8-2022

# Hiding in plain sight: accounting for rate heterogeneity in trait evolution models

James Boyko
*University of Arkansas, Fayetteville*

Hiding in plain sight: accounting for rate heterogeneity in trait evolution models

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Biological Sciences

by

James Boyko
University of Toronto
Bachelor of Science, 2015
University of Toronto
Master of Science, 2017

August 2022
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Jeremy Beaulieu, Ph.D.
Dissertation Director

_____
Andrew Alverson, Ph.D.
Committee Member

_____
Adam Siepelski, Ph.D.
Committee Member

_____
Brian O'Meara, Ph.D.
Committee Member (ex. offico)

_____
David McNabb, Ph.D.
Committee Member

## ABSTRACT

Within the last four decades, phylogenetic comparative methods have become the defacto method of analysis for comparative biologists. The availability of high-quality comparative datasets has been matched by an explosion of possible phylogenetic models. In large part, the efforts to increase the realism of phylogenetic comparative methods has been successful as evidenced by their widespread use. To this extensive literature, my contributions are modest. I have focused my dissertation work on two main themes. First, most phenotypic evolution is not independent of other phenotypes. Changes in a particular character may influence changes in another and modeling these characters in isolation can mislead our inferences. Second, evolutionary change is heterogeneous. Not all species are going to change in the same way at all times and failing to account for that will mislead our inferences. The intersection of these two themes, character dependence and rate heterogeneity, is more natural than it may first appear. This dissertation has four chapters addressing various issues in current phylogenetic comparative methods. In Chapter I, I extend discrete character models to allow for any number of characters with any number of observed or hidden states. In Chapter II, I apply hidden Markov models to the issue of false correlation between discrete character evolution. I demonstrate that allowing for character independent rate heterogeneity through the application of hidden Markov models, is one way to account for this statistical bias. In Chapter III, I develop a new model called hOUwie which detects correlation between discrete and continuous characters and estimates their joint evolution. In Chapter IV, I apply the hOUwie model to 33 clades of angiosperms and attempt to understand the evolutionary patterns of plant life history as it relates to climatic variation.

## LIST OF PUBLISHED PAPERS

**Chapter I:**
Boyko, J. D. & Beaulieu, J. M. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. *Methods Ecology and Evolution,* **12**, 468–478.

**Chapter II:**
Boyko, J.D. & Beaulieu, J.M. 2022. Reducing the biases in false correlations between discrete characters. *Systematic Biology*. In revision.

**Chapter III:**
Boyko, J.D., O'Meara, B.C., & Beaulieu, J.M. 2022. Jointly Modeling the Evolution of Discrete and Continuous Traits. *Evolution*. In prep.

**Chapter IV:**
Boyko J.D., Hagen E.R., Beaulieu J.M, &Vasconcelos T.N.C. Long-term responses of life-history strategies to climatic variability in flowering plants. *New Phytologist*. Invited paper, In prep.

## ACKNOWLEDGEMENTS

Academia tends to be a profession of high passionate peaks and deep emotional valleys and without the love and commiseration of my friends and family, the journey here would have been far more difficult. I would like to thank the members of the Beaulieu lab, past and present, for creating a wonderful environment to learn and grow. I'd especially like to thank Daniel Caetano, for his outstanding mentorship in the early days of my PhD. My committee members: Adam Siepielski, Andrew Alverson, and Brian O'Meara have provided invaluable mentorship and guidance which has left me a more thoughtful, thorough, and *pragmatic* scientist. I would like to thank my closest friends over the last 5 years: Conner, Eric, Peter, Simon, and Zach for their constant support and gentle ribbing. It is in their company that I was most able to relax and enjoy the beautiful state of Arkansas (this excludes 17 weeks of commercial free football which quickly became some of the most unpleasant moments of my life). I'd also like to thank my loving and caring parents, who drove 17 hours from Toronto (twice) to help me establish an incredibly comfortable base of operations here in Arkansas. I'd like to thank Thais for her love and support. Although the days writing this dissertation were often long and difficult, in her company I was able to write this dissertation at complete peace and I am so excited for our future together. Finally, I'd like to thank Jeremy Beaulieu. He has been a thoughtful, caring, and supportive mentor throughout my entire PhD. The highest compliment I can pay was that throughout all the ups and downs of my life in Arkansas, after failed ideas or sleepless nights, I never doubted joining the Beaulieu lab was the correct decision. If the tape of life were replayed, I'd hope to find myself here again.

*Dedicated to nonna, nonno, babushka, dedushka, and zia.*

*May they rest in peace.*

## TABLE OF CONTENTS

# INTRODUCTION

*It is abundantly evident that rates of evolution vary. They vary greatly from group to group, and even among closely related lineages there may be strikingly different rates. Differences in rates of evolution [...] are among the reasons for the great diversity of organisms on the earth.*

- *Simpson (1953)*

Within the last four decades, phylogenetic comparative methods have transitioned from a simple, but arguably infeasible tool for correcting non-independence (Felsenstein 1985, 2004; Huey et al. 2019) to the defacto method of analysis for comparative biologists (O'Meara 2012; Cooper et al. 2016; Huey et al. 2019). The issue of infeasibility, due to a lack of access to well resolved phylogenies, was quickly overcome by the flood of new molecular data and techniques for inferring phylogenies. In fact, so well resolved is this issue, that the amount and quality of molecular data is producing phylogenetic trees of such large size that new algorithms for calculating the likelihood of complex models are consistently being developed (e.g., (Pupko et al. 2000; Freckleton 2012; Ho and Ané 2014; Irvahn and Minin 2014; Hiscott et al. 2016; Louca and Pennell 2020; Mitov et al. 2020). The availability of high-quality phylogenies and comparative datasets have been matched by an explosion of possible phylogenetic models, often being developed with the hope of inferring details of the macroevolutionary process rather than only correcting statistical non-independence (e.g., Hansen 1997; Galtier 2001; Butler and King 2004; Housworth et al. 2004; Pagel and Meade 2006; Bokma 2008; Hansen et al. 2008; Hadfield and Nakagawa 2010; Eastman et al. 2011, 2013; Bartoszek et al. 2012; Freckleton 2012; Hansen and Bartoszek 2012; O'Meara 2012; Revell and Reynolds 2012; Slater et al. 2012; Thomas and

Freckleton 2012; Beaulieu et al. 2013; Ingram and Mahler 2013; Höhna et al. 2014, 2016; Revell 2014, 2021; Cybis et al. 2015; Beaulieu and O'Meara 2016; Bastide et al. 2017, 2018; Caetano and Harmon 2017, 2019; Mitov et al. 2019, 2020; Blomberg et al. 2020; Nürk et al. 2020; Boyko and Beaulieu 2021, 2022; Jhwueng 2021). In large part, the efforts to increase the realism of phylogenetic comparative methods has been successful as evidenced by their widespread use.

To this extensive literature, my contributions are modest. I have focused my dissertation work on two main themes. First, most phenotypic evolution is not independent of other phenotypes. Changes in a particular character may influence changes in another and modeling these characters in isolation can mislead our inferences. Second, evolutionary change is heterogeneous. Not all species are going to always change in the same way and failing to account for that will mislead our inferences. The intersection of these two themes, character dependence and rate heterogeneity, is more natural than it may first appear. This claim is particularly true in the context of phylogenetic comparative methods (PCMs), the primary subject of this manuscript.

Consider how we can infer one event influencing another. We may expect that the outcome or observation of event X (whether the outcome is X=0 or X=1) changes the probability of event Y (the chances of Y=0 or Y=1 occurring). For example, the probability of crossing a busy street may differ substantially from crossing a street with no traffic. There is a correlation or dependence between crossing a street (X = 0 = not crossing) and the amount of traffic (Y = 1 = a lot of traffic). For evolutionary biologists, dependencies are of particular interest since they often represent repeated outcomes of evolution and can give insights into the underlying evolutionary process. For example, repeatedly observing and quantifying a correlation between habitat type and morphology may indicate convergent evolution and give insights into the

processes which underly bursts of speciation when lineages are exposed to novel environments (Schluter 2000). In PCMs, we test for correlation by examining differences in rates – specifically, the difference of a rate in the presence of one background state versus another. Are the speciation rates of lineages with red flowers greater than those with blue flowers? Does a transition from annual to perennial occur more quickly in lineages which are already woody? Do rates of evolution depend on an animal's body size? Each of these questions poses that a character's particular state correlates not necessarily with the state of a second character, but with the rates of change within that second character. This is a subtle, but crucially important difference, because evolutionary change is heterogeneous as a rule (Gould 1989).

Of course, rates are not data. Rather, they follow from data when combined with a model of how phenotypes evolve (Bookstein 1987; Cooper and Purvis 2009). This makes choosing a reasonable evolutionary model, or set of models, important. The macroevolutionary models dealt with in this dissertation are some of the most common used in the field today. The models require as input an ultrametric and time calibrated phylogeny as well as a matrix of discrete or continuous characters measured at the species level. For example, a dataset could comprise a set of average limb lengths of individual *Anolis* species as well as their discretized ecomorph categorization (Losos 1990, 1992; Losos et al. 2006; Ingram and Mahler 2013). This dataset, in combination with a phylogenetic depiction of the relationships between species, are the foundation of making inferences about macroevolutionary patterns. However, any inference made from the data will only be as good as the model set being evaluated.

Our goal in phylogenetic comparative modeling is to describe the distribution of characters that we see along the tips of the phylogeny. There are two main classes of character that we will be discussing: discrete and continuous. To model the evolution of a single

continuous character, one common model is the Ornstein-Uhlenbeck (OU) model (Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014). Conceptually, this model combines the stochastic evolution of a trait through time with a deterministic component that models the tendency for a trait to evolve towards an "optimum." The optimum can represent the set of "selective regimes", "regimes", or Simpson's "adaptive zones" (Cressler et al. 2015), though it is consistent with a variety of true underlying microevolutionary models (Hansen 2014). An early attempt to model discrete character evolution was a stochastic model of correlated trait evolution (Pagel 1994). The model Pagel (1994) proposed was based on continuous-time Markov models of nucleic acid substitution (e.g. Felsenstein, 1981; Jukes and Cantor, 1969; Kimura, 1980), but applied to estimate transition rates in pairs of binary characters. Finally, phylogenetic models can often be distinguished by the number of discrete traits and characters being analyzed. For example, BiSSE (Maddison et al. 2007) and HiSSE (Beaulieu and O'Meara 2016) can be distinguished from MuHiSSE (Nakov et al. 2019) and SecSSE (Herrera-Alsina et al. 2019) by the former two models allowing only binary character data and the latter two allowing for both multiple traits and more than two characters per trait. Once the traits have been chosen and codified into character states, the observations are seen by the model as a sequence of data values associated with the tips of a phylogeny. These data values and the position in the phylogeny are all we know about the species. In fact, most of our information for inferring rates lies in explaining the distribution of these traits as they relate to other species in the phylogeny.

So, if we know that testing for correlation is done by comparing rates, then the natural rate heterogeneity of the evolutionary process may become an issue for our macroevolutionary models (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015). Decoupling rate variation

due to some focal character of interest will always need to be contrasted against rate heterogeneity due to some unobserved variable (Beaulieu and O'Meara 2016; Caetano et al. 2018; Boyko and Beaulieu 2022). Furthermore, in the opinion of the author, declaring a correlation between a background state and a heterogenous process is a very old and consequential error. Take something heterogenous, for example human personality, and attempt to explain it (Allport 1962; Lamiell 1987; Silvia 2006). One of the most common explanations of differences in personality is to relate individual idiosyncrasies with the position of stars and moons – a field known as astrology (Zarka 2009). If our only explanations for human behavior was planetary positions, then one would have no choice but to embrace the explanation because it is a given that human personalities differ. A lack of alternative explanations can mislead our understanding because one may view any explanation as better than none. Furthermore, if an explanation includes differences, it will, at some point, correctly predict and describe the heterogenous process. This is particularly true if alternative explanations rely on an assumption that human personality does not differ. Returning to the field of comparative biology, this problem has been encountered and described in two of the most commonly used phylogenetic comparative models: state-dependent speciation and extinction (SSE) models and Markov models testing for correlation between discrete characters (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015). In both cases, the issue lies in null explanations requiring a non-heterogenous process to explain the data and then being contrasted to a model in which a focal character is the source of rate variation.

The way I have attempted to address this problem, and account for character independent rate heterogeneity, is through the use of Hidden Markov models (HMMs). HMMs are a simplified way that we can bring realism to our modeling while also making statistically

consistent and unbiased estimates of evolutionary parameters. Hidden Markov models (HMMs) are models in which the distribution that generates an observation depends on the state of an underlying and unobserved Markov process (Zucchini et al. 2017). In phylogenetic comparative methods, HMMs are often discussed as unknown characters whose presence causes heterogeneity in the observed transition rates, diversification rates, or both (Beaulieu et al. 2013; Pennell et al. 2014; Revell 2014; Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Beaulieu and O'Meara 2016; Brown and Thomson 2018; Caetano et al. 2018; Folk et al. 2018; Ng and Smith 2018; Uyeda et al. 2018; Otero et al. 2019). When a Markov model is said to be hidden, it describes states which are not the same as the observations. Instead, an HMM is a way to link different processes that can help explain heterogeneity in the distribution of observations. HMMs are appropriate when a single transition rate matrix or diversification rate are not the same for all lineages. To address this, hidden Markov models link two or more processes. They will only be favored if there is signal in the data for these two or more separate processes.

This dissertation has four chapters addressing various issues in current phylogenetic comparative methods. In Chapter I, I extend discrete character models to allow for any number of characters with any number of observed or hidden states. This addresses the issue that phenotypic evolution, even when simplified to a discrete character, is better understood as the confluence of several characters evolving together, rather than a single character evolving independently. Furthermore, I demonstrate the some of the advantages of increasing state space from an information theoretic view. In Chapter II, I apply hidden Markov models to the issue of false correlation between discrete character evolution. As we have discussed, comparative biologists are often interested in whether to discrete characters are correlated with each other as a significant and repeated dependent relationship may give insight into the underlying evolutionary

process. However, it has been thrown that several commonly used comparative methods are susceptible to false correlations. In this Chapter, I demonstrate that allowing for character independent rate heterogeneity through the application of hidden Markov models, is one way to account for this statistical bias. In Chapter III, I develop a new model for linking discrete and continuous character evolution. The model called hOUwie, attempts to detect correlation between discrete and continuous characters and estimates their joint evolution. Furthermore, this model is developed with the issues of the previous chapter in mind and therefore allows for character independent rate heterogeneity as an alternative explanation. In Chapter IV, I apply the hOUwie model to 33 clades of angiosperms and attempt to understand the evolutionary patterns of plant life history. It has been theorized that certain climatic variables, such as precipitation and temperature, can explain the distribution of annuals and perennials globally. However, previous studies have been contradictory with results depending on the specific clade or climatic variable being analyzed. Here we demonstrate that a consistent driver of annual life history is the maximum temperature of the hottest month. Furthermore, we show that some of the most commonly used model systems for life history evolution have biased our perceptions of the evolutionary process and in our analyses were more often the exception than the rule.

## References

Allport G.W. 1962. The general and the unique in psychological science. Personality: Critical Concepts in Psychology.:274–288.

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. Journal of Theoretical Biology. 314:204–215.

Bastide P., Ané C., Robin S., Mariadassou M. 2018. Inference of Adaptive Shifts for Multivariate Correlated Traits. Systematic Biology. 67:662–680.

Bastide P., Mariadassou M., Robin S. 2017. Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 79:1067–1093.

Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling Stabilizing Selection: Expanding the Ornstein–Uhlenbeck Model of Adaptive Evolution. Evolution. 66:2369–2383.

Beaulieu J.M., O'Meara B.C. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Syst Biol. 65:583–601.

Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. Syst Biol. 62:725–737.

Blomberg S.P., Rathnayake S.I., Moreau C.M. 2020. Beyond Brownian Motion and the Ornstein-Uhlenbeck Process: Stochastic Diffusion Models for the Evolution of Quantitative Characters. The American Naturalist. 195:145–165.

Bokma F. 2008. Detection of "Punctuated Equilibrium" by Bayesian Estimation of Speciation and Extinction Rates, Ancestral Character States, and Rates of Anagenetic and Cladogenetic Evolution on a Molecular Phylogeny. Evolution. 62:2718–2726.

Bookstein F.L. 1987. Random walk and the existence of evolutionary rates. Paleobiology. 13:446–464.

Boyko J., Beaulieu J. 2022. A potential solution to the unresolved challenge of false correlation between discrete characters. .

Boyko J.D., Beaulieu J.M. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. Methods Ecol Evol. 12:468–478.

Brown J.M., Thomson R.C. 2018. Evaluating Model Performance in Evolutionary Biology. Annual Review of Ecology, Evolution, and Systematics. 49:95–114.

Butler M.A., King A.A. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist. 164:683–695.

Caetano D.S., Harmon L.J. 2017. ratematrix: An R package for studying evolutionary integration among several traits on phylogenetic trees. Methods in Ecology and Evolution. 8:1920–1927.

Caetano D.S., Harmon L.J. 2019. Estimating Correlated Rates of Trait Evolution with Uncertainty. Systematic Biology. 68:412–429.

Caetano D.S., O'Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models: HMM and the adequacy of SSE models. Evolution. 72:2308–2324.

Cooper N., Purvis A. 2009. What factors shape rates of phenotypic evolution? A comparative study of cranial morphology of four mammalian clades. Journal of Evolutionary Biology. 22:1024–1035.

Cooper N., Thomas G.H., FitzJohn R.G. 2016. Shedding light on the 'dark side' of phylogenetic comparative methods. Methods in Ecology and Evolution. 7:693–699.

Cybis G.B., Sinsheimer J.S., Bedford T., Mather A.E., Lemey P., Suchard M.A. 2015. ASSESSING PHENOTYPIC CORRELATION THROUGH THE MULTIVARIATE PHYLOGENETIC LATENT LIABILITY MODEL. Ann Appl Stat. 9:969–991.

Eastman J.M., Alfaro M.E., Joyce P., Hipp A.L., Harmon L.J. 2011. A Novel Comparative Method for Identifying Shifts in the Rate of Character Evolution on Trees. Evolution. 65:3578–3589.

Eastman J.M., Wegmann D., Leuenberger C., Harmon L.J. 2013. Simpsonian "Evolution by Jumps" in an Adaptive Radiation of Anolis Lizards. arXiv:1305.4216 [q-bio].

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. Am. Nat. 125:1–15.

Felsenstein J. 2004. Inferring phylogenies. Sinauer associates Sunderland, MA.

Folk R.A., Soltis P.S., Soltis D.E., Guralnick R. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. American Journal of Botany.:364–375.

Freckleton R.P. 2012. Fast likelihood calculations for comparative analyses. Methods in Ecology and Evolution. 3:940–947.

Galtier N. 2001. Maximum-Likelihood Phylogenetic Analysis Under a Covarion-like Model. Molecular Biology and Evolution. 18:866–873.

Gould S. 1989. Wonderful Life: The Burgess Shale and then Nature of History. WW Norton, New York.

Hadfield J.D., Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. Journal of Evolutionary Biology. 23:494–508.

Hansen T.F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. Evolution. 51:1341–1351.

Hansen T.F. 2014. Use and Misuse of Comparative Methods in the Study of Adaptation. In: Garamszegi L.Z., editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 351–379.

Hansen T.F., Bartoszek K. 2012. Interpreting the Evolutionary Regression: The Interplay Between Observational and Biological Errors in Phylogenetic Comparative Studies. Systematic Biology. 61:413–425.

Hansen T.F., Pienaar J., Orzack S.H. 2008. A Comparative Method for Studying Adaptation to a Randomly Evolving Environment. Evolution. 62:1965–1977.

Herrera-Alsina L., van Els P., Etienne R.S. 2019. Detecting the Dependence of Diversification on Multiple Traits from Phylogenetic Trees and Trait Data. Syst Biol. 68:317–328.

Hiscott G., Fox C., Parry M., Bryant D. 2016. Efficient Recycled Algorithms for Quantitative Trait Models on Phylogenies. Genome Biology and Evolution. 8:1338–1350.

Ho L. si, Ané C. 2014. A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. Syst Biol. 63:397–408.

Höhna S., Heath T.A., Boussau B., Landis M.J., Ronquist F., Huelsenbeck J.P. 2014. Probabilistic Graphical Model Representation in Phylogenetics. Syst Biol. 63:753–771.

Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. Syst Biol. 65:726–736.

Housworth E.A., Martins E.P., Lynch M. 2004. The Phylogenetic Mixed Model. The American Naturalist. 163:84–96.

Huey R.B., Garland T., Turelli M. 2019. Revisiting a Key Innovation in Evolutionary Biology: Felsenstein's "Phylogenies and the Comparative Method." The American Naturalist.:000–000.

Ingram T., Mahler D.L. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods in Ecology and Evolution. 4:416–425.

Irvahn J., Minin V.N. 2014. Phylogenetic Stochastic Mapping Without Matrix Exponentiation. Journal of Computational Biology. 21:676–690.

Jhwueng D.-C. 2021. Two Gaussian Bridge Processes for Mapping Continuous Trait Evolution along Phylogenetic Trees. Mathematics. 9:1998.

Jukes T.H., Cantor C.R. 1969. Evolution of protein molecules. Mammalian protein metabolism. 3:132.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16:111–120.

Lamiell J.T. 1987. The psychology of personality: An epistemological inquiry. Columbia University Press.

Losos J.B. 1990. The evolution of form and function: Morphology and locomotor performance in West Indian Anolis lizards. Evolution. 44:1189–1203.

Losos J.B. 1992. The Evolution of Convergent Structure in Caribbean Anolis Communities. Syst Biol. 41:403–420.

Losos J.B., Glor R.E., Kolbe J.J., Nicholson K. 2006. Adaptation, speciation, and convergence: A hierarchical analysis of adaptive radiation in Caribbean Anolis lizards. Annals of the Missouri Botanical Garden. 93:24–33.

Louca S., Pennell M.W. 2020. A General and Efficient Algorithm for the Likelihood of Diversification and Discrete-Trait Evolutionary Models. Systematic Biology. 69:545–556.

Maddison W.P., FitzJohn R.G. 2015. The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. Syst Biol. 64:127–136.

Maddison W.P., Midford P.E., Otto S.P., Oakley T. 2007. Estimating a Binary Character's Effect on Speciation and Extinction. Syst Biol. 56:701–710.

Mitov V., Bartoszek K., Asimomitis G., Stadler T. 2020. Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. Theoretical Population Biology. 131:66–78.

Mitov V., Bartoszek K., Stadler T. 2019. Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models. Proceedings of the National Academy of Sciences. 116:16921–16926.

Nakov T., Beaulieu J.M., Alverson A.J. 2019. Diatoms diversify and turn over faster in freshwater than marine environments. Evolution. 73:2497–2511.

Ng J., Smith S.D. 2018. Why are red flowers so rare? Testing the macroevolutionary causes of tippiness. Journal of Evolutionary Biology. 31:1863–1875.

Nürk N.M., Linder H.P., Onstein R.E., Larcombe M.J., Hughes C.E., Piñeiro Fernández L., Schlüter P.M., Valente L., Beierkuhnlein C., Cutts V., Donoghue M.J., Edwards E.J., Field R., Flantua S.G.A., Higgins S.I., Jentsch A., Liede-Schumann S., Pirie M.D. 2020. Diversification in evolutionary arenas—Assessment and synthesis. Ecology and Evolution. 10:6163–6182.

O'Meara B.C. 2012. Evolutionary Inferences from Phylogenies: A Review of Methods. Annual Review of Ecology, Evolution, and Systematics. 43:267–285.

Otero A., Jiménez-Mejías P., Valcárcel V., Vargas P. 2019. Being in the right place at the right time? Parallel diversification bursts favored by the persistence of ancient epizoochorous traits and hidden factors in Cynoglossoideae. American Journal of Botany. 106:438–452.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. B: Biol. Sci. 255:37–45.

Pagel M., Meade A. 2006. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. Am. Nat. 167:808–825.

Pennell M.W., Harmon L.J., Uyeda J.C. 2014. Is there room for punctuated equilibrium in macroevolution? Trends in Ecology & Evolution. 29:23–32.

Pupko T., Pe I., Shamir R., Graur D. 2000. A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences. Mol Biol Evol. 17:890–896.

Rabosky D.L., Goldberg E.E. 2015. Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. Syst Biol. 64:340–355.

Revell L.J. 2014. Ancestral Character Estimation Under the Threshold Model from Quantitative Genetics. Evolution. 68:743–759.

Revell L.J. 2021. A variable-rate quantitative trait evolution model using penalized-likelihood. PeerJ. 9:e11997.

Revell L.J., Reynolds R.G. 2012. A New Bayesian Method for Fitting Evolutionary Models to Comparative Data with Intraspecific Variation. Evolution. 66:2697–2707.

Schluter D. 2000. The ecology of adaptive radiation. OUP Oxford.

Silvia P.J. 2006. Exploring the psychology of interest. Oxford ; New York: Oxford University Press.

Simpson G.G. 1953. The major features of evolution. The Major Features of Evolution. Columbia University Press.

Slater G.J., Harmon L.J., Wegmann D., Joyce P., Revell L.J., Alfaro M.E. 2012. Fitting Models of Continuous Trait Evolution to Incompletely Sampled Comparative Data Using Approximate Bayesian Computation. Evolution. 66:752–762.

Thomas G.H., Freckleton R.P. 2012. MOTMOT: models of trait macroevolution on trees. Methods in Ecology and Evolution. 3:145–151.

Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst Biol. 67:1091–1109.

Zarka P. 2009. Astronomy and astrology. Proceedings of the International Astronomical Union. 5:420–425.

Zucchini W., MacDonald I.L., Langrock R. 2017. Hidden Markov models for time series: an introduction using R. Chapman and Hall/CRC.

# CHAPTER I

## Generalized hidden Markov models for phylogenetic comparative datasets

James Boyko and Jeremy Beaulieu


### Abstract

Hidden Markov models (HMM) have emerged as an important tool for understanding the evolution of characters that take on discrete states. Their flexibility and biological sensibility make them appealing for many phylogenetic comparative applications. Previously available packages placed unnecessary limits on the number of observed and hidden states that can be considered when estimating transition rates and inferring ancestral states on a phylogeny. To address these issues, we expanded the capabilities of the R package corHMM to handle *n*-state and *n*-character problems and provide users with a streamlined set of functions to create custom HMMs for any biological question of arbitrary complexity. We show that increasing the number of observed states increases the accuracy of ancestral state reconstruction. We also explore the conditions for when an HMM is most effective, finding that an HMM is an appropriate model when the degree of rate heterogeneity is moderate to high. Finally, we demonstrate the importance of these generalizations by reconstructing the phyllotaxy of the ancestral angiosperm flower. Partially contradicting previous results, we find the most likely state to be a whorled perianth, whorled androecium, whorled gynoecium. The difference between our analysis and previous studies was that our modeling explicitly allowed for the correlated evolution of several flower characters.

**Introduction**

Hidden Markov models (HMMs) are important centerpieces in many biological applications (Eddy, 2004; Yang Lou, 2017). They provide a natural framework for comparative biologists, particularly for relaxing assumptions about homogeneous evolution through time and across taxa without vastly increasing the number of parameters (e.g., Felsenstein & Churchill, 1996; Galtier, 2001; Penny, McComish, Charleston, & Hendy, 2001; Beaulieu, O'Meara, & Donoghue, 2013; Beaulieu & O'Meara, 2016). For instance, simple models of binary character evolution make sense for small, young clades, because a single set of transition rates seems like a reasonable assumption. However, homogeneous rates are unlikely to explain the evolution of the same character across a much larger and older clade in which transition rates may differ dramatically among subclades, perhaps due to correlations with traits that were not included in the model. This observation was the motivation for the development of the hidden rate model (HRM) of Beaulieu *et al.* (2013), which uses a hidden Markov approach to objectively locate regions of a phylogeny where hidden factors have either promoted or constrained the evolutionary process for a binary character.

Within comparative biology, HMMs have been applied as both standalone models (Beaulieu et al., 2013) and in combination with other phylogenetic models (e.g., hidden state-dependent speciation and extinction models, Beaulieu & O'Meara, 2016). Hidden Markov models can be used to address many problems in comparative biology (Siepel & Haussler, 2005) and their flexibility allows biologists to create models tailored to their specific hypotheses. However, previous implementations of HMMs for comparative methods have placed limitations on the number of observed and hidden states. For instance, the implementation of the HRM model of Beaulieu et al. (2013) is restricted only to the analysis of binary characters. There is no

mathematical basis for limiting the number of observed states or hidden states in an HMM, and such constraints necessitate a simplification of datasets and candidate models.

Here we describe a new version of corHMM that implements *n*-state HMMs. This does not require new algorithms or a different likelihood function. Instead, we optimized and generalized existing code so users can create custom HMMs for any biological question of arbitrary complexity. We have also added a number of "quality of life" improvements that make corHMM much easier to use and interpret, including an implementation of stochastic character mapping (simmap; Bollback, 2006). Additionally, we demonstrate the effectiveness of HMMs to identify rate heterogeneity when it is present, and we outline the informational advantages of increasing the number of observed and hidden states in discrete character data sets. Finally, to demonstrate the importance of this generalization, we apply corHMM to reconstruct the phyllotaxy of the ancestral angiosperm flower.

## Methods

### *Generalizing HMMs*

From a technical standpoint, hidden Markov models have a hierarchical structure that can be broken down into two components: a "state-dependent process" (Fig. 1a,b) and an unobserved "parameter process" (Fig. 1c)(Zucchini, MacDonald, & Langrock, 2017). In comparative biology, for characters that take on discrete states the standard "state-dependent process" is a continuous-time Markov chain with finite state-space (CTMC-FS). The benefit of a Markov model is its simplicity — to calculate the probabilities of observed discrete states at the tips of a phylogeny all that is required is a tree, a transition model describing transitions among a set of observed states, and frequencies at the root (O'Meara, 2012; Fig. 1a,b). The observed states

could be any discretized trait such as presence or absence of extrafloral nectaries (Marazzi et al., 2012), woody or herbaceous growth habit (Beaulieu et al., 2013), or diet state across all animals (Román-Palacios, Scholl, & Wiens, 2019). However, a simple Markov process that assumes homogeneity through time and across taxa is often not adequate to capture the variation of real datasets (e.g. Beaulieu *et al.,* 2013). Under an HMM, observations are generated by a given state-dependent process, which in turn depends on the state of the parameter process. In other words, the observed data are the product of several processes occurring in different parts of a phylogeny and the parameter process is way of linking them. It is initially unknown what the parameter process corresponds to biologically, hence the moniker "hidden" state. Nevertheless, the information for detecting hidden states comes from the differences in how the observed states change. As long as the transitions between observed states of different lineages are more adequately described by several Markov processes rather than a single process, there will be information to detect hidden states (see *3.1 Performance in Simulations*).

The likelihood of any HMM is obtained by maximizing the standard likelihood formula, $L = P(D|\mathbf{Q}, T)$, for observing character states , $D$, across a set of extant taxa, given the continuous-time Markov model **Q**, and a fixed topology with a set of branch lengths (denoted by $T$). For a binary character, **Q** is a 2×2 transition matrix representing the transition rates, whose entries define transitions between the character states, 0 and 1. To form an HMM, we expand **Q** to accommodate both observed and hidden states. Formally, the HMM can be generalized to include any number of observed states (e.g., 0, 1, 2), and hidden states (e.g., A, B, C). Following Beaulieu and O'Meara (2016), the state space is defined as $o$ being the index of the observed state, $o \in 0,1,\ldots,\alpha$, and $h$ as the index of the hidden state, $h \in A, B, \ldots, \beta$. Thus, a given model will have, in general, $|o| \times |h|$ states. In corHMM, the model **Q** is defined by amalgamating each

of the state-dependent processes and the parameter process specified in the model. For example, if we have state-dependent matrices, $\boldsymbol{R}$,

$$\boldsymbol{R}_1 = \begin{pmatrix} - & 1 & 2 \\ 1 & - & 3 \\ 2 & 3 & - \end{pmatrix}, \ \boldsymbol{R}_2 = \begin{pmatrix} - & 5 & 0 \\ 4 & - & 7 \\ 0 & 6 & - \end{pmatrix},$$

that are related by a parameter-process $\boldsymbol{P}$,

$$\boldsymbol{P} = \begin{pmatrix} - & 9 \\ 8 & - \end{pmatrix} = \begin{pmatrix} - & r_{R1\to R2} \\ r_{R2\to R1} & - \end{pmatrix},$$

where entries $r_{R1\to R2}$ and $r_{R2\to R1}$ define transition rates between the state-dependent processes, we can extend Eq. 2 of Tarasov (2019) to amalgamate these processes,

$$\mathbf{Q} = A(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{P}) = \begin{pmatrix} - & 1 & 2 & 9 & 0 & 0 \\ 1 & - & 3 & 0 & 9 & 0 \\ 2 & 3 & - & 0 & 0 & 9 \\ 8 & 0 & 0 & - & 5 & 0 \\ 0 & 8 & 0 & 4 & - & 7 \\ 0 & 0 & 8 & 0 & 6 & - \end{pmatrix}.$$

This matrix can be understood as a block matrix where the diagonal blocks are the state-dependent processes $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, and the off-diagonal blocks are the parameter process, $\boldsymbol{P}$, that describe transitions between $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$. It is important to note that although we amalgamate the matrices above, it is still possible for users to modify the transitions between hidden states.

The general formulation of an HMM can easily be extended to examine the correlated evolution of multiple characters (Pagel, 1994). For example, consider a case of two binary characters where trait 1 defines the presence or absence of fleshiness of fruits, and trait 2 defines whether or not the fruits are animal-dispersed. At most there are four binary combinations of these characters (i.e., 00, 01, 10, and 11). But, it can also be coded as a single multistate character, where 1=dry fruits not dispersed by animals, 2=dry fruits dispersed by animals, 3=fleshy fruits not dispersed by animals, and 4=fleshy fruits dispersed by animals. Therefore, transforming binary combinations to multistate characters also applies for two characters with a

different number of observed states. In other words, one character could be binary (e.g., dry vs. fleshy fruit) and the other could be multistate (e.g., fruits dispersed mechanically, by wind, or by animal).

*Simulation Study*

We conducted a set of simulations to address two broad goals. First we tested whether there is an informational advantage to increasing the number of observed states by comparing two-state, three-state, and four-state datasets. Our second goal was to test the ability of hidden Markov models to detect varying degrees of rate heterogeneity. We then linked these goals together by testing whether HMMs can recover some of the informational content of unobserved characters through the use of hidden states. These simulations are in no way exhaustive, but represent a set of reasonable questions that many beginning users might have about the behavior of HMMs.

*Increasing the number of observed characters or states*

To test the behavior of two-state, three-state, and four-state datasets we relied on ancestral state reconstruction (ASR) at nodes. ASR is a widely-utilized feature of corHMM, and it is important to know the accuracy of multistate ancestral reconstructions. Additionally, using ancestral states gives us a direct means to compare models with different datasets. A 250-tip phylogeny was simulated (birth rate set to 1 event $Myr^{-1}$, and death rate of 0.5 events $Myr^{-1}$) to be used as a fixed tree with a root age of 12.46 Myr and mean branch length of 0.89 Myr. Datasets were simulated using transition rates sampled from a truncated normal distribution ($\mu = 1$, $\sigma = 0.5$), which were then scaled to have mean rates of 0.1, 1.0, or 10 transitions $Myr^{-1}$ by dividing the rate matrix by the sum of the diagonal and then multiplying by the desired scalar. This resulted in a range of evolutionary rates where the expected number of transitions ranged from ~21, 210, or 2100

transitions across the entire tree depending on the mean rate. It should also be noted that although these rates are of interest in a theoretical setting, they may not be representative of most empirical settings where transition rates can be orders of magnitude lower than we are using (see *3.2: Case study: reconstructing the ancestral angiosperm flower)*. For each transition model, 100 datasets were simulated. The transition rates of each dataset were then estimated and their maximum likelihood estimates were used to infer marginal probabilities of each character state across the tree. This procedure was repeated 10 times.

An underappreciated concern with evaluating models that differ in the number of observed states is that the probability of guessing the correct state without any additional information is simply $1/k$ states. This could, in theory, inflate the accuracy of datasets with fewer states even though the tip states themselves provide no information about the ancestral states when the rates are high (Schultz, Cocroft, & Churchill, 1996; Sober & Steel, 2011, 2014). To deal with this issue, we also calculated the mutual information, measured in bits, about ancestral states from each dataset and model (Cover & Thomas, 1991; Sober & Steel, 2011). Specifically, mutual information is a measure of how much ancestral state uncertainty is reduced by knowing the tip states (details of our derivation are given in Supplementary materials). The initial uncertainty, or unconditional entropy, is set by the model – given a model of evolution and no knowledge of the extant tips, how uncertain is the best guess of the ancestral states? The remaining uncertainty after ASR, or conditional entropy, is given by the combination of the model and the tip states – given the model of evolution and knowledge of the extant tips, how uncertain is the best guess of the ancestral states? It is important to note that information, just like ancestral state reconstruction, is highly correlated with the model of evolution, and thus any results related to information will take on the assumptions of the model.

We define information as the difference between the unconditional entropy of the node states, $H(X_v)$, and the entropy of the node states conditioned on the data, $H(X_v|X_h = D)$ (Cover & Thomas, 1991). The unconditional entropy of node $v$ is defined as:

$$H(X_v) = -\sum_{i=1}^{k} \pi[X_v = i] log_2(\pi[X_v = i]),$$

where $\pi[X_v = i]$ is the prior probability of a node taking a particular state. For the root, the prior depends on user choice, as there are several options (Yang, Kumar, & Nei, 1995; Pagel, 1999; FitzJohn, Maddison, & Otto, 2009). Here we assume the prior probability on the root node is the expected equilibrium frequency, $\pi$, which is calculated directly from the transition model by solving $\pi \boldsymbol{Q} = 0$. This aligns our expectation of the root node with all other internal nodes such that, in the absence of information from the tips, the probability of a particular state is assumed to be drawn from the equilibrium frequencies. In other words, the information of the tip states decreases as rates increase and, ultimately, the probability of a node state becomes completely determined by the model. We define the conditional entropy as:

$$H(X_v|X_h = D) = -\sum_{i=1}^{k} P[X_v = i|X_h = D] log_2(P[X_v = i\ |X_h = D]),$$

where $P[X_v = i|X_h = D]$ is the conditional probability that a node is fixed as being in state $i$ given the probability of observing the tip data (which is just the marginal probability of state *i*). In particular, we are interested in the average entropy of a node for all states $i \ldots k$, given we observe a particular dataset, $X_h = D$. Thus, the conditional entropy will vary by node, but the unconditional entropy is set by the model. To produce a measure of mutual information between the observations at the tips and estimates at internal nodes, we take the difference between the conditional entropy and the unconditional entropy and average across all nodes. However, the

unconditional entropies will be greater for datasets that include more states because unconditional entropy sets the upper limit of what is possible to learn. This alone could contribute to large informational differences between models with different numbers of observed states. Therefore, we also measure the proportion of maximum information gained ($\frac{mutual\ information}{unconditional\ entropy} \times 100\%$).

*Evaluating hidden Markov models*

We evaluated the ability to detect rate heterogeneity by simulating data under an HMM. As outlined above (see *2.1 Generalizing HMMs*), there are two major axes along which an HMM differs from standard Markov models. First, we varied the magnitude of the difference in the state-dependent process by simulating data under a model where there was: (1) no difference between the state-dependent processes ($\boldsymbol{R}_1$=$\boldsymbol{R}_2$), (2) a 2-fold difference in rates between the state-dependent processes (e.g. if $\boldsymbol{R}_1$'s mean rate was 1 $Myr^{-1}$, $\boldsymbol{R}_2$ mean rate would be 2 $Myr^{-1}$), (3) a 10-fold difference between the state-dependent processes, and (4) a covarion-like trait model in which within $\boldsymbol{R}_1$ all transitions occur freely, but for $\boldsymbol{R}_2$ all transition rates are zero, and evolution is essentially "turned off" (Penny et al., 2001). For all simulation scenarios, we set the parameter-process to have equal transition rates between state-dependent processes. In addition to examining ancestral state reconstruction at nodes, we also used the new makeSimmap to assess how well the model captures the expected number of character changes within and among all branches in the tree. For each of the 150 datasets simulated above, we evaluated 100 simmaps per model by counting the number of transitions for a given simmap.

Next, we tested the impact of the magnitude of the asymmetry in the underlying parameter-process. We simulated data where the state-dependent process always differed by 100-fold, but for the underlying parameter-process there was: (1) no difference in transition rate

($r_{R1\rightarrow R2}$=$r_{R2\rightarrow R1}$), (2) 1.5× faster transition rate to the slower rate class ($r_{R1\rightarrow R2}$>$r_{R2\rightarrow R1}$), (3) 2× faster transition rate to the slower rate class, (4) 10× faster transition rate to the slower rate class. For each of the models described, we used simulated 500-tip phylogeny with a root age of 15.43 Myr (birth rate set to 1 event $Myr^{-1}$, and death rate of 0.7 events $Myr^{-1}$).

Finally, we examined how much information is available when we allow for hidden states to be observed at the tips. We used the same data generated from simulations examining state-dependent differences, but this time we did not remove the hidden state. We then fit a Markov model to this full dataset and compared it to models in which the "second character" remained unobserved.

*Case study: reconstructing the ancestral angiosperm flower*

Understanding the origin of flowering plants is widely considered to be one of the most important goals of systematic botany. In a recent study, (Sauquet et al., 2017) compiled an extensive database of floral characteristics and attempted to reconstruct the morphology of the ancestral angiosperm flower. Sauquet et al. (2017) did not present a single answer for the ancestral state because there were several possible state combinations depending on the method used and uncertainty associated with each of those estimates. Nonetheless, their hypothetical diagram of the ancestral flower as having a whorled perianth, whorled androecium, and spiral gynoecium proved controversial. For example, Sokoloff, Remizowa, Bateman, & Rudall (2018) disputed this depiction of the ancestral phyllotaxy, suggesting that some of the characters were scored incorrectly and that it seemed improbable that state combinations that are rare in the data could be the ancestral state. Sokoloff et al. (2018) instead prefer the hypothesis that the ancestral flower was either entirely whorled or entirely spiraled. In response, Sauquet et al. (2018) rescored the disputed characters and reanalyzed the dataset using the same methods as the

original study. Their Bayesian analyses conformed to the predictions of Sokoloff et al. (2018), but remained highly uncertain overall. A limitation of the original study was the fact that "no comparative method exists yet to account for the potential correlation of more than two discrete characters" (Sauquet et al., 2017, but see Beaulieu & Donoghue, 2013). Given that flowers are highly integrated structures with potentially several developmental constraints the correlation between states seems a necessary prerequisite to study their evolution (Sauquet et al., 2017, 2018; Sokoloff et al., 2018). Treating the phyllotaxy of the perianth, androecium, and gynoecium as independent represents a major assumption with potentially large consequences on the ancestral state reconstruction. Indeed, Sauquet et al. (2017) conducted several pairwise correlation analyses and found that the phyllotactic patterns of these focal characters were strongly correlated.

*Worked example and methods*

We limited ourselves to including only the characters related to the phyllotaxy of the perianth, androecium, and gynoecium. Although it is possible to include other characters, given the corresponding increase of parameter space, we suspect that we would not have the power to accurately infer the model and ancestral states (O'Meara et al., 2016). The dataset of Sauquet et al. (2018) has several polymorphic species as well as species for which some of the tip states are unknown. Therefore, we analyzed three separate datasets: (1) no uncertain taxa (n = 295), (2) polymorphic species included (n = 297), and (3) all taxa included (n = 780). We treated the phyllotaxy of the perianth, androecium, and gynoecium as either "whorled" or "spiral" and polymorphic species are coded to have both states. The choice of dataset has major implications for model performance because corHMM will exclude state combinations that are absent from the dataset. However, the inclusion of either polymorphic or unknown states for taxa will expand

state space and thus increase the number of parameters that need to be estimated. Finally, we use the C series phylogeny of Sauquet et al. (2017) in which *Amborella* is constrained as the sister to all remaining angiosperms and *Monocotyledoneae*, *Ceratophyllaceae*, and *Eudicotyledoneae* are constrained to form a monophyletic group.

In our case, we have three data columns each with two observed states. Because this dataset contains two or more columns of trait information, each column is automatically interpreted as an evolving character. In these cases, corHMM will also automatically remove dual transitions from the model since that would constitute two or more evolutionary events (Pagel, 1994; Maddison, Midford, Otto, & Oakley, 2007). However, previous work has suggested that dual transitions are possible and likely in this system (Sauquet et al. 2018; Sokoloff et al. 2018). Thus, we include both models in which dual transitions are allowed and disallowed. Our analysis without hidden states include three different model structures: model="ER" (equal rates), model="SYM" (symmetric rates), and model="ARD" (all rates differ). The other options used (rate.cat=1 and nstarts=10) specify that no hidden states are to be used and that the maximum likelihood search will be performed 10 additional times with different initial parameters.

We also include a set of analyses in which hidden states are present because it is likely that there are unobserved characters which influence the evolution of the angiosperm flower. We include four hidden state models: *ER/ER*, *SYM/SYM*, *ARD/ARD*, and *ER/ARD*. Each of these models allows for the possibility of rate heterogeneity through the inclusion of a hidden state, however the state-dependent processes differ. In the *ER/ER* model all changes between states occur at the same rate within a state-dependent process, but the magnitude of change can depend on the underlying parameter process. The *SYM/SYM* model specifies that there are equal rates of

reversible change among character states. The *ARD/*ARD model specifies that there could be a bias towards a particular state, but this state may differ depending on whether the lineage is in $\boldsymbol{R}_1$ or $\boldsymbol{R}_2$. Finally, *ER/ARD* is a hybrid model which includes aspects of the equal rates model and all rates differ model.

## Results

### *Performance in Simulation*

Overall, the accuracy of an ancestral state reconstruction is predominately a function of the transition rates, but there are regions of parameter space where the number of states is influential (Fig. 2a). For example, all datasets generally inferred the correct ancestral state at low rates and datasets with more states performed better at intermediate rates. However, when viewed in terms of information, datasets that contained just two states showed detectable informational loss when compared to the three- and four-state datasets. In fact, across all scenarios — low, intermediate, and especially at the highest rates — datasets with more states consistently showed more informational gain relative to the maximum information content for a given number of states (Fig. 2b). We suspect this largely reflects the impacts of homoplasy when the number of character states are restricted in the model (Sanderson & Donoghue, 1989; Steel & Penny, 2005). This is not to say that more character states are always necessary for accurate ASR. Rather, we demonstrate that there are cases when additional characters or character states enhance the accuracy of an ancestral state reconstruction and those datasets have a signal of increased information.

*Evaluating hidden Markov models*

The accuracy of ancestral state estimation, based solely on reconstructing character states at nodes, appears largely unaffected by the inclusion of hidden states regardless of differences in the state-dependent processes (Fig. 3a). However, the amount of information gained depends on both the use of an HMM and the presence of strong differences between the state-dependent processes (Fig. 3b). These seemingly contradictory results are a consequence of how we calculate information. Model uncertainty certainly comes from the increase in parameters of HMMs relative to standard Mk models and manifests in both increased model complexity and an increased number of potential ancestral states. The increase in possible ancestral states results in a higher unconditional entropy, which can actually lead to greater informational content even when the ancestral states are not known with as much certainty as a Mk model. However, as we show in section 3.1.1 increasing the number of observed states improved ancestral state reconstruction accuracy, despite a greater number of estimated parameters, and so this does not solely account for the greater ancestral state accuracy of a Mk over an HMM. We suspect it is also due to datasets fit under an HMM have added uncertainty applied to the tips because it is initially unknown which hidden state a particular taxon occupies. The greater uncertainty at the tips is likely the reason why we observe Mk models outperforming HMMs in ancestral state reconstruction, and the greater uncertainty of the model is likely why HMMs are able to extract more information from a given dataset.

We found that when the generating model does not have state-dependent differences, the HMM does not pickup significant rate variation and resembles the character history implied by the standard Markov model (Fig. 4a-c). When there was no difference between the state-dependent processes, 2.6% of datasets had an AICc difference greater than 2 in support of an

HMM over a Markov model. Whereas 100% of datasets supported an HMM when data was simulated under a covarion-like model. These findings suggest that HMMs are supported in datasets where rate heterogeneity is present and this can be seen qualitatively through simmap reconstructions (Fig. 4d-f). We found little effect of altering the transition rate bias of the parameter process on either ancestral state reconstruction or information content.

Unsurprisingly, observing the "second character" states increased the amount of information (Fig. 5). However, as the state-dependent processes became more distinguishable, the informational gap between an HMM and including the observed second character decreased. In other words, when the evolution of an observed character changes across the phylogeny, an HMM is able to extract additional information from a dataset.

*Case study: reconstructing the ancestral angiosperm flower*

Across all three datasets our best supported model was *ER/ER*, a two rate class model with both state-dependent processes being equal rates (Table 1; Table S1-S3). Because our modeling set included a wide range of complexity ranging from 1 estimated rate (ER) to 114 estimated rates (ARD/ARD), we used AIC weights to calculate the model averaged ancestral states for datasets individually (Fig. 6). For all three datasets we find that an entirely whorled angiosperm flower is the most likely state. However, we found that the preferred ancestral state is highly variable and dependent on the model and the entirely whorled angiosperm flower is likely a reflection of the *ER/ER* model's high AIC weight within the set of tested models (Table 1; Table S1). For several of the models we found that the parameter estimates reached the upper limit of the transition rates allowed. This could be reflection of a lack of adequate data, too many unknown and polymorphic state combinations, and/or unrealistic models included in the set.

However, none of the transition rates estimated reached the upper limit for any of the best supported models.

**Discussion**

Hidden Markov models are an essential tool for inferring character states across phylogenies. The new version of corHMM, expands the array of potential uses of HMMs by increasing the number of possible character states and allowing users to construct custom models. In addition, we demonstrated the informational advantages of using hidden Markov models versus simple Markov models. Users interested in hypothesis-driven model construction are encouraged to read through the vignette associated with the corHMM package. This vignette fully describes how to use the package and includes several examples of how to take a biological hypothesis and codify it into an explicit HMM.

Information theory has mainly been discussed in a theoretical context and rarely used in practice to understand empirical trait evolution (Mossel, 2003; Mossel & Peres, 2003; Townsend & Naylor, 2007; Sober & Steel, 2011, 2014; Gascuel & Steel, 2014). In this paper, we have introduced a measure for the amount of information that the tips provide the nodes during ancestral state reconstruction. Two important caveats of this measure of information. First, the data and model are taken as fixed. These are not uncommon assumptions in phylogenetic comparative methods. For example, if one is to interpret an ancestral state reconstruction it comes with the implicit assumption that the model accurately describes the evolution of the traits (Beaulieu & O'Meara, 2019). Second, mutual information, as we have defined it, only provides information relative to the specified model and specified tips. A model which is more uncertain about any ancestral state, such as an equal rates model, is likely to have a more informative

ancestral state reconstruction because any deviation from an uninformed ancestral state is due to the particular values of the tip states. This does not make the equal rates model better than alternatives nor do we advocate for the use of information to assist in model selection. Instead, mutual information provides insight into the interaction between the model and tip states. Higher information of particular nodes could be indicative of an area of the phylogeny where the model's equilibrium frequencies were different from the ancestral state reconstruction and thus the tips provided the major explanation of the ancestral state. Mutual information is also highly correlated with the rates of evolution and has the intuitive property that as rates of evolution (or time) increase the information that the tips provide to the nodes decreases (Sober & Steel, 2011).

It is important to have a biologically realistic model of trait evolution when conducting an ancestral state reconstruction. With the generalizations made to corHMM we have provided two distinct ways to increase the realism of phylogenetic comparative modeling. First, we have allowed for the correlated evolution of several characters and states. Whether traits are correlated because of underlying pleiotropy leading to genetic correlation (Conner et al., 2011) or selective covariance (Mahler, Revell, Glor, & Losos, 2010), at the macroevolutionary scale they are better understood in a holistic context rather than independently evolving subunits. Second, the inclusion of hidden states allows for more detailed descriptions of the evolutionary process. State-dependent processes can differ in both rate and structure and thus provide a description of heterogeneity in the tempo and mode of evolution. However, these generalizations do not exist without cost. Increased complexity of models leads to greater parameterization which can lead to poor model performance (Grundler & Rabosky, 2020). Thus, as others recommend, we suggest having multiple working hypotheses (Chamberlin, 1890; Platt, 1964; Mayr, 1997; Burnham & Anderson, 2002). The generalizations and tools available in corHMM allow for the construction

of a carefully defined set of candidate models which can be compared in an information theoretic context. The Akaike Information Criterion (AIC) applies the principles of parsimony and represents a trade-off between bias and variance as a function of the dimension of the model (Forster & Sober, 1994). Combining AIC with a carefully constructed set of models leads to multi-model inference. Rather than focusing on a single best model, we can focus on the parameters from the set of our best supported models (Burnham & Anderson, 2002; Caetano, O'Meara, & Beaulieu, 2018). It is, therefore, just as important to include Mk models alongside HMMs because in cases where the increased parameterization of an HMM are unnecessary, alternative models with less parameterization are available as simpler explanations.

We have demonstrated that there is a potential informational and accuracy advantage of including additional states and characters in a simulation setting (*3.1.1 Increasing the number of observed characters or states)*. However, it remained to be seen whether modeling the correlated evolution of multiple characters would impact the ancestral state in an empirical example and whether those results match biological expectations. The controversy surrounding the phyllotaxy of the ancestral angiosperm flower is a particularly appropriate case study for the generalized version of corHMM, as it not only allows for the dependent evolution of several discrete characters but also includes hidden states as a fitting addition to help describe the heterogeneous evolution of angiosperms. We presented three different datasets, each allowing for a different level of polymorphism, uncertainty, and number of tip states. In our first dataset, we excluded all polymorphic species and any species with an unknown tip state. In this case, we found a two rate class *ER/ER* model was favored and the most likely ancestral state of the floral phyllotaxy was entirely whorled. This is in contrast to previous work which used a similarly constrained dataset and suggested either an ambiguous state (Maximum Parsimony result), an entirely spiral floral

phyllotaxy (Maximum Likelihood result), or a spiral perianth, whorled androecium, and spiral gynoecium (reversible jump Monte Carlo Markov Chain result) (Sauquet et al. 2018 - Appendix S2, V7Csub1).

Previous work posed the question of whether we should dismiss ancestral state combinations not observed among living species (Sauquet et al. 2018). The first dataset we presented excluded any trait combinations not observed in the data. However, the other two datasets allow for an ancestral state combination that was never directly observed at the tips because these analyses include tips where the states are not completely known. In both cases, we found the hidden rates model *ER/ER* where dual transitions are allowed to be the best supported and the model averaged most likely ancestral state was an entirely whorled floral phyllotaxy. Thus, across all datasets we found that the ancestral state combination was one of the most common tip state combinations. However, this does not mean a combination of states unknown in any extant species is impossible. In fact, we find that the preferred ancestral state is highly variable and dependent on the model (Table 1; Table S1-S3) and the entirely whorled angiosperm flower is likely a reflection of the *ER/ER* model's high AIC weight within the set of tested models. This means that should a more realistic model be introduced in the set we could find a very different answer and highlights the importance of having a set of biologically realistic models.

**Conclusion**

Although there is a growing consensus that phylogenies and their associated methods are being used in ways that exceed what they can infer (Losos 2011; Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Cooper et al. 2016), we have shown that there is still under-utilized

information in phylogenetic comparative datasets. First, HMMs extract signals of rate heterogeneity when it is present and equally important, do not falsely locate signals where they are absent. Second, increased trait depth adds new information and consistently improves ancestral state reconstruction estimates. Indeed, as datasets continue to grow, so will the analytical power that biologists have for testing complex models of evolution. Finally, the inclusion of correlated trait evolution and hidden states is relevant beyond theoretical considerations, and we have shown that these generalizations can change the results of an ancestral state reconstruction in empirical datasets. There is still a great deal of uncertainty in the reconstruction of the ancestral phyllotaxy of angiosperms, but by using AIC weighted marginal probabilities we have been able to take into account different biological explanations of floral evolution, eventually finding support an entirely whorled perianth, androecium, and gynoecium. Although hidden Markov models are not a perfect substitute for real observation of a hidden character, they make for a tractable and a biologically reasonable description of heterogeneity in the evolutionary process over long time scales.

**References**

Beaulieu, J. M., & Donoghue, M. J. (2013). Fruit Evolution and Diversification in Campanulid Angiosperms. *Evolution*, *67*(11), 3132–3144. doi:10.1111/evo.12180

Beaulieu, J. M., & O'Meara, B. C. (2016). Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. *Systematic Biology*, *65*(4), 583–601. doi:10.1093/sysbio/syw022

Beaulieu, J. M., & O'Meara, B. C. (2019). Diversity and skepticism are vital for comparative biology: a response to Donoghue and Edwards (2019). *American Journal of Botany*, *106*(5), 613–617. doi:10.1002/ajb2.1278

Beaulieu, J. M., O'Meara, B. C., & Donoghue, M. J. (2013). Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. *Systematic Biology*, *62*(5), 725–737. doi:10.1093/sysbio/syt034

Bollback, J. P. (2006). SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, *7*(1), 88.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). New York: Springer-Verlag. Retrieved from https://www.springer.com/gp/book/9780387953649

Caetano, D. S., O'Meara, B. C., & Beaulieu, J. M. (2018). Hidden state models improve state-dependent diversification approaches, including biogeographical models: HMM AND THE ADEQUACY OF SSE MODELS. *Evolution*, *72*(11), 2308–2324. doi:10.1111/evo.13602

Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, *15*(366), 92–96.

Conner, J. K., Karoly, K., Stewart, C., Koelling, V. A., Sahli, H. F., & Shaw, F. H. (2011). Rapid Independent Trait Evolution despite a Strong Pleiotropic Genetic Correlation. *The American Naturalist*, *178*(4), 429–441. doi:10.1086/661907

Cooper, N., Thomas, G. H., & FitzJohn, R. G. (2016). Shedding light on the 'dark side' of phylogenetic comparative methods. *Methods in Ecology and Evolution*, *7*(6), 693–699. doi:10.1111/2041-210X.12533

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Eddy, S. R. (2004). What is a hidden Markov model? *Nature Biotechnology*, *22*(10), 1315–1316. doi:10.1038/nbt1004-1315

Felsenstein, J., & Churchill, G. A. (1996). A Hidden Markov Model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution*, *13*(1), 93–104. doi:10.1093/oxfordjournals.molbev.a025575

Forster, M., & Sober, E. (1994). How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science*, *45*(1), 1–35. doi:10.1093/bjps/45.1.1

Galtier, N. (2001). Maximum-Likelihood Phylogenetic Analysis Under a Covarion-like Model. *Molecular Biology and Evolution*, *18*(5), 866–873. doi:10.1093/oxfordjournals.molbev.a003868

Gascuel, O., & Steel, M. (2014). Predicting the Ancestral Character Changes in a Tree is Typically Easier than Predicting the Root State. *Systematic Biology*, *63*(3), 421–435. doi:10.1093/sysbio/syu010

Grundler, M. C., & Rabosky, D. L. (2020). Macroevolutionary analysis of discrete traits with rate heterogeneity. *BioRxiv*, 2020.01.07.897777. doi:10.1101/2020.01.07.897777

Losos, J. B. (2011). Seeing the Forest for the Trees: The Limitations of Phylogenies in Comparative Biology: (American Society of Naturalists Address). *The American Naturalist*, *177*(6), 709–727. doi:10.1086/660020

Maddison, W. P., & FitzJohn, R. G. (2015). The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic Biology*, *64*(1), 127–136. doi:10.1093/sysbio/syu070

Maddison, W. P., Midford, P. E., Otto, S. P., & Oakley, T. (2007). Estimating a Binary Character's Effect on Speciation and Extinction. *Systematic Biology*, *56*(5), 701–710. doi:10.1080/10635150701607033

Mahler, D. L., Revell, L. J., Glor, R. E., & Losos, J. B. (2010). Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution: International Journal of Organic Evolution*, *64*(9), 2731–2745.

Marazzi, B., Ané, C., Simon, M. F., Delgado-Salinas, A., Luckow, M., & Sanderson, M. J. (2012). Locating Evolutionary Precursors on a Phylogenetic Tree. *Evolution*, *66*(12), 3918–3930. doi:10.1111/j.1558-5646.2012.01720.x

Mayr, E. (1997). *This is biology: The science of the living world*. Universities Press.

Mossel, E. (2003). On the Impossibility of Reconstructing Ancestral Data and Phylogenies. *Journal of Computational Biology*, *10*(5), 669–676. doi:10.1089/106652703322539015

Mossel, E., & Peres, Y. (2003). Information flow on trees. *The Annals of Applied Probability*, *13*(3), 817–844. doi:10.1214/aoap/1060202828

O'Meara, B. C. (2012). Evolutionary Inferences from Phylogenies: A Review of Methods. *Annual Review of Ecology, Evolution, and Systematics*, *43*(1), 267–285. doi:10.1146/annurev-ecolsys-110411-160331

O'Meara, B. C., Smith, S. D., Armbruster, W. S., Harder, L. D., Hardy, C. R., Hileman, L. C., … Diggle, P. K. (2016). Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1830), 20152304. doi:10.1098/rspb.2015.2304

Pagel, M. (1994). Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B*, *255*(1342), 37–45. doi:10.1098/rspb.1994.0006

Penny, D., McComish, B. J., Charleston, M. A., & Hendy, M. D. (2001). Mathematical Elegance with Biochemical Realism: The Covarion Model of Molecular Evolution. *Journal of Molecular Evolution*, *53*(6), 711–723. doi:10.1007/s002390010258

Platt, J. R. (1964). Strong Inference. *Science*, *146*(3642), 347–353.

Rabosky, D. L., & Goldberg, E. E. (2015). Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. *Systematic Biology*, *64*(2), 340–355. doi:10.1093/sysbio/syu131

Román-Palacios, C., Scholl, J. P., & Wiens, J. J. (2019). Evolution of diet across the animal tree of life. *Evolution Letters*, *3*(4), 339–347. doi:10.1002/evl3.127

Sanderson, M. J., & Donoghue, M. J. (1989). Patterns of Variation in Levels of Homoplasy. *Evolution*, *43*(8), 1781–1795. doi:10.1111/j.1558-5646.1989.tb02626.x

Sauquet, H., von Balthazar, M., Doyle, J. A., Endress, P. K., Magallón, S., Staedler, Y., & Schönenberger, J. (2018). Challenges and questions in reconstructing the ancestral flower of angiosperms: A reply to Sokoloff et al. *American Journal of Botany*, *105*(2), 127–135. doi:10.1002/ajb2.1023

Sauquet, H., von Balthazar, M., Magallón, S., Doyle, J. A., Endress, P. K., Bailes, E. J., … Schönenberger, J. (2017). The ancestral flower of angiosperms and its early diversification. *Nature Communications*, *8*(1), 16047. doi:10.1038/ncomms16047

Schultz, T. R., Cocroft, R. B., & Churchill, G. A. (1996). The Reconstruction of Ancestral Character States. *Evolution*, *50*(2), 504–511. doi:10.1111/j.1558-5646.1996.tb03863.x

Siepel, A., & Haussler, D. (2005). Phylogenetic Hidden Markov Models. In *Statistical Methods in Molecular Evolution* (pp. 325–351). New York: Springer-Verlag. doi:10.1007/0-387-27733-1_12

Sober, E., & Steel, M. (2011). Entropy increase and information loss in Markov models of evolution. *Biology & Philosophy*, *26*(2), 223–250. doi:10.1007/s10539-010-9239-x

Sober, E., & Steel, M. (2014). Time and Knowability in Evolutionary Processes. *Philosophy of Science*, *81*(4), 558–579. doi:10.1086/677954

Sokoloff, D. D., Remizowa, M. V., Bateman, R. M., & Rudall, P. J. (2018). Was the ancestral angiosperm flower whorled throughout? *American Journal of Botany*, *105*(1), 5–15. doi:10.1002/ajb2.1003

Steel, M., & Penny, D. (2005). *Maximum parsimony and the phylogenetic information in multistate characters*. New York: Oxford University Press.

Tarasov, S. (2019). Integration of Anatomy Ontologies and Evo-Devo Using Structured Markov Models Suggests a New Framework for Modeling Discrete Phenotypic Traits. *Systematic Biology*. doi:10.1093/sysbio/syz005

Townsend, J. P., & Naylor, G. (2007). Profiling Phylogenetic Informativeness. *Systematic Biology*, *56*(2), 222–231. doi:10.1080/10635150701311362

Uyeda, J. C., Zenil-Ferguson, R., & Pennell, M. W. (2018). Rethinking phylogenetic comparative methods. *Systematic Biology*, *67*(6), 1091–1109. doi:10.1093/sysbio/syy031

Yang Lou, X. (2017). Hidden Markov Model Approaches for Biological Studies. *Biometrics & Biostatistics International Journal*, *5*(4). doi:10.15406/bbij.2017.05.00139

Zucchini, W., MacDonald, I. L., & Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC.

# Appendix

**Table 1** – Model rankings from the maximum-likelihood analysis of the ancestral angiosperm flower for dataset two (polymorphic species included and unknown states excluded). Models separated by a "/" indicate a hidden rates model and the split distinguishes between the two state-dependent process (for example, *ER/ARD*, represents a hidden rate model where R1 is an equal rates model and R2 is an all rates differ model). Dual describe whether the model allowed for multi-state transitions (for example, if dual transitions were TRUE, then changing from entirely whorled phyllotaxy to entirely spiral phyllotaxy is allowed). AICc is the sample size corrected Aikaike Information Criterion. AICcWt is the relative likelihood of each model and is used in model averaging. Mean rate is the average transition rate for a particular model. ASR is the most likely ancestral state reconstruction for a particular model and its marginal probability. *k* rates is the number of independent rate parameters being estimated for a given model.

| Model | Dual Transitions | AICc | AICcWt | Mean Rate | ASR | *k* rates |
|---|---|---|---|---|---|---|
| ER | FALSE | 208.44 | <0.01 | <0.001 | (1,R1) 93% | 1 |
| SYM | FALSE | 138.25 | <0.01 | 16.67 | (1,R1) 99% | 12 |
| ARD | FALSE | 125.53 | <0.01 | 13.56 | (5,R1) 100% | 24 |
| ER/ER | FALSE | 118.06 | 0.01 | 37.52 | (1,R2) 63% | 4 |
| SYM/SYM | FALSE | 147.3 | <0.01 | 6.25 | (1,R2) 93% | 26 |
| ER/ARD | FALSE | 132.6 | <0.01 | 4.72 | (8,R2) 87% | 27 |
| ARD/ARD | FALSE | 191.76 | <0.01 | 13.58 | (4,R1) 82% | 50 |
| ER | TRUE | 115.2 | 0.06 | <0.001 | (1,R1) 97% | 1 |
| SYM | TRUE | 193.77 | <0.01 | 5.51 | (5,R1) 100% | 50 |
| ARD | TRUE | 212.59 | <0.01 | 16.36 | (5,R1) 100% | 56 |
| **ER/ER** | **TRUE** | **109.72** | **0.93** | **<0.001** | **(1,R2) 91%** | **4** |
| SYM/SYM | TRUE | 383.59 | <0.01 | 5.47 | (8,R1) 92% | 102 |
| ER/ARD | TRUE | 220.93 | <0.01 | 3.13 | (8,R2) 93% | 59 |
| ARD/ARD | TRUE | 443.07 | <0.01 | 10.76 | (4,R1) 86% | 114 |

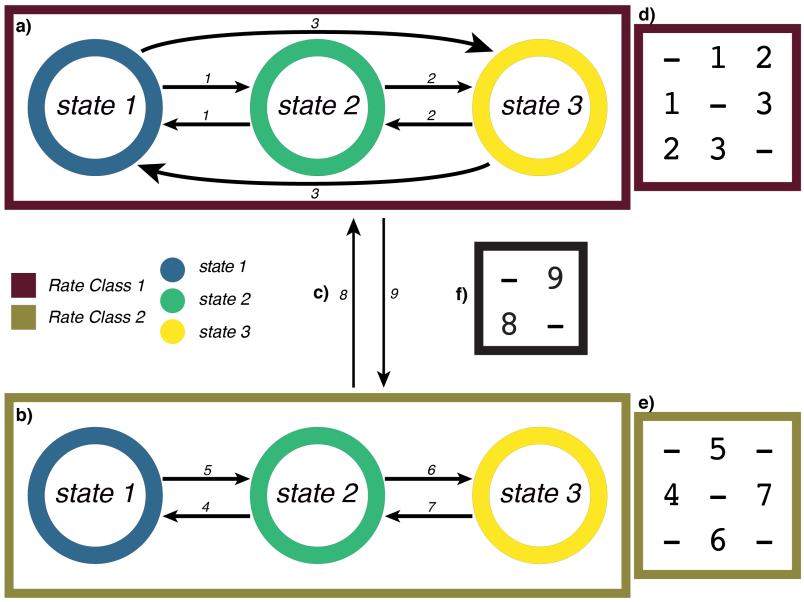


**Figure 1**. A decomposed HMM containing 3 observed states and 2 hidden rate classes. R1 is one state-dependent process that describes transitions to and from observed states as being equal (a, d), whereas R2 is a state-dependent process that describes state 2 as a necessary intermediate (b, e). The parameter process that relates R1 and R2 and describes the transitions between R1 and R2 (c, f).

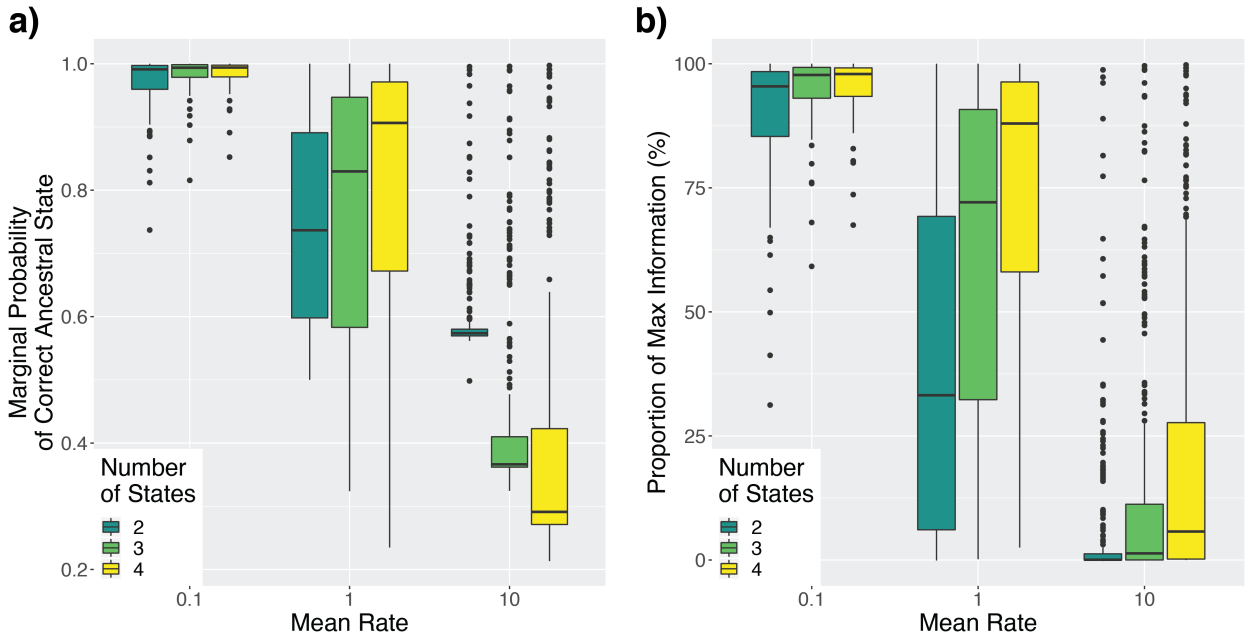


**Figure 2**. Performance of standard and hidden Markov models depending on the number of states in the dataset and mean rate. Each dataset was simulated under a mean rate of 0.1, 1, or 10 transitions $Myr^{-1}$, with 2, 3, or 4 observed states and no hidden states. a) The marginal probability of estimating the correct ancestral state. b) The proportion of information gained about ancestral states from each dataset and model.

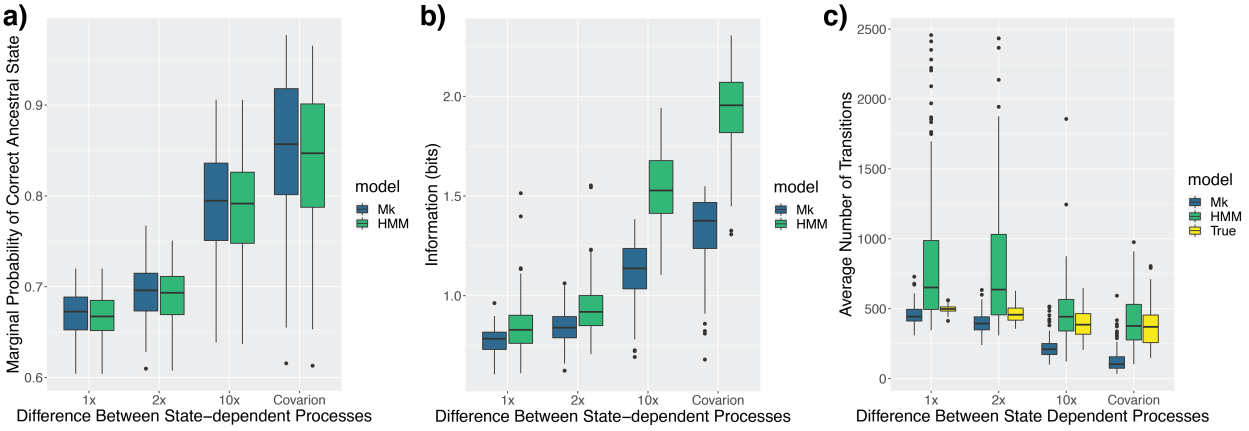


**Figure 3**. Comparison of fits from a Markov and HMM model when an HMM is the generating model. We vary the difference between state-dependent processes from no difference (1x) to complete asymmetry where state transitions occur in one state-dependent process only (i.e., "covarion" model; see *Performance in simulation*). a) The marginal probability of the correct ancestral state. b) The average amount of information (bits) for ancestral states from each dataset and model. c) The number of transitions averaged over 150 simmaps.

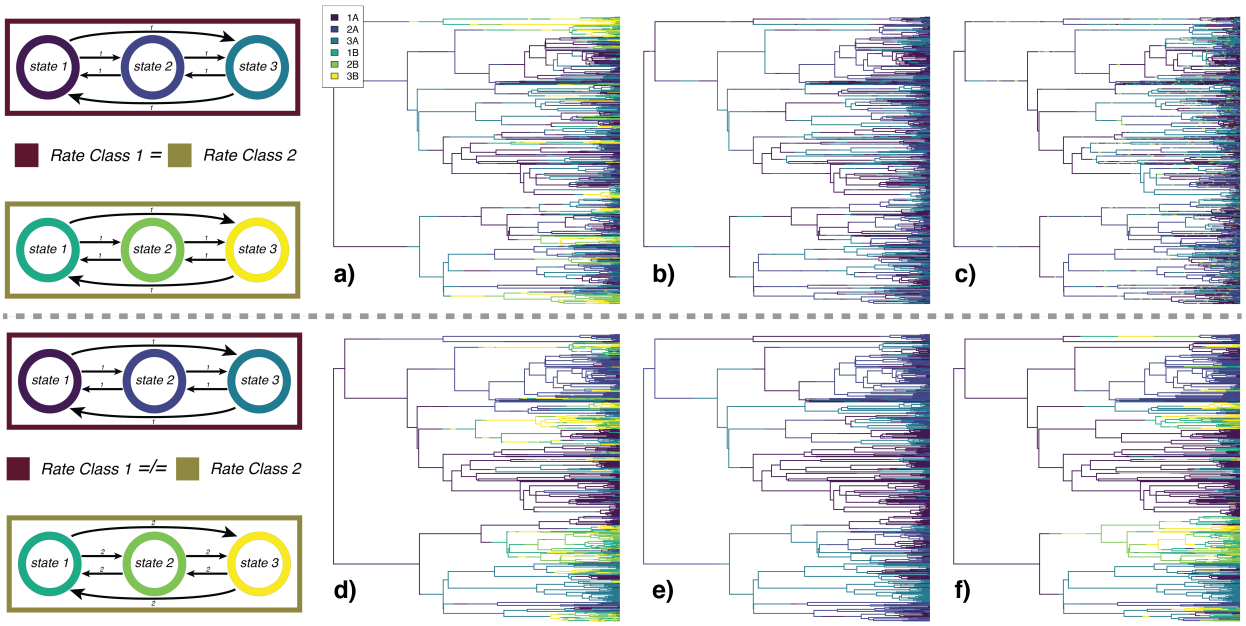


**Figure 4.** Stochastic maps demonstrating the effect of differences in the magnitude of two state-dependent processes. The first row shows data is simulated where there were no differences in the state-dependent processes, where (a) is the true generating model, (b) is one example of the character history simulated under the MLE from standard Markov model, and (c) is one example of the character history simulated under the MLE of an HMM. The second row is the same, but with data simulated with a 10-fold difference between the state-dependent processes. A Markov model does not contain a distinction between the hidden classes A and B, thus it is displayed only in terms of the states 1A, 2A, and 3A. Comparing the HMM in (c) and (f) demonstrates that an HMM will only detect a hidden state when it influences the observed, state-dependent, process.

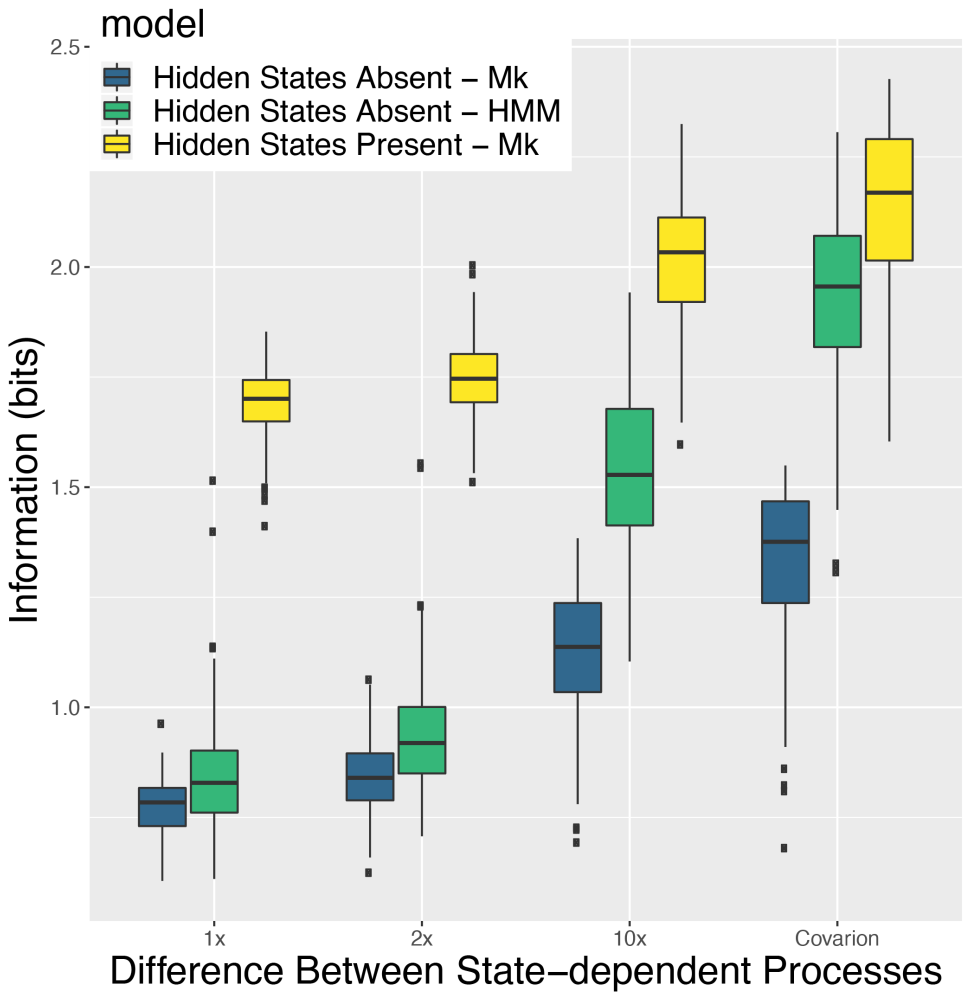


**Figure 5**. Average information when a hidden state is either directly observable or unobserved. If the hidden state is unobserved (*Hidden States Absent*), we compare the information gained when fitting a Markov model (Mk) or a hidden Markov model (HMM) to a dataset that was generated with a hidden state, but that hidden state was removed from the dataset. When the hidden state is directly observable (*Hidden States* Present) we fit a standard Mk to the full dataset that includes the potential hidden state. When the hidden state is directly observed, the datasets are comprised of 6 discrete states.

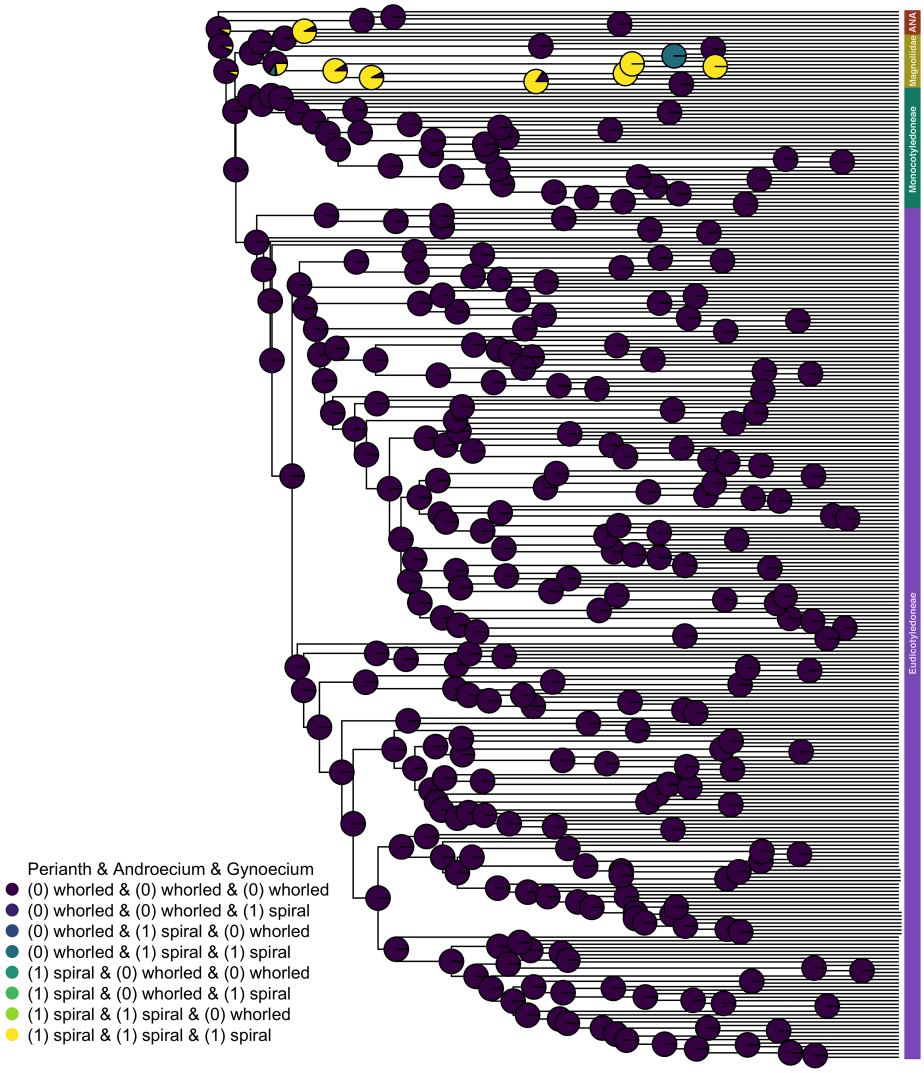


**Figure 6**. Model averaged ancestral state reconstructions of dataset two (polymorphic species are included, but species with unknown states are excluded). The marginal probability that the root state is entirely whorled is 91.7%.

## Derivation of Mutual Information

We define information as the difference between the unconditional entropy of the node states, $H(X_v)$, and the entropy of the node states conditioned on the data, $H(X_v|X_h = D)$ (Cover & Thomas, 1991). The unconditional entropy of node $v$ is defined as:

$$H(X_v) = -\sum_{i=1}^{k} \pi[X_v = i] log_2(\pi[X_v = i]),$$

where $\pi[X_v = i]$ is the prior probability of a node taking a particular state. For the root, the prior depends on user choice, as there are several options (Yang, Kumar, & Nei, 1995; Pagel, 1999; FitzJohn, Maddison, & Otto, 2009). Here we assume the prior probability on the root node is the expected equilibrium frequency, $\pi$, which is calculated directly from the transition model by solving $\pi \boldsymbol{Q} = 0$. This aligns our expectation of the root node with all other internal nodes such that, in the absence of information from the tips, the probability of a particular state is assumed to be drawn from the equilibrium frequencies. In other words, the information of the tip states decreases as rates increase and, ultimately, the probability of a node state becomes completely determined by the model. We define the conditional entropy as:

$$H(X_v|X_h = D) = -\sum_{i=1}^{k} P[X_v = i|X_h = D] log_2(P[X_v = i\ |X_h = D]),$$

where $P[X_v = i|X_h = D]$ is the conditional probability that a node is fixed as being in state $i$ given the probability of observing the tip data (which is just the marginal probability of state *i*). In particular, we are interested in the average entropy of a node for all states $i \ldots k$, given we observe a particular dataset, $X_h = D$. Thus, the conditional entropy will vary by node, but the unconditional entropy is set by the model. To produce a measure of mutual information between the observations at the tips and estimates at internal nodes, we take the difference between the

conditional entropy and the unconditional entropy and average across all nodes. However, the unconditional entropies will be greater for datasets that include more states because unconditional entropy sets the upper limit of what is possible to learn. This alone could contribute to large informational differences between models with different numbers of observed states. Therefore, we also measure the proportion of maximum information gained $(\frac{mutual\ information}{unconditional\ entropy} \times 100\%)$.

**Table S1** – Modeling results when all polymorphic and unknown taxa are removed. Note: When we say "unknown taxa" we mean that at least one of the states of that species is unknown.

| Model | k.rate | AICc | AICcWt | MeanRate | ASR |
|---|---|---|---|---|---|
| **ER** | 1 | 99.95 | 0.04 | <0.01 | (1,R1) 98% |
| **SYM** | 6 | 100.59 | 0.03 | <0.01 | (1,R1) 98% |
| **ARD** | 12 | 98.82 | 0.07 | <0.01 | (3,R1) 100% |
| **ER/ER** | 4 | 93.71 | 0.86 | <0.01 | (1,R2) 91% |
| **SYM/SYM** | 14 | 104.06 | <0.01 | <0.01 | (1,R2) 90% |
| **ER/ARD** | 15 | 104.48 | <0.01 | <0.01 | (4,R2) 93% |
| **ARD/ARD** | 26 | 128.79 | <0.01 | <0.01 | (3,R1) 99% |

**Table S2** – Modeling results when polymorphic are included and unknown taxa are removed.

| Model | Dual Transitions | AICc | AICcWt | MeanRate | ASR (%) | k.rate |
|---|---|---|---|---|---|---|
| **ER** | FALSE | 208.44 | <0.01 | <0.01 | (1,R1) 93 | 1 |
| **SYM** | FALSE | 138.25 | <0.01 | 16.67 | (1,R1) 99 | 12 |
| **ARD** | FALSE | 125.53 | <0.01 | 13.56 | (5,R1) 100 | 24 |
| **ER/ER** | FALSE | 118.06 | 0.01 | 37.52 | (1,R2) 63 | 4 |
| **SYM/SYM** | FALSE | 147.3 | <0.01 | 6.25 | (1,R2) 93 | 26 |
| **ER/ARD** | FALSE | 132.6 | <0.01 | 4.72 | (8,R2) 87 | 27 |
| **ARD/ARD** | FALSE | 191.76 | <0.01 | 13.58 | (4,R1) 82 | 50 |
| **ER** | TRUE | 115.2 | 0.06 | <0.01 | (1,R1) 97 | 1 |
| **SYM** | TRUE | 193.77 | <0.01 | 5.51 | (5,R1) 100 | 50 |
| **ARD** | TRUE | 212.59 | <0.01 | 16.36 | (5,R1) 100 | 56 |
| **ER/ER** | TRUE | 109.72 | 0.93 | <0.01 | (1,R2) 91 | 4 |
| **SYM/SYM** | TRUE | 383.59 | <0.01 | 5.47 | (8,R1) 92 | 102 |
| **ER/ARD** | TRUE | 220.93 | <0.01 | 3.13 | (8,R2) 93 | 59 |
| **ARD/ARD** | TRUE | 443.07 | <0.01 | 10.76 | (4,R1) 86 | 114 |

**Table S3** – Modeling results when polymorphic are included and unknown taxa are included.

| Model | k.rate | AICc | AICcWt | MeanRate | ASR |
|---|---|---|---|---|---|
| **No dual transitions** | | | | | |
| **ER** | 1 | 322.91 | <0.01 | <0.01 | (1,R1) 93% |
| **SYM** | 12 | 248.08 | <0.01 | 16.67 | (1,R1) 99% |
| **ARD** | 24 | 231.28 | 0.01 | 15.56 | (7,R1) 58% |
| **ER/ER** | 4 | 234.55 | <0.01 | 37.73 | (1,R2) 97% |
| **SYM/SYM** | 26 | 250.65 | <0.01 | 12.45 | (1,R2) 93% |
| **ER/ARD** | 27 | 237.47 | <0.01 | 17.44 | (7,R2) 55% |
| **ARD/ARD** | 50 | 282.92 | <0.01 | 11.1 | (8,R1) 82% |
| **Model** | **k.rate** | **AICc** | **AICcWt** | **MeanRate** | **ASR** |
| **Dual transitions** | | | | | |
| **ER** | 1 | 228.27 | 0.04 | <0.01 | (1,R1) 97% |
| **SYM** | 50 | 287.89 | <0.01 | 17.4 | (5,R1) 99% |
| **ARD** | 56 | 302.98 | <0.01 | 20.36 | (7,R1) 97% |
| **ER/ER** | 4 | 221.73 | 0.95 | <0.01 | (1,R2) 93% |
| **SYM/SYM** | 102 | 410.48 | <0.01 | 4.98 | (8,R2) 88% |
| **ER/ARD** | 59 | 309.76 | <0.01 | 11.71 | (2,R1) 90% |
| **ARD/ARD** | 114 | 445.36 | <0.01 | 5.99 | (8,R1) 76% |

# CHAPTER II

## A potential solution to the unresolved challenge of false correlation between discrete characters

James Boyko and Jeremy Beaulieu


### Abstract

The correlation between two characters is often interpreted as evidence that there exists a significant and biologically important relationship between them. However, Maddison and FitzJohn (2015) recently pointed out that in certain situations find evidence of correlated evolution between two categorical characters is often spurious, particularly, when the dependent relationship stems from a single replicate deep in time. Here we will show that there may, in fact, be a statistical solution to the problem posed by Maddison and FitzJohn (2015) naturally embedded within the expanded model space afforded by the hidden Markov model (HMM) framework. We demonstrate that the problem of single unreplicated evolutionary events manifests itself as rate heterogeneity within our models and that this is the source of the false correlation. Therefore, we argue that this problem is better understood as model misspecification rather than a failure of comparative methods to account for phylogenetic pseudoreplication. We utilize HMMs to develop a multi-rate independent model which, when implemented, drastically reduces support for correlation. The problem itself extends beyond categorical character evolution, but we believe that the practical solution presented here may lend itself to future extensions in other areas of comparative biology.

## Introduction

Correlated or dependent evolution on a macroevolutionary scale is defined as a change in a character state (e.g., plumage color) that is linked to the presence of a particular state in a separate character (e.g., beak color). In other words, the evolution of character $X$ can be said to be dependent on character $Y$ if in the presence of a particular state of $Y$ (e.g., $Y_0$), shifts within character $X$ occur in a different way from when the lineage is in an alternative state of $Y$ (e.g., $Y_1$). For example, a shift from $X_0$ to $X_1$ may occur more quickly when paired with $Y_1$ than with $Y_0$ resulting in a distribution with many character pairs $X_1Y_1$. It is often the case that these sorts of dependent relationships between characters seem obvious, especially if the observations of many individuals are consistent.

However, what happens when all observations of the pair come from, for example, one biogeographic region? In other words, there may have been many individual pairs of $X_1Y_1$ observed, but they all came from one population. Since the strength of the relationship is related to the number of individual observations, the non-independence of them raises concerns about the validity of the proposed correlation. This problem extends to interspecific comparisons too, but rather than observations being linked to one of two populations, they are associated with particular taxonomic groupings and shared histories. This fact was well understood as early as Darwin (1859) and the tools for dealing with the resulting statistical non-independence have been available to comparative biologists since the foundational work of Felsenstein (1985). Nevertheless, this issue of "phylogenetic pseudoreplication", where species are non-independent due to their shared ancestry, served as the basis for the concerns raised by Maddison and FitzJohn (2015) regarding tests of dependent character evolution.

Maddison and FitzJohn (2015) demonstrated that the most widely used phylogenetic method for detecting correlated evolution between categorical characters (Pagel 1994), almost always indicates strong evidence of correlation when singular events deep in time can account for the co-distribution of two characters. To demonstrate their point, they fit correlated models to datasets generated under their so-called, “Darwin's” and the “Unreplicated Burst” scenarios (Fig. 1). Darwin’s scenario results in the perfect co-distribution of two characters, which in practice, might occur when testing for correlations between two synapomorphies (e.g., presence/absence of middle ear bones and fur). Under the Unreplicated Burst scenario, only one of the two characters has phylogenetically replicated change. This scenario occurs when one of the characters is a synapomorphy for the clade, with the other character undergoing several changes within the focal clade. The issue is that, when applied to either Darwin’s or the Unreplicated Burst scenario, commonly used comparative methods (Pagel 1994) will almost always indicate strong evidence of correlation despite the dependent relationship arising from little more than a single event deep in time.

There is considerable interest in understanding and, ultimately, finding a resolution to the problem posed by Maddison and FitzJohn (2015). Recently, Uyeda et al. (2018) suggested that for Darwin’s scenario, the relatively long periods of stasis between the two characters (i.e., minimal trait change) is the primary cause for their significant dependent relationship. In fact, they showed that the probability of selecting a character-dependent model (i.e., a model of correlated evolution between the two characters) over a character-independent model (i.e., a model where the two characters are explicitly not correlated) was proportional to the ratio between the length of the branch where the shift occurred and the total length of the tree. The nature of this ratio ensured that a correlated model would always be supported in cases where

singular evolutionary events led to a co-distribution of characters. Another study, by Gardner and Organ (2021), tested a variety of correlated models beyond Markov models and examined the structure of datasets which are susceptible to the problem of false dependence. They found that all the tested comparative methods produced erroneous correlations when datasets were phylogenetically pseudoreplicated.

In both of these studies, the authors have addressed the problem by encouraging scientists to think critically about their models. While this recommendation is certainly admirable and correct, it is not a direct and satisfying solution to the statistical problems presented so far, as no amount of methodological vigilance will ever prevent analyses from being marred by phylogenetic pseudoreplication. However, prior analyses have limited model comparisons to only a few models, and have overlooked the very large set of alternative Markov models which can also be consistent with correlation or independence depending on the model's structure. These alternative models have been briefly discussed previously (Pagel 1994; Pagel and Meade 2006) and, as we will show, the inclusion of a few examples within the model set can play a crucial role in ensuring a fair test of correlation. These underrepresented models, in addition to enormous model space provided by hidden Markov models (HMMs) for addressing rate heterogeneity across the tree (Beaulieu et al. 2013; Boyko and Beaulieu 2021), form the basis of our putative statistical solution to the problem posed by Maddison and FitzJohn (2015). We acknowledge that the problem itself extends beyond categorical character evolution, but we believe that the practical solution presented here may lend itself to future extensions in other areas.

We draw on two important insights as they relate to models of categorical character evolution. The first is that model space is severely underexplored and that the inclusion of more

complex, character-independent models within our modeling set helps reduce evidence of false correlation. We note that estimates of transition rates to and from unobserved character states are not statistically identifiable, revealing that the canonical character-dependent model is over-parameterizationed in phylogenetically pseudoreplicated datasets like Darwin's scenario (Fig. 1a). When only two or three of the four possible character state combinations are observed, we produce models nested within the correlated and independent model that are overwhelmingly favored over both. Second, the issue of false dependent relationships is not one of stasis *per se*, but rather, a failure to account for rate heterogeneity. We demonstrate that an explicit character-independent hidden Markov model (HMM) provides significant evidence for models of independent evolution in cases where a correlated model would have previously been supported. This is because under the classic Pagel (1994) framework, support for correlation comes from both a dependent relationship between characters and a strong signal of rate heterogeneity. By amending the Pagel framework with a model which allows for rate heterogeneity independent of a focal character, we correct the bias towards correlation. We also reiterate that the relative support of each model should be considered when interpreting biologically sound results rather than examining tests of character dependence against "trivial" nulls (Beaulieu and O'Meara 2016; Caetano et al. 2018; O'Meara and Beaulieu 2021).

## Methods and Results

### *Correlated models depend on observations of intermediate states*

While much has been written about the specifics of Pagel's model, we briefly review aspects of it in order to better illustrate our point — namely, that certain transition rates are not estimatable and that their inclusion may be an additional cause of false correlations uncovered by Maddison

and FitzJohn (2015). The correlated or dependent model of discrete character evolution, introduced by Pagel (1994), uses a continuous-time Markov process to estimate the rate of transitions between character states (Fig. 2ab). With a single binary character, *X*, the transition rate matrix, denoted as **Q**, is a simple 2x2 matrix, which contains all the information necessary to estimate the probability of a transition occurring between two states of character *X* over a given period of time. At its most complex, **Q** would contain two transition rates: from state $X_0$ to state $X_1$, and from state $X_1$ to state $X_0$. If we introduce a second binary character, *Y*, the number of possible observed state combinations is expanded — that is, the possible observed state combinations become $X_0Y_0$, $X_0Y_1$, $X_1Y_0$, and $X_1Y_1$. Consequently, this requires an expansion of **Q** to a 4x4 matrix, to account for all the possible transitions between state combinations. This model is considerably more complex, as the number of transitions goes from a maximum of two to a maximum of 12. However, the model introduced by Pagel (1994) is constrained specifically for the purpose of detecting correlations between characters by examining whether the state of one variable affects the probability of change in the other. To do this, dual transitions (i.e., changes in both *X* and *Y* occurring in a single time step) are removed. As noted by Pagel (1994), setting dual transition rates to zero does not rule out dual transitions over long periods of time. Rather, a dual transition from $X_0Y_0$ must first pass through state $X_0Y_1$ or $X_1Y_0$, before finally transitioning to $X_1Y_1$. Equating the rates of transitions between particular pathways allows for the construction and testing of an independent model (Pagel and Meade 2006). A model of independent evolution is nested within the correlated model but assumes that the transition rates between states of a character are equal to one another regardless of the state of the other character (e.g., [$X_0$ to $X_1$ | $Y_0$] = [$X_0$ to $X_1$ | $Y_1$]; Fig. 2ab). In other words, if these two characters,

*X* and *Y*, are independent, the presence of one character will have no influence on the change of the other and thus model selection criteria should choose the simpler model.

Using this specific nested framework, we were able to replicate the results of Maddison and Fitzjohn (2015). Specifically, we generated 100 datasets for Darwin's scenario and the Unreplicated Bursts scenario. Phylogenies were simulated under a $\lambda=1$ and $\mu=0.5$ until 100 extant taxa were reached, and each resulting tree was then evaluated for a focal monophyletic group between 40 and 60 taxa. For Darwin's scenario, extant species within the focal clade were assigned $X_1Y_1$ and species outside the clade were assigned $X_0Y_0$. We simulated Unreplicated Bursts by assigning all species outside the focal clade $X_0$, and all species within the clade $X_1$. Next, character *Y* was simulated at a rate of 100 transitions per million years. Outside of the focal clade, species were assigned $Y_0$ whereas within the focal clade, the simulated data resulted in both $Y_0$ and $Y_1$. We used *corHMM* (Beaulieu et al. 2013; Boyko and Beaulieu 2021) to fit and compare the four-state independent model (Fig. 2a) against the four-state correlated model (Fig. 2b) using Akaike Information Criterion (AIC). In all cases, we found overwhelming support for the correlated model for both Unreplicated Bursts and Darwin's scenario datasets (See Supplemental Materials). The mean AIC weight for the correlated model under Darwin's scenario was 92.52% and under Unreplicated Bursts it was 99.96%. As expected, an independent model was never favored over a correlated model in either scenario.

For Darwin's scenario, setting aside the critical analytical issues regarding phylogenetic pseudoreplication, we had additional concerns with the structure of the data and how this might impact estimates of transition rates. Under any continuous-time Markov process, the estimates of the transition rates among all possible character combinations are reflective of the observed state frequencies and distribution at the tips. But, what if two of the four character combinations are

not observed at all? Here we are referring to the two combinations, $X_0Y_1$ and $X_1Y_0$, not observed in any of the tips under Darwin's scenario. There may be biological reasons for not observing intermediate state combinations. For example, these combinations may be at some selective disadvantage, resulting in rapid transitions to another, more viable character combination (e.g., $X_0Y_0$ or $X_1Y_1$). Alternatively, it could be that one or both combinations are never possible due to some underlying genetic or developmental reasons (e.g., certain fruit character combinations, see Beaulieu and Donoghue 2013). However, whatever biological meaning is attributed to the lack of intermediate character state observations, in this case, is beside the point. There seems to be obvious, and yet unrecognized, identifiability issues with including transitions to and from these unobserved state combinations in the model, calling into question fitting the correlated model to these types of data. That is to say, if we never see intermediate state combinations at the tips, how can the model ever favor one pathway over the other?

To illustrate this point, we examined the likelihood surface of one of the datasets simulated under Darwin's scenario and fit under Pagel's correlated model (Fig. 3). Whether starting from $X_0Y_0$ or $X_1Y_1$, transition rate estimates to either of the unobserved character combinations fall along a ridge of equal likelihood, where changing the rate of transition to one unobserved state determines the rate for the transitions to the other unobserved state. When a lineage transitions into one of the states, the likelihood surface for transitions out of these states to either state $X_0Y_0$ or $X_1Y_1$ are completely flat, with all rates ranging from 0.1 to 100 transitions per unit time all having nearly identical likelihoods. Taken together, the preferred model estimates for various transition rates arise simply by chance of the optimization procedure, but more importantly, there are parameters which are clearly unneeded to explain the data.

One obvious solution is to simply remove the unobserved character combinations from the model completely. From a modeling perspective, removing unobserved states removes the parameters that fall along the likelihood ridge and should lead to a model that ends up being well estimated. Consequently, the question of whether independent or dependent models better explain the data becomes irrelevant as the two models collapse into one another when unobserved states are removed (Fig. 2c,d). This is clearly seen when the collapsed model is applied to an Unreplicated Burst scenario. Whether one starts with an independent model (Fig. 2a) or a correlated model (Fig. 2b), once unobserved states are removed, comparing alternative transition pathways between $X_0Y_0$ and $X_1Y_1$ are no longer possible. For example, take transitions between states of character $X$. Both the correlated and independent models estimate transitions from $X_0$ to $X_1$ as depending only on $Y_0$, since $X_0Y_1$ is not observed in the dataset. Since it is not possible to compare the likelihood of alternative scenarios of dependence a comparison of correlation and independence becomes irrelevant.

Including a collapsed model as part of our model set drastically changes the results. We found complete support for a collapsed state model for both Darwin's scenario and Unreplicated Bursts (see Supplemental Materials). The average AIC weight for the collapsed model is 99.7% under Darwin's scenario and 100.0% under an Unreplicated burst scenario. This suggests that the support for the correlated models over simpler independent models is a result of an intuitive, but necessary parameter constraint. Specifically, in an independent model, transitions between observed states are constrained to be identical to transitions between unobserved states (e.g., $X_0Y_0$ to $X_0Y_1$ must be identical to $X_0Y_1$ to $X_1Y_1$, even if $X_0Y_1$ is never observed). In contrast, the correlated model is not subject to these constraints. This is, of course, the important distinction between the two models and what allows us to test for correlated evolution. In this case, the

support for dependence occurs because, in a sense, the correlated model is free to "throw away" the inestimatable transition rates which describe movement to and from intermediate states, while the independent model is forced to evaluate them. However, this issue becomes moot when exclusively modeling observed state combinations because the dependent and independent models become equivalent descriptions of the evolutionary process and are, therefore, indistinguishable for the given data.

*Rate heterogeneity is necessary when testing for correlation between categorical variables*

A major issue for the collapsed model described above is that in Darwin's scenario, a single observation of $X_0Y_1$ and $X_1Y_0$ removes the possibility of collapsing the model structure. This suggests that modeling only observed state combinations is not a generalizable solution to the phylogenetic pseudoreplication of categorical characters. As we will show, with only a single observation of intermediate character combinations, support for the correlated model over an independent model remains substantial. Even so, the results above highlight information limitations and that the strong evidence for dependent models may be due to a lack of viable alternative independent models rather than being irrefutable evidence of correlation.

It is worth considering again the possible explanations of the data under Darwin's scenario. One possibility is that the characters $X$ and $Y$ evolve slowly and that their co-distribution is the result of two independent events deep in time. The probability of this scenario has been explored in-depth and its implausibility is a major contributor to the recurrent issues of false correlation when comparing dependent and independent models (Uyeda et al. 2018). We propose a complementary explanation for the correlated model's support: the independent model structure fixes the transition $X_0$ to $X_1$ to always be the same rate in the context of the state of $Y$ (Fig. 2a), whereas a dependent model structure allows transitions from $X_0$ to $X_1$ to vary

depending on the state of *Y* (Fig. 2b). Support for the correlated model, therefore, comes from the fact that the best explanation of the data is not one that has a single slow transition rate for the characters. Instead, the most likely description of the process is one in which transitions between $X_0$ and $X_1$ or $Y_0$ and $Y_1$ are allowed to occur rapidly within the focal clade and occur slowly outside of the focal clade. The relative stasis of $X_0$ outside the focal clade and the rapid accumulation of $X_1$ within the clade suggests that changes in *X* are not consistent throughout the tree.

Hidden Markov models (HMMs) are a natural way to deal with this kind of rate heterogeneity across the tree. The underlying mathematical framework of an HMM is no different than a typical Markov model. They utilize a rate matrix, **Q**, to estimate the probabilities of transitioning between discrete states and arrive at the likelihood of the model given the observed dataset (Felsenstein and Churchill 1996). However, HMMs introduce a so-called "hidden-state", which can represent any number of unobserved factors, biological or otherwise. Based on the presence or absence of this hidden-state, changes between observed states are allowed to vary. In the most extreme cases, the absence of the hidden state may halt the evolutionary process and result in periods of stasis. For example, Marazzi et al. (2012) conceptualized the hidden-state as a "precursor" trait and only in its presence could extrafloral nectaries (EFNs) emerge. It is important to emphasize that the precursor state was never directly observed and that the information for its presence or absence of the hidden state came from the rate heterogeneity of EFNs transitions. In some parts of the tree, the model EFNs emerged rapidly and in others there were periods of stasis. Of course, HMMs are more general than either halting or actuating the evolutionary process and are used to quantify rate heterogeneity without the necessity of stasis (e.g., comparing fast, slow, or intermediate rates as in Beaulieu et al.

2013). The key point here is that they allow for rate heterogeneity that is unlinked to another observed character.

We developed and tested a hidden Markov independent model (HMIM) which accounts for rate heterogeneity while maintaining the independence of the observed focal characters $X$ and $Y$ (Fig. 4). In our view, the inclusion of our model within the evaluated set better levels the playing field between correlated and independent models. For example, if we focus on character $X$, our proposed model utilizes hidden states to vary transition rates between $X_0$ and $X_1$ based on an unobserved character. This is similar to the way that the correlated model allows transition rates between $X_0$ and $X_1$ to differ based on the observed state of $Y$. If the cause of false correlation was, as we suspect, not accounting for rate heterogeneity, then both the hidden state independent and correlated model should be preferable to the simple independent model and evidence of correlation between $X$ and $Y$ should be greatly reduced.

We first removed the possibility of collapsing the Markov model by modifying Darwin's scenario. We defined the focal clade as being the monophyletic group where all observations of $X_1Y_1$ occur and randomly add the intermediate state observations of $X_0Y_1$ and $X_1Y_0$ within the focal clade (which refer to as "inside" hereafter), outside of the focal clade (which we refer to as "outside" hereafter), and both within and outside the focal clade (which refer to as "both" hereafter) (Fig. 5). Next, we verified that this modified Darwin's scenario still suffers from the problems of the original Darwin's scenario by comparing the independent and correlated models *sensu* Pagel (1994). We then added the hidden Markov independent model to the model set and evaluated two questions: (1) when comparing independent models to one another, is there evidence of rate heterogeneity? and (2) is support for the correlated model reduced when compared to an independent model with rate heterogeneity? In addition to AIC weight, we

utilized evidence ratios (*ER*) to explore the relative likelihood of our models. Evidence ratios are a simple extension of AIC weights, but as a means of evaluation, are important here since they allow us to focus on evaluating the relative evidence of pairs of models irrespective of other models in the set (Burnham and Anderson 2002). The evidence for model *i* over model *j* is the ratio between their AIC weights: $ER = w_i / w_j$ and it can help quantify whether the best model in our comparison is convincingly best. With alternative samples, a convincingly best model is likely to be chosen again sample to sample. However, if evidence for a model is low, we expect model selection uncertainty to be high. Following Burnham and Anderson (2002), an evidence ratio of greater than 2.7 is used as a guide to justify judging support for one model being better than another. This also neatly corresponds to a $\Delta AIC = 2$. We emphasize that this value should *not* be misconstrued as a significant test in a frequentist sense since we are not evaluating the probability of rejecting a null hypothesis.

For all modified Darwin's scenarios, we found substantial evidence (ER > 2.7) for a correlated model over a single rate class independent model (Fig. 5). The geometric mean evidence ratio for the correlated model over the single rate independent model was $ER_{outside}$= 59.51, $ER_{inside}$= 78.16, $ER_{both}$= 11.44 (Fig. 5), thus we, again, successfully recreated the conditions of Maddison and FitzJohn (2015) under a modified Darwin's scenario. Next, we examined the evidence for rate heterogeneity by comparing a single rate independent model to the hidden Markov independent model. We found substantial evidence for rate heterogeneity across all scenarios, with all mean evidence ratios of the HMIM over the standard independent model well over 20, indicating substantial support for rate heterogeneity ($ER_{outside}$= 24.45, $ER_{inside}$= 24.33, $ER_{both}$= 50.45). Finally, we tested whether there is still conclusive evidence of correlation between characters if we include the hidden state independent model within our

modeling set. We found that the evidence for a correlated model over the hidden Markov independent model was greatly reduced when compared to the single rate class independent model (Fig. 5; $ER_{outside}$= 2.43, $ER_{inside}$= 3.21, $ER_{both}$= 0.22; Fig. 5). In fact, with only two observations of each intermediate state combination ($X_0Y_1$ and $X_1Y_0$), support for the hidden Markov independent model over the correlated model was substantial (evidence for HMIM over a correlated model: $ER_{both}$= 4.41). Taken together, these findings suggest that 1) there is indeed substantial evidence of rate heterogeneity, and that this is causing the signal of false correlation; and 2) including a hidden Markov independent model can, at least, muddle evidence for correlation.

*A (potentially) complete solution to biased correlation between synapomorphies*

It was still concerning to us that for the original and two of the modified Darwin's scenarios (specifically the "outside" and "inside" sets; see Fig. 5), support for the correlated model was still often greater than the hidden state independent model. Although the addition of character independent rate heterogeneity muddles support for the correlated model, in the most extreme cases the best model remained the dependent model. To deal with this issue, we applied what we learned thus far, with regards to the over-parameterization of models and the necessity of rate heterogeneity and added a new set of simpler and nested models within the set presented thus far to specifically address the issues of Darwin's scenario.

It is critical to emphasize that model space has been underexplored and that there are many nested model structures that are consistent with either independence or correlation depending on their constraints (see also Pagel and Meade 2006). Here we describe two constrained versions of the independent and correlated models that achieve the most efficient description of the data. One simplified version of the correlated model suggests that when either

character *X* or *Y* is in state 0, rates of change are slower or faster than when either character is in state 1 (Fig. 6b). We refer to this as the "simplified correlated" model and it represents the simplest way to model a dependent relationship between two binary characters. Next, we created a "simplified independent" model of equal parameterization to the simplified correlated model, which equates all changes from 0 to 1 regardless of the character and the same is done for changes from 1 to 0 (Pagel and Meade 2006; Fig. 6a).

The structures of these simplified models have certain qualities that may make them apt descriptions of data like Darwin's scenario. Primarily, these models suggest that changes between states 0 and 1 do not necessarily depend on the specific identity of character *X* or *Y* since they are constrained to be equal. When we consider the redundancy of a dataset composed of two synapomorphies, it is obvious that there is little to no information that distinguishes the two characters– that is, it makes no difference whether one analyzes character *X* or character *Y* since their distributions are identical. The simplified models make that assumption explicit. It is also important to note that the simplified independent model and simplified correlated model maintain independence and dependence *sensu* Pagel (1994). The background state of the unchanging character does not influence changes in the case of the simplified independent model, whereas the background state of the unchanging character will influence rates of change in the case of the simplified dependent model (Pagel and Meade 2006). Finally, we can introduce rate heterogeneity by modeling the simplified independent and correlated models as two rate class hidden Markov models (Fig. 6c).

Returning to the modified Darwin's scenario datasets, we found consistent and overwhelming support for the simplified hidden Markov independent model across all scenarios (Table 1). The average AIC weight of the simplified HMIM when fit to modified Darwin's

scenarios are $w_{outisde}$= 89.6%, $w_{inside}$= 90.2%, and $w_{both}$= 93.5%. The set of models applied to this data included all models discussed thus far as well as more complicated versions of those previously described (such as a standard correlated model with multiple rate classes). Additionally, to ensure that these models are not biased towards being favored across all datasets, we simulated data under a simplified correlated, simplified independent, and simplified hidden Markov independent models. We then fit each model to these datasets and found that the generating model is consistently chosen as the best fitting model (see Supplemental Materials). In summary, our findings suggest that when a complete model set is considered, the bias towards a correlation noted by Maddison and Fitzjohn (2015) disappears. The model which best describes data under a strict Darwin's scenario is not one of correlation, but a simplified independent model with character independent rate heterogeneity.

*Broadly applicable solutions*

The issue discussed herein is recognized as being broadly applicable to several comparative methods that test for associations between variables (FitzJohn 2010; Rabosky and Goldberg 2015; Uyeda et al. 2018; Nakov et al. 2019; Gardner and Organ 2021). It is concerning that such a significant issue has seemingly gone unresolved for so long given comparative methods are of critical importance for understanding macroevolutionary patterns. However, in our view, the prevalence of the problems identified over the past few years is due to a singular overarching cause, namely, model misspecification, which occurs when a model, or set of models, is incomplete. Within the context of their model sets, authors of previous studies have correctly portrayed and analyzed the correlation bias of modeling dependence between discrete characters (Maddison and FitzJohn 2015; Uyeda et al. 2018; Gardner and Organ 2021). However, the danger of model misspecification is that the inferences drawn from an incomplete set are highly

susceptible to unforeseen biases – a fact will hold true in both theoretical and empirical contexts. Here, we are arguing that the model set is incomplete without the inclusion of models that allow for rate heterogeneity that is independent of the focal characters. The canonical character independent model of Pagel (1994) has no way to account for multiple rates of evolution, whereas support for a correlated model can come from both evidence of correlation and evidence of rate heterogeneity. The additional support from explaining rate heterogeneity is not a feature exclusive to correlated characters, and thus accounting for independent rate heterogeneity is necessary to resolve the model set misspecification. This misspecified model set has led to consistently biased evidence towards correlation, and it is the same issue addressed by the inclusion of the character independent models within state-dependent speciation extinction models (Beaulieu and O'Meara 2016). In that case, the biased association was between diversification rates and phenotype (Rabosky and Goldberg 2015), but the cause is the same. Models in which there are no differences in diversification are compared to models which tested for the presence of a correlation between character and diversification rate (which necessarily allow for multiple rates of diversification).

One difference between the problem of false correlation in SSE models and the problems within simpler Markov models is the narrative surrounding them. In the case of SSE models, the problem was viewed as a high false positive rate (Rabosky and Goldberg 2015), whereas in the case of discrete character evolution we are led towards viewing rate heterogeneity through the lens of single unreplicated evolutionary events (Maddison and FitzJohn 2015). However, both points contribute to the same problem and if we view single evolutionary events as examples of where evolution has changed in tempo or mode, then the inclusion of hidden Markov models as a solution arises naturally from the problem.

Since we as comparative biologists are involved in a historical science, we will inevitably encounter single evolutionary events of large importance. However, it must be recognized that datasets which are susceptible to biases from singular events are not amenable to most phylogenetic comparative tests. Although here we have resolved the statistical biases associated with false correlations, there is no amount of methodological massaging that will allow for a satisfying test of macroevolutionary correlation between two synapomorphies. This is because comparative methods rely on several independent replicates of correlation such that the associations found between the variables may be considered robust even when extended beyond the dataset used for the analysis. If there is only one example of the correlation arising in the entire dataset, we should not have confidence in extending our inferences beyond the clade and should be wary of the correlation even within the focal clade. However, that is not to say there is no mechanistic reason for an association between synapomorphies. It is entirely possible that two characters which share identical evolutionary histories have an underlying biological link. Nonetheless, conclusions about the potential links between these characters cannot come from studies conducted at a macroevolutionary scale, and they should instead be investigated at a smaller scale (Beaulieu and O'Meara 2018, 2019; Donoghue and Edwards 2019). Additional lines of evidence and a more mechanistic explanation will be necessary in order for a conclusion of correlation to be satisfying (Gardner and Organ 2021). In a sense, the hidden rate classes of our proposed framework may represent lineage-specific factors that, once present, readily allow for a shift in the tempo and mode of a lineage's evolution (Maddison and FitzJohn 2015; Ogburn and Edwards 2015).

A broader methodological conclusion that can be drawn from our results, which have been echoed elsewhere (Beaulieu and O'Meara 2016; Caetano et al. 2018; O'Meara and

Beaulieu 2021), is that testing against simple null hypotheses is usually not a productive way to do science. Rather than testing for a binary outcome of whether or not correlation is present, it is often beneficial to examine what these models suggest about the evolutionary process. Utilizing model comparison and finding that correlation exists is certainly interesting, but the real utility of modeling macroevolutionary processes is interpreting parameters that could not have been identified from the pattern alone. Within reason, it is often possible to look at the distribution of two discrete characters and be able to say whether the two are correlated before doing any modeling. However, it is more difficult to specify numerical values for the rates at which these characters evolve. For example, neither a glance at the dataset nor summary statistics will be consistently informative as to how many orders of magnitude faster a lineage in state $Y_0$ evolves character $X$ than a lineage in state $Y_1$. Additionally, transition rates which are measured in changes per million years (more specifically, changes per time unit of the phylogenetic tree) are directly comparable across any comparative study. For instance, changes in flower color in one study can be compared directly to changes in mammalian diet in another, because the parameters of transition have the same unit (event per unit of time). With these parameter estimates we may more robustly test hypotheses based on a well-defined model of macroevolution (Pennell and Harmon 2013). Furthermore, an examination of parameter estimates applies to most commonly used macroevolutionary models. For example, Vasconcelos et al. (2021) tested a set of three hypotheses related to how the mode of seed dispersal related to climatic niche evolution using Ornstein-Uhlenbeck models. This is not new for these types of models, but a key point from this study is that the model support was not as important as the relative value of the parameters. Instead of examining whether model A was more supported than model B, they looked at how specific hypotheses (i.e., that abiotically dispersed seeds tend to have a more arid climatic

optima) were differentially supported across a diverse set of models. A focus on parameter estimates rather than relative model support underscores that we are uncertain about the best model, but we wish to estimate parameters which reflect that uncertainty and robustly relate them to our hypotheses. This insight spurred the adoption of model-averaging by comparative biologists, which is now recognized as vital for macroevolutionary studies (see Caetano et al. 2018).

**Conclusion**

Sparked by an appreciation of the limitations of PCMs, several commonly used phylogenetic comparative methods have seen critical challenges recently, which have led to advancements useful for both developers and users (Boettiger et al. 2012; Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Louca and Pennell 2020). Here, too, the critiques of classic tests of correlation (Pagel 1994) are not wrong, and the recommendations of past studies remain useful (Maddison and FitzJohn 2015; Uyeda et al. 2018; Gardner and Organ 2021). Instead, what we have demonstrated is that the statistical bias towards correlation is primarily due to a misspecification of the model set and a failure to account for character independent rate heterogeneity. We have highlighted that the inclusion of non-standard Markov models in the model set can be critical for the quality of the inferences being made. We acknowledge that choosing a diverse set of models *a priori* is not always straightforward, but both likelihood and Bayesian methods will only be as effective as the plausibility of the models set being analyzed (Burnham and Anderson 2002). We know that a homogeneous process over millions of years and across thousands of lineages is incorrect (Eldredge and Gould 1972) and that the individual parts of an organism do not evolve independently (Levins and Lewontin 1985). While we may

not be able to always specify each of these individual processes, we must try to incorporate them in our modeling. Accounting for rate heterogeneity through HMMs is a simplified way that we can bring realism to our modeling while also making statistically consistent and unbiased estimates of evolutionary parameters. From there, undoubtedly more work will be necessary (e.g., Goldberg and Foo 2020). But comparative analyses must at the very least attempt to account for what we know about macroevolution while making us aware of the wonderful idiosyncrasies of evolutionary history.

**References**

Beaulieu J.M., Donoghue M.J. 2013. Fruit Evolution and Diversification in Campanulid Angiosperms. Evolution. 67:3132–3144.

Beaulieu J.M., O'Meara B.C. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Syst Biol. 65:583–601.

Beaulieu J.M., O'Meara B.C. 2018. Can we build it? Yes we can, but should we use it? Assessing the quality and value of a very large phylogeny of campanulid angiosperms. Am. J. Bot. 105:417–432.

Beaulieu J.M., O'Meara B.C. 2019. Diversity and skepticism are vital for comparative biology: a response to Donoghue and Edwards (2019). Am. J. Bot. 106:613–617.

Beaulieu J.M., O'Meara B.C., Donoghue M.J. 2013. Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. Syst Biol. 62:725–737.

Boettiger C., Coop G., Ralph P. 2012. Is Your Phylogeny Informative? Measuring the Power of Comparative Methods. Evolution. 66:2240–2251.

Boyko J.D., Beaulieu J.M. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. Methods Ecol Evol. 12:468–478.

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Caetano D.S., O'Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models: HMM and the adequacy of SSE models. Evolution. 72:2308–2324.

Darwin C. 1859. On the Origin of Species, 1859. Routledge.

Donoghue M.J., Edwards E.J. 2019. Model clades are vital for comparative biology, and ascertainment bias is not a problem in practice: a response to Beaulieu and O'Meara (2018). Am. J. Bot. 106:327–330.

Eldredge N., Gould S.J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Models in paleobiology. 1972:82–115.

Felsenstein J. 1985. Phylogenies and the Comparative Method. Am. Nat. 125:1–15.

Felsenstein J., Churchill G.A. 1996. A Hidden Markov Model approach to variation among sites in rate of evolution. Mol. Biol. Evol. 13:93–104.

FitzJohn R.G. 2010. Quantitative Traits and Diversification. Syst Biol. 59:619–633.

Gardner J.D., Organ C.L. 2021. Evolutionary Sample Size and Consilience in Phylogenetic Comparative Analysis. Syst Biol. 70:1061–1075.

Goldberg E.E., Foo J. 2020. Memory in trait macroevolution. Am. Nat. 195:300–314.

Levins R., Lewontin R. 1985. The dialectical biologist. Harvard University Press.

Louca S., Pennell M.W. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature. 580:502–505.

Maddison W.P., FitzJohn R.G. 2015. The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. Syst Biol. 64:127–136.

Marazzi B., Ané C., Simon M.F., Delgado-Salinas A., Luckow M., Sanderson M.J. 2012. Locating Evolutionary Precursors on a Phylogenetic Tree. Evolution. 66:3918–3930.

Nakov T., Beaulieu J.M., Alverson A.J. 2019. Diatoms diversify and turn over faster in freshwater than marine environments. Evolution. 73:2497–2511.

Ogburn M.R., Edwards E.J. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm clade Montiaceae. Mol. Phylogenet. Evol. 92:181–192.

O'Meara B., Beaulieu J. 2021. Potential survival of some, but not all, diversification methods. Available from ecoevorxiv.org/w5nvd.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. B: Biol. Sci. 255:37–45.

Pagel M., Meade A. 2006. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. Am. Nat. 167:808–825.

Pennell M.W., Harmon L.J. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Annals of the New York Academy of Sciences. 1289:90–105.

Rabosky D.L., Goldberg E.E. 2015. Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. Syst Biol. 64:340–355.

Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst Biol. 67:1091–1109.

Vasconcelos T., Boyko J.D., Beaulieu J.M. 2021. Linking mode of seed dispersal and climatic niche evolution in flowering plants. J. Biogeogr. In press.

# Appendix

**Table 1.** Average ΔAIC values for 100 datasets with standard deviations shown in brackets. Each column represents a scenario described in the main text and each row represents a different Markov model structure which may be consistent with independence or correlation. For each scenario, 8 or 9 models were fit to the datasets. The collapsed model is fit only when not all potential state combinations are directly observed and therefore are not fit in modified scenarios. A ΔAIC of 0 indicates the best model and models within 2 AIC units of each other are generally considered good fits to the data (Burnham and Anderson 2002).

| *Scenario* | *Darwin's* | *Unreplicated bursts* | *Modified Darwin's (outside)* | *Modified Darwin's (inside)* | *Modified Darwin's (both)* |
|---|---|---|---|---|---|
| **Collapsed** | **0.0 (± 0.0)** | **0.0 (± 0.0)** | NA | NA | NA |
| **Independent** | 17.9 (± 12.3) | 36.8 (±9.0) | 14.3 (±3.6) | 14.8 (±4.0) | 15.6 (±4.7) |
| **Simplified independent** | 13.9 (±2.3) | 67.3 (±15.8) | 10.3 (±3.6) | 10.8 (±4.0) | 11.6 (±4.7) |
| **Correlated** | 12.0 (±0.2) | 8.0 (±0.1) | 6.1 (±0.5) | 6.1 (±0.7) | 10.8 (±2.6) |
| **Simplified correlated** | 13.9 (±2.3) | 30.0 (±8.2) | 9.8 (±3.6) | 10.4 (±4.1) | 11.6 (±4.7) |
| **Hidden Markov independent** | 20.8 (±6.8) | 9.2 (±0.4) | 7.9 (±1.2) | 8.4 (±3.3) | 7.8 (±2.2) |
| **Simplified hidden Markov independent** | 5.5 (±0.1) | 36.3 (±9.1) | **0.0 (±0.0)** | **0.0 (±0.0)** | **0.0 (±0.0)** |
| **Correlated hidden Markov** | 29.7 (±0.3) | 24.9 (±0.8) | 22.9 (±0.7) | 23.5 (±0.8) | 23.2 (±1.4) |
| **Simplified correlated hidden Markov** | 18.8 (±2.1) | 34.3 (±7.7) | 14.2 (±3.3) | 14.3 (±2.8) | 15.7 (±3.5) |

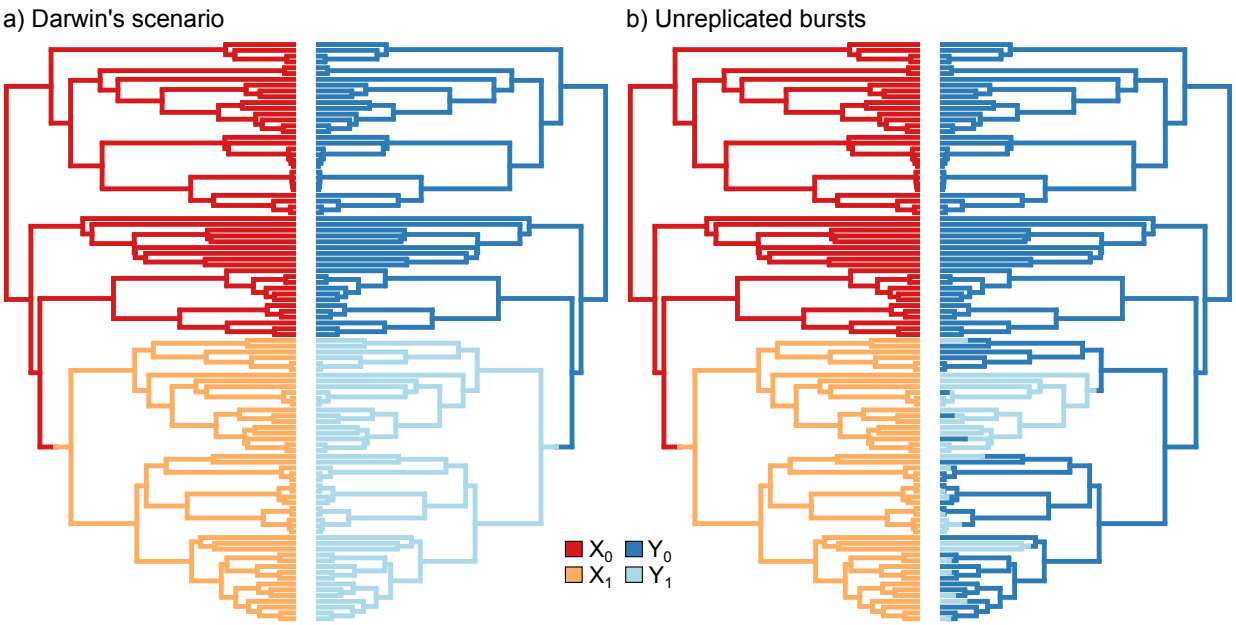


**Figure 1.** The two problematic scenarios from Maddison and FitzJohn (2015) for the evolution of characters $X$ and $Y$. Character $X$ is painted on the left phylogeny using red and orange for state $X_0$ and $X_1$, whereas character $Y$ is painted on the right phylogeny using dark blue and light blue for state $Y_0$ and $Y_1$. a) Darwin's scenario is depicted as a single event deep in time that has led to the co-distribution of $X_0Y_0$ outside of the focal clade and $X_1Y_1$ within the focal clade. b) Unreplicated bursts scenario is where a single event deep in time has led to the co-distribution of $X_0Y_0$ outside of the focal clade and $X_1Y_0$ and $X_1Y_1$ within the focal clade.

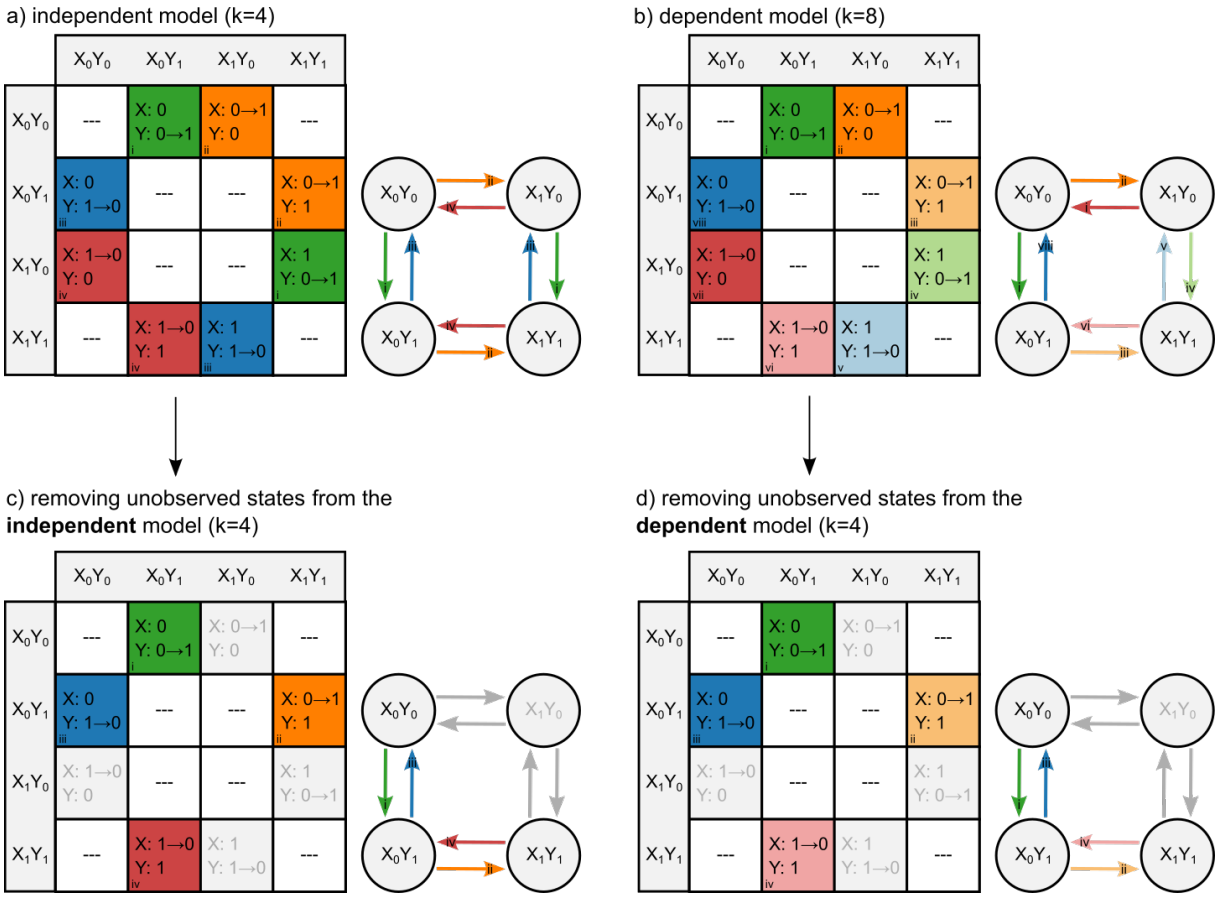


**Figure 2.** Representations of the different transition rate matrices, **Q**, with *k* number of parameters associated with each. Where transitions are fixed to occur at the same rate, the squares are colored to be the same. Unique parameters are also indicated with a roman numeral in the bottom left corner of the square. To the right of each matrix, a ball and stick representation of the model is presented with colors and parameter numbers matching the transitions indicated in the matrix, **Q**. The ball and stick representation is organized such that internal arrows represent transitions from 1 to 0, and external arrows represent transitions from 0 to 1. Additionally, arrows which cross the vertical midpoint indicate transitions in character *X*, whereas transitions across the horizontal midpoint indicate transitions in character *Y*. a) An independent model with four unique parameters, which fixes transitions within a character such that changes in *X* or *Y* do not depend on the state of the other character. b) A dependent model with eight unique parameters, whichs model allows transitions within a character to depend on the state of the other character. c) A model which removes transitions to and from an unobserved state from the independent model (a). d) A model that removes transitions to and from an unobserved state from the dependent model (b). In (c) and (d) the unobserved state is based on the Unreplicated Burst scenario where $X_0Y_1$ is not observed.

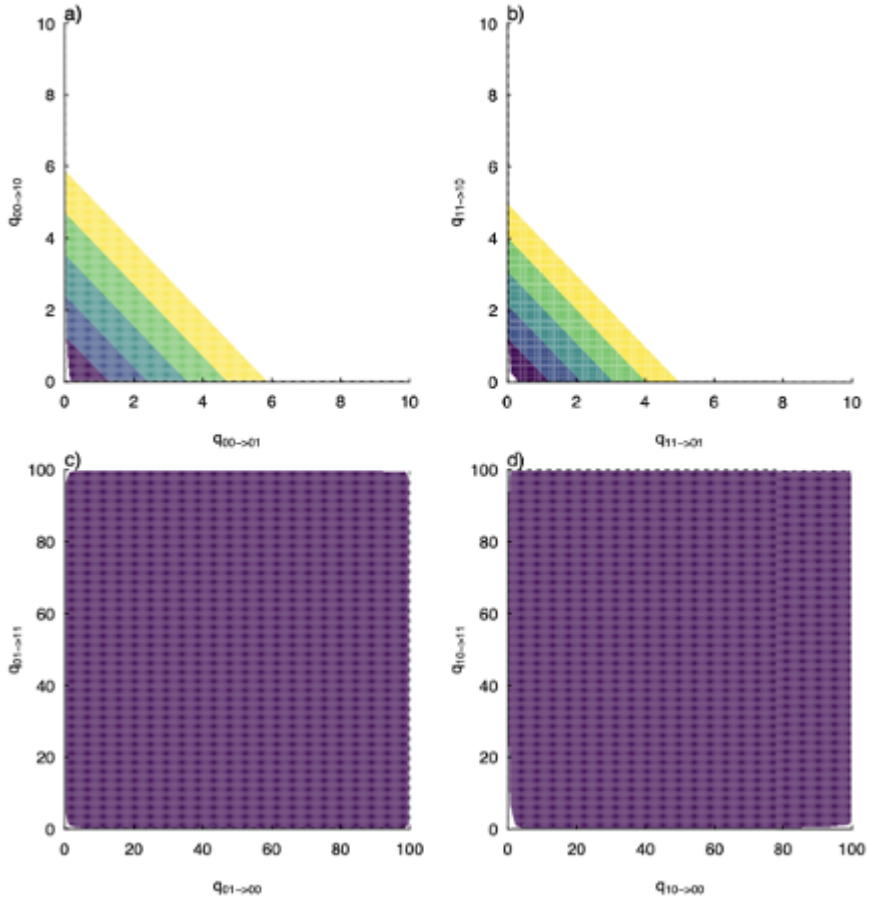


**Figure 3**. An example likelihood surface of a correlated model when applied to one of the 100 Darwin's scenario datasets. The color of the plot indicates the likelihood of a particular pair of parameters when the remaining transition rates are optimized. Thus, each point represents the maximum likelihood estimate when the transition rates indicated by the axes are fixed. a) Transitions from $X_0Y_0$ to an intermediate state result in several likelihood ridges. b) Transitions from $X_1Y_1$ to an intermediate state result in several likelihood ridges. c) Transitions from $X_0Y_1$ to either $X_0Y_0$ or $X_1Y_1$ result in a completely flat likelihood surface. d) Transitions from $X_1Y_0$ to either $X_0Y_0$ or $X_1Y_1$ result in a completely flat likelihood surface.

| | $X_{A0}Y_{A0}$ | $X_{A0}Y_{A1}$ | $X_{A1}Y_{A0}$ | $X_{A1}Y_{A1}$ | $X_{B0}Y_{B0}$ | $X_{B0}Y_{B1}$ | $X_{B1}Y_{B0}$ | $X_{B1}Y_{B1}$ |
|---|---|---|---|---|---|---|---|---|
| $X_{A0}Y_{A0}$ | --- | $X_A$: 0<br>$Y_A$: 0→1<br>i | $X_A$: 0→1<br>$Y_A$: 0<br>ii | --- | $X_0$: A→B<br>$Y_0$: A→B<br>ix | --- | --- | --- |
| $X_{A0}Y_{A1}$ | $X_A$: 0<br>$Y_A$: 1→0<br>iii | --- | --- | $X_A$: 0→1<br>$Y_A$: 1<br>ii | --- | $X_0$: A→B<br>$Y_1$: A→B<br>ix | --- | --- |
| $X_{A1}Y_{A0}$ | $X_A$: 1→0<br>$Y_A$: 0<br>iv | --- | --- | $X_A$: 1<br>$Y_A$: 0→1<br>i | --- | --- | $X_1$: A→B<br>$Y_0$: A→B<br>ix | --- |
| $X_{A1}Y_{A1}$ | --- | $X_A$: 1→0<br>$Y_A$: 1<br>iv | $X_A$: 1<br>$Y_A$: 1→0<br>iii | --- | --- | --- | --- | $X_1$: A→B<br>$Y_1$: A→B<br>ix |
| $X_{B0}Y_{B0}$ | $X_0$: B→A<br>$Y_0$: B→A<br>ix | --- | --- | --- | --- | $X_B$: 0<br>$Y_B$: 0→1<br>v | $X_B$: 0→1<br>$Y_B$: 0<br>vi | --- |
| $X_{B0}Y_{B1}$ | --- | $X_0$: B→A<br>$Y_1$: B→A<br>ix | --- | --- | $X_B$: 0<br>$Y_B$: 1→0<br>vii | --- | --- | $X_B$: 0→1<br>$Y_B$: 1<br>vi |
| $X_{B1}Y_{B0}$ | --- | --- | $X_1$: B→A<br>$Y_0$: B→A<br>ix | --- | $X_B$: 1→0<br>$Y_B$: 0<br>viii | --- | --- | $X_B$: 1<br>$Y_B$: 0→1<br>v |
| $X_{B1}Y_{B1}$ | --- | --- | --- | $X_1$: B→A<br>$Y_1$: B→A<br>ix | --- | $X_B$: 1→0<br>$Y_B$: 1<br>viii | $X_B$: 1<br>$Y_B$: 1→0<br>vii | --- |

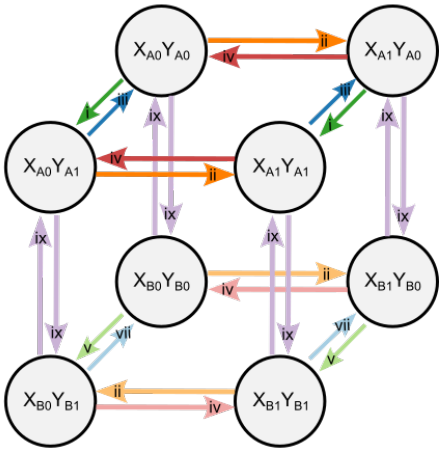


**Figure 4**. The hidden Markov independent model (HMIM), which allows transitions within a character to have rate heterogeneity without it necessarily being linked to an observed character. This matrix can be read as a block matrix, with 4x4 blocks representing transitions between observed characters following an independent model (top left and bottom right) and transitions between hidden rate classes *A* and *B* (top right and bottom left). The independent model is essentially duplicated in the top left (blue and green) and bottom right (red and orange) of the block matrix with transitions occurring between these different types of independent models (purple). Here, transition rates between the hidden states are fixed to be the same (parameter ix), but it is straightforward to allow the transition between rate class *A* and *B* to differ.

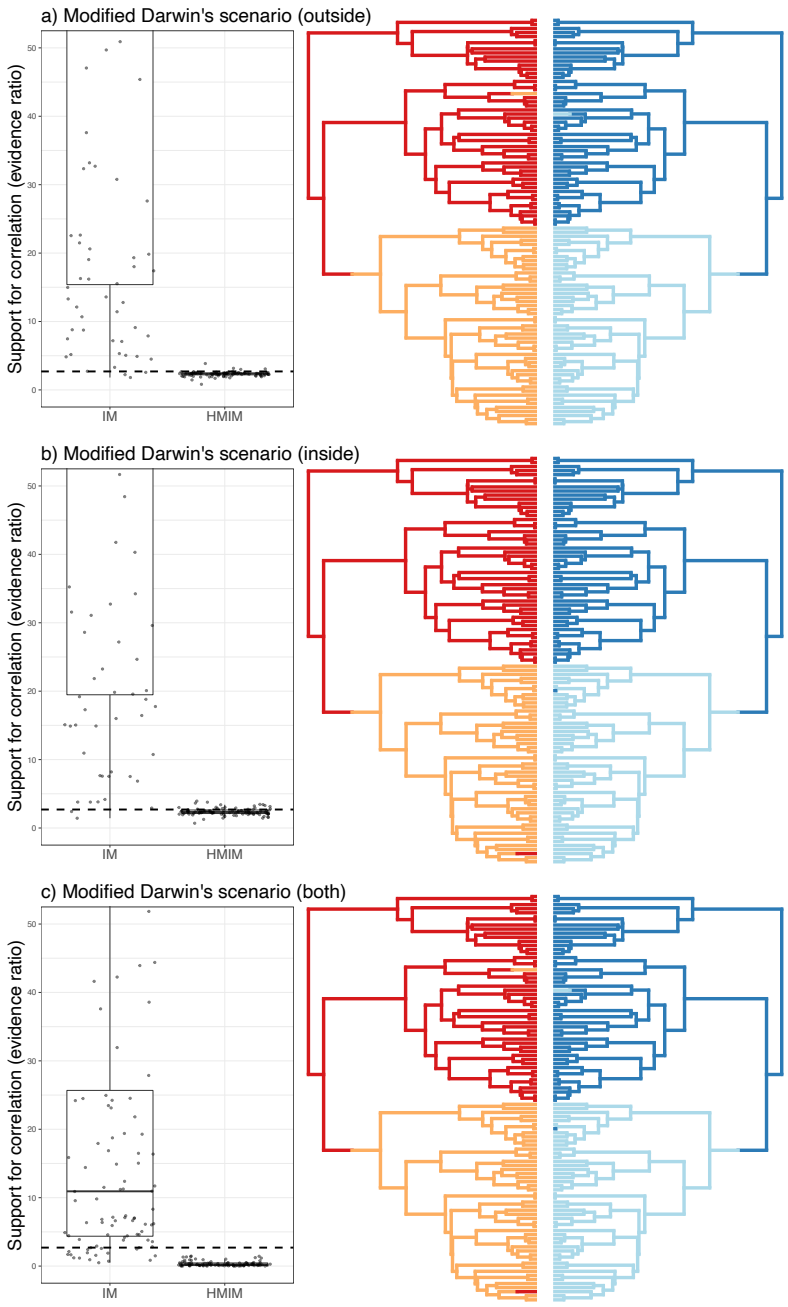


**Figure 5**. The amount of evidence for correlation when comparing a correlated model to ether an independent model (IM) or hidden Markov independent model (HMIM). The models are fit to data of the modified version of Darwin's scenario where a single observation of $X_0Y_1$ and $X_1Y_0$ is added outside of the focal clade (a), inside of the focal clade (b), and both within and outside of the focal clade (c). Evidence ratios for each m`odel comparison are plotted as boxplots to the left of the simulation scenario. In all cases, the evidence ratio of the correlated model over the independent model is substantially greater than 2.7 (left boxplot) but, the correlated model receives much less support over the hidden Markov independent model (right boxplot).

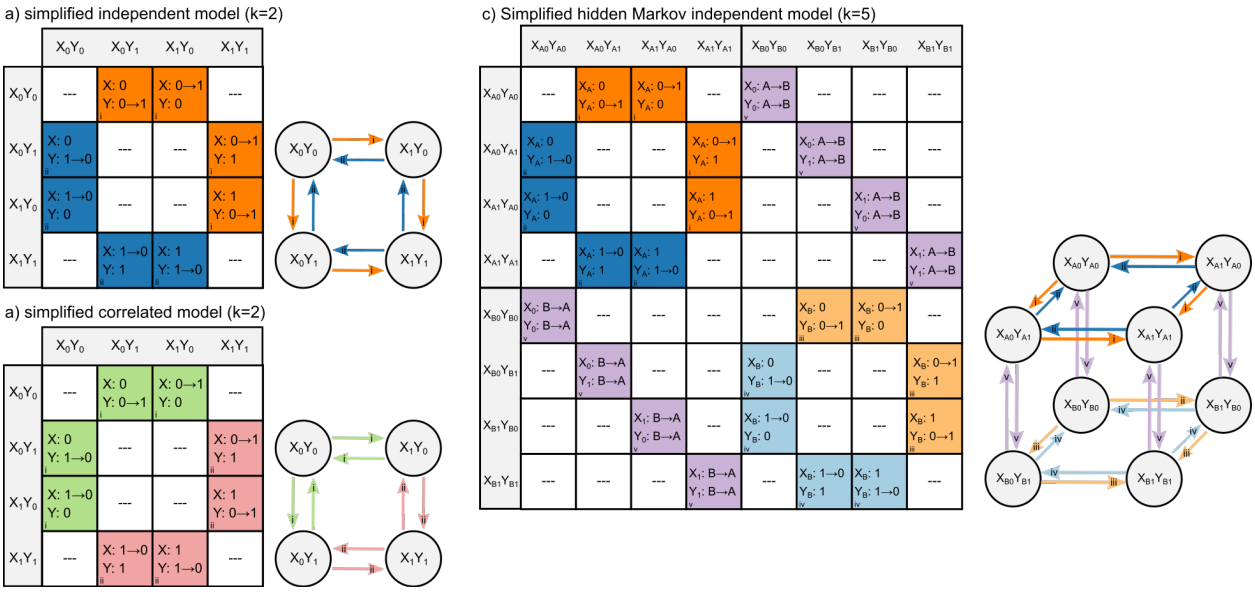


**Figure 6**. a) A simplified independent model. In this model, transitions from 0 to 1 all occur at the same rate and transitions from 1 to 0 all occur at the same rate. b) A simplified correlated model. Under this model, transitions between states of character *X* and *Y* depend on the background state of the other character. c) A simplified hidden Markov independent model, where the simple independent model of (a) is used in the hidden Markov framework which allows for rate heterogeneity independent of focal characters. The same can be done for the simple correlated model (not shown).

# Supplemental Figures

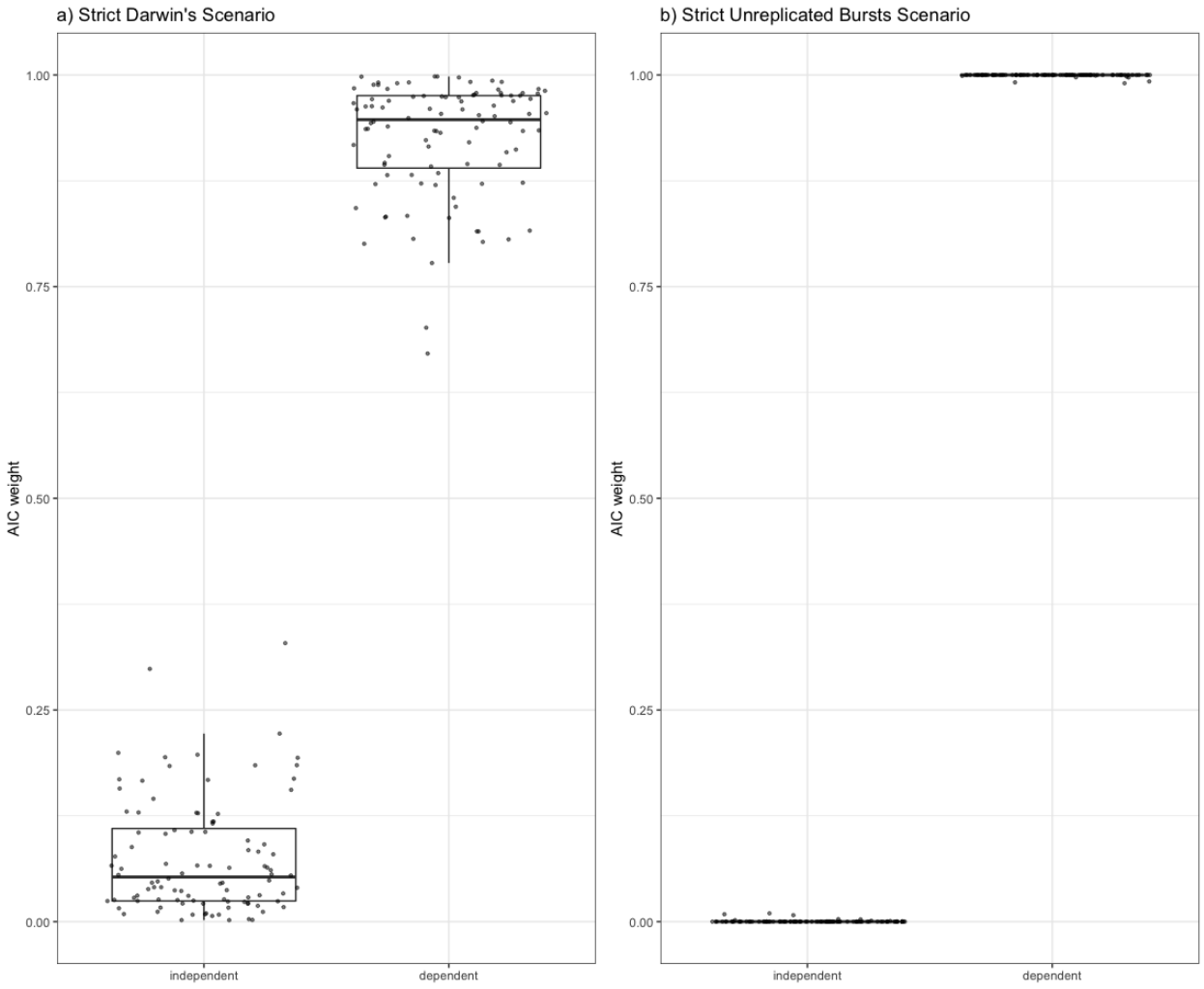


**Figure S1.** Replicated the Maddison and Fitzjohn (2015) result with our simulation and model fitting framework. Support for a dependent/ correlated model is consistently greater than an independent model.

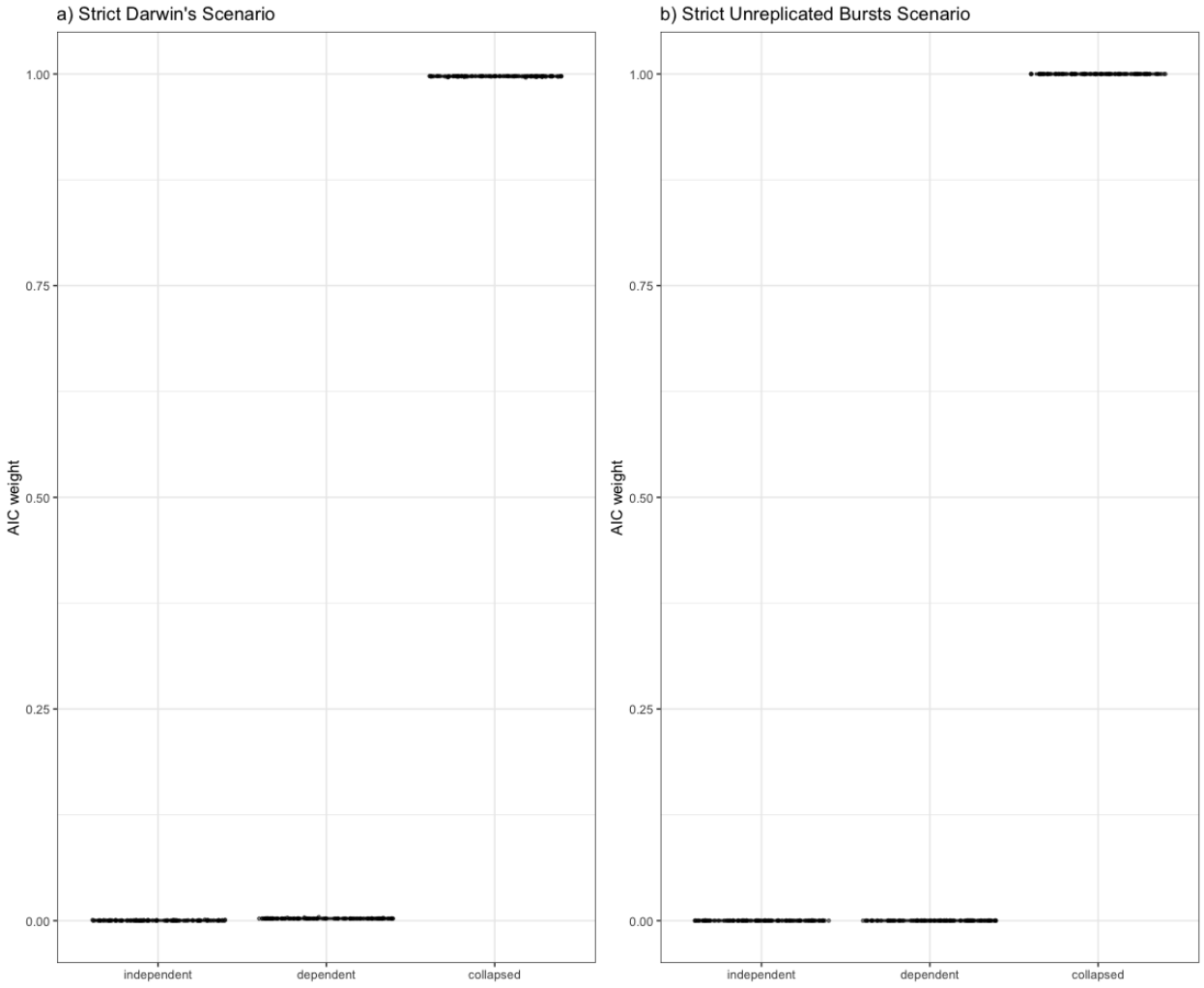


**Figure S2.** The same model set used by Maddison and Fitzjohn (2015), but with the inclusion of a collapsed model. Support for the collapsed model is overwhelming.

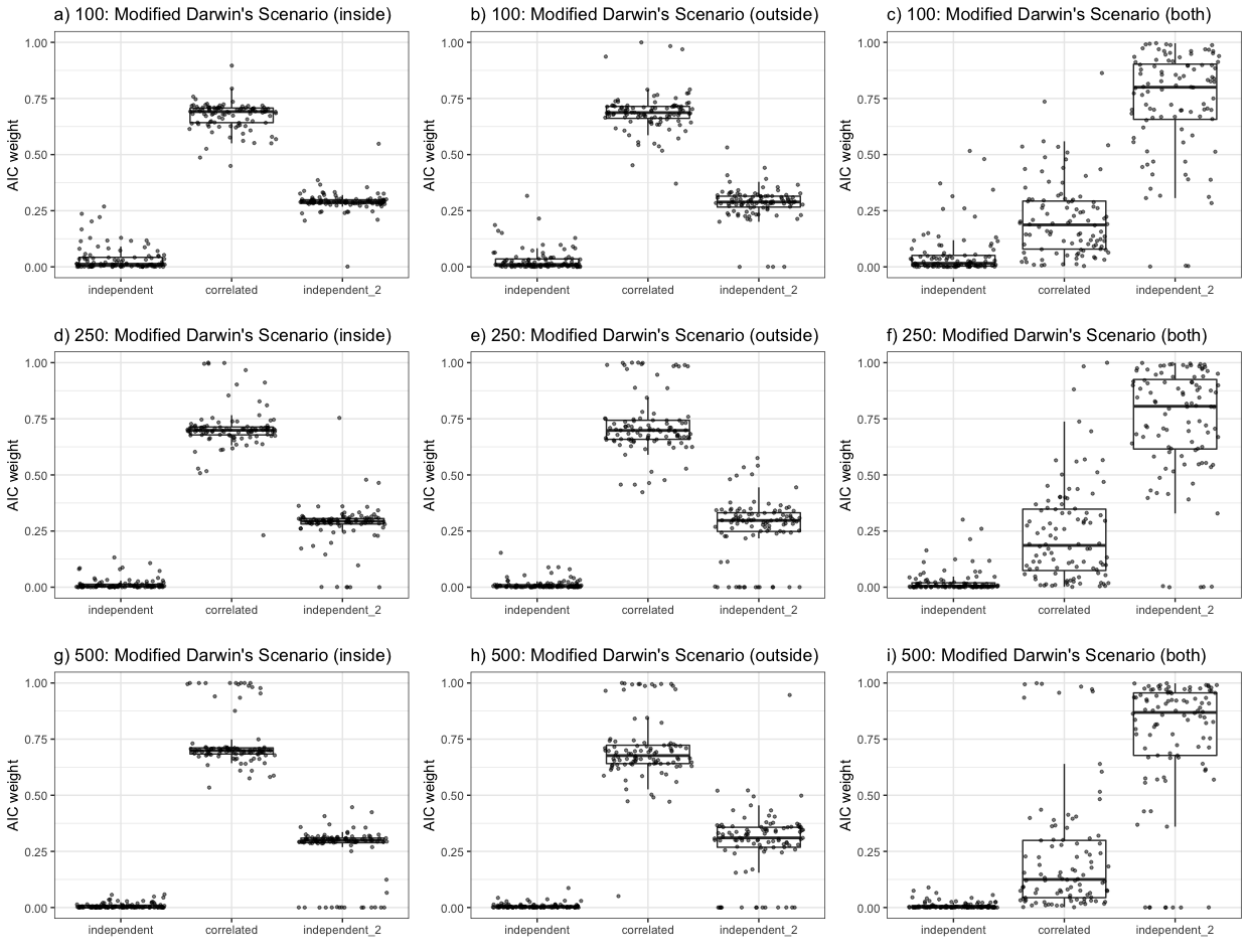


**Figure S3**. The effect of increasing the number of taxa on model support. Shown here are the two standard Pagel (1994) models (independent and correlated) as well as the unsimplified hidden state independent model. Support for the models is consistent across 100, 250, 500 taxa.

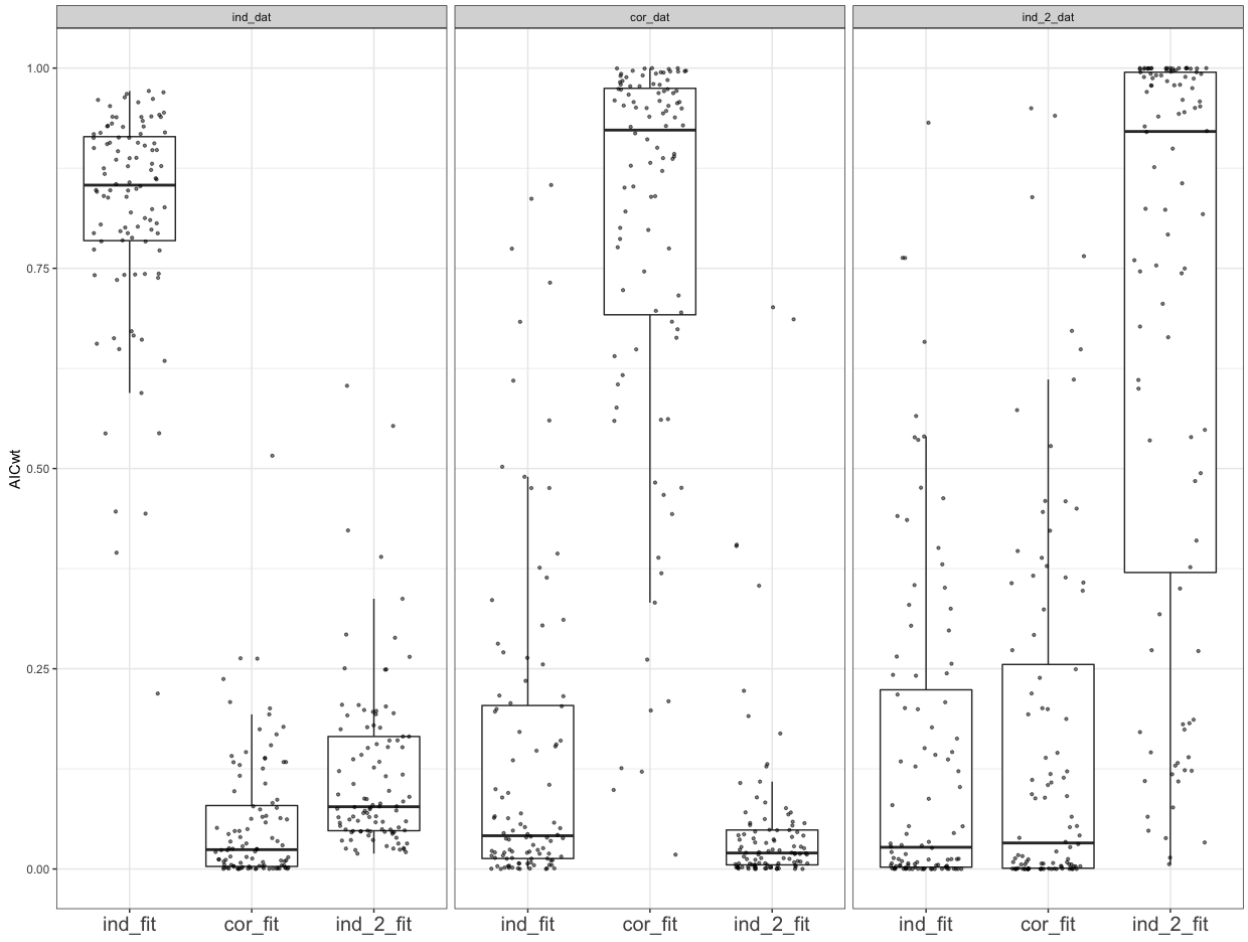


**Figure S4.** Akaike model weights are shown for data simulated under a simplified independent model (ind_dat), simplified correlated model (cor_dat) and simplified hidden Markov independent model (ind_2) for 100 unique datasets (See Figure 6 for model structure). For the simple independent and dependent models, the rates of evolution were 1 and 5 changes per million years. With the addition of the hidden states, we added rates of 2 and 10 for the second rate category as well as a transition rate of 4 between rate classes. Phylogenetic trees of 100 taxa were simulated with a birth rate of 1 and death rate of 0.75. Total branching time in the tree was rescaled to a total of 10 MY.

# CHAPTER III

## Jointly Modeling the Evolution of Discrete and Continuous Traits

James Boyko, Brian O'Meara, and Jeremy Beaulieu


## Abstract

Whether modeling the evolution of a discrete or continuous character, the focal trait of interest does not evolve in isolation, which requires comparative methods that can model multivariate evolution of several traits. However, most progress along these lines have involved multivariate evolution within the same class of character (i.e., either multivariate continuous or multivariate discrete) and there are significantly fewer options when jointly modeling traits when one trait is discrete and the other is a continuous character. Here we develop such a framework to explicitly estimate the joint likelihood for discrete and continuous characters. Specifically, our model combines the probability of observing the continuous character under a generalized OU process with the probability of the discrete character under a hidden Markov model, linked by a shared underlying regime. We use simulation studies to demonstrate that this approach, *hOUwie*, is able to accurately evaluate parameter values across a broad set of models. We then apply our model to test whether fleshy and dry fruits of Ericaceae lineages are correlated with their climatic niche evolution as represented by the aridity index. Consistent with our expectations, we find that dry fruits have higher rates of climatic niche evolution, that the climatic niche of fleshy fruits is more conserved and dry fruits have a more humid climatic optimum.

**Introduction**

A common theme in comparative biology is the detection of causal, or least mechanistic, factors that affect the evolution of quantitative characters. Questions of how plant life habit influence genome size evolution (Beaulieu et al. 2012), how substrate use alters limb length evolution (Mahler et al. 2013), or how tooth morphology slowly changes in response to habitat and diet (Toljagić et al. 2018) are all examples of testing whether evolutionary changes in a discrete variable may have altered evolutionary trajectories of a continuously varying trait. One very common phylogenetic comparative approach for these types of questions is to employ an Ornstein-Uhlenbeck (OU) model, which assumes distinct regimes, described by the evolution of a discrete character, are known completely *a priori* (e.g., Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012), or assumes that "shifts" in regimes can be inferred directly from the distribution of the continuous trait (e.g. Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al. 2016). While these approaches are practical, the discrete trait is assumed the driving force underlying the evolution of the continuous character. However, dependence rarely flows just one way in evolution, and we suspect that as often as a discrete character causes change in the continuous character, continuous characters also influence discrete character evolution, or at the very minimum, can provide information about how they may be evolving in tandem.

Progress along these lines has mostly involved acknowledging uncertainty in the evolution of the discrete character by fitting models over a large set of stochastically generated character mappings. That is, a large set of alternative reconstructions of the discrete character are obtained completely uninformed by the continuous trait's evolution, then the likelihood of the continuous character becomes the average of the likelihoods across these maps (e.g., Revell

2012). The advantage of this approach is that there is an explicit model for how regimes change through time, but the evolution of these regimes remains entirely independent of the continuous trait, and the probability of these regimes is not explicitly considered. For example, it is possible that the model that best fits the discrete data generates stochastic maps that does not provide a good fit to the continuous data.

A promising approach was recently described for detecting adaptive codon evolution (Jones et al 2020), where a set of maps obtained for a discrete phenotype under a standard Markov process is optimized along with parameters associated with genotype properties, thus forcing an emergent dependency between the two. Similarly, May and Moore (2020) developed a joint model for discrete and continuous characters under a state-dependent Brownian motion model. Their approach takes advantage of prior probabilities within a Bayesian framework to accommodate variation in the "background" rate of evolution in the continuous trait (i.e., rate variation across lineages that is independent of the discrete character under consideration). The novel Bayesian pipeline recently developed by Tribble et al. (2021) is the first attempt that we are aware of for jointly modeling discrete and continuous traits under an OU framework. Their approach samples discrete stochastic mappings informed by the discrete trait along with regime mappings which were informed by the continuous trait while accounting for the potential of hidden variation. While a more effective test of correlation between discrete and continuous characters, one drawback is that they do not explicitly account for the joint probability of the discrete and continuous parameter estimates together. They assume that the combination of independently estimated discrete and continuous models produces a joint estimate.

Here we develop and implement a framework that provides an explicitly joint estimate of the likelihood for a discrete and continuous character. Specifically, our model combines the

probability of the continuous character given a particular regime evolving under a generalized OU process, and the probability of that discrete regime painting obtained from an expanded set of Markov models, integrated over many regime paintings. We demonstrate how our framework, which we call *hOUwie*, can be used to test hypotheses of correlated evolution between discrete and continuous characters while also accounting for hidden character states and unobserved variation. Finally, we apply several *hOUwie* models to test the correlated dynamics of the mode of seed dispersal and climatic niche evolution and compare our results to those that did not account for the potential joint evolution of discrete and continuous variables.

## Methods

### *The hOUwie model*

Our model is composed of two processes: one describing the evolution of a discrete character and the other describing the evolution of a continuous character. To model the evolution of a single continuous character we use an Ornstein-Uhlenbeck (OU) model (Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012; Ho and Ané 2014a). Formally, the OU process is an Itô diffusion satisfying:

$$dX(t) = \alpha(\theta(t) - X(t)) + \sigma dB(t).$$

Conceptually, this model combines the stochastic evolution of a trait through time with a deterministic component that models the tendency for a trait to evolve towards an "optimum." In this model, the value of a trait, $X(t)$, is pulled towards an optimum, $\theta(t)$, at a rate scaled by the parameter $\alpha$. The optimum, $\theta(t)$, is a piecewise constant on intervals and takes values in a finite set $\{\theta_i\}$. This can represent the set of "selective regimes", "regimes", or Simpson's "adaptive zones" (Cressler et al. 2015), though it is consistent with a variety of true underlying

microevolutionary models (Hansen 2014). Additionally, random deviations are introduced by Gaussian white noise $dB(t)$, which is distributed as a normal random variable with mean zero and variance equal to $\sigma^2 dt$. Thus, $\sigma^2$ is a constant describing the rate of stochastic evolution away from the optimum. We use the set of extensions introduced by Beaulieu et al. (2012) and implemented in the R package *OUwie*, which allows for multiple primary optima $\theta(t)$ in which both the pull strength ($\alpha$) and the rate of stochastic evolution ($\sigma^2$) can vary across the phylogeny. However, the algorithm used to calculate the likelihood described in Beaulieu et al. (2012) involves a computationally costly matrix inversion procedure. Here we implement a linear-time computation of the likelihood of Gaussian trait models following (Ho and Ané 2014a). To do this, we first transform the phylogeny such that its variance covariance matrix, $V$, is 3-point structured. We can write the variance covariance matrix of the untransformed phylogeny as $V = D_u \tilde{V} D_u$, where following Beaulieu et al. (2012) and Ho and Ané (2014),

$$\tilde{V}_{ij} = \sum_{\gamma=1}^{\kappa(i,j)} \frac{\sigma^2_{ij,\gamma}}{2\alpha_{ij,\gamma}} (\mathrm{e}^{2\alpha_{ij,\gamma} s_{ij,\gamma}} - e^{2\alpha_{ij,\gamma} s_{ij,\gamma-1}}),$$

$$\text{and, } D_u = e^{\sum_{\gamma=1}^{\kappa(i)} \alpha_{i,\gamma}(s_{i,\gamma} - s_{i,\gamma-1})},$$

where, $s_\gamma$ is the distance from the root to the beginning of the selective regime ($\gamma$) for the $\kappa$ number of selective regimes along the path from the root to the last common ancestor of $i$ and $j$, $\kappa(i,j)$, or from the root to the terminal tip $i, \kappa(i)$. Our transformed phylogeny now has a variance covariance matrix $\tilde{V}_{ij}$ and diagonal matrix $D_u$. We can then calculate the quadratic quantities and determinant of $V$ (Ho and Ané 2014a). The probability of our continuous trait is given by

$$log(P(X|\, D, z, \vartheta, \psi)) = n \log(2\pi) + \log(det(V)) + \frac{P'V^{-1}P - 2P'V^{-1}Q + Q'V^{-1}Q}{2},$$

where $n$ is the number of tips in the phylogeny ($\psi$), $P$ is the continuous trait value of each species, and $Q$ is the expected value of each species given the continuous trait model calculated following equation (11) of Beaulieu et al. (2012), $D$ is the discrete character data, $z$ is a particular regime mapping, and $\vartheta$ are the parameters of the *hOUwie* model.

Next, we describe the calculation of the probability of the underlying regime structure, $\gamma$, that is the joint probability of discrete characters ($D$) and stochastic mapping ($z$). This calculation is analogous to the pathway likelihood of Steel and Penny (2000). Recently, May and Moore (2020) suggested that the joint probability of a regime structure and the discrete character is the product of the probabilities of exponentially distributed waiting times. By this definition, branch lengths are the sum of waiting times. But, when we calculate the probability of starting and ending a branch in state *i*, the likelihood of a regime structure was unaffected by the number of transitions (Supplemental Materials), even though the maximum likelihood estimate should be zero transitions (O'Meara 2008).

To calculate the probability of discrete characters ($D$) and stochastic mapping ($z$) we instead use an approximation. Our approximation relies on a finite number of degree-2 internodes and uses the standard Chapman-Kolmgorov equation to calculate the probabilities of beginning in a particular state $i$ and ending in state $j$ (Pagel 1994) and is identical to a joint probability of a set of state reconstructions (Yang 2006). As the number of internodes increase, the amount of time between nodes decreases and the approximation improves (Rao and Teh 2013). The joint probability of a regime structure and the discrete character *i*

$$P(D, z|Q, \psi) \;=\; P(x_0|Q, \psi) \prod_{\ell=1}^{n-1} P(z_\ell|Q, T_\ell),$$

where $\mathbf{Q}$ the instantaneous rate matrix ($\mathbf{Q} \in \vartheta$), $\psi$ is the phylogeny, $P(x_0|\mathbf{Q},\psi)$ is the root state probability (Pagel 1994; Yang 2006; Maddison et al. 2007), $n$ is the number of external nodes (tips), internal nodes, and internodes (degree-2 nodes) summed, $\ell$ indicates a particular branch, $P(z_\ell|\mathbf{Q},T_\ell) = e^{\mathbf{Q}T_\ell}\,\mathbb{1}_\gamma$, where $\mathbb{1}_\gamma$ is an indicator function which ensures that we only use the probability of states indicated by the specific the regime mapping instead of summing over all possible state combinations. The continuous character probability requires the discrete state(s) to be defined along the entire branch, thus we place transitions halfway between any two nodes.

For each set of parameters evaluated during the maximum likelihood search, a set of possible mappings of discrete states and continuous regimes are generated to evaluate the discrete and continuous likelihoods (Fig. 3). Ideally, we would calculate the likelihood by summing across all possible reconstructions (note that we want the sum across the reconstructions, not the single reconstruction with highest likelihood). The number of such reconstructions is very large (number of states ^ ((2*number of taxa-2)*(1+number of degree two internodes per edge))), which is particularly daunting as the sum must be calculated anew for every unique examined set of parameter values as part of search. We found in early work where we did look at this exhaustively that a few mappings made up the vast majority of the total likelihood, so we set up the analysis to focus on calculating total likelihood given the highest probability mappings.

To do this, we first approximate the conditional state probabilities at nodes. The conditional state probability, unlike the more common marginal reconstruction or joint state reconstruction (Pupko et al. 2000; Felsenstein 2004; Yang 2006), calculates the probability that a node has a particular state value conditioned only on the observations of its descendants. For a particular focal node, we calculate the probability of the observing all pairwise descendant values

given the OU model parameters, integrated over all possible rootward node states, and observed tipward discrete states (Fig 3d, see Supplemental for more detail). Although this is only an approximation of the conditional state probabilities, it proves to be an essential improvement over the typical procedure of sampling many stochastic maps based solely on the discrete process (Fig. 4). Next, the conditional probabilities of states at nodes are sampled starting with the root. Once the root is sampled, descendent states are sampled based on both the conditional ancestral values and the sampled ancestral state. This is achieved by multiplying the conditional probability of the node states by the probability of starting in the sampled rootward ancestral value and ending in any of the tipward states (the latter is calculated using familiar matrix exponentiation methods; e.g., Pagel 1994). Finally, under usual stochastic mapping procedures we would use rejection sampling (Nielsen 2002; Rao and I 2013) to simulate a path between the sampled rootward and tipward nodes. However, for increased computational efficiency, we opt to place transitions at pre-defined internodes. After nodes and internodes are sampled in step two, mappings are evaluated to ensure consistency with the discrete model (i.e., impossible transitions do not occur) and branches are painted based on the sampled nodes with transitions occurring half-way between nodes (and remember that a single edge may have multiple internodes placed on it).

Our function for the joint probability of a continuous and a discrete character is,

$$P(X, D|\vartheta, \psi) = \sum_{z} P(X|\, D, z, \vartheta, \psi) P(D, z|\vartheta, \psi),$$

where summing over all generated maps ($z$), $P(X|\, D, z, \vartheta, \psi)$ is the probability of the continuous character ($X$) given the discrete character data ($D$), mapping ($z$), hOUwie parameters ($\vartheta$), and phylogeny ($\psi$). $P(D, z|\vartheta, \psi)$ is the joint probability of the discrete character data ($D$) and stochastic mapping ($z$) given the *hOUwie* parameters ($\vartheta$) and phylogeny ($\psi$).

*The hOUwie model space*

Our simulation studies examined 22 possible *hOUwie* model structures for a binary discrete character, although the possible number of models is significantly higher because any number of discrete characters and states can be modeled together. For the discrete component of the model, we assumed that transitions between the observed characters were equal. We constrained transitions between hidden states to be the same for observed states, but this constraint can be relaxed if desired. The continuous model structures allowable in *hOUwie* are a generalized form of those allowed in *OUwie* and now include models in which only $\alpha$ varies (OUA), only $\sigma^2$ varies (OUV), and combinations of an OU and BM process (OUBM1 and OUBMV). We note that the OUBM1 model within *hOUwie* differs from The Ornstein–Uhlenbeck Brownian-motion (OUBM) model presented in Hansen et al. (2008) and Bartoszek et al. (2012) since the latter models are of multiple continuous characters, rather than different processes describing the same continuous character.

The potential model structures range from completely character-dependent to character-independent. Character-dependent (CD) models are models in which *any* continuous OU parameter differs between observed discrete state, whereas character-independent models (CID) test whether observed discrete states can be described by the same OU parameters. There are two types of character-independent model (Fig. 1). First, character-independent models include structures where there are no differences between any OU parameters. Under this model the entire evolutionary history of the clade can be described by a single $\alpha$, $\sigma^2$, and $\theta$ (Fig. 1a). To combat this unrealistic assumption we introduce a character-independent model which allows for differences in the OU parameters to depend upon an unobserved hidden state (CID+) and has been shown to correct for the bias towards detecting correlation (Boyko and Beaulieu 2022).

This addition allows for heterogeneity within the evolutionary process without the necessity of it being linked to a focal trait (Fig. 1c). In total we examine 22 unique model structures (2 CID, 10 CD, and 10CID+).

*Simulation study*

For each of the 22 *hOUwie* model structures, we simulated 50 datasets for phylogenies of 25, 100, and 250 taxa for a total of 3300 unique datasets. Phylogenies were pure birth phylogenetic trees with $\lambda = 1$, rescaled tree height to 1, and the root state was fixed to state 1. The parameters used to generate a phenotypic dataset depend on the structure of the generating model. For example, an OUM model and OU1 model can have identical $q_{ij}, \alpha,$ and $\sigma^2$, but they must differ in $\theta$ or else OUM will collapse into OU1 (model structures associated with model name are shown in Table 1). The simulating parameters were chosen to match Beaulieu et al. (2012) with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$. Once a phylogeny and phenotypic dataset were simulated, we fit our models to assess parameter estimation accuracy and model selection power. Although this represents a small subset of the potentially vast parameter space available to OU models, the behavior of these models has been thoroughly characterized and thus we chose parameters within the range of typical identifiability (Beaulieu et al. 2012; Ho and Ané 2014*a*; Cressler et al. 2015). Additionally, because *hOUwie* uses a variable number of mappings, we evaluate changing the number of stochastic maps. We fit each model using 25, 100, and 250 stochastic mappings per likelihood evaluation. Each dataset was evaluated using the true generating model, a BM1, an OU1, and either the character-dependent or character-independent counterpart to the generating model. For example, if the data were simulated under a character-dependent OUM model where the value of $\theta_1$ and $\theta_2$ depend on the observed character, a character-independent OUM model would also be fit as part of the model

set. Under the CID+ OUM model, a variable $\theta$ is still allowed, but it is unlinked to the focal character and thus should provide a more reliable character independent null hypothesis than BM1 or OU1 (Beaulieu and O'Meara 2016; Uyeda et al. 2018; May and Moore 2020; Boyko and Beaulieu 2022).

*The impact of climatic variables on seed dispersal*

For sedentary organisms, such as plants, dispersal is mainly limited to a brief stage of their life cycle and mediated mainly through the movement of seeds (Levin et al. 2003). Generally, the expectation is that seeds dispersed by frugivores are going to be dispersed to environments more like their parents' environment, whereas abiotically dispersed seeds are likely to be more erratic in their dispersal patterns (Schupp 1993; Westoby et al. 1996). Furthermore, it has been proposed that adaptations for frugivorous dispersal is linked to tropical and subtropical biomes, because in these warmer and wetter habitats, large trees create shady environments where competition for light is more important. A shadier habitat then imposes a selective pressure for larger seeds because more nutrients are needed for germination and initial survival (Foster and Janson 1985). However, the evolution of larger seeds comes with a tradeoff as they have a significantly lower dispersal potential (Howe and Smallwood 1982). Thus, we might expect that the climatic variables of a habitat influence the probability of transitioning between abiotic and biotic modes of dispersal, with transition rates from abiotic to biotic being greater in less arid environments.

Here we use dry or fleshy fruit morphology as a proxy for abiotic or biotic seed dispersal (Lorts et al. 2008) to evaluate three predictions outlined in Vasconcelos et al. (2021), but specifically measuring the aridity index. First, we expect that the climatic optima for fleshy fruits will be more humid compared to dry fruits ($\theta_{\mathrm{dry}} < \theta_{\mathrm{fleshy}}$). Second, we expect that dry fruits will have faster rates of climatic niche evolution ($\sigma^2_{\mathrm{dry}} > \sigma^2_{\mathrm{fleshy}}$). Finally, we expect that the

climatic niches of fleshy fruits will be more conserved through time ($\alpha_{dry} < \alpha_{fleshy}$). We apply several *hOUwie* models to test these hypotheses and compare our results to those discussed in Vasconcelos et al. (2021). We expect that any differences found between this study and Vasconcelos et al. (2021) are because we can explicitly account for the joint probability of the discrete and continuous characters. We focus our attention on Ericaceae specifically because Vasconcelos et al. (2021) found two counter-intuitive results. Namely, they found that the phenotypic optima of dry fruits were more humid than fleshy fruited lineages, and that the rate of climatic evolution was greater in fleshy fruits than dry fruits.

We included 25 *hOUwie* models within our model set: 2 CID, 10 CD, 10 CID+, and 3 HYB. *Gaultheria* is technically a dry-fruited genus within Ericaceae but has a persistent fleshy calyx that attracts frugivores (Stevens et al. 2004). However, since we are interested in the association between dispersal and fruit type, we code this as fleshy fruited within our dataset. Models are evaluated using the sample size corrected Akaike Information Criterion (AICc) and model averaging is conducted when discussing how our results relate to our hypotheses (Burnham and Anderson 2002). Measurement error is included for each model fit as within species variance (the sample-sized weighted average of the individual species variances following Labra et al. (2009) and Vasconcelos et al. (2021)). We evaluate then model averaged parameter estimates of $\theta$, $\sigma^2$, and $\alpha$ for fleshy and dry fruited lineages, as they relate to our hypotheses and compare our results to Vasconcelos et al. (2021). Finally, we conduct a parametric bootstrap of 100 simulated datasets to evaluate the standard error of our model averaged parameter estimates.

## Results

### *Simulation study*

For character-independent (CID) models, our heuristic adaptive sampling algorithm consistently produced more probable mappings than using purely discrete mappings for all models examined. On average, adaptive sampling produced mappings which were roughly 38 log likelihood units better than purely discrete sampling when examining joint probabilities. This was driven primarily by the improved continuous probabilities which were on average 38.4 log likelihood units better. In contrast, the discrete probability of each mapping was similar with discrete-only simulations producing maps that were on average 0.39 log likelihood units better (Table 1; Figure 3). For character-dependent models, the difference was negligible (not shown). This is because when the discrete and continuous character are linked, discrete-only mappings will match the continuous character's distribution quite well.

Most character-dependent models (CD) had lower overall deviations from the generating model across all model types. The RMSE was largest for alpha at 1.76 and 1.65 (if variable alpha) and errors were generally higher for more complex models. All other parameters had relatively similar RMSE, ranging from 0.1 for discrete the rate to 0.75 for $\sigma_2^2$. The BMV (BM with variable $\sigma$), OUV (OU with variable $\sigma$), OUA (OU with variable $\alpha$), and OUM (OU with variable $\theta$) models generally had the lowest errors, but there were some biases present. Most notably, alpha was biased upwards for OUM and OUV models and under variable alpha models (OUA, OUMA, OUVA, OUMVA), the difference between the alpha estimates tended to be larger than the generating parameter difference. The more complex models had larger error variances but showed similar biases as the simple models. Finally, OUBM models showed a significantly downward biased $\alpha$, suggesting BM like processes (Figure 5; Table 2).

Character-independent models with rate heterogeneity models generally performed well in terms of parameter estimates, but as expected, due to their inherit uncertainty, CID+ models had larger errors than CD models. The largest error was estimates of $\sigma_2^2$which had an RMSE of 8.5, although the median error value was only 0.03, suggesting that the large RMSE is driven by a long rightward tail of the estimates. Like CD models, $\alpha_1$ and $\alpha_2$ consistently showed the largest RMSE at 3.6 and 1.2. In general, $\alpha$ was underestimated with medians of -0.4 and -1.4 below the simulating values of 3 and 1.5. This means that models for CID+ models tended to be more BM like even under an OU generated data (Figure 5; Table 2).

Increasing the number of taxa examined improved both CD and CID+ performance. The RMSE for $\alpha$ was nearly cut in half between when moving from 25 tips to 250 tips from 5.2 to 2.8 under CID+ models. Nonetheless, some parameters continued to be estimated poorly, such as $\sigma_2^2$. Interestingly, increasing the number of stochastic maps improved CID+ performance, but did not substantially improve estimation under CD models (Fig. 5bc).

Generally, evidence of CD when it was the generating model was consistent across all model types. The lowest support for the OUA and OUBM1 models at an average AICwt of 0.31 and 0.13. For complex models, such as OUMVA, model support for was 0.81 and highest for OUMV at 0.97. CID+ models fared worse in terms of generating consistent support even when they were the generating model. Models which were difficult to estimate under character dependence were difficult to find consistent support for under character independence. The most extreme case was OUA model for which CID+ model was never chosen as the best supported model. However, models which performed well for CD tended to perform well under CID+. For example, OUM models garnered consistent support when with an average AICwt of 0.733 (Table 3; Figure 6).

For both CD and CID+ models, support improved when increasing the number of tips analyzed. Support for a CD model when CD was the generating model increased from $w_{CD} = 0.5$ to $w_{CD} = 0.67$ to $w_{CD} = 0.79$ for 25, 100, 250 tips and support for a CID+ model when it was the generating model increased from $w_{CID+} = 0.11$ to $w_{CID+} = 0.15$ to $w_{CID+} = 0.22$. Similarly, increasing the number of stochastic maps generally improved the fit, but not as much as increasing the number of tips. We found that the false evidence of correlation (as measured by the average AICwt of a character-dependent model when character-independence was the generating model) was generally not an issue for variable $\theta$ models (OUM*). Variable $\theta$ models had average AICwts for false character-dependence ranging from 0.03 to 0.23 and for none of our simulations models was a CD model best supported. Under a simple OUM model, CID+ models helped correct any potential bias with an average AICwt of 0.68. However, false evidence of correlation was an issue for variable $\sigma_i^2$and $\alpha_i$ models. False support for CD as measured by AIC weight ranged from 0.34 to 0.44 when $\theta$ was fixed and $\alpha_i$ and/or $\sigma_i^2$ varied. Although CID+ models did not garner much support when these models were fit, OU1 and BM1 models served as reasonable null hypotheses in these cases. In general, we found that when CID models were the generating model, evidence of CID was strongest and when CD models were the generating model, evidence of character dependence was strongest. This suggests that the effect of rate heterogeneity causing false correlations is not as pronounced as other comparative methods (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015).

*Seed dispersal and climatic evolution*

We found evidence of a character-dependent model over either a simple or hidden state character-independent model, suggesting a link between the climatic niche of Ericaceae lineages and their fruit type (Table 6). The best supported models were OUMVA and OUVA with AIC

weights of 0.41 and 0.32 respectively. This suggests that there were character dependent differences in phenotypic optima, rates of evolution, and overall phylogenetic signal. To evaluate support for our hypotheses we examined the model averaged parameter estimates (Table 7). The estimated optimum 0.81 for fleshy fruits suggests a more arid environment for their optimal habitat, and the 0.97 AI of dry fruits corresponds to a more humid environment (Middleton and Thomas 1997). However, both optima correspond to non-dryland humid environments. Both $\sigma^2$and $\alpha$ interact to create tip variance, so in addition to $\sigma^2$, we measured the stationary variance $V = \frac{\sigma^2}{2\alpha}$. As predicted, we found that Ericaceae lineages with dry fruits were more variable in their climatic niche evolution ($\sigma^2_{dry} = 0.011\ AI^2 MY^{-1}$, $V_{dry} = 0.37\ AI^2$) compared to fleshy fruits ($\sigma^2_{fleshy} = 0.007\ AI^2 MY^{-1}$, $V_{fleshy} = 0.15\ AI^2$). Additionally, the phylogenetic signal of fleshy fruited lineages was greater than dry fruited lineages ($\alpha_{fleshy} = 0.022 MY^{-1} > \alpha_{dry} = 0.014 MY^{-1}$). This corresponds to phylogenetic half-lives of $t_{1/2,dry} = 46.4$ MY and $t_{1/2,fleshy} = 30.3$ MY which are 38% and 25% of the total tree height respectively. Transitions to fleshy fruit occurred at 0.0015 transitions per million years which is more than 4.3 times faster than transitions to dry fruits (0.00035 transitions per million years). Given the total branch length in the tree is $10{,}120\ MY$, we would expect 15.6 transitions to fleshy fruit and 3.6 transitions to dry fruits to have occurred throughout the history of Ericaceae. Finally, on average lineages were in more arid environments than predicted by the model (average difference of 0.19 AI), with some species expected to be in much more humid environments (difference between current AI and optimal AI ranged from -4.4 to 0.85; Figure 7).

## Discussion

Phylogenetic comparative methods have been widely applied to study discrete and continuous characters separately. Due primarily to computational limitations there are few options which jointly evaluate both classes of character. The *hOUwie* framework proposed here overcomes these limitations, and we demonstrate how it is used to test hypotheses of correlated evolution between discrete and continuous characters while accounting for hidden character states and unobserved variation. Our model jointly models discrete and continuous characters by linking both via a common regime painting. However, unlike other similar methods, our likelihood formula explicitly calculates the probability of the underlying regimes. This has the advantage of describing the discrete character evolution probabilistically and allows information from the discrete and continuous characters to jointly contribute to the overall likelihood.

### *Relationship to existing methods*

Considerable progress has been made towards more realistic models of continuous character evolution within the last two decades. Continuous character models which initially relied on either single rate Brownian motion or simple Ornstein-Uhlenbeck models (Felsenstein 1985; Hansen 1997) have seen several extensions to allow for heterogeneity in the evolutionary process as well as the deterministic influence of underlying independent variables. Generally, these models can be classified as either being “hypothesis driven” or “data driven” (Martin et al. 2022). Hypothesis driven models are those which require *a priori* hypotheses regarding where evolutionary rates may differ throughout the phylogeny. These include models which have extended simple single-rate BM to incorporate rate variation based on discrete regime mappings (e.g., O’Meara et al. 2006; Thomas et al. 2006; Revell and Collar 2009; Caetano and Harmon 2017) or more generalized Ornstein-Uhlenbeck models where parameters are allowed to vary

based on an underlying regime mapping (e.g., Butler and King 2004; Bartoszek et al. 2012; Beaulieu et al. 2012). In contrast, several methods have focused on the development of data driven, shift-detection methods (which may indeed be used in testing hypotheses, but these hypotheses are not directly used in creating the regime map). These methods utilize an Ornstein-Uhlenbeck process to automatically detect where in the phylogeny evolutionary rates and phenotypic optima shift (Ingram and Mahler 2013b; Uyeda and Harmon 2014; Khabbazian et al. 2016; Bastide et al. 2017). Furthermore, some recently developed methods have allowed for rate variation without the assumption of constant regimes at all. Instead, these models assume the rates themselves evolve and change throughout the phylogeny under various Brownian motion-like processes (Lemey et al. 2010; Eastman et al. 2013; Revell 2021; Martin et al. 2022) or single optima Ornstein-Uhlenbeck processes (Hansen et al. 2008; Mitov et al. 2019). The method presented here is most like the latter group. *hOUwie* attempts to explicitly model the evolution of rate shifts according to regimes which jointly influence discrete and continuous character evolution. The regimes themselves are never fixed a priori and each is evaluated as a partial contribution to the overall probability of the data. The advantage of this approach is that it acknowledges the uncertainty in the underlying regime paintings and allows them to change through time.

Additionally, unlike *hOUwie*, the "hypothesis driven" or "data driven" models do not explicitly account for the joint modeling of the discrete and continuous characters. Most progress in this area has, until recently, been made via phylogenetic logistic regressions (Ives and Garland 2010) or threshold models in which the discrete character is modeled by a continuously varying unobserved lability (Felsenstein 2012; Revell 2014; Cybis et al. 2015). However, these models rely on more simplistic evolutionary models without character independent rate heterogeneity

(such as single rate Brownian motion). This lack of character independent rate heterogeneity has recently been recognized as a potential source of inflated correlation between discrete and continuous characters. Such was the reasoning for the MuSSCRat model (May and Moore 2020). Like *hOUwie*, MuSSCRat allows for character-independent rate heterogeneity following a multiple rate Brownian motion model to be directly contrasted against character correlation to correct for potential biases towards correlation. However, as we describe in detail above, the way the underlying discrete character is calculated in *hOUwie*, as well as how rate heterogeneity is modeled, differs substantially from May and Moore (2020). Finally, Tribble et al. (2021) has recently developed a method which is similar to the one presented here. One of the primary differences between *hOUwie* and the Bayesian pipeline discussed in Tribble et al. (2021) is how discrete character evolution is treated. Specifically, Tribble et al. (2021) assumed that character-independent mappings are generated under the same parameters which best fit their focal discrete character. In contrast, *hOUwie* allows the free estimation of character-independent discrete rates which best fit both discrete and continuous data. This difference may lead to biases against null models in the Tribble et al. (2021) approach since the character-independent regimes are forced to follow a character-dependent discrete model.

*Character-independent models and null hypotheses*

There is a growing appreciation that comparing constant-rate null models to variable-rate alternative models will consistently favor rate heterogeneity, regardless of whether there is a genuine association with a focal variable (Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Beaulieu and O'Meara 2016; Uyeda et al. 2018; O'Meara and Beaulieu 2021; Boyko and Beaulieu 2022). This problem, termed the "straw-man effect" by May and Moore (2020), has

been demonstrated to lead to nearly 100% error rates for evidence of discrete character correlation (Maddison and FitzJohn 2015; Boyko and Beaulieu 2022), and has severely biased evidence towards state-dependent speciation and extinction (Rabosky and Goldberg 2015; Beaulieu and O'Meara 2016). Given these often-overwhelming error rates in other comparative methods, we expected to find a similarly consistent bias towards correlation between discrete and continuous characters. However, we found that support for single rate character-independent null models was greater than character-dependent models even when simulated under character-independent models with rate heterogeneity. Although the inclusion of explicit multi-rate character independent models (CID+) models did help reduce evidence of false correlation in some cases, by and large, simplistic null models performed admirably. This is not to say that the error rates for discrete and continuous character correlation should be dismissed outright. If our simulations correctly assess that nearly one-third of results find false evidence of a correlation between continuous character rates of evolution and discrete characters, then better null models are certainly needed. But, in comparison to the profound effect that model misspecification has had in other comparative analyses (Beaulieu and O'Meara 2016; Boyko and Beaulieu 2022), the joint models tested here have substantially lower error rates.

We suspect that part of the reason that the correlation between discrete and continuous characters is less susceptible to "straw-man" effects than other PCMs is related to the inefficiency of sampling potential maps from the univariate stochastic mapping model. A common approach to fitting OU models involves simulating many stochastic maps to represent underlying regimes from parameters estimated only from the discrete character (Revell 2013). The resulting distribution of underlying regimes will therefore reflect a distribution appropriate for the discrete character, but not necessarily suitable for the continuous character. This is

especially true if the continuous character is unlinked to the focal discrete character. Indeed, we found that if the discrete and continuous characters are unlinked, most stochastic maps, even though good descriptions of the discrete characters, were completely inadequate representations of continuous regimes. Thus, any joint model with these maps contributed little to the overall likelihood. Under our simulation protocol, for a typical run, 90% of the total likelihood for the best set of parameters came from just 2% of the attempted maps.

In some ways the substantial contributions of only a few underlying regimes to the overall likelihood is good. First, it makes spurious links between a randomly distributed discrete character and a continuous character more unlikely since associations between regimes and continuous variables tend to be specific. This ultimately reduces the potential "straw-man" effect. Second, the continuous characters can inform the placement of shared regimes and therefore shift detection methods, where the continuous data are all that provides information about regimes shifts (Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al. 2016; Bastide et al. 2017), may be appropriate across a broad range of scenarios. However, this property also makes sampling a good set of regimes to get an accurate estimate of the likelihood difficult and is why the development of our adaptive sampling heuristics was necessary. Adaptive sampling, in combination with our approximation of the joint conditional distributions, helped make parameter estimation more accurate. Increasing the amount of sampled regime mappings is useful in improving precision (Fig 5c), at the cost of longer run time.

*Interplay of continuous, discrete, and hidden traits*

In many studies that deal with the correlation of discrete and continuous traits, it is often assumed that the discrete trait functions as the independent trait and the continuous trait as the

dependent trait. This assumption is baked into methods that map the discrete trait first and then analyze the continuous trait given these mappings, but it would be easy to fall into this form of thinking even with *hOUwie*, which does not have this assumption. Instead, *hOUwie* can help understand whether and how traits are correlated. For example, one could see if mammal body size correlates with trophic level: are hypercarnivores larger on average than herbivores? It could be that an herbivorous (discrete character) beaver evolves a taste for meat and then grows bigger (continuous character) so it can take down bigger prey; it could be that once things get to be the size of a bison (continuous character) they start adding more and more rodents to their diet, eventually becoming carnivores (discrete character). Causality can go both directions, and of course both traits may be evolving based on some other third trait and not functionally related to each other.

*hOUwie* is part of a series of hidden state models developed by our research groups (i.e., Beaulieu et al. 2013; Beaulieu and O'Meara 2016; Caetano et al. 2018; Boyko and Beaulieu 2021, 2022; Vasconcelos et al. 2022). One misconception we have noted in use of these methods is the thought that there is a single, discrete, hidden character in the biology. These models do model a single hidden character (with potentially many states), but this could be reflecting multiple characters evolving together or other factors that change in a heritable manner through time. It is a way to allow heterogeneity, especially by factors that vary by clades. With *hOUwie*, this heterogeneity can affect the discrete trait, the continuous trait, both, or neither.

*Seed dispersal and climatic niche evolution in Ericaceae*

Here we revaluated three hypotheses related to climatic niche evolution and seed dispersal and found that: (1) the climatic optima of dry fruits was more humid than fleshy fruits ($\theta_{fleshy} < \theta_{dry}$), (2) lineages with dry fruits had faster rates of climatic niche evolution ($\sigma^2_{dry} > \sigma^2_{fleshy}$),

and (3) climatic niches of fleshy fruits are more conserved through time ($\alpha_{dry} < \alpha_{fleshy}$). In contrast to previous findings, the higher rate and stationary variance of climatic niche evolution for dry seeds matched our original hypothesis (Vasconcelos et al. 2021). This is to be expected because abiotically dispersed seeds are likely to be more erratic in their dispersal patterns (Schupp 1993; Westoby et al. 1996). Additionally, that our results differ from previous findings (Vasconcelos et al. 2021) suggests that jointly modeling climatic niche evolution alongside fruit type changed our parameter estimation in a meaningful way.

Our final hypothesis, which stated that fleshy, biotically dispersed, seeds are more likely to be associated with humid environments, was not supported. However, it has been suggested that a trade-off between seed persistence, seed size, and dispersal strategies can be also common in arid environments (Venable and Brown 1988; Nunes et al. 2017). Specifically, large seed size may occasionally help withstand unfavorable conditions associated with increased aridity (Nunes et al. 2017). With an increased seed size, biotic seed dispersal and fleshy fruits, may become necessary for seed dispersal. This may be the case for Styphelieae, which is distributed in the arid Australian heathland and, of all predominately fleshy-fruited groups, lies the furthest from the inferred aridity optima. Additionally, it has been found that the proportion of abiotically dispersed seeds increases as elevation increases, due to the decreasing availability of frugivores (Chapman et al. 2016). Given that several radiations of Ericaceae lineages are associated with montane habitats (Schwery et al. 2015), it may be that the distribution of dry and fleshy fruits are a consequence of elevation rather than being directly linked to climatic niche evolution. Finally, it has been noted Ericaceae lineages are often found in well-leached soils and epiphytic habitats (Schwery et al. 2015). If associations with soil type are more important than links to climatic optima, we may expect that fruit-dependent climatic optima are consequence of unmodeled

factors. Although our modeling explicitly considers hidden variables that may lead to rate heterogeny, if the proposed hidden variable (soil condition) is closely linked to our modeled variable (aridity), then we may not be able to detect the presence of hidden variation. This may be the case between soil condition and aridity (Moreno-Jiménez et al. 2019).

*Caveats and possible extensions*

There are three important caveats to our proposed modeling framework. First, our discrete mapping probability, $P(D, z|\vartheta, \psi)$, is only an approximation. What we calculate is the probability of starting in a particular state $i$ and ending a particular state $j$, summed over all possible paths. However, the continuous model probability is based off a particular pathway history that is defined throughout the entire branch (Hansen 1997). Ultimately, this means that the underlying regimes are not treated identically for the continuous and discrete characters. The second caveat is that we do not force *hOUwie* to sum over all possible mappings $z$. This is because the number of mappings will grow exponentially as the number of nodes and internodes increases and the computation will quickly become infeasible (see Jones et al. 2020). Although this may not be entirely necessary since we have shown that only a small percentage of possible mappings contribute to the overall joint probability. Nonetheless, an ideal solution could be the use Markov-Modulated Ornstein-Uhlenbeck models (Huang et al. 2016) since this would remove the need for a regime mapping approach, but these have yet to be applied in phylogenetic comparative biology. *hOUwie* currently only deals with one discrete and one continuous trait at a time – a set of discrete traits can be handled by converting them to a single multistate character, but incorporating multiple continuous traits requires adding correlations between them. Finally, it is possible to extend *hOUwie* to include state-dependent speciation and extinction dynamics which have been shown to influence the distribution of discrete characters (Maddison 2006) and

would therefore influence continuous characters if the two were linked. However, this extension would require a different calculation of the underlying regime mapping probability. Approaches for stochastically mapping SSE models already exist (Freyman and Höhna 2019), so the largest remaining challenge of this extension would be generating high joint probability mappings.

**Conclusion**

The use of pre-defined discrete character mappings can be useful for testing hypotheses which rely on distinct, well-defined differences in the evolutionary histories of lineages. However, this approach assumes that the underlying mapping is known with complete accuracy and ignores the probabilistic nature of discrete regimes. *hOUwie*'s methodology integrates over the uncertainty of high probability character mappings and relies on the interpretation of parameter estimates from contrasting model structures to find evidence for hypotheses. Rather than assuming an *a priori* mapping, *hOUwie* can utilize the mutual information about the discrete and continuous characters to learn something about the underlying regimes evolution.

**References**

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. Journal of Theoretical Biology. 314:204–215.

Bastide P., Mariadassou M., Robin S. 2017. Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 79:1067–1093.

Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling Stabilizing Selection: Expanding the Ornstein–Uhlenbeck Model of Adaptive Evolution. Evolution. 66:2369–2383.

Beaulieu J.M., O'Meara B.C. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Syst Biol. 65:583–601.

Boyko J., Beaulieu J. 2022. A potential solution to the unresolved challenge of false correlation between discrete characters. .

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Butler M.A., King A.A. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist. 164:683–695.

Caetano D.S., Harmon L.J. 2017. ratematrix: An R package for studying evolutionary integration among several traits on phylogenetic trees. Methods in Ecology and Evolution. 8:1920–1927.

Chapman H., Cordeiro N.J., Dutton P., Wenny D., Kitamura S., Kaplin B., Melo F.P.L., Lawes M.J. 2016. Seed-dispersal ecology of tropical montane forests. Journal of Tropical Ecology. 32:437–454.

Cressler C.E., Butler M.A., King A.A. 2015. Detecting Adaptive Evolution in Phylogenetic Comparative Analysis Using the Ornstein–Uhlenbeck Model. Systematic Biology. 64:953–968.

Cybis G.B., Sinsheimer J.S., Bedford T., Mather A.E., Lemey P., Suchard M.A. 2015. ASSESSING PHENOTYPIC CORRELATION THROUGH THE MULTIVARIATE PHYLOGENETIC LATENT LIABILITY MODEL. Ann Appl Stat. 9:969–991.

Eastman J.M., Wegmann D., Leuenberger C., Harmon L.J. 2013. Simpsonian "Evolution by Jumps" in an Adaptive Radiation of Anolis Lizards. arXiv:1305.4216 [q-bio].

Felsenstein J. 1985. Phylogenies and the Comparative Method. Am. Nat. 125:1–15.

Felsenstein J. 2004. Inferring phylogenies. Sinauer associates Sunderland, MA.

Felsenstein J. 2012. A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model. The American Naturalist. 179:145–156.

Foster S., Janson C.H. 1985. The relationship between seed size and establishment conditions in tropical woody plants. Ecology. 66:773–780.

Freyman W.A., Höhna S. 2019. Stochastic Character Mapping of State-Dependent Diversification Reveals the Tempo of Evolutionary Decline in Self-Compatible Onagraceae Lineages. Systematic Biology. 68:505–519.

Hansen T.F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. Evolution. 51:1341–1351.

Hansen T.F. 2014. Use and Misuse of Comparative Methods in the Study of Adaptation. In: Garamszegi L.Z., editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 351–379.

Hansen T.F., Pienaar J., Orzack S.H. 2008. A Comparative Method for Studying Adaptation to a Randomly Evolving Environment. Evolution. 62:1965–1977.

Ho L. si, Ané C. 2014a. A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. Syst Biol. 63:397–408.

Ho L.S.T., Ané C. 2014b. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. Methods in Ecology and Evolution. 5:1133–1146.

Howe H.F., Smallwood J. 1982. Ecology of Seed Dispersal. Annual Review of Ecology and Systematics. 13:201–228.

Huang G., Jansen H.M., Mandjes M., Spreij P., Turck K.D. 2016. Markov-modulated Ornstein-Uhlenbeck processes. Advances in Applied Probability. 48:235–254.

Ingram T., Mahler D.L. 2013a. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods in ecology and evolution. 4:416–425.

Ingram T., Mahler D.L. 2013b. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods in Ecology and Evolution. 4:416–425.

Ives A.R., Garland T. 2010. Phylogenetic Logistic Regression for Binary Dependent Variables. Syst Biol. 59:9–26.

Jones C.T., Youssef N., Susko E., Bielawski J.P. 2020. A Phenotype–Genotype Codon Model for Detecting Adaptive Evolution. Systematic Biology. 69:722–738.

Khabbazian M., Kriebel R., Rohe K., Ané C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. Methods in Ecology and Evolution. 7:811–824.

Lemey P., Rambaut A., Welch J.J., Suchard M.A. 2010. Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time. Molecular Biology and Evolution. 27:1877–1885.

Levin S.A., Muller-Landau *Helene C., Nathan *Ran, Chave *Jérôme. 2003. The Ecology and Evolution of Seed Dispersal: A Theoretical Perspective. Annual Review of Ecology, Evolution, and Systematics. 34:575–604.

Lorts C.M., Briggeman T., Sang T. 2008. Evolution of fruit types and seed dispersal:A phylogenetic and ecological snapshot. Journal of Systematics and Evolution. 46:396.

Maddison W.P. 2006. Confounding Asymmetries in Evolutionary Diversification and Character Change. Evolution. 60:1743–1746.

Maddison W.P., FitzJohn R.G. 2015. The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. Syst Biol. 64:127–136.

Maddison W.P., Midford P.E., Otto S.P., Oakley T. 2007. Estimating a Binary Character's Effect on Speciation and Extinction. Syst Biol. 56:701–710.

Mahler D.L., Ingram T., Revell L.J., Losos J.B. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. Science. 341:292–295.

Martin B.S., Bradburd G.S., Harmon L.J., Weber M.G. 2022. Modeling the Evolution of Rates of Continuous Trait Evolution. :2022.03.18.484930.

May M.R., Moore B.R. 2020. A Bayesian Approach for Inferring the Impact of a Discrete Character on Rates of Continuous-Character Evolution in the Presence of Background-Rate Variation. Syst Biol. 69:530–544.

Middleton N., Thomas D. 1997. World atlas of desertification.. ed. 2. .

Mitov V., Bartoszek K., Stadler T. 2019. Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models. Proceedings of the National Academy of Sciences. 116:16921–16926.

Moreno-Jiménez E., Plaza C., Saiz H., Manzano R., Flagmeier M., Maestre F.T. 2019. Aridity and reduced soil micronutrient availability in global drylands. Nat Sustain. 2:371–377.

Nielsen R. 2002. Mapping Mutations on Phylogenies. Systematic Biology. 51:729–739.

Nunes A., Köbel M., Pinho P., Matos P., Bello F. de, Correia O., Branquinho C. 2017. Which plant traits respond to aridity? A critical step to assess functional diversity in Mediterranean drylands. Agricultural and Forest Meteorology. 239:176–184.

O'Meara B. 2008. Using Trees: Myrmecocystus Phylogeny and Character Evolution and New Methods for Investigating Trait Evolution and Species Delimitation (PhD Dissertation). Nat Prec.:1–1.

O'Meara B., Beaulieu J. 2021. Potential survival of some, but not all, diversification methods. .

O'Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for Different Rates of Continuous Trait Evolution Using Likelihood. Evolution. 60:922–933.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. B: Biol. Sci. 255:37–45.

Pupko T., Pe I., Shamir R., Graur D. 2000. A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences. Mol Biol Evol. 17:890–896.

Rabosky D.L., Goldberg E.E. 2015. Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. Syst Biol. 64:340–355.

Rao V., Teh Y.W. 2013. Fast MCMC Sampling for Markov Jump Processes and Extensions. :26.

Revell L.J. 2013. A Comment on the Use of Stochastic Character Maps to Estimate Evolutionary Rate Variation in a Continuously Valued Trait. Systematic Biology. 62:339–345.

Revell L.J. 2021. A variable-rate quantitative trait evolution model using penalized-likelihood. PeerJ. 9:e11997.

Revell L.J., Collar D.C. 2009. Phylogenetic Analysis of the Evolutionary Correlation Using Likelihood. Evolution. 63:1090–1100.

Schupp E.W. 1993. Quantity, quality and the effectiveness of seed dispersal by animals. Vegetatio. 107:15–29.

Schwery O., Onstein R.E., Bouchenak-Khelladi Y., Xing Y., Carter R.J., Linder H.P. 2015. As old as the mountains: the radiations of the Ericaceae. New Phytologist. 207:355–367.

Steel M., Penny D. 2000. Parsimony, Likelihood, and the Role of Models in Molecular Phylogenetics. Mol Biol Evol. 17:839–850.

Stevens P.F., Luteyn J., Oliver E.G.H., Bell T.L., Brown E.A., Crowden R.K., George A.S., Jordan G.J., Ladd P., Lemson K., Mclean C.B., Menadue Y., Pate J.S., Stace H.M., Weiller C.M. 2004. Ericaceae. In: Kubitzki K., editor. Flowering Plants · Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales. Berlin, Heidelberg: Springer. p. 145–194.

Thomas G.H., Freckleton R.P., Székely T. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proceedings of the Royal Society B: Biological Sciences. 273:1619–1624.

Toljagić O., Voje K.L., Matschiner M., Liow L.H., Hansen T.F. 2018. Millions of Years Behind: Slow Adaptation of Ruminants to Grasslands. Syst Biol. 67:145–157.

Tribble C.M., May M.R., Jackson-Gain A., Zenil-Ferguson R., Specht C.D., Rothfels C.J. 2021. Unearthing modes of climatic adaptation in underground storage organs across Liliales. .

Uyeda J.C., Harmon L.J. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Systematic biology. 63:902–918.

Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018. Rethinking phylogenetic comparative methods. Syst Biol. 67:1091–1109.

Vasconcelos T., Boyko J.D., Beaulieu J.M. 2021. Linking mode of seed dispersal and climatic niche evolution in flowering plants. J. Biogeogr. n/a.

Venable D.L., Brown J.S. 1988. The Selective Interactions of Dispersal, Dormancy, and Seed Size as Adaptations for Reducing Risk in Variable Environments. The American Naturalist. 131:360–384.

Westoby M., Leishman M., Lord J. 1996. Comparative ecology of seed size and dispersal. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 351:1309–1318.

Yang Z. 2006. Computational molecular evolution. Oxford University Press Oxford.

# Appendix

**Table 1:** A comparison of the effectiveness of the adaptive sampling procedure and standard discrete only sampling of maps. Regardless of the sampling procedure, all probabilities are calculated in the same way and so any differences in probabilities reflects each procedure's ability to generate appropriate mappings. 50 stochastic mappings are used to calculate the likelihood of the parameters. For each model type, data are simulated following our methods with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$. The generating parameters are used to evaluate probability of each dataset and thus the probabilities represented here are not necessarily the same as those derived from the MLE. Generally, adaptive sampling improves the joint estimate by improving the probability of the continuous character and is most effective for variable $\theta$ models. As expected, discrete only sampling produces regime paintings which better reflect the discrete character than adaptive sampling, but the difference is minor.

| Model class | Model type | Sampling procedure | Discrete marginal $log_e$ likelihood | Continuous marginal $log_e$ likelihood | Joint $log_e$ likelihood |
|---|---|---|---|---|---|
| CID+ | BMV | adaptive sampling | -16.48 | 10.54 | **-10.59** |
| | | discrete only | -16.43 | 9.19 | -10.59 |
| | OUA | adaptive sampling | -15.46 | 44.34 | **25.14** |
| | | discrete only | -15.53 | 43.11 | 24.96 |
| | OUV | adaptive sampling | -30.89 | 47.86 | **12.17** |
| | | discrete only | -30.14 | 46 | 12.11 |
| | OUVA | adaptive sampling | -11.88 | 36.91 | **21.14** |
| | | discrete only | -11.17 | 36.27 | 21.08 |
| | OUM | adaptive sampling | -11.94 | 57.57 | **39.08** |
| | | discrete only | -11.19 | 53.56 | 32.21 |
| | OUMA | adaptive sampling | -9.94 | 35.01 | **17.39** |
| | | discrete only | -9.38 | 2.19 | -20.48 |
| | OUMV | adaptive sampling | -19.96 | 20.77 | **-15.64** |
| | | discrete only | -14.76 | -2.92 | -25.83 |
| | OUMVA | adaptive sampling | -13.91 | 25.47 | **7.48** |
| | | discrete only | -13.23 | 26.36 | 4.48 |
| | OUBM1 | adaptive sampling | -14.26 | 42.2 | **24.39** |
| | | discrete only | -14.88 | 40.89 | 24.22 |
| | OUBMV | adaptive sampling | -19.17 | 49.1 | **18.84** |
| | | discrete only | -19.01 | 33.45 | 7.71 |

**Table 2:** The average accuracy of *hOUwie* parameter estimates across several model classes and types as measured by root-mean-square error (RMSE). RMSE is calculated for each model type by taking the square root of the mean squared error (MSE), where MSE is the average squared difference between the MLE and the simulating parameters. Data is generated with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$, and for phylogenies with 25, 100, and 250 taxa. Finally, model fits use either 25, 100, or 250 stochastic maps per likelihood iteration. The table shown here calculates RMSE integrating over all phylogenetic tree sizes and number of stochastic maps (n=8217). Dashes indicate a parameter that is not estimated for a given model type. Generally, character independent (CID+) models had higher errors than character dependent (CD) models. The greatest errors occurred when estimating alpha in variable alpha models for both CD and CID+ model classes. Estimates of the optimum and transition rates generally had the lowest errors.

| Model class | Model type | RMSE $q$ | RMSE $\alpha_1$ | RMSE $\alpha_2$ | RMSE $\sigma_1^2$ | RMSE $\sigma_2^2$ | RMSE $\theta_1$ | RMSE $\theta_2$ |
|---|---|---|---|---|---|---|---|---|
| CD | BMV | 0.12 | - | - | 0.1 | 0.28 | 0.22 | - |
| | OUV | 0.11 | 1.27 | - | 0.15 | 0.33 | 0.05 | - |
| | OUA | 0.12 | 1.55 | 1.63 | 0.11 | - | 0.06 | - |
| | OUM | 0.13 | 1.49 | - | 0.1 | - | 0.07 | 0.13 |
| | OUVA | 0.09 | 1.44 | 1.11 | 0.14 | 0.98 | 0.06 | - |
| | OUMV | 0.16 | 1.82 | - | 0.16 | 0.32 | 0.07 | 0.17 |
| | OUMA | 0.15 | 2.11 | 2.48 | 0.28 | - | 0.12 | 0.5 |
| | OUMVA | 0.18 | 1.62 | 1.12 | 0.12 | 1.07 | 0.76 | 1.06 |
| | OUBM1 | 0.1 | 2.64 | - | 0.08 | - | 0.08 | - |
| | OUBMV | 0.09 | 2.29 | - | 0.13 | 2.37 | 0.08 | - |
| CID+ | BMV | 0.05 | - | - | 0.27 | 10.11 | 0.24 | - |
| | OUV | 0.04 | 1.13 | - | 0.32 | 1.83 | 0.05 | - |
| | OUA | 0.05 | 2.93 | 1.34 | 0.33 | - | 0.07 | - |
| | OUM | 0.09 | 2.53 | - | 0.15 | - | 0.44 | 0.2 |
| | OUVA | 0.05 | 1.26 | 1.11 | 0.27 | 13.44 | 0.07 | - |
| | OUMV | 0.1 | 2.5 | - | 0.16 | 2.12 | 1.3 | 0.68 |
| | OUMA | 0.05 | 8.28 | 1.27 | 0.23 | - | 5.88 | 0.8 |
| | OUMVA | 0.07 | 5.54 | 1.24 | 0.2 | 9.37 | 8.76 | 1.35 |
| | OUBM1 | 0.05 | 3.33 | - | 0.32 | - | 0.14 | - |
| | OUBMV | 0.05 | 3.5 | - | 0.27 | 8.79 | 0.14 | - |

**Table 3**: AIC weights summarizing the average support for each model class when they are the generating model. Data is generated with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$ for phylogenies with 25, 100, and 250 taxa and model fits using either 25, 100, or 250 stochastic maps per likelihood iteration. When the generating model class is character dependent (CD) or character independent (CID+) we expect that the AICwt will be highest for that model when fit. Character dependent models generally show that pattern, however CID+ models generally perform poorly. An additional concern is datasets simulated by a character independent model with rate heterogeneity (datasets generated by a CID+ model) are best fit by CD models – which would be a spurious correlation. Although there was often some signal of character dependence in these models (AICwt of CD when CID+ is generating), most of the AIC weight was for simple character independent models (BM1 or OU1).

| Generating model class | Generating model type | AICwt of BM1 | AICwt of OU1 | AICwt of CD | AICwt of CID+ | Proportion generating model chosen as best |
|---|---|---|---|---|---|---|
| CD | BMV | 0.18 | 0.17 | **0.64** | 0.02 | 0.62 |
| | OUV | 0.03 | 0.22 | **0.74** | 0.02 | 0.73 |
| | OUA | 0.07 | **0.56** | 0.31 | 0.06 | 0.15 |
| | OUM | 0.04 | 0.02 | **0.9** | 0.04 | 0.92 |
| | OUVA | 0.04 | 0.21 | **0.7** | 0.06 | 0.7 |
| | OUMV | 0.02 | 0.02 | **0.93** | 0.03 | 0.95 |
| | OUMA | 0.12 | 0.15 | **0.64** | 0.09 | 0.66 |
| | OUMVA | 0.05 | 0.13 | **0.76** | 0.06 | 0.76 |
| | OUBM1 | 0.19 | **0.58** | 0.13 | 0.1 | 0.08 |
| | OUBMV | 0.07 | 0.2 | **0.71** | 0.02 | 0.73 |
| CID+ | BMV | **0.36** | 0.28 | 0.33 | 0.03 | 0.01 |
| | OUV | 0.04 | **0.49** | 0.43 | 0.04 | 0.01 |
| | OUA | 0.06 | **0.56** | 0.37 | 0.02 | 0 |
| | OUM | 0.21 | 0.09 | 0.03 | **0.67** | 0.71 |
| | OUVA | 0.07 | **0.55** | 0.35 | 0.04 | 0.03 |
| | OUMV | 0.24 | 0.19 | 0.14 | **0.44** | 0.44 |
| | OUMA | **0.41** | 0.4 | 0.13 | 0.06 | 0.06 |
| | OUMVA | 0.24 | **0.39** | 0.21 | 0.16 | 0.15 |
| | OUBM1 | 0.24 | **0.55** | 0.16 | 0.05 | 0.01 |
| | OUBMV | 0.23 | **0.37** | 0.3 | 0.1 | 0.08 |

**Table 4:** Average AIC weight as the number of taxa increases for each model class. Colored cells indicate the AIC weight of the generating model class. In general, as the number of taxa increases the average support for the generating model class increases.

| Generating model class | nTaxa | AICwt BM1 | AICwt OU1 | AICwt CD | AICwt CID+ |
|---|---|---|---|---|---|
| CD | 25 | 0.12 | 0.22 | 0.51 | 0.15 |
| | 100 | 0.06 | 0.22 | 0.7 | 0.02 |
| | 250 | 0.02 | 0.14 | 0.82 | 0.02 |
| CID+ | 25 | 0.28 | 0.35 | 0.24 | 0.14 |
| | 100 | 0.21 | 0.4 | 0.23 | 0.15 |
| | 250 | 0.11 | 0.34 | 0.32 | 0.22 |

**Table 5:** Modeling results from the 25 models fit to Ericaceae aridity index and fruit type data. Model classes are character independent without rate heterogeneity (CID), character dependence (CD), character independence with rate heterogeneity (CID+), and mixed character dependent and character independence (HYB). Character dependent models suggest that climatic niche evolution will be linked to the fruit type. We found substantial support for OUVA (variable $\sigma^2$ and $\alpha$) and OUMVA (variable $\sigma^2$, $\alpha$, and $\theta$) models. np is the number of freely estimated parameters. lnLik is the joint likelihood of the MLE. DiscLik and ContLik are the marginal likelihood of the discrete and continuous datasets respectively, given the maximum joint likelihood estimate of the parameters. AIC is the Akaike information criterion, ΔAIC is the difference from the best fit model measured as the difference between each model's AIC, and AICwt is the relative support for each model.

| Model class | Model type | np | lnLik | DiscLik | ContLik | AIC | ΔAIC | AICwt |
|---|---|---|---|---|---|---|---|---|
| CID | BM1 | 4 | -243.89 | -32.62 | -206.67 | 495.78 | 39.07 | 0 |
| | OU1 | 5 | -225.5 | -32.62 | -188.28 | 461.01 | 4.3 | 0.05 |
| CD | BMV | 5 | -243.78 | -32.62 | -207.08 | 497.56 | 40.85 | 0 |
| | OUV | 6 | -225.49 | -32.62 | -188.47 | 462.98 | 6.27 | 0.02 |
| | OUA | 6 | -224.95 | -32.58 | -189.48 | 461.9 | 5.19 | 0.03 |
| | OUM | 6 | -224.12 | -32.57 | -187.79 | 460.24 | 3.53 | 0.07 |
| | ***OUVA*** | ***7*** | ***-221.62*** | ***-32.58*** | ***-184.44*** | ***457.24*** | ***0.53*** | ***0.32*** |
| | OUMV | 7 | -224.05 | -32.62 | -188.15 | 462.1 | 5.39 | 0.03 |
| | OUMA | 7 | -223.21 | -32.58 | -187.97 | 460.42 | 3.71 | 0.06 |
| | ***OUMVA*** | ***8*** | ***-220.35*** | ***-32.6*** | ***-183.27*** | ***456.71*** | ***0*** | ***0.41*** |
| | OUBM1 | 5 | -243.84 | -32.57 | -206.67 | 497.68 | 40.97 | 0 |
| | OUBMV | 6 | -243.79 | -32.61 | -206.99 | 499.57 | 42.87 | 0 |
| CID+ | BMV | 7 | -244.8 | -33.11 | -205.78 | 503.59 | 46.89 | 0 |
| | OUV | 8 | -228.77 | -32.98 | -190.16 | 473.55 | 16.84 | 0 |
| | OUA | 8 | -226.42 | -33.17 | -188.53 | 468.84 | 12.13 | 0 |
| | OUM | 8 | -226.43 | -33.32 | -189.07 | 468.87 | 12.16 | 0 |
| | OUVA | 9 | -244.38 | -33.43 | -202.12 | 506.76 | 50.05 | 0 |
| | OUMV | 9 | -225.2 | -33.39 | -182.88 | 468.39 | 11.68 | 0 |
| | OUMA | 9 | -225.57 | -32.68 | -189.92 | 469.14 | 12.43 | 0 |
| | OUMVA | 10 | -227.39 | -33.13 | -185.15 | 474.79 | 18.08 | 0 |
| | OUBM1 | 7 | -244.44 | -33.16 | -206.67 | 502.88 | 46.17 | 0 |
| | OUBMV | 8 | -225.58 | -32.71 | -186.58 | 467.17 | 10.46 | 0 |
| HYB | BMS | 9 | -244.46 | -33.08 | -204.83 | 506.93 | 50.22 | 0 |
| | OUM | 10 | -224.12 | -32.67 | -188.99 | 468.23 | 11.52 | 0 |
| | OUMVA | 16 | -226.56 | -33.03 | -179.11 | 485.13 | 28.42 | 0 |

**Table 6:** Model averaged parameter estimates and standard errors for Ericaceae aridity index and fruit type data. Models with higher AIC weights contribute more overall to the parameter values. The units for $\alpha$, $\sigma^2$, and $\theta$ are $\frac{P}{PET} \div time$, $\left(\frac{P}{PET}\right)^2$, and $\frac{P}{PET}$ respectively. Where P is the average annual precipitation and PET is average annual potential evapotranspiration. Rates of $q$ are measured in transitions per million years.

| Continuous parameter estimates | | | | Discrete parameter estimates | |
|---|---|---|---|---|---|
| | $\alpha$ | $\sigma^2$ | $\theta$ | | |
| Dry | 0.015 (±0.0059) | 0.011 (±0.0043) | 0.97 (±0.011) | $q_{dry\ to\ fleshy}$ | 0.0015 (±0.00058) |
| Fleshy | 0.023 (±0.011) | 0.007 (±0.002) | 0.81 (±0.28) | $q_{fleshy\ to\ dry}$ | 00036 (±0.000086) |

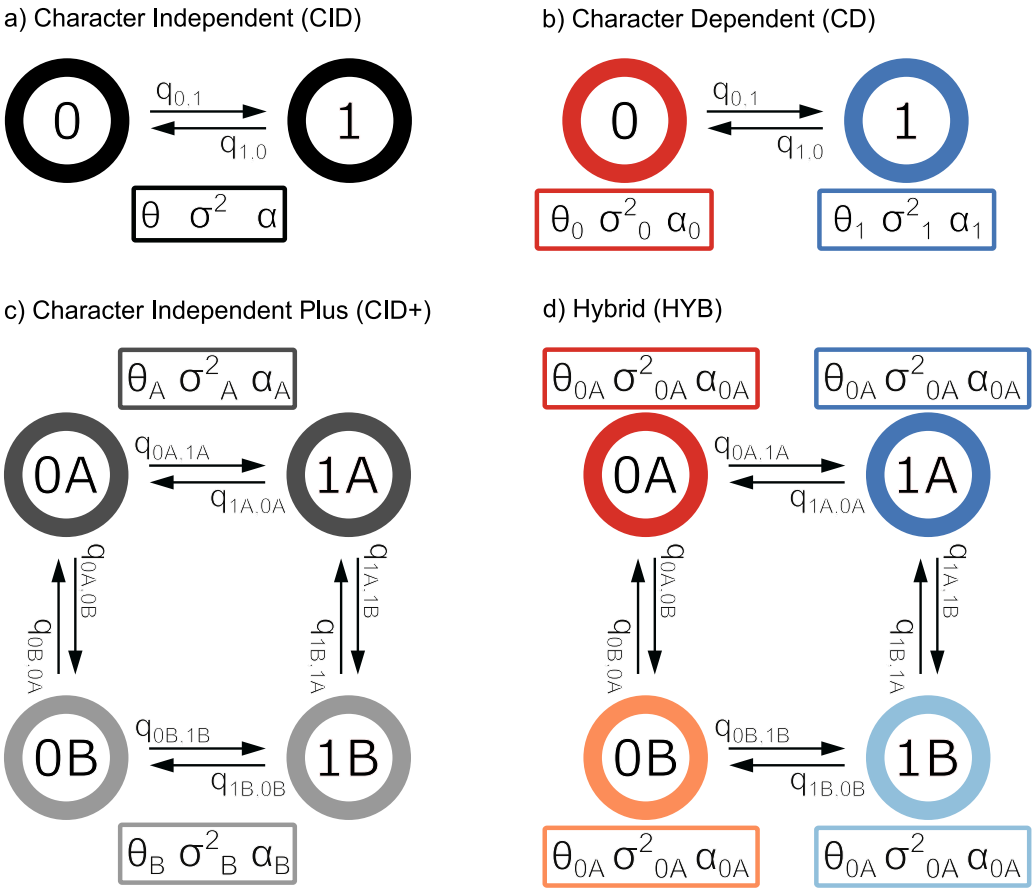


**Figure 1.** A state-transition diagram describing the model classes allowable in hOUwie. Each panel is comprised of observed discrete states 0 and 1 with possible hidden states A and B. Transitions between states are described with the $q$ parameter. Continuous model parameters appear in a box below the states they describe, and their association is displayed with a subscript specific to that state. a) A simple character independent model in which the two observed states do not influence the continuous character which will have the same $\theta, \sigma^2, \alpha$ throughout the phylogeny. b) A character dependent model in which the continuous character depends on the discrete character by virtue of $\theta, \sigma^2, \alpha$ being associated with a particular observed discrete state. c) A character independent model with rate heterogeneity. The two observed states (0 and 1) are not directly linked to the continuous character. However, the continuous character is still allowed to have multiple $\theta, \sigma^2, \alpha$ describing its evolution, but these parameters are associated with hidden states A and B. d) A hybrid model in which each combined observed and hidden state is allowed to have its own $\theta, \sigma^2, \alpha$. Under this model, the continuous character is linked to both character dependent differences (parameters associated with 0 and 1) and character independent differences (A and B).

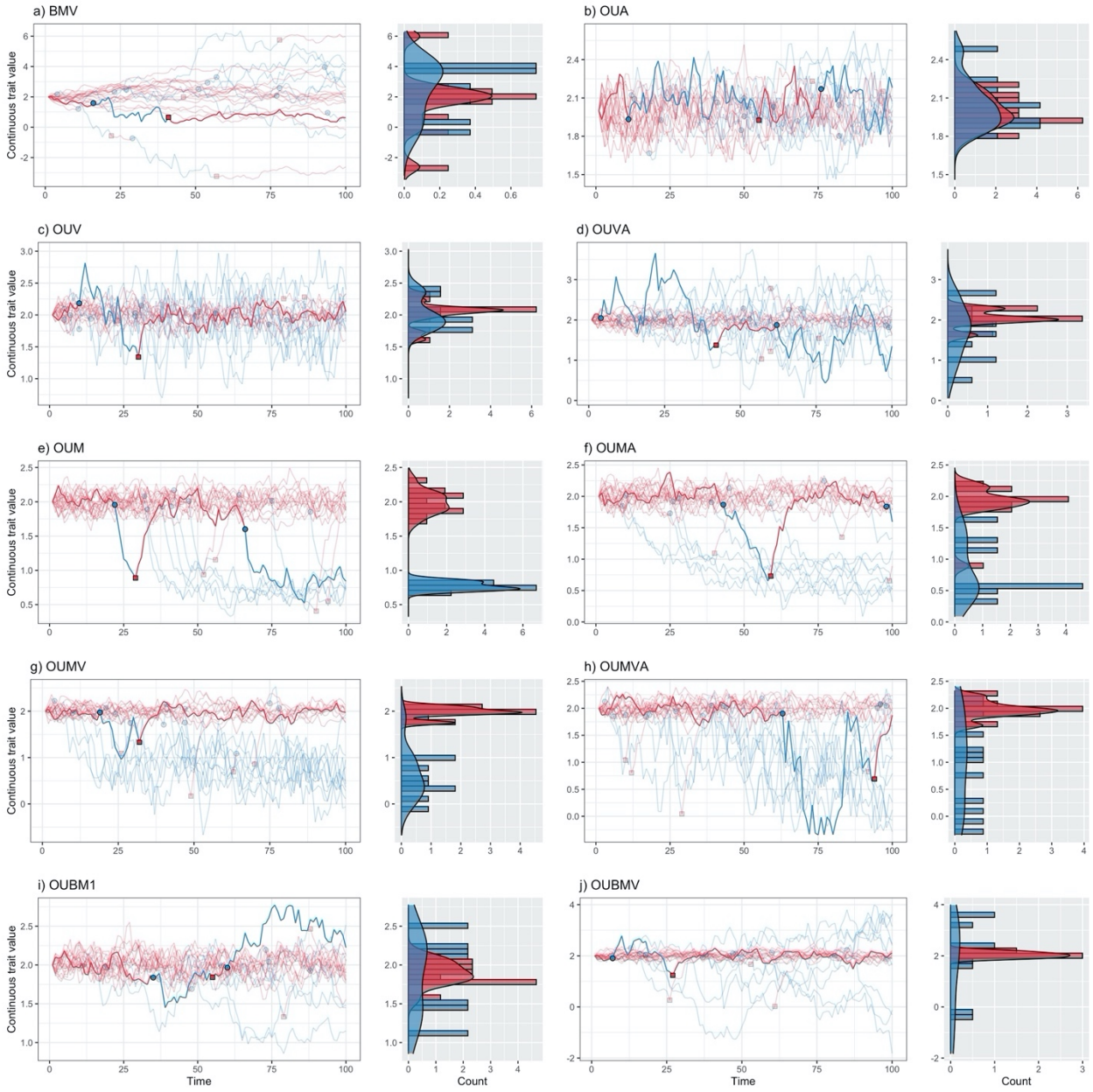


**Figure 2**. A visual representation of binary discrete character *hOUwie* model types. Discrete time forward simulations are conducted starting in the red state and the distribution of the continuous character is plotted on the right as a histogram and density plot. Each line represents a continuous character value at some time. Transitions occur at colored points and each line is colored by the current discrete state. 100 time-steps are simulated with the same parameters as our simulation study ($q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$). The highlighted line was randomly chosen from the set in which at least one discrete state transition occurred.

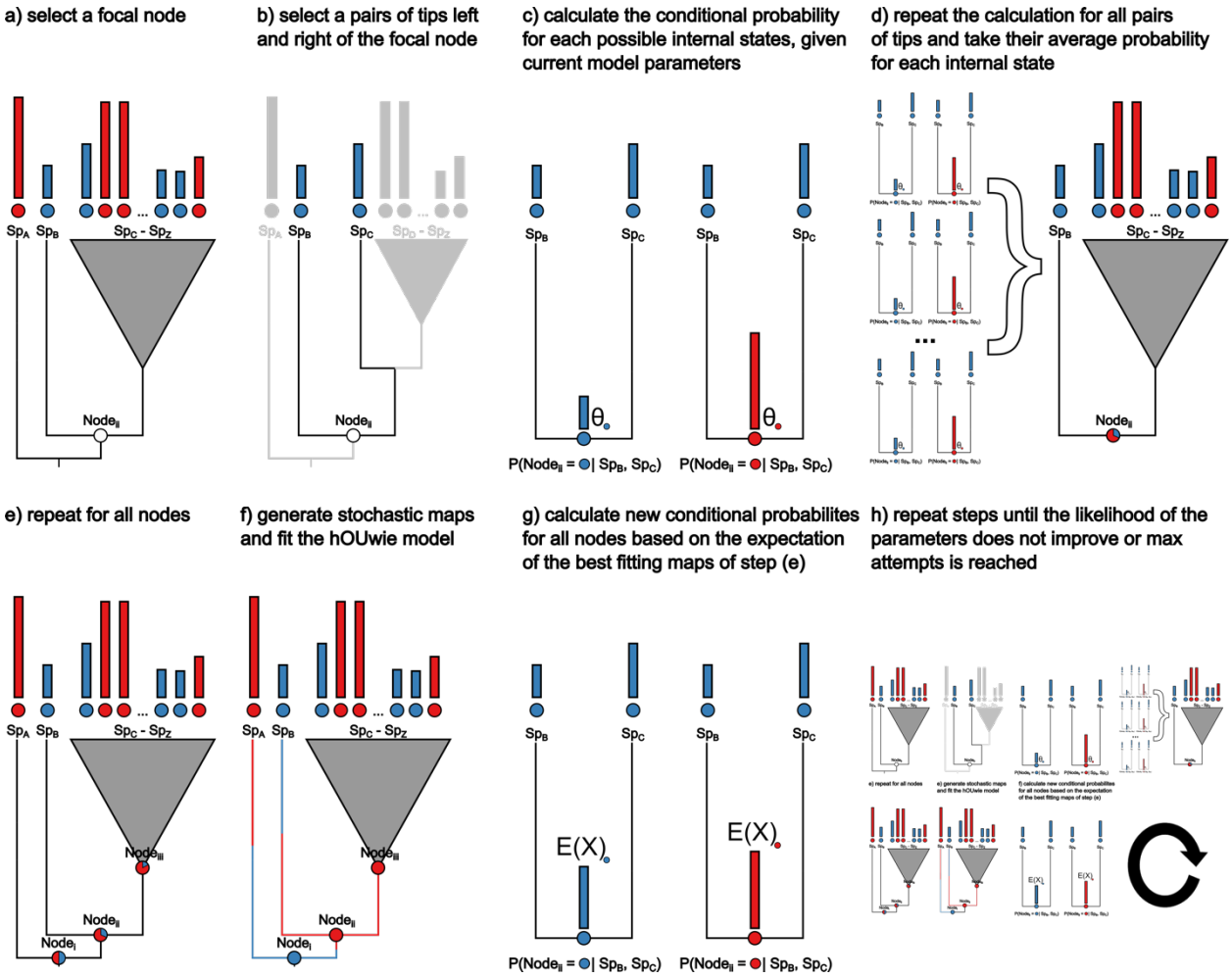


**Figure 3.** A visual representation of the algorithm underlying the calculation of conditional node probabilities and the adaptive sampling procedure. The goal of the procedure is to produce underlying regime paintings well suited to both the discrete and continuous character. a) select the focal node for which we will be calculating the joint conditional probabilities of the discrete and continuous characters. b) on each side of the node we select a pair of tips. c) the conditional probability of the observed discrete and continuous character is calculated for each discrete regime state with an ancestral continuous value equal to $\theta$ of that regime state. d) the conditional probability of the focal node is calculated as the average probability of each regime state for all pairs of observed tips. e) the conditional probabilities are calculated for all internal nodes. This can be turned off within *hOUwie* by setting the sample_nodes argument to false. f) A stochastic map is generating using forward simulation rejection sampling. g) adaptive sampling uses the highest joint probability of previously generated underling regimes to generate a set of ancestral continuous character values. This differs from previous ancestral values because instead of assuming the value $\theta$ for each regime state, it calculates the expected value given the root state and regime mapping for that particular node. h) we repeat steps d) through g) until the joint likelihood of the set of underlying regimes does not improve.

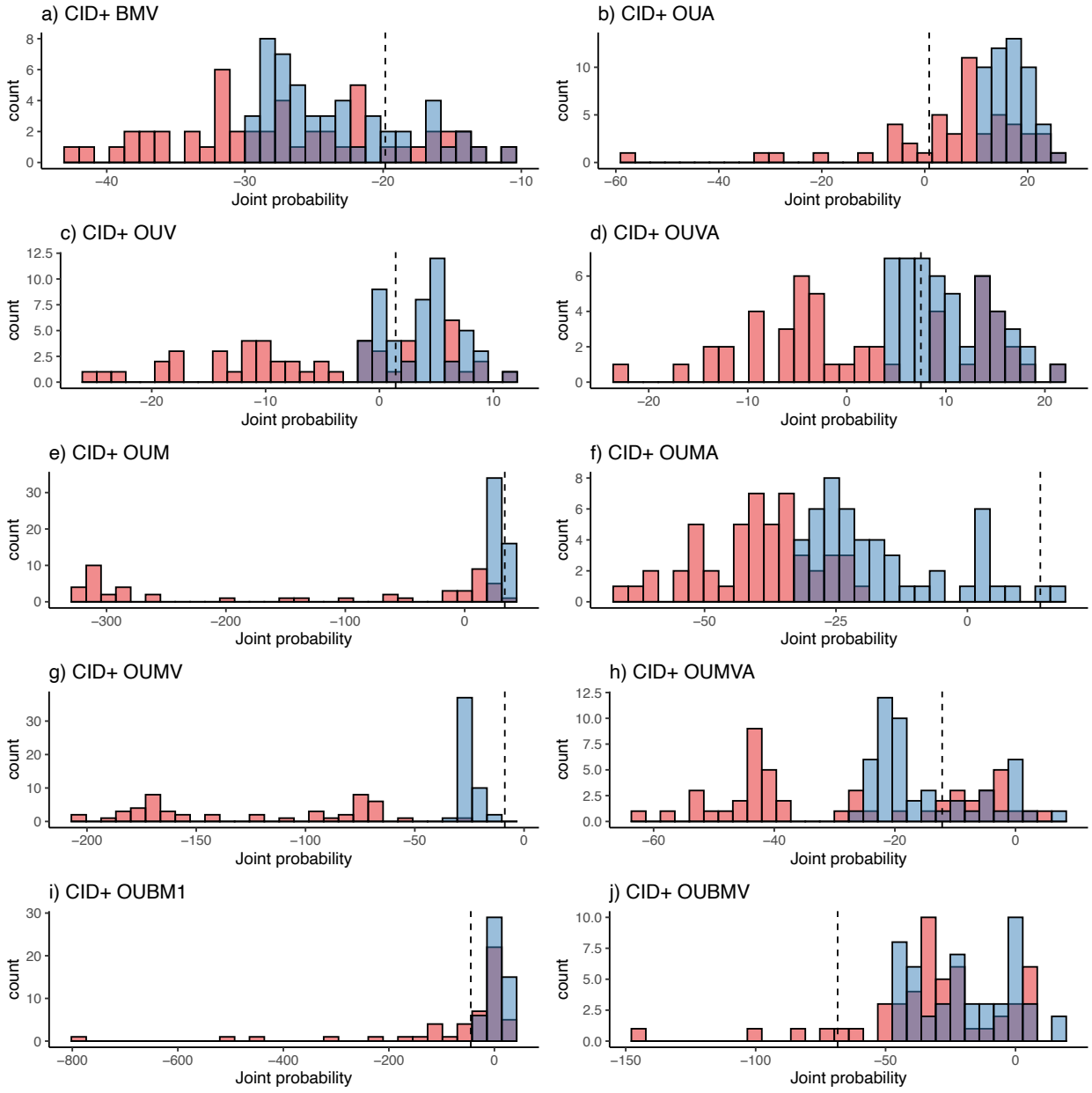


**Figure 4.** Overlapping histograms comparing the effectiveness of the adaptive sampling procedure (blue) and standard discrete only sampling (red) of maps. Regardless of the sampling procedure, all probabilities are calculated in the same way and so any differences in probabilities reflects each procedure's ability to generate appropriate mappings. 50 stochastic mappings are used to calculate the likelihood of the parameters. For each model type, data are simulated following our methods with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$. Dashed line likelihood under generating map.

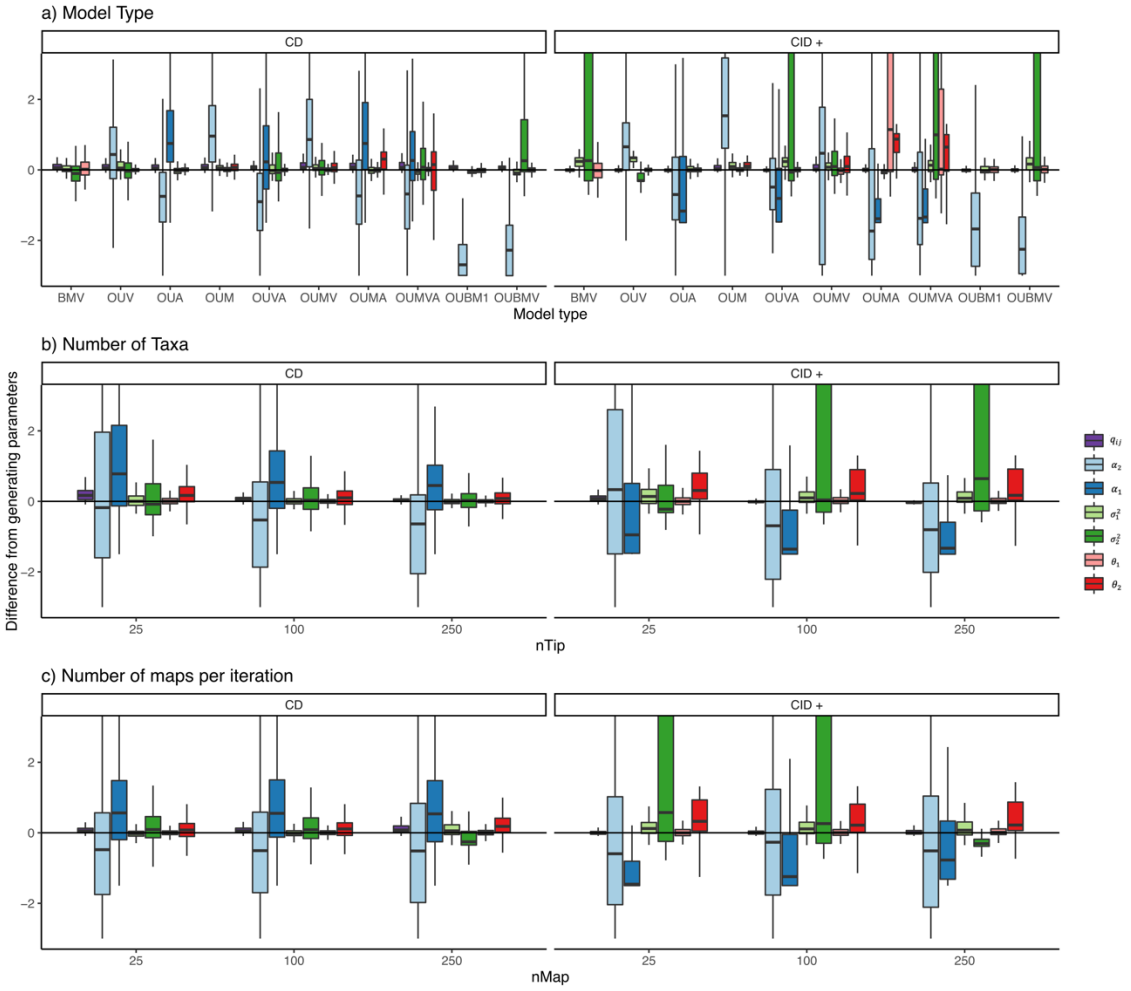


**Figure 5.** The raw difference of the maximum likelihood parameter estimates and the generating values depending on the a) model type, b) number of taxa in the dataset, and c) number of stochastic maps per iteration of the likelihood search. Generally, variable alpha models had the highest biases with alpha being consistently underestimated. As the number of taxa increased, estimation of CD model parameters was estimated with less error. The number of maps per iteration had the greatest effect on character independent models with rate heterogeneity.

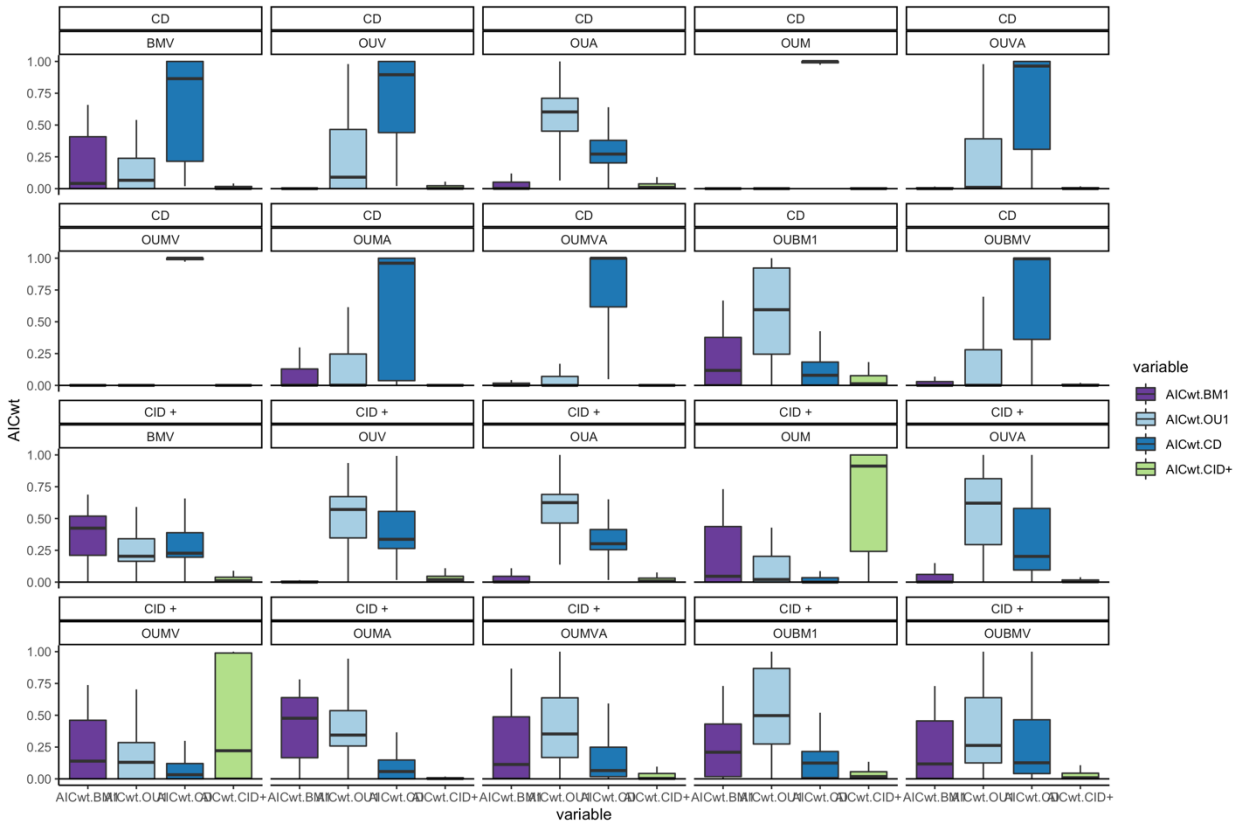


**Figure 6.** AIC weights summarizing the average support for particular model classes and model type when they are the generating model. Headings indicate the generating model type and model class. Data was generated with $q_{ij} = 0.1, \alpha_1 = 3, \alpha_2 = 1.5, \sigma_1^2 = 0.35, \sigma_2^2 = 1, \theta_1 = 2,$ and $\theta_2 = 0.75$ for phylogenies with 25, 100, and 250 taxa and model fits using either 25, 100, or 250 stochastic maps per likelihood iteration. When the generating model class is character dependent (CD) or character independent (CID+) we expect that the AICwt will be highest for that model when fit.

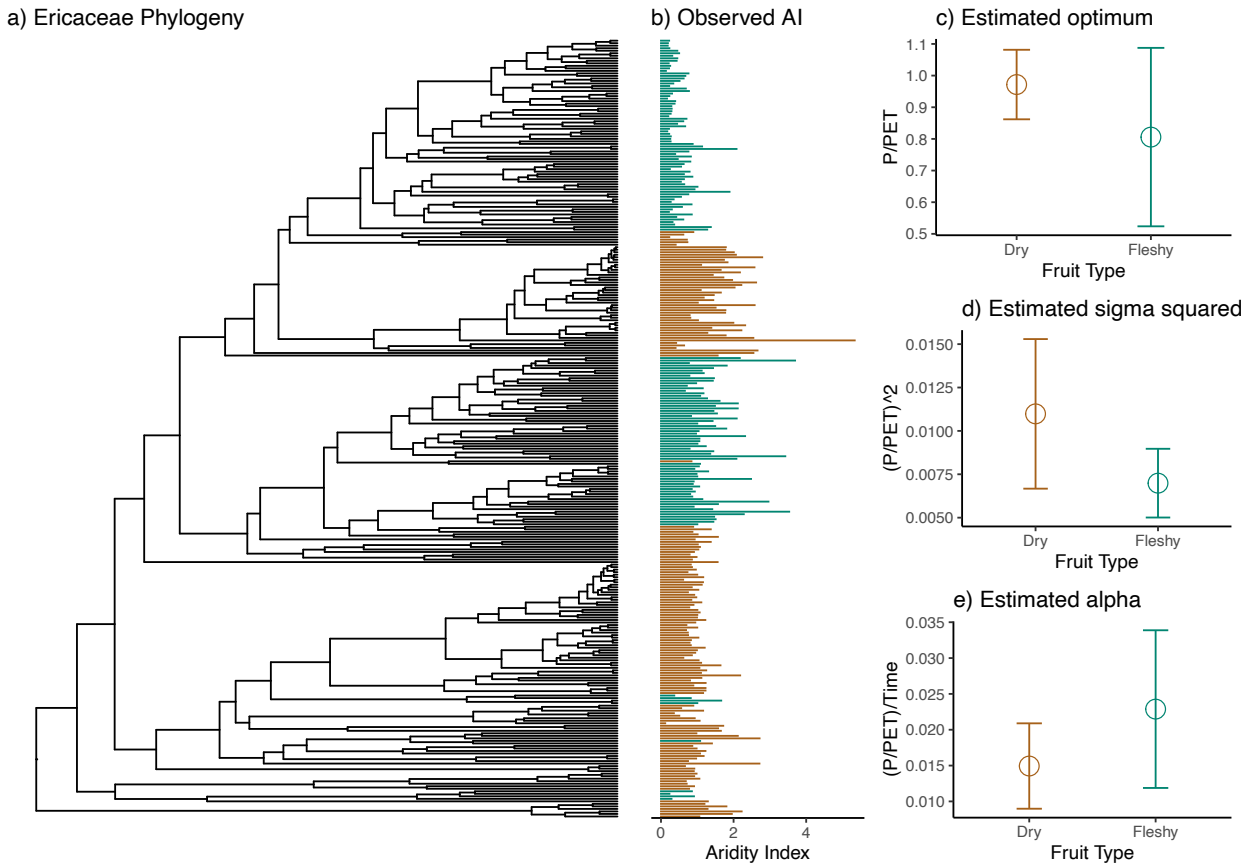


**Figure 7.** a) Ericaceae phylogeny for which we had data (n=309). b) Ln aridity index dataset where each bar is colored by dry (brown) and fleshy (green) fruit type. c) Model averaged parameter estimates with standard error calculated from 100 parametric bootstraps.

# CHAPTER IV

## Long-term responses of life-history strategies to climatic variability in flowering plants

James Boyko, Eric Hagen, Jeremy Beaulieu, and Thais Vasconcelos

## Abstract

Understanding the evolution of life history strategies within flowering plants is a long-standing goal in evolutionary biology. Increasingly, biologists have sought to explain the distribution of annuals and perennials based on their association with broad climatic variables such as temperature or precipitation. However, these efforts have focused on specific clades or geographic areas and, due to methodological limitations, have not allowed joint modeling the evolution of both climatic niches and life history strategies. Here, we combine data on life history strategy and geographic distribution for 9,993 flowering plant species and a recently developed modeling framework which accounts for rate heterogeneity and joint evolution of continuous and discrete traits to evaluate two hypotheses: (1) that annuals tend to evolve in highly seasonal regions prone to extreme heat and drought, because they can rapidly take advantage of short beneficial climatic conditions for reproduction, and (2) that annuals tend to have faster rates of climatic niche evolution than perennials, due to their higher vagility and shorter generation times. Temperature, more specifically the temperature during the warmest season of a year, is the main climatic factor influencing the evolution of annual life history strategy in flowering plants. Annuals are favored in this type of climate due to their ability to escape heat stress as seeds, but they are outcompeted by perennials in regions where extreme heat is uncommon or inexistent. Precipitation and seasonality are less important factors, perhaps due the existence of alternative mechanisms for drought tolerance in perennial species.

**Introduction**

Flowering plants have evolved into multiple types of life history strategies to survive environmental challenges (Grime 1977; Stearns 1992). For instance, resprouting plants can have underground systems to persist through fire and drought (e.g., Rando et al. 2016; Howard et al. 2019) and large trees can become deciduous or have scales to protect their growing buds during freezing conditions (Raunkiaer 1934; Edwards et al. 2017). Other plants have increasingly shortened their life cycles so that germination, fertilization, and seed release all happen through the favorable season of a single year, allowing their progeny to live through unfavorable seasonal weather as seeds (Mulroy and Rundel 1977). The latter describes the life history strategy of annual plants, which are semelparous (i.e., reproduce just once before death; Stearns 1992). This is opposite to the vast majority of flowering plant species, which are mostly iteroparous (i.e., reproduce multiple times and in multiple years) and are characterized by a perennial life history strategy with adaptations to survive an indefinite number of unfavorable seasons as sporophytes (plants) rather than gametophytes (seeds for angiosperms and gymnosperms; Raunkiaer 1934; Friedman 2020).

Botanists have long been interested in finding environmental correlates associated with the evolution of different life history strategies in flowering plants because species with annual and perennial strategies are unevenly distributed across the globe (Figure 1; Raunkiaer 1934; Ricklefs and Renner 1994; Friedman 2020). The distribution of perennials is nonlinear, as they are disproportionately diverse both in areas where freezing is constant, such as higher latitudes and alpine habitats (Billings and Mooney 1968; Givnish 2015), and in areas with warmer tropical climates (Grime 1977). On the other hand, annuals compose the greatest proportion of the flora in mid-latitude areas subject to prolonged drought, such as desert and Mediterranean habitats

(Mulroy and Rundel 1977). Though annuals are considerably less common than perennials across the angiosperm tree of life (Friedman 2020), they can represent over 50% of the floristic diversity in some of these regions (Figure 1b; Raunkiaer 1934).

Although the uneven distribution in the proportion of different life forms across the globe has long been recognized (Raunkiaer 1934; Stebbins 1974; Grime 1977; Friedman 2020), botanists still debate the relative importance of the historical drivers of this pattern, with special focus on the role of climate. For instance, according to the theory of life history strategies in plants, annuals are more likely to evolve where climate is seasonal because the annual strategy allows for rapid responses to short-lived favorable climatic conditions beneficial to reproduction (Cole 1954; Friedman 2020). Support for this has been found in clades typical of Mediterranean habitats, such as *Heliophila* (Brassicaceae) in Africa (Monroe et al. 2019), *Bellis* (Asteraceae) in Europe (Fiz et al. 2002), and in grasses (Poaceae) (Humphreys and Linder 2013). Others have argued that evolution of the annual life form is linked to occupation of generally warmer environments (Stearns 1992), and support for this has been found in temperate clades such as Montiaceae (Ogburn and Edwards 2015). Similarly, annuals would be excluded from alpine environments where frost is common due to high seedling mortality (Givnish, 2015). Finally, some have argued that both temperature and precipitation combined, as well as their seasonality throughout the year, are relevant in explaining the evolution of different strategies, as has been shown in *Oenothera* (Onagraceae) (Evans et al. 2005). In other words, temperature (particularly extremes of heat and frost), precipitation (especially drought), and seasonality have all been found to influence evolutionary transitions between annual and perennial strategies within flowering plants. However, studies have so far focused on specific clades or geographic areas,

and it is unclear which patterns are general enough to hold when multiple clades are considered in the same analytical framework.

In addition to low generality, previous work has paid insufficient attention to how the rates and directions of climatic niche evolution differ between clades possessing different life history strategies. Whereas the different life history strategies likely evolved in response to particular climatic pressures, they may also impact long term biogeographical patterns of lineages evolving under them. For example, the evolution of the annual habit is linked to a series of traits associated with securing reproduction and increased vagility, like selfing (Stebbins 1950; Aarssen 2000) and relatively high investment in seed production (Friedman 2020). For those reasons, annuals are considered to be generally good invaders (Pannell et al. 2015; Linder et al. 2018), a generalization supported by the observation that many of the worst invasive plants in the world are annuals (Holzmueller and Jose 2009). Furthermore, phenotypic evolution may be faster in annuals than in perennials due to their generally shorter generation times (e.g., Smith and Beaulieu 2009), which could make them able to adapt more quickly to changing environmental conditions (Andreasen and Baldwin 2001).

Here, we assess the dialectical relationship between climatic factors and the evolution of life history strategies in flowering plants. To that end, we apply recent theoretical developments in trait evolution models (Chapter III) to explicitly incorporate the impact of climatic niche variation on the evolution of life history strategies. We account for the heterogeneity of evolutionary histories in flowering plants and the habitats associated with them by analyzing a broad sample of clades where multiple transitions between annual and perennial strategies are observed. Two specific hypotheses are addressed: (1) annuals tend to evolve in warmer and drier climates, or where seasonality is stronger, more often than perennials; and (2) annuals tend to

have faster rates of climatic niche evolution than perennials, because of their higher invasibility and shorter generation times. We expect to find mixed support for our hypotheses due to clade specific evolutionary patterns. Some clades will undoubtedly have more heterogeneity in transition rates between life history strategies, whereas other clades may have exclusively unidirectional transitions, and yet others may have no heterogeneity at all. However, due to our large dataset and the ability to account for rate heterogeneity in our model, we expect that we can illuminate the generalities of the long-term responses of life-history strategies to climatic variability in flowering plants.

## Methods

### *Phylogenetic and life history datasets*

To build a dataset of life history strategies for a set of flowering plant clades, we used the recent release of the World Checklist of Vascular Plants dataset (WCVP, 2022*; note: this data is part of a dataset to be officially released in November 2022. The dataset was made available for manuscripts that are part of a *New Phytologist* special issue to be published in 2023), which includes life form data following the Raunkiaer (1934) system. The Raunkiaer system classifies different life history strategies in flowering plants based on the position of the buds in relation to the soil at the end of the growing season and on how plants protect growing buds during the unfavorable seasons. We scored as “annuals” all species marked as “Therophytes” (including combinations such as “Climbing therophyte” and “Semiaquatic therophyte”) or “Biennials” in the WCVP dataset. All other life forms, such as “Cryptophytes”, “Nanophanerophytes”, and “Phanerophytes”, were scored as “perennials”.

Following this scoring, the proportion of “annuals” to “perennials” in the WCVP dataset is around 1:5. In other words, annual plants are considerably less common than perennials and therefore it is more common to find clades where all species are perennials than clades where evolutionary transitions between annual and perennial strategies are observed. However, we restricted our set of clades to groups that presented multiple evolutionary transitions between different life history strategies. Selecting only groups where both life history states are present is not the ideal scenario, because excluding groups consisting only of perennials may bias our view of how different life histories and climatic niches impact each other across evolutionary time. On the other hand, our analytical framework accounts for hidden heterogeneity that would come from a character independent continuous trait evolution, partially mitigating this source of bias.

The set of clades selected for our analyses is not restricted to a single taxonomic rank and includes any clade that matched the criteria: (1) both annuals and perennial strategies are observed; (2) time calibrated phylogenetic tree is available in the literature; and (3) phylogenetic tree includes from c. 50 to c. 1000 tips and at least 10% of the known species diversity assigned to that clade. The clades selected were: the families Balsaminaceae (Rose et al. 2018), Gesneriaceae (Roalson and Roberts 2016), Onagraceae (Freyman and Höhna 2019), Orobanchaceae (Schneider and Moore 2017), Polemoniaceae (Rose et al., 2018), and Solanaceae (Särkinen et al. 2013). The Malvaceae subfamilies Eumalvoideae and Grewioideae (Hoorn et al. 2019), the Apiaceae subfamily Apioideae (Banasiak et al. 2013), the Poaceae subfamilies Pooideae and Panicoideae (Spriggs et al. 2014), and the Primulaceae subfamily Primuloideae (de Vos et al. 2014). The Asteraceae tribe Cardueae (Park and Potter 2015), the Brassicaceae tribes Alysseae, Arabideae, Brassiceae, Cardamineae, Chorisporeae, Erysimeae, Euclidieae, Heliophileae, Lepidieae, Thelypodieae (Huang et al. 2020) and Cremolobeae, Eudemeae, and

Schizopetaleae ("CES-clade"; Salariato et al. 2016), the Plantaginaceae tribe Antirrhineae (Gorospe et al. 2020), the Primulaceae tribe Lysimachieae (Yan et al. 2018), the Rubiaceae tribes Rubieae and Spermacoceae (Neupane et al. 2017; Ehrendorfer et al. 2018). The genera *Chamaecrista* (Fabaceae, Vasconcelos et al. 2020). *Croton* (Euphorbiaceae, Arévalo et al. 2017), *Hypericum* (Hypericaceae, Nürk et al. 2013), *Lupinus* (Fabaceae, Drummond et al. 2012) and *Salvia* (Lamiaceae, Kriebel et al. 2020). All clades combined sum 33 phylogenetic trees and 9,993 tips and lineages are distributed globally. We also completed the life form scoring by adding data collected from the literature, so that each clade had a maximum of 30% missing data.

*Distribution points and climatic data*

We standardized all species names in the 33 phylogenetic trees following the GBIF taxonomic backbone with the R packages taxize (Chamberlain and Szöcs 2013) and downloaded occurrence points that had preserved specimens associated with them using functions of the R package rgbif (Chamberlain and Boettiger 2017), resulting in a dataset of 3,158,632 occurrence points. This dataset was filtered according to the native distribution range of genera and species using the shapefiles of the Working Group on Taxonomic Databases for Plant Sciences (TDWG) for level 3 botanical countries (Brummitt et al. 2001) combined with the WCVP dataset. This filtering was particularly important to exclude the invasive range of several species, keeping only native ranges according to the expertise of taxonomists. Other irregularities such as points in the sea, outliers, duplicated coordinates for the same species and centroids of countries were also removed using a similar protocol as Vasconcelos et al. (2021).

Based on our hypotheses, and because there is no consensus in the literature of what type of climatic variables correlate with evolutionary transition of annual and perennial strategies, we used the climate data from CHELSA (Climatologies at high resolution for the earth's land

surface areas; Karger et al. 2017). In total, eight climatic variables divided into three groupings were tested (Table 1): (1) mean variables, including BIO 1: Mean Annual Temperature (MAT), BIO 12: Mean Annual Precipitation (MAP) and Aridity Index (AI; the higher the more humid); (2) seasonality variables, including BIO 4: Temperature Seasonality and BIO15: Precipitation Seasonality; and (3) variables associated with climatic extremes, including BIO17: Precipitation of Driest Quarter (drought), BIO5: Maximum Temperature of the Warmest Month (heat) and BIO6: Minimum Temperature of the Coldest Month (freezing conditions). All variables were analyzed in their finer scale of 30arc sec (1km in the equator?). To summarize climatic data for each species, we used functions of the R packages sp and rasters (Bivand et al. 2008; Hijmans et al. 2015) to extract a value for each filtered occurrence point based on the climatic layers we assembled. To mitigate the impact of collecting bias, we filtered these points so that no more than one occurrence point for every 1 x 1 degree cell for each species was included. The value of each remaining point was then log transformed and used to calculate mean and within species variance (Labra et al. 2009) for each species, which was used as error measurement in downstream analyses.

*Trait evolution analyses*

Our analysis is conducted with two complementary goals in mind. First, we wish to accurately model the potential link between climatic niche evolution and life history characters within each of our 33 clades. This is done by fitting a set of 10 hOUwie models with 50 stochastic mappings per iteration and adaptive sampling enabled. hOUwie is a recently developed model which explicitly models the joint evolution of discrete and continuous characters (Chapter III). Each of the fitted model structures can be parameterized such that it is either character dependent or character independent. Character dependent models test for an explicit difference in climatic

niche evolution between annual and perennial lineages whereas character independent model structures assume no difference. Furthermore, several models have a mixture of character dependent and independent processes, allowing some differences between parameters to depend on life-history and other parameters to be fixed as equal. Finally, we include character independent models which allow for trait-independent rate heterogeneity. These types of models have been shown to be important as robust null hypotheses and to account for the possibility that our model selection without HMIMs would be biased towards correlation as a consequence of detecting rate heterogeneity without true correlation (Chapter II). In the context of this study, these models account for the fact that climatic niche evolution is likely to be variable throughout the phylogeny regardless of potential correlation with life-history.

The parameters we allow to vary in our model are rates of transition between annual and perennial ($q$), the phenotypic optima of the climatic niche ($\theta$), and the rate of climatic niche evolution ($\sigma^2$). We conduct model averaging and compare several parameter estimates within hOUwie to test for: (1) a relationship between climatic optima and life history strategy ($\theta$), and (2) whether evolutionary rates of annuals are greater for annuals than perennials across all climatic variables ($\sigma^2$). The differences between these climatic niche optima of annuals and perennials are expected to depend on the particular climatic variable being modeled (Table 1). For each clade, we test whether there is a signal of correlation between the climatic variable and life history strategy by examining the differences between parameter estimates.

To determine the model averaged parameter estimates we first reconstructed the probabilities of each tip state. This step is done for every fitted model and is necessary because the potential inclusion of hidden states means that there may be additional uncertainty in the tip states. Second, we multiply the probability of each tip state by the parameter value associated

with that state. In the case of character independence, the parameter value associated with the observed tip states will always be the same. This step leaves us with a set of parameter values for each extant tip. Finally, we conduct model averaging by weighting each tip's parameter values by the AIC weight of the model fit it is associated with. These tip values are then categorized as either annual or perennial and the mean of each discrete category is taken for each clade. We note that any differences in parameter estimates within a clade are significant since they represent parameter estimates from a model set. Each tip will always have the same observed state (unless explicitly coded as unknown), but their hidden state may differ. Thus, all estimated parameters are averaged over hidden rate classes based on the associated observed character and joint probability of the underlying regime.

The second part of our analysis is conducted to test whether the associations we detect within clades are broadly consistent across clades. We use phylogenetic paired t-tests to assess whether model averaged parameter estimates associated with life history strategy are consistently different across clades (Revell 2012). We used the whole seed plant phylogeny based on molecular data from Smith and Brown (2018; "GBMB" tree) as a template to generate a backbone phylogeny that includes each of the 33 clades as individual tips (Figure 2a), using the R packages phangorn (Schliep 2011) and ape (Paradis et al. 2004) to prune out all other tips.

## Results

### *Multi-clade analysis with hOUwie: model selection*

In general, we find a mix of support for character dependence and independence depending on both the clade and climatic variable being analyzed (Figure 2b). Certain clades, such as *Lupinus* and Pooideae, have consistent support for some form of character dependence, whereas other

clades, such as Orobanchaceae and *Chamaecrista*, show little correlation between life history strategy and climatic niche evolution. However, these patterns are only broad overviews and do not distinguish between in what way that the character dependent relationship exists (i.e., whether a clade finds support for a variable theta model and others a variable sigma model, but both cases are character dependent). What is shown in Figure 2 is the sum of the AIC weights for all models with some relationship between the parameters of climatic niche evolution and life history strategy. To determine whether our hypotheses are supported by the modeling results, we examine the model averaged parameter estimates for annual and perennial lineages.

First, we outline the difference in estimates related to long term average temperature, precipitation, and aridity. For BIO1 (Table S1), the difference in climatic optima ranged from 12.4 degrees Celsius (°C) higher in annuals in Gesneriaceae to 7.37°C higher in perennials in *Croton*. All clades but Apioideae, Grewioideae, Solanaceae, *Hypericum*, Primulaceae, Balsaminaceae, and *Croton* had a pattern of higher temperature $\theta_{annuals}$ than $\theta_{perennials}$. For BIO12 (Table S5), the difference in climatic optima ranged from 492.1mm more precipitation in annuals in Gesneriaceae to 3601mm more precipitation in perennials in Polemoniaceae. All clades but Gesneriaceae, *Chamaecrista*, Spermacoceae, Thelypodieae, Brassiceae, Orobanchaceae, Lysimachieae, Cardamineae, and Lepidieae had a pattern of higher precipitation in $\theta_{perennials}$ than $\theta_{annuals}$. Finally, for AI (Table S8), the difference in climatic optima ranged from 930.5AI more humid in annual habitats in Thelypodieae to 834131.9AI more humid in perennial habitats in *Lupinus*. *Lupinus* was clearly an outlier in $\theta_{annuals}$ for humidity, the next closest difference was for Polemoniaceae with 8954.6AI more humidity in perennials. Nonetheless, all clades but Thelypodieae, Brassiceae, Spermacoceae, Orobanchaceae, and Gesneriaceae showed a higher humidity optima for $\theta_{perennials}$ than $\theta_{annuals}$.

Next, we outline the difference in estimates related to temperature and precipitation seasonality. For BIO4 (Table S2), and excluding outliers, the difference in climatic optima ranged from 660.9°C standard deviations higher in annuals in Lupinus to 31.6°C standard deviations higher in perennials in Spermacoceae. All clades but Cardamineae, Gesneriaceae, Rubieae, Cardueae, Chorisporeae, Orobanchaceae, and Spermacoceae had a pattern of greater temperature variability in annuals than perennials. For BIO15 (Table S7), the difference in climatic optima ranged from 27.8mm CV (coefficient of variation) more precipitation variation in annuals in Lupinus to 28.0mm CV more precipitation variability in perennials in *Croton*. All clades but Polemoniaceae, Primulaceae, Cardueae, Antirrhineae, Thelypodieae, Arabideae, Brassiceae, Apioideae, Orobanchaceae, Cardamineae, Erysimeae, and *Croton* had a pattern of more precipitation variability in annuals than perennials.

Finally, we outline the difference in the most extreme climatic conditions of a year as they are measured by temperature and precipitation. For BIO5 (Table S3), the difference in climatic optima ranged from 16°C higher max temperature in annuals in Chorisporeae to 0.28°C higher max temperature in perennials in Balsaminaceae. For BIO6 and ignoring outliers for which models failed to converge (Table S4), the difference in climatic optima ranged from 6.87°C minimum temperature in annuals in Pooideae to 2.65°C minimum temperature in perennials Primulaceae. All clades but Polemoniaceae, Croton, Hypericum, Lepidieae, Onagraceae, Erysimeae, Grewioideae, Solanaceae, Primulaceae, Antirrhineae, and Balsaminaceae had a pattern of lower minimum temperature of the coldest months in perennials than annuals. For BIO14 (Table S6), the difference in climatic optima ranged from 2.39mm of precipitation during the driest month in annuals in Erysimeae to 32.4 mm of precipitation in the driest month in perennials in *Hypericum*. All clades but Erysimeae, *Croton*, Brassiceae,

Thelypodieae, CES, Apioideae, Lepidieae, Gesneriaceae, Lupinus, Orobanchaceae, and Cardamineae had a pattern of lower minimum precipitation of the driest months in annuals than perennials.

*General patterns: parameter estimates*

When averaging all models, phylo.t.test comparisons between parameter estimates show slight differences in sigma squared across all variables (Figure 3-5), but none of these differences were significant (phylo.t.test $p > 0.05$). Significant differences in climatic optima between annuals and perennials ($\theta_{annuals} \neq \theta_{perennials}$) were observed for three out of the eight climatic variables analyzed (Figure 3): (1) mean annual temperature, where annuals tend to have higher values than perennials ($\theta_{annuals} > \theta_{perennials}$; phylo.t.test $p = 0.01$); (2) maximum temperature of the warmest month, where annuals tend to have higher values then perennials ($\theta_{annuals} > \theta_{perennials}$; phylo.t.test $p < 0.001$) and precipitation of the driest month, where annuals tend to have lower values than perennials ($\theta_{annuals} < \theta_{perennials}$; phylo.t.test $p = 0.01$). Though some parameter values of other variables are coherent with our hypotheses, no significant differences between climatic optima of annuals and perennials were observed for mean annual precipitation, aridity index, minimum temperature of the coldest month and for none of the variables representing climatic seasonality (phylo.t.test $p > 0.05$).

When looking at individual clades, the maximum temperature of the warmest month was the climatic variable where the strongest pattern was observed. For this variable, the only exception to the general pattern was Balsaminaceae (24.53°C in annuals, 24.82°C in perennials). Maximum difference in optima values is observed in the Brassicaceae tribes Chorisporeae (32°C

in annuals and 15.6°C in perennials), Euclidieae (27.59°C in annuals, 13.90°C in perennials), and Thelypodieae (32.79°C in annuals, 19.8°C in perennials).

Transition rates tended to be higher for annuals to perennials (0.045-0.097 transitions per million years) than perennials to annuals (0.036-0.085 transitions per million years). We note that in cases where the discrete character was influenced by the continuous character (character dependent models), there is the potential for a great deal of variation in the ancestral state (Figure 6). This is because, even though a purely discrete process may favor an entirely annual or perennial life history, when accounting for a reconstruction of the climatic niche, the most probable discrete state will depend on the continuous character distribution. For example, the ancestral state for Apioideae had a marginal probability of 99% annual when life history was reconstructed alongside the climatic optima for the warmest temperature of the hottest month, but a probability of 100% perennial when reconstructed alongside the optima for annual precipitation.

## Discussion

### *Ancestral state reconstruction can be sensitive to climatic associations*

Although not directly related to our main hypotheses, our results challenge, as others have (e.g., Carlquist 1974; Baldwin 2007), the traditional idea that annuality is always a “derived” condition in flowering plants. In fact, for 13 out of 33 clades we analyzed, the root state was recovered as an “annual” life form with greater than 50% certainty and several transitions to perennial life form. This, in and of itself, is interesting, but we note that there was a great deal of variation in the ancestral state reconstruction depending on the particular climatic variable. Some clades such as Apioideae, Rubieae, or Balsamiaceae could be reconstructed with a highly certain annual state

or highly certain perennial state depending on the climatic variable being modeled. This highlight both the importance of joint modeling and the inherent uncertainty of reconstructing ancestral states. In cases where uncertainty was highest, the best supported model was often a character dependent model in which the phenotypic optima was allowed to vary (OUM). This stark difference occurs because the probability of the OU model can be quite sensitive to root states (Butler and King 2004; Ho and Ané 2014). Under an OU model, the influence of the root state decays through time in proportion to the strength of selection (Hansen 1997) and thus the selection of the root state can have a large impact on the model's fit to the data.

This also highlights the sensitivity of ancestral state reconstruction to the particular model and dataset being fit, especially when conducted independently of factors that may influence the evolution of a discrete character. Climate, for example, has been found to be an important factor influencing the evolution of many different discrete traits in plants, such as fruit type (Vasconcelos et al. 2021) and underground storage organs (Tribble et al. 2021). In that way, we show the importance of joint modeling to understand the evolution of discrete traits that respond to climate. This finding suggests that a multivariate extension of *hOUwie*, where several continuous characters are modeled simultaneously, could be important for correctly reconstructing ancestral states. Finally, it is important to note that the high amount of uncertainty depending on the bioclimatic variable was not the case for all character dependent models. *Lupinus*, Heilophileae, Solanaceae, Pooideae and other clades showed high amounts of support for character dependence and highly certain ancestral state reconstructions across all climiatic variables. This shows that there is also the potential of increasing the overall certainty of the reconstruction if both the discrete and continuous character had the same likely regime reconstructed at the root.

*Annuals do not have faster rates of climatic niche evolution*

Previous literature point towards lineages with shorter generation times having faster rates of evolution (e.g., Mooers and Harvey 1994; Smith and Beaulieu 2009). We found that this is not the case for annuals, and there are some possible reasons for this. First, although annuals do tend to have a faster development in their post-germination phase (Grime, 1977; Friedman 2020), their generations are not necessarily shorter because annuals can also have relatively longer seed dormancy and can remain in the form of seeds for many years (Venable and Lawlor 1980; Nunney 2002; Kooyers 2015). In that way, their generation times can be in fact much longer in the pre-germination phase, leading to the incorrect assumption that the visible aboveground, post-germination phase represents the whole life cycle.

Second, many annuals are self-compatible due to the necessity of guaranteed fertilization in a single reproductive event (Aarssen, 2000). Selfing has long been considered an evolutionary dead-end in plants (Stebbins, 1950) because inbreeding depression reduces genetic diversity of selfing populations, precluding adaptation to changing environments (Takebayashi and Morrell 2001; Escobar et al. 2010; Shimizu and Tsuchimatsu 2015; but see Igic and Busch 2013) which may constraint rates of niche evolution in annuals despite their generally higher vagility. In areas of constant disturbance, such as in areas of anthropogenic influence, annuals will be favored due to their higher vagility and their short reproductive window between germination and seed dispersal (Grime, 1977). Though this may make them look like they are generally better invaders, they are poor competitors against perennials in more stable environments and thus are "confined" to habitats where heat is very extreme (Grime, 1977).

*Lack of general rules for most variables, including seasonality and precipitation*

As the accessibility of data and methods to test trait evolution hypotheses using phylogenetic comparative frameworks increased, multiple studies have found that temperature, precipitation and seasonality variables are relevant in explaining the evolution of different life history strategies in plants (Fiz et al. 2002; Evans et al. 2005; Humphreys and Linder 2013; Ogburn and Edwards 2015; Monroe et al. 2019). Our results show that some of these previously documented patterns are clade or area specific, but do not hold as a generality across all flowering plants. For instance, we found no significant difference in optima values for mean annual precipitation and aridity index across all clades, and the lack of strong signal for drought as an important factor in the evolution of annual strategy was unanticipated. We did recover a significant difference between $\theta_{annuals}$ and $\theta_{perennials}$ for precipitation of the driest month ($p < 0.05$) with annuals tending to have a drier optimum, but this pattern was not observed in one third of the clades analyzed. The reason for this lack of strong correlation with precipitation may be the existence of other forms of compensatory mechanisms to deal with extreme drought in perennial plants. Several mechanisms of vegetative tolerance to desiccation have evolved in perennials, including, but not restricted to, changes in photosynthesis pathways (Ehleringer et al. 1991), presence of subterraneous structures (Howard et al., 2019), succulence of leaves and stems (Ogburn and Edwards, 2010), and senescence of photosynthesis structures during dry seasons (Munné-Bosch and Alegre 2004). In that way, evolutionary pathways to survive drought are diverse and evolving an annual lifestyle is not the sole mechanism to escape drought available for plants.

A similar lack of significant association was found for all variables related to seasonality, and for minimum temperature of the coldest month, a variable associated with freezing

temperatures. In those cases, $\theta_{annuals}$ and $\theta_{perennials}$ are not significantly different from each other across all clades, meaning that there is little support for the role of these climatic variables as general rules governing how life history strategies evolve in plants. This suggests that the relevance of these variables to the evolution of life history strategies are probably clade specific and related to particularities of their geographical distributions. For example, in groups where species distribution varies from dry lowland to humid alpine environments, such as *Lupinus* (Drummond et al., 2012; Givnish, 2015) and the Brassicaeae tribe Arabideae (Koch et al. 2012), $\theta_{perennials}$ was found to be lower. In those cases, perennials may indeed be associated with a frost tolerance strategy, due to evenly distributed events of frost in mountains that lead to high seedling mortality in annuals ("winter by night and summer by day"; Givnish, 2015). However, in groups such as Balsaminaceae, Onagraceae, and Solanaceae, where their distribution ranges from tropical to temperate biomes (Wagner et al. 2007) and most perennial species are restricted to humid tropical forests where frost does not occur, annuals are the strategy found in areas where occasional events of frost are present, such as Mediterranean habitats (Pescador et al. 2018). In that way, our results do not support these variables as strong generalities for the whole of flowering plants, but we also do not discard their importance in some groups, depending on their geographical distribution.

*Annual strategy as a heat avoidance mechanism*

The one constant pattern we found across almost all analyzed clades relates to their response to extreme heat. In 32 out of the 33 clades, we found $\theta_{annuals}$ to be consistently higher for maximum temperature of the warmest month. This points towards a generality in the way flowering plants evolve in response to survival in areas subject to extreme heat, where adult mortality is high and surviving as a seed through the hottest seasons may be an option (Angert et al. 2007; Venable

2007). Both annuals and perennials are probably equally sensitive to heat stress in their adult form (Raunkiaer 1934; Teskey et al. 2015), but annuals can evade the hottest season in the form of seed, which is one of the most resistant plant structures (e.g., Janzen 1984). Annuality then becomes a type of heat avoidance mechanism.

In *Impatiens* (Balsaminaceae), the group that was constantly found to go against this general pattern, many of the annuals occur in temperate regions of North America, Europe, and Asia, whereas many perennials are native to the warmer tropical areas (Grey-Wilson 1980; Ruchisansakun et al. 2016). They are mainly summer annuals (i.e. complete their life cycle during the summer), in contrast to other species in our dataset which are winter annuals (complete life cycle during winter; e.g. Mulroy and Rundel, 1977). To our knowledge there is no list of species at a global scale that distinguish winter from summer annuals, nor are there any evolutionary studies comparing these two different types of life history strategies. However, we suspect the strong support for maximum temperature of the warmest month as an important variable means that most annuals, at least in our dataset, are likely to be summer annuals. That would be also consistent with the idea that Mediterranean and subtropical deserts, where summers are the most unfavorable season for plants, generally favor the evolution of annuals. From an evolutionary standpoint, this further supports the lack of alternative pathways for heat tolerance in vegetative structures in plants. This is a worrying scenario for most environments dominated by perennials, given that extreme heat and heat waves tend to become increasingly frequent (Teskey et al., 2015).

## Conclusion

This study provides the first broadscale analysis of life history strategy evolution in flowering plants in relation to their distribution across a climatic gradient. We show how multi-clade analyses can change previous ideas based on a few groups. As predicted, we found mixed support for most climatic variables tested due to clade-specific evolutionary patterns. However, this approach also allowed us to find at least one universality in the long-term responses of life history evolution in relation to climate. Temperature variables, especially extreme heat, were found to have a consistent effect in all clades, pointing towards a generality that the annual semelparous strategy probably often evolves as a mechanism of heat avoidance, possibly due to the lack of alternative evolutionary pathways to survive heat stress in plants. Finally, we also showed how climatic variables have a huge influence in the evolution of correlated discrete traits once a joint modeling approach is considered. Besides answering our research questions, this analysis also provides an example of how to use hOUwie for future users.

**References**

Aarssen L.W. 2000. Why Are Most Selfers Annuals? A New Hypothesis for the Fitness Benefit of Selfing. Oikos. 89:606–612.

Andreasen K., Baldwin B.G. 2001. Unequal Evolutionary Rates Between Annual and Perennial Lineages of Checker Mallows (Sidalcea, Malvaceae): Evidence from 18S–26S rDNA Internal and External Transcribed Spacers. Molecular Biology and Evolution. 18:936–944.

Angert A.L., Huxman T.E., Barron-Gafford G.A., Gerst K.L., Venable D.L. 2007. Linking growth strategies to long-term population dynamics in a guild of desert annuals. Journal of Ecology. 95:321–331.

Arévalo R., van Ee B.W., Riina R., Berry P.E., Wiedenhoeft A.C. 2017. Force of habit: shrubs, trees and contingent evolution of wood anatomical diversity using Croton (Euphorbiaceae) as a model system. Annals of Botany. 119:563–579.

Baldwin B.G. 2007. Adaptive radiation of shrubby tarweeds (Deinandra) in the California Islands parallels diversification of the Hawaiian silversword alliance (Compositae–Madiinae). American Journal of Botany. 94:237–248.

Banasiak Ł., Piwczyński M., Uliński T., Downie S.R., Watson M.F., Shakya B., Spalik K. 2013. Dispersal patterns in space and time: a case study of Apiaceae subfamily Apioideae. Journal of Biogeography. 40:1324–1335.

Billings W.D., Mooney H.A. 1968. The Ecology of Arctic and Alpine Plants. Biological Reviews. 43:481–529.

Bivand R.S., Pebesma E.J., Gómez-Rubio V., Pebesma E.J. 2008. Applied spatial data analysis with R. Springer.

Brummitt R.K., Pando F., Hollis S., Brummitt N.A. 2001. World geographical scheme for recording plant distributions. International working group on taxonomic databases for plant sciences (TDWG)

Butler M.A., King A.A. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist. 164:683–695.

Carlquist S.J. 1974. Island biology. Columbia University Press.

Chamberlain S.A., Boettiger C. 2017. R Python, and Ruby clients for GBIF species occurrence data. .

Chamberlain S.A., Szöcs E. 2013. taxize: taxonomic search and retrieval in R. F1000Research. 2.

Cole L.C. 1954. The Population Consequences of Life History Phenomena. The Quarterly Review of Biology. 29:103–137.

Drummond C.S., Eastwood R.J., Miotto S.T.S., Hughes C.E. 2012. Multiple Continental Radiations and Correlates of Diversification in Lupinus (Leguminosae): Testing for Key Innovation with Incomplete Taxon Sampling. Systematic Biology. 61:443–460.

Edwards E.J., Chatelet D.S., Chen B.-C., Ong J.Y., Tagane S., Kanemitsu H., Tagawa K., Teramoto K., Park B., Chung K.-F., Hu J.-M., Yahara T., Donoghue M.J. 2017. Convergence, Consilience, and the Evolution of Temperate Deciduous Forests. The American Naturalist. 190:S87–S104.

Ehleringer J.R., Sage R.F., Flanagan L.B., Pearcy R.W. 1991. Climate change and the evolution of C4 photosynthesis. Trends in Ecology & Evolution. 6:95–99.

Ehrendorfer F., Barfuss M.H.J., Manen J.-F., Schneeweiss G.M. 2018. Phylogeny, character evolution and spatiotemporal diversification of the species-rich and world-wide distributed tribe Rubieae (Rubiaceae). PLOS ONE. 13:e0207615.

Escobar J.S., Cenci A., Bolognini J., Haudry A., Laurent S., David J., Glémin S. 2010. An Integrative Test of the Dead-End Hypothesis of Selfing Evolution in Triticeae (poaceae). Evolution. 64:2855–2872.

Evans M.E.K., Hearn D.J., Hahn W.J., Spangle J.M., Venable D.L. 2005. Climate and Life-History Evolution in Evening Primroses (oenothera, Onagraceae): A Phylogenetic Comparative Analysis. Evolution. 59:1914–1927.

Fiz O., Valcárcel V., Vargas P. 2002. Phylogenetic position of Mediterranean Astereae and character evolution of daisies (Bellis, Asteraceae) inferred from nrDNA ITS sequences. Molecular Phylogenetics and Evolution. 25:157–171.

Freyman W.A., Höhna S. 2019. Stochastic Character Mapping of State-Dependent Diversification Reveals the Tempo of Evolutionary Decline in Self-Compatible Onagraceae Lineages. Systematic Biology. 68:505–519.

Friedman J. 2020. The Evolution of Annual and Perennial Plant Life Histories: Ecological Correlates and Genetic Mechanisms. Annual Review of Ecology, Evolution, and Systematics. 51:461–481.

Givnish T.J. 2015. Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. New Phytologist. 207:297–303.

Gorospe J.M., Monjas D., Fernández-Mazuecos M. 2020. Out of the Mediterranean Region: Worldwide biogeography of snapdragons and relatives (tribe Antirrhineae, Plantaginaceae). Journal of Biogeography. 47:2442–2456.

Grey-Wilson C. 1980. Impatiens of Africa. CRC Press.

Grime J.P. 1977. Evidence for the Existence of Three Primary Strategies in Plants and Its Relevance to Ecological and Evolutionary Theory. The American Naturalist. 111:1169–1194.

Hansen T.F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. Evolution. 51:1341–1351.

Hijmans R.J., Van Etten J., Cheng J., Mattiuzzi M., Sumner M., Greenberg J.A., Lamigueiro O.P., Bevan A., Racine E.B., Shortridge A. 2015. Package 'raster.' R package. 734.

Ho L.S.T., Ané C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. Methods in Ecology and Evolution. 5:1133–1146.

Holzmueller E.J., Jose S. 2009. Invasive plant conundrum: What makes the aliens so successful? Journal of Tropical Agriculture. 47:18–29.

Hoorn C., van der Ham R., de la Parra F., Salamanca S., ter Steege H., Banks H., Star W., van Heuven B.J., Langelaan R., Carvalho F.A., Rodriguez-Forero G., Lagomarsino L.P. 2019. Going north and south: The biogeographic history of two Malvaceae in the wake of Neogene Andean uplift and connectivity between the Americas. Review of Palaeobotany and Palynology. 264:90–109.

Howard C.C., Folk R.A., Beaulieu J.M., Cellinese N. 2019. The monocotyledonous underground: global climatic and phylogenetic patterns of geophyte diversity. American Journal of Botany. 106:850–863.

Huang X.-C., German D.A., Koch M.A. 2020. Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. Annals of Botany. 125:29–47.

Humphreys A.M., Linder H.P. 2013. Evidence for recent evolution of cold tolerance in grasses suggests current distribution is not limited by (low) temperature. New Phytologist. 198:1261–1273.

Igic B., Busch J.W. 2013. Is self-fertilization an evolutionary dead end? New Phytologist. 198:386–397.

Janzen D.H. 1984. Dispersal of Small Seeds by Big Herbivores: Foliage is the Fruit. The American Naturalist. 123:338–353.

Karger D.N., Conrad O., Böhner J., Kawohl T., Kreft H., Soria-Auza R.W., Zimmermann N.E., Linder H.P., Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. Scientific data. 4:1–20.

Koch M.A., Karl R., German D.A., Al-Shehbaz I.A. 2012. Systematics, taxonomy and biogeography of three new Asian genera of Brassicaceae tribe Arabideae: An ancient distribution circle around the Asian high mountains. TAXON. 61:955–969.

Kooyers N.J. 2015. The evolution of drought escape and avoidance in natural herbaceous populations. Plant Science. 234:155–162.

Kriebel R., Drew B., González-Gallegos J.G., Celep F., Heeg L., Mahdjoub M.M., Sytsma K.J. 2020. Pollinator shifts, contingent evolution, and evolutionary constraint drive floral disparity in Salvia (Lamiaceae): Evidence from morphometrics and phylogenetic comparative methods. Evolution. 74:1335–1355.

Labra A., Pienaar J., Hansen T.F. 2009. Evolution of Thermal Physiology in Liolaemus Lizards: Adaptation, Phylogenetic Inertia, and Niche Tracking. The American Naturalist. 174:204–220.

Linder H.P., Lehmann C.E.R., Archibald S., Osborne C.P., Richardson D.M. 2018. Global grass (Poaceae) success underpinned by traits facilitating colonization, persistence and habitat transformation. Biological Reviews. 93:1125–1144.

Monroe J.G., Gill B., Turner K.G., McKay J.K. 2019. Drought regimens predict life history strategies in Heliophila. New Phytologist. 223:2054–2062.

Mooers A.Ø., Harvey P.H. 1994. Metabolic rate, generation time, and the rate of molecular evolution in birds. Molecular phylogenetics and evolution. 3:344–350.

Mulroy T.W., Rundel P.W. 1977. Annual Plants: Adaptations to Desert Environments. BioScience. 27:109–114.

Munné-Bosch S., Alegre L. 2004. Die and let live: leaf senescence contributes to plant survival under drought stress. Functional Plant Biol. 31:203–216.

Neupane S., Lewis P.O., Dessein S., Shanks H., Paudyal S., Lens F. 2017. Evolution of woody life form on tropical mountains in the tribe Spermacoceae (Rubiaceae). American Journal of Botany. 104:419–438.

Nunney L. 2002. The Effective Size of Annual Plant Populations: The Interaction of a Seed Bank with Fluctuating Population Size in Maintaining Genetic Variation. The American Naturalist. 160:195–204.

Nürk N., Scheriau C., Madriñán S. 2013. Explosive radiation in high Andean Hypericum—rates of diversification among New World lineages. Frontiers in Genetics. 4.

Ogburn M.R., Edwards E.J. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm clade Montiaceae. Mol. Phylogenet. Evol. 92:181–192.

Pannell J.R., Auld J.R., Brandvain Y., Burd M., Busch J.W., Cheptou P.-O., Conner J.K., Goldberg E.E., Grant A.-G., Grossenbacher D.L., Hovick S.M., Igic B., Kalisz S., Petanidou T., Randle A.M., de Casas R.R., Pauw A., Vamosi J.C., Winn A.A. 2015. The scope of Baker's law. New Phytologist. 208:656–667.

Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Park D.S., Potter D. 2015. Why close relatives make bad neighbours: phylogenetic conservatism in niche preferences and dispersal disproves Darwin's naturalization hypothesis in the thistle tribe. Molecular Ecology. 24:3181–3193.

Pescador D.S., Sánchez A.M., Luzuriaga A.L., Sierra-Almeida A., Escudero A. 2018. Winter is coming: plant freezing resistance as a key functional trait for the assembly of annual Mediterranean communities. Annals of Botany. 121:335–344.

Rando J.G., Zuntini A.R., Conceição A.S., van den Berg C., Pirani J.R., de Queiroz L.P. 2016. Phylogeny of Chamaecrista ser. Coriaceae (Leguminosae) Unveils a Lineage Recently Diversified in Brazilian Campo Rupestre Vegetation. International Journal of Plant Sciences. 177:3–17.

Raunkiaer C. 1934. The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer. The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in ecology and evolution.:217–223.

Ricklefs R.E., Renner S.S. 1994. Species Richness Within Families of Flowering Plants. Evolution. 48:1619–1636.

Roalson E.H., Roberts W.R. 2016. Distinct Processes Drive Diversification in Different Clades of Gesneriaceae. Systematic Biology. 65:662–684.

Rose J.P., Kleist T.J., Löfstrand S.D., Drew B.T., Schönenberger J., Sytsma K.J. 2018. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. Molecular Phylogenetics and Evolution. 122:59–79.

Ruchisansakun S., van der Niet T., Janssens S.B., Triboun P., Techaprasan J., Jenjittikul T., Suksathan P. 2016. Phylogenetic Analyses of Molecular Data and Reconstruction of Morphological Character Evolution in Asian Impatiens section Semeiocardium (Balsaminaceae). Systematic Botany. 40:1063–1074.

Salariato D.L., Zuloaga F.O., Franzke A., Mummenhoff K., Al-Shehbaz I.A. 2016. Diversification patterns in the CES clade (Brassicaceae tribes Cremolobeae, Eudemeae, Schizopetaleae) in Andean South America. Botanical Journal of the Linnean Society. 181:543–566.

Särkinen T., Bohs L., Olmstead R.G., Knapp S. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol. 13:214.

Schliep K.P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27:592–593.

Schneider A.C., Moore A.J. 2017. Parallel Pleistocene amphitropical disjunctions of a parasitic plant and its host. American Journal of Botany. 104:1745–1755.

Shimizu K.K., Tsuchimatsu T. 2015. Evolution of Selfing: Recurrent Patterns in Molecular Adaptation. Annual Review of Ecology, Evolution, and Systematics. 46:593–622.

Smith S.A., Beaulieu J.M. 2009. Life history influences rates of climatic niche evolution in flowering plants. Proceedings of the Royal Society B: Biological Sciences. 276:4345–4352.

Smith S.A., Brown J.W. 2018. Constructing a broadly inclusive seed plant phylogeny. American Journal of Botany. 105:302–314.

Spriggs E.L., Christin P.-A., Edwards E.J. 2014. C4 Photosynthesis Promoted Species Diversification during the Miocene Grassland Expansion. PLOS ONE. 9:e97722.

Stearns S.C. 1992. The evolution of life histories. Oxford university press Oxford.

Stebbins G.L. 1950. Variation and evolution in plants. Variation and evolution in plants. Columbia University Press.

Stebbins G.L. 1974. Flowering plants: evolution above the species level. Cambridge, Mass: Belknap Press of Harvard University Press.

Takebayashi N., Morrell P.L. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. American Journal of Botany. 88:1143–1150.

Teskey R., Wertin T., Bauweraerts I., Ameye M., Mcguire M.A., Steppe K. 2015. Responses of tree species to heat waves and extreme heat events. Plant, Cell & Environment. 38:1699–1712.

Tribble C.M., May M.R., Jackson-Gain A., Zenil-Ferguson R., Specht C.D., Rothfels C.J. 2021. Unearthing modes of climatic adaptation in underground storage organs across Liliales. .

Vasconcelos T., Boyko J.D., Beaulieu J.M. 2021. Linking mode of seed dispersal and climatic niche evolution in flowering plants. J. Biogeogr. n/a.

Vasconcelos T.N.C., Alcantara S., Andrino C.O., Forest F., Reginato M., Simon M.F., Pirani J.R. 2020. Fast diversification through a mosaic of evolutionary histories characterizes the endemic flora of ancient Neotropical mountains. Proceedings of the Royal Society B: Biological Sciences. 287:20192933.

Venable D.L. 2007. Bet Hedging in a Guild of Desert Annuals. Ecology. 88:1086–1090.

Venable D.L., Lawlor L. 1980. Delayed germination and dispersal in desert annuals: Escape in space and time. Oecologia. 46:272–282.

de Vos J.M., Hughes C.E., Schneeweiss G.M., Moore B.R., Conti E. 2014. Heterostyly accelerates diversification via reduced extinction in primroses. Proceedings of the Royal Society B: Biological Sciences. 281:20140075.

Wagner W.L., Hoch P.C., Raven P.H. 2007. Revised classification of the Onagraceae. Systematic Botany Monographs.

WCSP (2022). 'World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://wcsp.science.kew.org/ [Note: this data is part of a dataset to be officially released in November 2022. The dataset was made available for manuscripts that are part of a special issue on the WCVP to be published by New Phytologist in 2023.]

Yan H.-F., Zhang C.-Y., Anderberg A.A., Hao G., Ge X.-J., Wiens J.J. 2018. What explains high plant richness in East Asia? Time and diversification in the tribe Lysimachieae (Primulaceae). New Phytologist. 219:436–448.

# Appendix

**Table 1.** Inequalities describing how expectations of climatic optima and variance will differ for each climatic variable. When $\theta_a > \theta_p$, we expect the climatic optima for that variable to be greater for annuals than perennials. When $\theta_a < \theta_p$, we expect the climatic optima for that variable to be greater for perennials than annuals. For all variables, we expect annuals to present higher rates of climatic niche evolution ($\sigma^2$) for annuals than perennials.

| | Mean vars | | | Seasonality vars | | Extreme vars | | |
|---|---|---|---|---|---|---|---|---|
| | BIO1 | BIO12 | AI | BIO4 | BIO15 | BIO5 | BIO6 | BIO14 |
| Estimated $\theta$ | $\theta_a > \theta_p$ | $\theta_a < \theta_p$ | | $\theta_a > \theta_p$ | | $\theta_a > \theta_p$ | | $\theta_a < \theta_p$ |
| Estimated $\sigma^2$ | $\sigma^2_a > \sigma^2_p$ | | | | | | | |

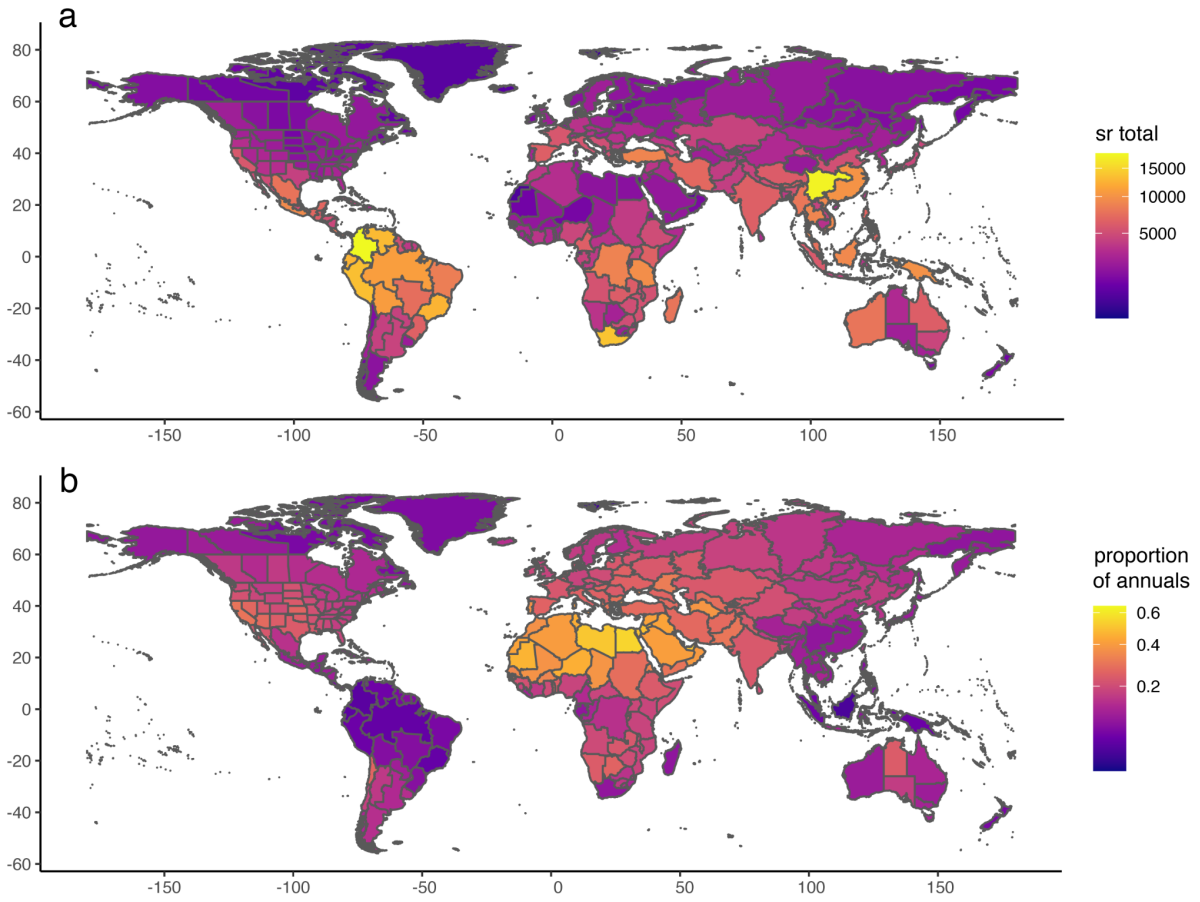


**Figure 1.** Global distribution of vascular plant diversity and proportion of annual plants. (a) Total species richness of vascular plants by botanical country according to the WCVP database (WCVP, 2022), and (b) Proportion of annual plants in relation to total species richness. Y-axis: longitude; x-axis: latitude.

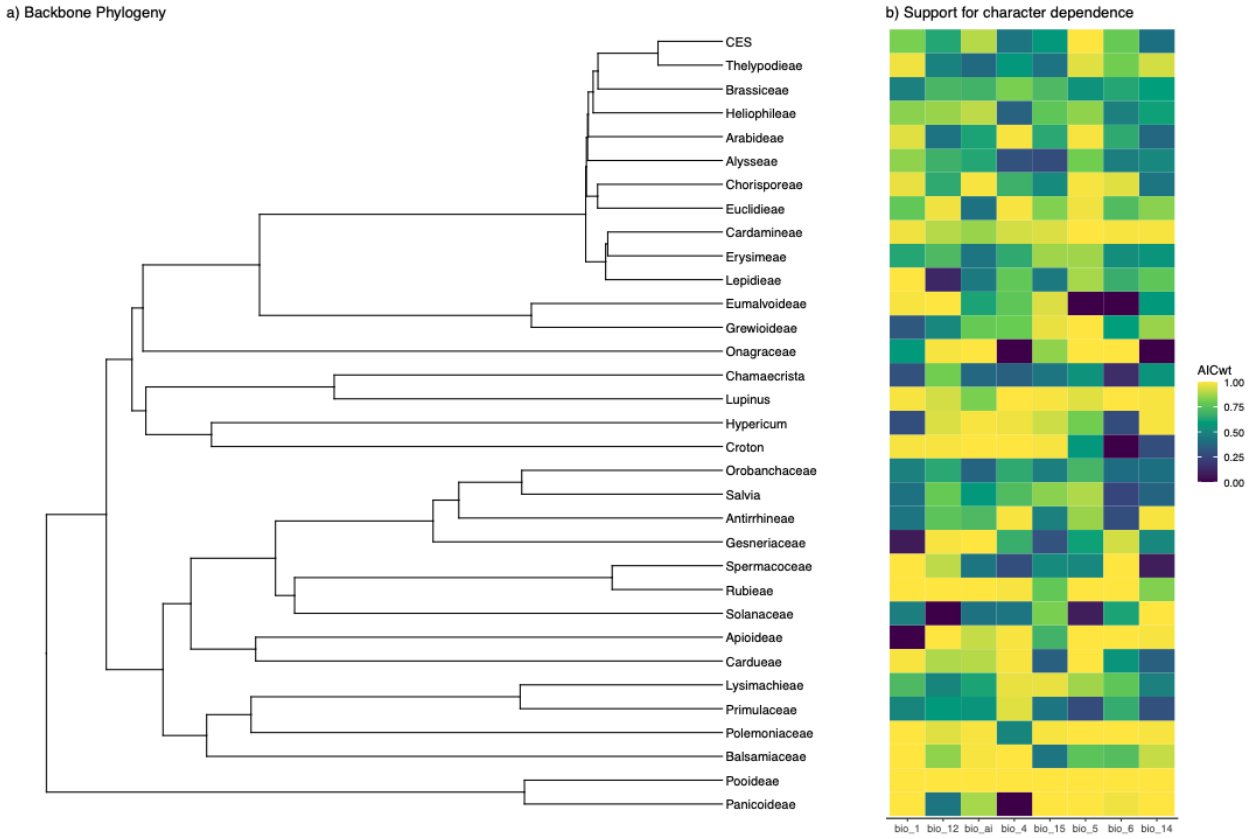


**Figure 2.** a) the backbone phylogeny based on Smith and Brown (2018) pruned to include groups analyzed in this paper. b) Each square of the heat map represents the summed AIC weight of models which support some form of character dependence. These do not distinguish the type of character dependence, so support for an OUM model (variable optimum model) will be summed with support for an OUV model (variable rate model). The columns are broken into groups with by (1) mean: BIO1 = Annual Mean Temperature, BIO12 = Annual Precipitation, BIOAI = Aridity Index, (2) seasonality: BIO4 = Temperature Seasonality (standard deviation ×100), BIO15 = Precipitation Seasonality (Coefficient of Variation), and (3) extreme season: BIO5 = Max Temperature of Warmest Month, BIO6 = Min Temperature of Coldest Month, BIO14 = Precipitation of Driest Month.

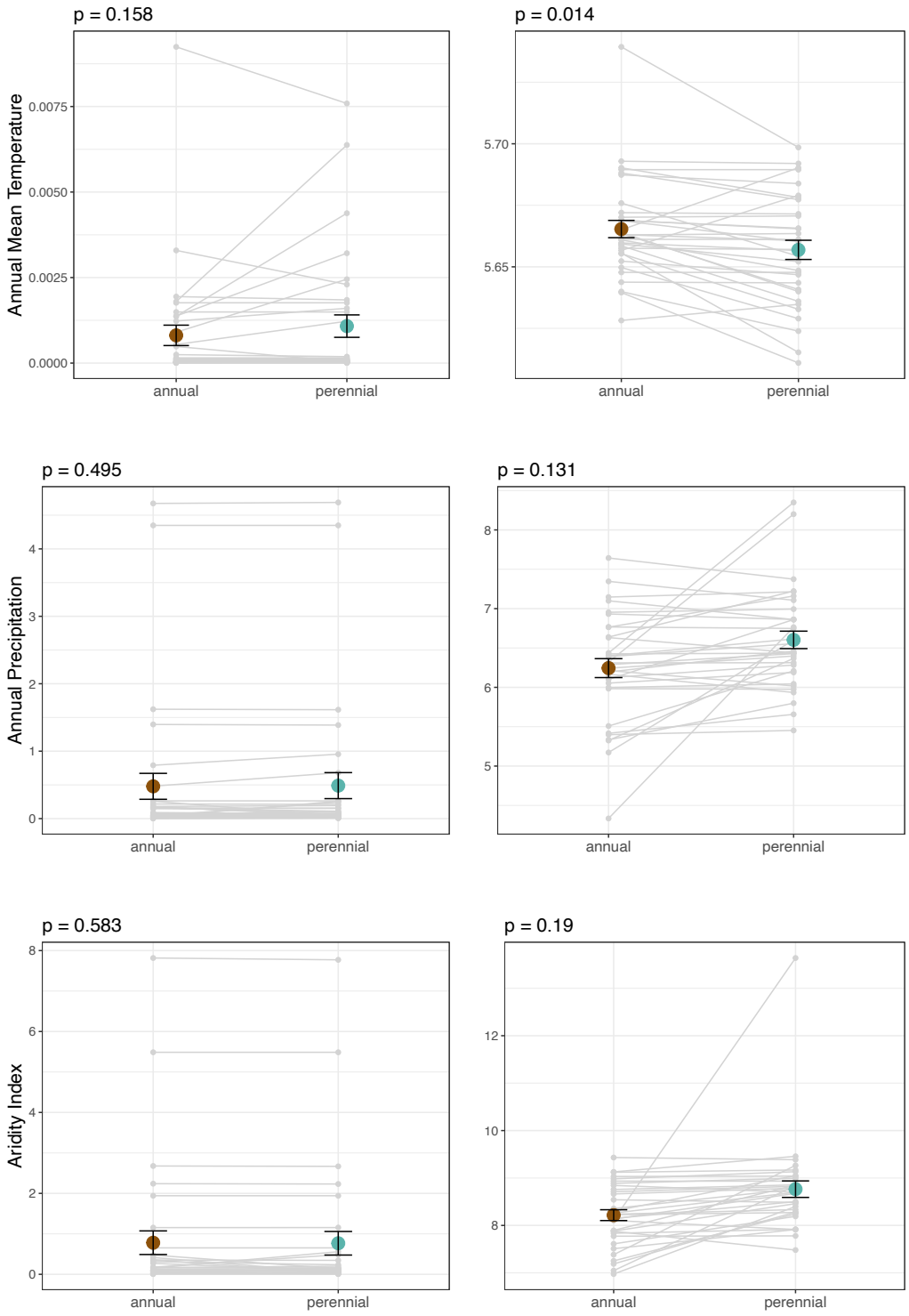


**Figure 3.** Each row represents a different measure of mean climatic niche. Model averaged parameter estimates for $\sigma^2$ (left column) and $\theta$ (right column) for each given observed state (annual and perennial) averaged over all clades. Grey lines represent individual clade comparisons between estimates associated with each observed state. Foreground points are the mean values of each parameter estimate.

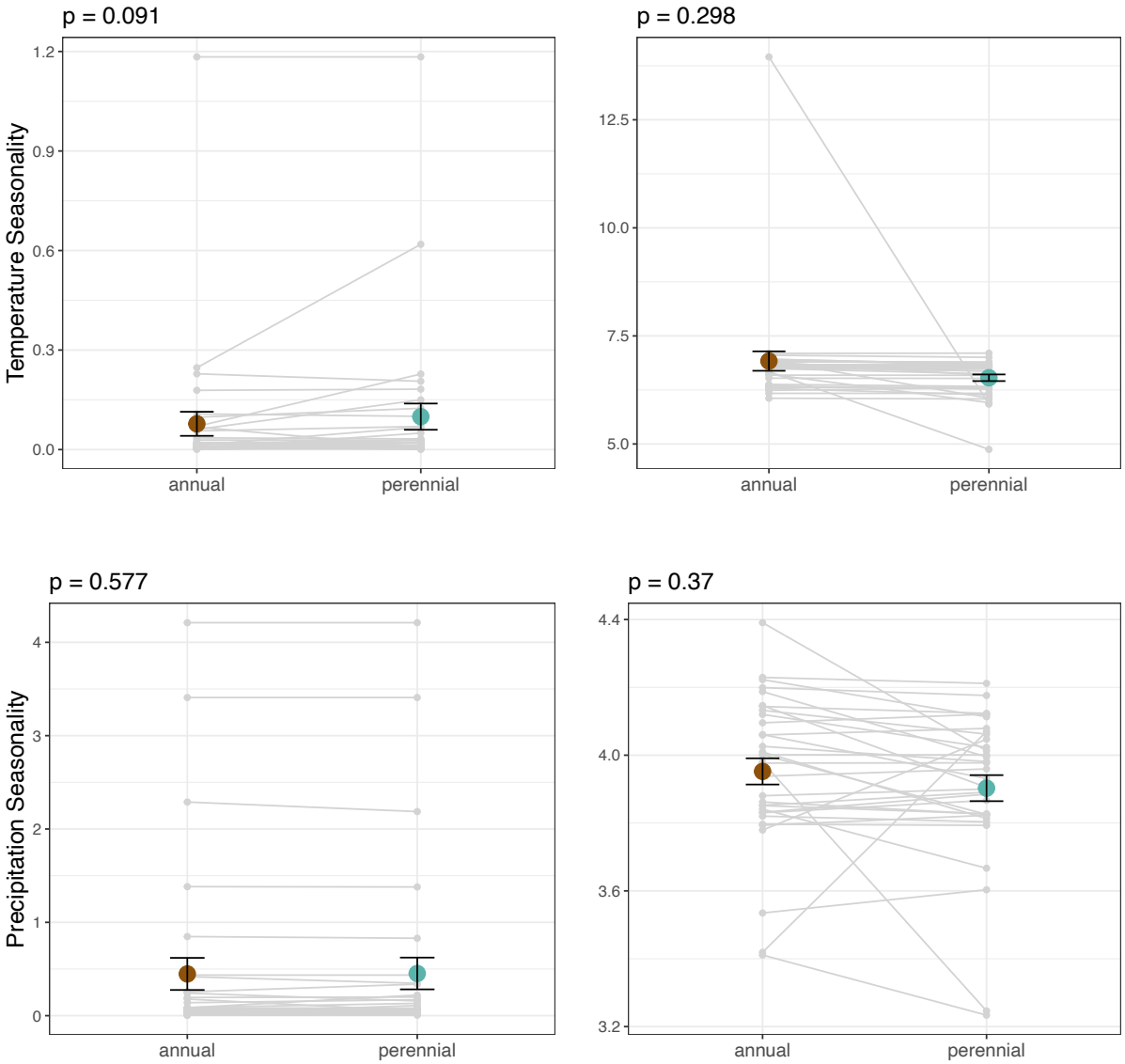


**Figure 4.** Each row represents a different measure of climatic variability. Model averaged parameter estimates for $\sigma^2$ (left column) and $\theta$ (right column) for each given observed state (annual and perennial) averaged over all clades. Grey lines represent individual clade comparisons between estimates associated with each observed state. Foreground points are the mean values of each parameter estimate.

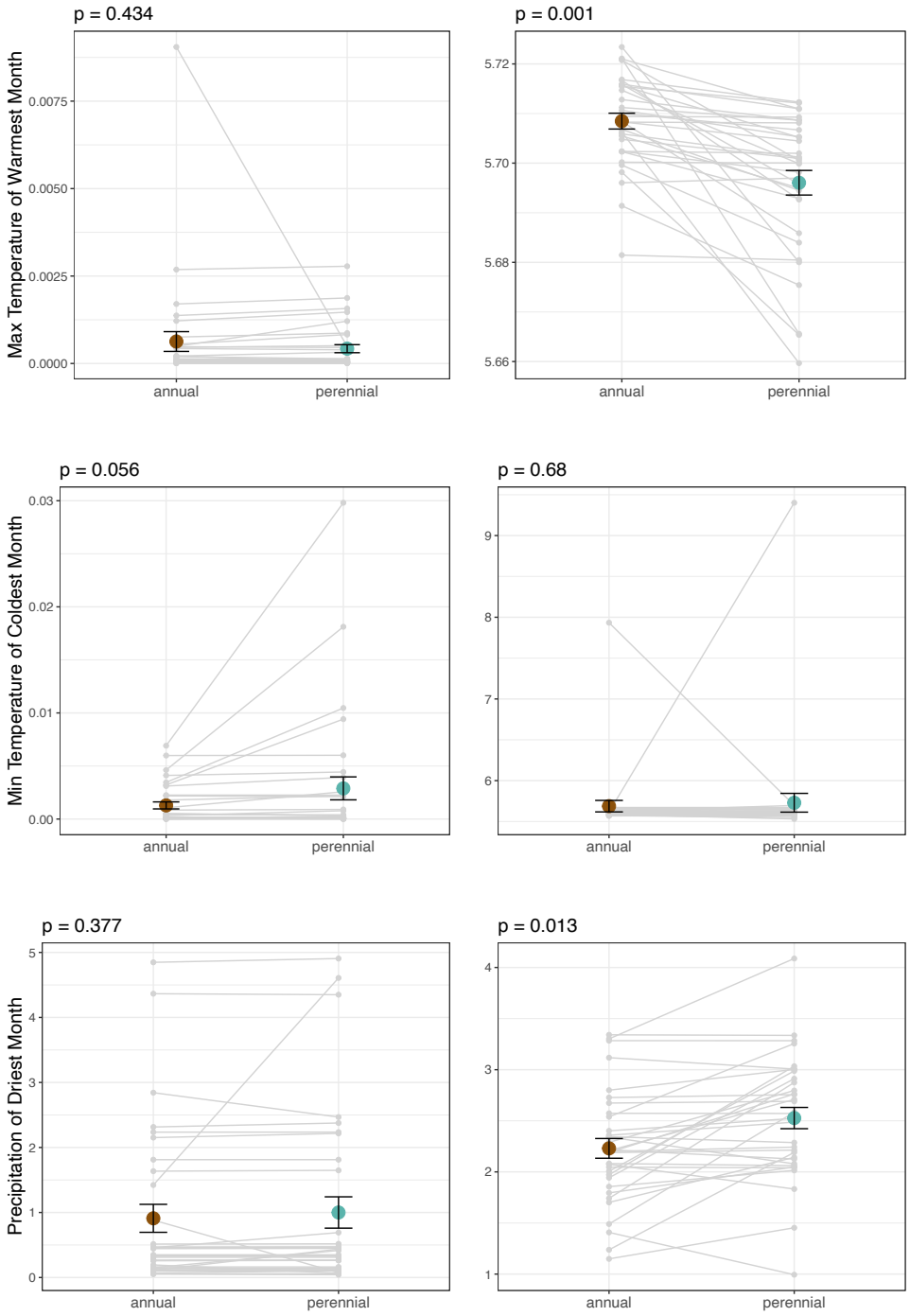


**Figure 5.** Each row represents a different measure of extreme climatic conditions. Model averaged parameter estimates for $\sigma^2$ (left column) and $\theta$ (right column) for each given observed state (annual and perennial) averaged over all clades. Grey lines represent individual clade comparisons between estimates associated with each observed state. Foreground points are the mean values of each parameter estimate.

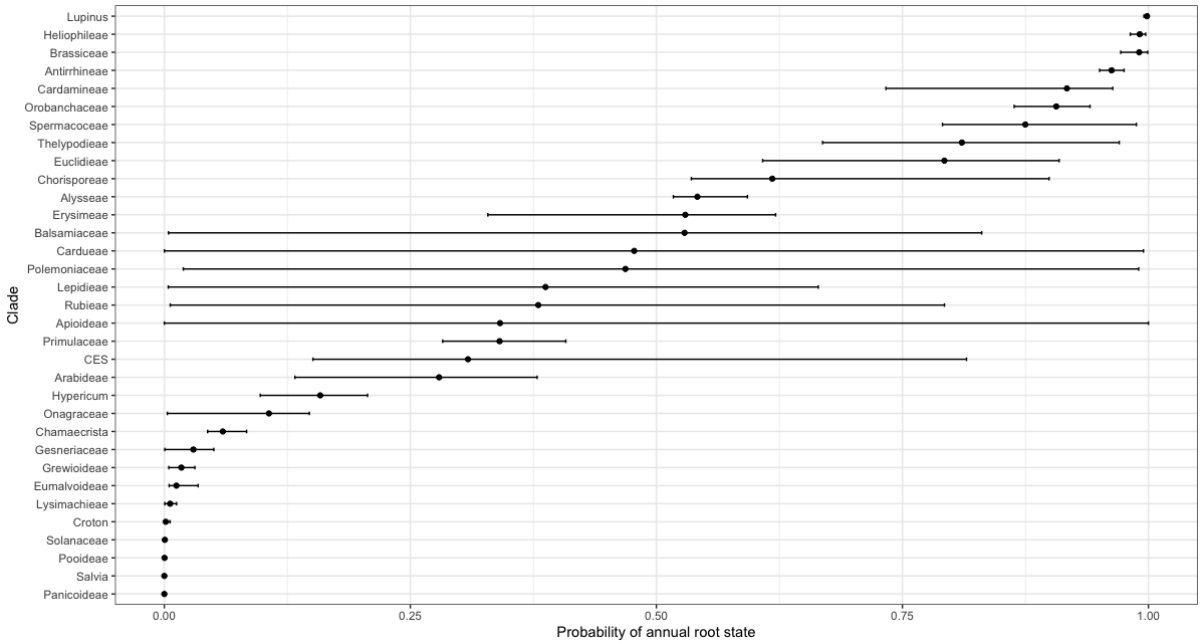


**Figure 6.** Probability of an annual root state averaged across all climatic variables. Error bars show the range of root probabilities depending on a given bioclimatic variable.

**Table S1.** Parameter estimates from the model averaged hOUwie fits for BIO1

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.0438796 | 6.66380991 | 0.00122882 | 5.65231669 |
| Alysseae | perennial | 0.02477475 | 6.66380991 | 0.00160141 | 5.64680254 |
| Antirrhineae | annual | 0.02940751 | 0.04643861 | 2.65E-05 | 5.66873603 |
| Antirrhineae | perennial | 0.01245086 | 0.04643861 | 2.68E-05 | 5.66024456 |
| Apioideae | annual | 0.00058539 | 0.05740227 | 5.01E-05 | 5.64775307 |
| Apioideae | perennial | 0.03378897 | 0.05740227 | 5.02E-05 | 5.64775307 |
| Arabideae | annual | 0.10684912 | 1.52370347 | 0.00149364 | 5.63994426 |
| Arabideae | perennial | 0.14433569 | 1.52370347 | 0.0014937 | 5.62388522 |
| Balsamiaceae | annual | 0.01410781 | 0.15320731 | 6.99E-05 | 5.65668698 |
| Balsamiaceae | perennial | 0.01362342 | 0.15320731 | 6.99E-05 | 5.67899648 |
| Brassiceae | annual | 0.04331724 | 4.8301285 | 0.00194311 | 5.66307004 |
| Brassiceae | perennial | 0.01707056 | 4.8301285 | 0.00184615 | 5.66067065 |
| Cardamineae | annual | 0.05677185 | 1.11450045 | 0.00054358 | 5.65534687 |
| Cardamineae | perennial | 0.0285243 | 1.11450045 | 0.00122269 | 5.63275777 |
| Cardueae | annual | 0.07259066 | 0.0707443 | 3.76E-05 | 5.66395545 |
| Cardueae | perennial | 0.01356082 | 0.0707443 | 3.76E-05 | 5.63997566 |
| CES | annual | 0.08029827 | 0.33840208 | 8.23E-05 | 5.6497311 |
| CES | perennial | 0.02859758 | 0.33840208 | 8.09E-05 | 5.6289667 |
| Chamaecrista | annual | 0.08542555 | 0.36907315 | 1.65E-05 | 5.69291157 |
| Chamaecrista | perennial | 0.01174591 | 0.36907315 | 2.00E-05 | 5.69199213 |
| Chorisporeae | annual | 0.08315248 | 6.46759564 | 0.00136669 | 5.6556097 |
| Chorisporeae | perennial | 0.09301251 | 6.46759564 | 0.00438094 | 5.61525076 |
| Croton | annual | 0.0062452 | 0.08819602 | 1.65E-05 | 5.66504307 |
| Croton | perennial | 0.00428844 | 0.08819602 | 1.65E-05 | 5.69025885 |
| Erysimeae | annual | 0.05569405 | 4.2867926 | 0.00090283 | 5.64377249 |
| Erysimeae | perennial | 0.06482172 | 4.2867926 | 0.00245118 | 5.64349519 |
| Euclidieae | annual | 0.03000192 | 6.63279819 | 0.00924364 | 5.63957406 |
| Euclidieae | perennial | 0.04529422 | 6.63279819 | 0.00758932 | 5.6109697 |
| Eumalvoideae | annual | 0.03829349 | 0.04342246 | 1.84E-05 | 5.69024117 |
| Eumalvoideae | perennial | 0.01247766 | 0.04342246 | 2.30E-05 | 5.67821104 |
| Gesneriaceae | annual | 0.01125012 | 1.46E-06 | 7.05E-06 | 5.73938276 |
| Gesneriaceae | perennial | 0.00243141 | 1.46E-06 | 6.83E-06 | 5.69851976 |
| Grewioideae | annual | 0.01822188 | 0.02790307 | 5.30E-06 | 5.6894835 |
| Grewioideae | perennial | 0.00155572 | 0.02790307 | 5.65E-06 | 5.68949456 |
| Heliophileae | annual | 0.03667496 | 6.09150422 | 0.00024601 | 5.66884398 |
| Heliophileae | perennial | 0.02186931 | 6.09150422 | 0.00018783 | 5.66569729 |
| Hypericum | annual | 0.08537842 | 0.38584394 | 0.00015109 | 5.66297032 |
| Hypericum | perennial | 0.01726038 | 0.38584394 | 0.00015248 | 5.6635204 |
| Lepidieae | annual | 0.28073252 | 0.00872247 | 0.00048522 | 5.66117101 |
| Lepidieae | perennial | 0.21380198 | 0.00872247 | 3.37E-05 | 5.66116977 |
| Lupinus | annual | 0.02506642 | 0.01243178 | 1.55E-05 | 5.67202581 |
| Lupinus | perennial | 0.06983453 | 0.01243178 | 9.55E-05 | 5.6714866 |
| Lysimachieae | annual | 0.03527644 | 0.07063979 | 3.79E-05 | 5.67589926 |
| Lysimachieae | perennial | 0.02933563 | 0.07063979 | 3.90E-05 | 5.65429022 |
| Onagraceae | annual | 0.15590832 | 2.27618099 | 0.00176227 | 5.65971482 |
| Onagraceae | perennial | 0.12897585 | 2.27618099 | 0.00176195 | 5.65664412 |
| Orobanchaceae | annual | 0.04551376 | 0.10480994 | 0.00011349 | 5.65928522 |
| Orobanchaceae | perennial | 0.03955944 | 0.10480994 | 0.0001157 | 5.65210788 |
| Panicoideae | annual | 0.02899001 | 9.22595692 | 0.00329369 | 5.68736146 |
| Panicoideae | perennial | 0.04371949 | 9.22595692 | 0.00229697 | 5.68384497 |
| Polemoniaceae | annual | 0.01151066 | 0.06820139 | 2.61E-05 | 5.65755837 |
| Polemoniaceae | perennial | 0.00894239 | 0.06820139 | 8.35E-05 | 5.65705563 |
| Pooideae | annual | 0.02100831 | 6.21567172 | 0.00179795 | 5.65854111 |
| Pooideae | perennial | 0.03167847 | 6.21567172 | 0.00637541 | 5.63597646 |
| Primulaceae | annual | 0.02420475 | 0.1861174 | 0.00013539 | 5.62820689 |
| Primulaceae | perennial | 0.01241557 | 0.1861174 | 0.00013539 | 5.63476058 |
| Rubieae | annual | 0.07887715 | 7.60596218 | 0.00137441 | 5.6609244 |
| Rubieae | perennial | 0.02367558 | 7.60596218 | 0.00321191 | 5.64852974 |
| Salvia | annual | 0.01863225 | 0.08485602 | 2.61E-05 | 5.66888468 |
| Salvia | perennial | 0.01559416 | 0.08485602 | 2.61E-05 | 5.66545115 |
| Solanaceae | annual | 0.03571926 | 0.13652036 | 7.70E-05 | 5.67014821 |
| Solanaceae | perennial | 0.01489461 | 0.13652036 | 7.70E-05 | 5.67068859 |
| Spermacoceae | annual | 0.03551673 | 0.11004689 | 3.39E-05 | 5.68812404 |
| Spermacoceae | perennial | 0.01325236 | 0.11004689 | 3.39E-05 | 5.67732921 |
| Thelypodieae | annual | 0.04539937 | 0.34730271 | 0.00010741 | 5.66352413 |
| Thelypodieae | perennial | 0.05954099 | 0.34730271 | 0.00010601 | 5.64077808 |

**Table S2.** Parameter estimates from the model averaged hOUwie fits for BIO4

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.04402529 | 0.50513189 | 0.0214897 | 6.90085056 |
| Alysseae | perennial | 0.02115548 | 0.50513189 | 0.0206555 | 6.8964374 |
| Antirrhineae | annual | 0.03074464 | 0.05615667 | 0.00202029 | 6.7449743 |
| Antirrhineae | perennial | 0.013287 | 0.05615667 | 0.00748483 | 6.62181024 |
| Apioideae | annual | 0.05385838 | 0.09432435 | 0.01597978 | 6.83677007 |
| Apioideae | perennial | 0.01000302 | 0.09432435 | 0.01366324 | 6.80066289 |
| Arabideae | annual | 0.52479212 | 0.67707888 | 0.06067607 | 6.90746927 |
| Arabideae | perennial | 0.3550588 | 0.67707888 | 0.14993046 | 6.88769237 |
| Balsamiaceae | annual | 0.01437388 | 0.00400375 | 0.00938075 | 13.9544985 |
| Balsamiaceae | perennial | 0.01525116 | 0.00400375 | 0.00052895 | 5.92161613 |
| Brassiceae | annual | 0.04402116 | 1.51465082 | 0.10740962 | 6.78182407 |
| Brassiceae | perennial | 0.02645235 | 1.51465082 | 0.1001567 | 6.78182379 |
| Cardamineae | annual | 0.05569111 | 2.01050027 | 0.24669257 | 6.81475042 |
| Cardamineae | perennial | 0.03214692 | 2.01050027 | 0.61875634 | 6.81510396 |
| Cardueae | annual | 0.05471326 | 0.164068 | 0.00997412 | 6.83333 |
| Cardueae | perennial | 0.03127461 | 0.164068 | 0.021815 | 6.8371473 |
| CES | annual | 0.0777941 | 0.00390449 | 0.00447383 | 6.75219765 |
| CES | perennial | 0.03003493 | 0.00390449 | 0.00591825 | 6.73773539 |
| Chamaecrista | annual | 0.08230412 | 0.20181379 | 0.00240117 | 6.05770015 |
| Chamaecrista | perennial | 0.01024206 | 0.20181379 | 0.00236646 | 6.04145922 |
| Chorisporeae | annual | 0.07476909 | 3.43309151 | 0.0975446 | 7.09512575 |
| Chorisporeae | perennial | 0.08212299 | 3.43309151 | 0.1243816 | 7.10139911 |
| Croton | annual | 0.03874914 | 0.10079235 | 0.01904948 | 6.93732722 |
| Croton | perennial | 0.00457246 | 0.10079235 | 0.01130981 | 6.05811266 |
| Erysimeae | annual | 0.05077356 | 0.92002517 | 0.05616603 | 6.95733832 |
| Erysimeae | perennial | 0.06039982 | 0.92002517 | 0.06972218 | 6.84685833 |
| Euclidieae | annual | 0.02911251 | 0.53731559 | 0.00364014 | 7.05320916 |
| Euclidieae | perennial | 0.04189974 | 0.53731559 | 0.04962295 | 7.00469944 |
| Eumalvoideae | annual | 0.04550747 | 0.07355345 | 0.02100277 | 6.37815164 |
| Eumalvoideae | perennial | 0.01531847 | 0.07355345 | 0.01325006 | 6.33094879 |
| Gesneriaceae | annual | 0.00318162 | 0.00073853 | 0.00285393 | 6.17085499 |
| Gesneriaceae | perennial | 0.00242287 | 0.00073853 | 0.00279004 | 6.17196296 |
| Grewioideae | annual | 0.01938833 | 0.05716908 | 0.0079496 | 6.28565391 |
| Grewioideae | perennial | 0.00189339 | 0.05716908 | 0.00879124 | 6.13622989 |
| Heliophileae | annual | 0.03419233 | 0.62867033 | 0.00727061 | 6.52207105 |
| Heliophileae | perennial | 0.01728062 | 0.62867033 | 0.004512 | 6.51685737 |
| Hypericum | annual | 0.09399561 | 0.02725369 | 0.06798064 | 6.93723798 |
| Hypericum | perennial | 0.01269791 | 0.02725369 | 0.00794815 | 6.83169659 |
| Lepidieae | annual | 0.25400584 | 1.14314901 | 0.17858675 | 6.79070592 |
| Lepidieae | perennial | 0.20257109 | 1.14314901 | 0.18199736 | 6.68723438 |
| Lupinus | annual | 0.04310612 | 0.04786416 | 0.00280172 | 6.67441324 |
| Lupinus | perennial | 0.11680511 | 0.04786416 | 0.03104991 | 4.87468226 |
| Lysimachieae | annual | 0.03684639 | 0.0181623 | 0.00073335 | 6.85649666 |
| Lysimachieae | perennial | 0.02895284 | 0.0181623 | 0.01024726 | 6.74457896 |
| Onagraceae | annual | 0.05946736 | 0.05839793 | 0.01966654 | 6.59235285 |
| Onagraceae | perennial | 0.00386964 | 0.05839793 | 0.02759087 | 6.59038337 |
| Orobanchaceae | annual | 0.04635661 | 0.1061356 | 0.01752305 | 6.75953878 |
| Orobanchaceae | perennial | 0.04025596 | 0.1061356 | 0.02163064 | 6.77726258 |
| Panicoideae | annual | 0.03284142 | 7.3301921 | 1.18361673 | 6.30995609 |
| Panicoideae | perennial | 0.04869404 | 7.3301921 | 1.18361673 | 6.30995609 |
| Polemoniaceae | annual | 0.01171329 | 0.0124009 | 0.00378928 | 6.61445969 |
| Polemoniaceae | perennial | 0.00820808 | 0.0124009 | 0.00447191 | 5.95634138 |
| Pooideae | annual | 0.02059462 | 0.91006226 | 0.06981983 | 6.79387246 |
| Pooideae | perennial | 0.02835526 | 0.91006226 | 0.22768235 | 6.79387224 |
| Primulaceae | annual | 0.02537533 | 2.75171986 | 0.22837268 | 6.96082126 |
| Primulaceae | perennial | 0.01166081 | 2.75171986 | 0.20583799 | 6.84584095 |
| Rubieae | annual | 0.11559292 | 0.73390042 | 0.01330869 | 6.78873939 |
| Rubieae | perennial | 0.01060442 | 0.73390042 | 0.06720687 | 6.79041683 |
| Salvia | annual | 0.01671827 | 0.04699644 | 0.01048412 | 6.89295788 |
| Salvia | perennial | 0.01561789 | 0.04699644 | 0.01224901 | 6.58687126 |
| Solanaceae | annual | 0.03175675 | 0.10939134 | 0.02884998 | 6.34154496 |
| Solanaceae | perennial | 0.01342037 | 0.10939134 | 0.0281287 | 6.26688753 |
| Spermacoceae | annual | 0.04342677 | 0.15412548 | 0.03526634 | 6.24154558 |
| Spermacoceae | perennial | 0.00688428 | 0.15412548 | 0.03213335 | 6.30131514 |
| Thelypodieae | annual | 0.04375139 | 0.00783391 | 0.00513416 | 6.7591833 |
| Thelypodieae | perennial | 0.07228538 | 0.00783391 | 0.00817083 | 6.68778877 |

Table S3. Parameter estimates from the model averaged hOUwie fits for BIO5

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.04425457 | 2.82044388 | 0.00046104 | 5.70589138 |
| Alysseae | perennial | 0.02518721 | 2.82044388 | 0.00051073 | 5.70120769 |
| Antirrhineae | annual | 0.03031724 | 0.09562457 | 3.33E-05 | 5.71280648 |
| Antirrhineae | perennial | 0.01284412 | 0.09562457 | 4.40E-05 | 5.7086574 |
| Apioideae | annual | 0.05876041 | 0.14644361 | 7.44E-05 | 5.70679809 |
| Apioideae | perennial | 0.00627024 | 0.14644361 | 7.43E-05 | 5.69449996 |
| Arabideae | annual | 0.11370213 | 2.85991638 | 0.00170142 | 5.69139876 |
| Arabideae | perennial | 0.14269825 | 2.85991638 | 0.0018734 | 5.67541569 |
| Balsamiaceae | annual | 0.01388103 | 0.58288565 | 0.00019675 | 5.69603103 |
| Balsamiaceae | perennial | 0.01108232 | 0.58288565 | 0.00013285 | 5.6969941 |
| Brassiceae | annual | 0.04699898 | 3.24276752 | 0.00137183 | 5.70940182 |
| Brassiceae | perennial | 0.01527256 | 3.24276752 | 0.00157252 | 5.70927405 |
| Cardamineae | annual | 0.05377991 | 3.2632822 | 0.00121778 | 5.69961339 |
| Cardamineae | perennial | 0.032428 | 3.2632822 | 0.0014688 | 5.68396766 |
| Cardueae | annual | 0.04871319 | 0.09687431 | 4.17E-05 | 5.72076313 |
| Cardueae | perennial | 0.03332827 | 0.09687431 | 4.17E-05 | 5.70053771 |
| CES | annual | 0.09100693 | 0.25476791 | 3.76E-05 | 5.69817177 |
| CES | perennial | 0.03158301 | 0.25476791 | 5.42E-05 | 5.66537602 |
| Chamaecrista | annual | 0.06983725 | 0.46858579 | 2.52E-05 | 5.71545846 |
| Chamaecrista | perennial | 0.01246693 | 0.46858579 | 2.55E-05 | 5.71211278 |
| Chorisporeae | annual | 0.07616816 | 1.53255972 | 0.00021348 | 5.72108751 |
| Chorisporeae | perennial | 0.08669855 | 1.53255972 | 0.00032482 | 5.66559455 |
| Croton | annual | 0.01895065 | 0.1323376 | 2.04E-05 | 5.71590343 |
| Croton | perennial | 0.00492704 | 0.1323376 | 2.04E-05 | 5.71097597 |
| Erysimeae | annual | 0.06234572 | 1.81727809 | 0.00054987 | 5.70232049 |
| Erysimeae | perennial | 0.05857594 | 1.81727809 | 0.00082668 | 5.69286933 |
| Euclidieae | annual | 0.02896242 | 4.95123364 | 0.00904794 | 5.70626809 |
| Euclidieae | perennial | 0.04619274 | 4.95123364 | 0.00048371 | 5.65967287 |
| Eumalvoideae | annual | 0.03944987 | 0.06236298 | 2.78E-05 | 5.7082662 |
| Eumalvoideae | perennial | 0.01354125 | 0.06236298 | 2.78E-05 | 5.70811508 |
| Gesneriaceae | annual | 0.00334229 | 0.05270552 | 9.74E-06 | 5.70236796 |
| Gesneriaceae | perennial | 0.00264402 | 0.05270552 | 9.73E-06 | 5.70198224 |
| Grewioideae | annual | 0.01893903 | 0.14106783 | 1.30E-05 | 5.72100168 |
| Grewioideae | perennial | 0.00151463 | 0.14106783 | 1.19E-05 | 5.71092692 |
| Heliophileae | annual | 0.0357692 | 2.34640908 | 0.00013408 | 5.70831532 |
| Heliophileae | perennial | 0.02175662 | 2.34640908 | 0.00010853 | 5.70446631 |
| Hypericum | annual | 0.08843732 | 0.11337648 | 0.00021589 | 5.71000005 |
| Hypericum | perennial | 0.0182061 | 0.11337648 | 4.25E-05 | 5.70668625 |
| Lepidieae | annual | 0.24857415 | 4.26456567 | 0.00268474 | 5.71064194 |
| Lepidieae | perennial | 0.20838006 | 4.26456567 | 0.00277999 | 5.705299 |
| Lupinus | annual | 0.05797418 | 0.03439922 | 2.65E-05 | 5.70236488 |
| Lupinus | perennial | 0.13153395 | 0.03439922 | 7.67E-05 | 5.69817026 |
| Lysimachieae | annual | 0.03323861 | 0.16572421 | 3.63E-05 | 5.71469912 |
| Lysimachieae | perennial | 0.02841232 | 0.16572421 | 3.71E-05 | 5.69988256 |
| Onagraceae | annual | 0.03641358 | 0.67560333 | 0.0004794 | 5.71586994 |
| Onagraceae | perennial | 0.00543767 | 0.67560333 | 0.00045825 | 5.6957791 |
| Orobanchaceae | annual | 0.04609974 | 0.34577101 | 0.00011566 | 5.70479609 |
| Orobanchaceae | perennial | 0.04319785 | 0.34577101 | 0.00013263 | 5.70094528 |
| Panicoideae | annual | 0.02473338 | 2.23586369 | 0.00042825 | 5.71684926 |
| Panicoideae | perennial | 0.03880131 | 2.23586369 | 0.00039091 | 5.71230186 |
| Polemoniaceae | annual | 0.01165917 | 0.09109349 | 3.75E-05 | 5.71668617 |
| Polemoniaceae | perennial | 0.00876047 | 0.09109349 | 4.22E-05 | 5.69266014 |
| Pooideae | annual | 0.01997802 | 1.79546201 | 0.00049092 | 5.70839621 |
| Pooideae | perennial | 0.0306571 | 1.79546201 | 0.00121053 | 5.68587639 |
| Primulaceae | annual | 0.02498328 | 0.21131656 | 0.00010666 | 5.68146218 |
| Primulaceae | perennial | 0.01184284 | 0.21131656 | 0.00010748 | 5.68045901 |
| Rubieae | annual | 0.05448645 | 4.40248025 | 0.00075548 | 5.70546933 |
| Rubieae | perennial | 0.04163527 | 4.40248025 | 0.00086976 | 5.69494975 |
| Salvia | annual | 0.02069279 | 0.08465005 | 2.74E-05 | 5.71564015 |
| Salvia | perennial | 0.01502724 | 0.08465005 | 3.43E-05 | 5.70519573 |
| Solanaceae | annual | 0.02891487 | 0.15523176 | 7.65E-05 | 5.7001863 |
| Solanaceae | perennial | 0.01337719 | 0.15523176 | 7.65E-05 | 5.70001728 |
| Spermacoceae | annual | 0.03506212 | 0.04059133 | 9.92E-06 | 5.71121359 |
| Spermacoceae | perennial | 0.01202575 | 0.04059133 | 1.00E-05 | 5.70865015 |
| Thelypodieae | annual | 0.04322226 | 0.11086913 | 6.73E-05 | 5.72341629 |
| Thelypodieae | perennial | 0.07238157 | 0.11086913 | 9.21E-05 | 5.68001463 |

Table S4. Parameter estimates from the model averaged hOUwie fits for BIO6

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.04472623 | 2.96503564 | 0.00178492 | 5.60076864 |
| Alysseae | perennial | 0.02478184 | 2.96503564 | 0.00219527 | 5.5984029 |
| Antirrhineae | annual | 0.02848683 | 0.03742949 | 4.22E-05 | 5.62559986 |
| Antirrhineae | perennial | 0.01220293 | 0.03742949 | 4.71E-05 | 5.69633689 |
| Apioideae | annual | 0.01190875 | 0.1005569 | 0.00016355 | 5.61684919 |
| Apioideae | perennial | 0.02026953 | 0.1005569 | 0.00016355 | 5.59624181 |
| Arabideae | annual | 0.1131574 | 1.00919909 | 0.00225588 | 5.58165672 |
| Arabideae | perennial | 0.14555317 | 1.00919909 | 0.00225588 | 5.57231951 |
| Balsamiaceae | annual | 0.01428589 | 0.00316519 | 2.64E-05 | 5.61634716 |
| Balsamiaceae | perennial | 0.01352837 | 0.00316519 | 4.68E-05 | 9.40136049 |
| Brassiceae | annual | 0.04830821 | 6.35900042 | 0.00411011 | 5.61942721 |
| Brassiceae | perennial | 0.02026394 | 6.35900042 | 0.00443601 | 5.61443573 |
| Cardamineae | annual | 0.05078855 | 9.75096861 | 0.00691248 | 5.60377737 |
| Cardamineae | perennial | 0.03177057 | 9.75096861 | 0.02980824 | 5.58540293 |
| Cardueae | annual | 0.06227451 | 0.06880442 | 0.00010617 | 5.61370299 |
| Cardueae | perennial | 0.02301537 | 0.06880442 | 0.0001068 | 5.58938693 |
| CES | annual | 0.0791401 | 0.00396446 | 7.98E-05 | 5.57178955 |
| CES | perennial | 0.03479396 | 0.00396446 | 1.07E-05 | 5.56875346 |
| Chamaecrista | annual | 0.09622943 | 0.11880452 | 1.82E-05 | 5.67148368 |
| Chamaecrista | perennial | 0.01146116 | 0.11880452 | 1.83E-05 | 5.6711077 |
| Chorisporeae | annual | 0.08033507 | 8.41196231 | 0.00321585 | 5.58656082 |
| Chorisporeae | perennial | 0.08755152 | 8.41196231 | 0.00941716 | 5.53147387 |
| Croton | annual | 0.02991996 | 0.05484939 | 4.80E-05 | 5.66243354 |
| Croton | perennial | 0.00530731 | 0.05484939 | 4.75E-05 | 5.66244962 |
| Erysimeae | annual | 0.05955994 | 0.39318421 | 0.00032684 | 5.58871394 |
| Erysimeae | perennial | 0.06206891 | 0.39318421 | 0.00078375 | 5.591116 |
| Euclidieae | annual | 0.03139276 | 1.67154646 | 0.00311077 | 5.57249656 |
| Euclidieae | perennial | 0.05188074 | 1.67154646 | 0.00392761 | 5.54840272 |
| Eumalvoideae | annual | 0.03826592 | 0.04090619 | 4.44E-05 | 5.66130735 |
| Eumalvoideae | perennial | 0.01349127 | 0.04090619 | 4.44E-05 | 5.64490202 |
| Gesneriaceae | annual | 0.00597533 | 0.00011239 | 1.81E-05 | 7.93344527 |
| Gesneriaceae | perennial | 0.00232387 | 0.00011239 | 1.74E-05 | 5.69702837 |
| Grewioideae | annual | 0.01761991 | 0.02822903 | 2.15E-05 | 5.66027413 |
| Grewioideae | perennial | 0.001786 | 0.02822903 | 2.67E-05 | 5.66292423 |
| Heliophileae | annual | 0.03500753 | 0.99048755 | 8.42E-05 | 5.62648548 |
| Heliophileae | perennial | 0.02276974 | 0.99048755 | 0.00010956 | 5.62437744 |
| Hypericum | annual | 0.07730112 | 0.42836443 | 0.00043481 | 5.62264575 |
| Hypericum | perennial | 0.01801538 | 0.42836443 | 0.00040052 | 5.62313186 |
| Lepidieae | annual | 0.25911108 | 0.13251048 | 0.00050075 | 5.61405362 |
| Lepidieae | perennial | 0.21785597 | 0.13251048 | 0.0003173 | 5.61464612 |
| Lupinus | annual | 0.04801391 | 0.00037044 | 1.77E-05 | 5.61983629 |
| Lupinus | perennial | 0.09070246 | 0.00037044 | 0.00019399 | 5.61983468 |
| Lysimachieae | annual | 0.03158252 | 0.00677689 | 2.49E-05 | 5.61532842 |
| Lysimachieae | perennial | 0.02977141 | 0.00677689 | 6.70E-05 | 5.61046734 |
| Onagraceae | annual | 0.03706599 | 1.35863169 | 0.00105959 | 5.61348578 |
| Onagraceae | perennial | 0.00323669 | 1.35863169 | 0.00257146 | 5.6142152 |
| Orobanchaceae | annual | 0.04676778 | 0.50912483 | 0.00084406 | 5.60796797 |
| Orobanchaceae | perennial | 0.04291609 | 0.50912483 | 0.00090528 | 5.60230897 |
| Panicoideae | annual | 0.02590142 | 7.02397466 | 0.00598784 | 5.65319941 |
| Panicoideae | perennial | 0.03932365 | 7.02397466 | 0.00600914 | 5.65127263 |
| Polemoniaceae | annual | 0.01191555 | 0.05554799 | 3.95E-05 | 5.60799034 |
| Polemoniaceae | perennial | 0.0085158 | 0.05554799 | 0.00013439 | 5.6079908 |
| Pooideae | annual | 0.03005168 | 7.02338554 | 0.00461929 | 5.61197783 |
| Pooideae | perennial | 0.02806981 | 7.02338554 | 0.01812177 | 5.58654015 |
| Primulaceae | annual | 0.0225558 | 1.4506934 | 0.00218386 | 5.56845748 |
| Primulaceae | perennial | 0.01196071 | 1.4506934 | 0.0021197 | 5.57852618 |
| Rubieae | annual | 0.03901116 | 9.60836014 | 0.00345269 | 5.61800702 |
| Rubieae | perennial | 0.03892015 | 9.60836014 | 0.01046649 | 5.60194233 |
| Salvia | annual | 0.01855879 | 0.08099552 | 7.45E-05 | 5.62297199 |
| Salvia | perennial | 0.01548408 | 0.08099552 | 7.45E-05 | 5.62271121 |
| Solanaceae | annual | 0.03769563 | 0.10215951 | 0.00014587 | 5.63775801 |
| Solanaceae | perennial | 0.01237181 | 0.10215951 | 0.00014587 | 5.64190815 |
| Spermacoceae | annual | 0.03534535 | 0.13736101 | 0.00015063 | 5.66189236 |
| Spermacoceae | perennial | 0.01316777 | 0.13736101 | 0.00012331 | 5.64154072 |
| Thelypodieae | annual | 0.04237487 | 1.00925814 | 0.00051586 | 5.61278281 |
| Thelypodieae | perennial | 0.07564032 | 1.00925814 | 0.00030275 | 5.59905695 |

**Table S5.** Parameter estimates from the model averaged hOUwie fits for BIO12

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.04116941 | 0.19036195 | 0.04527165 | 6.24275828 |
| Alysseae | perennial | 0.02025504 | 0.19036195 | 0.03815442 | 6.39814099 |
| Antirrhineae | annual | 0.02919324 | 0.05880125 | 0.04210127 | 6.05373627 |
| Antirrhineae | perennial | 0.01269128 | 0.05880125 | 0.05850862 | 6.18955257 |
| Apioideae | annual | 0.01602982 | 0.11423193 | 0.07621422 | 6.19452692 |
| Apioideae | perennial | 0.0221401 | 0.11423193 | 0.07621422 | 6.46129814 |
| Arabideae | annual | 0.10444553 | 7.3119644 | 4.34841792 | 6.42520345 |
| Arabideae | perennial | 0.14178575 | 7.3119644 | 4.34841442 | 6.43388472 |
| Balsamiaceae | annual | 0.01417992 | 0.01179961 | 0.00842918 | 6.43690021 |
| Balsamiaceae | perennial | 0.01378934 | 0.01179961 | 0.00791053 | 8.34905015 |
| Brassiceae | annual | 0.04707618 | 1.87664502 | 1.39611112 | 6.21384524 |
| Brassiceae | perennial | 0.01834658 | 1.87664502 | 1.38690848 | 6.01805817 |
| Cardamineae | annual | 0.03827605 | 1.29718208 | 0.48138907 | 6.76727827 |
| Cardamineae | perennial | 0.03585023 | 1.29718208 | 0.67981705 | 6.74828047 |
| Cardueae | annual | 0.06688424 | 0.09082574 | 0.02928048 | 6.31051667 |
| Cardueae | perennial | 0.01642933 | 0.09082574 | 0.06484279 | 6.31288359 |
| CES | annual | 0.07955082 | 0.10451031 | 0.15538542 | 5.33390817 |
| CES | perennial | 0.03046781 | 0.10451031 | 0.10954848 | 5.79888036 |
| Chamaecrista | annual | 0.08373911 | 0.05367876 | 0.00196485 | 7.34630367 |
| Chamaecrista | perennial | 0.01153733 | 0.05367876 | 0.01038718 | 7.10679122 |
| Chorisporeae | annual | 0.08350274 | 3.22152374 | 0.79087149 | 5.41609029 |
| Chorisporeae | perennial | 0.09160367 | 3.22152374 | 0.95533864 | 5.65707773 |
| Croton | annual | 0.03687071 | 0.08882436 | 0.25867727 | 6.76402145 |
| Croton | perennial | 0.0058538 | 0.08882436 | 0.03636092 | 7.16056622 |
| Erysimeae | annual | 0.0435169 | 0.62594525 | 0.14738825 | 6.12313936 |
| Erysimeae | perennial | 0.08007439 | 0.62594525 | 0.15498314 | 6.28319991 |
| Euclidieae | annual | 0.03156694 | 0.17283082 | 0.02370745 | 5.39913121 |
| Euclidieae | perennial | 0.04644634 | 0.17283082 | 0.25837963 | 5.45281885 |
| Eumalvoideae | annual | 0.02813053 | 0.05236832 | 0.17352251 | 5.17245971 |
| Eumalvoideae | perennial | 0.01664049 | 0.05236832 | 0.04625187 | 6.60057645 |
| Gesneriaceae | annual | 0.00514178 | 0.02075967 | 0.00419459 | 7.64338703 |
| Gesneriaceae | perennial | 0.00259023 | 0.02075967 | 0.00696187 | 7.37442153 |
| Grewioideae | annual | 0.01851069 | 0.01519887 | 0.01079399 | 6.63883528 |
| Grewioideae | perennial | 0.00204155 | 0.01519887 | 0.01047099 | 7.22264512 |
| Heliophileae | annual | 0.03148566 | 0.18267111 | 0.08926383 | 5.50873313 |
| Heliophileae | perennial | 0.01997662 | 0.18267111 | 0.07281015 | 6.20895515 |
| Hypericum | annual | 0.08474611 | 0.68310322 | 0.17384484 | 7.14698412 |
| Hypericum | perennial | 0.01717609 | 0.68310322 | 0.05788547 | 7.21427088 |
| Lepidieae | annual | 0.22678721 | 4.09428613 | 1.62184764 | 5.98357632 |
| Lepidieae | perennial | 0.21022811 | 4.09428613 | 1.61479762 | 5.97607088 |
| Lupinus | annual | 0.03000707 | 0.03296219 | 0.0470141 | 6.32075538 |
| Lupinus | perennial | 0.07829517 | 0.03296219 | 0.10536494 | 8.20029689 |
| Lysimachieae | annual | 0.03603358 | 0.13033183 | 0.02691997 | 6.9317815 |
| Lysimachieae | perennial | 0.02863371 | 0.13033183 | 0.03122399 | 6.86290627 |
| Onagraceae | annual | 0.0385471 | 0.12746318 | 0.16128676 | 6.10071335 |
| Onagraceae | perennial | 0.00435933 | 0.12746318 | 0.09164123 | 6.8609902 |
| Orobanchaceae | annual | 0.04836983 | 0.04741257 | 0.03653261 | 6.63191087 |
| Orobanchaceae | perennial | 0.04104038 | 0.04741257 | 0.04334687 | 6.44066335 |
| Panicoideae | annual | 0.02649464 | 0.53361681 | 0.26292268 | 6.95358999 |
| Panicoideae | perennial | 0.03891246 | 0.53361681 | 0.26292275 | 6.99399853 |
| Polemoniaceae | annual | 0.01107295 | 0.02045512 | 0.02278383 | 4.33439272 |
| Polemoniaceae | perennial | 0.00787351 | 0.02045512 | 0.02486642 | 6.76427556 |
| Pooideae | annual | 0.0188484 | 8.0067692 | 4.67402096 | 6.28878834 |
| Pooideae | perennial | 0.02946306 | 8.0067692 | 4.68880261 | 6.42808007 |
| Primulaceae | annual | 0.02361879 | 0.38999068 | 0.18046481 | 6.39076673 |
| Primulaceae | perennial | 0.01149745 | 0.38999068 | 0.19282366 | 6.55496722 |
| Rubieae | annual | 0.03459232 | 0.75828112 | 0.04039053 | 6.40171632 |
| Rubieae | perennial | 0.04682316 | 0.75828112 | 0.2410869 | 6.61440928 |
| Salvia | annual | 0.01898019 | 0.05582403 | 0.061728 | 5.3249917 |
| Salvia | perennial | 0.015896 | 0.05582403 | 0.06167412 | 6.36815166 |
| Solanaceae | annual | 0.01401341 | 0.06630001 | 0.15851054 | 5.99752213 |
| Solanaceae | perennial | 0.01109792 | 0.06630001 | 0.15154511 | 6.04384296 |
| Spermacoceae | annual | 0.03530102 | 0.05217572 | 0.02525213 | 7.09950546 |
| Spermacoceae | perennial | 0.01211311 | 0.05217572 | 0.039049 | 6.86151135 |
| Thelypodieae | annual | 0.04425106 | 0.17905675 | 0.21687839 | 6.16657138 |
| Thelypodieae | perennial | 0.07060306 | 0.17905675 | 0.20996241 | 5.93483687 |

**Table S6.** Parameter estimates from the model averaged hOUwie fits for BIO14

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.04128678 | 0.39587375 | 0.4441553 | 2.35840124 |
| Alysseae | perennial | 0.02371718 | 0.39587375 | 0.44781198 | 2.48680676 |
| Antirrhineae | annual | 0.03145354 | 0.25392107 | 0.51413319 | 1.70130112 |
| Antirrhineae | perennial | 0.01421217 | 0.25392107 | 0.51806035 | 2.14218543 |
| Apioideae | annual | 0.01524207 | 0.12422328 | 0.31137455 | 2.3450107 |
| Apioideae | perennial | 0.02096277 | 0.12422328 | 0.31137456 | 2.28645066 |
| Arabideae | annual | 0.1246393 | 2.50038113 | 4.36494717 | 2.6735432 |
| Arabideae | perennial | 0.14813544 | 2.50038113 | 4.34975364 | 2.69028869 |
| Balsamiaceae | annual | 0.01508213 | 0.07740779 | 0.14151924 | 2.53885605 |
| Balsamiaceae | perennial | 0.01425172 | 0.07740779 | 0.13577727 | 3.25555217 |
| Brassiceae | annual | 0.04329255 | 0.24011376 | 0.46438134 | 2.07425051 |
| Brassiceae | perennial | 0.02385256 | 0.24011376 | 0.47115288 | 1.8322232 |
| Cardamineae | annual | 0.04647318 | 2.30525389 | 1.42219497 | 3.28406468 |
| Cardamineae | perennial | 0.02832631 | 2.30525389 | 4.60777535 | 3.28371196 |
| Cardueae | annual | 0.07434588 | 0.13324756 | 0.26512498 | 2.18891674 |
| Cardueae | perennial | 0.01256012 | 0.13324756 | 0.2651275 | 2.21302382 |
| CES | annual | 0.19864708 | 0.2023053 | 0.25726829 | 2.21535711 |
| CES | perennial | 0.03586053 | 0.2023053 | 0.26335649 | 2.12561137 |
| Chamaecrista | annual | 0.09928236 | 0.07370525 | 0.06941213 | 2.72743857 |
| Chamaecrista | perennial | 0.01049807 | 0.07370525 | 0.10452338 | 2.75579867 |
| Chorisporeae | annual | 0.08930619 | 5.04648042 | 4.84786201 | 1.85570731 |
| Chorisporeae | perennial | 0.09086837 | 5.04648042 | 4.90602539 | 2.01469117 |
| Croton | annual | 0.02965713 | 0.13331075 | 0.26503343 | 3.11749788 |
| Croton | perennial | 0.00576384 | 0.13331075 | 0.26503813 | 3.00631608 |
| Erysimeae | annual | 0.05617456 | 0.90560808 | 1.63570551 | 2.34096316 |
| Erysimeae | perennial | 0.05681704 | 0.90560808 | 1.64801392 | 2.07878073 |
| Euclidieae | annual | 0.03052756 | 0.17899059 | 0.15139652 | 1.1490183 |
| Euclidieae | perennial | 0.04908825 | 0.17899059 | 0.41589176 | 1.45295695 |
| Eumalvoideae | annual | 0.03055092 | 0.06645048 | 0.13627013 | 1.23769843 |
| Eumalvoideae | perennial | 0.01301408 | 0.06645048 | 0.13627068 | 2.19076227 |
| Gesneriaceae | annual | 0.00960378 | 0.02763987 | 0.05832076 | 3.34129952 |
| Gesneriaceae | perennial | 0.00248544 | 0.02763987 | 0.05812546 | 3.33565852 |
| Grewioideae | annual | 0.02246705 | 0.05825947 | 0.11657273 | 2.0061262 |
| Grewioideae | perennial | 0.00299285 | 0.05825947 | 0.11656734 | 2.98738402 |
| Heliophileae | annual | 0.03505583 | 0.04611129 | 0.05166065 | 1.97158873 |
| Heliophileae | perennial | 0.01952112 | 0.04611129 | 0.04293603 | 3.03400287 |
| Hypericum | annual | 0.09775582 | 0.70354441 | 0.88301032 | 3.30162956 |
| Hypericum | perennial | 0.01262613 | 0.70354441 | 0.0737134 | 4.08757487 |
| Lepidieae | annual | 0.24148401 | 1.4627211 | 2.84283768 | 2.08167002 |
| Lepidieae | perennial | 0.20250524 | 1.4627211 | 2.46772765 | 2.06101929 |
| Lupinus | annual | 0.0299301 | 0.00045008 | 0.11232501 | 2.04600881 |
| Lupinus | perennial | 0.06609475 | 0.00045008 | 0.43654883 | 2.04027876 |
| Lysimachieae | annual | 0.04000742 | 0.32148745 | 0.33002989 | 2.79958746 |
| Lysimachieae | perennial | 0.02904615 | 0.32148745 | 0.33010595 | 3.00227727 |
| Onagraceae | annual | 1.47718469 | 0.26326998 | 0.45818026 | 1.94290395 |
| Onagraceae | perennial | 1.66286254 | 0.26326998 | 0.69082882 | 2.91186965 |
| Orobanchaceae | annual | 0.04082546 | 0.9540879 | 2.15192758 | 2.57300925 |
| Orobanchaceae | perennial | 0.04470213 | 0.9540879 | 2.21659985 | 2.57171532 |
| Panicoideae | annual | 0.01888656 | 1.29897197 | 2.23586442 | 2.28973178 |
| Panicoideae | perennial | 0.04030941 | 1.29897197 | 2.23586442 | 2.7961387 |
| Polemoniaceae | annual | 0.0132323 | 0.07515773 | 0.10438561 | 1.48991147 |
| Polemoniaceae | perennial | 0.01012157 | 0.07515773 | 0.15057698 | 2.58427997 |
| Pooideae | annual | 0.02001605 | 0.93384385 | 1.81217652 | 2.18967575 |
| Pooideae | perennial | 0.02696252 | 0.93384385 | 1.8121767 | 2.76232292 |
| Primulaceae | annual | 0.02057645 | 0.17474529 | 0.46749182 | 2.20033062 |
| Primulaceae | perennial | 0.01156039 | 0.17474529 | 0.46749183 | 2.24246698 |
| Rubieae | annual | 0.11367892 | 1.244906 | 2.31498746 | 2.20696365 |
| Rubieae | perennial | 0.03310733 | 1.244906 | 2.37652947 | 2.70892502 |
| Salvia | annual | 0.01870823 | 0.07861361 | 0.16211643 | 1.79452442 |
| Salvia | perennial | 0.0155215 | 0.07861361 | 0.162119 | 2.04769413 |
| Solanaceae | annual | 0.01715148 | 0.17187496 | 0.34899588 | 1.74133921 |
| Solanaceae | perennial | 0.01330192 | 0.17187496 | 0.34899592 | 2.87471849 |
| Spermacoceae | annual | 0.03407916 | 0.03791588 | 0.10042255 | 2.40158345 |
| Spermacoceae | perennial | 0.01287896 | 0.03791588 | 0.10057995 | 2.52361849 |
| Thelypodieae | annual | 0.05653152 | 0.02218699 | 0.19335883 | 1.40802748 |
| Thelypodieae | perennial | 0.05639596 | 0.02218699 | 0.06639772 | 0.99389082 |

Table S7. Parameter estimates from the model averaged hOUwie fits for BIO15

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.04155405 | 6.76094614 | 1.38224676 | 3.79661529 |
| Alysseae | perennial | 0.02486197 | 6.76094614 | 1.37899295 | 3.79316364 |
| Antirrhineae | annual | 0.02871898 | 0.18997829 | 0.0699409 | 3.93790739 |
| Antirrhineae | perennial | 0.0121807 | 0.18997829 | 0.07433216 | 3.95943992 |
| Apioideae | annual | 0.05266043 | 0.09778867 | 0.04944416 | 3.85233249 |
| Apioideae | perennial | 0.00934632 | 0.09778867 | 0.05590919 | 3.89088209 |
| Arabideae | annual | 0.13736519 | 5.98079291 | 2.28870901 | 3.79356842 |
| Arabideae | perennial | 0.13691382 | 5.98079291 | 2.18669782 | 3.82285345 |
| Balsamiaceae | annual | 0.01423208 | 0.04892724 | 0.01477393 | 4.13217229 |
| Balsamiaceae | perennial | 0.01342129 | 0.04892724 | 0.01544012 | 4.06214418 |
| Brassiceae | annual | 0.0452066 | 0.78788956 | 0.41823176 | 3.83045967 |
| Brassiceae | perennial | 0.01679124 | 0.78788956 | 0.34660596 | 3.86681622 |
| Cardamineae | annual | 0.05609979 | 0.34653582 | 0.08432257 | 3.53490268 |
| Cardamineae | perennial | 0.03290568 | 0.34653582 | 0.22058434 | 3.60329065 |
| Cardueae | annual | 0.05763683 | 0.2244015 | 0.05683454 | 3.8804587 |
| Cardueae | perennial | 0.03353348 | 0.2244015 | 0.05817862 | 3.90034307 |
| CES | annual | 0.07679611 | 0.00798582 | 0.00979194 | 4.06059801 |
| CES | perennial | 0.03143227 | 0.00798582 | 0.02070469 | 3.9313969 |
| Chamaecrista | annual | 0.0884329 | 0.00586828 | 0.00622706 | 4.14633571 |
| Chamaecrista | perennial | 0.01161986 | 0.00586828 | 0.00543901 | 3.90529362 |
| Chorisporeae | annual | 0.06861689 | 0.49866028 | 0.1389933 | 4.02607213 |
| Chorisporeae | perennial | 0.08217822 | 0.49866028 | 0.17045693 | 3.98113444 |
| Croton | annual | 0.01827458 | 0.18210444 | 0.05639281 | 3.41900189 |
| Croton | perennial | 0.00489446 | 0.18210444 | 0.05626612 | 4.06993011 |
| Erysimeae | annual | 0.05877242 | 0.51344609 | 0.24097373 | 3.77932462 |
| Erysimeae | perennial | 0.06446853 | 0.51344609 | 0.15767626 | 4.04711484 |
| Euclidieae | annual | 0.03326008 | 0.24037128 | 0.08779928 | 4.14366647 |
| Euclidieae | perennial | 0.04482494 | 0.24037128 | 0.13181966 | 4.12335769 |
| Eumalvoideae | annual | 0.03993844 | 0.0764752 | 0.00729255 | 4.22932158 |
| Eumalvoideae | perennial | 0.01353152 | 0.0764752 | 0.02191895 | 4.21162143 |
| Gesneriaceae | annual | 0.00385405 | 0.04958787 | 0.01776767 | 4.00122058 |
| Gesneriaceae | perennial | 0.00260615 | 0.04958787 | 0.01772866 | 4.00082011 |
| Grewioideae | annual | 0.02103246 | 0.07945596 | 0.00708682 | 4.39063068 |
| Grewioideae | perennial | 0.00164885 | 0.07945596 | 0.02138871 | 4.01319547 |
| Heliophileae | annual | 0.03635766 | 0.23313833 | 0.01802255 | 3.82037234 |
| Heliophileae | perennial | 0.02558159 | 0.23313833 | 0.05086611 | 3.80351134 |
| Hypericum | annual | 0.0993899 | 0.29426783 | 0.17792348 | 3.41088445 |
| Hypericum | perennial | 0.01401524 | 0.29426783 | 0.05157863 | 3.23368566 |
| Lepidieae | annual | 0.22685506 | 1.81377752 | 0.84738575 | 3.86200635 |
| Lepidieae | perennial | 0.20373133 | 1.81377752 | 0.8292531 | 3.82621227 |
| Lupinus | annual | 0.02919639 | 0.03739374 | 0.02477043 | 3.9796499 |
| Lupinus | perennial | 0.07787184 | 0.03739374 | 0.1085328 | 3.24632011 |
| Lysimachieae | annual | 0.03717362 | 0.11835388 | 0.00848736 | 4.00125334 |
| Lysimachieae | perennial | 0.0281207 | 0.11835388 | 0.0507377 | 3.82533477 |
| Onagraceae | annual | 0.03837558 | 0.66114535 | 0.43469802 | 4.00959811 |
| Onagraceae | perennial | 0.0045882 | 0.66114535 | 0.43503327 | 3.81326747 |
| Orobanchaceae | annual | 0.04441356 | 0.36790357 | 0.19490971 | 3.83165369 |
| Orobanchaceae | perennial | 0.04145397 | 0.36790357 | 0.19944175 | 3.88496486 |
| Panicoideae | annual | 0.02800291 | 9.99992238 | 3.4084159 | 4.18741051 |
| Panicoideae | perennial | 0.04026629 | 9.99992238 | 3.408369 | 3.99529444 |
| Polemoniaceae | annual | 0.0115838 | 0.07434326 | 0.01963483 | 3.97675337 |
| Polemoniaceae | perennial | 0.00885962 | 0.07434326 | 0.0531946 | 3.97746893 |
| Pooideae | annual | 0.02522916 | 8.71583652 | 4.21029626 | 3.84098247 |
| Pooideae | perennial | 0.0311285 | 8.71583652 | 4.21029627 | 3.66684044 |
| Primulaceae | annual | 0.02314268 | 0.12321535 | 0.06162424 | 4.05960667 |
| Primulaceae | perennial | 0.01178813 | 0.12321535 | 0.06479303 | 4.07902527 |
| Rubieae | annual | 0.10892769 | 0.78391979 | 0.25530172 | 3.85143834 |
| Rubieae | perennial | 0.02367202 | 0.78391979 | 0.33716985 | 3.82713992 |
| Salvia | annual | 0.01700832 | 0.09567735 | 0.01553976 | 4.19909233 |
| Salvia | perennial | 0.0150281 | 0.09567735 | 0.03452563 | 4.1760563 |
| Solanaceae | annual | 0.03925798 | 0.20868257 | 0.08293553 | 4.12002722 |
| Solanaceae | perennial | 0.01441309 | 0.20868257 | 0.08113076 | 4.0231765 |
| Spermacoceae | annual | 0.03412617 | 0.05841163 | 0.02295724 | 4.22272915 |
| Spermacoceae | perennial | 0.01240046 | 0.05841163 | 0.02422034 | 4.11327556 |
| Thelypodieae | annual | 0.04313355 | 0.03939268 | 0.0394036 | 4.09556299 |
| Thelypodieae | perennial | 0.07203443 | 0.03939268 | 0.0375932 | 4.12090552 |

**Table S8.** Parameter estimates from the model averaged hOUwie fits for AI

| clade | Group.1 | rate | alpha | sigma.sq | theta |
|---|---|---|---|---|---|
| Alysseae | annual | 0.04030233 | 0.34853482 | 0.16651319 | 8.23596117 |
| Alysseae | perennial | 0.02424038 | 0.34853482 | 0.20031482 | 8.33323375 |
| Antirrhineae | annual | 0.02992664 | 0.05268783 | 0.0665517 | 7.83521596 |
| Antirrhineae | perennial | 0.01289018 | 0.05268783 | 0.08899283 | 7.91692405 |
| Apioideae | annual | 0.02107216 | 0.07106027 | 0.0799759 | 8.13484316 |
| Apioideae | perennial | 0.01946203 | 0.07106027 | 0.0799759 | 8.67364721 |
| Arabideae | annual | 0.09698462 | 10 | 5.48687265 | 8.71047719 |
| Arabideae | perennial | 0.13109321 | 10 | 5.48687265 | 8.82531687 |
| Balsamiaceae | annual | 0.01416107 | 4.68829094 | 1.15412892 | 9.12859414 |
| Balsamiaceae | perennial | 0.01232552 | 4.68829094 | 1.15506067 | 9.45715807 |
| Brassiceae | annual | 0.04485486 | 2.05732314 | 2.67691747 | 8.1041908 |
| Brassiceae | perennial | 0.01930118 | 2.05732314 | 2.66802778 | 7.91405884 |
| Cardamineae | annual | 0.05377465 | 6.8854024 | 2.23948173 | 8.98243888 |
| Cardamineae | perennial | 0.03299381 | 6.8854024 | 2.23158769 | 9.13251343 |
| Cardueae | annual | 0.05541007 | 0.13570678 | 0.10129996 | 8.147476 |
| Cardueae | perennial | 0.02616704 | 0.13570678 | 0.1178579 | 8.50145217 |
| CES | annual | 0.0852525 | 0.25268969 | 0.4753827 | 7.50996351 |
| CES | perennial | 0.03088316 | 0.25268969 | 0.13867862 | 7.91360517 |
| Chamaecrista | annual | 0.08469111 | 0.00953602 | 0.00697957 | 9.03197266 |
| Chamaecrista | perennial | 0.01153876 | 0.00953602 | 0.00673797 | 9.03932254 |
| Chorisporeae | annual | 0.08161394 | 0.2744213 | 0.01678788 | 7.04724848 |
| Chorisporeae | perennial | 0.10328263 | 0.2744213 | 0.13663049 | 8.83609909 |
| Croton | annual | 0.03415848 | 0.0692009 | 0.41526537 | 8.93385665 |
| Croton | perennial | 0.00640381 | 0.0692009 | 0.03627557 | 8.9338686 |
| Erysimeae | annual | 0.06030963 | 0.71632778 | 0.35194845 | 8.22615218 |
| Erysimeae | perennial | 0.06051128 | 0.71632778 | 0.34481591 | 8.27159025 |
| Euclidieae | annual | 0.03019637 | 0.50928631 | 0.17665409 | 7.25212009 |
| Euclidieae | perennial | 0.05275321 | 0.50928631 | 0.48401792 | 8.24217854 |
| Eumalvoideae | annual | 0.0409016 | 0.04877877 | 0.09335003 | 7.18902524 |
| Eumalvoideae | perennial | 0.01614903 | 0.04877877 | 0.09015431 | 8.28117526 |
| Gesneriaceae | annual | 0.00387631 | 0.04328585 | 0.01111532 | 9.43186045 |
| Gesneriaceae | perennial | 0.00229391 | 0.04328585 | 0.01121134 | 9.3856122 |
| Grewioideae | annual | 0.01734095 | 0.02189404 | 0.02827933 | 8.30361448 |
| Grewioideae | perennial | 0.00186462 | 0.02189404 | 0.0181696 | 9.04959079 |
| Heliophileae | annual | 0.06634313 | 0.13605649 | 0.08969506 | 6.97503679 |
| Heliophileae | perennial | 0.10404222 | 0.13605649 | 0.06252826 | 8.39869433 |
| Hypericum | annual | 0.08162554 | 0.59004273 | 0.35475392 | 9.12085686 |
| Hypericum | perennial | 0.01643431 | 0.59004273 | 0.02609249 | 9.16897766 |
| Lepidieae | annual | 0.24377528 | 9.48539121 | 7.81624215 | 7.77321372 |
| Lepidieae | perennial | 0.20901694 | 9.48539121 | 7.77153297 | 7.77984526 |
| Lupinus | annual | 0.02805195 | 0.0327964 | 0.08562482 | 8.1163569 |
| Lupinus | perennial | 0.07866818 | 0.0327964 | 0.11880514 | 13.6381535 |
| Lysimachieae | annual | 0.03774993 | 0.11544436 | 0.02712105 | 8.88684662 |
| Lysimachieae | perennial | 0.02849096 | 0.11544436 | 0.02389185 | 8.99585704 |
| Onagraceae | annual | 0.03871019 | 0.10943066 | 0.27245281 | 7.89504685 |
| Onagraceae | perennial | 0.00416502 | 0.10943066 | 0.07267066 | 8.85193465 |
| Orobanchaceae | annual | 0.04398991 | 0.15740082 | 0.1580424 | 8.5411767 |
| Orobanchaceae | perennial | 0.04174612 | 0.15740082 | 0.15936042 | 8.48590867 |
| Panicoideae | annual | 0.02815515 | 0.81789315 | 0.65324868 | 8.66344107 |
| Panicoideae | perennial | 0.03949567 | 0.81789315 | 0.65324854 | 8.78505209 |
| Polemoniaceae | annual | 0.01167735 | 0.03361776 | 0.03668367 | 7.38397126 |
| Polemoniaceae | perennial | 0.00838739 | 0.03361776 | 0.03667393 | 9.26525933 |
| Pooideae | annual | 0.0270249 | 2.41066525 | 1.94154524 | 8.26179387 |
| Pooideae | perennial | 0.03013477 | 2.41066525 | 1.94154524 | 8.69603936 |
| Primulaceae | annual | 0.0251974 | 0.2483582 | 0.11808828 | 8.75717914 |
| Primulaceae | perennial | 0.01151704 | 0.2483582 | 0.15285536 | 8.85551769 |
| Rubieae | annual | 0.04169511 | 1.18607452 | 0.15363439 | 8.35687711 |
| Rubieae | perennial | 0.04908283 | 1.18607452 | 0.55882939 | 8.7696908 |
| Salvia | annual | 0.01577602 | 0.05665405 | 0.07639197 | 7.61132764 |
| Salvia | perennial | 0.015253 | 0.05665405 | 0.07636625 | 8.19096074 |
| Solanaceae | annual | 0.02411958 | 0.12608527 | 0.31342469 | 7.88138697 |
| Solanaceae | perennial | 0.014473 | 0.12608527 | 0.23621796 | 8.45916171 |
| Spermacoceae | annual | 0.03479382 | 0.03536571 | 0.03113017 | 8.84366586 |
| Spermacoceae | perennial | 0.01199718 | 0.03536571 | 0.03419885 | 8.69948668 |
| Thelypodieae | annual | 0.04566576 | 0.02670115 | 0.10177345 | 7.87572838 |
| Thelypodieae | perennial | 0.07724747 | 0.02670115 | 0.09840788 | 7.48079913 |

# CONCLUSION

*"I had said at the very end of the paper that if someone insisted on doing comparative methods without using phylogenies, that it might be more useful if they took up selling real estate."*

- *Joe Felsenstein (1985 & 2019)*

## Summary

In Chapter I, I generalized hidden Markov models of discrete character evolution to allow for any number of characters, observed states, or hidden states. This project, though not requiring any new mathematical contributions for the character evolution aspect, set the groundwork for several collaborations and deepened my understanding of discrete character evolution. I also demonstrated some of the informational advantages of allowing for additional characters in states when conducting ancestral state reconstructions. To do this, I developed a novel metric of information as probabilities pass towards the root of the phylogeny.

Next, in Chapter II, I applied some of the knowledge gained in Chapter I and found a way to help correct bias in correlation analyses between discrete characters. The method was analogous to that of Beaulieu and O'Meara (2016) and utilized hidden Markov models to allow for character independent rate heterogeneity. I found that part of the reason for biased correlation was that when a signal of dependence is tested at a macroevolutionary level, it is actually a test of rate heterogeneity tied to a focal character. By allowing for rate heterogeneity independent of a focal character, the signal for rate heterogeneity can be "removed" from the comparison leaving the model set to evaluate whether there is a signal of correlation.

In Chapter III, I develop a new model for linking discrete and continuous character evolution called hOUwie. This model is developed with the issues of the previous chapter in mind and therefore allows for character independent rate heterogeneity as an alternative explanation. This project required the implementation of several previously published algorithms (Ho and Ané 2014) as well as the development of novel heuristics.

Finally, in Chapter IV, I applied the model developed in chapter 3 to 33 clades of angiosperms and attempt to understand the evolutionary patterns of plant life history in relation to climate. We demonstrated that a consistent driver of annual life history is the maximum temperature of the warmest month. Furthermore, we showed that some of the most commonly used model systems for life history evolution have biased our perceptions of the evolutionary process and in our analyses were more often the exception than the rule. Interestingly, this chapter also demonstrated most clearly the importance of joint modeling through the highly variable ancestral state reconstructions. The joint parameter estimation could massively change which state an ancestral state was reconstructed as depending on the particular climatic variable associated.

**Future Work**

I now turn to ideas of future work by discussing several extensions of this dissertation. Based on my work in Chapter I and Chapter II, I think a program which automatically searches model space (the alternative ways that we can constrain or estimate our parameters within a model) could be quite useful and interesting. Most of the discussion when it comes to modeling has been focussed on finding the maximum likelihood parameter estimates for a fixed model structure and dataset. However, model structures can vary by dropping and equating parameters. A model with a parameter dropped or two parameters equated can make vastly different assumptions about the

evolutionary process. In Chapter II, I discussed the differences between independent and correlated models as being differences in parameter constraints. The closest practice to searching model space is multimodel inference (Burnham and Anderson 2002), but this relies on biologists choosing a set of models which are reasonable for the question. As we saw in Chapter II, even biologists who are extremely knowledgeable in models of discrete character evolution can choose model sets which introduce a severe bias in their results. My contention is that choosing a biologically and statistically reasonable set of models is easier said than done. Besides that, I am unconvinced that we do not want the best fit to the data given appropriate corrections for overfitting the model. My goal with this program would be to search discrete model space and find the model which best fits the data according to AIC (Akaike 1998). There are two immediate challenges to this proposed extension. First, model space is vast. Searching this space exhaustively would be computationally infeasible without some clever heuristic algorithm. As part of this first challenge I include redundancy as a major issue (Cover and Thomas 1991). A more complex model estimated with two independently estimated parameters numerically being equal may have an identical likelihood to a model in the set where those two parameters were not freely estimated and instead constrained to be equal. This is likely to occur and even desirable if the data better supports these simpler models. However, I worry that the frequency that certain model structures appear may be biased in some way. The second challenge is model interpretation. Once the program had dredged through all possible model structures, we are left with an issue of interpretation. In the current paradigm, we fit a set of exclusive hypotheses that have their reasoning in the biology of the system. That means each model structure can be interpreted as a particular hypothesis and the parameters may be discussed in that context. If we were to fit all models at once, then we may have no direct link to any biological hypotheses. This

would mean relying on unbiased interpretations of the parameter estimates, model structures, and developing tools that make it easier to understand the parameters in the context of biology.

I would like to develop a fundamentally different model of macroevolution not reliant on Markovian or Gaussian dynamics, but it is not clear what this kind of model would look like. For example, one issue with a dredge-like extension of corHMM is that it may, or may not, deal with the single unreplicated events described by Maddison and FitzJohn (2015). Although I am confident in the source of biased correlation is an inappropriate model set, the fact remains that evolution may not be Markovian at all. Many events in evolution are singular and unrepeatable. In contrast, Markov models rely on the repeated occurrence of events. And although there are many datasets that show repeated transitions between discrete states, who is to say that we are calling two things identical when, in fact, they have an entirely different set of causes and underlying genetic architectures. To say that the same transitions (X0 to X1) occurred at the same rate in different parts of the tree is quite the assumption to make. Even with hidden Markov models, we are limited in the number of ways that this transition can occur and will always constrained by Markovian dynamics. One issue with this paradigm, and a potential avenue for future work, is that the amount of time spent in a particular state has no influence on whether a transition is likely to occur. Markov models are memoryless processes which means that their current state will entirely determine their next state (Cover and Thomas 1991; Ephraim and Merhav 2002; Sober and Steel 2011; Gascuel and Steel 2018; Goldberg and Foo 2020; Boyko and Beaulieu 2021). We know that there are evolutionary lags and that genetics can be canalized such that the longer spent with a particular genetic architecture, the less likely you are to transition (Waddington 1953). If Markovian processes are truly inappropriate for evolution, then no amount of model structure searching will allow for an accurate depiction of evolution. There

are some options available for memoried processes in evolutionary models (Goldberg and Foo 2020), but there are likely to be identifiability issues with these models similar to those discussed in Louca and Pennell (2020). In my own exploration of these processes, I have found that many of these memoried models are applied to time series data (Zucchini et al. 2017), and although the implementation may be straightforward, their practicality is not as obvious.

There are several places hOUwie (Chapter III) could be improved and extended. The idea behind the model is important because we know things do not evolve independently (Levins and Lewontin 1985). Furthermore, I trust that our approximations are good and that the likelihood itself is correct since it operates using the same probabilities published previously (Beaulieu et al. 2012; Ho and Ané 2014). But the program can still be computationally costly even with the large speedups for any single iteration. It would be ideal to be able to directly calculate the likelihood without the need of simulation. One way forward would be to utilize Markov modulated OU processes to derive expected values, variances, and covariances (Galtier and Jean-Marie 2004; Huang et al. 2016; Behme and Sideris 2022). Once those quantities are derived we could either use standard matrix inversion to calculate the likelihood (Hansen 1997; Butler and King 2004) or, more excitingly, we should be able to transform phylogeny to match the appropriate variance and covariances and apply the three point algorithm of Ho and Ané (2014). This is what was done in Chapter III for rapidly calculating the continuous character probability and is now available for use in the R-package OUwie. The greatest difficulty in this approach would be deriving the moments of the Markov modulated Ornstein-Uhlenbeck process, but this is fodder for future collaboration.

Another potential improvement to jointly modeling discrete and continuous character evolution would come from the continuous value directly influencing the probability of the

discrete transition. Currently, the joint probability calculation relies on a shared underlying regime, but it should be theoretically possible to have the specific value of the continuous character directly influence the probability of a discrete transition. It is possible that the use of a threshold model would make the statistical assessment of this model more tractable. A threshold model (Felsenstein 2005, 2012; Revell 2014; Hiscott et al. 2016) assumes that underlying a discrete phenotype is a continuously varying character. This character may be said to represent the additive genetic variation underlying phenotypic expression. The advantage of this approach would be that modeling discrete and continuous characters jointly would simply be a multivariate Gaussian distribution and could utilize the multivariate statistics of multiple continuous characters (Bartoszek et al. 2012). Finally, an extension that incorporates the probability of the phylogeny may be necessary. State-dependent speciation and extinction (Maddison et al. 2007; Beaulieu and O'Meara 2016; Louca and Pennell 2020b) models fundamentally correct problems related to biases in character distributions at tips (Maddison 2006). If the distribution of several continuous and discrete characters are biased in some way by diversification dynamics, this could lead to an overestimation of correlation. There is also work showing that OU models are effected by the structure of the phylogeny itself (Cooper et al. 2016b). A joint discrete, continuous, speciation and extinction model (tentatively the "hOUSSE" model), could help resolve these potential biases. The immediate extension of hOUwie to hOUSSE should be straightforward requiring that discrete regimes are calculated following an SSE model instead of a discrete character evolution model. But, generating high probability underlying regimes may potentially be an issue.

Finally, there are a few assumptions that we made in Chapter IV that could be improved. For instance, we lumped several classes of perennial life histories together. It is possible that the

hidden rate variation we found in some cases is failing to recognize that, for example, a shrub, geophyte, and a tree all evolve in the same way and react to environmental stresses in different ways. Furthermore, a major shortcoming (that would require an extension to the hOUwie model) was that climatic variables were modeled independent of one another. Extending the hOUwie model such that the cumulative effect of all climatic variables would be an interesting avenue to explore. The ideal approach would be to model the covariances between the different continuous characters and their influence on one another. This approach should be better able to capture and model the species climatic niche breadth. However, this would be a mathematically challenging problem and a more pragmatic approach would be to simply model several continuous characters all at once and evaluate a given joint regime probability for all the continuous characters. This would be a simple way to allow all continuous characters to influence the likelihood of the parameters at once. However, this would massively expand the model space and ignore the covariances between the continuous characters.

**Concluding Remarks**

Phylogenetic comparative methods (PCMs) have been criticized for high false positive rates (e.g., Maddison and FitzJohn 2015; Rabosky and Goldberg 2015), a lack of power (Boettiger et al. 2012; Cooper et al. 2016a), a lack of identifiability (e.g., Kubo and Iwasa 1995; Louca and Pennell 2020), and even a lack of common sense (e.g., Westoby et al. 1995). However, each newly developed method must in part justify its existence by addressing the shortcomings of previous iterations. Given that PCMs development is such an active field of research, it seems natural that criticism would be fast and frequent. For my part, I entered my PhD under the assumption that the information contained in a phylogenetic dataset could not be adequate to

know the diversification dynamics through time, or the ancestral state of lineages billions of years old. Now, I conclude the dissertation with mixed feelings. On the one hand, macroevolutionary models, like all models, are severe simplifications of reality and, because we deal with a historical science, our models are borderline untestable. On the other hand, they do exactly as advertised. The assumptions of each model are generally available for anyone interested and the model parameters do have the occasional reasonable interpretation. Furthermore, I have seen in collaborations how the best supported set of models can often be useful in resolving long standing natural history debates (the first time that happened, I was genuinely shocked). So, yes, maybe the models are all horribly wrong. Maybe any similarity between our parameters estimates and an underlying biological reality are purely by chance, or a statistical bias, or the result of an invented narrative. But, refining these methods and attempting to glimpse deeper into evolutionary history is an incredibly exciting and worthwhile prospect. It is so easy, and often necessary, to get caught up in the esoteric minutia of the methods, but it is for a fundamental understanding of nature that we do. And ultimately, there are few things more gratifying in comparative methods than when this tool can *actually* lead us to discovering something new. So, it is worth trying.

## References

Akaike H. 1998. Information Theory and an Extension of the Maximum Likelihood Principle. :15.

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. Journal of Theoretical Biology. 314:204–215.

Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling Stabilizing Selection: Expanding the Ornstein–Uhlenbeck Model of Adaptive Evolution. Evolution. 66:2369–2383.

Beaulieu J.M., O'Meara B.C. 2016. Detecting Hidden Diversification Shifts in Models of Trait-Dependent Speciation and Extinction. Syst Biol. 65:583–601.

Behme A., Sideris A. 2022. Markov-modulated generalized Ornstein-Uhlenbeck processes and an application in risk theory. Bernoulli. 28:1309–1339.

Boettiger C., Coop G., Ralph P. 2012. Is Your Phylogeny Informative? Measuring the Power of Comparative Methods. Evolution. 66:2240–2251.

Boyko J.D., Beaulieu J.M. 2021. Generalized hidden Markov models for phylogenetic comparative datasets. Methods Ecol Evol. 12:468–478.

Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Butler M.A., King A.A. 2004. Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. The American Naturalist. 164:683–695.

Cooper N., Thomas G.H., FitzJohn R.G. 2016a. Shedding light on the 'dark side' of phylogenetic comparative methods. Methods in Ecology and Evolution. 7:693–699.

Cooper N., Thomas G.H., Venditti C., Meade A., Freckleton R.P. 2016b. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biological Journal of the Linnean Society. 118:64–77.

Cover T.M., Thomas J.A. 1991. Elements of information theory. New York: Wiley.

Ephraim Y., Merhav N. 2002. Hidden Markov processes. IEEE Transactions on Information Theory. 48:1518–1569.

Felsenstein J. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philosophical Transactions of the Royal Society B: Biological Sciences. 360:1427–1434.

Felsenstein J. 2012. A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model. The American Naturalist. 179:145–156.

Galtier N., Jean-Marie A. 2004. Markov-Modulated Markov Chains and the Covarion Process of Molecular Evolution. :7.

Gascuel O., Steel M. 2018. A Darwinian Uncertainty Principle: Supplementary Information - Mathematical proofs. bioRxiv.

Goldberg E.E., Foo J. 2020. Memory in trait macroevolution. Am. Nat. 195:300–314.

Hansen T.F. 1997. Stabilizing Selection and the Comparative Analysis of Adaptation. Evolution. 51:1341–1351.

Hiscott G., Fox C., Parry M., Bryant D. 2016. Efficient Recycled Algorithms for Quantitative Trait Models on Phylogenies. Genome Biology and Evolution. 8:1338–1350.

Ho L. si, Ané C. 2014. A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. Syst Biol. 63:397–408.

Huang G., Jansen H.M., Mandjes M., Spreij P., Turck K.D. 2016. Markov-modulated Ornstein-Uhlenbeck processes. Advances in Applied Probability. 48:235–254.

Kubo T., Iwasa Y. 1995. INFERRING THE RATES OF BRANCHING AND EXTINCTION FROM MOLECULAR PHYLOGENIES. Evolution. 49:694–704.

Levins R., Lewontin R. 1985. The dialectical biologist. Harvard University Press.

Louca S., Pennell M.W. 2020a. Extant timetrees are consistent with a myriad of diversification histories. Nature. 580:502–505.

Louca S., Pennell M.W. 2020b. A General and Efficient Algorithm for the Likelihood of Diversification and Discrete-Trait Evolutionary Models. Systematic Biology. 69:545–556.

Maddison W.P. 2006. Confounding Asymmetries in Evolutionary Diversification and Character Change. Evolution. 60:1743–1746.

Maddison W.P., FitzJohn R.G. 2015. The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. Syst Biol. 64:127–136.

Maddison W.P., Midford P.E., Otto S.P., Oakley T. 2007. Estimating a Binary Character's Effect on Speciation and Extinction. Syst Biol. 56:701–710.

Rabosky D.L., Goldberg E.E. 2015. Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. Syst Biol. 64:340–355.

Revell L.J. 2014. Ancestral Character Estimation Under the Threshold Model from Quantitative Genetics. Evolution. 68:743–759.

Sober E., Steel M. 2011. Entropy increase and information loss in Markov models of evolution. Biology & Philosophy. 26:223–250.

Waddington C.H. 1953. GENETIC ASSIMILATION OF AN ACQUIRED CHARACTER. Evolution. 7:118–126.

Westoby M., Leishman M.R., Lord J.M. 1995. On Misinterpreting the `Phylogenetic Correction'. Journal of Ecology. 83:531–534.

Zucchini W., MacDonald I.L., Langrock R. 2017. Hidden Markov models for time series: an introduction using R. Chapman and Hall/CRC.