There and Back Again: Exploring the Roles of Models and Natural History in Macroevolution

by

Michael C. Grundler

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in The University of Michigan
2020

Doctoral Committee:

Associate Professor Daniel L. Rabosky, Chair
Assistant Professor Alison R. Davis Rabosky
Professor Edward L. Ionides
Associate Professor Stephen A. Smith

Michael C. Grundler
mgru@umich.edu
ORCID iD: 0000-0003-0729-6687

Acknowledgements

I thank Bilbo Baggins for use of his evocative phrase *there and back again*, which is meant to connote the long journey and many helping hands it has taken to see this thesis over the finish line.

I thank my advisor, Dr. Dan Rabosky, for his advice and encouragement over many years. Thank you for so patiently helping me along, I learned a lot!

I thank my dissertation committee members whose scientific expertise I regrettably did not take advantage of the way I should have but who encouraged and supported me nonetheless: Dr. Stephen Smith, Dr. Alison Davis Rabosky, and Dr. Edward Ionides. They are not responsible for the many shortcomings in this work!

My time as a graduate student was immeasurably enriched by many happy hours spent in the field with Dan and Alison and members of their respective labs. To all those who participated in these trips (you know who you are!): thank you, thank you, thank you!

When not in the field or in front of a computer terminal, I was privileged to work with the Museum of Zoology's world-class herpetology collection, and I thank collection manager Greg Schneider for making this possible.

And finally, to my family, and to Iris Holmes. I wouldn't have made it here without you!

Table of Contents

List of Figures

Abstract

Ecological diversity in nature is tremendously complex. Evolutionary biologists and ecologists have sought to understand this complexity using foundational concepts like ecological niches, guilds, and adaptive zones. The merger of these concepts with stochastic models and phylogenies helped create the field of phylogenetic comparative methods, which has made fundamental contributions to our understanding of the evolutionary history of life's rich ecological variety and the role ecology plays in the diversification of species and phenotypes and the assembly of species-rich communities. Despite this progress, however, phylogenetic comparative methods have been slow to expand their data repertoire. There is a general rarity of comparative datasets that include primary natural history observations of organisms in nature and of comparative methods to work with such data. The main contribution of this dissertation is to address this shortfall. I do so in three main ways. First, in earlier chapters I study some simple stochastic models of ecological character state change, revealing unappreciated subtleties that complicate our ability to interpret their results in terms of historical events. Second, building off lessons learned from these early chapters, I develop a new method that uses primary natural

history observations to jointly infer the phylogenetic distribution of ecological niche states for individual species and their unsampled ancestors. Third, to demonstrate the flexibility of the new method, I conduct an empirical analysis on the diversification of snake feeding habits using a new comprehensive database of observations of prey acquisition by snakes that I compiled. Taken together, the research in this dissertation demonstrates how fundamental observations of organisms in nature can be used to make quantitative inferences about the macroevolution of complex ecological traits and suggests new ways of integrating natural history data into comparative biology.

Chapter 1
Hello, world

Introduction

Phylogenetic comparative methods are proximally concerned with three things: inferring (possibly correlated) sequences of phenotypic change, estimating rates of phenotypic evolution, and estimating speciation and extinction rates. With rare exceptions, none of these quantities can be directly observed and measured because the time spans over which the relevant processes occur are simply too vast. Instead, stochastic models are used to estimate evolutionary rates and sequences from two sources of information: 1) a phylogeny with temporally scaled branch lengths relating a set of (typically extant) species and 2) phenotypic measurements made on a subset of those species. In attempting to estimate evolutionary rates and sequences, the ultimate goal is to find answers to more interesting questions arising from observable patterns in nature, such as why more species coexist in the tropics than anywhere else on earth, or why most plants have flowers, or why snakes, which can neither walk nor chew food, are so ecologically diverse.

My dissertation research investigates a particular class of stochastic models known as continuous-time Markov chains (CTMC) that are commonly used in the field of macroevolution to study phenotypic evolution. Beginning with the early work of Harvey & Pagel (1991), CTMC models began to replace maximum parsimony as the primary comparative method for studying

phenotypic evolution of discrete characters. Since then, CTMC models have been enthusiastically embraced by the field, and new methods are continually being introduced into the literature and put to use by empirical studies faster than the full range of their behavior can be studied and understood. In some cases, this trajectory has resulted in significant setbacks, such as the revelation that state-dependent speciation-extinction models exhibit type I error rates that can approach 100 percent (Rabosky & Goldberg 2015).

Throughout this thesis I work with relatively simple CTMC models. A model may be simple because it economizes on the number of parameters, types of transitions, or both. I generally work with a model with only two states or with a fully symmetric model that has just a single transition rate (sometimes called the equal-rates model). This is in contrast to the much more elaborate models found in the literature but is motivated by what I consider to be our field's poor understanding of how even the simplest models use available information to estimate evolutionary transition rates. As an example, consider a model with just two states (A and B). How do we interpret the finding of a higher transition rate estimate from A to B than from B to A? Does such a finding support the claim that more transitions from A to B occurred over a clade's evolution? Or that state A is more evolutionary labile than state B and prone to change more quickly? These would be reasonable interpretations of such a finding and are prevalent in the empirical literature, but it must be remembered that the values are estimated from a dataset that contains no actual observations of character state change, unlike classical inference about Markov chains that depends on such observations (e.g. Anderson & Goodman 1957).

In the absence of direct observations of character state change, all we can really say is that estimated transition rate values merely confer high probability on the observed character state data at the tips of a phylogeny, under the assumption that the process that generated those

data was a CTMC. Direct interpretation of transition rate estimates in terms of actual historical events will depend on the general acceptability of that assumption and on evaluation of the sources of information that actually inform rate estimates. Given the incommensurate time-scales of micro- and macro-evolutionary processes, I do not find the use of a phenomenological model like a CTMC problematic. And as the simplest random model for a "stuff happens" viewpoint, I think the equal-rates model is generally a good choice given some of its close connections to the method of maximum parsimony (e.g. Tuffley & Steel 1997) and the diverse random phenomena that have Poisson approximations (e.g. Aldous 1989). However, during the course of this dissertation I have come to appreciate some details of the likelihood calculations and the information sources that influence rate estimates that have undermined some of my confidence in the use of asymmetric models in macroevolution.

In the appendix I present an argument that asymmetries in transition rate estimates, especially in large phylogenies, are most likely driven by asymmetries in character state stasis rather than asymmetries in character state change. The crux of the argument is best illustrated by example. Squamate reproductive modes (oviparity and viviparity) are parsimoniously explained by assigning the majority of ancestors and their descendants the oviparous state. This means that a substantial proportion of the overall phylogenetic lineage length displays a pattern of character state stasis (where ancestor and descendant share the same oviparous state). This, in turn, is interpreted as evidence by an asymmetric model for a several fold higher transition rate from the viviparous state to the oviparous state than the reverse, as this is a likely explanation (under an asymmetric CTMC) for observing so many ancestor-descendant pairs in the same oviparous state (i.e. any transitions to viviparity must revert back quickly). Although such an explanation may make sense for a nucleotide substitution process, I find it very hard to justify for the complex

traits like reproductive mode that are typically studied in macroevolution, and I think many empirical estimates of transition rate asymmetries may ultimately result from this behavior of the likelihood function, including the observation that rate estimates are frequently biased toward the majority tip-state (Nosil & Mooers 2005). I do not intend to argue that asymmetric Markov models have no place in macroevolution, and the arguments in the appendix apply only to the simplest of cases. But until we have a better understanding of how different sources of information in comparative datasets influence transition rate estimates, I think it is best to be wary of inferences resulting from asymmetric models, a recommendation that was made quite early on when these models were first being applied to comparative trait datasets (Schluter et al. 1997).

Chapter summaries

In Chapter 2, I describe a Bayesian approach for detecting phylogenetic shifts in the evolutionary rate of a two-state discrete character, and I implement the method in the BAMM software program (Rabosky 2014). At the initial time of writing, the likelihood equations, which were derived using the approach pioneered by Maddison et al. (2007), were a novel feature of the method since they directly solved a system of ordinary differential equations to obtain the conditional likelihoods for ancestral states, instead of using the more familiar algorithm introduced by Felsenstein (1981). The approach has since been generalized to characters with more than two states by Louca and Pennell (2019), although to my knowledge the closed-form solutions presented in Chapter 2 do not yet appear in the literature. One of the interesting aspects to emerge from this study was the realization that the rate-shift mechanism employed by the model essentially introduced a second way for the character to change state along a lineage. In the absence of constraints, the model would infer a rate-shift at each location where a maximum

parsimony reconstruction implied a change of state. Upon reflection, this finding was actually quite consistent with earlier mathematical results by Tuffley and Steel (1997) revealing some of the links between parsimony and likelihood[1], but it meant that the unconstrained model was not very useful for its intended purpose of detecting clade-level differences in rates of character state change. My ad hoc solution was to constrain the model so that forward and reverse transition rates were identical. I present some simulation results demonstrating that the method does a reasonable job detecting rate-shifts, even when data are simulated under an asymmetric process.

In Chapter 3, I explore an alternative estimator for transition rates that appear in continuous-time Markov chain models of discrete character evolution. This work was motivated by the question of whether maximum parsimony reconstructions could serve as a source of useful information for estimating transition rates, and by the observation that rate estimates reported in the literature on phenotypic macroevolution often appear (to my subjective prior expectation) unusually large. The basic approach was first used by Janson (1992) in a study of seed dispersal syndromes in plants, but it was not picked up on by the field. One of the fundamental differences between the estimator in Chapter 3 and the maximum average likelihood estimator used in standard practice is that in the latter variation in branch lengths influences the estimated relative rate values whereas in the former relative rate values are only affected by the pattern of ancestor-descendant character state changes. I use simulations to demonstrate that maximum average likelihood estimates of transition rates have higher mean squared errors than estimates based on maximum parsimony reconstructions. Quite unexpectedly, errors in maximum average likelihood estimates are accentuated in simulated

[1] In particular, that for a two-state character the method maximum parsimony agrees with the method of maximum likelihood under a fully-symmetric Markov model where the rate of change is 0 for all branches where ancestor and descendant are in the same state and is infinite for all branches where ancestor and descendant are in different states.

datasets where likelihood-ratio tests strongly favor the choice of a model with asymmetric transition rates over a model with symmetric transition rates. In these same situations, by contrast, parsimony-based estimates retain their good performance.

In Chapter 4, I describe a novel phylogenetic comparative method that I developed for working with observational count data sampled from a set of discrete resource categories. This work arose from the need to analyze the dataset on snake diets described in Chapter 5 but is applicable to multivariate count data more generally. The data described in Chapter 5 are highly heterogeneous: sample sizes vary dramatically among species and numerous zeros pervade the data matrix. Given these facts it seemed inappropriate to simply normalize observational counts to proportions and use existing Gaussian diffusion models for three reasons. First, numerous zeros mean that many data points are near the edges of the constrained simplex space where a normal approximation to the untransformed or square-root transformed proportions is unreliable (e.g., Cavalli-Sforza & Edwards 1967). Second, zeros also prevent the use of log-ratio transformations that would take the data into an unconstrained space where a Gaussian diffusion is appropriate (e.g., Aitchison 1986). And third, normalizing counts to proportions implies an equivalency among species with respect to how well characterized patterns of resource use are that does not, in fact, exist. These considerations motivated the development of the new method described in Chapter 4. I assume that individual species belong to a finite set of ecological states, which evolve along the branches of a phylogeny according to a fully symmetric continuous-time Markov chain. I imagine that ecological states assign each species a vector of proportions and that observational counts are sampled from multinomial distributions parameterized by these proportion vectors. The method seeks to estimate ecological state assignments and underlying

proportion vectors, and I present simulation results demonstrating that it achieves satisfactory performance in both these objectives.

In Chapter 5, I present a large, open-source data set on observations of prey acquisition by snakes. The data, which are fairly comprehensive with respect to snake phylogeny, synthesize numerous literature reports and incorporate some of my own observations from fieldwork and museum specimen dissections. This effort was motivated by the lack of comprehensive databases covering aspects of vertebrate trophic ecology. FishBase (Froese & Pauly 2000) is an outstanding exception, but among tetrapods comparable efforts are nowhere close. Widely used databases for birds (Sekercioglu et al. 2004) and mammals (Kissling et al. 2014; Wilman et al. 2014) do not actually contain any primary natural history observations on resource utilization. For example, the proportional utilizations of different foods by birds and mammals reported in Wilman et al. (2014) are not, in spite of appearances, quantitative measures (although they are sometimes treated this way [e.g. Pigot et al. 2016]). They are, in fact, ordinal ranks derived from the subjective impressions of data compilers when reading textual descriptions in field guides. These impressions can vary greatly from one observer to the next (e.g. Parravicini et al. 2020). The data described in Chapter 5 only include direct observations of organisms in nature and are subject to little ambiguity.

In Chapter 6, I present an empirical analysis of the diet data described in Chapter 5 using the method developed in Chapter 4. Variation in trophic ecology is a conspicuous feature of animal diversity, and numerous studies have now analyzed diet datasets within a phylogenetic context. In the main, these studies use discrete character encodings designed to capture major differences in feeding ecology among the focal organisms. In the macroevolutionary literature this tradition goes back at least to the influential study by Mitter et al. (1988) linking the

evolution of plant-feeding in insects to enhanced diversification. Even earlier antecedents, however, can be seen in the works of Simpson (1944), Van Valen (1971), and Root (1967), whose concepts of adaptive zones, ways of life, and ecological guilds align closely with how many researchers categorize species into different ecological states. The central motivation for the work in Chapter 6 is certainly not to argue that such concepts are without merit or that studies based on discrete character encodings are not useful. Rather, it is to explore how studies can approach phylogenetic comparative analysis of animal diets in a more data-driven fashion, using primary natural history observations, such that ecological state descriptions emerge as an output of analysis rather than an input.

Appendix

This appendix develops an argument that asymmetries in estimated transition rates are often more strongly informed by patterns of character state stasis rather than by patterns of character state change, thereby complicating their interpretation as descriptors of change, a viewpoint common in the empirical literature. The intent is to explore how the two patterns that inform the likelihood calculations – a net change of state across a branch and no net change of state across a branch (which I call stasis) – influence the maximum likelihood estimate of transition rate asymmetry. To a first order approximation, I find that the ratio of transition rates equals the ratio of the number of character state changes (i.e., when the pattern of asymmetry in rates matches the pattern of asymmetry in character change) only when the amount of time spent in stasis in each state is equal. When the amount of time in stasis in each state is unequal, the ratio of transition rates is biased in the direction of the state displaying the most stasis, sometimes to the point where it can overwhelm the pattern change. In other words, even if nearly all net change is from A to B the transition rate from B to A can be much higher than the reverse

rate if most of the tree is in state B. While these conclusions make sense given the model, I do not think they are widely appreciated. Indeed, the attempt in this appendix was motivated precisely because there is so little information in the literature about this topic.

For a two-state continuous-time Markov chain of discrete character evolution the lineage transition probabilities are,

$$p_{01}(t) = \frac{\varepsilon}{1+\varepsilon} - \frac{\varepsilon}{1+\varepsilon} e^{-\tau t}$$

$$p_{00}(t) = \frac{1}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon} e^{-\tau t}$$

$$p_{10}(t) = \frac{1}{1+\varepsilon} - \frac{1}{1+\varepsilon} e^{-\tau t}$$

$$p_{11}(t) = \frac{\varepsilon}{1+\varepsilon} + \frac{1}{1+\varepsilon} e^{-\tau t}$$

where $\tau = (q_{01} + q_{10})$, $\varepsilon = \frac{q_{01}}{q_{10}}$, and q_{01} and q_{10} are the forward and reverse transition rates, respectively. My focus concerns the behavior of the maximum likelihood estimator for the parameter ε and how it may be affected by the pattern of evolutionary stasis. I use the symbol $L_{\hat{X}}$ to represent the likelihood of $\hat{X}$, which represents an assignment of character states to all nodes in a phylogeny that is consistent with the observed state assignments at the terminal nodes. In standard practice, the likelihood L of the observed state assignments is computed by summing over all possible $\hat{X}$ so that $L = \sum_{\hat{X}} L_{\hat{X}}$. Because $L_{\hat{X}}$ is just a product of independent lineage transition probabilities we have,

$$\frac{\partial L}{\partial \varepsilon} = \sum_{\hat{X}} L_{\hat{X}} b(\hat{X})$$

$$b(\hat{X}) = \sum_{e=(u,v)} \frac{\frac{\partial}{\partial \varepsilon} p_{\hat{X}(u)\hat{X}(v)}(t_e)}{p_{\hat{X}(u)\hat{X}(v)}(t_e)}$$

Where the second sum runs over each edge $e = (u, v)$ connecting ancestral node *u* with descendant node *v*. We can gain insight into the behavior of the likelihood function by studying $b(\hat{X})$, which can be written as

$$b(\hat{X}) = \sum_{\substack{e=(u,v):\\ \hat{X}(u)=0\\ \hat{X}(v)=1}} \frac{1}{(1+\varepsilon)\varepsilon} - \sum_{\substack{e=(u,v):\\ \hat{X}(u)=1\\ \hat{X}(v)=0}} \frac{1}{1+\varepsilon} + \sum_{\substack{e=(u,v):\\ \hat{X}(u)=1\\ \hat{X}(v)=1}} \frac{1-e^{-\tau t_e}}{(1+\varepsilon)(\varepsilon+e^{-\tau t_e})} - \sum_{\substack{e=(u,v):\\ \hat{X}(u)=0\\ \hat{X}(v)=0}} \frac{1-e^{-\tau t_e}}{(1+\varepsilon)(1+\varepsilon e^{-\tau t_e})}$$

To a first order Taylor series approximation, which will be reasonably accurate whenever the rate of character evolution τ is low, this can be represented as,

$$b(\hat{X}) = \sum_{\substack{e=(u,v):\\ \hat{X}(u)=0\\ \hat{X}(v)=1}} \frac{1}{(1+\varepsilon)\varepsilon} - \sum_{\substack{e=(u,v):\\ \hat{X}(u)=1\\ \hat{X}(v)=0}} \frac{1}{1+\varepsilon} + \sum_{\substack{e=(u,v):\\ \hat{X}(u)=1\\ \hat{X}(v)=1}} \frac{\tau t_e}{(1+\varepsilon)^2} - \sum_{\substack{e=(u,v):\\ \hat{X}(u)=0\\ \hat{X}(v)=0}} \frac{\tau t_e}{(1+\varepsilon)^2}$$

Recalling that $\varepsilon = \frac{q_{01}}{q_{10}}$, the above tells us that the ML estimate of ε is an increasing function of the number of changes from ancestral state 0 to descendant state 1 *and* the total amount of time where ancestor and descendant are both in state 1 (stasis). Conversely, the ML estimate of ε is a decreasing function of the number of changes from ancestral state 1 to descendant state 0 and the amount of time where ancestor and descendant are both in state 0.

The ML estimate of ε for the average likelihood *L* (i.e., the likelihood that results from summing the individual character history likelihoods) is influenced by each character history $\hat{X}$. Each configuration $\hat{X}$ will tend to pull the estimate of ε towards its own ML estimate, and it will do so with a force equal to $L_{\hat{X}} b(\hat{X})$. In other words, high likelihood configurations influence the

ML estimate more than low likelihood configurations. The ML estimate of ε for each individual history is just the value of ε that solves $b(\hat{X}) = 0$. The behavior of these individual configuration ML estimates can be studied graphically as follows. I let $f_{10} = \frac{n_{10}}{n_{01}+n_{10}}$ be the fraction of all state changes from ancestor to descendant that represent reversals to state 0, and I let $\psi = \frac{1}{n_{01}+n_{10}}\left(\sum_{\substack{e=(u,v):\\ \hat{X}(u)=1\\ \hat{X}(v)=1}} \tau t_e - \sum_{\substack{e=(u,v):\\ \hat{X}(u)=0\\ \hat{X}(v)=0}} \tau t_e\right)$, which represents the between state difference in summed lineage length displaying stasis scaled by the number of net changes from ancestor to descendant. Positive values correspond to an excess of stasis in state 1 and negative values correspond to an excess of stasis in state 0. Solutions to $b(\hat{X}) = 0$ can then be expressed as,

$$f_{10} = \frac{1}{1+\varepsilon} + \frac{\varepsilon}{(1+\varepsilon)^2}\psi$$

Approximate ML estimates of ε occur where the horizontal line f_{10} intersects the curve $\frac{1}{1+\varepsilon} + \frac{\varepsilon}{(1+\varepsilon)^2}\psi$. Examples of these solution curves are shown in Figure 1.1. The overall picture that emerges is that the pattern of stasis can dramatically impact the ML estimate of the ratio of transition rates. For example, even if nearly 100 percent of state changes between ancestor and descendant represent gains of state 1, the transition rate from state 1 to state 0 can be higher, sometimes much higher, than the transition rate from state 0 to state 1 if the majority of tree length is in state 0.

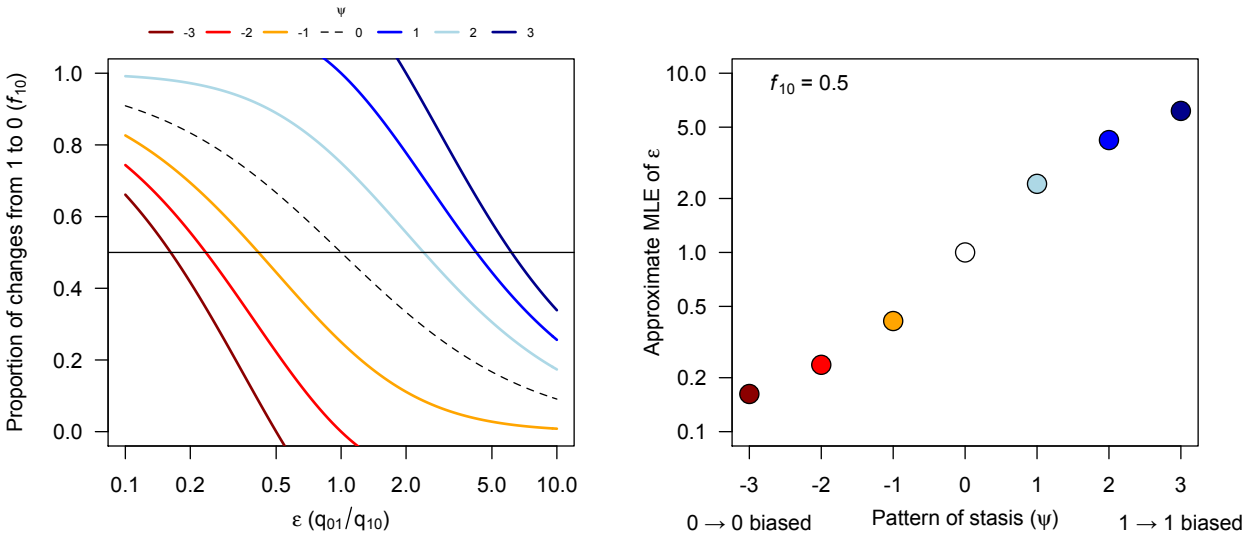

Figure 1.1 *The role of stasis in maximum likelihood estimates of transition rate asymmetry*

Left: Approximate ML estimates of the ratio of transition rates (ε) occur where horizontal lines intersect the different stasis (ψ) curves. *Right*: Example illustrating the behavior of the approximate ML estimate of ε as a function of stasis for a fixed proportion of changes from state 1 to state 0 ($f_{10} = 0.5$).

To verify the accuracy of this approximation I applied it to the squamate reproductive mode dataset published by Pyron & Burbrink (2014). I sampled 100 maximum parsimony histories of reproductive mode evolution and for each of these numerically estimated the maximum likelihood value of ε and τ. I then calculated the predicted ML estimate of ε using the approximation method discussed above while holding τ fixed to the ML estimate found earlier. The correspondence between true and predicted estimates is quite good and is illustrated in Figure 1.2. Thus, even though more than three-quarters of the parsimony implied events of character state change represent origins of viviparity from an oviparous ancestor, the transition rate from viviparity to oviparity is upwards of twice the transition rate from oviparity to

viviparity, simply because the majority of lineage length in maximum parsimony reconstructions is in the oviparous state.

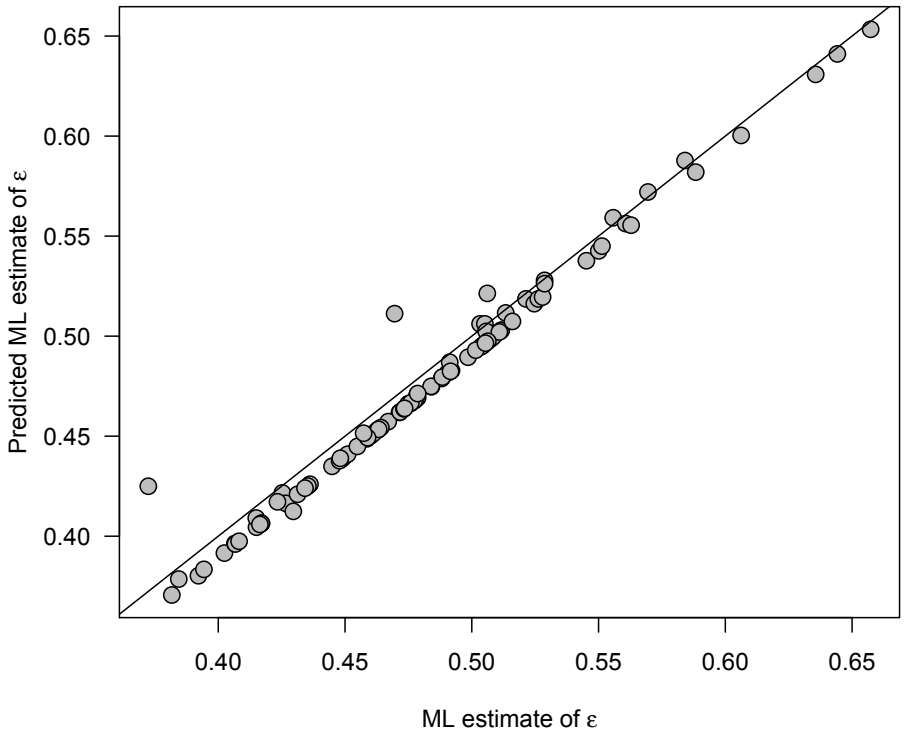

Figure 1.2 *Accuracy of approximation to ML estimate of transition rate asymmetry*

Correspondence between the predicted ML estimate of ε obtained using the approximation method developed above and the true ML estimate of ε obtained using numerical optimization with the squamate reproductive mode dataset.

Chapter 2
Macroevolutionary Analysis of Discrete Traits with Rate Heterogeneity[2]

Abstract

Organismal traits show dramatic variation in phylogenetic patterns of origin and loss across the tree of life. Understanding the causes and consequences of this variation depends critically on accounting for heterogeneity in rates of trait evolution among lineages. Here, we describe a method for modeling among-lineage evolutionary rate heterogeneity in a trait with two discrete states. The method assumes that the present-day distribution of a binary trait is shaped by a mixture of stochastic processes in which the rate of evolution varies among lineages in a phylogeny. The number and location of rate changes, which we refer to as rate-shift events, are inferred automatically from the data. Simulations reveal that the method accurately reconstructs rates of trait evolution and the number and location of rate-shift events, even when simulated data violate model assumptions. We apply the method to an empirical dataset of mimetic coloration in snakes and find elevated rates of trait evolution in two clades of harmless snakes that are broadly sympatric with dangerously venomous New World coral snakes,

[2] Grundler, M.C. and Rabosky, D.L. (2020). Macroevolutionary analysis of discrete traits with rate heterogeneity. *BioRxiv*, 897777, https://doi.org/10.1101/2020.01.07.897777.

recapitulating an earlier analysis of the same dataset. Although the method performed well on many simulated data sets, we caution that overall power for inferring heterogeneous dynamics of single binary traits is low.

Introduction

Organismal traits evolve recurrently across the tree of life, and the frequency with which traits evolve varies widely among clades. Among extant amniotes, for example, viviparity has evolved at least 100 times but only a single origin of it (in mammals) occurs outside squamate reptiles (Blackburn 1982, 1985). Similarly, sugar-secreting glands known as extrafloral nectaries have arisen hundreds of times among plants, mainly within the legumes, but not once among gymnosperms (Weber and Keeler 2013). Identifying the causes and consequences of repeated convergence depends critically on accounting for heterogeneity in rates of trait evolution. Ancestral trait reconstructions reveal evidence of convergence and can identify other traits enabling the repeated evolution of a convergent trait (Maddison 1990; de Queiroz and Rodríguez-Robles 2006; Marazzi et al. 2012; Christin et al. 2013). However, inferences of ancestral states may be seriously misled by models that fail to account for rate heterogeneity (King and Lee 2015). Several evolutionary theories also link how quickly traits evolve to how quickly lineages diversify (Vermeij 1973; Stanley 1979), and the repeated evolution of a trait together with methods that model among-lineage variation in evolutionary rates allows for direct tests of this coupling (Rabosky et al. 2013; Igea et al. 2017).

Methods to account for evolutionary rate heterogeneity have grown in number and in sophistication over recent years (O'Meara 2006; Revell and Collar 2009; Eastman et al. 2011; Lloyd et al. 2012; Marazzi et al. 2012; Beaulieu et al. 2013; Landis et al. 2013; Rabosky et al.

2014; Uyeda and Harmon 2014), but much of this methodological progress has focused on continuous traits such as body mass or seed size. Phylogenetics deals extensively with rate heterogeneity in discrete character evolution because of the challenge it poses for inference of evolutionary relationships (e.g. Heath et al. 2012), and some of these methods carry over to the study of non-molecular traits. The "hidden rates" (Beaulieu et al. 2013) and "precursor" (Marazzi et al. 2012) models, first introduced to study variation in growth form and extrafloral nectary production in plants, are closely related to covarion models of nucleotide substitution (Fitch and Markowitz 1970; Galtier 2001; Penny et al. 2001). The "random local clock" model (Drummond and Suchard 2010) used to study variation in the rate of nucleotide substitution has also been used to study evolutionary rate variation in mimetic color pattern (Davis Rabosky et al. 2016) and reproductive mode parity (King and Lee 2015) in squamate reptiles.

In this paper, we describe a method for modeling evolutionary rate heterogeneity in a trait with two discrete states and implement it using the Bayesian Analysis of Macroevolutionary Mixtures (BAMM) framework (Rabosky 2014; Rabosky et al. 2014). The model discussed here is closely related to several existing phylogeny inference methods that model among-lineage substitution rate variation (Huelsenbeck 2000; Drummond and Suchard 2010) but differs in details of likelihood calculation and implementation. The general approach assumes that the present-day distribution of a binary state character is shaped by a mixture of stochastic processes in which the rate of evolutionary transition between the two states experiences shifts under a Poisson process across the branches of a phylogeny. The number and location of rate changes, which we refer to as rate-shift events, are inferred automatically from the data. Simulations reveal that the method accurately infers rates of evolution and the number and location of rate-shift events, even when simulated data violate model assumptions.

Materials & Methods

Likelihood of a binary state character

We assume that each branch of a phylogeny evolves independently of the others and that the probability of a lineage transitioning to a character state different from its current state does not depend on its prior history of trait evolution (Pagel 1994). For a discrete character with two states, state 0 and state 1, trait evolution is modeled by a "forward" transition rate (denoted q_{01}), which governs how frequently a lineage in state 0 changes to state 1, and a "reverse" transition rate (denoted q_{10}), which governs how frequently a lineage in state 1 changes to state 0. For example, over a sufficiently small interval time Δt, the probability of observing a transition from state 0 to state 1 is approximately $q_{01}\Delta t$.

To write down a likelihood function for estimating the transition rate parameters, we follow the approach of Maddison et al. (2007) and define $D_{N0}(t)$ to be the probability that lineage N evolves the distribution of character states observed among its descendants given that it is in state 0 at time t. We define $D_{N1}(t)$ analagously. Next, consider what can happen in the short interval of time between t and $t+h$, where $t+h$ is closer to the root and where h is taken to be an interval of time small enough that the probability of more than one character state transition in the interval is negligible. There are only two possibilities. Either the lineage remains in the state that it was in at time $t+h$ or it switches to the other state. We can therefore write $D_{N0}(t+h)$ and $D_{N1}(t+h)$ as functions of $D_{N0}(t)$, $D_{N1}(t)$, and the transition rate parameters q_{01} and q_{10},

$$D_{N0}(t+h) = (1 - q_{01}h)D_{N0}(t) + q_{01}hD_{N1}(t)$$

$$D_{N1}(t+h) = (1 - q_{10}h)D_{N1}(t) + q_{10}hD_{N0}(t)$$

By rearranging and letting h approach zero we form two coupled differential equations that describe how these probabilities change through time,

$$\frac{dD_{N0}(t)}{dt} = -q_{01}D_{N0}(t) + q_{01}D_{N1}(t)$$

$$\frac{dD_{N1}(t)}{dt} = -q_{10}D_{N1}(t) + q_{10}D_{N0}(t)$$

These can be solved (see appendix) to give the closed form solutions,

$$D_{N0}(t) = A + B_0 e^{-(q_{01}+q_{10})(t-t_0)} \quad (1)$$

$$D_{N1}(t) = A + B_1 e^{-(q_{01}+q_{10})(t-t_0)} \quad (2)$$

Where t_0 is the initial time and,

$$A = \pi_0 D_{N0}(t_0) + \pi_1 D_{N1}(t_0)$$

$$B_0 = \pi_1 D_{N0}(t_0) - \pi_1 D_{N1}(t_0)$$

$$B_1 = \pi_0 D_{N1}(t_0) - \pi_1 D_{N0}(t_0)$$

Where π_i is the equilibrium frequency of state *i*. Equations (1) and (2) are evaluated for each branch of the phylogeny proceeding from the tips to the root in a post-order traversal. The quantities $D_{N0}(t_0)$ and $D_{N1}(t_0)$ are the initial conditions used to begin evaluation for each branch. If we have these values, we can compute the conditional likelihood of the data for a branch's stem group by simply setting t equal to the time at the base (rootward) of the branch and t_0 equal to the time at the head (tipward) of the branch and evaluating equations (1) and (2). At each internal node, we form a new set of initial conditions by multiplying the $D_{.0}$ and $D_{.1}$ at the base of the node's left descendant branch with those of its right descendant. When we reach the root, R, of the tree $D_{R0}(t_R)$ and $D_{R1}(t_R)$ yield the probability of the data given the transition rate parameters and conditional on the root being in state 0 or state 1, respectively. To get the unconditional likelihood we must combine these two values, but doing so requires knowing the

probability of the root being in state 0 or in state 1. We assume that the root is in state 0 or state 1 with probabilities implied by their conditional likelihoods and compute the full likelihood as $\frac{D_{R0}(t_R)}{D_{R0}(t_R)+D_{R1}(t_R)} D_{R0}(t_R) + \frac{D_{R1}(t_R)}{D_{R0}(t_R)+D_{R1}(t_R)} D_{R1}(t_R)$. This is the same weighting scheme used in some models of trait-dependent speciation and extinction (FitzJohn et al. 2009). To begin computation, the initial conditions for each tip in the tree are set to $D_{\cdot 0}(0) = 1$ and $D_{\cdot 1}(0) = 0$ if the tip is in state 0 and vice versa if the tip is in state 1.

Rate-shift model for a binary state character

At broad phylogenetic scales, it may be unreasonable to assume that the transition rates remain constant over all regions of phylogeny. We therefore allow regions of a phylogeny to belong to different macroevolutionary "rate-regimes", which are independent sets of transition rate parameters that describe the evolution of a binary state character over the regions of phylogeny to which they pertain. Transitions between rate-regimes, which we refer to as "rate-shift events", are assumed to occur along the branches of a phylogeny according to a compound Poisson process (Huelsenbeck 2000; Blanquart and Lartillot 2006; Rabosky 2014). To compute the likelihood of the data under a rate-shift model requires minimal modification of the process described in the previous section. Rather than traversing directly from a node to its ancestor when evaluating equations (1) and (2), we pause at each rate-shift event that occurs on the branch and use the values of (1) and (2) at that point as starting values for an additional evaluation of (1) and (2) under the new set of transition rates.

This model of rate variation is unsuitable when forward and reverse transition rates are asymmetric (Fig. 2.1). By proposing a new rate-shift event for every event of character state change and simply making one rate arbitrarily large and the other rate arbitrarily small it is

possible to fit any data set with probability 1 (when conditioned on the occurrence of the rate-shift events). This is because we can guarantee with probability 1 the origin and persistence of a derived character state by making the transition rate toward the derived character state arbitrarily large and the reverse rate arbitrarily small. In such a scenario, rate-shift events become decoupled from broad-scale among-lineage variation in the rate of trait evolution. Our implementation of an asymmetric version of this rate-shift model yielded results consistent with these expectations for several empirical datasets. For this reason, the default implementation in BAMM constrains forward and reverse transition rates to be identical. All analyses presented below use this symmetric implementation.

Implementation

The rate-shift model described above is implemented in the Bayesian software program BAMM using reversible-jump Markov Chain Monte Carlo simulation (Rabosky 2014). Briefly, BAMM assumes that the number of rate-shift events on a phylogeny is drawn from a Poisson distribution with a rate parameter that is itself drawn from an exponential distribution. This formulation implies that the number of rate-shifts is drawn from a geometric distribution and that the expected number of rate-shifts, denoted by Λ, is simply the mean of the exponential hyperprior placed on the Poisson prior (Mitchell and Rabosky 2017). We extend the BAMM implementation for binary data by placing an exponential prior on the transition rate and use a proportional shrinking-expanding proposal mechanism to update its value. All other details remain the same and are described elsewhere (Mitchell and Rabosky 2017).

Simulation study

We conducted a simulation study to assess how well the method estimates branch specific rates of evolution and infers the presence of rate-shift events.

(1) Choice of phylogeny

To carry out the simulations, we generated 100 phylogenies having between 50 and 1,000 tips by randomly sampling internal nodes from the 3,962-tip ultrametric squamate reptile phylogeny of Pyron and Burbrink (2014). Nodes were assigned weights such that all sizes (measured as the number of living descendants) of extracted clades had an equal probability of being selected. We chose to select subsets of a large empirical phylogeny, rather than simulated phylogenies, to introduce more realistic distributions of branch lengths than might be obtained using simple tree simulation models (e.g. Yule or constant-rate birth-death models).

(2) Simulating trait evolution

For each phylogeny, we simulated evolution of a binary state trait 10 times using 2 different procedures. In the first case, we determined the number of rate-shift events to place on the tree by drawing a random integer from a Poisson distribution with a rate of 1. We determined the locations for these rate-shifts by selecting that number of internal nodes randomly without replacement, again using weights that gave all sizes of subtrees an equal probability of being chosen, and choosing a uniform random point along each chosen node's branch. We chose the transition rate for each rate-regime by drawing a random number from a log-normal distribution with a mean of $\log 0.01$ and a standard deviation of $-\frac{1}{3}\log 0.01$. This corresponds to a log-normal distribution with a spread that position rates of 0.0001 and 1 three standard deviations below and above the mean, respectively, and these values were chosen simply for their potential to generate datasets having a range of phylogenetic signals. The second case was identical to the

first except that we allowed transition rates to be asymmetric. The degree of asymmetry was determined by drawing a random number from a log-normal distribution with a mean of $\log 1$ and a standard deviation of $\frac{1}{3}\log 100$. This corresponds to a log-normal distribution centered on unbiased transition rates with a spread that positions a 100-fold bias in transition rates 3 standard deviations from the mean.

(3) Information content of rate-shift events

Due to the stochastic nature of the simulations we expected rate-shift events to vary in their degree of detectability. To quantify this, we followed Rabosky et al. (2017) and calculated the "information content" of each rate-shift, which is a measure of how strongly the data support a model with rate variation over a model with no rate variation. For each simulated trait distribution, we optimized the value of the transition rate parameter to maximize the likelihood of the full data under a model with no rate-shifts. We denote this maximum-likelihood parameter estimate by θ_{CR}. Next, we denote by D_i the trait data contained in the subtree formed by the set of all nodes and branch segments belonging to the *i*-th rate-regime and by θ_i the transition rate that generated those data. We calculated the information content of the *i*-th rate regime as the difference in log likelihoods of D_i under the two parameter sets,

$$\Delta \operatorname{Log} L_i = \log L(D_i|\theta_i) - \log L(D_i|\theta_{CR})$$

Where $L(D_i|\theta_i)$ denotes the likelihood of the data in the *i*-th rate regime given the generating parameter, and θ_{CR} indicates the transition rate obtained by maximizing the likelihood for the full data under a model with no rate-shifts. If the data support a model with rate-shift events, the $\Delta\text{Log } L$ statistic must be greater than 0. In general, we expect that this statistic must be substantially greater than 0 for BAMM to detect a rate-shift event. If $\Delta\text{Log } L$ is expressed using the Akaike Information Criterion it can be rewritten as $\Delta\text{Log } L = {}^{S}/_{2} + k$, where *S* is the

difference in AIC scores needed to accept a model with an additional rate-shift and k is the number of extra parameters required to fit an additional rate-shift (in our implementation $k = 2$, corresponding to the location of the rate-shift and its rate parameter). If $S \geq 0$ is interpreted as support for a model with an extra rate-shift, the minimum $\Delta\text{Log } L$ needed to detect an event is 2.

BAMM analysis

We analyzed each simulation with BAMM using a single Markov chain that ran for ten million generations with Λ set equal to 1. The transition rates governing trait evolution need starting values before BAMM can proceed. For each simulation, we divided the total number of observed character state transitions by the summed branch length of the simulation's phylogeny and used this value for the initial transition rate and for the median of the exponential prior placed on the transition rate. This number is unavailable in empirical datasets but will be close to the number implied by parsimony when rates of character evolution are low.

Performance assessment

For each dataset simulated under a symmetric or asymmetric rate-shift model we assessed BAMM's ability to estimate rates of trait evolution and to detect rate-shift events. We performed each assessment using the estimated posterior distribution after discarding the first ten-percent of samples.

(1) Estimating the rate of trait evolution

To determine how accurately BAMM infers rates of trait evolution we scaled each branch length to correspond to its average estimated rate of trait evolution. We performed this branch-by-

branch calculation for each sample in the simulated posterior distribution using the BAMM-estimated transition rate parameters and assigned each branch an overall length equal to,

$$v_i = \frac{\sum_s v_{i,s}}{N}$$

Where $v_{i,s}$ denotes the rate of evolution of branch *i* in the *s*-th posterior sample, and N indicates the number of samples in the posterior distribution. Using these values, we computed the tree-wide proportional branch rate error for a simulation as,

$$R = \sum_i \frac{t_i}{T}(\log v_i - \log V_i)$$

Where V_i denotes the true rate of evolution of branch *i*, and the summation is taken over all branches in the phylogeny. The error of the *i*-th branch estimate was weighted by the proportional contribution of its branch length, t_i, to the sum of all branch lengths, T, in the phylogeny. For unbiased transition rate estimates this equation is equal to 0.

(2) Detection of rate-shift events

For each simulated rate-shift event, we computed the mean detection accuracy as $\frac{\sum_s J_s}{N}$, where J_s indicates the detection accuracy in the *s*-th posterior sample and N indicates the number of samples in the posterior distribution. Letting C_s denote the set of rate-shift events detected by BAMM in the *s*-th posterior sample, we measured the detection accuracy of a generating rate-shift event as,

$$J_s = \max\left(\left\{\frac{|N \cap N_i|}{|N \cup N_i|}, i \in C_s\right\}\right)$$

Where N is the set of tips descended from the generating rate-shift event and N_i is the set of tips descended from the *i*-th rate-shift detected by BAMM in the *s*-th posterior sample. A value of 1 occurs when BAMM identifies the precise node that corresponds to the generating rate-shift

event, and a value of 0 occurs when BAMM fails to identify the generating event at all. Intermediate values occur when BAMM identifies rate-shift events that correspond to nodes above or below the node where the generating rate-shift event occurred.

(3) Empirical application

Finally, we analyzed an empirical dataset of mimetic coloration in snakes previously analyzed with several related methods (Davis Rabosky et al. 2016). Red-black banded coloration arises repeatedly among harmless colubrid snakes that occur in broad sympatry with dangerously venomous red-black banded coral snakes in the Neotropics and parts of the North temperate zone. The incidence of red-black banded coloration is particularly high among dipsadine snakes, a highly diverse clade of colubrid snakes that occur in local and regional sympatry with coral snakes across the Neotropics. Using the random local clock model (Drummond and Suchard 2010) and a Medusa-like (Alfaro et al. 2009) model of discrete trait evolution, Davis Rabosky et al. (2016) inferred elevated rates of mimetic color evolution within dipsadine snakes resulting from repeated independent origins of red-black banding coincident with the diversification of coral snakes in the Neotropics. We repeated their macroevolutionary rate analysis of this dataset using the method developed in this paper. We divided the number of parsimony-implied character state changes by the total branch length of the phylogeny to obtain a median rate for the transition rate prior and ran BAMM for 10 million generations with Λ set equal to 10. We repeated this analysis 3 times under 3 different transition rate prior specifications corresponding to a 2-, 5-, and 10-fold speedup over the median rate implied by parsimony.

Results

(1) Estimating the rate of trait evolution

Tree-wide proportional branch rate errors were low regardless of whether the data were simulated under a symmetric (mean = 0.09, median = 0.08) or asymmetric (mean = -0.27, median = -0.17) rate-shift model (Fig. 2.2). The correlations between estimated and true branch rates were high when data were simulated using a symmetric model (Pearson's $\rho = 0.64$, Spearman's $\rho = 0.90$) but were substantially noisier when data were simulated using an asymmetric model (Pearson's $\rho = 0.26$, Spearman's $\rho = 0.60$). When branch rates were multiplied by the temporal duration of each branch to convert them into the expected number of character state changes these correlations improved, particularly for the asymmetric model (*symmetric*: Pearson's $\rho = 0.66$, Spearman's $\rho = 0.93$; *asymmetric*: Pearson's $\rho = 0.37$, Spearman's $\rho = 0.75$). Spearman rank correlations were substantially higher than Pearson product-moment correlations for both models, indicating that relative branch rates are generally better estimated than absolute branch rates.

(2) Detection of rate-shift events

BAMM's ability to detect rate-shift events was generally low due to the limited information content of rate-shift events in the simulated data (Fig. 2.4). Despite relative rate differences between ancestral and derived rate-shift events that varied over 6 orders of magnitude, only 400 of 2045 simulated rate-shift events had an information content above 2, the theoretical minimum above which BAMM is expected to have power to detect them (see Discussion). This is due in part to the small average number of tips in simulated rate-regimes but also to the legitimate difficulty of simulating binary character data that reveal strong evidence for rate heterogeneity (cf. Fig. 2.4 rightmost panel). When BAMM has information to detect rate-shift events, however, it does reasonably well, and its performance improves monotonically as the information content

of rate-shifts increases (cf. Fig. 2.4 leftmost panel). For example, BAMM detected the locations of 36 rate-shifts with an information content of at least 10 with a mean accuracy of 88%.

(3) Empirical application

Analysis of the empirical dataset of red-black banded coloration in snakes largely recapitulated previous results showing an increased rate of trait evolution in Neotropical dipsadine snakes (but also revealed high rates in a clade of North temperate colubrine snakes) (Fig. 2.5). When analyzed under a strong, well-informed transition rate prior, rates of trait evolution ranged from a low of 0.00065 My^{-1} in basal snake lineages to a high of 0.0093 My^{-1} in dipsadine snakes, with an overall mean of 0.0024 My^{-1} that closely matched the overall rate of 0.002 My^{-1} obtained by dividing the number of parsimony-inferred state changes by the total branch length (cf. Fig. 2.5 leftmost panel). More liberal priors did not change this general picture, however rates of trait evolution in the upper tail of the estimated branch rate distribution tended to creep upward as the transition rate prior flattened. Flatter transition rate priors were associated with a higher number of posterior rate-shifts caused by the partitioning of single rate-shift events inferred under steeper transition rate priors into multiple smaller rate-shift events (i.e. lineages left out of these new rate-shift events fell back into the ancestral root regime). These new rate-shift events tended to have estimated rates of evolution that were higher than estimates made using steeper transition rate priors.

Discussion

In this paper, we describe a Bayesian method, implemented in the BAMM software program, for studying among-lineage variation in the rate of evolution of a binary character. Overall, our results show that the method accurately infers rates of trait evolution and the

presence and location of among-lineage evolutionary rate variation, even when simulated data violate model assumptions. Although the method performed well on many simulated data sets, we caution that overall power for inferring heterogeneous dynamics of single binary traits may be low.

The ability of the method to detect rate-shift events depends on the size of the clade belonging to a rate-regime but also on how much information the data contain with respect to the parameters of the rate-regime. In our simulations, we estimated this information content using a log-likelihood ratio that measures the likelihood of a given rate-shift event under the true parameters relative to the corresponding likelihood under a simple model where the rate is set to the whole-tree average. This information content can be surprisingly low even for large clades. In retrospect, this is not necessarily surprising because with only two character states there is a limit to how different data generated by two different rates can become. A 10- or even 100-fold rate increase will not be detectable if the ancestral rate is already high enough as to leave no phylogenetic signal. Similarly, a 10- or 100-fold rate decrease will not be detectable if the ancestral rate is already so low as to make character change highly improbable. In general, we expect detectability of rate-shifts to depend strongly on how distinct one rate's phylogenetic signal is from another. This will have a strong stochastic component to it, and for binary data will likely have a low signal-to-noise ratio making detection of rate-shift events difficult.

While we did not explore prior sensitivity exhaustively in this study, the empirical results indicate that branch rate estimates (and ancestral state reconstructions by implication) are sensitive to the transition rate prior. In the empirical example, disagreements among the different priors occur in regions of phylogeny having elevated rates of evolution. This is not surprising given that the method works with only a single binary character and that the fast-evolving clades

in the empirical phylogeny have relatively few taxa, but it does call for vigilance. A sensible rule of thumb is to treat with caution any result where the overall mean rate estimate disagrees strongly with the rate implied by parsimony. Encouragingly, the different prior specifications are in broad agreement on where relative rate differences occur in the empirical phylogeny despite disagreements over absolute rate estimates. The results also indicate that the transition rate prior interacts with the prior on the number of rate-shifts, suggesting that the method's ability to estimate the precise location of a rate-shift and its ability to estimate the associated rate of evolution may trade-off. Finally, the combination of a prior with a high number of expected rate-shift events and a relatively flat transition rate prior can lead to an apparent abundance of single tips with derived character states having elevated rates of evolution. This is because dropping a high rate of evolution on such a branch entails no penalty. It makes the derived state more probable and with only a single lineage does not suffer from the likelihood penalty that a larger clade fixed for a derived state would suffer from if given an elevated rate. Users should remain alert to this scenario and treat its presence as an indication that the priors are exerting an undue influence. For all these reasons, we recommend the use of a strong, well-informed prior on the transition rate and setting the median rate of the prior equal to the parsimony-implied rate seems like a sensible choice.

A fundamental difficulty facing macroevolutionary models of discrete character evolution is that the data contain low information content with respect to rates of evolution since actual events of character state change are not directly observed. This challenge is exacerbated further in asymmetric Markov models which must estimate the rate of evolution while simultaneously inferring the equilibrium frequencies of the character states. From this fact alone, we should expect asymmetric Markov models to have lower information content with respect to

transition rates than symmetric Markov models, which a priori assume equilibrium frequencies of character states are equal. In a discrete-time version of the Markov model used by this and many other studies, Sanderson (1993) has shown that maximum likelihood parameter estimates are intrinsically biased upwards for an asymmetric model with two character states but are unbiased for a symmetric model. The extent to which these conclusions apply in continuous-time or generalize to more than two states is not known to us but warrants further study.

One consequence of the low information content of binary data are ancestral state reconstructions that may appear nonsensical when performed with asymmetric models. For example, Pagel (1999) presented a comb phylogeny in which every tip in the clade was fixed for one of two possible character states and showed that an asymmetric Markov model reconstructed the root as belonging to either state with equal probability. From an asymmetric model's perspective, a clade that is almost entirely fixed for a single character state is likely to have been generated by a process with a very high transition rate toward the majority state and a very low transition rate away from it, which makes the tip states mostly independent of the ancestral states because the asymmetry in rates will yield the same outcome at the tips regardless of the assignment of states to internal nodes. By contrast, from a symmetric model's perspective the clade simply has a very low rate of evolution and the tip states resemble their ancestors as a result. In fact, transition rates are nearly always biased in the direction of the state appearing most frequently among the tips of the tree (Nosil and Mooers 2005; Maddison 2006). This means that in an analysis that infers a high rate of transition from state 0 to state 1 and low rate of transition from state 1 to state 0, it will sometimes be the case that a majority of the evolutionary transitions were from state 1 to state 0. This runs counter to intuition and indicates the need for

caution when using transition rates to infer directionality in the history of trait evolution (Nosil and Mooers 2005; Goldberg and Igic 2008).

Appendix

Derivation of likelihood equations (1) and (2)

Begin with the differential equations,

$$\frac{dD_{N0}(t)}{dt} = -q_{01}D_{N0}(t) + q_{01}D_{N1}(t) \quad \text{(A1)}$$

$$\frac{dD_{N1}(t)}{dt} = -q_{10}D_{N1}(t) + q_{10}D_{N0}(t) \quad \text{(A2)}$$

By solving (A1) for $D_{N1}(t)$ and equating its derivative to (A2) we can form the second-order differential equation,

$$\frac{d^2D_{N0}(t)}{dt^2} + (q_{01} + q_{10})\frac{dD_{N0}(t)}{dt} = 0$$

Which, after finding the roots of its auxiliary equation, has the general solution,

$$D_{N0}(t) = c_1 + c_2e^{-(q_{01}+q_{10})(t-t_0)} \quad \text{(A3)}$$

Where t_0 is a starting time on node N's branch, $c_1$and c_2 are constants, and t is larger than t_0 but smaller than the time at the base of node N's branch.

To solve for c_1 note that at our initial condition when $t = t_0$,

$$c_2 = D_{N0}(t_0) - c_1 \quad \text{(A4)}$$

Furthermore, differentiating equation (A3) gives,

$$\frac{dD_{N0}(t)}{dt} = -(q_{01} + q_{10})c_2e^{-(q_{01}+q_{10})(t-t_0)} \quad \text{(A5)}$$

And equating (A1) with (A5) and setting $t = t_0$ yields, after some algebra,

$$q_{01}c_1 - q_{10}c_2 = q_{01}D_{N1}(\quad_0)$$

(A6)

By substituting the right-hand side of (A4) for c_2 in (A6) and solving for c_1 we find,

$$c_1 = \frac{q_{01}D_{N1}(t_0)+q_{10}D_{N0}(t_0)}{q_{01}+q_{10}} \quad \text{(A7)}$$

Finally, by substituting the right-hand side of (A7) for c_1 in (A6) and solving for c_2 we find,

$$c_2 = \frac{q_{01}D_{N0}(t_0)-q_{01}D_{N1}(t_0)}{q_{01}+q_{10}} \quad \text{(A8)}$$

Thus, equation (A3) is equivalent to equation (1) after substitution of (A7) and (A8), noting that $\frac{q_{01}}{q_{01}+q_{10}} = \pi_1$ and that $\frac{q_{10}}{q_{01}+q_{10}} = \pi_0$. An equivalent derivation of equation (2) is performed by starting with $D_{N1}(t) = c_1 + c_2 e^{-(q_{01}+q_{10})(t-t_0)}$. When $\pi_0 = \pi_1 = \frac{1}{2}$ as in the BAMM implementation this is simply a special case of the JC69 model (Jukes and Cantor 1969) with two states.

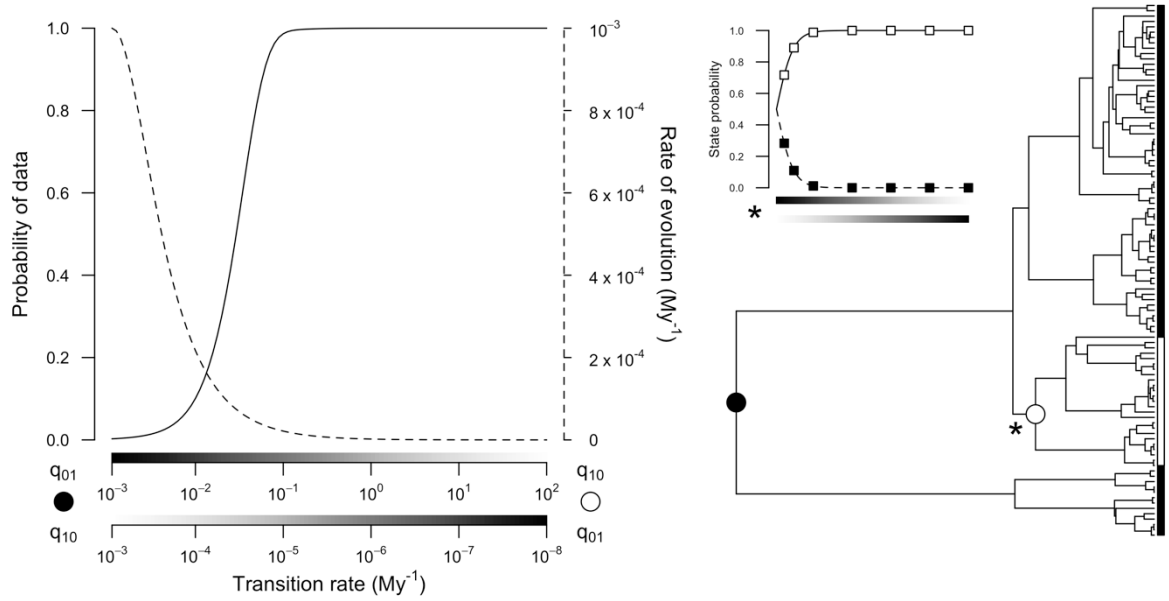

Figure 2.1 *Why a rate-shift mechanism is unsuitable for asymmetric Markov models*

A worked example showing why asymmetric Markov models are unsuitable for BAMM-type rate-shift models. A single rate-shift event (denoted by a white circle) is placed along the branch subtending the clade fixed for character state 0. The *x*-axes in the left plot depict transition rates for this rate-shift and for the root event (denoted by a black circle). Note that they run in opposite directions, e.g. the upper *x*-axis is q_{01} for the root event but is q_{10} for the rate-shift event. As the asymmetry in transition rates is increased the probability of the data (denoted by the solid line) rises to 1 while the overall rate of character evolution (denoted by the dashed line) falls to 0. Simultaneously, as the rate of character evolution falls to 0 the asymmetry in transition rates causes the prior probability of being in one or the other character state to rise to 1, e.g. the inset plot denoted by an asterisk depicts the probability of being in state 0 or state 1 as a function of the asymmetric transition rates of the rate-shift event. Thus, at the extreme a "rate-shift" event simply introduces a second way to observe a stochastic event of character state change and does not correspond to among-lineage heterogeneity in rates of character evolution.

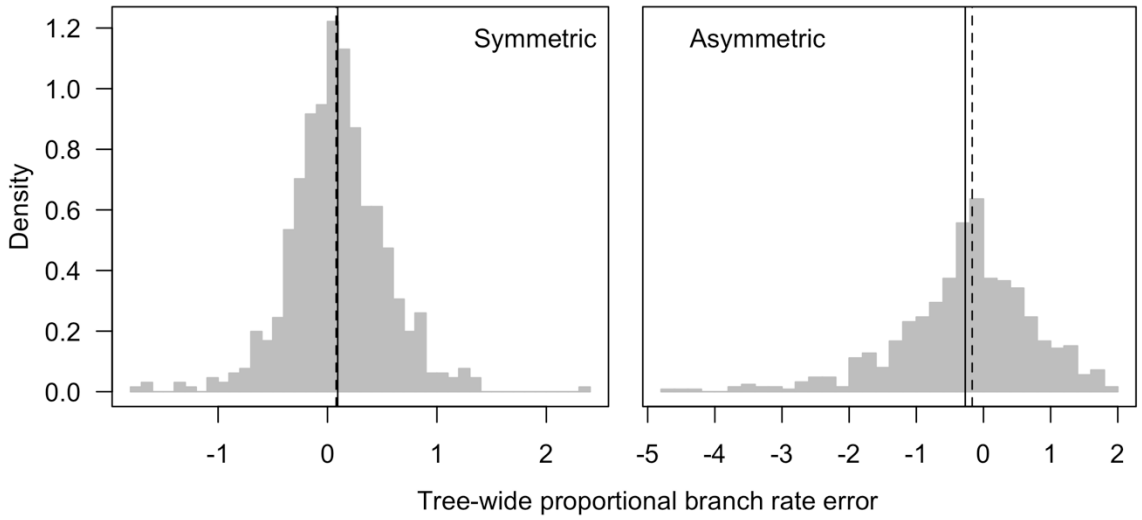

Figure 2.2 *Tree-wide error in branch rate estimates*

Tree-wide proportional branch rate error in BAMM-estimated branch rates when data are simulated under a symmetric rate-shift model (*left*) or an asymmetric rate-shift model (*right*). Branch rates are measured as the expected number of character transitions per million years. Tree-wide proportional branch rate error is a weighted sum of the logarithmic difference between estimated branch rates and true branch rates over all branches in a phylogeny. Errors on short branches are down-weighted relative to errors on long branches. Unbiased estimates have an error of 0, negative and positive values correspond to under- and over-estimation errors, respectively. Each histogram depicts the distribution of tree-wide proportional branch rate error over simulations with at least one rate-shift event. The vertical lines show the mean (solid) and median (dashed) of each distribution.

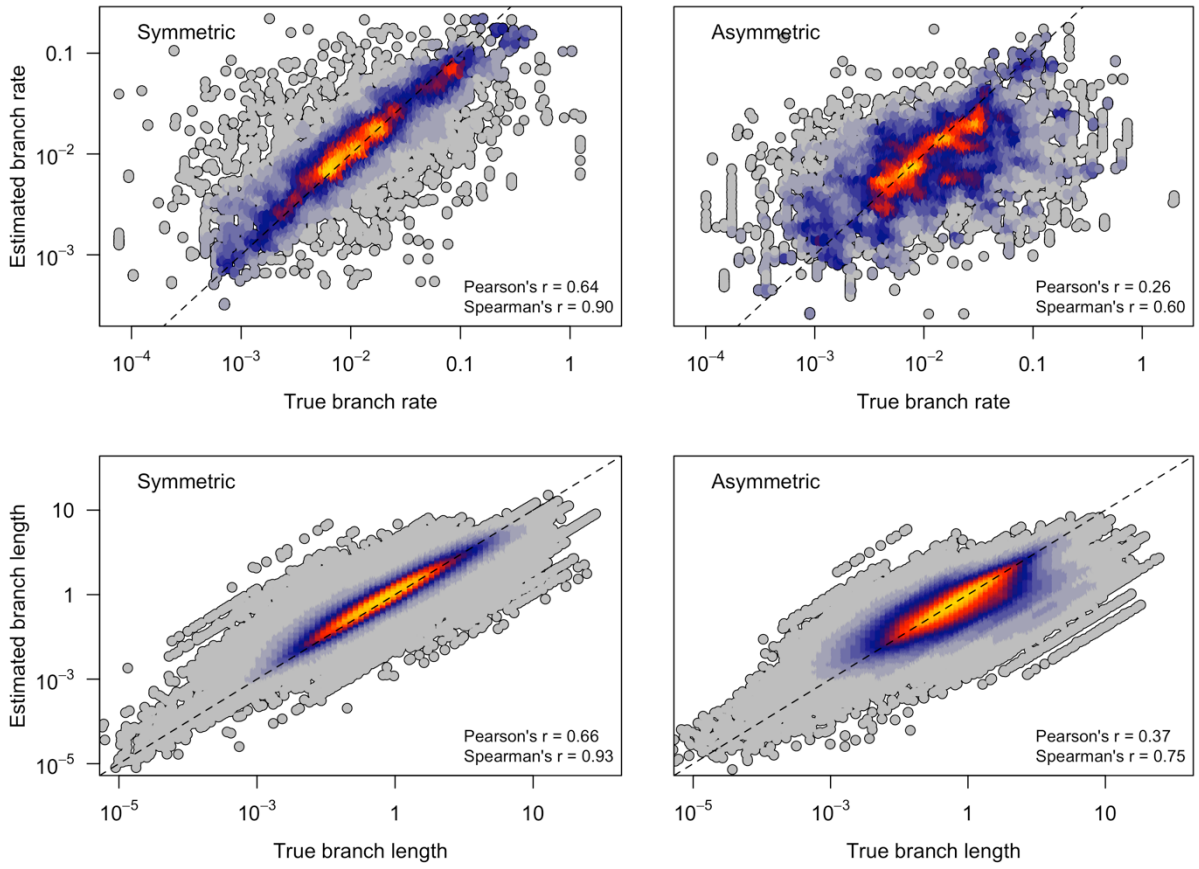

Figure 2.3 *Correlation of branch rate estimates with true values*

Correlation between true branch rates/lengths and BAMM-estimated branch rates/lengths when data are simulated under a symmetric rate-shift model (*left panels*) or an asymmetric rate-shift model (*right panels*). Branch rates (*top panels*) are measured as the expected number of character transitions per million years. Branch lengths (*bottom panels*) are measured as the expected number of character transitions occurring along the length of a branch. Each point depicts the average posterior rate estimate for a single branch of a phylogeny in one simulation, and the color of each point corresponds to the density of neighboring points (warm colors indicate high densities). Only branches from phylogenies with at least one simulated rate-shift are represented. The one-to-one line is dashed.

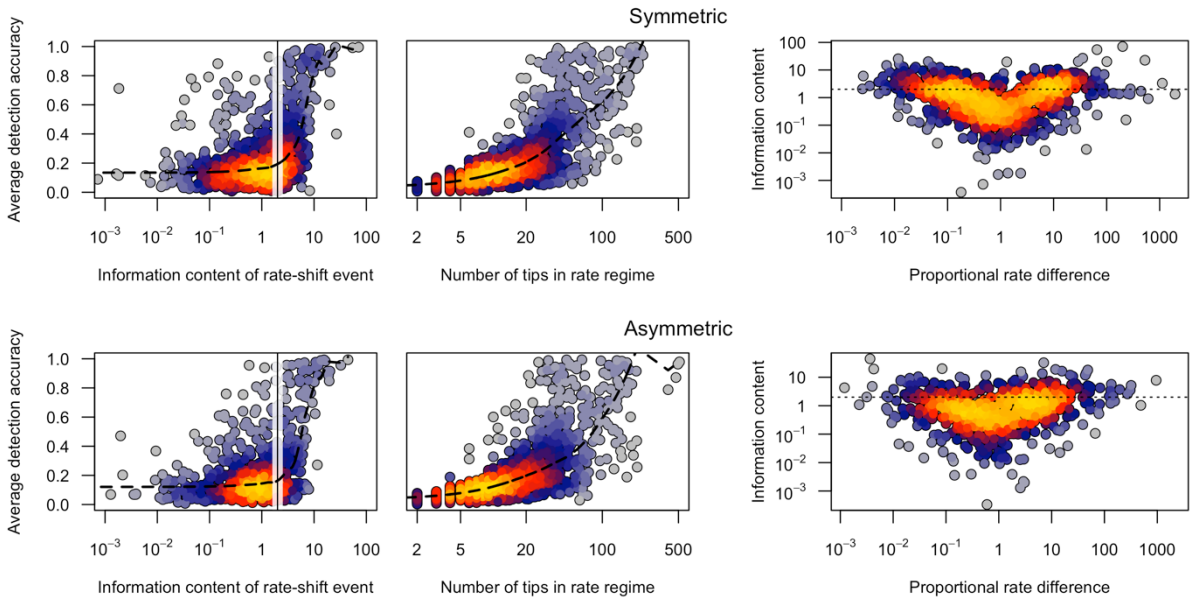

Figure 2.4 *Rate-shift detection accuracy*

Rate-shift detection accuracy with BAMM when data are simulated under a symmetric rate-shift model (*top panels*) or an asymmetric rate-shift model (*bottom panels*). The average detection accuracy measures how accurately BAMM inferred a rate-shift's location. The information content of a rate-shift event is a log-likelihood ratio that measures the likelihood of a given rate-shift event under the true parameters relative to the corresponding likelihood under a simple model where the rate is set to the whole-tree average. In the leftmost panels, the vertical line is the theoretical minimum log-likelihood ratio above which BAMM is expected to have power to detect a rate shift event, and it closely coincides with the upward inflection of the LOWESS regression lines (dashed). The rightmost panels plot an event's information content against its proportional rate difference. The proportional rate difference of an event is the ratio of its rate of evolution to the ancestral rate preceding it. In all panels, each point represents a simulated rate-shift event, and the color of each point corresponds to the density of neighboring points (warm colors indicate high densities).

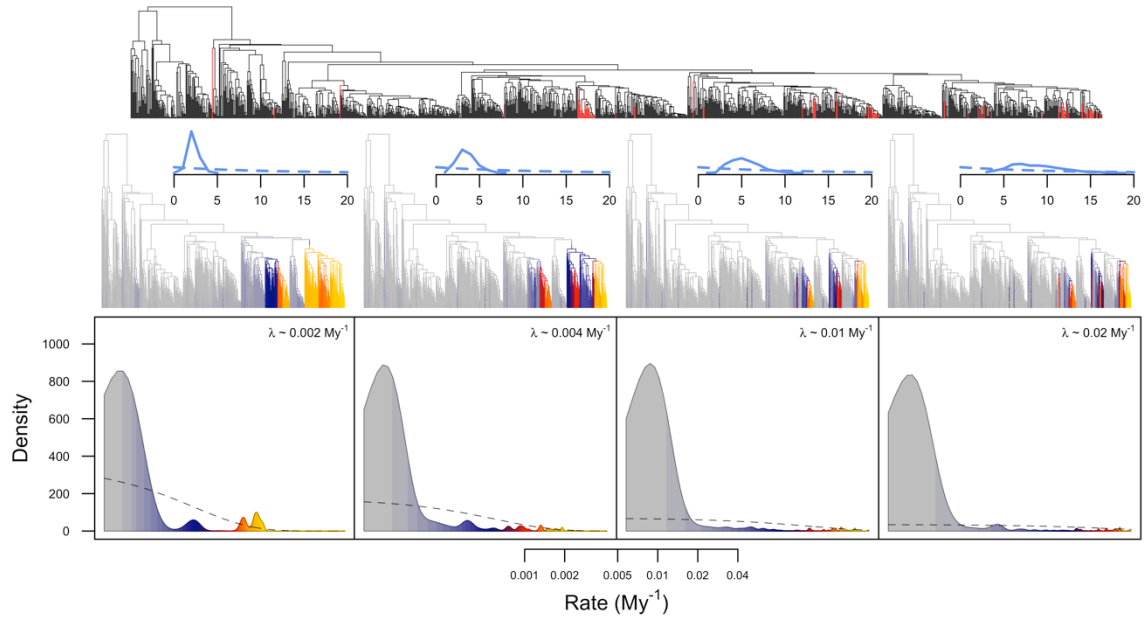

Figure 2.5 *Empirical estimates of rate variation in red-black banded coloration in snakes* Evolution of mimetic coloration in snakes under four different transition rate priors in BAMM. A high rate of evolution of red-black banded mimetic coloration in snakes is inferred in Neotropical dipsadine snakes and North American colubrines, although precise locations of rate-shifts and rate estimates differ across prior specifications. In the lower panels, the *x*-axis is the average posterior rate of trait evolution (expected number of character state changes per million years) and the *y*-axis is a kernel density estimate of the branch-specific rate distribution over 4 different prior specifications. The dashed line shows the prior distribution of the rate of evolution and the prior median is indicated by the inset expression. Average posterior branch rates are mapped onto the phylogeny in the center panels. The line graphs above each phylogeny show the posterior distribution of the number of rate-shift events (solid line) in the credible shift set compared to the prior distribution (dashed line). Branches colored red in the topmost panel show the distribution of red-black banded mimetic coloration among tips in the empirical phylogeny. Phylogeny from Pyron and Burbrink (2014).

Chapter 3
Parsimony-based Transition Rate Estimates Outperform Maximum Average Likelihood Estimates for a Two-state Character[3]

Abstract

Rates of character evolution in macroevolutionary datasets are typically estimated by maximizing the likelihood function of a continuous-time Markov chain (CTMC) model of character evolution over all possible histories of character state change, a technique known as maximum average likelihood. An alternative approach is to estimate ancestral character states independently of rates using parsimony and to then condition likelihood-based estimates of transition rates on the resulting ancestor-descendant reconstructions. We use maximum parsimony reconstructions of possible pathways of evolution to implement this alternative approach for single-character datasets simulated on empirical phylogenies using a two-state CTMC. We find that transition rates estimated using parsimonious ancestor-descendant reconstructions have lower mean squared error than transition rates estimated by maximum average likelihood. Although we use a binary state character for exposition, the approach remains valid for an arbitrary number of states. Finally, we show how this method can be used to

[3] Grundler, M.C. and Rabosky, D.L. (2020). Macroevolutionary analysis of discrete character evolution using parsimony-informed likelihood. *BioRxiv*, 897603, https://doi.org/10.1101/2020.01.07.897603.

rapidly and easily detect phylogenetic variation in tempo and mode of character evolution with two empirical examples from squamates. These results highlight the mutually informative roles of parsimony and likelihood when testing hypotheses of character evolution in macroevolution.

Introduction

Continuous-time Markov chains (CTMC) are commonly used in macroevolution to model the evolution of discrete characters to make inferences about evolutionary rates and patterns of change. Because the historical sequence of ancestor-descendant character state changes in most cases is not directly observed, statistical inference about Markov chain models is typically performed by summing over all possible histories of character evolution that could have resulted in the observed distribution of character states among living species using the well-known peeling algorithm (Felsenstein 1981). This technique, sometimes referred to as "maximum average likelihood" (Steel & Penny 2000), weights each possible evolutionary pathway by its probability of generating the observed data, and the resulting parameter estimates are therefore not conditioned on any particular history of character evolution. Pagel (1999) terms these "global estimators" and they are recommended as best practice by Mooers & Schluter (1999) and the only implementation available in most commonly used software packages in macroevolution (e.g., Paradis et al. 2004; Revell 2012; FitzJohn 2014).

An alternative approach to maximum average likelihood, termed "most-parsimonious likelihood", is to estimate both the unknown internal node states and the unknown transition rates simultaneously using likelihood (Barry and Hartigan 1987). Goldman (1990) refers to internal node states as "incidental parameters", since they are realizations of a random process, and raises concerns about the statistical consistency of such estimates (because the number of

parameters grows with the addition of new species to the dataset). Perhaps as a result, most-parsimonious likelihood techniques do not appear to be used in common practice. Alternatively, internal node states can be estimated independently (using parsimony), and likelihood estimates of transition rate parameters can be made by conditioning on these possible parsimonious pathways of evolution. In a macroevolutionary context, Janson (1992) appears to have been the first to use such an approach in a study of seed dispersal syndromes in plants. Some properties of these estimators were later studied by Sanderson (1993). The approach has received little attention since these two early works.

Here we revisit the approach of conditioning transition rate estimates on parsimonious ancestral state reconstructions, an approach we term parsimony-informed likelihood. We find that transition rates estimated using parsimony-informed likelihood have lower mean-squared error than rates estimated using maximum average likelihood and that this difference is accentuated in simulations where likelihood-ratio tests support an asymmetric model over a symmetric model of character evolution. We then show how the simple form of the parsimony-informed likelihood estimator leads to a rapid way of exploring datasets for phylogenetic variation in tempo and mode of character evolution.

Materials & Methods

General theoretical background

Given an assignment of character states to each node in a phylogeny the likelihood of the observed character state data at the terminal nodes is the product of state-to-state transition probabilities over all ancestor-descendant pairs. The full likelihood is computed by summing

these evolutionary pathway likelihoods over all possible histories of character evolution that could have given rise to the observed character state data using Felsenstein's (1981) peeling algorithm. We refer to transition rates estimated with this method as maximum average likelihood rates, abbreviated MAV. In the general multistate case, the transition probabilities used in the peeling algorithm are computed by numerically solving the matrix exponential $\exp(\boldsymbol{Q}t)$, where $\boldsymbol{Q}$ is the matrix of transition rates and t is the branch length separating ancestor and descendant. However, the simple two-state case like the kind used in this study permits an analytic solution given by,

$$p_{01}(t) = \frac{\varepsilon}{1+\varepsilon} - \frac{\varepsilon}{1+\varepsilon} e^{-\tau t}$$

$$p_{00}(t) = \frac{1}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon} e^{-\tau t}$$

$$p_{10}(t) = \frac{1}{1+\varepsilon} - \frac{1}{1+\varepsilon} e^{-\tau t}$$

$$p_{11}(t) = \frac{\varepsilon}{1+\varepsilon} + \frac{1}{1+\varepsilon} e^{-\tau t}$$

where $\tau = (q_{01} + q_{10})$, $\varepsilon = \frac{q_{01}}{q_{10}}$, and q_{01} and q_{10} are the forward and reverse transition rates, respectively.

Parsimony-informed likelihood begins by restricting attention only to the histories of character of evolution inferred with maximum parsimony (MP). There will often be more than one MP history and the set of MP histories will be a subset of the character histories used in the

computation of the maximum average likelihood. By treating these ancestor-descendant state reconstructions as "observations" sampled from a CTMC at random time points we can estimate the transition probability matrix of the discrete-time Markov chain embedded within the CTMC. This is,

$$\widehat{\boldsymbol{P}} = \begin{bmatrix} 1 - \hat{p}_{01} & \hat{p}_{01} \\ \hat{p}_{10} & 1 - \hat{p}_{10} \end{bmatrix}$$

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{ii} + n_{ij}} \quad i, j \in \{0, 1\}$$

Here, $\hat{p}_{ij}$ is the probability of a gain or loss and is calculated as the number of branches where a change from $i \rightarrow j$ occurred divided by the number of branches where such a change could have occurred. In practice, these probabilities are calculated by sampling MP histories of character evolution using Fitch's three algorithms (1971). We estimated this matrix by sampling 1000 MP histories such that each permissible state-to-state transition had uniform weight. In actuality, the number of MP histories can be much larger than this. If we further assume that the random time points at which we "observe" the CTMC are independent of the process of character evolution and distributed according to a Poisson process, we can use $\widehat{\boldsymbol{P}}$ to estimate the transition rate matrix $\boldsymbol{Q}$ of the CTMC. This is,

$$\widehat{\boldsymbol{Q}} = \hat{\lambda}(\boldsymbol{I} - \widehat{\boldsymbol{P}}^{-1})$$

Where $\hat{\lambda}$ is the estimated rate of the Poisson process that generates the time points at which we observe the chain's state. In this study, we take $\hat{\lambda} = \frac{N}{\sum_{i=1}^{N} t_i}$, the number of nodes divided by the summed branch length of the phylogeny. For a two-state model $\widehat{\boldsymbol{Q}}$ is simply,

$$\widehat{\boldsymbol{Q}} = \frac{\hat{\lambda}}{1 - \hat{p}_{10} - \hat{p}_{01}} \begin{bmatrix} -\hat{p}_{01} & \hat{p}_{01} \\ \hat{p}_{10} & -\hat{p}_{10} \end{bmatrix}$$

This estimator is derived by Karr (1991, pp. 384-386). It was used by Janson (1992) in a study of plant dispersal syndromes that remains a lucid but under cited discussion of the role of Markov models in macroevolutionary inference. Although we present results for only two states the method is applicable to any number of states (i.e., for K states $\hat{p}_{ij} = \frac{n_{ij}}{\sum_k n_{ik}} \quad i, j \in \{0, \ldots, K-1\}$). Note that $\hat{\lambda}$ does not influence the relative values of the estimated rate matrix. Only the pattern of transitions (including apparent stasis) affects the relative values. This is an important difference from the MAV estimator used in standard practice, which is also affected by variation in branch lengths over which transitions occur. We refer to rates estimated with this method as parsimony-informed likelihood transition rates, abbreviated PIL.

From a computational perspective PIL has a clear advantage over MAV because it requires only a single matrix inversion instead of repeated matrix exponentiation, which makes PIL more feasible for datasets with many character states. A disadvantage of PIL is that it can lead to undefined transition rates when change is rare enough that not all possible transitions are observed in parsimony reconstructions. A second potential advantage is that $\widehat{\boldsymbol{Q}}_{PIL}$ may be less variable and therefore less prone to error than $\widehat{\boldsymbol{Q}}_{MAV}$ because $\widehat{\boldsymbol{Q}}_{PIL}$ is conditioned on a minimal set of character state changes needed to explain the data. We explore this possibility in the simulations described below.

Simulations

We simulated 100 single-character datasets on each of 50 empirical phylogenies using a two-state CTMC model of discrete character evolution. To ensure simulated character datasets were evolved on phylogenies with realistic branch length distributions, we obtained empirical

phylogenies by randomly sampling clades containing between 100 and 1,000 terminal nodes from a recent phylogeny of ray-finned fishes (Rabosky et al. 2018). For each simulation we chose model parameters ε and τ such that the probability of state identity between every ancestor and descendant pair was at least one-half. In other words, simulations were parameterized such that a change of state from ancestor to descendant was never more probable than retaining the ancestral state. This property does not preclude simulations from having a fast rate of evolution, but it does require that transition rates become more symmetrical as the rate of evolution increases. For each simulated dataset we estimated ε and τ using the MAV and PIL estimators discussed above. When performing MAV estimates we used both an unconstrained and a constrained ($\varepsilon = 1$) model. Our rationale for fitting a constrained model to simulated datasets is that in practice researchers will often select an unconstrained model for analysis only when a likelihood ratio test indicates it substantially improves model fit over a constrained model. We conducted all simulations and parameter estimation in R 3.5.2 (R Core Team 2018) using the package macroevolution (available from https://github.com/blueraleigh/macroevolution).

Variation in tempo and mode

Because $\widehat{\boldsymbol{Q}}_{PIL}$ can be computed relatively rapidly it presents a simple method for detecting phylogenetic variation in tempo and mode of character evolution. First, compute $\widehat{\boldsymbol{Q}}_{PIL}$ for the entire tree and for each clade showing some minimum number of parsimony-inferred character state changes. If the minimum number of parsimony changes required for the test is too low rates cannot be estimated, but if it is too high there will be no opportunity to test for variation in tempo and mode. We used a value of 10 in our analysis. Second, for each of these clades compute the maximum average likelihood of their data using their clade-specific $\widehat{\boldsymbol{Q}}_{PIL}$ as

well as the tree-wide $\widehat{\boldsymbol{Q}}_{PIL}$. The magnitude of the log-ratio formed from these two likelihoods will be an indicator of how strongly the rate and pattern of character state changes in the clade differs from the rest of the phylogeny. For a two-state character the log-ratio should be at least 4 before concluding that a clade displays a significantly different tempo and mode of character evolution than the rest of the phylogeny. This is only a heuristic since we do not actually fit a model with multiple rate matrices. Its value can be derived by assuming the clade-specific log-likelihood ratio represents the difference in log-likelihoods between a model with and without a clade-specific rate-matrix and by requiring a minimum difference of zero in the AIC between these two models before accepting the model with the clade-specific rate-matrix.

Results

Across all simulations the mean squared error (MSE) in $\hat{\varepsilon}$ for MAV (MSE = 1.86) and PIL (MSE = 1.82) estimators was very similar (Fig. 3.1). However, comparing just the subset of simulations for which an unconstrained model outperforms a constrained model by a likelihood-ratio test reveals dramatic differences, with MSE for PIL estimates being substantially lower (MSE = 1.82) than for MAV estimates (MSE = 6.99) (Fig. 3.1). The cause of this difference is evidently from the tendency of the likelihood-ratio test to select a subset of simulations that show clearly biased MAV estimates. When comparing $\hat{\tau}$ across all simulations, unconstrained MAV (MSE = 1.38) and constrained MAV (MSE = 1.18) estimates had higher MSE than PIL (MSE = 0.0012) estimates (Fig. 3.2). These differences also persisted in the subset of simulations for which an unconstrained model outperforms a constrained model by a likelihood-ratio test, with unconstrained MAV estimates (MSE = 2.80) having the highest MSE, followed by constrained

MAV estimates (MSE = 0.35), and with PIL estimates (MSE = 0.00081) again achieving the lowest MSE.

When we used PIL to detect phylogenetic variation in tempo and mode in datasets simulated with no variation in tempo and mode, from 0% – 11% of simulations showed clades with significant departures in tempo and mode from the rest of the phylogeny. Pooling simulations across all phylogenies, only 2.6% of simulations detected significant phylogenetic variation in tempo and mode. Using stricter criteria (i.e. requiring more character state changes and a higher log-likelihood ratio) would further lower these percentages. These values indicate that the test does not readily identify variation in tempo and mode where none exists, although they do not provide insight into the converse question. When we applied the test to two previously analyzed empirical datasets it identified sets of clades showing variable tempo and mode that were in broad agreement with previous analyses that used computationally intensive MCMC techniques for the same purpose (Fig. 3.3, 3.4). We note that each of these two empirical applications of the test requires only a few seconds, permitting rapid exploration of datasets for clades with variable tempo and mode.

Discussion

In this study we show that maximum average likelihood estimates of transition rates for continuous-time Markov chain models of single-character evolution can be improved by using parsimonious ancestral state reconstructions to inform their calculation. We also show how this approach leads to a simple, fast, and effective means of exploring datasets for among-clade variation in tempo and mode of character evolution.

To see how combining parsimony and likelihood might be applied to other problems in macroevolutionary inference consider Pagel's (1994) test for correlated evolution. If we perform this test without using parsimonious ancestor reconstructions a single co-distributed event of character change can lead to a highly significant result (Maddison and FitzJohn 2015; Uyeda et al. 2018). On the other hand, because the data can be explained by a single co-distributed event of character change such a test using a parsimony-informed likelihood estimator would lead to undefined transition rates because (under parsimony) no reconstructed ancestors would have the requisite states (i.e., 01 and 10) needed to compute the transition probabilities. In some respects, the specter of undefined transition rates is an undesirable property of the PIL approach, but the explicit dependence of transition probabilities on independent evolutionary transitions is an important reminder of the need to consider phylogenetic replication when making statements about correlated evolution.

One of the reasons a parsimony-informed approach may be underutilized is that it conditions transition rate estimates on parsimony reconstructions that assume an equally-weighted cost matrix, and parsimony does not provide any criteria for justifying this assumption (Ree and Donoghue 1999). However, the use of simple parsimony makes the approach conservative. If we detect differences in transition rates among states based on reconstructions derived from a cost matrix that assumes there are no differences, we can reasonably conclude that differences exist (Maddison 1994). Furthermore, the alternative, which is to ignore parsimony, conditions transition rate estimates on histories of character evolution for which there is no empirical evidence, or which could be rejected on the grounds of implausibility. For many macroevolutionary datasets we are not interested in those histories as they are unlikely to affect our conclusions, and it is reasonable to ignore them in the computation of the likelihood. In some

cases, conditioning rate estimates on the universe of possible histories may actually be misleading. For example, if we attempt to reconstruct the reproductive mode of the most recent common ancestor of living squamates using a Markov model that ignores parsimony it favors a viviparous ancestor at almost two-to-one odds (Pyron and Burbrink 2014). This result is almost certainly incorrect given what we know about the biology of squamate reproductive mode (Shine 2014). The conventional reason given for the model's failure in this case is the presence of phylogenetic variation in tempo and mode that the model does not account for (King and Lee 2015). While this may be part of the reason, it is worth noting that if we fit the exact same model but condition transition rate estimates on parsimony reconstructions it now favors an oviparous ancestor at almost twenty-to-one odds, a result that is much more plausible but, from a (unconditioned) Markov chain model's perspective, not as probable.

So far, the discussion has not addressed why mean-squared error in transition rates is lower using parsimony-informed likelihood compared to maximum average likelihood. We suggest this is in part because conditioning rate estimates on parsimony reconstructions helps to reduce the variance in parameter estimates that comes from incorporating the contributions of all possible pathways of evolution. This seems to be most clearly the case for estimates of τ, the sum of forward and reverse transition rates (Fig. 3.2). The parsimony-informed likelihood estimates show a clear negative bias compared to the constrained maximum average likelihood estimates but also less variance, which drives overall mean-squared down. Interestingly, unconstrained maximum average likelihood estimates of τ show a strong positive bias and substantially more variance than either of the two other estimators. This positive bias has been noted before (Beaulieu & O'Meara 2016) and is an intrinsic property of the related discrete-time ML estimators studied by Sanderson (1993).

An unexpected result was the increase in mean-squared error we observed among the unconstrained MAV estimates when we examined only the subset of simulations where an unconstrained (asymmetric) model was favored over a constrained (symmetric) model by a likelihood-ratio test. This is cause for concern because it indicates that the rule we use to select a more complicated model over a simpler model also selects for situations that lead to greater estimation errors. Precisely why we observe a difference in performance of the two estimators is not known, but it appears that conditioning estimates on parsimony reconstructions acts to shrink the estimate of ε toward 1, which prevents the more extreme biased estimates observed with the MAV estimator.

Finally, we conclude with some remarks on how the results of this study might be affected by a different simulation procedure. Recall that all simulated datasets were generated such that the probability of a net state change from ancestor to descendant was never more probable than retaining the ancestral state. This does not preclude simulations from having a high rate of character evolution, but it does mean that when the rate of character evolution is high the generating model is close to symmetrical. In this case, the likelihood of the model is not influenced by the polarity of state changes and is a decreasing function of the number of character state changes (Tuffley and Steel 1997; Lewis 2001). This helps explain the good performance of parsimony-informed likelihood even when rates of character evolution are high. The possible pathways of character state change that the model assigns a high likelihood closely align with MP pathways. If we did not impose conservative behavior on our simulations it would create a situation where the outcome of a net state change from ancestor to descendant was more likely than the outcome of retaining the ancestral state for at least some, and potentially a large fraction, of the branches in a phylogeny. We do not believe that this is a plausible description of

the macroevolutionary process of character evolution given the tendency of researchers to study traits that are often fixed for large clades. Nonetheless, even in such a situation it is not immediately clear that the performance of the parsimony-informed likelihood approach discussed here would decrease relative to maximum average likelihood because the loss in phylogenetic signal engendered by such a process seems likely to increase the variance of parameter estimates made without the information added by parsimony.

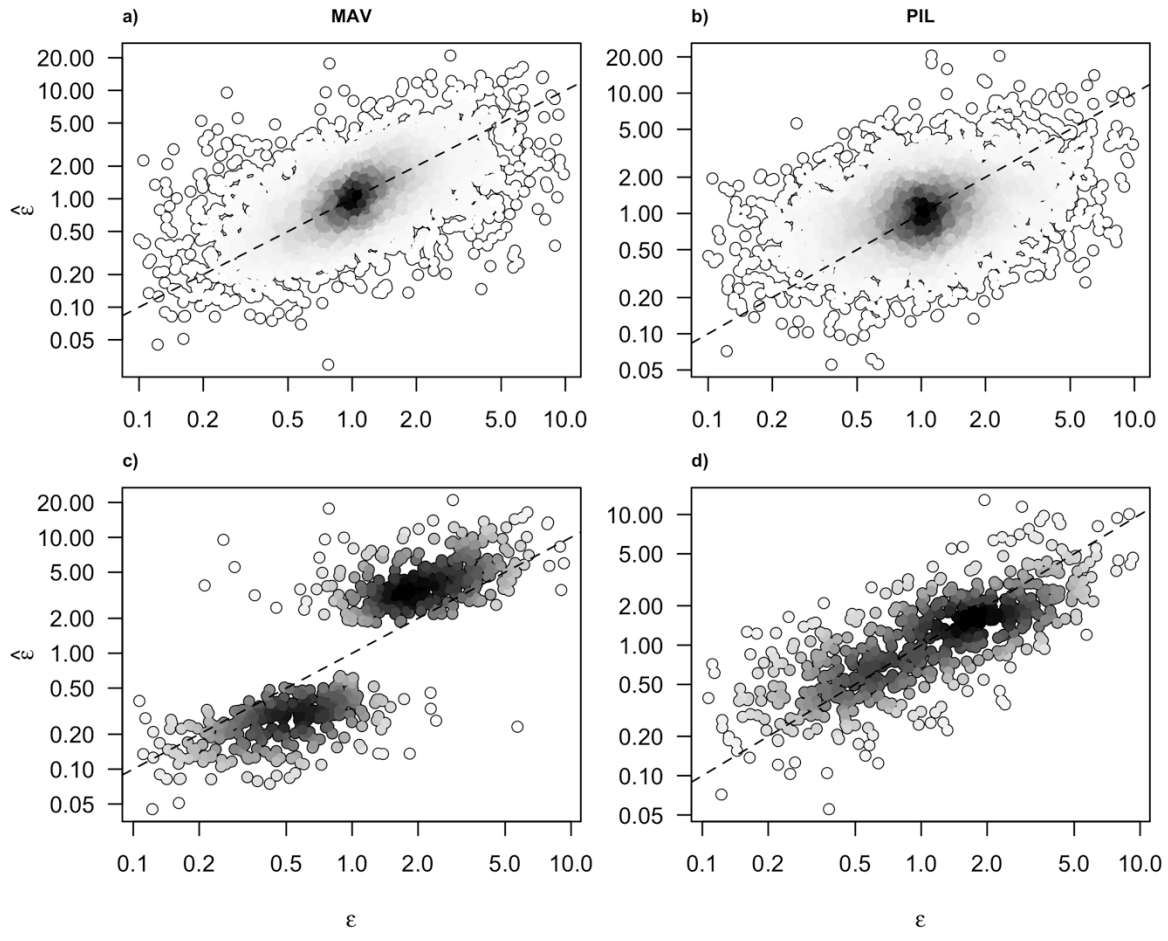

Figure 3.1 *Correlation between true and estimated rate ratios with maximum average likelihood and parsimony-informed likelihood*

Relationship between the true (ε) and estimated ($\hat{\varepsilon}$) ratio of forward and reverse transition rates for maximum average likelihood (MAV) and parsimony-informed likelihood (PIL) estimators. Darker shading indicates a higher density of points. The top panels show estimates for the full set of simulations. The bottom panels show estimates for the subset of simulations for which a likelihood ratio test determined an unconstrained model offered a significantly better fit than a constrained ($\varepsilon = 1$) model.

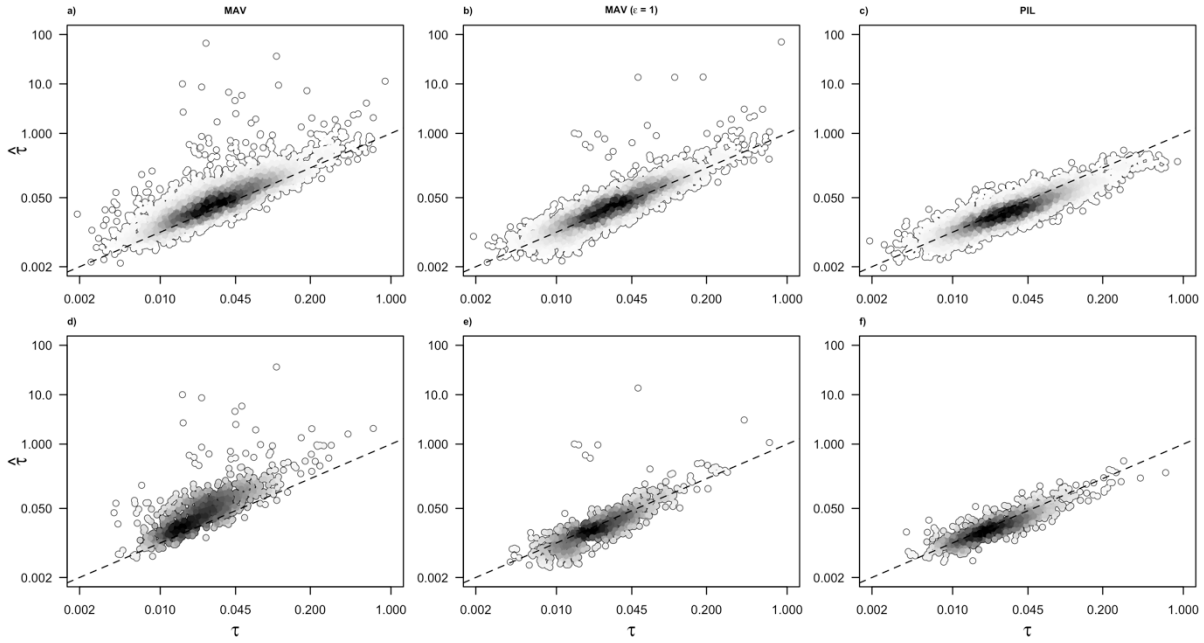

Figure 3.2 *Correlation between true and estimated rates with maximum average likelihood and parsimony-informed likelihood*

Relationship between the true (τ) and estimated ($\hat{\tau}$) sum of forward and reverse transition rates for maximum average likelihood (MAV) and parsimony-informed likelihood (PIL) estimators. Darker shading indicates a higher density of points. The top panels show estimates for the full set of simulations. The bottom panels show estimates for the subset of simulations for which a likelihood ratio test determined an unconstrained model offered a significantly better fit than a constrained ($\varepsilon = 1$) model.

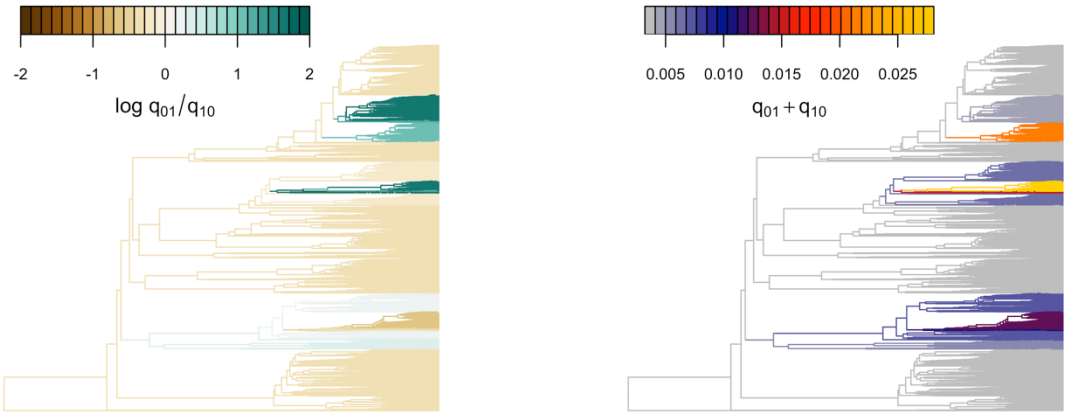

Figure 3.3 *Variation in tempo and mode of squamate reproductive mode detected with parsimony-informed rate estimates*

Evolutionary gains and losses of viviparity show variable tempo and mode across the squamate tree of life. Rate matrices were estimated using maximum likelihood conditioned on histories of character evolution inferred with parsimony for clades with at least 10 parsimony-inferred changes. Each clade-specific rate matrix was then compared to the rate-matrix for the whole tree. If the log-likelihood of the data in the subtree formed by a clade was at least 4 log-likelihood units greater under its clade-specific rate matrix than under the tree-wide rate matrix its branches were colored according to the parameters of that matrix. Data from Pyron and Burbrink (2014).

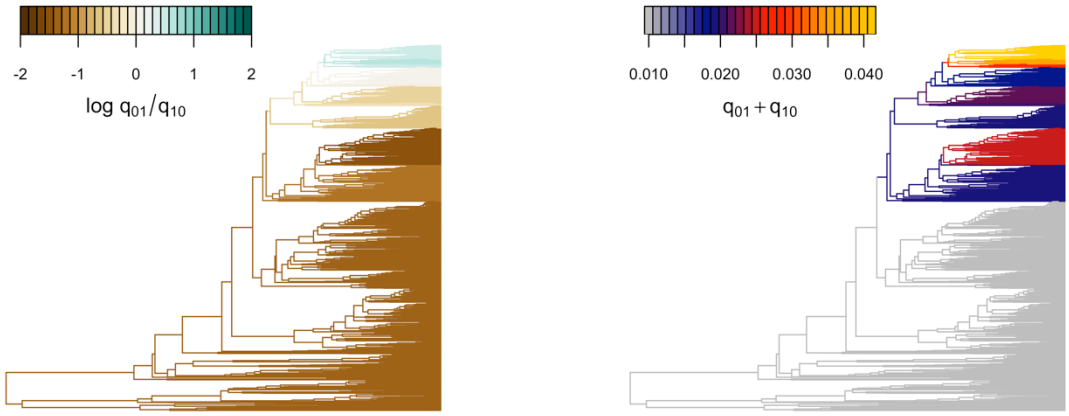

Figure 3.4 *Variation in tempo and mode of red-black banded coloration detected with parsimony-informed rate estimates*

Evolutionary gains and losses of red-black banded coloration show variable tempo and mode across the snake tree of life. Rate matrices were estimated using maximum likelihood conditioned on histories of character evolution inferred with parsimony for clades with at least 10 parsimony-inferred changes. Each clade-specific rate matrix was then compared to the rate-matrix for the whole tree. If the log-likelihood of the data in the subtree formed by a clade was at least 4 log-likelihood units greater under its clade-specific rate matrix than under the tree-wide rate matrix its branches were colored according to the parameters of that matrix. Data from Davis Rabosky et al. (2016).

Chapter 4
Complex Ecological Phenotypes on Phylogenetic Trees: A Markov Process Model for Comparative Analysis of Multivariate Count Data[4]

Abstract

The evolutionary dynamics of complex ecological traits – including multistate representations of diet, habitat, and behavior – remain poorly understood. Reconstructing the tempo, mode, and historical sequence of transitions involving such traits poses many challenges for comparative biologists, owing to their multidimensional nature. Continuous-time Markov chains (CTMC) are commonly used to model ecological niche evolution on phylogenetic trees but are limited by the assumption that taxa are monomorphic and that states are univariate categorical variables. A necessary first step in the analysis of many complex traits is therefore to categorize species into a pre-determined number of univariate ecological states, but this procedure can lead to distortion and loss of information. This approach also confounds interpretation of state assignments with effects of sampling variation because it does not directly incorporate empirical observations for individual species into the statistical inference model. In this study, we develop a Dirichlet-multinomial framework to model resource use evolution on

[4] Grundler, M.C. and Rabosky, D.L. (2020). Complex ecological phenotypes on phylogenetic trees: a Markov process model for comparative analysis of multivariate count data. *Systematic Biology*, accepted, https://doi.org/10.1093/sysbio/syaa031

phylogenetic trees. Our approach is expressly designed to model ecological traits that are multidimensional and to account for uncertainty in state assignments of terminal taxa arising from effects of sampling variation. The method uses multivariate count data across a set of discrete resource categories sampled for individual species to simultaneously infer the number of ecological states, the proportional utilization of different resources by different states, and the phylogenetic distribution of ecological states among living species and their ancestors. The method is general and may be applied to any data expressible as a set of observational counts from different categories.

Introduction

Most species in the natural world make use of multiple, categorically-distinct types of ecological resources. Many butterfly species use multiple host plants, for example (Ehrlich & Raven 19645/14/20 9:07:00 AM; Robinson 1999). Insectivorous warblers in temperate North America use multiple distinct microhabitats and foraging behaviors (MacArthur 1958), as do honeyeaters in mesic and arid Australia (Miller et al. 2017). The evolution of novel patterns of resource use can impact phenotypic evolution (Martin & Wainwright 2011; Davis et al. 2016), diversification (Mitter et al. 1988; Givnish et al. 2014), community assembly (Losos et al. 2003; Gillespie 2004), and ecosystem function (Harmon et al. 2009; Bassar et al. 2010). Consequently, there has been substantial interest in understanding how ecological traits related to resource use evolve and for exploring their impacts on other evolutionary and ecological phenomena (Vrba 1987; Futuyma & Moreno 1988; Forister et al. 2012; Price et al. 2012; Burin et al. 2016).

Making inferences about the evolutionary dynamics of resource use, however, first requires summarizing the complex patterns of variation observed among taxa into traits that can

be modeled on phylogenetic trees. Typically, this is done by explicit or implicit projection of a multidimensional ecological phenotype into a greatly simplified univariate categorical space (e.g. “carnivore”, “omnivore”, “herbivore”). It is widely recognized that the real-world complexities of resource use are not adequately described by a set of categorical variables (Hardy & Linder 2005; Hardy 2006). Nonetheless, it is also true that major differences in resource use can sometimes be summed up in a small set of ecological states, a point made by Mitter et al. (1988) in their study of phytophagy and insect diversification. For this reason, continuous-time Markov chain (CTMC) models, which require classifying species into a set of character states, have become commonplace in macroevolutionary studies of ecological trait evolution (Kelley & Farrell 1998; Nosil 2002; Price et al. 2012; Hardy & Otto 2014; Cantalapiedra et al. 2014; Burin et al. 2016). CTMC models describe a stochastic process for evolutionary transitions among a set of character states and are used to infer ancestral states and evolutionary rates, and to perform model-based hypothesis tests (O’Meara 2012). Many complex ecological traits, however, will be poorly characterized by a set of univariate categorical states. Consequently, species classified in single state can nonetheless exhibit substantial differences in patterns of resource use, creating challenges for interpreting evolutionary transitions among character states as well as for understanding links between character state evolution and diversification.

An additional limitation of currently used continuous-time Markov chains for modeling resource use evolution emerges from the fact that species are classified into ecological states without regard for the quality and quantity of information available to perform the classification exercise. As an example, species with few ecological observations might be classified as specialists for a particular resource, when their apparent specialization is strictly a function of the small number of ecological observations available for the taxon. More generally, by failing to

use a statistical model for making resource state assignments, we neglect a major source of uncertainty in our data: the uneven and incomplete knowledge of resource use across different taxa. This uncertainty, in turn, has substantial implications for how we project patterns of resource use onto a set of resource states. By failing to account for uneven and finite sample sizes characteristic of empirical data on resource use we cannot be certain if state assignments reflect true similarities or differences in resource use or are merely the expected outcome of sampling variation.

Consider the simple example in Figure 4.1, consisting of four species and three resources. Panels (a) and (b) illustrate the true resource states and their phylogenetic distribution across a set of four species and their ancestors. Here, an ancestral specialist evolved a generalist diet via a single transition (panel b), such that there are two extant species with the ancestral specialist diet (species X and Y) and two with the derived generalist diet (species P and Q). Panel (c) illustrates the observation process: empirical observations on diet are collected for each taxon, which are influenced by sampling variation. For each taxon, the data consist of a set of counts, each count recording the number of times the taxon was observed to use a specific resource. In panel (d), these empirical observations are used to classify each species into one of two diet states, based on the sampled relative importance of the three food resource categories (e.g., panel c). These relative importance estimates are based on uneven, and in some cases quite small, sample sizes, consistent with many empirical datasets (Vitt & Vangilder 1983; Shine 1994; Alencar et al. 2013). In panel (e) we imagine repeating the state assignment process on independent datasets while holding the samples sizes fixed to those in panel (c), which reveals that both the initial state assignments and the number of states from (d) are highly sensitive to real-world levels of sampling variation. This has obvious implications for downstream macroevolutionary analyses.

There is a serious risk that incorrect classification, and therefore spurious diet state variation, will emerge from sampling variation alone. In the analyses that underlie Figure 1, we find that more than 70 percent of tip state classifications do not match the true pattern of resource use.

Is this a problem in practice? This issue is difficult to assess because few studies provide information about the sample sizes that underlie state assignments. In most cases, ecological states are simply asserted as known. It is also important to emphasize that the specific problem in Figure 1 is an outcome of a more general problem: standard CTMC models have a limited ability to model complex ecological phenotypes because of the assumption that states in the model are univariate categorical variables (Hardy & Linder 2005; Hardy 2006). While it is true that CTMC models operate on a countable state space, it is not true that the states of the system must be univariate categorical variables. In this paper we treat states as multivariate probability distributions to develop a model for studying the evolutionary dynamics of ecological resource use on phylogenetic trees that avoids the need to classify taxa into univariate categorical states.

Our approach is explicitly designed to model resource traits that are multidimensional and to account for uncertainty in ecological state assignments of terminal taxa arising from effects of sampling variation. Antecedents to our approach can be found in the use of hidden Markov models in macroevolution (Felsenstein & Churchill 1996; Marazzi et al. 2012; Beaulieu et al. 2013; Beaulieu & O'Meara 2016; Caetano et al. 2018), which assume a deterministic many-to-one mapping from a set of hidden states to a set of observable states. By contrast, we assume that each hidden state is a multinomial distribution and that observed data are sampled outcomes from these distributions (see panels (a) through (c) of Fig. 1). The number of states in the model and the states themselves are not directly observed and are estimated from the data. Using simulations and an empirical dataset of snake diets, we show how the method can use

observational counts from a set of discrete resource categories to simultaneously infer the number of resource states, the proportional utilization of resources by different states, and the phylogenetic distribution of ecological states among living species and their ancestors. The method is general and applicable to any data expressible as a set of observational counts from different resource categories.

Materials & Methods

General overview

We assume there is a discrete set of resource niches, which we refer to as "resource states" or, more simply, as "states". We use the term resource in a broad sense, with the understanding that it may refer to behavior, habitat, prey, predators, etc. Although the set of resource states is finite, we imagine that a state assigns a taxon a continuously-valued parameter that determines its proportional utilization of different resources. Taxa that share a resource state therefore exhibit similar patterns of resource utilization. The data for each taxon consist of a set of counts, with each count recording the number of times a specific taxon was observed to use a specific resource. We refer to different resources as "resource categories". The set of resource states and their associated distributions over resource categories are therefore not directly observed but must be estimated from empirical count data. We assume that resource states evolve along a phylogenetic tree according to a CTMC process and that empirical counts are observations drawn from a multinomial distribution associated with each resource state. Our approach is similar to phylogenetic threshold models that combine a probability model for the evolution of an unobserved variable and a probability model for sampling the observed data

conditioned on the set of unobserved variables (Felsenstein 2012; Revell 2014). The general approach of treating states as probability distributions is due to Baum and Petrie (1966).

Model description

Let $\boldsymbol{D} = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_n\}$ denote the data observed for a set of *n* related taxa. Each datum $\boldsymbol{d}_i = \{d_{i1}, \ldots, d_{iJ}\}$ records the number of times the *i*-th taxon was observed to use each of *J* possible resource categories. For example, in MacArthur's (1958) study on foraging behaviors in five species of wood-warblers, he counted the number of times he observed each species use different foraging orientations (radial, tangential, vertical). For Blackburnian warblers, observations (counts) for each of these behaviors were 11, 1, and 3; the resource vector for this species is thus d = {11, 1, 3}, for a total of 15 observations across all three "resource categories". We assume that each taxon belongs to one of *K* possible states, and we let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, where each $X_i \in \{1, \ldots, K\}$ denotes the state of a particular taxon. States are unobserved but are assumed to represent distinct resource niches so that $\boldsymbol{X}$ partitions $\boldsymbol{D}$ into subsets of taxa with homogeneous patterns of resource utilization. Our target of inference is the posterior distribution of resource state assignments $p(\boldsymbol{X}|\boldsymbol{D}) \propto p(\boldsymbol{D}|\boldsymbol{X})p(\boldsymbol{X})$. To that end we define a probability model for the data as

$$p(\boldsymbol{D}|\boldsymbol{X}) = \prod_{k=1}^{K} h(\{\boldsymbol{d}_{i|X_i=k}\})$$

where $h(\{\boldsymbol{d}_{i|X_i=k}\})$ is the marginal likelihood function of the data for taxa assigned to state k. We assume that

$$h\left(\{\boldsymbol{d}_{i|X_i=k}\}\right) = \int_{\boldsymbol{\theta}} \prod_{i|X_i=k} f\left(\boldsymbol{d}_{i|X_i=k}|\boldsymbol{\theta}_k\right) g(\boldsymbol{\theta}_k|\boldsymbol{\beta})\, d\boldsymbol{\theta}$$

where f is a multinomial density with parameter $\boldsymbol{\theta}_k$ and g is a Dirichlet prior on $\boldsymbol{\theta}_k$ with hyperparameter $\boldsymbol{\beta}$. Because the Dirichlet distribution is conjugate to the multinomial distribution this parameterization allows us to analytically marginalize over the unknown multinomial parameters. Concretely, this means that $h\left(\left\{\boldsymbol{d}_{i|X_i=k}\right\}\right)$ is a Dirichlet-multinomial density, which, ignoring multinomial coefficients associated with each $\boldsymbol{d}_i$, is equal to

$$h\left(\left\{\boldsymbol{d}_{i|X_i=k}\right\}\right) = \frac{\Gamma\left(\sum_{j=1}^{J}\beta_j\right)}{\Gamma\left(\sum_{j=1}^{J}\sum_{i|X_i=k}d_{ij}+\sum_{j=1}^{J}\beta_j\right)}\frac{\prod_{j=1}^{J}\Gamma\left(\beta_j+\sum_{i|X_i=k}d_{ij}\right)}{\prod_{j=1}^{J}\Gamma\left(\beta_j\right)}$$

Formally, this model equates a resource niche (state) to a multinomial density and thereby assumes that resource use observations for taxa belonging to the same state are independent and identically distributed. As a model for count data it is closely related to topic models of word composition in a collection of text documents (Blei et al. 2003; Yin and Wang 2014) and to population genetic models of allele frequency composition in a set of populations (Pritchard et al. 2000). The key difference here is the specification of a prior model for $\boldsymbol{X}$. Because taxa are related, $\boldsymbol{X}$ is the outcome of evolution and individual X_i are not independent. We model evolution as a CTMC process where the rate of change is the same between all states (i.e. there is no evolutionary trend in the model) but varies independently among lineages. Our prior model for $\boldsymbol{X}$ is then the familiar phylogenetic likelihood function (Felsenstein 1981). Following Huelsenbeck et al. (2008), we treat lineage-specific rates of evolution as nuisance parameters drawn independently from a Gamma prior with shape parameter α and rate parameter fixed to 1. This model induces the same distribution on $\boldsymbol{X}$ as a model where the number of expected state changes along a branch is the same for all branches. Steel (2011) refers to this as the ultra-common mechanism model (UCM) to mark its contrast with the no-common

mechanism model (Tuffley and Steel 1997) from which it derives. Concretely, this means that the transition probabilities used in the pruning algorithm to compute $p(\boldsymbol{X})$ are equal to

$$p_{qr} = \frac{1}{K} - \frac{1}{K}\left(\frac{K}{K-1} + 1\right)^{-\alpha}$$

when ancestor and descendant states differ and to

$$p_{qq} = \frac{1}{K} + \frac{K-1}{K}\left(\frac{K}{K-1} + 1\right)^{-\alpha}$$

when ancestor and descendant states are the same. The full specification of the model is therefore

$$\boldsymbol{X} \sim \mathrm{UCM}_K(\alpha)$$

$$\boldsymbol{\theta}_{k=1,\ldots,K} \sim \mathrm{Dirichlet}_J(\boldsymbol{\beta})$$

$$\boldsymbol{d}_{i|X_i=k} \sim \mathrm{Multinomial}_J\left(\boldsymbol{\theta}_k, \sum\nolimits_j d_{ij}\right)$$

Interpretation of hyperparameters

The hyperparameter α of UCM controls phylogenetic signal and equals the expected number of state changes along any branch separating ancestor and descendant. As $\alpha \to 0$, phylogenetic signal approaches 1 because descendants almost surely resemble their ancestors. As $\alpha \to \infty$, phylogenetic signal approaches 0 because a descendant's state becomes independent of its ancestor's state and resembles a random draw from a discrete uniform distribution. The hyperparameter $\boldsymbol{\beta}$ acts as a vector of pseudo-counts for each resource category. Larger values place more prior weight on observing a particular resource category in a set of counts than smaller values. This has implications for how the model makes decisions about when to separate taxa into different resource states. Differences among resource categories assigned a low prior

weight contribute more toward separating taxa into distinct resource states than differences among resource categories assigned a high prior weight.

Posterior inference

Inference under the model results in a sequence of samples of resource state assignments $\boldsymbol{X}$ for terminal taxa that approximate the posterior distribution $p(\boldsymbol{X}|\boldsymbol{D})$. Each sample partitions terminal taxa into some number $c(\boldsymbol{X}) \leq K$ of groups, and each such group implies a posterior distribution over the parameter $\boldsymbol{\theta}_k$ of the multinomial density that the group represents. Specifically, due to the conjugacy of the multinomial and Dirichlet distributions the posterior distribution of each $\boldsymbol{\theta}_k$ is also Dirichlet with parameter $\{\beta_1 + \sum_{i|X_i=k} d_{i1}, \ldots, \beta_J + \sum_{i|X_i=k} d_{iJ}\}$. Therefore, inference of $\boldsymbol{X}$ implicitly yields posterior information about both the number of distinct resource niches and the relative importance of each resource category to each resource niche. Furthermore, because $p(\boldsymbol{X})$ accounts for phylogenetic relatedness among terminal taxa, each sample of $\boldsymbol{X}$ induces a posterior distribution over ancestral resource states which we can explore using, e.g., stochastic character mapping (Nielsen 2002) or marginal ancestral state reconstructions (Schluter et al. 1997).

Influence of K

We define phylogenetic signal as the quantity

$$p_{qq} - p_{rq} = \left(\frac{K}{K-1} + 1\right)^{-\alpha}$$

which ranges from 0 to 1 and quantifies how much information a descendant's state provides about the state of its ancestor (Royer-Carenzi et al. 2013). A value of 1 means the

ancestor was almost surely in the same state while a value of 0 means the ancestor was just as likely to have been in a different state than in the same state. Viewed as a function of K, phylogenetic signal approaches the curve $2^{-\alpha}$ as $K \to \infty$. Compared to a smaller value of K, a larger value of K therefore results in greater phylogenetic signal under the prior model for any given value of α. Because a large phylogenetic signal imposes a relatively greater prior penalty on events of character state change than a small phylogenetic signal, increasing K is expected to decrease the range of values of $c(\boldsymbol{X})$ observed in posterior samples of $\boldsymbol{X}$ because closely related taxa will preferentially be placed into a shared resource state unless the data provide strong evidence against such a grouping. We note that the symmetry in the prior model means that the algorithmic complexity of evaluating $p(\boldsymbol{X})$ is related to $c(\boldsymbol{X})$ and not to K. This means that larger values of K may actually speed up computation relative to smaller values. As $K \to \infty$ the prior model approaches the random cluster model studied by Mossel and Steel (2004).

Implementation

A Bayesian implementation of the model is available as an R package from github.com/blueraleigh/macroevolution. A Gibbs sampler for $p(\boldsymbol{X}|\boldsymbol{D})$ is derived using results from Yin and Wang (2014) and Schadt et al. (1998) and presented in the Appendix. The hyperparameters α and $\boldsymbol{\beta}$ are updated using slice sampling (Neal 2003).

Simulation study

To illustrate application of the method we designed a simulation study using an empirical dataset on pseudoboine snake diets (Alencar et al. 2013). Our rationale for basing simulations on

an empirical dataset is to ensure that properties of the data used to evaluate performance of the method are consistent with real studies, especially the distribution of observations per taxon and the distribution of resource specialization. Pseudoboine snakes are common members of the squamate reptile communities found in lowland rainforests of South America. Predominantly terrestrial or semi-arboreal, these snakes mainly eat small mammals, lizards, and other snakes. The dataset includes 606 observations of prey items from 8 prey categories for 32 species of pseudoboine snakes. Per species sample sizes range from 1 to 56 observations. We reanalyzed these data using a 33-species pseudoboine phylogeny extracted from the posterior distribution of trees in Tonini et al. (2016). We chose to use a different phylogeny from the original authors because their phylogeny excluded many species represented in the empirical dataset by small sample sizes, including some with the potential to impact state reconstructions at deeper nodes (Fig. 4.2). Using a more fully sampled phylogeny allowed us to better explore the influence of sample size on parameter estimation. The original publication coded each species with at least 8 diet observations into a set of 5 specialist diet states and 1 generalist diet state. Species were considered specialists if the prey resource represented at least 70 percent of recorded prey items (as in our Figure 4.1). When we applied the resampling procedure illustrated in Figure 4.1 to this empirical dataset under the assumption that the original state assignments represented the "truth", we found that in approximately 20 percent of resampled datasets at least one original state (not always the same state) was not present and that in about 84 percent of cases at least one species was coded incorrectly (although overall coding accuracy was high, ranging from 0.77 to 1). Thus, this dataset illustrates some of the concerns raised in our introduction but is also well-sampled enough and shows enough variation to facilitate the estimation of separate multinomial distributions.

We first analyzed the empirical dataset assuming a prior model with $K = 20$ states, and we ran a Markov chain for 160000 iterations sampling every 128 iterations while holding the $\boldsymbol{\beta}$ hyperparameter at a constant value of $\{1, \ldots, 1\}$. We subsequently extracted the maximum a posteriori estimate of $\boldsymbol{X}$, which implied a total of 6 resource states, and set the multinomial parameters for each of these states to their posterior mean value. We then simulated 20 datasets at each of 7 different levels of phylogenetic signal (0.1, 0.3, 0.5, 0.6, 0.7, 0.8, and 0.9) under the UCM model assuming K = 2, 3, 4, 5 and 6 resource states, and used the empirically estimated multinomial parameters and the empirical sample size distribution to generate prey resource use observations for each dataset. We analyzed the 700 simulated datasets assuming a prior model with $K = 20$ resource states, and we ran each Markov chain for 160000 iterations sampling every 128 iterations, again keeping the $\boldsymbol{\beta}$ hyperparameter at a constant value of $\{1, \ldots, 1\}$.

To assess model performance, we compute a per-taxon accuracy index that measures how closely the multinomial density estimated for a taxon matches the multinomial density that generated the taxon's data. Specifically, for terminal taxa accuracy is computed as

$$T_i = 1 - \frac{1}{M}\sum\nolimits_m \sup\nolimits_j \left|\hat{\theta}_{ijm} - \theta_{ij}\right|$$

where $\hat{\theta}_{ijm}$ is the posterior mean estimate of the *j*-th multinomial proportion for the *i*-th taxon in the *m*-th posterior sample and θ_{ij} is the true proportion and the average is taken over *M* posterior samples. The summand measures the largest absolute difference in proportions assigned by true and estimated multinomial densities to the same resource category. When the estimated and true multinomial densities perfectly coincide, the summand will equal 0 and accuracy will be 1. We can compute a similar metric for internal nodes using marginal ancestral state reconstructions as

$$A_i = 1 - \frac{1}{M}\sum_m \sum_k w_{ik} \sup_j \left|\hat{\theta}_{ijkm} - \theta_{ij}\right|$$

where w_{ik} is the marginal posterior probability that the *i*-th internal node is assigned to resource state *k*.

Results

Results from the empirical dataset suggest at least five distinct trophic modalities among pseudoboine snakes with a posterior mode of six, corresponding to three highly specialized and three rather generalized diets (Fig. 4.3). Marginal posterior probabilities from ancestral state reconstructions reveal a scenario where a specialized diet of lizards was likely ancestral for the clade. Two subsequent dietary niche shifts occurred: one to a diet of lizards + mammals, and another to lizards + mammals + snakes. Interestingly, there is no evidence for a stepwise expansion of diet to first include mammals and then snakes. Rather, it appears these two states evolved independently: the ancestor of the lizard + mammal + snake diet is inferred to feed on lizards alone, rather than lizards + mammals. The lizard + mammal expansion occurred in the ancestor of the genus *Oxyrhopus*, and the lizard + mammal + snake expansion occurred in the ancestor of the clade containing *Pseudoboa* and *Mussurana*. Several lineages in this latter clade appear to have then undergone niche contractions to specialist diets of lizards (*Pseudoboa nigra*), lizard eggs (*Drepanoides anomalus*), and bird eggs (*Rhachidelus brazili*).

Analysis of simulated datasets show that posterior frequency distributions of the number of resource states among terminal taxa were highly concentrated or centered on the number of resource states used to generate simulated datasets, indicating that the data were informative about the number of distinct resource use patterns among terminal taxa (Fig. 4.4). For taxa with

moderate sample sizes, the posterior mean estimate of each taxon's multinomial density was highly similar to the true distribution in the generating model (Fig. 4.5). This high similarity persisted across varying levels of phylogenetic signal, but small sample sizes resulted in lower accuracy when phylogenetic signal was low, creating in a distinctly triangular distribution of accuracy scores for terminal taxa. By contrast, accuracy scores for internal nodes decreased monotonically with decreasing phylogenetic signal and showed a close correspondence to the lower boundary observed for terminal node accuracy scores (Fig. 4.6). Because terminal nodes are associated with observational data, their states can be reasonably well-estimated even when phylogenetic signal is low. When signal is low, however, it is very difficult to estimate states at internal nodes because the observational data at tip nodes contain little information about ancestral states in this case.

Discussion

We developed a comparative method for macroevolutionary analysis of multivariate count data. The method is general and may be applied to any data expressible as a set of observational counts from different categories. Such datatypes arise frequently in community ecology and behavior. Potential applications include the comparative analysis of diet, foraging behavior, activity patterns, and habitat preferences. The method is similar to standard continuous-time Markov chain (CTMC) models of phenotypic evolution but differs in several important respects. First, the number of states in the model and the states themselves are unobserved and must be estimated from empirical data on resource use. Second, each state is a multinomial distribution rather than a categorical variable. This latter property enables researchers to model complex multidimensional phenotypes that cannot readily be expressed in

univariate space without considerable loss of information. Moreover, this property allows the method to accommodate uncertainty in the state assignments of terminal taxa that arises from the effects of sampling variation.

Simulations revealed that the new method is generally able to determine the correct number of states and that it provides accurate estimates of the underlying multinomial distributions, both for terminal taxa as well as internal nodes. We designed simulations around empirical patterns of resource use in a dataset on snake diets (Alencar et al. 2013). Therefore, caution is warranted in generalizing the good performance observed in the current study to other datasets. In particular, performance of the model will depend on the idiosyncrasies of individual datasets, including the distribution of sample sizes and the distribution of overlaps in resource use among species. We expect that states represented by few observations will be difficult to infer, especially if those states show appreciable overlap with other states.

Our empirical analysis identified at least two feeding modalities among the set of species Alencar et al. (2013) recognize as "generalists": species that feed predominantly on snakes but that regularly eat lizards and mammals, and species that feed predominantly on mammals and lizards. Ancestral state estimates strongly suggest that each of these feeding modalities arose from a more specialized diet comprised almost entirely of lizards. This is in contrast to the results of Alencar et al. (2013), which imply that nearly all origins of specialized feeding modalities occurred from a generalist ancestor. Interestingly, the model finds evidence that the data-poor species *Mussurana bicolor* occupies a distinct dietary niche (cf. Fig. 4.3). This is because the relatively high proportion of amphibians in *M. bicolor* diet samples, although supported by few observations, is highly unlikely given the complete absence of amphibian prey items from its close relatives. However, because *M. bicolor* has such a poorly characterized diet

a more conservative prior model (larger K) can easily cause *M. bicolor* to fall back into the state of its close relatives. The model also fails to recognize the small mammal and snake "specialist" states recognized by Alencar et al. (2013) even though the species representing these states have larger sample sizes than *M. bicolor*. Our explanation for this is that both mammals and snakes are commonly observed prey items in the diets of species closely related to these putative mammal and snake specialists. As a result, the amount of data needed to recognize a specialist on one of these prey resources is substantially greater than what would be needed to recognize a specialist on a prey resource that is rarely or never observed in the diet of close relatives.

As currently implemented, the approach described here does not directly model evolutionary gains and losses or substitutions of different resources. Indeed, under the prior model no resource is ever truly absent from the reconstructed states (although its proportional representation may approach zero as values in $\boldsymbol{\beta}$ become small). This contrasts with biogeographic-type models that explicitly model resource use expansions, contractions, and substitutions (e.g. see Hardy (2017) for application of Ree and Smith's (2008) dispersal-extinction-cladogenesis model to binary encoded diet data). Although these types of changes are implicit in the sequence of reconstructed states derived from the model, future studies might want to explore how to combine more complex evolutionary models with the current model for count data. Nonetheless, the advantage of a simple evolutionary model is that it has broad scope. It would be possible, for example, to apply the current approach to continuous characters by keeping the same evolutionary model but changing the model for observations from a multinomial to a multivariate-normal distribution, which could then be applied to other data types used for quantifying resource use such as stable isotope ratios of carbon and nitrogen.

One challenge for comparative methods is their limited ability to model ecological phenotypes that cannot be neatly summarized by a single value (Hardy & Linder 2005; Hardy 2006). Recent years have seen progress in this direction for continuous traits, including models that accommodate intraspecific variation, function-valued traits, and other non-gaussian data (Ives et al. 2007; Felsenstein 2008; Evans et al. 2009; Jones & Moriarty 2013; Goolsby 2015; Quintero et al. 2015). The general approach developed here, where each state is a multinomial distribution rather than a categorical variable, extends this progress to traits like diet and habitat that are typically treated as univariate categorical variables. By placing an emphasis on individual natural history observations, the method draws attention to the central role such observations play in evolutionary biology (Greene 1986) and to the many remaining opportunities for developing comprehensive ecological databases that advance our understanding of biodiversity (Hortal et al. 2015).

Summary

We described a novel methodological framework for studying the evolutionary dynamics of complex ecological traits on phylogenetic trees. Previous approaches to this problem have assumed that ecological states are categorical variables and that species are monomorphic for particular states. We relaxed this assumption through the use of a CTMC model that treats ecological states as probability distributions from which observed data are sampled. Results from our model provide a much richer understanding of macroevolutionary patterns than past approaches and can help illuminate the "phylogenetic natural history" of particular systems (Uyeda et al. 2018). Although our method is designed for the analysis of multivariate count data,

we suggest that the approach of treating states as probability distributions has wide applicability and will greatly facilitate the comparative analysis of novel sources of ecological data.

Appendix

If we update $\boldsymbol{X}$ one terminal taxon at a time by sampling a new state from the proposal distribution $p(X_i = k|\boldsymbol{X}_{-i}, \boldsymbol{D})$, the Metropolis-Hastings acceptance ratio will equal 1. We derive this distribution below using results from Yin and Wang (2014) and Schadt et al. (1998) and present an algorithm for its implementation. First, we have

$$\begin{aligned} p(X_i = k|\boldsymbol{X}_{-i}, \boldsymbol{D}) &= \frac{p(\boldsymbol{D}, \boldsymbol{X}_{-i}, X_i = k)}{p(\boldsymbol{D}, \boldsymbol{X}_{-i})} \\ &\propto \frac{p(\boldsymbol{D}, \boldsymbol{X}_{-i}, X_i = k)}{p(\boldsymbol{D}_{-i}, \boldsymbol{X}_{-i})} \\ &= \frac{p(\boldsymbol{D}|\boldsymbol{X}_{-i}, X_i = k)}{p(\boldsymbol{D}_{-i}|\boldsymbol{X}_{-i})} \frac{p(\boldsymbol{X}_{-i}, X_i = k)}{p(\boldsymbol{X}_{-i})} \\ &\propto \frac{p(\boldsymbol{D}|\boldsymbol{X}_{-i}, X_i = k)}{p(\boldsymbol{D}_{-i}|\boldsymbol{X}_{-i})} p(\boldsymbol{X}_{-i}, X_i = k) \end{aligned}$$

The symbols $\boldsymbol{D}_{-i}$ and $\boldsymbol{X}_{-i}$ is refer to the data matrix and resource state vector with terminal taxon *i* excluded. The first proportionality follows because $p(\boldsymbol{D}, \boldsymbol{X}_{-i})$ can be written as $p(\boldsymbol{d}_i, \boldsymbol{D}_{-i}, \boldsymbol{X}_{-i}) = p(\boldsymbol{D}_{-i}, \boldsymbol{X}_{-i}|\boldsymbol{d}_i)p(\boldsymbol{d}_i)$. But $\boldsymbol{d}_i$ contains no information about $(\boldsymbol{D}_{-i}, \boldsymbol{X}_{-i})$ so $p(\boldsymbol{D}_{-i}, \boldsymbol{X}_{-i}|\boldsymbol{d}_i) = p(\boldsymbol{D}_{-i}, \boldsymbol{X}_{-i})$ and we simply ignore the constant $p(\boldsymbol{d}_i)$. The second proportionality follows because $p(\boldsymbol{X}_{-i})$ does not depend on *i* and hence is the same for all values of X_i. Yin and Wang (2014) derive $\frac{p(\boldsymbol{D}|\boldsymbol{X}_{-i}, X_i = k)}{p(\boldsymbol{D}_{-i}|\boldsymbol{X}_{-i})}$ which can be written as

$$\frac{p(\boldsymbol{D}|\boldsymbol{X}_{-i}, X_i = k)}{p(\boldsymbol{D}_{-i}|\boldsymbol{X}_{-i})} \propto \frac{\Gamma\left(\sum_j n_{kj\neg i} + \beta_j\right)}{\Gamma\left(\sum_j n_{kj\neg i} + d_{ij} + \beta_j\right)} \frac{\prod_j \Gamma\left(n_{kj\neg i} + d_{ij} + \beta_j\right)}{\prod_j \Gamma\left(n_{kj\neg i} + \beta_j\right)}$$

where $n_{kj} = \sum_{h|X_h=k} d_{hj}$ and $n_{kj\neg i} = n_{kj} - d_{ij}$. To compute $p(\boldsymbol{X}_{-i}, X_i = k)$ we make use of the recurrence relations described Schadt et al. (1998). Let *a(m)* denote *m*'s ancestor and let *s(m)* denote *m*'s sibling. We write *s(m)=r* and *s(m)=l* when *m* is a left, respectively right, descendant of *a(m)*. With $\boldsymbol{T}$ representing the entire phylogeny, we let $\boldsymbol{T}_m$ denote the subtree rooted at node *m* and we let $\boldsymbol{T}_{-m}$ denote the complementary subtree that results from pruning $\boldsymbol{T}_m$ from $\boldsymbol{T}$. The notation $\{X_h, h \in \boldsymbol{T}_m\}$ is used to indicate the set of resource states in $\boldsymbol{X}$ where *h* is a terminal node in $\boldsymbol{T}_m$. We implicitly augment the state vector $\boldsymbol{X}$ to $\hat{\boldsymbol{X}}$ so that it include states at terminal and internal nodes and use the notation $\hat{X}_m$ to refer to the state of any node. If *m* is a terminal node $\hat{X}_m = X_m$. Let $\boldsymbol{F}_m = \{F_m(1), \ldots, F_m(K)\}$ where each $F_m(r) = p(\{X_h, h \in \boldsymbol{T}_m\} | \hat{X}_m = r)$ and let $\boldsymbol{S}_m = \{S_m(1), \ldots, S_m(K)\}$ where each $S_m(q) = \sum_r p_{qr} F_m(r)$ and p_{qr} denotes the transition probability. $\boldsymbol{F}_m$ and $\boldsymbol{S}_m$ are computed with a postorder traversal of $\boldsymbol{T}$ using the standard peeling algorithm (Felsenstein 1981). Now let $\boldsymbol{U}_m = \{U_m(1), \ldots, U_m(K)\}$ where each $U_m(r) = p\big(\{X_h, h \in \boldsymbol{T}_{-m}\} \big| \hat{X}_m = r\big) p(\hat{X}_m = r) = \sum_q U_{a(m)}(q)\, p_{qr} S_{s(m)}(q)$. $\boldsymbol{U}_m$ can be computed with a preorder traversal of $\boldsymbol{T}$ after defining $\boldsymbol{U}_{ROOT} = \{\frac{1}{K}, \ldots, \frac{1}{K}\}$. When *m* is a terminal node the joint probability $U_m(r) = p\big(\{X_h, h \in \boldsymbol{T}_{-m}\} \big| \hat{X}_m = r\big) p(\hat{X}_m = r)$ can be written as $U_m(r) = p(\boldsymbol{X}_{-m} | X_m = r) p(X_m = r) = p(X_m = r, \boldsymbol{X}_{-m})$. Putting these threads together gives

$$p(X_i = k | \boldsymbol{X}_{-i}, \boldsymbol{D}) \propto \frac{\Gamma\big(\sum_j n_{kj\neg i} + \beta_j\big)}{\Gamma\big(\sum_j n_{kj\neg i} + d_{ij} + \beta_j\big)} \frac{\prod_j \Gamma\big(n_{kj\neg i} + d_{ij} + \beta_j\big)}{\prod_j \Gamma\big(n_{kj\neg i} + \beta_j\big)} U_i(k)$$

With $\boldsymbol{F}$, $\boldsymbol{S}$, and $\boldsymbol{U}$ precomputed from an initial configuration of $\boldsymbol{X}$ we use algorithm A.1 to update $\boldsymbol{X}$, which assumes each internal node has only two immediate descendants and that a preorder traversal always visits a left descendant before visiting a right descendant.

Algorithm A.1

for each node *i* in a preorder traversal of $\boldsymbol{T}$:

if $\boldsymbol{U}_{a(i)}$ is marked:

update $\boldsymbol{U}_i$

if $\boldsymbol{U}_i$ is marked:

update $\boldsymbol{S}_{s(i)=l}$

update $\boldsymbol{U}_i$

if *i* is a tip:

set $z = X_i$

for each resource category *j*:

set $n_{zj} = n_{zj} - d_{ij}$

sample $z^* \sim \frac{1}{W} p(X_i = k|\boldsymbol{X}_{-i}, \boldsymbol{D})$ with $W = \sum_r p(X_i = r|\boldsymbol{X}_{-i}, \boldsymbol{D})$

set $X_i = z^*$

if z^* does not equal *z*:

set $F_i(z) = 0$

set $F_i(z^*) = 1$

if *s(i)=r*:

mark $\boldsymbol{U}_{s(i)=r}$

else:

update $\boldsymbol{S}_i$

mark $\boldsymbol{F}_{a(i)}$

Algorithm A.1 ensures that whenever we visit a terminal node the local $\boldsymbol{U}_i$ needed to compute $p(X_i = k|\boldsymbol{X}_{-i}, \boldsymbol{D})$ are correct. When the first preorder traversal of $\boldsymbol{T}$ has finished $\boldsymbol{F}$ and $\boldsymbol{S}$ correctly specify the conditional likelihood arrays for the new $\boldsymbol{X}$. However, the update operations leave the joint likelihood arrays in $\boldsymbol{U}$ out of date even though they are locally correct for each individual update. Therefore, a second preorder traversal of $\boldsymbol{T}$ to update $\boldsymbol{U}$ is necessary before updating $\boldsymbol{X}$ again.

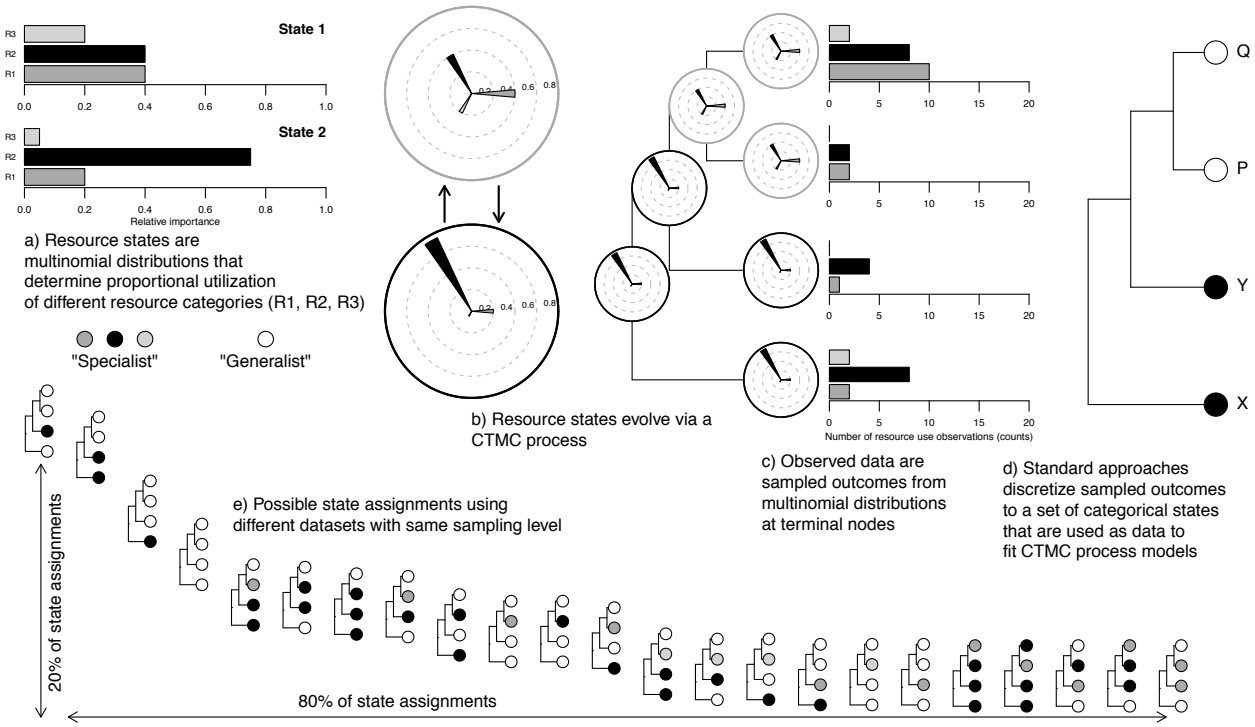

Figure 4.1 *Why neglecting sampling variation can mislead state assignments*

Distribution and representation of multivariate ecological phenotypes (a, b, c), data as sampled by researchers (c), and sampled states as typically represented by univariate categorical traits (d, e). Loss and distortion of information associated with complex phenotypes motivates the development of the Dirichlet-multinomial model described in this article. a) True resource states are multinomial distributions that determine the proportional utilization of three dietary resource categories by four species. Species X and Y are primarily specialists on resource category R2 whereas species P and Q exhibit generalized use of all three resource categories. b) The resource state of a species is the outcome of evolution via a CTMC process where the states correspond to multinomial distributions over a set of resource categories. Here, the multinomial distributions from (a) are represented as rose plots: the direction of a spoke identifies the resource category and the length of a spoke is equal to the proportional utilization of that category. The phylogeny depicts the true evolutionary history of change. c) Sampling: empirical data are sampled

outcomes from these multinomial distributions, and the number of observations can differ among species. Here, for example, species Q has 20 observations and species P has just 4. d) Univariate projection: typically, these multivariate outcomes are projected onto univariate categorical resource states, resulting in loss of information and sensitivity to sampling variation. In this example, a species is considered a "specialist" on a particular category if the sampled proportion of the category exceeds 0.7. Otherwise, it is considered a "generalist". In this case, the dataset and cutoff value align to match each species with its correct modal resource category. e) Simulation illustrating how univariate projection and sampling effects can lead to spurious variation in state assignments and incorrect evolutionary histories. The sampling and projection process illustrated above was repeated 1000 times, holding true resource distributions (a) and sample sizes (c) identical. State assignments are sorted along the *x*-axis according to their frequency of occurrence in the simulated datasets. Note that the procedure correctly matches all species with their modal food resource in a minority of cases and results in a variable number of states across datasets. The implication for macroevolutionary studies is that we cannot be certain whether state assignments are reflective of true patterns of resource use or are merely the expected outcome of sampling variation and projection.

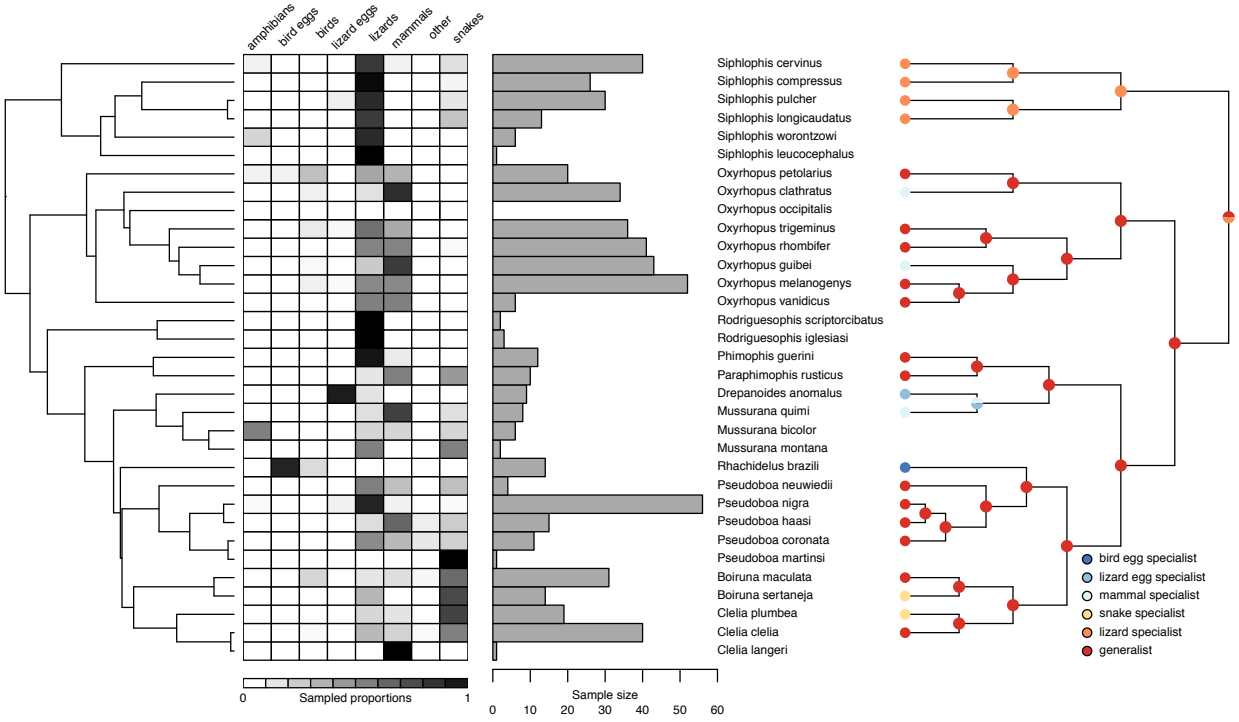

Figure 4.2 *Empirical dataset of pseudoboine snake diets and original character state encodings*

Summary attributes of the snake dietary dataset used to parameterize the simulation study, including phylogeny of pseudoboine snakes from Tonini et al. (2016) (left), and relative prey frequencies (middle left), total numbers of food observations per snake species (middle right), and cladogram with univariate ecological state encodings plus maximum parsimony reconstruction from Alencar et al. (2013) (right). Dark colors in the prey frequency matrix indicate higher sampled proportions of a particular prey item in a given diet. Our decision to use a different phylogeny from Alencar et al. (2013) was motivated by the observation that the original study excluded many species with low sample sizes, including species in phylogenetically informative positions with the potential to impact reconstructed evolutionary scenarios at deeper nodes (i.e., *Rodriguesophis*). In the present approach, ecological states are assumed to be multivariate probability distributions from which observed data are sampled, and

the model estimates these distributions from the empirical data. In this case, even species with small sample sizes are informative and are therefore not excluded.

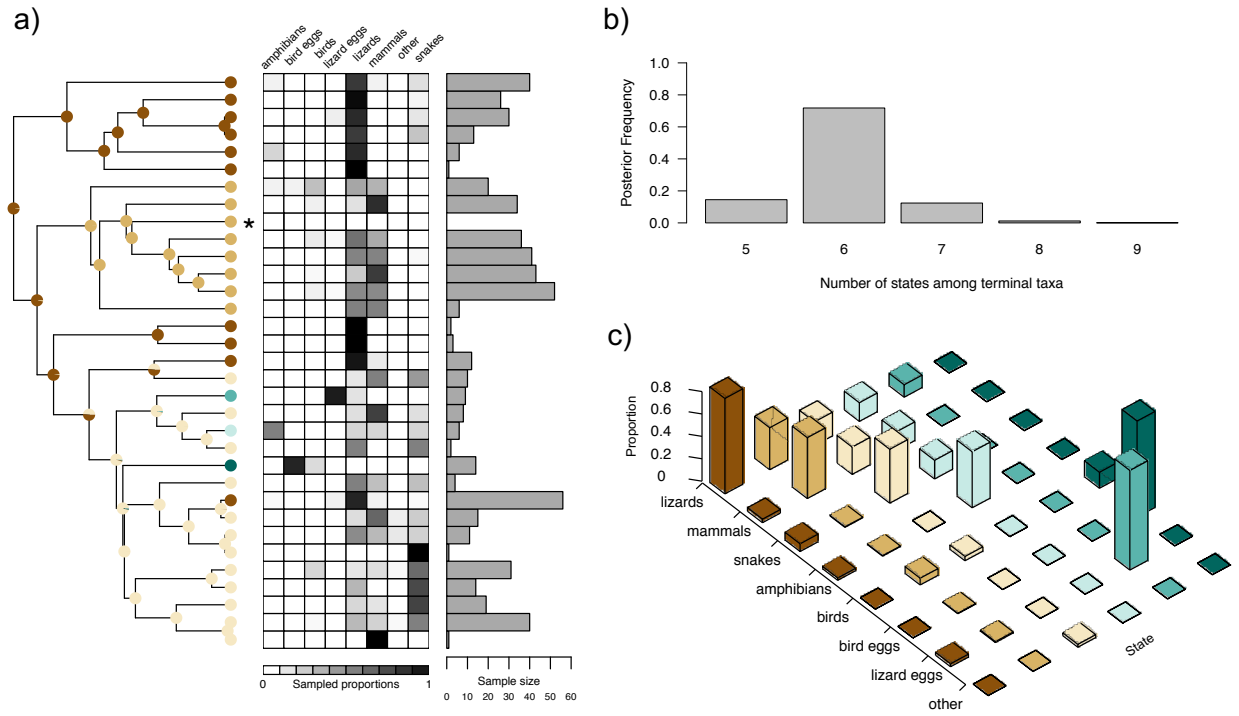

Figure 4.3 *Results of reanalysis of original pseudoboine dataset with model-identified diet states*

Reanalysis of the dataset presented in Figure 2 using the new model. a) The phylogeny and sampled diet observations were used to infer the diet states (colored circles) and their evolutionary history (marginal ancestral state probabilities shown as pie charts on the phylogeny). b) Posterior frequency distribution of the number of diet states among terminal taxa, computed from the posterior distribution of tip state assignments. The tip states in panel (a) depict the maximum a posteriori estimate from the posterior distribution of tip state assignments. Note that diet states are not observed directly, even at the tips of the tree; rather, all observed data are assumed to be sampled from a set of multinomial distributions. c) The maximum a posteriori estimate of the multinomial distributions for the diet states depicted on the phylogeny

in panel (a). Here, the model inferred 6 states, corresponding to 3 specialist (> 70% specificity for a single prey group) and 3 generalist diets. Note that the terminal node marked with an asterisk is missing data, and information about its probable diet state is drawn only from what the model has learned about the states of its neighbors and the likelihood of evolutionary change.

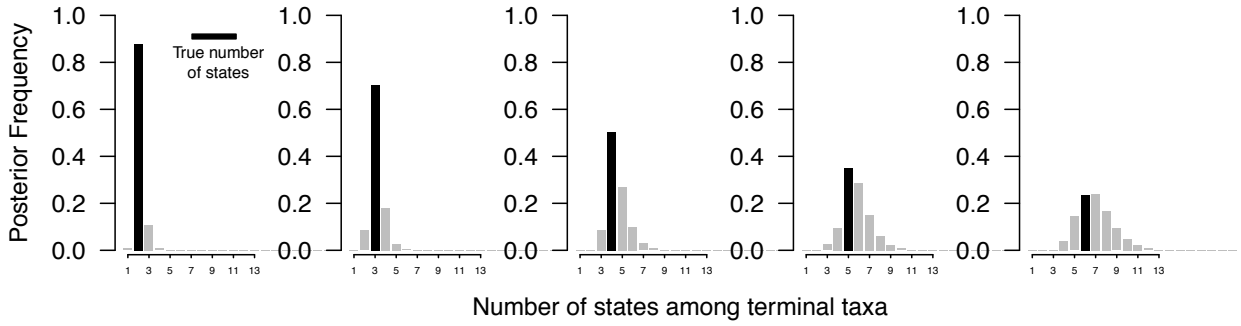

Figure 4.4 *Posterior distributions of the number of model-identified resource states*
Each bar chart depicts the posterior frequency distribution of the number of resource states observed among the set of extant taxa; black bars highlight the number of resource states in the generating model. In all cases the prior model assumed there were 20 resource states that terminal taxa could potentially occupy.

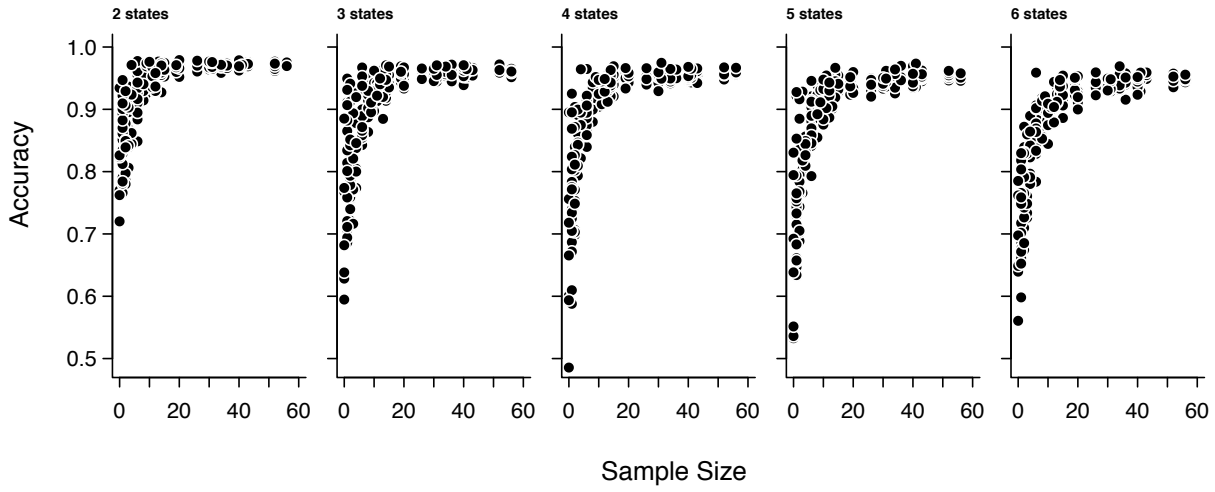

Figure 4.5 *Accuracy of estimated resource use distributions as a function of sample size*
Accuracy of model-inferred resource use distributions for terminal taxa as a function of the number of resource use observations available. Each point corresponds to a terminal taxon in one simulation. Accuracy is a measure of the difference between true and estimated multinomial densities. An accuracy of 1 means that the posterior mean estimate of the multinomial density and the true multinomial density assign precisely the same proportions to each resource category. See main text for more details. Accuracy is consistently high across all simulations for taxa with moderate sample sizes.

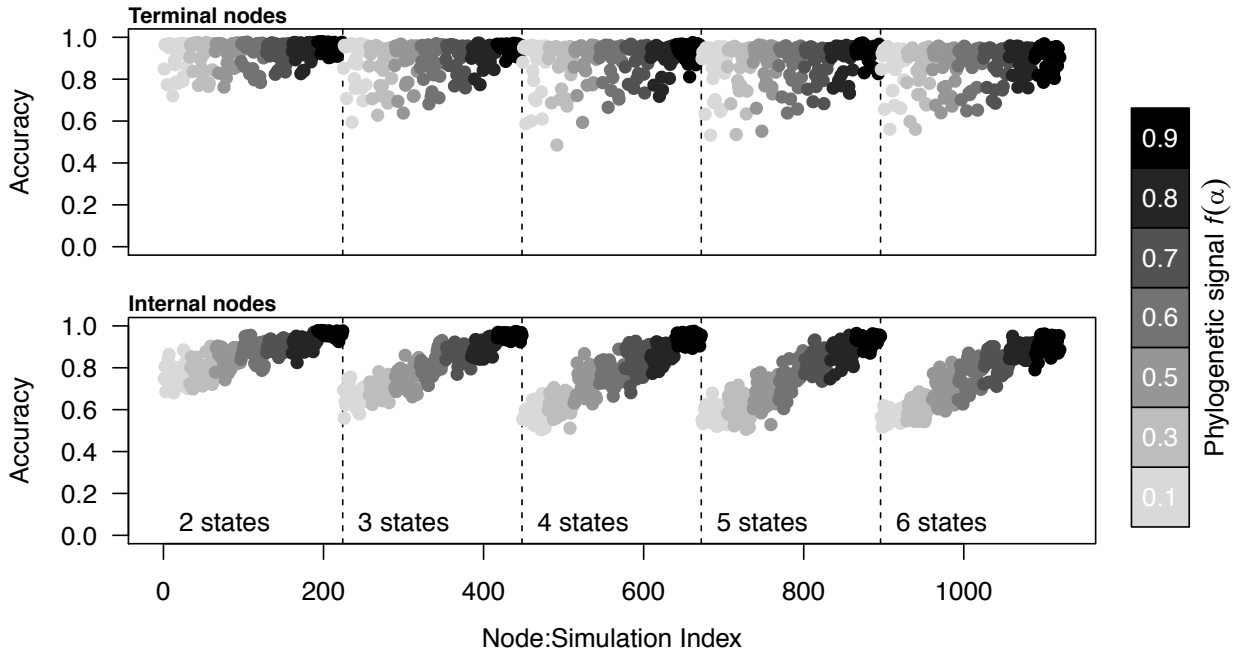

Figure 4.6 *Accuracy of estimated resource use distributions as a function of phylogenetic signal for terminal and internal nodes*

Accuracy of model-inferred resource use distributions for terminal taxa and internal nodes. Each point corresponds to a node in one simulation and points are colored according to the phylogenetic signal that was used to generate the simulated dataset. Accuracy is a measure of the difference between true and estimated multinomial densities. An accuracy of 1 means that the posterior mean estimate of the multinomial density and the true multinomial density assign precisely the same proportions to each resource category. In the case of internal nodes, differences are also averaged using the marginal posterior probabilities assigned to the different resource use distributions. See main text for more details. The triangular distribution of accuracy for terminal nodes is evidently caused by small sample sizes (cf. Fig. 5). Accuracy for internal nodes decreases monotonically with decreasing phylogenetic signal and appears to track the lower bound for accuracy of terminal taxa. Because tip nodes are associated with observational data, their states can be estimated with reasonable accuracy even when phylogenetic signal is

low. However, low signal makes it very difficult to reconstruct internal node states, as expected because tip states provide very little information about the states of their ancestors in this case.

Chapter 5
SquamataBase: A Natural History Database and R Package for Comparative Biology of Snake Feeding Habits[5]

Abstract

Public databases in taxonomy, phylogenetics, and geographic and fossil occurrence records are key research tools that provide raw materials, on which broad-scale analyses and synthesis in their respective fields are based. Comparable repositories for natural history observations are rare. Publicly available natural history data on traits like diet, habitat, and reproduction are scattered across an extensive primary literature and remain relatively inaccessible to researchers interested in using these data for broad-scale analyses in macroecology and macroevolution. In this paper, I introduce SquamataBase, an open-source R package and database of predator-prey records involving the world's snakes. SquamataBase facilitates the discovery of natural history observations for use in comparative analyses and synthesis and, in its current form, contains observations of at least 18,304 predator individuals comprising 1,227 snake species and at least 58,633 prey items comprising 3,231 prey taxa. To facilitate integration with comparative analysis workflows, the data are distributed inside an R

[5] Grundler, M.C. (2020). SquamataBase: a natural history database and R package for comparative biology of snake feeding habits. *Biodiversity Data Journal* 8: e49943.

package, which also provides basic functionality for common data manipulation and filtering operations. Moving forward, the continued development of public natural history databases and their integration with existing digitization efforts in biodiversity science should become a priority.

Introduction

Understanding how organisms interact with their environment lies at the heart of evolutionary biology and ecology. The data that furnish this understanding come from the practice of natural history. Careful observations about diet, habitat, reproduction, behavior, and a range of other ecological traits are vital, not only for a basic understanding of an organism's way of life, but also for a broad array of more general questions in evolutionary biology and ecology. This is not a new perspective (e.g. Greene 2005). Natural history is fundamental to our understanding of a broad variety of phenomena, from diversity gradients to adaptive radiation to community assembly (Futumya 1998, Stroud and Losos 2016). Yet despite this central role for natural history, the growth of online public repositories for natural history data lags far behind comparable repositories for other types of data, such as nucleotide sequences and geographic occurrence records. Whereas large, specimen-based databases are available for these latter data types, they are woefully lacking for natural history observations (but see databases in Toledo et al. 2007; Schalk and Cove 2018; and Jobe et al. 2019).

This is surprising owing to the fact that such observations ultimately furnish the raw data used to test and challenge theoretical predictions (Greene 1986). Unexpected or unusual natural history observations, often dismissed out of hand as “anecdotal”, can also reveal novel patterns and spur new lines of enquiry when carefully catalogued (Boero 2013). For example, researchers who analyzed thousands of anecdotal reports of unusual feeding behavior in birds discovered

that a clade's rate of behavioral innovation is positively correlated with the ability of species to expand their geographic range (Sol et al. 2002), as well as with a clade's species richness (Nicolakakis et al. 2003), lending support to the hypothesis that behavioral flexibility can drive accelerated rates of evolution and, more generally, to the idea that evolvability is an important driver of macroevolutionary patterns.

At a more fundamental level, publicly recorded natural history is essential for revealing the extent of our knowledge about the lives of other organisms. The widespread availability of field guides, carrying concise species accounts, can lead to the perception that much of the autecology of organisms is already known. This assumption is probably premature for the majority of life on Earth and our ability to identify knowledge gaps rests on the availability of natural history data (Hortal et al. 2015, Poisot et al. 2016). For example, after reviewing species accounts in a major compendium of mammal biology, only 38% of terrestrial mammals had recorded diet preferences (Kissling et al. 2014). The situation is undoubtedly worse for less charismatic groups of organisms.

Recognizing the importance of and the need for repositories of publicly recorded natural history, standards-based frameworks capable of aggregating natural history observations from diverse sources are beginning to emerge (Poelen et al. 2014). Ideally, such initiatives will help identify and fill shortfalls in our knowledge of biodiversity and facilitate the discovery of natural history observations for use in comparative analyses and synthesis. In practice, the limited number of providers that maintain high-resolution natural history datasets make the realization of these goals difficult.

Existing natural history datasets are generally derived from coarse summaries of the primary or secondary literature. For example, recent studies have used species accounts in major

compendiums of bird and mammal biology to assemble global-scale datasets on traits like diet and foraging mode and these data have been used to address a range of questions in macroecology and macroevolution (Kissling et al. 2014, Pigot et al. 2016, Price et al. 2012, Wilman et al. 2014). However, the coarseness of such datasets can mask patterns that are apparent at finer scales (e.g. Borries et al. 2013), potentially limiting our ability to identify knowledge gaps and to develop novel lines of inquiry and analysis into how patterns of intraspecific trait variability are related to patterns at broader interspecific scales.

Natural history observations, like geographic occurrences or nucleotide sequences, are inherently tied to individual organisms, but unlike these latter, data can seldom be queried and downloaded at a specimen-based level. In the sections below, I briefly introduce and describe SquamataBase, an open-source R package and specimen-based database of predator-prey observations involving the world's snakes.

Installation

The development version of SquamataBase is hosted on Github and can be installed with the aid of the package devtools from within R, using the command

```
devtools::install_github("blueraleigh/squamatabase")
```

The source code and commit history of the project can be viewed at

https://github.com/blueraleigh/squamatabase

Each stable release (including data and code) is also automatically archived with the Zenodo data repository. The current stable version is archived at

https://doi.org/10.5281/zenodo.3667777

Data Model

The core of SquamataBase is a database for storing data on specimens and trophic interactions between specimens. In the context of the SquamataBase data model, a "specimen" is a set of individual organisms (or components thereof) belonging to the same taxon (e.g. species, genus, family etc.). Each set has two measures of size (count and mass) and can be fleshed out with additional attributes if they are available, such as age, sex and body length. This generalized definition of a specimen to include multiple individuals is necessary because many publications present aggregate observations (e.g. 12 *Thamnophis sirtalis* ate 34 *Anaxyrus americanus* tadpoles), lacking individual-specific data. A generalized definition allows us to easily incorporate these observations alongside more specific observations.

A predator-prey interaction, or “food record" in SquamataBase terminology, is an observation of a snake specimen eating or attempting to eat a prey specimen. SquamataBase does not impose any particular categorization of prey specimens, instead it simply records their taxonomic identities as stated by the original authors. Categorization of prey specimens into a smaller number of groups for analysis is left to users (see below) because, in general, there will be many possible ways to categorize the original prey specimen taxonomic identities into a smaller number of prey types.

Each food record is linked to a reference publication where the data originate. Numerous contextual details are associated with a food record, including the basis for the record, whether the interaction was directly observed or inferred from evidence, the spatiotemporal context of the interaction, its outcome, and details regarding habitat, ingestion direction, and foraging strategy.

To ensure standardization, all taxonomic names reported in reference publications are matched against the taxonomy provided from the Catalogue of Life (Roskov et al. 2016).

Detailed documentation about each of the database fields, as well as the methods used to compile the data, are available in the package help documentation and can be accessed with the command `help(diet)`. In its current form, the database contains observations of at least 18,304 predator individuals comprising 1,227 snake species and at least 58,633 prey items comprising 3,231 prey taxa. These observations originate from a broad sample of geographic regions and phylogenetic lineages (Fig. 5.1).

Approximately 1,700 different scientific publications currently serve as the source of observations recorded in SquamataBase. Relevant publications were located through the use of keyword queries in academic search engines and by a systematic review of table of contents for well-known herpetological journals (e.g. Herpetological Review, Herpetology Notes). I also located additional relevant articles by consulting the references in reviewed articles. Every effort was made to ensure that the same observation, reported in two different publications, was not also duplicated in SquamataBase (e.g. Gaiarsa et al. 2013 and Ferreto Fiorillo et al. 2013 report on the same specimen of *Mussurana bicolor* preying on a watersnake). The majority of observations in the database result from papers describing (1) dissections of fluid preserved museum specimens and (2) direct encounters with snakes in the field that were actively consuming a prey or had recently consumed a prey item that could be regurgitated by forced palpation. Glaudas et al. 2017 have noted that these sources of information can provide different pictures of the prey spectrum for *Bitis arietans* (Puff Adder).

Filtering Records

SquamataBase provides functionality for filtering records by taxonomy and geography via the `filter_records` function. Taxonomic filtering can be performed on both predator and

prey. For example, filtering records to only include observations from the snake genus *Chironius* is performed as:

```
> diet <- filter_records(predator_taxon = "Chironius")
```

To constrain this record further, we can pass the returned object to `filter_records` again with an additional criterion. For example, if we only wanted records involving prey items of the frog genus *Scinax* we would do:

```
> diet <- filter_records(diet, prey_taxon = "Scinax")
```

Geographic filtering can be performed with country level administrative names or with a bounding box. For example, the following line constrains the existing record set to only include records from Ecuador and Peru:

```
> diet <- filter_records(diet, locality_adm0_name = c("Ecuador",
"Peru"))
```

Whereas the next line will constrain the existing record set to only include records lying between 80°W longitude and 60°W longitude and between 10°S latitude and the equator:

```
> diet <- filter_records(diet, xmin = -80, xmax = -60, ymin = -10,
ymax = 0)
```

Prey Classification

There are many ways to categorize prey items into different groups, but a relatively common categorization scheme is simply to use higher prey taxonomy. SquamataBase therefore provides two out-of-the-box categorization schemes that can be used to group prey specimens

into a relatively small number of prey types according to higher taxonomy. These two built-in schemes also serve as examples of how users may programmatically devise their own categorization schemes using the taxonomic metadata associated with each data record. The function that performs prey categorization is `group_prey` and we invoke it on a record set like so:

```
> diet <- group_prey(diet, grouping = "coarse")
```

If the argument "grouping" is a character mode, then it must be one of "coarse" or "detailed", which correspond to the two built-in categorization schemes alluded to above. In either case, the function returns a modified record set that contains an additional field identifying the prey category into which each prey specimen has been placed.

The `group_prey` function also allows users to define their own prey categorization scheme and pass it to the function through the grouping argument. In this case, the argument must be a named list of functions, each one of which must return either `TRUE` or `FALSE`. For each record in the record set, each function in the list is tried, in order, until a `TRUE` value is returned. The name of the first function that returns `TRUE` is then the name of the prey group applied to the record. Arguments to these functions are expected to be fields that are present in the record set to which the prey grouping is being applied. Users can study the two built-in examples by inspecting the function bodies for the commands `prey_coarse` and `prey_detailed`.

Aggregating Records

SquamataBase provides several options for aggregating records to create higher level summaries of the recorded prey items for snakes in a record set. These are available through the `aggregate_records` function. By default, the function will create a 3-column data frame with each row comprising a tuple of the form *(q, r, n)*, where *q* is a snake species, *r* is a prey group and *n* is the number of recorded instances of *r* appearing in the diet of *q*. The optional "by" argument to the `aggregate_records` function serves to disaggregate this default layout by specifying a set of additional fields to preserve as columns in the result. For example, invoking the command `aggregate_records(diet, by = "locality_adm0_name")` will return tuples of the form *(q, r, p, n)* and *n* is now the number of recorded instances of *r* appearing in the diet of *q* in country *p*. Due to the nature of the data, there are several ways the value for *n* can be computed, because each data record contains the number *npred* of predator and the number *nprey* of prey individuals involved in the trophic interaction. The default behavior of the function computes *n* by taking `min(npred, nprey)`, but this can be changed by the user through the use of function arguments.

Conclusion

Shortfalls in our knowledge of species interactions and species trait distributions pose significant challenges to the study and understanding of biodiversity (Hortal et al. 2015). Specimen-based natural history databases can help delimit knowledge gaps and provoke solutions for their resolution (Poisot et al. 2016). By developing SquamataBase, my goal is simultaneously to facilitate the discovery and reuse of natural history data in comparative analyses and to encourage researchers to continue to publish and make available their observations. There is considerable scope for expanding the development of specimen-based

natural history databases and integrating them with existing digitisation initiatives in biodiversity informatics and I suggest that this is a promising area in which to invest more effort.

Acknowledgements

A.R. Davis Rabosky suggested the name SquamataBase. D. Rabosky, I. Holmes, P. Title, M.R. Grundler, J. Larson, J. Shi, S. Singhal, R. von May, J. Mitchell, M. Harvey, A. Macdonald and an anonymous reviewer offered comments that greatly improved the manuscript.

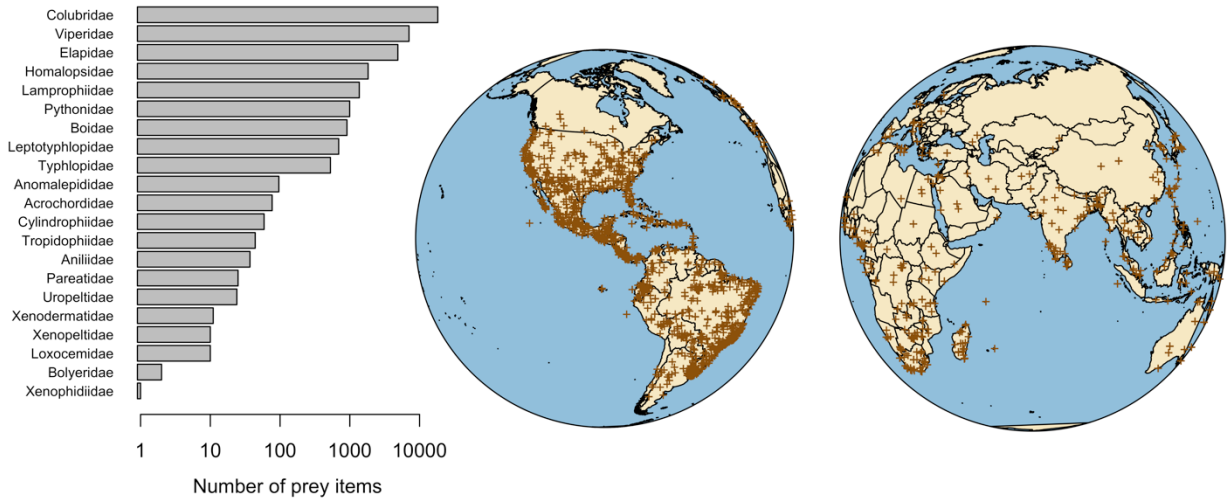

Figure 5.1 *Taxonomic and geographic distribution of prey records in SquamataBase*

Predator-prey observations recorded in SquamataBase originate from a broad sampling of geographic regions and phylogenetic lineages. The bar graph illustrates the number of prey items currently recorded from major snake families. Many of these observations are georeferenced and their locations are illustrated as marks on the orthographic projections.

Chapter 6
Diversification of Snake Diets Revealed by Primary Natural History Observations

Abstract

The origin of novel trophic ecologies underlies the diversification of many different animal groups and is frequently investigated at the macroevolutionary scale by using coarse character state encodings. These simple approaches seldom reflect the complex patterns evident in nature, posing a major challenge for constructing data-driven evolutionary scenarios of trophic niche evolution and ecological diversification. Primary natural history observations from the field and museum provide the most direct link to organismal ecology and can help meet this challenge but are seldom used in macroevolutionary studies, either because data are unavailable or because existing comparative methods are unsuitable. Here, we show how to use such observations for investigations of macroevolutionary tempo and mode by inferring evolutionary histories of trophic niche evolution across the global radiation of extant snakes. We assembled a dataset on snake diets (33,925 primary natural history observations of prey acquisition for 880 species) and used these data together with a novel analytical framework that allows us to directly integrate observations of prey use into the comparative inference model. Our results reveal repeated convergences of feeding strategies across the snake tree of life and differences in the

rate of trophic innovation associated with major lineages of snakes and colonization of new biogeographic theaters. Most feeding modalities in snakes appear after the K-Pg boundary during the Eocene, suggesting that ecological diversification was facilitated by new opportunities arising during the recovery and reorganization of biotic communities in the wake of mass extinctions. Our results demonstrate how fundamental observations of organisms in nature can be used to make quantitative inferences about the macroevolution of complex ecological traits and suggest new ways of integrating natural history data into comparative biology.

Introduction

Animal diversity harbors an enormous variety of foraging behaviors and dietary preferences. To make sense of this variety, biodiversity researchers group species into a smaller number of adaptive zones or ecological guilds based on broadly similar patterns of resource use (Simpson 1944; Root 1967; Van Valen 1971; Simberloff & Dayan 1991). This practice is fundamental to our understanding of the evolutionary history of animal feeding and ecological diversification (Losos & Mahler 2010). Reducing natural complexity into a small but meaningful number of groups has allowed researchers to address how shifts in trophic strategy influence the diversification of species and phenotypes and the assembly of species-rich communities (Mitter et al. 1988; Lovette et al. 2002; Seehausen 2006; Reddy et al. 2012).

The discretization of animal trophic diversity into a small number of groups, however, poses two primary challenges. First, the actual classification of taxa into groups is frequently guided by researcher experience and intuition rather than by quantitative data. This results in ambiguity and disagreement about where to place certain species when boundaries between groups are fuzzy, as they often are (Parravicini et al. 2020). Second, researcher concepts of

adaptive zones or ecological guilds are often deliberately simplified, focusing on first-order patterns of variation evident in nature. As a result, infrequent variation that is nonetheless important for understanding aspects of trophic diversity, such as why African cichlids with narrow diet breadths still exhibit functional and behavioral capacities for trophic generalism, is missed (Liem 1980; Golcher-Benevides & Wagner 2019).

One promising strategy for addressing these challenges is to directly incorporate primary natural history observations into comparative inference methods. In principal, such approaches accommodate greater natural complexity without sacrificing the useful idea of a supra-specific niche embodied in concepts of adaptive zones and ecological guilds, which then emerge as outputs of data-driven, phylogenetically informed analyses rather than as inputs. In practice, however, the rarity of comparative diet datasets that include firsthand observations of organisms in nature, and the limited data repertoire of comparative methods, means that most studies continue to rely on highly simplified character state encodings to describe trophic differences among species (e.g., Price et al. 2012; Cantalapiedra et al. 2014; Burin et al. 2016; but see, e.g., Vitt & Pianka 2005; Colston 2010; Hardy & Otto 2014; Forister et al. 2015).

In this study we describe the dynamics of trophic niche evolution across extant snakes, a global radiation of more than 3,500 species that accounts for over 10% of terrestrial vertebrate diversity. We assembled a dataset on snake diets (33,925 primary natural history observations of prey acquisition for 880 species) and used these data together with a new stochastic model-based comparative method that we developed to reconstruct evolutionary histories of dietary change from primary natural history data (Grundler & Rabosky 2020). The analytical framework we use here enables us to derive all results directly from observations as they might be recorded in field notebooks, without a priori categorization of snakes into highly simplified ecological states.

Snake evolution has given rise to a remarkable diversity of feeding habits, many of which are highly specialized and dramatically different from the diets of other squamate reptiles (lizards). While the ecological origins of early snakes are uncertain, a number of lines of comparative evidence indicate that the diversity of modern feeding styles is descended from a burrowing ancestor with a restricted gape and limited ability to ingest large prey (Greene 1983; Hsiang et al. 2015; Shine & Wall 2008; Wiens et al. 2006; Yi & Norell 2015). Subsequently, numerous functional innovations facilitated the evolutionary expansion of snake diets, including the origin of novel prey subjugation behaviors (Greene & Burghardt 1978; Jayne et al. 2002), highly kinetic skulls with complex musculature (Gans 1961; Cundall & Greene 2000), and sophisticated venom delivery systems (Kardong 1980; Savitzky 1980; Fry et al. 2008). These innovations dramatically increased the breadth of ecological opportunity available to snakes, ultimately leading to the many species we observe today.

We find that after an initial shift away from eating invertebrates, the diversity of snake feeding habits increased rapidly after the K-Pg boundary, with substantial increases in the rate of trophic innovation associated with more recent radiations resulting from the colonization of new biogeographic theaters. Our results demonstrate how fundamental observations of organisms in nature can be used to make quantitative inferences about the macroevolution of complex ecological traits and suggest new ways of integrating natural history data into comparative biology.

Results and Discussion

The merged phylogenetic and diet dataset contains 880 species representing 354 genera from nearly all snake families (the only exceptions being Anomochilidae and Gerrhopilidae).

Per-species sample sizes (number of observed prey items) range from 1 to 746 with a mean of 38 and a median of 12, while per-genus sample sizes range from 1 to 2749 with a mean of 96 and a median of 25, for a total of 33,925 observations (Fig. 6.1). Most observations in the database are from direct encounters with snakes in the field or from dissections of preserved museum specimens. Combining these two sources of data results in a more complete picture of the prey spectrum consumed by any given species, as field and museum specimens sometimes differ in relative frequencies of recorded prey types (Glaudas et al. 2017). Snake diets can vary within species, driven by sexual and ontogenetic differences in body size and by geographic variation in available prey types (Shine 1991). Our compilation records these details when possible, but the dataset used for analysis in the present study aggregates all records available for a given species, thereby creating a composite picture of the prey spectrum sampled by individual species.

The nature of the snake diet data poses analytical challenges because sampling is highly uneven across species, and the data matrix consists of unnormalized observational counts of sampled prey items with numerous zeroes. We analyzed the data with a new stochastic model-based inference method that uses phylogeny and the observed counts of sampled prey items to jointly infer continuous dietary niche states for each species and their unsampled ancestors (Methods). The resulting trophic network structure is informed by both the observed diet data and the phylogenetic relationships of sampled species, and these two sources of information allow us to incorporate observations from species with highly variable sample sizes because the model can use information from well-sampled phylogenetic relatives to estimate dietary niches for species with poorly characterized diets.

A striking visual aspect of the inferred trophic network structure is the variation in connectivity observed among different categories of prey (Fig. 6.2). Nearly all prey groups have

an associated set of specialist predators, but more generalized predators occur almost exclusively among snakes that feed on terrestrial vertebrates. The relative absence of generalized diets that include invertebrates and fishes may stem from the unique adaptations required to subdue and consume these prey and the constraints imposed by small body size and specific macrohabitat associations (Savitzky 1981). Even among more generalized species, however, sampled diets rarely include more than two or three distinct kinds of prey. Of these, there are clear tendencies for some combinations of prey items to co-occur more commonly in sampled diets than others, reinforcing prior concepts of snake feeding guilds (Ditmars 1912; Arnold 1972; Toft 1985). How these associations arise is uncertain. Proposed mechanisms include the correlated co-occurrence of prey items in the environment as well as chemical and functional similarity of exploited prey (Greene 1983; Cadle & Greene 1993; de Queiroz & Rodríguez-Robles 2006; Cooper 2008).

The inferred trophic structure suggests that vertebrate prey resources can be loosely arranged along a primary gradient with terrestrial endotherms (birds and mammals) on one end and aquatic ectotherms (fishes) on the other (Fig. 6.2). Along this gradient, terrestrial ectotherms (mainly frogs and squamates) occupy intermediate positions, with amphibians closer to fishes and squamates closer to birds and mammals. At the broadest scale, these associations are likely to be driven, in part, by effects of body size and macrohabitat. Aquatic snakes that prey on fishes regularly encounter frogs and tadpoles that rely on water for reproduction and larval development, for example, and only larger snakes can safely subdue and consume birds and mammals. However, the network structure also makes clear that no single prey resource gradient adequately summarizes dietary variation. Moreover, differences in sampled diets among many closely-sized snakes with similar macrohabitat associations indicate that trophic niche variation defies any simple explanation in terms of habitat and body size alone.

An interesting aspect of the inferred trophic network concerns terrestrial snakes that eat invertebrates and the extent to which their feeding habits differ from lizards. Most of the commonly recorded invertebrate prey items in sampled snake diets (annelids, mollusks, centipedes, spiders and scorpions, various insects and their larvae) appear rarely in sampled lizard diets or are recorded along with many other invertebrate prey groups, contrasting with the extreme level of specialization observed in snakes (Pianka 1986; Vitt & Pianka 2005). One possible explanation for the contrasting lizard and snake patterns is that many of the prey items used by snakes (centipedes and arachnids) are themselves dangerous predators that are important sources of mortality in squamate reptiles (Schalk & Cove 2018; von May et al. 2019), requiring large body size and venom to subdue (Rodríguez-Robles 1994). Likewise, several groups that are disproportionately used by snakes are heavily defended by shells (snails) and defensive mucosal production (slugs and annelids) and require specialized behaviors, dentition, and oral gland secretions to surmount (Sazima 1989; Rossman & Myer 1990; Hoso et al. 2007; Zaher et al. 2014). Although it is clear that most lizard diets do not resemble the diets of vertebrate-feeding snakes, more detailed comparisons of lizard and snake diets are clearly warranted to determine the extent to which the feeding habits of invertivorous snakes have diverged from lizards.

Our analysis reconstructs the most recent common ancestor of all snakes as feeding exclusively on insects with high probability, followed by an early shift to a diet comprised predominantly of lizards (Fig. 6.2). An insect-feeding ancestor makes sense in light of the molecular phylogeny: the earliest diverging snake lineages, scolecophidians (blind snakes), comprise a paraphyletic assemblage of species that feed almost exclusively on eusocial insects. Feeding on insects or lizards is inferred to have characterized many of the Mesozoic ancestors of extant snakes, which subsequently diversified into elaborate feeding modalities after the K-Pg

boundary during the Eocene, a time when squamate communities were beginning to recover from end-Cretaceous extinctions (Longrich et al. 2012) (Fig. 6.4). There is little direct evidence about the diets of ancestral snakes, but what exists is consistent with an early shift to vertebrate feeding. The controversial snake-like fossil *Tetrapodophis* (possible a marine varanoid lizard) from the Early Cretaceous contains unidentified vertebrate remains (Martill et al. 2015; Lee et al. 2016); the snake fossil *Sanajeh* from the Late Cretaceous may have eaten hatchling dinosaurs or their eggs (Wilson et al. 2010); an unidentified boa fossil from the Eocene contains a crocodilian (Greene 1983); an Eocene python fossil contains a lizard (Smith & Scanferla 2016). Beyond these few examples, inferences about ancestral snake diets rely primarily on phylogenetic comparisons among present-day species.

How morphologically and ecologically representative blind snakes are of early ancestral snakes is contentious. Some have considered them to be surviving representatives of the earliest snakes (Bellairs & Underwood 1951; Miralles et al. 2018), but others have argued that they are highly derived and cannot be considered morphological or ecological analogs of snake ancestors (Evans 1955; Chretien et al. 2019). The situation is further obscured by conflicting placements of blind snakes with respect to fossil snakes and other crown-group snakes across different datasets and analysis techniques (Garberoglio et al. 2019). Regardless, it is clear that the shift to vertebrate feeding happened early in snake evolution, maybe even facilitated by the consequent increase in gut volume resulting from adaptive morphological changes in response to fossorial habits (Shine & Wall 2008).

Our results imply that a striking diversity of trophic modalities, some quite distinct from one another, are inferred to have originated from a lizard-eating ancestor (Fig. 6.3). The preponderance of lizards in reconstructed ancestral diets stems, in part, from the fact that they are

the single most commonly recorded prey type in the current dataset. However, lizards are also abundant in the same terrestrial environments as snakes and their generally small body size compared to most snakes makes them suitable prey for a broad range of snake body sizes and gape widths. Indeed, many snakes that feed on birds and mammals as adults have juvenile diets comprised of lizards (Brito 2004), and lizards may have been the target of early selection during the shift to a vertebrate diet.

This finding also rejects the idea that many specialized feeding modalities originated from more generalized ancestors. In fact, dietary niche breadths are highly heritable (phylogenetically), suggesting that "specialists" and "generalists" are neither more nor less evolutionary labile than the other, and increases and decreases in reconstructed dietary niche breadths are equally common (Fig S6.2). This is not to say that no dietary shifts toward highly specialized feeding modalities were preceded by generalized ancestors. For example, many egg-eating snakes arose from ancestors inferred to occasionally eat eggs as part of a broader diet, consistent with previous findings (de Quieroz & Rodríguez-Robles 2006). However, in other cases such stepping-stone-like patterns are less clear. For example, 13 of the 15 recorded prey items for the Neotropical dipsadine *Rhachidelus brazili* are bird eggs, but no bird eggs are recorded in 139 prey items from the diets of its five closest relatives, which consist largely of lizards, snakes, and mammals. Such observations are speculative because in our method transitional diets can only be inferred for ancestors if they are preserved among living species. However, these results suggest the potential for occasional dramatic (rapid) dietary shifts, an inference that is supported by some observations from present-day snake populations. For example, the dipsadine snake genus *Pseudalophis*, present on the Galapagos Islands and the western coast of South America, preys predominantly on lizards and occasionally birds (Ortiz-

Catedral et al. 2019). However, at least one population of Galapagos snakes has taken to intertidal foraging on coastal fishes (Merlen & Thomas 2013), a behavior unknown from any other populations or close relatives that lends support to the claim that niche shifts are frequently initiated by changes in behavior (Mayr 1963).

Our analysis recovers numerous independent origins of similar feeding strategies across the global snake radiation. Notably, independent origins of specialized mammal-eaters first appear unambiguously in ancestral states with the most recent common ancestors of vipers, boas, and pythons during the Eocene, a time when rodents (the predominant mammals recorded in snake diets) were spreading and diversifying around the world (Wu et al. 2012) and consistent with prior suggestions that the rise of mammals, particularly rodents, provided ecological opportunity for the diversification of some snake clades (Maria Albino 1993; Rodríguez-Robles et al. 1999). Perhaps most remarkably, vermivory (earthworm feeding) has arisen independently in nearly all major snake lineages, including typhlopids (*Acutotyphlops subocularis*), xenodermids (*Achalinus*), pareids (*Xylophis*), viperids (*Atheris barbouri*), homalopsids (*Brachyorrhos*), elapids (*Toxicocalamus, Ogmodon*), uropeltids, natricids, pseudoxenodontids (*Plagiopholis*), dipsadids, and colubrids. Phylogenetic autocorrelation (a clustering metric) in the proportion of annelids in sampled snake diets is the lowest of all prey categories: despite a similar number of reconstructed gains, annelids in sampled snake diets show considerably lower phylogenetic clustering than vertebrate prey items like mammals and fishes (Fig. S6.3). Such differences hint at the possibility that feeding strategies differ in evolutionary accessibility and versatility, and earthworm feeding may be among the most evolutionary and ecologically accessible feeding strategies available to snakes.

Reconstructed ancestor-descendant diet sequences reveal evidence of elevated rates of change among colubroid snakes (after Zaher et al. 2009), a cosmopolitan clade comprising just over one-half of living snake diversity (Fig. 6.2). Some of the fastest rates of change are associated with colonization of the nearctic and neotropical regions by the families Natricidae and Dipsadidae. Macroevolutionary rates often show time dependency: rates calculated over shorter time intervals are higher than rates calculated over longer intervals (Gingerich 1983). This time dependency also affects rates calculated in this study (Fig. S6.1). The cause and significance of this pattern is debated, but in our case part of the explanation appears simple: rates calculated over longer time intervals include more lineages with no dietary changes, which drive net rates of change down. A similar explanation has been proposed to explain time dependency observed in diversification rates (Henao Diaz et al. 2019). However, we do not exclude the possibility that clade-level differences in dietary lability also play a role in driving rate variation. For example, owing to their greater diversity, colubroid clades show systematically higher net rates of change than non-colubroids even when controlling for overall phylogenetic lineage length (Fig. S6.1), suggesting the possibility of a general coupling between rates of lineage diversification and rates of trophic evolution. In spite of generally higher rates of trophic innovation in colubroids, however, nearly all feeding modalities observed in the current dataset are represented by lineages that diverged prior to the origin of colubroids, indicating that the colubroid mega-radiation has been facilitated more by an ability to exploit existing ecological opportunities rather than by invasion of previously inaccessible trophic niches.

Conclusion

We used a comprehensive dietary dataset and a novel stochastic framework to demonstrate how primary natural history observations can be integrated with stochastic-model based comparative methods to form quantitative inferences about macroevolutionary tempo and mode. Our results highlight the complex evolutionary dynamics of trophic niche evolution in snakes, and complement prior analyses emphasizing the deep historical roots of dietary differences observed in lizard (Vitt & Pianka 2005) and snake (Colston et al. 2010) communities. The data included in the present study represent the combined effort and observations of numerous field workers, yet there remain considerable opportunities for expanding the dataset to fill taxonomic and geographic sampling gaps and for methodological innovations to accommodate different sources of variation contributing to the tremendous diversity of snake feeding habits.

Materials & Methods

Data acquisition and prey categorization

We used a recent compilation of prey items observed in sampled snake diets to characterize feeding habits for as many snakes as possible (Grundler 2020). Observations in the database were compiled from an extensive review of the primary literature and subsequently categorized into 22 different prey groups according to higher level taxonomy. The analyzed dataset uses a composite picture of the prey spectrum sampled by individual species by aggregating records across different sources of intraspecific variation (e.g. age, sex, geography). This decision was motivated by the broad phylogenetic scope of the current study and uneven sampling across species, but there remain important opportunities for incorporating different

sources of intraspecific variation into comparative analyses of this sort. For comparative analysis we used the phylogenetic hypothesis from Tonini et al. (2016).

Probabilistic reconstruction of diet states and trophic niche evolution

We used a recently described comparative method to reconstruct the history of trophic niche evolution (Grundler & Rabosky 2020). Briefly, each extant species in a phylogeny can be characterized by a vector $\boldsymbol{x}$ that describes its proportional utilization of the 22 different diet categories. If we denote the full set of such vectors by the symbol $\boldsymbol{X}$, the fundamental goal of the method is to both estimate $\boldsymbol{X}$ and to extend $\boldsymbol{X}$ so that hypothetical ancestral species, represented by internal nodes of the phylogeny, are also assigned proportional utilization vectors. We denote such extensions by $\widehat{\boldsymbol{X}}$, and the method probabilistically reconstructs all $\widehat{\boldsymbol{X}}$ that are consistent with $\boldsymbol{X}$. By assumption, the proportional utilization vectors reconstructed for ancestors must be represented among the set observed in $\boldsymbol{X}$. This is enforced by assuming that $\boldsymbol{X}$ contains at most K unique proportional utilization vectors corresponding to distinct trophic niche states (the actual number discovered by the method may be less than K). Because the observed data consist only of counts, the method samples a range of $\boldsymbol{X}$ that confer high probability on the observed count data under a Dirichlet-multinomial sampling model with hyperparameter α. As a result, the method returns posterior distributions for both $\boldsymbol{X}$ and $\widehat{\boldsymbol{X}}$, thereby naturally accommodating uncertainty in both observed and ancestral trophic niche states. We used an uninformative Dirichlet prior (α=1) and set K=1000, a relatively high value that favors more parsimonious evolutionary histories due to the stronger penalty incurred with each event of character state change. Gibbs sampling was performed for 30,000 iterations, and every 10-th sample was recorded for a total of 3000

posterior samples. Posterior averages reported in the main text were computed using the last 500 samples. Likelihood and parameter traces are provided in the supplement (Fig. S6.5).

Estimating evolutionary fluxes and net rates of change

We estimated average evolutionary fluxes between diet categories for each branch in the phylogeny under an optimal transport model. Specifically, for a given ancestor-descendant configuration on edge *e = (u, v)* we computed the matrix $\boldsymbol{U}$ that transformed an ancestral diet $\hat{\boldsymbol{x}}_i(u)$ into a derived diet $\hat{\boldsymbol{x}}_i(v)$ (where $\hat{\boldsymbol{x}}_i(u) \neq \hat{\boldsymbol{x}}_i(v)$) at a minimum total cost $d_{\boldsymbol{U}}(\hat{\boldsymbol{x}}_i(u), \hat{\boldsymbol{x}}_i(v))$. Note that the rows of $\boldsymbol{U}$ must sum to $\hat{\boldsymbol{x}}_i(u)$ and the columns to $\hat{\boldsymbol{x}}_i(v)$. The optimal transformation cost is defined as

$$d_{\boldsymbol{U}}(\hat{\boldsymbol{x}}_i(u), \hat{\boldsymbol{x}}_i(v)) = \min_{\boldsymbol{U}} \sum_{jk} U_{jk} C_{jk}$$

where the sum runs over all pairs of diet categories. The elements of $\boldsymbol{U}$ describe how much proportional utilization of each diet category in an ancestor must be transformed into proportional utilization of each diet category in a descendant. The matrix $\boldsymbol{C}$ assigns a value to the cost of transforming a unit of one diet category into a unit of another and must be directly specified. Matrices $\boldsymbol{U}$ were computed for each branch using the Sinkhorn-Knopp algorithm (Cuturi 2013) and averaged over the posterior. The matrix $\boldsymbol{C}$ was considered fixed, with diagonal elements set to 0 and off-diagonal elements set to 1. With this specification, the elements of $\boldsymbol{U}$ can be thought of as the effective number of gains and losses that take place during an evolutionary transition from one diet to another. In the extreme case, when ancestor and descendant are pure specialists on different prey groups, only a single element of $\boldsymbol{U}$ is non-zero and that element will equal 1.

We estimated the net rate of evolution for each clade by calculating the average total number of gains and losses of prey categories within a clade divided by the total branch length of all lineages in the clade. We used these clade rates to assign each terminal node a tip rate, which we computed as the weighted average of all clade rates on the phylogenetic path leading from the tip to the root. The weights attached to each clade rate increase as they approach the root, which relaxes each tip rate back to the overall average clade rate computed from the entire phylogeny. The strength of relaxation depends on how quickly weights increase toward the root, but the overall qualitative picture of variation in tip rates is unchanged by different weighting schemes (Fig. S6.4).

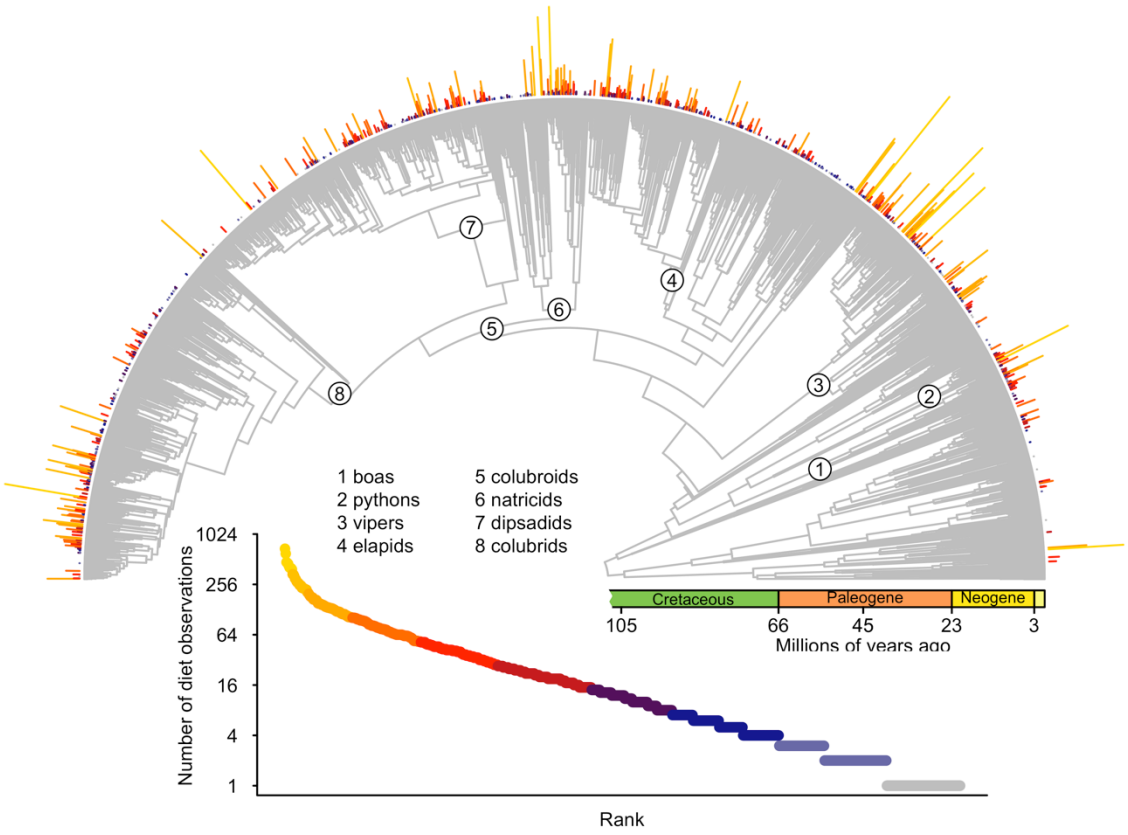

Figure 6.1 *Phylogenetic distribution and sample size variation of snake diets included in analysis*

Complete species-level phylogeny of snakes from Tonini et al. (2016) highlighting major clades, evolutionary time-scale, and sample size distribution for number of prey use observations. Rank-abundance curve below the phylogeny and segments along the outer semicircle depict the sample size distribution for all snakes with diet observations. Gaps along the outer semicircle occur for species with no diet observations, and these species were pruned from the phylogeny prior to analysis. A total of 33,952 primary natural history observations of prey acquisition by 880 species of snakes were available for analysis.

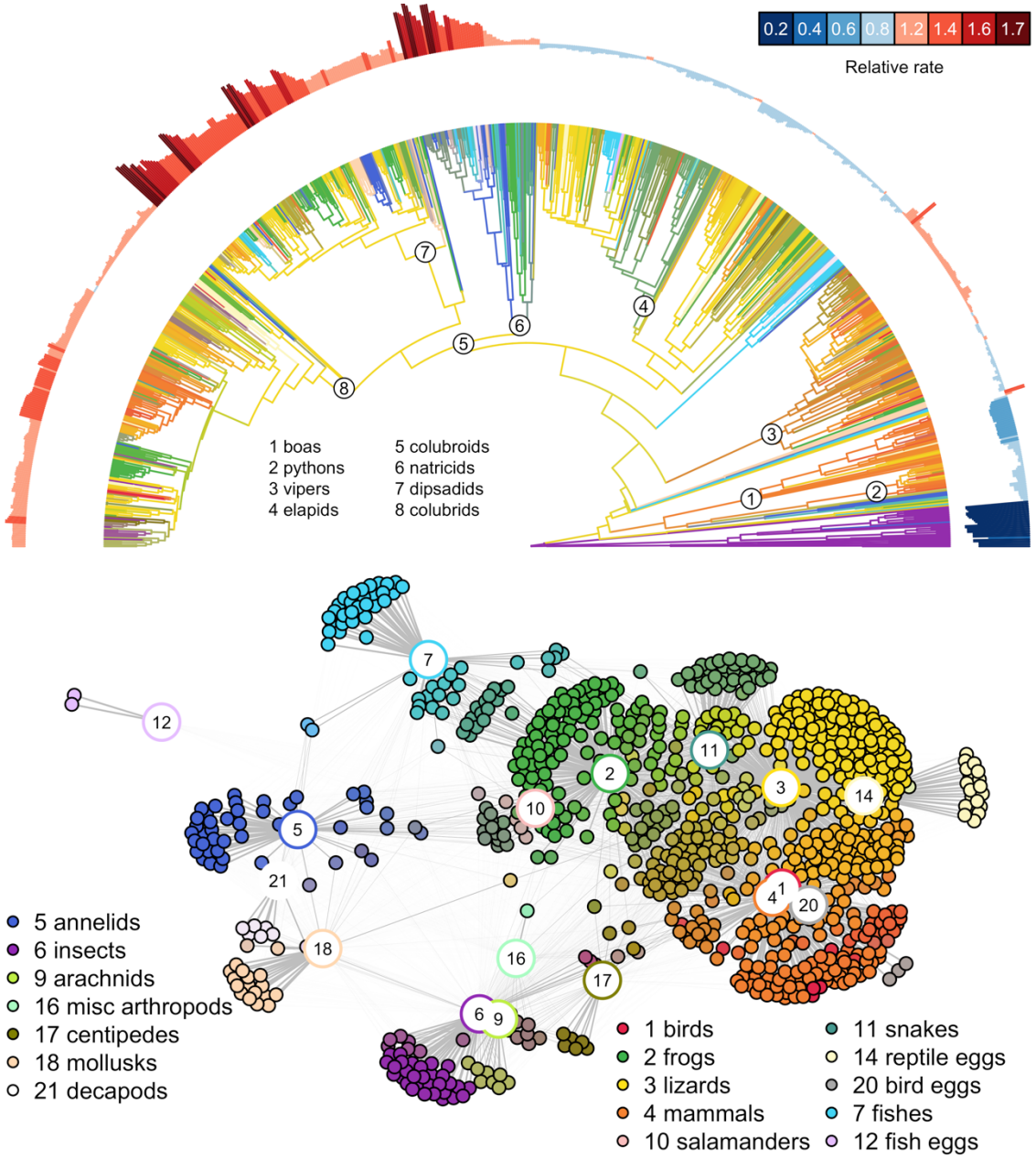

Figure 6.2 *Model inferred trophic network structure and evolutionary histories of diet evolution across extant snakes*

Model inferred trophic network structure (bottom) and evolutionary dynamics of diet evolution (top) across the radiation of extant snakes estimated from 33,952 primary natural history observations of prey acquisition by snakes. ***Bottom:*** Numbered circles represent the prey

categories used for analysis. Filled circles represent individual snake species in the dataset. Each species is connected to the prey items it feeds on using line widths proportional to the estimated relative importance of the item in its diet. Each prey item is represented by a color (shown by the borders of numbered circles) and the color assigned to an individual snake species is an additive mixture of the colors of the prey items it feeds on. Prey items that commonly co-occur in snake diets are positioned near one another, as are snakes with similar diets. Some prey items are rare enough that they do not appreciably alter the appearance of sampled species (e.g. crocodilians, turtles) and are omitted from the plot. ***Top*:** Reconstructions of ancestral snake diets (branch colors) and evolutionary rates of prey switching (outer semicircle) were inferred using a Dirichlet-multinomial Markov model for multivariate ecological trait evolution. Branch colors denote reconstructed patterns of resource use and are colored according to the same scheme used in the bottom panel. Outer semicircle denotes average rates of diet evolution for each taxon expressed relative to the average for all snakes (see text for details). Evolutionary rates are higher for the colubroid mega-radiation, which accounts for just over half of global snake diversity. Despite generally lower evolutionary rates, however, non-colubroid snakes display a similar breadth of feeding modalities as colubroids. Time-calibrated phylogeny for the 880 species for which field-based diet observations were available was taken from Tonini et al. (2016).

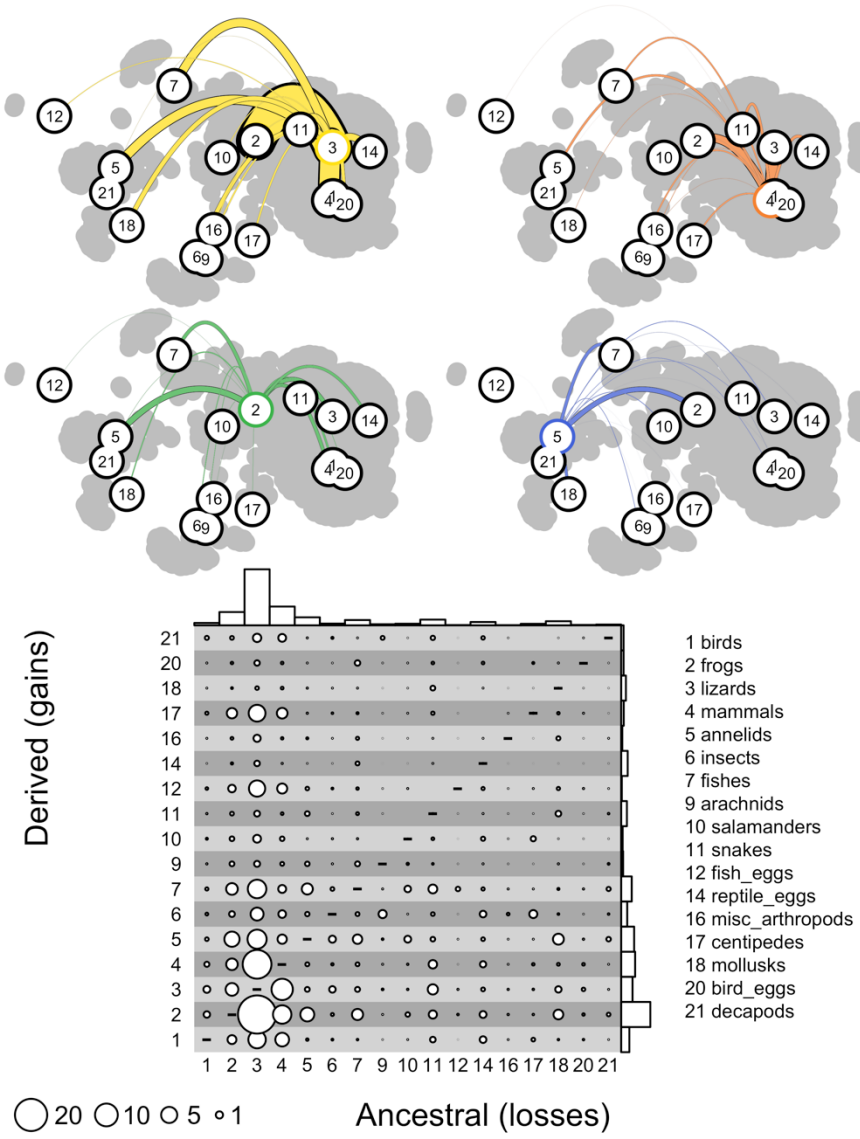

Figure 6.3 *Evolutionary dynamics (gains and losses) of diet evolution across extant snakes*

Average number of evolutionary gains and losses among different prey categories in snake diets. Reconstructions of ancestral snake diets were inferred using a Dirichlet-multinomial Markov model, and gains and losses between ancestors and descendants were computed under an optimal transport model that minimized the cost of transformation (see text for details). Colors and numbers follow the same scheme used in Figure 6.2. Top panels depict evolutionary gains of different prey categories from the ancestral prey category highlighted in color. Bottom panel

depicts gains and losses for all categories. Line widths and point sizes are proportional to the total number of gains/losses. Numerous independent origins of similar feeding strategies occur across the snake tree of life, often from a lizard-eating ancestor. Gains and losses are unequally distributed among prey categories, and some feeding strategies show much greater turnover than others, suggesting that feeding strategies differ in evolutionary accessibility and versatility.

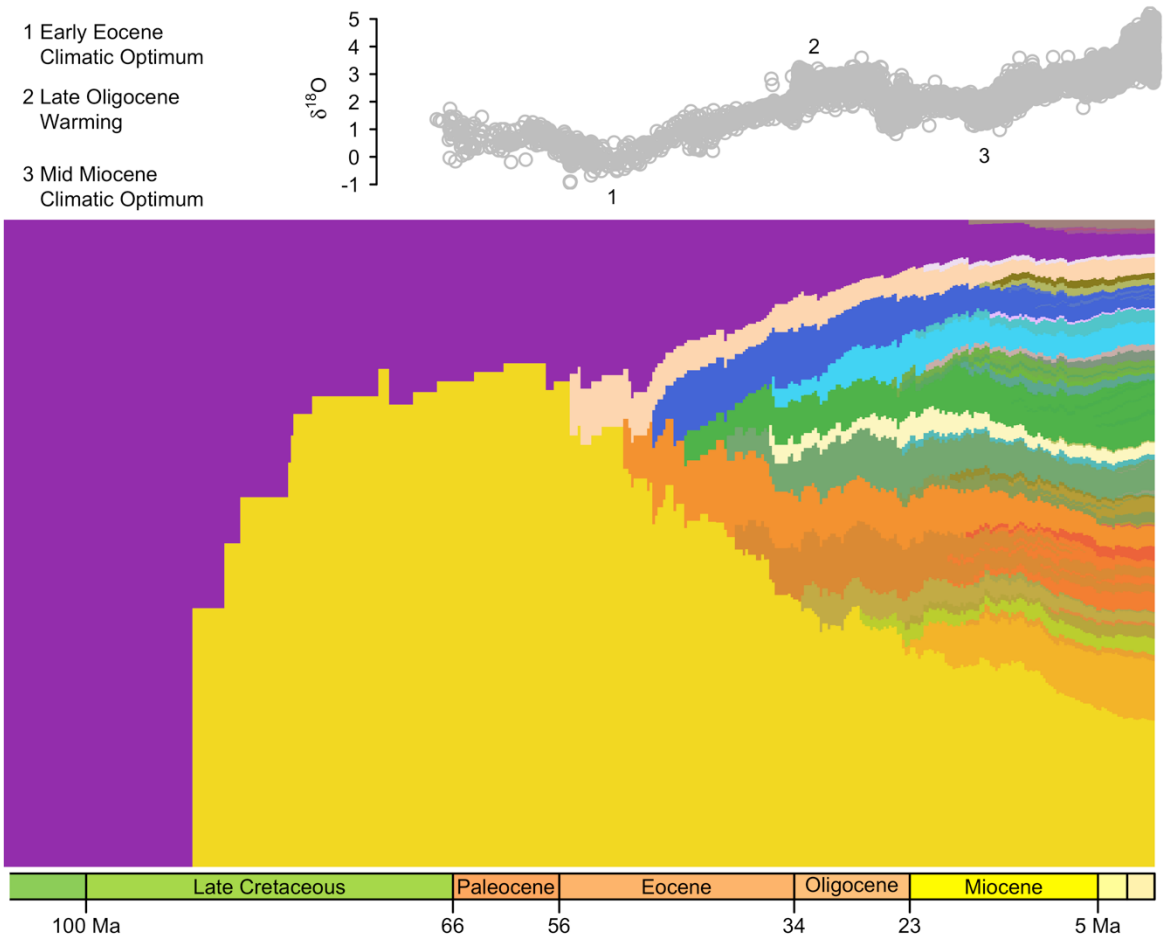

Figure 6.4 *Snake diets through time*

Proportional lineage representation of different feeding strategies during the global radiation of extant snakes. Diet-through-time profile is formed by plotting the most probable dietary niche state for each lineage at consecutive time points from the root to the present. Diet states are represented by distinct colors following the same scheme used in Figure 6.2, and dominant components of each state are labelled. Each diet state is plotted so that it originates at the crown clade age of the reconstructed ancestor where it first appears (using stem clade age pushes first appearances of some states back to the late Cretaceous). Stable isotope data from Zachos et al. (2001) are shown in the top panel to provide a timeline of major earth history events. The

diversity of snake dietary niches begins to flourish during the Eocene, when reconstructed cladogenetic events mark the origin of many higher taxonomic snake lineages.

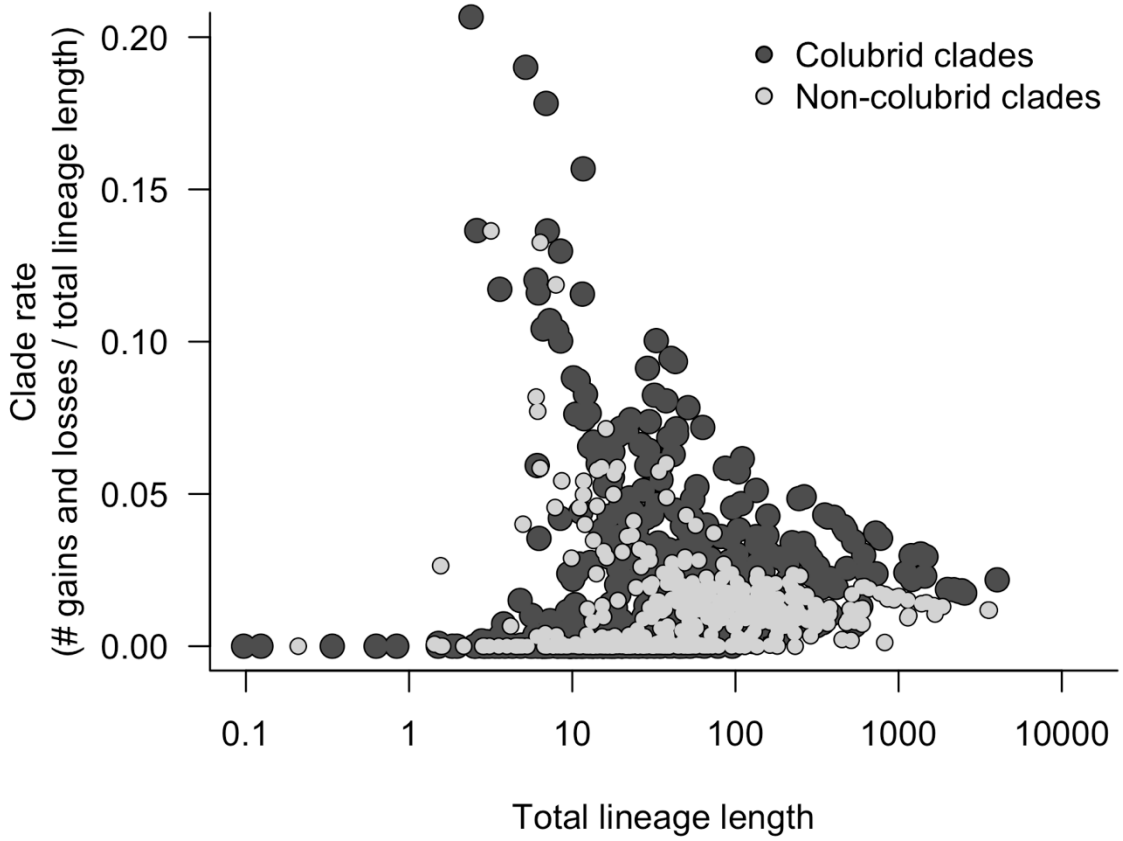

Figure S6.1 *Evolutionary rates of dietary changes*

Net rates of diet evolution for all clades in the Tonini et al. (2016) phylogeny used in the present study show that, in general, colubroids have the highest rates of all snakes, even after controlling for effects of overall lineage length. This pattern arises mainly from the leveraging effects of dipsadines and natricines, which evolved lots of dietary diversity and speciated quickly upon arrival in the Nearctic and Neotropics.

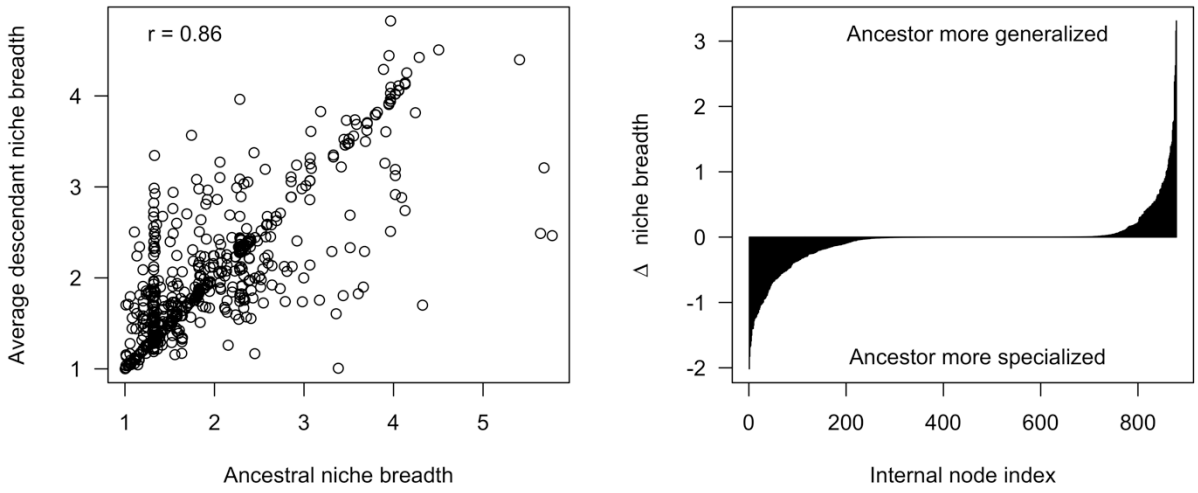

Figure S6.2 *Changes in dietary niche breadth from ancestor to descendant*

Average descendant niche breadths (effective number of different prey categories in diet) strongly resemble ancestral niche breadths (left) and increases and decreases in niche breadth are about equally common (right). These patterns suggest that "generalist" and "specialist" snakes do not differ in terms of overall evolutionary lability.

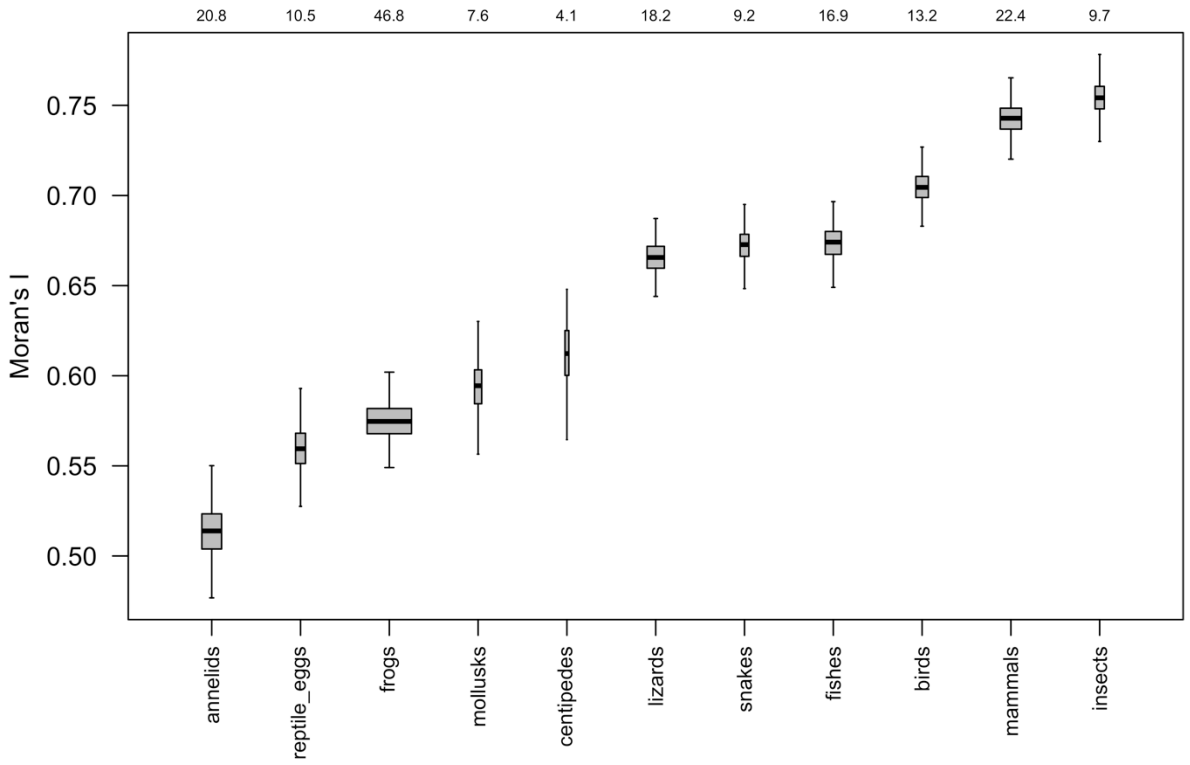

Figure S6.3 *Phylogenetic clustering of proportional prey use*

Phylogenetic clustering (Moran's I) of the relative importance of different prey categories. Boxplot widths for different prey categories are proportional to the estimated average number of evolutionary gains (numbers along the top margin). Annelids show the lowest levels of clustering, a consequence of their widespread phylogenetic distribution. More restricted prey categories have higher levels of clustering (e.g. mammals and insects). These numbers are potentially impacted by sampling effects. For example, some worm-eating clades have many species (e.g. *Atractus*, *Calamaria*) but are under-represented in the dataset. In this case, more complete sampling of these clades would be expected to increase Moran's I.

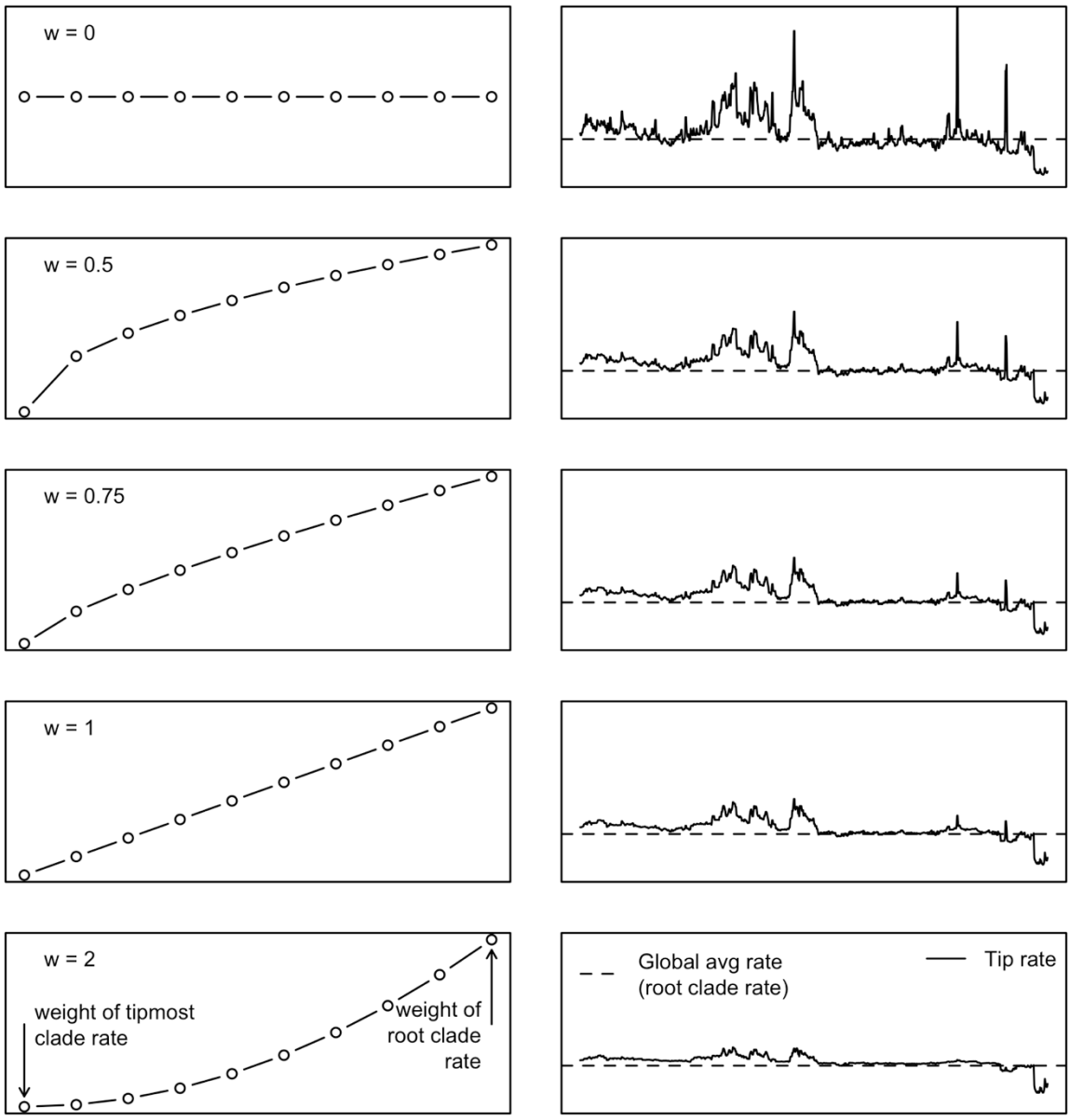

Figure S6.4 *Weighting schemes used for tip-rate calculation*

Different weighting schemes (left) applied to the clade rates shown in Figure S6.1 and their effect on tip rates (right). Tip rates are calculated as a weighted average of all clade rates on the phylogenetic path leading back to the root. While the weighting scheme influences the strength of relaxation toward the overall average rate, the qualitative pattern of tip rates is relatively

unchanged by choice of weighting scheme. Tip rates in the main text (Figure 6.2) correspond to scheme $w = 1$.

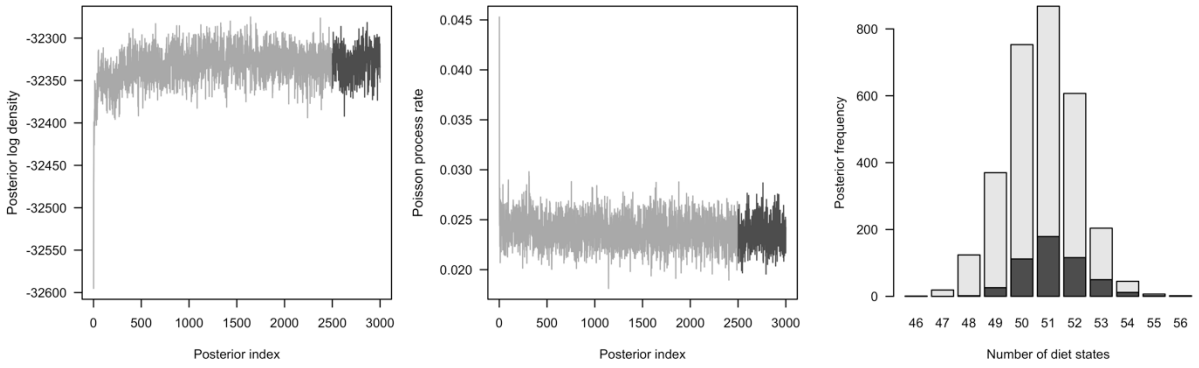

Figure S6.5 *Likelihood and parameter traces during Gibb's sampling*

Likelihood (left) and parameter (middle) traces reveal good mixing of the Gibb's sampler. The number of distinct dietary niches (right) sampled during the run is far less than the number of species (880), indicating that many sampled snake diets are indistinguishable from one another. The highlighted portions correspond to the final 500 samples that were used to form posterior average summaries mentioned in the main text.

Bibliography

1. Aitchison J. 1986. The statistical analysis of compositional data. London: Chapman and Hall.
2. Aldous D. 1989. Probability Approximations via the Poisson Clumping Heuristic. Springer-Verlag.
3. Alencar L.R.V., Gaiarsa M.P., Martins M. 2013. The evolution of diet and microhabitat use in pseudoboine snakes. South American Journal of Herpetology. 8:60–66.
4. Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences. 106:13410–13414.
5. Anderson T.W., Goodman L.A. 1957. Statistical Inference about Markov Chains. The Annals of Mathematical Statistics. 28:89–110.
6. Arnold S.J. 1972. Species densities of predators and their prey. The American Naturalist. 106:220–236.
7. Barry D., Hartigan J.A. 1987. Statistical Analysis of Hominoid Molecular Evolution. Statistical Science. 2:191–207.

8. Bassar R.D., Marshall M.C., López-Sepulcre A., Zandona E., Auer S.K., Travis J., Pringle C.M., Flecker A.S., Thomas S.A., Fraser D.F., Reznick D.N. 2010. Local adaptation in Trinidadian guppies alters ecosystem processes. PNAS. 107:3616–3621.

9. Baum L.E., Petrie T. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. The Annals of Mathematical Statistics. 37:1554–1563.

10. Beaulieu J.M., O'Meara B.C. 2016a. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Systematic Biology. 65:583–601.

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

12. 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. 62:725–737.

13. Bellairs A.D., Underwood G. 1951. The origin of snakes. Biological Reviews. 26:193–237.

14. Blackburn D.G. 1982. Evolutionary Origins of Viviparity in the Reptilia. I. Sauria. Amphibia-Reptilia. 3:185–205.

15. Blanquart S., Lartillot N. 2006. A Bayesian Compound Stochastic Process for Modeling Nonstationary and Nonhomogeneous Sequence Evolution. Molecular Biology and Evolution. 23:2058–2071.

16. Blei D.M., Ng A.Y., Jordan M.I. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research. 3:993–1022.

17. Boero F. 2013. Observational articles: a tool to reconstruct ecological history based on chronicling unusual events. F1000Research. 2.

18. Borries C., Gordon A.D., Koenig A. 2013. Beware of Primate Life History Data: A Plea for Data Standards and a Repository. PLoS ONE. 8:e67200.

19. Brito J.C. 2004. Feeding ecology of Vipera latastei in northern Portugal: ontogenetic shifts, prey size and seasonal variations. Herpetological Journal. 14:13–19.

20. Burin G., Kissling W.D., Guimarães P.R., Sekercioglu C.H., Quental T.B. Omnivory in birds is a macroevolutionary sink. Nature Communications. 7:11250.

21. Cadle J.E., Greene H.W. 1993. Phylogenetic patterns, biogeography, and the ecological structure of Neotropical snake assemblages. In: Ricklefs R.E., Schluter D., editors. Species Diversity in Ecological Communities: Historical and Geographical Perspectives. Chicago: University of Chicago Press. p. 281–293.

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

23. Cantalapiedra J.L., FitzJohn R.G., Kuhn T.S., Hernández Fernández M., DeMiguel D., Azanza B., Morales J., Mooers A.Ø. 2014. Dietary innovations spurred the diversification of ruminants during the Caenozoic. Proceedings of the Royal Society B: Biological Sciences. 281:20132746.

24. Cavalli-Sforza L.L., Edwards A.W. 1967. Phylogenetic analysis. Models and estimation procedures. American Journal of Human Genetics. 19:233–257.

25. Chretien J., Wang-Claypool C.Y., Glaw F., Scherz M.D. 2019. The bizarre skull of Xenotyphlops sheds light on synapomorphies of Typhlopoidea. Journal of Anatomy. 234:637–655.

26. Christin P.-A., Osborne C.P., Chatelet D.S., Columbus J.T., Besnard G., Hodkinson T.R., Garrison L.M., Vorontsova M.S., Edwards E.J. 2013. Anatomical enablers and the evolution of C4 photosynthesis in grasses. Proceedings of the National Academy of Sciences. 110:1381–1386.

27. Colston T.J., Costa G.C., Vitt L.J. 2010. Snake diets and the deep history hypothesis. Biological Journal of the Linnean Society. 101:476–486.

28. Cooper W. 2008. Tandem evolution of diet and chemosensory responses in snakes. Amphib Reptilia. 29:393–398.

29. Cundall D., Greene H.W. 2000. Feeding in Snakes. Feeding: Form, Function, and Evolution in Tetrapod Vertebrates. Elsevier. p. 293–333.

30. Cuturi M. 2013. Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances. Advances in Neural Information Processing Systems. 26:2292–2300.

31. Davis A.M., Unmack P.J., Vari R.P., Betancur-R R. 2016. Herbivory promotes dental disparification and macroevolutionary dynamics in grunters (Teleostei: Terapontidae), a freshwater adaptive radiation. The American Naturalist. 187:320–333.

32. Davis Rabosky A.R., Cox C.L., Rabosky D.L., Title P.O., Holmes I.A., Feldman A., McGuire J.A. 2016. Coral snakes predict the evolution of mimicry across New World snakes. Nature Communications. 7:1–9.

33. Ditmars R.L. 1912. The feeding habits of serpents. Zoologica. 1:197–238.

34. Drummond A.J., Suchard M.A. 2010. Bayesian random local clocks, or one rate to rule them all. BMC Biology. 8:114.

35. 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.

36. Ehrlich P.R., Raven P.H. 1964. Butterflies and plants: a study in coevolution. Evolution. 18:586–608.

37. Evans H.E. 1955. The osteology of a worm snake, Typhlops jamaicensis (Shaw). The Anatomical Record. 122:381–396.

38. Evans M.E.K., Smith S.A., Flynn R.S., Donoghue M.J. 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses (*Oenothera*, Sections *Anogra* and *Kleinia*). The American Naturalist. 173:225–240.

39. Felsenstein J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution. 17:368–376.

40. Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. The American Naturalist. 171:713–725.

41. Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the threshold model. The American Naturalist. 179:145–156.

42. Felsenstein, J., Churchill G.A. 1996. A Hidden Markov Model approach to variation among sites in rate of evolution. Molecular biology and evolution. 13:93–104.

43. Ferreto Fiorillo B., Robson Souza P., Martins M. 2013. Mussurana bicolor (Serpentes: Dipsadidae: Pseudoboini) preying upon Helicops leopardinus (Serpentes: Dipsadidae: Hydropsini). Notas de Historia Natural. 2:40–41.

44. Fitch W.M. 1971. Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. Systematic Biology. 20:406–416.

45. Fitch W.M., Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochemical Genetics. 4:579–593.

46. FitzJohn R.G. 2014. Diversitree: comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution.:1084–1092.

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

48. Forister M.L., Dyer L.A., Singer M.S., Stireman III J.O., Lill J.T. 2012. Revisiting the evolution of ecological specialization, with emphasis on insect-plant interactions. Ecology. 93:981–991.

49. Forister M.L., Novotny V., Panorska A.K., Baje L., Basset Y., et al. 2015. The global distribution of diet breadth in insect herbivores. PNAS. 112:442–447.

50. Froese R., Pauly D., editors. 2000. FishBase: concepts, design and data sources. Phillipines: ICLARM.

51. Fry B.G., Scheib H., van der Weerd L., Young B., McNaughton J., Ramjan S.F., Vidal N., Poelmann R.E., Norman J.A. 2008. Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia). Molecular & Cellular Proteomics. 7:215–246.

52. Futumya D.J. 1998. Wherefore and Whither the Naturalist? The American Naturalist. 151:1–6.

53. Futuyma D.J., Moreno G. 1988. The evolution of ecological specialization. Annual Review of Ecology and Systematics. 19:207–233.

54. Gaiarsa M., de Alencar L., Martins M. 2013. Natural history of Pseudoboine snakes. Papéis Avulsos de Zoologia. 53:261–283.

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

56. Gans C. 1961. The feeding mechanism of snakes and its possible evolution. American Zoologist. 1:217–227.

57. Garberoglio F.F., Apesteguía S., Simões T.R., Palci A., Gómez R.O., Nydam R.L., Larsson H.C.E., Lee M.S.Y., Caldwell M.W. 2019. New skulls and skeletons of the Cretaceous legged snake Najash, and the evolution of the modern snake body plan. Science Advances. 5:eaax5833.

58. Gillespie R. 2004. Community assembly through adaptive radiation in Hawaiian spiders. Science. 303:356–359.

59. Gingerich P.D. 1983. Rates of evolution: effects of time and temporal scaling. Science. 222:159–161.

60. Givnish T.J., Barfuss M.H.J., Van Ee B., Riina R., Schulte K., Horres R., Gonsiska P.A., Jabaily R.S., Crayn D.M., Smith J.A., Winter K., Brown G.K., Evans T.M., Holst B.K., Luther H., Till W., Zizka G., Berry P.E., Sytsma K.J. 2014. Adaptive radiation, correlated and contingent evolution, and net diversification in Bromeliaceae. Molecular Phylogenetics and Evolution. 71:55–78.

61. Glaudas X., Kearney T., Alexander G. 2017. Museum Specimens Bias Measures of Snake Diet: A Case Study Using the Ambush-Foraging Puff Adder (Bitis arietans). Herpetologica. 73:121–128.

62. Godfrey M., Thomas R.A. 2013. A Galapagos Ectothermic Terrestrial Snake Gambles a Potential Chilly Bath for a Protein-rich Dish of Fish. Herpetological Review. 44:415–417.

63. Golcher-Benavides J., Wagner C.E. 2019. Playing out Liem's Paradox: Opportunistic Piscivory across Lake Tanganyikan Cichlids. The American Naturalist. 194:260–267.

64. Goldberg E.E., Igić B. 2008. On Phylogenetic Tests of Irreversible Evolution. Evolution. 62:2727–2741.

65. Goldman N. 1990. Maximum Likelihood Inference of Phylogenetic Trees, with Special Reference to a Poisson Process Model of DNA Substitution and to Parsimony Analyses. Systematic Biology. 39:345–361.

66. Goolsby E.W. 2015. Phylogenetic Comparative Methods for Evaluating the Evolutionary History of Function-Valued Traits. Systematic Biology. 64:568–578.

67. Greene H. 2005. Organisms in nature as a central focus for biology. Trends in Ecology & Evolution. 20:23–27.

68. Greene H.W. 1983. Dietary correlates of the origin and radiation of snakes. American Zoologist. 23:431–441.

69. Greene H.W. 1986. Natural history and evolutionary biology. In: Feder M.E., Lauder G.V., editors. Predator–prey relationships: perspectives and approaches from the study of lower vertebrates. University of Chicago Press. p. 198.

70. Greene H.W., Burghardt G.M. 1978. Behavior and phylogeny: constriction in ancient and modern snakes. Science. 200:74–77.

71. Grundler M.C. 2020. SquamataBase: a natural history database and R package for comparative biology of snake feeding habits. Biodiversity Data Journal. 8:e49943.

72. Grundler M.C., Rabosky D.L. 2020. Complex ecological phenotypes on phylogenetic trees: a Markov process model for comparative analysis of multivariate count data. Systematic Biology.

73. Hardy C.R. 2006. Reconstructing ancestral ecologies: challenges and possible solutions. Diversity and Distributions. 12:7–19.

74. Hardy C.R., Linder H.P. 2005. Intraspecific variability and timing in ancestral ecology reconstruction: a test case from the cape flora. Systematic Biology. 54:299–316.

75. Hardy N.B. 2017. Do plant-eating insect lineages pass through phases of host-use generalism during speciation and host switching? Phylogenetic evidence. Evolution. 71:2100–2109.

76. Hardy N.B., Otto S.P. 2014. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. Proceedings of the Royal Society B: Biological Sciences. 281:20132960.

77. Harmon, L. J., Matthews B., Des Roches S., Chase J.M., Shurin J.B., Schluter D. 2009. Evolutionary diversification in stickleback affects ecosystem functioning. Nature. 458:1167–1170.

78. Harvey P.H., Pagel M.D. 1991. The comparative method in evolutionary biology. Oxford University Press.

79. Heath T.A., Holder M.T., Huelsenbeck J.P. 2012. A Dirichlet Process Prior for Estimating Lineage-Specific Substitution Rates. Molecular Biology and Evolution. 29:939–955.

80. Henao Diaz L.F., Harmon L.J., Sugawara M.T.C., Miller E.T., Pennell M.W. 2019. Macroevolutionary diversification rates show time dependency. PNAS. 116:7403–7408.

81. Hortal J., de Bello F., Diniz-Filho J.A.F., Lewinsohn T.M., Lobo J.M., Ladle R.J. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics. 46:523–549.

82. Hoso M., Asami T., Hori M. 2007. Right-handed snakes: convergent evolution of asymmetry for functional specialization. Biology Letters. 3:169–172.

83. Hsiang A.Y., Field D.J., Webster T.H., Behlke A.D., Davis M.B., Racicot R.A., Gauthier J.A. 2015. The origin of snakes: revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. BMC Evolutionary Biology. 15.

84. Huelsenbeck J.P., Ané C., Larget B., Ronquist F. 2008. A Bayesian perspective on a non-parsimonious parsimony model. Systematic Biology. 57:406–419.

85. Huelsenbeck J.P., Larget B., Swofford D. 2000. A Compound Poisson Process for Relaxing the Molecular Clock. Genetics. 154:1879–1892.

86. Igea J., Miller E.F., Papadopulos A.S.T., Tanentzap A.J. 2017. Seed size and its rate of evolution correlate with species diversification across angiosperms. PLOS Biology. 15:e2002792.

87. Ives A.R., Midford P.E., Garland Jr T. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Systematic Biology. 56:252–270.

88. Janson C.H. 1992. Measuring Evolutionary Constraints: A Markov Model for Phylogenetic Transitions Among Seed Dispersal Syndromes. Evolution. 46:136–158.

89. Jayne B.C., Voris H.K., Ng P.K.L. 2002. Snake circumvents constraints on prey size. Nature. 418.

90. Jobe K., Montaña C., Schalk C. 2019. Emergent patterns between salamander prey and their predators. Food Webs. 21:e00128.

91. Jones N.S., Moriarty J. 2013. Evolutionary inference or function-valued traits: Gaussian process regression on phylogenies. Journal of the Royal Society Interface. 10:20120616.

92. Kardon K.V. 1980. Evolutionary patterns in advanced snakes. American Zoologist. 20:269–282.

93. Karr A.F. 1991. Point processes and their statistical inference. New York: M. Dekker.

94. Kelley S.T., Farrell B.D. 1998. Is specialization a dead end? The phylogeny of host use in *Dendroctonus* bark beetles (Scolytidae). Evolution. 52:1731–1743.

95. King B., Lee M.S.Y. 2015. Ancestral State Reconstruction, Rate Heterogeneity, and the Evolution of Reptile Viviparity. Systematic Biology. 64:532–544.

96. Kissling W.D., Dalby L., Fløjgaard C., Lenoir J., Sandel B., Sandom C., Trøjelsgaard K., Svenning J.-C. 2014. Establishing macroecological trait datasets: digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. Ecology and Evolution. 4:2913–2930.

97. Landis M.J., Schraiber J.G., Liang M. 2013. Phylogenetic Analysis Using Lévy Processes: Finding Jumps in the Evolution of Continuous Traits. Systematic Biology. 62:193–204.

98. Lee M.S.Y., Palci A., Jones M.E.H., Caldwell M.W., Holmes J.D., Reisz R.R. 2016. Aquatic adaptations in the four limbs of the snake-like reptile Tetrapodophis from the Lower Cretaceous of Brazil. Cretaceous Research. 66:194–199.

99. Lewis P.O. 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. Systematic Biology. 50:913–925.

100. Liem K. 1980. Adaptive significance of intra- and interspecific differences in the feeding repertoires of cichlid fishes. American Zoologist. 20:295–314.

101. Lloyd G.T., Wang S.C., Brusatte S.L. 2012. Identifying Heterogeneity in Rates of Morphological Evolution: Discrete Character Change in the Evolution of Lungfish (sarcopterygii; Dipnoi). Evolution. 66:330–348.

102. Longrich N.R., Bhullar B.-A.S., Gauthier J.A. 2012. Mass extinction of lizards and snakes at the Cretaceous-Paleogene boundary. Proceedings of the National Academy of Sciences. 109:21396–21401.

103. Losos J.B., Leal M., Glor R.E., de Queiroz K., Hertz P.E., Rodríguez Schettino L., Chamizo Lara A., Jackman T.R., Larson A. 2003. Niche lability in the evolution of a Caribbean lizard community. Nature. 424:542–545.

104. Losos J.B., Mahler D.L. 2010. Adaptive radiation: the interaction of ecological opportunity, adaptation, and speciation. Evolution Since Darwin: The First 150 Years. Sunderland, MA: Sinauer Associates, Inc. p. 382–420.

105. Louca S., Pennell M.W. 2019. A General and Efficient Algorithm for the Likelihood of Diversification and Discrete-Trait Evolutionary Models. Systematic Biology. 0:1–12.

106. Lovette I.J., Bermingham E., Ricklefs R.E. 2002. Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. Proceedings of the Royal Society B: Biological Sciences. 269:37–42.

107. MacArthur R.H. 1958. Population ecology of some warblers of northeastern coniferous forests. Ecology. 39:599–619.

108. Maddison D.R. 1994. Phylogenetic Methods for Inferring the Evolutionary History and Processes of Change in Discretely Valued Characters. Annual Review of Entomology. 39:267–292.

109. Maddison W.P. 1990. A Method for Testing the Correlated Evolution of Two Binary Characters: Are Gains or Losses Concentrated on Certain Branches of a Phylogenetic Tree? Evolution. 44:539–557.

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

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

112. Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a Binary Character's Effect on Speciation and Extinction. Systematic Biology. 56:701–710.

113. 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.

114. Maria Albino A. 1993. Snakes from the Paleocene and Eocene of Patagonia (Argentina): Paleoecology and coevolution with mammals. Historical Biology. 7:51–69.

115. Martill D.M., Tischlinger H., Longrich N.R. 2015. A four-legged snake from the Early Cretaceous of Gondwana. Science. 349:416–419.

116. Martin C.H., Wainwright P.C. 2011. Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of *Cyprinodon* pupfish. Evolution. 65:2197–2212.

117. von May R., Biggi E., Cárdenas H., Isabel Dias M., Alarcón C., Herrera V., Santa-Cruz R., Tomasinelli F., Westeen E.P., Sánchez-Paredes C.M., Larson J.G., Title P.O.,

Grundler M.R., Grundler M.C., Davis Rabosky A.R., Rabosky D.L. Ecological interactions between arthropods and small vertebrates in a lowland Amazon rainforest. Amphibian & Reptile Conservation. 13:65–77.

118. Mayr E. 1963. Animal species and evolution. Cambridge, MA: Harvard University Press.

119. Miller E.T., Wagner S.K., Harmon L.J., Ricklefs R.E. 2017. Radiating despite a lack of character: ecological divergence among closely related, morphologically similar honeyeaters (Aves: Meliphagidae) co-occurring in arid Australian environments. The American Naturalist. 189:E14–E30.

120. Miralles A., Marin J., Markus D., Herrel A., Hedges S.B., Vidal N. 2018. Molecular evidence for the paraphyly of Scolecophidia and its evolutionary implications. Journal of Evolutionary Biology. 31:1782–1793.

121. Mitchell J.S., Rabosky D.L. 2017. Bayesian model selection with BAMM: effects of the model prior on the inferred number of diversification shifts. Methods in Ecology and Evolution. 8:37–46.

122. Mitter C., Farrell B., Wiegmann B. 1988. The Phylogenetic Study of Adaptive Zones: Has Phytophagy Promoted Insect Diversification? The American Naturalist. 132:107–128.

123. Mooers A.Ø., Schluter D. 1999. Reconstructing Ancestor States with Maximum Likelihood: Support for One-and Two-Rate Models. Systematic Biology. 48:623–633.

124. Mossel E., Steel M. 2004. A phase transition for a random cluster model on phylogenetic trees. Mathematical Biosciences. 187:189–203.

125. Neal, Radford M. 2003. Slice Sampling. The Annals of Statistics. 31:705–767.

126. Nicolakakis N., Sol D., Lefebvre L. 2003. Behavioural flexibility predicts species richness in birds, but not extinction risk. Animal Behaviour. 65:445–452.

127. Nielsen R. 2002. Mapping mutations on phylogenies. Systematic Biology. 51:729–739.

128. Nosil P. 2002. Transition rates between specialization and generalization in phytophagous insects. Evolution. 56:1701–1706.

129. Nosil P., Mooers A.Ø. 2005. Testing Hypotheses About Ecological Specialization Using Phylogenetic Trees. Evolution. 59:2256–2263.

130. O'Meara B.C. 2012. Evolutionary inferences from phylogenies: a review of methods. Annual Review of Ecology and Systematics. 43:267–285.

131. 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.

132. Ortiz-Catedral L., Christian E., Skirrow M.J.A., Rueda D., Sevilla C.R., Kumar K., Reyes E.M.R., Daltry J.C. 2019. Diet of six species of Galapagos terrestrial snakes (Pseudalsophis spp.) inferred from faecal samples. Herpetology Notes. 12:701–704.

133. Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society of London. Series B: Biological Sciences. 255:37–45.

134. Pagel M. 1999. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. Systematic Biology. 48:612–622.

135. Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 20:289–290.

136. Parravicini V., Casey J.M., Schiettekatte N.M.D., Brandl S.J., Pozas-Schacre C., Carlot J., Edgar G.J., Graham N.A.J., Harmelin-Vivien M., Kulbicki M., Strona G., Stuart-Smith

R.D., Vii J. 2020. Global gut content data synthesis and phylogeny delineate reef fish trophic guilds. .

137. 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. 53:711–723.

138. Pianka E.R. 1986. Ecology and natural history of desert lizards; analyses of the ecological niche and community structure. Princeton, New Jersey: Princeton University Press.

139. Pigot A.L., Trisos C.H., Tobias J.A. 2016a. Functional traits reveal the expansion and packing of ecological niche space underlying an elevational diversity gradient in passerine birds. Proceedings of the Royal Society B: Biological Sciences. 283:20152013.

140. Pigot A.L., Trisos C.H., Tobias J.A. 2016b. Functional traits reveal the expansion and packing of ecological niche space underlying an elevational diversity gradient in passerine birds. Proceedings of the Royal Society B: Biological Sciences. 283:20152013.

141. Poelen J.H., Simons J.D., Mungall C.J. 2014. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. Ecological Informatics. 24:148–159.

142. Poisot T., Gravel D., Leroux S., Wood S.A., Fortin M.-J., Baiser B., Cirtwill A.R., Araújo M.B., Stouffer D.B. 2016. Synthetic datasets and community tools for the rapid testing of ecological hypotheses. Ecography. 39:402–408.

143. Price S.A., Hopkins S.S.B., Smith K. K., Roth V.L. 2012. Tempo of trophic evolution and its impact on mammalian diversification. PNAS. 109:7008–7012.

144. Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 155:945–959.

145. Pyron R.A., Burbrink F.T. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. Ecology Letters. 17:13–21.

146. de Queiroz A., Rodríguez-Robles J.A. 2006. Historical Contingency and Animal Diets: The Origins of Egg Eating in Snakes. The American Naturalist. 167:684–694.

147. Quintero I., Keil P.K., Jetz W., Crawford F.W. 2015. Historical biogeography using species geographical ranges. Systematic Biology. 64:1059–1073.

148. Rabosky D.L. 2014. Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. PLOS ONE. 9:e89543.

149. Rabosky D.L., Chang J., Title P.O., Cowman P.F., Sallan L., Friedman M., Kaschner K., Garilao C., Near T.J., Coll M., Alfaro M.E. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 559:392–395.

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

151. Rabosky D.L., Grundler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H., Larson J.G. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution. 5:701–707.

152. Rabosky D.L., Mitchell J.S., Chang J. 2017. Is BAMM Flawed? Theoretical and Practical Concerns in the Analysis of Multi-Rate Diversification Models. Systematic Biology. 66:477–498.

153. Rabosky D.L., Santini F., Eastman J., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nature Communications. 4:1–8.

154. Reddy S., Driskell A., Rabosky D.L., Hackett S.J., Schulenberg T.S. 2012. Diversification and the adaptive radiation of the vangas of Madagascar. Proceedings of the Royal Society B: Biological Sciences. 279:2062–2071.

155. Ree R.H., Donoghue M.J. 1999. Inferring Rates of Change in Flower Symmetry in Asterid Angiosperms. Systematic Biology. 48:633–641.

156. Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology. 57:4–14.

157. Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution. 3:217–223.

158. Revell L.J. 2014. Ancestral character estimation under the threshold model from quantitative genetics. Evolution. 68:743–759.

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

160. Robinson G.S. 1999. HOSTS - a database of the hostplants of the world's Lepidoptera. Nota Lepidopterologica. 22:35–47.

161. Rodríguez-Robles J.A. 1994. Are the Duvernoy's Gland Secretions of Colubrid Snakes Venoms? Journal of Herpetology. 28:388–390.

162. Rodríguez-Robles J.A., Bell C., Greene H.W. 1999. Gape size and evolution of diet in snakes: Feeding ecology of erycine boas. Journal of Zoology. 248:49–58.

163. Root R.B. 1967. The niche exploitation pattern of the blue-gray gnatcatcher. Ecological Monographs. 37:317–350.

164. Roskov Y., Abucay L., Orrell T., Nicolson D., Flann C., Bailly N., Kirk P., Bourgoin T., DeWalt R., Decock W., De Wever A., editors. 2016. Species 2000 & ITIS Catalogue of Life, 2016 Annual Checklist. Leiden, the Netherlands: Naturalis.

165. Rossman D.A., Myer P.A. 1990. Behavioral and Morphological Adaptations for Snail Extraction in the North American Brown Snakes (Genus storeria). Journal of Herpetology. 24:434–438.

166. Royer-Carenzi M., Pontarotti P., Didier G. 2013. Choosing the best ancestral character state reconstruction method. Mathematical Biosciences. 242:95–109.

167. Sanderson M.J. 1993. Reversibility in Evolution: A Maximum Likelihood Approach to Character Gain/Loss Bias in Phylogenies. Evolution. 47:236–252.

168. Savitzky A.H. 1980. The Role of Venom Delivery Strategies in Snake Evolution. Evolution. 34:1194–1204.

169. Savitzky A.H. 1983. Coadapted character complexes among snakes: fossoriality, piscivory, and durophagy. American Zoologist. 23:397–409.

170. Sazima I. 1989. Feeding behavior of the snail-eating snake, Dipsas indica. Journal of Herpetology. 23:464–468.

171. Schadt E.E., Sinsheimer J.S., Lange K. 1998. Computational Advances in Maximum Likelihood Methods for Molecular Phylogeny. Genome Res. 8:222–233.

172. Schalk C., Cove M. 2018. Squamates as prey: predator diversity patterns and predator-prey size relationships. Food Webs. 17:e00103.

173. Schluter D., Price T., Mooers A.Ø., Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. Evolution. 51:1699–1711.

174. Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research. Proceedings of the Royal Society B: Biological Sciences. 273:1987–1998.

175. Sekercioglu C.H., Daily G.C., Ehrlich P.R. 2004. Ecosystem consequences of bird declines. PNAS. 101:18042–18047.

176. Shine R. 1991. Intersexual Dietary Divergence and the Evolution of Sexual Dimorphism in Snakes. The American Naturalist. 138:102–122.

177. Shine R. 1994. Allometric patterns in the ecology of Australian snakes. Copeia. 1994:851–867.

178. Shine R. 2014. Evolution of an Evolutionary Hypothesis: A History of Changing Ideas about the Adaptive Significance of Viviparity in Reptiles. Journal of Herpetology. 48:147–161.

179. Shine R., Wall M. 2008. Interactions between locomotion, feeding, and bodily elongation during the evolution of snakes. Biological Journal of the Linnean Society. 95:293–304.

180. Simberloff D., Dayan T. 1991. The guild concept and the structure of ecological communities. Annual Review of Ecology and Systematics. 22:115–143.

181. Simpson G.G. 1944. Tempo and Mode in Evolution. New York: Columbia University Press.

182. Smith K.T., Scanferla A. 2016. Fossil snake preserving three trophic levels and evidence for an ontogenetic dietary shift. Palaeobiodiversity and Palaeoenvironments. 96:589–599.

183. Sol D., Timmermans S., Lefebvre L. 2002. Behavioural flexibility and invasion success in birds. Animal Behaviour. 63:495–502.

184. Stanley S.M. 1979. Macroevolution, pattern and process. San Francisco: W. H. Freeman.

185. Steel M. 2011. Can we avoid “SIN” in the house of “no common mechanism”? Systematic Biology. 60:96–109.

186. Steel M., Penny D. 2000. Parsimony, Likelihood, and the Role of Models in Molecular Phylogenetics. Molecular Biology and Evolution. 17:839–850.

187. Stroud J.T., Losos J.B. 2016. Ecological Opportunity and Adaptive Radiation. Annual Review of Ecology, Evolution, and Systematics. 47:507–532.

188. Toft C.A. 1985. Resource partitioning in amphibians and reptiles. Copeia. 1985:1–21.

189. Toledo L., Ribeiro R., Haddad C. 2007. Anurans as prey: an exploratory analysis and size relationships between predators and their prey. Journal of Zoology. 271:170–177.

190. Tonini J., Beard K.H., Barbosa Ferreira R., Jetz W., Pyron R.A. 2016. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological Conservation. 204:23–31.

191. Tuffley C., Steel M. 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bulletin of Mathematical Biology. 59:581–607.

192. 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.

193. Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018a. Rethinking phylogenetic comparative methods. Systematic Biology. 67:1091–1109.

194. Uyeda J.C., Zenil-Ferguson R., Pennell M.W. 2018b. Rethinking phylogenetic comparative methods. Systematic Biology. 67:1091–1109.

195. Van Valen L. 1971. Adaptive zones and the orders of mammals. Evolution. 25:420–428.

196. Vermeij G.J. 1973. Adaptation, Versatility, and Evolution. Systematic Biology. 22:466–477.

197. Vitt L.J., Pianka E.R. 2005. Deep history impacts present-day ecology and biodiversity. PNAS. 102:7877–7881.

198. Vitt L.J., Vangilder L.D. 1983. Ecology of a snake community in northeastern Brazil. Amphibia-Reptilia. 4:273–296.

199. Vrba E.S. 1987. Ecology in relation to speciation rates: some case histories of Miocene-Recent mammal clades. Evolutionary Ecology. 1:283–300.

200. Weber M.G., Keeler K.H. 2013. The phylogenetic distribution of extrafloral nectaries in plants. Annals of Botany. 111:1251–1261.

201. Wiens J.J., Brandley M.C., Reeder T.W. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. Evolution. 60:123–141.

202. Wilman H., Belmaker J., Simpson J., de la Rosa C., Rivadeneira M.M., Jetz W. 2014. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals: Ecological Archives E095-178. Ecology. 95:2027–2027.

203. Wilson J.A., Mohabey D.M., Peters S.E., Head J.J. 2010. Predation upon Hatchling Dinosaurs by a New Snake from the Late Cretaceous of India. PLOS Biology. 8:e1000322.

204. Wu S., Wu W., Zhang F., Ye J., Ni X., Sun J., Edwards S.V., Meng J., Organ C.L. 2012. Molecular and Paleontological Evidence for a Post-Cretaceous Origin of Rodents. PLOS One. 7:e46445.

205. Yi H., Norell M.A. 2015. The burrowing origin of modern snakes. Science Advances. 1:e1500743.

206. Yin J., Wang J. 2014. A dirichlet multinomial mixture model-based approach for short text clustering. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’14.:233–242.

207. Zaher H., Grazziotin F.G., Cadle J.E., Murphy R.W., de Moura-Leite J.C., Bonatto S.L. 2009. Molecular phylogeny of advanced snakes (Serpentes, Caenophidia) with an emphasis on South American Xenodontines: a revised classification and descriptions of new taxa. Papéis Avulsos de Zoologia. 49:115–153.

208. Zaher H., de Oliveira L., Grazziotin F.G., Campagner M., Jared C., Antoniazzi M.M., Prudente A.L. 2014. Consuming viscous prey: a novel protein-secreting delivery system in neotropical snail-eating snakes. BMC Evolutionary Biology. 14.