). Thus, our simulations include realizations with zero, one or many rate shifts and explore the robustness of `Pesto` in a more agnostic but also more challenging scenario. Finally, the insights of our simulation study should extrapolate to other methods that rely on the same or similar underlying models and theory (see also Martínez-Gómez et al., 2024).

Robustness in estimates of the shift rate.— We estimated the shift rate both when we considered the speciation and extinction rate categories as known and unknown. First, we will discuss the known case, where we set the speciation and extinction rate categories $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ to their true values (see Table 4.1). The results indicate that it is almost impossible to accurately estimate the shift rate for small trees (< 100 taxa). For large trees ($> 10,000$ taxa), the shift rate estimates converge to the true value. Fig. 4.9 broadly shows a bimodal distribution of estimates for the shift rate: one mode close to zero, and a second mode clearly larger than zero. Most of the trees with an estimated shift rate near zero either did not have any true shifts, or experienced shifts close to the present (e.g., on a terminal branch) during the simulation. Note that some shift rate estimates converged to our pre-defined boundary of the allowed parameter range (upper limit equal to one half times the maximum speciation rate). For these situations we noticed that the likelihood curve is flat and the hill-climbing algorithm can diverge.

Not surprising, when we consider the speciation and extinction rate categories $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ as unknown, it becomes more difficult to estimate the shift rate accurately (Fig. 4.9b, bottom row). Overall, the estimation error in η is larger if we also want to estimate the parameters of base distributions. For large trees, the shift rate estimates converge closer to the true values. However, estimates of the shift rate are biased even for very large trees ($50,000$ taxa) — with underestimates of about one order of magnitude. This is likely due to the diversification rates being estimated using the constant-rate process $(\hat{\lambda}, \hat{\mu})$, which are not representative when the true tree has undergone rate shifts.

Robustness of estimation in the number of diversification rate shifts.— We also assessed how well `Pesto` can estimate the number of diversification rate shifts from simulated phylogenies. Altogether, 54.7% of the simulated trees experienced at least one diversification rate shift. Among these, shift histories with a single diversification rate shift were the most common (27.1%), and more numerous shifts less common. The mean number of diversification rate shifts (among the trees with at least one rate shift) was 8.3, while the median was 3. Before we discuss the quantitative results, we first present two hand-picked simulation-and-inference pairs that are useful in guiding the understanding of how the model behaves (Fig. 4.10). As a brief reminder, we considered a branch to have experienced a diversification rate shift if the Bayes factor was larger than 10 and the number of diversification rate shifts $(\hat{N})$ was larger than 0.5.

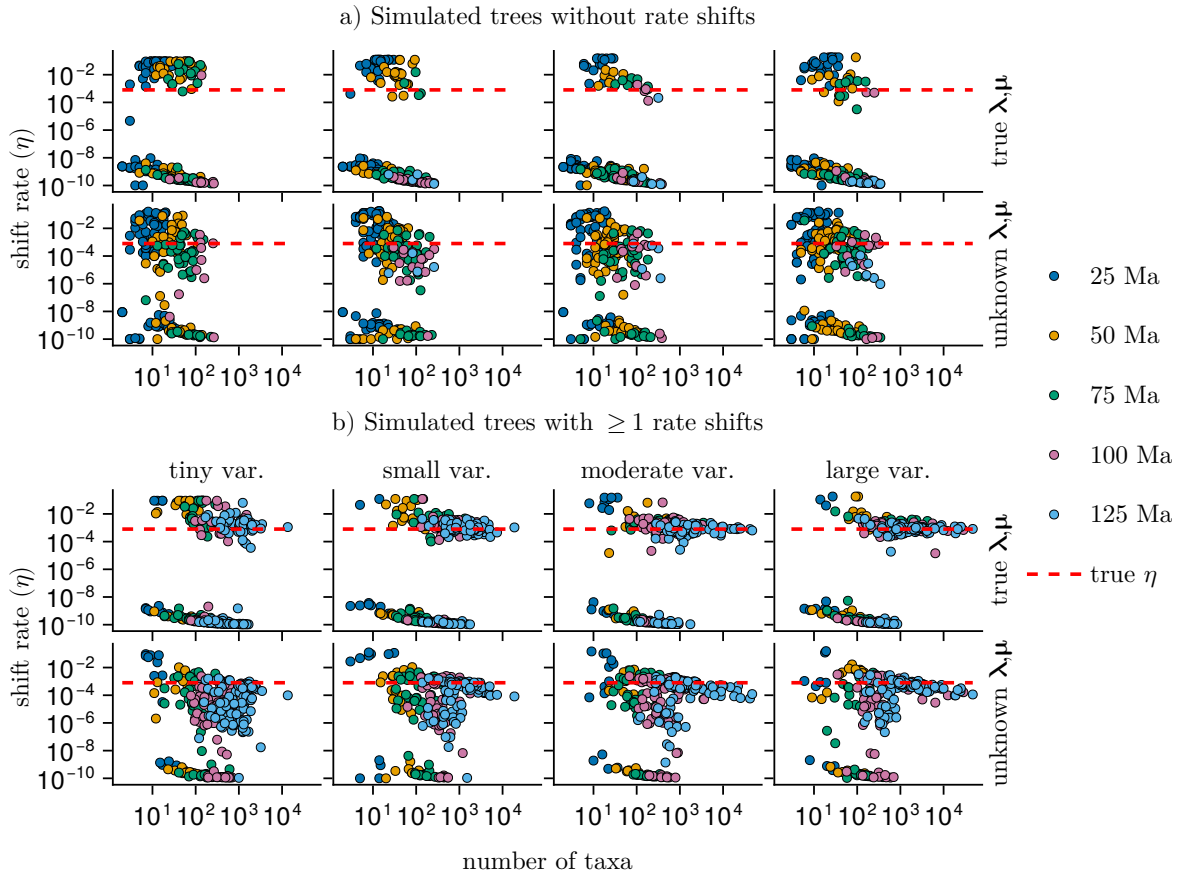


Figure 4.9: Maximum-likelihood estimates of the shift rate (η). Each point represents a phylogeny, simulated under the models in Table 4.1. The true shift rate is chosen such that diversification rate shift events are much less frequent than branching or extinction events (red line, $\eta = 0.0008$). The phylogenies are split by whether or not a rate shift event happened (panels a and b). We estimated the shift rate both with the true rate categories (i.e., the three-category models in Table 4.1) as well as using the two-step approach (i.e. first estimating $\hat{\lambda}, \hat{\mu}$ under the constant-rate birth-death process and secondly estimating the shift rate using maximum likelihood). The allowed rate variation ranges from tiny to large, and we simulated trees with a range of 25 to 125 Ma tree heights. Simulating for a longer time and increasing the rate variation has a similar effect in reducing the estimation error, since both will result in more species-rich trees.

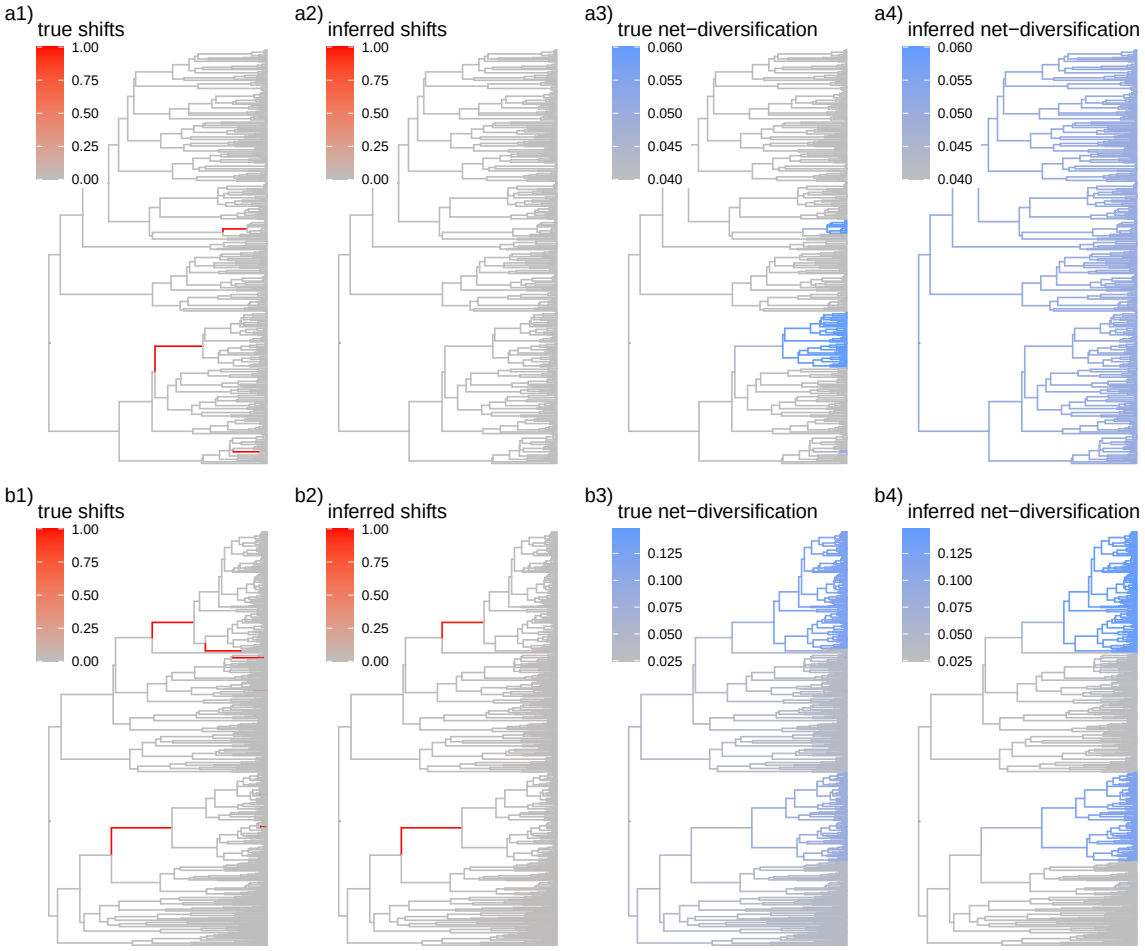


Figure 4.10: Two examples of simulated trees under the birth-death-shift model. First where `Pesto` did not recover the true shifts (a), and second where the `Pesto` analyses recovered the two oldest shifts (b). The first tree is simulated under tiny rate variation, and the second under large rate variation (models A and D in Table 4.1), for a period of 100 million years. We used the two-step approach in `Pesto` (with $n^2 = K = 36$ categories) to estimate the rate shifts and the branch-specific diversification rates. The color scales are linked per row.

In Fig. 4.10, we show a tree simulated under a model that has tiny diversification rate variation, and a tree that has large diversification rate variation (models A and D, Table 4.1). These two trees are representative of the major trends across all 2,000 simulated trees. In the first tree, there were three diversification rate shifts, and `Pesto` incorrectly inferred zero shifts, i.e. no false positives and three false negative shifts. In the second tree, there were seven diversification rate shifts, and `Pesto` correctly inferred two major diversification rate shift events, i.e., no false positives and five false negative shifts. There are several reasons why `Pesto` might not be able to recover diversification rate shifts such as the ones in Fig. 4.10. As a general rule, we believe that the power to detect diversification rate shifts is high when the true diversification rate variation is large, and that the power is low when the true diversification rate variation is small. This is corroborated in Fig. 4.10 where we did not recover any diversification rate shifts in the tree with tiny diversification rate variation ($r_1 = 0.04$ vs. $r_3 = 0.06$), while we were able to detect some diversification rate shifts in the tree with large diversification rate variation ($r_1 = 0.04$ vs. $r_3 = 0.12$). For the diversification rate shifts that were not recovered in the second tree (despite there being a large amount of diversification rate variation), they i) occurred closer to the present, and some of the diversification rate shifts ii) led to a decrease in the net-diversification rate. Both of these scenarios make it unlikely for the descendant lineages to experience more than a few speciation events, and the resulting subclade is relatively poor in terms of the number of the species. When there are relatively few species in a particular clade, the information available for detecting a diversification rate shift is limited. In turn, this reduces the power of the method to detect diversification rate shifts.

Overall (Fig. 4.11), we found that the accuracy of diversification rate shift detection was very high (mean = 99.6%), the false positive ratio was tiny (mean = 0.004%), and the false negative ratio was high (mean = 94.4%). The inference of diversification rate shift events is more robust when the tree size is larger, and when the diversification rate variation is higher. We only saw substantial false positive shifts for small trees with fewer than 250 taxa. For all larger trees, the number of false positives were negligible. This means that `Pesto` is conservative by underestimating the number of diversification rate shifts. If `Pesto` infers a diversification rate shift, then there is strong evidence that this diversification rate shift corresponds to a true diversification rate shift because of the extremely low false positive ratio.

Robustness of branch-specific diversification rate estimates.— Next, we assessed the robustness of the branch-specific diversification rate estimates. First, we consider again the same two example trees (Fig. 4.10). For the tree simulated under tiny diversification rate variation, `Pesto` did not correctly infer the diversification rate shift, and thus the background net-diversification was overestimated, and the diversification rate was underestimated in the clades that shifted to a higher rate. For the second tree, `Pesto` recovered the two major

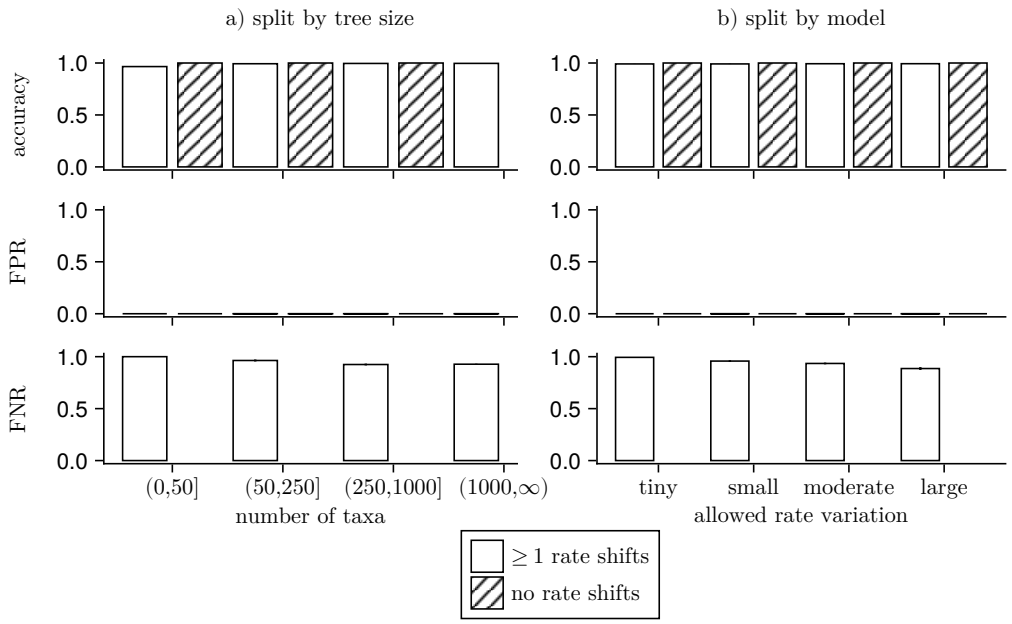


Figure 4.11: Classification of diversification rate shift inferences for 2000 simulated trees, under the two-step inference procedure, split by tree size (a), rate variation (b) and whether or not the true rate shift history contained at least one diversification rate shift. For each tree, we interpreted a diversification rate shift to be present if the Bayes factor was greater than 10 and $\hat{N}$ was greater than 0.5 per branch. Sorting each branch into true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), we calculated the accuracy ((TP+TN)/(TP+TN+FP+FN)), the false positive ratio (FPR = FP/(FP+TN)) and the false negative ratio (FNR = FN/(TP+FN)) for each tree. Each bar represents an average, and the error bar represents the standard error. We only calculated the false negative ratio for trees that had at least one diversification rate shift, since if TP+FN = 0, then the ratio is not defined. The false positive ratio is overwhelmingly small (mean = 0.004%), meaning that if we inferred a diversification rate shift on a branch, it was almost always present in the true tree. The false negative ratio is large, and often close to 1 (mean = 94.4%), meaning that we often inferred too few or no diversification rate shift even when there were shifts in the simulated tree. Different y-axis limits can be seen in Fig. S4.11.

shifts. Thus, `Pesto` correctly found a diversification rate difference in the two major clades that underwent a rate shift. However, the branch-specific diversification rates are slightly overestimated in the two high-rate clades, and slightly underestimated in the low-rate part of the tree.

To evaluate the estimation error in the full set of simulated trees, we calculated the proportional error in the branch-specific speciation rates (see Rabosky, 2014, eq. 15). For small trees (< 100 taxa), the average proportional error is roughly in the range of 0.5 to 2.0, meaning that the branch-specific diversification rates were underestimated by half, or overestimated by twice the true value (Fig. 4.12b, bottom row). When the true diversification rate category parameters ($\boldsymbol{\lambda}, \boldsymbol{\mu}$) are used (Fig. 4.12, top and middle row), the proportional error converges to one when the number of taxa increases, indicating that the branch-specific diversification rate estimates are unbiased, and the precision of the method increases with larger trees. When we also estimate the diversification rate categories via the base distribution parameters,

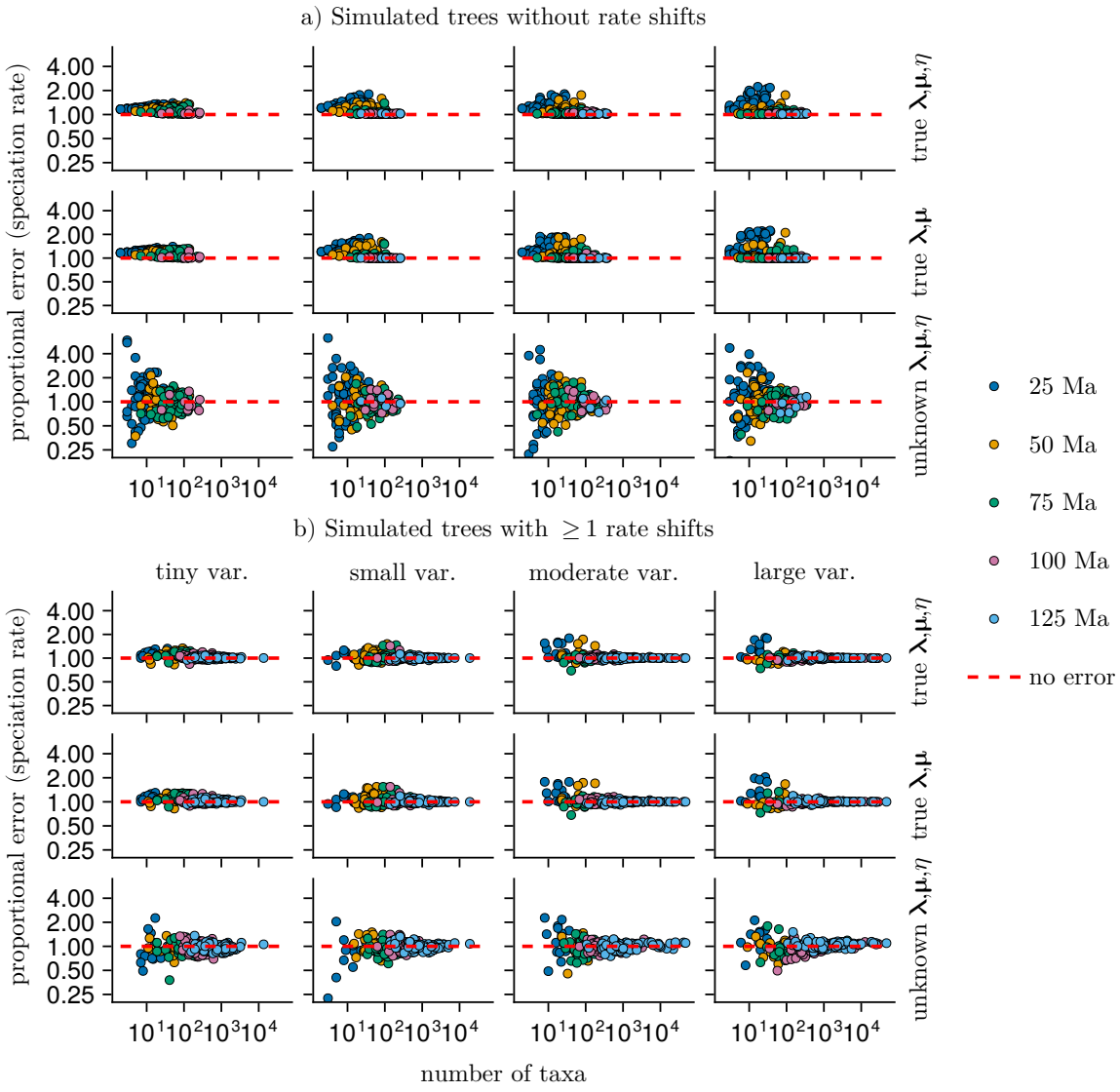


Figure 4.12: The estimation error in branch-specific speciation rates, as proportional errors, in simulated trees. Panel a) includes trees where no rate shifts happened, and panel b) includes trees where at least one rate shift happened. We calculated the proportional errors per branch, and averaged across the tree: $\exp(\frac{1}{n_{\text{branches}}} \sum_M \log(\bar{\lambda}_M) - \log(\bar{\lambda}_{M,\text{true}}))$, where $\bar{\lambda}_M$ is the posterior mean speciation rate on branch M (Eq. 9). A value of one means that the branch-specific speciation rates are unbiased, on average across the tree. Larger than one means overestimation, and less than one means underestimation. See Figs. S4.7 and S4.8 for estimation errors in extinction and net-diversification rates.

then the proportional error still converges, however with a small bias towards overestimating branch-specific diversification rates.

4.5 Discussion

In this study we developed a new approach called `Pesto` for estimating branch-specific diversification rates. The key feature of our approach is the computational speed; we can estimate branch-specific diversification rates on phylogenies with thousands of taxa without algorithmic stochasticity in minutes or hours. In this discussion, we will first discuss general aspects about our ability to infer branch-specific diversification rates and the number of diversification rate shifts that are not exclusive to `Pesto`. Then, we briefly contrast `Pesto` with other existing methods and indicate the advantages and disadvantages. Finally, we provide some general recommendation about how to use `Pesto` for inferring branch-specific diversification rates.

Inferring diversification rate shifts

To explain how one can best interpret the branch-specific results obtained from `Pesto`, it can be useful to consider how Höhna et al. (2019) investigated the posterior distribution by simulating stochastic mappings, similar to how stochastic mappings of discrete character evolution can be simulated (Huelsenbeck et al., 2003). Our approach is conceptually the same. Consider a stochastic variable Z that represents the history of diversification rate shift events on the phylogeny. The posterior distribution of Z is determined by the model (i.e., the base distribution and its parameters, and the shift rate) and the observations (i.e., the reconstructed phylogeny). What we have done is to develop a fast algorithm that directly computes summaries of Z, for example the branch-specific posterior mean number of diversification rate shifts. While this approach is much faster than a stochastic mapping approach, it comes with the disadvantage that we disregard any information about the variation in Z, in other words how uncertain the branch-specific estimates are. If one wishes to know the extent of the uncertainty in the branch-specific estimates, we recommend to use stochastic mapping, for example as in `LSBDS` (Höhna et al., 2019).

Sensitivity to the shift rate.— Moore et al. (2016) argued based on the performance of `BAMM` that the posterior distribution of the number of diversification rate shift events is strongly related to the prior expectation, which in our case is controlled by the diversification shift rate η. Overall, we agree that there is strong sensitivity to the shift rate parameter, or in other words that the diversification rate shift events are only weakly identifiable. Despite this, our exploration of the primates phylogeny revealed that the posterior mean number of diversification rate shifts is at minimum about one, even if the shift rate is assumed to be arbitrarily small. Therefore, the hypothesis of there being no diversification rate shifts

can confidently be excluded, and more generally this indicates that there is some (but weak) information in a phylogeny to detect branch-specific shifts in the diversification rate (Fig. 4.7c).

Assessing significant diversification rate shifts events.— Visually inspecting the branch-specific diversification rates can highlight the branches where a diversification rate shift event has occurred. Therefore, it might seem obvious to say how many branches experienced a diversification rate shift (Fig. 4.4). However, quantitatively assessing how many rate shifts occurred (and on which branches) has remained more challenging (Rabosky, 2014; Shi and Rabosky, 2015; May and Moore, 2016; Mitchell and Rabosky, 2017; Höhna et al., 2019). In principle, Bayes factors should be able to detect the branches for which the data and not only the prior increases the probability of a diversification rate shift. We note that the true prior distribution on the number of rate shifts is unknown, and we used a simplistic approximation for the prior. As with any statistical significance test, we recommend to assess both the hypothesis test (e.g., the Bayes factor) as well as the effect size (e.g., $\hat{N}$, or the absolute change in net-diversification rate across the branch). In other words, if the statistical hypothesis test is significant, but the effect size is minimal, then we argue that the diversification rate shift is not biologically meaningful (Nakagawa and Cuthill, 2007). Therefore, we suggest to compute both the posterior mean number of diversification rate shifts ($\hat{N}$) and the Bayes factor for there being at least one or more rate shift events. If the posterior mean number of diversification rate shifts on a branch ($\hat{N}$) is substantially larger than 0, say 0.5, and the Bayes factor indicates strong support (say 10), then we are confident in that a diversification rate shift event actually occurred on the branch. These criteria are designed to be conservative, such that we get few false positives, and high accuracy, but not necessarily few false negatives (see also May and Moore, 2016).

In our simulation study, we observed that `Pesto` has high power to detect diversification rate shift events when there were sufficiently many descendant lineages (i.e., due to an increase in the net-diversification rate), and lower power in other cases (i.e., due to a decrease in net-diversification rate). Since roughly half of the branches in a phylogeny are terminal lineages with only one descendant, it is not surprising that `Pesto` and similar methods have low power to detect all diversification rate shift events. Finally, as expected, we find that larger diversification rate shifts are easier to detect.

Comparison with other methods

The birth-death-shift model as implemented here and in `LSBDS` (Höhna et al., 2019) is conceptually similar to other birth-death models that allow for among-lineage rate variation, including `BAMM` (Rabosky, 2014), `CLaDS2` (Maliet et al., 2019), `MTBD` (Barido-Sottani et al., 2020), `MiSSE` (Vasconcelos et al., 2022) or `RPANDA` (Mazet et al., 2023), and we provide a comparison in Table 4.2. For example, `CLaDS2` models the diversification rate after a rate shift event as being correlated with the ancestral diversification rate, while we assume that the new

Table 4.2: A comparison of methods[a] for inferring branch-specific diversification rates and diversification rate shift events.

	`Pesto`	`LSBDS`	`BAMM`	`CLaDS2`	`MTBD`	`MiSSE`	`RPANDA`
Autocorrelated shifts	no	no	no	yes	no	no	no
Base distribution	discrete	discrete	continuous	continuous	discrete	discrete	-
Base dist. shape	flexible	flexible	exponential	log-normal	categorical	categorical	-
Base dist. parameters	mean	flexible	mean	mean & variance	-	-	-
Div. rate decay	no	no	yes	no	no	no	yes
Shifts at extinct/ unsampled lineages	yes	yes	no	yes	no	yes	no
Phylogeny	fixed	flexible	fixed	fixed[b]	fixed[b]	fixed	fixed
Location of rate shifts	branch	branch	branch	node	branch	branch	node
Shift hypothesis test	Bayes factor	Bayes factor	Bayes factor	-	-	-	AICc
Focal estimate[c]	tip & branch	branch	tip & branch	branch	branch	tip & node	branch
Ancestral reconstruction[d]	dynamic, $O(NK)$	sampling	sampling	sampling	sampling	dynamic, $O(N^2K^2)$	stepwise AICc

[a]Methods include `Pesto`, `LSBDS` (Höhna et al., 2019), `BAMM` (Rabosky, 2014), `CLaDS2` (Maliet et al., 2019), `MTBD` (Barido-Sottani et al., 2020), `MiSSE` (Vasconcelos et al., 2022) and `RPANDA` (Mazet et al., 2023). [b]See Barido-Sottani and Morlon (2023) where the phylogeny is allowed to be jointly inferred. [c]All MCMC based methods and `Pesto` can infer, with adequate post-processing of the results, estimates at internal nodes and tips. [d]N represents the number of nodes in a phylogeny, and K represents the number of rate categories.

diversification rate is drawn independently. Furthermore, `BAMM`, `MTBD`, `MiSSE` and our birth-death-shift model allow for diversification rates to shift along branches, unlike in `CLaDS2` and `RPANDA`, where diversification rate shift events happen at branching events. `BAMM` is perhaps most similar to `Pesto` and `LSBDS`, although `BAMM` allows for the diversification rates to decay over time, while we assume the diversification rates to be constant within a diversification rate category. There are two main reasons for why `Pesto` is more computationally efficient than other methods. First, `Pesto` uses fixed values for the diversification rate categories (Fig. 4.1). In other implementations, e.g., the `LSBDS` model in `RevBayes`, the values of the diversification rate categories can be estimated using hierarchical models. Second, unlike `LSBDS`, `BAMM`, `CLaDS2` and `MTBD`, we do not make use of Monte Carlo algorithms for the inference. Instead, we use a dynamic programming algorithm to compute the marginal probabilities of the diversification rate categories across the tree, which only requires two tree traversals. This is in contrast to `MiSSE`, which uses the marginal ancestral state reconstruction algorithm of Yang et al. (1995) to obtain marginal probabilities at the nodes in the tree. The disadvantage of this marginal ancestral state reconstruction algorithm is that it requires a tree traversal for every node and every diversification rate category (Caetano et al., 2018), which is considerably more than the two in `Pesto`.

Advantages and Disadvantages of `Pesto`

When designing how the parameters and the branch-specific metrics were to be estimated, our overall goal was for `Pesto` to be suitable to be run on exceptionally species-rich phylogenies. The main advantages of `Pesto` include i) its fast run time, ii) it is based on a generative stochastic process (unlike e.g. Chan and Moore 2005; Alfaro et al. 2009), and iii) quantitative tests of the hypothesis of whether there was a diversification rate shift event. However, there

are also some disadvantages with the estimation approaches that we used. Recall that we estimated the parameters of the base distributions (i.e., $\hat{\lambda}$ and $\hat{\mu}$) using the lineage-homogeneous, constant-rate birth-death process. While this gives a rough estimate of the overall scale of the birth and death rates, we expect that if a phylogeny has undergone diversification rate shifts, that the estimated rates obtained by fitting this simple model are biased. We suspect that the reason we see i) upward biased speciation rates and ii) downward biased shift rates, in particular for very large simulated trees, is partly due to the choice of using the lineage-homogeneous birth-death model for estimating the parameters of the base distributions. Furthermore, we investigated the posterior distribution of the rate shift histories conditionally on the parameters of the base distributions, and conditionally on the shift rate. We expect that i) joint estimation of the parameters and ii) joint inference of the parameters and the rate shift history would be more statistically robust (as in Höhna et al., 2019), however more research is needed for this to be computationally feasible for very large trees.

In keeping with most contemporaneous phylogenetic diversification rate studies (e.g. Egan et al., 2024; Quintero et al., 2024), we have assumed that the phylogeny is known without error. Even though Barido-Sottani and Morlon (2023) demonstrated that joint inference of phylogeny and a birth-death-shift model is possible, this type of joint inference is computationally challenging even for clades with few taxa, and is virtually impossible for species-rich clades. As a quicker but less robust alternative to the proper joint inference, however, one can independently apply `Pesto` to several samples from the posterior distribution of phylogenies, and interpret the variation in the results as estimation error that is due to uncertainty in the phylogeny.

General Recommendations on how to use `Pesto`

Both simple and more detailed instructions on how to use `Pesto` are provided at `https://kopperud.github.io/Pesto.jl/dev/`. As when fitting any probabilistic model, we recommend to inspect and consider the parameter estimates and their units. The speciation rates, extinction rates, and shift rates are in units of "number of events per lineage per time". If the phylogenetic tree is a species tree, then the time unit is typically one million years. In our experience, the shift rate should be much smaller than the mean speciation and extinction rates, by at least one or two orders of magnitude. It is biologically plausible that there are more speciation events than diversification rate shift events (note that not all speciation events are included in a reconstructed tree but all rate shift events would have to be included). If the estimated shift rate is too large, this might indicate that there was a numerical problem. For this reason, we put limits on the maximum allowed shift rate when finding the maximum likelihood value, however this hard boundary is not guaranteed to always avoid the issue.

In general, we recommend to use the log-normal distribution for the base distribution of diversification rates. Moreover, users of `Pesto` will have to decide how many rate classes to

use. In principle, more rate classes are better than fewer classes as the underlying continuous distribution is better approximated. However, additional rate classes come with computational demands. For clades that are similar to the primates in terms of time span and species heterogeneity, we recommend to use at least six or more rate classes, and to use all pairwise combinations of extinction and speciation rates (i.e., at least $K = n^2 = 36$ rate categories). If a phylogeny is particularly heterogeneous, we recommend to test whether six diversification rate classes is sufficient, and to use more if necessary. This can for example be tested by fitting the model with seven or eight classes, and assessing whether the results remain similar or are substantially different.

4.6 Conclusion

We have shown that we are able to estimate posterior mean branch-specific diversification rates and number of diversification rate shift events more efficiently than before (one computation vs. thousands or millions of simulations). The birth-death-shift model can be fitted to small trees (< 100 taxa) in about one second. Larger trees ($> 10,000$ taxa) take slightly more time, on the order of a few minutes on a standard desktop computer. Compared with the established estimation technique of using Markov-chain Monte Carlo and sampling stochastic mapping of the diversification rate shift history, which can take hours, days or even weeks for large trees, our new method and implementation is several orders of magnitude more efficient. This improvement in run time is not only more convenient, but enables empiricists to broaden their taxonomic resolution. We expect that through `Pesto`, investigations of diversification rate variation across the tree of life will become more practically feasible and accessible.

Reproducibility

The analyses and figures presented here can be reproduced by running the scripts in the accompanying github repository (`https://github.com/kopperud/pesto_ms_analyses`). The simulated trees and results are not committed to the repository due to file size restrictions, however they can be downloaded from datadryad DOI: 10.5061/dryad.mgqnk995b

Chapter 5

The nature and prevalence of diversification rate shifts across the Tree of Life

Pesto is a sauce made from basil leaves, toasted pine nuts, garlic, Parmesan, olive oil and salt. A bowl of pasta and pesto makes for a quick and tasty meal. Perhaps now they will begin selling bifurcating spaghetti in convenience stores.

5.1 Abstract

Strong disparity of species richness in major clades is well documented, but how the underlying process that generated biodiversity varies is less understood. We investigate clade-specific diversification rate shifts in many species-rich phylogenies spanning the Tree of Life. Diversification rate shifts are ubiquitous across the studied phylogenies, indicating that the process of diversification is highly variable. Our results show more frequent changes in younger clades and an overall excess of upshifts (rapid radiations), resulting in an acceleration of net-diversification rates. Heterogeneity in diversification appears also to be related to the overall tempo of diversification itself. The nature of diversification rate shifts (i.e., upshifts vs downshifts and speciation rate vs extinction rate shifts) remains inconclusive due to data insufficiency and inference challenges.

The disparity in species richness across major groups of life is one of the most studied aspects of biodiversity research. Some organisms, including the tuatara reptile, coelacanth fishes or the ginkgo tree, often nicknamed 'living fossils', represent species-poor groups with only one or a few living taxa despite the old age of their corresponding lineages. Other groups such as beetles, bats, cichlids and orchids, are represented by thousands of living species and are much more species-rich than their sister clades or other clades of comparable age. How these stark differences in species richness arise and if these examples represent exceptions from the general diversification process remains a significant topic of debate in evolutionary biology (Sanderson and Donoghue, 1994; Jetz et al., 2012; Hedges et al., 2015).

To statistically evaluate differences in species richness, one can obtain diversification rates (i.e., speciation and extinction rates) for different study groups (Sanderson and Donoghue, 1994, 1996). When considering the whole Tree of Life, and thus the general process of historical biodiversity, then subclades with differential rates of diversification must have experienced shifts in the rate of diversification, either through many small rate shifts (Davies et al., 2004; Budd and Mann, 2025) or few but large rate shift events (Alfaro et al., 2009; Jetz et al., 2012; Rabosky et al., 2018). Considerable effort has been devoted over the last decades into developing robust and efficient approaches, including the likelihood-based birth-death-shift process, to model lineage-specific diversification rates with diversification rate shifts (e.g., Alfaro et al., 2009; Rabosky, 2014; Barido-Sottani et al., 2020; Maliet and Morlon, 2022; Quintero et al., 2024). Speciation events (births) are represented by bifurcations in the phylogeny, increasing the diversity, whereas extinction events (deaths) reduce the diversity of the group, and diversification rate shift events lead to different rates of speciation and extinction of the subtending lineage (Morlon et al., 2024). For example, a shift towards increased speciation rate may lead to the generation of a successful, species-rich clade whereas a shift towards a lower speciation rate (and at the same time low extinction rate), may lead to the situation of a 'living fossil'.

While several studies have investigated diversification rate variation across the Tree of Life (Jetz et al., 2012; Rabosky et al., 2018; Martínez-Gómez et al., 2024), these studies were limited in two ways. First, due to computational constraints of the birth-death-shift model (and variants thereof, Rabosky, 2014; Barido-Sottani et al., 2020; Vasconcelos et al., 2022; Maliet and Morlon, 2022; Quintero et al., 2024), large phylogenetic trees could not be investigated in a reasonable time. Second, previous studies have been concerned with investigating specific study groups, and less so with the overall pattern of diversification rate heterogeneity.

A central hypothesis is whether the number of rate shifts varies across the Tree of Life. That is, we are interested in whether some groups experienced more diversification rate shifts than other groups, and if these differences can be explained by stochastic variation. We expect the number of rate shifts to be related to the time span of the phylogeny, since over a longer time span there have been more opportunities to accumulate rate shifts. When the

number of shifts are counted per time unit and per lineage, we can compare phylogenies of different size and timescale. Overall, there may be several factors that are correlated with variable rates of diversification rate shifts and we will focus primarily on age dependency and net diversification rates. An age-dependency of diversification rate shifts could indicate some fundamental property of older vs. younger clades or indicate some overall temporal variation in when diversification rate shifts are more prominent. Net-diversification rates are an indicator of volatility and turnover in clades which could be correlated with the amount of diversification rate shifts.

We also investigate whether diversification rate shifts led to an upward or downward change in net diversification. Prior empirical evidence suggests that diversification rates in a clade have a tendency to dwindle (Scantlebury, 2013; Moen and Morlon, 2014), although these patterns were found across time and not necessarily among clades. Alternatively, it is possible that older clades could have experienced a rapid radiation due to an upwards diversification rate shift event, followed by a gradual decline in diversification rate (Lovette and Bermingham, 1999; Kozak et al., 2006; Rabosky and Lovette, 2008b). Importantly, most studies have focused on rapid radiations with an increase in diversification rates. Our study will elucidate whether rapid radiations (i.e., upward shifts) are truly a more common pattern, or whether looking broadly over many large-scale phylogenies shows a more balanced pattern of upward and downward diversification rate shifts.

We test our hypotheses by estimating lineage-specific diversification rates and shifts of diversification rates along lineages. The recently developed `Pesto` framework (Kopperud and Höhna, 2025) enabled us to fit the birth-death-shift model to numerous and large phylogenies obtained from previous publications. Overall, we observe that diversification rate shifts are present in all studied empirical phylogenies. However, some phylogenies experienced far more diversification rate shifts than others. In particular, younger clades appear to exhibit more frequent rate shifts than older clades. Assessing the nature of diversification rate shifts is complicated by intrinsic limitations in detecting such events. Specifically, we can readily detect upshifts in net diversification, but we are less able to detect downshifts, which could explain why we "see" more rapid radiations. Finally, we discuss the prospect of assessing whether a rate shift event is primarily due to a change in extinction rate or speciation rate.

5.2 Estimating branch rates and rate shift events across the Tree of Life

We compiled a set of 24 published empirical phylogenies. Overall, we targeted species-rich phylogenies that were reconstructed using probabilistic methods, and we selected groups such that there were no taxonomic overlaps (see Table S5.1 in the Supplementary Materials). These phylogenies span across many multicellular forms of life, including marine vertebrates, terres-

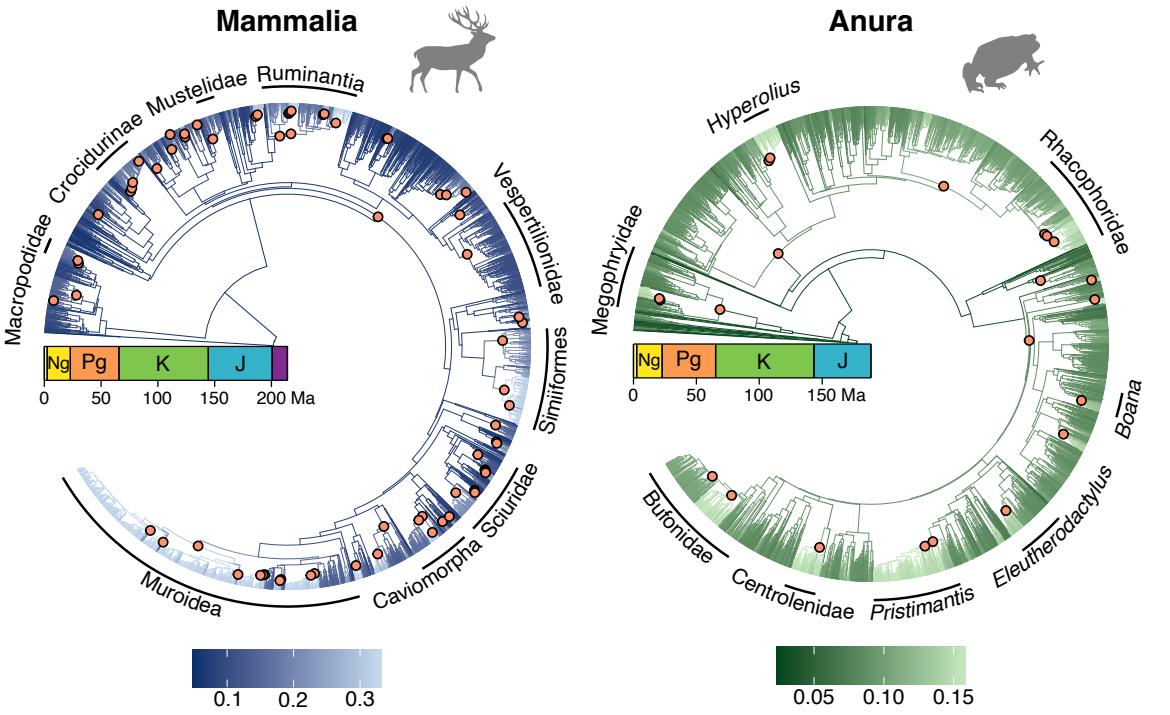


Figure 5.1: Branch-specific net-diversification rate estimates for phylogenies of mammals (Mammalia, Álvarez-Carretero et al., 2022) **and frogs** (Anura, Portik et al., 2023). Branches are colored according to their mean net-diversification rates, expressed in number of events per million year per lineage (see Eq. S5.2). Orange dots mark nodes whose parent branches showed strong support for a rate shift event (Bayes factor > 10, see Eq. S5.4). Some notable clades including at least one strongly supported rate shift event are labeled on the phylogenies. Silhouettes were obtained from `www.phylopic.org`.

trial animals, flowering plants, ferns, mushroom-forming fungi and lichen-forming fungi. In total, our dataset includes 9 phylogenies of animals, 13 for plants and 2 for fungi. Thus, many of the major groups in the Tree of Life are represented, covering a total diversity of over 300,000 species. The time spans for each phylogeny range from older groups originating at 420 Ma to radiations as recent as 12 Ma, whereas the number of taxa represented ranges from about 250 to 20,000. Overall, this representation of major groups covers the process generating historical and present-day multicellular biodiversity, thus allowing us to learn how variable the process of diversification is.

We assessed heterogeneity in the process of diversification by estimating branch-specific diversification rates and assessing shifts in diversification rates along lineages (Fig. 5.1). To estimate branch-specific diversification rates and shifts in diversification rates, we used a recently developed method called `Pesto` (Kopperud and Höhna, 2025). `Pesto` implements the stochastic birth-death-shift model (Höhna et al. 2019, see also Rabosky 2014; Barido-Sottani et al. 2020), however, it uses a novel and more efficient algorithm to infer branch-specific diversification rates. A more efficient inference algorithm is advantageous in two ways. First, it allows us to rapidly estimate branch-specific diversification rates and diversification rate shift

events for numerous large phylogenies. Second, it makes it feasible to carry out simulation studies with thousands of replicates, helping us learn about the robustness and potential biases of the analyses.

When estimating the number of diversification rate shifts, we found that all phylogenies had significant support for lineage-specific diversification rate variation (Fig. 5.2). Specifically, we used approximate Bayes factors to determine whether there was significant support for a rate shift event on a branch (Shi and Rabosky, 2015; Kopperud and Höhna, 2025). Our estimates show that lineages experienced a shift in diversification rate ranging from as frequent as every 50, to as rare as every 10,000 million years (Fig. 5.3c), corresponding to one shift for every 13 to 285 tips (Fig. S5.2c). These results indicate that the process of diversification is heterogeneous across the Tree of Life, i.e., varying diversification rates among lineages are the norm rather than the exception.

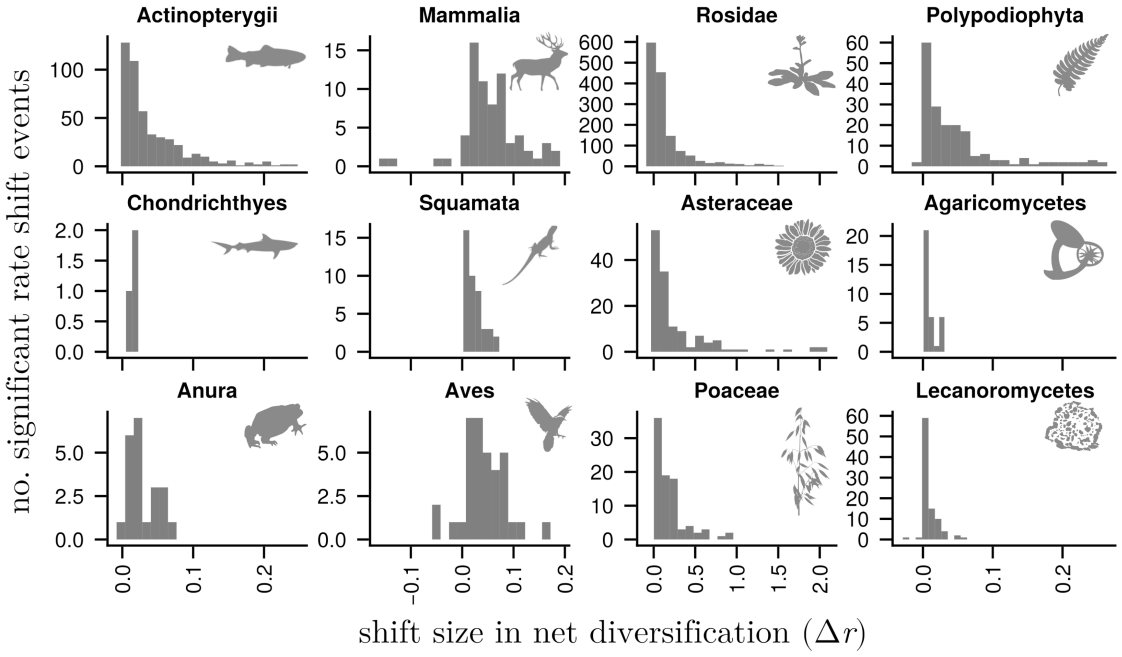


Figure 5.2: Number of strongly supported rate shift events in selected empirical phylogenies. The number of significant diversification rate shift events (N^*) is broken down by the branch-specific change in net-diversification rate (Δr, Eq. S5.3). The represented phylogenies are ray-finned fishes (Rabosky et al., 2018), cartilaginous fishes (Stein et al., 2018), frogs (Portik et al., 2023), mammals (Álvarez-Carretero et al., 2022), lizards and snakes (Zheng and Wiens, 2016), birds (tree from Quintero et al. 2022, original analysis by Jetz et al. 2012), rosids (Sun et al., 2020), daisies (Palazzesi et al., 2022), grasses (Spriggs et al., 2014), ferns (Nitta et al., 2022), agaricomycetes (Sánchez-García et al., 2020) and a large group of lichen-forming fungi (lecanoromycetes, Nelsen et al., 2020). The x-axes are linked for each column. Note that the flowering plants exhibit larger shift sizes than any other groups. See Fig. S5.1 for all phylogenies. Silhouettes were obtained from `www.phylopic.org`.

5.3 Upshifts vs. downshifts in net diversification

When examining individual phylogenies, we discovered a strong imbalance in terms of whether branch-specific rate shift events resulted in an upward or downward change in net-diversification rate. For instance, we inferred 70 strongly supported rate shift events in mammals (Fig. 5.1, left). Of these, 66 were upshifts and only 4 were downshifts. Similarly, we inferred 22 strongly supported rate shifts in frogs, and only one was a downshift (Fig. 5.1, right). Thus, the number of shifts increasing in net-diversification rate vs. decreasing in net-diversification rate appears to be completely imbalanced in favor of increasing net-diversification rates.

Overall, we recovered many upshifts and few downshifts across all of the empirical phylogenies we studied (Fig. 5.2, contrary to Moen and Morlon, 2014; Harmon et al., 2021). Taken at face value, a general tendency of ever-increasing diversification rates would lead to an ever more rapidly growing biodiversity (see also Morlon et al., 2012; Rabosky and Benson, 2021). From a paleobiological perspective, however, we believe that downshifts can and do occur. After all, the fossil record tells us that entire clades have gone completely or nearly extinct (e.g., Benoit et al., 2024).

To assess whether the absence of inferred downshifts is due to a methodological bias, we performed several simulation and inference experiments that are presented in the Supplementary Materials. Specifically, we simulated several phylogenies that had only one rate shift event, and the rate shift event was either an upward or a downward shift (Figs. S5.4 to S5.7). The inferences on the simulated phylogenies revealed that we are able to correctly detect upshifts, as well as to recover constant diversification. However, when there were downshift events in the true phylogenetic history, we were almost never able to infer the events. This coincides with the statistical level of support for upshifts and downshifts in the empirical analyses. In contrast to upshifts, downshifts were in general recovered with moderately strong support (Bayes factor slightly higher than 10), and almost never with decisive support (Bayes factor higher than 100). While our model allows for diversification rates to decrease, we see that inferred rate shift decreases are usually tiny, and spread out over time and different branches. Coupled with our conservative criterion for assessing whether there was a rate shift event on a branch, this results in a methodological bias where downshifts are almost never detected (Figs. S5.7 and S5.8). This lack of power is not surprising. Our data includes speciation events, thus giving power to infer upshifts in net diversification. On the other hand, a downwards shift in net diversification (either due to lower speciation or increased extinction) is fundamentally difficult to infer from reconstructed phylogenies of only extant taxa, as the signal in the data would be a lack of speciation events or increased extinction on a single lineage.

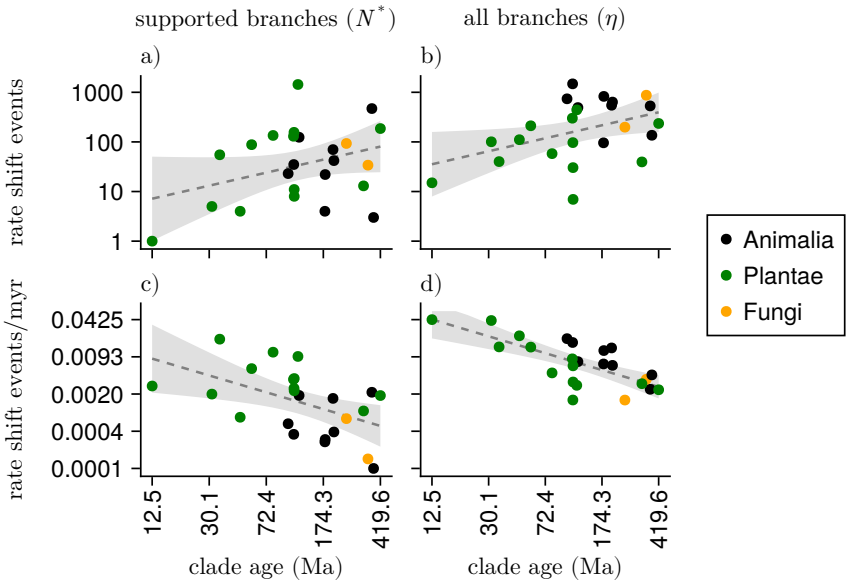


Figure 5.3: Diversification rate shift events vs. clade age in empirical phylogenies. Panels a,c show the number of branches (N^*) that showed strong support (Bayes factor > 10) for there being a rate shift event. Panels b,d show the number of diversification rate shifts according to the shift rate parameter (η). The clade age is the time from the present until the most-recent common ancestor of the group. The dashed lines are ordinary least-squares regressions on the log-transformed variables, and the shaded areas represent two standard error deviations from the line.

5.4 Variation in diversification shift rate across the Tree of Life

Age scaling effect

Next, we explored whether some phylogenies showed more frequent diversification rate shifts than other phylogenies. A priori, we expect species-rich or older phylogenies to have undergone more diversification rate shifts, because they have had more time to accumulate diversification rate shifts than younger phylogenies. We confirm this expectation of older phylogenies having undergone more diversification rate shifts than younger phylogenies (Fig. 5.3a,b). However, our results show a negative age scaling effect in the tempo of diversification rate shifts. Specifically, the number of significant diversification rate shifts per time unit (Fig. 5.3c,d) is negatively correlated with the age of the phylogeny.

At face value, younger groups (i.e., originated at 10 Ma) shift their diversification rates about 100 times more often than older groups (i.e., originated at 400 Ma). Similar age scaling effects have also been demonstrated with other types of evolutionary rates, notably molecular substitution rates (Ho et al., 2007), rates of morphological evolution (Estes and Arnold, 2007; Hansen, 2024) and diversification rates (Henao Diaz et al., 2019; Harmon et al., 2021). There are several possible methodological explanations for why the age scaling effect could occur.

For example, a negative power scaling can arise due to the use of ratios with a common term (Gingerich, 1983; Sheets and Mitchell, 2001), and the effect becomes more pronounced when the range of time intervals being considered is large (De Lisle and Svensson, 2023).

In order to evaluate whether the age scaling pattern is a real pattern, as opposed to being indicative of a methodological artefact, we performed several auxiliary analyses that are presented in the Supplementary Materials. They show that within-phylogeny estimates of diversification rate shifts are higher towards the present (Fig. S5.12), shift rates in very young phylogenies are overestimated (Fig. S5.10a), and sub-sampled clades from a larger phylogeny also exhibit a negative age scaling effect (Fig. S5.13). These analyses indicate that the observed empirical age scaling in diversification rate shifts (Fig. 5.3d) can partially be a consequence of the inference procedure employed for this study. However, when looking only at strongly supported shift events, our simulation study shows that the rate (N^*/t) is consistently underestimated in very young phylogenies (Fig. S5.10b), while it is approximately unbiased for older phylogenies. This result goes in the opposite direction of the empirical age scaling pattern in the rate of strongly supported shift events (Fig. 5.3c). Thus, our observed age-scaling effect indicates a true biological phenomenon which cannot be explained alone by methodological causes.

Variability and tempo in diversification

We additionally observed that the number of diversification rate shifts per time unit is positively correlated to net-diversification rate (Fig. 5.4c,d). In general, plant phylogenies diversify at a higher pace, and experience diversification rate shifts more often, compared to animals and fungi (Fig. 5.4c). This indicates that the variability and tempo of diversification are inherently linked. We conjecture that high diversification rates are associated with high rates of phenotypic evolution (e.g., Rabosky et al., 2013; Folk et al., 2019; Crouch and Tobias, 2022; Cerezer et al., 2023). If phenotypic traits that can induce diversification rate shifts (e.g., body size, ecological niche breadth, life history traits) also evolve more rapidly due to faster diversification, then high diversification rates will lead to more frequent rate shifts. Likewise, if low diversification rates are associated with low rates of phenotypic evolution, then a slowly diversifying clade would have fewer opportunities to evolve phenotypic traits that can shift its diversification rate.

5.5 Shifts in speciation vs. extinction rate

A related question is whether a diversification rate shift event is due to a change in the speciation rate, the extinction rate, or both. In principle, our analyses allow for all three possibilities. In `Pesto`, it is possible to inspect the posterior mean number of rate shifts, and classify them into the three types of rate shifts (see the Supplementary Materials for

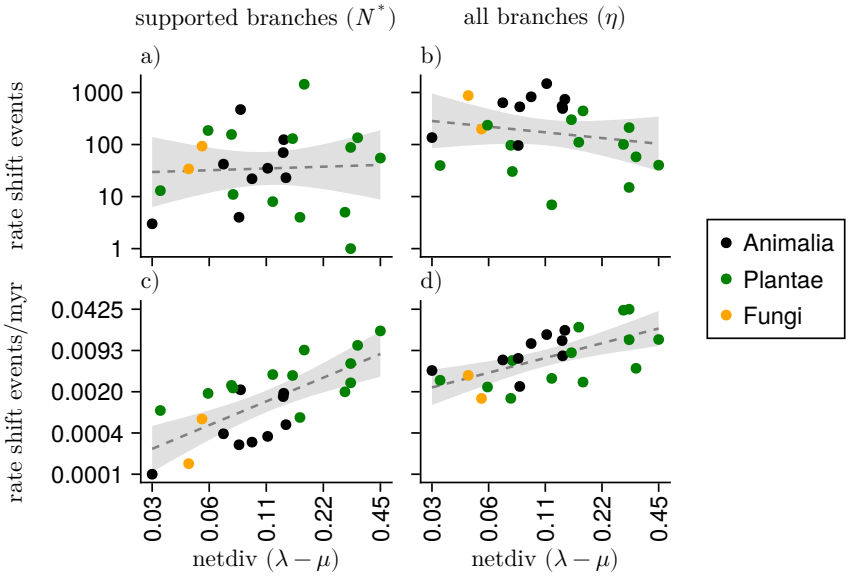


Figure 5.4: Diversification rate shift events vs. net diversification in empirical phylogenies. Panels a,c show the number of branches (N^*) that showed strong support (Bayes factor > 10) for there being a rate shift event. Panels b,d show the number of diversification rate shifts according to the shift rate parameter (η). The net-diversification rate is the average branch-specific rate, weighted by branch length. The dashed lines are ordinary least-squares regressions on the log-transformed variables, and the shaded areas represent two standard error deviations from the line.

more details). Our exploration of the inferred type of diversification rate shift revealed two intriguing aspects. First, the number of estimated joint shifts corresponds rather well to the number expected from the prior (Fig. S5.3), which in turn is determined by the number of diversification rate categories. This is unfortunate, as the discretization procedure was intended to approximate the birth-death-shift process with continuous rate variation. In principle, as the number of rate categories approaches infinity, the discretized model should converge to the fully continuous model. However, as the number of rate categories increase, the number of joint speciation+extinction shifts converges to the total, while extinction and speciation shifts separately converge to zero. Second, we examined the shifts that resulted in a change in the extinction rate. Even though the prior expectation of speciation and extinction rate shifts is equal, we see in the empirical data that speciation rate shifts are inferred more often than expected, and extinction rate shifts are inferred less often than expected (Fig. S5.3). This is in agreement with previous reports of extinction rates being harder or not possible to infer (Nee et al., 1994a; Liow et al., 2010; Rabosky, 2010; Beaulieu and O'Meara, 2015). The challenge we see is also similar to (Barido-Sottani and Morlon, 2025), who found it was difficult to separate whether a rate shift event was due to a change in speciation, extinction or fossilization rates, or whether there was a combination of shifts.

5.6 Caveats in diversification rate estimates

Our estimates of lineage-specific diversification rates and shifts in diversification rates rely on the birth-death-shift process (Höhna et al., 2019; Barido-Sottani et al., 2020) and specifically its implementation in `Pesto` (Kopperud and Höhna, 2025). As with any statistical analysis, the assumptions of the underlying model – and any violations of those assumptions – are key to assessing the robustness of parameter estimates. First, our model uses the conventional assumption for taxon sampling, in that every species in the phylogeny has been sampled with equal probability (also called uniform taxon sampling, Höhna et al., 2011; Höhna, 2014). When selecting phylogenies, we aimed for those that had high taxon sampling, while at the same time trying to avoid artificially grafted trees. The mean taxon-sampling fraction was about 45% across the phylogenies, and the lowest was about 11%. Thus, we believe that most of the phylogenetic analyses are robust with respect to taxon sampling. Second, we assume that speciation and extinction rates are allowed to vary by about one order of magnitude within a phylogeny. While this may be a reasonable assumption on short timescales, it is arguably less reasonable for phylogenies on longer timescales (hundreds of millions of years). Third, our model assumes that diversification rates are constant over time within a category. If in reality an upward diversification shift is followed by a concerted slowdown of all lineages within this clade, then our estimates could be biased. Assuming that diversification rates decay within a rate category is possible (Rabosky et al., 2014b), yet inferring other types of time-varying patterns is considerably more challenging, and allowing for any type of time-varying function may result in the model not being identifiable (Louca and Pennell, 2020b). Fourth, our branch-specific inferences rely on fixed parameter estimates, i.e., the rate of diversification rate shifts η, the mean speciation rate $\hat{\lambda}$ and the mean extinction rate $\hat{\mu}$. We inferred these parameters using maximum likelihood before estimating diversification rate shifts and lineage-specific diversification rates. Thus, if the shift rate η is overestimated, this can lead to an overestimated number of strongly supported diversification rate shifts. For example, broomrapes (Orobanchaceae Mortimer et al., 2022), daisies (Asteraceae, Palazzesi et al., 2022) and rosids (Rosidae, Sun et al., 2020) had the most extreme estimates with about 0.01 number of strongly supported rate shift events per million years (Fig. 5.3c). However, since all phylogenies were treated equally, we believe that no systematic bias was introduced. Finally, our analyses relied on published phylogenies, which we treated as if they were known without error. Although this is the common approach in large-scale diversification rate inferences (e.g., Quintero et al., 2024), the quality of larger phylogenies might be worse as more ad-hoc procedures are required to place taxa and infer divergence times, e.g., placing taxa based on taxonomy without molecular or morphological data (Smith and O'Meara, 2012).

5.7 Conclusions

Biodiversity has changed over time and across lineages. Our results show that the net-diversification rate varies among lineages for every single analyzed phylogeny. We examined how diversification rates vary across many large-scale phylogenies, spanning major groups of multicellular organisms. Younger phylogenies appear to have experienced diversification rate shifts more often than older phylogenies, similar to how rates of molecular and phenotypic evolution scale with time interval. Furthermore, we found more diversification rate shifts in clades with generally higher diversification rates, indicating general patterns of lability (higher number of diversification rate shifts together with higher number of speciation events) and conservatism (stable or low numbers of diversification rate shifts together with fewer speciation events). Overall, we inferred many rate shift events where net diversification increased, indicating that rapid radiations are common in contrast to 'living fossil' lineages. Understanding whether net-diversification rates tend to shift upwards or downwards is complicated by the fact that downward shifts are extremely difficult to detect. Finally, our results provide the foundations for future studies on the nature and prevalence of diversification rate shifts in exceptionally species-rich clades.

Chapter 6

Commonly used Bayesian diversification methods lead to biologically meaningful differences in branch-specific rates on empirical phylogenies

Bonsai is the Japanese art of growing a miniature tree. By trimming, pruning and grafting, the artist introduces shifts in the branching pattern.

6.1 Abstract

Identifying along which lineages shifts in diversification rates occur is a central goal of comparative phylogenetics; these shifts may coincide with key evolutionary events such as the development of novel morphological characters, the acquisition of adaptive traits, polyploidization or other structural genomic changes, or dispersal to a new habitat and subsequent increase in environmental niche space. However, while multiple methods now exist to estimate diversification rates and identify shifts using phylogenetic topologies, the appropriate use and accuracy of these methods is hotly debated. Here we test whether five Bayesian methods—Bayesian Analysis of Macroevolutionary Mixtures (`BAMM`), two implementations of the Lineage-Specific Birth-Death-Shift model (`RevBayes` and `Pesto`), the approximate Multi-Type Birth-Death model (`MTBD`; implemented in `BEAST2`), and the cladogenetic diversification rate shift model (`CLaDS2`)—produce comparable results. We apply each of these methods to a set of 65 empirical time-calibrated phylogenies and compare inferences of speciation rate, extinction rate, and net diversification rate. We find that the five methods often infer different speciation, extinction, and net-diversification rates. Consequently, these different estimates may lead to different interpretations of the macroevolutionary dynamics. The different estimates can be attributed to fundamental differences among the compared models. Therefore, the inference of shifts in diversification rates is strongly method-dependent. We advise biologists to apply multiple methods to test the robustness of the conclusions or to carefully select the method based on the validity of the underlying model assumptions to their particular empirical system.

6.2 Lay summary

Understanding why some groups of organisms have more species than others is key to understanding the origin of biodiversity. Theory and empirical evidence suggest that multiple distinct historical events—such as the evolution of particular morphological features (e.g., the flower, the tetrapod limb) and competition amongst species—can produce this pattern of divergent species richness. Identifying when and where on the tree of life shifts in diversification rates occur is important for explaining the origin of modern-day biodiversity and understanding how disparity among species evolves. Several statistical methods have been developed to infer diversification rates and identify these shifts. While these methods each attempt to make inferences about changes in the tempo of diversification, they differ in their underlying statistical models and assumptions. Here we test if these methods draw similar conclusions using a dataset of 65 time-calibrated phylogenies from across multicellular life. We find that inferences of where rate shifts occur strongly depends on the chosen method. Therefore, biologists should choose the model whose assumptions they believe to be the most valid and justify their model choice *a priori*, or consider using several independent methods to test an evolutionary hypothesis.

Key Words: diversification-rate analyses, Phylogeny, BAMM, RevBayes, BEAST, ClaDS2, macroevolution

6.3 Introduction

Understanding the patterns and processes that shape the tree of life is one of the central pursuits of biology. However, inferring the tempo of evolution among lineages—the patterns of speciation and extinction that gave rise to our extant biodiversity—remains a difficult problem both theoretically and computationally (Rabosky, 2010; Moore et al., 2016; Louca and Pennell, 2020b).

Several methods estimate diversification rates (speciation and extinction rates, individually) assuming that rates are constant across the tree (Morlon, 2014). Recently developed methods have built upon constant-rate models by allowing diversification parameters to vary depending on the state of a focal character (Maddison et al., 2007) or, even more recently, among branches of the phylogeny, which allows for lineage-specific diversification rate estimates (*e.g.*, Rabosky, 2014; Höhna et al., 2019; Barido-Sottani et al., 2020; Maliet and Morlon, 2022).

Such lineage-specific methods have the potential to offer powerful insights into our understanding of evolution, such as the potential time-dependency of macroevolutionary diversification (Henao Diaz et al., 2019), the macroecological and macroevolutionary causes of the latitudinal diversity gradient (Givnish et al., 2018; Rabosky et al., 2018), and macroevolutionary support of Darwinian and Simpsonian theories of microevolution within adaptive zones (Cooney et al., 2017).

The application of diversification-rate estimation methods, however, has been marred by controversy over their implementation (Moore et al., 2016; Rabosky et al., 2017; Meyer and Wiens, 2018; Meyer et al., 2018; Rabosky, 2018) and by theoretical findings that seemingly undermine the general reliability of inferring diversification parameters from phylogenies of extant species (Louca and Pennell, 2020b; Helmstetter et al., 2022; Kopperud et al., 2023). These issues are liable to discourage empiricists, who may wonder if the disagreements among model developers and theorists correspond with biologically relevant inference differences in empirical studies.

To address this question, we assess how inferences under five leading contemporary Bayesian methods—Bayesian Analysis of Macroevolutionary Mixtures (`BAMM`; Rabosky, 2014); the Lineage-Specific Birth-Death-Shift model (`RevBayes`; Höhna et al., 2019) and its MCMC-free implementation: Phylogenetic Estimation of Shifts in the Tempo of Origination (`Pesto`; Kopperud and Höhna, 2025); the approximate Multi-Type Birth-Death model (`MTBD`; Barido-Sottani et al., 2020); and the Cladogenetic Diversification Rate Shift model (`CLaDS2`; Maliet et al., 2019)—compare to each other.

While all five methods aim to estimate lineage-specific diversification rates, they differ in how and where rate shifts are allowed to occur.

1. `BAMM` models diversification rates as varying across lineages by testing among models that include different numbers of diversification-rate regimes (sets of speciation and

extinction parameters) and different placements of those regimes in the tree; however `BAMM` does not model rate shifts on extinct (thus unobserved) branches (Rabosky, 2014).

2. The `RevBayes` model, as implemented in `RevBayes` (Höhna et al., 2016a), samples rate regimes from a prior distribution discretized into a fixed number of rate categories; this discretization facilitates computation and allows the method to model shifts on extinct branches (Höhna et al., 2019).

3. `Pesto` is a new implementation of the `RevBayes` model that analytically computes the posterior mean speciation and extinction rates conditional on a set of hyperparameters without the need for Monte Carlo sampling (Kopperud and Höhna, 2025).

4. The `MTBD` method is based on a multitype birth-death process that infers the number of rate regimes as well as the transition rate γ between rate regimes (Barido-Sottani et al., 2020). This approach allows for the same rate regime to be present in different parts of the tree. The approximate `MTBD`, tested here, assumes that no rate changes occur in the extinct parts of the tree; this approximation, when applied with a high transition rate prior, has been found to not substantially differ from the exact `MTBD` method, which allows rates changes along extinct lineages (Barido-Sottani et al., 2020).

5. Finally, in the `CLaDS2` model, diversification rates only change at speciation events. Descendant lineages inherit the speciation rate via a stochastic process that is influenced by the α parameter, which represents the long-term trend (i.e., increase or decreases) of the speciation rate (Maliet et al., 2019). This model results in many small and frequent shifts in diversification rates regimes, unlike the other methods, which tend to infer a few large shifts in rate regimes (Maliet et al., 2019; Maliet and Morlon, 2022). Another aspect of `CLaDS2` is that extinction rates are not inferred per branch. Instead, the model estimates a global turnover parameter ($\epsilon = \mu_i/\lambda_i$). However, shifts are allowed to occur along extinct branches.

Other methods, not tested here, leverage hidden states using a maximum likelihood framework (*e.g.*, Vasconcelos et al., 2022).

To assess whether the theoretical and computational differences among these methods result in biologically meaningful differences, we reanalyze 65 empirical datasets, compiled from Henao Diaz et al. (2019), using each of `BAMM`, `RevBayes`, `Pesto`, `MTBD`, and `CLaDS2`. We address the question: do different analytical methods for estimating branch-specific diversification rates produce significantly different results across an array of empirical datasets?

6.4 Methods

6.4.1 Empirical Data

Our empirical data are derived from the set of 104 chronograms compiled and analyzed with `BAMM` by Henao Diaz et al. (2019). From the Henao Diaz et al. set we excluded trees with fewer than 30 extant taxa in order to concentrate on more informative datasets, resulting in our final set of 76 chronograms.

6.4.2 Model Settings

Our goal was to apply each method as a typical diligent user might. For each chronogram, we used the incomplete-sampling fraction collected from the original study by Henao Diaz et al. (2019), and applied that sampling fraction when we ran each of the five inference methods. While the methods differ in their specific parameterizations of the birth-death process, we attempted to use comparable settings and priors across methods.

For `BAMM` analyses, we modified the control files from Henao Diaz et al. (2019). We set lambda to be time-constant rather than time-variable in order to more closely match the inferences of other methods. We set `BAMM` priors for each phylogeny using the `setBAMMpriors()` function in the `BAMMtools R` package (Rabosky et al., 2014b). This function computes dataset-specific priors by estimating metrics from the dataset such as the root age of the chronogram and then estimating reasonable and broad expectations for shifts and rates. We ran `BAMM` v2.5.0 using the BAMMtools priors and control files, which determined the phylogeny-specific number of generations for a single MCMC chain. We removed the first 10% of the MCMC samples as burnin and assessed convergence by computing estimates of effective sample size (ESS) using the R package Coda (Plummer et al., 2006). We specifically looked for convergence of the log-likelihood parameter and the 'number of distinct regimes' parameter, as is recommended (Rabosky et al., 2014a). Analyses that did not reach convergence were run for additional generations until they converged.

For `RevBayes` analyses we used the same set of priors for all phylogenies (except for sampling fraction) with eight categories for speciation and for extinction (64 total rate categories). The number of rate categories was chosen after performing a test on one representative phylogeny, which found that increasing the rate categories above 64 did not result in a significant change in model fit. For each chronogram we ran four MCMC chains for 5,000 generations. Convergence was assessed for each chain by checking that the ESS values for all model parameters in the log files were greater than 200 using the R package Coda (Plummer et al., 2006). Chains that did not reach convergence were restarted and run for an additional 5000 generations. We merged the posteriors, retaining the last 4000 generations from the MCMC (10% burnin for non-restarts and 60% burnin for restarts).

We applied `Pesto` in a three-step fashion. First, we estimated the parameters of a constant-rate birth-death process and treated these as hyperparameters: the speciation rate (λ) and the extinction rate (μ). Second, we set up a state-dependent speciation-extinction (SSE) model. In this model, we used rate values that correspond to ten quantiles of two lognormal distributions with medians λ and μ, and standard deviation 0.587. In the SSE model, we used all pairwise combinations of these (i.e. 100 rate categories). Further, we fixed the shift rate parameter (η) such that we would expect 10 rate shifts across the phylogenies (i.e. $\eta = 10/\text{tree-length}$). Third, we calculated the posterior state probabilities along each branch. Finally, we plotted the posterior mean rates averaged over the time span for each individual branch.

We ran the `MTBD` model under default priors (implemented in `BEAST2`; Bouckaert et al., 2014; Barido-Sottani et al., 2020). We ran three MCMC chains for 100,000,000 generations per phylogeny. We removed the first 25% as burnin and assessed MCMC convergence by checking that ESS values were higher than 200 for all rates.

We ran `CLaDS2` using the default priors (as described in Maliet and Morlon, 2022). We ran three MCMC chains for each dataset and took a 25% burnin, as is the default setting for `CLaDS2`. We assessed convergence by calculating the Gelman statistic (Gelman et al., 2013) every 1000^{th} generation and stopping the analysis once it achieved a Gelman statistic of 1.05, following the standard guidelines for `CLaDS2`. While we set up our `BAMM` analyses to be time constant, `CLaDS2` does not allow users to fix the inherited speciation-rate parameter (α). Thus our analyses estimate α and infer trends of increasing or decreasing speciation rate through time.

All methods are conditioned on "survival:" the occurrence of the most recent common ancestor and survival of the two descendent lineages (Nee et al., 1994b). However, there are method-specific differences in the details how methods condition for different rate-shift scenarios that could potentially contribute to different rate estimates across methods.

Convergence analysis

In cases where MCMC convergence was difficult, we aimed to determine the potential underlying cause. To assess whether the subset of trees where one or more method failed to converge was substantially different from the subset that did converge, we compared descriptive metrics including phylogeny size, phylogeny age, incomplete sampling fraction, branch length variance, and multidimensional scaling (MDS) via Robinson-Foulds (RF Robinson and Foulds, 1981) and Kuhner-Felsenstein (KF, Kuhner and Felsenstein, 1994) distances.

Processing Model Output

We obtained estimates of the relevant diversification parameters (*e.g.*, speciation rate, extinction rate, etc.) from each model. `BAMM` posterior estimates of speciation rate and extinction

rate were extracted using the `getMarginalBranchRateMatrix()` function in the `BAMMtools` `R` package (Rabosky et al., 2014b).

We extracted LSBDS posterior distributions from the stochastic branch rate log file produced by the `mnStochasticBranchRate()` function in `RevBayes`. In the `Pesto` analyses, we computed the branch rates averaged across the branch. If λ_k and μ_k are the rate values in state k, and $P_k(t)$ is the posterior probability of being in state k at time t, then the average net-diversification rate along a branch is

$$\frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \big[\sum_k (\lambda_k - \mu_k) P_k(t) \big] dt, \tag{1}$$

where t_0 is the youngest and t_1 is the oldest end point of the branch.

The posterior distributions of speciation and extinction rates of the MTBD model were obtained from the extended Newick file produced by BEAST2 using a modified read.beast() function from the treeio package (Wang et al., 2020). As CLaDS2 does not directly infer extinction rates, we calculated extinction rates per branch by multiplying the inferred global turnover value (ϵ) by the branch-specific speciation rates ($\mu_i = \lambda_i * \epsilon$). For all branches and models, we calculated net diversification by subtracting extinction rate from speciation rate ($\lambda_i - \mu_i$) per MCMC generation.

6.4.3 Comparing Model Inferences

To compare inferences among the five models, we (1) visualized rate estimates on individual chronograms, (2) summarized inferences across all chronograms in the dataset to reveal systematic differences, (3) identified differences in the location and magnitude of inferred shifts among methods, and 4) tested for overlap in the 95% HPD interval of the posterior distributions (described in Appendix S6.4).

Visualizing rates on trees

The canonical way of presenting the results of branch-specific diversification-rate analyses is by coloring the branches of the tree by the estimated rates. For each tree, we colored each branch by the posterior median estimate of speciation, extinction, and net diversification to visualize if the methods inferred similar shifts in similar locations on the tree.

Comparing rate estimates by method

To understand whether the methods displayed any consistent differences across the chronograms, we calculated six summary statistics for each tree. For each diversification rate (*i.e.*, speciation rate, extinction rate, and net diversification rate) we calculated the posterior medians for each branch, and from those posterior medians we calculated the tree-wide mean and

variance in branch rates for each phylogeny. For each of the six summary statistics (mean and variance for each of the three rates), we set up a linear mixed-effect model:

$$\log(\text{summary statistic}) = \mathbf{X}\beta + \mathbf{Z}u + r, \tag{2}$$

with inference method as a fixed-effect categorical predictor (effect sizes β), phylogeny as a random effect categorical predictor (u), and an error term r. $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for the fixed and random effects. We visually checked that the residuals (r) were normally distributed and did not suffer from heteroscedasticity; phylogenies that violated these assumptions were excluded from this analysis. For each linear model, we tested if the least-square means of each pair of methods were statistically different using Tukey's corrected p-value for multiple comparisons.

Location and magnitude of rate shifts

We additionally tested whether the methods inferred consistent locations and magnitudes of rate shifts, using the rooted version of the Kuhner-Felsenstein distance (Kuhner and Felsenstein, 1994). To do this, we first replaced branch lengths of each timetree with the posterior median rate estimate, from a given method, then scaled each branch by the total tree height. This produces a method-dependent tree with branch lengths that provide information regarding the magnitude and location of rate shifts but with identical topology. We calculated KF distances between the rescaled trees from each pair of methods; this distance is equivalent to the mean square error (MSE) given that the two trees being compared have the same topology, as they do in our analyses. For each tree and for each diversification parameter, we computed the mean square error among the different methods:

$$\text{MSE} = \frac{1}{N}\sum_{i}^{N}(\lambda_i - \lambda_i')^2, \tag{3}$$

where λ_i (or similarly μ_i, or $(\lambda_i - \mu_i)$) is the scaled diversification rate parameter for branch i.

A large MSE tells us that the two methods being compared infer different rate magnitudes and/or rate shifts in different locations. A small MSE, however, indicates that the two methods give us similar results.

Diversification rates may vary according to clade age, potentially due to sampling (Henao Diaz et al., 2019; Louca et al., 2022); in particular, younger clades may exhibit faster diversification rates. To pull apart the potentially compounding effects of method and clade age on our results, we also we tested whether the method-specific patterns we observe still hold for phylogenies of different ages by splitting our dataset into older and younger phylogenies (where older and younger trees are defined as having a root age greater or less than the median root

age) and repeating the above-described procedure for both groups.

6.4.4 Computation

We ran all diversification analyses either locally, on the Savio HPC at UC Berkeley, or using the CIPRES Science Gateway V. 3.3 (Miller et al., 2010).

We performed all comparison analyses in `R` version 3.6.0 (R Core Team, 2020). We performed data manipulation with the R packages `phytools`, (Revell, 2012), `tidyverse` (Wickham et al., 2019), `reshape2` (Wickham, 2012), `readr` (Wickham and Hester, 2020), `plyr` (Wickham, 2011), and `coda` (Plummer et al., 2006). We generated plots with R packages `see` (Lüdecke et al., 2021), `ggplot2` (Wickham, 2016), `ggpubr` (Kassambara, 2018), `ggtree` (Yu et al., 2018), `ggsignif` (Ahlmann-Eltze, 2017), `ggExtra` (Attali and Baker, 2016), `cowplot` (Wilke, 2016) and `pdftools` (Ooms, 2020). We fit linear mixed models using the `R` package *lmer* (Bates et al., 2015) and obtain emeans estimates using the R package `eemeans` (Lenth, 2020). We additionally used `smacof` (Mair et al., 2022) and `phangorn` (Schliep, 2011) to perform MDS and to calculate RF and KF distances. Citations for R packages were generated with `RefManageR` (McLean, 2014).

6.5 Results and discussion

6.5.1 Convergence

Our full dataset contains 76 chronograms from multicellular organisms, with 31 – 4161 extant tips, root ages of 4.9 – 349.8 MYA, and 0.014% – 100% of extant species sampled (Fig. S6.1A and Table S6.1). All methods converged for 43 trees (the "complete subset"; Fig. S6.1B). Of the methods tested, `RevBayes` had the most difficulty achieving MCMC convergence (it converged for 46 trees). All methods except `RevBayes` converged in 65 trees (the "partial subset"; Fig. S6.1C); `Pesto` directly computes the posterior mean and thus "convergence" does not apply. Trees that did not converge have poorer taxon sampling (i.e. the ratio of sampled species to total species richness; P-value = .039), older root ages (P-value = 0.0001), and greater branch length variance (P-value = .00006) than the converged trees, but sample size (number of tips) was not an important factor (P-value = .076; Fig. S6.2C–F). The branch length variance is consistent with the degree of spread between the KF and RF MDS analyses (Fig. S6.2A–B); the KF MDS—which accounts for branch lengths as well as topology—has a larger spread then the RF MDS. Overall these results fit with our intuitive understanding of the challenges in inferring shifts in diversification rates. We expect that older trees and trees with greater variation in branch lengths should undergo more rate shifts than younger trees and those with less variation in branch lengths. Thus inferring the diversification-rate parameters of these trees should be generally more challenging for the MCMC. These results suggest that users should be particularly attentive to MCMC convergence if their chronogram(s) are

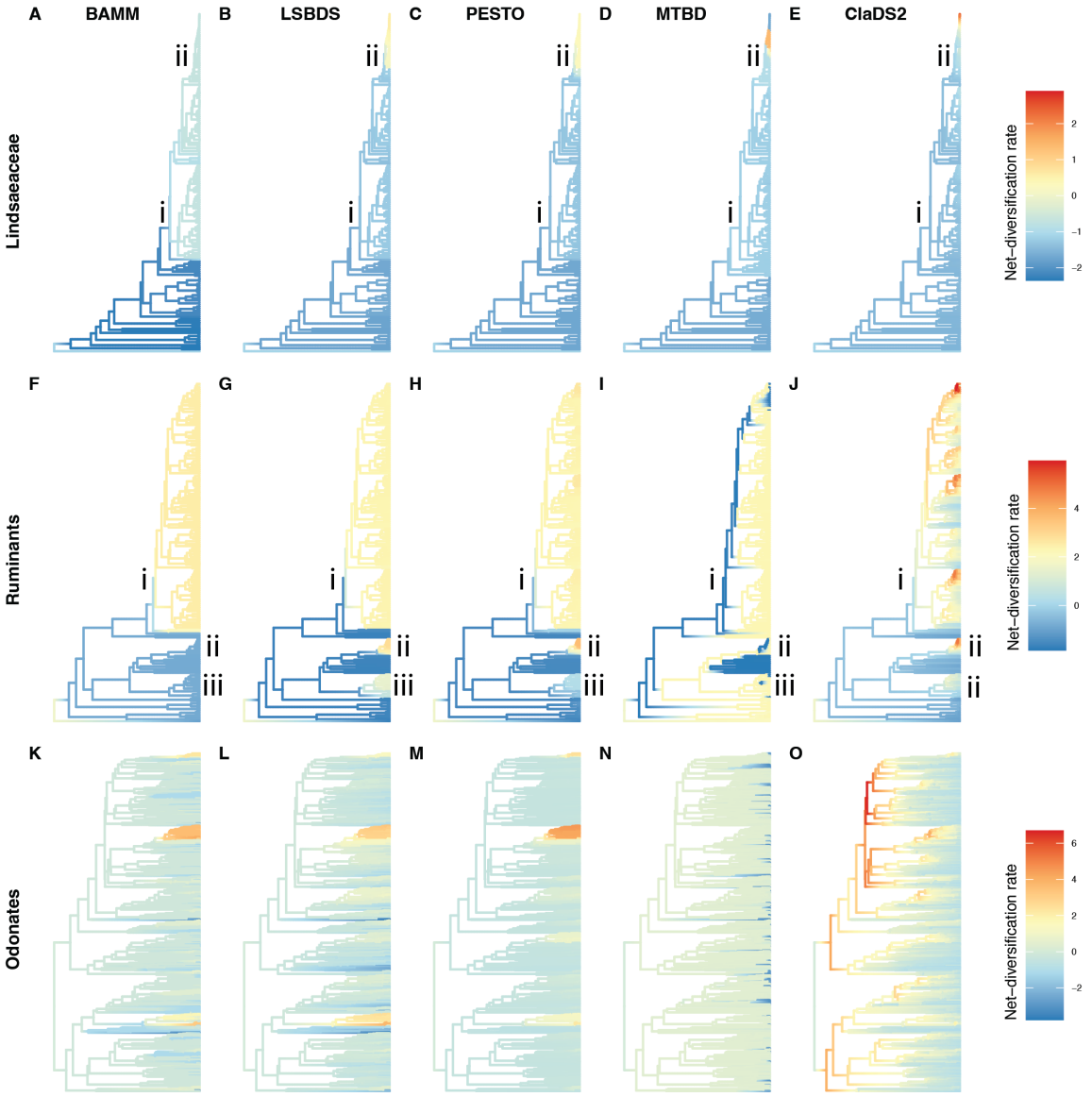


Figure 6.1: Three representative phylogenies with Z-transformed (mean centering and scaling to unit variance) posterior median estimate of net diversification painted on the branches. Columns show estimates from `BAMM` (A,F,K), `RevBayes` (B,G,L), `Pesto` (C,H,M), `MTBD` (D,I,N) and `CLaDS2` (E,J,O). (A-E) Phylogeny of Lindsaeaceae (necklace ferns; Testo and Sundue, 2016), (F-J) Phylogeny of Ruminants (tetrapod; Toljagić et al., 2018), and (K-O) Phylogeny of Odonates (dragonflies and damselflies; Waller and Svensson, 2017). The rate values are in units of events per lineage per million years.

poorly sampled, old, or have a lot of branch-length variation, and especially so if they are using `RevBayes`. In these cases, even more so than usual, it is important to run each MCMC multiple times independently, to assess both stationarity and convergence.

6.5.2 Comparison of Methods

Visualizing rates on trees

None of the 43 phylogenies in our "complete subset" had concordant estimates among all methods given our evaluation criteria (Fig. 6.1). For some phylogenies, the methods inferred similar shifts in net diversification (*e.g.*, Fig. 6.1A–E), whereas for others the inferred shifts differed slightly (*e.g.*, Fig. 6.1F–J) or strongly (*e.g.*, Fig. 6.1K–O).We would expect some differences when comparing different modeling approaches, as there are patterns to the differences in our results that can be attributed to the fundamental differences between the models. We illustrate these patterns using a few example phylogenies, which are representative of the patterns one will find when perusing the full set of trees in the supplemental materials (Appendix S6.5).

Occasionally two methods generally identified similar patterns. For example, `BAMM` and `RevBayes` identified a similar shift of about the same magnitude in speciation rates for some clades, *e.g.*, the Lindsaeaceae (necklace ferns, clade i; Fig. 6.1A,B). Nonetheless, there are still differences between the two methods, *e.g.*, a second nested rate shift in the `RevBayes` and `Pesto` estimates (clade ii).

In the ruminants (tetrapods) phylogeny (Fig. 6.1F–J), we find that even for results that overall appear similar between methods, there are meaningful differences between their estimates. For example, `BAMM`, `RevBayes`, and `Pesto` inferred a shift around the ancestor of clade i, but `RevBayes` and `Pesto` also find approximately two more shifts (Fig. 6.1G, clades ii and iii). Likewise, `MTBD` differs from the latter two as it infers several shifts in the largest clade and low net diversification rates on the backbone of that lineage (Fig. 6.1I). Similarly, `CLaDS2` infers a slightly different history from all of them, including multiple slow downs as well as an increase in net-diversification within clade i. `BAMM`, `RevBayes`, and `Pesto` identify a shift in approximately the same node (indicated by i) while `MTBD` infers many replicated increases in rate within clade i. `RevBayes` and `Pesto` infer the same shift, which is expected as they are based on the same underlying model and assumptions.

Multiple diversification shifts across a phylogeny is common to many of the `MTBD` trees (Fig. 6.1I,N; Barido-Sottani et al., 2020). This pattern is caused by the bimodal posterior distribution commonly inferred by this method (Barido-Sottani et al., 2020). Point-estimate summary statistics (*e.g.*, posterior median) of these types of distributions are susceptible to small variation between the ancestor-descendant branches, which causes point estimates to switch between the two optima producing the rapid switching pattern (Fig. 6.1I,N).

Likewise, `CLaDS2` is the only model in our analysis that directly models trends in rates through time. The inherited speciation rate (α) determines if daughter lineages inherit generally faster or slower rates than the parental lineage. (α) only changes at cladogenic (speciation) events, which results in many small changes at cladogenetic events, rather than the few large changes that characterize the other methods (Maliet et al., 2019; Maliet and Morlon, 2022).

When $\alpha < 1$ evolutionary slowdowns occur where the ancestral lineages have higher net diversification rates than the descendant lineages, a pattern observed in our data (Fig. 6.1O; Moen and Morlon, 2014).

On the other hand, `BAMM`, `RevBayes`, and `Pesto` are similar models and therefore we may expect them to infer similar diversification rates and shifts (Ronquist et al., 2021). While this is sometimes true (*e.g.*, `BAMM` and `RevBayes` agree: Fig. 6.1 K,L), other times there are pronounced differences (*e.g.*, `BAMM` is different from both `RevBayes` and `Pesto`: Fig. 6.1F,G). Differences between `BAMM` and the other two models may be due to the well-known differences between the models, namely assuming either rate shifts can (`RevBayes` and `Pesto`) or cannot (`BAMM`) occur on extinct lineages or unsampled lineages (Moore et al., 2016).

Likewise, `RevBayes` and `Pesto` are not always in agreement (*e.g.*, Fig. 6.1L,M). While they are mathematically equivalent models, their inferences may differ for two reasons: (1) The overall method of parameter inference and prior distributions. LSBDS infers the mean speciation rate, mean extinction rate, and the shift rate jointly using full Bayesian inference via MCMC using stochastic character mapping. In contrast, `Pesto` uses empirical Bayes hyperparameters (similar to `BAMM`), where the mean speciation and extinction rates are first estimated using a constant-rate birth-death process. This means that branch rates estimated using `Pesto` are conditional on the parameters instead of integrating over the parameter values in the MCMC as in `RevBayes`. (2) We used a fixed shift rate in `Pesto`, ($\eta = 10$/tree length) instead of estimating it from the data. While the estimates of the shift rate in `RevBayes` are sensitive to the shift rate prior, estimates of branch-specific speciation and extinction rates are more robust to their priors (Höhna et al., 2019). We argue that the inference method (empirical Bayes vs. full Bayesian hierarchical model) between `Pesto` and `RevBayes` as well as using 64 vs. 100 rate categories largely explains the differences in the branch-specific rate estimates.

Comparisons of rate estimates by method

To gain a global perspective of the differences between these models, we calculated two tree-wide summary statistics and distance metrics in order to compare these methods across the entire dataset.

We ran all comparisons on the complete subset (the 43 trees that converged for all methods; Fig. S6.4) and on the partial subset (the 65 trees that converged for all methods except `RevBayes`; Fig. 6.2). Comparisons between these two subsets reveal only one small difference (*e.g.*, compare BAMM vs. MTBD—Fig. 6.2C vs. Fig. S6.4C), and the most significant differences did not change. Given the large number of datasets that did not converge for `RevBayes` but converged for all other methods (unconverged datasets = 22) as well as the theoretical similarities between `Pesto` and `RevBayes`, we report the following results for the partial dataset (but see Fig. S6.4A–F for summaries from the complete dataset).

Table 6.1: Post-hoc pairwise comparisons of inference methods using the tree-wide average of summary statistics: speciation, extinction, and net-diversification rates. Columns contain the summary statistics, contrasts of inference methods, the ratios of geometric means, standard errors, degrees of freedom, t-ratios, Tukey-adjusted p-values, significances, and the percent variances explained by the random effect.

Summary Statistic	Contrasts	Means Ratio	SE	DF	T-Ratio	Adj. P-Value	Sig.	% Var.
Speciation	BAMM / ClaDS2	1.0114	0.0336	189	0.3402	0.986	N.S.	94.57
	BAMM / PESTO	0.8833	0.0294	189	-3.7326	0.001	**	
	BAMM / MTBD	1.0352	0.0344	189	1.041	0.726	N.S.	
	ClaDS2 / PESTO	0.8733	0.029	189	-4.0728	0.000	***	
	ClaDS2 / MTBD	1.0236	0.034	189	0.7009	0.897	N.S.	
	PESTO / MTBD	1.172	0.039	189	4.7736	0.000	***	
Extinction	BAMM / ClaDS2	1.2616	0.3006	189	0.9752	0.764	N.S.	58.65
	BAMM / PESTO	3.0505	0.7268	189	4.6808	0.000	***	
	BAMM / MTBD	1.119	0.2666	189	0.4718	0.965	N.S.	
	ClaDS2 / PESTO	2.418	0.5761	189	3.7056	0.002	**	
	ClaDS2 / MTBD	0.887	0.2113	189	-0.5034	0.958	N.S.	
	PESTO / MTBD	0.3668	0.0874	189	-4.209	0.000	***	
Net Diversification	BAMM / ClaDS2	0.6467	0.0406	188.067	-6.9392	0.000	***	81.83
	BAMM / PESTO	0.6666	0.0419	188.067	-6.4555	0.000	***	
	BAMM / MTBD	0.8278	0.052	188.067	-3.0073	0.016	*	
	ClaDS2 / PESTO	1.0309	0.0644	188.0003	0.4863	0.962	N.S.	
	ClaDS2 / MTBD	1.2802	0.08	188.0003	3.9523	0.001	***	
	PESTO / MTBD	1.2419	0.0776	188.0003	3.466	0.004	**	

We recover consistent differences in rate estimates among methods, particularly between `Pesto` and all other models; `CLaDS2` also was an outlier, albeit to a lesser extent (Tables 6.1 and S2). In contrast, `BAMM` and `MTBD` tended to infer similar speciation and extinction rates. We find that tree-wide average speciation and extinction estimates of `Pesto` are statistically different from all other methods (Fig. 6.2A-B).

While `Pesto` inferred higher tree-wide average speciation values, the magnitude of the differences is small (ratio of means < 1.2 for all significant contrasts; Table Table 6.1). Conversely, `Pesto` inferred lower tree-wide averages of extinction rates with larger magnitude changes (ratio of geometic means > 1.2; Table Table 6.1). The significant difference between `Pesto` and other methods holds for tree-wide average net-diversification as well, except for the comparison between `Pesto` and `CLaDS2` (Fig. 6.2C), which is not significant.

Additionally, `CLaDS2` tree-wide average net-diversification estimates are significantly different from `BAMM` and `MTBD` (Fig. 6.2C). A significant difference in net-diversification could be driven by the `CLaDS2` parameterization of extinction: extinction is not directly estimated in `CLaDS2`. Therefore the net diversification rates of `CLaDS2` are scaled speciation rates. Alternatively, the differences between methods could be due to the wider variance of net diversification estimates that both `BAMM` and `MTBD` have compared to `CLaDS2` (Fig. S6.3A–C). However, similar to tree-wide average speciation, the magnitude of difference between the contrast is not large (Table Table 6.1). There is also a weakly significant difference between speciation rates of `BAMM` and `MTBD` in our partial subset that was not found in the smaller complete subset. Despite these differences, the methods tend to agree on the top three datasets with fastest net-diversification rate (Fig. S6.5, generally inferring that the same set of lineages are evolving

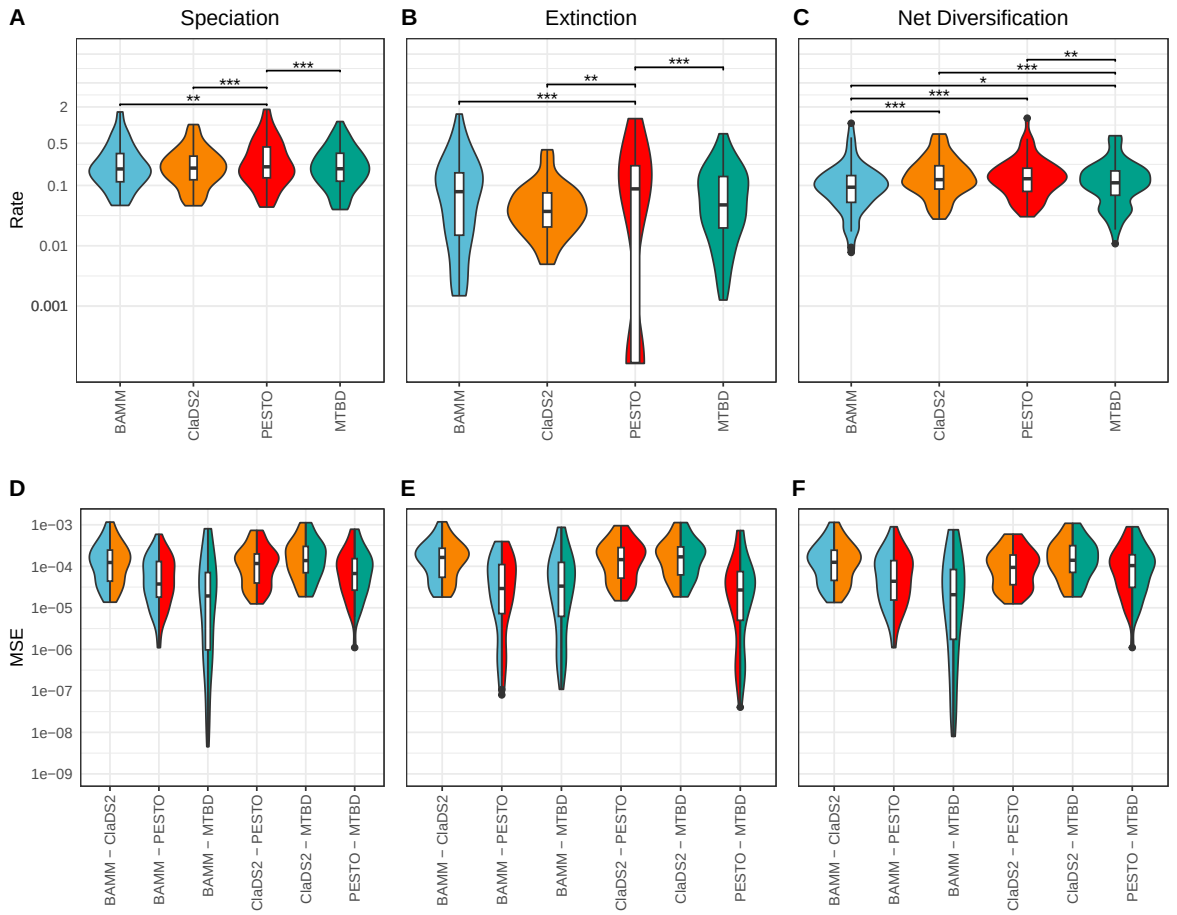


Figure 6.2: Comparison of tree-wide summary statistics across methods for the partial subset (n=65). (A–C) Tree-wide mean of posterior median estimates of the branch-specific rate parameters, plotted on a log scale. Asterisks correspond to the p-value of linear mixed model, calculated on the natural log of the rates (*: 0.05 > P-value > 0.01; **: 0.01 > P-value > 0.001; ***: 0.001 > P-value). (D–F) Pairwise mean squared error (MSE) between inference methods of phylogenies with branch lengths scaled by rates (speciation, extinction, and net diversification), plotted on a log scale. Split colors correspond to inference method color in A–C. Distributions closer to zero indicate that the inference methods produced more similar rate estimates, whereas higher values indicate greater dissimilarity. (D) MSE of speciation-scaled phylogenies; (E) MSE of extinction-scaled phylogenies; (F) MSE of net-diversification-scaled phylogenies.

remarkably rapidly.

All methods generally had comparable tree-wide average extinction-rate estimates with the exception of `Pesto`, which sometimes infers much lower extinction rates for some trees than the other methods (though, on average, it infers higher extinction rates). The inference of extinction rate has been the subject of substantial debate, particularly in how failures to account for diversification shifts along extinct branches can impact the likelihood function (Moore et al., 2016; Rabosky et al., 2017). Regardless of the theoretical importance of correctly inferring extinction rates, we demonstrate that differences between extinction and speciation rates manifest in statistically different estimates of net diversification in empirical

studies. Therefore, our results indicate that method-dependent tree-wide bias in diversification parameter inference may influence the interpretation of evolutionary shifts in diversification rates.

We find discrepancies between results derived from tree-wide summary statistics and our visual inspection of trees (see section "Visualizing rates on trees"). For example we find that `CLaDS2` and `Pesto` show no statistical difference in average net diversification (Fig. 6.2C). However, visual inspection of many trees suggests that `CLaDS2` and `Pesto` often differ greatly in the number and position of inferred rate shifts (*e.g.*, Fig. 6.1). Conversely, `BAMM` and `RevBayes` often look very similar when we assess individual phylogenies and yet significantly differ when we compare speciation, extinction, and diversification averages (Fig. S6.3A–C). This discrepancy reveals the difficulty of summarizing diversification rate estimates across phylogenies to reveal general patterns, and motivates the topology-informed rate comparisons, discussed in the following section.

Location and magnitude of rate shifts

We also test whether the models recover similar locations and magnitudes of rate shifts by comparing the mean squared error (MSE) of branch rates; this metric bridges the discrepancies between the global metrics and the observed patterns across the trees (both described above; Fig. 6.2).

When quantifying differences in the location and magnitude of shifts in speciation and net diversification rates, `CLaDS2` differs the most (larger MSE), compared with the other methods (Fig. 6.2D,F) and it is by far the biggest outlier across the models when visually inspecting the trees (Fig. 6.1E,J,O). This result indicates that `CLaDS2` estimates differ strongly from those of the other methods in the degree of the shifts it infers, and in their location. This result is in contrast to the tree-wide averages presented above (see also Fig. 6.2A–C), where `CLaDS2` is unexceptional. These results are also corroborated by analyses that take into account uncertainty in rate estimates (see Appendix S6.4, especially Fig. S6.6 and Table S6.3). We recover the same pattern when we subset our data by older and younger trees (Fig. S6.7), indicating that our comparisons are robust to clade age.

Tools for Assessing Methods

Inferred rates generally differ depending on the analysis method; how then should an empirical biologist choose which method to use? A reasonable conclusion from our analyses is that these methods are inherently unreliable. However, given the continued interest (both via use and development) in these methods, we advise empirical users who do wish to conduct diversification-rate analyses to take one of two paths.

The first path is to carefully select a method based on the model assumptions. The methods presented in this analysis have theoretical differences in their approach, which appear to

produce corresponding differences in results. For example, methods differ in whether shifts in diversification rates are allowed on extinct or unsampled lineages (`RevBayes`, `Pesto`, and the "exact `MTBD`" not tested here), whether diversification rates of each regime are drawn from a continuous distribution (`BAMM`, `MTBD`, and `CLaDS2`) or from a set of discrete rate categories (`RevBayes` and `Pesto`), and if shifts occur at cladogenetic events (`CLaDS2`) or along lineages (`BAMM`, `MTBD`, `RevBayes`, and `Pesto`). The models make additional assumptions, such as whether shifts in diversification rates affect the process-intrinsic parameters (the speciation and extinction rates) or transformations thereof (*e.g.*, the net diversification or turnover rate) and whether shifts affect single parameters or combinations of parameters. These assumptions lead to notably different interpretations of how values change through time. Choice of method can be supported by taxon-specific data such as species distribution, fossil record, or phenotypic data (Morlon, 2014). Thus, users should also familiarize themselves with how these models parameterize and estimate diversification rates and ensure that these modeling choices reflect the user's assumptions about biological processes.

The second path is to critically compare multiple methods when performing diversification analyses. We have shown that—despite the difference in models—in some cases multiple methods produce results with similar biological interpretations. To facilitate the adoption of this practice, we provide R code to easily visualize the results of multiple diversification-rate models across the same phylogeny: `https://github.com/Jesusthebotanist/CompDiv_processing_and_plotting`.

Regardless of the path taken, these methods are known to be sensitive to prior choice (e.g. `RevBayes`: prior on number of rate shifts η, Höhna et al. 2019, `MTBD`: prior on state change rate γ, Barido-Sottani et al. 2020, `BAMM`: prior on rate of shifts Moore et al. 2016; Mitchell and Rabosky 2017). Method developers generally recommend making inferences under a range of priors to characterize the impact of prior choice on the analysis. While assessing impact of the prior choice is outside of the scope of this paper, we strongly recommend researchers investigate prior sensitivity when conducting using these methods.

The Future of Diversification Analyses

The rise of methods aiming to identify shifts in diversification speaks to the importance of these analyses for understanding the drivers and impacts of important evolutionary events. However, we advocate for caution, for two reasons described below.

First, taking a cautious approach is especially important in light of the many potential problems with these methods, including the controversy surrounding the identifiability of birth-death models (Louca and Pennell, 2020b, but see also Helmstetter et al. 2022; Legried and Terhorst 2022; Morlon et al. 2022; Kopperud et al. 2023, among others).

Louca and Pennell (2020b) presented a class of birth-death models that are unidentifiable if the rate functions are time-varying (but homogeneous across lineages) and allowed to take

any continuous shape. Nonetheless, hypothesis-driven approaches are not allowed to take any arbitrary shape. Since the rate shapes are designed to test diversification scenarios, defined *a priori*, it has been shown that this approach is not prone to the identifiability issue (Morlon et al., 2022). Even time-varying models that are more agnostic about prior hypotheses are not typically allowed to take any continuous rate shape. Among the "agnostic" models, the piecewise-constant model is the most eminent (Stadler, 2011; Magee et al., 2020), and this model has been proven to be asymptotically identifiable provided there are not too many pieces (Legried and Terhorst, 2022).

However, in spite of the non-identifiability, inferences of rapidly changing speciation and extinction rates are still typically robust (Kopperud et al., 2023). The issue of non-identifiability remains to be investigated thoroughly in lineage-heterogeneous models. These models are more parameter-rich than their homogeneous cousins, and so we do not expect the issue of non-identifiability to be any simpler here.

Second, we caution against relying too heavily on the estimates from a single method without justifying the assumptions encoded into the model's choices regarding parameterization and estimation, as we describe in detail in "Tools for Assessing Methods".

Some users may wish to use model testing or model adequacy analyses to choose among methods, but we currently do not recommend these approaches. Model testing—especially across models implemented in different software packages—is likely to pick up signal from differences in prior specification that may mislead or otherwise confuse the results. In addition, Bayes factors for these models, if calculated from marginal likelihoods (e.g., by stepping-stone or path sampling), are unreliable due to the conflation of prior and likelihood in the tree model: such Bayes factors may offer decisive support for the wrong model (May and Rothfels, 2023). Model adequacy approaches have not yet been implemented (to our knowledge) in any of the software for these types of analyses. For example, posterior predictive simulations in `Pesto` or `RevBayes` would not condition on the observed rate shifts and instead would simulate from the prior distribution of rates. The development of appropriate methods for model adequacy under these models would be a highly useful tool for future research. However—for now—we recommend that users look towards other sources of data on diversification rates (*e.g.*, the fossil record, expected time to speciation, etc.) to verify if the models' estimated rates fall within reasonable expectations given the particularities of the system.

The methods investigated in this paper vary in their underlying model and assumptions, but are theoretically related (Ronquist et al., 2021). Due to these model differences, we expect differences in inferences which, in turn, could translate into different biological interpretations. Using a set of empirically derived phylogenies, we show that this is true (Fig. 6.1): no two methods inferred the same shifts for any phylogeny. In some cases, methods generally agreed upon the location and timing of inferred shifts, but in other cases methods strongly disagreed. Method-dependent differences of individual trees were corroborated by tree-wide summary

statistics, which indicated small but significant differences between methods (Fig. 6.2; Table Table 6.1). While these results hold up when we take into account the uncertainty in rate estimates, we also urge caution in relying too heavily on summary statistics and encourage users to carefully examine their posterior distributions, as 95% HPD intervals vary among methods and distributions may be bimodal, which may mislead common summary statistics (Fig. S6.8, see also Barido-Sottani et al., 2020).

Regardless, it is clear there will be a continued interest in using diversification analysis with a renewed appreciation for the complexities of these methods and the details of how rates are parameterized and estimated.

6.5.3 Acknowledgements

We thank Dr. Francisco Henao Diaz and Dr. Matthew Pennell for providing materials necessary to replicate their analyses. We thank Dr. Matthew Pennell and two anonymous reviewers for extensive feedback on the manuscript. We thank Dr. Michael R. May, members of the Rothfels Lab *sensu lato*, and the Specht lab for comments on the manuscript, we thank the Bonnie Lane Brain Trust for rapid brainstorming and the Cornell University Statistical consulting unit provided valuable advice on linear mixed models.

6.5.4 Funding

This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650441 for JM-G and No. DGE-1752814 for CMT. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This material is based upon work supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2209159 to JM-G and No. 2109835 to CMT. JM-G was additionally supported by a Cornell Provost Diversity Fellowship. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Emmy Noether-Program (Award HO 6201/1-1 to S.H.) and by the European Union (ERC, MacDrive, GA 101043187). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The authors declare no conflicts of interest.

6.5.5 Data accessibility

All scripts, data and outputs can be found at on Dryad at (doi:10.6078/D18Q68) upon publication. A set of R functions to help user analyze outputs of studied Baysian methods can be found at (`https://github.com/Jesusthebotanist/CompDiv_processing_and_plotting`)

Chapter 7

RevGadgets: an R package for visualizing Bayesian phylogenetic analyses from RevBayes

"The Italian mathematician **Gerolamo Cardano** (1501–1576) is believed to be the first to have formulated the notion of probability in gambling in terms of the number of distinguishable ways that events may occur."

(Pearl, 1988, p.70)

7.1 Abstract

1. Statistical phylogenetic methods are the foundation for a wide range of evolutionary and epidemiological studies. However, as these methods grow increasingly complex, users often encounter significant challenges with summarizing, visualizing, and communicating their key results.

2. We present `RevGadgets`, an `R` package for creating publication-quality figures from the results of a large variety of phylogenetic analyses performed in `RevBayes` (and other phylogenetic software packages).

3. We demonstrate how to use `RevGadgets` through a set of vignettes that cover the most common use cases that researchers will encounter.

4. `RevGadgets` is an open-source, extensible package that will continue to evolve in parallel with `RevBayes`, helping researchers to make sense of and communicate the results of a diverse array of analyses.

7.2 Introduction

Beyond being a graphical representation of the Tree of Life, phylogenetic trees provide a rigorous basis for a wide range of evolutionary and epidemiological inferences. Phylogenetic methods allow researchers to understand how molecular and morphological traits evolve (Nei, 1987; Yang, 2014; Felsenstein, 1985; Harvey and Pagel, 1991), how lineages disperse over geographic space (Ronquist and Sanmartín, 2011), and how lineages diversify over time (Morlon, 2014), among other evolutionary phenomena. Additionally, phylogenetic methods can be used to inform conservation decisions (Faith, 1992) and are powerful epidemiological tools (Volz et al., 2013; Baele et al., 2017).

Phylogenetic methods are increasingly based on explicit probabilistic models with parameters that describe underlying evolutionary processes. As datasets grow and evolutionary hypotheses become more nuanced, these models necessarily become more complex. `RevBayes` (Höhna et al., 2016a) is a Bayesian phylogenetic inference program that was developed to accommodate this increasing complexity and allows users to explore a vast space of phylogenetic models. Models in `RevBayes` are specified as probabilistic graphical models (Höhna et al., 2014), which are graphical representations of the underlying dependencies among parameters (and their corresponding prior distributions), similar to individual Legos being used to build a complex city. Using this graphical modeling framework, users can design customized models and tailor analyses to their particular datasets and research questions. However, this flexibility comes at a cost: because of the nearly infinite variety of possible models (and model combinations) that users can explore in `RevBayes`, the results of these analyses are often challenging to summarize and visualize using standard software. This is a significant limitation for `RevBayes` users because, in addition to being the primary method for reporting results of phylogenetic analyses, graphical summaries are a valuable tool for making sense of scientific results (Tufte, 2001), and for diagnosing modeling and analytical problems (Kerman et al., 2008).

Historically, `RevBayes` users have had to process and plot their results using *ad hoc* scripts written for each analysis, which imposed a significant barrier to entry for users not familiar with the structure of `RevBayes` output or comfortable with developing their own graphical summaries. To address these challenges, we developed `RevGadgets`. `RevGadgets` is an `R` package (R Core Team, 2020) that adds to the diverse ecosystem of phylogenetic visualization tools—*e.g.*, `ape` (Paradis and Schliep, 2019), `Tracer` (Rambaut et al., 2018), `phytools` (Revell, 2012), `ggtree` (Yu et al., 2017), `FigTree` (Rambaut, 2014), `IcyTree` (Vaughan, 2017), among many others—but is specialized for output produced by `RevBayes`. `RevGadgets` serves as a bridge between `RevBayes` analyses and existing tools for phylogenetic data processing and plotting in `R`, especially the `ggtree` package suite, which includes the `ggtree`, `tidytree`, and `treeio` packages (Wang et al., 2020; Yu et al., 2017). `RevGadgets` provides tools for plotting summary trees (including summaries of parameters for each branch), ancestral-state

estimates, and posterior distributions of parameters for a variety of models. Using the general framework of `ggplot2`, the `tidyverse`, and associated packages (Wickham, 2016; Wickham et al., 2019), plotting functions return plot objects with default, but customizable, aesthetics. Here, we present five vignettes demonstrating how to use `RevGadgets` to summarize results for a variety of phylogenetic analyses.

```
# specify the tree file
file <- "bears.mcc.tre"

# read the tree
tree <- readTrees(paths = file)

# plot the tree
plotFBDTree(tree = tree,
            timeline = TRUE,
            geo_units = "epochs",
            tip_age_bars = TRUE,
            node_age_bars = TRUE,
            age_bars_colored_by = "posterior",
            label_sampled_ancs = TRUE) +
  ggplot2::theme(legend.position=c(0.05, 0.55))
```

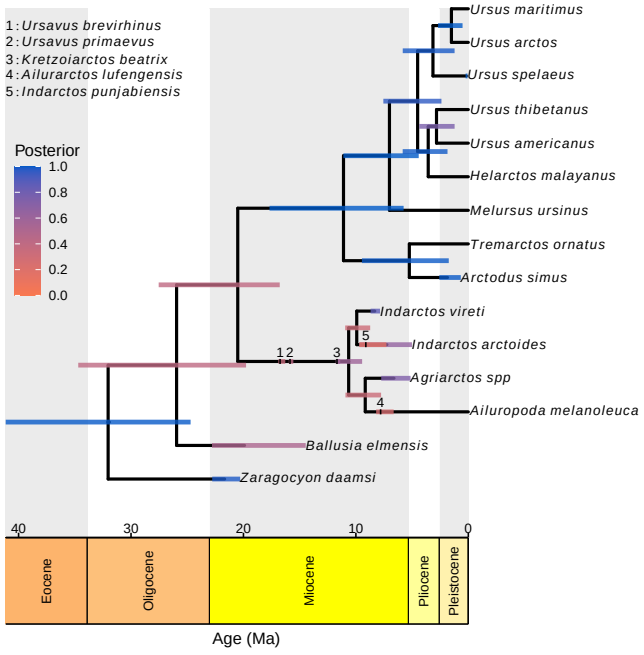


Figure 7.1: Plotting a time-calibrated phylogeny of extinct and extant taxa. Top) `RevGadgets` code for reading in and plotting a time-calibrated phylogeny of extant and extinct bears. We use the `theme` function from `ggplot2` to add the posterior-probability legend. Bottom) The maximum sampled-ancestor clade-credibility (MSACC) tree for the bears. Sampled ancestors are indicated by numbers along the branches (legend, top left). Bars represent the 95% credible interval of the age of the node, tip or sampled ancestor in millions of years (geological timescale, x-axis); the color of the bar corresponds to the posterior probability (legend, middle left) of that a clade exists, the posterior probability that a fossil is a sampled ancestor, or the posterior probability that a tip is not a sampled ancestor. (Data from Abella et al., 2012; Heath et al., 2014.)

7.3 Phylogenies

Phylogenies are central to all analyses in `RevBayes`, so accurate and information-rich visualizations of evolutionary trees are critical. In this case study, we demonstrate the tree-plotting functionality of `RevGadgets`, with methods to visualize phylogenies and their associated posterior probabilities, divergence-time estimates, and branch-specific parameter estimates.

`RevGadgets` provides paired functions for (1) reading in and processing data, and (2) summarizing and visualizing results. For phylogenies, the function `readTrees()` loads trees (either individual trees, or sets of trees) in either Newick or `NEXUS` (Maddison et al., 1997) formats, then processes associated branch or node annotations, and finally stores the tree(s) as `treedata` object(s) (as defined by `treeio`; Wang et al., 2020). Users can then visualize the `treedata` object using either `plotTree()` or `plotFBDTree()`, as we demonstrate below. Alternatively, users may choose to write custom plotting code using existing `ggtree` functions.

`RevGadgets` can plot both unrooted and rooted trees, and creates plots that are compatible with plotting options from `ggtree`. Additionally, `RevGadgets` provides extensive functionality for plotting trees with non-contemporaneous tips, such as those produced by total-evidence analyses under the fossilized birth-death [FBD] process (Heath et al., 2014; Zhang et al., 2016). The fossilized birth-death process (and the related serially-sampled birth-death process; Stadler, 2010) produces sampled ancestors (samples that are directly ancestral to another sampled taxon and thus are not represented as tips in the tree), and the ages of the samples are often subject to uncertainty (*e.g.*, because of imperfect knowledge about the age of the strata from which the samples were collected). As a consequence, conventional tree plotting tools are unsuitable for plotting FBD trees. We demonstrate how to use `RevGadgets` to plot the results of an FBD analyses of living and extinct bears (Fig. 7.1; data from Abella et al., 2012 and Heath et al., 2014). We include age bars colored by the posterior probability of the corresponding node, a geological time scale and labeled epochs from the package `deeptime` (Gearty, 2021), and fossils estimated to be direct ancestors of other samples (*i.e.*, sampled ancestors).

In addition to visualizing trees themselves, `RevGadgets` allows researchers to visualize branch-specific parameters, for example rates of evolution or diversification for each branch in the phylogeny. In Fig. 7.2, we demonstrate how to use `plotTree()` to visualize the estimated optimal body size as it varies across the cetacean phylogeny under a relaxed Ornstein-Uhlenbeck process (Butler and King, 2004; Uyeda and Harmon, 2014; data from Steeman et al., 2009; Slater et al., 2010). Under this model, a quantitative character evolves towards an adaptive optimum that changes along the branches of the tree, and thus the optimum associated with each branch is a focal inference.

The `plotTree()` function can also visualize unrooted or circular phylogenies, and users may add text annotations to denote posterior probabilities or other quantities. Users can apply `ggtree` functions to modify the `RevGadgets` plot, *e.g.*, to highlight certain clades with

```
# specify the annotated tree file
file <- "relaxed_OU_MAP.tre"

# read the tree
tree <- readTrees(paths = file)

# plot the tree
plotTree(tree = tree,
         color_branch_by = "branch_thetas",
         line_width = 1.7) +
  ggplot2::theme(legend.position=c(0.1, 0.9))
```

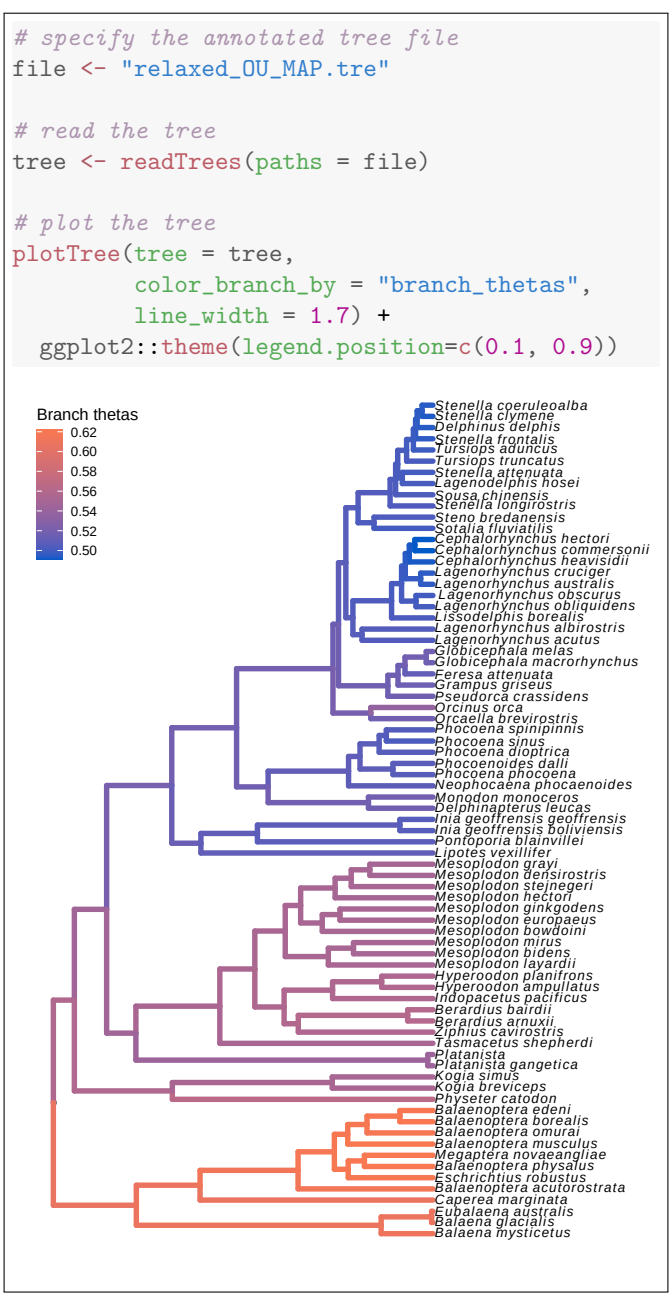


Figure 7.2: Plotting branch-specific parameter values across a phylogeny. Top) `RevGadgets` code for reading in and plotting the cetacean phylogeny that has been annotated with branch-specific adaptive optima (θ) inferred under a relaxed Ornstein-Uhlenbeck model. Bottom) The cetacean phylogeny with branches colored according to the posterior-mean estimate of the inferred branch-specific optimum body size, θ (legend, top left). (Phylogeny from Steeman et al., 2009; body size data in units of natural log-transformed meters from Slater et al., 2010.)

`geom_hilight()` or to add phylopics (`http://phylopic.org/`) using `geom_phylopic()`. Together, these functions provide user-friendly and customizable tree-plotting functionality for a variety of core research questions in evolutionary biology.

7.4 Posterior estimates of numerical parameters

`RevGadgets` provides several tools to visualize posterior distributions of numerical parameters. The output produced by most `RevBayes` analyses is a (typically tab-delimited) text file where rows correspond to samples from sequential iterations of an MCMC analysis, and columns correspond to parameters in the model. Most information of interest to researchers—*e.g.*, most probable parameter values (maximum *a posteriori*, or MAP, estimates), 95% credible intervals (CIs), or full posterior distributions—requires processing this raw MCMC output. Here, we demonstrate methods for processing and visualizing MCMC output for both quantitative and qualitative parameters.

We illustrate the core functions for reading, summarizing and visualizing posterior distributions of specific parameters with an example analysis of chromosome number evolution (Fig. 7.3; data from Freyman and Höhna, 2018). We use `readTrace()` to read in parameters sampled during one or more MCMC analyses. We then use `summarizeTrace()` to calculate the posterior mean and 95% credible interval for the focal parameters. Finally, we plot the marginal posterior distributions of the focal parameters using `plotTrace()`.

Plots of the posterior distributions of parameter values are key to a thorough understanding of the results of any Bayesian analysis. These tools encourage users to explore their results thoroughly rather than relying on single summary statistics. These summaries and plots may also be useful as tools for science communication and education on statistical phylogenetics, as they can easily be used to demonstrate differences in parameter estimates that result from changes to basic phylogenetic models. Additionally, the output of `readTrace()` may be passed to `R` packages specializing in MCMC diagnosis, *e.g.*, `convenience` (Fabreti and Höhna, 2022) or `coda` (Plummer et al., 2006). These functions are compatible with any delimited text file of MCMC samples, and can be used with the output of most Bayesian phylogenetic programs.

7.5 Ancestral-state estimates

In addition to making inferences about the underlying process of evolution, researchers may be interested in studying how particular characters evolved across the branches of the phylogeny. Ancestral-state estimation is a method for inferring that history.

`RevGadgets` offers two different types of summaries for ancestral-state estimates: (1) maximum *a posterior* (MAP) estimates, *i.e.*, the states with the highest posterior probability at each node, and; (2) pie charts that represent each state in proportion to its probability at each node. Ancestral-state estimates may be represented as text annotations rather than colored symbols. Additionally, `RevGadgets` can summarize and visualize ancestral-state estimates at internal nodes and at the "shoulders", *i.e.*, at the beginning of each branch. Plotting the states at internal nodes is appropriate for standard evolutionary models of anagenetic (within-lineage) change. However, models of evolution that include a cladogenetic compo-

```
# specify the log files with rates of
# chromosome evolution
files <- c("chromevol_simple.log")

# read the trace
trace <- readTrace(path = files)

## Reading in log file 1

# summarize the traces of the parameters
# gamma: chromosome gain rate
# delta: chromosome loss rate
summary <- summarizeTrace(
    trace = trace,
    vars  = c("gamma", "delta"))

# report the summary for the gain rate
summary[["gamma"]]

## $trace_1
##          mean        median           MAP
##    0.28853914    0.23994560    0.13586156
##  quantile_2.5 quantile_97.5
##    0.03347161    0.80123940

# plot the posteriors of each parameter
plotTrace(trace = trace,
    vars  = c("gamma", "delta"))[[1]]
```

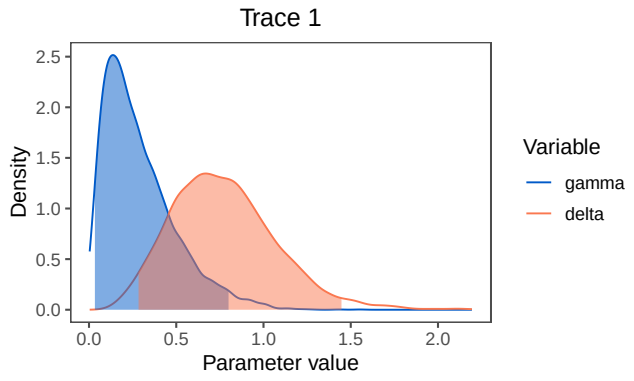
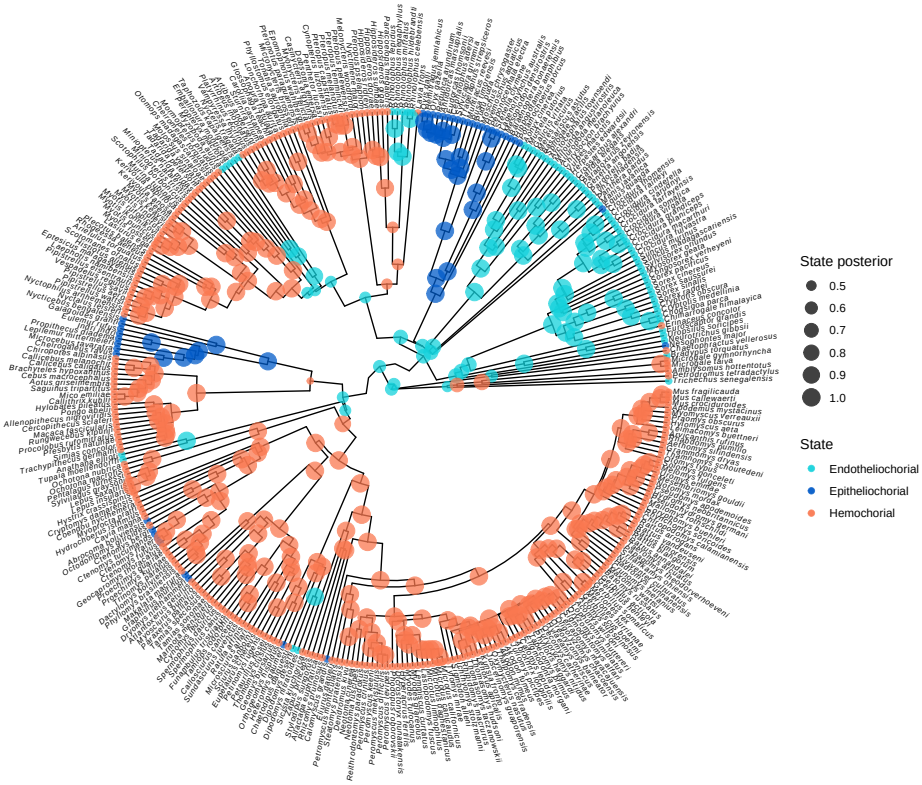


Figure 7.3: Plotting posterior distributions of numerical parameter values. Top) `RevGadgets` code for reading in and plotting the posterior distributions of rates of chromosome evolution in *Aristolochia*. Bottom) Marginal posterior distributions of the two rate parameters. Shaded regions represent the 95% credible interval of each posterior distribution. (Data from Freyman and Höhna, 2018.)

nent (*e.g.*, models of biogeographic or chromosome-number evolution; Ree and Smith, 2008; Goldberg and Igić, 2012; Freyman and Höhna, 2018) also allow states to change at speciation events. In this case, researchers may also want to plot the shoulder states, which represent the ancestral-state estimates for each daughter lineage immediately following the speciation

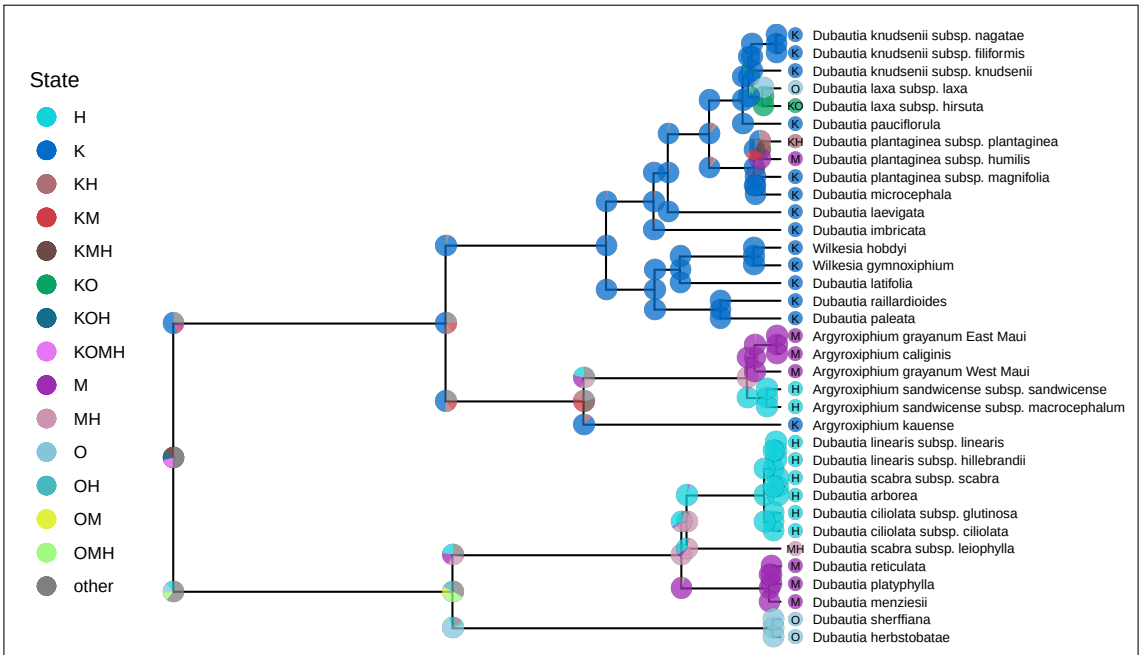



```
# specify the tree file
file <- "ase_freeK.tree"

# read in the tree annotated with ancestral states
freeK <- processAncStates(file, state_labels = c("1" = "Epitheliochorial",
                                                 "2" = "Endotheliochorial",
                                                 "3" = "Hemochorial"))

# plot the tree with MAP ancestral states
plotAncStatesMAP(t = freeK, tree_layout = "circular") +
    ggplot2::theme(legend.position = c(0.96,0.5))
```

Figure 7.4: Plotting maximum *a posterior* (MAP) estimates of ancestral states on a circular phylogeny. Top) MAP estimates of ancestral placental states across the phylogeny of mammals. Each node is colored by the MAP state (legend, bottom right); the size of each symbol is proportional to the posterior probability of the map state (legend, top right). Bottom) `RevGadgets` code for reading in and plotting the MAP estimates for ancestral placental states across the mammals phylogeny. (Data from Elliot and Crespi, 2006.)

event.

We demonstrate how to plot ancestral-state estimates of placenta type across the mammal phylogeny under an asymmetric model of character evolution (Fig. 7.4; data from Elliot and Crespi, 2006). First, we use `processAncStates()` to read in and parse the phylogeny and ancestral-state estimates inferred using `RevBayes`. Second, we use `plotAncStatesMAP()` to color each node symbol according to the state with the highest posterior probability, and make the area of the symbol proportional to that state's posterior probability. Because of the size of




```
# specify the annotated tree file
file <- "simple_ase.tre"

# define the state labels
labs <- c("1"  = "K",   "2"  = "O",   "3"  = "M",  "4"  = "H",  "5"  = "KO",  "6"  = "KM",
          "7"  = "OM",  "8"  = "KH",  "9"  = "OH", "10" = "MH", "11" = "KOM", "12" = "KOH",
          "13" = "KMH", "14" = "OMH", "15" = "KOMH")

# read the annotated tree file
dec_example <- processAncStates(file, state_labels = labs)

# plot the tree with pie charts for ancestral states
plotAncStatesPie(t = dec_example, cladogenetic = TRUE, tip_labels_states = TRUE,
                 tip_label_italics = FALSE, tip_labels_offset = 0.2,
                 tip_pie_nudge_x = 0.15, tip_pie_size = 1.0,
                 node_pie_size = 1.5, tip_labels_states_offset = 0.05) +
    ggplot2::theme(legend.position = c(-0.05, 0.5), plot.margin = unit(c(0,0,0,2),"cm"))
```

Figure 7.5: Plotting posterior distributions of ancestral states under a cladogenetic model. Top) The posterior estimates ancestral biogeographic states of the Hawaiian silverswords estimated under a DEC model. The size of each pie slice is proportional to the posterior probability of a given state (legend, top left) for a particular lineage. Pies at nodes represent the state of the ancestral lineage immediately before speciation; pies at "shoulders" represent the states of each daughter lineage immediately following the speciation event. Bottom) `RevGadgets` code for reading in and plotting the posterior estimates for ancestral geographic range across the phylogeny of Hawaiian silverswords. (Data from Landis et al., 2018.)

the phylogeny, we choose to plot the estimates on a circular tree by changing the tree layout parameter.

Next, we demonstrate plotting estimates of ancestral ranges of the Hawaiian silversword alliance that were generated by a Dispersal-Extinction-Cladogenesis (DEC) model (Fig. 7.5; data from Landis et al., 2018). Since the DEC model features a cladogenetic component, we include shoulder-state estimates. Because of the large number of states in this analysis (15

possible ranges and one "other" category), more pre-processing is necessary. As before, we pass the appropriate state names to `processAncStates()`; however, in this case we plot pie charts representing the probability of each state using `plotAncStatesPie()`, and plot states at shoulders using `cladogenetic = TRUE`.

Beyond the above examples, these versatile plotting tools can visualize any discrete ancestral-state estimates reconstructed by `RevBayes`, including the results of chromosome count estimations (Freyman and Höhna, 2018) and discrete state-dependent speciation and extinction (SSE) models (Freyman and Höhna, 2019; Zenil-Ferguson et al., 2019).

7.6 Diversification rates

The processes of speciation and extinction (*i.e.*, lineage diversification) is of great interest to evolutionary biologists (Morlon, 2014). Rates of speciation and extinction may be modeled as constant over time and among branches (as in a constant-rate birth-death process; Kendall, 1948; Nee et al., 1994b), or allowed to vary over time (Stadler, 2011; May et al., 2016), across branches of a phylogeny (Rabosky, 2014; Höhna et al., 2019), or based on the character states of the evolving lineages (Maddison et al., 2007; Freyman and Höhna, 2019). For example, rates that vary across branches of the phylogeny can be visualized using `plotTree()` to color the branches by their inferred rate. State-dependent diversification models provide estimates of the speciation and extinction rates associated with each character state, and may also be used to estimate ancestral states. `plotTrace()` or specific processing and plotting functions for diversification rates—`processSSE()`, `plotMuSSE`, and `plotHiSSE`—may be used to visualize the estimated rates. `plotAncStatesMAP()` or `plotAncStatesPie()` may be used to visualize the ancestral-state estimates.

We demonstrate how to plot the results of a time-varying model—the episodic birth-death process (Stadler, 2011; Höhna, 2015)—applied to primate phylogeny (Fig. 7.6; Springer et al., 2012). The episodic birth-death analysis in `RevBayes` produces separate trace files each type of rate. We read these output files using `processDivRates()` and plot the resulting parameter estimates over time using `plotDivRates()`.

Together with the aforementioned functions for plotting diversification parameter estimates, `plotDivRates()` allows users to visualize the outputs of nearly all diversification analyses available in `RevBayes`. Stochastic character mapping of diversification estimates, in which the timing and location of diversification rate shifts are painted along the branches of the tree, will be added in the future (Freyman and Höhna, 2019; Höhna et al., 2019).

7.7 Model adequacy

In addition to visualizing the results of phylogenetic inferences with a specific model, `RevGadgets` provides tools for exploring the adequacy of the model (*i.e.*, whether the model provides an

```
# specify the log files with diversification
# rates and rate-change times
speciation_time_file <- "speciation_times.log"
speciation_rate_file <- "speciation_rates.log"
extinction_time_file <- "extinction_times.log"
extinction_rate_file <- "extinction_rates.log"

# read the log files
rates <- processDivRates(
    speciation_time_log = speciation_time_file,
    speciation_rate_log = speciation_rate_file,
    extinction_time_log = extinction_time_file,
    extinction_rate_log = extinction_rate_file)

# plot the diversification rates
plotDivRates(rates = rates)
```

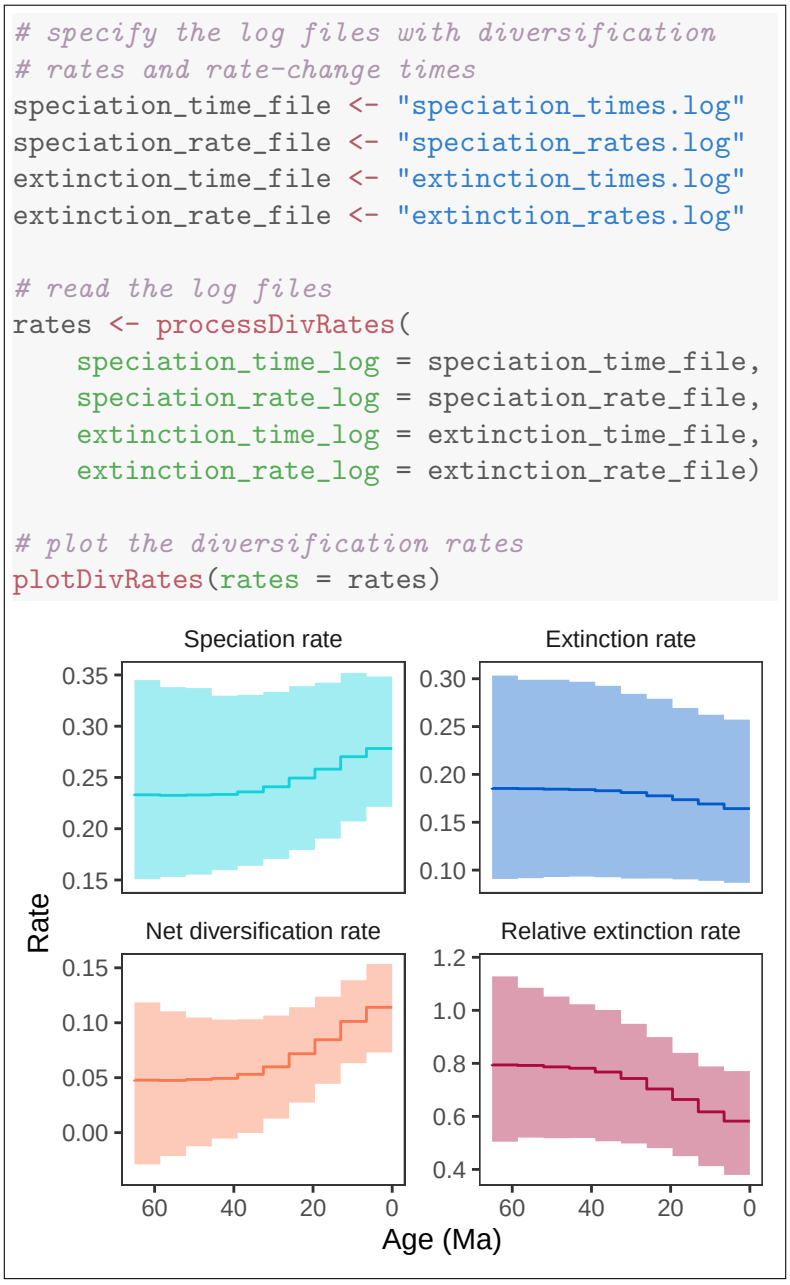


Figure 7.6: Plotting posterior distributions of diversification rates over time. Top) `RevGadgets` code for reading in and plotting the posterior estimates of diversification rates over time inferred from the primate phylogeny. Bottom) Posterior distributions of speciation and extinction rates over time, as well as the net diversification rate (speciation minus extinction) and the relative extinction rate (extinction divided by speciation). Dark lines correspond to the posterior-mean estimate of each parameter for each time interval, and shaded regions correspond to the 95% credible interval. (Data from Springer et al., 2012.)

adequate description of the data-generating process; Bollback, 2002; Gelman et al., 2013; Brown, 2014; Höhna et al., 2018). Posterior-predictive analysis tests whether a fitted model simulates (predicts) data that are similar to the observed data. This process is distinct from model testing, in which one model is chosen from a set of possible models, as the best model

```
# specify the simulated statistics file
sim <- "simulated_data_pps.csv"

# specify the empirical statistics file
emp <- "empirical_data_pps.csv"

# read the statistics files
stats <- processPostPredStats(path_sim = sim,
                              path_emp = emp)

# create the posterior-predictive plots
plots <- plotPostPredStats(data = stats)

# plot some of the statistics
plots[c(1,3,5,7)]
```

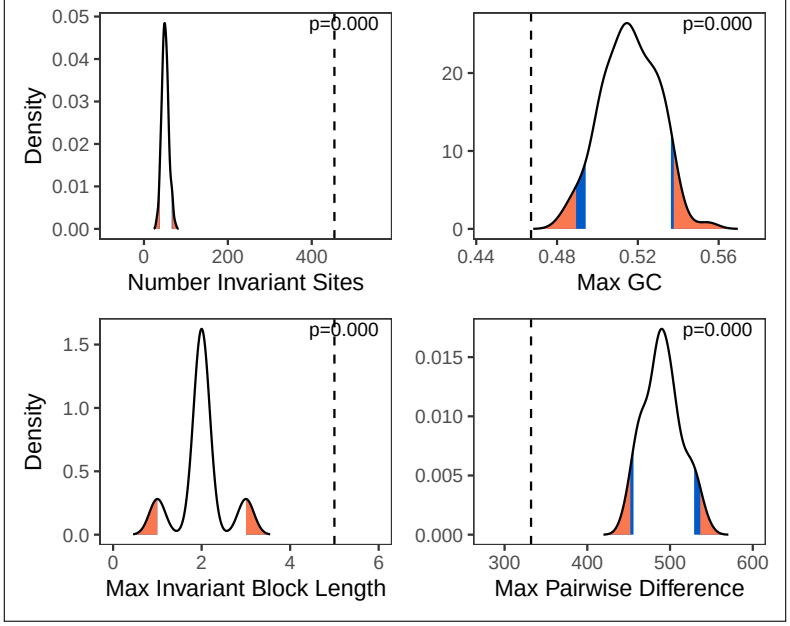


Figure 7.7: Plotting simulated posterior-predictive distributions to assess model adequacy. Top) `RevGadgets` code for reading in and plotting the distributions of summary statistics generated using posterior-predictive simulation posterior. Bottom) Posterior-predictive distributions (black curves) of four statistics simulated under the Jukes-Cantor model fit to primate *cytb*, compared to the same statistics computed on the observed data (dashed vertical lines). The posterior-predictive p-value (upper right of each panel) is the fraction of simulated statistics that are as or more extreme than the observed statistic. If the observed statistic falls in or beyond the orange region, we deem the model as inadequate at the 5% significance level; if the observed statistic falls in the blue region, the model is marginally adequate at the 10% significance level. In this case, the Jukes-Cantor model provides an inadequate description of the true generating process according to every summary statistic. (Data from Springer et al., 2012.)

of the set may still provide an inadequate description of the underlying process.

First, users analyze their data with the model of interest and then use the inferred posterior

distribution to simulate a number of new data sets. The user then selects test statistics that describe important features of the data (*e.g.*, the number of invariant sites in a nucleotide alignment) and calculates these statistics for both the observed data and the simulated data. If the statistic from the empirical data is reasonably included within the distribution of statistics from simulated datasets (posterior-predictive p-value > 0.05), the model is considered an adequate description of the process that produced the tested data feature.

Here, we demonstrate the workflow for a posterior-predictive analysis to test model adequacy of the Jukes-Cantor model for nucleotide sequence evolution (Jukes and Cantor, 1969) in a single gene across a sample of 23 primates (Fig. 7.7; data from Springer et al., 2012). First, we perform an analysis in `RevBayes` under a Jukes-Cantor model of nucleotide sequence data. Second, we use `RevBayes` to simulate datasets under the posterior distributions estimated in the first step. Third, we use `RevBayes` to calculate statistics from the simulated and empirical datasets. These statistics should describe aspects of the data that we hope capture a meaningful component of model performance. Finally, we use `RevGadgets` to plot those statistics and compute posterior-predictive p-values.

Despite being computationally inexpensive compared to Bayesian model comparison methods (*i.e.*, Bayes factor calculation), posterior-predictive approaches remain relatively uncommon in empirical phylogenetic studies. As genome-scale datasets and increasingly complex statistical methods become more accessible to researchers, posterior-predictive simulation will be critical to testing how well our models describe the underlying generative processes. This component of `RevGadgets` functionality and the associated clear workflows for performing and interpreting posterior-predictive tests will hopefully increase the adoption of this important tool.

7.8 Conclusions

`RevBayes` is a flexible platform for performing Bayesian phylogenetic evolutionary inferences. Because of the almost endless possibilities for building unique combinations of models in `RevBayes`, these analyses are often challenging to visualize using standard plotting software. We have developed an `R` package, `RevGadgets`, to produce publication-quality visualizations of phylogenetic analyses performed in `RevBayes`. The case studies described above illustrate some of the core functionality available in `RevGadgets` and demonstrate how to produce plots of the most commonly-performed `RevBayes` analyses. `RevBayes` is open source software that is actively maintained and developed. Likewise, `RevGadgets` is also open source and will continue to provide new plotting tools to meet new visualization challenges as they arise. `RevGadgets` and any future updates will be available on CRAN (`https://cran.r-project.org/web/packages/RevGadgets/index.html`) and on GitHub at `https://github.com/cmt2/RevGadgets`. Additionally, we provide thorough documentation for all functionality in the package and

maintain numerous tutorials demonstrating how to use `RevGadgets` on the `RevBayes` website at `https://revbayes.github.io/tutorials/`. Together, the modular modeling tools from `RevBayes` and the visualization gadgets in `RevGadgets` will help researchers make sense of and communicate the results of a diverse array of sophisticated phylogenetic analyses.

7.8.1 Dependencies

`RevGadgets` depends on many `R` packages, in particular: `ape` (Paradis and Schliep, 2019); `phangorn` (Schliep, 2011); `phytools` (Revell, 2012); `ggplot2` (Wickham, 2016); `ggtree` (Yu et al., 2017); `treeio` (Wang et al., 2020); `deeptime` (Gearty, 2021); `dplyr` (Wickham et al., 2021); `treeplyr` (Uyeda and Harmon, 2020); `tidytree` (Yu, 2021b); `reshape` (Wickham, 2007); `ggthemes` (Arnold, 2021); `tidyr` (Wickham, 2021); `tibble` (Müller and Wickham, 2021); `gginnards` (Aphalo, 2021a); `ggimage` (Yu, 2020); `ggplotify` (Yu, 2021a); `png` (Urbanek, 2013); and `ggpp` (Aphalo, 2021b).

Appendix A

Supplementary Material for Chapter 2

S2.1 Calculation of the equivalent rate functions within the congruence class

The pulled net-diversification (Louca and Pennell, 2020b) is one of the identifiable parameters in the reconstructed birth-death model and the central component in `CRABS`. It is defined as a differential equation

$$r_p(t) = \lambda(t) - \mu(t) + \frac{1}{\lambda(t)}\frac{d\lambda(t)}{dt} \quad . \tag{S2.1}$$

The pulled rates are calculated automatically in `CRABS` for any diversification model regardless of the shape of speciation and extinction rate functions provided. In principle, if the specific form of the speciation rate function is known, then $\frac{d\lambda(t)}{dt}$ and thus $\frac{d\lambda(t)}{dt}$ could be computed analytically. In `R`, such knowledge of the form of equation representing a function is not known, and therefore we compute $\frac{d\lambda(t)}{dt}$ numerically.

Next, recall that the pulled net-diversification rate, together with $\lambda(0)$, completely characterizes the congruence class. In order to construct a congruent model, we can propose an alternative rate function $\mu'(t)$, and solve for $\lambda(t)$. Solving Eq. S2.1 for $\mu(t)$ is trivial, as $\mu(t) = \lambda(t) - r_p(t) + \frac{1}{\lambda(t)}\frac{d\lambda(t)}{dt}$. In order to solve for $\lambda(t)$, we first rearrange the equation as

$$\frac{d\lambda(t)}{dt} = -\lambda(t)^2 + \lambda(t)(\mu(t) + r_p(t)) \quad . \tag{S2.2}$$

Then, we use an ODE (Ordinary Differential Equation) solver to compute the solution for $\lambda(t)$ numerically. More specifically, we use a Runge-Kutta method of order five, implemented in Fortran by Hairer and Wanner (1996). We use the function `radau` in the R-package `deSolve` (Soetaert et al., 2010) as an interface to access the ODE solver in R.

We set of functions to compute the alternative speciation and extinction rates. However, these functions are evaluated at specific time-points provided by the users. The alternative speciation and extinction rates correspond to a model within the congruence class, although

only at the provided time-points. In between these time-points the function are, as in our examples, interpolated. This makes these functions, in a strict mathematical sense, not part of the congruence class. Nevertheless, the resemblance to the true function within the congruence class can be made arbitrarily high, so that these functions are visually indistinguishable, see below.

S2.2 Accuracy of the method

S2.2.1 Lineages through time

We evaluate the rate functions on a pre-specified grid of time points. The idea is that, for sufficiently many time points, or sufficiently small Δt, we can approximate any continuously varying rate function. How many time points is considered sufficient? To assess this question, we require some sort of benchmark. One property of the congruence class is that all models must have the same deterministic lineage-through-time (dLTT) curve. Thus, we can check if our transformation to another model within the congruence class worked. We used the R-package TESS (Höhna et al., 2016b) to calculate the dLTT for a selection of models where we varied the amount of time points on the grid (Fig. S2.1). In the original analysis, we used 100 epochs or intervals to estimate the piecewise-constant diversification rates. It appears that 10, 20, and even 50 time points are inadequate. In our view, 500 or more time points are sufficient for precisely setting up the primates congruence class, and for proposing alternative congruent models.

S2.2.2 Likelihood

We also evaluated how the likelihood behaves in relation to the number of grid points that we used to set up the model. For all models, we used a reference model with speciation and extinction defined on a 100-sized grid. We defined the alternative model as having a constant extinction rate of $\mu'(t) = 0.1$. When calculating $r_p(t)$, and calculating the alternative $\lambda'(t)$, we used a grid of varying size, from 100 up to 10000. First, we solved the differential equations:

$$\begin{aligned} \frac{dE(t)}{dt} &= \mu(t) - E(t)(\lambda(t) + \mu(t)) + E(t)^2\lambda(t) \\ \frac{dM(t)}{dt} &= M(t)\lambda(t)(E(t) - 1) \qquad , \end{aligned} \tag{S2.3}$$

where $E(t)$ is the probability that a lineage at time t either went extinct before the present, or did not get sampled. $M(t)$ is the total number of lineages at time t in the reconstructed phylogeny. The initial conditions are $E(0) = 0.33$ (67% of described primate species were sampled at the present), and $M(0) = 233$ number of tips at the present. We follow Louca and Pennell (2020b) in calculating the log likelihood, conditioning upon that at least two lineages must have survived to the present:

$$\log(L) = 2\log(M(t_1)) - (n+1) \times \log(M(0)) + \sum_{i=2}^{n} \log\Big(- \frac{dM}{dt}\Big|_{t_i} \Big) \qquad , \tag{S2.4}$$

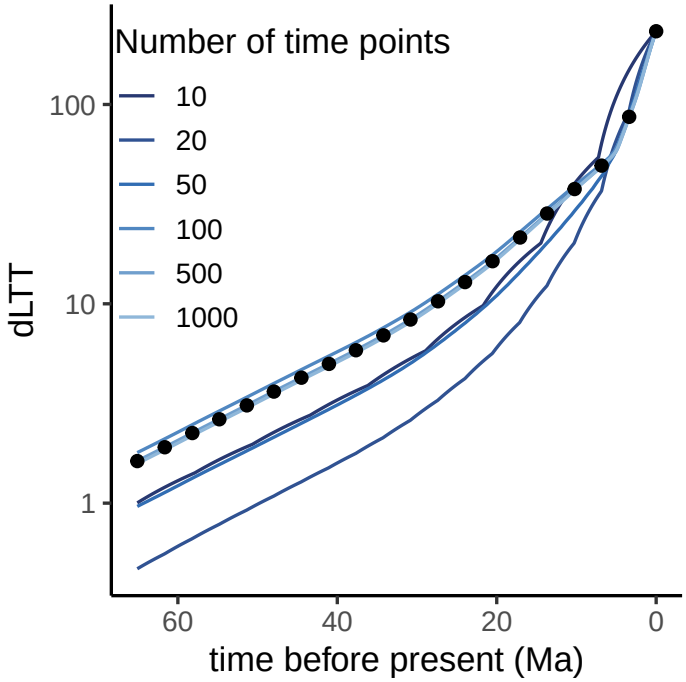


Figure S2.1: The deterministic lineage-through-time (dLTT) for two congruent models generated using a range of number of time points on the grid. If the transformation worked, i.e., we obtained truly another diversification rate model within the same congruence class, then the dLTT must be identical. We computed the dLTT curves using TESS (Höhna et al., 2016b). The dotted line represents the reference model (i.e. the dLTT based on the posterior median speciation and extinction rate estimates for the Primates phylogeny), where we used 1000 time points. The solid lines depict congruent models where we proposed a new extinction rate $\mu'(t) = 2 \times \sup(\lambda(t)) - \lambda(t)$, and inferred the congruent speciation rate. The proposed models approach the dLTT of the reference model as the number of time points increase. 500 and 1000 time points are nearly indistinguishable.

where t_1 is the time of the oldest branching event (the root), and t_i are the times of the subsequent branching events. The likelihood comparisons between the reference and the alternative model can be seen in Fig. S2.2. If two models are truly congruent, they should have exactly the same likelihood. The likelihood comparison between the two models at 100 time points is rather small (left-most dot), however there is substantial error with few number of time points. As we increase the number of time points, the error decreases. Although the likelihoods appear to converge, there is some fluctuation around zero, meaning the likelihoods are not quite stable for the dynamically changing primates model.

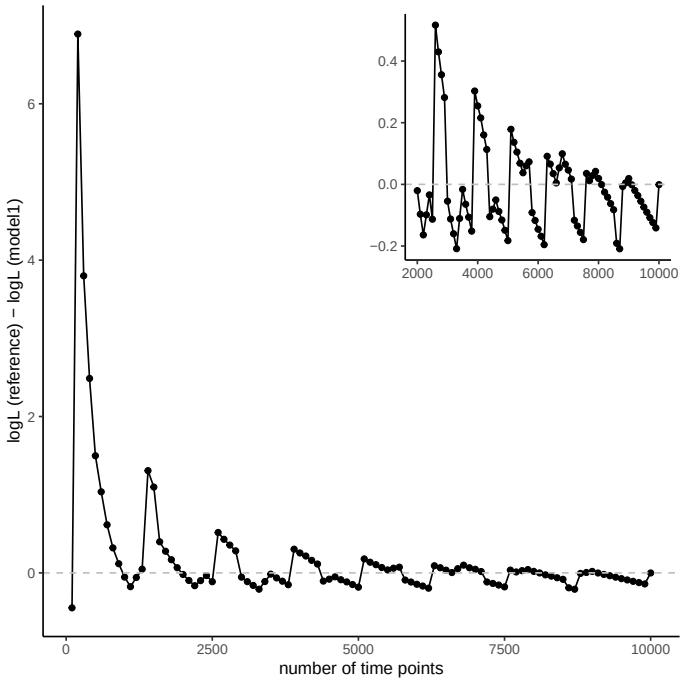


Figure S2.2: Comparison of the log likelihood between the primate reference model (main text Fig. 2.1), and an alternate model where we proposed a new extinction rate of $\mu'(t) = 0.1$. We used an equally-spaced vector going from 0 to the tree height for the time grid, with number of time points varying from 100 up to 10000. The inset is a cut-out of the likelihoods for more than 2000 time points. We used `CRABS` to compute $r_p(t)$ and $\lambda'(t)$, and calculated the likelihood as described in the text.

Appendix B

Supplementary Material for Chapter 3

S3.1 Extended datasets

In this section we plot the estimated reference model and summary of the directional trends for an additional nine datasets. These datasets provide more empirical examples about the robustness of estimated diversification patterns and trends within the congruence class. In addition to the trends in speciation rate, we also plot the trends in the net-diversification rate for the congruent models. Overall, the results corroborate our main findings in the main text: (1) sharp diversification rate changes are robust and unambiguously supported, (2) trends in the near-present show the strongest conflicting signal, and (3) reference models with comparably flat diversification rates provide ambiguous signal within the congruence class.

We first plot the figure from the main text, with the added information of the trends in the net diversification rate (Fig. S3.1). For the additional empirical examples, we chose to first include the finches (Fringillidae), manakins (Pipridae), and the most species-rich family of tubenose birds (Procellariidae) (Fig. S3.2). In Fig. S3.3 we depict the equivalent congruent models for another three families: mice (Muridae), white-eyes birds (Zosteropidae), and the evening primroses (Onagraceae). The diversification histories of the indigobirds, whydahs, and the cuckoo-finch (Viduidae), the New World warblers (Parulidae), and free-tailed bats (Molossidae) are comparatively more recent and their rates change more abruptly (Fig. S3.4). For this reason, we assessed the directional trends with a more conservative threshold of $\epsilon = 0.08$ rate units per million years (Fig. S3.4). The tree for the evening primroses was taken from Freyman & Höhna Freyman and Höhna (2019), while the remaining trees were taken from Condamine et al. Condamine et al. (2019).

The significant directional trends in the speciation rates are consistent across the congruence class when the slope of the speciation rate is steep (Figs. S3.2 to S3.4). When the speciation rate is shallow, it is trivial to construct a model that is in disagreement with the reference model.

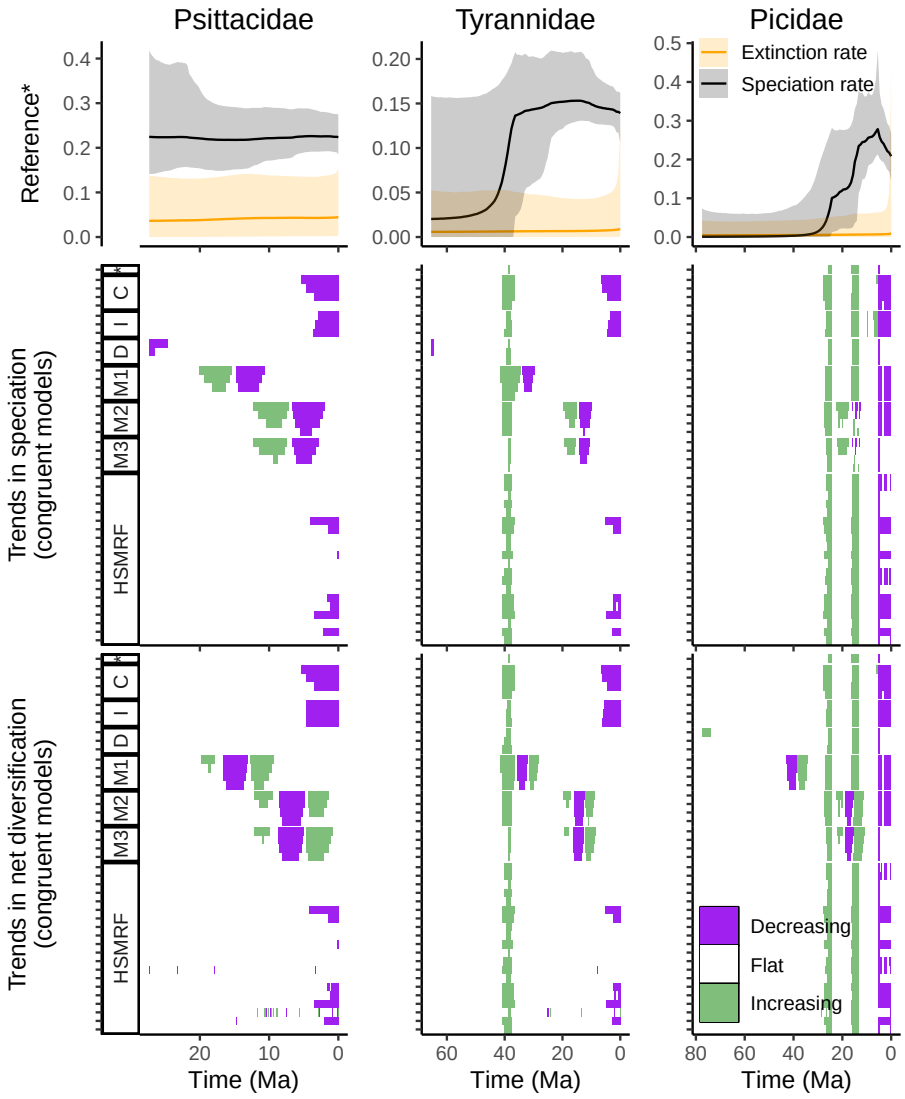


Figure S3.1: A summary of the congruence class for three bird families: New World/African true parrots (Psittacidae), tyrant flycatchers (Tyrannidae), and woodpeckers (Picidae). This figure is equal to the main text figure, except it also includes the trends in the net-diversification rates of the congruent models. The top row depicts the posterior median and 95% credible interval speciation and extinction rates, for the episodic birth-death models implemented in `RevBayes`. The two bottom panels depict summaries of directional trends in the congruence class, where each row is a model. The models include alternative extinction rates that are constant (C), exponentially increasing (I), exponentially decreasing (D), modal (M1-M3), HSMRF-distributed (horseshoe Markov random field), and the reference model (*). If the slope of the rate function is greater than the threshold ϵ (in rate units per million years), or less than $-\epsilon$, we say that the function is increasing, or decreasing, respectively.

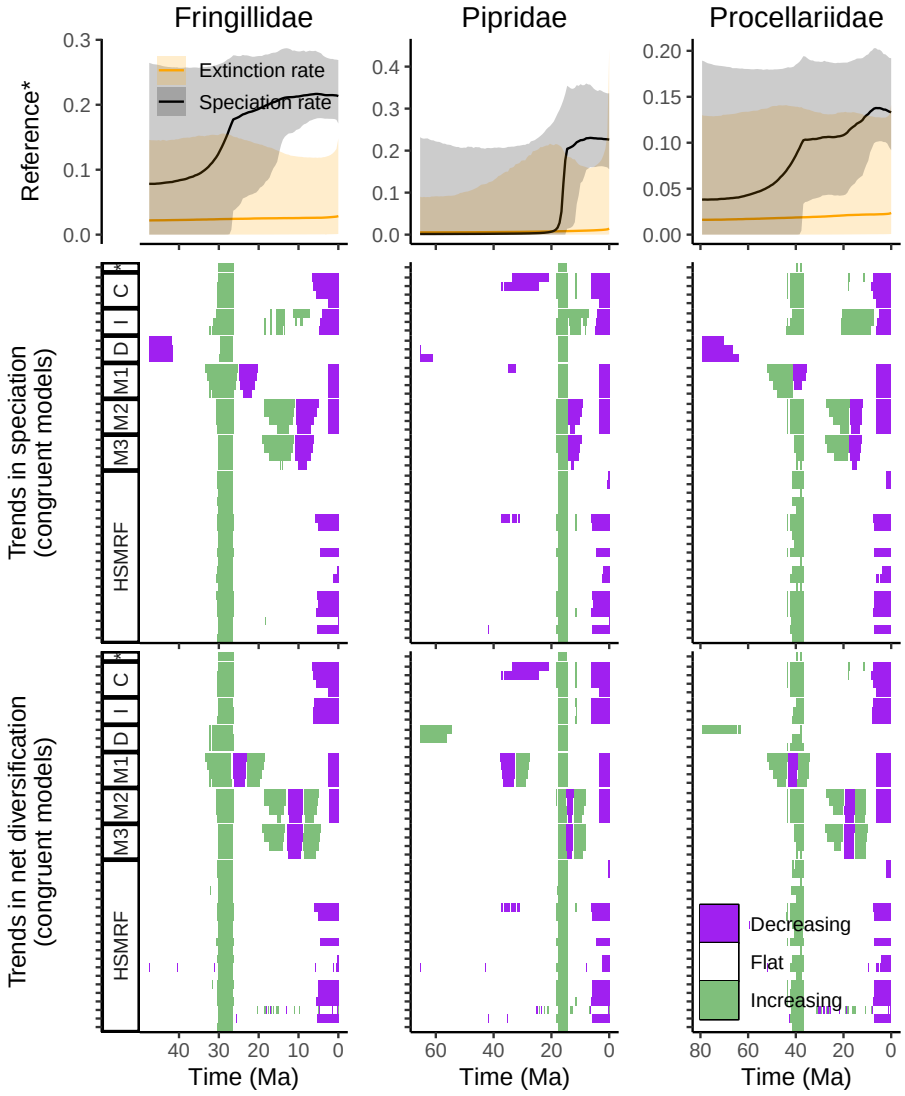


Figure S3.2: A summary of the congruence class for three bird families: finches (Fringillidae), manakins (Pipridae), and the most species-rich family of tubenoses (Procellariidae). The top row depicts the posterior median and 95% credible interval speciation and extinction rates, for the episodic birth-death models implemented in `RevBayes`. The two bottom panels depict summaries of directional trends in the congruence class, where each row is a model. The models include alternative extinction rates that are constant (C), exponentially increasing (I), exponentially decreasing (D), modal (M1-M3), HSMRF-distributed (horseshoe Markov random field), and the reference model (*). If the slope of the rate function is greater than the threshold ϵ (in rate units per million years), or less than $-\epsilon$, we say that the function is increasing, or decreasing, respectively. We used a threshold of $\epsilon = 0.01$ for the finches and manakins, and $\epsilon = 0.005$ for the tubenoses.

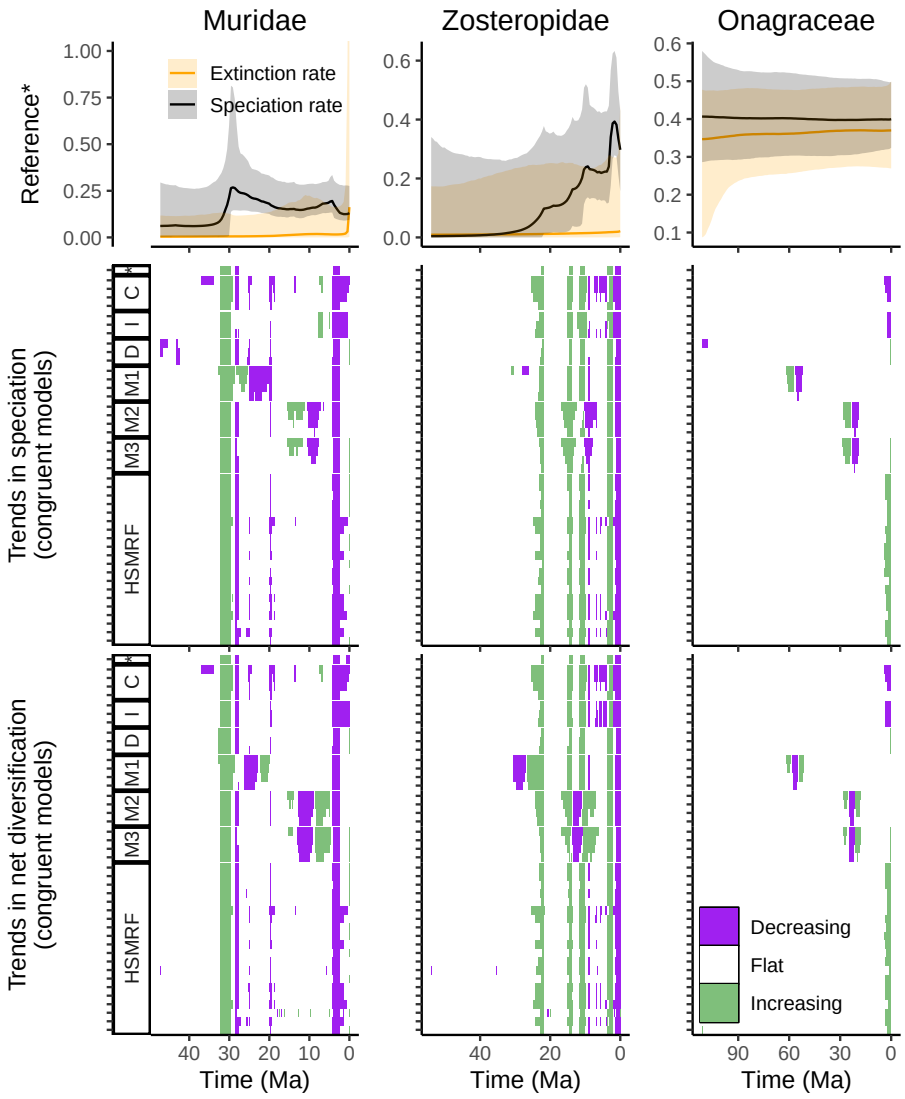


Figure S3.3: A summary of the congruence class for three families: mice (Muridae), white-eyes birds (Zosteropidae), and the evening primroses (Onagraceae). The top row depicts the posterior median and 95% credible interval speciation and extinction rates, for the episodic birth-death models implemented in `RevBayes`. The two bottom panels depict summaries of directional trends in the congruence class, where each row is a model. The models include alternative extinction rates that are constant (C), exponentially increasing (I), exponentially decreasing (D), modal (M1-M3), HSMRF-distributed (horseshoe Markov random field), and the reference model (*). If the slope of the rate function is greater than $\epsilon = 0.02$ (in rate units per million years), or less than $-\epsilon$, we say that the function is increasing, or decreasing, respectively. For the evening primroses, we added a constant of 0.3 such that $\mu'(t) = \mu(t) + 0.3$ for all proposed extinction rates. This accommodates their overall higher diversification rates, and prevents artifacts in the near-present due to the congruence class requiring λ_0 to be equal across all models.

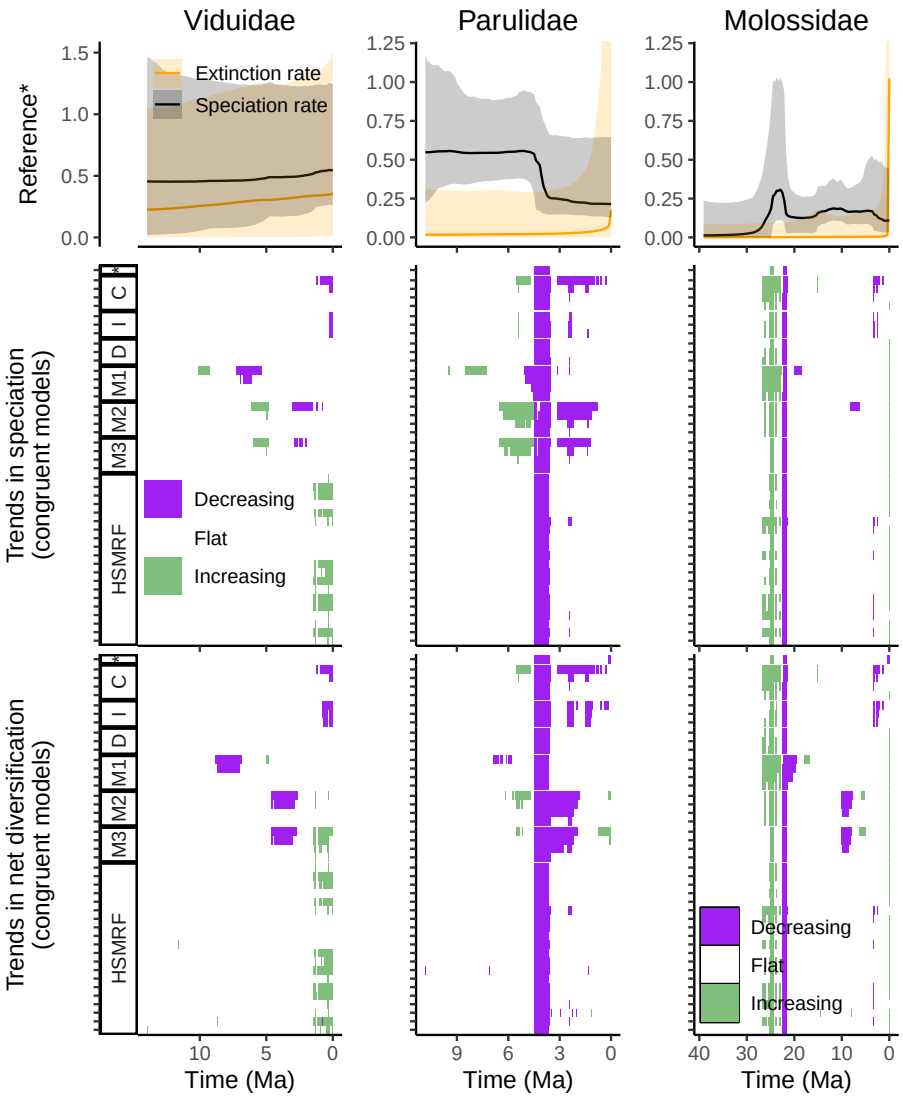


Figure S3.4: A summary of the congruence class for three families: indigobirds, whydahs, and the cuckoo-finch (Viduidae), the New World warblers (Parulidae), and free-tailed bats (Molossidae). The top row depicts the posterior median and 95% credible interval speciation and extinction rates, for the episodic birth-death models implemented in `RevBayes`. The two bottom panels depict summaries of directional trends in the congruence class, where each row is a model. The models include alternative extinction rates that are constant (C), exponentially increasing (I), exponentially decreasing (D), modal (M1-M3), HSMRF-distributed (horseshoe Markov random field), and the reference model (*). If the slope of the rate function is greater than $\epsilon = 0.08$ (in rate units per million years), or less than $-\epsilon$, we say that the function is increasing, or decreasing, respectively. For the Viduidae, we added a constant of 0.25 such that $\mu'(t) = \mu(t) + 0.25$ for all proposed extinction rates. This accommodates their overall higher diversification rates, and prevents artifacts in the near-present due to the congruence class requiring λ_0 to be equal across all models.

S3.2 Detailed exploration of the congruent models for three empirical datasets

In this section we plot the extended constructed congruent models set for the for the New World/African true parrots (Fig. S3.5), the tyrant flycatchers (Fig. S3.6), and the woodpeckers (Fig. S3.7). That is, we show the alternative rate functions within the congruence to provide a more detailed picture of the congruence class. These alternative rate functions were used in the main text to assess the robustness of the inferred diversification rate trends.

Most of the woodpecker alternative models exhibit the same three major trends in the woodpecker speciation history: an increase around 25 Ma, and another increase around 18 Ma, and a decrease starting around 8 Ma (Fig. S3.7). All alternative models within the congruence class agree on the overall signal of two increases, which is consistent with previous findings of a burst in speciation for the true woodpeckers (Picinae), and a later increase for the piculets (Picumninae, Shakya et al., 2017). Most alternative models show no other significant trends, with the exception of the modal models. For those, we were able to induce a short-lived increase, followed by a short-lived decrease in speciation rate (Fig. S3.7, M2-M3). Similarly, in the diversification rates of the flycatcher dataset, there is one strong rate shift at around 40 Ma, which is unanimously recovered by the congruence class (Fig. S3.6). The parrot diversification history exhibits relatively constant rates and the alternative models showed conflicting trends (Fig. S3.5). When the rates are constant, it is trivial to propose an alternative congruent model that contradicts the diversification pattern of the reference model.

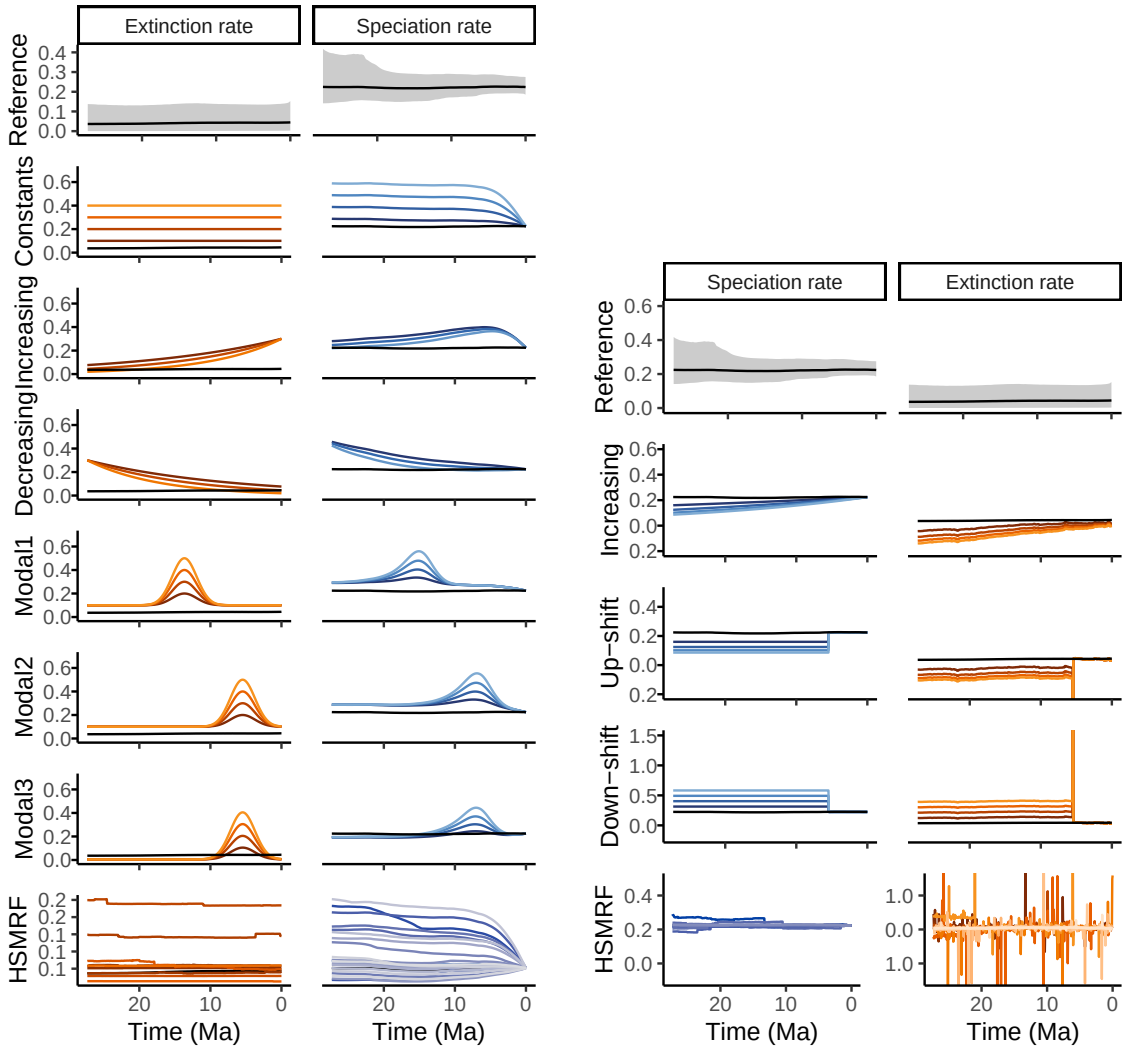


Figure S3.5: Congruent birth-death models for the New World/African true parrots (Psittacidae), based on a time-calibrated phylogenetic tree from Condamine et al. Condamine et al. (2019). Top row: the episodic birth-death model fitted in `RevBayes`, where the posterior median is in black, and the 95% credible interval in grey. The left panel exhibits models that are constructed by proposing alternative extinction rates. The right panel exhibits models that are constructed using alternative speciation rates. Each color-shade pair is one model: for example dark orange and dark blue. All models have the same likelihood. Included extinction rate shapes are exponentially increasing, exponentially decreasing, and a modal extinction event. The speciation rate shapes are linear increases, an instantaneous up-shift, and an instantaneous down-shift. We also proposed rates that are randomly drawn from the horseshoe Markov random field (HSMRF) distribution.

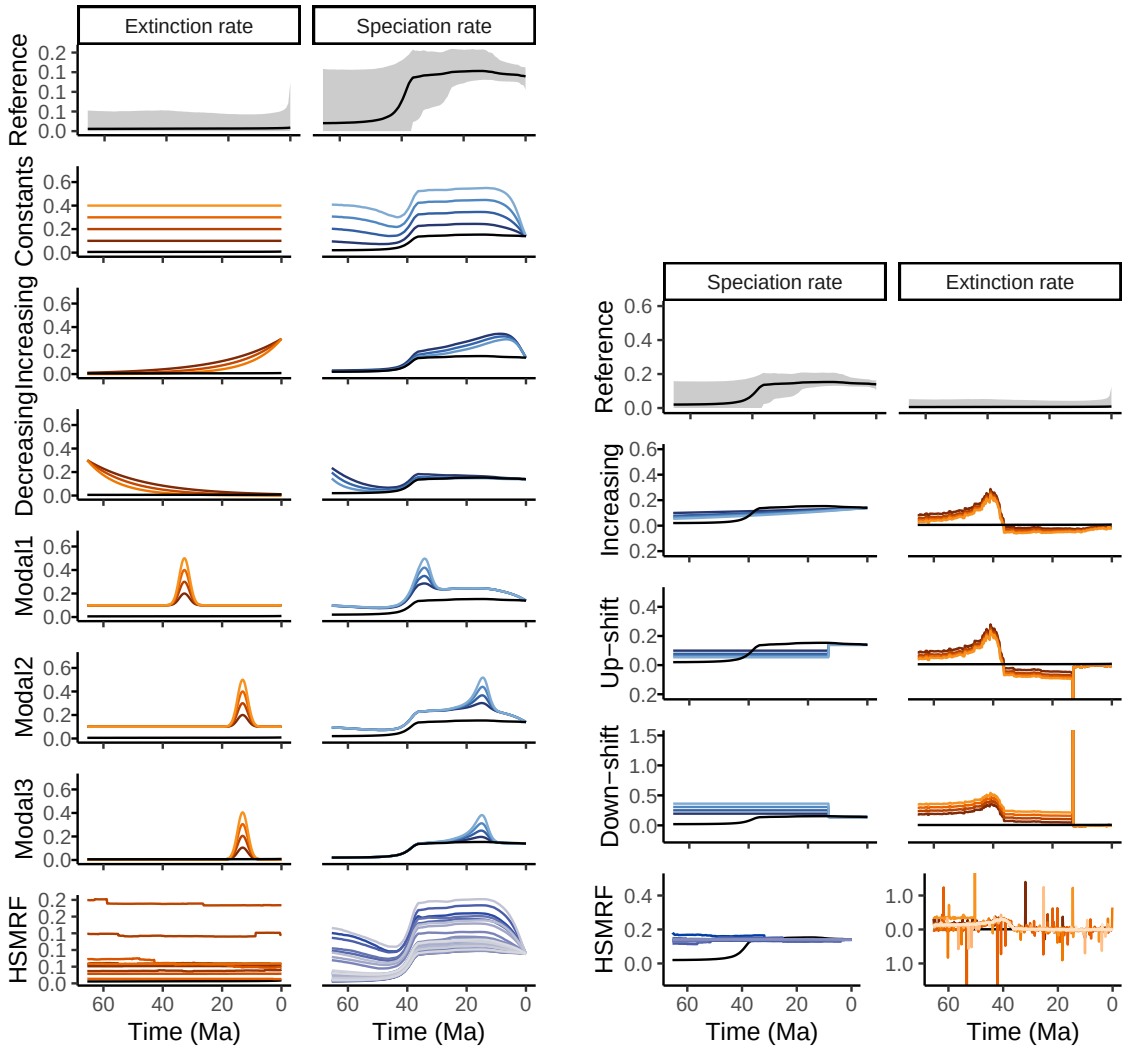


Figure S3.6: Congruent birth-death models for the tyrant flycatchers (Tyrannidae), based on a time-calibrated phylogenetic tree from Condamine et al. Condamine et al. (2019). Top row: the episodic birth-death model fitted in `RevBayes`, where the posterior median is in black, and the 95% credible interval in grey. The left panel exhibits models that are constructed by proposing alternative extinction rates. The right panel exhibits models that are constructed using alternative speciation rates. Each color-shade pair is one model: for example dark orange and dark blue. All models have the same likelihood. Included extinction rate shapes are exponentially increasing, exponentially decreasing, and a modal extinction event. The speciation rate shapes are linear increases, an instantaneous up-shift, and an instantaneous down-shift. We also proposed rates that are randomly drawn from the horseshoe Markov random field (HSMRF) distribution.

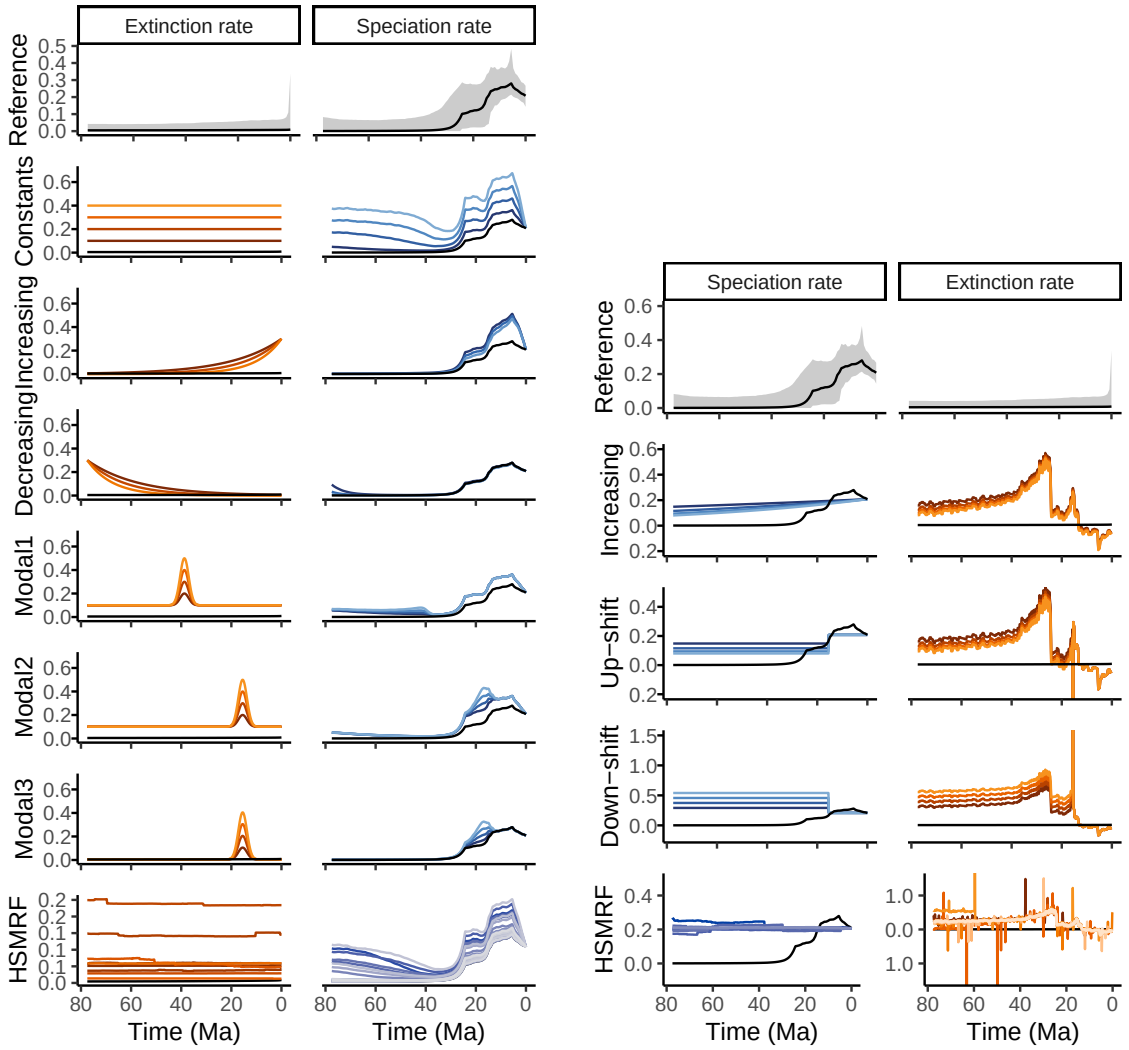


Figure S3.7: Congruent birth-death models for the tyrant woodpeckers (Picidae), based on a time-calibrated phylogenetic tree from Condamine et al. (2019). Top row: the episodic birth-death model fitted in `RevBayes`, where the posterior median is in black, and the 95% credible interval in grey. The left panel exhibits models that are constructed by proposing alternative extinction rates. The right panel exhibits models that are constructed using alternative speciation rates. Each color-shade pair is one model: for example dark orange and dark blue. All models have the same likelihood. Included extinction rate shapes are exponentially increasing, exponentially decreasing, and a modal extinction event. The speciation rate shapes are linear increases, an instantaneous up-shift, and an instantaneous down-shift. We also proposed rates that are randomly drawn from the horseshoe Markov random field (HSMRF) distribution.

S3.3 Evaluating alternative trend thresholds

In the main text, and following (Höhna et al., 2022), we proposed to assess the trends in the rate functions using the slope of the rate function. The slope is a continuous parameter, however, and can be a challenge to interpret. We chose to set a threshold of ϵ rate units per Ma, where if the slope was above the threshold, we considered the function to be significantly increasing (and analogously for decreasing rates). The choice of threshold is to some extent arbitrary. How much is a significant increase or decrease in the rate function? Choosing a threshold too small can result in picking up every small change (Fig. S3.8), and the trends will be difficult to interpret. Choosing a too large threshold (Fig. S3.10) will result in almost no changes being considered at all, and we are left with a blank, informationless figure. We recommend the researcher to test a few different thresholds that are suitable for their analysis, dataset, and relevant time scale.

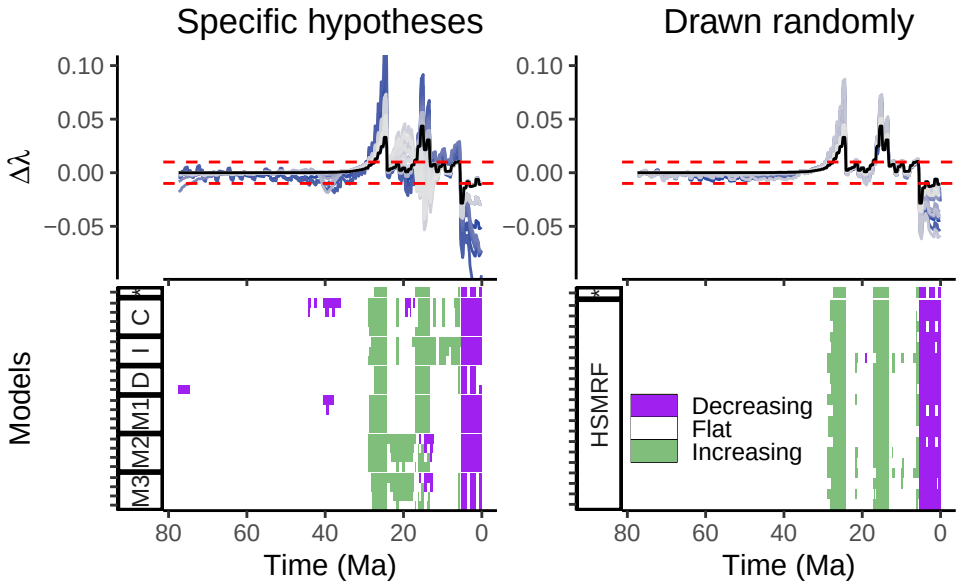


Figure S3.8: Congruent birth-death models for the tyrant woodpeckers (Picidae), based on a time-calibrated phylogenetic tree from Condamine et al. (Condamine et al., 2019), with a different threshold value ϵ to assess diversification rate trends. Top row: the slope of the speciation rate: $\Delta\lambda_i = (\lambda_i - \lambda_{i-1})/\Delta t$. The dashed red line indicates the threshold value of $\epsilon = 0.01$ rate units per Ma. Bottom row: a summary of whether each $\Delta\lambda$ is decreasing, increasing, or flat, as indicated by the threshold value. The extinction rate names are shortened, C: Constants, I: Increasing, D: Decreasing, M1-M3: Modals, HSMRF: horseshoe Markov random field, *: the reference model.

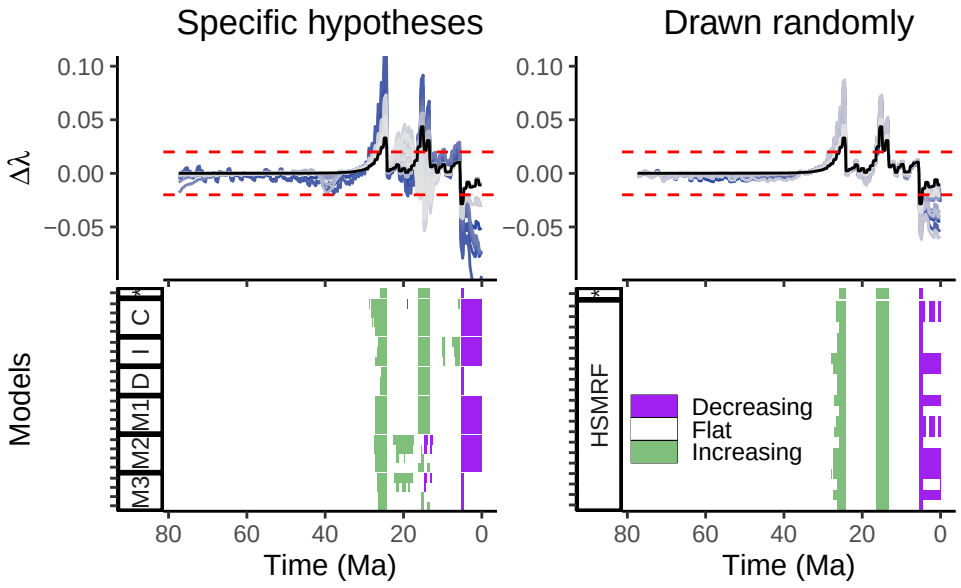


Figure S3.9: Equivalent to Fig. S3.8, however with a different threshold of $\epsilon = 0.02$ rate units per Ma. This threshold is the one we used for all other analyses and figures, unless otherwise specified.

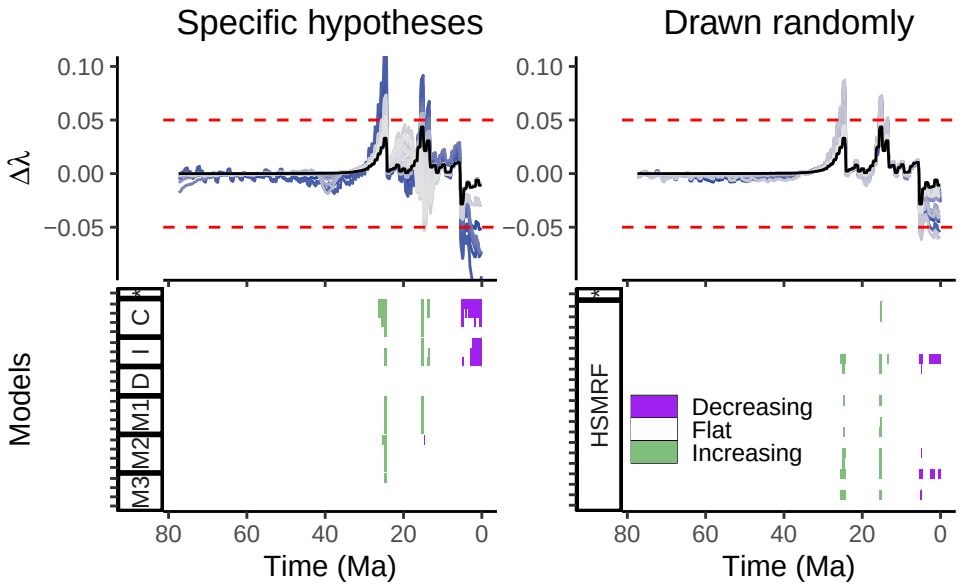


Figure S3.10: Equivalent to Fig. S3.8 and Fig. S3.9, however with a different threshold of $\epsilon = 0.05$ rate units per Ma. We did not use this threshold for any other analyses.

S3.4 Posterior samples

In this section we plot the first twenty samples from the posterior distribution (Fig. S3.11) that we used for the analyses in the main text (Fig. 3.3). This gives an indication of how variable the main temporal patterns in the rates are. For example, we can see that the samples from the woodpecker (Picidae) analysis largely display a strong increase in speciation rate, but the timing of when the first shift happened is uncertain, between 35 and 15 Ma (Fig. S3.11, bottom left panel). For this reason, we used a larger temporal window to calculate the directional trends from the posterior-distribution analyses, as opposed to the analyses using the posterior median. We computed the directional trend as $\Delta\lambda_i = (\lambda_i - \lambda_{i-k})/(t_i - t_{i-k})$, where k is the window size. In Fig. S3.12 we show how different choices of window sizes ($k \in 10, 25, 55, 80$) impact how the directional trends are portrayed.

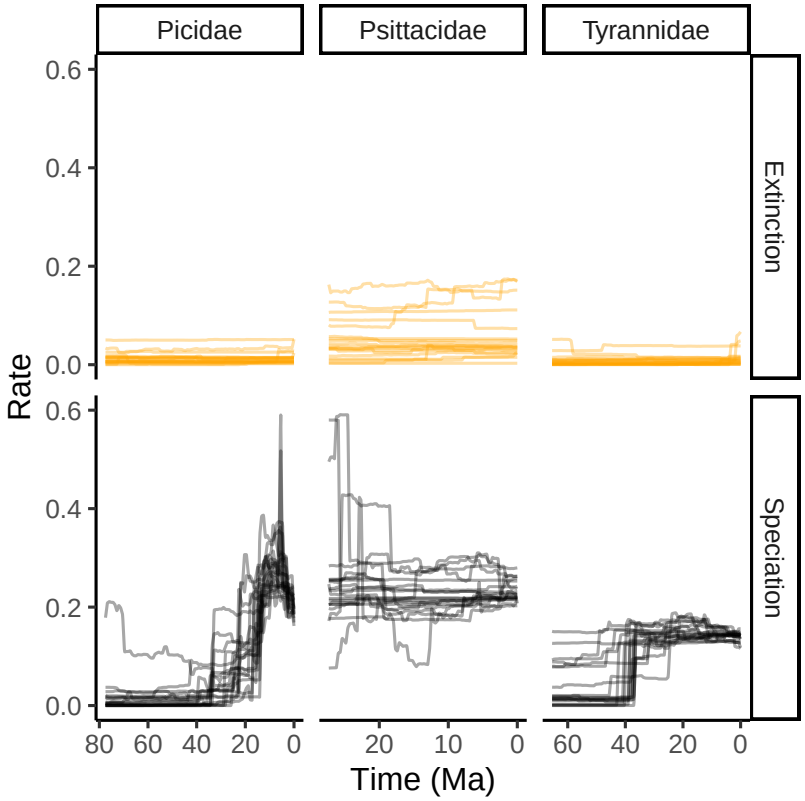


Figure S3.11: Speciation and extinction rate trajectories for the first twenty posterior samples used in assessing the trends in the congruence class when we take into consideration the uncertainty of the posterior estimates (main text Fig. 3.3).

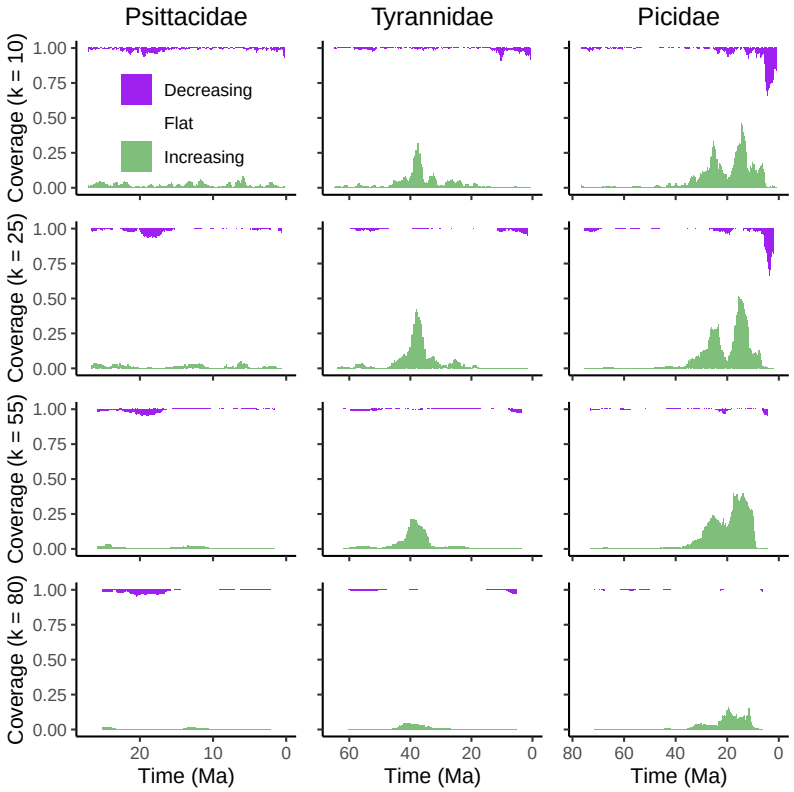


Figure S3.12: Directional trends in the speciation rate under 110 equidistant samples from the posterior distribution. Equal to main text Fig. 3.3, except with a variety of choices for the window size k, used when computing the slope in the speciation rate ($\Delta\lambda_i = (\lambda_i - \lambda_{i-k})/(t_i - t_{i-k})$)).

S3.5 Convergence assessment

In this section we assess whether the Markov chain Monte Carlo (MCMC) simulation converged for the three datasets used in the main text: the parrots (Fig. S3.13), the tyrant flycatchers (Fig. S3.14), and the woodpeckers (Fig. S3.15). We plot the speciation and extinction rate estimates for each replicate run (Fig. S3.13 right), (Fig. S3.14 right) and (Fig. S3.15 right), and we perform the Kolmogorov-Smirnov (KS) test to assess whether the samples drawn from the posterior are reproducible across replicates. The KS-test is a measure of whether the samples could be considered to be drawn from the same underlying distribution, that is, the MCMC sampler has converged to the same posterior distribution (Fabreti and Höhna, 2022). For two equally distributed random variables, with 625 independent samples from each, we expect the KS-test to be false positive ($D > D_{\text{crit}} = 0.09$) in 1% of repeated comparisons (Fabreti and Höhna, 2022). Our computed KS scores are well below D_{crit}, which indicates that our samples have converged. The posterior median and credible intervals are similar to the across all four replicates for the woodpeckers, the tyrant flycatchers, and the parrots (Figs. S3.13 to S3.15).

For the analyses presented in the main text and otherwise in the supplementary materials,

we pooled the four runs by concatenating the samples before computing the posterior medians and the credible intervals. This ensures more precise estimates in comparison to using each replicate individually. We conclude that our MCMC simulations converged towards the same parameter space and that our posterior median estimates are sufficiently precise for identifying the main time-varying patterns in the model.

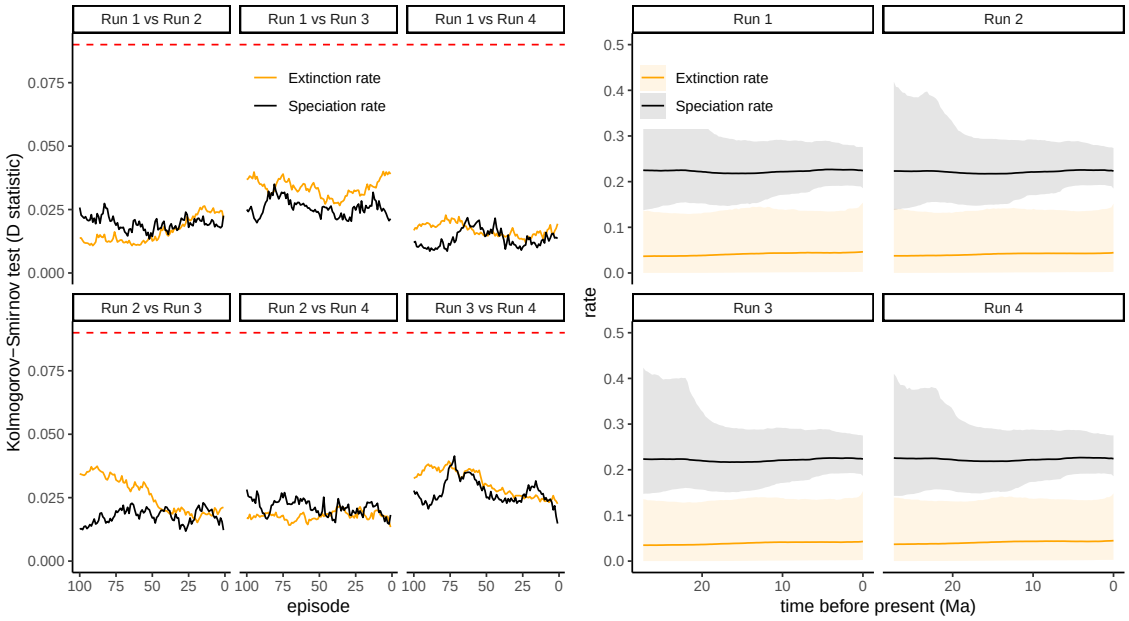


Figure S3.13: Convergence assessment for the New World/African true parrots, where we fitted an episodic birth-death model with a HSMRF (horseshoe Markov random field) distribution to model temporal autocorrelation in the rates. The analysis was performed in `RevBayes` with four replicates. We discarded the first 5000 iterations as burnin, ran the MCMC for another 50,000 iterations, and sampled every 10 iterations. The pairwise Kolmogorov-Smirnov test is an assessment of whether the two set of samples are drawn from the same distribution. Since we have $D < D_{\text{crit}} = 0.09$ (dashed red line), our replicates are similar, and the runs converged.

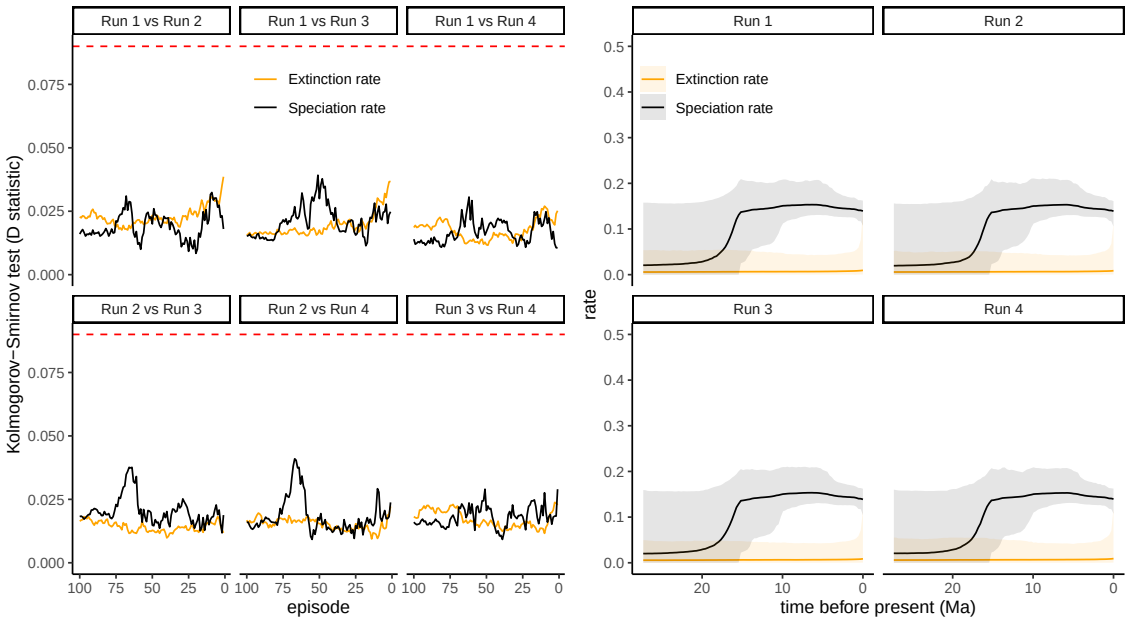


Figure S3.14: Convergence assessment for the tyrant flycatcher analysis, where we fitted an episodic birth-death model with a HSMRF (horseshoe Markov random field) distribution to model temporal autocorrelation in the rates. The analysis was performed in `RevBayes` with four replicates. We discarded the first 5000 iterations as burnin, ran the MCMC for another 50,000 iterations, and sampled every 10 iterations. The pairwise Kolmogorov-Smirnov test is an assessment of whether the two set of samples are drawn from the same distribution. Since we have $D < D_{\text{crit}} = 0.09$ (dashed red line), our replicates are similar, and the runs converged.

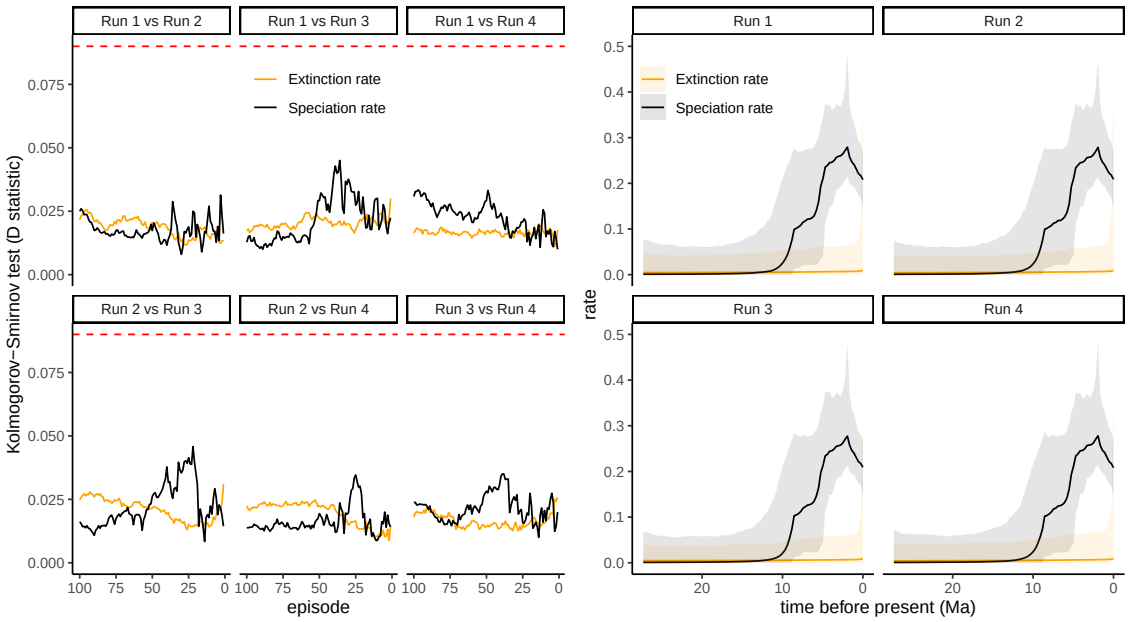


Figure S3.15: Convergence assessment for the woodpecker analysis, where we fitted an episodic birth-death model with a HSMRF (horseshoe Markov random field) distribution to model temporal autocorrelation in the rates. The analysis was performed in `RevBayes` with four replicates. We discarded the first 5000 iterations as burnin, ran the MCMC for another 50,000 iterations, and sampled every 10 iterations. The pairwise Kolmogorov-Smirnov test is an assessment of whether the two set of samples are drawn from the same distribution. Since we have $D < D_{\text{crit}} = 0.09$ (dashed red line), our replicates are similar, and the runs converged.

S3.6 Congruent speciation rates are more constrained than extinction rates

In the main text, and elsewhere in the supplementary material, we assess how the trends in rates are distributed throughout the congruence class. In doing so, we looked at many different samples of the congruence class. Our general impression was that, when we proposed speciation rates and inferred extinction rates ($\lambda' \rightarrow \mu'$), the variation in trends and absolute rates was greater than the reverse approach ($\mu' \rightarrow \lambda'$). Due to the restriction that all speciation rates must be equal at the present (λ_0) for all models in the congruence class, we suspected this to be the reason for why the inferred speciation rates were more constrained (λ') than the inferred extinction rates (μ').

In order to test our suspicion of μ' being more variable than λ', we simulated rate functions to generate a large number of congruent model, as can be seen in Fig. S3.16. We selected a reference model with constant extinction rate ($\mu(t) = 0.35$), and a time-varying speciation rate that increases sigmoidally through time ($\lambda(t) = 0.1 + 0.5/[1 + 0.1^{-t+h/2}]$, where h is the total time span). We drew rate functions from three distributions: a Brownian motion, an Ornstein-Uhlenbeck process, and independently from a lognormal distribution. For example, we would simulate speciation rates under a Brownian motion ("Proposed λ'", Fig. S3.16), and set up the congruent models and thereby get the corresponding extinction rate (μ'). Next, we calculated the standard deviation among the extinction rate values in the 1000 congruent models, for each period in time, and plot these (SD[μ']). In all of our models, the variance of the inferred extinction rates (μ') is greater than the variance of the inferred speciation rates (λ'), supporting our view that the distribution of speciation rates is more narrow within the congruence class, compared to extinction rates. Further, the variance of the inferred speciation rates are smaller at the change-point in the sigmoidal reference model (where $d\lambda/dt$ in the reference is the greatest). This provides further evidence that a period of rapid change is a reliable signal, without the need to invoke the topic of trends in rates. Notice when we proposed extinction rates (μ'), we restricted our samples to have a fixed value at the present ($\mu'(0) = 0.6$). When proposing speciation rates, this is necessary for the models to be congruent ($\lambda'(0) = \lambda_0$), however when proposing extinction rates this is not necessary. We imposed this constraint so as to make both proposal methods comparable. If we had removed this restriction, for example by drawing the initial value of the Brownian motion from a random distribution, we expect the variance of the corresponding speciation rates ($\mu' \rightarrow \lambda'$), and thus the difference between the two approaches, to be even greater.

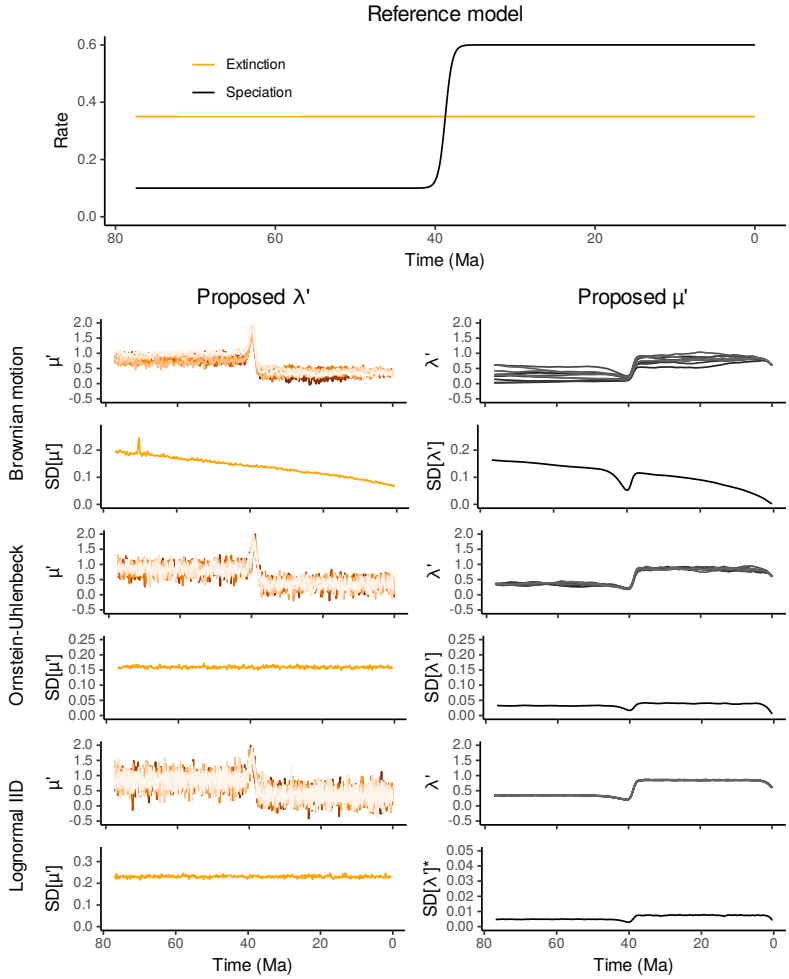


Figure S3.16: We set up a hypothetical reference model (top panel) with a sigmoidally increasing speciation rate ($\lambda(t) = 0.1 + 0.5/(1 + 0.1^{-t+h/2})$, where h is the total time span), and a constant extinction rate ($\mu(t) = 0.35$). In the following rows we proposed rates that were drawn from a Brownian motion (with $\sigma = 0.02$), an Ornstein-Uhlenbeck process (with $\sigma = 0.05$, $\alpha = 0.5$, $\theta = 0.6$), and rate values drawn independently from a lognormal distribution (with mean = 0.6, logsd = 0.05). For all proposals we set the initial value at $t = 0$ equal to 0.6. The exact rate proposals are not shown, however we show the corresponding rates of models that are congruent to the reference model (μ' and λ'). We also show the standard deviation (SD) of the inferred rates, for each time interval. When we proposed alternative extinction rates (μ') and inferred the corresponding speciation rate (λ'), the standard deviation is always smaller than the reverse (proposing λ' and inferring μ'). Further, the inferred speciation rates are less varied at the shift in the sigmoidal reference function (where $d\lambda/dt$ is the smallest), compared to the neighbouring regions in time. The Brownian motion diverges unbounded backwards in time, so we expect to see increasing variance with time. The Ornstein-Uhlenbeck process is in an equilibrium state, and so we expect the variance of the inferred rates not to vary with time. The independent lognormal draws also do not vary systematically with time. We show fifteen sampled rate functions, however to compute the variance we used a more extensive sample of 1000 rate functions for each configuration. We used a piecewise-linear function with 500 time points to represent the rate functions. *: Note the altered y-axis limits.

S3.7 Hypothetical models: alternative extinction rate

In the main text (Fig. 3.3) we showed three hypothetical diversification histories, and we proposed eight alternative categories of rate shapes that we used to construct congruent models. An extended version of Fig. 3.3 can be found in Fig. S3.17, where we also show the trends in the net-diversification rate. In this section we expand the three classes to eight sets of congruent models (Fig. S3.18). In all cases we used a constant extinction rate of $\mu = 0.28$. These congruent models exhibit a variety of inferred speciation rate trajectories, which we depict in this section (Figs. S3.19 to S3.26). For all datasets that show a rapid change in the speciation rate in the reference model, then all alternative models also show the same rapid change in speciation rates as shown by our assessment of directional trends. Conversely, if the reference speciation rate does not show a directional trend (i.e., the rate is approximately constant), then either the alternative speciation rates also showed no directional trend or the trends among alternative models were in disagreement.

We observed the strongest disagreement of trends for the speciation rate intervals near the present, especially for the constant reference model. The conflicting trends correspond directly to the scale of the proposed extinction rate relative to the reference speciation rate ($\lambda = 0.35$, panel C). The proposed extinction rates that are larger than the reference extinction rate $\mu = 0.28$ at the near-present (rates L+, S+, E+, and some constant rate functions) infer models with decreasing speciation rates near the present (Fig. S3.18, panel C and Fig. S3.19). Conversely, proposed extinction rates that are at present smaller than $\mu = 0.28$ infer models with increasing speciation rates near the present (Fig. S3.18, panel C and Fig. S3.19). Similar patterns can be seen at the near-present for all of the other congruence classes (Figs. S3.20 to S3.26). As we observed in our empirical datasets, if the reference speciation rate model was constant, then we observe no fully supported trends in the congruence class (Fig. S3.18, panel C). We see that models with moderate speciation rate changes do not show shared patterns across the congruence class (panels L+, L–). Reference models with rapidly changing rates show unambiguous agreement for their correspond trend at the shift time (panels S+, S–). The robustness of the trends regarding the congruence class weakens when the rate change is less steep (panels M, E+, E–).

The overall results of the hypothetical cases confirm the conclusions from the empirical datasets: changes that happen slowly over time are not robust to the congruence class. However, changes that happen abruptly in a small time window are indeed robust.

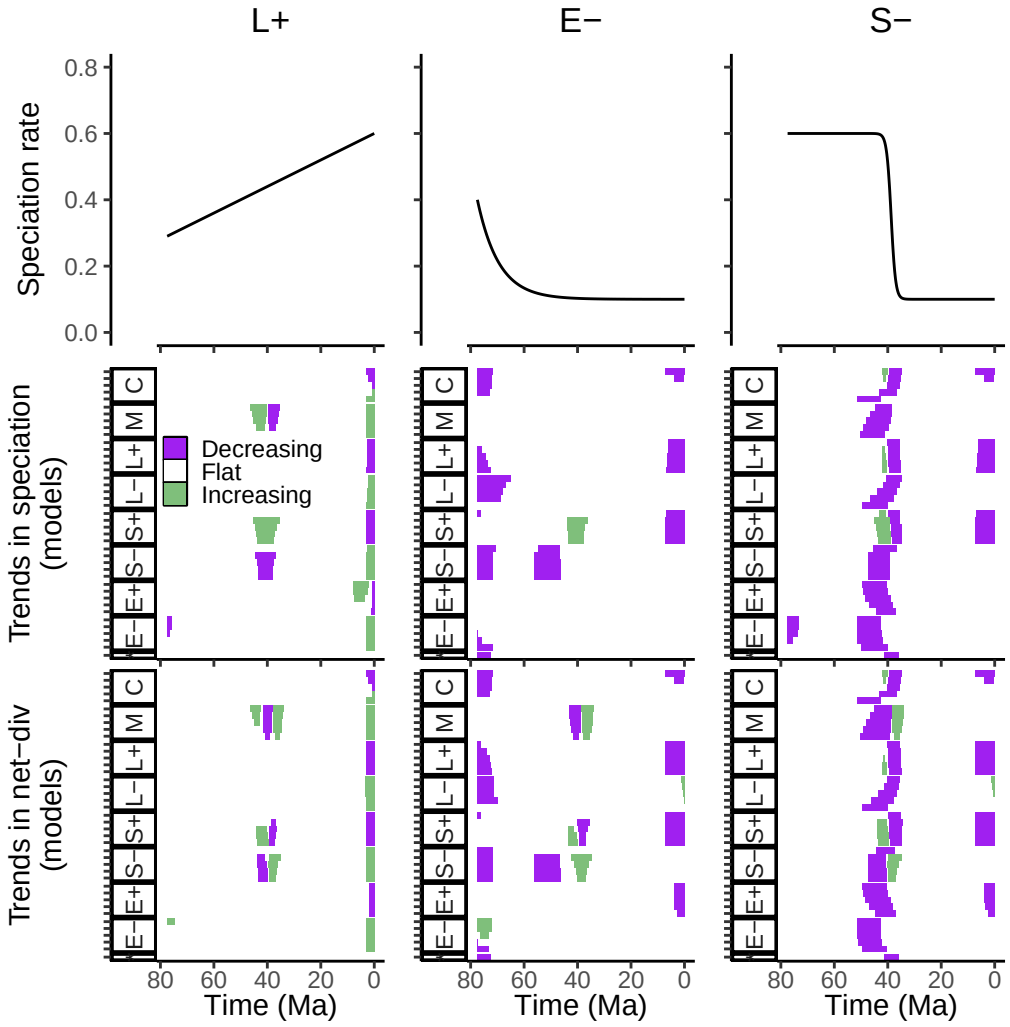


Figure S3.17: This figure corresponds to the main text Fig. 3.3, except we have added a third row of panels, where we also show the trends in the net-diversification rates in the sample of the congruence class $(\lambda(t) - \mu(t))$.

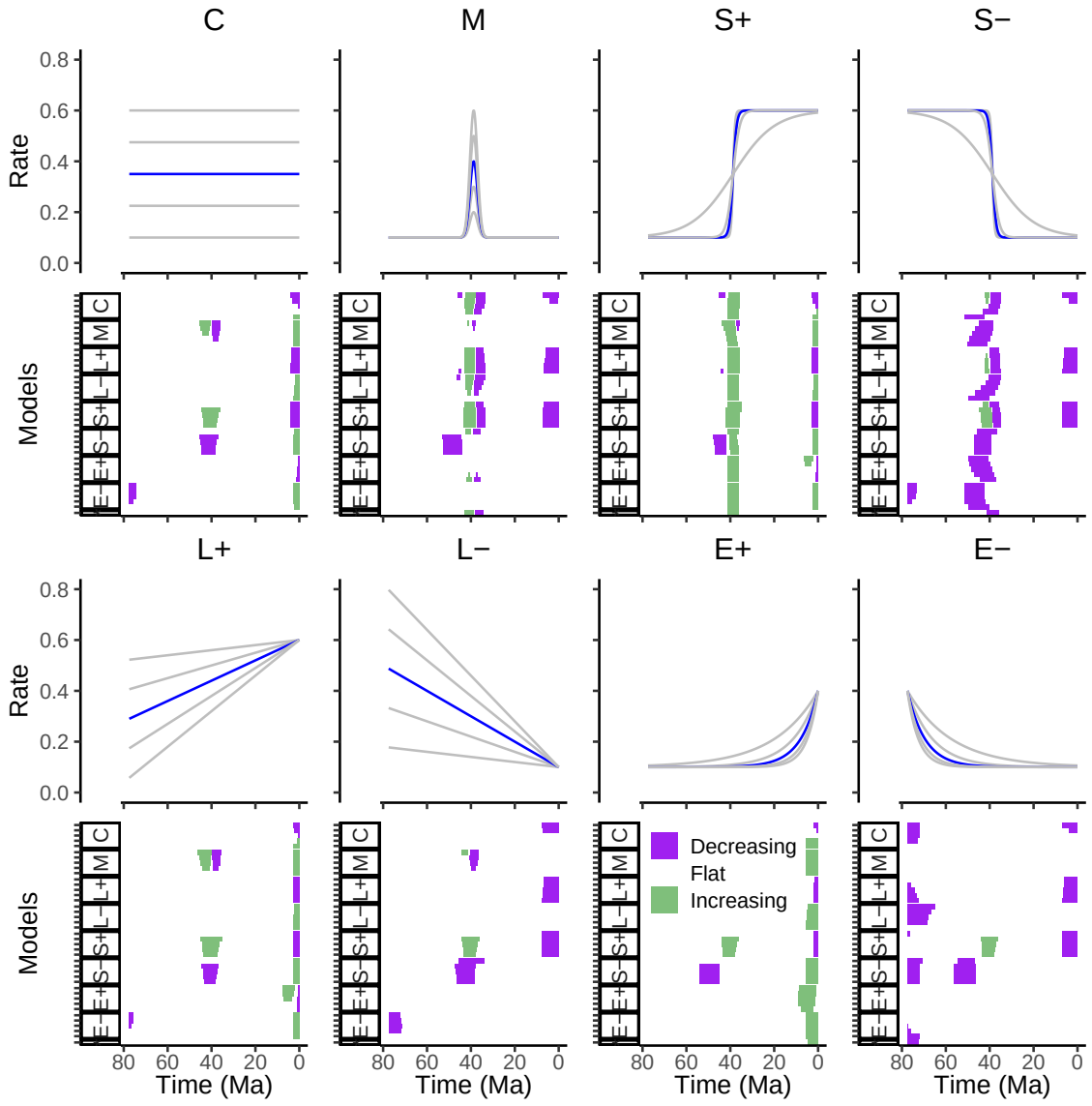


Figure S3.18: A collection of hypothetical scenarios. Each panel-pair represents a separate congruence class. The reference model for the congruence class is the speciation rate depicted in blue, and in all cases a constant extinction rate of $\mu = 0.28$. For each congruence class, we proposed various alternative extinction rates: constants (C), modals (M), sigmoidal increase (S+), sigmoidal decrease (S–), linear increase (L+), linear decrease (L–), exponential increase (E+) and exponential decrease (E–). Each congruence class results in a set of models with inferred speciation rate curves (Figs. S3.19 to S3.26). We computed the significant directional trends for the inferred speciation rates (with a threshold of $\epsilon = 0.02$ rate units per Ma) for each congruence class. We selected a time scale similar to the woodpecker phylogeny for interpretability.

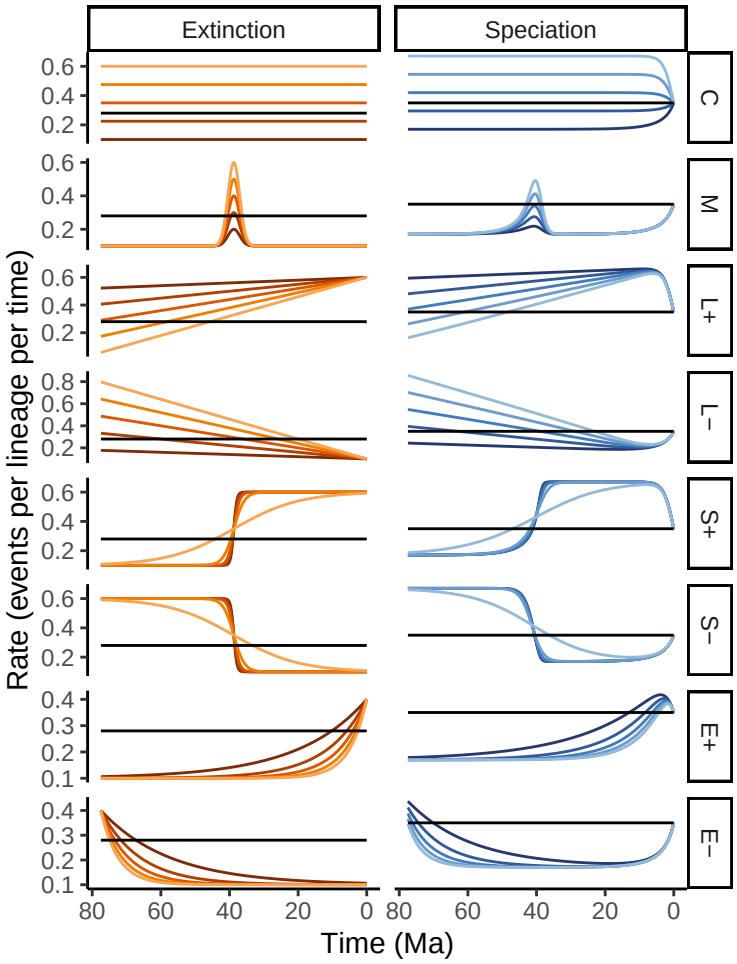


Figure S3.19: Hypothetical rate scenarios for the constant birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and a constant speciation rate function corresponding to (Fig. S3.18, C, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

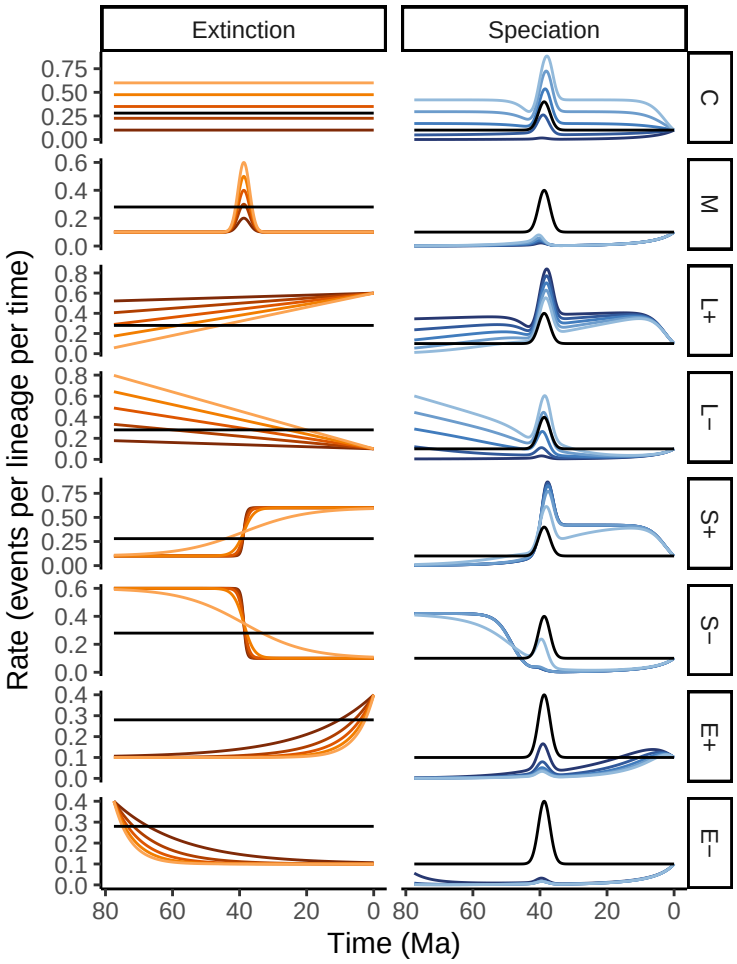


Figure S3.20: Hypothetical rate scenarios for the modal birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and a modal speciation rate function corresponding to (Fig. S3.18, M, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

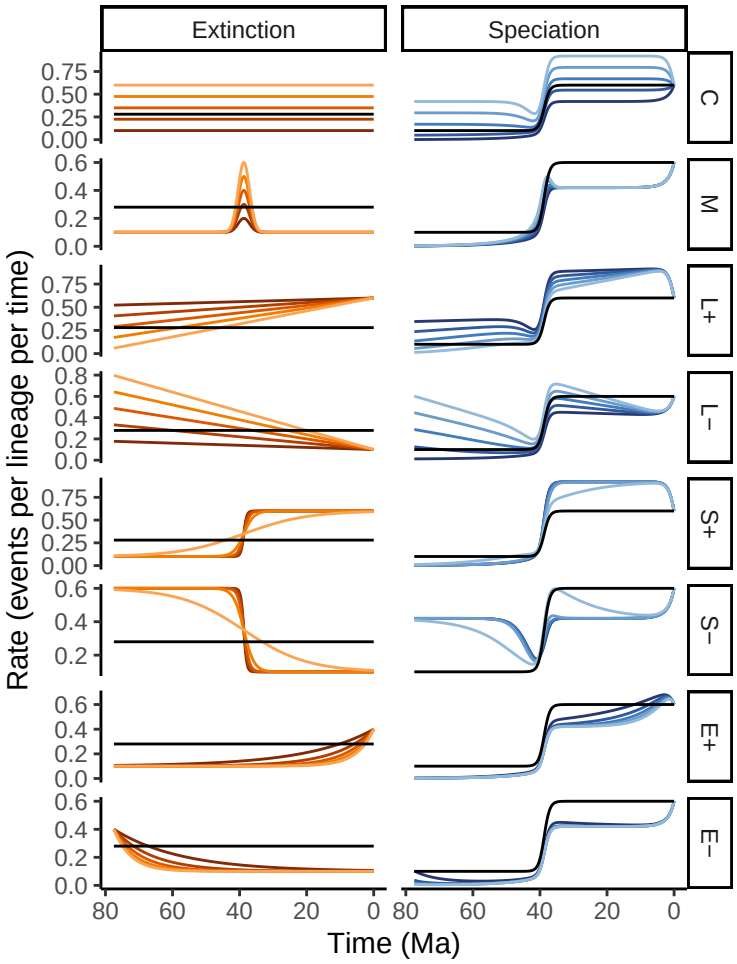


Figure S3.21: Hypothetical rate scenarios for the sigmoidally increasing birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and a sigmoidally increasing speciation rate function corresponding to (Fig. S3.18, S+, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

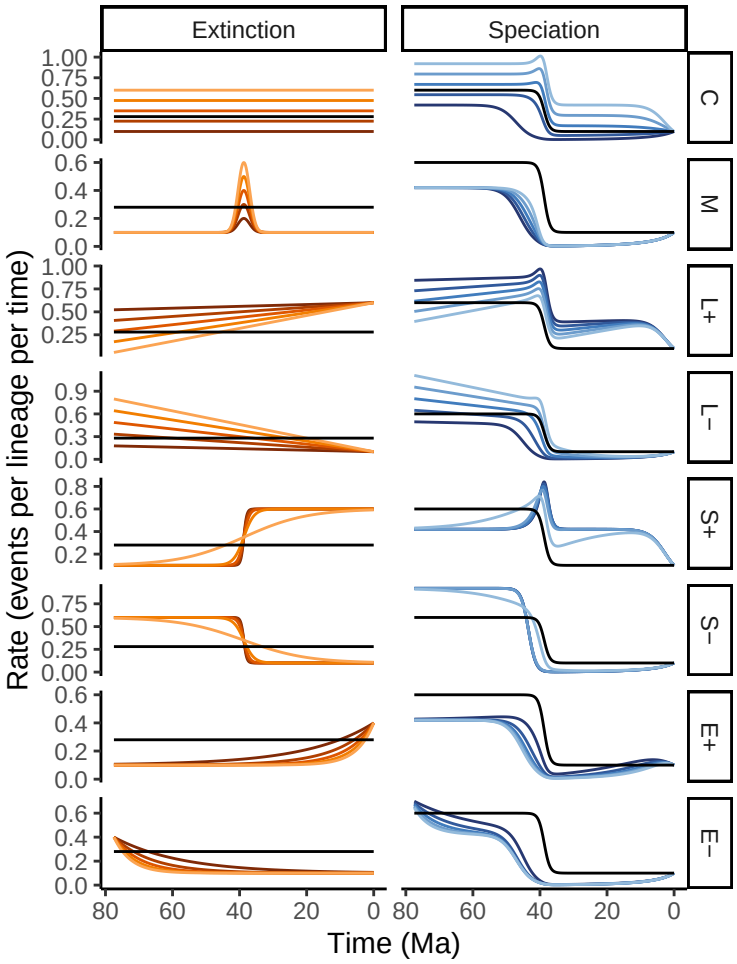


Figure S3.22: Hypothetical rate scenarios for the sigmoidally decreasing birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and a sigmoidally decreasing speciation rate function corresponding to (Fig. S3.18, S–, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

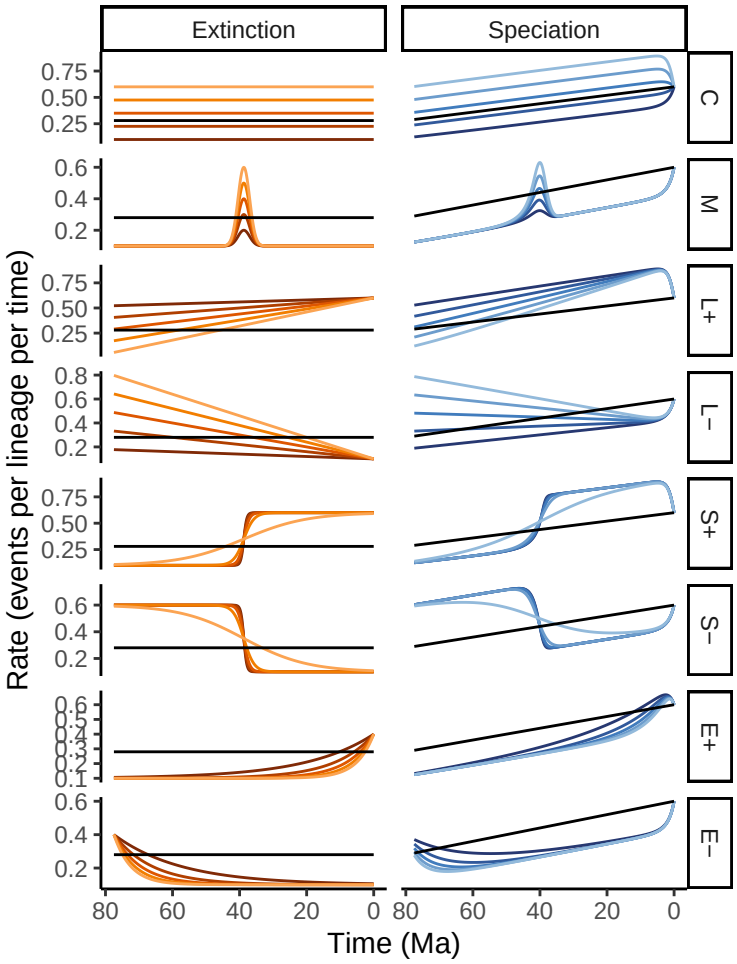


Figure S3.23: Hypothetical rate scenarios for the linearly increasing birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and a linearly increasing speciation rate function corresponding to (Fig. S3.18, L+, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

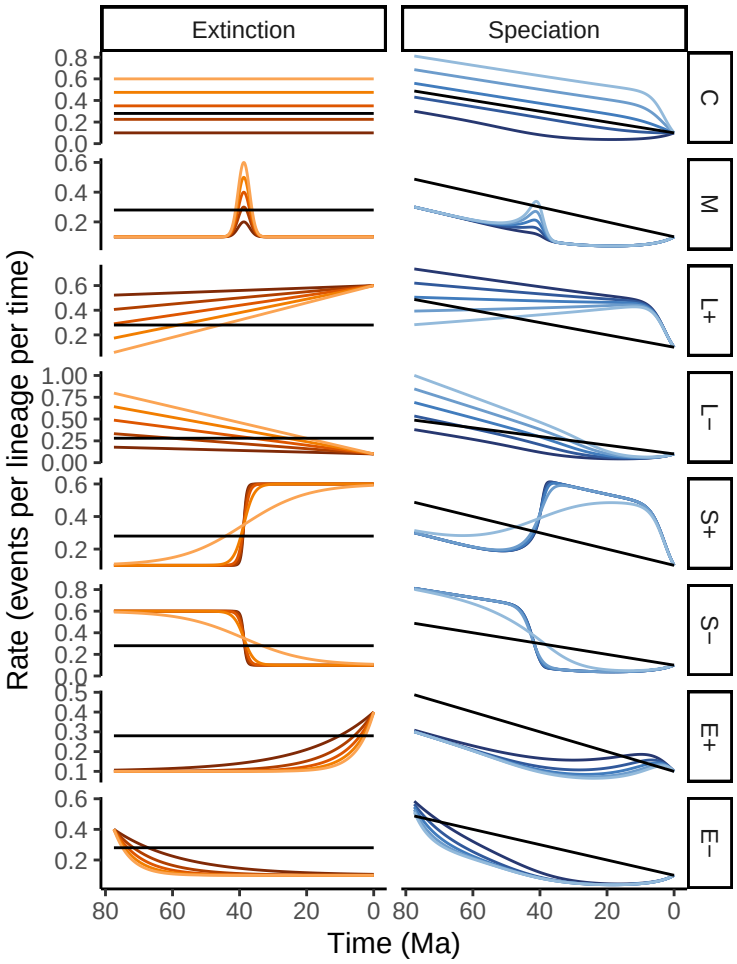


Figure S3.24: Hypothetical rate scenarios for the linearly decreasing birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and a linearly decreasing speciation rate function corresponding to (Fig. S3.18, L–, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

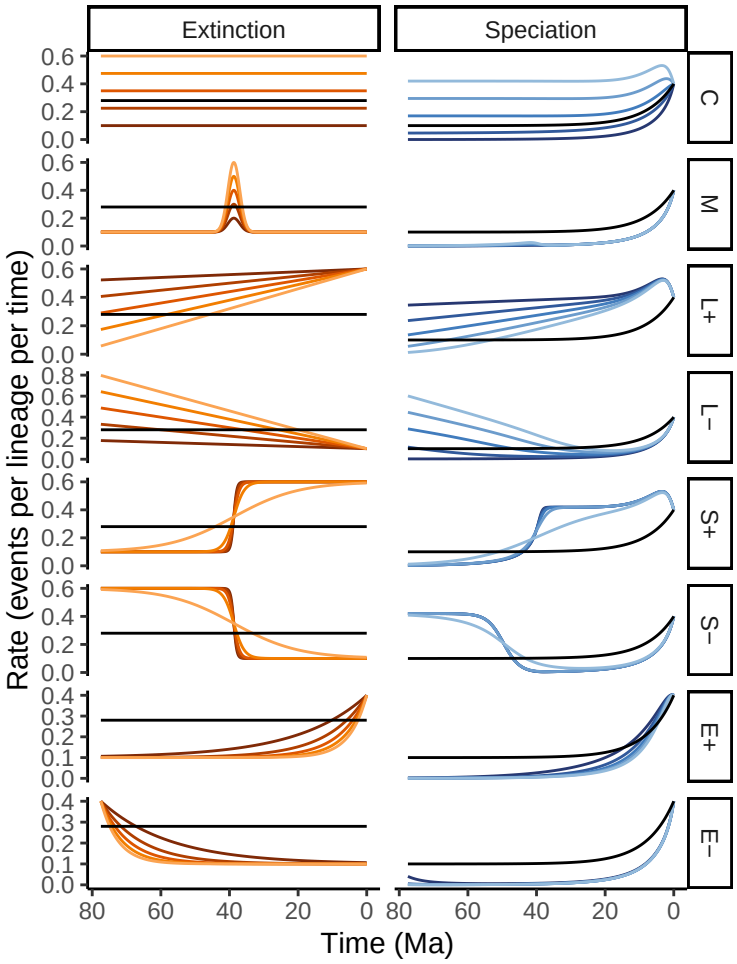


Figure S3.25: Hypothetical rate scenarios for the exponentially increasing birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and an exponentially increasing speciation rate function corresponding to (Fig. S3.18, E+, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

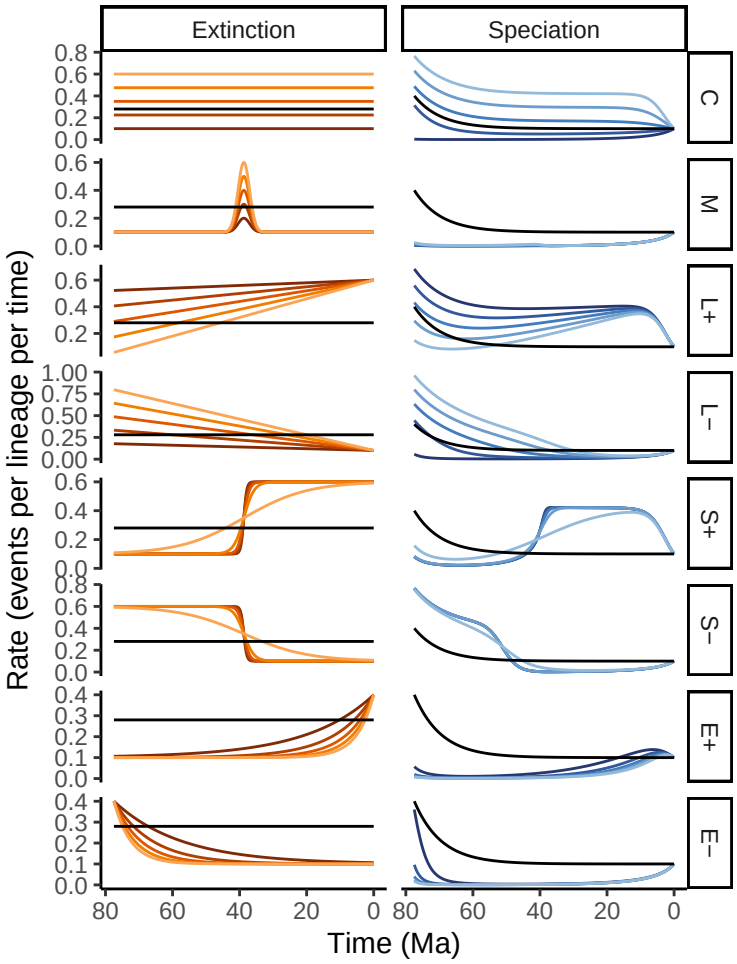


Figure S3.26: Hypothetical rate scenarios for the exponentially decreasing birth-death model in Fig. S3.18. The reference model is depicted in black in all panels: a constant extinction rate ($\mu = 0.28$), and an exponentially decreasing speciation rate function corresponding to (Fig. S3.18, E-, red line). Five alternative extinction rates are proposed in each row, including constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent speciation rates.

S3.8 Hypothetical models: alternative speciation rate

In the previous section, we explored hypothetical models where we proposed alternative extinction rate functions. In this section, we set up hypothetical models where we instead propose alternative speciation rate functions. As discussed in the main text, there is an additional constraint to be considered when proposing speciation rate functions. Namely, the speciation rate at the present (λ_0) must be equal for all models in the congruence class. Since we could not vary the constant-rate model with different intercepts according to this constraint, we omitted this group of rate functions. In all seven congruence classes, we constructed a reference with constant speciation rate ($\lambda = 0.28$), and for the extinction rate the red trajectories seen in (Fig. S3.27). Next, we proposed a variety of speciation rate trajectories, including a modal event (Fig. S3.28), sigmoidally increasing (Fig. S3.29), sigmoidally decreasing (Fig. S3.30), linearly increasing (Fig. S3.31), linearly decreasing (Fig. S3.32), exponentially increasing (Fig. S3.33), and exponentially decreasing (Fig. S3.34).

In comparison to our empirical examples, we were able to propose more valid alternative speciation rate functions (that did not result in negative extinction rate functions). This may be due to our reference extinction rates (Fig. S3.27, red lines) being larger than those estimated from our empirical datasets, and thus there is more room for variation in the extinction rate before it goes negative. Note that we used a constant reference speciation rate, and because all models have the same λ_0, the linear and exponentially increasing alternative speciation rate functions (L+ and E+) as well as the up-shift (sigmoidal, S+) produce necessarily lower alternative speciation rate functions.

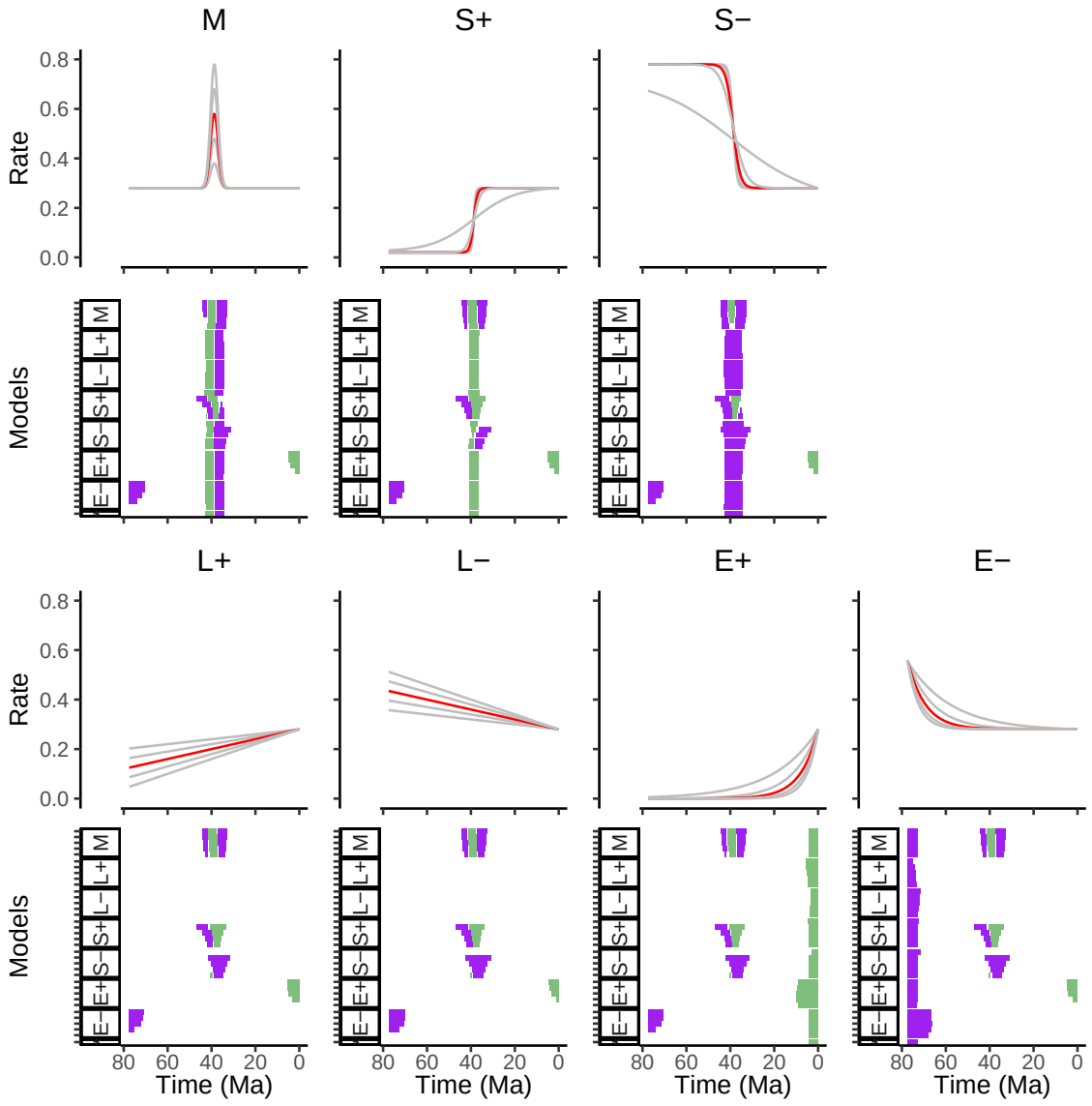


Figure S3.27: A collection of hypothetical models. Each panel-pair represents a separate congruence class. The reference model for the congruence class is the extinction rate depicted in red, and in all cases a constant speciation rate of $\lambda = 0.28$. For each congruence class, we proposed various alternative speciation rates: modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+) and exponential decrease (E–). Each congruence class results in a set of models with inferred speciation rate curves (See Figs. S3.28 to S3.34). We computed the significant directional trends for the inferred extinction rates (with a threshold of $\epsilon = 0.02$ rate units per Ma) for each congruence class. We selected a time scale similar to the woodpecker phylogeny for interpretability.

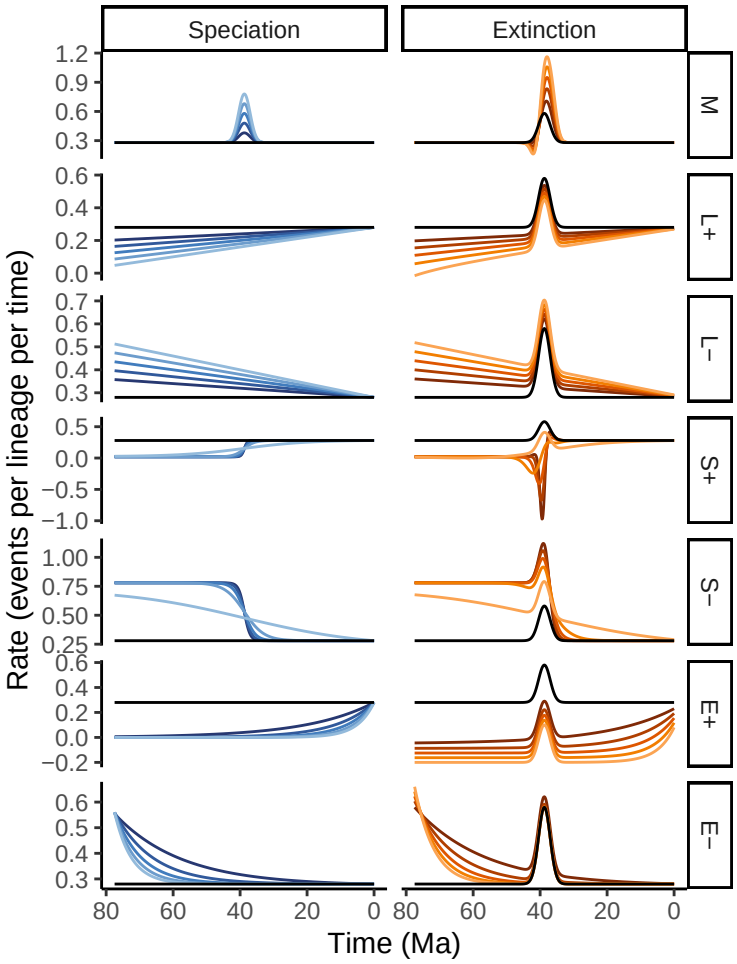


Figure S3.28: Hypothetical rate scenarios for the modal birth-death model in Fig. S3.27. The reference model is depicted in black in all panels: a constant speciation rate ($\lambda = 0.28$), and a modal extinction rate function corresponding to (Fig. S3.27, M, red line). Five alternative speciation rates are proposed in each row, including modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent extinction rates.

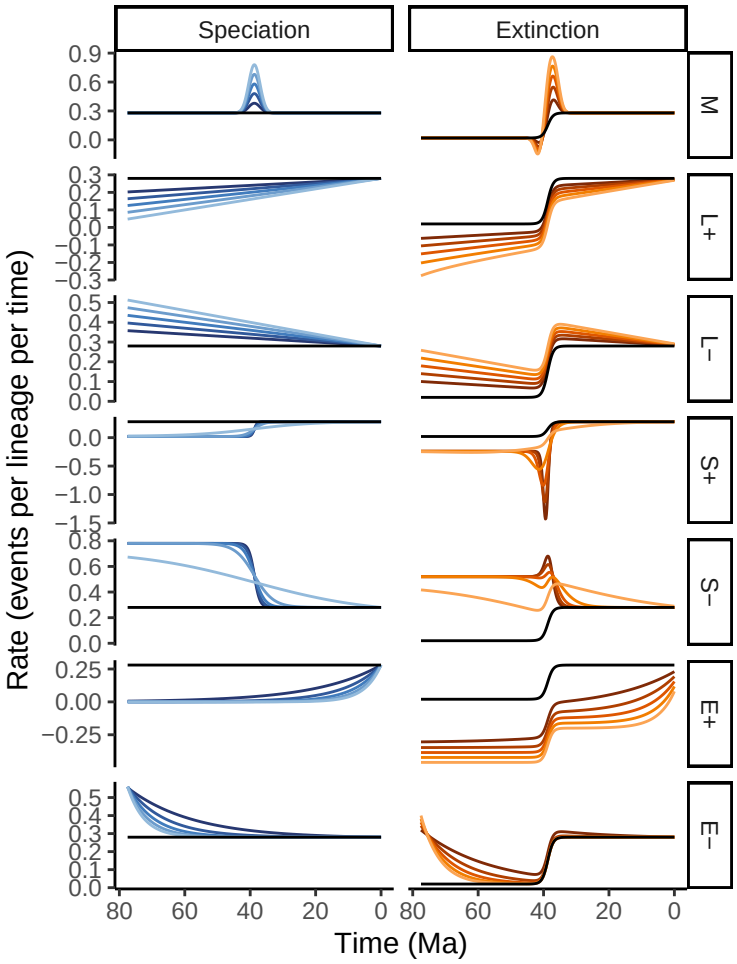


Figure S3.29: Hypothetical rate scenarios for the sigmoidally increasing birth-death model in Fig. S3.27. The reference model is depicted in black in all panels: a constant speciation rate ($\lambda = 0.28$), and a sigmoidally increasing extinction rate function corresponding to (Fig. S3.27, S+, red line). Five alternative speciation rates are proposed in each row, including modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent extinction rates.

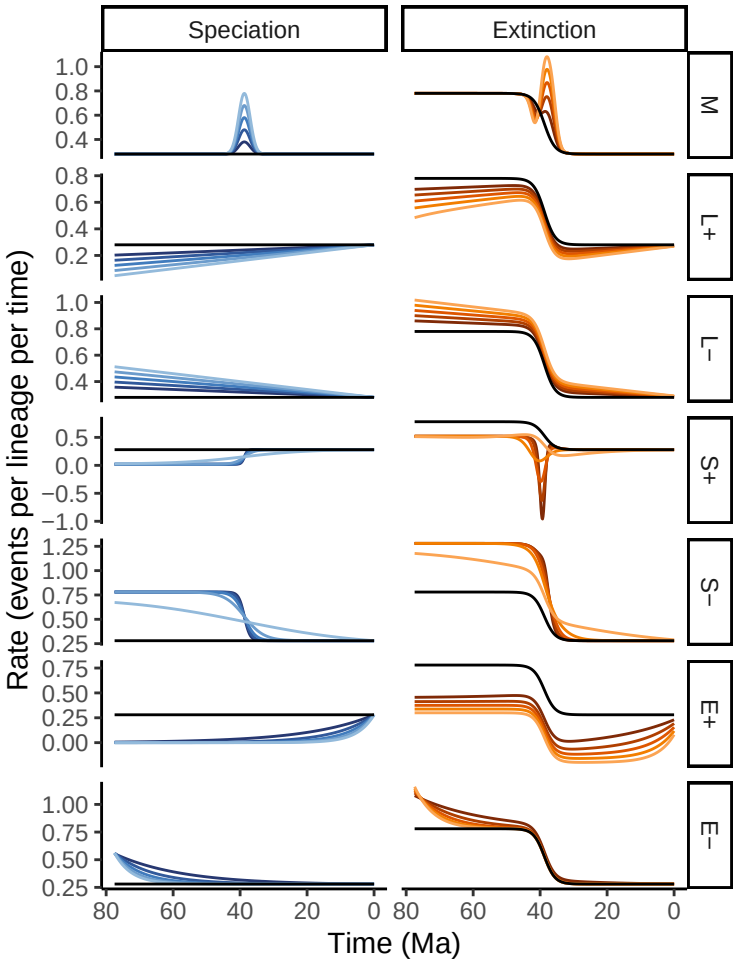


Figure S3.30: Hypothetical rate scenarios for the sigmoidally decreasing birth-death model in Fig. S3.27. The reference model is depicted in black in all panels: a constant speciation rate ($\lambda = 0.28$), and a sigmoidally decreasing extinction rate function corresponding to (Fig. S3.27, S–, red line). Five alternative speciation rates are proposed in each row, including modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent extinction rates.

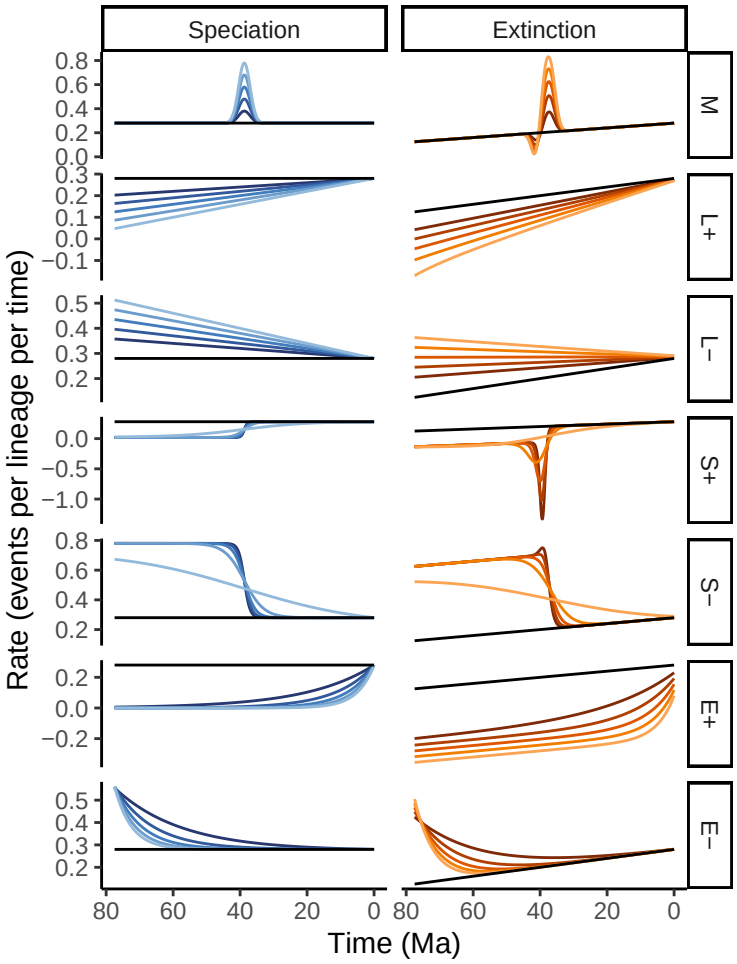


Figure S3.31: Hypothetical rate scenarios for the linearly increasing birth-death model in Fig. S3.27. The reference model is depicted in black in all panels: a constant speciation rate ($\lambda = 0.28$), and a linearly increasing extinction rate function corresponding to (Fig. S3.27, L+, red line). Five alternative speciation rates are proposed in each row, including modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent extinction rates.

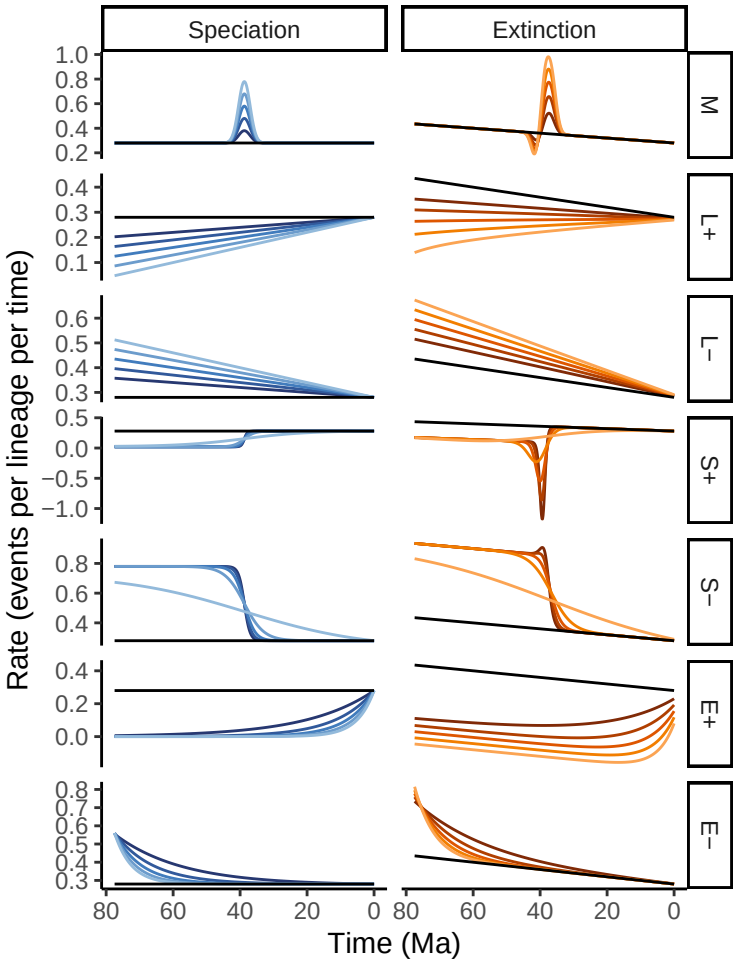


Figure S3.32: Hypothetical rate scenarios for the linearly decreasing birth-death model in Fig. S3.27. The reference model is depicted in black in all panels: a constant speciation rate ($\lambda = 0.28$), and a linearly decreasing extinction rate function corresponding to (Fig. S3.27, L–, red line). Five alternative speciation rates are proposed in each row, including modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent extinction rates.

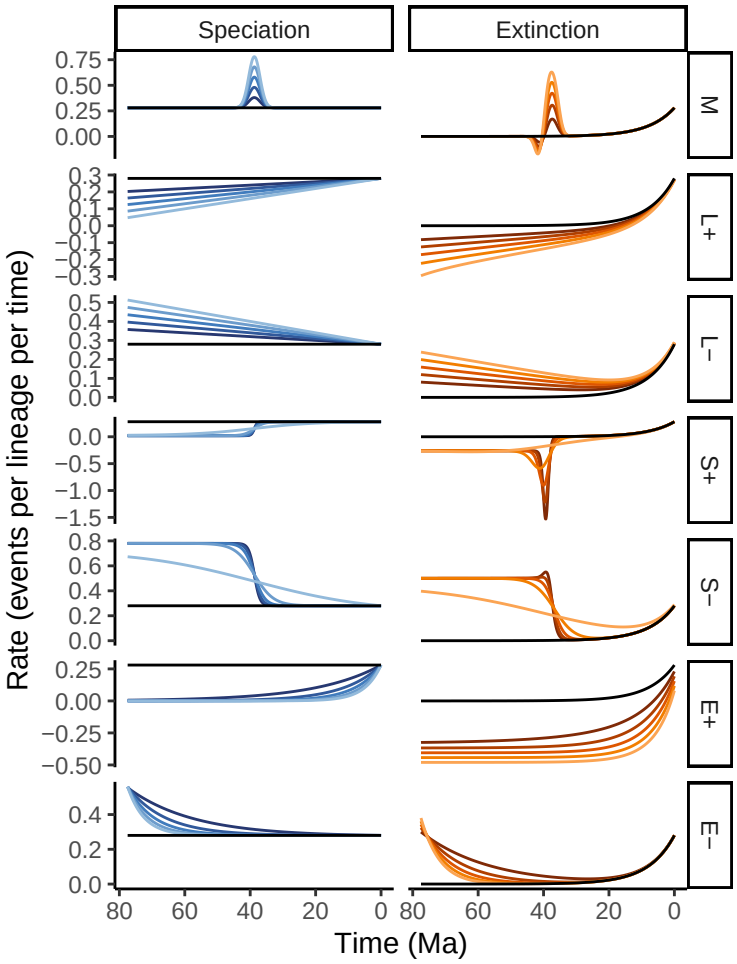


Figure S3.33: Hypothetical rate scenarios for the exponentially increasing birth-death model in Fig. S3.27. The reference model is depicted in black in all panels: a constant speciation rate ($\lambda = 0.28$), and an exponentially increasing extinction rate function corresponding to (Fig. S3.27, E+, red line). Five alternative speciation rates are proposed in each row, including modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent extinction rates.

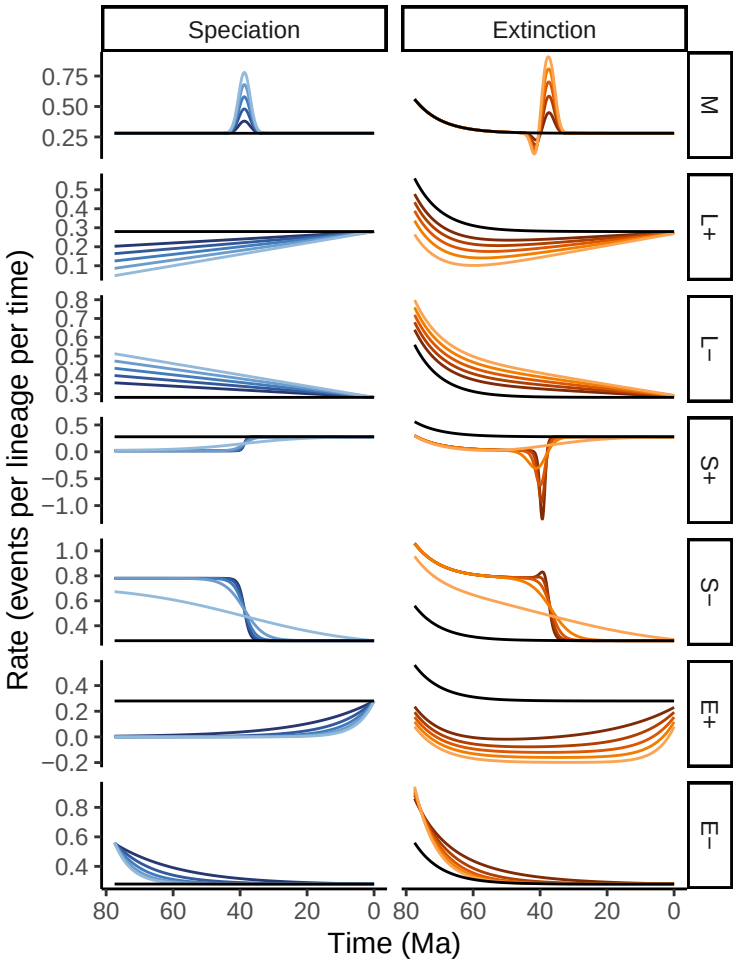


Figure S3.34: Hypothetical rate scenarios for the exponentially decreasing birth-death model in Fig. S3.27. The reference model is depicted in black in all panels: a constant speciation rate ($\lambda = 0.28$), and an exponentially decreasing extinction rate function corresponding to (Fig. S3.27, E–, red line). Five alternative speciation rates are proposed in each row, including modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). The right column depicts the congruent extinction rates.

S3.9 How steep is sufficiently steep?

In the main text and supplement we argued that the directional trends in the rates are robust to the congruence class, if the rate is sufficiently steep. We showed examples using various visualizations of different rate shapes. However, how steep is sufficiently steep in order for the signal to be robust to the congruence class? In Fig. S3.35, we examine five reference models where the minimum slope of the speciation rate function varies in two orders of magnitude, from -0.150 to -0.006 speciation rate units per million years. For the steepest model, we see agreement among the congruent models if we select a threshold of $\epsilon \leq 0.04$. In the next, less steep model ($d\lambda/dt = -0.059$), we see agreement only if we select a threshold of $\epsilon \leq 0.02$. If we select too large a threshold, then the agreement among congruent models disappears. If we examine a model that is more shallowly changing ($d\lambda/dt = -0.006$), then there is no agreement whatever the selection of thresholds. Note that, even though the slope of the next-shallowest model ($d\lambda/dt = -0.013$) is steeper than the smallest threshold ($\epsilon = 0.01$), there is still not consensus among the congruent models.

In our selection, it appears that we need a slope at or less than $d\lambda/dt = -0.028$ in order to have consensus among the congruent models. This is, however, based on a limited set of examples, whereas the function space for birth-death models is infinitely large. The answer to the question, "how steep is sufficiently steep?", may change depending on timing of the shift, on the extinction rate (here set to $\mu(t) = 0.28$), or on more recent, auxiliary time-variation in the speciation rate. Nevertheless, it remains a fact that rapidly changing rates are more agreed upon by the trends in the congruent models, as opposed to those for gradually changing rates.

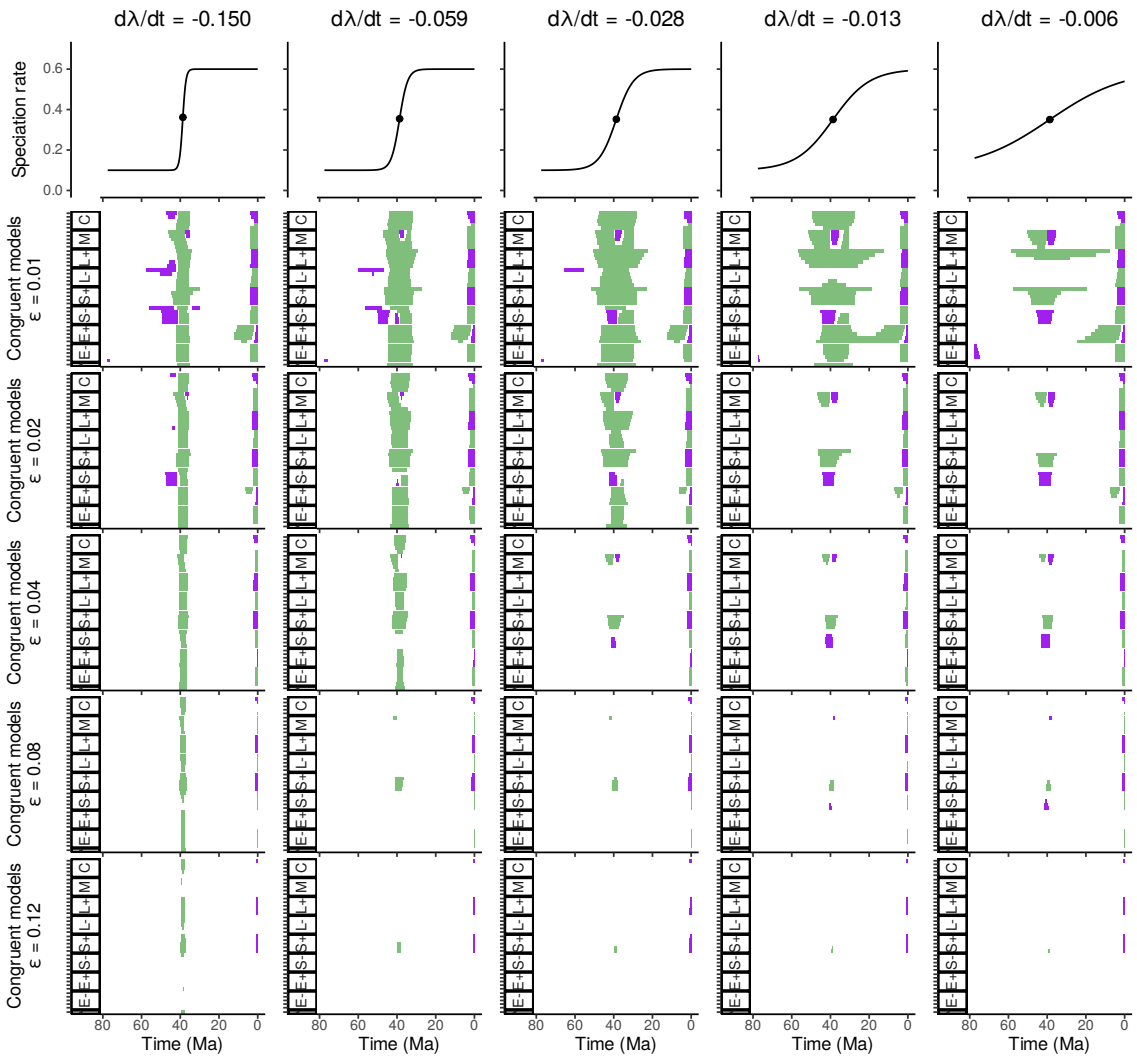


Figure S3.35: The top row depicts five reference models, each of which corresponds to its column. The reference models have a sigmoidally increasing speciation rate ($\lambda(t) = 0.1 + 0.5/(1 + s^{-t+h/2})$, where h is the tree height and $s \in \{0.3, 0.625, 0.8, 0.9, 0.95\}$ is the steepness coefficient). All of the reference models have an extinction rate of $\mu(t) = 0.28$. We plotted the steepest part of the speciation rate curve as a black dot, and display the minimum slope on the top (e.g. $d\lambda/dt = -0.150$). The subsequent five rows show the significant directional trends in congruent speciation rates. We sampled the congruence class by proposing alternative extinction rate functions. The proposal rate shapes include constants (C), modals (M), linear increase (L+), linear decrease (L–), sigmoidal increase (S+), sigmoidal decrease (S–), exponential increase (E+), and exponential decrease (E–). For steep speciation rates ($\lambda(t)$), and/or a small significance threshold for the trends (ϵ), there is consensus among the trends for the timing of the rate shift. For a more shallowly increasing reference model, or if a larger threshold is used, then the agreement among trends becomes unclear or disappears.

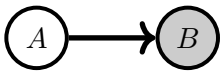
Appendix C

Supplementary Material for Chapter 4

S4.1 Statistical inference of ancestral diversification rates

Statistical inference of ancestral diversification rates under the birth-death-shift process poses some interesting questions. We note that several previous approaches that used an underlying stochastic process used a Bayesian statistical inference framework (Rabosky, 2014; Barido-Sottani et al., 2020; Höhna et al., 2019; Martínez-Gómez et al., 2024), or at least claimed to be doing so.

The simplest possible Bayesian inference problem can be depicted like so where the state



of the random variable A influences the state of the random variable B. Using Bayes theorem we can pose the question, "what is the probability of A, given that we know B?"

$$\overbrace{\Pr(A|B)}^{\text{Posterior}} = \frac{\overbrace{\Pr(B|A)}^{\text{Likelihood}} \times \overbrace{\Pr(A)}^{\text{Prior}}}{\underbrace{\Pr(B)}_{\text{Marginal Likelihood}}}. \tag{S4.1}$$

Here, the prior probability reflects the strength of belief in A before observing B, and the likelihood function tells us the probability of observing B given A. When more complex hierarchical models are considered, for example $B \leftarrow A \rightarrow C$, they are referred to more generally as *Bayesian belief networks* (Pearl, 1988). Bayesian networks that represent the splitting behaviour of a phylogeny are well established in macroevolution (e.g. Felsenstein, 1981; Pagel, 1994; Yang et al., 1995; Lewis, 2001; Höhna et al., 2014). There is also a generalized version of Bayes theorem for trees (Pearl, 1982), which we can use to ask questions such as "what is the probability of an internal node A given my observations?" Despite this, we argue that Bayesian inference is not well understood in the context of birth-death-shift models.

If we are interested in Bayesian estimates of ancestral diversification rates under the birth-death-shift process, we must represent the process as a Bayesian network. Conceptually, we would like to know the probability of a hypothesis Z, representing a specific diversification rate history, given that we know the phylogeny Ψ. In the next two sections, we discuss two problems that arise when we attempt to represent the birth-death-shift process as a Bayesian network. In the first approach, we are not able to separate the diversification rate history Z from the phylogeny Ψ. In the second approach, the "evidence" provided by the phylogeny is not encoded in the same way as for standard Bayesian networks. These two problems imply that the probabilistic representation of the birth-death-shift process is incompatible with established theory on Bayesian networks (Pearl, 1988; Russell and Norvig, 2010). Nevertheless, we provide a probabilistic method for inferring ancestral diversification rates that is inspired by Bayesian inference of ancestral discrete characters (Yang et al., 1995; Nielsen, 2002). Due to the incompatibilities, we describe our approach as being "pseudo-Bayesian", highlighting that our analysis is motivated by Bayesian inference.

S4.1.1 Problem 1: Branch rates and phylogeny are not separable

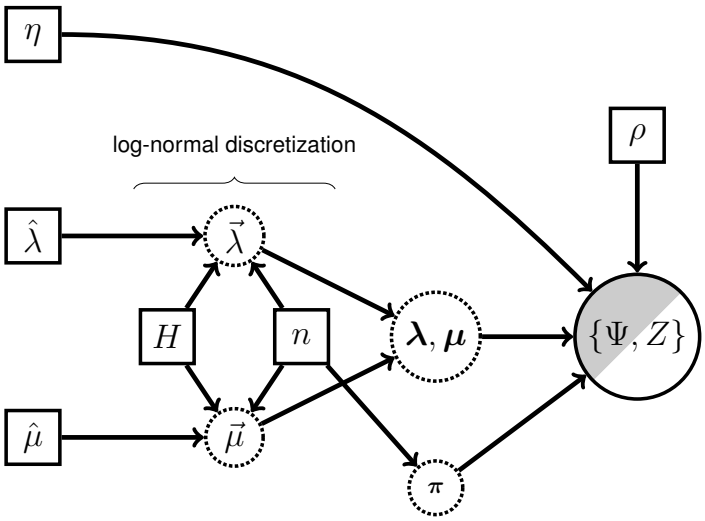


Figure S4.1: Bayesian network of the birth-death-shift model as used in `Pesto`. The network contains a node that is partially observed. This node is a stochastic variable representing the outcome of the birth-death-shift process. The node is partially shaded, meaning that we observe the reconstructed phylogeny Ψ, but we do not observe the diversification rate shift history Z. The $\hat{\lambda}$ and $\hat{\mu}$ parameters control the scale of the speciation and extinction rates, the number H controls the spread of the log-normal distribution, n represents the number of rate class discretizations, η is the diversification shift rate, ρ is the taxon sampling fraction, and $\boldsymbol{\pi}$ are the prior probabilities of the diversification rate categories at the root node. In `Pesto`, we estimate $\hat{\lambda}, \hat{\mu}$ using the lineage-homogeneous birth-death process, and we estimate η using the birth-death-shift process, while we use a priori fixed values for H, n, ρ.

Finding Bayesian estimators of branch-specific diversification rates on a phylogeny requires us to define what the hypothesis is and what the observation is. This is where the challenge for statistical inference of branch-specific diversification rate estimation arises. The hypotheses

that we are most interested in, concerns what the ancestral diversification rates were, including the possibility that different clades were diversifying at a different pace (Höhna et al., 2019). If we simulate under a birth-death-shift process, then a realization is a specific phylogenetic tree Ψ, including the history Z (Fig. S4.1). The phylogeny and the diversification rate history are linked, and it is not possible to separate the one from the other. With empirical phylogenies, however, the history of the rate shift events on the phylogeny are not recorded. This means that the process is only partially observed (hence the partially shaded node in Fig. S4.1). Thus, we are trying to infer the unobserved variable Z, and not a variable on which the data depends. This is incompatible with established theory on Bayesian inference, because there is no separation between the prior of Z and the likelihood of Z (see also May and Rothfels, 2023).

S4.1.2 Problem 2: Evidence is not encoded as evidence nodes

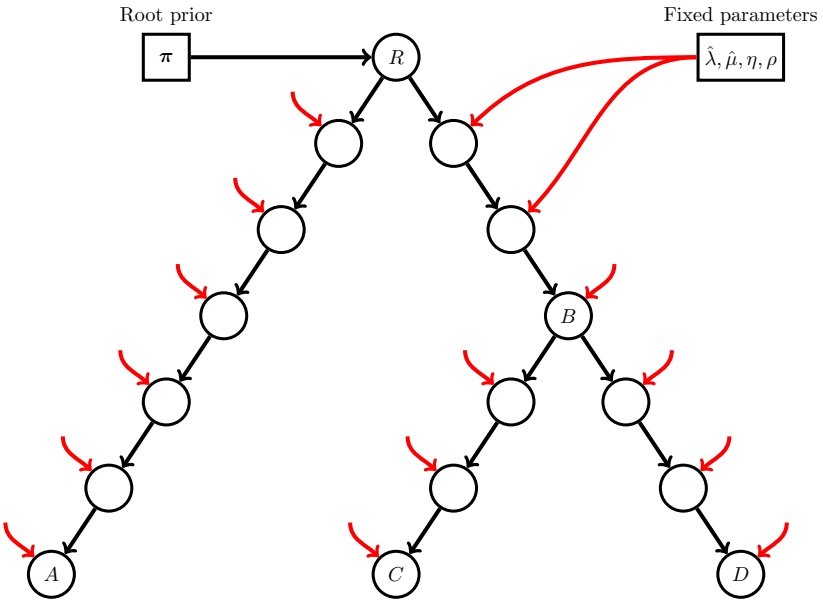


Figure S4.2: A graphical model representation of the distribution of the rate categories (circle nodes representing stochastic variables) along a three species phylogeny. The square nodes are fixed (they are not stochastic). The realization of the tree parameter Ψ is represented by the branching structure of the stochastic nodes, as well as the divergence times. What we are interested in calculating, is the probability of the diversification rate category some node B, conditional on i) the prior probability at the root, ii) the observed phylogeny, and iii) the fixed model parameters.

Suppose we consider a more specific hypothesis, by asking "what is the diversification rate on a particular branch and at a particular time?" In order to apply Bayesian statistical inference on this problem, we need to be able to represent the relationships of the hypothesized ancestral diversification rate, with respect to the hypothesis at other branches in the tree, or at the same branch but at a different time. Specifically, we use capital letters $R, A, B, \ldots$ to represent the hypothesized diversification rate category at different nodes in the phylogeny

(i.e., that of the root node, and descendants of the root node). The symbol Z then represents the set of all the nodes $\{R, A, B, \dots\}$. A graphical representation of the Bayesian network is shown in Fig. S4.2, where the diversification rate of a descendant B depends on the diversification rate of its ancestor R.

The difference between the Bayesian network in Fig. S4.2 and that of a standard Bayesian network lies in how empirical evidence is encoded in the network. Typically, this is encoded as a specific evidence node (depicted as a stochastic node with gray shading, Höhna et al. 2014), where the node is initialized to for example $[1, 0, \dots]$ (as in BiSSE, Maddison et al. 2007). However, in the birth-death-shift model, the "evidence" is not encoded by initializing stochastic nodes in this way. Instead, it is encoded in the topology and branch lengths of the phylogeny. This encoding of empirical evidence is different to that of established theory on Bayesian graphical networks (Pearl, 1988). Consequently, it is not possible to pose the question "what were the ancestral diversification rate categories, given that I know the phylogeny?" as a Bayesian inference problem, in the same manner as for standard problems.

S4.1.3 Joint probability of ancestral diversification rates

As we have shown, the Bayesian network of the birth-death-shift process is incompatible with established theory on Bayesian inference. Nevertheless, we attempt to represent the hypotheses about the ancestral diversification rates $Z \in \{R, A, B, \dots\}$ in a probabilistic manner. For the root node R in the phylogeny, the probability depends on the prior probability $\boldsymbol{\pi}$. It also depends on whether or not the root node is a speciation event

$$\Pr(R = r) = \begin{cases} \pi_r \Delta t \lambda_r & \text{if } R \text{ is a speciation event} \\ \pi_r (1 - \Delta t \lambda_r) & \text{if } R \text{ is a not speciation event,} \end{cases} \tag{S4.2}$$

where, $\Delta t \lambda_r$ is the probability of a speciation event occurring in category r in a small time interval Δt, and $(1 - \Delta t \lambda_r)$ is the probability of there not being a speciation event in the interval. Since Δt is an arbitrarily small constant, the established practice is to simplify $\Delta t \lambda_r$ to λ_r, and to omit $1 - \Delta t \lambda_r$ entirely (Maddison et al., 2007). Suppose node R has a descendant node A, then the probability of A conditionally on R also depends on whether or not A is a speciation event

$$\Pr(A = a | R = r) = \begin{cases} D_{a|r}(t_{\text{old}}) \lambda_a & \text{if } A \text{ is a speciation event} \\ D_{a|r}(t_{\text{old}}) & \text{if } A \text{ is a not speciation event,} \end{cases} \tag{S4.3}$$

where $D_{a|r}(t_{\text{old}})$ represents the rth entry in the solution to the dD/dt differential equation, when it is initialized to $D_a(t_{\text{young}}) = 1$ and $D_{i \neq a}(t_{\text{young}}) = 0$. Note that Eq. S4.3 does not sum to one over A. This is because not all possible outcomes are considered. For example, the process could have either gone extinct, or ended with two or more lineages (in which case

node A does not exist).

Eqs. S4.2 and S4.3 allow us to write the joint probability of the diversification rate categories $Z \in \{R, A, B, \dots\}$. If we use the shorthand notation $\Pr(A = a|R = r) = P(a, r)$, then the joint probability of the belief network in Fig. S4.2 is

$$\Pr(z) = \Pr(r, a, b, c, d) = \alpha \Pr(r) \Pr(a|r) \Pr(b|r) \Pr(c|b) \Pr(d|b), \tag{S4.4}$$

where α is a normalizing factor such that the joint probability sums to one. If we are interested in the probability distribution of a node, regardless of the other nodes in the phylogeny, we would analyze its marginal distribution. There are several methods one could use for calculating the marginal probabilities. The most common approach is to sample the distribution by Markov chain Monte Carlo simulation (Rabosky, 2014; Barido-Sottani et al., 2020) or stochastic mapping (Höhna et al., 2019). However, it is not necessary to use MCMC simulation methods to characterize the marginal distribution. Perhaps the simplest method is the enumeration method, in which all possible configurations of the Bayesian network are considered, and their probabilities are summed over (Russell and Norvig, 2010, p. 523). If we want to calculate the marginal probability of node B using the enumeration method, we would compute

$$\Pr(B = b|\Psi) = \alpha \sum_r \sum_a \sum_c \sum_d \Pr(r) \Pr(a|r) \Pr(b|r) \Pr(c|b) \Pr(d|b), \tag{S4.5}$$

i.e., by marginalizing out the other nodes R, A, C and D. If we separate and move the summation signs as far to the right as possible, as well as decompose $\Pr(b|r)$, it becomes more clear which part of the equation corresponds to the "likelihood"

$$\begin{aligned}
\Pr(B = b|\Psi) &= \alpha \sum_r \Pr(r) \left(\sum_a \Pr(a|r) \right) \Pr(b|r) \left(\sum_c \Pr(c|b) \right) \left(\sum_d \Pr(d|b) \right) \\
\Pr(B = b|\Psi) &= \alpha \sum_r \Pr(r) \left(\sum_a \Pr(a|r) \right) D_{b|r}(t_R) \underbrace{\lambda_b \left(\sum_c \Pr(c|b) \right) \left(\sum_d \Pr(d|b) \right)}_{\text{likelihood}},
\end{aligned} \tag{S4.6}$$

meaning the probability of observing the tree descended from node B, conditional on that $B = b$. Observe that we introduce a summation sign for each node that we marginalize over. Therefore, the number of operations scales exponentially with the number of nodes. While this is not an issue for small Bayesian networks, it quickly becomes a problem for larger networks. Fortunately, more computationally efficient algorithms exist. Pearl (1982) demonstrated how marginal probabilities could be calculated efficiently using a dynamic programming algorithm (see also Gallager, 1962). By efficient we mean that the amount of operations scales linearly with the amount of nodes in the tree. While we call this algorithm the backward-forward

algorithm, it is also known as a belief propagation algorithm, a message passing algorithm, or the sum-product algorithm (Teo et al., 2025). The following product rule describes the marginal probability of the diversification rate category at node B

$$\Pr(B = b|\Psi) = \alpha \times \Pr(\Psi^{\text{Downstream}}|B = b) \times \Pr(B = b|\Psi^{\text{Upstream}}) \tag{S4.7}$$

where the phylogeny Ψ has been split into two parts, $\Psi^{\text{Downstream}}$ (the part of the tree that is descended from node B) and Ψ^{Upstream} (the part of the tree that is not descended from node B). This product rule is fundamentally based on Bayes theorem, see Pearl (1982) for the original presentation, and (Pearl, 1988, pp. 162–174) for a more in-depth explanation. This is the same equation as in the main text, it is only the notation that is different

$$S_j(t) = \alpha D_j(t) F_j(t). \tag{S4.8}$$

We emphasize that Eqs. S4.5 to S4.8 give identical results for the marginal probabilities.

S4.1.4 Inference motivated by Bayesian thinking

Let us now take a step back to assess what type of statistical inference we used to estimate the ancestral diversification rate categories. We are, in fact, more interested in inferring these unobserved diversification rate categories than we are interested in any of the other parameters of the birth-death-shift process. The inference of ancestral (or branch-specific) diversification rate categories could be considered to be a similar problem to that of ancestral character estimation (e.g. Yang et al., 1995). The main differences are i) we assumed (perhaps unjustifiably) that we could characterize the Bayesian belief network of Z conditionally on Ψ, and that ii) the evidence is not encoded as specific evidence nodes in the network. We specify a prior probability at the root node of the phylogeny, in accordance with our prior beliefs. The marginal probability of the diversification rate category at some node B is determined by i) the prior on the root node, ii) the branching structure of the phylogeny, and iii) the fixed parameters $\boldsymbol{\lambda}, \boldsymbol{\mu}, \eta, \rho$. Since the marginal distribution depends on the prior and on the observation, we intuitively think of it as being a posterior distribution. There is one challenge, in that the Bayesian network for the birth-death-shift process is not compatible with standard theory on Bayesian inference (there is no data if the phylogeny is represented by the Bayesian network structure). Therefore, the joint probability (Eq. S4.4) may not be a rigorous probabilistic representation of the birth-death-shift model. Hence, we suggest the terms "pseudo-Bayesian inference", and "pseudo-posterior probability," indicating that our approach is motivated by Bayesian inference. Overall, we suggest that Bayesian motivated analysis is intrinsically consistent, but more research might shed more light on this particular statistical problem.

S4.2 Notation

We have tried to use notation that is consistent with the literature history, albeit with a few exceptions. Scalars are written in plain typeface, and vectors and matrices are in bold. Vectors are matrices with one column. Transposed vectors are matrices with one row.

Symbol	Description
K	The total number of rate categories in the model, indexed by $i, j \in \{1, 2, 3, \dots K\}$.
M	The branch index variable, i.e. $M \in \{1, 2, 3, \dots 2(n_{\text{tips}} - 1)\}$.
$\odot$	Element-wise multiplication.
$\oslash$	Element-wise division.
$\boldsymbol{\lambda}$	A vector with elements $\lambda_1, \lambda_2 \dots, \lambda_j$.
$\boldsymbol{\mu}$	A vector with elements $\mu_1, \mu_2 \dots, \mu_j$.
η	The shift rate from any category j to any other category i.
$\frac{\eta}{K-1}$	The shift rate from any category j a different category i.
$\mathbf{Q}$	The matrix representation of the shift rates, with $-\eta$ on the diagonal elements, and $\eta/(K-1)$ on the offdiagonal elements.
$\mathbf{E}(t)$	A vector function with elements $E_j(t)$. Evaluated at age t in the past, $E_j(t)$ gives the probability of the lineage not being in the sampled taxa at the present ($t = 0$). If there is complete taxon sampling, then $E_j(t)$ is the probability of the lineage going extinct before the present.
$\mathbf{D}(t)$	A vector function with elements $D_j(t)$. The elements $D_j(t)$ represent the probability densities of observing the subtending clade at the present given that it is in category j at age t.
$\mathbf{F}(t)$	A vector function with elements $F_j(t)$. $F_j(t)$ is similar to $D_j(t)$, as in it is a probability density of observing the subtending clade. It is, however, given conditional on the ancestral values of $D_j(t)$.
$\mathbf{S}(t)$	A vector function with elements $S_j(t)$ that represent the marginal probability of being in category j at time t. The elements of $\mathbf{S}(t)$ sum to one.
$\mathbf{A}(t)$	A matrix function which when evaluated at time t yields a square matrix of size (K, K). $\mathbf{A}(t)$ determines how the branch probability densities change with time.
$\mathbf{P}(t, t + \Delta t)$	A square matrix of size (K, K). P_{ij} is the ith row and jth column of $\mathbf{P}$, and represents the probability of going from category j to category i in a short time interval from an older time t to a younger time $t + \Delta t$, conditionally on that the process began in category j at time t.
$\hat{\mathbf{N}}(t)$	A square matrix of size (K, K). $\hat{\mathbf{N}}(t)$ is the estimator for the number of rate shifts from the beginning of the branch until time t. Zero on the diagonal entries, and with the number of rate shifts $\hat{N}_{ij}(t)$ on the offdiagonal.

Table S4.1: A summary and descriptions of the symbols used in the methods section.

S4.3 Extended methods

S4.3.1 Number of rate shifts

The idea is to calculate how many rate shifts there were along a branch, in the time span of a small time interval. The time interval must be sufficiently short, such that the assumption that only event can happen in the time interval is reasonable. We begin by considering the rate shift probabilities. The transition probability between time t and time $t+0$ is the identity matrix, i.e. the probability of changing category is zero, and the probability of remaining in the same category is one:

$$\mathbf{P}(t, t+0) = \mathbf{I}. \tag{S4.9}$$

When using the matrix representation of the transition probabilities, we will follow the notation of Felsenstein (2004, page 205), in that the element P_{ij} is the ith row and jth column of $\mathbf{P}$, and represents the transition probability of going from category j to category i in the time interval. P_{ij} in a small time interval ($\Delta t < 0$, going from old to young) is approximately (Höhna et al., 2019)

$$P_{ij}(t, t+\Delta t) \approx F_{ij}(t, t+\Delta t) D_i(t+\Delta t), \tag{S4.10}$$

where F_{ij} has been initialized as 1 in category j, and as 0 in all other categories $i \neq j$. Hence, $F_{ij}(t, t+\Delta t)$ is

$$F_{ij}(t, t+\Delta t) = \begin{cases} 1 - \Delta t[-(\lambda_j + \mu_j + \eta) + 2\lambda_j E_j(t)], & \text{if } j = i \\ 0 - \Delta t \frac{\eta}{K-1}, & \text{if } j \neq i \end{cases} \tag{S4.11}$$

Alternatively, $F_{ij}(t+\Delta t)$ can be expressed using matrix notation as $F_{ij}(t+\Delta t) = (\mathbf{e}_j - \Delta t \mathbf{A}(t)\mathbf{e}_j)_i$, where the vector $\mathbf{e}_j$ is a one-hot vector, i.e. it is 1 at index j, and 0 elsewhere. The matrix $\mathbf{A}(t)$ is (Louca and Pennell, 2020a)

$$\mathbf{A}(t) = \mathrm{diag}(-\boldsymbol{\lambda} - \boldsymbol{\mu} + 2\boldsymbol{\lambda} \odot \mathbf{E}(t)) + \mathbf{Q}. \tag{S4.12}$$

The diag(...) function constructs a diagonal matrix from a vector argument. The elements of $\mathbf{A}(t)$ are

$$\mathbf{A}(t) = \begin{bmatrix} -(\lambda_1 + \mu_1 + \eta) + 2\lambda_1 E_1(t) & \dots & \frac{\eta}{K-1} \\ \vdots & \ddots & \vdots \\ \frac{\eta}{K-1} & \dots & -(\lambda_K + \mu_K + \eta) + 2\lambda_K E_K(t) \end{bmatrix}. \tag{S4.13}$$

It follows that the matrix form of $P_{ij}(t, t+\Delta t)$ is

$$\mathbf{P}(t, t+\Delta t) \approx (\mathbf{I} - \Delta t \mathbf{A}(t)) \odot (\mathbf{D}(t+\Delta t)\mathbf{1}^\top), \tag{S4.14}$$

where $\mathbf{1}^\top$ is a row-vector of ones. For clarity, we populate $\mathbf{P}(t, t+\Delta t)$ with an example where we have $K = 3$ categories:

$$\mathbf{P}(t, t+\Delta t) = \begin{pmatrix} P_{11}(t, t+\Delta t) & P_{12}(t, t+\Delta t) & P_{13}(t, t+\Delta t) \\ P_{21}(t, t+\Delta t) & P_{22}(t, t+\Delta t) & P_{23}(t, t+\Delta t) \\ P_{31}(t, t+\Delta t) & P_{32}(t, t+\Delta t) & P_{33}(t, t+\Delta t) \end{pmatrix}, \tag{S4.15}$$

or approximately

$$\begin{pmatrix} (1-\Delta t[-(\lambda_1+\mu_1+\eta)+2\lambda_1 E_1(t)]) \\ D_1(t+\Delta t) \end{pmatrix} & -\Delta t\tfrac{\eta}{K-1} D_1(t+\Delta t) & -\Delta t\tfrac{\eta}{K-1} D_1(t+\Delta t) \\ -\Delta t\tfrac{\eta}{K-1} D_2(t+\Delta t) & \begin{pmatrix} (1-\Delta t[-(\lambda_2+\mu_2+\eta)+2\lambda_2 E_2(t)]) \\ D_2(t+\Delta t) \end{pmatrix} & -\Delta t\tfrac{\eta}{K-1} D_2(t+\Delta t) \\ -\Delta t\tfrac{\eta}{K-1} D_3(t+\Delta t) & -\Delta t\tfrac{\eta}{K-1} D_3(t+\Delta t) & \begin{pmatrix} (1-\Delta t[-(\lambda_3+\mu_3+\eta)+2\lambda_3 E_3(t)]) \\ D_3(t+\Delta t) \end{pmatrix} \end{pmatrix}. \tag{S4.16}$$

Since the columns of Eq. S4.14 don't sum to one, they do not represent a valid probability distribution. To normalize them, we divide each element by the corresponding column sum:

$$\mathbf{P}(t, t+\Delta t) = \Big[(\mathbf{I} - \Delta t\mathbf{A}(t)) \odot (\mathbf{D}(t+\Delta t)\mathbf{1}^\top)\Big] \oslash \mathbf{C}, \tag{S4.17}$$

where $\mathbf{C} = \mathbf{1}\mathbf{1}^\top[(\mathbf{I} - \Delta t\mathbf{A}(t)) \odot (\mathbf{D}(t+\Delta t)\mathbf{1}^\top)]$. The elements in $\mathbf{P}(t, t+\Delta t)$ are as follows:

$$P_{ij}(t, t+\Delta t) = \begin{cases} \dfrac{D_i(t+\Delta t)(1-\Delta t[-(\lambda_i+\mu_i+\eta)+2\lambda_i E_i(t)])}{D_j(t+\Delta t)(1-\Delta t[-(\lambda_j+\mu_j+\eta)+2\lambda_j E_j(t)]) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)} & \text{if } j = i \\ \dfrac{-\Delta t\frac{\eta}{K-1} D_i(t+\Delta t)}{D_j(t+\Delta t)(1-\Delta t[-(\lambda_j+\mu_j+\eta)+2\lambda_j E_j(t)]) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)} & \text{if } j \neq i. \end{cases} \tag{S4.18}$$

Apart from the normalization, we are not interested in the diagonal elements, since they do not represent rate shifts. If we weigh by the probability of beginning in ancestral category j at time t, we can get the number of shifts in an interval from time t to $t+\Delta t$:

$$\text{Shifts}_{ij}(t, t+\Delta t) = S_j(t)\frac{-\Delta t\frac{\eta}{K-1} D_i(t+\Delta t)}{D_j(t+\Delta t)(1-\Delta t[-(\lambda_j+\mu_j+\eta)+2\lambda_j E_j(t)]) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)}, \text{if } j \neq i. \tag{S4.19}$$

Let $\hat{N}_{ij}(t)$ be the estimator for the accumulated number of rate shifts from the beginning of the branch until time t. Suppose we know $\hat{N}_{ij}(t)$, or $\hat{N}_{ij}(t_0) = 0$. Then, we can calculate the number of shifts after some time Δt:

$$\hat{N}_{ij}(t+\Delta t) = \hat{N}_{ij}(t) + S_j(t)\frac{-\Delta t\frac{\eta}{K-1} D_i(t+\Delta t)}{D_j(t+\Delta t)(1-\Delta t[-(\lambda_j+\mu_j+\eta)+2\lambda_j E_j(t)]) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)}, \text{if } j \neq i. \tag{S4.20}$$

Subtract $\hat{N}_{ij}(t)$ and divide by Δt:

$$\frac{\hat{N}_{ij}(t+\Delta t) - \hat{N}_{ij}(t)}{\Delta t} = \frac{-\eta}{K-1} S_j(t)\frac{D_i(t+\Delta t)}{D_j(t)(1-\Delta t[-(\lambda_j+\mu_j+\eta)+2\lambda_j E_j(t)]) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)}, \text{if } j \neq i. \tag{S4.21}$$

If we take the limit as $\Delta t \to 0$, then the expression simplifies (and we set the derivative to 0 if $j = i$ since we don't need these):

$$\lim_{\Delta t \to 0} \frac{\hat{N}_{ij}(t + \Delta t) - \hat{N}_{ij}(t)}{\Delta t} = \frac{d\hat{N}_{ij}(t)}{dt} = \begin{cases} -S_j(t) \dfrac{D_i(t)}{D_j(t)} \dfrac{\eta}{K-1} & \text{if } j \neq i \\ 0 & \text{if } j = i. \end{cases} \tag{S4.22}$$

In matrix form, this is

$$\frac{d\hat{\mathbf{N}}(t)}{dt} = \frac{-\eta}{K-1} \odot (\mathbf{1}\mathbf{1}^\top - \mathbf{I}) \odot \mathbf{D}(t)\mathbf{S}(t)^\top \oslash \mathbf{1}\mathbf{D}(t)^\top. \tag{S4.23}$$

We can solve Eq. S4.23 using the same numerical integration techniques as before, over the time interval of the branch, with initial condition $\hat{\mathbf{N}}(t_0) = \mathbf{0}$. The number of shifts among all pairs (i, j) for all $j \neq i$, in the time direction of old to young, are on the off-diagonals. This represents $K(K-1)$ non-zero differential equations, and so if the state space K is large, then computing the number of rate shifts $\hat{\mathbf{N}}$ can be slow. If we sum over all entries in $\hat{\mathbf{N}}$, we obtain an estimate of the total number of shifts in the time interval from t_0 until t:

$$\hat{N}(t) = \mathbf{1}^\top \hat{\mathbf{N}}(t) \mathbf{1}, \tag{S4.24}$$

and we can compare with the results from simulated mappings of rate category histories, as implemented in RevBayes (Fig. S4.3). The posterior number of rate shifts $\hat{N}$ can also be summed across all of the branches, in order to get an overall idea of how many shifts there were in the entire phylogeny.

S4.3.2 Probability of a shift

We want to calculate the probability that there was at least one shift on a branch:

$$P_{\geq 1 \text{ shifts}} = \ldots, \tag{S4.25}$$

or equivalently, the probability that there were zero shifts

$$P_{0 \text{ shifts}} = 1 - P_{\geq 1 \text{ shifts}}. \tag{S4.26}$$

One way to do so would be to begin by calculating the probability that there were no rate shifts in rate category j. Remember that the probabilities in Eq. S4.18 are conditional on that the current category is j, i.e. they sum to one over the arrival categories i. Conditional on that the process began in the departure category j, we can calculate the probability that at

least no shifts with departure category j occurred as follows

$$\begin{aligned} P_{0\text{ shifts},j}(t, t+\Delta t) &= P_{jj}(t, t+\Delta t) \\ P_{0\text{ shifts},j}(t, t+\Delta t) &= \frac{D_i(t+\Delta t)(1-\Delta t[-(\lambda_i+\mu_i+\eta)+2\lambda_i E_i(t)])}{D_j(t+\Delta t)(1-\Delta t[-(\lambda_j+\mu_j+\eta)+2\lambda_j E_j(t)]) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)}. \end{aligned} \tag{S4.27}$$

Next we want to calculate the probability that there were no rate shifts away from category j along the entire branch, meaning we need to take the product of several probabilities. We can write it recursively as follows, by introducing a dummy variable $X_j(t)$ that represents the probability that there were no rate shifts departing from category j until time t. We know the probability $X_j(t)$ at the beginning of a time span, it is simply set to one. Then, the probability of no shifts after some small time interval Δt has elapsed is simply the previous probability times the transition probability

$$X_j(t+\Delta t) = Xj(t)P_{jj}(t, t+\Delta t), \tag{S4.28}$$

representing the progression along the branch. Because sums are typically more numerically stable than products, we will instead use the log transformed probability

$$\begin{aligned} \log X_j(t+\Delta t) &= \log X_j(t) + \log P_{jj}(t, t+\Delta t) \\ \log X_j(t+\Delta t) - \log X_j(t) &= \log P_{jj}(t, t+\Delta t), \end{aligned} \tag{S4.29}$$

and next we populate $\log P_{jj}(t, t+\Delta t)$ using Eq. S4.18. Because there will be a lot of symbols, we use a substitute c_j representing $-(\lambda_j+\mu_j+\eta)+2\lambda_j E_j(t)$ and shorten the equations a bit. Then, we have

$$\log X_j(t+\Delta t) - \log X_j(t) = \log\left[\frac{D_j(t+\Delta t)(1-\Delta t c_j)}{D_j(t+\Delta t)(1-\Delta t c_j) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)}\right] \tag{S4.30}$$

and we can divide by Δt on both sides and take the limit

$$\begin{aligned} \frac{\log X_j(t+\Delta t) - \log X_j(t)}{\Delta t} &= \frac{1}{\Delta t}\log\left[\frac{D_j(t+\Delta t)(1-\Delta t c_j)}{D_j(t+\Delta t)(1-\Delta t c_j) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)}\right] \\ \lim_{\Delta t\to 0}\frac{\log X_j(t+\Delta t) - \log X_j(t)}{\Delta t} &= \lim_{\Delta t\to 0}\frac{1}{\Delta t}\log\left[\frac{D_j(t+\Delta t)(1-\Delta t c_j)}{D_j(t+\Delta t)(1-\Delta t c_j) - \Delta t\frac{\eta}{K-1}\sum_{i\neq j} D_i(t+\Delta t)}\right]. \end{aligned} \tag{S4.31}$$

Note that the limits of $1/\Delta t$ and the log term both converge to zero. In this case we can use L'Hôpital's rule, differentiating the numerator and the denominator with respect to Δt. The

left term vanishes, and it can be shown that the derivative for the log-term is

$$\frac{d}{d\Delta t}\log\left[\frac{D_j(t+\Delta t)(1-\Delta t c_j)}{D_j(t+\Delta t)(1-\Delta t c_j)-\Delta t\frac{\eta}{K-1}\sum_{i\neq j}D_i(t+\Delta t)}\right]=\frac{\frac{\eta}{K-1}\sum_{i\neq j}D_i(t+\Delta t)}{D_j(t+\Delta t)+\alpha\Delta t+\beta\Delta t^2}, \tag{S4.32}$$

where α and β is represent some terms that are tedious to write. When evaluating the limit of $\Delta t \to 0$, then α and β vanish, and we get the following differential equation

$$\frac{d\log X_j(t)}{dt}=\frac{\eta}{K-1}\frac{\sum_{i\neq j}D_i(t)}{D_j(t)}, \tag{S4.33}$$

which we can solve for numerically using the initial condition $\log X_j(t_{\text{old}}) = 1$. Note that dt is a negative number, since we are integrating from the past to the present. Recall that all this is conditional on that the process began in rate category j. In order to get the posterior probability that there were no rate shifts in any rate category, we need to take the weighted average as

$$P_{0\text{ shifts}}=\sum_j S_j(t_{\text{old}})e^{\log X_j(t_{\text{young}})}, \tag{S4.34}$$

where $S_j(t_{\text{old}})$ is the posterior probability that the process was in rate category j at the beginning of the branch, and $X_j(t_{\text{young}})$ is the probability that there were no rate shifts departing from rate category j in the time interval $\{t_{\text{old}}, t_{\text{young}}\}$ (i.e. the branch time). It follows that the posterior probability that there was at least one rate shift is

$$P_{\geq 1\text{ shifts}}=1-P_{0\text{ shifts}}. \tag{S4.35}$$

The probability obtained by this equation is validated by comparing with an alternative simulation approach in Fig. S4.3c.

S4.3.3 Numerical solutions

We use the algorithm by Tsitouras (2011), a fourth-order Runge-Kutta method, implemented in the Julia module `DifferentialEquations.jl` (Rackauckas and Nie, 2017) in order to numerically solve the systems of differential equations. Conditional on some tolerance values (we used abstol $= 10^{-3}$, reltol $= 10^{-6}$), the algorithm adaptively selects a set of time points t_i to solve the system of differential equations. It will fit the trajectory of the probability densities at times $t_i \in [t_0, t_{\text{end}}]$, where t_0 is the beginning, and t_{end} is the end. We store the exact solutions at the time points t_i. To get values at intermediate time points (between t_i and t_{i+1}), we fit and evaluate an interpolating function (a fourth-order polynomial) to get an approximate result. First, we solve $d\mathbf{E}/dt$ across the time span of the phylogeny, without regard to the phylogeny. In a second step, we iterate in a postorder traversal of the tree, and solve $d\mathbf{D}/dt$ for each branch, storing the resulting $\mathbf{D}(t)$, evaluating the polynomial represen-

tation of $\mathbf{E}(t)$ where necessary. In a third step, we iterate over the tree in a preorder traversal. We solve $d\mathbf{F}/dt$ for each branch, using the polynomial representations of $\mathbf{E}(t)$ and $\mathbf{D}(t)$ where necessary, and store the result $\mathbf{F}(t)$. At the end of each branch, we normalize by a scaling factor:

$$\begin{aligned} \mathbf{F}_M(t_{\text{youngest}}) &:= \frac{\mathbf{F}_M(t_{\text{youngest}})}{\theta_{M,F}}, \\ \mathbf{D}_M(t_{\text{oldest}}) &:= \frac{\mathbf{D}_M(t_{\text{oldest}})}{\theta_{M,D}}. \end{aligned} \tag{S4.36}$$

We follow the implementation of `diversitree` (FitzJohn, 2012) and use a scaling factor such that the normalized probabilities sum to unity: $\theta_{M,D} = \mathbf{1}^\top \mathbf{D}_M(t_{\text{oldest}})$. Since the proportionality of the probabilities are preserved, we can use the normalized values in the initial conditions for the next step in the tree iteration. This avoids numerical problems where the products of $\mathbf{D}$ and $\mathbf{F}$ across many branches become too tiny, and floating point operations become inaccurate. When we compute the log-likelihood, we add the scaling factors back in ($\theta_{M,D}$).

S4.4 RevBayes equivalence

The birth-death-shift model implemented in `Pesto` is fundamentally the same model as the one used in `RevBayes` (Höhna et al., 2019), with a few technical caveats. In `Pesto`, the branch-specific rate estimates and the shift inferences are calculated conditionally on the diversification rate categories (i.e. $\boldsymbol{\lambda}, \boldsymbol{\mu}$) and the shift rate ($\eta$). If the diversification rate categories and the shift rate are estimated jointly using a hierarchical model, as in `RevBayes`, then the two approaches are not identical. If all rate categories and parameters are kept equal, however, then the `RevBayes` and `Pesto` implementations are theoretically equivalent, and should give the same result. In Fig. S4.3, we calculated a) the posterior mean branch-specific speciation rates, b) the posterior mean number of rate shift events, and c) the posterior probability that there was at least one rate shift — all per branch, using both `Pesto` and `RevBayes`. We used the same model in both scenarios, with the rate categories being fixed to $\boldsymbol{\lambda} = [0.1, 0.2, 0.4, 0.1]$, $\boldsymbol{\mu} = [0.05, 0.15, 0.10, 0.2]$, and the shift rate fixed to $\eta = 0.00248$. Given sufficiently many stochastic character maps (we used 25000), the branch-specific estimates should correspond exactly. In a qualitative sense, our comparison shows that the branch estimates are corresponding. There are two parts of the tree with different rates, the Old World Monkeys clade vs the rest. There is also one branch where there was strong support for a rate shift ($N > 0.8$).

In a quantitative sense, however, there are some differences between the `RevBayes` and the `Pesto` analysis results, as there are a few deviations from the one-to-one line. We attribute this to numerical differences in solving the ordinary differential equations, and the stochastic character maps, which are both approximate methods. Both `Pesto` and `RevBayes` use a Runge-

Kutta algorithm with adaptive step size. `Pesto` uses the `Tsit5` method (Tsitouras, 2011) implemented in the `DifferentialEquations.jl` Julia library (Rackauckas and Nie, 2017), while `RevBayes` uses the Dormand–Prince method (Dormand and Prince, 1980) implemented in the `boost` C++ library. Both of these algorithms are fundamentally similar, and give similar solutions. `RevBayes` uses a fixed step size on top of this, to simulate stochastic character maps forward in time (with standard settings `nTimeSlices = 500`, i.e. on average 500 time slices on a lineage spanning from the root to the tip of the phylogeny), and saves the result only on these specific time points. `Pesto` also saves the intermediate solutions (between the beginning and the end of the branch), but not nearly as many, and the number of saved states are determined dynamically instead of a priori. If the probability densities are relatively flat along a branch, then `Pesto` saves only a handful of time points, e.g. 3 or 4. If the probability densities vary drastically along a branch, i.e. the derivative is different to zero, then `Pesto` will save several more, e.g. 20-40. In practice, this means that `RevBayes` uses far more steps (i.e. evaluations of the difference equation) than `Pesto`, and so it is slower. `Pesto` also saves the solution for every branch, instead of discarding it. There is a trade-off in that `Pesto` saves more information, and the drawback is that `Pesto` requires more memory. Furthermore, `RevBayes` has to repeat the stochastic character maps thousands of times before the branch rate estimates converge, whereas `Pesto` only needs to calculate the branch rates once.

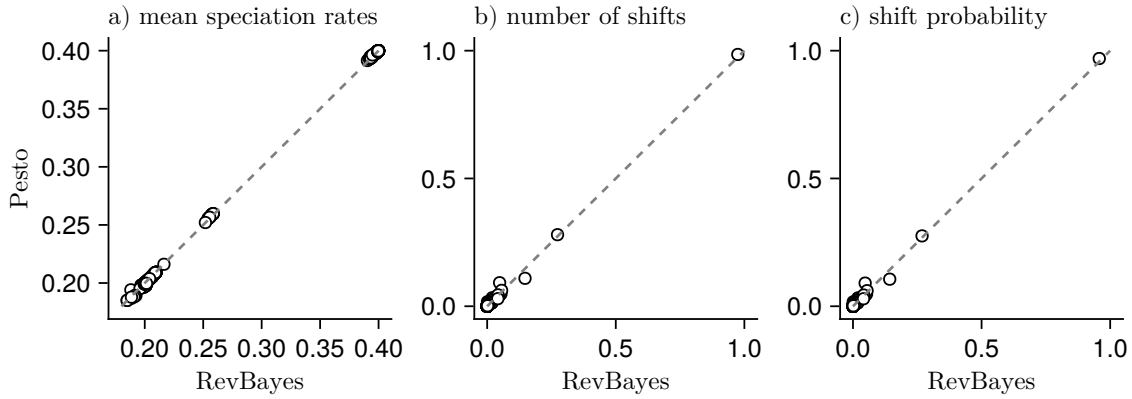


Figure S4.3: Comparisons of branch-specific calculations between `Pesto` and `RevBayes`. We calculated a) the posterior mean speciation rates, b) the posterior mean number of rate shift events, and c) the posterior probability that there was at least one rate shift. Each point represents one branch in the primates tree. The rate categories were set to $\boldsymbol{\lambda} = [0.1, 0.2, 0.4, 0.1]$, $\boldsymbol{\mu} = [0.05, 0.15, 0.10, 0.2]$, and the shift was was set to $\eta = 0.00248$ (i.e. the maximum-likelihood shift rate, conditional on $\boldsymbol{\lambda}, \boldsymbol{\mu}$. We conditioned on survival in both cases. In `RevBayes`, we used the setting `nTimeSlices = 500`, and drew 25000 independent stochastic character maps. The value on the x-axis is the mean across the 25000 stochastic character maps. The dashed line represents the one-to-one line.

S4.5 Variance of branch-specific speciation rates

Whereas in the previous section we plotted branch-specific speciation rates that were averaged across the branch length, it is also possible to compute branch-specific diversification rates that are evaluated at a specific point t in time. Since the `RevBayes` implementation does not by default print speciation rates that are specific to a particular time, we re-implemented the stochastic mapping in a `Julia` script. Suppose we are given an initial probability vector $\mathbf{S}(t)$, where t is the oldest time of the branch, and we use some small step size Δt. Note that we are going from the past to the present, and therefore Δt is a negative number. The simulation algorithm can be described as follows

1. Draw a rate category j according to the probability vector $\mathbf{S}(t)$.
2. Initialize $F_j(t) = 1$ and set $F(t)$ to 0 for all other categories.
3. Find a solution to $\mathbf{F}(t + \Delta t)$ using numerical integration.
4. Update the probability vector $\mathbf{S}(t + \Delta t) := \alpha \mathbf{F}(t + \Delta t) \odot \mathbf{D}(t + \Delta t)$, where α is a normalizing constant such that $\mathbf{S}(t + \Delta t)$ sums to one.
5. Set $t := t + \Delta t$.
6. Repeat steps 1–5 until the youngest time point of the branch is reached.

This will give us a vector of the diversification rate categories, for example $[1, 1, 3, 3, 3, \dots]$, which we can use to calculate the mean and variance of the speciation rate for each time bin. Suppose that $\lambda(t)$ is a discrete random variable with possible values $\lambda_1, \lambda_2, \dots$ and probabilities $S_1(t), S_2(t), \dots$, then we can calculate the theoretical expectation and variance for a particular time t

$$\begin{aligned} \mathbb{E}[\lambda(t)] &= \sum_j \lambda_i S_j(t) \\ \mathrm{Var}[\lambda(t)] &= \mathbb{E}[\lambda(t)^2] - \mathbb{E}[\lambda(t)]^2 \\ \mathrm{Var}[\lambda(t)] &= \sum_j \lambda_i^2 S_j(t) - \left(\sum_j \lambda_j S_j(t) \right)^2 . \end{aligned} \tag{S4.37}$$

The simulated and theoretical approaches to compute the mean and variance is shown in Fig. S4.4, and the estimates are very similar, up to some simulation error (we used 100 time bins and 5000 replicates in the simulated approach). Note that unlike `RevBayes`, this specific simulation does not include estimation error that accumulates from older branches in the tree, as we fixed $\mathbf{S}(t)$ to be equal for this branch for both the simulated and theoretical approaches. If the whole tree were used to simulate the diversification rate shift history, then we expect

that more time bins and/or more replicates would be needed to achieve errors of similar scale.

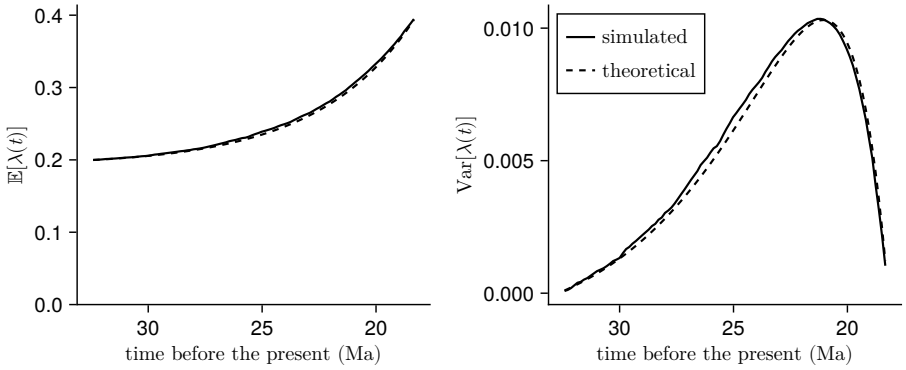


Figure S4.4: Estimates of the posterior mean speciation rate $\lambda(t)$ for the branch in the primates phylogeny that showed the strongest support for a branch-specific diversification rate shift (the branch leading up to the Old World Monkeys clade). The speciation rate categories are $\boldsymbol{\lambda} = [0.1, 0.2, 0.4, 0.1]$, the extinction rate categories are $\boldsymbol{\mu} = [0.05, 0.15, 0.10, 0.2]$, and the diversification shift rate is $\eta = 0.00248$. At the beginning of the branch, the second rate category is the most probable (0.994), whereas the third category is the most probable at the end of the branch (0.968).

Suppose one is interested in calculating the variance of the branch-wide average speciation rate, in addition to the mean (as in the previous section). Conceptually, if $\lambda(t)$ is the discrete random variable for the speciation rate at time t, we can imagine the branch-average $\bar{\lambda}$ as a weighted sum of several of time bins, where $t_1, t_2, \ldots$ correspond to the time of the bins

$$\bar{\lambda} = \frac{1}{c}\lambda(t_1) + \frac{1}{c}\lambda(t_2) + \ldots, \tag{S4.38}$$

where c is the number of time bins. In other words, we have created a new random variable that is a weighted sum of other random variables. Computing the expectation of $\bar{\lambda}$ is straight forward

$$\mathbb{E}[\bar{\lambda}] = \frac{1}{c}\sum_{a=1}^{c}\lambda(t_a). \tag{S4.39}$$

Computing the variance is more tricky, however, as we also need to take into the account the covariance among the time bins

$$\mathrm{Var}[\bar{\lambda}] = \frac{1}{c^2}\sum_{a=1}^{c}\mathrm{Var}[\lambda(t_a)] + \frac{1}{c^2}\sum_{a=1}^{c}\sum_{b>a}^{c}\mathrm{Cov}[\lambda(t_a), \lambda(t_b)], \tag{S4.40}$$

where t_a is some time point and t_b is time point younger than t_a. The variance $\mathrm{Var}[\lambda(t)]$ we had earlier, but this expression also contains the covariance across different time bins. Assuming that the covariance is approximately zero is not a reasonable assumption, as the variance then converges to zero with increasing number of time bins. Consequently, to theoretically calculate

the across-branch variance in speciation rates, we must somehow calculate the covariance across time bins. One way of calculating the covariance is

$$\begin{aligned}\mathrm{Cov}[\lambda(t_a), \lambda(t_b)] = &\sum_{j=1}^{K}\sum_{i=1}^{K} S_j(t_a) \times P_{ij}(t_a, t_b) \\ &\times (\lambda_j - \mathbb{E}[\lambda(t_a)]) \times (\lambda_i - \mathbb{E}[\lambda(t_b)]),\end{aligned} \tag{S4.41}$$

where $S_j(t_a)$ represents the marginal probability of the rate category being j at time t_a, and $P_{ij}(t_a, t_b)$ is the transition probability from category j at time t_a to the category j at a younger time t_b.

Specifically, $P_{ij}(t_a, t_b)$ is calculated by setting $\mathbf{F}(t_a)$ to a one-hot vector, and numerically integrating with $d\mathbf{F}/dt$ to a younger time t_b, and then element-wise multiplying with $\mathbf{D}(t_b)$, all repeated for all possible one-hot vectors of $\mathbf{F}(t_a)$, and assembling them into a matrix $\mathbf{P}$. Finally, we normalize the matrix $\mathbf{P}$ such that the columns sum to one. Note that we can not use Eq. S4.18 to calculate $P_{ij}(t_a, t_b)$, as Eq. S4.18 is only a good approximation for very short time intervals t to $t + \Delta t$. All taken together, the expression for the variance becomes

$$\begin{aligned}\mathrm{Var}[\bar{\lambda}] = \frac{1}{c^2}\sum_{a=1}^{c}\mathrm{Var}[\lambda(t_a)] + \frac{1}{c^2}\sum_{a=1}^{c}\sum_{b>a}^{c}\sum_{j=1}^{K}\sum_{i=1}^{K} &S_j(t_a) \times P(Z(t_b) = i | Z(t_a) = j) \\ &\times (\lambda_j - \mathbb{E}[\lambda(t_a)]) \times (\lambda_i - \mathbb{E}[\lambda(t_b)]).\end{aligned} \tag{S4.42}$$

When comparing $\mathrm{Var}[\bar{\lambda}]$ between the simulated, stochastic mapping approach and the theoretical approach (Fig. S4.5), we can see that the theoretical approach converges when more time bins are used, and 20 or 30 time bins are perhaps sufficient for a reasonable approximation. One caveat of the example in Fig. S4.5 is, however, that it was made with a toy model

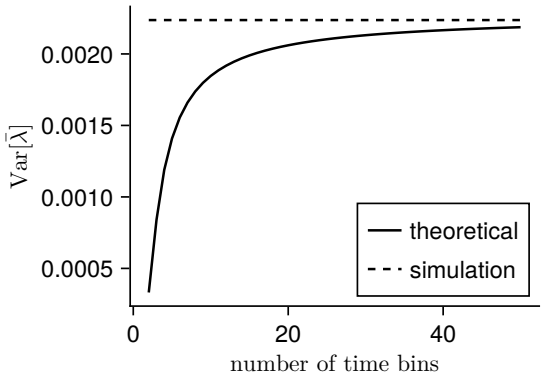


Figure S4.5: The posterior variance of the across-branch mean speciation rate, using i) a simulated, stochastic mapping approach, and ii) a theoretical approach that uses a time-bin approximation. The units of $\mathrm{Var}[\bar{\lambda}]$ are "number of speciation events per lineage per Ma" squared. When sufficiently many number of time bins are used, then the variance converges.

that only has four diversification rate categories. In a more typical setting, where we want more robust estimates of the branch-specific diversification rates, we would like to use more diversification rate categories, for example $n^2 = K = 36$ or $n^2 = K = 100$. The equation to calculate $\mathrm{Var}[\bar{\lambda}]$ contains four nested sums, and we expect that for a larger model it will become cumbersome to calculate the variance in this way. Therefore, if the across-branch variance in diversification rates is of interest, we recommend to use the stochastic mapping approach as used in Höhna et al. (2019).

S4.6 The likelihood surface

The likelihood surface, when plotted as a function of $\hat{\lambda}$ and $\hat{\mu}$, has several ridges with local optima (Fig. S4.6). This type of surface can be difficult to traverse using numerical methods. Optimization algorithms can find a local maximum instead of the global maximum likelihood parameter values. Similarly, Markov chain Monte Carlo samplers can have a difficult time switching between the ridges. This means that it is difficult to sample the posterior distribution, and in practice it can lead to problems with convergence. In `Pesto` we used a two-step approach to find estimates for $\hat{\lambda}$, $\hat{\mu}$ and η. Based on our exploration of the likelihood surface for the primates tree, this will not lead to the global maximum likelihood parameter estimates, but instead something that is reasonably close (Fig. S4.6, white dot), within a few log-likelihood units of the global maximum.

S4.7 Estimation error in branch-specific diversification rates

In the main text, we showed the proportional error in the inference of branch-specific speciation rates. Here, we show the equivalent in branch-specific extinction rates Fig. S4.7. The figure shows that most proportional errors are between 0.33 and 3.0, meaning that branch-specific extinction rate estimates are underestimated by a third, or overestimated by a factor of three. However, as the tree size increases, the proportional errors converge to approximately zero, meaning that the method is approximately unbiased. We also calculated branch-specific estimation errors for net-diversification rates. Since branch-specific net-diversification rates are often negative, however, it made less sense to use the geometric mean as a summary method per tree. For this reason, we plot the arithmetic mean of the branch-specific estimation errors in Fig. S4.8.

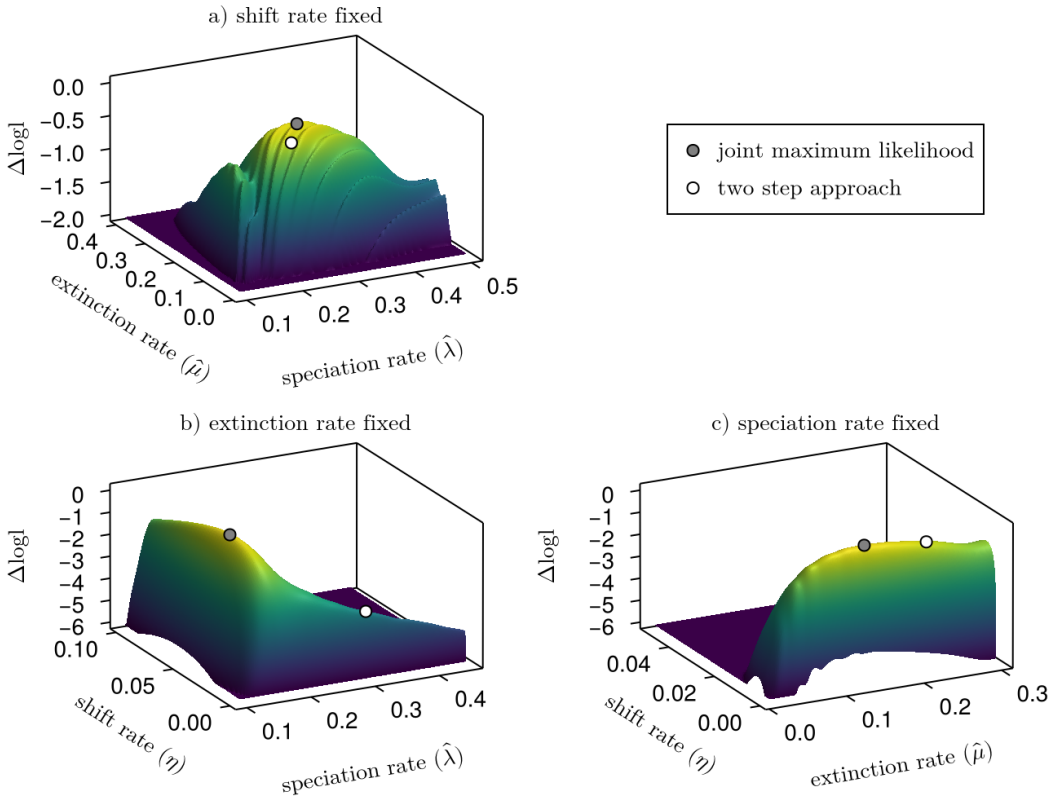


Figure S4.6: The likelihood surface of the birth-death-shift model as a function of the three parameters $\hat{\lambda}$, $\hat{\mu}$ and η. From $\hat{\lambda}$, $\hat{\mu}$, we take 10 quantiles from the distributions LogNormal($\log(\hat{\lambda})$, sd = 0.587) and LogNormal($\log(\hat{\mu})$, sd = 0.587), and use all pairwise comparisons to form $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. The log likelihoods are offset such that the maximum is zero. The data is the primates phylogeny. The three panels depict grid searches where we held one parameter constant in each panel. In panel a) we set $\eta = 1/\sum_i b_i$, where b_i is the length of branch i. In b) we set $\hat{\mu} = 0.22$, and in c) we set $\hat{\lambda} = 0.30$ (the maximum-likelihood parameter values under the constant-rate birth-death process). Notice that the parameter values inferred using the two-step approach is within a few log-likelihood units of the joint maximum likelihood approach, however they are not identical.

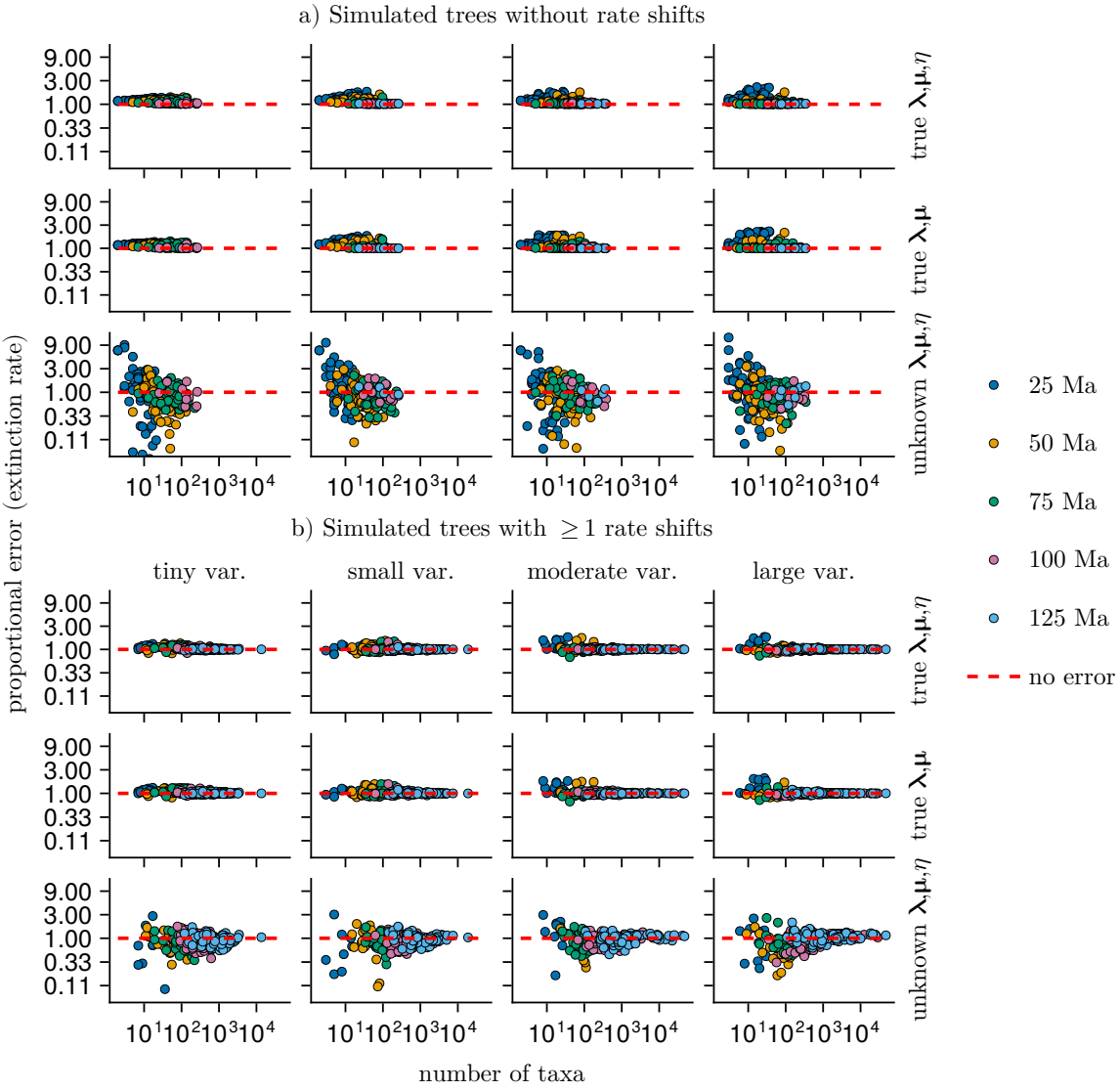


Figure S4.7: The estimation error in branch-specific extinction rates, as proportional errors. For each tree simulated tree, we calculated the proportional errors per branch, and averaged across the tree: $\exp(\frac{1}{n_{\text{branches}}} \sum_M \log(\bar{\mu}_M) - \log(\bar{\mu}_{M,\text{true}}))$, where $\bar{\mu}_M$ is the estimate for the average extinction rate on branch M. A value of one means that the branch-specific extinction rates are unbiased, on average across the tree. Larger than one means overestimation, and less than one means underestimation. The proportional errors for extinction rates are slightly larger than those for speciation rates.

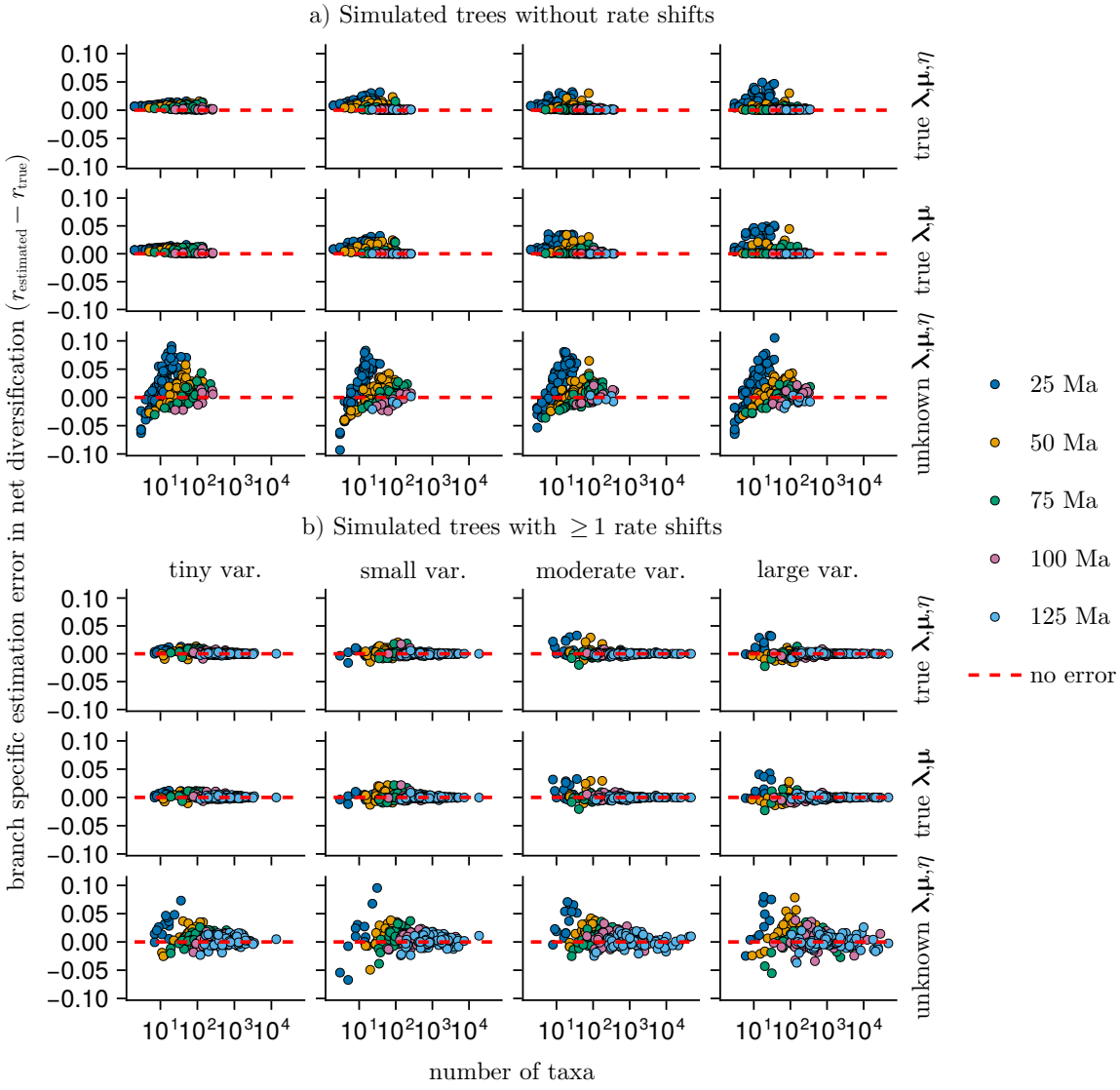


Figure S4.8: The estimation error in branch-specific net-diversification rates. For each tree simulated tree, we calculated the estimation errors per branch, and averaged across the tree: $\frac{1}{n_{\text{branches}}} \sum_M (\bar{r}_M - \bar{r}_{M,\text{true}})$, where $\bar{r}_M$ is the estimate for the average net-diversification rate on branch M. Note that this plot shows arithmetic means, unlike geometric means in the other figures of estimation error. A value of zero means that the branch-specific extinction rates are unbiased, on average across the tree. Larger than zero means overestimation, and less than zero means underestimation. Seven data points with an error of about −0.61 are omitted from the figure due to the choice of y-axis limits.

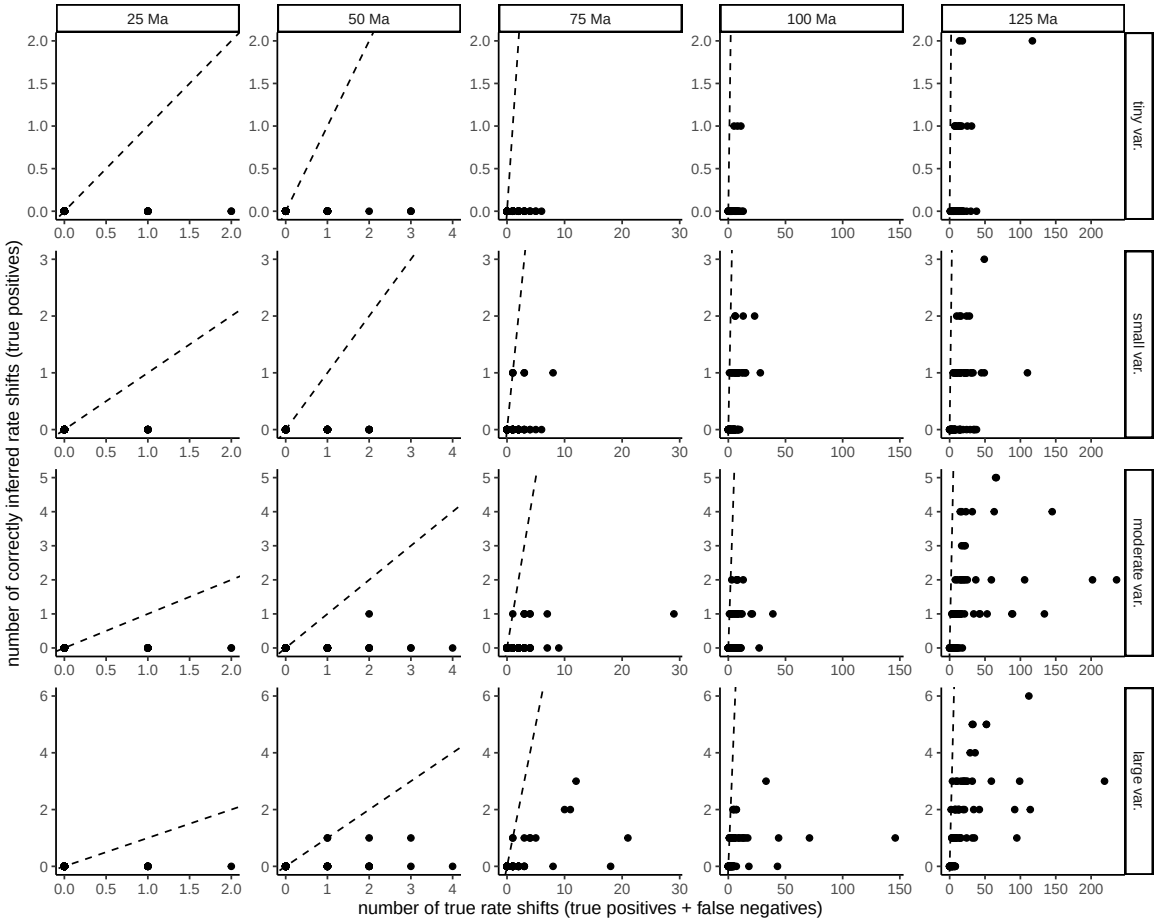


Figure S4.9: The number of correctly inferred rate shifts (i.e. true positives) versus the number of rate shifts that happened (true positives + false negatives) in the simulated trees. Each point represents one phylogeny. The criteria for that a shift was inferred was that the Bayes factor had to be larger than 10, and $\hat{N}$ had to be larger than 0.5. The panels are split by simulation time (from 25 to 125 Ma) and rate variation in the true model (i.e. models A to D in Table 1). The dashed lines represent the one-to-one relationship.

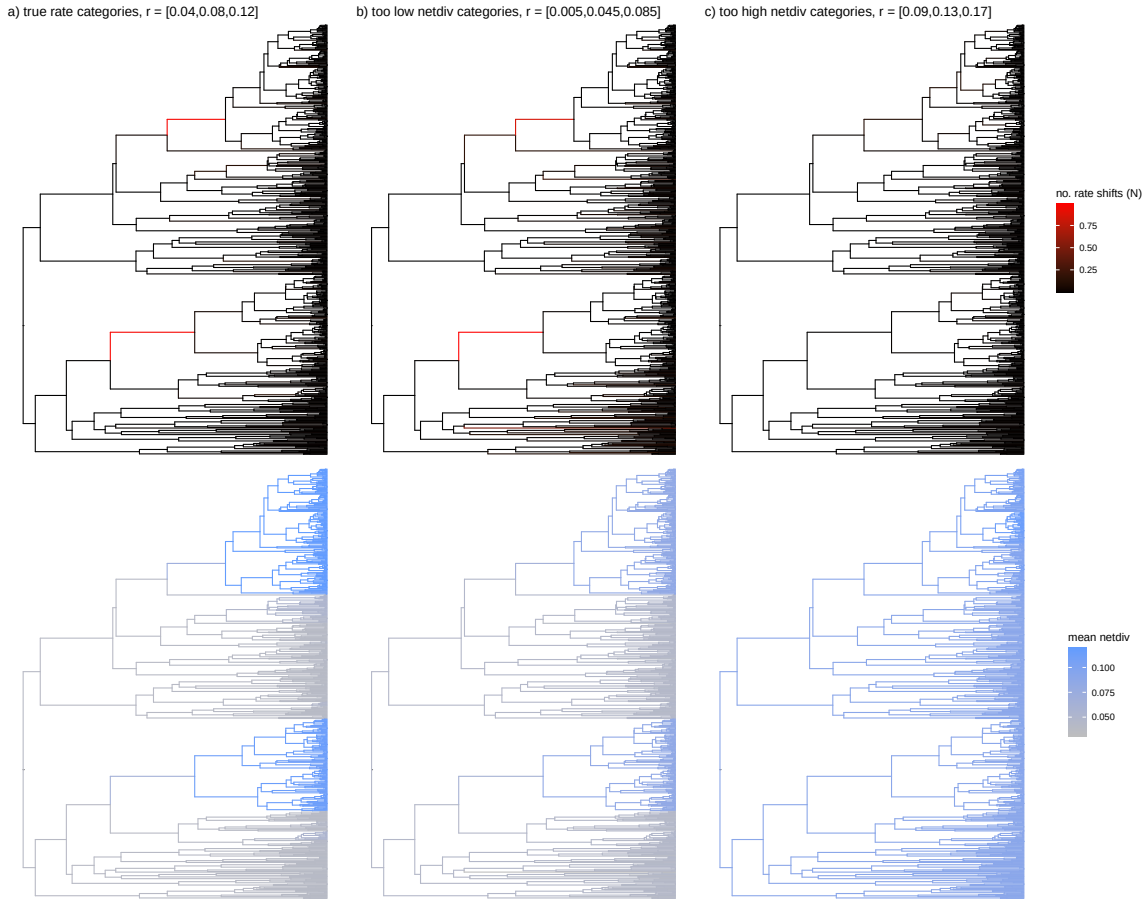


Figure S4.10: A brief example of what the result might be if the rate categories are mis-specified. To emulate this, we set the net-diversification rates to a) their true values, b) too low and c) too high. The relative extinction (i.e. $\mu/\lambda = 2/3$) is equal across all three models and across the rate categories. The phylogeny is the same that was used in the main text in Fig. 4.10b, simulated from the model with large rate variation. In all three models we estimated the shift rate using maximum likelihood. In a) and b), the two rate shifts that led to the radiation of species-rich clades are recovered. However, in b) the net-diversification rates in the two in-groups are underestimated, and there is one false positive. In c) the shift rate estimate is about zero, no shifts are recovered, and the net-diversification rates are overestimated in what would be the "outgroup".

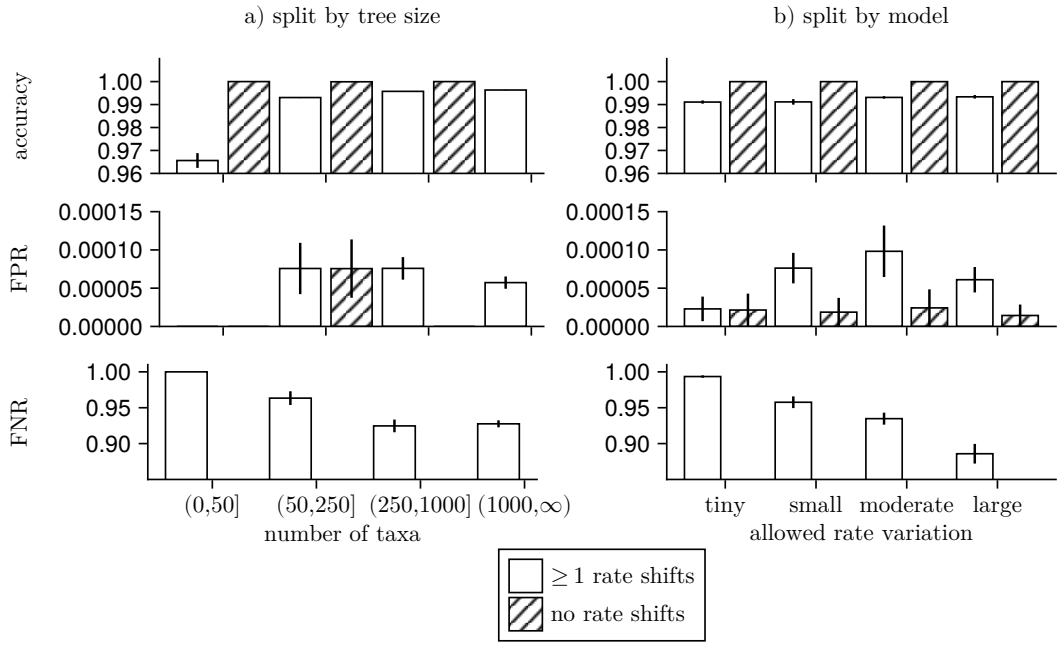


Figure S4.11: Equivalent to Fig. 4.11 in the main manuscript, however with the y-axis limits chosen on a much smaller interval than {0,1}, better show-casing the variation within each panel.

Appendix D

Supplementary Material for Chapter 5

Table S5.1: A summary of the datasets. The age (in millions of years) is the time from the most-recent common ancestor to the present. The sampling fraction is the ratio of species in the tree to the number of described species. The net-diversification rate is the branch-specific net-diversification rate, averaged (but weighted by branch length) over the phylogeny. N^* is the number of branches that showed strong support (Bayes factor > 10) for that there was a rate shift event.

Name	Tax. level	Reference	Age (Ma)	#Taxa	Netdiv	N^*	Sampl.	Sampling reference
Actinopterygii	Class	(Rabosky et al., 2018)	368	11638	0.0811	470	0.37	(Rabosky et al., 2018)
Agaricomycetes	Class	(Varga et al., 2019)	347	5135	0.0429	34	0.26	(Varga et al., 2019)
Anthophila	Unranked	(Henríquez-Piskulich et al., 2024)	120	4586	0.138	124	0.23	(Henríquez-Piskulich et al., 2024)
Anura	Order	(Portik et al., 2023)	179	5242	0.0933	22	0.69	(Portik et al., 2023)
Aristolochiaceae	Family	(Allio et al., 2021)	111	247	0.120	8	0.40	(Mulder, 2003)
Asteraceae	Family	(Palazzesi et al., 2022)	80	2723	0.341	135	0.12	(Palazzesi et al., 2022)
Aves	Class	(Quintero et al., 2022)[1]	111	9993	0.113	35	1.00	(Quintero et al., 2022)
Caryophyllales	Order	(Smith et al., 2018)	111	5036	0.0726	156	0.40	(Smith et al., 2018)
Chondrichthyes	Class	(Stein et al., 2018)	378	610	0.0274	3	0.51	(Stein et al., 2018)
Cornales	Order	(Rose et al., 2018)	111	410	0.074	11	0.68	(Thomas et al., 2021)
Cycadaceae	Family	(Liu et al., 2021)	13	103	0.313	1	0.88	(Liu et al., 2021)
Ericales	Order	(Rose et al., 2018)	110	4532	0.154	130	0.36	(Rose et al., 2018)
Gesnerioideae	Subfamily	(Serrano-Serrano et al., 2017)	44	588	0.168	4	0.49	(Serrano-Serrano et al., 2017)
Lecanoromycetes	Class	(Nelsen et al., 2020)	249	3373	0.0505	93	0.25	(Nelsen et al., 2020)
Mammalia	Class	(Álvarez-Carretero et al., 2022)	202	4705	0.137	70	0.70	(Upham et al., 2024)
Orobanchaceae	Family	(Mortimer et al., 2022)	35	917	0.451	55	0.40	(Mortimer et al., 2022)
Phasmatodea	Order	(Bank and Bradler, 2022)	179	513	0.0796	4	0.15	(Bank and Bradler, 2022)
Pinophyta	Division	(Leslie et al., 2018)	324	578	0.0303	13	0.90	(Leslie et al., 2018)
Poaceae	Family	(Spriggs et al., 2014)	58	3595	0.313	88	0.29	(Spriggs et al., 2014)
Polypodiophyta	Division	(Nitta et al., 2022)	420	5685	0.0544	186	0.46	(Nitta et al., 2022)
Rhopalocera	Suborder	(Kawahara et al., 2023)	101	2244	0.142	23	0.12	(Kawahara et al., 2023)
Rosidae	Subclass	(Sun et al., 2020)	118	19740	0.177	1437	0.19	(Sun et al., 2020)
Salvia	Genus	(Kriebel et al., 2019)	31	519	0.292	5	0.52	(Kriebel et al., 2019)
Squamata	Order	(Zheng and Wiens, 2016)	205	4162	0.0657	42	0.43	(Zheng and Wiens, 2016)

[1] We obtained the tree file from (Quintero et al., 2022), but the phylogenetic reconstruction was originally done by (Jetz et al., 2012).

S5.1 Materials and methods

Empirical phylogenies

When selecting phylogenies, we focused on clades that were species-rich and studies that used probabilistic methods to infer the phylogeny. Specifically, we only used trees that had positive branch lengths in units of time (Ma), were ultrametric, and binary splitting. When a publication provided a summary tree from the tree inference procedure, such as a MAP (maximum *a posteriori*) or MCC (maximum clade credibility) summary tree, we used that summary tree. For publications in which a summary tree was not provided, we used the posterior distribution of trees and computed the MCC tree with common ancestor heights using `TreeAnnotator` in `BEAST` (Drummond et al., 2012a).

Modeling clade-specific shifts in diversification rates

We used the birth-death-shift model (Höhna et al., 2019) implemented in the Julia module `Pesto` (Kopperud and Höhna, 2025). The birth-death-shift model is similar to the state-dependent speciation extinction (SSE) model (Maddison et al., 2007; FitzJohn, 2012), except that the states are considered unknown or unobserved. The method is conceptually similar to BAMM (Rabosky, 2014) and the multi-type birth-death model (Barido-Sottani et al., 2020), however the results of an analysis are most similar to those of BAMM (see Martínez-Gómez et al., 2024, for a review). The advantage of `Pesto` over the aforementioned methods is that `Pesto` uses a deterministic algorithm instead of time-consuming Markov-chain Monte Carlo sampling to compute the branch-specific diversification rates, and the number of rate shifts.

The diversification rate categories

When a rate shift event happens, the new speciation and extinction rates are independent of the ancestral rates. The new speciation and extinction rates follow some base distributions (Höhna et al., 2019), where we chose log-normal distributions:

$$\begin{aligned} &\text{LogNormal}(\log(\hat{\lambda}), 0.587) \\ &\text{LogNormal}(\log(\hat{\mu}), 0.587), \end{aligned} \tag{S5.1}$$

which have parameters $\hat{\lambda}, \hat{\mu}$ controlling the scale, and a standard deviation of 0.587. The standard deviation of 0.587 is chosen such that the 2.5–97.5% quantile interval spans one order of magnitude on the linear rate scale (for example, a speciation rate ranging from 0.1 to 1.0). While the birth-death-shift model is conceptually continuous, we use a discrete approximation in order for the model to be mathematically and computationally tractable. Thus, we discretize the log-normal distributions into quantiles and store the middle point of each quantile as $\vec{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]$, and likewise for $\vec{\mu}$. Using all pairwise combinations of $\vec{\lambda}$ and $\vec{\mu}$, we set up two more vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, comprised of $n^2 = K$ elements. The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the rate categories. The number of rate categories is decided before the model is

fitted. In the empirical analyses, we used $n^2 = K = 100$ rate categories. The rate categories represent the allowed rate variation in the birth-death-shift model. As the number of rate categories increases, the model converges to the full, continuous distribution of diversification rate variation (Kopperud and Höhna, 2025).

Parameter estimation

The birth-death-shift model has three parameters: $\hat{\lambda}$ controls the mean of the speciation rate distribution, $\hat{\mu}$ controls the mean of the extinction rate distribution, and η controls how often the process shifts from one category to another. Using maximum likelihood, we obtain joint estimates of $\hat{\lambda}, \hat{\mu}$ and η. We do this by setting up a series of constraints on the model fitting procedure, for example, by forcing η to be smaller than the other two parameters. We use a numerical optimization algorithm, specifically Newton's method to iteratively improve the likelihood (Mogensen and Riseth, 2018), until a solution is reached. See the Supplementary text for a more detailed description of the optimization procedure.

Branch-specific rates and rate shift support

When calculating the likelihood of the model, and therefore also when we are fitting the parameters $(\hat{\lambda}, \hat{\mu}, \eta)$, we are integrating over all possible rate shift histories. Thus, the rate shift history does not appear as a parameter in our model. Despite this, the rate shift history is a random, unobserved variable that has a prior distribution (determined by the model) and a posterior distribution (determined by the model and the data). Mapping the posterior distribution of a rate shift history can be done using stochastic mapping (Huelsenbeck et al., 2003; Freyman and Höhna, 2019; Höhna et al., 2019), however this can be a computational burden (Martínez-Gómez et al., 2024). Instead of using stochastic maps to estimate branch-specific diversification rates, we use a deterministic algorithm (Kopperud and Höhna, 2025). Specifically, we calculate the posterior mean diversification rate per branch as

$$r = \frac{1}{t_1 - t_2} \int_{t_1}^{t_2} \sum_{i=1}^{K} S_{M,i}(t)(\lambda_i - \mu_i) dt, \tag{S5.2}$$

where $S_{M,i}(t)$ is the marginal probability of the rate category being i at time i on branch M, and t_1, t_2 are the oldest and youngest times of branch M, respectively. $S_{M,i}(t)$ is normalized, meaning that it sums to 1 over the rate categories i for any particular time t. We calculate the change in the net-diversification rate along a branch as follows

$$\Delta r = \sum_{i=1}^{K} S_{M,i}(t_2)(\lambda_i - \mu_i) - \sum_{i=1}^{K} S_{M,i}(t_1)(\lambda_i - \mu_i), \tag{S5.3}$$

i.e., as the difference between the means of the oldest and the youngest times of the branch.

In order to assess the statistical support for rate shift events, we use an approximate Bayes factor approach (Shi and Rabosky, 2015). This takes into account both the prior probability

(π) and posterior probability (P) of whether or not there was a rate shift event on branch M

$$\text{Bayes factor} = \frac{\frac{P_M(\geq 1 \text{ shifts})}{\pi_M(\geq 1 \text{ shifts})}}{\frac{P_M(0 \text{ shifts})}{\pi_M(0 \text{ shifts})}}. \tag{S5.4}$$

We approximate the prior probability π by calculating the probability of zero events given a Poisson distribution with rate $(t_1 - t_2)\eta$. For the posterior probability, see (Kopperud and Höhna, 2025). We consider a branch to have strong statistical support for there being a rate shift event, if the Bayes factor is 10 or higher. To summarize how many branches had strong support across a phylogeny, we use the symbol N^* to represent the number of such branches.

Note that the branch rate estimates, as well as the Bayes factors, are given conditionally on the point estimates of $\hat{\lambda}, \hat{\mu}, \eta$. In effect, we are ignoring estimation error in these parameters. While there is no doubt large uncertainty in our parameter estimates, this omission is a compromise that allows us to obtain branch-specific estimates fast and efficiently even for large trees.

Validation studies

In order to test whether the method can robustly infer the patterns we found from the empirical data, we set up a series of simulation studies as part of a validation procedure. In the Supplementary Materials, we provide several sections where we present the true rate shift quantities from simulated phylogenies, the best estimates from our inference procedure, and the estimation error.

Software

We analysed the phylogenies using `Pesto` (Kopperud and Höhna, 2025), performed post-processing using R packages for phylogenetic manipulation (Paradis and Schliep, 2019; Wang et al., 2020), and plotted the figures using `ggtree` (Yu et al., 2017) and `Makie.jl` (Danisch and Krumbiegel, 2021). We also used `TreeAnnotator` to create summary trees from posterior distributions of trees (Drummond et al., 2012a).

S5.2 Supplementary figures

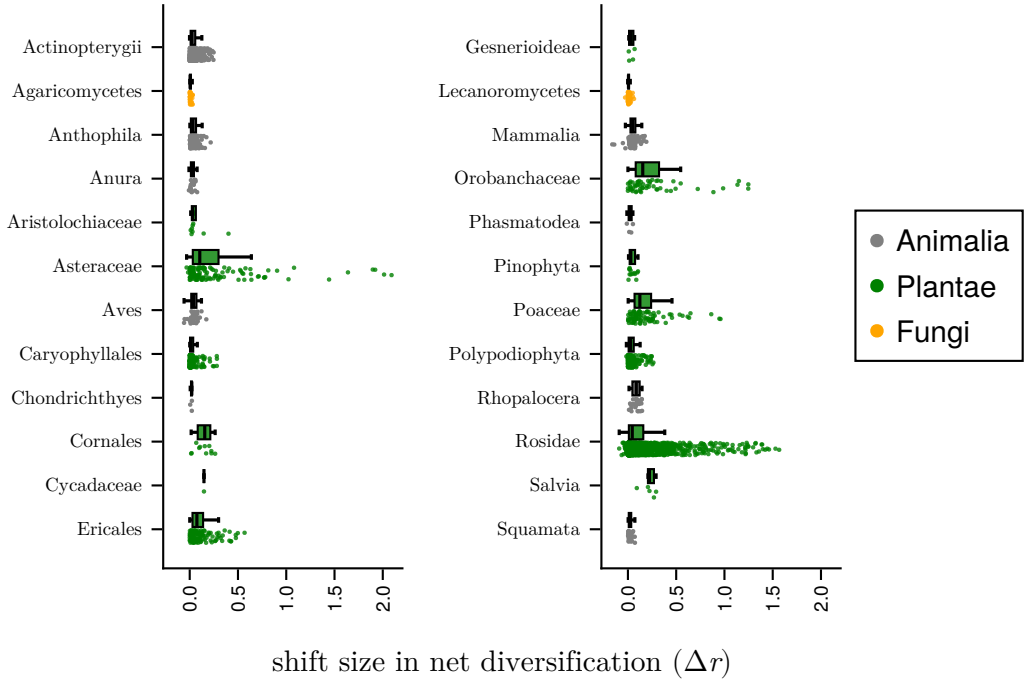


Figure S5.1: An extended version of Figure 5.2, where all of the empirical datasets are included. Each dot represents a branch in the phylogeny that has strong support (Bayes factor > 10) for that there was at least one rate shift event. Note that the majority of shifts are in positive direction, however there are some that are negative. The box plots show the median, with the crossbar spanning the interquartile range, and the length of the whiskers are 1.5 times the interquartile range.

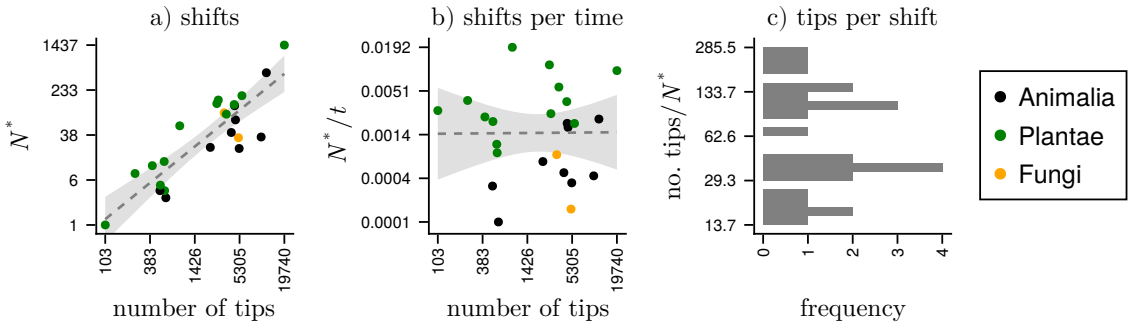


Figure S5.2: The number of diversification rate shift events, and the number of rate shifts per time (Ma) versus the number of tips included in the reconstructed phylogeny. Each dot represents one empirical phylogeny (see Extended Data Table S5.1). The lines are ordinary least-squares regressions, and the shaded areas represents two standard error deviations from the lines.

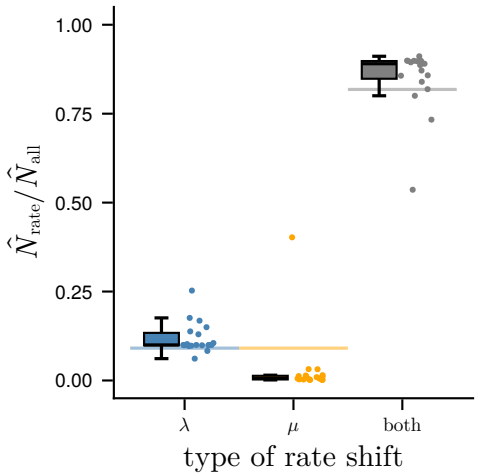


Figure S5.3: The fraction of number of shifts in speciation rate ($\hat{N}_\lambda$), extinction rate ($\hat{N}_\mu$), and both ($\hat{N}_{\lambda+\mu}$) over all rate shifts ($\hat{N}_{\text{all}} = \hat{N}_\lambda + \hat{N}_\mu + \hat{N}_{\lambda+\mu}$). For more details on the posterior number of rate shift events $\hat{N}$, see the Supplementary text. We summed $\hat{N}_{\text{rate}}$ across the branches that showed strong support for there being a rate shift event (Bayes factor > 10). One empirical phylogeny is represented by three dots (blue, orange, gray), summing to one. We expect *a priori* (horizontal lines) that there are $(n-1)/(n^2-1)$ shifts in a single rate, and $(n-1)^2/(n^2-1)$ joint shifts, or $1/11$ and $9/11$ in this case (since the number of speciation and extinction rate classes was $n = 10$). The box plots show the median, with the crossbar spanning the interquartile range, and the length of the whiskers are 1.5 times the interquartile range.

S5.3 Supplementary text

S5.3.1 Inferring the type of diversification rate shifts using simulated phylogenies

We were interested in assessing whether `Pesto` is able to detect what kind of rate shift event that happened on a branch. Specifically, we investigated whether a rate shift event is due to i) a change in the speciation rate, or ii) a change in the extinction rate. In order to test if our inference is able to detect which of the three scenarios that occurred, we simulated several phylogenies where there was exactly zero or one rate shift event. In expectation, the simulation procedure results in a phylogeny that has a moderate "backbone" tempo of diversification. In the subclade that is descended from the rate shift event, it either resulted in an upshift or downshift in net diversification. The specifics of the simulated phylogenies are discussed in detail below.

Table S5.2: Summary of the models we used to simulate grafted trees (Figs. S5.4 to S5.8).

Model	μ	λ	$\lambda - \mu$	age (Ma)	$\mathbb{E}$[#tips|survival]
backbone	0.325	0.235	0.09	60	1593.8
upshift (λ)	0.52	0.235	0.285	15	260.7
upshift (μ)	0.325	0.0	0.325	15	261.9
downshift	0.13	0.235	−0.105	40	4.4

$\mathbb{E}[\#\text{tips}|\text{survival}] = 2(1 + \frac{\lambda}{\lambda-\mu}(e^{t(\lambda-\mu)} - 1))$ is the expected number of tips for a phylogeny simulated from a constant-rate birth-death process with birth rate λ, death rate μ, over a time period of t, conditional on that both the left and the right subtrees survived until the present (Höhna, 2015, appendix B). Note that the process begins with a speciation event at the origin, i.e., it has two active lineages from the start.

We simulated the phylogenies using the R-package `TESS` (Höhna et al., 2016b) under various settings, see Table S5.2 for a summary. Specifically, we conditioned on that both the left and right lineages that descended from the root node survived until the present. Using a backbone model with a moderate net-diversification rate ($\lambda - \mu = 0.09$), we simulated 350 trees over a time span of 60 Ma. In each of these 350 backbone trees, we randomly selected a branch that was alive at 15 Ma in the reconstructed tree. At this branch, we pruned the original subtree and replaced it with a separate simulated tree, which we simulated from a model with a higher net-diversification rate. The upshift can either be due to an increase in speciation rate, or a decrease in extinction rate. We selected the change in speciation and extinction rates such that the expected number of tips in the surviving tree was about 260, over the time span of 15 Ma. We performed a similar prune-and-regraft to assess downshifts in diversification, however at an earlier time (at 40 Ma instead of at 15 Ma).

Next, we used `Pesto` to infer branch-specific diversification rates. Ten trees selected at random are plotted in Figs. S5.4 to S5.7, and the speciation, extinction and net-diversification rates are mapped with colors on the branches. When there are no diversification rate shift

events, `Pesto` tends to correctly infer that there were no shift events Fig. S5.4. This is expected, as we previously confirmed that `Pesto` has a low false positive rate (Kopperud and Höhna, 2025). For the trees in which there was an upshift event, `Pesto` is usually able to recover the main rate shift event, although not always Figs. S5.5 and S5.6. When there was an upshift event, `Pesto` rarely recovers a different upshift as a false positive (e.g., Fig. S5.5, tree 4). When there was a downshift event, as in Fig. S5.7, `Pesto` is almost never able to detect it. Thus, we are able to correctly recover i) constancy, ii) upshifts in net-diversification, but not iii) downshift in net-diversification.

Inferring whether the upshift event is due to i) a change in the speciation rate, or ii) a change in the extinction rate, appears to be a more difficult task. We first motivate with color-mapped phylogenies, and follow up with a more quantitative assessment. In some phylogenies (e.g., Fig. S5.5 trees 1,2), `Pesto` correctly infers that there was a change in the speciation rate. In many other phylogenies (e.g., Fig. S5.5 trees 3,5,6,8), `Pesto` incorrectly infers that the rate shift event was a joint increase in speciation and decrease in extinction. Similar results can be seen in Fig. S5.6, where `Pesto` is sometimes but usually not able to correctly infer that the rate shift event was due to a decrease in extinction.

The results for the number of diversification rate shifts across the full 350 simulated phylogenies is summarized in Fig. S5.8. Here, we make use of the quantity $\hat{N}_M$, which represents the posterior mean number of diversification rate shift events, for a particular branch M. We will recap how to calculate and interpret $\hat{N}_M$, but see (Kopperud and Höhna, 2025) for further explanations including how it is derived. Moreover, we specify $\hat{N}_{M,ij}$ for the number of diversification rate change events from rate category j a different rate category i. We use the following differential equation to calculate $\hat{N}_{M,ij}$

$$\frac{d\hat{N}_{M,ij}(t)}{dt} = \begin{cases} -S_{M,j}(t)\dfrac{D_{M,i}(t)}{D_{M,j}(t)}\dfrac{\eta}{K-1} & \text{if } j \neq i \\ 0 & \text{if } j = i, \end{cases} \tag{S5.5}$$

where $D_{M,i}(t)$ is the probability of observing the clade descended from branch M at time i, given that the process was in rate category i at time t. The quantity $S_{M,j}(t)$ is the marginal probability that the rate category was j at time t on branch M. The initial value for $\hat{N}(t)$ is set to 0 at the oldest time point of the branch, and we use a numerical ODE solver to find a solution at the youngest time point. Solving this for all combinations of i, j gives us a matrix $\hat{\boldsymbol{N}}_M$. Next, we sum over the set of branches $M \in \mathcal{M}$ to obtain

$$\hat{\boldsymbol{N}} = \sum_{M \in \mathcal{M}} \hat{\boldsymbol{N}}_M, \tag{S5.6}$$

where $\mathcal{M}$ is the set of branches that showed strong statistical support (with Bayes factor > 100, more conservative than in the main text). This gives us a summary of the number of rate

change events that can be plotted without the phylogeny. Furthermore, we can decompose $\hat{\boldsymbol{N}}$ and categorize the number of rate shift event by whether it represents a change in the speciation rate ($\hat{N}_\lambda$), the extinction rate ($\hat{N}_\mu$) or a change in both rates simultaneously ($\hat{N}_{\lambda+\mu}$). If we drop the subscripts ij, and use an example with few rate classes (here $n = 2$, $K = 4$), it becomes easier to write the full matrix

$$\hat{\boldsymbol{N}} = \begin{bmatrix} - & \hat{N}_\lambda & \hat{N}_\mu & \hat{N}_{\lambda+\mu} \\ \hat{N}_\lambda & - & \hat{N}_{\lambda+\mu} & \hat{N}_\mu \\ \hat{N}_\mu & \hat{N}_{\lambda+\mu} & - & \hat{N}_\lambda \\ \hat{N}_{\lambda+\mu} & \hat{N}_\mu & \hat{N}_\lambda & - \end{bmatrix}. \tag{S5.7}$$

Note that the inferences depicted in Figs. S5.4 to S5.8 were made using $n = 10$ rate classes. Therefore, the actual $\hat{\boldsymbol{N}}$ matrix in this simulation study has 100 rows and 100 columns.

In order to summarize whether a diversification rate change is due to a change in the speciation or the extinction rate, we use the posterior mean number of diversification rate shifts $\hat{N}_{ij}$ to construct three summary metrics. The metrics describe the net change in the rate (speciation, extinction or net-diversification), weighted by the number of rate change events

$$\begin{aligned} \Delta\lambda &= \sum_{i,j} (\lambda_i - \lambda_j)\hat{N}_{ij} \\ \Delta\mu &= \sum_{i,j} (\mu_i - \mu_j)\hat{N}_{ij} \\ \Delta r &= \sum_{i,j} (r_i - r_j)\hat{N}_{ij}, \end{aligned} \tag{S5.8}$$

where $r_i = \lambda_i - \mu_i$ represents the net-diversification rate in rate category i. The results are depicted in Fig. S5.8. Each simulated phylogeny represents a count in the histogram, and the vertical dashed line represents the true change in each rate.

There are five striking results. First, the diversification-rate changes in the constant-rate backbone trees are inferred almost perfectly, with only a handful of false positive upshifts in net diversification. Second, the downshifts are never recovered correctly — the trees that include downshifts appear as if they have been simulated by a constant-rate process. Third, the upshift scenarios (both due to changes in μ and λ) appear to be somewhat recovered, if we consider the inferred change in net-diversification (Δr, Fig. S5.8, right column), although there is considerable estimation error. Fourth, it seems to be very difficult to infer whether a rate shift event was due to a change in the speciation or the extinction rate. For the scenario where the true upshift was due to a change in the speciation rate, the estimates appear to be approximately centered around the true change (vertical dashed line). When the true upshift was due to a change in the extinction rate (third row), however, there is a mismatch between the inference and the true change. While the inferred change in speciation has the correct mode ($\Delta\lambda = 0$), the inferred change in extinction is strongly biased. As a fifth point, we do not

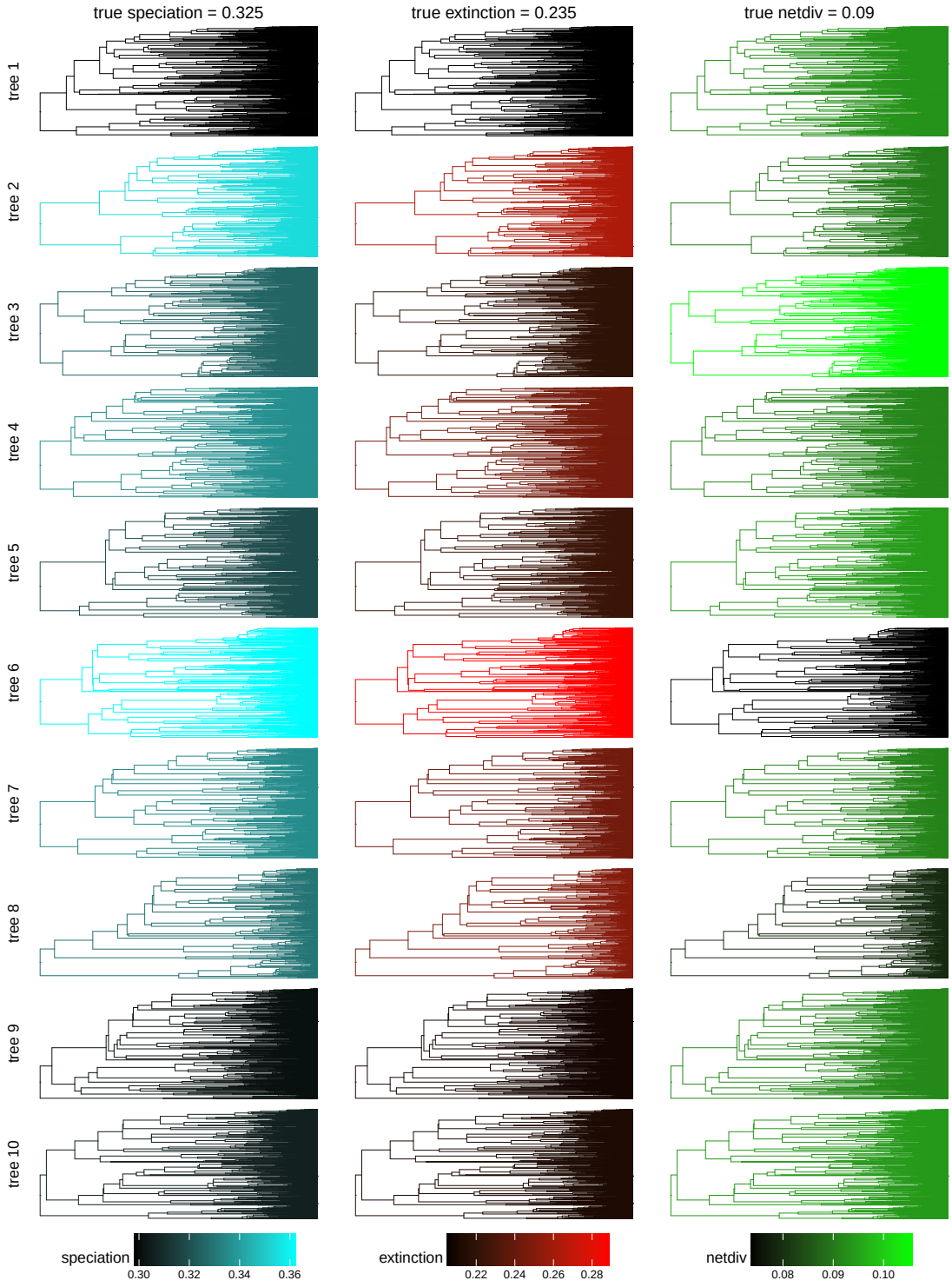


Figure S5.4: Ten trees where we simulated a backbone tree under the constant-rate birth-death model ($\lambda = 0.325, \mu = 0.235$) for a time period of 60 Ma. We inferred branch-specific speciation, extinction and net-diversification rates (shown in color). There is some estimation error in the rates, but we almost never recovered a rate shift event (low false positive ratio).

interpret these results as it being impossible to detect whether the upshift was due to a change in the speciation or the extinction rate. If the patterns in the data were identical, we would expect to see equal or very similar distributions in the second and third rows in Fig. S5.8. The difference can not be attributed to difference in total species-richness, as we conditioned both grafted in-groups to have about 260 expected number of tips. As the distributions are clearly different, we think that there is some signal available in the divergence times that gives information about the nature of the diversification rate shifts.

Telling whether an upshift is due to a change in the speciation or extinction rate is no doubt a difficult task. The ability of inferring which rate that shifted is perhaps limited by the size of the phylogeny. It may be the case that we need more species-rich trees to reliably infer the correct type of rate shift events (these trees had on average about 2000 tips). However, our results could also be driven by the model assumptions. In the birth-death-shift model that we set up, we implicitly assumed that the going from any rate category j to any other rate category i is happening at an equal rate. This implies i) that the rate of speciation and extinction shifts are equal, but also that ii) joint speciation+extinction rate shifts are more common than single shifts, and the ratio increases with the number of rate categories. We envision that the overall hypothesis could be addressed more elegantly if we i) allowed the rate of speciation and extinction shifts to be different, and ii) disallowed joint shifts entirely. Nevertheless, assessing whether a rate shift event is due to a change in the speciation or the extinction rate requires further research.

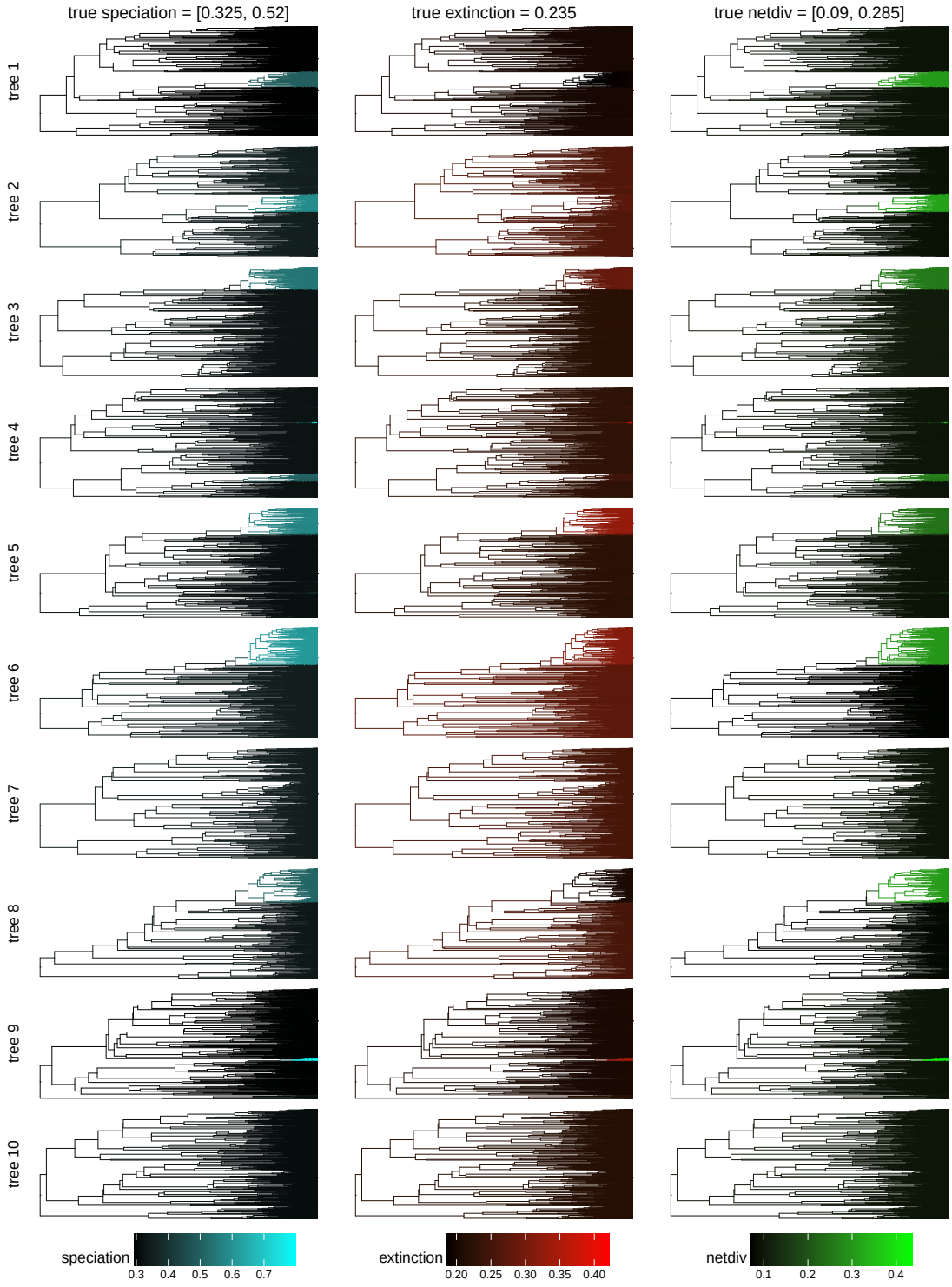


Figure S5.5: Ten trees where we simulated a backbone tree under the constant-rate birth-death model ($\lambda = 0.325, \mu = 0.235$) for a time period of 60 Ma, and replaced a randomly selected subtree at 15 Ma with a tree that had a higher speciation rate ($\lambda = 0.52$). We inferred branch-specific speciation, extinction and net-diversification rates (shown in color). The shift in net-diversification rate is usually recovered, but not always. When a shift in speciation rate occurs, it can often be mis-identified as a shift in the extinction rate, or as a joint shift (speciation and extinction rate).

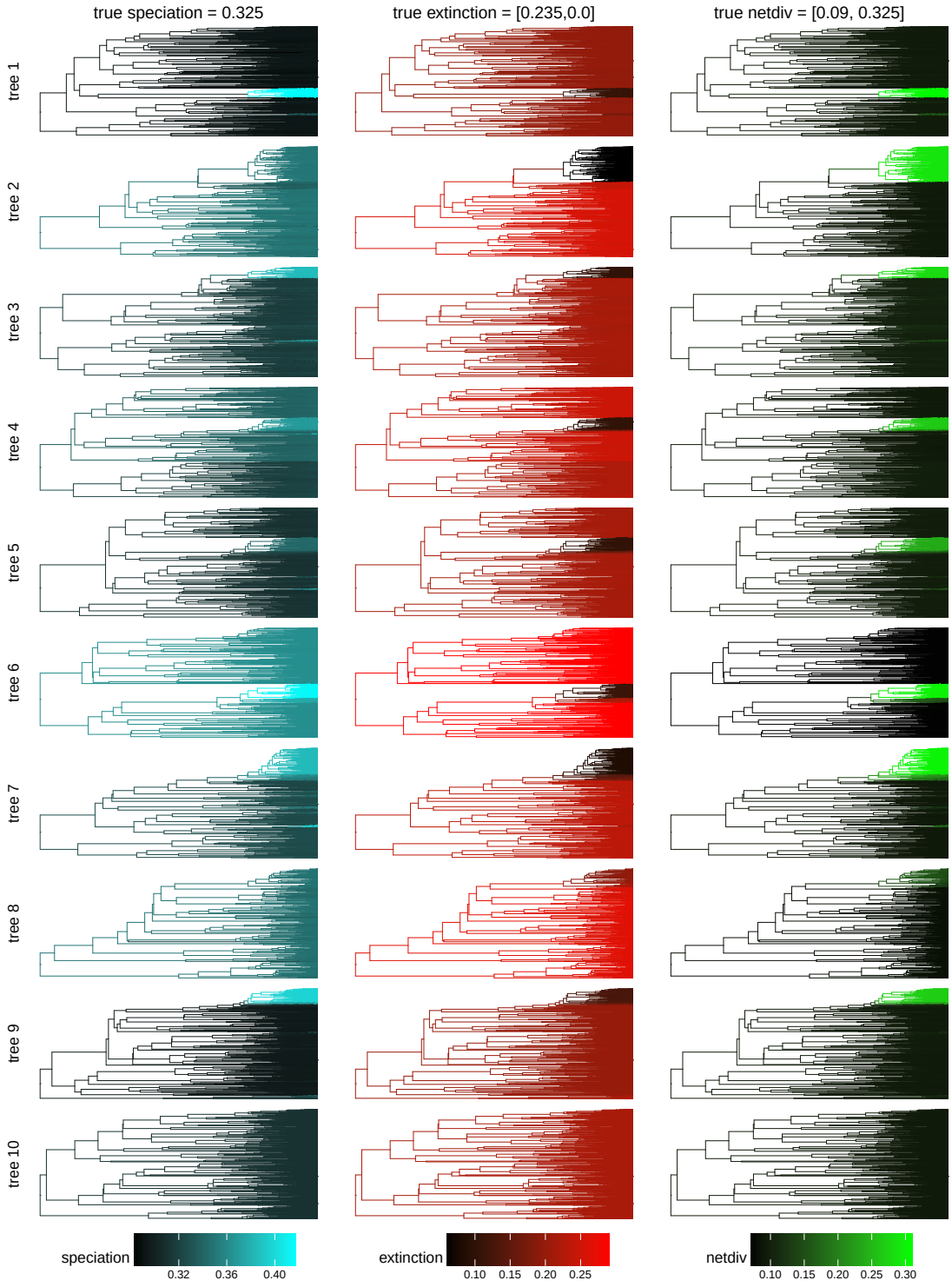


Figure S5.6: Ten trees where we simulated a backbone tree under the constant-rate birth-death model ($\lambda = 0.325, \mu = 0.235$) for a time period of 60 Ma, and replaced a randomly selected subtree at 15 Ma with a tree that had a lower extinction rate ($\mu = 0.0$). We inferred branch-specific speciation, extinction and net-diversification rates (shown in color). The shift in net-diversification rate is usually recovered, but not always. When a shift in extinction rate occurs, it can often be mis-identified as a shift in the speciation rate, or as a joint shift (speciation and extinction rate).

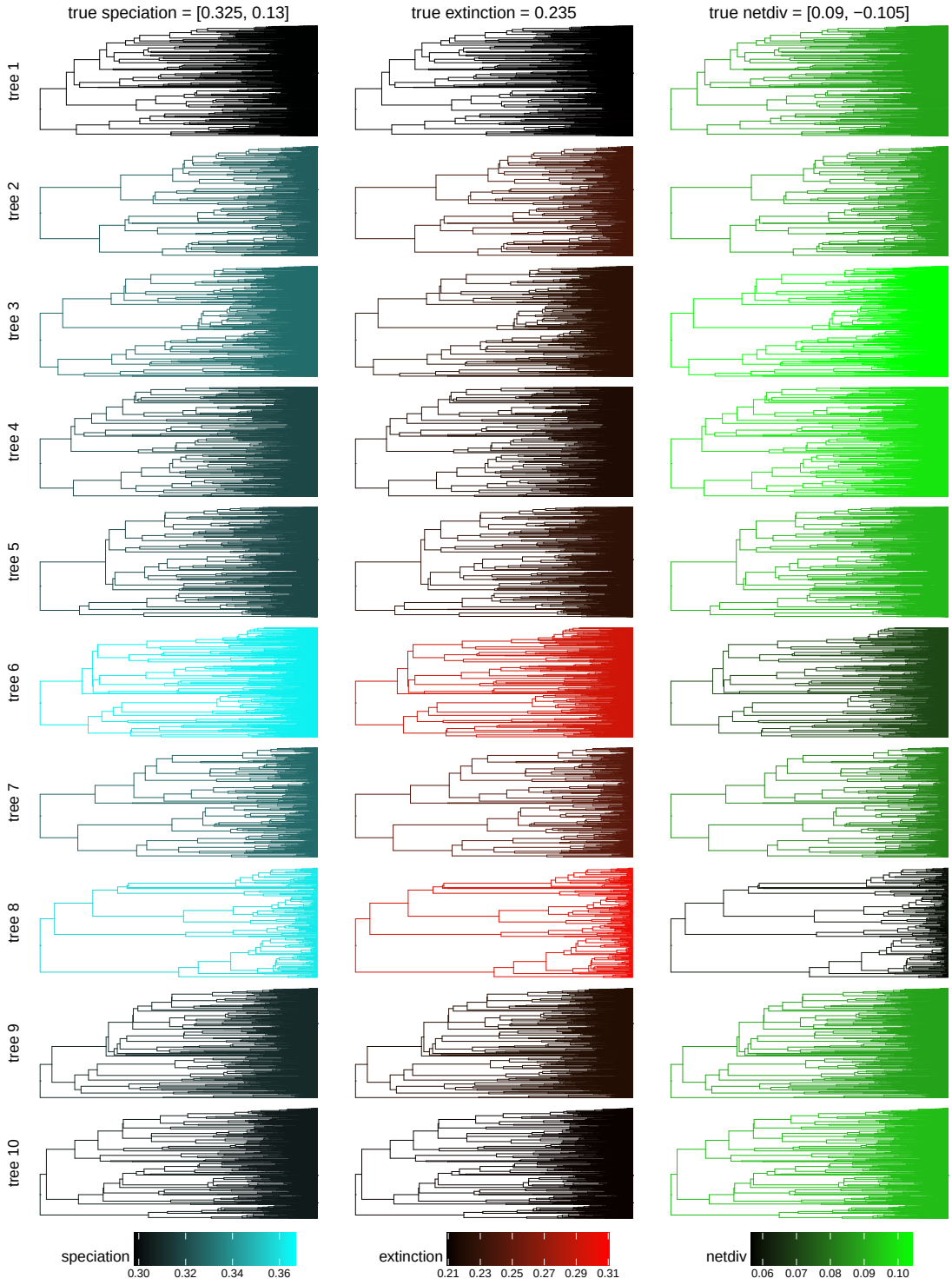


Figure S5.7: Ten trees where we simulated a backbone tree under the constant-rate birth-death model ($\lambda = 0.325, \mu = 0.235$) for a time period of 60 Ma, and replaced a randomly selected subtree at 40 Ma with a tree that had a lower speciation rate ($\lambda = 0.13$). We inferred branch-specific speciation, extinction and net-diversification rates (shown in color). The rate shift event that leads to a reduction in net-diversification rate is virtually never recovered.

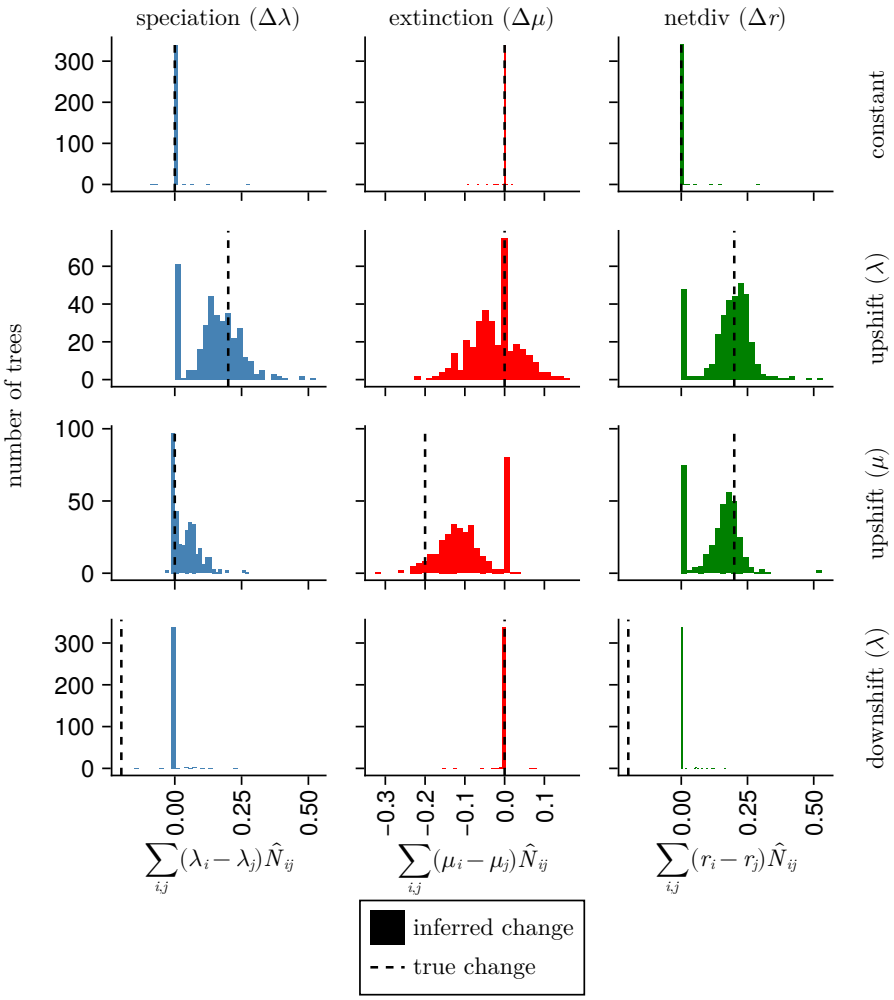


Figure S5.8: An assessment of the inferred change in branch-specific speciation, extinction and net-diversification rates, on simulated phylogenies with either no rate shift events (constant) or exactly one rate shift event (either upshift or downshift). We used the models in Table S5.2 to simulate various trees with various species richness. We simulated 350 backbone trees, which in combination with simulated in-group trees, we used to construct 350 upshift trees (due to an increase in λ), 350 upshift trees (due to a decrease in μ), and 350 downshift trees. Each data point in a histogram represents one tree. Each row represents the same set of trees.

Validating the number of rate shift inferences using simulated phylogenies

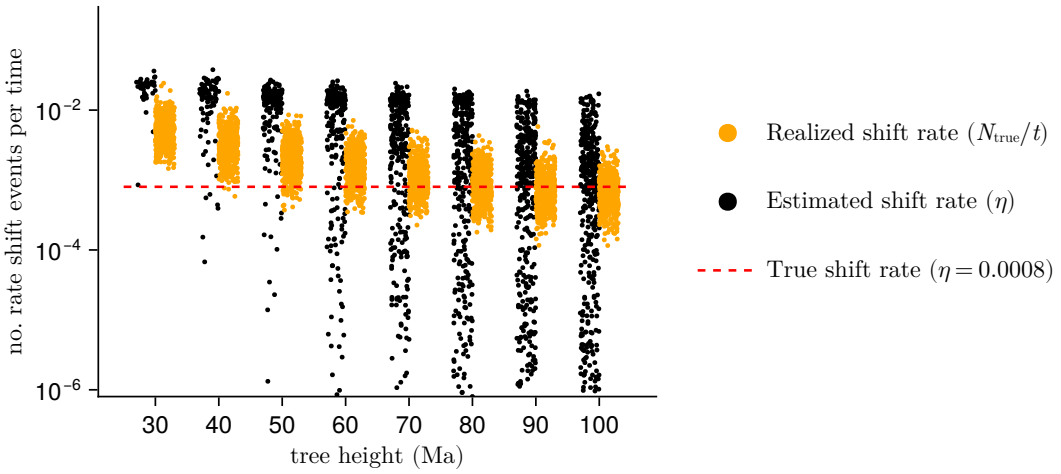


Figure S5.9: The shift rate (in units of rate shift events per time), estimated for several trees across a range of tree heights. Each dot represents one phylogeny. For each tree height in the range of 30 to 100 million years, we simulated 500 reconstructed trees and inferred the number of rate shifts. The position of the dots on the x-axis is jittered, i.e., moved randomly to the left or the right, for visibility purposes. The y-axis has been cut off at 10^{-6}, however there are many phylogenies with a smaller estimated shift rate (practically zero) that are not visible.

In the main text, we present analyses of empirical phylogenies, where the results indicate that there is a strong age-scaling effect for the number of rate shifts per time. We were concerned that this age-scaling effect was a result of a methodological artefact. To test whether this was an artefact, we set up a simulation study under a birth-death-shift model, where we can evaluate the performance of the birth-death-shift inference in a known scenario. Specifically, we simulated phylogenies under the following three-state model

$$\begin{aligned} \boldsymbol{\lambda} &= [0.12, 0.21, 0.30] \\ \boldsymbol{\mu} &= [0.08, 0.14, 0.20] \\ \eta &= 0.0008, \end{aligned} \tag{S5.9}$$

meaning that the relative extinction rate (μ/λ) was $2/3$, and the shift rate was selected such that we would get far fewer rate shift events than branching events. All simulations started in the category with the lowest net-diversification rate. We carefully hand-picked these rates in order to be able to simulate trees for a range of tree heights, without running into computational problems. Since extremely large trees are impractical, we set a maximum number of tips ($50,000$), and we rejected simulated trees that exceeded this threshold. In our experience, the hand-picked rates above generated trees with reasonable size for a range of tree heights (in the range of 30 to 100 million years, in increments of 10 myr), allowing us to test the inference of `Pesto` for a large number of tree replicates (500 per tree height),

while keeping the simulation rejection rate at a reasonable level. This yielded 4000 simulated reconstructed trees in total, for which we also knew the true shift history along each branch.

Next, we inferred branch-specific diversification and branch-specific diversification rate events using `Pesto`. Specifically, specifically the following model setup

$$\begin{aligned}
&(\hat{\lambda}, \hat{\mu}, \eta) && \text{estimated by ML under the birth-death-shift model} \\
&\vec{\lambda} && \text{six quantiles from LogNormal}(\log(\hat{\lambda}), \text{sd} = 0.587) \\
&\vec{\mu} && \text{six quantiles from LogNormal}(\log(\hat{\mu}), \text{sd} = 0.587) \\
&(\boldsymbol{\lambda}, \boldsymbol{\mu}) && \text{all pairwise combinations of } \vec{\lambda}, \vec{\mu}.
\end{aligned} \tag{S5.10}$$

In other words, there are 36 rate categories (represented by $\boldsymbol{\lambda}, \boldsymbol{\mu}$). In Fig. S5.9, we see that the number of realized shifts per time (N_{true}/t) is higher than what is expected from the true shift rate (η). This is due to sampling biases, both since we i) conditioned on that the phylogenies must have at least one rate shift event, and ii) we conditioned on survival of the left and right subtrees descending from the root. When assessing the robustness of the method, we therefore investigated the estimation error as the difference between the estimates (η or N^*) and the realized number of rate shift events (N_{true}, based on the specific simulated rate shift histories).

In Fig. S5.10, we see that there is some estimation error for the shift rate (a) and the number of strongly inferred rate shifts events per time (b). Specifically, the shift rate is overestimated for young phylogenies, but is almost unbiased for older phylogenies. The number of strongly supported shift events per time (N^*/t) is underestimated for young phylogenies, however. This metric is also underestimated for older phylogenies, which we expect is due to that we are not able to infer downshifts reliably. In Fig. S5.11 we see the same data except plotted as a function of the number of tips. The patterns are broadly similar, in that the shift rate is overestimated, while the number of strongly supported rate shift events is underestimated, when considering phylogenies with few taxa.

The results tell us that the age scaling effect in the shift rate (η) could be partially explained by estimation error in young phylogenies. However, the age scaling effect in the number of strongly inferred rate shift events cannot be explained by estimation error in young phylogenies, as the direction of the bias is opposite to the empirical age scaling pattern. Thus, we do not believe that estimation error on its own can explain the age scaling pattern in the empirical analyses.

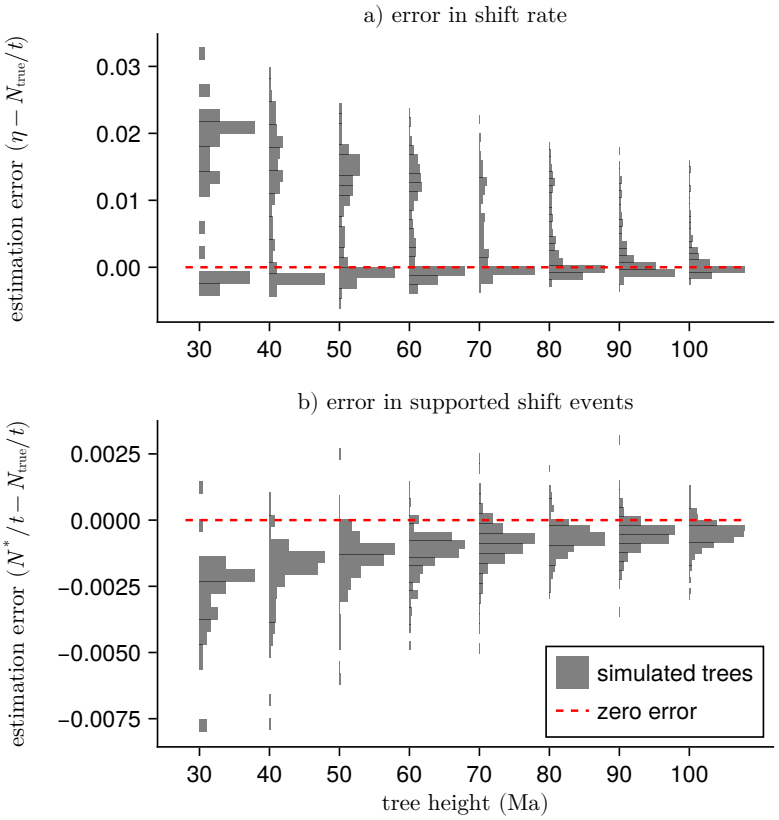


Figure S5.10: Estimation error in the number of rate shifts, as a function of tree height. The estimation error is the difference between the estimated value and the true value realized in the simulated trees. We used either the shift rate parameter (a) or the number of strongly supported branches per time (N^*/t) as the metric for the number of rate shift events. There are fewer trees for short tree heights (about 50 trees for 30 Ma), and more trees for longer tree heights (about 450 trees for 100 Ma). The shift rate η is overestimated for some young phylogenies, and it is approximately unbiased for old phylogenies.

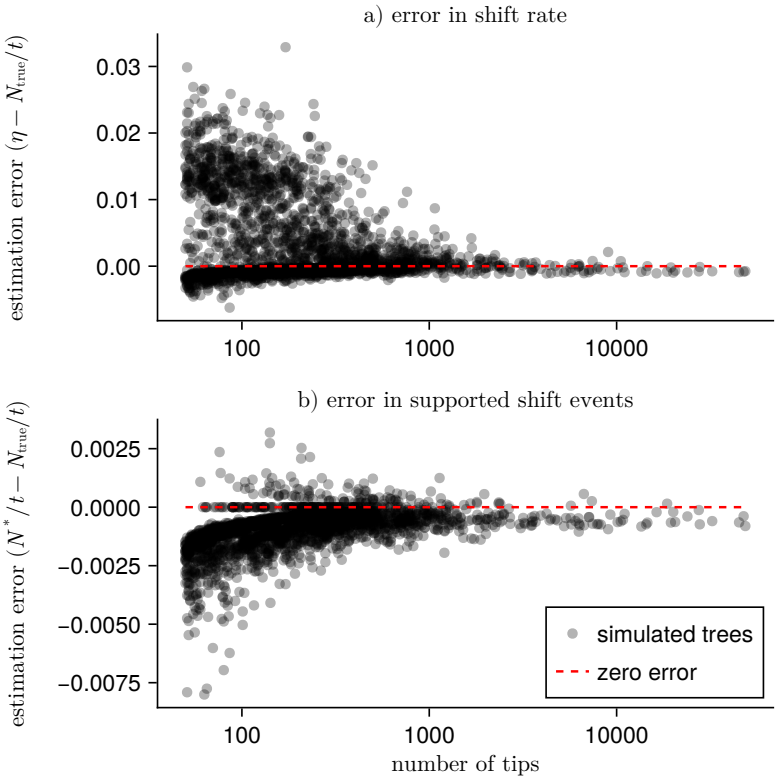


Figure S5.11: The same data as in Fig. S5.10 but as a function of the number of tips in the phylogeny. The shift rate η is overestimated for phylogenies with few tips, and the number of supported shifts per time (N^*/t) is underestimated for phylogenies with few tips.

Within-phylogeny estimates of diversification shift rate

In the main text, we saw that there was an apparent age scaling effect of shift rate (η) with the clade age of the phylogeny. At first sight, it appears as if the diversification rate shifts more often for younger phylogenies than older phylogenies. We argued that this is a time averaging effect, which is caused by lack of information in the oldest parts of the phylogeny with which to infer shifts in the diversification rate. In the main text, we only compared the number of diversification rate shifts among phylogenies. In this section, we calculate and present posterior estimates of the number of diversification rate within phylogenies. If there is less evidence for diversification rate shifts in more ancient parts of the phylogenies, then we expect that the estimate for the number of rate shifts should be smaller in the past than at the present.

In order to calculate the number of rate shifts per time, we used our expression for the derivative of $\hat{N}$ with respect to time, i.e., $\frac{d\hat{N}_M}{dt(t)}(t)$ for a particular branch with index M. This quantity represents the branch-specific estimate for the instantaneous shift rate, i.e., the number of diversification rate shifts per time, specifically on branch M and at time t. Suppose that on a particular time t, say at the root of the phylogeny, t_{MRCA}, there are two active lineages in a set A. After the first branching event has occurred, there are three active lineages in the set B, and so on. In order to get a sense of the change across time, we discretized time from the root age of the phylogeny until the present. Next, we calculated the geometric mean of the shift rate across the number of active lineages like so

$$\text{mean shift rate} = \exp\left(\frac{1}{\sum_A 1} \sum_{M \in A} \log\left(\frac{d\hat{N}_M}{dt}(t)\right)\right), \tag{S5.11}$$

for each time point t with its corresponding set of active lineages A. In other words, the mean posterior shift rate is an average of the instantaneous shift rate for all branches that were alive at a particular time in the past (and that survived until the present).

The results of this procedure is visualized in Fig. S5.12. In panel b) we show the mean shift rate through time for simulated phylogenies, whereas in c) we show the same for empirical phylogenies. Both panels b) and c) are a subsample of a greater number of phylogenies. When assessing each phylogeny individually, it appears at first as if the shift rate through time can change erratically, and it is different from phylogeny to phylogeny. Some shift histories may seem to match with prior knowledge in the group, for example in mammals the shifts were most prevalent briefly before the Cretaceous-Paleogene boundary, and in rosids there is a steady amount of rate shifts until very close to the present. Without specific knowledge of the particular group, however, it may be difficult to interpret what the rate shift history means.

One common pattern we noticed, was that the estimates of the rate shifts tended to increase through time. In order to summarize this, we computed whether the mean shift rate was higher at the present (i.t. $t = 0$) than at the root (i.e., $t = t_{\text{MRCA}}$). This is shown in

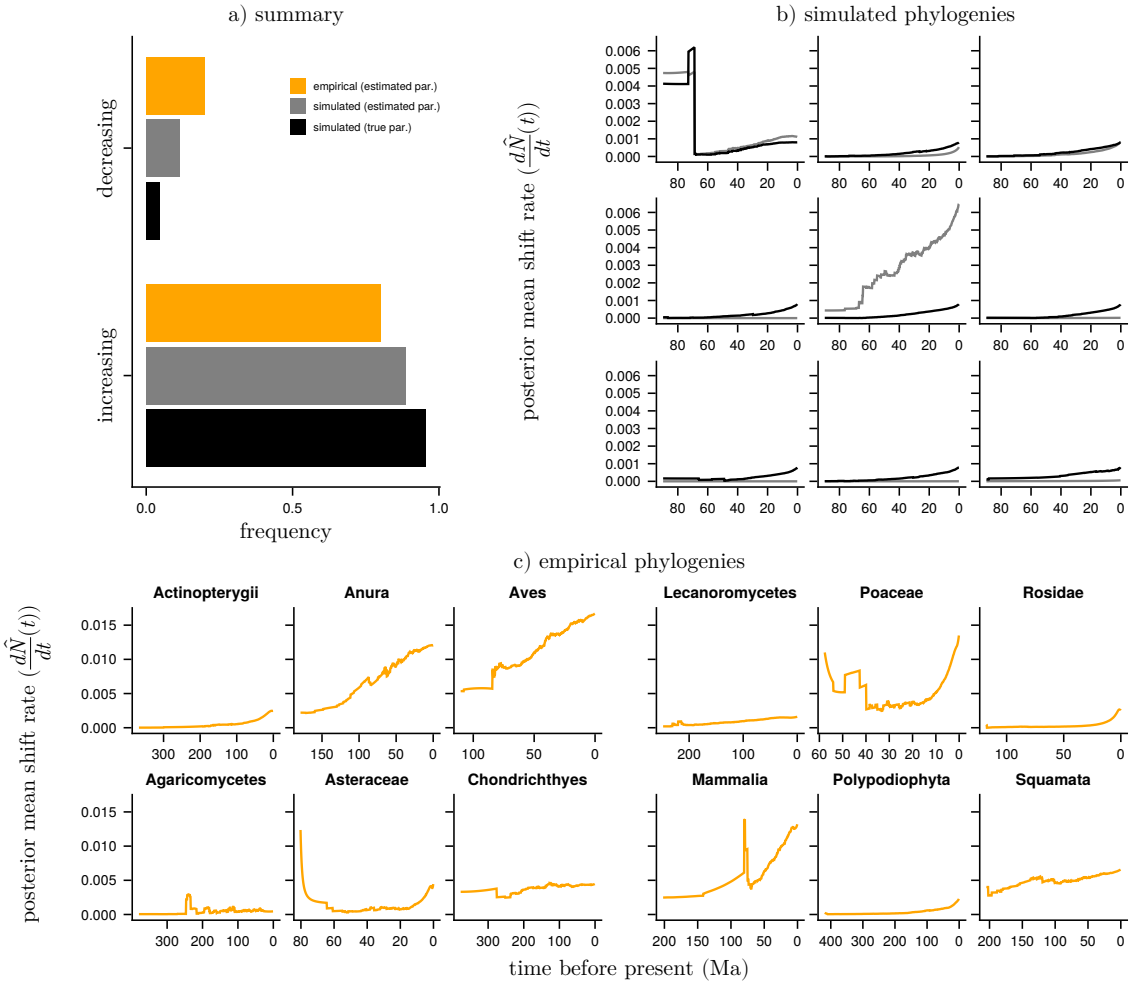


Figure S5.12: The posterior estimate for the shift rate, averaged (geometric mean) across the number of active lineages per time slice. a) represents a summary of whether the tempo of rate shifts overall increased or increased. If the most recent shift rate ($d\hat{N}/dt(t=0)$) was greater than the shift rate at the time of the most-recent common ancestor (i.e., t_{MRCA}), then we counted the phylogeny as increasing. If not, we regarded the model as decreasing. b) represents 9 examples from simulated trees, with true parameters $\boldsymbol{r} = [0.04, 0.07, 0.10]$, $\epsilon = 2/3$ and $\eta = 0.0008$. The (in total 500) trees were simulated for a period of 90 Ma, see the previous section for more details on the simulation setup. Panel c) shows 12 of the empirical phylogenies, specifically the ones featured in Fig. 2. Note that the posterior estimate for the shift rate at the present (i.e., $dN/dt(t=0)$) is equal to the estimate for the shift rate parameter η. In about 80–96% of the phylogenies, the shift rate was overall increasing through time.

Fig. S5.12a, where we recorded between 80–96% phylogenies as having a shift rate estimate that was increasing through time within the phylogeny. This includes simulated phylogenies where the true shift rate is known, and is constant throughout all lineages and across time. For almost all (96%) of the simulated phylogenies we inferred a pattern of increasing shift rate through time.

In part this may seem reasonable, as if one were to consider a realized birth-death-shift history, then at the root of the tree there has not elapsed enough time for any rate shift event

to occur, and therefore the estimate for the shift rate should be small. Interestingly, the mean shift rate at the present (i.e., $t = 0$) matches almost perfectly with the true shift rate when we set the parameter to the true value ($\eta = 0.0008$). However, the posterior estimate for the instantaneous shift rate tends to be smaller in older parts of the phylogeny. We argue that the branch-specific estimates of the shift-rate corroborates the pattern for the among-phylogeny estimates.

Assessing the age-scaling effect by analyzing subtrees of the ray-finned fish tree

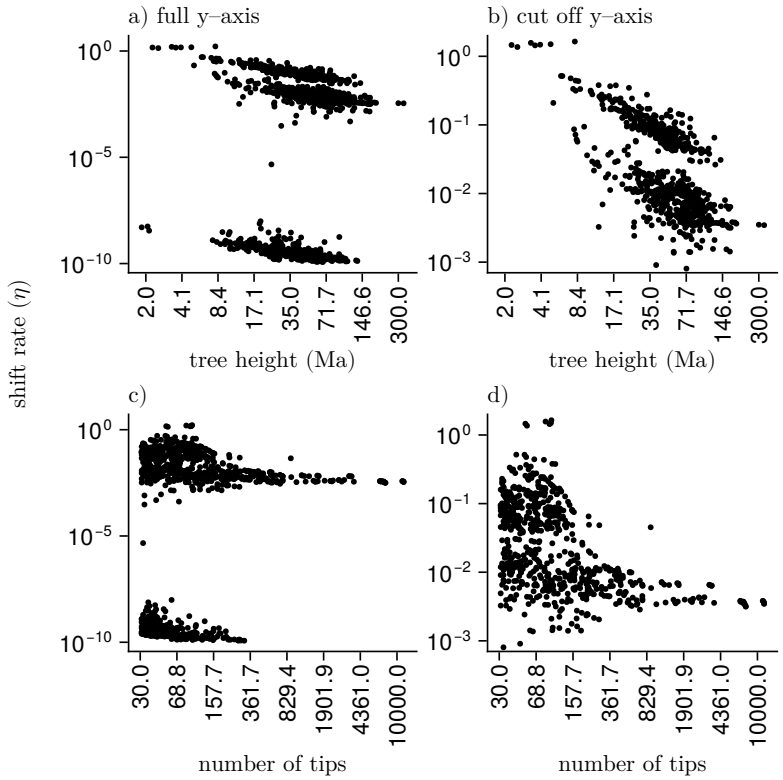


Figure S5.13: Each point represents a subtree in the ray-finned fish phylogeny (Rabosky et al., 2018). We included all subtrees that were represented by at least 30 species. Panels a) and b) depict the same scatter points, however in b) the y-axis has been truncated to remove the bottom portion of the points. We cut off the y-axis likewise in the bottom row, panels c) and d). Disregarding the clades for which we estimated none or negligibly small diversification rate variation (i.e., $\eta < 10^{-5}$), there is a negative scaling effect, in that young and species-rich clades exhibit far more diversification rate shifts per time than older and more established clades.

We also investigated the number of rate shift changes and its time-scaling effect on an empirical tree, the ray-finned fish phylogeny (Rabosky et al., 2018). To do so, we found all subtrees for the fish phylogeny that had at least 30 taxa represented at the present. This resulted in a total of 1035 subtrees. They ranged in tree height from approximately 2 million years (a cichlid subclade) to 330 million years, representing one of the earliest diverging lineages. For 310 subtrees, we estimated that there was none or negligibly small diversification rate variation (i.e., a shift rate of $\eta < 10^{-5}$). These trees were all represented by relatively few taxa (< 360 no. taxa, Fig. S5.13).

The fish phylogeny is relatively large, and most splits starting from the root have many more than 30 living descendants. This means that many of the subtrees in Fig. S5.13 are also represented by their immediate left and right descendant clades. In other words, the dots in Fig. S5.13 are pseudoreplicates and are strongly correlated in a phylogenetic manner.

Young and species-rich clades like cichlids and icefishes appear to exhibit far more diversification rate shifts than older and more established clades. In particular, one clade showed an extreme amount of diversification rate shifts with $\eta \approx 20$. There are also a few other clades with a similar estimate for the number of rate shifts per time (Fig. S5.13, panel b). We argue that these extreme estimates are not robust, and arise due to how the phylogeny was reconstructed. Several internal branches in these clades have lengths of near zero, as if a polytomy resolving technique had been used, and do not appear to be distributed across a range of branch lengths as would be expected under a birth-death (or birth-death-shift) process. Thus, while cichlids, icefishes and other highly-diversifying clades certainly exhibit significant diversification rate variation, we consider the specific estimates for branch-specific diversification rates, and diversification rate shifts in these clades to be unreliable.

Nevertheless, the big picture from Fig. S5.13 points to a strong negative relationship between the diversification shift rate (η) and the tree height, even within a single empirical phylogeny. Young and species-rich clades appear to have undergone far more diversification rate shifts per time than older and more established clades.

Technical details on estimating the parameters

In the birth-death-shift model, there are three parameters that govern how the branching process behaves. These are $\hat{\lambda}, \hat{\mu}$, which control the mean of the speciation and extinction rate distributions, and the shift rate η which controls how often diversification rate shifts occur (Table S5.3).

$\hat{\lambda}$	the parameter controlling the mean of the speciation rate distribution
$\hat{\mu}$	the parameter controlling the mean of the extinction rate distribution
η	the parameter controlling how often diversification rate shifts occur
$\vec{\lambda}$	n quantiles from LogNormal(log($\hat{\lambda}$), sd=0.587)
$\vec{\mu}$	n quantiles from LogNormal(log($\hat{\mu}$), sd=0.587)
($\boldsymbol{\lambda}$,$\boldsymbol{\mu}$)	all pairwise combinations of $\vec{\lambda}, \vec{\mu}$.

Table S5.3: Summary of parameters in the birth-death-shift model. The number of rate quantiles used is n, and the number of rate categories in ($\boldsymbol{\lambda}$,$\boldsymbol{\mu}$) is $K = n^2$. The standard deviation of the log-normal distributions is set to 0.587 unless otherwise specified, which results in a log-normal distribution whose 2.5%–97.5% quantile spans one order of magnitude. The rate categories ($\boldsymbol{\lambda}$,$\boldsymbol{\mu}$) are entirely determined by $\hat{\lambda}, \hat{\mu}$ and n.

In a simulation study, the parameters ($\hat{\lambda}, \hat{\mu}$,η) are known without error. For empirical phylogenies, however, these are not known and must be estimated from the phylogeny. We chose to estimate the parameters by finding the parameter values that maximize the likelihood of the parameters given the phylogeny. We explored several ways of finding the maximum likelihood estimates using numerical optimization algorithms. In doing so, we discovered several scenarios in which the estimation procedure does not work well, or produced incoherent results. In a short list, these include:

(a) If the extinction rate ($\hat{\mu}$) is greater than the speciation rate ($\hat{\lambda}$)

(b) If the shift rate (η) is too high, for example greater than the speciation or extinction rate ($\hat{\lambda}$ or $\hat{\mu}$)

(c) If any of the parameters are too large

(d) If any of the parameters are too small

(e) If there are too few species in the phylogeny (e.g. < 25)

For this reason, we decided to impose several constraints on the optimization procedure. First, we imposed the constraint that $5\eta < \hat{\lambda} > \hat{\mu}$, by introducing three dummy variables, x_1, x_2 and x_3. We defined $\eta = x_1$, $\mu = x_2$, and $\lambda = \text{maximum}(5x_1, x_2) + x_3$, and optimizing over those transformed variables instead. This has an effect of imposing that the overall net-diversification is positive (remedying problem a), and that the shift rate is relatively small (remedying problem b). In the final rate categories $\boldsymbol{\lambda}, \boldsymbol{\mu}$, however, the net-diversification is allowed to be negative.

For the optimization algorithm, we decided to use Newton's method, since it is known to have fast (quadratic) convergence. Newton's method is a stepwise updating approach, where the next step $\vec{y}_{i+1}$ is updated as

$$\vec{y}_{i+1} = \vec{y}_i - kH(\vec{y}_i)^{-1}\nabla f(\vec{y}_i), \tag{S5.12}$$

where $\vec{y}_i$ is the current position, and k is some step size coefficient. $\nabla f(\vec{y}_i)$ is the gradient of the target function (i.e., the first derivatives of the negative log likelihood) evaluated at $\vec{y}_i$, and $H(\vec{y}_i)$ is the Hessian matrix (i.e., the second derivatives of the negative log likelihood) evaluated at $\vec{y}_i$. $\vec{y}_0$ is some starting point. We used automatic differentiation to compute the derivatives (Revels et al., 2016), and we used the implementation of (Mogensen and Riseth, 2018) for Newton's method. Newton's method is more computationally expensive than gradient-free methods or gradient descent per iteration, as it requires one to compute and invert the Hessian matrix. The extra computational cost is in practice not a problem, as it is outweighed by the fast convergence of Newton's method, and we typically need few (< 50) iterations before convergence is reached. One challenge with Newton's method is, however, that it only works for a problem where the parameters are unconstrained, i.e., $\vec{y}$ must be allowed to vary between $-\infty$ and $+\infty$. As a fix, we re-transformed our parameters x_1, x_2, x_3 using a logistic function with soft boundaries:

$$g(y) = \frac{\mathrm{U} - \mathrm{L}}{1 + e^{-s(y - \mathrm{M}))}} + \mathrm{L}. \tag{S5.13}$$

Here, U is the upper limit, L is the lower limit, $M = (\mathrm{U} + \mathrm{L})/2$ is the midpoint, and $s = 1/2$ is a steepness coefficient. The inverse function is

$$h(x) = \mathrm{M} - \frac{1}{s}\log\left(\frac{\mathrm{U} - \mathrm{L}}{x - \mathrm{L}} - 1\right), \tag{S5.14}$$

and we used lower limits $\mathrm{L} \in \{10^{-8}, 10^{-4}, 10^{-4}\}$, upper limits $\mathrm{U} \in \{0.3, 1.0, 1.0\}$ for x_1, x_2, x_3, respectively. Since this transform effectively maps our constrained parameters to an unconstrained space, we can use Newton's method to optimize for the maximum likelihood parameters. By imposing lower and upper limits for our parameters, we remedy problems c) and d).

The choice of starting point $\vec{y}_0$ also has an impact on the optimization procedure. We observed that, if different starting points $\vec{y}_0$ are chosen, the optimization routine will find different (local) maximum likelihood regions. In other words, if the procedure is run a few times, it may reach two or more different solutions. Although the log-likelihood values may not be too different, the inferred position, size and number of rate shift events on the phylogeny may vary. For example, one may get a solution A which represents one large rate shift on one branch, and a solution B which represents several smaller diversification rate shifts on several branches in different parts of the phylogeny. When running the analyses on an empirical

phylogeny, we therefore strongly recommend to repeat the inference procedure several times, and to pick the solution that maximizes the (global) likelihood.

In the inference procedure, we opted for using a random starting position $\vec{x}$. First, we estimated the net-diversification rate ($r_{\text{CBD}} > 0$) and the extinction rate ($\mu_{\text{CBD}} > 0$) under the constant-rate (i.e., lineage-homogeneous) birth-death model (CBD). With these estimates, we set up three distributions:

$$\begin{aligned} d_1 &\sim \text{LogNormal}(\log(0.01), \text{sd} = 0.5) \\ d_2 &\sim \text{LogNormal}(\log(\mu_{\text{CBD}}), \text{sd} = 0.5) \\ d_3 &\sim \text{LogNormal}(\log(r_{\text{CBD}}), \text{sd} = 0.5). \end{aligned} \tag{S5.15}$$

We drew random starting points x_1, x_2, x_3 from the distributions d_1, d_2, d_3, which we transformed to the parameters y_1, y_2, y_3 using Eq. S5.14.

For problem e), we simply decided not to assess the birth-death-shift model for phylogenies with fewer than 50 taxa. This is not a big problem, as inferences made from such phylogenies will anyways be unreliable, and it is perhaps better to make the simplifying assumption that there is no among-lineage rate variation.

Appendix E

Supplementary Material for Chapter 6

S6.1 Empirical datatsets

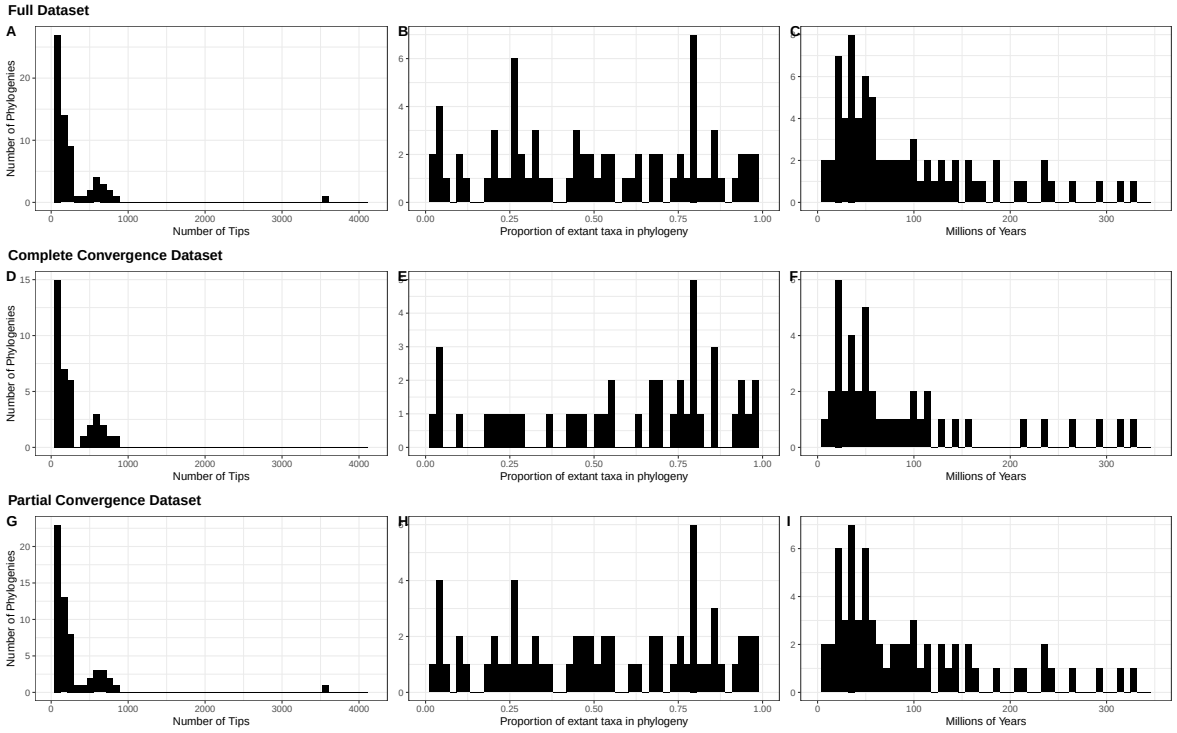


Figure S6.1: Summary statistics on the chronograms used in analyses from (Henao Diaz et al., 2019). A,D,G) Distribution of the number of tips per phylogeny. B,E,H) Distribution of incomplete-sampling percentage per phylogeny. C,F,I) Distribution of the age of the most recent comment ancestor. A-C) Summary statistics for the full dataset in this study. D-F) Summary statistics for the 'complete converged dataset' (only includes trees that reached convergence for all methods, n=43). G-I) Summary statistics for the 'Complete convergence dataset' (only includes tree that reached convergence for all methods except LSBDS, n = 65)

S6.2 Convergence

Table S6.1: The chronograms that form part of the complete and partial convergence data subsets. The complete convergence subset includes trees that converged across all methods: LSBDS, BAMM, ClaDS2, and MTBD. The partial convergence subset includes trees that converged across all methods *but* LSBDS. The abbreviations are borrowed from the Henao Diaz et al. (2019) paper.

Abbrev.	Clade	Complete convergence	Partial convergence	Citation
A2016	Viperidae	TRUE	TRUE	Alencar et al. (2016)
A2017	Costaceae	TRUE	TRUE	André et al. (2016)
B2016	Onthophagus	TRUE	TRUE	Breeschoten et al. (2016)
BR2017	Actinopterygii	FALSE	FALSE	Betancur-R et al. (2017)
C2015	Ovalentaria	TRUE	TRUE	Campanella et al. (2015)
C2017	Coronellini	TRUE	TRUE	Chen et al. (2017)
D2018	Dytiscidae	FALSE	TRUE	Désamoré et al. (2018)
F2017	Cichlidae	TRUE	TRUE	Missing in Henao Diaz et al. (2019)
G2011	Ceanothus	TRUE	TRUE	Goldberg et al. (2011)
G2017	Scolytinae	FALSE	FALSE	Gohli et al. (2017)
H2007	Pinnipedia	TRUE	TRUE	Higdon et al. (2007)
H2016	Cracidae	TRUE	TRUE	Hosner et al. (2016)
H2017	Ctenitis	TRUE	TRUE	Hennequin et al. (2017)
HI2017	Quercus	TRUE	TRUE	Hipp et al. (2018)
I2011	Sebastes	TRUE	TRUE	Ingram (2011)
I2016	Anolis	FALSE	TRUE	Ingram et al. (2016)
IL2017	Heliconia	TRUE	TRUE	Iles et al. (2017)
JO2016	Corvides	TRUE	TRUE	Jønsson et al. (2016)
L2012	Coniferophyta; Pinidae	FALSE	FALSE	Leslie et al. (2012)
L2013	Rhododendron section Vireya	TRUE	TRUE	Neupane et al. (2017)
L2014	Agama	TRUE	TRUE	Leaché et al. (2014)
L2016	Odonata	FALSE	TRUE	Letsch et al. (2016)
L2017	Lobelioidae	TRUE	TRUE	Lagomarsino et al. (2017)
MC2016a	Balistidaea	TRUE	TRUE	McCord and Westneat (2016)
MC2016b	Monacanthidae	FALSE	TRUE	McCord and Westneat (2016)
N2017a	Spermacoceae	TRUE	TRUE	Neupane et al. (2017)
P2016	Cephalotes	TRUE	TRUE	Price et al. (2016)
P2017	Testudinata	FALSE	FALSE	Pereira et al. (2017)
PB2014	Squamata	FALSE	FALSE	Pyron and Burbrink (2014)
pg1646	Pinnipedia	FALSE	TRUE	Higdon et al. (2007)
pg1953	Furnariidae	TRUE	TRUE	Derryberry et al. (2011)
pg2575	Passeriformes	TRUE	TRUE	Barker et al. (2013)
pg2576	Actinopterygii	FALSE	FALSE	Betancur-R et al. (2017)
pg2659	Otophysi	TRUE	TRUE	Chen et al. (2013)
pg2689	Lupinus	TRUE	TRUE	Drummond et al. (2012b)
pg2850	Columbidae	TRUE	TRUE	Cibois et al. (2014)
pg2853	Trochilidae	TRUE	TRUE	Désamoré et al. (2018)
PRE2017	Temnothrorax	TRUE	TRUE	Prebus (2017)
R2013	Mormoopidae & Phyllostomidae	TRUE	TRUE	Rojas et al. (2013)
R2018	Phymaturus	FALSE	TRUE	Reaney et al. (2018)

Continued on next page

Table S6.1 – continued from previous page

Abbrev.	Clade	Complete convergence	Partial convergence	Citation
S2012	Carnivora	FALSE	TRUE	Slater et al. (2012)
S2014	Poaceae	FALSE	FALSE	Spriggs et al. (2014)
S2015	Galliformes	TRUE	TRUE	Stein et al. (2015)
S2018.01	Elasmobranchii	FALSE	FALSE	Stein et al. (2018)
SA2016	Columbiformes	TRUE	TRUE	Soares et al. (2016)
SamplePrimates	Primates	TRUE	TRUE	Missing in Henao Diaz et al. (2019)
SampleWhale	Ballenidae	TRUE	TRUE	Steeman et al. (2009)
SH2015	Chiroptera	FALSE	TRUE	Shi and Rabosky (2015)
SH2016	Pectinidae	FALSE	TRUE	Sherratt et al. (2016)
SO2017	Aedes	TRUE	TRUE	Soghigian et al. (2017)
ST2017	Astacoidea & Parastacoidea	FALSE	TRUE	Stern et al. (2017)
SW2014	Pergidae	TRUE	TRUE	Schmidt and Walter (2014)
T2015	Myrtaceae	TRUE	TRUE	Thornhill et al. (2015)
T2018	Cetartiodactyla	TRUE	TRUE	Toljagić et al. (2018)
TE201602	Aspleniaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201603	Athyriaceae	TRUE	TRUE	Testo and Sundue (2016)
TE201604	Blechnaceae	TRUE	TRUE	Testo and Sundue (2016)
TE201607	Cyatheaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201608	Cystopteridaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201610	Dennstaedtiaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201614	Dryopteridaceae	FALSE	FALSE	Testo and Sundue (2016)
TE201616	Gleicheniaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201617	Hymenophyllaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201619	Lindsaeaceae	TRUE	TRUE	Testo and Sundue (2016)
TE201620	Lomariopsidaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201624	Marattiaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201625	Marsileaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201631	Ophioglossaceae	FALSE	TRUE	Testo and Sundue (2016)
TE201634	Polypodiaceae	FALSE	FALSE	Testo and Sundue (2016)
TE201636	Pteridaceae	TRUE	TRUE	Testo and Sundue (2016)
TE201641	Tectariaceae	TRUE	TRUE	Testo and Sundue (2016)
TE201642	Thelypteridaceae	FALSE	TRUE	Testo and Sundue (2016)
UC2015	Rhinantheae	FALSE	TRUE	Uribe-Convers and Tank (2015)
VA2017b	Myrteae	FALSE	FALSE	Vasconcelos et al. (2017)
W2013	Vitis	TRUE	TRUE	Wan et al. (2013)
WS2017.01	Anisoptera	TRUE	TRUE	Waller and Svensson (2017)
WS2017.02	Zygoptera	FALSE	FALSE	Waller and Svensson (2017)

S6.3 Methods comparison

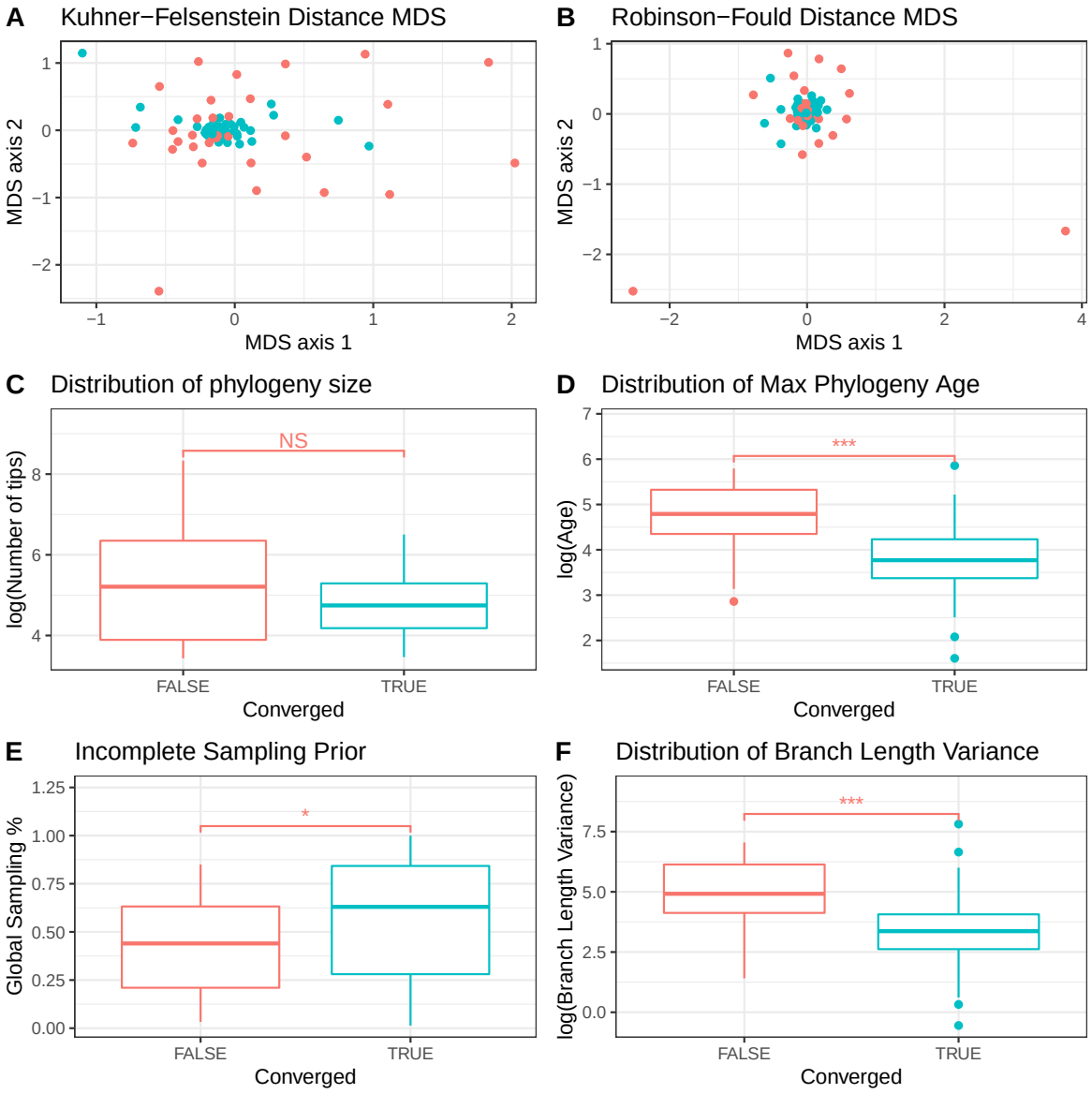


Figure S6.2: Comparison of convergence. Phylogeneies where converged was reached across all five methods n = 43 (Blue); Phylogeneies where converge was not reached for one or more methods n=33 (Red). A-B) Multidimensional scaling of phylogeny by, A) Kuhner-Felsenestin distance matrix and by, B) Robinson-Fould distance matrix. C) Size of the phylogeny in terms of number of tips. D) Age of the oldest nodes in the phylogeny. E) Percent incomplete sampling incorporated into diversification analysis. F) Variance of branch lengths of phylogenies. C-F) T-test performed between converged and unconverged phylogenies. P-value of T-test displayed (*: 0.05>P-value>0.01; **: 0.01>P-value>0.001; ***: 0.001>P-value).

Table S6.2: Post-hoc pairwise comparisons of variance summary statistics

Summary statistic (rate)	Contrast	Ratio of Geometric Means	Standard error	t. ratio	Tukey adjusted p-value	Variance explained by Random Effect (%)	Signif.
Speciation	BAMM / ClaDS2	7.00E-07	4.00E-04	-11.8382	0	60.12	***
	BAMM / PESTO	0.0033	0.0021	-9.2785	0		***
	BAMM / MTBD	0.4084	0.2509	-1.4577	0.465		N.S.
	ClaDS2 / PESTO	4.8196	2.9612	2.5597	0.054		N.S.
	ClaDS2 / MTBD	588.653	361.6703	10.3805	0		***
	PESTO / MTBD	122.1364	75.041	7.8208	2.13E-12	60.12	***
Extinction	BAMM / ClaDS2	0.0041	0.0041	-5.5293	6.32E-07		***
	BAMM / PESTO	15.2876	15.1724	2.7478	0.033		*
	BAMM / MTBD	1.2546	1.2451	0.2285	0.996		N.S.
	ClaDS2 / PESTO	3694.5857	3666.7289	8.277	1.17E-13		***
	ClaDS2 / MTBD	303.192	300.9059	5.7578	2.03E-07		***
	PESTO / MTBD	0.0821	0.0814	-2.5193	0.060		N.S.
Diversification	BAMM / ClaDS2	5.00E-04	3.00E-04	-11.6464	0	54.99	***
	BAMM / PESTO	0.0014	9.00E-04	-9.9236	0		***
	BAMM / MTBD	0.401	0.2649	-1.3832	0.511		N.S.
	ClaDS2 / PESTO	3.1214	2.0623	1.7228	0.315		N.S.
	ClaDS2 / MTBD	880.9356	582.0436	10.2632	0		***
	PESTO / MTBD	282.2281	186.4711	8.5404	7.44E-15		***

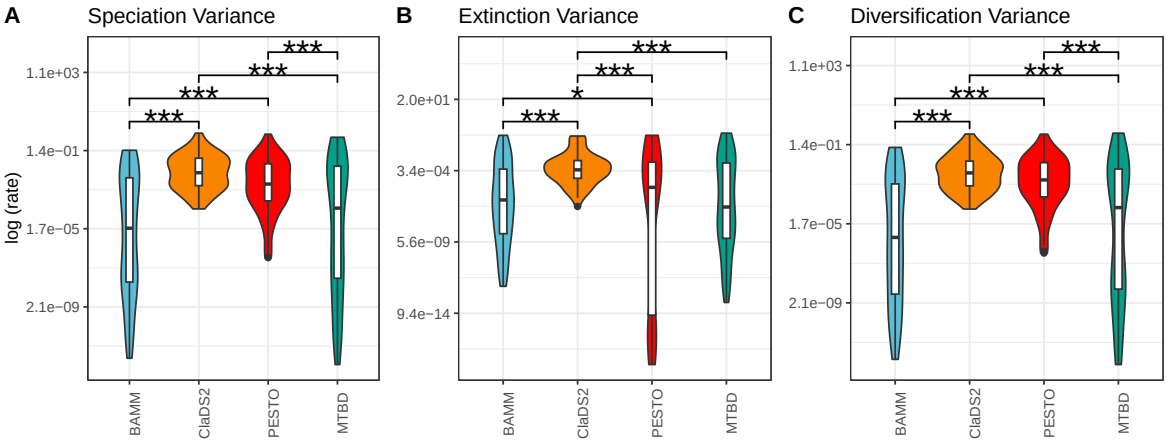


Figure S6.3: Comparison of summary statistics across the Partial Convergence dataset which excludes LSBDS. The distributions of averages is shown in Fig. 6.2A–C in the main text. (A) Variances of speciation rates, (B) Variance of extinction rate, (C) Variance of net diversification rate. P-value of linear mixed model displayed (*: 0.05>P-value>0.01; **: 0.01>P-value>0.001; ***: 0.001>P-value)

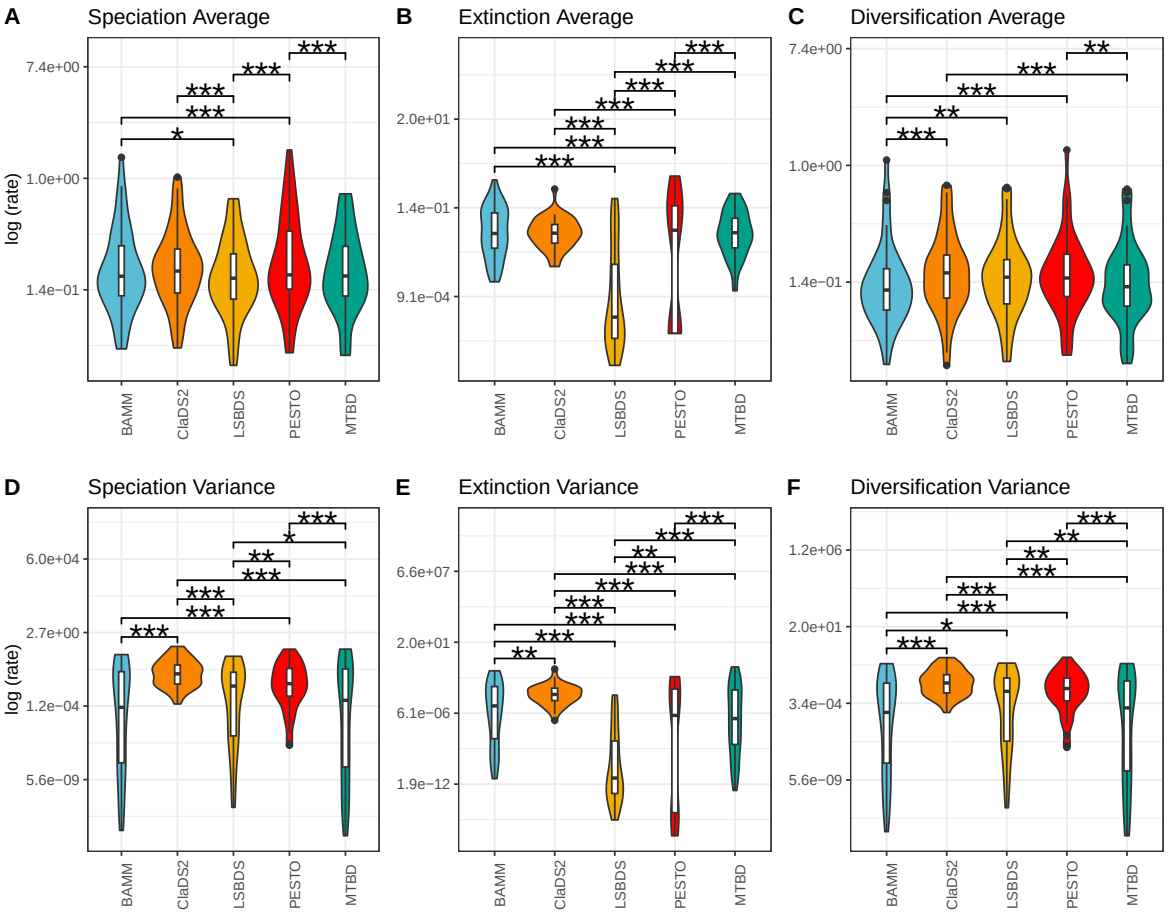


Figure S6.4: Comparison of summary statistics across methods for the complete convergence subset which includes all methods. (A) Average speciation rate, (B) Average extinction rate, (C) Average net diversification rate, (D) Variances of speciation rates, (E) Variance of extinction rate, (F) Variance of net diversification rate. P-value of linear mixed model displayed (*: 0.05>P-value>0.01; **: 0.01>P-value>0.001; ***: 0.001>P-value).

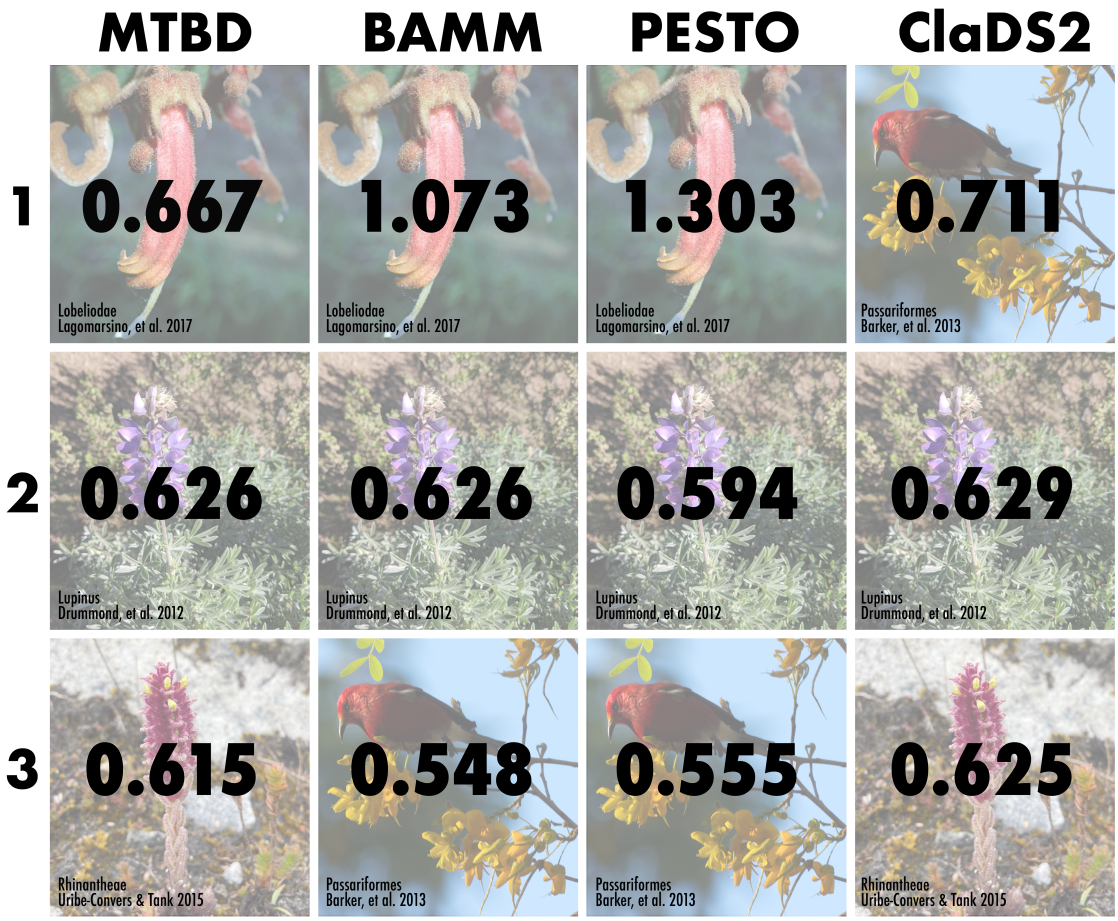


Figure S6.5: Matrix showing the three fastest lineages for each method in the partial convergence subset. Numbers over the photos indicate the estimated tree-wide mean net-diversification rate. All photos by CMT except for Rhinantheae, by Fabien Anthelme (CC).

S6.4 Quantifying uncertainty in rate estimates

S6.4.1 Uncertainty Methods

To quantify how methods differ in terms of the uncertainty of their branch rate estimates, we calculated the 95% HPD interval for all net diversification branch estimates for all trees. We used the complete convergence subset in order to include LSBDS, however our results hold for the partial subset as well. This analysis excludes PESTO, as PESTO directly infers the posterior mean and thus does not estimate a posterior distribution from which an HPD interval could be inferred. Using the HPD intervals, we calculated two metrics. First, we calculated the HPD interval overlap ratio (HPD IOR) between pairs of methods by taking the length of overlap between two HPD intervals and dividing it by the length of the union of the two HPD intervals, where the union is the total "distance" spanned by the two intervals. An HPD IOR of 1 is signifies complete overlap while 0 signifies no overlap. This measure quantifies the similarity of HPD intervals across all branches and methods. Second, we quantified the average precision of the uncertainty—in other words, the breadth of the 95% HPD interval for estimates from `RevBayes`, `CLaDS2`, `BAMM`, and `MTBD`. We used a linear mixed model to test for statistical differences between these interval breadths:

$$\log(95\%\text{ HPD interval breadth}) = \mathbf{X}\beta + \mathbf{Z}i + r, \tag{S6.1}$$

with inference method as a fixed-effect categorical predictor (effect sizes β), branch as a random effect categorical predictor (i), and an error term r, we tested if the least-squares means of each pair of methods were statistically different using Tukey's corrected p-value for multiple comparisons.

S6.4.2 Uncertainty results

When comparing the HPD IOR between methods we observed the same general trends that we observed in our MSE result (Fig. S6.6, Fig. 6.2B). `CLaDS2` differed the most, on average a lower overlap ration when compared to all other methods comparisons (S6.6. `BAMM-RevBayes` and `BAMM-MTBD` show high overlap, and `MTBD-RevBayes` shows higher overlap than any `CLaDS2` comparison but is lower than `BAMM-RevBayes` and `BAMM-MTBD`.

On average HPD IOR for all comparison was less than 20% (S6.6). This indicates that on average most of the methods are inferring different rates for a particular branch. However, for some branches—excluding `CLaDS2-RevBayes` and `CLaDS2-MTBD`—there was nearly complete overlap in HPD intervals (*i.e.*, HPD IOR near 1) suggesting that in these cases the methods infer nearly the same rate and uncertainty range. Branch-dependent agreement in method is generally consistent with our qualitative assessment of painted trees (Fig. 6.1 in the main text). There were certain trees where methods inferred the same values, but generally for

most phylogenies the methods inferred different rates and rate shifts.

When considering HPD interval length in net-diversification estimates, a post-hoc pairwise comparison of methods found that significant difference amongst all contrast except for `RevBayes` and `MTBD` (Table S6.3). Interestingly, `CLaDS2` has 95% HPD intervals are at least 1.97 standard deviation longer than all other methods (ClaDS2 vs. BAMM: d = 2.3240, SD = 0.00994; ClaDS2 vs. MTBD: d = 2.0002, SD = 0.00994; ClaDS2 vs.LSBDS: d = 1.9778, SD = 0.00994) This was surprising given that comparisons that included `CLaDS2` had the lowest overlap. Collectively, these corroborate our results based on summary statistics of method dependent difference. Moreover, our results highlight that often times methods are inferring partially overlapping distribution and therefore the entire posterior distribution should be examined when drawing conclusion.

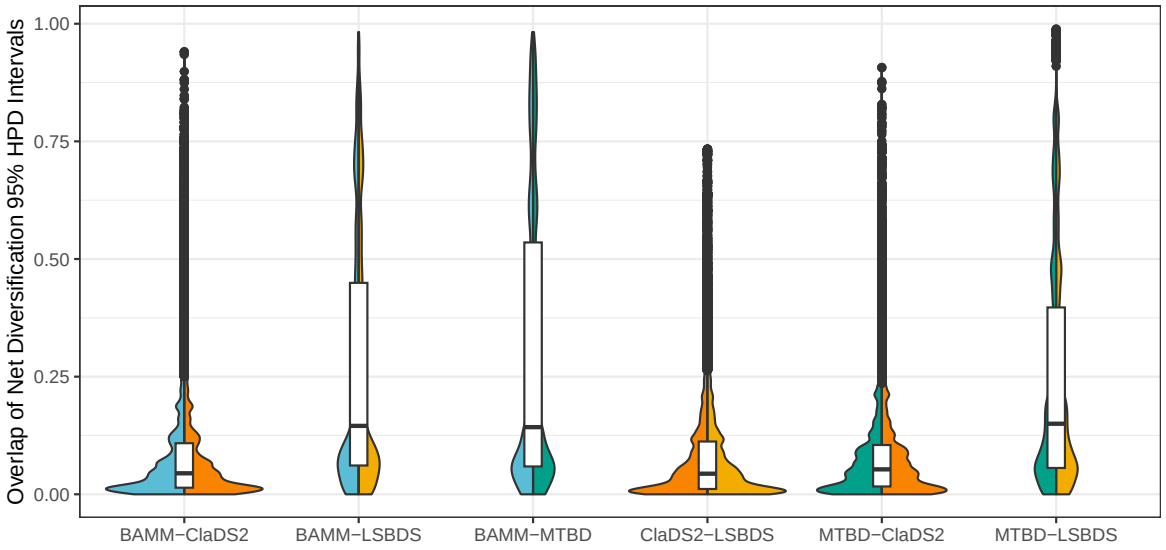


Figure S6.6: The extent to which the 95 % HPD intervals for net-diversification estimates overlap of the partial convergence subset. Violin plots correspond to pairs of methods (*eg.* BAMM and LSBDS).

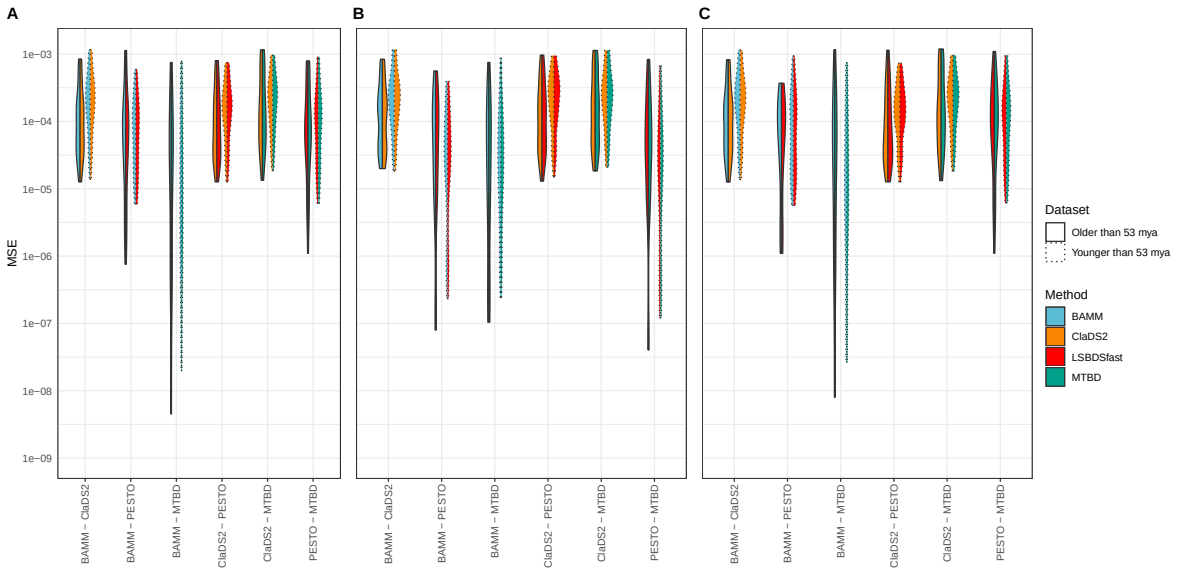


Figure S6.7: Pairwise mean squared error (MSE) between inference methods of phylogenies with branch lengths scaled by rates (A,D) speciation, (B,E) extinction, and (C,F) net diversification), plotted on a log scale. Split colors correspond to inference method color in FIG 2A–C. Distributions closer to zero indicate that the inference methods produced more similar rate estimates, whereas higher values indicate greater dissimilarity. (A-C) Data subsetted by median age of the phylogeny into older and younger phylogenies.

Table S6.3: Post-hoc pairwise comparisons of inference methods on HPD interval length. Peformed on the "complete convergence dataset" Columns contain the, contrasts of inference methods, the ratios of geometric means, standard errors, degrees of freedom, t-ratios, Tukey-adjusted p-values, significances.

Contrasts	Means Ratio	SE	DF	Z-Ratio	Adj. P-Value	Sig.
ClaDS2 / BAMM	4.431	0.02822	Inf	233.787	<.0001	***
ClaDS2 / MTBD	3.601	0.02293	Inf	201.219	<.0001	***
ClaDS2 / LSBDS	3.550	0.02260	Inf	198.960	<.0001	***
BAMM / MTBD	0.813	0.00518	Inf	-32.568	<.0001	***
BAMM / LSBDS	0.801	0.00510	Inf	-34 827	<.0001	***
MTBD / LSBDS	0.986	0.00628	Inf	-2.258	0.1080	***

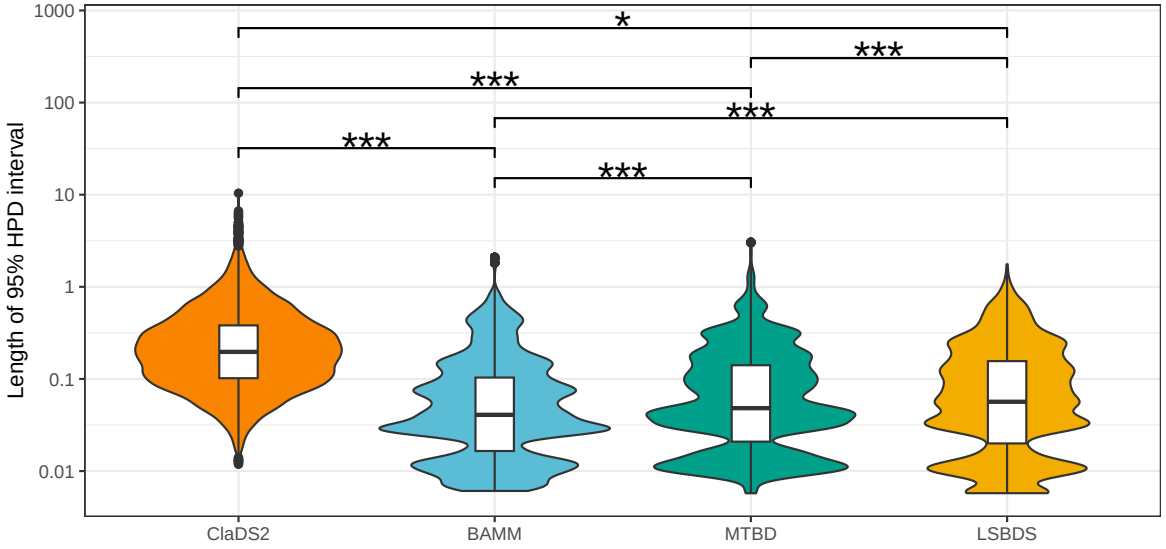


Figure S6.8: Comparison of HPD length of net-diversification across all branches of all trees in the partial convergence subset. P-value of linear mixed model with branch as random effect displayed (*: 0.05>P-value>0.01; **: 0.01>P-value>0.001; ***: 0.001>P-value).

S6.5 Results for individual chronograms

In the published version of the manuscript, we included a figure for each chronogram in the partial convergence subset. To avoid excessive page count, the chronograms are omitted in this thesis.

Bibliography

Abella, J., D. M. Alba, J. M. Robles, A. Valenciano, C. Rotgers, R. Carmona, P. Montoya, and J. Morales. 2012. *Kretzoiarctos gen. nov.*, the oldest member of the giant panda clade. PLoS One 7:e48985.

Ahlmann-Eltze, C. 2017. ggsignif: Significance Brackets for "ggplot2". R package version 0.4.0.

Alencar, L. R., T. B. Quental, F. G. Grazziotin, M. L. Alfaro, M. Martins, M. Venzon, and H. Zaher. 2016. Diversification in vipers: Phylogenetic relationships, time of divergence and shifts in speciation rates. Molecular Phylogenetics and Evolution 105:50–62.

Alexander, H. K., A. Lambert, and T. Stadler. 2016. Quantifying age-dependent extinction from species phylogenies. Systematic Biology 65:35–50.

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

Allen, A. P. and J. F. Gillooly. 2006. Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. Ecology letters 9:947–954.

Allio, R., B. Nabholz, S. Wanke, G. Chomicki, O. A. Pérez-Escobar, A. M. Cotton, A.-L. Clamens, G. J. Kergoat, F. A. Sperling, and F. L. Condamine. 2021. Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. Nature Communications 12:354.

Alroy, J., C. Marshall, R. Bambach, K. Bezusko, M. Foote, F. Fürsich, T. Hansen, S. Holland, L. Ivany, D. Jablonski, et al. 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification. Proceedings of the National Academy of Sciences 98:6261–6266.

Álvarez-Carretero, S., A. U. Tamuri, M. Battini, F. F. Nascimento, E. Carlisle, R. J. Asher, Z. Yang, P. C. Donoghue, and M. Dos Reis. 2022. A species-level timeline of mammal evolution integrating phylogenomic data. Nature 602:263–267.

André, T., S. Salzman, T. Wendt, and C. D. Specht. 2016. Speciation dynamics and biogeography of Neotropical spiral gingers (Costaceae). Molecular Phylogenetics and Evolution 103:55–63.

Andréoletti, J. and H. Morlon. 2023. Exploring congruent diversification histories with flexibility and parsimony. Methods in Ecology and Evolution 14:2931–2941.

Andréoletti, J., A. Zwaans, R. C. Warnock, G. Aguirre-Fernández, J. Barido-Sottani, A. Gupta, T. Stadler, and M. Manceau. 2022. The occurrence birth–death process for combined-evidence analysis in macroevolution and epidemiology. Systematic Biology 71:1440–1452.

Aphalo, P. J. 2021a. gginnards: Explore the Innards of 'ggplot2' Objects. R package version 0.2.0.

Aphalo, P. J. 2021b. ggpp: Grammar Extensions to 'ggplot2'. R package version 0.5.8.

Arnold, J. B. 2021. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 5.1.0.

Attali, D. and C. Baker. 2016. ggExtra: Add marginal histograms to 'ggplot2', and more 'ggplot2' enhancements. R package version 0.3.

Baele, G., M. A. Suchard, A. Rambaut, and P. Lemey. 2017. Emerging concepts of data integration in pathogen phylodynamics. Systematic Biology 66:e47–e65.

Bank, S. and S. Bradler. 2022. A second view on the evolution of flight in stick and leaf insects (Phasmatodea). BMC ecology and evolution 22:62.

Barido-Sottani, J., G. Aguirre-Fernández, M. J. Hopkins, T. Stadler, and R. Warnock. 2019. Ignoring stratigraphic age uncertainty leads to erroneous estimates of species divergence times under the fossilized birth–death process. Proceedings of the Royal Society B 286:20190685.

Barido-Sottani, J. and H. Morlon. 2023. The ClaDS rate-heterogeneous birth–death prior for full phylogenetic inference in BEAST2. Systematic Biology 72:1180–1187.

Barido-Sottani, J. and H. Morlon. 2025. Integrating fossils samples with heterogeneous diversification rates: a combined Multi-Type Fossilized Birth-Death model. bioRxiv `https://www.biorxiv.org/content/10.1101/2025.02.03.636219v1`.

Barido-Sottani, J., T. G. Vaughan, and T. Stadler. 2020. A multitype birth–death model for Bayesian inference of lineage-specific birth and death rates. Systematic Biology 69:973–986.

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

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

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

Beaulieu, J. M., B. C. O'Meara, and M. J. Donoghue. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Systematic Biology 62:725–737.

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

Benoit, J., C. F. Kammerer, K. Dollman, D. P. Groenewald, and R. M. Smith. 2024. Did gorgonopsians survive the end-Permian "Great Dying"? A re-appraisal of three gorgonopsian specimens (Therapsida, Theriodontia) reported from the Triassic Lystrosaurus declivis Assemblage Zone, Karoo Basin, South Africa. Palaeogeography, Palaeoclimatology, Palaeoecology 638:112044.

Betancur-R, R., E. O. Wiley, G. Arratia, A. Acero, N. Bailly, M. Miya, G. Lecointre, and G. Orti. 2017. Phylogenetic classification of bony fishes. BMC Evolutionary Biology 17:1–40.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. 2017. Julia: A fresh approach to numerical computing. SIAM Review 59:65–98.

Bokma, F. 2009. Problems detecting density-dependent diversification on phylogenies. Proceedings of the Royal Society B: Biological Sciences 276:993–994.

Bollback, J. P. 2002. Bayesian model adequacy and choice in phylogenetics. Molecular Biology and Evolution 19:1171–1180.

Boschman, L. M. and F. L. Condamine. 2022. Mountain radiations are not only rapid and recent: ancient diversification of South American frog and lizard families related to Paleogene Andean orogeny and Cenozoic climate variations. Global and Planetary Change 208:103704.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:e1003537.

Breeschoten, T., C. Doorenweerd, S. Tarasov, and A. P. Vogler. 2016. Phylogenetics and biogeography of the dung beetle genus *Onthophagus* inferred from mitochondrial genomes. Molecular Phylogenetics and Evolution 105:86–95.

Brown, J. M. 2014. Predictive approaches to assessing the fit of evolutionary models. Systematic Biology 63:289–292.

Budd, G. E. and R. P. Mann. 2025. Evolutionary tempo, supertaxa and living fossils. Systematic Biology Page syaf020.

Burress, E. D. and M. M. Muñoz. 2021. Ecological Opportunity from Innovation, not Islands, Drove the Anole Lizard Adaptive Radiation. Systematic Biology 71:93–104.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. The American Naturalist 164:683–695.

Caetano, D. S., B. C. O'Meara, and J. M. Beaulieu. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution 72:2308–2324.

Campanella, D., L. C. Hughes, P. J. Unmack, D. D. Bloom, K. R. Piller, and G. Ortí. 2015. Multi-locus fossil-calibrated phylogeny of Atheriniformes (Teleostei, Ovalentaria). Molecular Phylogenetics and Evolution 86:8–23.

Carvalho, C. M., N. G. Polson, and J. G. Scott. 2010. The horseshoe estimator for sparse signals. Biometrika 97:465–480.

Cerezer, F. O., C. S. Dambros, M. T. Coelho, F. A. Cassemiro, E. Barreto, J. S. Albert, R. O. Wüest, and C. H. Graham. 2023. Accelerated body size evolution in upland environments is correlated with recent speciation in south american freshwater fishes. Nature Communications 14:6070.

Chan, K. M. A. and B. R. Moore. 2005. SYMMETREE: whole-tree analysis of differential diversification rates. Bioinformatics 21:1709–1710.

Chen, W.-J., S. Lavoué, and R. L. Mayden. 2013. Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). Evolution 67:2218–2239.

Chen, X., A. R. Lemmon, E. M. Lemmon, R. A. Pyron, and F. T. Burbrink. 2017. Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. Molecular Phylogenetics and Evolution 111:206–218.

Cibois, A., J.-C. Thibault, C. Bonillo, C. E. Filardi, D. Watling, and E. Pasquet. 2014. Phylogeny and biogeography of the fruit doves (Aves: Columbidae). Molecular Phylogenetics and Evolution 70:442–453.

Condamine, F. L., J. Rolland, S. Höhna, F. A. H. Sperling, and I. Sanmartín. 2018. Testing the role of the Red Queen and Court Jester as drivers of the macroevolution of Apollo butterflies. Systematic Biology 67:940–964.

Condamine, F. L., J. Rolland, and H. Morlon. 2013. Macroevolutionary perspectives to environmental change. Ecology letters 16:72–85.

Condamine, F. L., J. Rolland, and H. Morlon. 2019. Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. Ecology letters 22:1900–1912.

Cook, O. F. 1906. Factors of species-formation. Science 23:506–507.

Cooney, C. R., J. A. Bright, E. J. Capp, A. M. Chira, E. C. Hughes, C. J. Moody, L. O. Nouri, Z. K. Varley, and G. H. Thomas. 2017. Mega-evolutionary dynamics of the adaptive radiation of birds. Nature 542:344–347.

Crouch, N. M. and J. A. Tobias. 2022. The causes and ecological context of rapid morphological evolution in birds. Ecology letters 25:611–623.

Culshaw, V., T. Stadler, and I. Sanmartín. 2019. Exploring the power of Bayesian birth-death skyline models to detect mass extinction events from phylogenies with only extant taxa. Evolution 73:1133–1150.

Danisch, S. and J. Krumbiegel. 2021. Makie.jl: Flexible high-performance data visualization for Julia. Journal of Open Source Software 6:3349.

Davies, T. J., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, and V. Savolainen. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. Proceedings of the National Academy of Sciences 101:1904–1909.

De Lisle, S. P. and E. I. Svensson. 2023. Evolutionary change is remarkably time-independent across scales. bioRxiv `https://www.biorxiv.org/content/10.1101/2023.12.02.569704v3`.

Derryberry, E. P., S. Claramunt, G. Derryberry, R. T. Chesser, J. Cracraft, A. Aleixo, J. Pérez-Emán, J. Remsen, Jr, and R. T. Brumfield. 2011. Lineage diversification and morphological evolution in a large-scale continental radiation: the Neotropical ovenbirds and woodcreepers (Aves: Furnariidae). Evolution 65:2973–2986.

Désamoré, A., B. Laenen, K. B. Miller, and J. Bergsten. 2018. Early burst in body size evolution is uncoupled from species diversification in diving beetles (Dytiscidae). Molecular Ecology 27:979–993.

Devore, J. L. and K. N. Berk. 2012. Modern mathematical statistics with applications vol. 285. 2nd ed. Springer, New York, NY.

Didier, G., M. Fau, and M. Laurin. 2017. Likelihood of tree topologies with fossils and diversification rate estimation. Systematic Biology 66:964–987.

do Rosario Petrucci, B., M. R. May, and T. A. Heath. 2025. Fossils improve extinction-rate estimates under state-dependent diversification models. Philosophical Transactions B 380:20230313.

Donoghue, P. C. J. and Z. Yang. 2016. The evolution of methods for establishing evolutionary timescales. Philosophical Transactions of the Royal Society B: Biological Sciences 371:20160020.

Dormand, J. R. and P. J. Prince. 1980. A family of embedded Runge-Kutta formulae. Journal of computational and applied mathematics 6:19–26.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012a. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29:1969–1973.

Drummond, C. S., R. J. Eastwood, S. T. Miotto, and C. E. Hughes. 2012b. Multiple continental radiations and correlates of diversification in Lupinus (Leguminosae): testing for key innovation with incomplete taxon sampling. Systematic Biology 61:443–460.

Edwards, A. W. F. 1972. Likelihood. Cambridge University Press.

Edwards, A. W. F. and L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. Pages 67–76 *in* Phenetic and phylogenetic classification (V. Heywood and J. McNeil, eds.). Systematics Association Publication.

Egan, J. P., A. M. Simons, M. S. Alavi-Yeganeh, M. P. Hammer, P. Tongnunui, D. Arcila, R. Betancur-R, and D. D. Bloom. 2024. Phylogenomics, lineage diversification rates, and the evolution of diadromy in Clupeiformes (anchovies, herrings, sardines, and relatives). Systematic Biology 73:683–703.

Elliot, M. G. and B. J. Crespi. 2006. Placental invasiveness mediates the evolution of hybrid inviability in mammals. The American Naturalist 168:114–120.

Estes, S. and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. The American Naturalist 169:227–244.

Etienne, R., B. Haegeman, T. Stadler, T. Aze, P. Pearson, A. Purvis, and A. Phillimore. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proceedings of the Royal Society B: Biological Sciences 279:1300–1309.

Etienne, R. S., B. Haegeman, Á. Dugo-Cota, C. Vilà, A. Gonzalez-Voyer, and L. Valente. 2023. The phylogenetic limits to diversity-dependent diversification. Systematic biology 72:433–445.

Fabreti, L. G. and S. Höhna. 2022. Convergence assessment for Bayesian phylogenetic analysis using MCMC simulation. Methods in Ecology and Evolution 13:77–90.

Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61:1–10.

Feller, W. 1939. Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in warscheinlichkeitstheoretischer Behandlung. Acta Biotheoretica 5:11–40.

Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. Systematic Biology 22:240–249.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist 125:1–15.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, MA.

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

FitzJohn, R. G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

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

Fogg, J., E. S. Allman, and C. Ané. 2023. Phylocoalsimulations: a simulator for network multispecies coalescent models, including a new extension for the inheritance of gene flow. Systematic Biology 72:1171–1179.

Folk, R. A., R. L. Stubbs, M. E. Mort, N. Cellinese, J. M. Allen, P. S. Soltis, D. E. Soltis, and R. P. Guralnick. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. Proceedings of the National Academy of Sciences 116:10874–10882.

Freyman, W. A. and S. Höhna. 2018. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Systematic Biology 67:195–215.

Freyman, W. A. and S. Höhna. 2019. Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. Systematic Biology 68:505–519.

Gallager, R. 1962. Low-density parity-check codes. IRE Transactions on information theory 8:21–28.

Gearty, W. 2021. deeptime: Plotting Tools for Anyone Working in Deep Time. R package version 2.1.0.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian data analysis. CRC press.

Geman, S. and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence 6:721–741.

Gingerich, P. D. 1983. Rates of evolution: effects of time and temporal scaling. Science 222:159–161.

Givnish, T. J., A. Zuluaga, D. Spalink, M. Soto Gomez, V. K. Lam, J. M. Saarela, C. Sass, W. J. Iles, D. J. L. De Sousa, J. Leebens-Mack, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. American Journal of Botany 105:1888–1910.

Gohli, J., L. R. Kirkendall, S. M. Smith, A. I. Cognato, J. Hulcr, and B. H. Jordal. 2017. Biological factors contributing to bark and ambrosia beetle species diversification. Evolution 71:1258–1272.

Goldberg, E. E. and B. Igić. 2008. On phylogenetic tests of irreversible evolution. Evolution: International Journal of Organic Evolution 62:2727–2741.

Goldberg, E. E. and B. Igić. 2012. Tempo and mode in plant breeding system evolution. Evolution: International Journal of Organic Evolution 66:3701–3709.

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

Grabowski, M., B. T. Kopperud, M. Tsuboi, and T. F. Hansen. 2023a. Both diet and sociality affect primate brain-size evolution. Systematic Biology 72:404–418.

Grabowski, M., J. Pienaar, K. L. Voje, S. Andersson, J. Fuentes-González, B. T. Kopperud, D. S. Moen, M. Tsuboi, J. Uyeda, and T. F. Hansen. 2023b. A cautionary note on "A

cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies". Systematic Biology 72:955–963.

Groves, C. P. 2005. Order Primates. *in* Mammal species of the world: a taxonomic and geographic reference (D. Wilson and D. Reeder, eds.) 3rd ed. Johns Hopkins University Press.

Gupta, A., M. Manceau, T. Vaughan, M. Khammash, and T. Stadler. 2020. The probability distribution of the reconstructed phylogenetic tree with occurrence data. Journal of Theoretical Biology 488:110115.

Hairer, E. and G. Wanner. 1996. Solving Ordinary Differential Equations II. Stiff and Differential-algebraic problems. Springer series in computational mathematics 14. Springer-Verlag, London.

Hansen, T. F. 2024. Three modes of evolution? Remarks on rates of evolution and time scaling. Journal of Evolutionary Biology 37:1523–1537.

Harmon, L. J., M. W. Pennell, L. F. Henao-Diaz, J. Rolland, B. N. Sipley, and J. C. Uyeda. 2021. Causes and consequences of apparent timescaling across all estimated evolutionary rates. Annual Review of Ecology, Evolution, and Systematics 52:587–609.

Harvey, P. H. and M. D. Pagel. 1991. The Comparative Method in Evolutionary Biology. Oxford University Press.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences 111:E2957–E2966.

Hedges, S. B., J. Marin, M. Suleski, M. Paymer, and S. Kumar. 2015. Tree of life reveals clock-like speciation and diversification. Molecular Biology and Evolution 32:835–845.

Hein, J. 1990. Reconstructing evolution of sequences subject to recombination using parsimony. Mathematical biosciences 98:185–200.

Helmstetter, A. J., S. Glemin, J. Käfer, R. Zenil-Ferguson, H. Sauquet, H. de Boer, L.-P. M. Dagallier, N. Mazet, E. L. Reboud, T. L. Couvreur, et al. 2022. Pulled diversification rates, lineages-through-time plots, and modern macroevolutionary modeling. Systematic Biology 71:758–773.

Henao Diaz, L. F., L. J. Harmon, M. T. Sugawara, E. T. Miller, and M. W. Pennell. 2019. Macroevolutionary diversification rates show time dependency. Proceedings of the National Academy of Sciences 116:7403–7408.

Hennequin, S., G. Rouhan, A. Salino, Y.-F. Duan, M.-C. Lepeigneux, M. Guillou, S. Ansell, T. E. Almeida, L.-B. Zhang, and H. Schneider. 2017. Global phylogeny and biogeography of the fern genus Ctenitis (Dryopteridaceae), with a focus on the Indian Ocean region. Molecular Phylogenetics and Evolution 112:277–289.

Henríquez-Piskulich, P., A. F. Hugall, and D. Stuart-Fox. 2024. A supermatrix phylogeny of the world's bees (Hymenoptera: Anthophila). Molecular Phylogenetics and Evolution 190:107963.

Hernández-Hernández, T., E. C. Miller, C. Román-Palacios, and J. J. Wiens. 2021. Speciation across the tree of life. Biological Reviews 96:1205–1242.

Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. Evolution 46:627–640.

Higdon, J. W., O. R. Bininda-Emonds, R. Beck, and S. H. Ferguson. 2007. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. BMC Evolutionary Biology 7:1–19.

Hipp, A. L., P. S. Manos, A. González-Rodríguez, M. Hahn, M. Kaproth, J. D. McVay, S. V. Avalos, and J. Cavender-Bares. 2018. Sympatric parallel diversification of major oak clades in the americas and the origins of mexican species diversity. New Phytologist 217:439–452.

Ho, S. Y., B. Shapiro, M. J. Phillips, A. Cooper, and A. J. Drummond. 2007. Evidence for time dependency of molecular rate estimates. Systematic Biology 56:515–522.

Hodges, S. A. and M. L. Arnold. 1995. Spurring plant diversification: are floral nectar spurs a key innovation? Proceedings of the Royal Society of London. Series B: Biological Sciences 262:343–348.

Höhna, S. 2014. Likelihood Inference of Non-Constant Diversification Rates with Incomplete Taxon Sampling. PLoS one 9:e84184.

Höhna, S. 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. Journal of Theoretical Biology 380:321–331.

Höhna, S., L. M. Coghill, G. G. Mount, R. C. Thomson, and J. M. Brown. 2018. P3: Phylogenetic posterior prediction in RevBayes. Molecular Biology and Evolution 35:1028–1034.

Höhna, S., W. A. Freyman, Z. Nolen, J. P. Huelsenbeck, M. R. May, and B. R. Moore. 2019. A Bayesian approach for estimating branch-specific speciation and extinction rates. BioRxiv `https://www.biorxiv.org/content/10.1101/555805v1`.

Höhna, S., T. A. Heath, B. Boussau, M. J. Landis, F. Ronquist, and J. P. Huelsenbeck. 2014. Probabilistic graphical model representation in phylogenetics. Systematic Biology 63:753–771.

Höhna, S., B. T. Kopperud, and A. F. Magee. 2022. CRABS: Congruent rate analyses in birth–death scenarios. Methods in Ecology and Evolution 13:2709–2718.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist. 2016a. Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology 65:726–736.

Höhna, S., M. R. May, and B. R. Moore. 2016b. TESS: an R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. Bioinformatics 32:789–791.

Höhna, S., T. Stadler, F. Ronquist, and T. Britton. 2011. Inferring speciation and extinction rates under different species sampling schemes. Molecular Biology and Evolution 28:2577–2589.

Hosner, P. A., E. L. Braun, and R. T. Kimball. 2016. Rapid and recent diversification of curassows, guans, and chachalacas (Galliformes: Cracidae) out of Mesoamerica: Phylogeny inferred from mitochondrial, intron, and ultraconserved element sequences. Molecular Phylogenetics and Evolution 102:320–330.

Hua, X., T. Herdha, and C. J. Burden. 2022. Protracted speciation under the state-dependent speciation and extinction approach. Systematic Biology 71:1362–1377.

Hua, X. and J. J. Wiens. 2013. How does climate influence speciation? The American Naturalist 182:1–12.

Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic mapping of morphological characters. Systematic Biology 52:131–158.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

Iles, W. J., C. Sass, L. Lagomarsino, G. Benson-Martin, H. Driscoll, and C. D. Specht. 2017. The phylogeny of Heliconia (Heliconiaceae) and the evolution of floral presentation. Molecular Phylogenetics and Evolution 117:150–167.

Ingram, T. 2011. Speciation along a depth gradient in a marine adaptive radiation. Proceedings of the Royal Society B: Biological Sciences 278:613–618.

Ingram, T., A. Harrison, D. L. Mahler, M. d. R. Castañeda, R. E. Glor, A. Herrel, Y. E. Stuart, and J. B. Losos. 2016. Comparative tests of the role of dewlap size in Anolis lizard speciation. Proceedings of the Royal Society B: Biological Sciences 283:20162199.

Jeffreys, H. 1961. Theory of Probability. 3rd ed. Oxford University Press.

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

Jønsson, K. A., P.-H. Fabre, J. D. Kennedy, B. G. Holt, M. K. Borregaard, C. Rahbek, and J. Fjeldså. 2016. A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). Molecular Phylogenetics and Evolution 94:87–94.

Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules. Pages 21–132 *in* Mammalian Protein Metabolism (H. N. Munro, ed.) vol. 3. Academic Press, New York, NY.

Kass, R. E. and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90:773–795.

Kassambara, A. 2018. ggpubr: "ggplot2" based publication ready plots. R package version 0.1.

Kawahara, A. Y., C. Storer, A. P. S. Carvalho, D. M. Plotkin, F. L. Condamine, M. P. Braga, E. A. Ellis, R. A. St Laurent, X. Li, V. Barve, et al. 2023. A global phylogeny of butterflies reveals their evolutionary history, ancestral hosts and biogeographic origins. Nature Ecology & Evolution 7:903–913.

Kendall, D. G. 1948. On the generalized "birth-and-death" process. The Annals of Mathematical Statistics 19:1–15.

Kerman, J., A. Gelman, T. Zheng, and Y. Ding. 2008. Visualization in Bayesian data analysis. Pages 709–724 *in* Handbook of Data Visualization. Springer.

Knuth, D. 1997. The Art of Computer Programming vol. 1. 3rd ed. Addison-Wesley, Reading, MA.

Kopperud, B. T. and S. Höhna. 2025. Phylogenetic Estimation of Branch-Specific Shifts in the Tempo of Origination. Systematic Biology Page syaf041.

Kopperud, B. T., S. Lidgard, and L. H. Liow. 2022a. Enhancing georeferenced biodiversity inventories: automated information extraction from literature records reveal the gaps. PeerJ 10:e13921.

Kopperud, B. T., A. F. Magee, and S. Höhna. 2022b. afmagee/CRABS: CRABS v1.0. `https://doi.org/10.5281/zenodo.7079514`.

Kopperud, B. T., A. F. Magee, and S. Höhna. 2023. Rapidly changing speciation and extinction rates can be inferred in spite of nonidentifiability. Proceedings of the National Academy of Sciences 120:e2208851120.

Kozak, K. H., D. W. Weisrock, and A. Larson. 2006. Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: Plethodon). Proceedings of the Royal Society B: Biological Sciences 273:539–546.

Kriebel, R., B. T. Drew, C. P. Drummond, J. G. González-Gallegos, F. Celep, M. M. Mahdjoub, J. P. Rose, C.-L. Xiang, G.-X. Hu, J. B. Walker, et al. 2019. Tracking temporal shifts in area, biomes, and pollinators in the radiation of *Salvia* (sages) across continents: leveraging anchored hybrid enrichment and targeted sequence data. American Journal of Botany 106:573–597.

Kubo, T. and Y. Iwasa. 1995. Inferring the rates of branching and extinction from molecular phylogenies. Evolution 49:694–704.

Kuhner, M. K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Molecular Biology and Evolution 11:459–468.

Lagomarsino, L. P., E. J. Forrestel, N. Muchhala, and C. C. Davis. 2017. Repeated evolution of vertebrate pollination syndromes in a recently diverged Andean plant clade. Evolution 71:1970–1985.

Landis, M. J., W. A. Freyman, and B. G. Baldwin. 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. Evolution 72:2343–2359.

Leaché, A. D., P. Wagner, C. W. Linkem, W. Böhme, T. J. Papenfuss, R. A. Chong, B. R. Lavin, A. M. Bauer, S. V. Nielsen, E. Greenbaum, et al. 2014. A hybrid phylogenetic–phylogenomic approach for species tree estimation in African Agama lizards with applications to biogeography, character evolution, and diversification. Molecular Phylogenetics and Evolution 79:215–230.

Legried, B. and J. Terhorst. 2022. A class of identifiable phylogenetic birth–death models. Proceedings of the National Academy of Sciences 119:e2119513119.

Legried, B. and J. Terhorst. 2023. Identifiability and inference of phylogenetic birth–death models. Journal of Theoretical Biology 568:111520.

Lenth, R. V. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.3.

Leslie, A. B., J. Beaulieu, G. Holman, C. S. Campbell, W. Mei, L. R. Raubeson, and S. Mathews. 2018. An overview of extant conifer evolution from the perspective of the fossil record. American Journal of Botany 105:1531–1544.

Leslie, A. B., J. M. Beaulieu, H. S. Rai, P. R. Crane, M. J. Donoghue, and S. Mathews. 2012. Hemisphere-scale differences in conifer evolutionary dynamics. Proceedings of the National Academy of Sciences 109:16217–16221.

Letsch, H., B. Gottsberger, and J. L. Ware. 2016. Not going with the flow: a comprehensive time-calibrated phylogeny of dragonflies (Anisoptera: Odonata: Insecta) provides evidence for the role of lentic habitats on diversification. Molecular Ecology 25:1340–1353.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50:913–925.

Liow, L. H. and J. A. Finarelli. 2014. A dynamic global equilibrium in carnivoran diversification over 20 million years. Proceedings of the Royal Society B: Biological Sciences 281:20132312.

Liow, L. H., T. B. Quental, and C. R. Marshall. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? Systematic Biology 59:646–659.

Liu, J., A. J. Lindstrom, N. S. Nagalingum, J. J. Wiens, and X. Gong. 2021. Testing the causes of richness patterns in the paleotropics: time and diversification in cycads (Cycadaceae). Ecography 44:1606–1618.

Louca, S. and M. Doebeli. 2018. Efficient comparative phylogenetics on large trees. Bioinformatics 34:1053–1055.

Louca, S., L. F. Henao-Diaz, and M. Pennell. 2022. The scaling of diversification rates with age is likely explained by sampling bias. Evolution 76:1625–1637.

Louca, S., A. McLaughlin, A. MacPherson, J. B. Joy, and M. W. Pennell. 2021. Fundamental Identifiability Limits in Molecular Epidemiology. Molecular Biology and Evolution 38:4010–4024.

Louca, S. and M. W. Pennell. 2020a. A general and efficient algorithm for the likelihood of diversification and discrete-trait evolutionary models. Systematic Biology 69:545–556.

Louca, S. and M. W. Pennell. 2020b. Extant timetrees are consistent with a myriad of diversification histories. Nature 580:502–505.

Louca, S. and M. W. Pennell. 2021. Why extinction estimates from extant phylogenies are so often zero. Current Biology 31:3168–3173.

Louca, S., P. M. Shih, M. W. Pennell, W. W. Fischer, L. W. Parfrey, and M. Doebeli. 2018. Bacterial diversification through geological time. Nature Ecology & Evolution 2:1458–1467.

Lovette, I. J. and E. Bermingham. 1999. Explosive speciation in the new world dendroica warblers. Proceedings of the Royal Society of London. Series B: Biological Sciences 266:1629–1636.

Lüdecke, D., I. Patil, M. S. Ben-Shachar, B. M. Wiernik, P. Waggoner, and D. Makowski. 2021. see: An R package for visualizing statistical models. Journal of Open Source Software 6:3393.

Machado, F. A., C. S. Mongle, G. Slater, A. Penna, A. Wisniewski, A. Soffin, V. Dutra, and J. C. Uyeda. 2023. Rules of teeth development align microevolution with macroevolution in extant and extinct primates. Nature Ecology & Evolution 7:1729–1739.

MacPherson, A., S. Louca, A. McLaughlin, J. B. Joy, and M. W. Pennell. 2022. Unifying phylogenetic birth–death models in epidemiology and macroevolution. Systematic Biology 71:172–189.

Maddison, D. R., D. L. Swofford, and W. P. Maddison. 1997. NEXUS: an extensible file format for systematic information. Systematic Biology 46:590–621.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56:701–710.

Magee, A. F. and S. Höhna. 2021. Impact of K-Pg Mass Extinction Event on Crocodylomorpha Inferred from Phylogeny of Extinct and Extant Taxa. bioRxiv `https://www.biorxiv.org/content/10.1101/2021.01.14.426715v1`.

Magee, A. F., S. Höhna, T. I. Vasylyeva, A. D. Leaché, and V. N. Minin. 2020. Locally adaptive Bayesian birth-death model successfully detects slow and rapid rate shifts. PLOS Computational Biology 16:e1007999.

Mair, P., P. J. Groenen, and J. de Leeuw. 2022. More on multidimensional scaling and unfolding in r: smacof version 2. Journal of Statistical Software 102:1–47.

Maliet, O., F. Hartig, and H. Morlon. 2019. A model with many small shifts for estimating species-specific diversification rates. Nature Ecology & Evolution 3:1086–1092.

Maliet, O. and H. Morlon. 2022. Fast and accurate estimation of species-specific diversification rates using data augmentation. Systematic Biology 71:353–366.

Martínez-Gómez, J., M. J. Song, C. M. Tribble, B. T. Kopperud, W. A. Freyman, S. Höhna, C. D. Specht, and C. J. Rothfels. 2024. Commonly used Bayesian diversification methods lead to biologically meaningful differences in branch-specific rates on empirical phylogenies. Evolution Letters 8:189–199.

May, M. R., S. Höhna, and B. R. Moore. 2016. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. Methods in Ecology and Evolution 7:947–959.

May, M. R. and B. R. Moore. 2016. How Well Can We Detect Lineage-Specific Diversification-Rate Shifts? A Simulation Study of Sequential AIC Methods. Systematic Biology 65:1076–1084.

May, M. R. and C. J. Rothfels. 2023. Diversification models conflate likelihood and prior, and cannot be compared using conventional model-comparison tools. Systematic Biology 72:713–722.

Mazet, N., H. Morlon, P.-h. Fabre, and F. L. Condamine. 2023. Estimating clade-specific diversification rates and palaeodiversity dynamics from reconstructed phylogenies. Methods in Ecology and Evolution 14:2575–2591.

McCord, C. L. and M. W. Westneat. 2016. Phylogenetic relationships and the evolution of BMP4 in triggerfishes and filefishes (Balistoidea). Molecular Phylogenetics and Evolution 94:397–409.

McLean, M. W. 2014. Straightforward Bibliography Management in R Using the RefManager Package. arXiv `https://arxiv.org/abs/1403.2036`.

Melo, B. F., B. L. Sidlauskas, T. J. Near, F. F. Roxo, A. Ghezelayagh, L. E. Ochoa, M. L. Stiassny, J. Arroyave, J. Chang, B. C. Faircloth, et al. 2022. Accelerated diversification explains the exceptional species richness of tropical characoid fishes. Systematic Biology 71:78–92.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21:1087–1092.

Meyer, A. L., C. Román-Palacios, and J. J. Wiens. 2018. BAMM gives misleading rate estimates in simulated and empirical datasets. Evolution 72:2257–2266.

Meyer, A. L. and J. J. Wiens. 2018. Estimating diversification rates for higher taxa: BAMM can give problematic estimates of rates and rate shifts. Evolution 72:39–53.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pages 1–8 *in* 2010 gateway computing environments workshop (GCE) Ieee.

Mitchell, J. S., R. S. Etienne, and D. L. Rabosky. 2019. Inferring diversification rate variation from phylogenies with fossils. Systematic Biology 68:1–18.

Mitchell, J. S. and D. L. Rabosky. 2017. 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.

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

Mogensen, P. K. and A. N. Riseth. 2018. Optim: A mathematical optimization package for Julia. Journal of Open Source Software 3:615.

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

Morlon, H. 2014. Phylogenetic approaches for studying diversification. Ecology letters 17:508–525.

Morlon, H., J. Andréoletti, J. Barido-Sottani, S. Lambert, B. Perez-Lamarque, I. Quintero, V. Senderov, and P. Veron. 2024. Phylogenetic insights into diversification. Annual Review of Ecology, Evolution, and Systematics 55.

Morlon, H., B. D. Kemps, J. B. Plotkin, and D. Brisson. 2012. Explosive radiation of a bacterial species group. Evolution 66:2577–2586.

Morlon, H., E. Lewitus, F. L. Condamine, M. Manceau, J. Clavel, and J. Drury. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Methods in Ecology and Evolution 7:589–597.

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

Morlon, H., S. Robin, and F. Hartig. 2022. Studying speciation and extinction dynamics from phylogenies: addressing identifiability issues. Trends in Ecology & Evolution 37:497–506.

Mortimer, S. M., J. Boyko, J. M. Beaulieu, and D. C. Tank. 2022. Synthesizing existing phylogenetic data to advance phylogenetic research in orobanchaceae. Systematic Botany 47:533–544.

Mulder, C. 2003. Aristolochiaceae. Review of Palaeobotany and Palynology 123:47–55.

Müller, K. and H. Wickham. 2021. tibble: Simple Data Frames. R package version 3.1.2.

Nakagawa, S. and I. C. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological reviews 82:591–605.

Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 344:77–82.

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

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press.

Nelsen, M. P., R. Lücking, C. K. Boyce, H. T. Lumbsch, and R. H. Ree. 2020. The macroevolutionary dynamics of symbiotic and phenotypic diversification in lichens. Proceedings of the National Academy of Sciences 117:21495–21503.

Neupane, S., P. O. Lewis, S. Dessein, H. Shanks, S. Paudyal, and F. Lens. 2017. Evolution of woody life form on tropical mountains in the tribe Spermacoceae (Rubiaceae). American Journal of Botany 104:419–438.

Nielsen, R. 2002. Mapping mutations on phylogenies. Systematic Biology 51:729–739.

Nitta, J. H., E. Schuettpelz, S. Ramírez-Barahona, and W. Iwasaki. 2022. An open and continuously updated fern tree of life. Frontiers in Plant Science 13:909768.

Oka, S.-i., T. Tomita, and K. Miyamoto. 2016. A mighty claw: pinching force of the coconut crab, the largest terrestrial crustacean. PLoS One 11:e0166108.

Olroyd, S. L. and B. T. Kopperud. 2025. Allometry of sound reception structures and evidence for a mandibular middle ear in non-mammalian synapsids. Evolution 79:905–921.

O'Meara, B. and J. Beaulieu. 2021. Potential survival of some, but not all, diversification methods. EcoEvoRxiv `https://doi.org/10.32942/osf.io/w5nvd`.

Ooms, J. 2020. pdftools: Text Extraction, Rendering and Converting of PDF Documents. R package version 2.3.1.

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

Palazzesi, L., O. Hidalgo, V. D. Barreda, F. Forest, and S. Höhna. 2022. The rise of grasslands is linked to atmospheric CO2 decline in the late Palaeogene. Nature Communications 13:293.

Paradis, E. and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Pawitan, Y. 2001. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press.

Pearl, J. 1982. Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach. Pages 133–136 *in* Proceedings of the National Conference on Artificial Intelligence AAAI, Pittsburgh, PA.

Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. 1st ed. Morgan Kaufmann Publishers, San Francisco, CA.

Peoples, N., M. D. Burns, M. Mihalitsis, and P. C. Wainwright. 2025. Evolutionary lability of a key innovation spurs rapid diversification. Nature 639:962–967.

Pereira, A. G., J. Sterli, F. R. Moreira, and C. G. Schrago. 2017. Multilocus phylogeny and statistical biogeography clarify the evolutionary history of major lineages of turtles. Molecular Phylogenetics and Evolution 113:59–66.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. R news 6:7–11.

Portik, D. M., J. W. Streicher, and J. J. Wiens. 2023. Frog phylogeny: a time-calibrated, species-level tree based on hundreds of loci and 5,242 species. Molecular Phylogenetics and Evolution 188:107907.

Powers, D. M. 2011. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. International Journal of Machine Learning Technology 2:37–63.

Prebus, M. 2017. Insights into the evolution, biogeography and natural history of the acorn ants, genus Temnothorax Mayr (Hymenoptera: Formicidae). BMC Evolutionary Biology 17:1–22.

Price, S. L., R. S. Etienne, and S. Powell. 2016. Tightly congruent bursts of lineage and phenotypic diversification identified in a continental ant radiation. Evolution 70:903–912.

Pyron, R. A. and F. T. Burbrink. 2013. Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. Trends in Ecology & Evolution 28:729–736.

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

Quintero, I., J. Andréoletti, D. Silvestro, and H. Morlon. 2025. The rise, decline and fall of clades. bioRxiv https://www.biorxiv.org/content/10.1101/2025.03.20.644316v1.

Quintero, I., N. Lartillot, and H. Morlon. 2024. Imbalanced speciation pulses sustain the radiation of mammals. Science 384:1007–1012.

Quintero, I., M. A. Suchard, and W. Jetz. 2022. Macroevolutionary dynamics of climatic niche space. Proceedings of the Royal Society B 289:20220091.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria.

Rabiner, L. R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77:257–286.

Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164.

Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64:1816–1824.

Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PloS one 9:e89543.

Rabosky, D. L. 2018. BAMM at the court of false equivalency: a response to Meyer and Wiens. Evolution 72:2246–2256.

Rabosky, D. L. and R. B. Benson. 2021. Ecological and biogeographic drivers of biodiversity cannot be resolved using clade age-richness data. Nature Communications 12:2945.

Rabosky, D. L., J. Chang, P. O. Title, P. F. Cowman, L. Sallan, M. Friedman, K. Kaschner, C. Garilao, T. J. Near, M. Coll, et al. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature 559:392–395.

Rabosky, D. L., S. C. Donnellan, M. Grundler, and I. J. Lovette. 2014a. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. Systematic Biology 63:610–627.

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

Rabosky, D. L. and I. J. Lovette. 2008a. Density-dependent diversification in North American wood warblers. Proceedings of the Royal Society B: Biological Sciences 275:2363–2371.

Rabosky, D. L. and I. J. Lovette. 2008b. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution 62:1866–1875.

Rabosky, D. L., J. S. Mitchell, and J. Chang. 2017. Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. Systematic Biology 66:477–498.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature communications 4:1958.

Rackauckas, C. and Q. Nie. 2017. DifferentialEquations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia. Journal of open research software 5.

Rambaut, A. 2014. FigTree 1.4. 2 software. Institute of Evolutionary Biology, Univ. Edinburgh `https://tree.bio.ed.ac.uk/software/figtree/`.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67:901–904.

Raup, D. M. 1986. Biological extinction in earth history. Science 231:1528–1533.

Raup, D. M., S. J. Gould, T. J. Schopf, and D. S. Simberloff. 1973. Stochastic models of phylogeny and the evolution of diversity. The Journal of Geology 81:525–542.

Raup, D. M. and J. J. Sepkoski. 1982. Mass extinctions in the marine fossil record. Science 215:1501–1503.

Reaney, A. M., M. Saldarriaga-Córdoba, and D. Pincheira-Donoso. 2018. Macroevolutionary diversification with limited niche disparity in a species-rich lineage of cold-climate lizards. BMC Evolutionary Biology 18:1–12.

Ree, R. H. and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57:4–14.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Revell, L. J. 2025. Ancestral state reconstruction of phenotypic characters. Evolutionary Biology 52:1–25.

Revels, J., M. Lubin, and T. Papamarkou. 2016. Forward-mode automatic differentiation in Julia. arXiv `https://arxiv.org/abs/1607.07892`.

Ricklefs, R. E. 2007. Estimating diversification rates from phylogenetic information. Trends in Ecology & Evolution 22:601–610.

Roberts-Hugghis, A. S., C. M. Martinez, K. A. Corn, and P. C. Wainwright. 2025. A classic key innovation constrains oral jaw functional diversification in fishes. Evolution Letters 9:24–40.

Robinson, D. F. 1971. Comparison of labeled trees with valency three. Journal of combinatorial theory, Series B 11:105–119.

Robinson, D. F. and L. R. Foulds. 1981. Comparison of phylogenetic trees. Mathematical biosciences 53:131–147.

Rojas, D., C. Mancina, J. Flores-Martínez, and L. Navarro. 2013. Phylogenetic signal, feeding behaviour and brain volume in Neotropical bats. Journal of Evolutionary Biology 26:1925–1933.

Ronquist, F., J. Kudlicka, V. Senderov, J. Borgström, N. Lartillot, D. Lundén, L. Murray, T. B. Schön, and D. Broman. 2021. Universal probabilistic programming offers a powerful approach to statistical phylogenetics. Communications Biology 4:1–10.

Ronquist, F. and I. Sanmartín. 2011. Phylogenetic methods in biogeography. Annual Review of Ecology, Evolution, and Systematics 42:441–464.

Rose, J. P., T. J. Kleist, S. D. Löfstrand, B. T. Drew, J. Schönenberger, and K. J. Sytsma. 2018. Phylogeny, historical biogeography, and diversification of angiosperm order ericales suggest ancient neotropical and east asian connections. Molecular Phylogenetics and Evolution 122:59–79.

Russell, S. J. and P. Norvig. 2010. Artificial intelligence: a modern approach. 3rd ed. Pearson, New Jersey.

Sánchez-García, M., M. Ryberg, F. K. Khan, T. Varga, L. G. Nagy, and D. S. Hibbett. 2020. Fruiting body form, not nutritional mode, is the major driver of diversification in mushroom-forming fungi. Proceedings of the National Academy of Sciences 117:32528–32534.

Sanderson, M. J. and M. J. Donoghue. 1994. Shifts in Diversification Rate with the Origin of Angiosperms. Science 264:1590–1593.

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

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences 74:5463–5467.

Saulsbury, J. G., C. T. Parins-Fukuchi, C. J. Wilson, T. Reitan, and L. H. Liow. 2024. Age-dependent extinction and the neutral theory of biodiversity. Proceedings of the National Academy of Sciences 121:e2307629121.

Scantlebury, D. P. 2013. Diversification rates have declined in the Malagasy herpetofauna. Proceedings of the Royal Society B: Biological Sciences 280:20131109.

Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.

Schmidt, S. and G. Walter. 2014. Young clades in an old family: major evolutionary transitions and diversification of the eucalypt-feeding pergid sawflies in Australia (Insecta, Hymenoptera, Pergidae). Molecular Phylogenetics and Evolution 74:111–121.

Sepkoski, J. J. 1998. Rates of speciation in the fossil record. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 353:315–326.

Serrano-Serrano, M. L., J. Rolland, J. L. Clark, N. Salamin, and M. Perret. 2017. Hummingbird pollination and the diversification of angiosperms: an old and successful association in Gesneriaceae. Proceedings of the Royal Society B: Biological Sciences 284:20162816.

Shakya, S. B., J. Fuchs, J.-M. Pons, and F. H. Sheldon. 2017. Tapping the woodpecker tree for evolutionary insight. Molecular Phylogenetics and Evolution 116:182–191.

Sheets, H. D. and C. E. Mitchell. 2001. Uncorrelated change produces the apparent dependence of evolutionary rate on interval. Paleobiology 27:429–445.

Sherratt, E., A. Alejandrino, A. C. Kraemer, J. M. Serb, and D. C. Adams. 2016. Trends in the sand: directional evolution in the shell shape of recessing scallops (Bivalvia: Pectinidae). Evolution 70:2061–2073.

Shi, J. J. and D. L. Rabosky. 2015. Speciation dynamics during the global radiation of extant bats. Evolution 69:1528–1545.

Silvestro, D., G. Zizka, and K. Schulte. 2014. Disentangling the effects of key innovations on the diversification of Bromelioideae (Bromeliaceae). Evolution 68:163–175.

Slater, G. J., L. J. Harmon, and M. E. Alfaro. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. Evolution 66:3931–3944.

Slater, G. J., S. A. Price, F. Santini, and M. E. Alfaro. 2010. Diversity versus disparity and the radiation of modern cetaceans. Proceedings of the Royal Society B: Biological Sciences 277:3097–3104.

Slotte, T., H. Huang, M. Lascoux, and A. Ceplitis. 2008. Polyploid speciation did not confer instant reproductive isolation in Capsella (Brassicaceae). Molecular Biology and Evolution 25:1472–1481.

Smith, S. A., J. W. Brown, Y. Yang, R. Bruenn, C. P. Drummond, S. F. Brockington, J. F. Walker, N. Last, N. A. Douglas, and M. J. Moore. 2018. Disparity, diversity, and duplications in the Caryophyllales. New Phytologist 217:836–854.

Smith, S. A. and B. C. O'Meara. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28:2689–2690.

Soares, A. E., B. J. Novak, J. Haile, T. H. Heupink, J. Fjeldså, M. T. P. Gilbert, H. Poinar, G. M. Church, and B. Shapiro. 2016. Complete mitochondrial genomes of living and extinct pigeons revise the timing of the columbiform radiation. BMC Evolutionary Biology 16:1–9.

Soetaert, K., T. Petzoldt, and R. W. Setzer. 2010. Solving differential equations in R: Package deSolve. Journal of Statistical Software 33:1–25.

Soghigian, J., T. G. Andreadis, and T. P. Livdahl. 2017. From ground pools to treeholes: convergent evolution of habitat and phenotype in aedes mosquitoes. BMC Evolutionary Biology 17:1–13.

Spriggs, E. L., P.-A. Christin, and E. J. Edwards. 2014. C4 photosynthesis promoted species diversification during the miocene grassland expansion. PloS one 9:e97722.

Springer, M. S., R. W. Meredith, J. Gatesy, C. A. Emerling, J. Park, D. L. Rabosky, T. Stadler, C. Steiner, O. A. Ryder, J. E. Janečka, et al. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PloS one 7.

Stadler, T. 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. Journal of Theoretical Biology 261:58–66.

Stadler, T. 2010. Sampling-through-time in birth–death trees. Journal of Theoretical Biology 267:396–404.

Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences 108:6187–6192.

Steeman, M. E., M. B. Hebsgaard, R. E. Fordyce, S. Y. Ho, D. L. Rabosky, R. Nielsen, C. Rahbek, H. Glenner, M. V. Sørensen, and E. Willerslev. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. Systematic Biology 58:573–585.

Stein, R. W., J. W. Brown, and A. Ø. Mooers. 2015. A molecular genetic time scale demonstrates Cretaceous origins and multiple diversification rate shifts within the order Galliformes (Aves). Molecular Phylogenetics and Evolution 92:155–164.

Stein, R. W., C. G. Mull, T. S. Kuhn, N. C. Aschliman, L. N. Davidson, J. B. Joy, G. J. Smith, N. K. Dulvy, and A. O. Mooers. 2018. Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. Nature Ecology & Evolution 2:288–298.

Stern, D. B., J. Breinholt, C. Pedraza-Lara, M. López-Mejía, C. L. Owen, H. Bracken-Grissom, J. W. Fetzner Jr, and K. A. Crandall. 2017. Phylogenetic evidence from freshwater crayfishes that cave adaptation is not an evolutionary dead-end. Evolution 71:2522–2532.

Sun, M., R. A. Folk, M. A. Gitzendanner, P. S. Soltis, Z. Chen, D. E. Soltis, and R. P. Guralnick. 2020. Recent accelerated diversification in rosids occurred outside the tropics. Nature communications 11:3333.

Taylor, P. D. and A. Waeschenbach. 2015. Phylogeny and diversification of bryozoans. Palaeontology 58:585–599.

Teo, B., P. Bastide, and C. Ané. 2025. Leveraging graphical model techniques to study evolution on phylogenetic networks. Philosophical Transactions B 380:20230310.

Testo, W. and M. Sundue. 2016. A 4000-species dataset provides new insight into the evolution of ferns. Molecular Phylogenetics and Evolution 105:200–211.

Thomas, S. K., X. Liu, Z.-Y. Du, Y. Dong, A. Cummings, L. Pokorny, Q.-Y. Xiang, and J. H. Leebens-Mack. 2021. Comprehending cornales: phylogenetic reconstruction of the order using the angiosperms353 probe set. American Journal of Botany 108:1112–1121.

Thompson, E. A. 1975. Human evolutionary trees. Cambridge University Press.

Thomson, R. C., P. Q. Spinks, and H. B. Shaffer. 2021. A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. Proceedings of the National Academy of Sciences 118:e2012215118.

Thornhill, A. H., S. Y. Ho, C. Külheim, and M. D. Crisp. 2015. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Molecular Phylogenetics and Evolution 93:29–43.

Title, P. O. and D. L. Rabosky. 2019. Tip rates, phylogenies and diversification: what are we estimating, and how good are the estimates? Methods in Ecology and Evolution 10:821–834.

Toljagić, O., K. L. Voje, M. Matschiner, L. H. Liow, and T. F. Hansen. 2018. Millions of years behind: slow adaptation of ruminants to grasslands. Systematic Biology 67:145–157.

Tribble, C. M., W. A. Freyman, M. J. Landis, J. Y. Lim, J. Barido-Sottani, B. T. Kopperud, S. Höhna, and M. R. May. 2022. RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes. Methods in Ecology and Evolution 13:314–323.

Tribble, C. M., J. I. Márquez-Corro, M. R. May, A. L. Hipp, M. Escudero, and R. Zenil-Ferguson. 2025. Macroevolutionary inference of complex modes of chromosomal speciation in a cosmopolitan plant lineage. New Phytologist 245:2350–2361.

Tsitouras, C. 2011. Runge–Kutta pairs of order 5(4) satisfying only the first column simplifying assumption. Computers & Mathematics with Applications 62:770–775.

Tsuboi, M., B. T. Kopperud, M. Matschiner, M. Grabowski, C. Syrowatka, C. Pélabon, and T. F. Hansen. 2024. Antler Allometry, the Irish Elk and Gould Revisited. Evolutionary Biology Pages 1–17.

Tsuboi, M., B. T. Kopperud, C. Syrowatka, M. Grabowski, K. L. Voje, C. Pélabon, and T. F. Hansen. 2020. Measuring complex morphological traits with 3D photogrammetry: a case study with deer antlers. Evolutionary Biology 47:175–186.

Tufte, E. 2001. The visual display of quantitative information. Cheshire: Graphic Press.

Upham, N., B. C, W. J, M. Becker, H. Handika, J. Zijlstra, and D. Huckaby. 2024. Mammal diversity database (version 1.13). `https://doi.org/10.5281/zenodo.10595931` accessed: 2024-11-27.

Urbanek, S. 2013. png: Read and write PNG images. R package version 0.1-8.

Uribe-Convers, S. and D. C. Tank. 2015. Shifts in diversification rates linked to biogeographic movement into new areas: An example of a recent radiation in the andes. American Journal of Botany 102:1854–1869.

Uyeda, J. and L. Harmon. 2020. treeplyr: 'dplyr' Functionality for Matched Tree and Data Objects. R package version 0.1.10.

Uyeda, J. C. and L. J. Harmon. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Systematic Biology 63:902–918.

Varga, T., K. Krizsán, C. Földi, B. Dima, M. Sánchez-García, S. Sánchez-Ramírez, G. J. Szöllősi, J. G. Szarkándi, V. Papp, L. Albert, et al. 2019. Megaphylogeny resolves global patterns of mushroom evolution. Nature Ecology & Evolution 3:668–678.

Vasconcelos, T., B. C. O'Meara, and J. M. Beaulieu. 2022. A flexible method for estimating tip diversification rates across a range of speciation and extinction scenarios. Evolution 76:1420–1433.

Vasconcelos, T. N., C. E. Proença, B. Ahmad, D. S. Aguilar, R. Aguilar, B. S. Amorim, K. Campbell, I. R. Costa, P. S. De-Carvalho, J. E. Faria, et al. 2017. Myrteae phylogeny,

calibration, biogeography and diversification patterns: increased understanding in the most species rich tribe of Myrtaceae. Molecular Phylogenetics and Evolution 109:113–137.

Vaughan, T. G. 2017. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. Bioinformatics 33:2392–2394.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, et al. 2001. The sequence of the human genome. science 291:1304–1351.

Veron, P., J. Andréoletti, T. Giraud, and H. Morlon. 2025. Speciation completion rates have limited impact on macroevolutionary diversification. Philosophical Transactions B 380:20230317.

Volz, E. M., K. Koelle, and T. Bedford. 2013. Viral phylodynamics. PLoS Computational Biology 9:e1002947.

Vos, R. and A. Mooers. 2006. A new dated supertree of the primates. *in* Inferring large phylogenies: the big tree problem (R Vos, Phd thesis). Simon Fraser University, Burnaby, British Columbia.

Wald, A. 1949. Note on the consistency of the maximum likelihood estimate. The Annals of Mathematical Statistics 20:595–601.

Waller, J. T. and E. I. Svensson. 2017. Body size evolution in an old insect order: no evidence for Cope's Rule in spite of fitness benefits of large size. Evolution 71:2178–2193.

Wan, Y., H. R. Schwaninger, A. M. Baldo, J. A. Labate, G.-Y. Zhong, and C. J. Simon. 2013. A phylogenetic analysis of the grape genus (*Vitis* l.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. BMC Evolutionary Biology 13:1–20.

Wang, L.-G., T. T.-Y. Lam, S. Xu, Z. Dai, L. Zhou, T. Feng, P. Guo, C. W. Dunn, B. R. Jones, T. Bradley, et al. 2020. treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Molecular Biology and Evolution 37:599–603.

Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7:256–276.

Weir, J. T. and D. Schluter. 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. Science 315:1574–1576.

Wickham, H. 2007. Reshaping data with the reshape package. Journal of Statistical Software 21:1–20.

Wickham, H. 2011. The split-apply-combine strategy for data analysis. Journal of Statistical Software 40:1–29.

Wickham, H. 2012. reshape2: Flexibly reshape data: a reboot of the reshape package. R package version 1.4.4.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H. 2021. tidyr: Tidy Messy Data. R package version 1.3.1.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, et al. 2019. Welcome to the Tidyverse. Journal of Open Source Software 4:1686.

Wickham, H., R. François, L. Henry, and K. Müller. 2021. dplyr: A Grammar of Data Manipulation. R package version 1.1.4.

Wickham, H. and J. Hester. 2020. readr: Read Rectangular Text Data. R package version 1.4.0.

Wilke, C. O. 2016. cowplot: streamlined plot theme and plot annotations for 'ggplot2'. R package version 1.1.3.

Yang, Z. 2014. Molecular Evolution: A Statistical Approach. Oxford University Press.

Yang, Z., S. Kumar, and M. Nei. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650.

Yu, G. 2020. ggimage: Use Image in 'ggplot2'. R package version 0.3.3.

Yu, G. 2021a. ggplotify: Convert Plot to 'grob' or 'ggplot' Object. R package version 0.1.2.

Yu, G. 2021b. tidytree: A Tidy Tool for Phylogenetic Tree Data Manipulation. R package version 0.4.6.

Yu, G., T. T.-Y. Lam, H. Zhu, and Y. Guan. 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. Molecular Biology and Evolution 35:3041–3043.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8:28–36.

Yule, G. U. 1924. II.—A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S. Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character 213:21–87.

Zenil-Ferguson, R., J. G. Burleigh, W. A. Freyman, B. Igić, I. Mayrose, and E. E. Goldberg. 2019. Interaction among ploidy, breeding system and lineage diversification. New Phytologist 224:1252–1265.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist. 2016. Total-evidence dating under the fossilized birth-death process. Systematic Biology 65:228–249.

Zheng, Y. and J. J. Wiens. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Molecular Phylogenetics and Evolution 94:537–547.