Species Richness in Time and Space:
a Phylogenetic and Geographic Perspective

by

Pascal Olivier Title

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
2018

Doctoral Committee:

        Assistant Professor and Assistant Curator Daniel Rabosky, Chair
        Associate Professor Johannes Foufopoulos
        Professor L. Lacey Knowles
        Assistant Professor Stephen A. Smith

Pascal O Title

ptitle@umich.edu

ORCID iD: 0000-0002-6316-0736

DEDICATION

To Judge Julius Title, for always encouraging me to be inquisitive.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Dan Rabosky, for taking me on as one of his first graduate students. I have learned a tremendous amount under his guidance, and conducting research with him has been both exhilarating and inspiring. I am also grateful for his friendship and company, both in Ann Arbor and especially in the field, which have produced experiences that I will never forget.

I would also like to thank my dissertation committee members: Dr. Johannes Foufopoulos, Dr. Lacey Knowles, and Dr. Stephen Smith. Their support, advice and encouragement has been greatly appreciated.

I would like to acknowledge my collaborators, both within and outside of my dissertation. In particular, I thank Michael Alfaro, Jordan Bemmels, Jonathan Chang, Alison Davis Rabosky, Lacey Knowles, Dan Rabosky, Sonal Singhal and Miriam Zelditch, for involving me on exciting projects, and for being such thoughtful coauthors. It has been a real pleasure

University, as well as from the Museum of Vertebrate Zoology community at the University of California, Berkeley.

I would also like to thank my family for nurturing a love for wildlife and the outdoors, and for encouraging my pursuit of a career in the sciences.

Last of all, but absolutely not least, I would like to thank Tara Smiley for her love and unwavering support, for challenging me intellectually, for being my adventure partner, and my best friend.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

ABSTRACT

Biodiversity varies dramatically across geographic space and across the tree of life, yet active debate among biologists remains regarding the underlying causes of these diversity patterns. By integrating phylogenies with species geographic range information and environmental or climatic datasets, we can explore questions relating to the assembly of communities and diversity gradients at continental to global scales.

In Chapter 1, I introduce major themes uniting macroevolution and macroecology. I describe the underlying conceptual framework that links my different research chapters together. I explain how these efforts advance our understanding of large-scale patterns of diversity, while providing critical assessments of tools and resources that facilitate the study of diversification across environmental and geographic gradients.

In Chapter 2, I highlight the recent availability of several large-scale phylogenies for squamate reptiles, and explore how they might affect macroevolutionary research. Using Australian squamates as a case study, I find that a great deal of conflict exists across phylogenies, both in terms of divergence times and topology. I demonstrate that these differences can be severe enough to alter conclusions drawn from downstream macroevolutionary analyses. I further explore the potential sources of and solutions for these discrepancies.

To properly test hypotheses pertaining to limits on species' geographic distributions, we need accurate geographic range estimates. A majority of studies currently rely on a set of

19 bioclimatic variables for species distribution modeling and related ecological research. In Chapter 3, I assemble a new bioclimatic dataset from variables described in the literature, in order to increase the number of predictors that are easily accessible to ecologists and evolutionary biologists. I find that incorporating these predictors into species distribution modeling workflows leads to noticeably improved models, and I anticipate that they will prove useful in macroecological studies as well.

In Chapter 4, I evaluate the performance of a number of approaches for estimating species-specific "tip rates" of speciation. These metrics, which quantify recent variation in rates of speciation across a phylogeny, are key for the study of trait-dependent diversification as well as spatial variation in rates across biomes and latitudinal gradients. Under a number of simulation scenarios, I assess the performance of three model-free tip rate metrics, and compare them to BAMM, a Bayesian model-based approach for estimating diversification rates. I find that BAMM exhibits the least amount of error in speciation rates in all diversification scenarios evaluated. One of the model-free metrics, DR, also performs well, although its performance is hampered by high variance in rate estimates.

Finally, in Chapter 5, I explore how biogeographic rates of dispersal have contributed to the latitudinal diversity gradient in marine fishes. There are dramatically more species in the tropics than at high latitudes, but prior research has found that speciation rates exhibit an inverse relationship with latitude, with the lowest rates in the tropics. I sought to determine whether or not global patterns in biogeographic immigration in marine fishes conform to an "out of the tropics" scenario, where lineages disperse out from the tropics and enrich higher latitude assemblages. I find that dispersal rates are strongly biased in a poles-to-tropics direction. However, given the strong latitudinal species richness gradient, estimated per-lineage rates of dispersal translate to greater net movement from the tropics to high latitudes, confirming that high latitude assemblages are enriched by tropical diversity over macroevolutionary timescales.

CHAPTER I

Introduction

Phylogenetic patterns coupled with the geographic distributions of species unlock a critical dimension of biodiversity. Spatial patterns in species richness provide a wealth of information that allows us to test a range of hypotheses pertaining to the ecological controls on the size of regional species pools (Belmaker and Jetz 2012) as well as the environmental correlates of species richness gradients (Hawkins et al. 2003, Currie et al. 2004). Integrating phylogenetic information makes it possible to additionally examine the historical and biotic factors, as well as the macroevolutionary processes, shaping the accumulation of diversity across space and through time (Graham and Fine 2008, Freckleton and Jetz 2009, Morlon et al. 2011). By evaluating both ecological and evolutionary hypotheses, biologists can assess the relative roles of a number of potential predictors contributing to the spatial biodiversity patterns that we observe today (Mittelbach et al. 2007, Belmaker and Jetz 2015, Holt et al. 2017).

Until relatively recently, the scale of analysis has been limited by the size of available phylogenies, typically restricted to families or orders, as well as by the availability of large-scale primary occurrence data. However, over the last decade, the emergence of new datasets and analytical methods has led to a "big data" revolution in ecology and evolutionary biology, with exciting opportunities for scientific inquiry. Natural history museums and collections have made tremendous efforts to digitize their holdings, and digital aggregators have been

brought online, making it possible to query and download millions of species occurrence records on a global scale (Constable et al. 2010, Jetz et al. 2012, Robertson et al. 2014). Additionally, high resolution, global environmental and climatic datasets have been made available. Although global climatic data have been available for some time (Booth et al. 2014), a number of new and complementary datasets have been published in the last few years (Tuanmu et al. 2015, Wilson and Jetz 2016, Karger et al. 2017, Hengl et al. 2017, Fick and Hijmans 2017, Amatulli et al. 2018) that should make it possible to test more targeted hypotheses and construct better geospatial models.

Alongside the increased availability of museum data and environmental data products, great strides have been made in computational biology and phylogenetics. Improvements in genetic sequence data acquisition and alignment (Smith et al. 2009, Thomson and Shaffer 2009, Hinchliff and Roalson 2013), as well as advances in computational phylogenetics (Zwickl 2006, Stamatakis 2014) have led to the recent inference of phylogenies with thousands, if not tens of thousands of species. The development of supermatrix and megaphylogeny inference methods, in particular, which typically involve large but mostly incomplete data matrices, has made it possible to take greater advantage of sequence data repositories like GenBank. Although concerns regarding the quality of these very large phylogenies have been raised (Misof et al. 2013, Hinchliff and Smith 2014), the incorporation of genomic data is a very active and promising topic of research (Zheng and Wiens 2016).

As large phylogenies become available for different groups of organisms and primary occurrence data become more accessible and complete, it is increasingly feasible to test hypotheses at broad spatial scales. This allows us to explore questions that were until recently out of reach, or for which researchers previously had to rely on sister species pairs to control for time since divergence (Cardillo 1999, Cardillo et al. 2005, Ricklefs 2006). Different combinations of phylogenetic, trait and environmental turnover can help distinguish the relative roles of conservatism versus lability in trait and niche evolution, *in situ* diversification and environmental filtering. The power of such analyses emerges at the continental to global

scale, where the replication of large biomes and spatial diversity patterns of large clades enables direct comparison. For example, to examine the relationship between phylogeny, traits and environment in desert ecosystems requires several large, independent desert biomes across a continent (e.g., Sonoran vs. Chihuahuan vs. Mojave Deserts of North America). For analyses that require different species communities and/or phylogenetic clades, a global approach would further be necessary (e.g., North American deserts vs. Australian deserts) to acquire a sufficient number of observations and thus data points for robust analyses. Large-scale approaches such as these allow us to extract meaningful generalities about the underlying macroecological and macroevolutionary processes shaping diversity across the Earth's major biomes.

Expanding analyses to a global scale with more inclusive phylogenies also has the potential to fundamentally change our understanding of major patterns of biodiversity. For example, the tropical Indo-Pacific has long been presumed to be a center of origination for marine fishes (Briggs 2000, Briggs 2003, Cowman and Bellwood 2013), with greater rates of speciation associated with coral reefs (Alfaro et al. 2007, Siqueira et al. 2016). However, equipped with a large phylogeny and geographic data for thousands of marine fish species, Rabosky et al. (2018) recovered a striking pattern of elevated speciation rates in the high latitudes, and depressed speciation rates in the tropics. This inverse relationship between species richness and speciation rates directly contradicts other studies, and leads us to re-think our understanding of the factors generating and maintaining the latitudinal diversity gradient over geologic timescales (Jablonski et al. 2006, Mittelbach et al. 2007, Weir and Schluter 2007).

In order to test hypotheses that relate phylogenetic information to spatial diversity patterns, appropriate metrics are required to summarize the most relevant aspects of species' evolutionary history. Although the specifics will depend on the particular hypotheses being tested, the unit of analysis is typically either taxonomic or geographic. If the unit of analysis is taxonomic, then geographic or environmental data are summarized by species or by

clade (e.g., species-level mean annual temperature then averaged across a clade or latitudinal range of the clade). From these clade-level data, diversification parameters can then be calculated across the phylogenetic tree. There have been considerable methodological developments in the estimation of speciation or net diversification rates for clades. Simple metrics, such as "method-of-moments" estimators of net diversification rate (Magallón and Sanderson 2001) that rely only on stem age and species richness to estimate rates have been used in a number of studies (Adams et al. 2009, Castro-Insua et al. 2018, and many others). Although these metrics are valuable, especially when the goal is to estimate rates for clades that are phylogenetically highly incomplete, their performance can be unreliable (Stadler et al. 2014). In such cases, model-based approaches to estimating diversification rates are preferable, given a phylogeny with reasonably complete species sampling (Rabosky 2017). A number of model-based approaches exist for inferring diversification rates (Alfaro et al. 2009, Morlon et al. 2011, Etienne and Haegeman 2012, Rabosky 2014, Lewitus and Morlon 2016), and the particulars of the phylogeny and the question at hand will determine which approach is most appropriate.

If the unit of analysis is geographic, such as with ecoregions or grid cells, then the goal is to summarize phylogenetic information for assemblages of species in each geographic area, which will most likely not be monophyletic groups. Within this framework, biologists have most often relied on phylogenetic diversity indices derived from either the variance-covariance matrix (e.g., phylogenetic species variability; Helmus et al. 2007) or from the patristic distance matrix (e.g., mean patristic distance, nearest neighbor distance; Webb et al. 2002, Graham and Fine 2008) of the phylogeny to acquire species-specific or pairwise values. These measures reflect different aspects of the phylogenetic relationships between the species in a geographic region of interest as well as characteristics of those species in relation to the full phylogeny (Fritz and Rahbek 2012, Tucker et al. 2016).

Certain biological questions are therefore best addressed at large spatial scales and across large taxonomic groups. By querying phylogenetic relationships, rates of diversification, and

the spatial configuration of species on a continental or global scale, we can test fundamental hypotheses pertaining to factors driving the generation and maintenance of species assemblages, the roles of equilibrial and non-equilibrial forces in shaping species richness patterns, and the ways in which traits or environmental attributes might promote or hinder diversification. I explore these concepts in my dissertation in four major data chapters, as described in greater detail below.

1.1   Overview of chapters

The primary goal of the dissertation research presented herein is to robustly evaluate various aspects of spatial macroevolution and macroecology – both methodological and empirical – that relate to the different components required in the study of diversity at continental and global scales. In this context, I first perform critical assessments of large phylogenies in terms of their value for macroevolutionary study. I then produce and apply a novel global bioclimatic dataset to the study of species distributions today and in the past. I further explore and evaluate macroevolutionary methods, focusing on approaches for estimating speciation rates. Finally, I investigate how biogeographic rates of speciation and dispersal have shaped the latitudinal diversity gradient in marine fishes.

An essential component to the study of geographic patterns of diversification is, of course, the phylogeny. Time-calibrated phylogenies provide the historical framework from which we can make inferences regarding the tempo of diversification. Furthermore, relationships between species and within geographic units, can tell us a great deal about the dispersal and colonization history of a group. Due in part to methodological advances in phylogenetic inference, and to ongoing sequencing efforts by many research labs, a number of large phylogenies have been published over the last decade, primarily for vertebrate groups (birds: Jetz et al. 2012, Burleigh et al. 2105; amphibians: Pyron and Wiens 2011, Jetz and Pyron 2018; squamates: Pyron et al. 2013, Tonini et al. 2016, Zheng and Wiens 2016; fish: Rabosky et

al. 2013, Rabosky et al. 2018), plant groups (flowering plants: Zanne et al. 2014, Smith et al. 2018; seed plants: Smith and Brown 2018), and even all of life (Hedges et al. 2015, Hinchliff et al. 2015). With the availability of such phylogenies growing, biologists interested in diversification, biogeography and trait evolution, have found these to be very attractive resources for both large-scale analyses and targeted analysis of particular clades of interest. However, many tend to take these trees at face value, unaware of topological constraints that may have been applied or of the conflict that may exist between these trees and other studies. A number of studies that use these trees have now been published on a vast range of topics, making a critical evaluation of the consequences of the differences between these trees for macroevolutionary analyses important and timely.

In **Chapter 2** (Title and Rabosky 2017) we evaluated all large phylogenies that have been published for squamate reptiles (Pyron and Burbrink 2014, Wright et al. 2015, Hedges et al. 2015, Zheng and Wiens 2016, Tonini et al. 2016), to explore and highlight these potential issues. We focused on 12 *in situ* radiations of squamates that have occurred in Australia. We first explored topological differences among these trees, as well as compared the crown clade ages of these radiations to those reported in the Australian squamate literature. We then examined how differences in clade ages translate to differences in diversification rates. Furthermore, we assessed whether or not the choice of phylogeny would influence the results of macroevolutionary tests, such as whether or not species richness can be explained by the estimated amount of time clades have diversified.

We found discordance in terms of the crown clade ages of Australian squamate radiations, in particular when the clade ages of these large trees were compared to the more targeted literature on Australian squamates. These differences in clade age resulted in significant differences in net diversification rate estimates. Thus, hypotheses regarding the role of time since divergence or diversification rate evaluated with different phylogenies had the potential to lead to different results and interpretations. We also found some disagreement in topology, with the phylogenies from Tonini et al. (2016) and Hedges et al. (2015) having the greatest

number of differences from the other trees as well as from each other. Ultimately, a number of factors contribute to the differences in the phylogenies that we explored. Two factors with a large impact on tree inconsistencies were 1) the lack of overlap in calibration data used to time-calibrate phylogenies and 2) differences in topological constraints imposed on the backbone of some of the trees. At least some of the issues discussed in the context of squamate reptiles are likely also present in large phylogenies for other groups (for instance, topological constraints were imposed for the bird phylogeny in Jetz et al. 2012). It is important that biologists using these trees be aware of the benefits and potential drawbacks that accompany these otherwise fantastic resources. Over time, we expect that many of the issues discussed in this chapter will be resolved or lessened, as we continue to make progress in developing more robust phylogenetic inference approaches, and as fossil calibration datasets are improved, evaluated and assessed.

At macroecological scales, environmental and climatic conditions have been hypothesized to play a role in generating variation in species richness (Hawkins et al. 2003, Currie et al. 2004). Ecological hypotheses have been proposed that suggest a role for climate in determining the number of individuals or the number of niches that various regions can support (Currie et al. 2004). whereas macroevolutionary hypotheses have been proposed where higher temperatures can lead to greater rates of speciation (Rohde 1992). At finer spatial scales, differences in environmental and climatic conditions are thought to play a role in allopatric speciation via niche conservatism (Peterson et al. 1999, Kozak and Wiens 2010, Hua and Wiens 2013, Jezkova and Wiens 2018). Different species may have different climatic preferences and physiological tolerances, and their ranges may be limited by different factors (Barbet-Massin and Jetz 2014). Therefore, it is important to have access to relevant climatic and environmental datasets in order to properly characterize species' environmental niches.

The vast majority of studies that have modeled species distributions, or that have employed climatic data in macroecological analyses, have relied on the WorldClim dataset (Hijmans et al. 2005, Fick and Hijmans 2017), and in particular on a set of 19 bioclimatic

variables. Although this climatic dataset has and continues to be an important resource, recent research on species distribution modeling has pointed to better performance for models built with variables that are a priori considered to be ecologically relevant to the species in question (Kearney et al. 2008, Doswald et al. 2009, Rödder et al. 2009, Synes and Osborne 2011).

In **Chapter 3** (Title and Bemmels 2018), we viewed this reliance on the 19 bioclimatic variables as a limiting factor in biologists' ability to select the most ecologically relevant variables for species distribution modeling. Other climatic and environmental indices have been described and are used in the literature (Synes et al. 2011, Braunisch et al. 2013, Metzger et al. 2013), but the advantage with WorldClim is the ready availability of global, high resolution datasets under past, present and future climatic conditions. We therefore identified an additional set of 16 climatic and two topographic indices that have been described in the literature and built a comprehensive dataset that makes these variables accessible for multiple spatial resolutions and time periods, globally. We named this dataset ENVIREM (ENVIronmental Rasters for Ecological Modeling).

Using 20 North American vertebrate species as case studies, we then assessed whether or not the availability of the ENVIREM dataset in the pool of potential variables resulted in improved species distribution modeling performance. Through the use of several evaluation metrics, we found that the inclusion of this new dataset led to improvements in a majority of cases. It is worth noting that an improvement in even a single species' distribution model should be viewed as justifying the value of the ENVIREM dataset, as the goal is to provide a greater range of predictor options.

In testing hypotheses regarding the relationship between diversification and geographic or environmental gradients, it is becoming increasingly commonplace to quantify relevant patterns in phylogenetic measures in terms of geographic units, such as grid cells or ecoregions. Although phylogenetic indices based on pairwise patristic distance matrices have been useful for quantifying geographic patterns in phylogenetic relationships and branch length

distributions (Graham and Fine 2008, Tucker et al. 2016), it may be desirable to geographically represent speciation or net diversification rates. This enables us to more explicitly test hypotheses that relate diversification processes to patterns of species richness, such as the latitudinal diversity gradient (Mittelbach et al. 2007). A number of approaches now exist to estimate species-specific "tip rates" of diversification that can be summarized for geographic regions. Despite the growing appeal and use of such approaches (Freckleton et al. 2008, Jetz et al. 2012, Kennedy et al. 2016, Harvey and Rabosky 2017, Quintero and Jetz 2018, Rabosky et al. 2018), there is significant confusion in the literature regarding whether these tip rates represent net diversification or speciation rates, and there has, as of yet, not been a thorough evaluation of the relative performance of available tip rate metrics. A commonly utilized metric, the DR statistic (Jetz et al. 2012) is a model-free metric based on the number of splitting events and internode distances from the root to the tips of a phylogeny. This metric was originally described as a measure of net diversification rate (speciation minus extinction rate); however, Belmaker and Jetz (2015) have since found it to be a better measure of speciation rate. Despite this, studies still continue to use the DR statistic to represent net diversification rate, a fundamentally different measure.

In **Chapter 4**, we compared a number of tip rate metrics, including the inverse of the terminal branch lengths, the node density metric (Freckleton et al. 2008), the DR statistic (Redding and Mooers 2006, Jetz et al. 2012) and BAMM, a model-based approach (Bayesian Analysis of Macroevolutionary Mixtures; Rabosky 2014). We evaluated whether or not the model-free metrics more tightly tracked the rate of speciation or net diversification, and then evaluated the performance of each tip rate approach across a broad range of simulated phylogenies. We found that all tip rate metrics are more accurately tracking speciation rate than net diversification rate. This has implications for the interpretation of large-scale diversity dynamics, as high speciation rates can be coupled with high extinction rates to lead to low net diversification rates. In terms of performance, we found that in all tests, BAMM performed better than the model-free tip rate metrics, exhibiting the greatest accuracy and

the lowest amount of error. The DR statistic also performed reasonably well, and may perform best for very small clades, where BAMM lacks the statistical power to accurately detect rate shifts.

Finally, in **Chapter 5**, we explore the factors that have shaped the latitudinal diversity gradient (LDG) in marine fishes. This group exhibits a strong richness gradient across latitudes, with an order of magnitude more species in the tropics than in the polar regions (Tittensor et al. 2010, Stuart-Smith et al. 2013, Rabosky et al. 2018). Geographic patterns of species richness across regions are thought to have been influenced by the interplay between speciation, extinction and dispersal (Ricklefs 2004, Wiens and Donoghue 2004, Goldberg et al. 2005), as well as by variation in effective carrying capacities (MacArthur 1969, Mittelbach et al. 2007). Several hypotheses have been proposed to explain how these different factors may have generated and continue to maintain this richness gradient. One prominent model is the "out of the tropics" model (Jablonksi et al. 2006), which suggests that speciation rates are highest in the tropics and that there is a net movement of species out from the tropics towards the poles. Rabosky et al. (2018) inferred a large phylogeny for ray-finned fishes and acquired marine fish distribution data for thousands of marine taxa. They found that, paradoxically, rates of speciation exhibit an inverse relationship with latitude, where the highest rates are in the regions with the lowest species richness. Although Jablonski's "out of the tropics" model involves greater tropical speciation rates, the net movement of species is still a core feature of the model, thereby marking the tropics as a major source of diversity shaping the LDG. To test for the predominance of poleward movement of taxa over evolutionary timescales, we modeled transition rates between tropical, temperate and polar regions, based on the phylogenetic and geographic dataset of Rabosky et al. (2018).

In addition to latitudinal transitions, there is reason to believe that global source-sink dynamics in marine fish biogeography might be different for shallow-water and deep-water species. The shallow waters of the oceans exhibit a strong thermal gradient across latitudes, but this gradient becomes weaker with ocean depth, as the environment becomes increas-

ingly homogeneous. Evolutionary transitions from shallow to deep-water are also thought to be relatively infrequent as they require significant adaptations to the lack of light, greater pressure and other major environmental differences that deep-water taxa experience (Brown and Thatje 2014, Priede 2017). Therefore, we might expect that the migration of taxa across tropical, temperate and polar regions would be different at depth.

We found that rates of dispersal, both in shallow and deep-water, are generally biased in a poles-to-tropics direction. Each rate of dispersal in deep-water was faster than its shallow-water counterpart, which lends support to the notion that deep-water species experience a more environmentally homogeneous landscape, with fewer biogeographic barriers than at the surface. In particular, we found that rates of dispersal were greatest out of the Arctic (Briggs 2003, Mecklenburg et al. 2011). This region has a long history of acting as a region of biotic interchange between major ocean basins, especially before climate cooling in the Middle Miocene, thus fueling southward species dispersal to temperate regions and to deeper waters (Mecklenburg et al. 2011).

We additionally found that if we quantify dispersal events, rather than rates, through ancestral state reconstruction, net movement of species does follow an "out of the tropics" scenario, both for shallow and deep-water. Taken together, we view the net movement of species as reflecting tropical inertia, where even with a slow rate of dispersal poleward, the tremendous richness of the tropics increases the frequency of dispersal events over geologic time, thus overcoming the expected pattern based on per-lineage rates alone.

CHAPTER II

Do macrophylogenies yield stable macroevolutionary inferences? An example from squamate reptiles[1]

2.1 Abstract

Advances in the generation, retrieval and analysis of phylogenetic data have enabled researchers to create phylogenies that contain many thousands of taxa. These "macrophylogenies" – large trees that typically derive from megaphylogeny, supermatrix, or supertree approaches – provide researchers with an unprecedented ability to conduct evolutionary analyses across broad phylogenetic scales. Many studies have now used these phylogenies to explore the dynamics of speciation, extinction, and phenotypic evolution across large swaths of the tree of life. These trees are characterized by substantial phylogenetic uncertainty on multiple levels, and the stability of macroevolutionary inferences from these datasets has not been rigorously explored. As a case study, we tested whether five recently published phylogenies for squamate reptiles – each consisting of more than 4000 species – yield congruent inferences about the processes that underlie variation in species richness across replicate evolutionary radiations of Australian snakes and lizards. We find discordance across the five focal phylogenies with respect to clade age and several diversification rate metrics, and in the effects of clade age on species richness. We also find that crown clade ages reported in

---

the literature on these Australian groups are in conflict with all of the large phylogenies examined. Macrophylogenies offer an unprecedented opportunity to address evolutionary and ecological questions at broad phylogenetic scales, but accurately representing the uncertainty that is inherent to such analyses remains a critical challenge for our field.

2.2   Introduction

It is increasingly possible to conduct macro- evolutionary analyses across broad phylogenetic scales, thanks to the recent development of phylogenies that include thousands of species. These data sets enable biologists to explore patterns that may be missed at smaller scales and to test long-standing hypotheses that pertain to continental or global patterns. For the purposes of this article, we use the term "macrophylogeny" to describe phylogenies that (i) are typically produced via supermatrix (typically very large and often sparse genetic data matrices; Driskell et al. 2004), supertree (the grafting of multiple phylogenies to one another; Sanderson et al. 1998), or megaphylogeny (the use of automated pipelines to assemble genetic data matrices; Smith et al. 2009) methods, (ii) include several thousand or more species-level taxa, and (iii) are sufficiently large that it is challenging or impossible to adequately account for numerous sources of phylogenetic uncertainty during tree construction and time calibration. Macrophylogenies provide standardized phylogenetic frameworks from which clades can be extracted and compared and several such trees have been used by many hundreds of studies as a starting point for "downstream" comparative analyses. Such macrophylogenies have been generated for birds (Jetz et al. 2012; Burleigh et al. 2015), mammals (Bininda-Emonds et al. 2007; Faurby and Svenning 2015), amphibians (Pyron and Wiens 2011), squamate reptiles (Pyron et al. 2013; Pyron and Burbrink 2014; Tonini et al. 2016; Zheng and Wiens 2016), ray-finned fishes (Rabosky et al. 2013), flowering plants (Zanne et al. 2014), and all of life (Hedges et al. 2015; Hinchliff et al. 2015). The appearance of such large phylogenies for a broad range of taxa within the last few years can be attributed to advances in sequence data acquisition and alignment (Smith et al. 2009;

Thomson and Shaffer 2009; Hinchliff and Roalson 2013) and computational improvements in phylogeny estimation (Zwickl 2006; Stamatakis 2014).

In this article, we ask a simple question: do different macrophylogenies yield congruent macroevolutionary inferences? Our question is motivated by the observation that the phylogenies listed above have been used by hundreds of subsequent studies involving character evolution, biogeography, comparative analysis, and species diversification. We focus on a single group of organisms—squamate reptiles—because multiple large- scale phylogenies now exist for the group (Pyron and Burbrink 2014; Wright et al. 2015; Hedges et al. 2015; Tonini et al. 2016; Zheng and Wiens 2016).

Although most researchers acknowledge that accommodating phylogenetic uncertainty is important, phylogenies produced by different research groups may differ in fundamental ways, and these differences may not be captured by simply considering posterior distributions of phylogenies (when available) produced by a single research group. In addition, phylogenetic uncertainty is itself rather poorly defined at the scale of macrophylogenies, even when researchers have made comprehensive distributions of phylogenies available for subsequent analyses. For example, a number of studies have used Kuhn et al.'s (2011) distribution of phylogenetic trees for all mammalian species for macroevolutionary analyses (e.g., Price et al. 2012; Rolland et al. 2014). However, this distribution of phylogenies accounts for a very weak form of uncertainty, as the only variation among trees comes from imputation, or the randomized resolution of nodes using taxonomic constraints, for which there were polytomies in the original Bininda-Emonds (2007) tree. Moreover, macrophylogenies are often distinct from smaller phylogenies, in that their size has required researchers to implement strong constraints on taxon monophyly (e.g., Rabosky et al. 2013; Zanne et al. 2014) or to fix the topological backbone of their phylogenies (e.g., Jetz et al. 2012). Finally, computational considerations can lead to challenges in validating tree optimizations due to the size of the data sets (Misof et al. 2013; Wright et al. 2015). Particular genetic samples can cause instability in phylogenetic inference ("rogue taxa"; Thomson and Shaffer 2009) and inference

complications can arise from the fact that data matrices for macrophylogenies constructed with supermatrix approaches often contain mostly missing data, leading to the presence of "tree terraces", or regions of tree space that cause ambiguity in phylogenetic inference (Misof et al. 2013; Hinchliff and Smith 2014; Sanderson et al. 2015).

In this article, we compare macroevolutionary correlates of species richness using five macrophylogenies (Table 1), that have recently been generated for squamate reptiles, to test whether these phylogenies yield congruent results. The Pyron, Wright, Zheng and Tonini trees were similarly inferred via supermatrix approaches. Specifically, Wright et al. (2015) provided a reanalysis of the DNA sequence alignment from Pyron et al. (2013), which they then further optimized in terms of both topology and branch length, thereby generating several alternative phylogenies based on the same sequence data, fossil calibrations and time calibration methodology (in this study, we use their "best" phylogeny with optimized topology and branch lengths). Hedges et al. (2015) produced a timetree of life (TTOL), which was generated by taking a tree representation of the NCBI taxonomy and repeatedly applying time and topological constraints to nodes, iteratively moving from the tips of the tree to the root. These constraints were taken from a database of phylogenies and divergence times that Hedges et al. (2015) compiled from the scientific literature. Although the TTOL has been presented as a resource for studying all of life, a number of studies have used taxonomic subsets for phylogenetic analysis (see Oliveira et al. 2016; Marin and Hedges 2016; Rolland and Salamin 2016). Zheng and Wiens (2016) combined the genetic data matrix from Pyron et al. (2013) with the matrix from another study (Wiens et al. 2012) that sampled up to 44 nuclear genes for 161 squamate species, to generate their phylogeny. Finally, Tonini et al. (2016) generated a squamate phylogeny for 9574 species, 5415 of which had genetic data, the rest of which were imputed using PASTIS (Thomas et al. 2013). In this study, we focus on a posterior distribution of 1000 trees for those taxa with genetic data only, where the topology has been constrained to the maximum likelihood estimate, but where divergence times vary.

Multiple studies have now used these phylogenies as "point estimates" for studying macroevolutionary patterns (Pyron and Burbrink 2014, Hedges et al. 2015, Scharf et al. 2015), and our goal in this article is to address whether these macrophylogenies yield congruent inferences about evolutionary and ecological processes. As a focal question, we investigated the determinants of continental-scale patterns of species richness, a conceptual issue of broad interest to evolutionary biologists and ecologists alike (Mittelbach et al. 2007; Fritz and Rahbek 2012; Kennedy et al. 2014). We focus on Australia, because it is home to multiple distinct radiations of squamate reptiles that differ greatly in species richness (Table 2). To cross-reference these phylogenetic datasets, we also compiled phylogenetic and biogeographic information on Australian squamate clades from the literature (Table 2). We assessed the influence that dataset choice might have on evaluating two hypotheses that pertain to drivers of diversity: the relationship between species richness and clade age and between species richness and diversification rate. Correlations between clade age and species richness have often been examined to assess support for the "time-for-speciation" effect (Stephens and Wiens 2003), which would imply that non-equilibrial factors play an important role in maintaining diversity. The second hypothesis follows from the simple assumption that clades with higher speciation rates should be more diverse than clades with lower speciation rates, although correlations between speciation and extinction rates can potentially weaken or even eliminate such relationships.

In evaluating these hypotheses, we show that these phylogenies are characterized by considerable discord in clade age, with important consequences for macroevolutionary inference. The incongruence that we find appears to be due to many factors, including time-calibration methodology and topological differences. Phylogenetic uncertainty is typically highly conditional on specific datasets and phylogenetic methodology, and our findings suggest an acute need to both quantify and conceptualize uncertainty in its absolute sense.

2.3 Methods

We identified 12 clades of Australian squamates that have radiated *in situ* by identifying groups in which the majority of the species occur on the Australian continent (Greer 1979; Hugall et al. 2008; Rawlings et al. 2008; Sanders et al. 2008; Oliver and Sanders 2009; Skinner et al. 2011; Vidal et al. 2012; Marin et al. 2013; Rabosky et al. 2014a; Sistrom et al. 2014). These clades account for roughly 93 percent of squamate species that occur in Australia (Reptile Database, Uetz and Hošek 2015). For each of these squamate radiations, we identified analogous clades in each of the five phylogenies, as well as in the literature (taxa used to define these clades can be found in Table S1, available on Dryad at `http://dx.doi.org/10.5061/dryad.60js5`). In most cases, we were able to identify equivalent clades across the three phylogenies that represent the Australian radiations. Phylogenies for each clade and for each phylogenetic dataset can be found in the supplement (Figures S1 – S12). The ease with which we identified clades across phylogenies can be categorized into three scenarios.

In the first scenario, we identified in each phylogeny an equivalent node that represents the Australian radiation, and that contains the same set of species (barring sampling disparities). This was the case for Agamidae, Carphodactylidae, Diplodactylidae, Pygopodidae, Sphenomorphinae and Varanidae. In the second scenario, a node was identified that represents the Australian radiation, but due to topological differences, the group was not always monophyletic. This situation arose for the *Egernia* group, the *Eugongylus* group, *Gehyra* and Typhlopidae. Therefore, the node that identifies the clade with the most Australian species was found, sometimes at the expense of either leaving out Australian species or by including a few non-Australian species. Details regarding how we selected nodes for these clades can be found in the supplementary materials (Figures S13 – S16). Finally, the third scenario involved more problematic clades. For Elapidae and Pythonidae, Australian radiations were easily identified for the Pyron, Wright, Zheng and Tonini phylogenies, but equivalent nodes could not be found in the TTOL, where the topology was greatly different from the other

macrophylogenies, and from the literature on these groups, such that no Australian radiation node existed. For the TTOL, we chose the node that is the most recent common ancestor to the species that are thought to belong to the Australian radiation, however, as a result, other non-Australian species were included (Figures S5 and S9 for Elapidae and Pythonidae, respectively). When calculating diversification metrics for these two clades from the TTOL, we used sampled richness rather than known species richness (154 for Elapidae and 78 for Pythonidae) as these clades no longer represent a subset of the set of species described in the literature.

Once equivalent clades were identified, we then extracted the crown clade age of each of these clades, and paired these ages with the known species richness of these clades (as opposed to species richness as sampled in the phylogenies). Where species richness was not found in the literature, we relied on species listings from the Reptile Database (Uetz and Hošek 2015).

We extracted crown clade ages from the five macrophylogenies, and identified crown clade ages from clade-specific literature (hereafter referred to as the "by-clade literature" dataset; Table 2), for a total of six datasets. As Tonini et al. (2016) generated a distribution of 10000 trees, we calculated the mean crown clade age across 1000 trees for each Australian radiation for our analyses, and report both the mean and 95 percent confidence interval (Table 2). We compared these crown clade ages across phylogenetic datasets and used a t-test to assess significance in Pearson correlation coefficients.

For each clade and for each dataset, we calculated three diversification metrics (Table S2). We computed per-lineage net diversification rate, as per equation seven in Magallón and Sanderson (2001), with a relative extinction rate of 0.5. As these clades might not be diversifying under a constant-rate scenario, we also estimated speciation rates using BAMM v2.5 (Rabosky 2014). BAMM is a Bayesian approach which requires an ultrametric phylogeny and identifies shifts in diversification, while allowing for temporal rate heterogeneity. We performed separate BAMM analyses on each Australian squamate clade as extracted

from each of the three phylogenies. We identified appropriate priors for speciation and extinction with BAMMtools (Rabosky et al. 2014b), and defined the prior for the distribution of rate shifts by setting the expected number of shifts at ten. We accounted for incomplete sampling by setting the global sampling fraction according to the known richness for each clade (Table 2), and ran BAMM for 30 million generations (100 million generations for Sphenomorphinae, the largest clade). As the divergence times (but not the topology) vary in the Tonini et al. (2016) distribution of trees, we extracted the clades of interest from 100 trees from the posterior distribution, and ran BAMM on each separately. To get an overall estimate of speciation rate for a given clade from the BAMM analyses, we calculated mean time-integrated rates across each clade phylogeny, averaged across the posterior distribution of BAMM results (Rabosky et al. 2014b). For the Tonini dataset, the clade-specific estimate was simply the median time-integrated rate taken from the distribution of 100 such rates that were estimated for each focal clade. Finally, we calculated the DR statistic, a species-specific measure of speciation rate at the tips of the tree (Jetz et al. 2012; Belmaker and Jetz 2015). We predicted that the DR statistic would be more sensitive to variation in branch lengths near the tips of the tree rather than to uncertainty in crown age. We calculated the DR statistic on phylogenies pruned to the set of taxa with matching names across all phylogenies in order to avoid any influence of sampling intensity. We made one manual adjustment, where we changed the genus of the Australian blindsnakes in all phylogenies to *Anilios* in order to avoid the loss of all Australian blindsnakes in the common set. For the Tonini dataset, we took the average of the DR statistic, calculated across 1000 trees from the posterior distribution. We then calculated the mean DR statistic for each Australian radiation. We examined the congruence across datasets in several predictors of species richness (clade age, diversification rate) and tested whether the relationship between these macroevolutionary predictors and species richness differed across the focal phylogenies.

Many researchers are interested in the relationship between clade age and species richness as well as the effect of time on lineage diversification within geographic regions (McPeek

and Brown 2007; Rabosky et al. 2012, Tank et al. 2015), but the power to detect this relationship from imperfect data on clade ages has rarely been addressed. We explored the effects of uncertainty in clade age on our ability to recover true correlations between crown clade age and species richness by performing a set of power simulations where we empirically parameterized the error variance in clade age from the variance in ages observed across the six empirical datasets. We simulated crown clade ages with fixed correlations to the observed species richness (Table 2), such that the observed variance in age among clades was equal to the mean among-clade variance across the six datasets in this study (variance = 94.595). We then computed the variance in ages for each clade across the focal datasets (e.g., elapids: variance = 182.82). We treated these clade-specific variances as the error distribution for "true" clade age, and – for each simulation – added noise to each simulated age by drawing normal random variables from these distributions. We performed this test across 20 true correlations (1000 simulations per correlation), ranging from zero to 0.95, and tabulated the frequency with which we observed a significant correlation between (log-transformed) richness and clade age.

Topological differences across these macrophylogenies might influence comparisons of crown clade ages, particularly if these differences lead to inconsistencies in how equivalent clades are identified across phylogenies. Furthermore, topological differences can impact time calibration and ultimately diversification analyses, as fossils or secondary calibrations will interact with tree topology in the calibration process. We compared the topologies of the macrophylogenies examined in this study to each other as well as to a maximum likelihood, 161-taxon phylogeny of squamates that was inferred from up to 46 genes (Reeder et al. 2015). We pruned all phylogenies to the set of common taxa and calculated pairwise Robinson-Foulds symmetric distances (Robinson and Foulds 1981) with the phangorn package v2.0.4 (Schliep 2011) in R. This metric determines the total number of branches that would need to be removed or added in order to transform one phylogeny into the other. We then projected these pairwise distances into two-dimensional space using multidimensional scaling.

20

2.4   Results

2.4.1   Pairwise comparisons of phylogenetic datasets

The pairwise relationships in crown clade ages exhibit large amounts of noise (Figure 1), with some pairs exhibiting negative correlations. Significant positive correlations were observed in only four of 15 possible comparisons. The TTOL and the by-clade literature crown clade ages were not found to be significantly positively correlated with any other dataset. Even if we omit Pythonidae – a clade that was highlighted as being problematic during the analogous clade selection process – four pairwise comparisons retain negative correlation coefficients. As Tonini et al. (2016) generated a distribution of trees, we were also able to compare the by-clade literature clade ages to the 95 percent confidence interval from the divergence times of the Tonini trees. Only six of the 12 clades showed overlap in these two datasets (Figure S17).

Pairwise comparisons of diversification metrics exhibit a similar pattern to the comparison of clade ages (Figure 1). This is expected, as crown clade age is a key component of diversification metrics. Net diversification rates for the Pyron, Wright, Zheng and Tonini phylogenies were significantly correlated (or nearly so with Wright – Tonini). Net diversification rates for the TTOL and clade literature were also significantly correlated, likely due to some of this literature being incorporated in the construction of the TTOL. Similar patterns were found with speciation rates from BAMM (Figure S18), and net diversification rates were highly correlated with BAMM speciation rates (Pearson's correlation $r = 0.88$). Mean clade values for the DR statistic were poorly correlated across phylogenies (Figure S19), although individual species values showed relatively high correlations (Figure S20). The DR statistic assumes that phylogenies are fully sampled, but we found that the metric is relatively robust to levels of incomplete sampling in the focal phylogenies (Figure S21).

2.4.2 Implications for drivers of diversity

We observed some conflict across the focal datasets regarding the roles of clade age and diversification rate in the generation of species richness patterns. As the Pyron and Wright trees generally behaved similarly to Zheng and Tonini, we present a subset of the datasets (Figure 2), but all are presented in the supplementary material (Figure S22). Only the Wright phylogeny led to significant relationships between crown clade age and species richness for Australian squamates (Figure S22a). Net diversification rate had a significant positive relationship with species richness for the TTOL and by-clade literature datasets (Figure 2b). Time-variable speciation rates from BAMM exhibited a lack of a relationship with species richness across datasets (Figure 2c), and the DR statistic was positively related to species richness for the TTOL exclusively (Figure 2d).

2.4.3 Power analysis of the richness – clade age relationship

We found that a true correlation between clade age and species richness must be relatively high to detect such a relationship in the presence of estimation error in clade age (Figure 3). For example, even with a true correlation of 0.8, which would be considered a strong relationship in the empirical literature, we would have failed to recover a significant correlation in at least 50 percent of datasets, given the discordance in clade ages across the focal datasets.

2.4.4 Comparisons of topology

We calculated Robinson-Foulds distances for two sets of trees: the macrophylogenies presented throughout this study (3487 taxa in common) and these phylogenies in addition to a backbone phylogeny from Reeder et al. (2015) for 118 taxa in common, representing 113 genera and 57 families. Topological discordance is highlighted across all taxonomic levels in the first analysis, and across deeper parts of the trees in the second analysis. We found that for both tree sets, the TTOL and Tonini trees tended to be most distant from each other

and from all other datasets (Figure 4).

2.5   Discussion

We found that macrophylogenies for squamate reptiles that have been produced by different research groups do not lead to predictable and consistent inferences on the causes of macroevolutionary patterns. Moreover, divergence times generally do not agree with those found in the literature on particular squamate groups, nor are they consistent across macrophylogenies. Our motivation for this study is the observation that published phylogenies – with or without uncertainty – are widely used as primary data for "downstream" macroevolutionary analyses involving diversification, phenotypic evolution, and comparative analyses. It is widely appreciated that calibrating phylogenies to an absolute timescale is a challenging task (Graur and Martin 2004; Hugall et al. 2007; Lee et al. 2009; Smith et al. 2010), although many significant advances have been made (Pyron 2011; Heath et al. 2014; Warnock et al. 2015). Additionally, the inference of macrophylogenies poses inherent difficulties because of the typical sparseness of genetic data for large taxon sets (e.g., Hinchliff and Smith 2014), and the computational challenges of optimizing topologies and branch lengths when the universe of possible trees is large (but see Smith et al. 2010; Sanderson et al. 2015; Wright et al. 2015).

We documented a lack of consistency in both absolute and relative clade ages for Australian squamates across several recent large phylogenetic datasets and the literature. These differences can have a significant impact on macroevolutionary analyses, as shown here with evaluations of the "time-for-speciation" effect and of the potential correlation between species richness and diversification rates, where conclusions varied across datasets (e.g., Figure 2). Similar inconsistencies would likely manifest themselves in the application of comparative methods with trait data. For example, after Pyron and Burbrink (2014) found support for viviparity as the reconstructed root state in squamates, Wright et al. (2015) showed that an improved phylogeny – obtained from the same sequence alignment – led to decreased support

for this controversial finding. If the primary difference between the five macrophylogenies considered here was simply the relative divergence time of a common set of clades, we would have observed highly concordant inferences across all datasets. However, this is not what we find. In the by-clade literature, the 95 percent confidence interval on the crown clade age has been reported for ten clades. Interestingly, out of 50 comparisons (ten clades and five macrophylogenies), we find 25 cases where the macrophylogeny clade age is outside of the 95 percent confidence interval from the clade-specific literature.

Even if clade age was a dominant contributor to species richness patterns across the Australian squamate clades, our analyses suggest that power to infer this relationship would be relatively low given the variance in clade ages observed across the focal datasets (Figure 3). This lack of statistical power would presumably influence measures of diversification. The rank ordering of Australian squamate clades by net diversification rate varies considerably across datasets (Figure S23), which would likely impact any analyses relating traits to diversification across the focal clades.

Why do we observe such discrepancies in clade ages across these datasets? This is a difficult question to answer as the fossil calibrations, genetic markers, calibration methodology, tree topology and error associated with each of these has the potential to lead to differences in node ages. Pyron and Burbrink (2014) constrained the divergence time for Lepidosauria, and applied point estimate constraints for six suprafamilial groups, applying secondary calibrations as inferred by Wiens et al. (2006). Wright et al. (2015) applied the same constraints as Pyron and Burbrink (2014). Zheng and Wiens (2016) applied 13 primary fossil calibrations, mostly as minimum age constraints, which were summarized and employed in a previous study (Mulcahy et al. 2012). Tonini et al. (2016) applied uniform prior distributions on the 95 percent highest posterior densities for ten clade ages reported by Jones et al. (2013). Pyron and Burbrink (2014), Wright et al. (2015) and Zheng and Wiens (2016) used treePL (Smith and O'Meara 2012) to render their phylogenies ultrametric and infer divergence times with these constraints. Tonini et al. (2016) time-calibrated their

phylogenetic backbone using the independent gamma rates model (Ronquist et al. 2012b) in MrBayes (Ronquist et al. 2012a), a relaxed-clock model that is similar to the one implemented in BEAST. As Pyron and Burbrink (2014) and Wright et al. (2015) employed the same genetic data, calibrations and calibration method, we wanted to determine how much of the differences in clade ages between these two trees is due to treePL optimization versus tree topology and branch length differences. We therefore re-calibrated the Pyron phylogeny with the same treePL parameters used by Wright et al. (2015), as provided in their supplementary material. We found that the resulting re-calibrated Pyron phylogeny has virtually identical crown clade ages to the Wright phylogeny (Figure 5) and exhibited very similar patterns and results to the Wright phylogeny in all analyses. Presumably, we would have found an equivalent result had we recalibrated the Wright phylogeny with Pyron et al.'s treePL parameters. This indicates that the majority of the Pyron – Wright discrepancies appears to be due to how the different research teams optimized and ran treePL, or to differences in the versions of treePL that were used for analysis. However, these differences are consequential: clade ages differ by up to 25 million years between these trees; the rank order of clades by diversification rate is in conflict; and clade age was a significant predictor of species richness in Wright but not Pyron.

Although the Pyron, Zheng and Tonini phylogenies have been inferred from similar data matrices of GenBank sequence data, the information used for time calibration is quite dissimilar. Zheng and Wiens applied primary fossil calibrations, whereas Pyron and Burbrink (2014) and Tonini et al. (2016) applied secondary calibrations, as they used clade ages derived in Wiens et al. (2006) and Jones et al. (2013), respectively. If we compare the fossil calibrations used by the source publications – Wiens et al. (2006), Mulcahy et al. (2012) and Jones et al. (2013) – only one fossil was shared in all three, two were shared by Wiens et al. (2006) and Mulcahy et al. (2012), and one was shared by Mulcahy et al. (2012) and Jones et al. (2013). However, as the Pyron and Tonini phylogenies were calibrated with secondary calibrations, use of the same fossil did not lead to use of the same date, or calibration of

the same node. Therefore, despite highly overlapping genetic data, we would not necessarily expect the node ages to be highly concordant.

Overall, there is little overlap between the set of fossil calibrations used in the Pyron, Wright, Zheng and Tonini phylogenies and those used in the clade-specific literature (the TTOL was not calibrated in the same way and therefore cannot be directly compared). The original clade-specific studies for six out of 12 Australian clades (Diplodactylidae, Carphodactylidae, Pygopodidae, Pythonidae, Typhlopidae, Elapidae) did not use any of the same fossil calibrations as the macrophylogenies. Skinner et al. (2011), Sistrom et al. (2014) and Vidal et al. (2012) shared some fossil calibrations with Wiens et al. (2006), and Hugall et al. (2008) shared a fossil calibration with Jones et al. (2013), however the dates used were not always the same. Most significantly, a fossil anguimorph, *Parviraptor*, representing the split between Iguania and Anguimorpha, was used by Skinner et al. (2011) as well as by Sistrom et al. (2014) in the calibration of four of the clades (*Egernia*, *Eugongylus*, Sphenomorphinae and *Gehyra*), and by Wiens et al. (2006), the study from which Pyron and Burbrink (2014) acquired their age constraints. However, Wiens et al. (2006) applied an age that is substantially younger (24 million years) than that used by the other studies, and the identity of *Parviraptor* has since become controversial (Hugall et al. 2007, Sanders and Lee 2008; Caldwell et al. 2015). The use of this potentially problematic earlier date has therefore propagated to the Pyron and Wright phylogenies. The overlap in fossil calibrations between datasets also does not necessarily lead to more or less congruence in clade ages, as *Gehyra* and Varanidae, which share calibrations with Wiens et al. (2006) are not particularly more stable in age across phylogenies. If we were to calculate the standard deviation of the clade ages across datasets as a rough measure of stability of clade age, *Gehyra* and Varanidae would rank tenth and fourth out of 12, respectively.

A worrisome finding is that none of the five macrophylogenies examined here, which represent all of the available large-scale species-level phylogenies for squamates at the time of writing, have crown clade ages that correspond to those found in the literature on these

particular groups of Australian squamates (Figure 1). The lack of congruence with the literature is such that analyses of diversification or trait evolution will likely be at odds with the literature on these groups, for example relating to the timing of biogeographic events. Although detailed analyses focused on particular clades will be necessary to gain a full understanding of the source of these discrepancies, the minimal overlap in calibrations used by the macrophylogeny studies in comparison to the clade-specific studies might play a role. Overall, the TTOL had clade ages that were most consistent with the literature on particular clades, although the relationship was not statistically significant (Figure 1). This is not surprising, as the divergence times in the TTOL are taken directly from the literature. However, in our examination of the TTOL topology, we found many phylogenetic relationships that are at odds with current understanding of squamate relationships, and this may be due to the fact that construction of the TTOL started with a tree representation of the NCBI taxonomy, with the subsequent random resolution of polytomies using a birth-death polytomy resolver (Hedges et al. 2015). Two of the more extreme examples that we identified are Pythonidae and Elapidae (the two clades in our "third scenario"). It has been established that Indo-Australian pythons form a monophyletic group (Rawlings et al. 2008; Reynolds et al. 2014). In the TTOL, we found that the Australian pythons are polyphyletic; the MRCA of Australian pythons in this tree defines a clade that also contains a biogeographically disparate set of taxa from another family (Boidae, Figure S9). Similarly, all previous analyses have suggested monophyly of Australian elapid snakes (Keogh 1998). However, in the TTOL, we found South American coral snakes, African and Asian cobras, and other non-Australian elapid species interspersed throughout the Australian elapids, rendering this group polyphyletic (Figure S5). We found a number of other surprising relationships in Typhlopidae and *Eugongylus*-clade skinks. These issues in the TTOL were also captured by our tree topology analyses, with the TTOL having the greatest Robinson-Foulds distance from all other trees (Figure 4). Considering that all trees exhibit such large distances with the TTOL, it is quite possible that other major topological problems exist in the TTOL

27

outside of the Australian clades.

It is critical to recognize that "phylogenetic uncertainty", as presented in the empirical literature, is a metric that is highly conditional on the data, models, and other constraints that enter a particular analysis. The distinction between absolute and conditional uncertainty is likely to be especially acute for macrophylogenies, which frequently utilize a number of constraining assumptions to ensure computational tractability. To illustrate this point, we compared the Tonini phylogeny with the squamate phylogeny from Reeder et al. (2015). Tonini et al. (2016) generated a distribution of 10000 trees to accommodate phylogenetic uncertainty. However, the backbone of these phylogenies is constrained in terms of topology (but not divergence times) and as a result does not vary across the posterior distribution. The topology of this constrained backbone is different from Reeder et al. (2015; Figure 4b). Therefore, any diversification or phenotypic evolution study that integrates across the posterior distribution of trees from Tonini et al. (2016) will not include the topology inferred by Reeder et al. (2015). A similar situation can be found with recent phylogenies for birds, where Jetz et al. (2012) also constrained the backbone of their tree to a topology of higher-level avian relationships that ultimately was not recovered by a more recent genomic study of avian phylogenetics (Prum et al. 2015). These two cases provide examples of phylogenetic uncertainty that reflect built-in constraints (in this case, of the backbone) and therefore fail to capture topologies that ultimately are being found to be more probable with larger or more complete data matrices. Incorporating uncertainty in backbone topologies into the final distribution of trees would allow one to account more thoroughly for phylogenetic uncertainty. Approaches that assess absolute phylogenetic uncertainty can potentially help assuage these issues (Brown 2014b), including the use of posterior predictive simulations (Brown 2014a), but the sheer size of the datasets considered here may render such approaches impractical in many cases.

2.6   Conclusions

We found that differences in timing and topology across the phylogenies we examined led to considerable variation in the crown clade ages of Australian squamate groups, both in an absolute and relative sense. This variation in age, in turn, influences our ability to recover macroevolutionary determinants of species richness. As Australian squamates belong to distantly related clades that span the squamate tree, it is very likely that similar problems exist for comparisons of other groups as well. Depending on the phylogenetic breadth of the group being analyzed, and the ages of the nodes involved, the severity of such problems might vary, as the age of nodes of interest can become less reliable with distance from the calibration nodes if molecular rate variation is high (Duchêne et al. 2014). As we found that the fossil calibrations used in the clade-specific literature were almost entirely non-overlapping with the calibrations used by the macrophylogeny studies, it would appear that there is an opportunity to evaluate and incorporate more of these calibrations into large-scale phylogenetic analyses for squamates, as incorporating calibration nodes throughout the tree should lead to more reliable estimates of node ages at both deep and shallow timescales (Duchêne et al. 2014). Ultimately, however, fossil calibrations need to be critically evaluated in terms of both their placement and age, and further research into identifying the most appropriate fossils for time calibration of phylogenies should be a priority (Near et al. 2005; Warnock et al. 2015). Additionally, the vast majority of sensitivity and simulation-based studies on divergence dating has focused on the program BEAST (Drummond and Rambaut 2007), whereas phylogenies like those discussed here are too large to be calibrated with this program. Simulation studies are needed to assess the performance and behavior of programs that can work with large phylogenies, such as treePL (Smith and O'Meara 2012). We suspect that, on account of constraints and other factors commonly used in macrophylogeny construction, phylogenetic uncertainty is generally more conditional than typically acknowledged. The conditional nature of this uncertainty can give a false sense of confidence in both phylogenies and inferences derived from those phylogenies, as we have shown with the

comparison of the macrophylogenies to Reeder et al. (2015).

Ultimately, the issues discussed in this study are likely to be resolved with the careful evaluation and placement of fossil calibrations, larger, more complete molecular data matrices, and a more rigorous presentation of phylogenetic uncertainty in the absolute sense, for example through the use of posterior predictive simulations (Brown 2014a). In the meantime, we recommend that the Tonini or Zheng phylogenies be used over the Pyron or Wright phylogenies, as the Wright tree was demonstrated to be an improvement over the Pyron tree, and the Tonini and Zheng trees were inferred from larger genetic data matrices and improved fossil information. Additionally, given the method of construction of the TTOL and the discrepancies in topology observed here, we generally do not recommend use of this phylogeny for downstream comparative analyses involving squamates. Finally, in conducting analyses with macrophylogenies, concordance with the taxon-specific literature should be evaluated if the timing of biogeographic events is important for the interpretation of results.

2.7   Data Archiving

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.60js5.

2.8   Acknowledgements

2.9 Funding
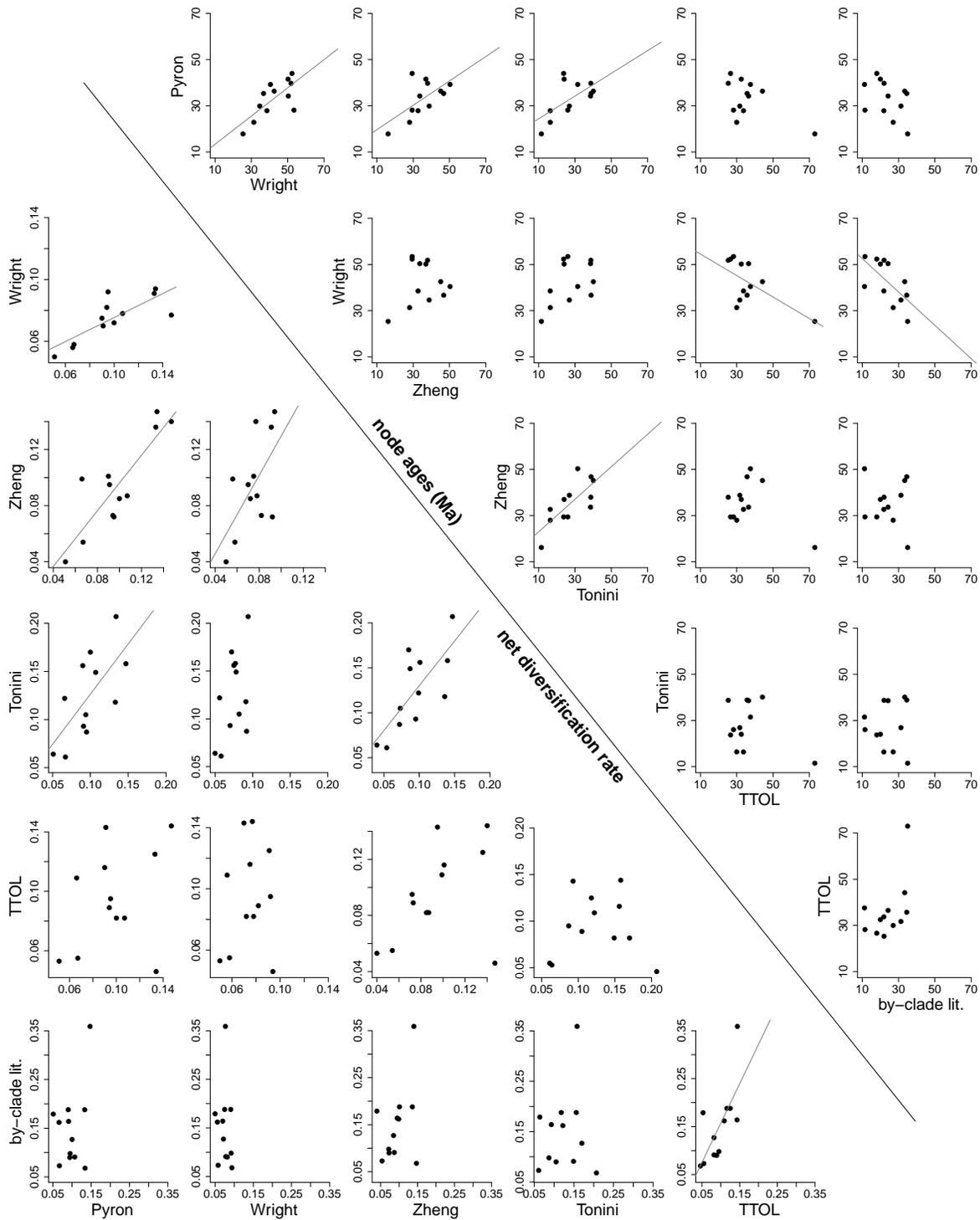
Figure 2.1: Pairwise comparisons of crown clade ages (Ma) and net diversification rates (species per million years), for Australian squamate clades in each of six different phylogenetic datasets. A line of best fit was plotted when the t-test p-value was below 0.05. We generally find noisy relationships between datasets, with negative trends in several cases.
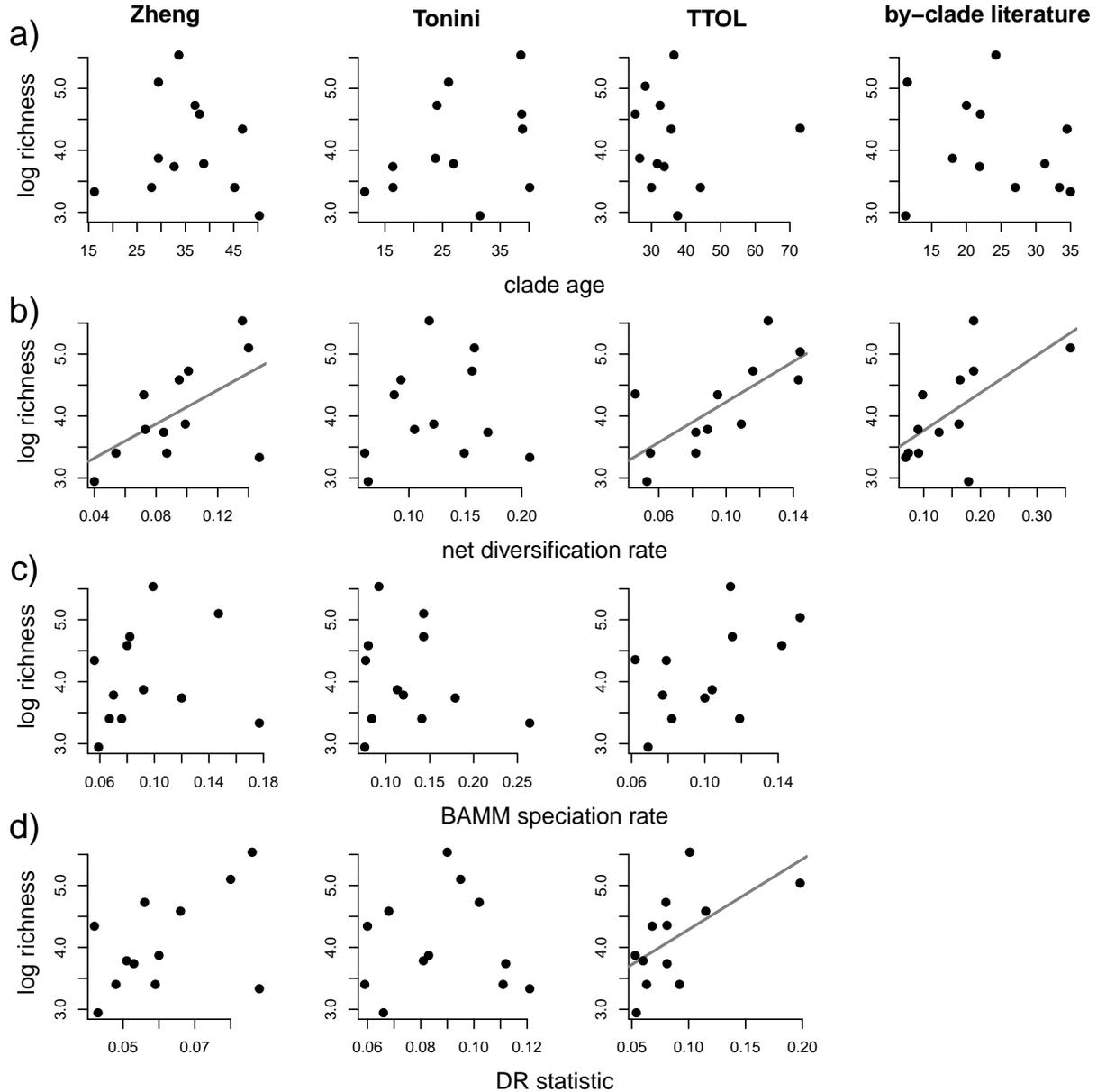
Figure 2.2: Influence of dataset selection on macroevolutionary hypotheses. Examination of the influence of dataset selection on **(a)** the relationship between clade age, **(b)** net diversification rate, **(c)** BAMM speciation rates and **(d)** the DR net diversification rate statistic on log species richness, for a subset of datasets (all datasets are shown in Figure S22). Each column represents the same phylogenetic dataset. BAMM speciation rates and the DR statistic are not available for the by-clade literature as separate, well-sampled clade phylogenies were not available. The line of best fit is only plotted for those relationships that are statistically significant, as determined through linear regression. The choice of phylogeny can have an appreciable impact on the outcome of these hypotheses regarding controls on species richness; see text for discussion.
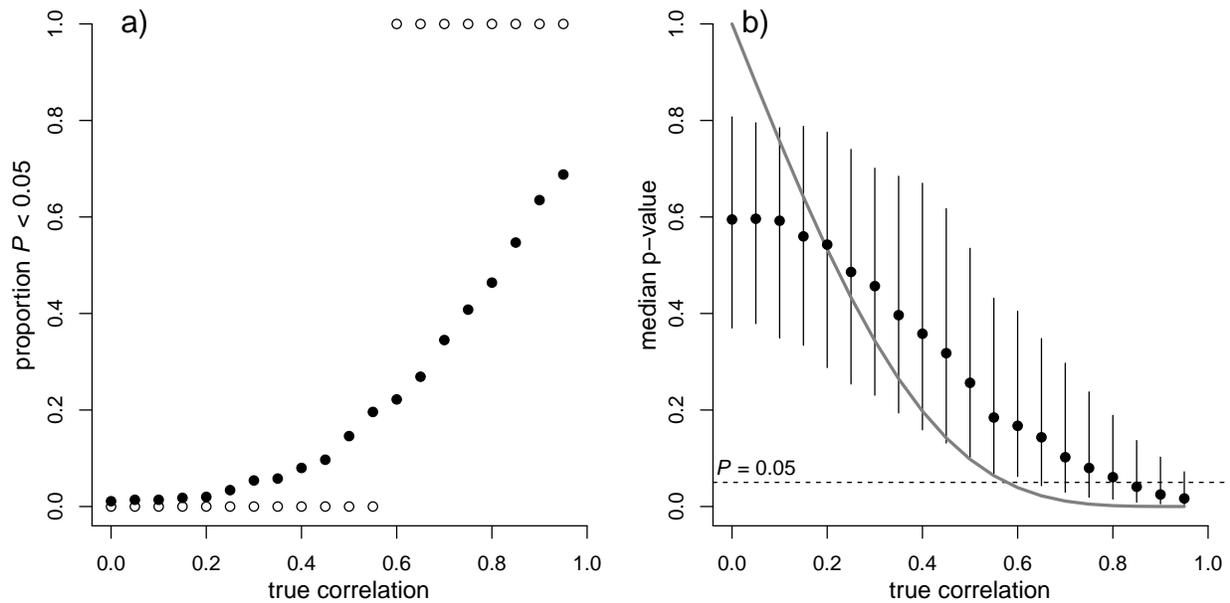
Figure 2.3: Impact of uncertainty in clade age on the analysis of age-richness relationships, across a range of true correlations. Age-richness datasets were simulated with fixed correlations and uncertainty was parameterized from the observed variation in age across the focal clades. **(a)** Proportion of simulations that recover a significant relationship between the two variables without (white points) and with (black points) noise added to the crown clade ages. **(b)** Median p-value and associated interquartile range across 1000 simulations of crown clade ages with added noise; gray line illustrates theoretical relationship (e.g., no uncertainty in age).
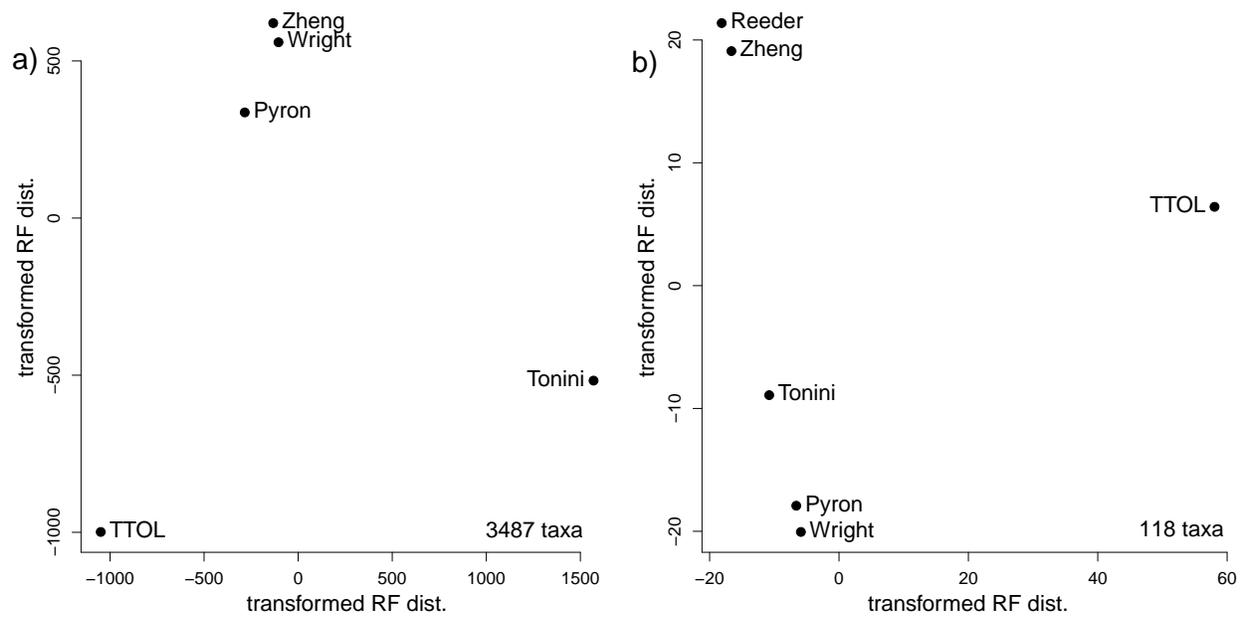
Figure 2.4: Comparisons of phylogenies in terms of Robinson-Foulds distances. Relative positions in terms of tree topology of the five macrophylogenies alone **(a)**, and **(b)** with the inclusion of the phylogeny from Reeder et al. (2015). Pairwise Robinson-Foulds distances have been projected to two-dimensional space via multi-dimensional scaling. Several phylogenies tend to cluster together, and the Tonini and TTOL phylogenies have greater topological differences relative to all other trees.
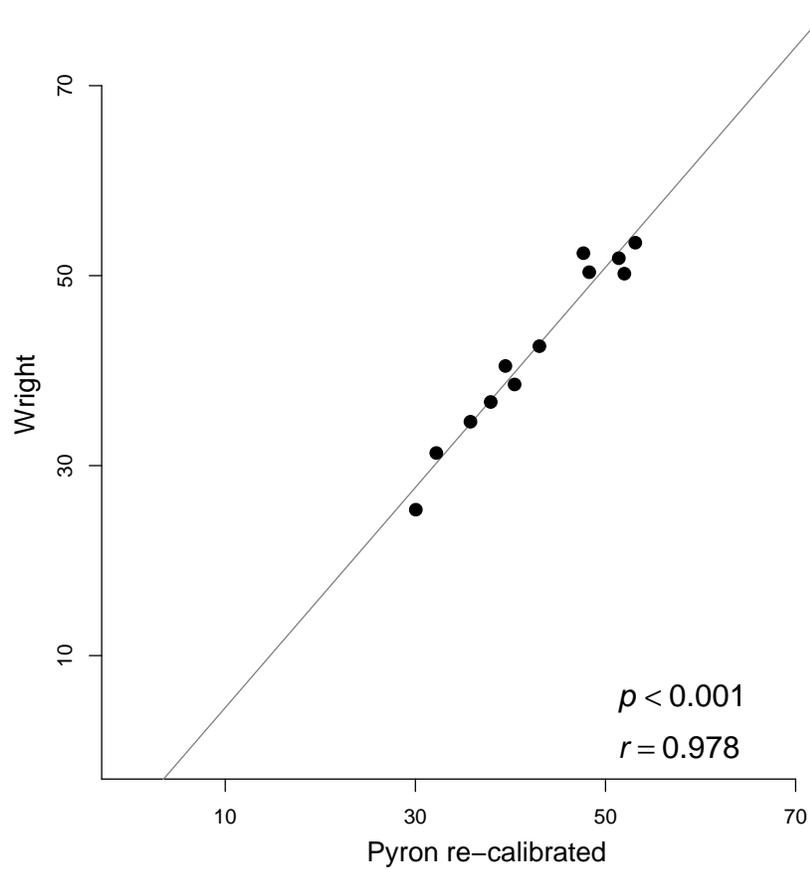
Figure 2.5: Relationship between divergence time estimates for Australian squamate clades using Pyron and Wright phylogenies, after re-calibrating the Pyron tree with treePL parameters used in the Wright tree calibration. Pearson correlation coefficient and t-test p-value are reported in the corner of the scatterplot. It is clear that the differences in crown clade ages between the Pyron and Wright phylogenies are mostly due to details of the calibration approach rather than the topology *per se*.

References	Tree name	Tips	Max # genes/sp	% sites with missing data	Inference method	Time calibration model	Calibration method	Calibrations	Calibration source
Pyron et al. 2013, Pyron & Burbrink 2014	Py-ron	4161	12	81	RAxML	Penalized likelihood rate smoothing	treePL	1 primary, 6 seconday	Wiens et al. 2006
Wright et al. 2015	Wright	4161	12	81	RAxML, Garli	Penalized likelihood rate smoothing	treePL	1 primary, 6 secondary	Wiens et al. 2006
Hedges et al. 2015	TTOL	4220	-	-	NCBI taxon-omy	-	Divergence time database	-	-
Zheng & Wiens 2016	Zheng	4172	52	92	RAxML	Penalized likelihood rate smoothing	treePL	13 primary	Mulcahy et al. 2012
Tonini et al. 2016	Tonini	5415	17	83	RAxML	MrBayes	MrBayes	10 secondary	Jones et al. 2013

Table 2.1: Summary of the macrophylogenies examined in this study.

clade	Pyron	Wright	Zheng	Tonini	TTOL	cladeLit	richness	source for date	source for richness
Agamidae	39.74	51.84	37.94	38.74 [28.5 - 51.2]	25.30	22.00	98	Hugall et al. 2008	Reptile Database
Carphodactylidae	36.35	42.58	45.19	40.13 [25.3 - 57.3]	44.15	33.40	30	Oliver et al. 2009	Reptile Database
Diplodactylidae	35.34	36.70	46.81	38.88 [24.6 - 49.8]	35.70	34.50	77	Oliver et al. 2009	Reptile Database
Egernia	44.01	52.37	29.43	23.76 [16.4 - 32.1]	26.60	18.00	48	Skinner et al. 2011	Reptile Database
Elapidae	28.13	53.47	29.44	26.04 [18.5 - 35]	28.20	11.50	164	Sanders et al. 2008	Reptile Database
Eugongylus	41.54	50.21	37.00	24.03 [16.4 - 31.5]	32.50	20.00	113	Skinner et al. 2011	Reptile Database
Gehyra	39.25	40.49	50.31	31.5 [23.1 - 39.4]	37.56	11.24	19	Sistrom et al. 2014	Sistrom et al. 2014
Pygopodidae	29.89	34.62	38.81	26.9 [17.4 - 38.0]	31.69	31.30	44	Oliver et al. 2009	Reptile Database
Pythonidae	17.83	25.36	16.19	11.5 [6.5 - 15.8]	72.98	35.00	28	Rawlings et al. 2008	Reptile Database
Sphenomorphinae	34.24	50.37	33.65	38.6 [29.9 - 47.8]	36.50	24.24	254	Rabosky et al. 2014	Rabosky et al. 2014
Typhlopidae	27.87	38.55	32.69	16.36 [11.0 - 22.5]	33.69	21.90	42	Marin et al. 2012	Marin et al. 2013
Varanidae	22.85	31.33	28.01	16.39 [10.8 - 21.8]	29.97	27.04	30	Vidal et al. 2012	Reptile Database

Table 2.2: Crown clade ages and species richness for each clade. 'Source for date' lists the reference from which crown clade age was taken for the clade literature age. 'Source for richness' lists the reference from which species richness was taken. The 95 percent confidence interval is also reported for 1000 trees from Tonini et al. (2016).

Figure S2.1: Phylogenies for Agamidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
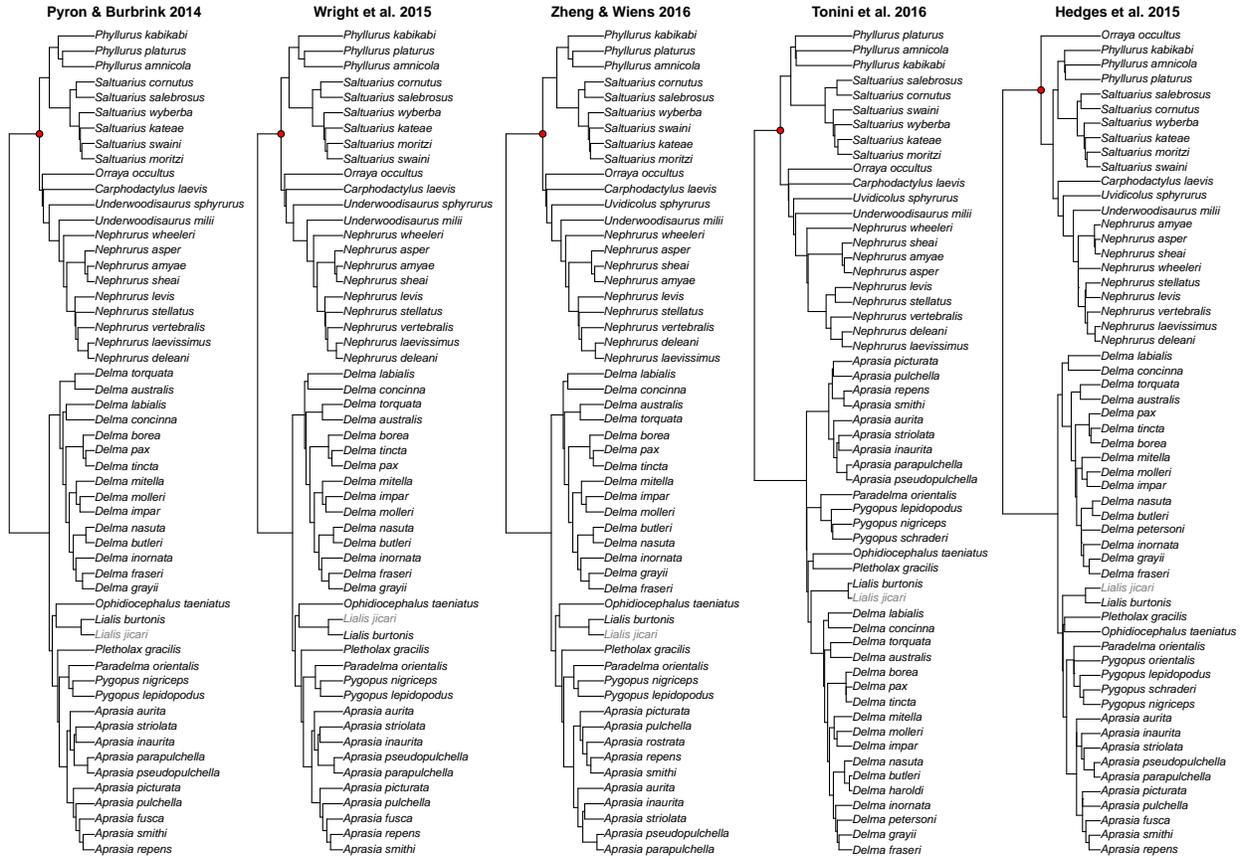
Figure S2.2: Phylogenies for Carphodactylidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
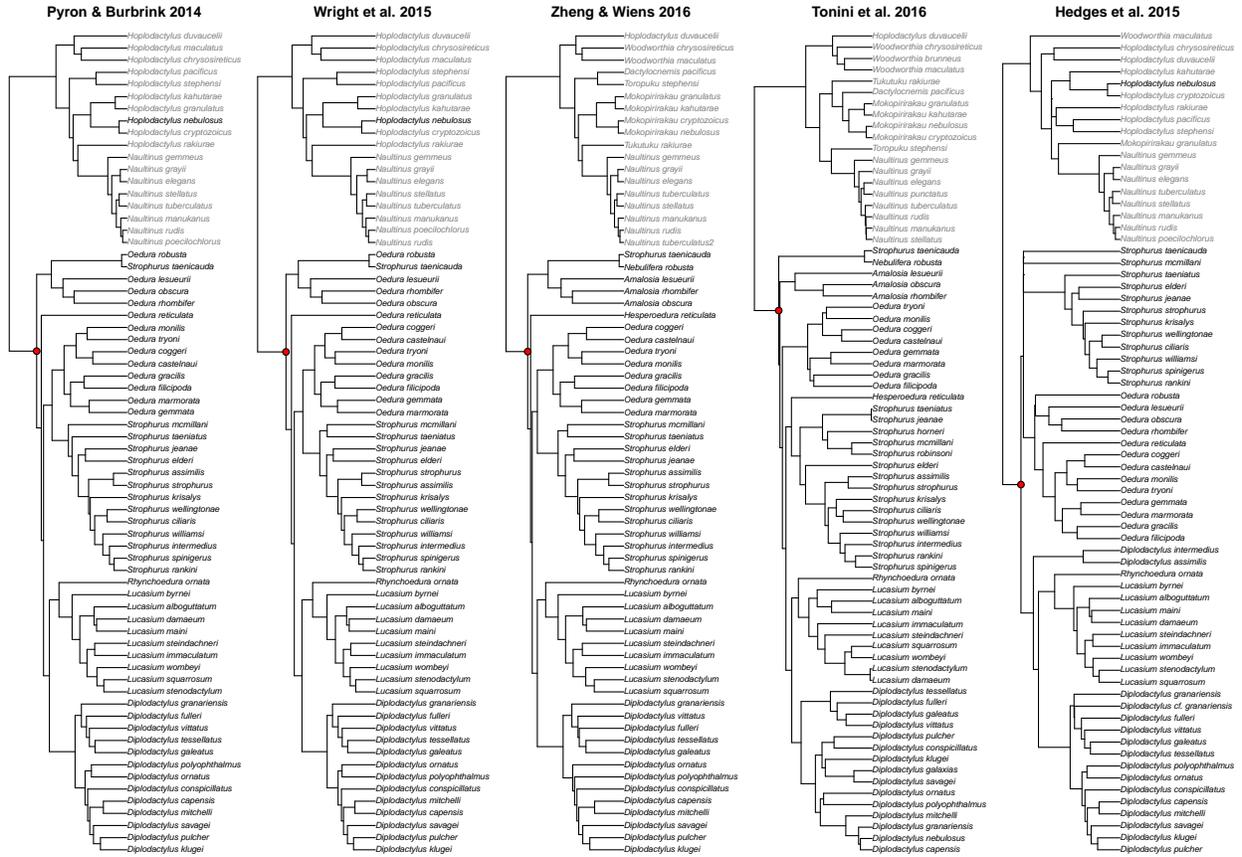
Figure S2.3: Phylogenies for Diplodactylidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
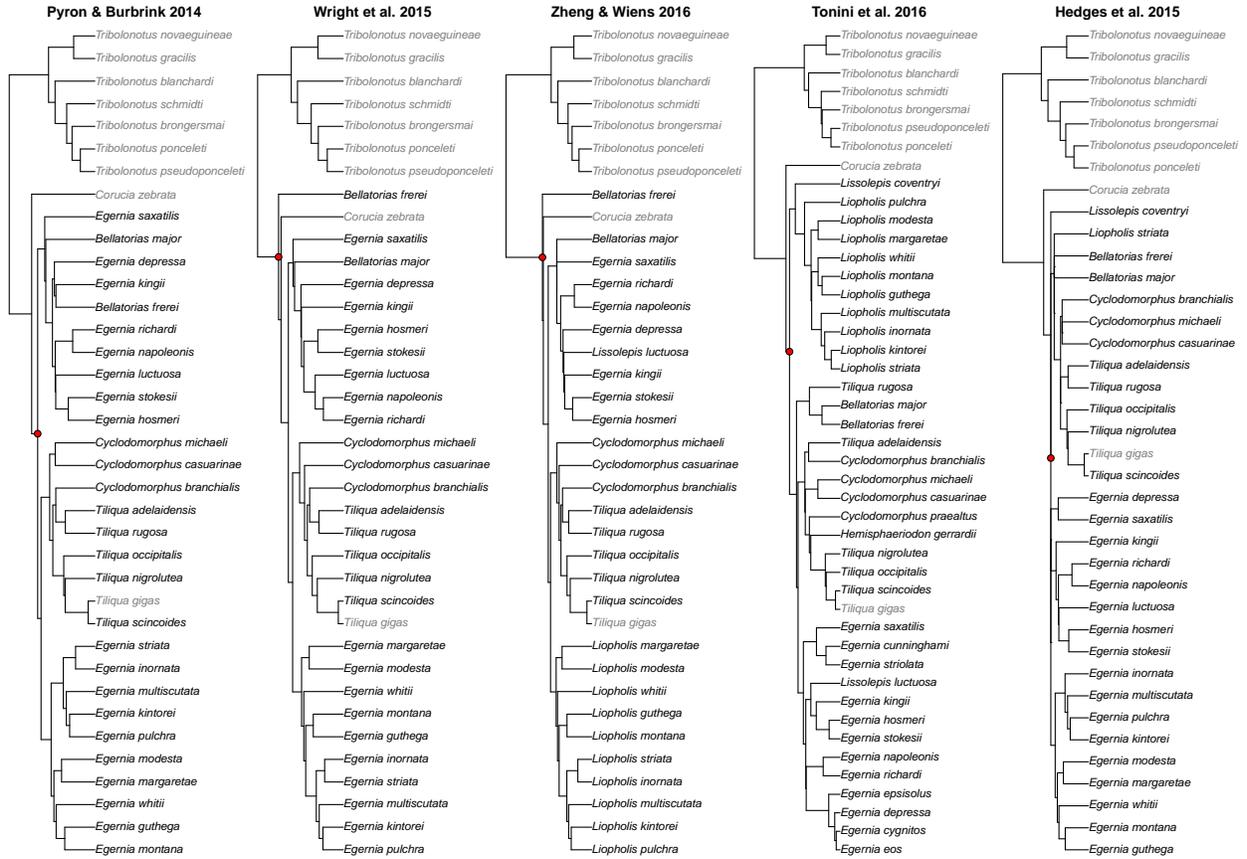
Figure S2.4: Phylogenies for *Egernia*. The node used to define the clade is indicated in red, and taxa in black occur in Australia.

Figure S2.5: Phylogenies for Elapidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
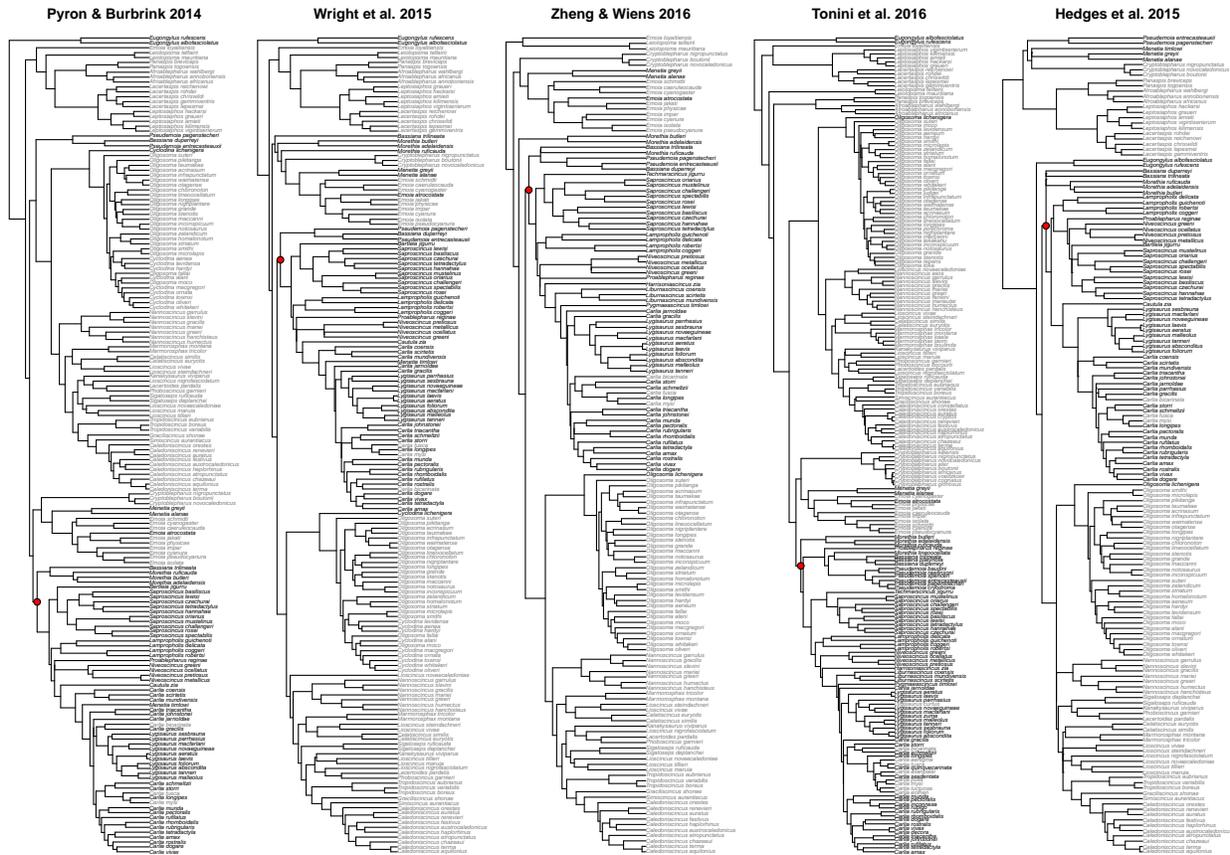
Figure S2.6: Phylogenies for *Eugongylus*. The node used to define the clade is indicated in red, and taxa in black occur in Australia.

44
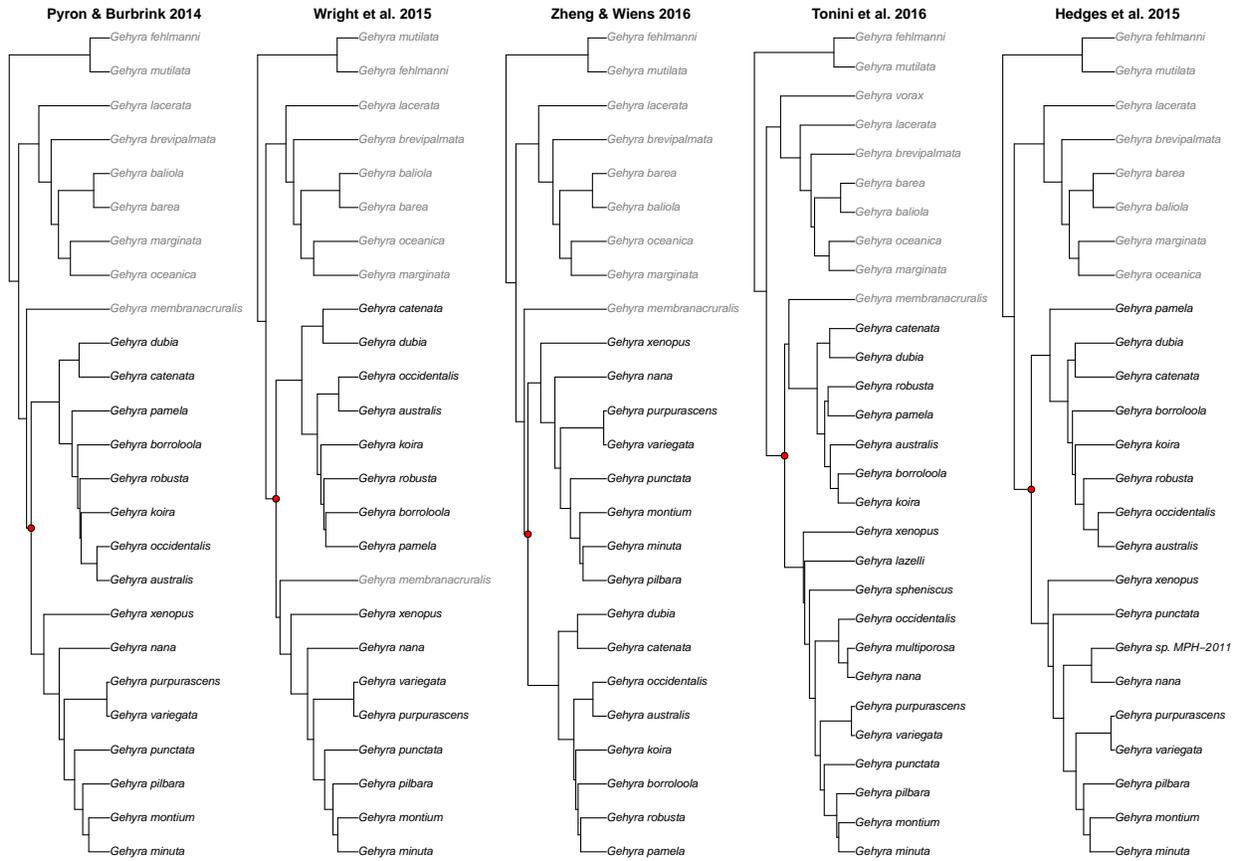
Figure S2.7: Phylogenies for *Gehyra*. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
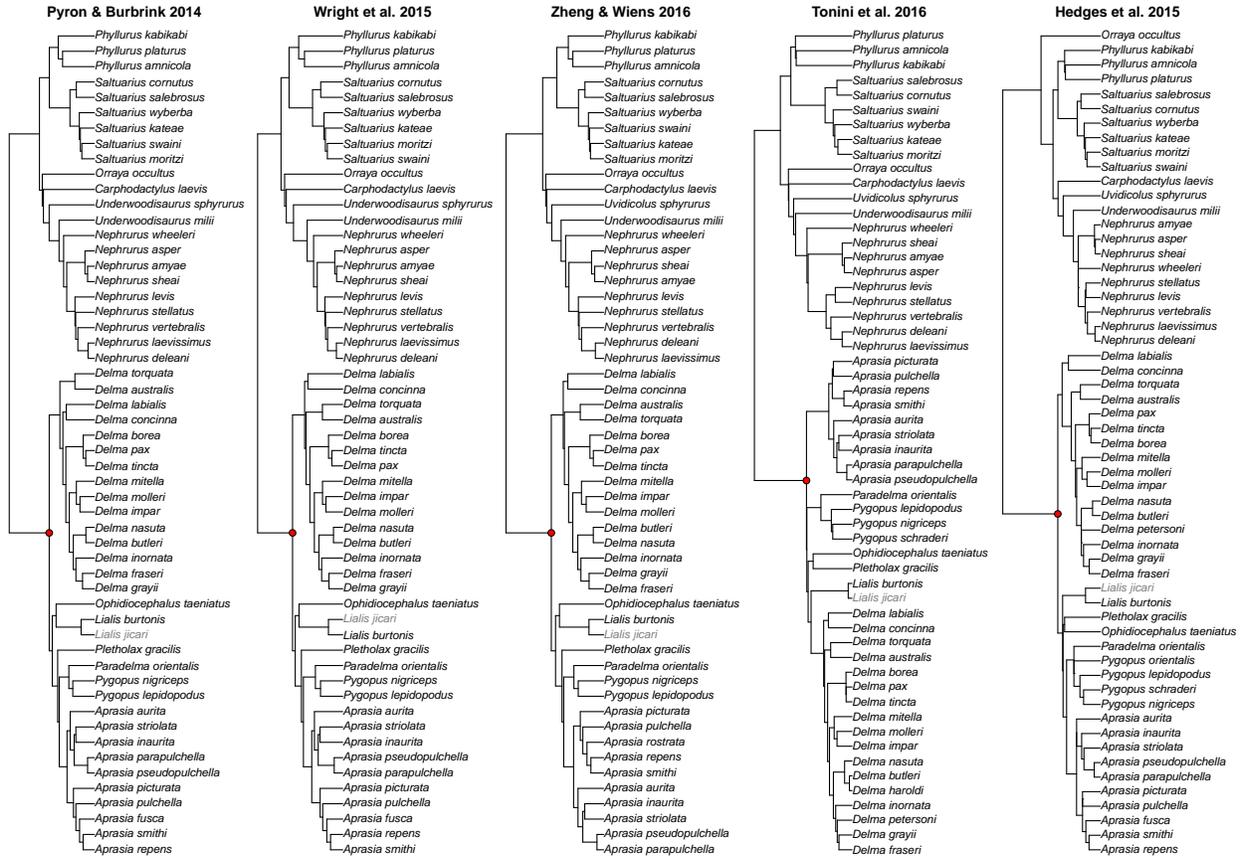
Figure S2.8: Phylogenies for Pygopodidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
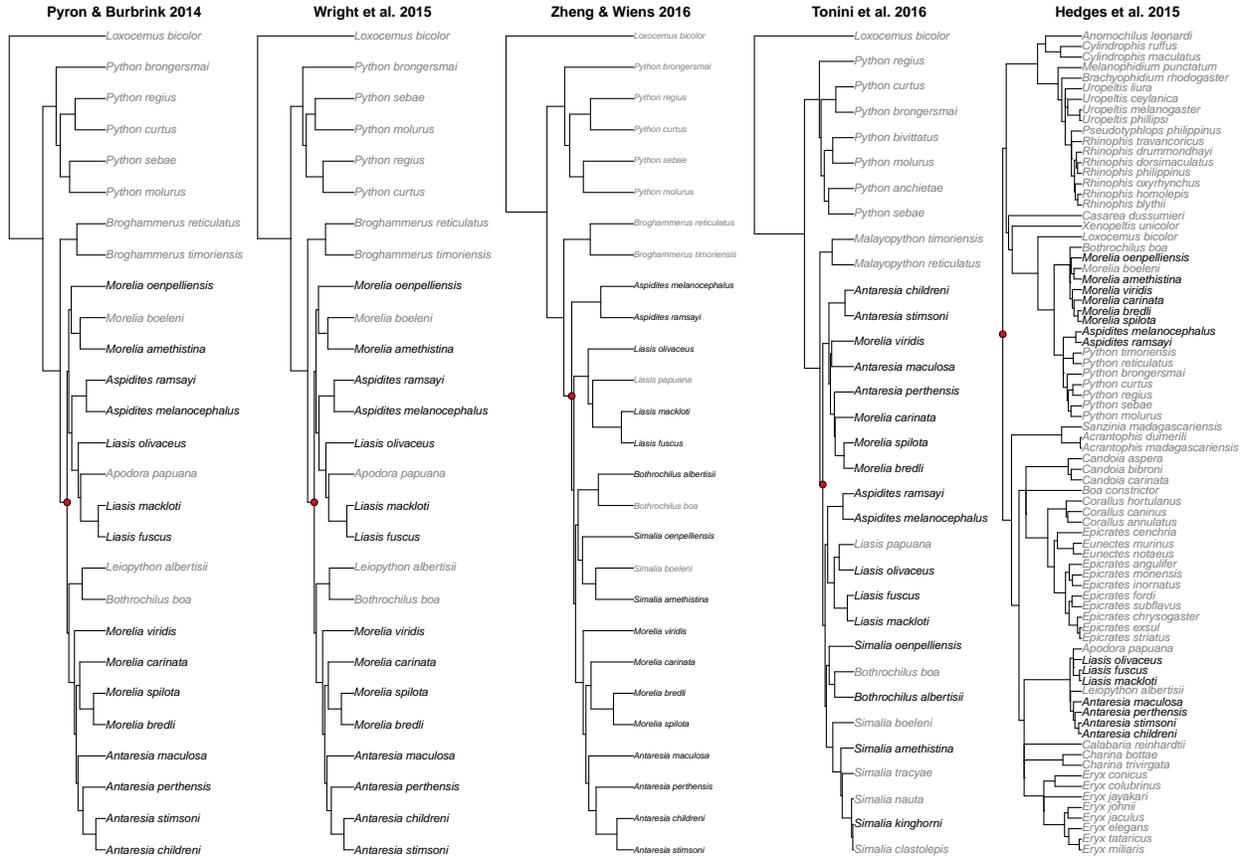
Figure S2.9: Phylogenies for Pythonidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
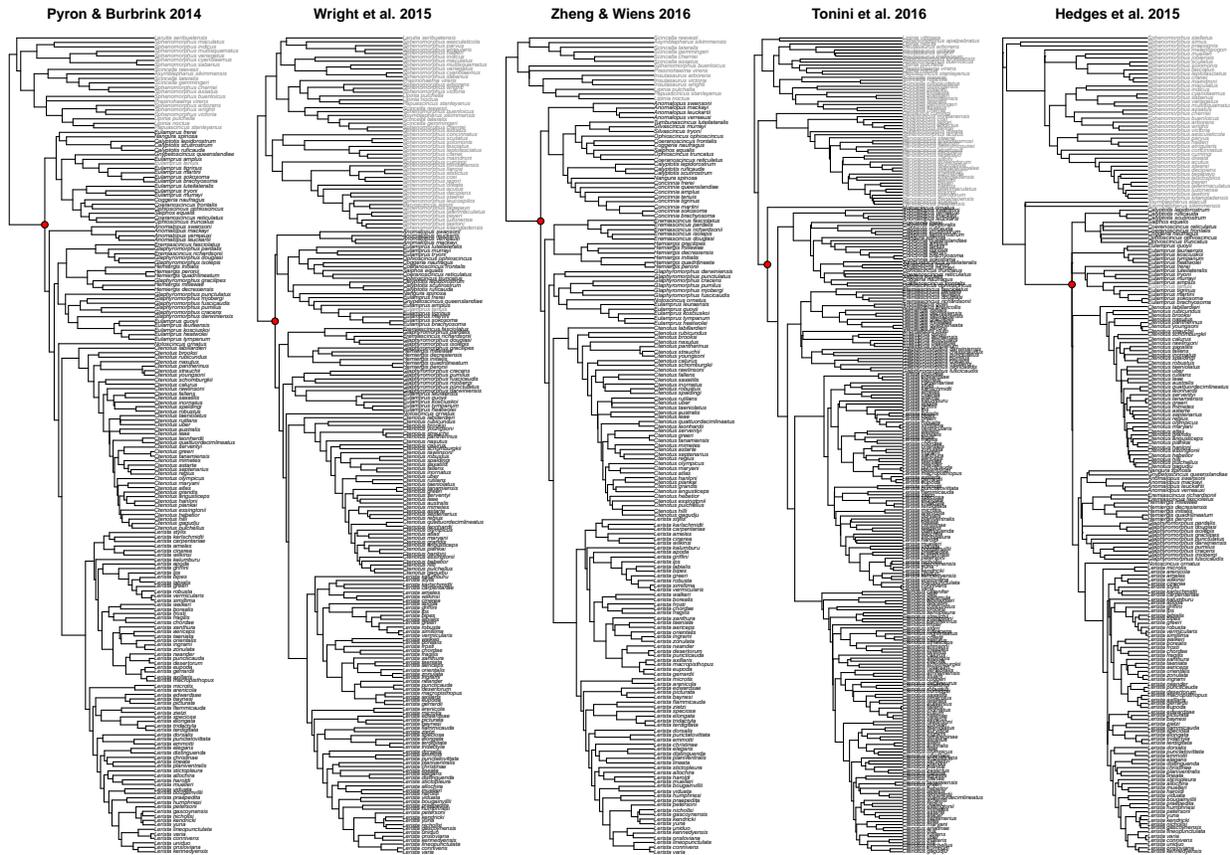
Figure S2.10: Phylogenies for Sphenomorphinae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.

48
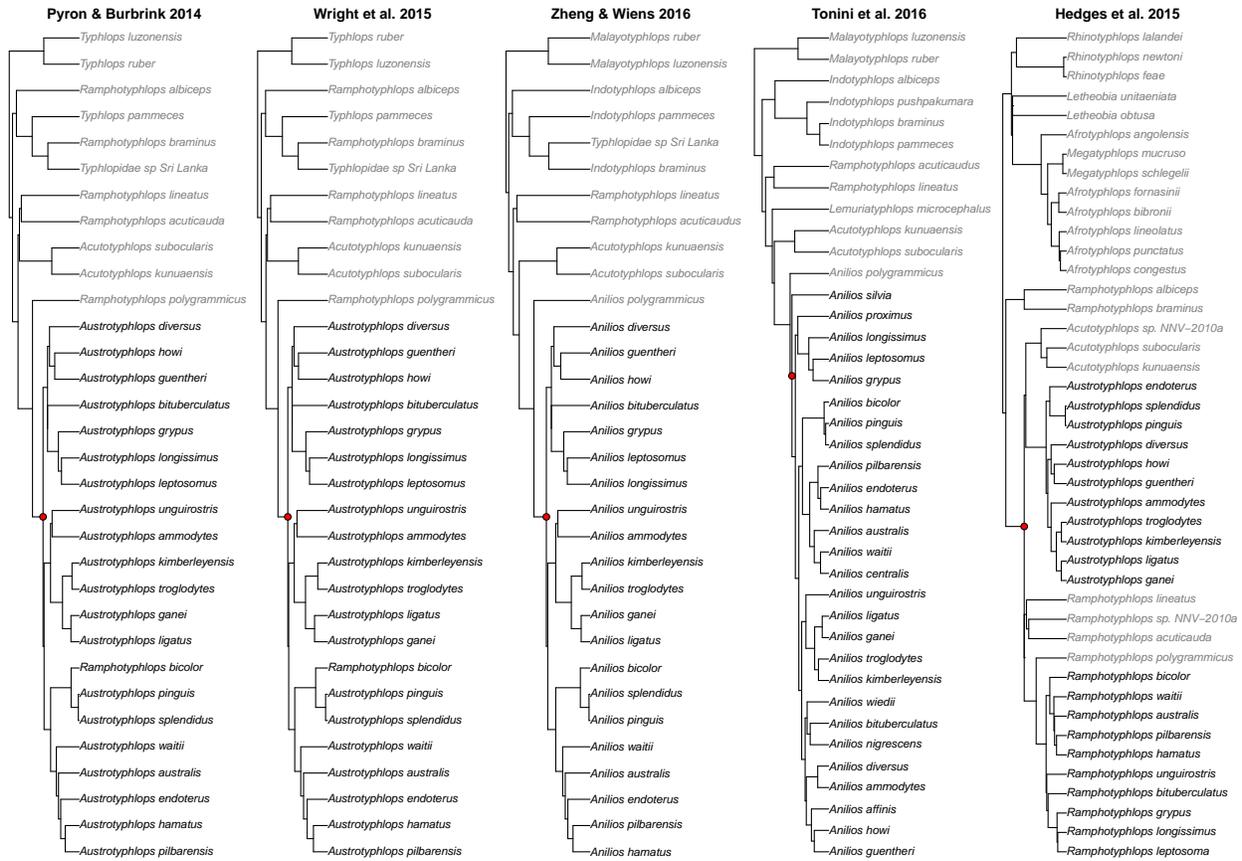
Figure S2.11: Phylogenies for Typhlopidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.

Figure S2.12: Phylogenies for Varanidae. The node used to define the clade is indicated in red, and taxa in black occur in Australia.
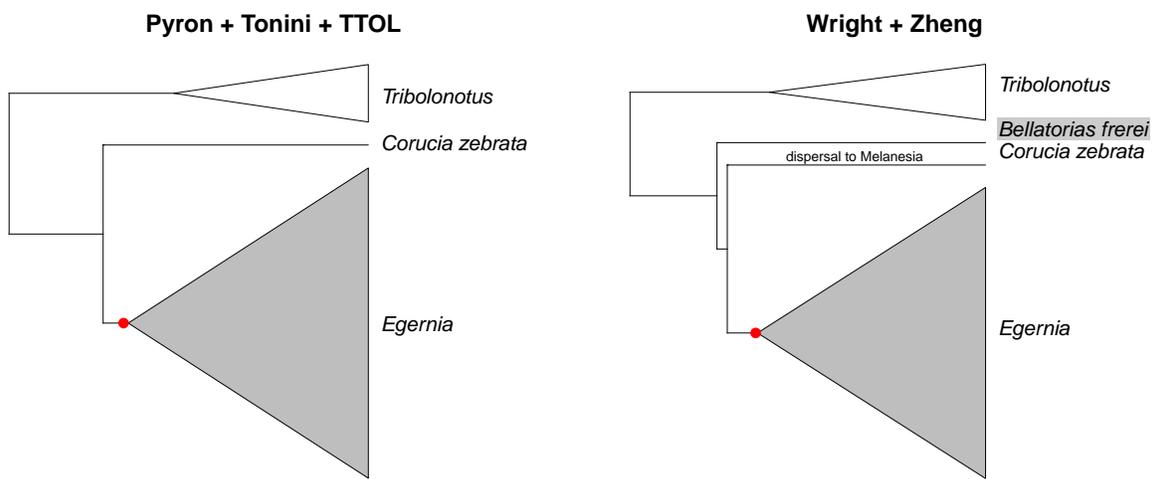
Figure S2.13: Geographic affinities of species belonging to, and closely related to the *Egernia* group clade. The node highlighted in red is the node that was used to define the Australian radiation. Different biogeographic scenarios are inferred from different phylogenies, such as the dispersal of *Corucia zebrata* out of Australia with the Wright and Zheng phylogenies. Clades and species in gray occur in Australia.
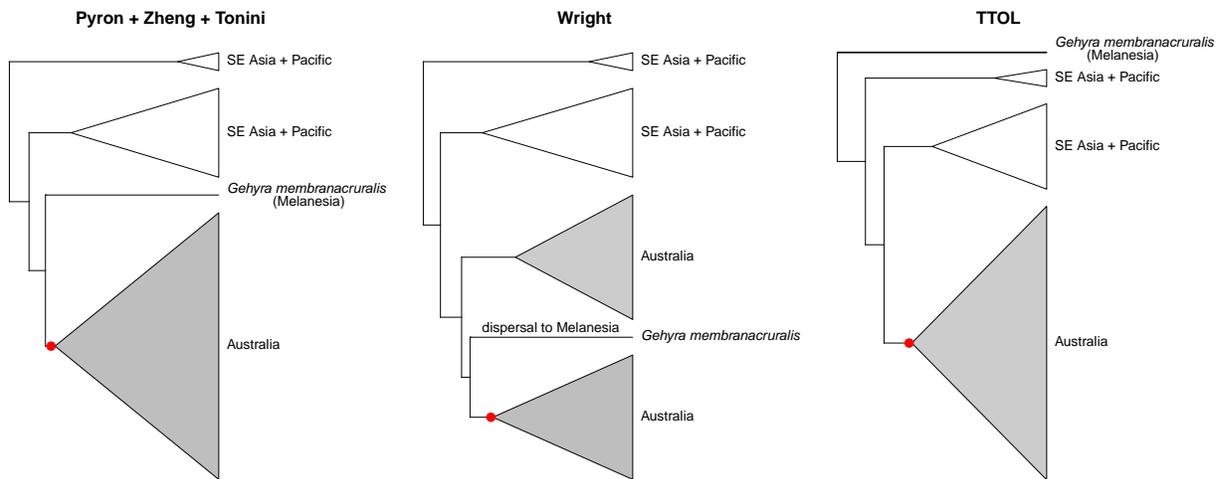
Figure S2.14: Geographic affinities of species belonging to, and closely related to the *Gehyra* clade. The node highlighted in red is the node that was used to define the Australian radiation. Different biogeographic scenarios are inferred from different phylogenies, such as the dispersal of *Gehyra membranacruralis* out of Australia with the Wright phylogeny.
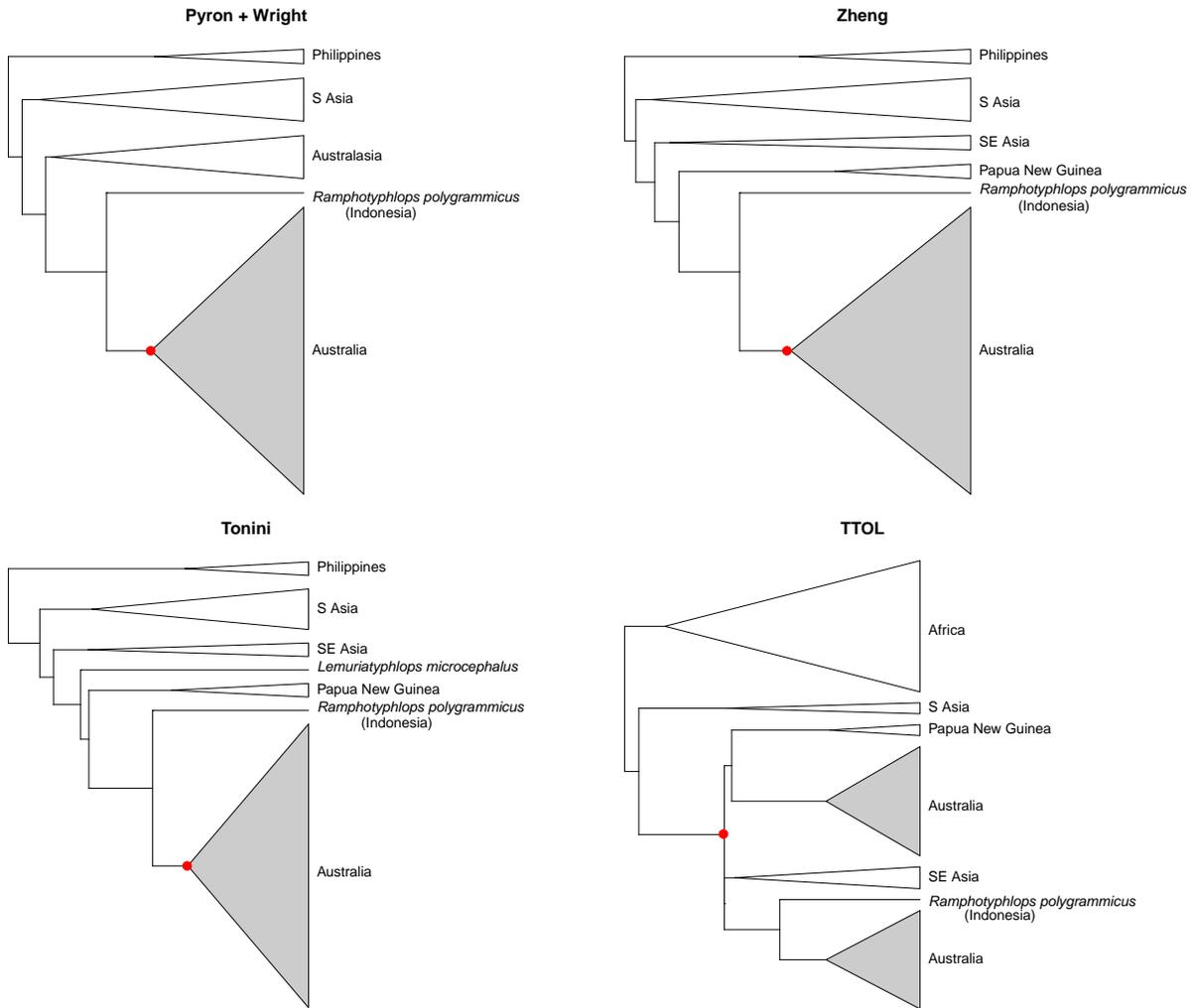
Figure S2.15: Geographic affinities of species belonging to, and closely related to the Typhlopidae clade. The node highlighted in red is the node that was used to define the Australian radiation. Different biogeographic scenarios are inferred from different phylogenies. In the TTOL, the Australian blind snakes are separated into two clades. Clades in gray occur on the Australian mainland.
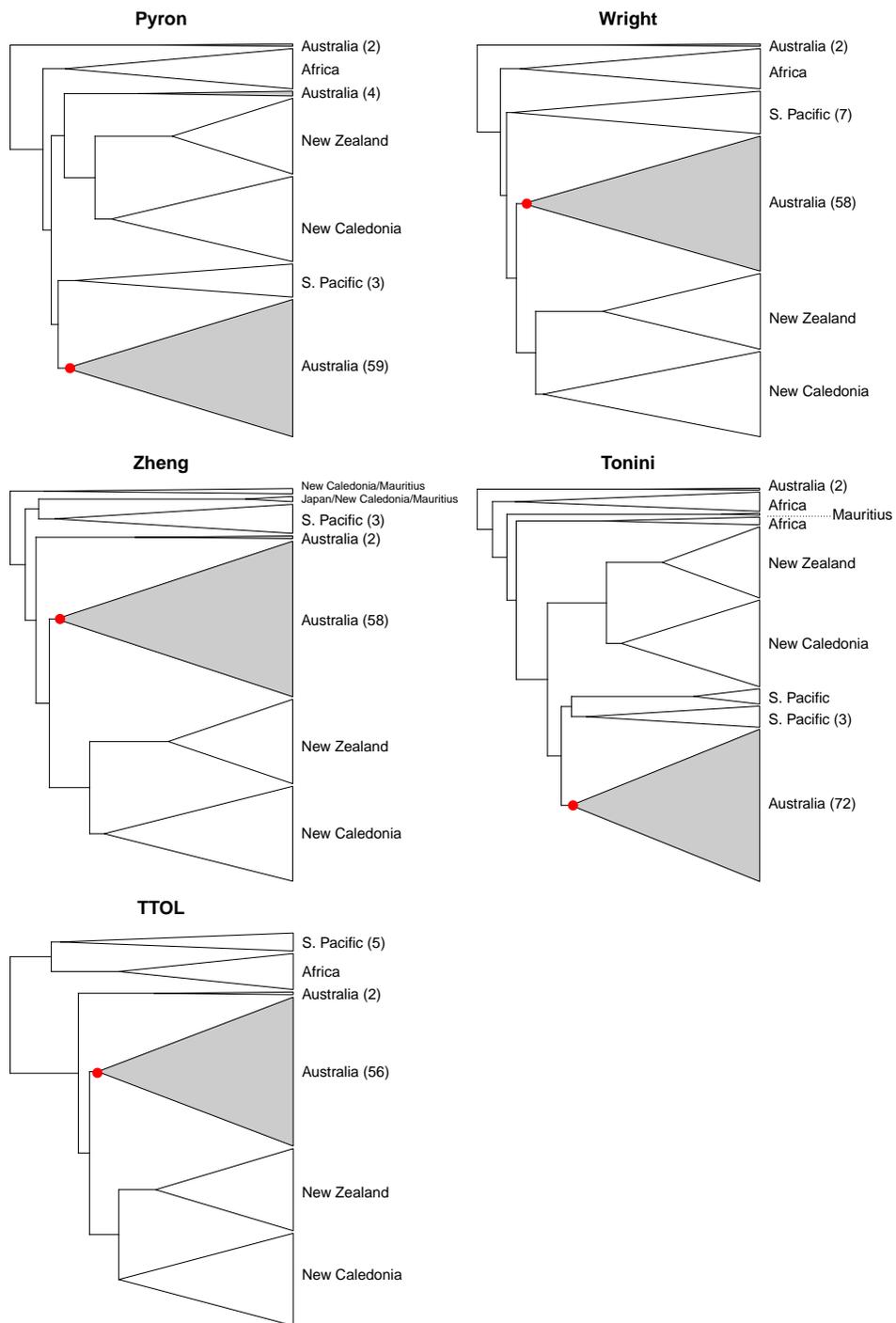
Figure S2.16: Geographic affinities of species belonging to, and closely related to the *Eugongylus* group clade. The node highlighted in red is the node that was used to define the Australian radiation. Different biogeographic scenarios are inferred from different phylogenies, specifically the phylogenetic placement of non-core-clade Australian species. Species numbers in parentheses indicate the number of Australian species. Clades in gray occur on the Australian mainland.
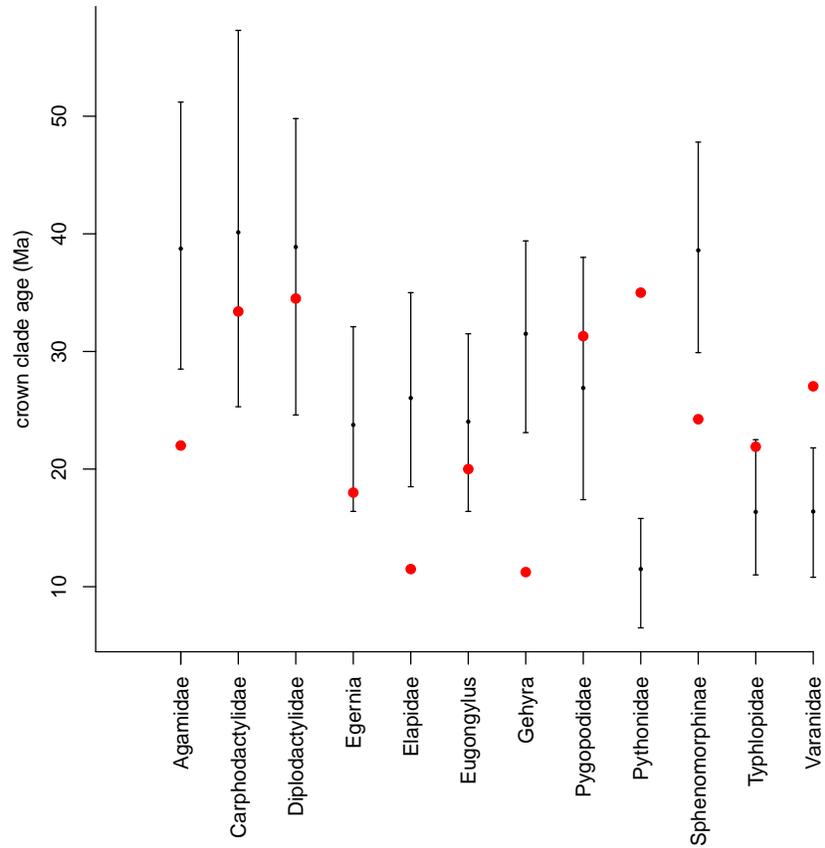
Figure S2.17: Comparison of crown clade ages between clade literature and Tonini et al. Clade ages from the by-clade literature are shown as red points, and are compared to the 95 percent confidence interval of crown clade ages from Tonini et al. (2016), as summarized from 1000 trees. In a majority of cases, the divergence dates from the clade-literature are outside of, or close to the edge of the 95 percent confidence intervals.
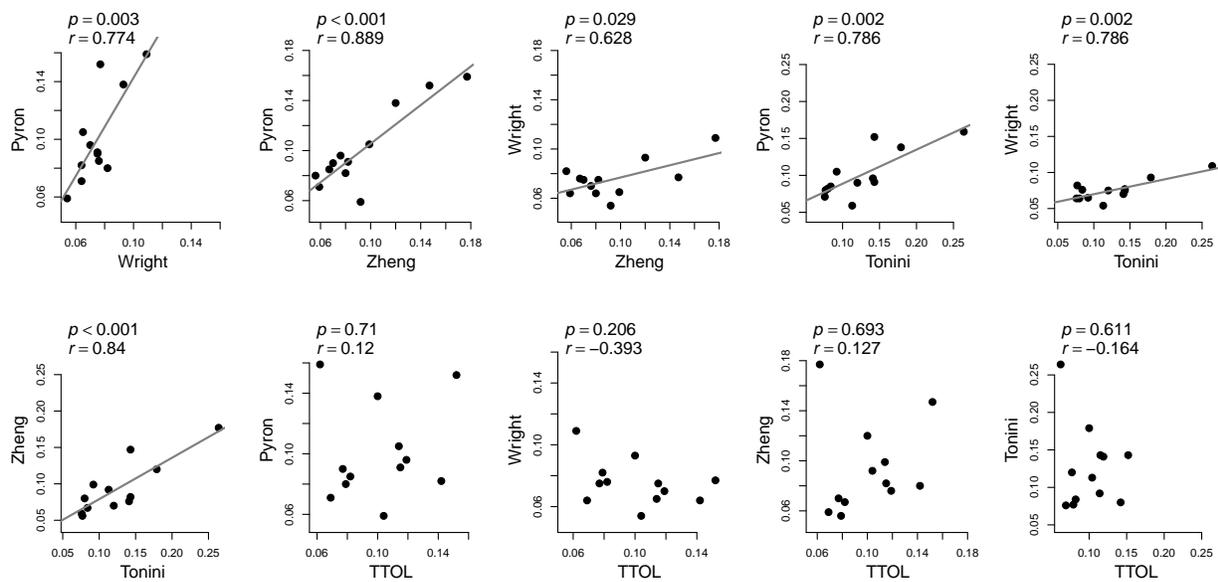
Figure S2.18: Pairwise comparisons of BAMM speciation rates for Australian squamate clades. Pearson correlation statistics are displayed in the top left corner of each plot.
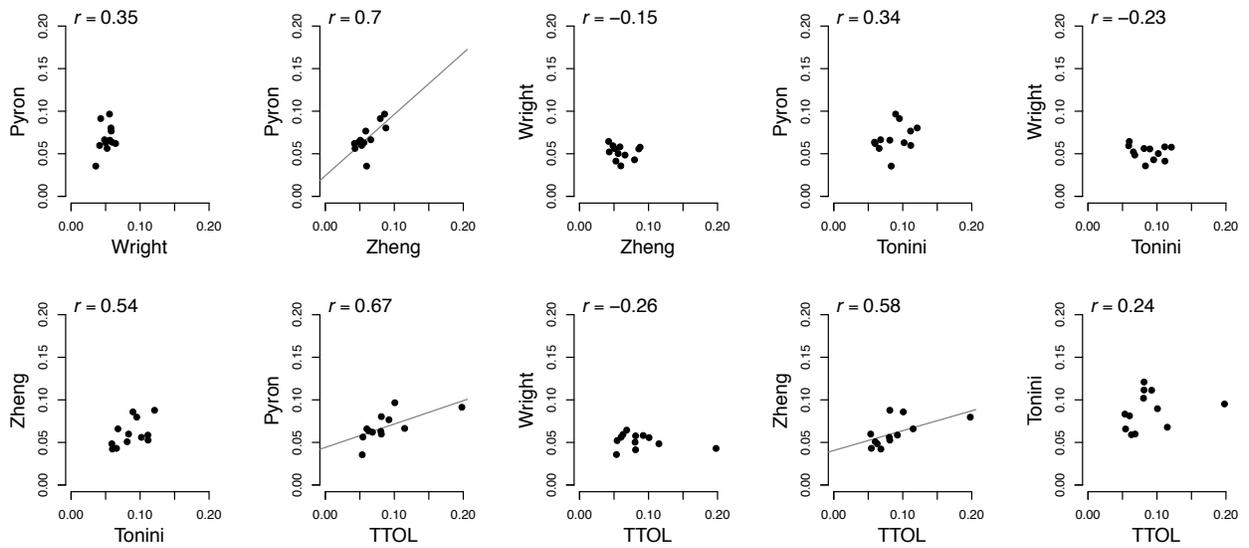
Figure S2.19: Pairwise comparisons of the DR statistic, averaged by Australian squamate clade. Pearson correlation statistics are displayed in the top left corner of each plot.
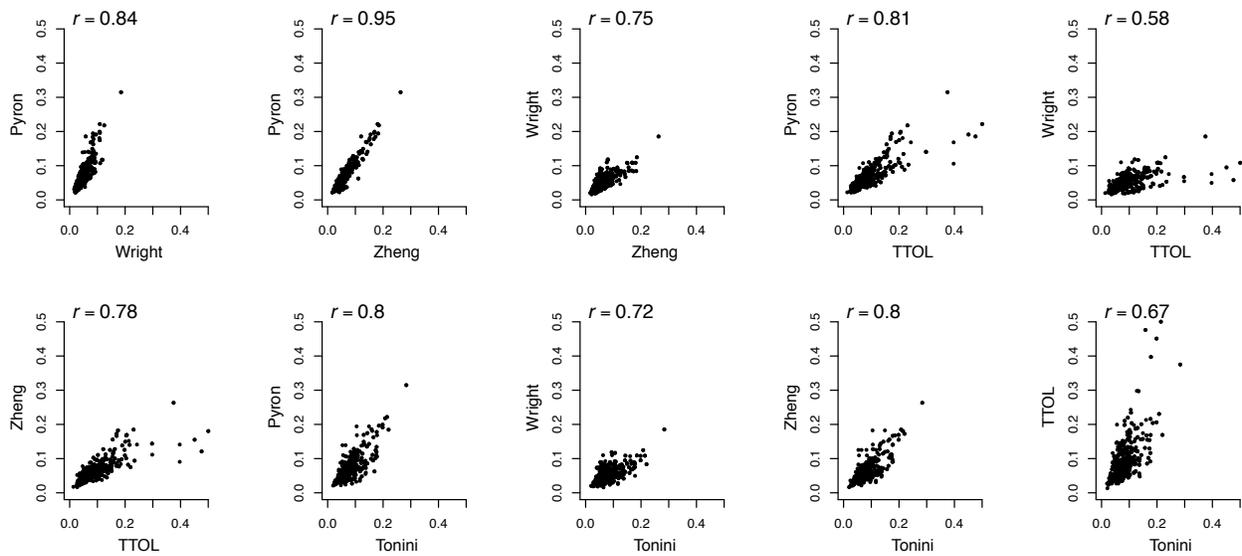
Figure S2.20: Pairwise comparisons of the per-species DR statistic for all Australian taxa. Correlation coefficients are listed in the top left corner.
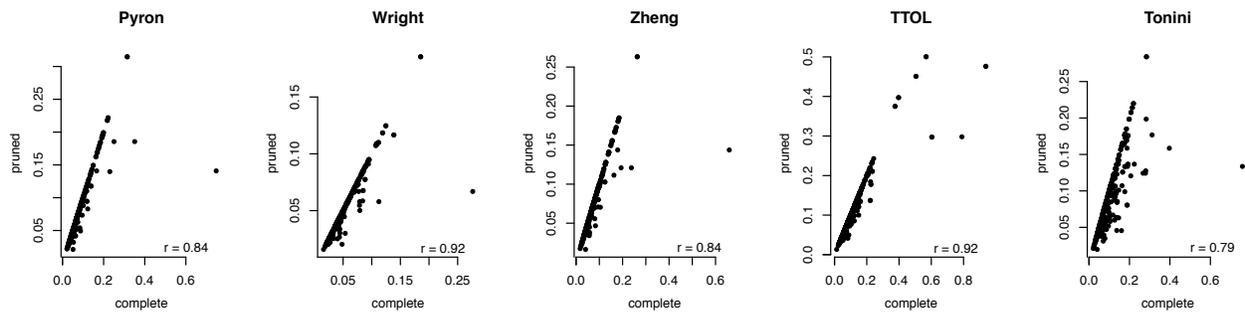
Figure S2.21: Comparison of per-species DR statistic for Australian taxa from complete and pruned trees. The pruned phylogenies were redued to a common set of 3487 taxa. Correlation coefficients are listed in the bottom right corner of each plot.
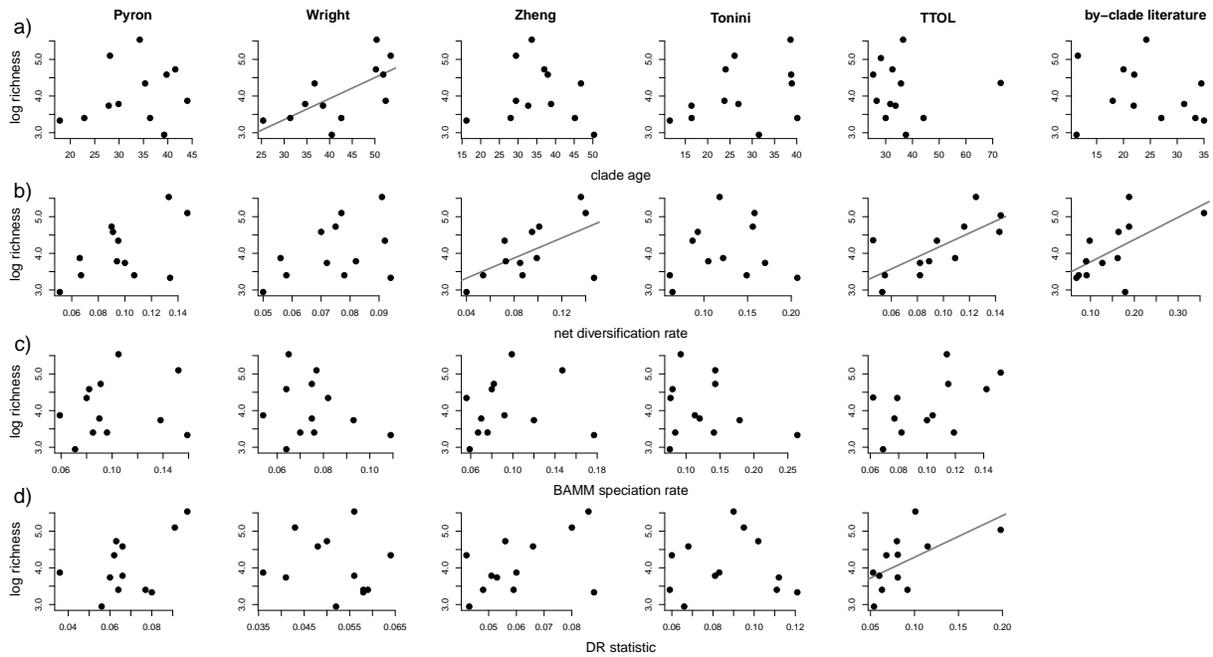
Figure S2.22: Examination of the influence of dataset selection on the relationship between species richness and various predictors: a) the relationship between clade age, (b) net diversification rate, (c) BAMM speciation rates and (d) the DR net diversification rate statistic on log species richness. Each column represents the same phylogenetic dataset. BAMM speciation rates and the DR statistic are not available for the by-clade literature as separate, well-sampled clade phylogenies were not available. The line of best fit is only plotted for those relationships that are statistically significant, according to linear regression.
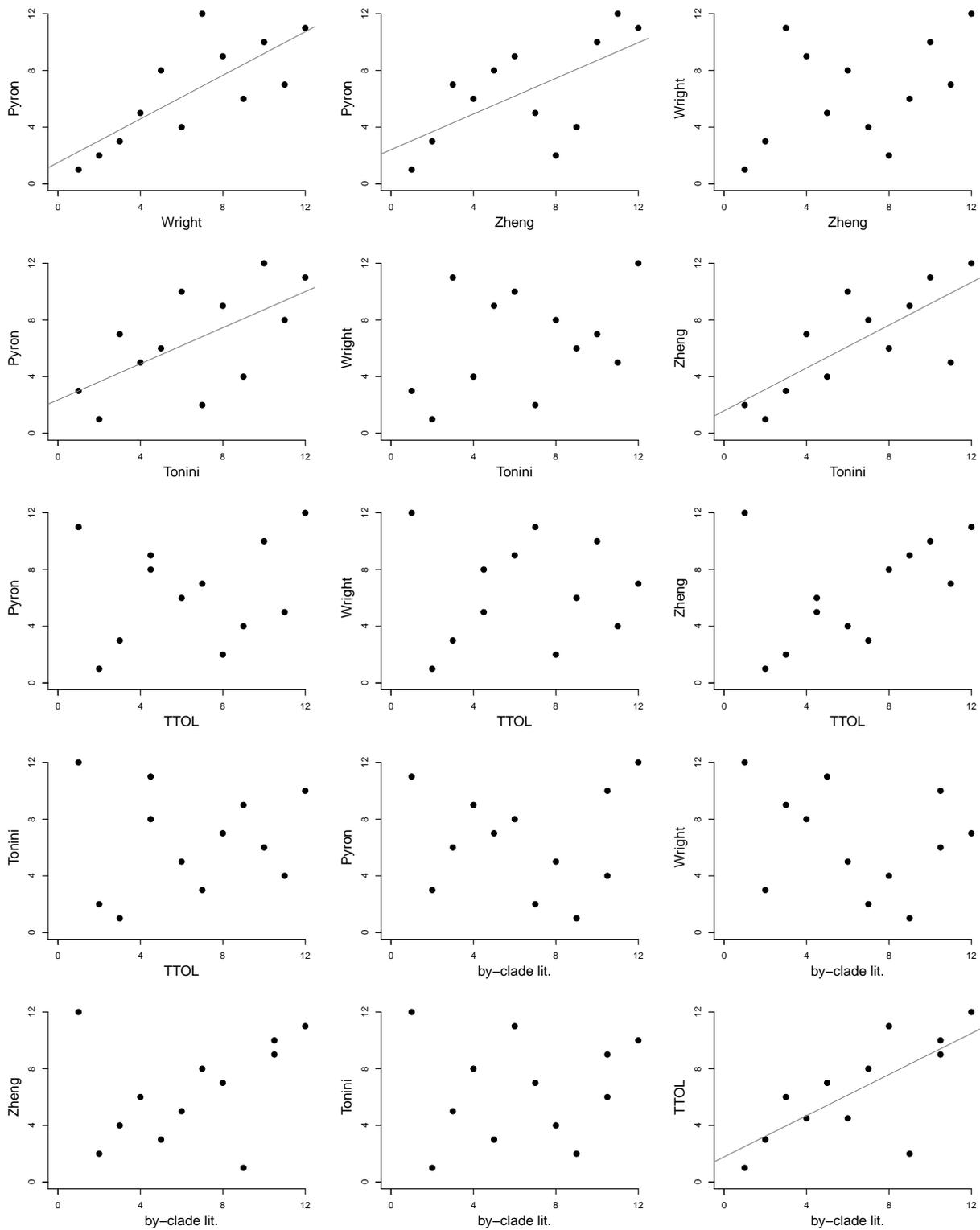
Figure S2.23: Pairwise comparisons of the rank order positions of different clades in terms of net diversification rate. Best-fit lines have been plotted when the Spearman's rank correlation test was statistically significant ($p < 0.05$).

clade	Pyron sp1	Pyron sp2	Wright sp1	Wright sp2	Zheng sp1	Zheng sp2	Tonini sp1	Tonini sp2	TTOL sp1	TTOL sp2
Agamidae	*Moloch horridus*	*Diporiphora arnhemica*	*Moloch horridus*	*Diporiphora arnhemica*	*Moloch horridus*	*Diporiphora arnhemica*	*Moloch horridus*	*Diporiphora arnhemica*	*Moloch horridus*	*Physignathus lesueurii*
Eugongylus	*Morethia butleri*	*Carlia vivax*	*Bassiana duperreyi*	*Carlia vivax*	*Bassiana duperreyi*	*Carlia vivax*	*Morethia butleri*	*Carlia vivax*	*Morethia butleri*	*Carlia vivax*
Egernia	*Bellatorias frerei*	*Egernia montana*	*Bellatorias frerei*	*Egernia montana*	*Bellatorias frerei*	*Liopholis montana*	*Bellatorias frerei*	*Liopholis montana*	*Bellatorias frerei*	*Egernia montana*
Sphenomorphinae	*Lerista varia*	*Calyptotis ruficauda*	*Lerista varia*	*Calyptotis ruficauda*	*Lerista varia*	*Calyptotis ruficauda*	*Notoscincus ornatus*	*Lerista varia*	*Lerista varia*	*Calyptotis ruficauda*
Gehyra	*Gehyra dubia*	*Gehyra minuta*	*Gehyra dubia*	*Gehyra minuta*	*Gehyra dubia*	*Gehyra minuta*	*Gehyra dubia*	*Gehyra minuta*	*Gehyra dubia*	*Gehyra minuta*
Diplodactylidae	*Oedura robusta*	*Diplodactylus pulcher*	*Oedura robusta*	*Diplodactylus pulcher*	*Nebulifera robusta*	*Diplodactylus pulcher*	*Nebulifera robusta*	*Diplodactylus pulcher*	*Oedura robusta*	*Diplodactylus pulcher*
Pygopodidae	*Delma labialis*	*Aprasia smithi*	*Delma labialis*	*Aprasia smithi*	*Delma labialis*	*Aprasia smithi*	*Delma labialis*	*Aprasia smithi*	*Delma labialis*	*Aprasia smithi*
Carphodactylidae	*Phyllurus kabikabi*	*Nephrurus deleani*	*Phyllurus kabikabi*	*Nephrurus deleani*	*Phyllurus kabikabi*	*Nephrurus deleani*	*Phyllurus kabikabi*	*Nephrurus deleani*	*Phyllurus kabikabi*	*Orraya occultus*
Pythonidae	*Aspidites ramsayi*	*Antaresia childreni*	*Aspidites ramsayi*	*Antaresia childreni*	*Aspidites ramsayi*	*Antaresia childreni*	*Aspidites ramsayi*	*Antaresia childreni*	*Aspidites ramsayi*	*Antaresia childreni*
Typhlopidae	*Austrotyphlops waitii*	*Austrotyphlops howi*	*Austrotyphlops waitii*	*Austrotyphlops howi*	*Anilios waitii*	*Anilios howi*	*Anilios waitii*	*Anilios silvia*	*Ramphotyphlops waitii*	*Austrotyphlops howi*
Elapidae	*Demansia vestigiata*	*Hydrophis ornatus*	*Demansia vestigiata*	*Hydrophis ornatus*	*Demansia vestigiata*	*Hydrophis ornatus*	*Demansia vestigiata*	*Micropechis ikaheka*	*Demansia vestigiata*	*Hydrophis ornatus*
Varanidae	*Varanus mertensi*	*Varanus storri*	*Varanus mertensi*	*Varanus storri*	*Varanus mertensi*	*Varanus storri*	*Varanus mertensi*	*Varanus storri*	*Varanus mertensi*	*Varanus storri*

Table S2.1: Taxon pairs used to identify the nodes defining each Australian squamate clade, for each macrophylogeny.

clade	dataset	cladeAge	richness	net div. rate	BAMM λ	DR
Agamidae	Pyron	39.74	98	0.09	0.082	0.066
	Wright	51.84		0.07	0.064	0.048
	Zheng	37.94		0.10	0.08	0.066
	Tonini	38.74		0.09	0.08	0.068
	TTOL	25.30		0.14	0.142	0.115
	cladeLiterature	22.00		0.16		
Carphodactylidae	Pyron	36.35	30	0.07	0.085	0.064
	Wright	42.58		0.06	0.076	0.059
	Zheng	45.19		0.05	0.067	0.048
	Tonini	40.13		0.06	0.084	0.059
	TTOL	44.15		0.06	0.082	0.063
	cladeLiterature	33.40		0.07		
Diplodactylidae	Pyron	35.34	77	0.10	0.08	0.062
	Wright	36.70		0.09	0.082	0.064
	Zheng	46.81		0.07	0.056	0.042
	Tonini	38.88		0.09	0.077	0.06
	TTOL	35.70		0.10	0.079	0.068
	cladeLiterature	34.50		0.10		
Egernia	Pyron	44.01	48	0.07	0.059	0.036
	Wright	52.37		0.06	0.054	0.036
	Zheng	29.43		0.10	0.092	0.06
	Tonini	23.76		0.12	0.113	0.083
	TTOL	26.60		0.11	0.104	0.053
	cladeLiterature	18.00		0.16		
Elapidae	Pyron	28.13	164	0.15	0.152	0.091
	Wright	53.47		0.08	0.077	0.043
	Zheng	29.44		0.14	0.147	0.08
	Tonini	26.04		0.16	0.143	0.095
	TTOL	28.20		0.14	0.152	0.198
	cladeLiterature	11.50		0.36		
Eugongylus	Pyron	41.54	113	0.09	0.091	0.063
	Wright	50.21		0.07	0.075	0.05
	Zheng	37.00		0.10	0.082	0.056
	Tonini	24.03		0.16	0.143	0.102
	TTOL	32.50		0.12	0.115	0.08
	cladeLiterature	20.00		0.19		
Gehyra	Pyron	39.25	19	0.05	0.071	0.056
	Wright	40.49		0.05	0.064	0.052
	Zheng	50.31		0.04	0.059	0.043
	Tonini	31.50		0.06	0.076	0.066

clade	dataset	cladeAge	richness	net div. rate	BAMM λ	DR
	TTOL	37.56		0.05	0.069	0.054
	cladeLiterature	11.24		0.18		
Pygopodidae	Pyron	29.89	44	0.09	0.09	0.066
	Wright	34.62		0.08	0.075	0.056
	Zheng	38.81		0.07	0.07	0.051
	Tonini	26.90		0.10	0.12	0.081
	TTOL	31.69		0.09	0.077	0.06
	cladeLiterature	31.30		0.09		
Pythonidae	Pyron	17.83	28	0.13	0.159	0.08
	Wright	25.36		0.09	0.109	0.058
	Zheng	16.19		0.15	0.177	0.088
	Tonini	11.50		0.21	0.264	0.121
	TTOL	72.98		0.05	0.062	0.081
	cladeLiterature	35.00		0.07		
Sphenomorphinae	Pyron	34.24	254	0.13	0.105	0.097
	Wright	50.37		0.09	0.065	0.056
	Zheng	33.65		0.14	0.099	0.086
	Tonini	38.60		0.12	0.092	0.09
	TTOL	36.50		0.12	0.114	0.101
	cladeLiterature	24.24		0.19		
Typhlopidae	Pyron	27.87	42	0.10	0.138	0.06
	Wright	38.55		0.07	0.093	0.041
	Zheng	32.69		0.09	0.12	0.053
	Tonini	16.36		0.17	0.179	0.112
	TTOL	33.69		0.08	0.1	0.081
	cladeLiterature	21.90		0.13		
Varanidae	Pyron	22.85	30	0.11	0.096	0.077
	Wright	31.33		0.08	0.07	0.058
	Zheng	28.01		0.09	0.076	0.059
	Tonini	16.39		0.15	0.141	0.111
	TTOL	29.97		0.08	0.119	0.092
	cladeLiterature	27.04		0.09		

Table S2.2: Diversification metrics for all clades and all phylogenetic datasets. "Clade age" is in millions of years, "richness" is in number of species, "net div. rate", "BAMM λ" and "DR stat" are in number of species per million years. "Net div. rate" is net diversification rate, assuming a relative extinction rate of 0.5.

CHAPTER III

ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling[1]

3.1 Abstract

Species distribution modeling is a valuable tool with many applications across ecology and evolutionary biology. The selection of biologically meaningful environmental variables that determine relative habitat suitability is a crucial aspect of the modeling pipeline. The 19 bioclimatic variables from WorldClim are frequently employed, primarily because they are easily accessible and available globally for past, present and future climate scenarios. Yet, the availability of relatively few other comparable environmental datasets potentially limits our ability to select appropriate variables that will most successfully characterize a species' distribution. We identified a set of 16 climatic and two topographic variables in the literature, which we call the ENVIREM dataset, many of which are likely to have direct relevance to ecological or physiological processes determining species distributions. We generated this set of variables at the same resolutions as WorldClim, for the present, mid-Holocene, and Last Glacial Maximum (LGM). For 20 North American vertebrate species, we then assessed

---

[1]Title, P.O. and Bemmels, J.B. (2017). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, 41, 291–307.

whether including the ENVIREM variables led to improved species distribution models compared to models using only the existing WorldClim variables. We found that including the ENVIREM dataset in the pool of variables to select from led to substantial improvements in niche modeling performance in 13 out of 20 species. We also show that, when comparing models constructed with different environmental variables, differences in projected distributions were often greater in the LGM than in the present. These variables are worth consideration in species distribution modeling applications, especially as many of the variables have direct links to processes important for species ecology. We provide these variables for download at multiple resolutions and for several time periods at envirem.github.io. Furthermore, we have written the 'envirem' R package to facilitate the generation of these variables from other input datasets.

3.2   Introduction

The ability to model a species' geographic distribution, given occurrence records and environmental information, is based on the assumption that abiotic factors directly or indirectly control species distributions (Austin 2002). Species distribution modeling (SDM) has led to a surge in research on topics such as species' potential invasiveness (Thuiller et al. 2005), the impacts of climate change on species distributions (Thuiller 2004, Hijmans and Graham 2006, Morin and Thuiller 2009), the relative importance of various predictors in determining species range boundaries (Glor and Warren 2011), historical reconstructions of species distributions (Svenning et al. 2011), conservation applications such as the identification of suitable habitats for undiscovered populations or reintroductions (Martínez-Meyer et al. 2006), analysis of broadscale patterns of species richness (Pineda and Lobo 2009), and spatially-explicit demographic simulations (Chan and Brown 2011, He et al. 2013). The ability to conduct such analyses at increasingly broad taxonomic and spatial scales has largely been facilitated by successful efforts to digitize museum specimen records, georeference associated localities (Guralnick et al. 2006, Ellwood et al. 2015) and provide this

information in a standardized format through easily accessible data portals (Constable et al. 2010, Wieczorek et al. 2012). While progress has been made in these efforts to make high quality occurrence records widely available (e.g. Global Biodiversity Information Facility, `www.gbif.org`), additional progress is still needed in providing and exploring the utility of different environmental datasets for modeling geographic distributions. In particular, it is unknown if currently available and widely used environmental datasets are sufficient and optimal for modeling distributions of terrestrial species.

The generation and projection of species distribution models requires data layers of environmental information that provide discriminatory power regarding presence and absence of species. As we typically do not know the true distribution of a species, it can be challenging to determine when an appropriate set of environmental variables has been chosen. Ideally, these variables should have direct relevance to ecological or physiological processes determining species distributions, but for many species this information is not generally available (Alvarado-Serrano and Knowles 2014). Correlative niche modeling approaches that rely on statistical associations between species occurrences and environmental variables are frequently used (Peterson et al. 2011, Alvarado-Serrano and Knowles 2014), in which the environmental determinants of habitat suitability are not known a priori. The 19 bioclimatic variables from WorldClim (Hijmans et al. 2005) are perhaps the most broadly employed set of environmental data layers for this purpose, on account of their high resolution, global coverage, and availability for both historical and future climate scenarios. However, the biological suitability of these bioclimatic variables and other such environmental datasets for modeling the distribution of the species in question is often not thoroughly assessed.

In the absence of specific knowledge about the environmental variables most likely to determine species distributions, it may be tempting to construct models using a large number of predictor variables, but such models run the risk of poor performance. For example, models built with several highly collinear variables are at an increased risk of overfitting and overparameterization (Dormann et al. 2012, Wright et al. 2014), and may behave

unexpectedly when projected to other time periods or geographic regions where they may encounter combinations of variables that have no analog in model training (Dormann et al. 2012, Owens et al. 2013, Warren et al. 2014). Additionally, whether large sets of environmental variables or smaller subsets of environmental data are used can greatly impact model predictions (Rödder et al. 2009, Synes and Osborne 2011, Braunisch et al. 2013). Variable reduction approaches can reduce model overfitting and improve model transferability (Warren et al. 2014, Wright et al. 2014), yet the relative merits of various approaches are poorly characterized and continue to be explored (Araújo and Guisan 2006, Braunisch et al. 2013). In general, variables may be reduced either statistically, or by selecting variables from ecological theory that are likely to be important given the physiology of the organism in question (Kearney et al. 2008, Doswald et al. 2009, Rödder et al. 2009, Synes and Osborne 2011).

Given the recognized importance of variable selection in constructing ecological niche models (Synes and Osborne 2011, Braunisch et al. 2013), increasing the availability of easily accessible datasets of environmental variables that may be ecologically and physiologically important to a variety of organisms should be a priority for improving flexibility and performance of SDM. Several environmental datasets are already available with which to perform SDM (e.g. WorldClim (Hijmans et al. 2005), PRISM (`www.prism.oregonstate.edu`; Daly et al. 2002), ClimateNA (Wang et al. 2012, Hamann et al. 2013, Wang et al. 2016)), but not all of these datasets are transferable among time periods or geographic regions or easily integrated with other variables. Additional environmental data layers that conceptually complement and are formatted for easy use alongside the 19 bioclimatic variables from WorldClim (Hijmans et al. 2005) – one of the most widely used environmental datasets for SDM – would broaden the options available for selection of environmental variables (whether based on ecological theory or through statistical variable reduction) and may lead to improved model performance for some species. Despite the description in the literature of formulae for many such variables that could be computed for particular regions or time

periods (see Synes and Osborne 2011 as an example), the use of such variables is limited to those researchers with the GIS skills necessary to generate these datasets and the desire to assemble them from several disparate sources.

To help satisfy this need, we introduce the ENVIREM dataset (ENVIronmental Rasters for Ecological Modeling): specifically, we provide a set of biologically relevant climatic and topographic variables (all of which have previously been described in the literature) at multiple resolutions and time periods. The variables we include were selected in particular because we hypothesize they are likely to have direct relevance to ecological or physiological processes determining distributions of many species. They should therefore facilitate ecologically-informed variable selection, and may also result in improved model performance using statistical variable-thinning approaches. As these variables are intended to complement the existing WorldClim dataset (Hijmans et al. 2005), we provide the ENVIREM dataset at the same extents and resolutions as WorldClim, for the present, mid-Holocene, and Last Glacial Maximum (LGM). We also provide an R package (R Core Team) that will enable users to generate these variables from primary sources for any resolution, geographic area, or time period, including for future time periods of interest (for which we have not provided static rasters due to the large number of climate change models in existence that are continually updated as climate-change projections improve). Finally, through several case studies, we show that the ENVIREM variables can improve model performance and be valuable additions to the set of variables that are currently widely used in species distribution modeling.

3.3   Methods

We compiled a list of biologically relevant climatic variables (Table1) that could be derived from monthly temperature and precipitation data (WorldClim ver. 1.4, Hijmans et al. 2005) and monthly extraterrestrial solar radiation (available from `www.cgiar-csi.org`). These variables are described by Thornthwaite (1948), Daget (1977), Hargreaves and Hargreaves

69

(1985), Willmott and Feddema (1992), Vörösmarty et al. (2005), Zomer et al. (2006, 2008), Rivas-Martínez and Rivas-Sáenz (2009), Sayre et al. (2009) and Metzger et al. (2013). We additionally produced two elevation-derived topographic variables, terrain roughness index (Wilson et al. 2007) and topographic wetness index (Boehner et al. 2002, Conrad et al. 2015), generated from a global 30 arc-second elevation and bathymetry digital elevation model (Becker et al. 2009). All variables were produced at the same resolutions as the bioclimatic variables that are currently available through WorldClim: 30 arc-seconds, and 2.5, 5 and 10 arc-minutes. Topographic variables were produced at a 30 arc-second resolution, and subsequently coarsened to match the lower resolutions, rather than constructed directly from lower-resolution elevation data. As such, the topographic variables of large grid cells at coarser scales represent the average fine-scale (i.e. 30 arc-second) values within each grid cell. Calculating the topographic variables in this manner was particularly important to avoid loss of information regarding terrain roughness index when scaling up to coarser resolutions. For the two climate variables related to growing degree-days (GDD), we note that GDD are accumulated on a daily basis, whereas our estimates are approximations based on mean monthly temperature (Table 1).

We generated rasters for all variables at multiple spatial resolutions for current climatic conditions, the mid-Holocene (approximately 6000 yr ago) and the Last Glacial Maximum (LGM, approximately 22000 yr ago). For the paleoclimate datasets, we generated variables from three global general circulation models (GCMs): the Community Climate System Model ver. 4 (CCSM4, Collins et al. 2006), the Model for Interdisciplinary Research On Climate (MIROC-ESM, Hasumi and Emori 2004), and the model of the Max Planck Inst. for Meteorology (MPI-ESM-P, Stevens et al. 2013). Fine-scale monthly rasters for these paleoclimate scenarios were generated from coarse-resolution GCM output using the delta downscaling method (Ramirez-Villegas and Jarvis 2010, and `www.worldclim.org/downscaling`) and are available with the WorldClim dataset. As the formulae for some ENVIREM variables require mean monthly temperature, which is available from the WorldClim dataset in the present

but not for other time periods, we calculated mean monthly temperature in all time periods as the mean of the maximum and minimum temperatures. In the present, this calculation is highly correlated with the available mean monthly temperatures (Pearson correlation coefficient > 0.99). All raster manipulation and variable creation was carried out in R with the raster package 2.5-2 (Hijmans et al. 2016).

Additional variables derived from and complementing the 19 bioclimatic variables from WorldClim (Hijmans et al. 2005) will only be of value in SDM applications if they represent information not currently contained in the 19 bioclimatic variables. To assess the degree of novelty of these new variables, we calculated the Pearson correlation coefficient between each of the ENVIREM variables and the 19 bioclimatic variables from WorldClim, at a global scale (10 arc-minute resolution), and also by biogeographic realm (Olson et al. 2001, Table 2, Table S2), for both the present and the past (CCSM4 global circulation model). Similarly, we also calculated correlation coefficients between terrain roughness index and topographic wetness index with elevation (Table 3) to explore whether these variables contain topographic information not captured by elevation alone.

3.3.1 Case studies

To investigate how the inclusion of the ENVIREM variables could affect the performance and predictions of species distribution models, we generated species distribution models with Maxent ver. 3.3.3k (Phillips et al. 2006) for 20 North American terrestrial vertebrate species, using the curated occurrence dataset from Waltari et al. (2007). Specifically, we generated niche models using three different sets of initial environmental predictor variables. Firstly, we generated models using only the 19 bioclimatic variables from WorldClim (referred to hereafter as the *bioclim model*). Secondly, we built models using the 19 bioclimatic variables plus 14 of the climatic ENVIREM variables (hereafter referred to as the *bioclim + envirem-clim* model). Finally, we generated niche models with the 19 bioclimatic variables and 16 ENVIREM variables, including 14 climatic variables and the two topographic vari-

ables (the *bioclim + envirem-all* model). Note that none of the models, including *bioclim + envirem-all*, included elevation as a predictor variable. We chose not to include two variables, aridityIndexThornthwaite as it was conceptually redundant with the climaticMoistureIndex, and monthCountByTemp10 because it is a categorical variable that would not have been amenable to the variable selection procedure that we applied. Finally, we did not generate any models using only the ENVIREM variables without the 19 bioclimatic WorldClim variables, as the ENVIREM variables are intended to supplement, not replace, the bioclimatic variables. All distribution modeling was performed in the dismo package ver. 1.0-15 in R (Hijmans 2016) from rasters at a 2.5 arc-minute resolution. This resolution is likely a reasonable match to the locational accuracy of the species occurrences, as these come primarily from museum collections, and is the resolution used for SDM in the original study (Waltari et al. 2007).

To construct each model, we first spatially thinned the occurrence records, retaining only occurrences that were greater than ten kilometers in proximity to one another, using the spThin package in R (Aiello-Lammens et al. 2015). For each species individually, we defined the model-training region by adding a 1000 km buffer around all occurrence records (Figure S1). All occurrence data and rasters were transformed and projected to the North America Albers Equal Area Conic projection, as it has been shown that a failure to account for changing grid-cell area across latitudes can negatively impact SDM results (Budic et al. 2015). We statistically thinned variables to include in each model for each species using the 'corSelect' function in the fuzzySim package ver. 1.6.3 in R (Barbosa 2015) where each pair of variables that is correlated above a set threshold is tested against the response variable (species presence and absence) with a bivariate model. The variable with a better fit as measured with AIC is selected while the other is dropped, and the procedure is repeated until all pairwise correlations are below the threshold. We applied a correlation threshold of 0.75, and generated pseudo-absences from 10000 randomly sampled points throughout the training region (excluding grid cells with known occurrence records) because there were no

true absence records in our data.

For each species, we measured SDM performance for the bioclim, the *bioclim + envirem-clim* and the *bioclim + envirem-all* models (with reduced sets of variables via statistical thinning as described above, Table 4) using three threshold-independent evaluation metrics: AUC_{TEST}, AUC_{DIFF}, and the size-corrected Akaike information criterion (AIC_c). AUC_{TEST} is a metric that measures the discriminatory ability of the species distribution model at test localities withheld during model construction, and thus represents the ability of the model to predict species presence (Peterson et al. 2011). AUC_{DIFF} is the difference between the AUC calculated from training localities and AUC_{TEST}, and is a measure of model overfitting, with higher values of AUC_{DIFF} representing more overfit models (Warren and Seifert 2011). AIC_c is an information theoretic metric that balances model fit against degrees of freedom from parameterization (i.e. model complexity), such that lower values of AIC_c correspond to models with better goodness-of-fit accounting for model complexity (Burnham and Anderson 2004, Warren and Seifert 2011). For AUC metrics, we partitioned calibration and evaluation data via the masked geographically-structured partitioning scheme described by Radosavlje-vic and Anderson (2014), implemented in the R package ENMeval ver. 0.2.1 (Muscarella et al. 2014), which leads to more realistic and less biased estimates of SDM performance than the more traditionally used random k-fold partitioning scheme. This partitioning scheme divides occurrence records into four geographic regions with an equal number of occurrence records, and calculates AUC metrics as the average of those metrics calculated individually using each of the four possible partitions of geographic regions into one region of evaluation data and three regions of calibration data. AIC_c was calculated from the full, non-partitioned models.

The complexity of SDMs built with Maxent can be adjusted with the regularization multiplier, increased values of which lead to less parameterized models, as well as with the inclusion of additional feature classes (i.e. transformations of the original predictor variables) that allow for increasingly complex models. We evaluated distribution models across different

sets of permissible feature classes, and for each of these, across a range of regularization multiplier values. The evaluation metrics described above were used to determine optimal feature class and model complexity for each model individually (Muscarella et al. 2014).

After selecting optimal feature class and model complexity for each model, we also compared performance of the optimal models across each of the three variable sets (i.e. *bioclim*, *bioclim + envirem-clim*, and *bioclim + envirem-all*) using the same evaluation metrics. The AUC metrics describe absolute performance of the models (ranging from 0 to 1). AIC_c, however, describes relative performance of candidate models. For this metric, we define a model as having substantial support over another if it has a difference in AIC_c greater than or equal to four, as models with AIC_c values more similar than this are generally considered to have equivalent support (Burnham and Anderson 2004). Although we present results for all evaluation metrics, we ultimately favor AIC_c for selecting the optimal model and variable set for each species, as the focus of our case studies is on model comparison, and AIC_c has been shown to perform better than AUC metrics according to a range of criteria, including the selection of optimal levels of model complexity, model transferability in space and time, and the relative ranking of variable importance (Warren and Seifert 2011, Warren et al. 2014, Moreno-Amat et al. 2015).

Theimpact of using different environmental variables in niche modeling may not be apparent if two sets of variables lead to similar projected distributions in the present. However, if the degree of correlation between two different sets of variables differs in the past compared to in the present, then variable choice might have a greater effect on SDM projections to other time periods. To explore this possibility, we calculated niche similarity in the present and in the LGM using Schoener's D (Schoener 1968, Warren et al. 2008), a metric that quantifies the degree of niche overlap in geographic space. Values of D range from 0 (completely different niches across geographic space) to 1 (identical niches over geographic space). Overlap was quantified with the fuzzySim package in R (Barbosa 2015). For each case-study species we focused the niche overlap calculation on the geographic regions of the model pro-

jections where comparisons among models are most meaningful, rather than across broad regions of the continent where all models predict low habitat suitability and are thus very similar. In particular, we calculated niche overlap statistics only over the geographic region predicted to contain suitable habitat in at least one of the models. To define this region, we first reduced the geographic extents of interest for both the projected *bioclim* and *bioclim + envirem-clim* models individually using a habitat suitability threshold that preserved 95% of the training presences. We further excluded areas outside the model training region, except for a few species where the majority of the predicted LGM distribution lay outside the training region. Finally, we combined these regions for both the *bioclim* and *bioclim + envirem-clim* models and calculated niche overlap from (non-thresholded) model projections within this combined region. We did not project the *bioclim + envirem-all* model to the LGM, because topographic variables are difficult to interpret for the LGM in glaciated regions of North America. These regions have experienced substantial changes in topography since the LGM due to glacial erosion (Bell and Laine 1985). However, we note that models using topographic variables could be projected to the LGM in particular regions of interest where topographic variables can be assumed to have remained static since the LGM (e.g. unglaciated regions of California, Bemmels et al. 2016).

3.3.2 Data deposition

The ENVIREM dataset has been deposited through the Univ. of Michigan Deep Blue Data repository `http://dx.doi.org/doi:10.7302/Z2BR8Q40` (Title and Bemmels 2017), and can be accessed through the project website at `www.envirem.github.io`. The 'envirem' R package is available on CRAN.

3.4 Results

The ENVIREM dataset comprises variables that were generated for three time periods (present, mid-Holocene and the LGM), using several different general circulation models

(CCSM4, MIROC-ESM, MPI-ESM-P) at multiple resolutions, so as to facilitate integration with rasters from WorldClim (Hijmans et al. 2005). All rasters are available for download at envirem.github.io. To enable users to generate these variables from other circulation models or time periods, we have provided all code in an R package 'envirem', available from CRAN.

At a global scale, most new climatic variables were highly correlated with at least one of the 19 bioclimatic variables from WorldClim (Table 2). The aridity-related variables (i.e. climatic moisture index and Thornthwaite's aridity index) and some of the PET-related variables were the least redundant at the global scale. However, many of the new variables were less highly correlated with the 19 bioclimatic variables within specific biogeographic realms. Oceania and the Afrotropics were the realms with the greatest number of new variables with lower maximum correlation coefficients ($\leqslant 0.85$), indicating that niche models of species from those regions may benefit most from the inclusion of these new variables. More often than not, correlations were lower during the mid-Holocene and LGM than in the present (Table S2, Table 2), which indicates that even if specific sets of variables are redundant in the present, they may not necessarily be redundant in other time periods and variable choice could have greater impacts on model projections to other time periods. All new climatic variables had a maximum correlation of $\leqslant 0.85$ in at least one biogeographic realm during at least one time period, with the exception of continentality, thermicity index, maximum temperature of the coldest month and minimum temperature of the warmest month. Some new variables were consistently most highly correlated with the same bioclimatic variable from WorldClim across regions, while other new variables were most highly correlated with different bioclimatic variables across different regions (Table S1).

In terms of topographic variables derived from elevation, terrain roughness index was not highly correlated with elevation globally or in any biogeographic region (Table 3). Topographic wetness index was also not highly correlated with elevation (Table 3), even though higher values of topographic wetness are conceptually associated with lower elevations at a local scale (i.e. within a given watershed; Boehner et al. 2002).

3.4.1 Case studies

Statistical thinning of the sets of variables prior to ecological niche modeling substantially reduced the number of variables, with three to 11 variables retained in each model (Table 4 S3, S4). For all species, at least one ENVIREM variable was retained in the *bioclim + envirem-clim* models. For the *bioclim + envirem-all* models, at least one topographic variable was retained for 19 of 20 species. For most species, one or more bioclimatic variables that were retained in the *bioclim* model were dropped from the *bioclim + envirem-clim* and *bioclim + envirem-all* models and were replaced by one or more of the ENVIREM variables, indicating that these variables are more strongly predictive of the presence and absence of the species than the dropped bioclim variables (Table S3, S4). The impact of including ENVIREM variables on model performance varied among species, but models containing ENVIREM variables performed substantially better (according to the AIC_c metric) than the *bioclim* model in 13 of 20 species.

In Figure 1, we highlight results for four species that show particularly distinct improvement with the ENVIREM variables: the spotted salamander *Ambystoma maculatum*, the blue grouse *Dendragapus obscurus*, the California gnatcatcher *Polioptila californica* and the mountain chickadee *Poecile gambeli*. In these four species, inclusion of ENVIREM variables led to improvements in all metrics of model performance, although differences in AIC_c values were more substantial than differences in AUC metrics for these species. Across the 16 other case study species (Figure S2-S5), an improvement in performance when including ENVIREM variables was found for ten species according to greater AUCTEST values (*Arborimus longicaudus*, *Chamaea fasciata*, *Desmognathus wrighti*, *Dicamptodon tenebrosus*, *Elaphe obsoleta*, *Glaucomys sabrinus*, *Glaucomys volans*, *Lampropeltis zonata*, *Martes americana* and *Myodes gapperi*). However, substantial improvements in model performance (improvement by more than four AIC_c units) were found for all but seven species according to AIC_c values, with no substantial difference for *Dicamptodon tenebrosus*, *Elaphe obsoleta* and *Lepus arcticus*, and a substantial decrease in performance for four species (*Crotalus atrox*,

Dicrostonyx groenlandicus, *Glaucomys volans* and *Myodes gapperi*). Inclusion of ENVIREM topographic variables specifically led to especially notable improvements in AIC_c scores for *Poecile gambeli* (Figure 1), *Eumeces fasciatus*, *Blarina brevicauda* and *Plethodon idahoensis* (Figure S2-S5).

The optimal Maxent parameters identified by the model evaluation metrics were typically not concordant across the *bioclim*, *bioclim + envirem-clim*, and *bioclim + envirem-all* models (Table S5). Similarly, as the different metrics evaluate the niche models using conceptually different criteria, AUC-based evaluations did not identify the same Maxent parameters as AIC_c-based evaluations (Table S5). As the focus of our case studies is on the choice of variables employed, an in-depth examination of the differences between AUC and AIC_c-based optimization of Maxent is beyond the scope of our study. We therefore focus the rest of our results and discussion on comparing predictions of models that were optimized based on AIC_c (see Methods).

Projections of the AIC_c-optimized species distribution models constructed with and without the ENVIREM variables generally did not differ greatly at continental scales for the current time period, but regional-scale differences in habitat suitability were observed. For the four case-study species showing greatest improvement in all evaluation metrics, the overall suitable ranges are very similar, though not identical, at the continental scale (Figure 2). In finer-scale maps focusing on a particular region of interest, however, there are more substantial differences in suitability across the landscape at a regional scale (Figure 2). For example, suitability of the California Central Valley for *Polioptila californica* is much higher in the *bioclim* model than in the *bioclim + envirem-clim* model. Similarly, regions of the California coast and northwestern Great Basin for *Dendragapus obscurus* are also considerably different across models, as well as large areas of the interior range of *Poecile gambeli*. Niche overlap (Schoener's D) between the two models averaged 0.9 for these four species and 0.91 across all modelled species (Figure S6, Table S6).

Differences between the predictions of the AIC_c-optimized *bioclim* and *bioclim + envirem-*

clim models become more pronounced when projected to the LGM (Figure 3, Table S5). In particular, Schoener's D niche overlap scores are much lower in the LGM (mean = 0.71, 0.71 and 0.72 for GCM CCSM4, MPI-ESM-P and MIROC-ESM, respectively) compared to the present, and for many species there are considerable differences between models in predicted distribution in the LGM (Figure 3). For *Ambystoma maculatum*, habitat suitability in the *bioclim* model was highest on exposed continental shelf off the coast of North Carolina, whereas in the *bioclim + envirem-clim* model the highest habitat suitability was in the Lower Mississippi River Valley. For *Dendragapus obscurus*, connectivity between regions was greater in the *bioclim + envirem-clim* model, and areas of high habitat suitability included the Columbia Plateau and northern Cascades. Both models for this species also showed marginally to moderately suitable habitat in western Canada and Alaska, although this may be an overprediction as at least part of this region was covered by the Cordilleran ice sheet during the LGM (Dyke et al. 2002). For *Polioptila californica*, the *bioclim* model predicted large regions of California to be suitable, including California's Central Valley, whereas in the *bioclim + envirem-clim* model, higher suitability was primarily restricted to Baja California and coastal regions of southern California. For *Poecile gambeli*, visual differences between model projections were even greater, with high habitat suitability in the Rocky Mountains in the *bioclim + envirem-clim* model only, and much higher habitat suitability throughout most of the species' range overall, and the Great Basin in particular.

3.5 Discussion

We have generated 18 climatic and topographic variables that will be valuable in a broad array of applications for species distribution modelling, and have made these variables easily available and complementary to an existing widely-used environmental dataset. Although they are largely derived from the same underlying dataset as the bioclimatic variables from WorldClim, we have demonstrated that including the ENVIREM variables in SDM can lead to notable improvements in performance and differences in projections of species distribution

models. Inclusion of these new variables led to substantial improvement in SDM performance (AIC$_c$ metric) in 13 out of 20 species, and substantially worse performance in only four species. Although inclusion of the ENVIREM variables did not always lead to significantly improved performance, the fact that they were beneficial to many species indicates that they are generally worth consideration when constructing species distribution models. The species-specific nature of our results also highlight the importance of following best practices for variable selection and parameter optimization, as we have done here. The importance of particular variables in SDM will be a function of the species under study, its distribution in geographic and climatic space, the time period and geographic region of interest, and the ultimate question being addressed. Nonetheless, the links to ecological and physiological processes represented in many of the ENVIREM variables mean that they will likely be particularly useful for a wide variety of applications.

3.5.1 Potential applications

As we have showcased here, the ENVIREM dataset will be of immediate value in SDM applications and will potentially lead to the generation of better species distribution models. If variable selection is done via statistical approaches, then inclusion of these variables will allow researchers to start with a larger pool of biologically relevant options, thereby increasing the odds that variables that are highly informative regarding the presence and absence of a species will be discovered. If the goal is to select variables a priori based on the ecology and natural history of the organism, then the ENVIREM variables will provide valuable options, as they are likely to be ecologically relevant to certain species and may have specific ties to biological processes for many species. SDM has been employed as a tool in a large variety of studies, and the inclusion of new variables has the potential to impact their conclusions. Identifying better sets of predictor variables for certain species could, among other things, potentially alter projections of species' invasiveness for particular regions (Peterson and Nakazawa 2008), alter our understanding of potentially suitable habitat

for species introductions (Martínez-Meyer et al. 2006), lead to identification of new areas of high habitat suitability for conservation interest, affect predictions of shifts in habitat suitability in response to future climate change (Thuiller 2004, Hijmans and Graham 2006, Morin and Thuiller 2009), lead to new phylogeographic hypotheses about where species may have been distributed in the past (Chan and Brown 2011, He et al. 2013, Bemmels et al. 2016), and impact our understanding of the evolution of climatic tolerances across related species (Title and Burns 2015, Kozak and Wiens 2016).

With these additional variables, ecologists and evolutionary biologists will also be able to craft more specific hypotheses that are informed by the ecology of the organisms under study. For example, in an integrative distributional, demographic and coalescent (iDDC) framework (Knowles and Alvarado-Serrano 2010, Brown and Knowles 2012, He et al. 2013), these variables will allow for the specification of competing hypotheses pertaining to the relative importance of different climatic and topographic variables in constraining the distribution of species over time (Bemmels et al. 2016), giving researchers greater flexibility than currently exists in modeling spatial and genetic patterns over time. Another example would be the inclusion of these additional variables in the spatial mapping of the velocity of climate change, which can tell us how organisms must move to track their current climatic conditions (Hamann et al. 2015). To our knowledge, this is the only existing multi-variable dataset that is truly complementary to WorldClim in its geographic breadth, application and accessibility. The Climond dataset (Kriticos et al. 2011) provides an extended suite of bioclimatic variables only at 10 and 30 arc-minutes for current and future climate scenarios, while the Ecoclimate dataset (Lima-Ribeiro et al. 2015) provides only the standard 19 bioclimatic variables for multiple past, present and future time periods at 30 arc-minutes. Other variables potentially useful for biodiversity modeling have been released, such as habitat heterogeneity (Tuanmu and Jetz 2015), global cloud cover (Wilson and Jetz 2016) and region-specific variables, such as ClimateNA (Wang et al. 2012, Hamann et al. 2013, Wang et al. 2016), but these variables are either not transferrable to other time periods, not available globally or not available at

finer spatial resolutions. In contrast, the ENVIREM dataset includes additional variables (some of which overlap with the Climond dataset) at all of the resolutions currently available from WorldClim, for past and current time periods. If researchers wish to perform SDM using occurrences that have high spatial precision in areas where region-specific datasets for all desired time periods are available, then alternatives to the ENVIREM dataset may prove most useful (e.g. ClimateNA; Wang et al. 2016). However, such a situation is likely to represent only a small minority of SDM applications, making the ENVIREM dataset more generally applicable. In addition, the envirem R package makes it possible to generate these variables for other time periods, or from alternative input datasets (for example PRISM; Daly et al. 2002), allowing users to easily customize their use of these variables.

3.5.2 Biological relevance of ENVIREM variables

Although the potential applications of these variables to SDM are vast, one unique benefit of the ENVIREM variables is their potential for improving our ability to construct niche models informed by ecological knowledge and natural history. Biologically informed niche models may be constructed for species for which the conceptual relationships between particular variables and biological processes relevant to determining a species' distribution are known a priori (Kearney et al. 2008, Doswald et al. 2009, Rödder et al. 2009, Synes and Osborne 2011), or may be constructed with the intention of exploring and testing different hypotheses about these relationships (Bemmels et al. 2016).

The potential mechanisms by which the ENVIREM variables may determine distributions are numerous and will be specific to the species of interest. In general, subsets of the ENVIREM variables may directly control species distributions, or (more commonly) may impact other processes that in turn determine distributions (Austin 2002). The particular variables included in the ENVIREM dataset were selected because of their clear conceptual links to particular ecological processes and indices. For example, growing degree-days are predictive of plant phenology and growth rate (McMaster and Wilhelm 1997), processes

which impact species range limits (Morin et al. 2007) and drive local adaptation (Howe et al. 2003). Evapotranspiration not only describes climate generally, but is also physiologically linked to plant growth potential due to its impact on gas exchange with the atmosphere and temperature regulation (Thornthwaite 1948, Katul et al. 2012). The more complex climatic indices included in the ENVIREM variables (e.g. thermicity, aridity, moisture, Emberger's pluviothermic quotient) may characterize environmental conditions that are more directly physiologically relevant to given species than simple descriptors of climate such as temperature or precipitation alone (Daget 1977). Finally, the topographic ENVIREM variables could conceivably be important predictors of habitat types associated with local- to regional-scale relief that may be key predictors of species distributions at these spatial scales (Lassueur et al. 2006, Austin and Van Niel 2011). We have provided just a few examples of potential links to biological factors that could determine species distributions, but the ecological relevance of any of the ENVIREM variables is likely to be species-specific and different species' distributions may be associated with environmental variables because of different mechanisms. Nonetheless, it is this type of conceptual relevance and these potential links to physiological and ecological processes that will make the ENVIREM variables particularly useful for many SDM applications.

3.5.3 Incorporating ENVIREM variables into SDM best practices

Ideally, the choice of variables for niche modeling should be informed by knowledge of the natural history and ecology of the organism under study, as this approach has been shown to produce more realistic niche models (Rödder et al. 2009, Saupe et al. 2012). However, it is most often the case that such information is not readily known (Alvarado-Serrano and Knowles 2014). How one should go about choosing bioclimatic variables is still an open question, the impact of which can be considerable (Peterson and Nakazawa 2008, Synes and Osborne 2011, Braunisch et al. 2013). It is generally not considered best practice to include all bioclimatic variables, as they exhibit a high degree of collinearity. This collinearity tends

to lead to overly complex, overfit models (Rodda et al. 2011). Additionally, the nature of the correlation between bioclimatic variables may differ across time periods, potentially leading to unexpected behavior in SDM projections (Rodda et al. 2011, Synes and Osborne 2011, Dormann et al. 2012, Warren et al. 2014). While we expect that many researchers will find the ENVIREM variables extremely useful for a variety of applications, we recommend that the merits of including all or some of the ENVIREM variables should be carefully considered relative to the specific application, and that variable thinning, model optimization, and other best practices in ecological niche modeling should be followed (Merow et al. 2013, Alvarado-Serrano and Knowles 2014). For example, as we do not have in-depth ecological information about the species whose ecological niches were modeled in our case studies, we employed a statistical approach to variable thinning in order to reduce the number of correlated variables, while retaining the variables with the greatest explanatory power.

An important finding of our case studies was that the difference between the *bioclim* and *bioclim + envirem-clim* models, as measured with Schoener's D, was small in the present, but greater in the LGM. Choice of predictor variables has previously been shown to have large impacts on model projections to other time periods or geographic regions (Peterson and Nakazawa 2008, Synes and Osborne 2011, Braunisch et al. 2013). The impact of variable selection points both to the utility of additional variables for developing and testing hypotheses about shifts in species distributions across different time periods and in novel spatial contexts, but also to the need for caution when making modeling decisions. Ideally, models could be evaluated in past time periods with independent fossil occurrences (Davis et al. 2014, Gavin et al. 2014, Moreno-Amat et al. 2015), but their availability will depend on the taxon under study.

In addition to the question of which environmental variables to use, a growing number of studies have demonstrated that species-specific tuning of virtually all steps in the niche modeling pipeline can lead to improved results, and that Maxent's default behavior is often not sufficient to achieve optimal performance (Anderson and Gonzalez 2011, Warren and

Seifert 2011, Merow et al. 2013, Radosavljevic and Anderson 2014, Moreno-Amat et al. 2015). Although we could have held all aspects save the predictor variables constant in the generation of niche models in order to be able to compare the results directly, generating models in this way is considered poor practice. Instead, we chose to independently generate the best possible models, given current best practices. We found that Maxent's default parameters were rarely optimal (Table S5), which echoes the findings of others that parameter tuning is an important step toward generating less overfit and more transferable species distribution models (Anderson and Gonzalez 2011, Warren and Seifert 2011, Merow et al. 2013, Radosavljevic and Anderson 2014, Moreno-Amat et al. 2015). Different evaluation metrics most often did not lead to the selection of the same optimized parameters (Table S5). This is expected, as AIC_c is intended to minimize the number of necessary parameters, while AUC metrics are not. Regardless of the environmental variables selected for SDM, the optimization of model parameters should always be considered, as model parameters can have a large impact on model performance and predictions (Figure 2, Figure S2-S5).

3.5.4 Utility of topographic variables in SDM

In addition to climatic variables, we also generated two topographic indices: topographic roughness and topographic wetness. These variables offer novel information as they are not redundant with elevation (Table 3), an environmental variable which is already broadly available for SDM. The use of elevation in SDM has been controversial (Hof et al. 2012), and may be particularly problematic when projecting to other time periods or geographic contexts where relationships between elevation and the climatic factors determining a species' niche may be different than the relationships in the context in which the model was built. However, the topographic roughness and topographic wetness indices are less likely to suffer from this complication because they are less causally linked than elevation to regional-scale climate, and they contain topographic information that may be useful for determining species distributions independent of climate. In particular, topographic roughness index may be

a reasonable surrogate for habitat heterogeneity and microsite availability that could be relevant to determining geographic distributions of some species, and topographic wetness index may help distinguish between areas that experience similar regional climate but differ markedly in microhabitat due to relative drainage position within a watershed.

However, it is important to consider whether topographic variables are available at an appropriate geographic scale for predicting species distributions. Variation in topographic features associated with microhabitats may occur at a much finer scale than that at which topographic variables are assessed, which could reduce their utility for SDM (Lassueur et al. 2006, Austin and Van Niel 2011, Pradervand et al. 2014). Since all topographic ENVIREM variables at all resolutions are ultimately averaged from values calculated from the finest-scale (30 arc-second) elevational model (see Methods), we have minimized concerns about the potential mismatch between the scale at which the indices were generated and at which topography is relevant to a species. However, it is still important to consider whether variation in topographic roughness and wetness at the 30 arc-second scale (approximately 926 m at the equator) is likely to be meaningful for the species in question for the particular geographic region of interest and intended modeling application.

Nonetheless, our case studies revealed that including topographic variables led to distinct improvement in SDM performance for several species, in some cases significantly exceeding the improvement gained by adding only the climatic ENVIREM variables (Figure 1, Figure S2-S5). These results once again emphasize the species-specific nature of the degree of utility of any new variable. Topographic variables are likely to be particularly useful for exploring competing hypotheses regarding whether local- to regional-scale factors such as microsite availability are important in determining species' distributions (Bemmels et al. 2016).

Beyond general considerations about whether or not topographic variables are important for modeling a species' distribution, care should also be taken in assessing whether or not static variables (i.e. variables that do not change over time) are appropriate to use for a given SDM application. The topographic variables we derive can be assumed to be largely static

through time (especially in unglaciated regions, with the exception of changes in coastline reflecting sea-level changes). Stanton et al. (2012) explored the inclusion of static variables in SDM and found that including such variables when projecting to future climate-change scenarios typically improved, and rarely hindered, SDM performance when the variables were known to influence species distributions. Nonetheless, we recommend particular caution when projecting to contexts where topography may have changed substantially over the timescale of interest, for example due to Pleistocene glacial erosion in North America (Bell and Laine 1985).

3.6   Conclusions

The ENVIREM variables constitute a valuable dataset for species distribution modeling for a variety of applications. Although they are complementary to and largely derived from the WorldClim database that is already widely in use, they contain novel information not captured by this database. In particular, the ENVIREM variables include conceptually novel climatic variables that may more closely reflect specific ecological and physiological processes, as well as topographic variables distinct from elevation that may represent non-climatic local- to regional-scale aspects of a species' niche. In our exploration of case studies for 20 North American vertebrate species, the impact of including the ENVIREM variables was species-specific: in 13 out of 20 cases model performance substantially improved compared to a model using only WorldClim variables, particularly when topographic ENVIREM variables were included; in seven cases model performance was not substantially different or declined. In general, models built with and without the ENVIREM variables produced habitat suitability predictions differing only modestly and at local scales in the current time period, but sometimes resulted in dramatic regional-scale differences in predicted habitat suitability when projected to a different time period. Overall, our results highlight how the ENVIREM variables often improve model performance, even when biological information about the variables that are most relevant to determining habitat suitability for a given

species is not known a priori. Furthermore, when knowledge about the determinants of species distributions is available from ecological theory, the ENVIREM variables may be particularly useful for developing and testing the predictions of species-specific hypotheses. The significant improvements in model performance we observed for many species when following best practices in species distribution modeling suggest that the ENVIREM variables are worth general consideration for SDM, as their main benefit is providing a more comprehensive set of environmental variables to choose from, whether through statistical variable thinning or variable selection informed by ecological knowledge.

3.7 Acknowledgements

3.8 Funding
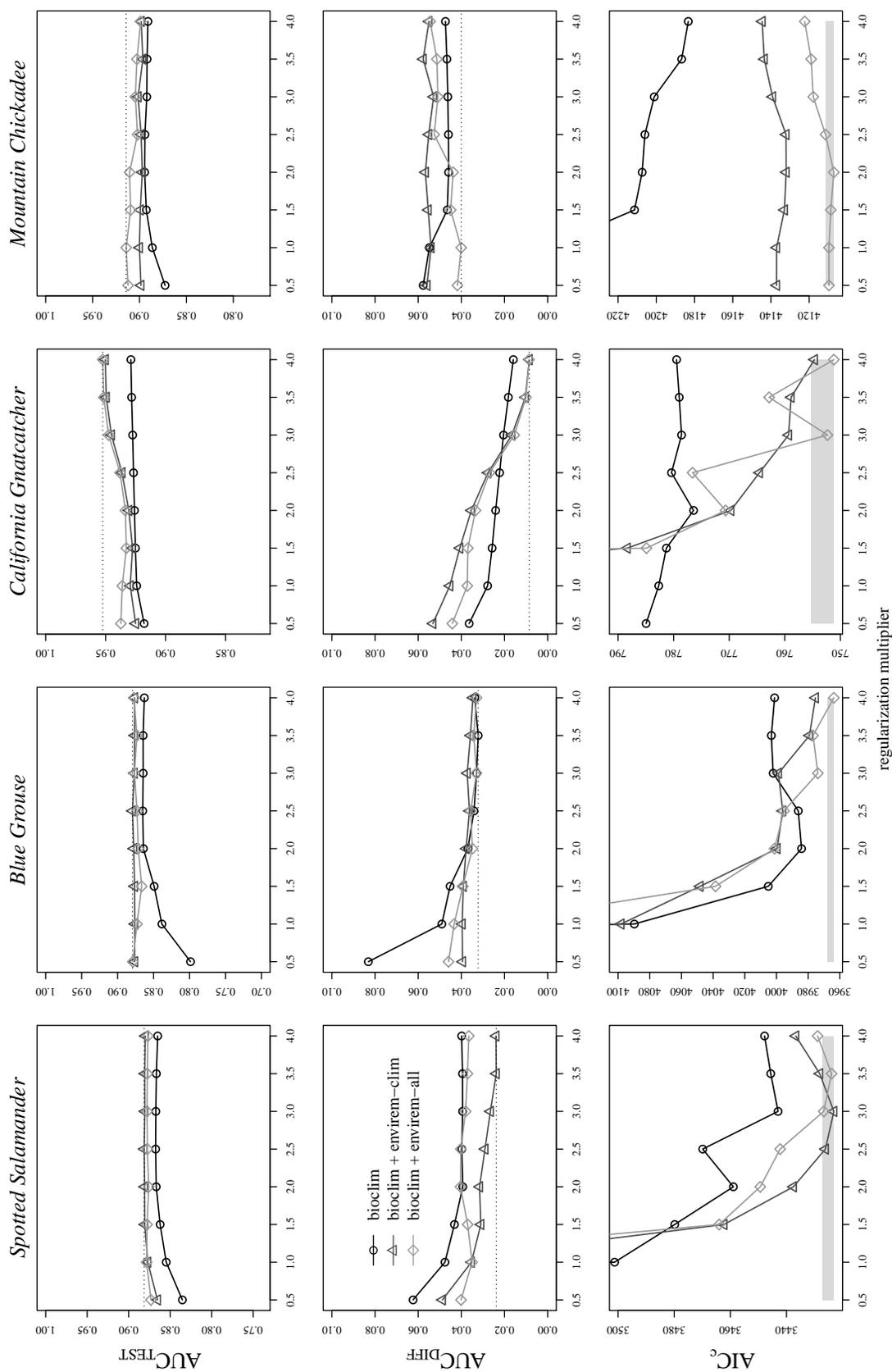
Figure 3.1: Ecological niche model performance with and without the ENVIREM variables for four selected case study species. Each line represents the set of feature classes that led to the best performance according to either AUC_{TEST} (top and middle panels) or AIC_c (bottom panel), with performance evaluated across a range of regularization multiplier values (Table S5). AUC_{DIFF} is a measure of model overfitting for the model selected by maximizing AUC_{TEST}. In the AUC plots, the dotted line represents the value for the best-performing model. In the AIC_c plots, the grey shading represents a ΔAIC_c of 4 from the best (lowest) AIC_c score. Performance of models within the grey polygon is not considered to be substantially different (Burnham and Anderson 2004).
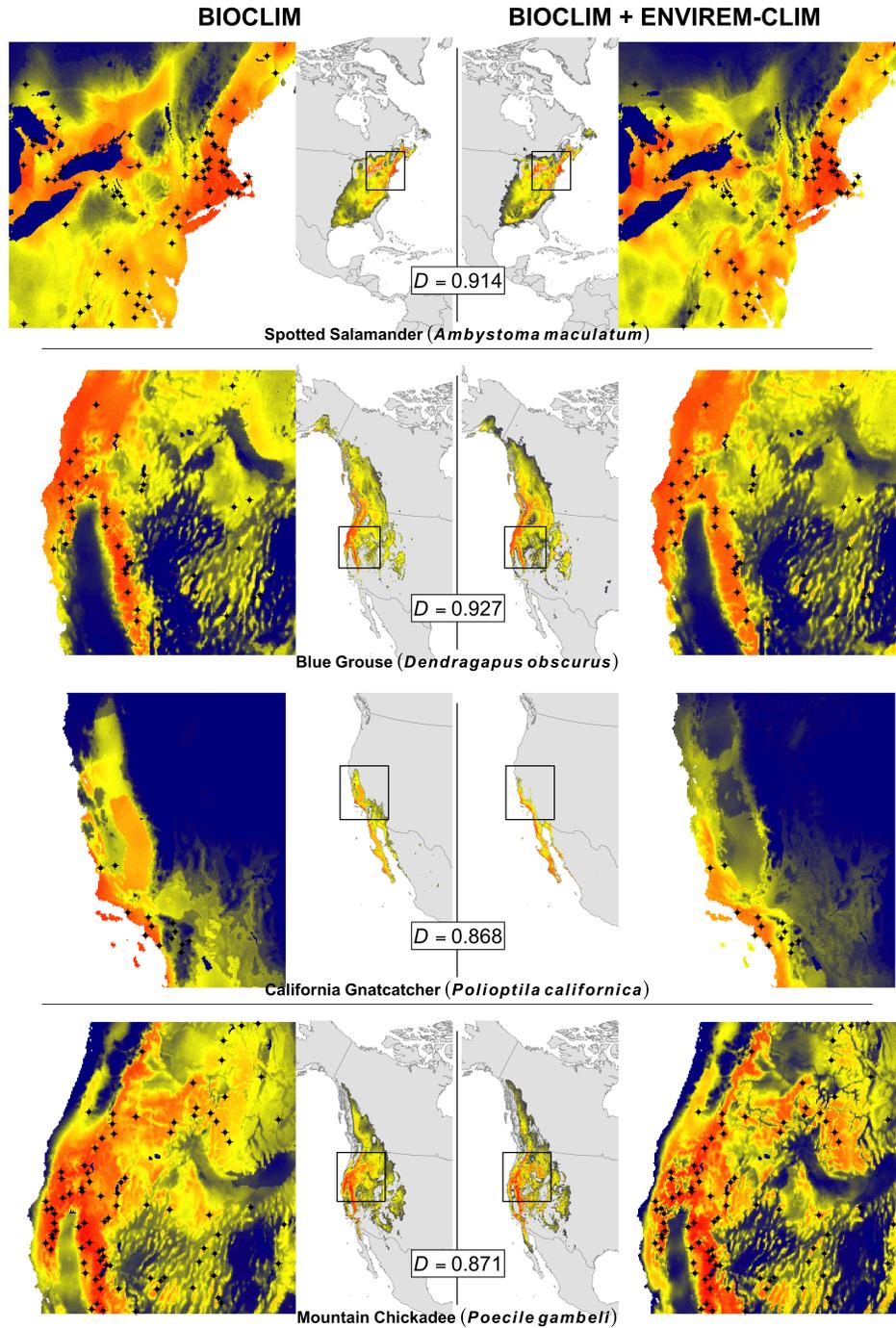
Figure 3.2: Predicted habitat suitability during the current time period for four case study species, from Maxent models optimized in terms of feature class and regularization parameter according to the AIC$_\text{c}$ metric, for models constructed with and without the ENVIREM variables. Suitability scores range from 0 (blue) to 1 (red). The central, continental-scale maps show habitat suitability within the training region only (see text for explanation), with predicted habitat suitability below a 95% training presence threshold considered to be unsuitable (grey). The outer maps show detail from the region within the box on the continental maps, selected to highlight local-scale differences between the models. Occurrence records are shown as black points. Schoener's *D* niche overlap is calculated between the *bioclim* and the *bioclim + envirem-clim* models, exclusively within the thresholded training regions (Figure S1; see the Methods section for additional details).

BIOCLIM **BIOCLIM + ENVIREM-CLIM**

$D = 0.841$

Spotted Salamander (*Ambystoma maculatum*)

$D = 0.786$

Blue Grouse (*Dendragapus obscurus*)

$D = 0.849$

California Gnatcatcher (*Polioptila californica*)

$D = 0.733$

Mountain Chickadee (*Poecile gambeli*)
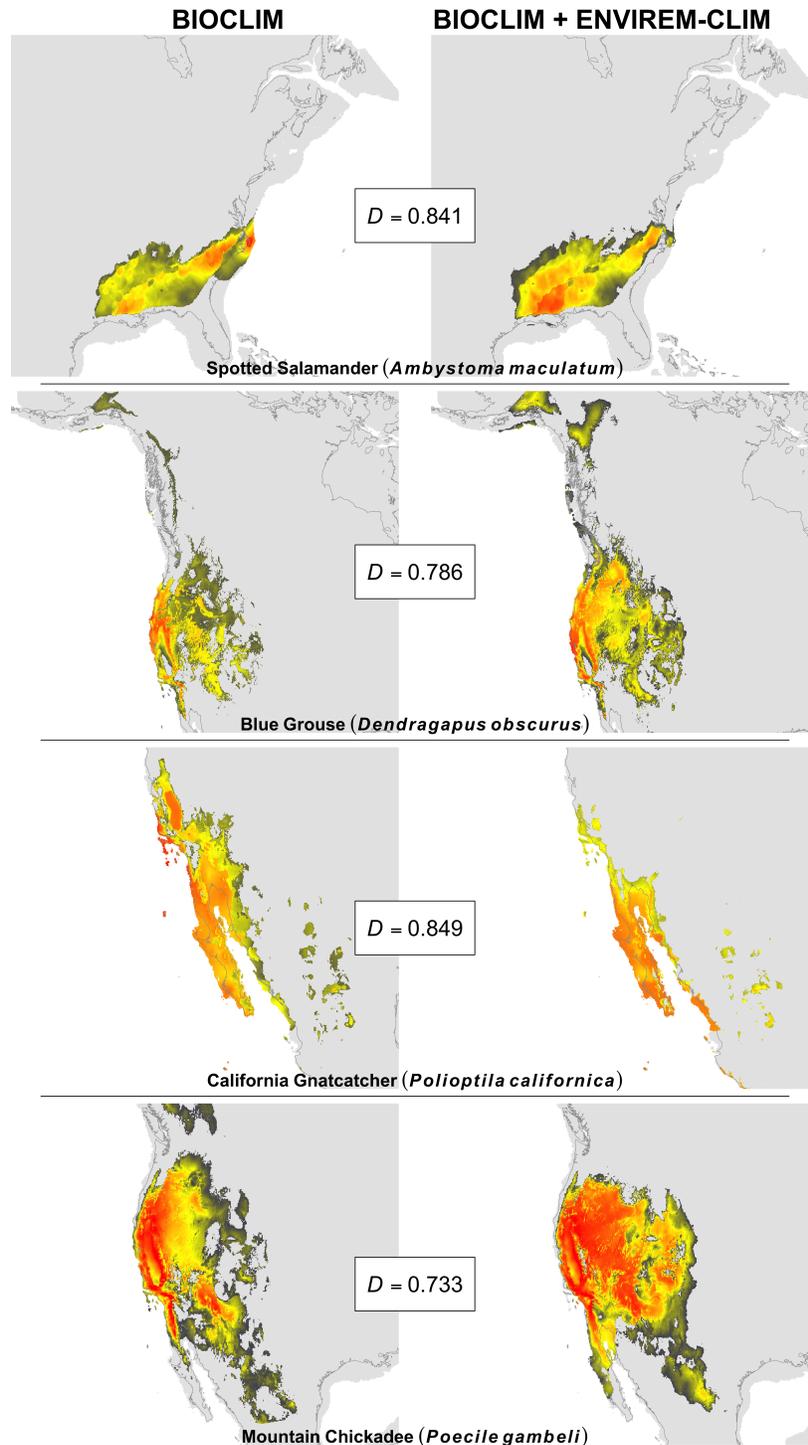
Figure 3.3: Predicted habitat suitability during the Last Glacial Maximum for four case study species, for models constructed with and without the ENVIREM variables. Suitability scores range from 0 (blue) to 1 (red). Optimization of model parameters and thresholding are as in Figure 2. Schoener's D niche overlap is calculated between the *bioclim* and the *bioclim + envirem-clim* models, exclusively within the thresholded training regions (Figure S1; see the Methods section for additional details). Habitat suitability is shown within the training region only, with predicted habitat suitability below a 95% training presence threshold considered to be unsuitable (grey).

variable abbreviation	brief description	units	source
annualPET	annual potential evapotranspiration: a measure of the ability of the atmosphere to remove water through evapotranspiration processes, given unlimited moisture	mm/year	A, B
aridityIndexThornthwaite	Thornthwaite aridity index: Index of the degree of water deficit below water need	-	C
climaticMoistureIndex	a metric of relative wetness and aridity	-	D, E
continentality	average temp. of warmest month - average temp. of coldest month	°C	F, G
embergerQ	Emberger's pluviothermic quotient: a metric that was designed to differentiate among mediterranean type climates	-	H
growingDegDays0	sum of mean monthly temperature for months with mean temperature greater than 0°C multiplied by number of days	-	I
growingDegDays5	sum of mean monthly temperature for months with mean temperature greater than 5°C multiplied by number of days	-	I
maxTempColdestMonth	max. temp. of the coldest month	°C·10	I
minTempWarmestMonth	min. temp. of the coldest month	°C·10	I
monthCountByTemp10	count of the number of months with mean temp greater than 10°C	months	I
PETColdestQuarter	mean monthly PET of coldest quarter	mm/month	I
PETDriestQuarter	mean monthly PET of driest quarter	mm/month	I
PETseasonality	monthly variability in potential evapotranspiration	mm/month	I
PETWarmestQuarter	mean monthly PET of warmest quarter	mm/month	I
PETWettestQuarter	mean monthly PET of wettest quarter	mm/month	I
thermInd	compensated thermicity index: sum of mean annual temp., min. temp. of coldest month, max. temp. of the coldest month, · 10, with compensations for better comparability across the globe	°C	F, G
tri	terrain roughness index	-	J
topoWet	SAGA-GIS topographic wetness index	-	K, L

Table 3.1: Summary of the variables in the ENVIREM dataset. Citations for variable sources are as follows: A: Zomer et al. (2006, 2008); B: Hargreaves and Hargreaves (1985); C: Thornthwaite (1948); D: Willmott and Feddema (1992); E: Vörösmarty et al. (2005); F: Sayre et al. (2009); G: Rivas-Martínez and Rivas-Sáenz (2009); H: Daget (1977); I: Metzger et al. (2013); J: Wilson et al. (2007); K: Boehner et al. (2002); L: Conrad et al. (2015).

Table 3.2: Pearson correlations between ENVIREM and WorldClim variables. Each realm cell shows the present correlation (bottom-left) and the LGM correlation (top-right).

variable	neotropical (present)	neotropical (LGM)	palearctic (present)	palearctic (LGM)	nearctic (present)	nearctic (LGM)	indo–malay (present)	indo–malay (LGM)	afrotropic (present)	afrotropic (LGM)	oceania (present)	oceania (LGM)	australasia (present)	australasia (LGM)	global (present)	global (LGM)
annualPET	0.88	0.93	0.94	0.93	0.96	0.94	0.8	0.85	0.9	0.93	0.83	0.55	0.94	0.93	0.91	0.94
aridityIndexThornthwaite	−0.84	−0.81	−0.78	−0.62	−0.73	−0.25	0.89	0.85	−0.83	−0.79	−0.8	−0.71	−0.91	−0.86	−0.81	−0.67
climaticMoistureIndex	0.93	0.91	−0.82	−0.83	−0.59	−0.63	0.91	0.91	0.98	0.98	0.89	0.88	0.95	0.96	0.81	0.6
continentality	1	1	1	1	0.99	0.99	0.99	1	0.99	0.99	1	1	1	1	1	1
embergerQ	0.95	0.95	0.91	0.9	0.94	0.87	0.88	0.85	0.94	0.95	0.92	0.85	0.97	0.94	0.93	0.91
growingDegDays0	1	1	0.93	0.87	0.89	0.77	1	1	1	1	1	1	1	1	0.97	0.94
growingDegDays5	1	1	0.91	0.84	0.85	0.74	1	1	1	1	1	1	1	0.99	0.96	0.92
maxTempColdest	0.98	0.98	1	1	1	0.99	0.97	0.97	0.92	0.93	0.98	0.97	0.97	0.97	1	1
minTempWarmest	0.98	0.98	0.99	0.99	0.98	0.99	0.96	0.93	0.96	0.95	1	0.98	0.96	0.97	0.98	0.99
monthCountByTemp10	0.87	0.88	0.95	0.89	0.93	0.81	0.74	0.86	0.49	0.73	0.62	0.77	0.7	0.92	0.95	0.94
PETColdestQuarter	0.93	0.95	0.87	0.87	0.87	0.77	0.91	0.94	0.82	0.85	0.58	0.52	0.93	0.94	0.9	0.89
PETDriestQuarter	0.83	0.84	0.92	0.91	0.87	0.87	0.89	0.9	0.84	0.87	−0.74	−0.57	0.74	0.73	0.87	0.87
PETseasonality	0.98	0.97	0.73	0.85	0.91	0.97	0.97	0.94	0.93	0.94	−0.76	−0.58	0.96	0.95	0.7	0.48
PETWarmestQuarter	0.74	0.8	0.98	0.98	0.99	0.98	0.94	0.94	0.84	0.88	0.84	0.7	0.94	0.9	0.91	0.95
PETWettestQuarter	0.79	0.85	0.89	0.92	0.95	0.94	0.68	0.76	0.62	0.76	−0.82	−0.1	0.91	0.93	0.83	0.91
thermicityIndex	1	0.99	0.98	0.98	0.98	0.99	0.96	0.97	0.99	0.99	1	1	0.97	0.98	0.98	0.99

Table 3.2: Pearson correlations between ENVIREM and WorldClim variables. The correlation is shown between each of the climatic ENVIREM variables and the WorldClim bioclimatic variable with which the ENVIREM variable is most strongly correlated (Table S1), globally and in separate biogeographic realms. For each variable and realm, the bottom-left triangle contains the correlation coefficient in the present, and the top-right triangle contains the correlation coefficient in the LGM (CCSM4) for the same bioclimatic variable. Grey shading indicates that the absolute value of the correlation is $\leqslant 0.85$.

	neotropical	palearctic	nearctic	indo-malay	afrotropic	oceania	australasia	global
terrain roughness	0.65	0.58	0.48	0.83	0.41	0.19	0.65	0.46
topographic wetness	-0.59	-0.45	-0.42	-0.67	-0.37	-0.49	-0.53	-0.39

Table 3.3: Pearson correlation coefficients between ENVIREM topographic variables and elevation, at a global scale as well as in different biogeographic realms.

	Spotted Salamander | bioclim	Spotted Salamander | all-vars	Spotted Salamander | all-vars+topo	Blue Grouse | bioclim	Blue Grouse | all-vars	Blue Grouse | all-vars+topo	California Gnatcatcher | bioclim	California Gnatcatcher | all-vars	California Gnatcatcher | all-vars+tc	Mountain Chickadee | bioclim	Mountain Chickadee | all-vars	Mountain Chickadee | all-vars+topo
annual mean temp [bio1]												
mean diurnal temp range [bio2]	+			+			+	+	+	+	+	+
isothermality [bio3]		+										
temp seasonality [bio4]										+		
max temp warmest month [bio5]												
min temp coldest month [bio6]												
temp annual range [bio7]	+	+	+	+	+	+						
mean temp of wettest quarter [bio8]	+			+			+			+	+	+
mean temp of driest quarter [bio9]							+	+	+			
mean temp of warmest quarter [bio10]	+			+			+			+		
mean temp of coldest quarter [bio11]												
annual precip [bio12]							+					
precip of wettest month [bio13]												
precip of driest month [bio14]							+	+	+	+	+	+
precip seasonality [bio15]	+	+	+	+	+	+				+	+	+
precip of wettest quarter [bio16]	+	+	+									
precip of driest quarter [bio17]				+								
precip of warmest quarter [bio18]				+	+	+	+	+	+	+	+	+
precip of coldest quarter [bio19]				+	+	+	+			+	+	+
annualPET												
climaticMoistureIndex		+	+		+	+		+	+			
continentality											+	+
embergerQ												
growingDegDays0												
growingDegDays5												
maxTempColdest												
minTempWarmest					+	+		+	+			
PETColdestQuarter												
PETDriestQuarter											+	+
PETseasonality		+	+		+	+					+	+
PETWarmestQuarter								+	+			
PETWettestQuarter		+	+		+	+		+	+			
thermicityIndex												
topoRoughness												
topoWetness			+			+			+			+

Table 3.4: Variables included in final models for four case study species. Variables included in each model were selected using a statistical variable selection approach (see Methods section for additional details).

Ambystoma maculatum   *Arborimus longicaudus*   *Blarina brevicauda*   *Chamaea fasciata*

Crotalus atrox   *Dendragapus obscurus*   *Desmognathus wrighti*   *Dicamptodon tenebrosus*

Dicrostonyx groenlandicus   *Elaphe obsoleta*   *Eumeces fasciatus*   *Glaucomys sabrinus*

Glaucomys volans   *Lampropeltis zonata*   *Lepus arcticus*   *Martes americana*

Myodes gapperi   *Plethodon idahoensis*   *Poecile gambeli*   *Polioptila californica*
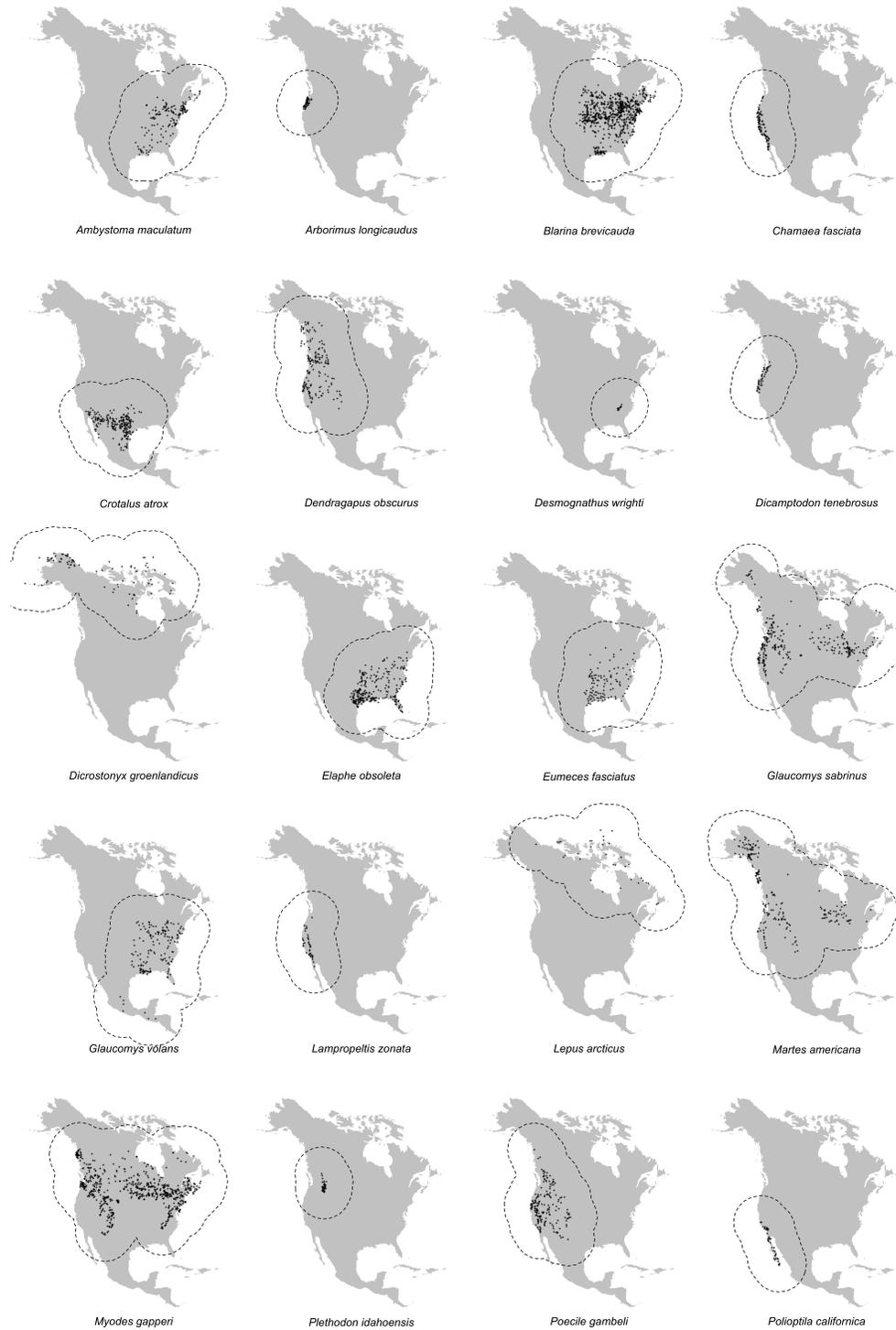
Figure S3.1: Occurrence records and training regions, for the 20 case study species (occurrence records from Waltari et al. 2007).

Figure S3.2: Model performance for 16 case study species. Model performance is shown for those 16 case study species not highlighted in the main text. Each line represents the set of feature classes that led to the best performance according to either AUC_{TEST} (top and middle panels) or AIC_c (bottom panel), with performance evaluated across a range of regularization multiplier values. AUC_{DIFF} is a measure of model overfitting for the model selected by maximizing AUC_{TEST}. In the AUC plots, the dotted line represents the value for the best-performing model. In the AIC_c plots, the grey shading represents a ΔAIC_c of 4 from the best (lowest) AIC_c score. Performance of models within the grey polygon is not considered to be substantially different (Burnham and Anderson 2004).

Figure S3.3: Model performance for 16 case study species. Model performance is shown for those 16 case study species not highlighted in the main text. Each line represents the set of feature classes that led to the best performance according to either AUC_{TEST} (top and middle panels) or AIC_c (bottom panel), with performance evaluated across a range of regularization multiplier values. AUC_{DIFF} is a measure of model overfitting for the model selected by maximizing AUC_{TEST}. In the AUC plots, the dotted line represents the value for the best-performing model. In the AIC_c plots, the grey shading represents a ΔAIC_c of 4 from the best (lowest) AIC_c score. Performance of models within the grey polygon is not considered to be substantially different (Burnham and Anderson 2004).

Figure S3.4: Model performance for 16 case study species. Model performance is shown for those 16 case study species not highlighted in the main text. Each line represents the set of feature classes that led to the best performance according to either AUC_{TEST} (top and middle panels) or AIC_c (bottom panel), with performance evaluated across a range of regularization multiplier values. AUC_{DIFF} is a measure of model overfitting for the model selected by maximizing AUC_{TEST}. In the AUC plots, the dotted line represents the value for the best-performing model. In the AIC_c plots, the grey shading represents a ΔAIC_c of 4 from the best (lowest) AIC_c score. Performance of models within the grey polygon is not considered to be substantially different (Burnham and Anderson 2004).

Figure S3.5: Model performance for 16 case study species. Model performance is shown for those 16 case study species not highlighted in the main text. Each line represents the set of feature classes that led to the best performance according to either AUC_{TEST} (top and middle panels) or AIC_c (bottom panel), with performance evaluated across a range of regularization multiplier values. AUC_{DIFF} is a measure of model overfitting for the model selected by maximizing AUC_{TEST}. In the AUC plots, the dotted line represents the value for the best-performing model. In the AIC_c plots, the grey shading represents a ΔAIC_c of 4 from the best (lowest) AIC_c score. Performance of models within the grey polygon is not considered to be substantially different (Burnham and Anderson 2004).

Figure S3.6: Predicted habitat suitability in the present for 16 case study species not highlighted in the main text, from Maxent models optimized in terms of feature class and regularization parameter according to the AIC_c metric, for models constructed with and without the ENVIREM variables. Suitability scores range from 0 (blue) to 1 (red). Habitat suitability is shown within the training region only, with predicted habitat suitability below a 95% training presence threshold considered to be unsuitable (grey).

	neotropical	palearctic	nearctic	indo-malay	afrotropic	oceania	australasia	global
annualPET	bio 1	bio 10	bio 1	bio 1	bio 5	bio 9	bio 10	bio 1
aridityIndexThornthwaite	bio 17	bio 12	bio 18	bio 15	bio 17	bio 18	bio 12	bio 17
climaticMoistureIndex	bio 12	bio 2	bio 2	bio 12	bio 12	bio 12	bio 12	bio 12
continentality	bio 4	bio 4	bio 4	bio 4	bio 4	bio 4	bio 4	bio 4
embergerQ	bio 12	bio 12	bio 12	bio 12	bio 12	bio 12	bio 12	bio 12
growingDegDays0	bio 1	bio 1	bio 1	bio 1	bio 1	bio 1	bio 1	bio 1
growingDegDays5	bio 1	bio 1	bio 1	bio 1	bio 1	bio 1	bio 1	bio 1
maxTempColdest	bio 11	bio 11	bio 11	bio 11	bio 11	bio 9	bio 11	bio 11
minTempWarmest	bio 10	bio 10	bio 10	bio 10	bio 10	bio 10	bio 10	bio 10
monthCountByTemp10	bio 10	bio 10	bio 1	bio 1	bio 11	bio 5	bio 1	bio 1
PETColdestQuarter	bio 11	bio 3	bio 3	bio 11	bio 11	bio 9	bio 11	bio 11
PETDriestQuarter	bio 9	bio 9	bio 9	bio 9	bio 9	bio 18	bio 9	bio 9
PETseasonality	bio 4	bio 5	bio 5	bio 7	bio 4	bio 3	bio 4	bio 7
PETWarmestQuarter	bio 5	bio 5	bio 5	bio 5	bio 5	bio 5	bio 2	bio 5
PETWettestQuarter	bio 8	bio 8	bio 8	bio 8	bio 8	bio 2	bio 8	bio 8
thermicityIndex	bio 1	bio 10	bio 10	bio 11	bio 1	bio 1	bio 11	bio 1

Table S3.1: Bioclimatic variables with the strongest correlation with ENVIREM variables, within each biogeographic realm and globally. Descriptions of the WorldClim bioclimatic variables can be found in Table 4.

	neotropical (present)	neotropical (mid-Hol.)	palearctic (present)	palearctic (mid-Hol.)	nearctic (present)	nearctic (mid-Hol.)	indo–malay (present)	indo–malay (mid-Hol.)	afrotropic (present)	afrotropic (mid-Hol.)	oceania (present)	oceania (mid-Hol.)	australasia (present)	australasia (mid-Hol.)	global (present)	global (mid-Hol.)
annualPET	0.88	0.88	0.94	0.93	0.96	0.96	0.8	0.79	0.9	0.86	0.83	0.87	0.94	0.93	0.91	0.91
aridityIndexThornthwaite	-0.84	-0.85	-0.78	-0.74	-0.73	-0.73	0.89	0.88	-0.83	-0.81	-0.8	-0.86	-0.91	-0.9	-0.81	-0.81
climaticMoistureIndex	0.93	0.93	-0.82	-0.8	-0.59	-0.59	0.91	0.9	0.98	0.98	0.89	0.9	0.95	0.95	0.81	0.81
continentality	1	1	1	1	0.99	0.99	0.99	0.99	0.99	0.99	1	0.99	1	1	1	1
embergerQ	0.95	0.95	0.91	0.92	0.94	0.94	0.88	0.88	0.94	0.93	0.92	0.92	0.97	0.97	0.93	0.92
growingDegDays0	1	1	0.93	0.92	0.89	0.88	1	1	1	1	1	1	1	1	0.97	0.97
growingDegDays5	1	1	0.91	0.9	0.85	0.84	1	0.99	1	1	1	1	1	1	0.96	0.96
maxTempColdest	0.98	0.98	1	1	1	1	0.97	0.97	0.92	0.92	0.98	0.98	0.97	0.96	1	1
minTempWarmest	0.98	0.98	0.99	0.99	0.98	0.97	0.96	0.93	0.96	0.94	1	1	0.96	0.96	0.98	0.97
monthCountByTemp10	0.87	0.87	0.95	0.94	0.93	0.92	0.74	0.79	0.49	0.58	0.62	0.73	0.7	0.74	0.95	0.95
PETColdestQuarter	0.93	0.93	0.87	0.87	0.87	0.88	0.91	0.92	0.82	0.79	0.58	0.43	0.93	0.92	0.9	0.89
PETDriestQuarter	0.83	0.84	0.92	0.92	0.87	0.86	0.89	0.85	0.84	0.8	-0.74	-0.8	0.74	0.71	0.87	0.87
PETseasonality	0.98	0.97	0.73	0.75	0.91	0.93	0.97	0.97	0.93	0.91	-0.76	-0.62	0.96	0.95	0.7	0.69
PETWarmestQuarter	0.74	0.75	0.98	0.97	0.99	0.99	0.94	0.93	0.84	0.8	0.84	0.82	0.94	0.94	0.91	0.9
PETWettestQuarter	0.79	0.8	0.89	0.89	0.95	0.95	0.68	0.66	0.62	0.45	-0.82	-0.82	0.91	0.89	0.83	0.81
thermicityIndex	1	1	0.98	0.98	0.98	0.98	0.96	0.97	0.99	0.98	1	1	0.97	0.97	0.98	0.98

Table S3.2: Pearson correlations between ENVIREM and WorldClim variables for current and mid-Holocene climate. Correlations are shown between each of the climatic ENVIREM variables and the WorldClim bioclimatic variable with which the ENVIREM variable is most strongly correlated (Table S1), globally and in separate biogeographic realms. For each variable and realm, the bottom-left triangle contains the correlation coefficient in the present, and the top-right triangle contains the correlation coefficient in the mid-Holocene (CCSM4) for the same bioclimatic variable. Grey shading indicates that the absolute value of the correlation is $\leqslant 0.85$.

Variable	*Arborimus longicaudus* | bioclim	*Arborimus longicaudus* | all-vars	*Arborimus longicaudus* | all-vars+topo	*Blarina brevicauda* | bioclim	*Blarina brevicauda* | all-vars	*Blarina brevicauda* | all-vars+topo	*Chamaea fasciata* | bioclim	*Chamaea fasciata* | all-vars	*Chamaea fasciata* | all-vars+topo	*Crotalus atrox* | bioclim	*Crotalus atrox* | all-vars	*Crotalus atrox* | all-vars+topo	*Desmognathus wrighti* | bioclim	*Desmognathus wrighti* | all-vars	*Desmognathus wrighti* | all-vars+topo	*Dicamptodon tenebrosus* | bioclim	*Dicamptodon tenebrosus* | all-vars	*Dicamptodon tenebrosus* | all-vars+topo	*Dicrostonyx groenlandicus* | bioclim	*Dicrostonyx groenlandicus* | all-vars	*Dicrostonyx groenlandicus* | all-vars+topo	*Elaphe obsoleta* | bioclim	*Elaphe obsoleta* | all-vars	*Elaphe obsoleta* | all-vars+topo
annual mean temp [bio1]																								
mean diurnal temp range [bio2]	+			+			+	+	+	+	+	+	+	+	+	+						+	+	+
isothermality [bio3]	+	+	+	+	+	+	+	+	+					+	+	+		+	+			+		
temp seasonality [bio4]							+			+						+								
max temp warmest month [bio5]	+												+	+	+				+					
min temp coldest month [bio6]	+	+	+				+	+	+															
temp annual range [bio7]				+	+	+													+	+	+		+	+
mean temp of wettest quarter [bio8]	+			+			+			+	+	+	+	+	+	+						+	+	+
mean temp of driest quarter [bio9]	+	+	+							+	+	+										+		
mean temp of warmest quarter [bio10]				+						+	+	+										+		
mean temp of coldest quarter [bio11]																								
annual precip [bio12]				+	+	+																		
precip of wettest month [bio13]																						+		
precip of driest month [bio14]							+	+	+	+	+	+												
precip seasonality [bio15]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
precip of wettest quarter [bio16]										+	+	+												
precip of driest quarter [bio17]		+	+													+						+	+	+
precip of warmest quarter [bio18]							+	+	+				+	+	+	+	+	+	+	+	+			
precip of coldest quarter [bio19]	+	+	+				+	+	+	+						+	+	+						
annualPET																				+	+			
climaticMoistureIndex					+	+					+	+		+	+					+	+			
continentality								+	+											+	+			
embergerQ																							+	+
growingDegDays0																								
growingDegDays5																								
maxTempColdest																								
minTempWarmest		+	+		+	+																	+	+
PETColdestQuarter																				+	+			
PETDriestQuarter					+	+														+	+			
PETseasonality					+	+		+	+		+	+		+	+									
PETWarmestQuarter											+	+												
PETWettestQuarter		+	+		+	+		+	+								+	+						
thermicityIndex																								
topoRoughness						+																		+
topoWetness			+						+			+			+			+			+			

Table S3.3: Variables included in final models for 16 case study species. Variables to include in each model were selected using a statistical variable selection approach (see Methods section for additional details).

	Eumeces fasciatus | bioclim	Eumeces fasciatus | all-vars	Eumeces fasciatus | all-vars+topo	Glaucomys sabrinus | bioclim	Glaucomys sabrinus | all-vars	Glaucomys sabrinus | all-vars+topo	Glaucomys volans | bioclim	Glaucomys volans | all-vars	Glaucomys volans | all-vars+topo	Lampropeltis zonata | bioclim	Lampropeltis zonata | all-vars	Lampropeltis zonata | all-vars+topo	Lepus arcticus | bioclim	Lepus arcticus | all-vars	Lepus arcticus | all-vars+topo	Martes americana | bioclim	Martes americana | all-vars	Martes americana | all-vars+topo	Myodes gapperi | bioclim	Myodes gapperi | all-vars	Myodes gapperi | all-vars+topo	Plethodon idahoensis | bioclim	Plethodon idahoensis | all-vars	Plethodon idahoensis | all-vars+topo
annual mean temp [bio1]																						+		
mean diurnal temp range [bio2]	+						+			+	+	+	+	+	+	+			+			+		
isothermality [bio3]							+	+	+	+	+	+												
temp seasonality [bio4]				+	+	+				+														
max temp warmest month [bio5]				+																				
min temp coldest month [bio6]										+	+	+												
temp annual range [bio7]													+	+	+	+	+	+	+	+	+			
mean temp of wettest quarter [bio8]	+			+			+			+						+	+	+	+	+	+	+		
mean temp of driest quarter [bio9]		+	+													+	+	+				+	+	+
mean temp of warmest quarter [bio10]	+						+									+			+			+		
mean temp of coldest quarter [bio11]																								
annual precip [bio12]							+	+	+															
precip of wettest month [bio13]																+	+	+						
precip of driest month [bio14]																			+	+	+			
precip seasonality [bio15]				+	+	+	+	+	+	+	+	+			+	+	+	+	+	+	+	+	+	+
precip of wettest quarter [bio16]	+																							
precip of driest quarter [bio17]	+	+	+	+	+	+										+	+	+				+	+	+
precip of warmest quarter [bio18]				+	+	+				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
precip of coldest quarter [bio19]				+	+	+				+	+	+										+	+	+
annualPET																								
climaticMoistureIndex		+	+		+	+		+	+					+	+		+	+		+	+			
continentality											+	+												
embergerQ		+	+																					
growingDegDays0																								
growingDegDays5														+	+					+	+		+	+
maxTempColdest																								
minTempWarmest		+	+		+	+		+	+								+	+					+	+
PETColdestQuarter															+									
PETDriestQuarter														+	+									
PETseasonality		+	+		+	+					+	+					+	+		+	+			
PETWarmestQuarter																								
PETWettestQuarter		+	+		+	+		+	+		+	+					+	+					+	+
thermicityIndex																								
topoRoughness		+																						
topoWetness						+			+			+			+			+			+			+

Table S3.4: Variables included in final models for 16 case study species. Variables to include in each model were selected using a statistical variable selection approach (see Methods section for additional details).

		AUC		AIC$_c$	
Species	Variable Set	Feature Class	RM	Feature Class	RM
Ambystoma maculatum	bioclim	LQH	2.50	LQHPT	3.00
	bioclim + envirem-clim	LQHP	2.50	LQHPT	3.00
	bioclim + envirem-all	LQHP	1.00	LQHP	3.50
Dendragapus obscurus	bioclim	H	2.50	LQHPT	2.00
	bioclim + envirem-clim	LQ	2.50	LQHP	4.00
	bioclim + envirem-all	LQHP	0.50	LQHP	4.00
Polioptila californica	bioclim	L	4.00	LQ	2.00
	bioclim + envirem-clim	LQHP	4.00	LQHP	4.00
	bioclim + envirem-all	LQHP	4.00	LQHP	4.00
Poecile gambeli	bioclim	H	2.00	LQHPT	4.00
	bioclim + envirem-clim	LQ	3.00	LQ	2.00
	bioclim + envirem-all	LQ	1.00	LQ	2.00
Arborimus longicaudus	bioclim	LQH	0.50	LQ	0.50
	bioclim + envirem-clim	H	1.00	LQHP	2.00
	bioclim + envirem-all	H	1.50	LQHPT	2.00
Blarina brevicauda	bioclim	H	2.00	LQHPT	1.00
	bioclim + envirem-clim	H	2.50	LQHPT	1.00
	bioclim + envirem-all	H	2.00	LQHPT	1.00
Chamaea fasciata	bioclim	LQHPT	3.50	LQHP	3.00
	bioclim + envirem-clim	LQH	3.00	LQHP	4.00
	bioclim + envirem-all	LQHP	4.00	LQHPT	3.50
Crotalus atrox	bioclim	H	2.50	LQHP	4.00
	bioclim + envirem-clim	LQH	4.00	LQ	2.00
	bioclim + envirem-all	LQHPT	3.50	LQ	2.50
Desmognathus wrighti	bioclim	LQH	1.00	LQH	2.00
	bioclim + envirem-clim	LQ	0.50	LQ	1.50
	bioclim + envirem-all	LQH	2.00	LQ	2.00
Dicamptodon tenebrosus	bioclim	H	4.00	LQ	1.00
	bioclim + envirem-clim	LQH	3.00	LQ	1.00
	bioclim + envirem-all	LQH	3.00	LQ	1.00
Dicrostonyx groenlandicus	bioclim	H	1.00	LQ	0.50
	bioclim + envirem-clim	H	1.00	LQ	0.50
	bioclim + envirem-all	H	1.50	LQ	1.00
Elaphe obsoleta	bioclim	H	4.00	LQHPT	1.50
	bioclim + envirem-clim	H	4.00	LQHPT	2.50
	bioclim + envirem-all	LQHP	4.00	LQHPT	1.50
Eumeces fasciatus	bioclim	LQ	3.00	LQHPT	2.00
	bioclim + envirem-clim	L	4.00	LQH	3.50
	bioclim + envirem-all	L	4.00	LQ	2.50
Glaucomys sabrinus	bioclim	LQH	4.00	LQHP	3.00
	bioclim + envirem-clim	LQHPT	1.00	LQHPT	2.50
	bioclim + envirem-all	LQHPT	1.00	LQHPT	3.00
Glaucomys volans	bioclim	H	4.00	LQHPT	2.00
	bioclim + envirem-clim	LQHP	0.50	LQHPT	1.00
	bioclim + envirem-all	LQHP	1.00	LQHP	4.00
Lampropeltis zonata	bioclim	LQHP	3.50	LQ	2.50
	bioclim + envirem-clim	LQHP	4.00	LQHPT	3.50
	bioclim + envirem-all	LQHP	4.00	LQ	3.50
Lepus arcticus	bioclim	LQHP	4.00	LQ	1.00
	bioclim + envirem-clim	H	4.00	LQ	2.00
	bioclim + envirem-all	H	0.50	LQ	4.00
Martes americana	bioclim	LQHP	3.50	LQHPT	2.50
	bioclim + envirem-clim	LQ	1.00	LQHPT	4.00
	bioclim + envirem-all	LQHPT	3.00	LQHPT	4.00
Myodes gapperi	bioclim	LQHP	1.50	LQHPT	1.00
	bioclim + envirem-clim	LQHPT	1.00	LQHPT	1.50
	bioclim + envirem-all	LQHPT	1.00	LQHPT	1.00
Plethodon idahoensis	bioclim	H	3.00	LQ	0.50
	bioclim + envirem-clim	H	4.00	LQ	1.00
	bioclim + envirem-all	H	4.00	LQ	1.00

Table S3.5: Optimized Maxent parameters for all 20 case study species, using AUC and AIC$_c$. The possible feature classes are linear (L), quadratic (Q), hinge (H), product (P) and threshold (T). The regularization multiplier (RM) controls the complexity of the model.

Species	Schoener's D			
	Current	LGM CCSM4	LGM MPI-ESM-P	LGM MIROC-ESM
Ambystoma maculatum	0.91	0.84	0.76	0.89
Arborimus longicaudus	0.88	0.61	0.47	0.55
Blarina brevicauda	0.94	0.78	0.80	0.77
Chamaea fasciata	0.97	0.96	0.97	0.97
Crotalus atrox	0.93	0.76	0.89	0.75
Dendragapus obscurus	0.93	0.79	0.79	0.83
Desmognathus wrighti	0.82	0.26	0.18	0.41
Dicamptodon tenebrosus	0.96	0.93	0.85	0.86
Dicrostonyx groenlandicus	0.92	0.77	0.79	0.72
Elaphe obsoleta	0.93	0.86	0.83	0.84
Eumeces fasciatus	0.93	0.81	0.77	0.80
Glaucomys sabrinus	0.88	0.68	0.77	0.66
Glaucomys volans	0.88	0.60	0.65	0.62
Lampropeltis zonata	0.88	0.77	0.86	0.86
Lepus arcticus	0.95	0.89	0.90	0.90
Martes americana	0.90	0.78	0.76	0.73
Myodes gapperi	0.90	0.62	0.67	0.53
Plethodon idahoensis	0.91	0.01	0.01	0.02
Poecile gambeli	0.87	0.73	0.54	0.79
Polioptila californica	0.87	0.85	0.89	0.84

Table S3.6: Schoener's D niche overlap for all case study species, between the *bioclim* and *bioclim + envirem-clim* models, in both the present and during the LGM with three GCMs.

CHAPTER IV

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

4.1   Abstract

Species-specific diversification rates, or "tip rates", can be computed quickly from phylogenies and are widely used to study diversification rate variation in relation to geography, ecology, and phenotypes. These tip rates provide a number of theoretical and practical advantages, such as the relaxation of assumptions of rate homogeneity in trait-dependent modeling approaches. However, there is significant confusion in the literature regarding whether these metrics estimate speciation or net diversification rates. Additionally, no study has yet compared the relative performance and accuracy of tip rate metrics.

We compared the statistical performance of three model-free rate metrics (inverse terminal branch lengths; node density metric; DR statistic) and a model-based approach (BAMM). We applied each method to a large set of simulated phylogenies that had been generated under different diversification processes; scenarios included multi-regime time-constant and diversity-dependent trees, as well as trees where the rate of speciation evolves under a diffusion process. We summarized performance in relation to the type of rate variation, the magnitude of rate heterogeneity and rate regime size. We also compared the ability of the metrics to estimate both speciation and net diversification rates.

We show decisively that model-free tip rate metrics estimate the rate of speciation and not net diversification. Error in net diversification rate estimates is high and increases dramatically as a function of the relative extinction rate. In contrast, error in speciation rate estimates is low and relatively insensitive to extinction. Across all diversification scenarios, BAMM inferred the most accurate tip rates and exhibited lower error than non-model-based approaches. DR was highly correlated with true speciation rates but exhibited high sample variance, and was the best metric for very small rate regimes.

We found that DR and BAMM are useful metrics for studying speciation rate dynamics and trait-dependent diversification. Although BAMM was more accurate than DR overall, the two approaches have complementary strengths. Because tip rate metrics are more reliable estimators of speciation rate, we recommend that empirical studies using these metrics exercise caution when drawing biological interpretations in any situation where the distinction between speciation and net diversification is important.

4.2 Introduction

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

Although rates of diversification have traditionally been quantified for clades, there has been a growing interest in estimating species-specific rates of diversification, which we refer

to here as "tip rates". Tip rates are increasingly used to describe patterns of geographic and trait-associated variation in diversification (Freckleton et al. 2008, Jetz et al. 2012, Kennedy et al. 2016, Harvey and Rabosky 2017, Quintero and Jetz 2018, Rabosky et al. 2018). It may seem strange to view evolutionary rates as a property of individual lineages, but such rates emerge naturally from the birth-death model we typically use to conceptualize the diversification process (Nee et al. 1992, Nee et al. 1994). Under the birth-death process, individuals (species) are characterized by per-lineage rates of species origination (speciation, λ) and extinction (μ). For the purposes of inference, these rates are typically assumed to be constant among contemporaneous members of a focal clade. However, tip rates can be viewed as our best estimate of the present-day rate of speciation or extinction for an individual lineage, conditional on past (usually recent) evolutionary history. As such, they provide information about the expected amount of time that will elapse before a lineage splits or becomes extinct.

A number of approaches have been used to estimate tip rates, including both model-based and non-model-based approaches. These approaches vary in terms of how much information they derive from a focal species relative to the amount of information they incorporate from other regions of the phylogeny. On one end of the spectrum, tree-wide estimates of speciation and extinction rates under a constant-rate birth-death (CRBD) model provides tip rates that are maximally auto-correlated across species in the clade; such rates for any given species are not independent of rates for any other species in the group of interest. On the other end of the spectrum, terminal branch lengths can be used to derive a censored estimate of the rate of speciation that is minimally autocorrelated with rates for other species in the focal clade. Terminal branch lengths are largely unique to each species (rates might be identical only for sister taxa), but provide a noisy measure of speciation, due to the stochasticity inherent in the diversification process (Nee et al. 1994). In contrast to single (terminal) branch estimates, tree-wide estimates should be less susceptible to stochastic noise, because they incorporate information from the entirety of the tree (e.g., multiple branches are used

in the estimates). Of course, the tree-wide estimate necessarily assumes that all tips share a common underlying diversification process. Other tip rate metrics fall somewhere between these two extremes, incorporating some tree-wide information but relaxing the assumption of homogeneous rates across all lineages (node density metric: Freckleton et al. 2008, DR: Jetz et al. 2012). The estimation of tip-specific rates thus entails a tradeoff between the precision of individual estimates and the stochastic error associated with those estimates.

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

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

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

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

Despite the potential utility of tip rates in geographic and trait-based analyses of speciation rate heterogeneity (Jetz et al. 2012, Belmaker and Jetz 2015, Oliveira et al. 2017, Quintero and Jetz 2018), there has yet been no comparative assessment of the accuracy and precision of the estimates. BAMM has low power to infer small rate regimes (Rabosky et al. 2017, Meyer and Wiens 2017), leading to the possibility that other approaches might perform better for smaller phylogenies or when the variation in rates among clades is subtle. However, DR and related methods will always identify variation in tip rates, even when

none exists, provided there is stochastic variation in branch lengths. A goal of this study is therefore to evaluate the trade-off between the stochastic noise inherent in non-model-based approaches, and the conservative but less noisy estimates from model-based metrics. We compare the performance of these metrics across a range of simulation scenarios, which include both discrete and continuous variation in rates.

4.3   Methods

4.3.1   Tip rate metrics

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

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

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

We also considered a simpler metric, node density (Freckleton et al. 2008, denoted by λ_{ND}). This is simply the number of splitting events along the path between the root and tip of a phylogeny, divided by the age of the phylogeny. While λ_{DR} down-weights the contribution

of branch lengths that are closer to the root, λ_{ND} equally weights the contributions of all branches along a particular root-to-tip path, regardless of where they occur in time. Under a pure-birth model ($\mu = 0$), both λ_{DR} and λ_{ND} should yield unbiased estimates of the rate of speciation.

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

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

4.3.2 Tip rate metrics estimate speciation, not net diversification

As suggested previously (Belmaker and Jetz 2015), DR and presumably other tip-based measurements, more accurately estimate the rate of speciation than the rate of net diversifi-

cation. However, numerous studies continue to refer to DR as a measure of net diversification (Marin and Hedges 2016, Oliveira et al. 2016, Cai et al. 2017, Quintero and Jetz 2018, and many others). This is incorrect and it is straightforward to demonstrate that λ_{TB}, λ_{ND} and λ_{DR} are more reliable measures of speciation rates and not net diversification rates, at least when extinction is moderate to high.

To illustrate this property of the metrics, we applied all approaches to constant-rate birth-death phylogenies simulated across a range of relative extinction rates ($\varepsilon = \lambda/\mu$) , including pure-birth trees ($\varepsilon = 0$) as well as trees exhibiting very high turnover ($\varepsilon = 1$). To evaluate accuracy of speciation estimates as a function of ε, we generated 1000 phylogenies with 100 tips each, where λ and ε were drawn from uniform distributions (λ: [0.05, 0.3]; ε: [0, 1]). Importantly, when λ is sampled uniformly with respect to ε, the distribution of r is not uniform: the mean, range and variance in r decrease dramatically as ε increases. To evaluate the accuracy of r as a function of ε, we thus generated a second set of trees by sampling r and ε from uniform distributions (r: [0.05, 0.3], ε [0, 1]). As a result, λ has constant mean and variance with respect to ε in the first set of simulations, and the same is true for r in the second set of simulations (Figure S1). All phylogeny simulations were conducted with the TreeSim package in R (Stadler 2011).

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

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

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

where λ_i is the estimated tip rate for species i out of N total species, λ_{TRUE} is the true tip rate. Mean absolute error captures the magnitude in error in tip rates, and mean

proportional error quantifies the bias in tip rates, as a function of the true tip rates (Rabosky et al. 2014).

4.3.3   Assessment of tip rate metrics

We tested the performance of the metrics by compiling publicly-available datasets from a number of simulation-based studies (Table 1). These simulated trees include rate heterogeneity in time and across lineages. Together, these phylogenies present a wide range of tree sizes and diversification rate shifts, providing an ideal comparative dataset for our purposes. To more easily distinguish between these tree types in the text, we refer to the BAMM-type, multi-regime time-constant phylogenies simply as "multi-regime", and the multi-regime diversity-dependent phylogenies simply as "diversity-dependent", even though discrete rate shifts are present in both types of trees. In addition to discrete-shift scenarios (e.g., BAMM-type process), we simulated phylogenies under an "evolving rates" model of diversification (Rabosky 2010; as corrected in Beaulieu and O'Meara 2015) to explore performance of tip rate metrics when diversification rates change continuously and independently along branches, as might occur if diversification rates are correlated with an underlying continuous trait (FitzJohn 2010). In these simulations, we allowed the logarithm of λ to evolve across the tree under a Brownian motion process, while holding ε constant. The magnitude of rate heterogeneity among branches is controlled by the diffusion parameter σ, where greater values lead to greater heterogeneity in speciation rates. Although published phylogenies with rate data were unavailable for this simulation scenario, we used simulation code and parameters taken directly from Beaulieu and O'Meara (2015) to generate trees with similar statistical properties to those in their study. Simulations were performed with the following parameters: $\lambda = 0.078, 0.103, 0.145, 0.249$ and $\varepsilon = 0.0, 0.25, 0.50, 0.75$. We simulated 100 phylogenies for each (λ, ε) pair, and for three values of σ ($\sigma = 0.03, 0.06, 0.12$). We evaluated tip rate accuracy by comparing estimated to true tip rates, using the absolute and proportional error metrics described above. We also examined the correlation between true

116

and estimated tip rates, combining tip rates from all phylogenies generated under the same class of diversification process, and visualizing these data as density scatterplots, generated with the LSD package in R (Schwalb et al. 2018), where colors indicate the density of points.

Size of diversification rate regimes might be an important factor in a tip rate metric's ability to accurately estimate rates. For example, BAMM's statistical power in detecting a shift to a new rate regime is a function of the number of taxa in that rate regime, and tip rates for taxa from small regimes will more likely be parameterized according to the larger parent regime or the tree-wide average rate (Rabosky et al. 2017); this is the expected behavior when BAMM fails to identify a rate shift. However, non-model-based approaches such as those examined in this study might be more accurate for small regimes. To explore how rate regime size influences the accuracy of tip rate metrics, we calculated the mean tip rate for each true rate regime from all multi-regime phylogenies (simulation datasets from Moore et al. 2016, Rabosky et al. 2017, Meyer and Wiens 2017, Mitchell et al. 2018). We then calculated the Pearson correlation coefficient and the slope of a linear model between true and estimated mean regime rates. We explored the performance of all metrics with respect to regime sample size, as in Rabosky et al. (2017: Figure 13). For comparison, we repeated all performance summaries on tip rates estimated by applying a simple constant-rate birth-death (CRBD) process to each simulated phylogeny. This exercise is an important control, because it indicates how much error we would expect for each simulated phylogeny under the simplifying (incorrect) assumption that rates are constant among lineages and through time for each dataset.

4.4 Results

4.4.1 Speciation or net diversification?

As expected, the tip rate metrics examined in this study are more accurate estimators of the rate of speciation (λ) and not the net rate of species diversification (r). Mean abso-

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

4.4.2 Tip rate accuracy across rate-variable phylogenies

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

rates, with the exception of λ_{ND}, which was not at all correlated with true rates (Figure S6).

In terms of mean per-tip error, λ_{BAMM} consistently outperformed the other metrics for multi-regime, diversity-dependent and evolving rates trees (Figure 3). Error in λ_{BAMM} increased as a function of rate heterogeneity for evolving rate phylogenies, but was largely independent of the magnitude of rate heterogeneity for the other scenarios. λ_{DR} generally exhibited greater error than λ_{BAMM}, and this error increased as a function of the level of heterogeneity for both the evolving rates and multi-regime trees. Error in λ_{DR} was generally invariant to the number of rate regimes for the diversity-dependent scenarios. However, λ_{DR} tended to have greater error than tree-wide estimates of speciation rates from a simple model that assumes no variation in rates through time or among lineages (λ_{CRBD}). λ_{ND} performed somewhat similarly to λ_{DR} for constant-rate and evolving rates trees, but worse for diversity-dependent trees. Error in λ_{TB} increased with increasing rate heterogeneity for constant-rate and evolving rates trees, but was relatively unaffected by rate heterogeneity in diversity-dependent trees (Figure S7). However, error for this metric was far greater than for all other tip metrics.

4.4.3   Effects of regime size on performance

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

Absolute error in regime mean tip rates was lowest for λ_{DR} and λ_{BAMM}, regardless of the size of the rate regime (Figure 5). BAMM's ability to accurately estimate tip rates improved with regime size, whereas absolute error was relatively consistent across regime sizes for λ_{DR} for regimes greater than 10 species. We also found that λ_{DR} slightly outperformed λ_{BAMM} for small rate regimes.

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

4.5   Discussion

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

In terms of accuracy, we found that BAMM performed better than non-model-based metrics across all datasets we considered: estimated tip rates were most highly correlated

with true tip rates, and mean per-tip error in rates was lower across a range of rate-variable simulation scenarios. BAMM is expected to perform well for phylogenies with discrete shifts in diversification rates as this type of rate variation is most consistent with BAMM's assumptions (Rabosky 2014, Mitchell and Rabosky 2016, Rabosky et al. 2017, Mitchell et al. 2018). However, BAMM performed surprisingly well for the evolving rates phylogenies, which conform poorly to the assumptions of the inference model. In these trees, the rate of speciation changes continuously under a diffusion process, and as a result, the phylogeny exhibits rate heterogeneity without discrete rate shifts.

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

Why do λ_{BAMM} and λ_{DR} exhibit such striking differences in performance across the simulation scenarios considered here? To illustrate the differences between inference under these metrics, we compared true tip rates to λ_{BAMM} and to λ_{DR} on a simulated birth-death tree with a single rate shift (Figure 6), as well as on one evolving rates tree simulated for this study (Figure 7). It is clear that if BAMM has the statistical power to detect true rate shifts, then it will perform well under rate shift scenarios. In contrast, λ_{DR} tracks true rate shifts but exhibits high sample variance. With an evolving rates tree (Figure 7), the simulation model is very different from the inference model in BAMM. However, it conservatively places rate shifts in order to accommodate rate heterogeneity that is spread across the phylogeny under a rather different model of rate variation. λ_{DR} also broadly tracks the overall pattern of the true rates, but the variance in the corresponding estimates is so high that performance is negatively affected. If we calculate mean (absolute) per-tip error in λ_{BAMM} and λ_{DR}, the error is relatively similar between λ_{BAMM} and λ_{DR}, but the variance in per-tip error for λ_{DR} is higher. Overall, BAMM exhibited substantially lower error than λ_{DR} under precisely this scenario (Figure 3).

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

Regardless of the performance summaries presented in this article, important questions remain with respect to how well tip rate metrics can estimate the true rate of speciation from empirical phylogenies. The phylogenies analyzed in this study were simulated under idealized processes and neglect potential biases and sources of uncertainty that are present in real datasets. For example, if the process of speciation takes time to complete, as is generally believed to be the case (i.e., the protracted speciation process; Rosindell et al. 2010, Etienne and Rosindell 2012), then the most recent speciation events may still be on-going at the present and typical species-level molecular phylogenies may fail to recognize these events. This will lead to an overestimation of terminal branch lengths, as some terminal branches potentially include incipient species. A related bias might arise due to incomplete taxon sampling, which disproportionately affects the length of terminal (or otherwise recent) branch lengths (Pybus and Harvey 2000). Likewise, variation in taxonomic practice across a phylogeny might lead to spurious rate variation, particularly if different species concepts are used, or if some clades in the phylogeny – but not others – have been subject to population genetic analysis or screens for cryptic species diversity. Additionally, it has been shown that BAMM and other methods may fail to infer accurate speciation rate dynamics if the phy-

logeny is in diversity decline – that is, when extinction rates increase towards the present and ultimately exceed speciation rates (Quental and Marshall 2011, Burin et al. 2018). A major, if obvious, caveat in the interpretation of tip rates is that they apply to recent speciation rates and are necessarily limited with respect to inferences about historical variation in speciation rate.

The greater the importance of the terminal branches in tip rate metrics, the greater the impact these biases might have on tip rate estimates. On one end of the spectrum, metrics such as λ_{TB} will be very sensitive to such biases as they rely exclusively on terminal branch lengths. Such approaches may retain utility as summary statistics (e.g., Bromham et al. 2016), but we did not find λ_{TB} to be an accurate estimate of speciation rates in any of our analyses. On the other end of the spectrum, a metric like λ_{ND} would be minimally impacted as this metric is attempting to capture an average speciation rate over an entire root-to-tip path and does not upweight the contribution of recent branch lengths. λ_{DR} is likely somewhere in the middle of this spectrum, as it gives decreasing weight to branches towards the root. λ_{BAMM} is potentially sensitive to such issues as well, although it may be possible to analytically correct for some biases in the mechanics of the model itself (e.g., Rosindell et al. 2010, Etienne and Rosindell 2012).

Potential empirical biases aside, tip rates present a number of practical advantages in the study of diversification rate variation. First, tip rates can be summarized and compared across non-monophyletic assemblages of species (Jetz et al. 2012, Kennedy et al. 2016, Belmaker and Jetz 2015, Oliveira et al. 2017, Quintero and Jetz 2018, Rabosky et al. 2018), making it possible to summarize rate characteristics of entire communities or regional assemblages of species. Second, estimation of rates at the present should be more robust to the influence of extinction, as extinction can erase the history of lineage splitting deeper in the phylogeny (Nee et al. 1994, Nee et al. 1994, Rabosky and Lovette 2008). Third, tip-specific rates can be paired with species-specific trait values or geographic attributes in order to test potential trait- or geography-dependent speciation rates (Freckleton et al. 2008,

Jetz et al. 2012, Rabosky and Goldberg 2017, Harvey and Rabosky 2017). Tip rates make it possible to relax strong assumptions of rate homogeneity within character states, which are inherent to certain trait-dependent models, including BiSSE and GeoSSE (Maddison et al. 2007, Goldberg et al. 2011, Ng and Smith 2014). Recent work has provided a conceptually rich and robust interpretive framework for SSE models that does not assume rate-constancy within character states (Beaulieu and O'Meara 2016, Caetano et al. 2018), but tip rates nonetheless can provide an important check on results obtained with SSE models by providing a direct means of visualizing the relationship between branch lengths and character states (Bromham et al. 2016, Hua and Bromham 2016, Harvey and Rabosky 2017). Visual inspection of data in this fashion has the potential to reduce false positives by calling attention to potential outliers and other sources of model inadequacy (Maddison and FitzJohn 2014, Rabosky and Goldberg 2015). A final advantage for non-model based tip rates, especially λ_{DR}, is that they can profitably be applied to extremely large phylogenies: there are few computational limits to using them on phylogenies with tens of thousands of tips or more, in contrast to formal model-based approaches for which BAMM, HiSSE, and other methods are poorly suited.

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

4.5.1 Acknowledgements
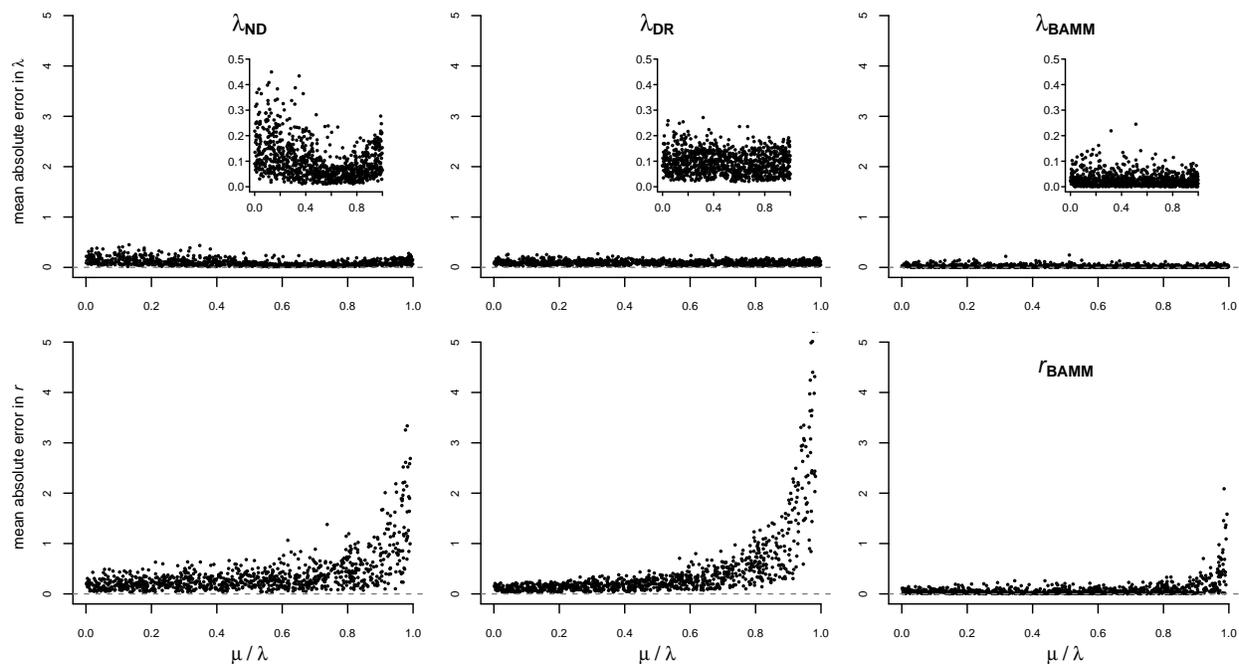
Figure 4.1: Mean absolute error in tip rate metrics for speciation and net diversification rate. Error in λ is shown in top panels, and error in r in bottom panels, for three different tip rate metrics, across a range of relative extinction rates. For BAMM, the estimated speciation and net diversification rates are presented in the top and bottom panels, respectively. Absolute error of zero implies perfect accuracy. Inset plots show error in λ with truncated y-axis scale to facilitate comparison among metrics. All tip rate metrics track λ more accurately than they track r. See Figure S3 for λ_{TB}, which performed much worse than the other metrics.
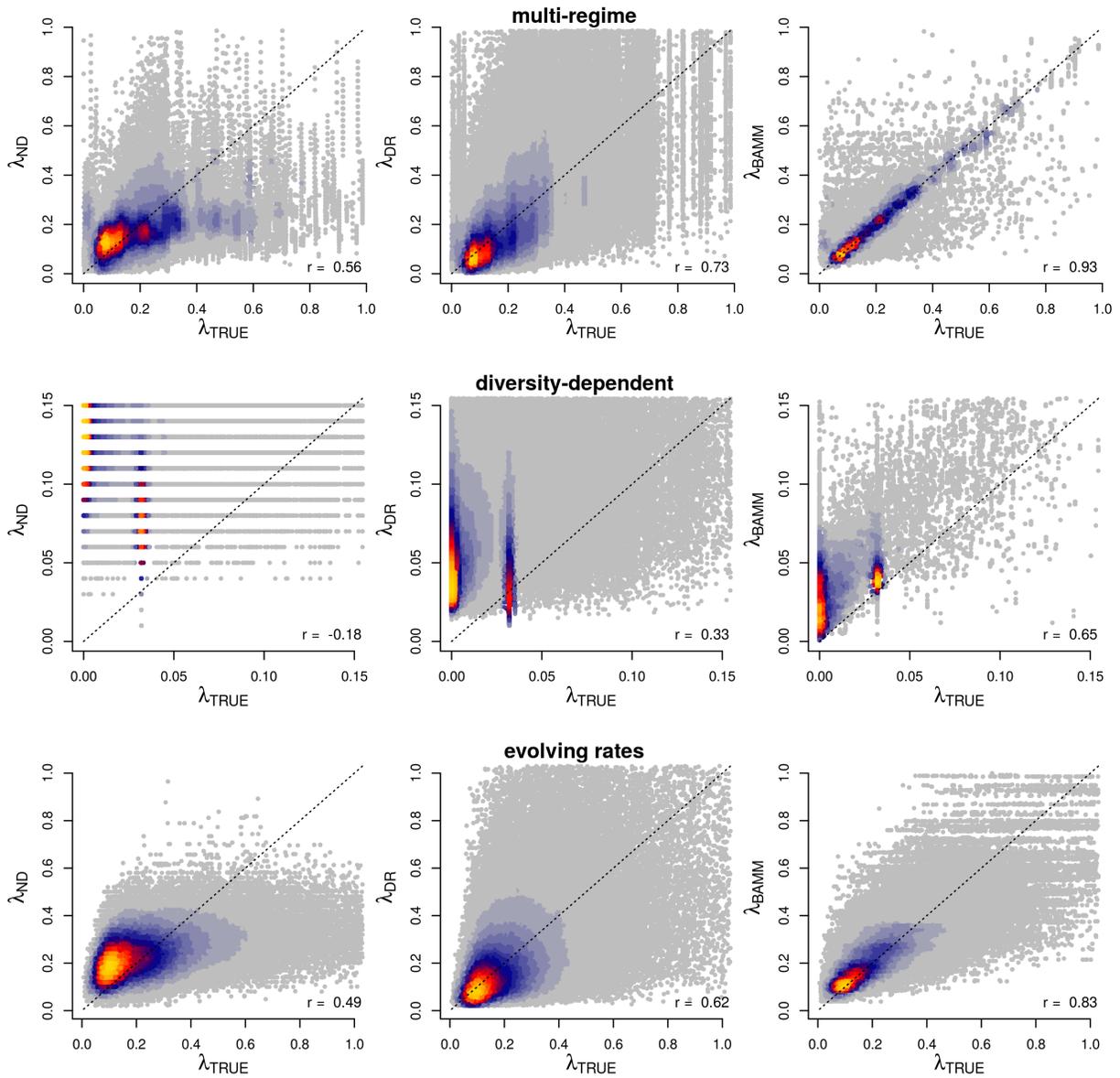
Figure 4.2: True tip rates (λ_{TRUE}) in relation to estimated tip rates. Tip rates were compared separately for different major categories of phylogeny simulations (rows). Plotting region is restricted to the 99th percentile of true rates, but Spearman correlations between true and estimated rates (lower right of each figure panel) are based on the full range of the data. Colors indicate the density of points in the scatter plots. The horizontal gaps in λ_{ND} for diversity-dependent trees are an artefact of all trees having the same crown age. λ_{BAMM} exhibited the strongest correlation with true rates for all simulation categories.
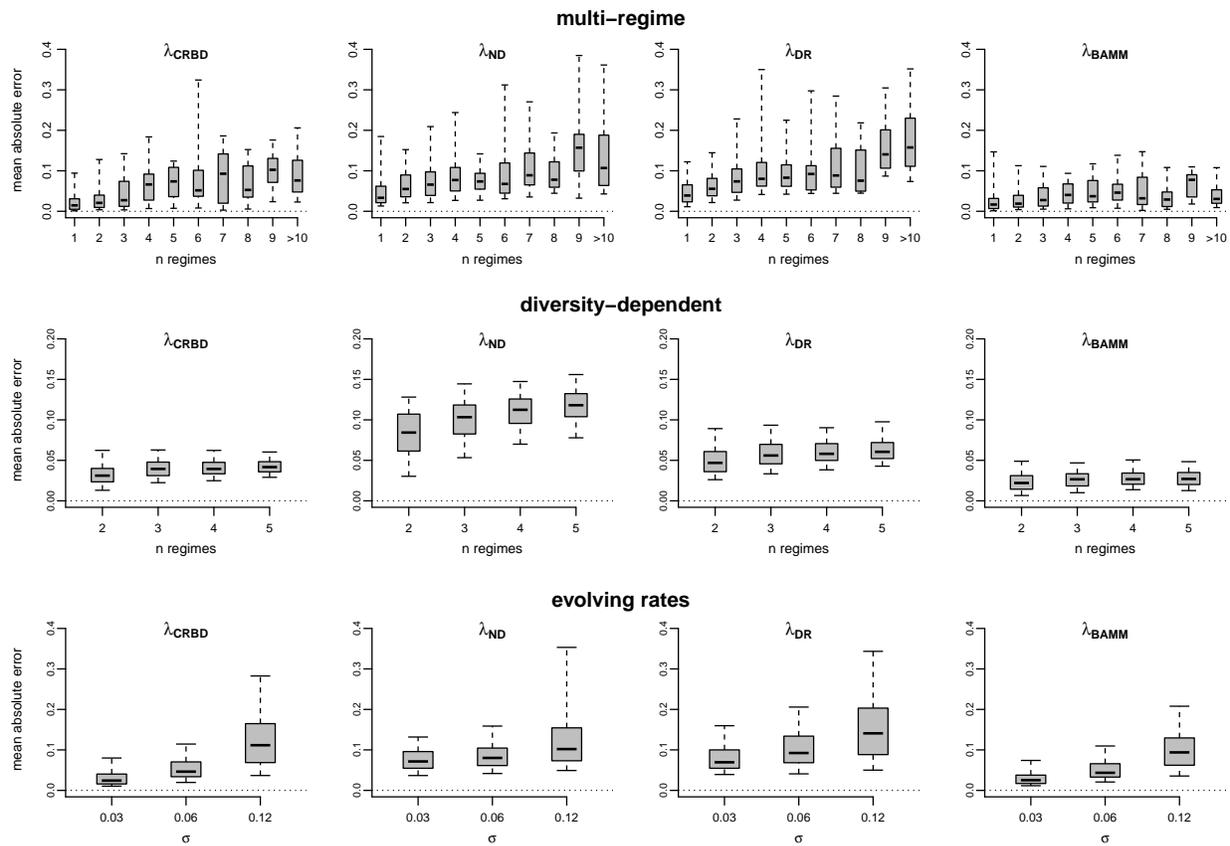
Figure 4.3: Mean per-tip absolute error in speciation rates as a function of the magnitude of rate heterogeneity in each simulated phylogeny. Results are presented separately for different categories of rate variation (Table 1); left column shows estimates from a constant-rate birth-death model for reference. The boxes and whiskers represent the $0.25 - 0.75$, and the $0.05 - 0.95$ quantile ranges, respectively. In some cases, λ_{ND} and λ_{DR} had more error than a simple CRBD model with no variation in tip rates. λ_{BAMM} had the least amount of error across all amounts of rate heterogeneity. See Figure S7 for λ_{TB}.
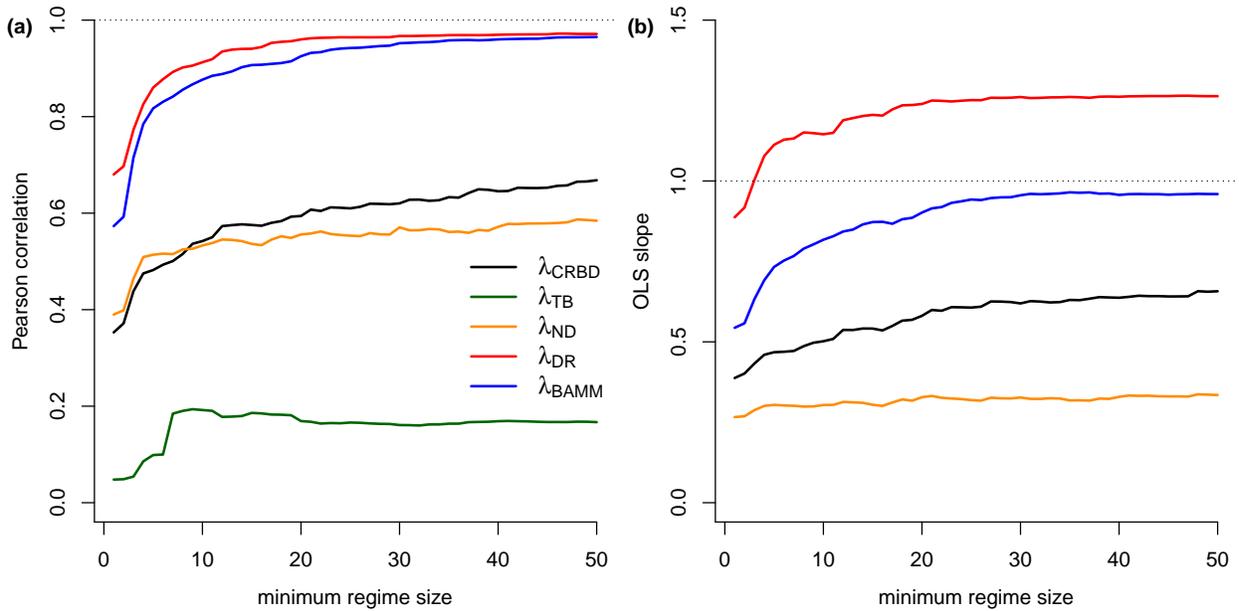
Figure 4.4: Performance of tip rate metrics as a function of minimum regime size, including Pearson correlation **(a)** and OLS regression slope **(b)** for mean rates with respect to λ_{TRUE}. λ_{DR} and λ_{BAMM} outperform the other metrics when summarized in this fashion, although λ_{DR} tends to overestimate the rate of speciation. The x-axis denotes the minimum regime size across which performance was summarized. For example, x = 20 corresponds to the correlations and slopes computed for all regimes with 20 or more tips; a value of x = 1 is the corresponding results for all regimes. The OLS slope for λ_{TB} is not visible as it ranges between 15 and 23.
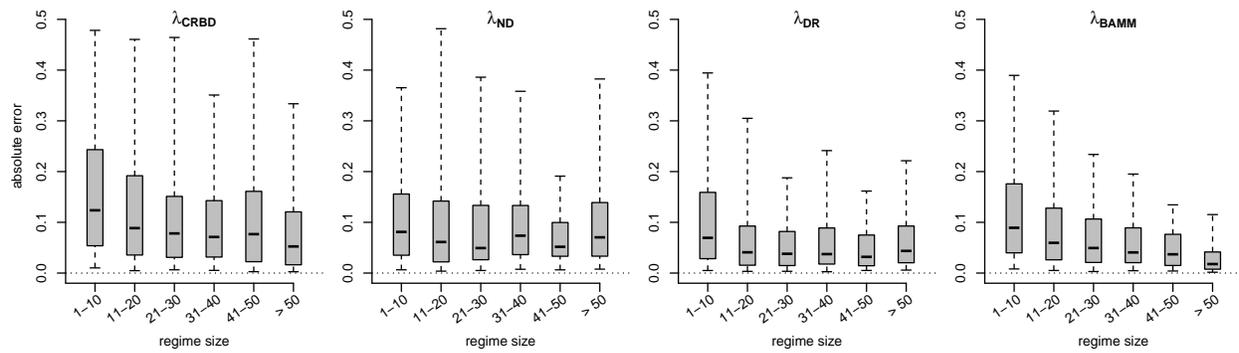
Figure 4.5: Mean per-regime absolute error in relation to true rate regime size, as binned into 10 size categories. The boxes and whiskers represent the $0.25 - 0.75$, and the $0.05 - 0.95$ quantile ranges, respectively. Perfectly estimated rates have an error of zero. λ_{DR} and λ_{BAMM} exhibit the least error when averaged by regimes, and λ_{DR} does slightly better for small clades (10-clade median error 0.07 for λ_{DR}, and 0.08 for λ_{BAMM}).
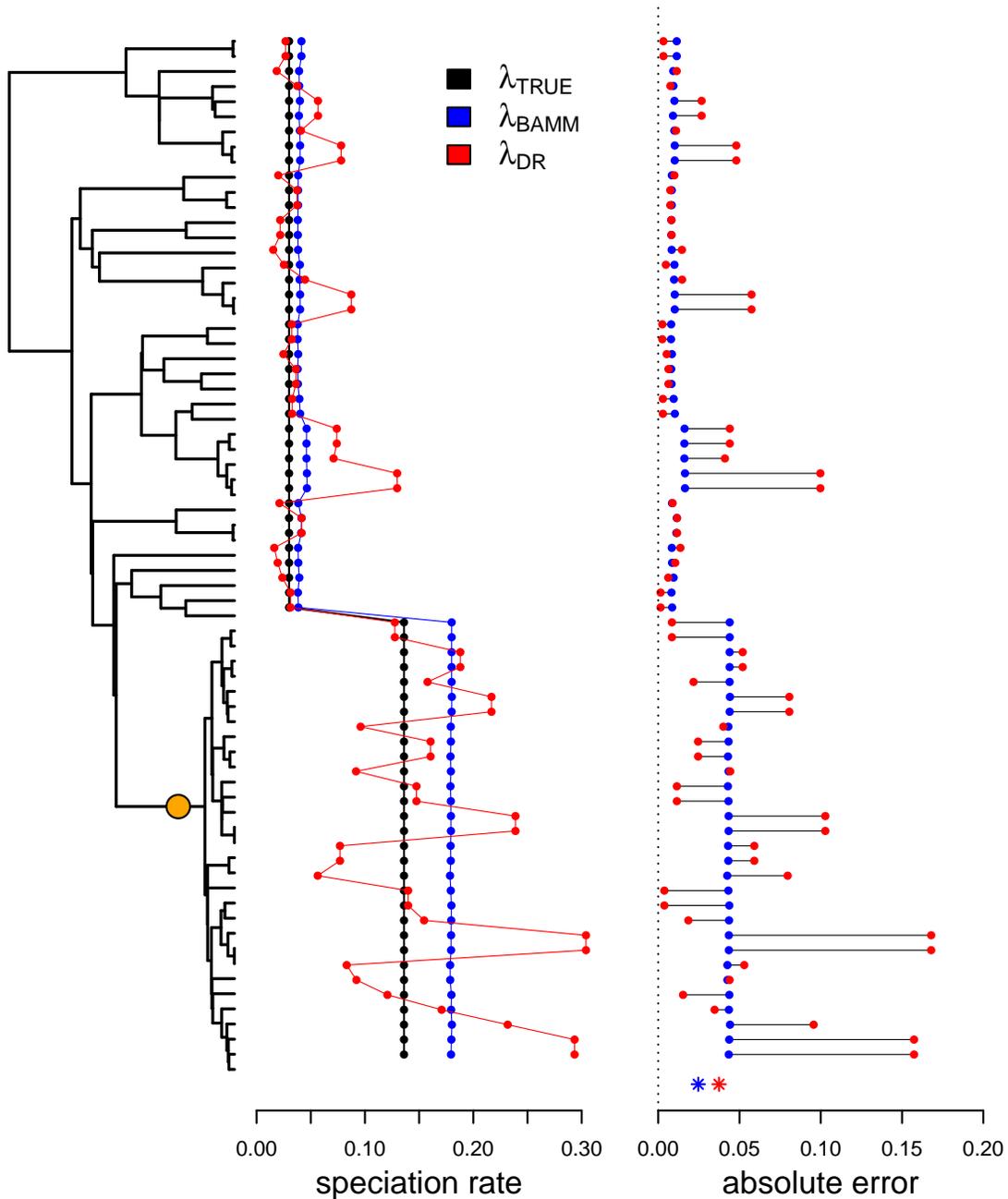
Figure 4.6: Examination of true and estimated tip rates in a single rate-shift tree. The location of the rate shift is represented by an orange circle. Subplots to the right of the tree illustrate true and estimated rates for each tip (left) and corresponding absolute error (right). Asterisks at the bottom denote mean per-tip error in tip rate metrics. Mean per-tip error is relatively low and similar between λ_{DR} and λ_{BAMM}, but the sample variance in λ_{DR} tip rates is high. In this example, the variance in absolute per-tip error in λ_{DR} is 0.002 versus 0.0003 for λ_{BAMM}. On average, λ_{DR} tends to either overestimate or underestimate rates relative to λ_{BAMM}, even if the mean per-tip error is relatively low for both metrics.

131
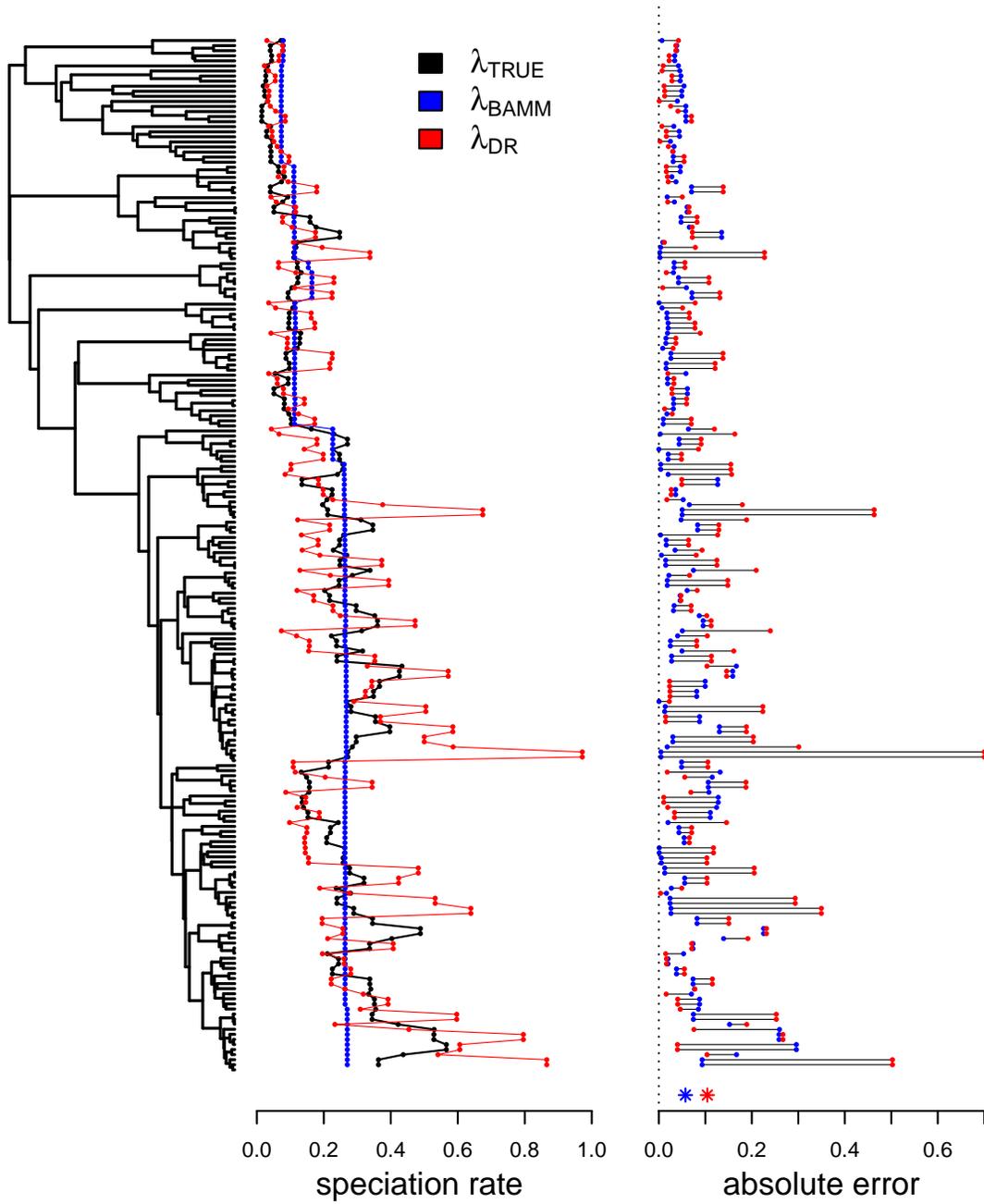
Figure 4.7: Examination of true and estimated tip rates in a simulated "evolving rates" tree, such that the speciation rate itself varies under a diffusion model. See Figure 6 for additional details. Neither metric is particularly well equipped to infer the true rate variation in this case. However λ_{BAMM}'s conservative estimates are still more accurate relative to λ_{DR}, which is negatively impacted by high variance in tip rates. Here, variance in absolute per-tip error in λ_{DR} is 0.012 versus 0.003 for λ_{BAMM}.

132

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

Table 4.1: Summary of simulated phylogenies.

simulations for evaluating speciation rate
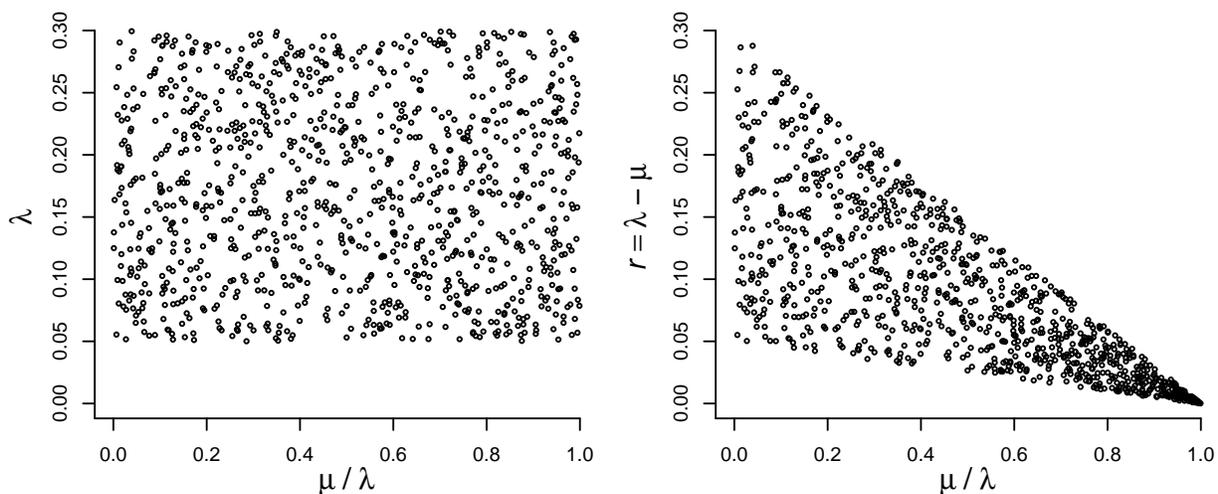
**simulations for evaluating net diversification rate**
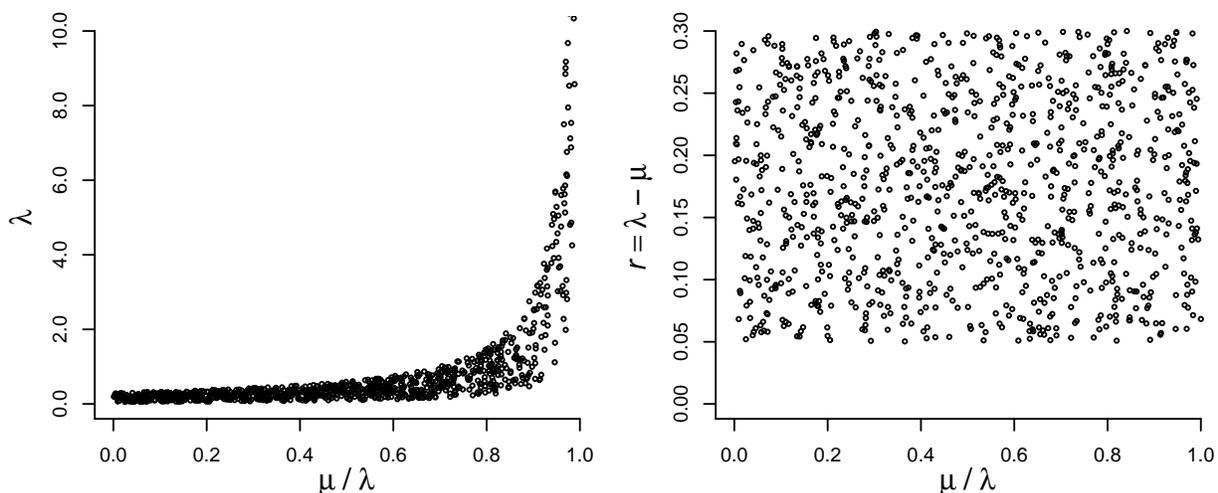
Figure S4.1: Details of simulations for disentangling speciation from net diversification rate. From the top row, it is clear that when λ is sampled uniformly with respect to ε, the distribution of r is not uniform: the mean, range and variance in r decrease dramatically as ε increases. The reverse is true for the distribution of λ when r is sampled uniformly with respect to ε (bottom row). Our simulation design ensures that λ and r are sampled from identical uniform distributions with respect to ε and ensures comparability of the resulting error estimates.

134
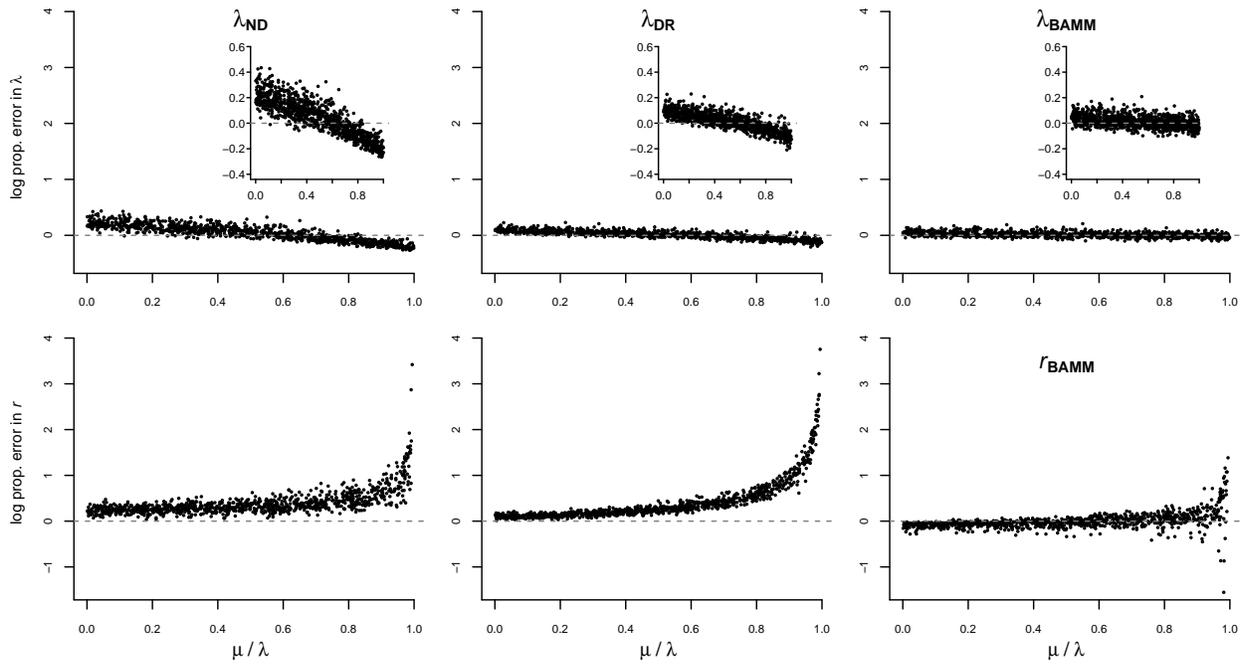
Figure S4.2: Log proportional accuracy in λ (top) and r (bottom) for different tip rate metrics, across a range of relative extinction rates. For BAMM, the estimated net diversification rate is presented. Proportional error of 0 implies perfect accuracy. Inset plots reveal greater detail in error for λ to ease metric comparison. All tip metrics track λ much more accurately than they track r, and λ_{BAMM} does so with the least amount of error. See Figure S3 for λ_{TB}.
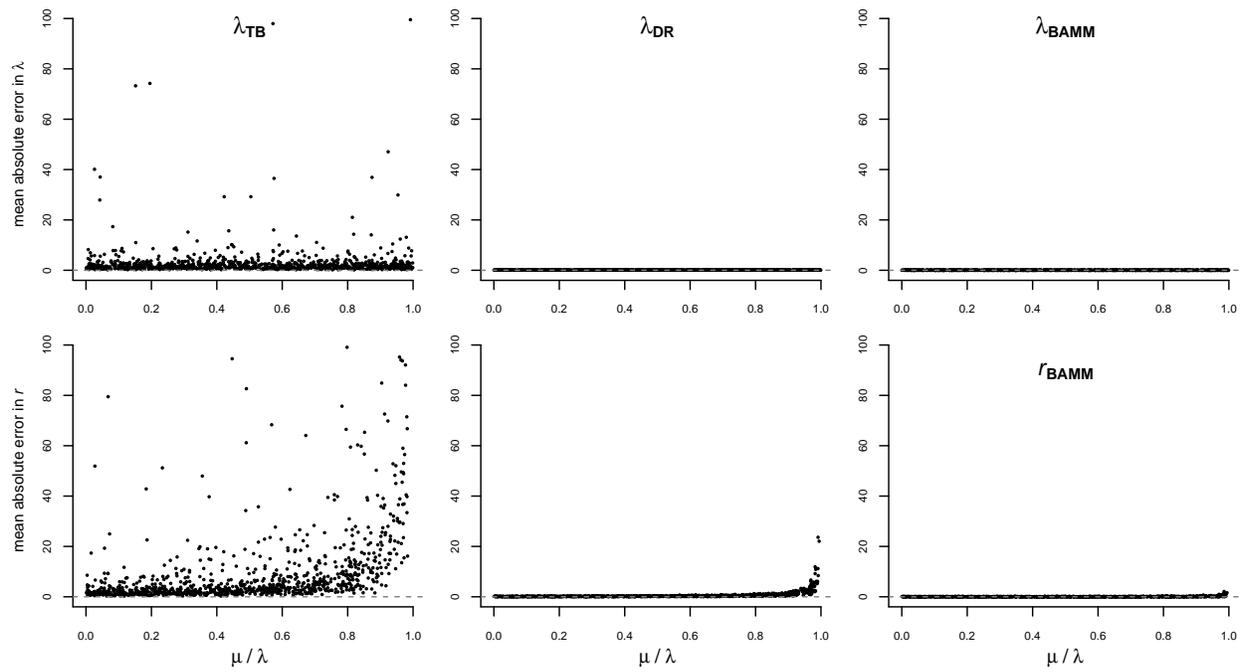
Figure S4.3: Mean absolute error in λ (top) and r (bottom) for λ_{TB}, with λ_{DR} and λ_{BAMM} on the same scale for comparison. For BAMM, the estimated net diversification rate is presented. λ_{TB} more accurately tracks λ than r, but the amount of error is orders of magnitude greater than for other metrics.
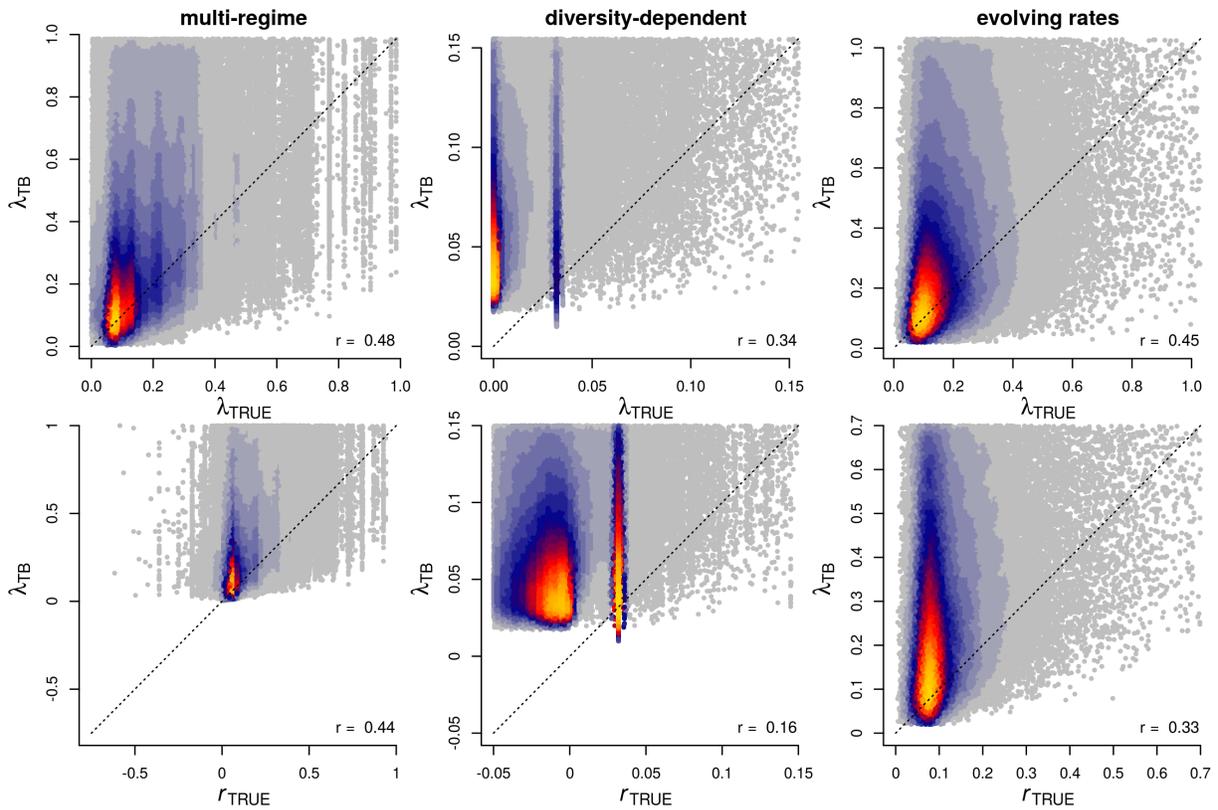
Figure S4.4: True tip rates (top row: λ_{TRUE}, bottom row: r_{TRUE}) in relation to λ_{TB}. Tip rates were compared separately for different major categories of phylogeny simulations (rows). Plotting region is restricted to the 99th percentile of true rates, but Spearman correlations between true and estimated rates (lower right of each figure panel) are based on the full range of the data. Colors indicate the density of points in the scatter plots. λ_{TB} is not particularly correlated with true tip rates.
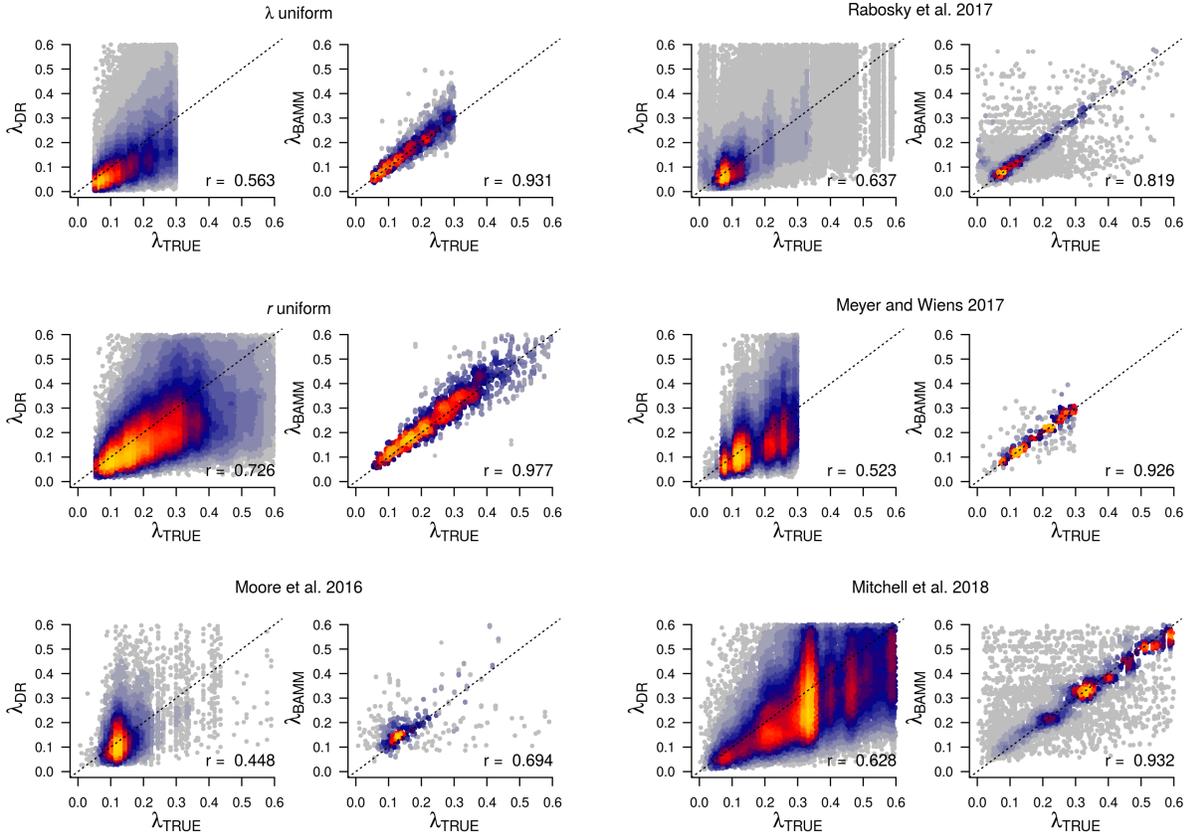
Figure S4.5: Comparison of λ_{DR} and λ_{BAMM} to true tip rates for separate simulation datasets. Data are separated by source, to confirm that patterns described in the main text are not driven by any one simulation study. Spearman's correlation is presented in the bottom right corner. Colors indicate the density of points in the scatter plots. Regardless of the dataset, λ_{BAMM} performs noticeably better than λ_{DR}.
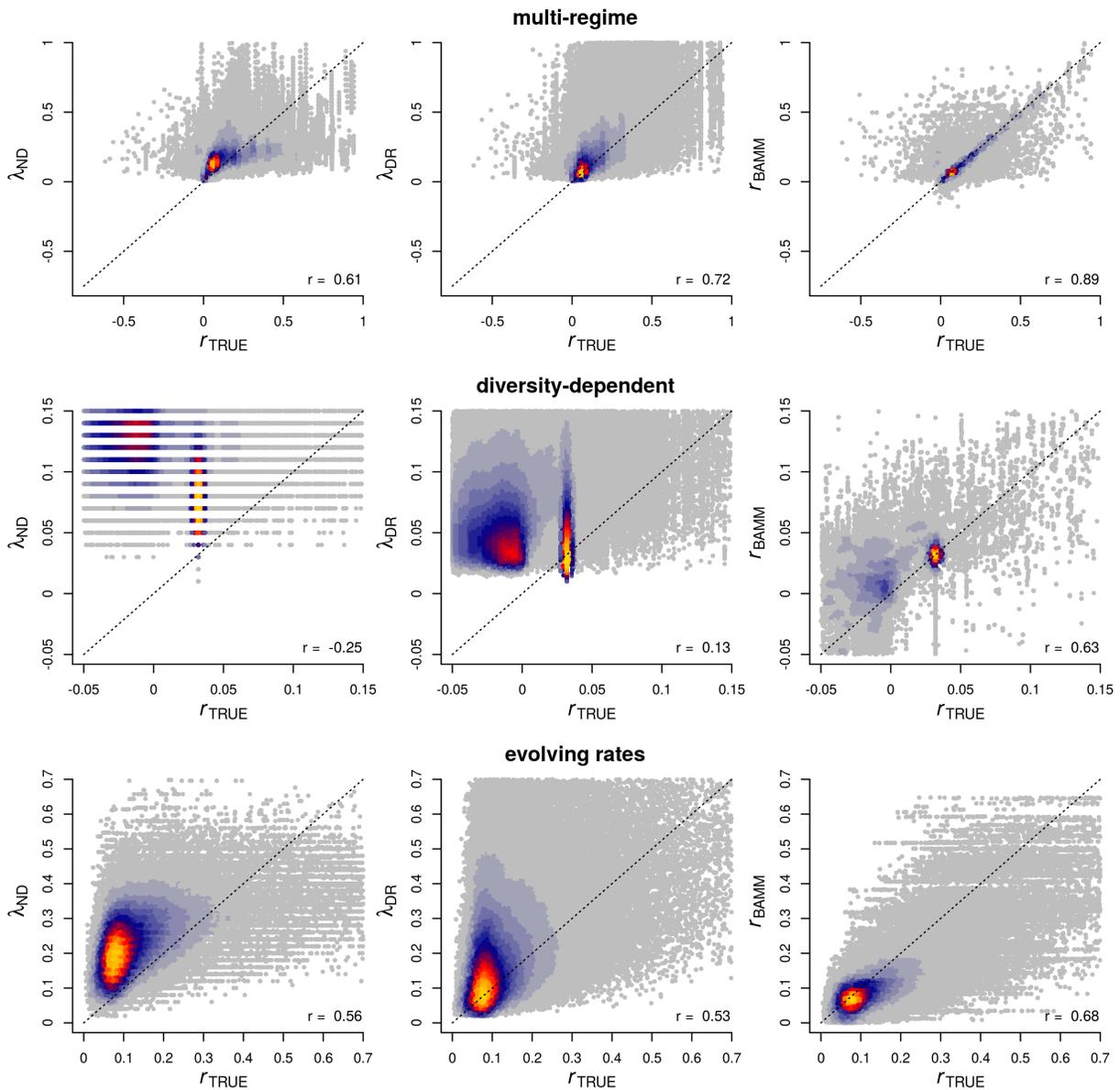
Figure S4.6: True net diversification tip rates (r_{TRUE}) in relation to estimated tip rates. Tip rates were compared separately for different major categories of phylogeny simulations (rows). Plotting region is restricted to the 99th percentile of true rates, but Spearman correlations between true and estimated rates (lower right of each figure panel) are based on the full range of the data. Colors indicate the density of points in the scatter plots. The horizontal gaps in λ_{ND} for diversity-dependent trees are an artefact of all trees having the same crown age. Relative performance comparison aside, correlations with r_{TRUE} are lower than with λ_{TRUE} (Figure 2).
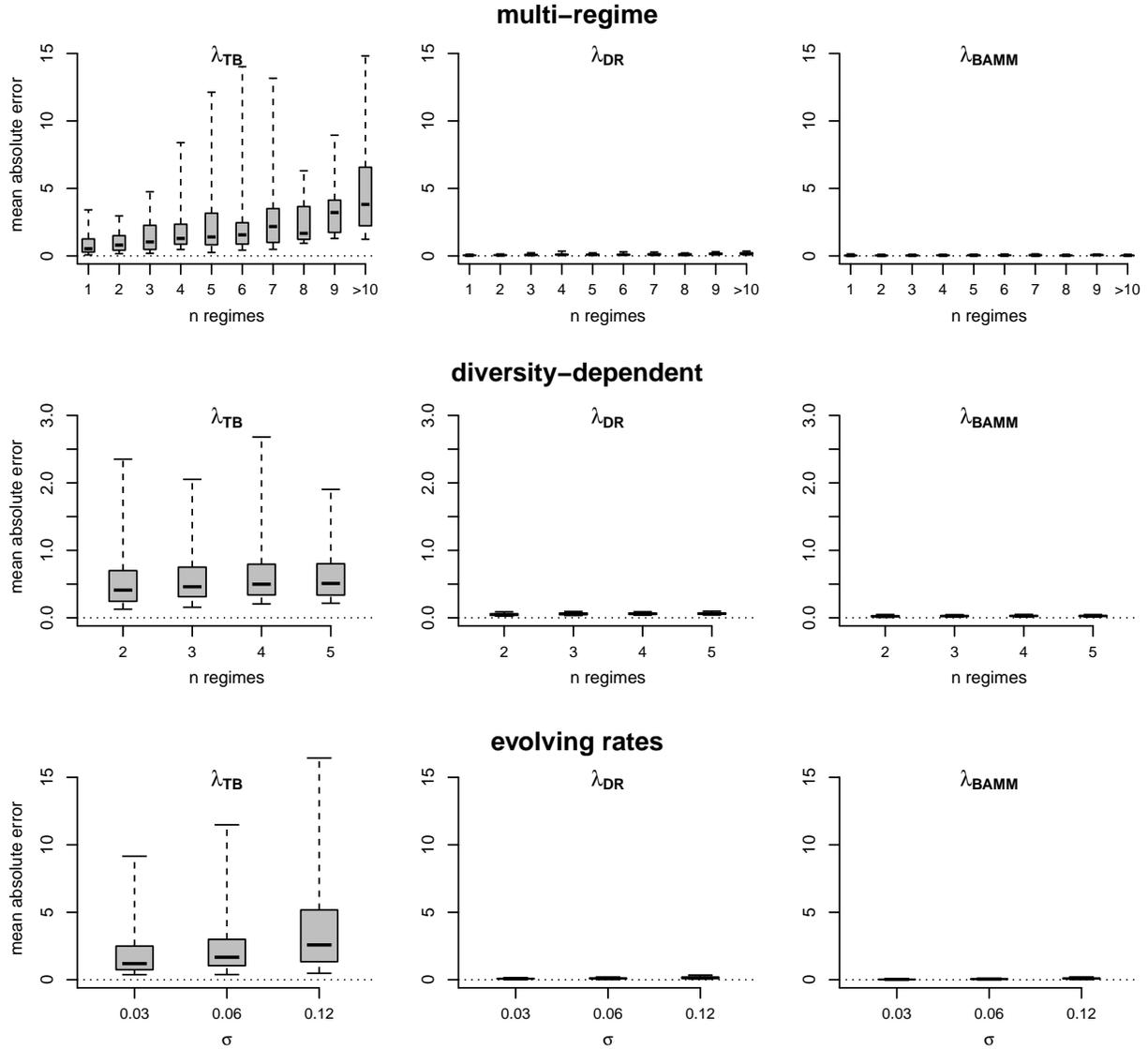
Figure S4.7: Mean per-tip absolute error in λ_{TB} as a function of the magnitude of rate heterogeneity in each simulated phylogeny. λ_{DR} and λ_{BAMM} are included on the same scale for comparison. Results are presented separately for different categories of rate variation (Table 1). The boxes and whiskers represent the $0.25 - 0.75$, and the $0.05 - 0.95$ quantile ranges, respectively. Error in λ_{TB} generally increases with increasing rate heterogeneity, and this error is almost two orders of magnitude greater than error in other tip rate metrics.
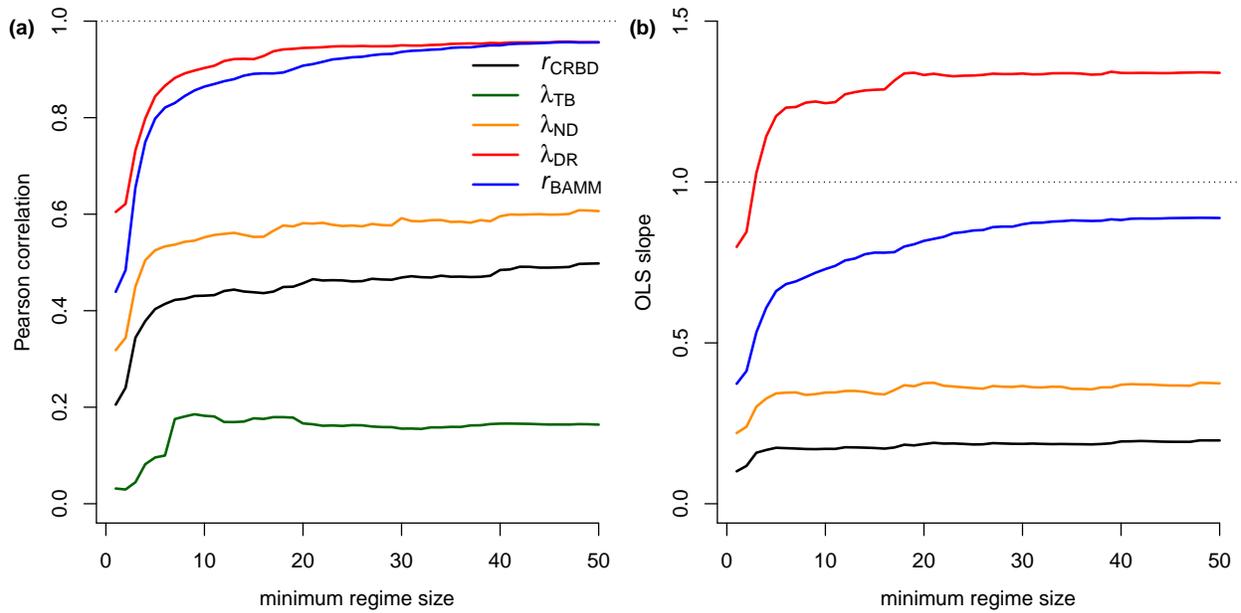
Figure S4.8: Performance of tip rate metrics as a function of regime size, including Pearson correlation **(a)** and OLS regression slope **(b)** for mean rates with respect to r_{TRUE}. λ_{DR} and r_{BAMM} outperform the other metrics when summarized in this fashion, although λ_{DR} overestimates the rate of net diversification (more so than it overestimated λ_{TRUE}, Figure 4). The x-axis denotes the minimum regime size across which performance was summarized. For example, x = 20 corresponds to the correlations and slopes computed for all regimes with 20 or more tips; a value of x = 1 is the corresponding results for all regimes. The OLS slope for λ_{TB} is not visible as it ranges between 10 and 25.

CHAPTER V

Dispersal and the latitudinal diversity gradient in marine fishes

5.1 Abstract

Marine fishes exhibit a striking latitudinal diversity gradient (LDG), with far more species occurring in the tropics than in extratropical regions. Speciation rates are substantially elevated in polar and high-latitude temperate regions relative to the tropics, suggesting that faster tropical speciation cannot explain the LDG in marine fish diversity. However, we do not yet understand the role of lineage dispersal over macroevolutionary timescales in generating and maintaining the LDG. The "out of the tropics" model posits that tropical taxa expand their ranges and disperse out, thereby enriching high latitude regions. However, extratropical diversification and environmental niche conservatism might be expected to influence the strength of the gradient as well as the movement of species from polar-temperate to tropical regions. We assessed latitudinal source-sink dynamics in marine fishes by estimating biogeographic transition rates and dispersals between tropical, temperate and polar regions while distinguishing between taxa with predominately shallow versus deep-water distributions. We find that biogeographic transition rates are greatest out of the Arctic and towards the tropics. Although rates are strongest in the opposing direction, the total number of dispersal events out of the tropics exceeds that of dispersal events out of the poles.

These results indicate that, even with relatively low dispersal rates, high species richness and 'tropical' inertia will drive macroevolutionary source-sink dynamics. We also find a strong pattern of greater movement of deep-water lineages than shallow-water lineages in either direction, suggesting that environmental conservatism and the depth distribution of biogeographic corridors likely play important roles in shaping global patterns of marine fish diversity.

5.2   Introduction

The latitudinal diversity gradient (LDG), whereby species richness is highest in the tropics and declines towards the poles, is one of the best-known yet poorly-understood biodiversity patterns observed today (Hillebrand 2004). Over macroevolutionary timescales, geographic patterns of species richness and endemism have been shaped by the interplay between variation in speciation and extinction rates, regional carrying capacities (MacArthur 1969, Mittelbach et al. 2007, Rabosky and Hurlbert 2015), and biogeographic dispersal (Ricklefs 2004, Wiens and Donoghue 2004, Goldberg et al. 2005, Fine 2015, Antonelli et al. 2018). Due to the spatial configuration and environmental characteristics of geographic regions, some will tend to exchange species more readily than others (Donoghue and Edwards 2014). Certain regions can thus behave as macroevolutionary sources, or centers of origination which then export species to other regions (Briggs 2003, Goldberg et al. 2005, Jablonski et al. 2006, Roy and Goldberg 2007). Likewise, other regions may be characterized as macroevolutionary sinks, where a significant proportion of the within-region diversity has originated elsewhere and arrived through dispersal.

Jablonski et al. (2006) described a "out of the tropics" model (OTT), where they proposed that species origination is greater in the tropics, and that, through tropical species range expansion, there is a net migration of species out of the tropics and into extratropical regions. This model assumes that 1) speciation rates are greater in the tropics, 2) extinction rates are no greater in the tropics than outside of the tropics, and 3) dispersal out of the

143

tropics is greater than dispersal into the tropics. An implication of this model is also that most extratropical lineages will have tropical ancestors. A related model – the "tropical conservatism hypothesis" (TCH; Wiens and Donoghue 2004) – suggests that, like the OTT model, there is greater speciation in, and greater dispersal out of the topics, but that dispersal is less frequent and limited to those lineages that are capable of adapting to novel environmental conditions (Smith et al. 2012, Kerkhoff et al. 2014).

The geography of marine fish richness is characterized by a strong LDG (Tittensor et al. 2010, Stuart-Smith et al. 2013, Rabosky et al. 2018). However, marine richness is not equally distributed near the equator, but tends to be concentrated in shallow tropical continental shelf environments (Briggs and Bowen 2013). The tropical Indo-Pacific in particular has the greatest species richness globally (Tittensor et al. 2010, Rabosky et al. 2018) and has been described as a center of origination for reef-associated fishes, where species diversity has both accumulated and been exported to neighboring regions (Briggs 2003, Alfaro et al. 2007, Cowman and Bellwood 2013, Siqueira et al. 2016). In comparison, the highly endemic fish fauna of the Antarctic has been characterized by origination at near off-shore islands and export to other regions of the Southern Ocean (Briggs 2003, Dornburg et al. 2017).

A number of hypotheses for the LDG predict faster speciation in the tropics (Rohde 1992, Allen et al. 2002, Jablonski et al. 2006, Mittelbach et al. 2007), yet Rabosky et al. (2018) found that rates of speciation for marine fishes exhibit an inverse latitudinal gradient. Significantly elevated rates at high latitudes indicate that, paradoxically, the regions with the fewest species are those characterized by the highest rates of speciation. The OTT and TCH hypotheses were framed around the assumption that rates of speciation are highest in the tropics. An additional core feature of these models is higher net movement of species from the tropics to higher latitudes and thus an appreciable fraction of extratropical diversity with tropical origins. Even if rates of speciation are not systematically higher in the tropics (Rabosky et al. 2018), the tropics might still be a dominant source of high-latitude species richness, thus helping to "flatten" what would otherwise be an even more severe LDG in

species richness.

Regardless of regional variation in diversification, a comprehensive understanding of the contributions of tropical conservatism and OTT dynamics to the LDG requires that we characterize the dynamics of inter-regional dispersal in the marine realm. It is thus important to characterize the extent to which tropical and extratropical regions serve as sources or sinks for species originating elsewhere. In this study, we quantify the magnitude and directionality of dispersal between major latitudinal zones to assess the roles of speciation rates and immigration in the shaping of the LDG. To do so, we characterize both the numbers of dispersals and regional transition rates, and the number of within-region speciation events.

We maintain a distinction between shallow and deep-water species composition in our analyses in order to account for the major environmental, ecological and biogeographic differences these marine regions entail. Species can be found in all regions of the ocean, from the surface to the abyssal depths, where the temperature, pressure, dissolved oxygen and light environment that those species experience is so dramatically different that it requires major physiological adaptations for survival (Portner 2002, Rogers 2015, Priede 2017). Ocean temperature is greatest at the surface, but drops precipitously at the thermocline and then exhibits relatively little variation throughout the remaining depth. Because of this depth stratification, the surface waters within the first 200 meters exhibit a strong latitudinal gradient in temperature, but this gradient is markedly weaker at depth. Deep-water ocean temperatures are therefore notably more homogeneous on a global scale (standard deviation of global marine temperature at 0, 200, and 500m: 11.3, 7.0 and 3.9°C, respectively; Figure S1; Boyer et al. 2013). The environmental differences between shallow and deep-water translate into strong ecological and physiological barriers for marine fishes (Brown and Thatje 2014, Priede 2017). As a result, colonization of deeper ocean waters is likely to be relatively rare, as evidenced by the fact that over 70 percent of extant marine fish diversity occurs within the first 500m from the surface (Priede and Froese 2013). As the deep-water marine environment is globally more environmentally homogeneous (Figure S1), it is thought that

species will experience fewer biogeographic barriers at depth; indeed, 82 percent of marine fish species with circumglobal distributions are bathypelagic or bathydemersal (Gaither et al. 2016).

If the OTT model is contributing to the LDG, then we expect to observe a high ratio of dispersals from low to high latitudes relative to within-region speciation events in temperate and polar regions, as well as greater transition rates from the tropics to those extratropical regions. Given the greater environmental homogeneity in deep water, we expect to maintain the same predictions for deep-water regions, only with greater magnitude.

5.3   Methods

5.3.1   Data acquisition

We obtained geographic range data for 12,018 out of approximately 15,500 known marine fish species (Mora et al. 2008) at a grid cell resolution of 150 x 150 km from Rabosky et al. (2018). The majority of these ranges were acquired from AquaMaps (Ready et al. 2010, Kaschner et al. 2016) in the form of vetted species distribution models, and were expanded upon by incorporating geographic range data from IUCN as well as from other literature sources (Coll et al. 2010, Mecklenburg et al. 2016, Rabosky et al. 2018). We also obtained a phylogeny of ray-finned fishes from Rabosky et al. (2018), containing 11,638 taxa with genetic data and 5,231 marine species (available at https://fishtreeoflife.org). We obtained depth classifications from FishBase (http://fishbase.org), where species were classified as shallow-water if the FishBase descriptor included pelagic, pelagic-neritic, pelagic-oceanic, reef-associated or demersal. Conversely, species were classified as deep-water if they were described as bathy-demersal, bathy-pelagic or bentho-pelagic. Taken together, we were able to combine geographic, phylogenetic and depth data for 4,987 marine fish species.

5.3.2 Geographic partitioning

We partitioned the globe into tropical, north temperate, south temperate, north polar and south polar regions. We opted to not use strict latitudinal thresholds, as major oceanic currents (e.g., the Gulf Stream) cause extensive regional variation in the latitudinal extent of both warm-water and cold-water; it is thus difficult to select any single latitudinal threshold that remains appropriate on a global scale. Additionally, as marine environmental conditions and fish taxonomic composition differ markedly between continental shelf regions and open ocean, we combined two marine regionalization datasets: Marine Ecoregions of the World (MEOW; Spalding et al. 2007) and Pelagic Provinces of the World (PPOW; Spalding et al. 2012). We manually modified pelagic province boundaries for the Leeuwin Current, Agulhas Current and the Non-gyral Southwest Pacific pelagic provinces by dividing them at 25 degrees latitude into tropical (northern) and temperate (southern) halves so that they more naturally align with the latitudinal ranges of the neighboring MEOW realms (Figure 1).

We rasterized these latitudinal zones to the same equal area grid as the species geographic range data, and classified each species in terms of which latitudinal zones it occurs in. A species was coded as occupying a region if its geographic range occupied 20 percent or more of the region grid cells, or if 50 percent or more of the species' range was found in that region. The latter criterion allowed us to account for small-ranged endemics. In the rare case where a species did not match either criterion, we assigned it to the region which overlapped with a minimum of ten percent of its range. This assignment was done for shallow and deep-water species separately, and a species could only be in shallow or deep regions, but never both, due to the FishBase depth categories.

5.3.3 Biogeographic transition rates

We developed a biogeographic transition model for tropical, temperate and polar regions (Figure 2) that closely follows the logic for the dispersal-extinction-cladogenesis (DEC)

model as developed by Ree and Smith (2008). We chose to model transitions between these states for the northern and southern hemispheres separately, in order to avoid producing a model with an unwieldy number of parameters. Specifically, we allow transitions between regions in single-step increments; i.e., a tropical species must first disperse to the temperate region, and then undergo local extinction (range contraction) in the tropics in order to become a purely temperate species, rather than entirely shift its range in one step. As discussed above, we maintain a distinction between shallow and deep-water species. Therefore, tropical-shallow and tropical-deep are separate states in our model. A species can only be either shallow-water or deep-water, but can undergo an evolutionary shift between those states. As species can occur in multiple geographic regions simultaneously, our complete list of states that species can occupy is: tropical-shallow, temperate-shallow, polar-shallow, tropical+temperate-shallow, temperate+polar-shallow, tropical+temperate+polar-shallow, and those same states for deep-water, totaling 12 states. We did not include a tropical+polar state, as this would have necessitated a discontinuity in the species' range, and no species in our dataset appear to show such a distribution.

To differentiate between transitions to and from particular regions, we parameterized the model such that transition rates included directional information to track "gain of the temperate state", "gain of the tropical state" and "gain of the polar state". These parameters were defined separately for shallow-water and deep-water transitions between geographic states. We included a range contraction parameter (local extinction), which defines the loss of a region, but constrained it to be identical for all regions, as Ree and Smith (2008) found that this parameter had relatively low accuracy. We also defined two pairs of vertical transition rates – shallow to deep, and deep to shallow – for tropical/temperate regions and for polar regions. We opted to define separate vertical transition rates for the polar regions, as the difference in temperature between shallow and deep water is minimized and therefore, we might expect transitions to be more frequent than in tropical or temperate waters, where the vertical temperature gradient is stronger.

We implemented this model as an asymmetric Markov k-state model (Mk; Lewis 2001) for discrete character evolution using the diversitree package (FitzJohn 2012) in R (R Core Team 2018). We supplied a 12x12 state matrix which defined parameter constraints (Table S1). We also considered multiple constrained submodels (Table 1) to test specific hypotheses about the directionality of inter-regional dispersal. In particular, we designed several models to specifically represent OTT scenarios, where transition rates from tropical to temperate and temperate to polar were constrained to be greater than transition rates from temperate to tropical and polar to temperate regions. By fitting these models, we were able to estimate per-lineage biogeographic transition rates between regions, and also assess support for particular scenarios through a statistical model selection framework. We fitted these models with a bounded Nelder-Mead optimization approach, as implemented in the dfoptim R package (Varadhan et al. 2018). In order to identify reasonable starting parameters, we first sampled 2000 initial values from a uniform distribution [0.001, 2], and calculated the likelihood for those parameters. We then selected the 100 sets of starting values that returned the highest log likelihood, and performed the optimization with lower and upper bounds of 1e-6 and 10, in order to find the maximum likelihood parameter estimates. By performing 100 optimizations per model, we could ensure with reasonable confidence that the optimization had in fact found the global optimum in parameter space.

A concern has been raised that maximum likelihood methods to estimate transition rates will tend to be biased by the frequency and distribution of states at the tips (Nosil and Mooers 2005), with a tendency for greater transition rates toward the more frequent state. Unlike state-dependent diversification models, such as BiSSE and GeoSSE (Maddison et al. 2007, Goldberg et al. 2011), the Mk models we employ do not incorporate speciation and extinction, which can lead to biased transition rate estimates (Goldberg and Igic 2008). Given that a majority of marine fish species occur in the tropics, it is possible that these issues might manifest themselves in our analyses. However, certain comparisons of rates should still be appropriate.

5.3.4 Ancestral state reconstruction

Although per-lineage transition rates between geographic regions are informative, they do not necessarily bear on the net exchange of species counts between those regions, as the actual number of events will depend on species richness patterns. It is entirely possible, for example, for the per-lineage rate of dispersal from temperate to tropical regions to be greater than tropical to temperate, but for the total number of species transition events to be greater out of the tropics, simply because there are more species in the tropics (i.e., the same realized dispersal of 5 species can hypothetically result from a low per-lineage dispersal rate of 0.001 from a 5000-species region, or from a high dispersal rate of 0.5 from a 10-species region). We therefore reconstructed biogeographic ancestral states under both parsimony and likelihood (ML) in order to enumerate speciation events within a region (no dispersal) and speciation coupled with dispersal events. Parsimony approaches have been shown to perform well, especially when there is heterogeneity in the rate of character evolution across the phylogeny (Tuffley and Steel 1997, King and Lee 2015, Davis Rabosky et al. 2016). We implemented a parsimony-based version of our biogeographic transition matrix through Sankoff parsimony (Sankoff 1975), which makes use of a cost matrix. This allowed us to define the number of steps required to move from one state to another. For example, a shift from tropical-shallow to temperate-shallow has a cost of 2: a new region must be gained by dispersal (temperate), and an ancestral region must be lost (tropical). We assumed that gain or loss of any single region (tropical, temperate, polar) entails unit cost. For these analyses, we reconstructed ancestral geographic states for northern and southern hemisphere regions within the same parsimony framework. To account for uncertainty in ancestral character states, we generated 1,000 equally parsimonious reconstructed histories. Parsimony analyses were run with the R package rbor (`https://www.github.com/blueraleigh/rbor`).

From the parsimony reconstructions, we visited each internal node and examined the reconstructed state of that node (the parent node) and of its two descendant nodes. We counted the number of within-region speciation events by tallying the number of times a

biogeographic state was shared between the node and both descendant nodes. We also counted the number of dispersal events, by tracking states present in a descendant node but not the parent node. If the parent node had multiple states that were not present in the descendant node, then we employed the cost matrix described above for parsimony in order to identify the more likely dispersal source. If more than one region was equally likely, the count was split (for instance, if a parent node was found in tropical+temperate and a descendant node had states tropical+temperate+polar, then the cost matrix would identify temperate-to-polar as more likely than tropical-to-polar, adding a tally to temperate $\rightarrow$ polar).

We generated 1000 joint ancestral reconstructions, under the best-fit Mk transition rate model, with the asr.joint function in diversitree (FitzJohn 2012). Similarly to the event counting approach under parsimony, we then visited all reconstructed internal node states and tracked dispersal and within-region speciation events. As the transition rate models were fit for each hemisphere separately, we counted events for each hemisphere separately as well. If multiple dispersal sources were possible, we again used the cost matrix.

We also calculated, for each region, the ratio of dispersal events to within-region speciation events, as counted through ancestral state reconstruction. This quantity allows us to more explicitly evaluate the influence of an OTT scenario, where dispersal events should be an important contributor to species assembly in the extratropics.

5.3.5 Sister pairs

As an additional check on relative counts of dispersal and within-region speciation, we also identified all sister species pairs in the phylogeny, and tallied the regions they occur in. This exercise has the benefit of not requiring reconstructed states at internal nodes. We would expect that the counts of sister pairs within the same region would follow the pattern we recover in within-region speciation event counts from ancestral state reconstruction. Additionally, although there is no directionality of dispersal in the sister pair data, we would expect the region pairs that share more sister species to also be the region pairs that

exchange species more frequently.

5.4 Results

The intersection of the geographic dataset for marine fishes, the phylogenetic sampling, and the depth classifications from FishBase resulted in a total of 4987 species, with 4475 occurring in the tropics and northern hemisphere, and 3916 species occurring in the tropics and southern hemisphere. Of the 4987 species, 3918 were classified as shallow-water and 1069 as deep-water, according to the FishBase depth categories. A majority of species were found to occur exclusively in tropical, temperate or polar regions, with 92.6 percent of species occupying a single geographic state, 3.9 percent occurring in two states (tropical+temperate or temperate+polar) and 3.3 percent occurring in all three states. Species richness as summarized by region reflected the general pattern observed in the full gridded dataset (Figure 1, Table 1), with large, successive drops in richness from the tropics to the poles for shallow-water, and a less pronounced drop in richness between tropics and temperate regions for deep-water. When considering the intersection of species with both geographic and phylogenic data, the deep-water north temperate region had slightly greater species richness than the deep-water tropical region. Mean speciation rates, computed as the average species-specific rate from a BAMM analysis performed in Rabosky et al. (2018), for those species in each geographic region, produced an inverse relationship with richness (Figure S2; correlation test of richness and mean speciation rates for both hemispheres combined: Spearman's ρ = -0.866).

5.4.1 Biogeographic modeling

We fit 10 biogeographic transition models to the northern and southern hemisphere subsets of the dataset. In both cases, we found overwhelming support for the most parameter-rich model (depicted in Figure 2, Table 2, Table S2), for which the transition rates were unconstrained in both direction and value, with the exception of vertical transitions for

tropical and temperate regions, which were constrained to have the same value. Notably, models that enforced poleward dispersal, and therefore emulated the OTT model, were rejected with significantly lower model fit than the best-fit model.

Dispersal rates from poles-to-tropics were each greater than their corresponding tropics-to-poles transition rates except for the rate from the shallow southern polar to temperate region, which was zero (Figure 3, Figure S3, Table 3). Dispersal rates out of the shallow-water Arctic were estimated to be 20 times faster than dispersal into the Arctic region. For deep-water, dispersal out of the Arctic was over 100 times greater than in the opposite direction. The bias towards poleward dispersal was not quite as extreme in the southern hemisphere, but rates out of the deep-water Southern Ocean were still several times greater than rates in the opposite direction.

Statistical model selection also favored the best-fit model over a model where vertical transition rates from shallow to deep were constrained to be identical to rates from deep to shallow (Table 2). Parameter estimates indicate that dispersals at high latitudes are more likely from shallow to deep (Figure 3, Table 3), but the opposite in tropical and temperate waters. The greatest vertical transition rates were the polar shallow to deep rates, which were many times the equivalent tropical/temperate rates.

5.4.2 Ancestral state reconstructions

We generated ancestral state reconstructions using Sankoff parsimony and maximum likelihood. By generating 1000 ancestral reconstructions, and tracking the frequency with which each state transitioned to another, we were able to summarize dispersal event and within-region speciation event counts, while accounting for uncertainty in identity and placement of the state transitions.

With parsimony, we recovered dispersal event counts that were consistently biased towards movement away from the tropics, for both shallow and deep-water species (Figure 3, S4). Dispersal counts were greatest for both shallow and deep-water dispersals from the

tropics to both northern and southern temperate regions. With the maximum likelihood reconstructions (Figure S5), there wasn't as much of a clear pattern in direction or magnitude of counts. Generally speaking, dispersals occurred more frequently in deep-water, except into and out of the north polar region. Parsimony counts for vertical dispersals were more numerous from shallow to deep waters for all regions, with the exception of the southern temperate region, where the opposite was found. Vertical transition counts were not as consistent with the ML-based counts (Figure S5).

The number of within-region speciation events from either reconstruction approach was highly correlated with species richness, with the greatest number in the tropics, and decreasing towards the poles (Table 4). In deep-water, the number of speciation events within temperate regions was much more similar to the tropics, with slightly greater counts for the north temperate, which reflects our species counts (Table 1).

Ratios of dispersals to within-region speciation events led to significant differences between regions. From our parsimony analysis, we found that all regions were more influenced by OTT dispersals than by dispersals in the other direction (Figure 4), in both shallow and deep water. The Arctic is heavily dominated by dispersal out of the north temperate. In contrast, dispersal plays a much more minor role for the Southern Ocean, especially in shallow water. Ratios based on the ML reconstructions were very different (Figure S6), with the pattern in deep-water essentially opposite from what was found with parsimony.

5.4.3 Sister pairs

We identified 1414 sister pairs. Of these pairs, 75 percent were found to occur within the same geographic region, and 25 percent were found to occur in non-overlapping geographic regions. The shallow tropics and shallow temperate regions contained the greatest number of same-region sister pairs (Table 4). Vicariant sister pairs were most frequently found in the shallow-tropics / shallow-temperate and shallow-tropics / deep-tropics (Table S3). We also normalized the sister species counts by region, by dividing the number of within-region

154

sister species by the number of sister species pairs that were in that region and any other (Table 4). The shallow-water tropics still had more within-region sister pairs than pairs that were not both within the tropics. However, the deep-water north temperate had the most within-region sister pairs, relative to the number of pairs shared with other regions.

5.5   Discussion

The biogeographic flux of marine fishes at a global scale is characterized by a pattern of greater rates of dispersal from the poles to the tropics. However, parsimony-derived counts of realized dispersal events depict the opposite pattern, with the net movement of species out from the tropics, towards the poles. This pattern was evident for both shallow-water and deep-water species. Model selection and parameter estimates strongly favored the poles-to-tropics directionality of rates over OTT models, as well as over other models that constrained various rates to be identical (Table 2). Caution is warranted in interpreting the importance of directionality in transition rates however, as we generally find greater transition rates towards the more frequent state (the tropics).

The most striking pattern to emerge from our transition rate estimates was the magnitude of the asymmetry in immigration and emigration rates into and out of the polar regions (Figure 3, Figure S3), for both the shallow and deep ocean, and in particular the Arctic.

The high latitudes are currently harsh environments dominated by ice and high seasonality; however, the northern and southern polar regions have been shaped by independent geologic and climatic histories, leaving indelible marks on the current makeup of the regional fish faunas. When the Bering seaway opened up 3.0-3.5 mya, the Arctic was ice-free and characterized by a temperate climate, thus the seaway provided a habitable dispersal route between the northern Pacific and northern Atlantic Ocean basins (DeVries and Steffensen 2005, Mecklenburg et al. 2011). The region was then dominated by a number of young and phylogenetically disparate families (such as the zoarcid eelpouts). As the Arctic cooled, species either adapted to these colder conditions, shifted their distributions to the northern

155

temperate regions, shifted to warmer, deeper waters, or shifted their habitat preferences to the mouths of rivers (Mecklenburg et al. 2011). In contrast, the Southern Ocean has likely maintained its current state of freezing water temperatures and continental glaciers for approximately the last 25 mya (Kennett 1982). Whereas Arctic marine environments are connected to temperate regions via shallow continental shelves, the Antarctic continental shelf and the Southern Ocean are separated from the southern temperate region by deep waters and the Antarctic Circumpolar Current. The fish fauna in the Antarctic is highly endemic, with the majority of species belonging to the notothens (icefishes), as well as representatives from the Liparidae (snailfishes) and Zoarcidae (eelpouts). Physiological studies have found that a number of notothen species have extremely narrow temperature and salinity tolerance (Somero and DeVries 1967, O'Grady and DeVries 1982), as they have evolved in isolation in an environmentally stable region for tens of millions of years. In contrast, the Arctic fauna is characterized by broader environmental tolerances and lower endemicity (DeVries and Steffensen 2005).

The differences in dispersal rates out of the high latitudes and into the temperate regions are consistent with these contrasting biogeographic histories. The Southern Ocean's highly endemic, environmentally specialized fauna does not have a history of dispersing out to warmer waters, except in a few rare cases (Cheng et al. 2003, Eastman 2005). In contrast, the cooling of the Arctic region has led to many dispersals out to the north temperate Pacific and Atlantic (Mecklenburg et al. 2011, Briggs and Bowen 2012). The contrast in migration rates from the polar to temperate regions in the northern and southern hemispheres might also reflect the greater temperature seasonality of the Arctic, compared to the relative stability of the Antarctic (Clarke and Crame 2010). Arctic species' broad thermal tolerance ranges may be related to wide annual range in temperature in the region, predisposing them to the warmer temperatures of the north temperate waters and imparting greater dispersal capability. In contrast, Antarctic species, and in particular the perciform notothens, have developed highly specialized adaptations (such as antifreeze mechanisms; Portner 2002,

Eastman 1993, Near et al. 2012) to inhabit an extreme, narrow range of thermal conditions. Such adaptations may have lessened these species' ability to disperse across the Antarctic Circumpolar Current into warmer southern temperate waters. We see further evidence of these hemisphere differences in the ratios of dispersal to speciation events in the polar regions (Figure 4). The Arctic has had a dynamic history of acting as a region of species exchange, whereas the Southern Ocean has been a stable region with relatively low immigration and emigration, and the host of *in situ* diversification. These parsimony-based ratios show greater influence of dispersal in the Arctic and greater influence of within-region speciation in the Southern Ocean.

There are a number of reasons why the transition rates estimated in this study may provide an inadequate description of the tempo and directionality of inter-regional dispersal. Major sources of bias include state-dependent diversification, non-equilibrium dynamics of trait evolution, and heterotachy. By employing Mk models to estimate transition rates, we do not account for the possibility of regional (e.g., "state-dependent") differences in diversification rates that might bias both transition rates and ancestral area reconstructions (Maddison et al. 2007, Goldberg and Igic 2008, Goldberg et al. 2011). For instance, if species in the tropical state have a greater net diversification rate than species outside of the tropics, then we would potentially overestimate the transition rate toward the tropical state, as that state is rising in frequency due to increased diversification rates, not dispersals. Likewise, if high extinction rates prevent the accumulation of diversity at high latitudes, we might find low transition rates towards the high latitudes.

Additionally, the distribution of character states might not be at equilibrium. Transition rate models, as well as trait-dependent diversification models, generally assume that the distribution of character states is the product of long-term rates of character evolution, speciation and extinction rates (Goldberg and Igic 2008, O'Meara et al. 2016). If species richness is still accumulating in the high latitudes, then marine fish diversity is not at equilibrium, and estimated transition rates can be biased as a result.

Transition rate models also do not explicitly account for heterogeneity in rates of character evolution (King and Lee 2015). However, as analyses are conducted on trees of increasing size, the likelihood that rates do vary across the tree increases as well. Across a phylogeny of global fish diversity that spans almost 200 mya, it is more likely than not that rates of dispersal have not been constant, and that heterotachy is present.

Given these potential issues, we hesitate to attach too much importance to the poles-to-tropics directionality of rates in our analyses, and some of the unexpected patterns in ML ancestral reconstructions (Figure S6) might be related to these issues. However, certain contrasts in rates probably do represent biological results. The strong contrast in rates of exchange between the Arctic and Southern Ocean are consistent with the climatic history of these regions, as discussed above. Furthermore, the differences in rates between shallow and deep-water regions are also not likely to be an artefact.

We recovered a very clear pattern of greater dispersal rates for deep-water species, relative to shallow-water species (Figure 3, S4). Every inter-regional deep-water rate was greater than its shallow-water counterpart, save for the dispersal rate from the north temperate to polar region. This is consistent with the idea that with greater depth, marine environmental conditions are more homogeneous (Figure S1; Gaither et al. 2016, Priede 2017). Features such as ocean fronts, that act as significant barriers to dispersal in surface waters, are more permeable at depth (as was found, for example, with the North Atlantic Subpolar Front, Vecchione et al. 2015), and the overall latitudinal gradient in temperature is less extreme at depth. This is supported by the finding that a majority of species with circumglobal distributions are deep-water species (Gaither et al. 2016). Additionally, there have been multiple instances of interhemispheric dispersal events of deep-water lineages, particularly in the families Scorpaenidae, Liparidae and Zoarcidae, dispersing from the North Pacific to the south temperate and Southern Ocean, through the tropics via cold deep waters ("isothermic submersion", Briggs 2003, Mecklenburg et al. 2011). Furthermore, in the Southern Ocean, the composition of species in the shallow-waters is dominated by the endemic notothen

icefishes. In deep-water, however, representatives of other families are a larger percentage of the diversity, thus reducing Arctic endemism at depth.

Rates of dispersal from shallow to deep water were much greater in the high latitudes than all other vertical dispersal rates. In the Arctic, this is consistent with adaptation to deeper, ice-free waters documented for a number of lineages in response to regional cooling over the last several million years (Mecklenburg et al. 2011). We might also expect rates of dispersal to be greater in the high latitudes simply because the sea temperature of shallow and deep water is more similar in these regions than anywhere else on the planet. Evidence of this can be found in the average depth ranges of tropical and polar species, where polar species generally have broader depth ranges, according to FishBase depth data (median tropical and polar depth range: 50m and 700m, across 8291 and 282 species, respectively).

Taken together, we find that rates of dispersal seem to reflect historical and current environmental similarity across latitudinal regions. Rates are low between the tropics and the high latitudes, where the environmental conditions are exceedingly harsh and colonization necessitates specialized adaptations. In contrast, rates between the high latitudes and the temperate regions are drastically higher for the Arctic. We see lower immigration and emigration rates out of the shallow Southern Ocean, which is inhabited primarily by species that have evolved narrow environmental niches. Perhaps most convincing is that dispersal rates are substantially higher for deep-water species, which experience greater environmental homogeneity than at the surface. Additionally, rates from shallow to deep-water in the high latitudes, which are most similar in terms of temperature, are the highest vertical dispersal rates.

Counts of dispersal events via parsimony analysis capture a pattern that is in stark contrast to the pattern in transition rates across latitudinal zones (Figure 3). These event counts are strongly correlated with species richness of the dispersal source regions (Spearman $r = 0.833$). If we take speciation as an example, the tropics have significantly lower rates of speciation than the higher latitudes (Figure 1, Rabosky et al. 2018); however, we count a

much greater number of speciation events in the tropical region (Figure S4). If a lineage has on average a relatively low probability of undergoing speciation, but there are many such lineages, then the realized outcome will be many speciation events. Likewise, if there is a low probability that a species would disperse out of the tropics, but there are thousands of tropical taxa, the realized outcome will be more dispersal events than in the opposite direction. Therefore, we may find a strong pattern of rates of dispersal biased in a "poles to tropics" direction, but tropical inertia due to high species richness results in a net realized dispersal pattern out of the tropics. In their original description of the OTT model, Jablonski et al. (2006) considered counts of range expansion out of the tropics, and commented on the events that allowed for extratropical expansion likely being infrequent (Jablonski et al. 2013). Our results are thus in agreement with the dispersal aspect of the OTT model.

Our results add to a growing number of studies that have found support for niche conservatism as a key concept in the generation and maintenance of the latitudinal diversity gradient. In similar studies that explored the LDG in different groups of organisms, dispersal rates as inferred with GeoSSE (Goldberg et al. 2011) were found to be greater from extratropical to tropical regions in amphibians (Pyron and Wiens 2013), squamates (Pyron 2014), and in the *Pheidole* genus of ants (Economo et al. 2018). Although rates of dispersal were not explicitly estimated, a significant role for niche conservatism in the LDG was found for birds (Duchene and Cardillo 2015) and for new world woody angiosperms (Kerkhoff et al. 2014). Additionally, Jablonski et al. (2013) found in marine bivalves that species with broad latitudinal ranges appeared to be tracking regions of similar temperature, providing another indication that thermal niche conservatism may be an important mechanism in the shaping of the LDG. No other study has examined the role of dispersal in the generation and maintenance of the LDG for marine fishes as a whole. Siqueira et al. (2016) focused on four coral reef-associated fish families, and found that speciation rates were greater in the tropics, while dispersal rates were greater out of the tropics. Cowman and Bellwood (2013) characterized the Indo-Australian Archipelago as a region of diversity accumulation

and a subsequent source of species movement into neighboring regions; however, their analyses focused only on three reef-associated families and longitudinal lineage exchange between major ocean basins. These studies do not contradict our findings, but rather differ from our analyses in both taxonomic and geographic scope.

5.6   Conclusion

The tropics contain the greatest number of marine fish species, and over the evolutionary history of the group, more species have dispersed out of the tropics, into the temperate, and ultimately into the polar regions. However, biogeographic rates of dispersal are greatest in the opposite direction: from the poles towards the tropics. The frequency and direction of dispersal events reflect the realized movement of lineages, resulting from species richness inertia despite the directionality of the biogeographic transition rates. But high rates of dispersal associated with the Arctic, indicate that the assembly of regional communities in the northern high latitude regions is dominated by dispersal. In contrast, regional assembly in the Southern Ocean has been dominated by *in situ* diversification. This is likely due to the contrast in biogeographic and climatic histories of these regions.

Our results also highlight the importance of environmental niche conservatism in shaping global patterns of diversity, and in determining the most likely dispersal routes. We found that, the more environmentally similar regions are, the greater the rate of transition between them. This was particularly clear in the distinction between shallow and deep-water transition rates. With recent anthropogenic climate change, the Arctic is warming faster than the global average (Hoegh-Guldberg and Bruno 2010). With this shift, there are already documented cases of northern temperate species dispersing north into regions that were once outside their tolerance ranges, and this is expected to increase dramatically over the next century (Fossheim et al. 2015, Wisz et al. 2015). The biogeographic response to climate change points to the sensitivity of marine species to sea temperature as determining geographic range limits and provides an example of the global biogeographic repercussions

that are likely to accompany a rapidly changing climate.

5.6.1 Acknowledgements
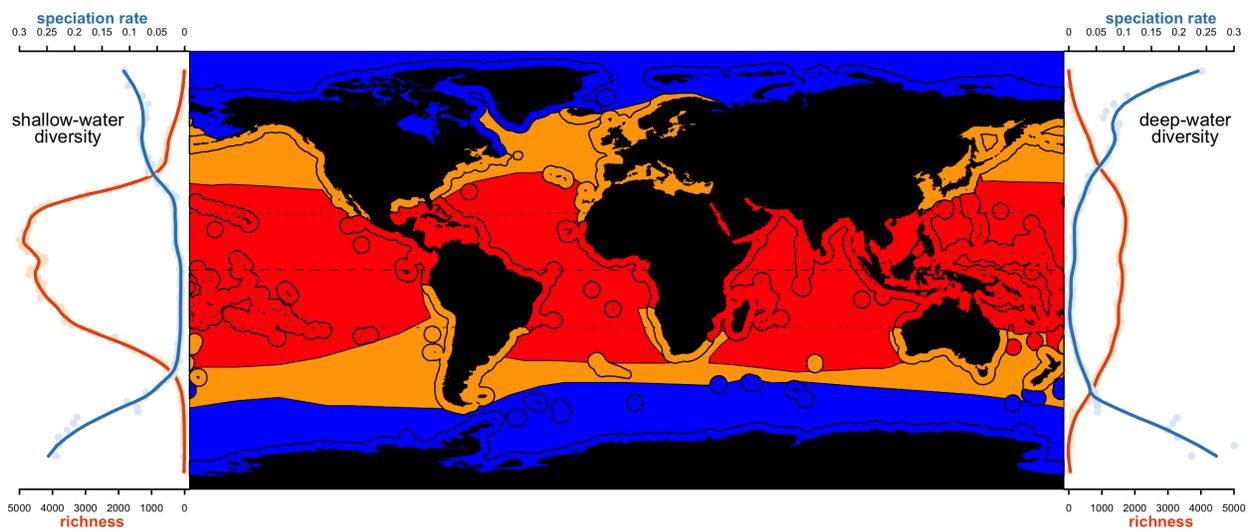
Figure 5.1: Latitudinal regions and latitudinal diversity gradients. Tropical (red), temperate (orange) and polar (blue) regions, as delineated from the MEOW (Spalding et al. 2007) and PPOW (Spaling et al. 2012) datasets. Polygons visible within the colored regions show the MEOW marine realms that separate marine shelf from open ocean regions. Latitudinal gradients for species richness (dark orange) and speciation rates (light blue) are shown for shallow-water species (left) and for deep-water species (right).
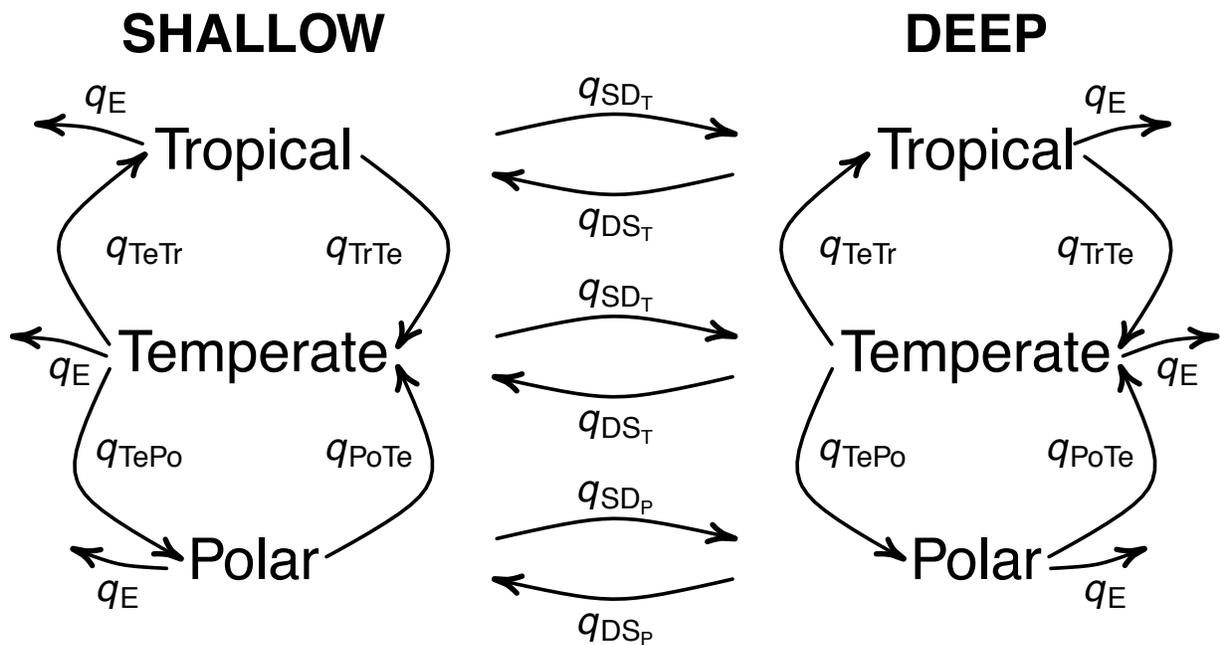
Figure 5.2: Conceptual diagram of the biogeographic transition model that had the best model fit. Transition rate parameters are labeled, and rates that are constrained to the same value have the same label. This model was fit for the northern and southern hemispheres separately. Tr = tropical, Te = temperate, Po = polar, SD = shallow $\to$ deep, DS = deep $\to$ shallow, E = local extinction.
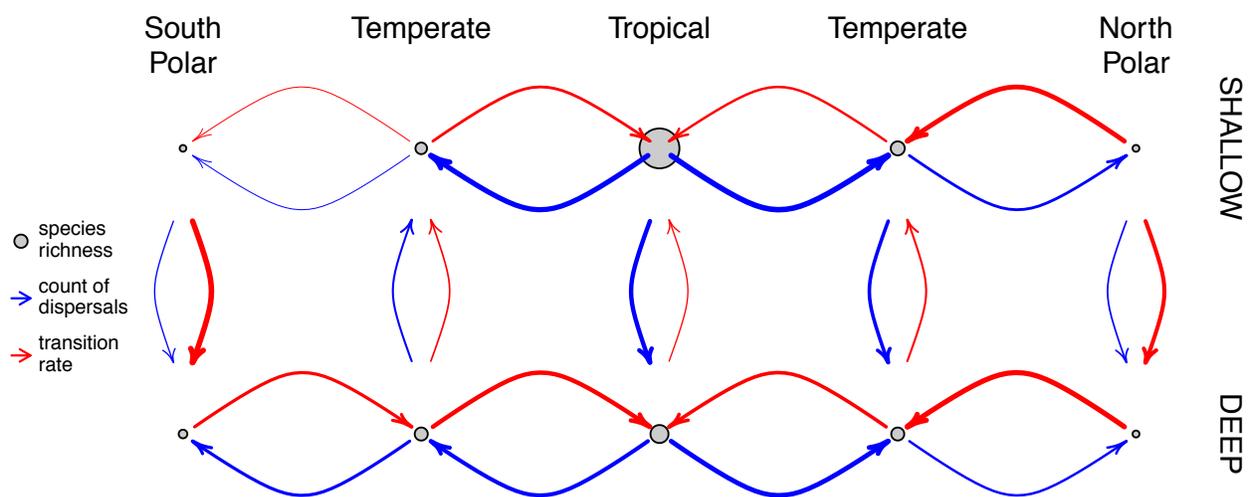
Figure 5.3: Rate and count differentials across latitudes and depth. Each arrow represents the difference in the transition rates or parsimony counts in one direction versus the other. The arrows have been scaled to reflect the magnitude of the differences, following a quantile transformation. Actual rate and count values can be found in Figure S3 and S4. The circles representing each region have been scaled according to the region's species richness. Transition rates dominate in a poles-to-tropics direction, whereas net movement based on dispersal events is in a tropics-to-poles direction.
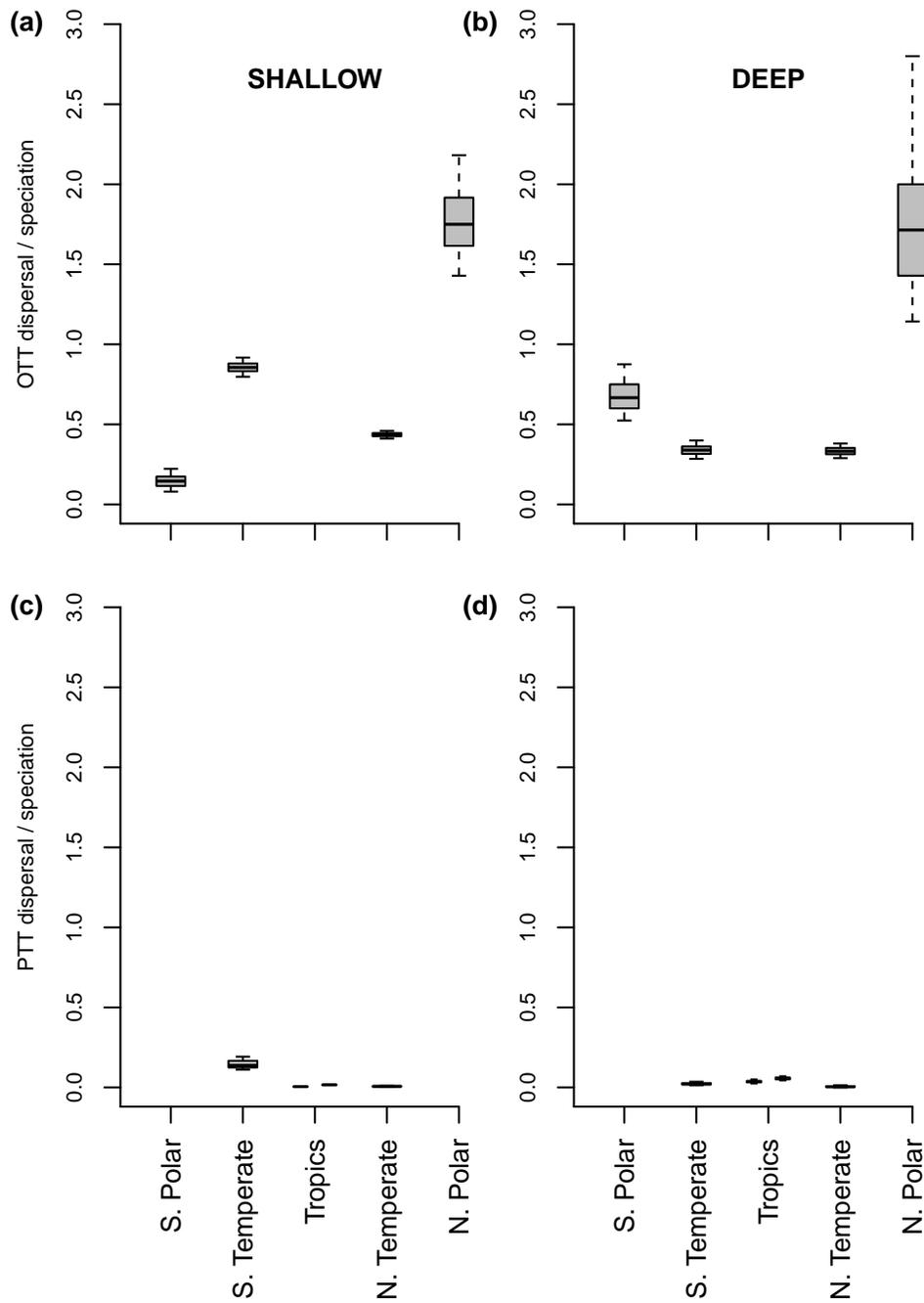
Figure 5.4: Ratios of dispersal to within-region speciation, based on parsimony. In **(a)** and **(b)**, out-of-the-Tropics (OTT) dispersals were used, therefore there is no ratio for the Tropics. In **(c)** and **(d)**, poles-to-Tropics (PTT) dispersals were used, hence there are no ratios for the polar regions, and there is both a south Temperate → Tropics and a north Temperate → Tropics ratio. Boxplots represent the distribution of ratios from 1000 parsimony-based ancestral state reconstructions, in terms of their 5-95 and 25-75% quantiles. The Arctic exhibits a signature of dispersal-dominated assembly, whereas in the Southern Ocean, within-region speciation is the dominant process of lineage accumulation.

hemisphere	region	depth	full richness	subset richness	mean speciation rate
northern	tropical	shallow	6167	2796	0.09
northern	temperate	shallow	1412	968	0.14
northern	polar	shallow	96	78	0.14
northern	tropical	deep	2149	579	0.09
northern	temperate	deep	1230	591	0.13
northern	polar	deep	108	75	0.17
southern	tropical	shallow	6167	2796	0.09
southern	temperate	shallow	971	452	0.10
southern	polar	shallow	67	52	0.20
southern	tropical	deep	2149	579	0.09
southern	temperate	deep	1242	458	0.09
southern	polar	deep	430	185	0.17

Table 5.1: Richness and speciation rate values for the different latitudinal regions. Richness was calculated as the number of species whose range is in each region (but not exclusively). Counts from our geographic dataset are shown as 'full richness', and as 'subset richness' when intersected with the taxa in the phylogeny.

| hemisphere	model	K	logLik	AICc	ΔAICc	wtAICc
north	**unconstrained**	**13**	**-3898.55**	**7823.18**	**0.00**	**1.00**
north	all shallow transitions equal	10	-3915.05	7850.15	26.96	0.00
north	OTT model enforced for shallow	13	-3915.05	7856.18	32.99	0.00
north	deep → shallow = shallow → deep for tropical/temperate only	12	-3916.63	7857.33	34.15	0.00
north	deep → shallow = shallow → deep	11	-3917.75	7857.55	34.37	0.00
north	OTT model enforced for deep	13	-3921.88	7869.83	46.65	0.00
north	shallow polar import = export, deep polar import = export	11	-3952.11	7926.27	103.09	0.00
north	OTT model enforced for shallow and deep	13	-3964.81	7955.70	132.52	0.00
north	all deep transitions equal	10	-4056.26	8132.57	309.39	0.00
north	shallow transitions equal, deep transitions equal	7	-4102.63	8219.29	396.11	0.00
south	**unconstrained**	**13**	**-2993.12**	**6012.31**	**0.00**	**1.00**
south	deep → shallow = shallow → deep for tropical/temperate only	12	-2999.86	6023.77	11.47	0.00
south	deep → shallow = shallow → deep	11	-3007.91	6037.87	25.56	0.00
south	shallow polar import = export, deep polar import = export	11	-3014.66	6051.37	39.06	0.00
south	all shallow transitions equal	10	-3016.47	6052.98	40.67	0.00
south	OTT model enforced for shallow	13	-3013.72	6053.51	41.20	0.00
south	OTT model enforced for deep	13	-3025.80	6077.68	65.37	0.00
south	OTT model enforced for shallow and deep	13	-3058.92	6143.92	131.61	0.00
south	all deep transitions equal	10	-3102.04	6224.12	211.81	0.00
south	shallow transitions equal, deep transitions equal	7	-3117.34	6248.69	236.38	0.00

Table 5.2: Model fit comparison for the northern and southern hemispheres. We found strong support in each hemisphere for the unconstrained model.

hemi-sphere	model	shallow				deep				vertical				range contraction
		OTT	OTT	PTT	PTT	OTT	OTT	PTT	PTT					
		$q_{Tr_S \to Te_S}$	$q_{Te_S \to Po_S}$	$q_{Te_S \to Tr_S}$	$q_{Po_S \to Te_S}$	$q_{Tr_D \to Te_D}$	$q_{Te_D \to Po_D}$	$q_{Te_D \to Tr_D}$	$q_{Po_D \to Te_D}$	q_{SD_T}	q_{SD_P}	q_{DS_T}	q_{DS_P}	q_E
northern	uncon-strained	0.007	0.0071	0.0158	0.1555	0.1243	0.0041	0.181	0.4845	0.0024	0.076	0.006	0	0.287
southern	uncon-strained	0.0024	0.0006	0.0403	0	0.0771	0.008	0.1818	0.0631	0.0024	10*	0.0048	4.3476	0.13

Table 5.3: Transition rate parameters from the best fit, unconstrained model. All poles-to-tropics (PTT) rates are greater than tropics-to-poles (OTT) rates, save one. *This parameter reached the upper bound during optimization.

hemisphere	region	depth	within-region speciation events		within-region sister species	normalized sister pairs
			parsimony	ML		
	tropical	shallow	2701.49	2580.73	702	0.60
	tropical	deep	403.4	272.35	93	0.20
northern	temperate	shallow	543.72	547.14	133	0.25
northern	polar	shallow	12.53	25.9	3	0.07
northern	temperate	deep	292.21	197.21	82	0.21
northern	polar	deep	6.47	11.02	3	0.08
southern	temperate	shallow	159.83	264.91	53	0.20
southern	polar	shallow	25.15	4.04	5	0.15
southern	temperate	deep	223.67	146.1	46	0.15
southern	polar	deep	40.44	32.93	11	0.12

Table 5.4: Counts of within-region speciation events and sister pairs. We also calculated normalized sister pair counts, which represent the percent of within-region pairs, compared to the number of pairs pairs that include that region. There is a latitudinal gradient in all of these quantities.
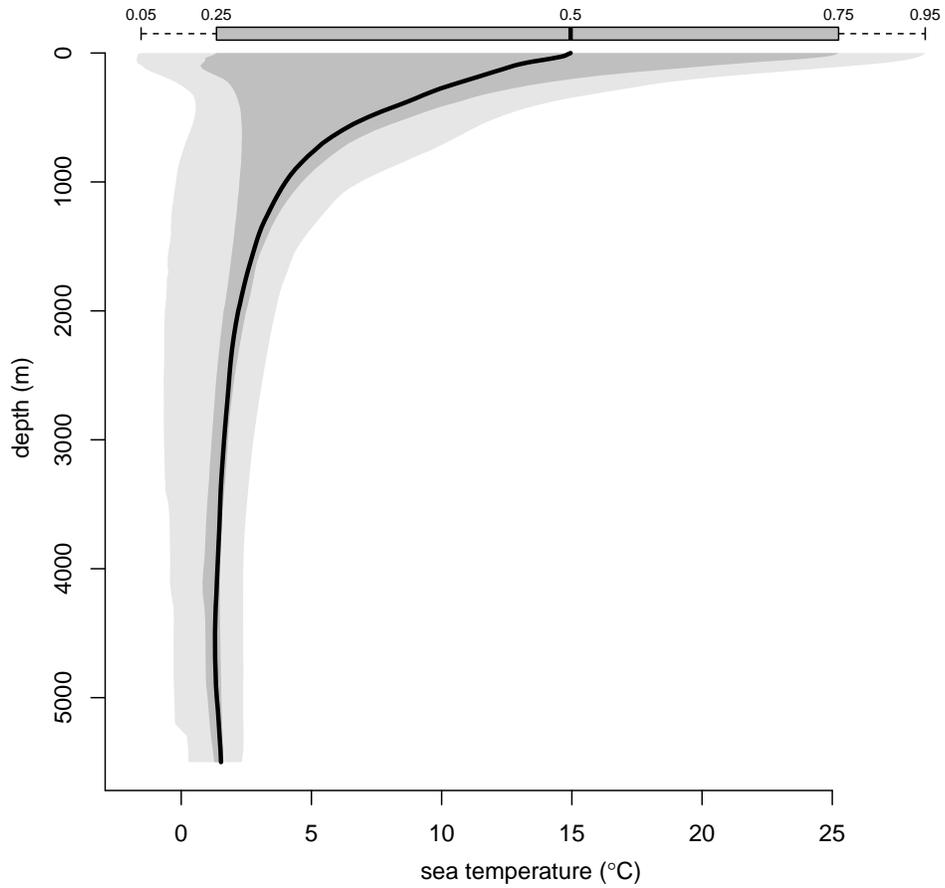
Figure S5.1: Global variation in sea temperature, as a function of depth. The light gray and dark gray polygons delineate the 5 to 95%, and 25 to 75% quantile ranges in temperature, respectively, and the black line represents the median. Temperature information from the NOAA World Ocean Atlas 2013 v.2 (Boyer et al. 2013).
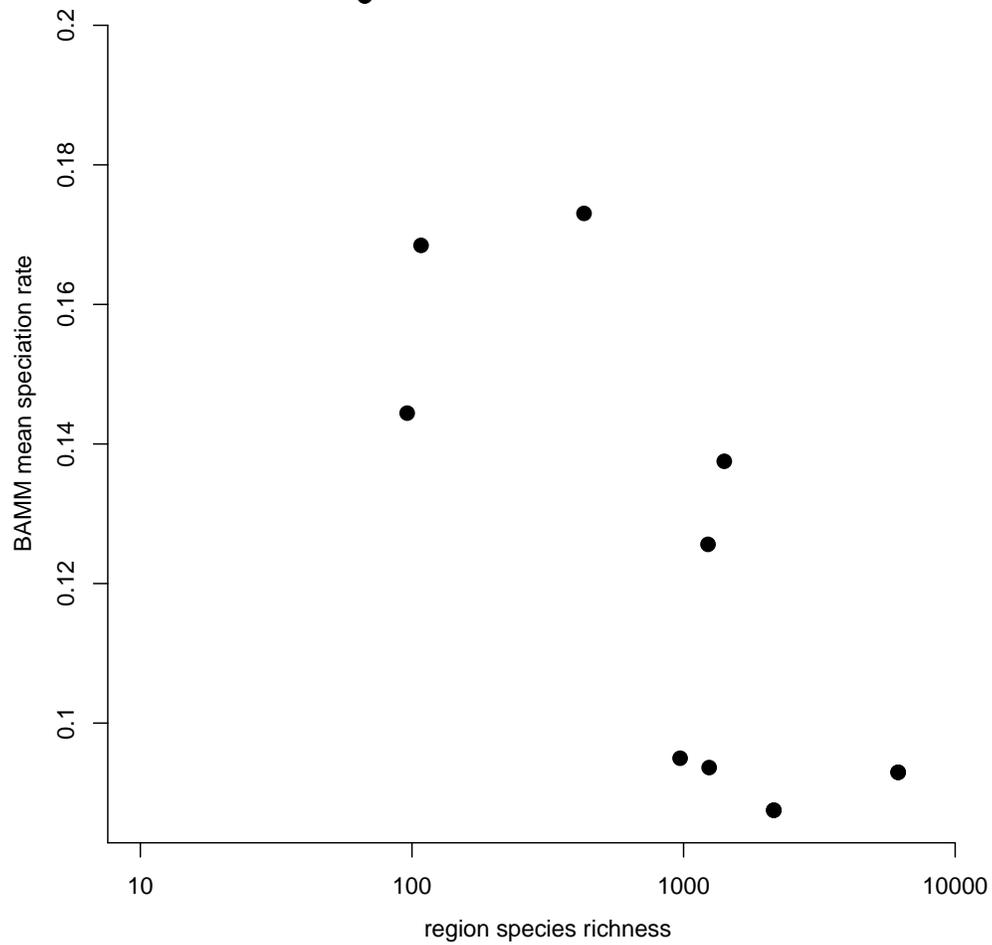
Figure S5.2: Relationship between species richness and speciation rates, for both northern and southern hemispheres. Species richness is on a log10 scale.
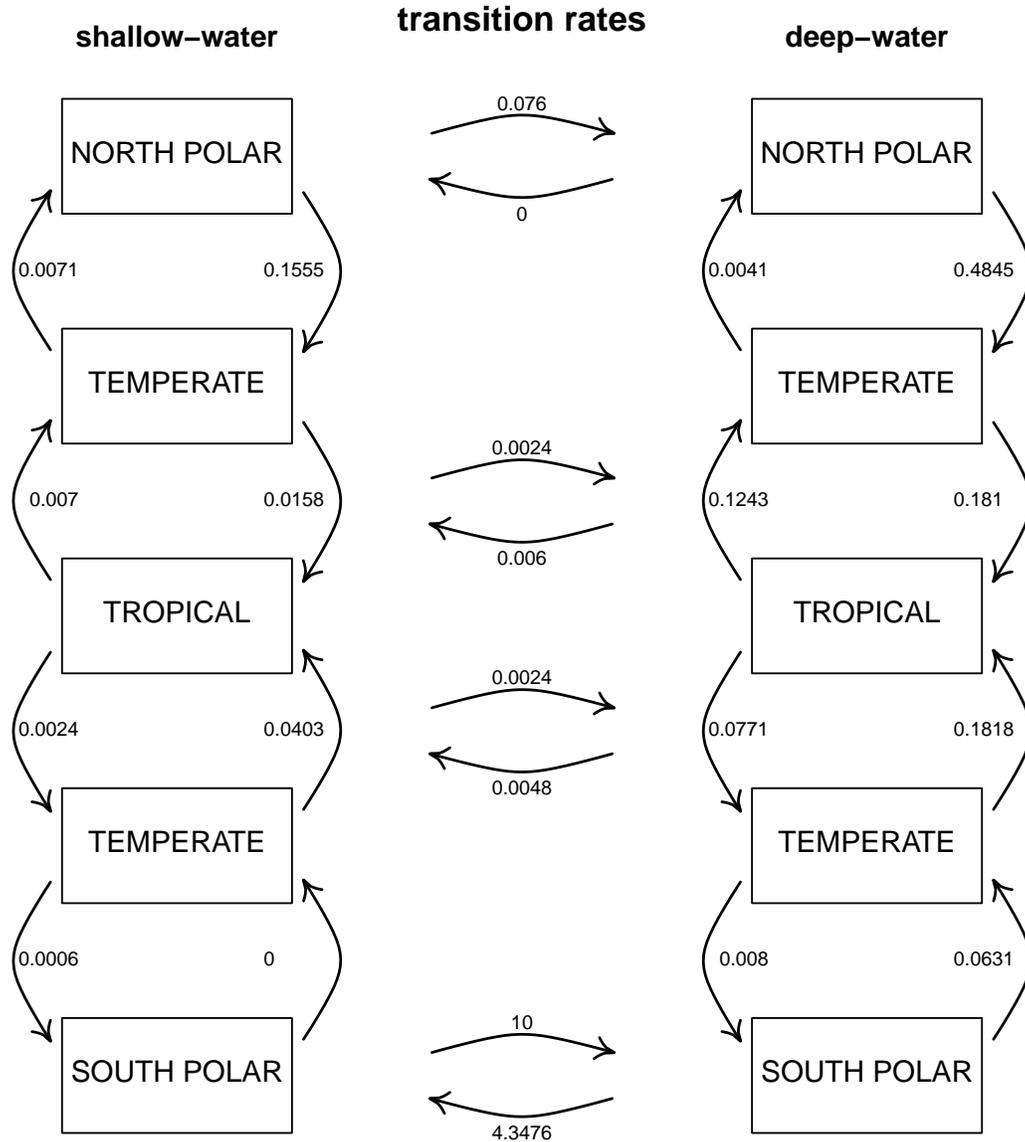
Figure S5.3: Transition rates from the best-fit biogeographic model. Vertical transition rates for tropical and temperate regions were constrained to the same value, which is why those arrows are placed midway between those regions.

Sankoff parsimony

shallow–water

deep–water
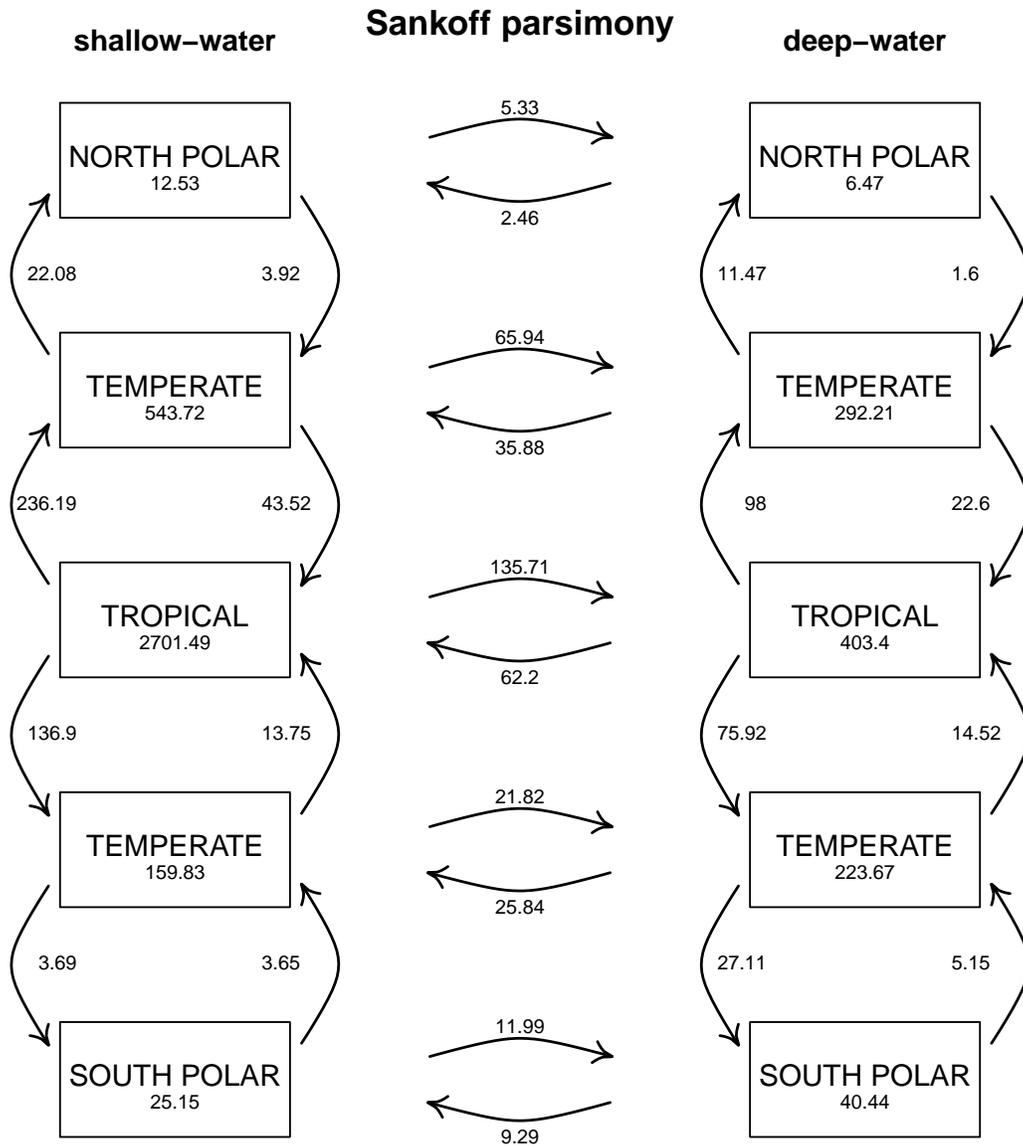
Figure S5.4: Dispersal and speciation event counts from parsimony. Arrows represent counts of dispersal events from one region to another. Counts within boxes represent the number of within-region speciation events. All counts are averaged across 1000 ancestral reconstructions.
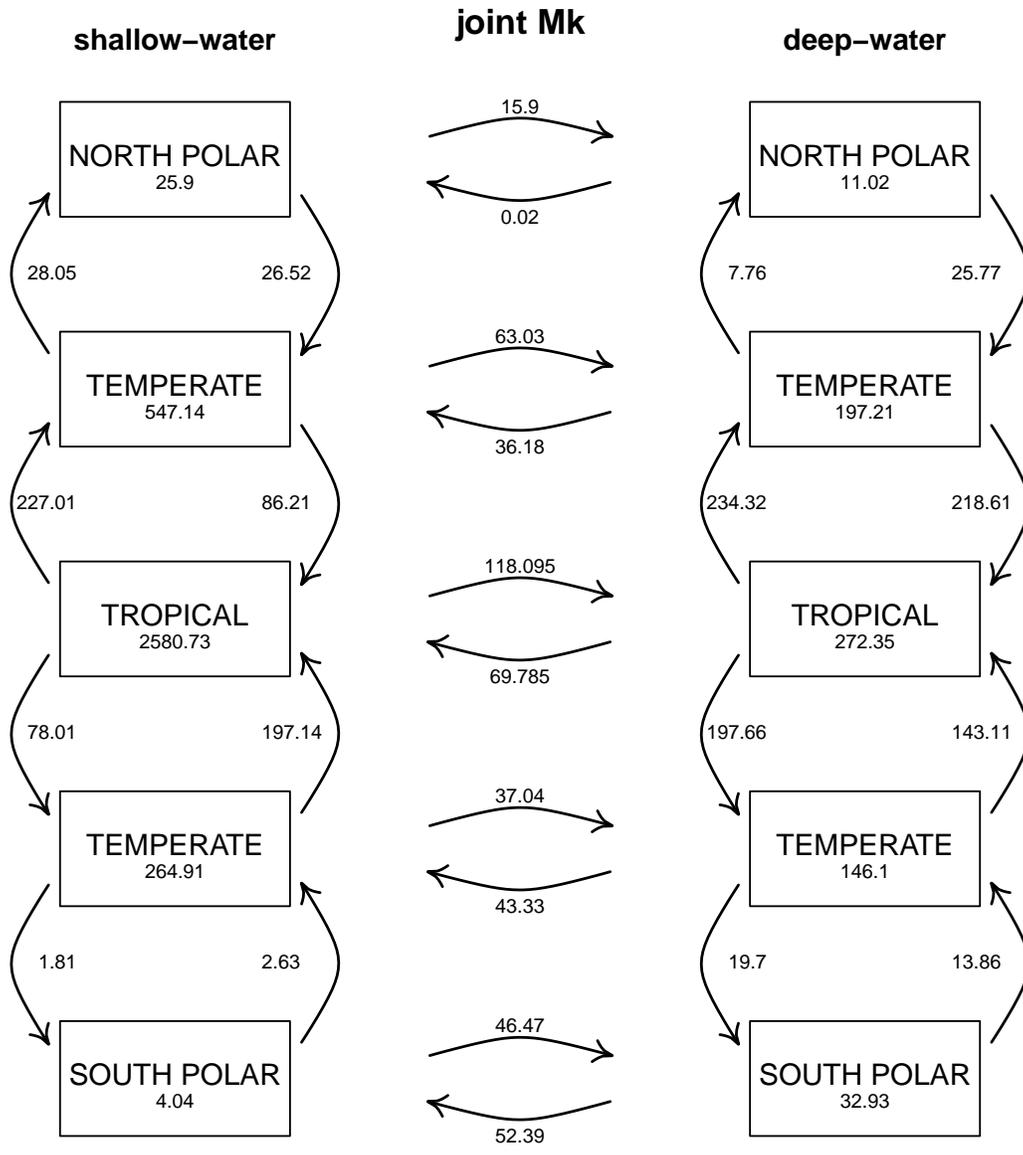
Figure S5.5: Dispersal and speciation event counts from maximum likelihood. Arrows represent counts of dispersal events from one region to another. Counts within boxes represent the number of within-region speciation events. Counts within boxes represent the number of within-region speciation events. All counts are averaged across 1000 ancestral reconstructions.
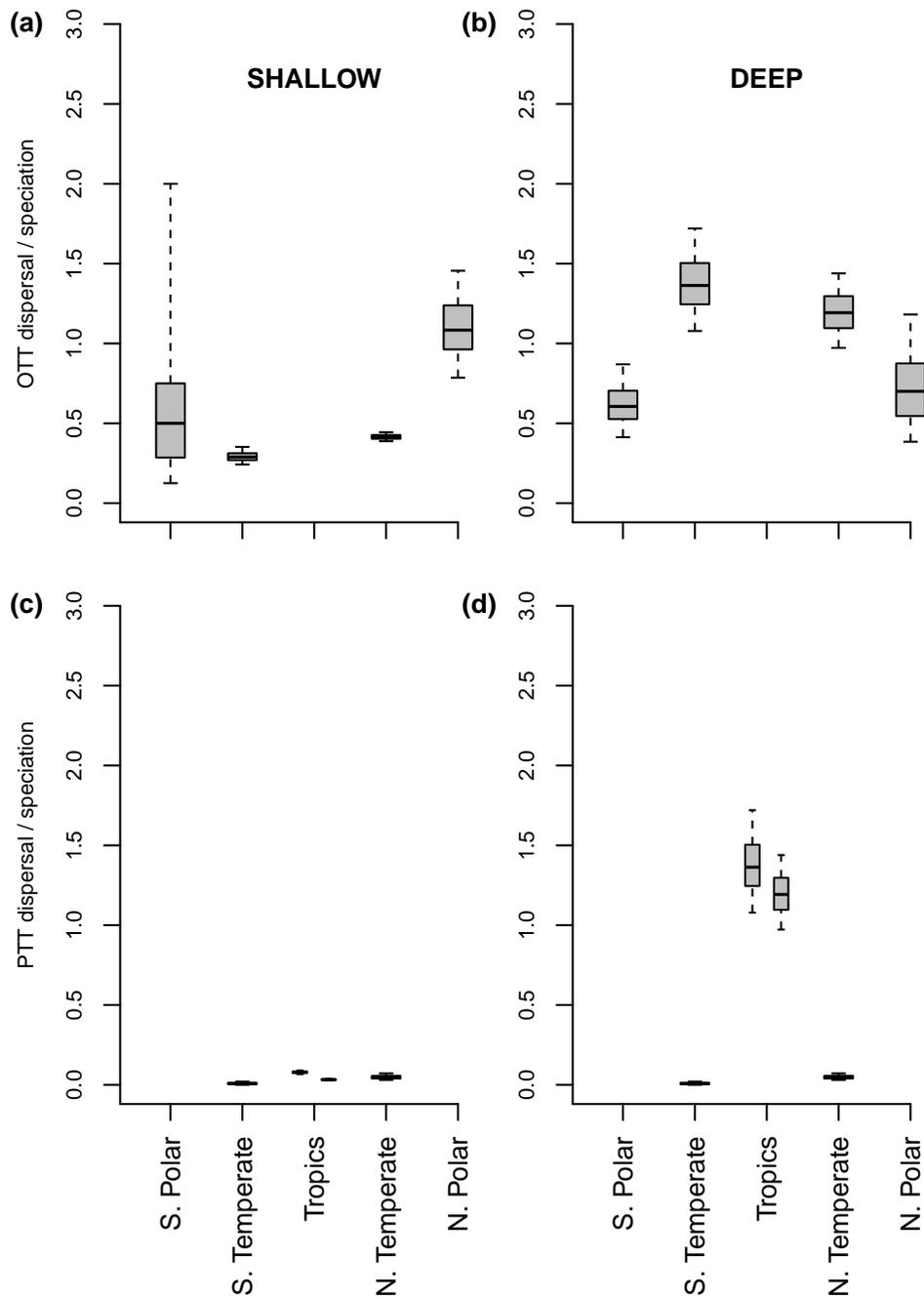
Figure S5.6: Relative importance of dispersal vs *in situ* speciation events from ML joint ancestral state reconstruction. In **(a)** and **(b)**, out-of-the-Tropics (OTT) dispersals were used, therefore there is no ratio for the Tropics. In **(c)** and **(d)**, poles-to-Tropics (PTT) dispersals were used, hence there are no ratios for the polar regions, and there is both a south Temperate → Tropics and a north Temperate → Tropics ratio. Boxplots represent the distribution of ratios from 1000 ML ancestral state reconstructions.

			S Tr 1	S Te 2	S Po 3	S Tr,Te 4	S Te,Po 5	S Tr,Te,Po 6	D Tr 7	D Te 8	D Po 9	D Tr,Te 10	D Te,Po 11	D Tr,Te,Po 12
S	Tr	1	-	q0102 ~ 0	q0103 ~ 0	q0104	q0105 ~ 0	q0106 ~ 0	q0107	q0108 ~ 0	q0109 ~ 0	q0110 ~ 0	q0111 ~ 0	q0112 ~ 0
S	Te	2	q0201 ~	-	q0203 ~ 0	q0204	q0205	q0206 ~ 0	q0207 ~ 0	q0208 ~ / q0107	q0209 ~ 0	q0210 ~ 0	q0211 ~ 0	q0212 ~ 0
S	Po	3	q0301 ~ 0	q0302 ~ 0	-	q0304 ~ 0	q0305	q0306 ~ 0	q0307 ~ 0	q0308 ~ 0	q0309	q0310 ~ 0	q0311 ~ 0	q0312 ~ 0
S	Tr,Te	4	q0401	q0402 ~ / q0401	q0403 ~ 0	-	q0405 ~ 0	q0406 ~ / q0205	q0407 ~ 0	q0408 ~ 0	q0409 ~ 0	q0410 ~ / q0107	q0411 ~ 0	q0412 ~ 0
S	Te,Po	5	q0501 ~ 0	q0502 ~ / q0401	q0503 ~ / q0401	q0504 ~ 0	-	q0506 ~ / q0204	q0507 ~ 0	q0508 ~ 0	q0509 ~ 0	q0510 ~ 0	q0511 ~ / q0309	q0512 ~ 0
S	Tr,Te,Po	6	q0601 ~ 0	q0602 ~ 0	q0603 ~ 0	q0604 ~ / q0401	q0605 ~ / q0401	-	q0607 ~ 0	q0608 ~ 0	q0609 ~ 0	q0610 ~ 0	q0611 ~ 0	q0612 ~ / q0309
D	Tr	7	q0701	q0702 ~ 0 / q0701	q0703 ~ 0	q0704 ~ 0	q0705 ~ 0	q0706 ~ 0	-	q0708 ~ 0	q0709 ~ 0	q0710	q0711 ~ 0	q0712 ~ 0
D	Te	8	q0801 ~ 0	q0802 ~ / q0701	q0803 ~ 0	q0804 ~ 0	q0805 ~ 0	q0806 ~ 0	q0807 ~ 0	-	q0809 ~ 0	-	q0811	q0812 ~ 0
D	Po	9	q0901 ~ 0	q0902 ~ 0	q0903	q0904 ~ 0	q0905 ~ 0	q0906 ~ 0	q0907 ~ 0	q0908 ~ 0	-	q0910 ~ 0	q0911	q0912 ~ 0
D	Tr,Te	10	q1001 ~ 0	q1002 ~ 0	q1003 ~ 0	q1004 ~ / q0701	q1005 ~ 0	q1006 ~ 0	q1007 ~ / q0401	q1008 ~ / q0401	q1009 ~ 0	-	q1011 ~ 0	q1012 ~ / q0811
D	Te,Po	11	q1101 ~ 0	q1102 ~ 0	q1103 ~ 0	q1104 ~ 0	q1105 ~ / q0903	q1106 ~ 0	q1107 ~ 0	q1108 ~ / q0401	q1109 ~ / q0401	q1110 ~ 0	-	q1112 ~ / q0810
D	Tr,Te,Po	12	q1201 ~ 0	q1202 ~ 0	q1203 ~ 0	q1204 ~ 0	q1205 ~ 0	q1206 ~ / q0903	q1207 ~ 0	q1208 ~ 0	q1209 ~ 0 / q0401	q1210 ~ / q0401	q1211 ~ / q0401	-

Table S5.1: Transition matrix Q as defined for the Mk model. Parameter constraints are shown as associations between a parameter and either another parameter, or 0. Tr = Tropical, Te = Temperate, Po = Polar, S = shallow, D = deep.

hemisphere	model	K	logLik	AICc	ΔAICc	wtAICc	q0104	q_q0107	q_q0204	q0205	q0305	q0309	q0401	q0701	q0710	q0810	q0811	q0903	q0911
north	unconstrained	13	-3898.55	7823.18	0.00	1.00	0.01	0.00	0.02	0.01	0.16	0.08	0.29	0.01	0.12	0.18	0.00	0.00	0.48
north	all shallow transitions equal	10	-3915.05	7850.15	26.96	0.00	0.01	0.00	0.01	0.01	0.01	0.18	0.24	0.01	0.10	0.15	0.00	0.07	0.65
north	OTT model enforced for shallow	13	-3915.05	7856.18	32.99	0.00	0.01	0.00	0.01	0.01	0.01	0.18	0.24	0.01	0.10	0.15	0.00	0.07	0.65
north	deep→shallow = shallow→deep for tropical/temperate only	12	-3916.63	7857.33	34.15	0.00	0.01	0.00	0.02	0.01	0.16	0.07	0.29	0.00	0.13	0.18	0.00	0.02	0.46
north	deep→shallow = shallow→deep	11	-3917.75	7857.55	34.37	0.00	0.01	0.00	0.02	0.00	0.17	0.04	0.28	0.00	0.12	0.17	0.01	0.04	0.38
north	OTT model enforced for deep	13	-3921.88	7869.83	46.65	0.00	0.01	0.00	0.02	0.01	0.50	10.00	0.29	0.01	0.15	0.15	0.00	10.00	0.00
north	shallow polar import = export, deep polar import = export	11	-3952.11	7926.27	103.09	0.00	0.01	0.00	0.01	0.00	0.00	0.09	0.17	0.01	0.08	0.12	0.00	0.00	0.00
north	OTT model enforced for shallow and deep	13	-3964.81	7955.70	132.52	0.00	0.01	0.00	0.01	0.00	0.00	0.09	0.17	0.01	0.10	0.10	0.00	0.00	0.00
north	all deep transitions equal	10	-4056.26	8132.57	309.39	0.00	0.01	0.00	0.02	0.00	0.52	8.87	0.18	0.00	0.03	0.03	0.03	10.00	0.03
north	shallow transitions equal, deep transitions equal	7	-4102.63	8219.29	396.11	0.00	0.01	0.00	0.01	0.01	0.01	0.09	0.12	0.01	0.03	0.03	0.03	0.03	0.03
south	unconstrained	13	-2993.12	6012.31	0.00	1.00	0.00	0.00	0.04	0.00	0.00	10.00	0.13	0.01	0.08	0.18	0.01	4.35	0.06
south	deep→shallow = shallow→deep for tropical/temperate only	12	-2999.86	6023.77	11.47	0.00	0.00	0.00	0.04	0.00	0.00	10.00	0.14	0.00	0.08	0.20	0.01	4.80	0.07
south	deep→shallow = shallow→deep	11	-3007.91	6037.87	25.56	0.00	0.00	0.00	0.04	0.00	0.03	0.09	0.14	0.00	0.08	0.20	0.01	0.09	0.06
south	shallow polar import = export, deep polar import = export	11	-3014.66	6051.37	39.06	0.00	0.00	0.00	0.04	0.00	0.00	10.00	0.12	0.01	0.07	0.17	0.01	3.96	0.01
south	all shallow transitions equal	10	-3016.47	6052.98	40.67	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.09	0.01	0.05	0.13	0.01	4.37	0.05
south	OTT model enforced for shallow	13	-3013.72	6053.51	41.20	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.08	0.01	0.05	0.13	0.01	4.34	0.05
south	OTT model enforced for deep	13	-3025.80	6077.68	65.37	0.00	0.00	0.00	0.04	0.00	0.13	10.00	0.12	0.01	0.10	0.10	0.01	4.20	0.01
south	OTT model enforced for shallow and deep	13	-3058.92	6143.92	131.61	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.06	0.01	0.05	0.05	0.01	4.07	0.01
south	all deep transitions equal	10	-3102.04	6224.12	211.81	0.00	0.00	0.00	0.02	0.00	0.06	10.00	0.06	0.00	0.03	0.03	0.03	3.97	0.03
south	shallow transitions equal, deep transitions equal	7	-3117.34	6248.69	236.38	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.05	0.01	0.02	0.02	0.02	3.90	0.02

Table S5.2: Model fit and parameter estimates for all transition rate models. Transition rate label numbers are listed in Table S1.

	N. Polar_S	N. Temperate_S	Tropical_S	S. Temperate_S	S. Polar_S	N. Polar_D	N. Temperate_D	Tropical_D	S. Temperate_D	S. Polar_D
N. Polar_S	3	23	0	0	0	6	6	2	2	0
N. Temperate_S		133	199	39	0	4	70	44	26	0
Tropical_S			702	67	3	0	28	146	36	0
S. Temperate_S				53	5	0	20	28	50	4
S. Polar_S						3	2	2	6	16
N. Polar_D							8	8	8	0
N. Temperate_D							82	76	68	28
Tropical_D								93	50	15
S. Temperate_D									46	17
S. Polar_D										11

Table S5.3: Dispersal and within-region sister species counts. On-diagonal counts represent the number of times sister species share that state. Off-diagonal counts represent the number of times sister species pairs have vicariant distributions in terms of those two states.

CHAPTER VI

Conclusion

The work that I have presented in this dissertation touches on a wide range of topics that are linked in their relevance to the study of spatial patterns of biodiversity. In addition to the findings presented in these four research chapters, it is clear that there are countless potential extensions of the topics discussed herein and many exciting opportunities to further interweave phylogenetic history and geographic information across taxonomic, temporal and spatial scales.

In Chapter 2 (Title and Rabosky 2017), we drew attention to a number of potential issues that are in many ways unique to the large phylogenies currently being produced. With Australian squamates as a case study, we demonstrated that differences in estimated divergence times and topologies across existing large squamate phylogenies result in discrepancies that can have a meaningful impact on downstream analyses. This is largely due to the inherent difficulties in phylogenetic inference with thousands of taxa when the genetic data are sparse, and when the tree space that needs to be searched is large. However, this field is rapidly experiencing improvements and methodological advances that will mitigate some of these issues in the future (Smith et al. 2010, Sanderson et al. 2015, Wright et al. 2015, Smith and Brown 2018).

The different squamate phylogenies also differed substantially in terms of crown clade ages for the different Australian groups. Perhaps more worrisome, none of the large phylogenies

were consistent with the crown clade ages reported in the literature on these groups. A closer look at the fossil calibration datasets employed in the time calibration of each phylogeny revealed that only one fossil calibration was shared amongst all of them, and there was little overlap with the fossil datasets used in the Australian squamate literature. Given these findings, the conflict in divergence times may not be particularly surprising. Fortunately, a number of recent works have since been published that present promising fossil calibrations for use in squamate phylogenetics (Head et al. 2015, Head et al. 2016), including the oldest known fossil stem squamate (Simões et al. 2018).

In Chapter 3 (Title and Bemmels 2018), we developed a comprehensive, open-source bioclimatic resource that we believe will be useful for a number of applications. As we demonstrated with several case studies, simply having a greater number of climatic variables to select from gives us greater flexibility in choosing predictors appropriate to the species under study, either through ecologically-informed manual selection or statistical variable reduction (Warren et al. 2014). Additionally, some of the climatic variables we generated have more direct relevance to ecological and physiological processes that may be important in determining species range limits (for example, growing degree-days is linked to plant phenology and growth rate; McMaster and Wilhelm 1997). These variables also make it possible to craft more targeted hypotheses in a statistical phylogeographic framework. For instance, Bemmels et al. (2016) used the ENVIREM dataset, in conjunction with the typical 19 bioclimatic variables from WorldClim (Hijmans et al. 2005), to test spatially-explicit phylogeographic hypotheses.

Perhaps equally important, but somewhat lacking, is the development of evaluation metrics that can accurately identify biologically relevant predictors. Although it is essential to have an adequate set of predictors to select from, current approaches are hindered by issues of variable collinearity and spatial autocorrelation, where species distribution models will appear to perform well simply because true presences that are close in geographic space were correctly classified as present (Lobo et al. 2008, Veloz 2009). Fourcade et al. (2017) went

so far as to demonstrate that predictors entirely generated from classical paintings, and not environmental data, could lead to species distribution models with reasonable performance, using current approaches. Therefore, although we view the ENVIREM dataset as a valuable new resource, other advances are desperately needed so that we may take full advantage of these climatic predictors and fit biologically meaningful models.

In Chapter 4, we evaluated the accuracy of a number of model-free tip rate metrics, as well as a Bayesian model-based approach, BAMM (Rabosky 2014). We tested these approaches against a number of different simulated diversification scenarios, including time-constant trees with multiple rate regimes, diversity-dependent trees, and trees where the rate of speciation evolves continuously, rather than via discrete shifts (Rabosky 2010, Beaulieu and O'Meara 2015). We demonstrated that model-free approaches (inverse of the terminal branch lengths, node density from Freckleton et al. 2008, DR from Jetz et al. 2012) all clearly track the rate of speciation and not diversification rates. This became notably apparent when the rate of lineage turnover was high. In such a scenario, the rate of speciation can theoretically take on a range of values, but net diversification (speciation minus extinction) must be low. This has important implications for the interpretation of these metrics in diversification dynamic studies, as a large tip rate value can still imply that equilibrial dynamics are in effect (it can be matched by high rates of extinction), whereas if tip rates represent net diversification rates, this would not be true.

We found that BAMM performed best in all of our evaluations. Furthermore, BAMM had the least amount of error with trees where the rate of speciation changes continuously, which is quite different from BAMM's inference model. The DR metric also performed reasonably well in our evaluations, outperforming BAMM for small rate regimes. However, whereas BAMM might be conservative in its placement of diversification rate shifts, DR suffers from high variance in tip rate estimates.

Overall, tip rates estimated with both BAMM and DR are proving to be valuable approaches to quantifying diversification rate variation across phylogenies. Of obvious relevance

to the major themes of this dissertation, tip rates lend themselves well to the study of diversification in a geographic context, as these rates can be summarized across species assemblages by grid-cell (Jetz et al. 2012, Kennedy et al. 2016, Oliveira et al. 2017, Quintero and Jetz 2018, Rabosky et al. 2018). Additionally, tip rates can be useful in the study of trait-dependent diversification (Rabosky and Goldberg 2017, Harvey and Rabosky 2017), as they make it possible to relax certain assumptions regarding rate homogeneity within character states. They are also easy to calculate, and can therefore be applied to very large phylogenies. Future research could explore how DR performs for very large phylogenies, where BAMM is incapable of converging due to the extremely large number of potential rate shift locations. Although the same issues with DR that were found in this study would persist, it would be important to assess whether or not this has any notable effect on hypothesis testing, especially when the phylogeny contains thousands of species.

Several biases may exist in empirical phylogenies that were not explored in this study. For instance, phylogenies rarely contain complete species representation, and there may be bias in terms of the phylogenetic placements of those missing species. BAMM can account for missing species analytically by incorporating the fraction of included species for different nodes in the tree (Rabosky 2014). Despite these analytical corrections, BAMM may still have lower statistical power to detect rate shifts, as power is related to the number of species in a potential rate regime (Rabosky et al. 2017). Stochastic polytomy resolution, whereby missing species are added to the phylogeny according to taxonomic constraints, may provide a way to improve performance. By generating many possible trees with complete species sampling, tip rate metrics can be calculated while integrating across alternative placements of missing taxa. Estimates of DR should improve (Rabosky et al. 2018) and BAMM should benefit from increased statistical power to identify rate shifts.

In Chapter 5, we explored how biogeographic dispersal has contributed to the latitudinal diversity gradient in marine fishes. We found that rates of dispersal are higher in a poles-to-tropics direction. Additionally, we recovered greater dispersal rates in deep-water, compared

to shallow-water lineages. This supports the notion that biogeographic barriers in shallow versus deep-water are not equivalent, and that there is greater environmental homogeneity on a global scale at depth (Vecchione et al. 2015, Gaither et al. 2016, Priede 2017). We found a strong contrast in rates between the northern and southern high latitudes, with dispersal rates out of the Arctic being the greatest, and relatively low dispersal rates assoeciated with the Southern Ocean, thus reflecting the Arctic's history as an area of biotic interchange between the north Pacific and north Atlantic oceans (Mecklenburg et al. 2011).

Counts of dispersal, as estimated through ancestral state reconstruction, show that directionality of dispersal rates does not imply realized net movement of species in that direction. Rather, the strong latitudinal gradient in species richness is such that a lower rate of migration from a large species pool still leads to more dispersals than a higher rate of migration from a small pool of species.

Overall, we find that the net migration of species across latitudes does support a key prediction of the "out of the tropics" model of Jablonski et al. (2006); however, latitudinal gradients in speciation rates do not (Rabosky et al. 2018). A critical piece of the puzzle that is currently missing is an understanding of how rates of extinction vary across latitudes. Previous research indicates that rates of extinction are likely to be elevated at high latitudes (Weir and Schluter 2007, Clarke and Crame 2010, Botero et al. 2014, Weir 2014) and that the high latitudes might therefore be characterized as regions of increased lineage turnover. A number of studies have investigated the differences in diversification and dispersal in and out of the tropics in a number of groups of organisms (Pyron and Wiens 2013, Pyron 2014, Rolland et al. 2014, Economo et al. 2018, Pulido-Santacruz and Weir 2016) and have found that rates of extinction are typically higher outside of the tropics. Although it is difficult to accurately estimate rates of extinction from extant-only molecular phylogenies (Rabosky 2010, Mitchell et al. 2018), information from groups with a robust fossil record (such as marine bivalves, Jablonski et al. 2017) and knowledge of past environmental change (Clarke and Crame 2010, Mecklenburg et al. 2011) support the possibility that rates of extinction

might be high at high latitudes, but a thorough understanding of how rates of extinction and net diversification rate vary with latitude is needed.

The dispersal rate patterns that we present appear to indicate that environmental tolerances might be important in determining which species are more likely to disperse between particular regions (Kerkhoff et al. 2014, Duchêne and Cardillo 2015), as we find greater dispersal in deep waters. Additional analyses that estimate dispersal rates for different groups within marine fishes would allow us to take a more detailed look at these patterns, and incorporate information on evolutionary lability and conservatism in species environmental preferences as well as ecological attributes of these diverse groups.

This body of work, along with other studies, strengthens the call to link phylogenetic information with species distributions. These studies also highlight the insights that can be gained through the formal evaluation of phylogenetic methods and geospatial datasets in a comparative framework. Finally, this work demonstrates the importance of global scale and phylogenetically comprehensive analyses. Surprising findings that arise from this integrated viewpoint challenge some of our former conceptions about the processes underlying major biodiversity patterns across broad spatial and taxonomic scales.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adams, D., Kozak, K.H. & Wiens, J.J. (2009) Are rates of species diversification correlated with rates of morphological evolution? *Proceedings of the Royal Society of London Series B: Biological Sciences*, **276**, 2729–2738.

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B. & Anderson, R.P. (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 1–5.

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

Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G. & Harmon, L.J. (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*, **106**, 13410–13414.

Alfaro, M.E., Santini, F. & Brock, C.D. (2007) Do reefs drive diversification in marine teleosts? Evidence from the pufferfishes and their allies (order Tetraodontiformes). *Evolution*, **61**, 2104–2126.

Allen, A.P., Brown, J.H. & Gillooly, J.F. (2002) Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, **297**, 1545–1548.

Alvarado-Serrano, D.F. & Knowles, L.L. (2014) Ecological niche models in phylogeographic studies: applications, advances and precautions. *Molecular Ecology Resources*, **14**, 233–248.

Amatulli, G., Domisch, S., Tuanmu, M.N., Parmentier, B., Ranipeta, A., Malczyk, J. & Jetz, W. (2018) A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, **5**, 180040.

Anderson, R.P. & Gonzalez Jr, I. (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling*, **222**, 2796–2811.

Antonelli, A., Zizka, A., Carvalho, F.A., Scharn, R., Bacon, C.D., Silvestro, D. & Condamine, F.L. (2018) Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences*, **60**, 201713819–6.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.

Austin, M.P. & Van Niel, K.P. (2011) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography*, **38**, 1–8.

Barbet-Massin, M. & Jetz, W. (2014) A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions*, **20**, 1285–1295.

Barbosa, A.M. (2015) fuzzySim: applying fuzzy logic to binary similarity indices in ecology. *Methods in Ecology and Evolution*, **6**, 853–858.

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

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

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

Becker, J.J., Sandwell, D.T., Smith, W.H.F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G. & Weatherall, P. (2009) Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS. *Marine Geodesy*, **32**, 355–371.

Bell, M. & Laine, E.P. (1985) Erosion of the Laurentide region of North America by glacial and glaciofluvial processes. *Quaternary Research*, **23**, 154–174.

Belmaker, J. & Jetz, W. (2012) Regional Pools and Environmental Controls of Vertebrate Richness. *The American Naturalist*, **179**, 512–523.

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

Bemmels, J.B., Title, P.O., Ortego, J. & Knowles, L.L. (2016) Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Molecular Ecology*, **25**, 4889–4906.

Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L. & Purvis, A. (2007) The delayed rise of present-day mammals. *Nature*, **446**, 507–512.

Böhner, J., Köthe, R., Conrad, O., Gross, J., Ringeler, A. & Selige, T. (2002) *Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation*, volume EUR 20398 EN, pp. 213–222. The European Soil Bureau, Joint Research Centre, Ispra.

Booth, T.H., Nix, H.A., Busby, J.R. & Hutchinson, M.F. (2014) bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEntstudies. *Diversity and Distributions*, **20**, 1–9.

Botero, C.A., Dor, R., McCain, C.M. & Safran, R.J. (2014) Environmental harshness is positively correlated with intraspecific divergence in mammals and birds. *Molecular Ecology*, **23**, 259–268.

Boyer, T., Antonov, J.I., Baranova, O.K., Coleman, C., Garcia, H.E., Grodsky, A., Johnson, D.R., Locarnini, R.A., Mishonov, A.V., O'Brien, T.D., Paver, C.R., Reagan, J.R., Seidov, D., Smolyar, I.V. & Zweng, M.M. (2013) World ocean database. Technical Report 209, NOAA, Silver Spring, MD.

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H. & Bollmann, K. (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography*, **36**, 971–983.

Briggs, J.C. (2000) Centrifugal speciation and centres of origin. *Journal of Biogeography*, **27**, 1183–1188.

Briggs, J.C. (2003) Marine centres of origin as evolutionary engines. *Journal of Biogeography*, **30**, 1–18.

Briggs, J.C. & Bowen, B.W. (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography*, **39**, 12–30.

Briggs, J.C. & Bowen, B.W. (2013) Marine shelf habitat: biogeography and evolution. *Journal of Biogeography*, **40**, 1023–1035.

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

Brown, A. & Thatje, S. (2013) Explaining bathymetric diversity patterns in marine benthic invertebrates and demersal fishes: physiological contributions to adaptation of life at depth. *Biological Reviews*, **89**, 406–426.

Brown, J.L. & Knowles, L.L. (2012) Spatially explicit models of dynamic histories: examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Molecular Ecology*, **21**, 3757–3775.

Brown, J.M. (2014a) Detection of Implausible Phylogenetic Inferences Using Posterior Predictive Assessment of Model Fit. *Systematic Biology*, **63**, 334–348.

Brown, J.M. (2014b) Predictive approaches to assessing the fit of evolutionary models. *Systematic Biology*, **63**, 289–292.

Budic, L., Didenko, G. & Dormann, C.F. (2015) Squares of different sizes: effect of geographical projection on model parameter estimates in species distribution modeling. *Ecology and Evolution*, **6**, 202–211.

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

Burleigh, J.G., Kimball, R.T. & Braun, E.L. (2015) Building the avian tree of life using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, **84**, 53–63.

Burnham, K.P. & Anderson, D.R. (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, **33**, 261–304.

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

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

Caldwell, M.W., Nydam, R.L., Palci, A. & Apesteguía, S. (2015) The oldest known snakes from the Middle Jurassic-Lower Cretaceous provide insights on snake evolution. *Nature Communications*, **6**, 1–11.

Cardillo, M. (1999) Latitude and rates of diversification in birds and butterflies. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **266**, 1221–1225.

Cardillo, M., Orme, C.D.L. & Owens, I.P.F. (2005) Testing for Latitudinal Bias in Diversification Rates: An Example Using New World Birds. *Ecology*, **86**, 2278–2287.

Castro-Insua, A., Gómez-Rodriguez, C., Wiens, J.J. & Baselga, A. (2018) Climatic niche divergence drives patterns of diversification and richness among mammal families. *Scientific Reports*, p. 8781.

Chan, L.M. & Brown, J.L. (2011) Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Molecular Phylogenetics and Evolution*, **59**, 523–537.

Cheng, C.H.C. (2003) Functional Antifreeze Glycoprotein Genes in Temperate-Water New Zealand Nototheniid Fish Infer an Antarctic Evolutionary Origin. *Molecular Biology and Evolution*, **20**, 1897–1908.

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

Clarke, A. & Crame, J.A. (2010) Evolutionary dynamics at high latitudes: speciation and extinction in polar marine faunas. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3655–3666.

Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., Ballesteros, E., Bianchi, C.N., Corbera, J., Dailianis, T., Danovaro, R., Estrada, M., Froglia, C., Galil, B.S., Gasol, J.M., Gertwagen, R., Gil, J., Guilhaumon, F., Kesner-Reyes, K., Kitsos, M.S., Koukouras, A., Lampadariou, N., Laxamana, E., López-Fé de la Cuadra, C.M., Lotze, H.K., Martin, D., Mouillot, D., Oro, D., Raicevich, S., Rius-Barile, J., Saiz-Salinas, J.I., San Vicente, C., Somot, S., Templado, J., Turon, X., Vafidis, D., Villanueva, R. & Voultsiadou, E. (2010) The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. *PLoS ONE*, **5**, e11842–36.

Collins, W.D., Bitz, C.M., Blackmon, M.L., Bonan, G.B., Bretherton, C.S., Carton, J.A., Chang, P., Doney, S.C., Hack, J.J. & Henderson, T.B. (2006) The community climate system model version 3 (CCSM3). *Journal of Climate*, **19**, 2122–2143.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. & Böhner, J. (2015) System for automated geoscientific analyses (saga) v. 2.1.4. *Geoscientific Model Development*, **8**, 1991–2007.

Constable, H., Guralnick, R.P., Wieczorek, J., Spencer, C. & Peterson, A.T. (2010) VertNet: A new model for biodiversity data sharing. *PLoS Biology*, **8**, e1000309.

Cowman, P.F. & Bellwood, D.R. (2013) The historical biogeography of coral reef fishes: global patterns of origination and dispersal. *Journal of Biogeography*, **40**, 209–224.

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

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guegan, J.F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E. & Turner, J.R.G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121–1134.

Daget, P. (1977) Le bioclimat méditerranéen: analyse des formes climatiques par le système d'Emberger. *Vegetatio*, **32**, 87–103.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L. & Pasteris, P. (2002) A knowledge-based approach to the statistical mapping of climate. *Climate research*, **22**, 99–113.

Davis, E.B., McGuire, J.L. & Orcutt, J.D. (2014) Ecological niche models of mammalian glacial refugia show consistent bias. *Ecography*, **37**, 1133–1138.

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**, 11484.

Developers, K..M. (2004) K-1 coupled gcm (miroc) description. Technical Report K-1 Technical Report 1, Univ. of Tokyo, Cent. for Clim. Syst. Res., Tokyo.

DeVries, A.L. & Steffensen, J.F. (2005) The Arctic and Antarctic Polar Marine Environments. *Physiology of Polar Fishes*, pp. 1–24. Elsevier, New York.

Donoghue, M.J. & Edwards, E.J. (2014) Biome Shifts and Niche Evolution in Plants. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 547–572.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D. & Lautenbach, S. (2012) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.

Dornburg, A., Federman, S., Lamb, A.D., Jones, C.D. & Near, T.J. (2017) Cradles and museums of Antarctic teleost biodiversity. *Nature Ecology & Evolution*, pp. 1–6.

Doswald, N., Willis, S.G., Collingham, Y.C., Pain, D.J., Green, R.E. & Huntley, B. (2009) Potential impacts of climatic change on the breeding and non-breeding ranges and migration distance of European Sylviawarblers. *Journal of Biogeography*, **36**, 1194–1208.

Driskell, A.C., Ané, C., Burleigh, J.G., McMahon, M.M., O'Meara, B.C. & Sanderson, M.J. (2004) Prospects for building the tree of life from large sequence databases. *Science*, **306**, 1172–1174.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Duchêne, D.A. & Cardillo, M. (2015) Phylogenetic patterns in the geographic distributions of birds support the tropical conservatism hypothesis. *Global Ecology and Biogeography*, **24**, 1261–1268.

Duchêne, S., Lanfear, R. & Ho, S.Y.W. (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Molecular Phylogenetics and Evolution*, **78**, 277–289.

Dyke, A.S., Andrews, J.T., Clark, P.U. & England, J.H. (2002) The Laurentide and Innuitian ice sheets during the last glacial maximum. *Quaternary Science Reviews*, **21**, 9–31.

Eastman, J.T. (1993) *Antarctic fish biology: evolution in a unique environment*. Academic Press.

Eastman, J.T. (2005) The nature of the diversity of Antarctic fishes. *Polar Biology*, **28**, 93–107.

Economo, E.P., Huang, J.P., Fischer, G., Sarnat, E.M., Narula, N., Janda, M., Guénard, B., Longino, J.T. & Knowles, L.L. (2018a) Evolution of the latitudinal diversity gradient in the hyperdiverse ant genus Pheidole. *biorxivorg*, pp. 1–41.

Economo, E.P., Narula, N., Friedman, N.R., Weiser, M.D. & Guénard, B. (2018b) Macroecology and macroevolution of the latitudinal diversity gradient in ants. *Nature Communications*, p. 1778.

Ellwood, E.R., Dunckel, B.A., Flemons, P., Guralnick, R.P., Nelson, G., Newman, G., Newman, S., Paul, D., Riccardi, G., Rios, N., Seltmann, K.C. & Mast, A.R. (2015) Accelerating the Digitization of Biodiversity Research Specimens through Online Public Participation. *BioScience*, **65**, 383–396.

Etienne, R.S. & Haegeman, B. (2012) A Conceptual and Statistical Framework for Adaptive Radiations with a Key Role for Diversity Dependence. *The American Naturalist*, **180**, E75–E89.

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

Faurby, S. & Svenning, J.C. (2015) A species-level phylogeny of all extant and late Quaternary extinct mammals using a novel heuristic-hierarchical Bayesian approach. *Molecular Phylogenetics and Evolution*, **84**, 14–26.

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **21**, 455–14.

Fine, P.V.A. (2015) Ecological and Evolutionary Drivers of Geographic Variation in Species Diversity. *Annual Review of Ecology, Evolution, and Systematics*, **46**, 369–392.

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.

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

FitzJohn, R.G. (2012) Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, pp. no–no.

Fossheim, M., Primicerio, R., Johannesen, E., Ingvaldsen, R.B., Aschan, M.M. & Dolgov, A.V. (2015) Recent warming leads to a rapid borealization of fish communities in the Arctic. *Nature Climate Change*, **5**, 673–677.

Fourcade, Y., Besnard, A.G. & Secondi, J. (2017) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, **27**, 245–256.

Freckleton, R. & Jetz, W. (2009) Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **276**, 21.

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

Fritz, S.A. & Rahbek, C. (2012) Global patterns of amphibian phylogenetic diversity. *Journal of Biogeography*, **39**, 1373–1382.

Gaither, M.R., Bowen, B.W., Rocha, L.A. & Briggs, J.C. (2015) Fishes that rule the world: circumtropical distributions revisited. *Fish and Fisheries*, **17**, 664–679.

Gaither, M.R., Violi, B., Gray, H.W.I., Neat, F., Drazen, J.C., Grubbs, R.D., Roa-Varón, A., Sutton, T. & Hoelzel, A.R. (2016) Depth as a driver of evolution in the deep sea: Insights from grenadiers (Gadiformes: Macrouridae) of the genus Coryphaenoides. *Molecular Phylogenetics and Evolution*, **104**, 73–82.

Gavin, D.G., Fitzpatrick, M.C., Gugger, P.F., Heath, K.D., Rodríguez-Sánchez, F., Dobrowski, S.Z., Hampe, A., Hu, F.S., Ashcroft, M.B., Bartlein, P.J., Blois, J.L., Carstens, B.C., Davis, E.B., de Lafontaine, G., Edwards, M.E., Fernandez, M., Henne, P.D., Herring, E.M., Holden, Z.A., Kong, W.s., Liu, J., Magri, D., Matzke, N.J., McGlone, M.S., Saltré, F., Stigall, A.L., Tsai, Y.H.E. & Williams, J.W. (2014) Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, **204**, 37–54.

Glor, R.E. & Warren, D.L. (2011) Testing ecological explanations for biogeographic boundaries. *Evolution*, **65**, 673–683.

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

Goldberg, E.E., Roy, K., Lande, R. & Jablonski, D. (2005) Diversity, endemism, and age distributions in macroevolutionary sources and sinks. *The American Naturalist*, **165**, 623–633.

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

Graham, C.H. & Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.

Graur, D. & Martin, W. (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics*, **20**, 80–86.

Greer, A.E. (1979) A phylogenetic subdivision of Australian skinks. *Records of the Australian Museum*, **32**, 339–371.

Guralnick, R.P., Wieczorek, J., Beaman, R. & Hijmans, R.J. (2006) BioGeomancer: automated georeferencing to map the world's biodiversity data. *PLoS Biology*, **4**, e381.

Hamann, A., Roberts, D.R., Barber, Q.E., Carroll, C. & Nielsen, S.E. (2014) Velocity of climate change algorithms for guiding conservation and management. *Global Change Biology*, **21**, 997–1004.

Hamann, A., Wang, T., Spittlehouse, D.L. & Murdock, T.Q. (2013) A Comprehensive, High-Resolution Database of Historical and Projected Climate Surfaces for Western North America. *Bulletin of the American Meteorological Society*, **94**, 1307–1309.

Hargreaves, G.L. & Hargreaves, G.H. (1985) Irrigation water requirements for Senegal River basin. *Journal of Irrigation and Drainage Engineering*, **111**, 265–275.

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

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., OBrian, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, Water, and Broad-Scale Geographic Patterns of Species Richness. *Ecology*, **84**, 3105–3117.

He, Q., Edwards, D.L. & Knowles, L.L. (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, **67**, 3386–3402.

Head, J.J. (2015) Fossil calibration dates for molecular phylogenetic analysis of snakes 1: Serpentes, Alethinophidia, Boidae, Pythonidae. *Palaeontologia Electronica*, **18**, 1–17.

Head, J.J., Mahlow, K. & Müller, J. (2016) Fossil calibration dates for molecular phylogenetic analysis of snakes 2: Caenophidia, Colubroidea, Elapoidea, Colubridae. *Palaeontologia Electronica*, **19**, 1–21.

Heath, T.A., Huelsenbeck, J.P. & Stadler, T. (2014) The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, **111**, E2957–E2966.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. (2015) Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, **32**, 835–845.

Helmus, M.R., Bland, T.J., Williams, C.K. & Ives, A.R. (2007) Phylogenetic measures of biodiversity. *The American Naturalist*, **169**, E68–83.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S. & Kempen, B. (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, **12**, e0169748–40.

Hijmans, R.J. (2016a) *dismo: species distribution modeling*. R package version 1.0-15.

Hijmans, R.J. (2016b) *raster: Geographic Data Analysis and Modeling*. R package version 2.5-8.

Hijmans, R.J., Cameron, S., Parra, J.L. & Jones, P. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hijmans, R.J. & Graham, C.H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12**, 2272–2281.

Hillebrand, H. (2004) On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist*, **163**, 192–211.

Hinchliff, C.E. & Roalson, E.H. (2013) Using supermatrices for phylogenetic inquiry: an example using the sedges. *Systematic Biology*, **62**, 205–219.

Hinchliff, C.E. & Smith, S.A. (2014) Some Limitations of Public Sequence Data for Phylogenetic Inference (in Plants). *PLoS ONE*, **9**, e98986.

Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., Crandall, K.A., Deng, J., Drew, B.T., Gazis, R., Gude, K., Hibbett, D.S., Katz, L.A., Laughinghouse IV, H.D., McTavish, E.J., Midford, P.E., Owen, C.L., Ree, R.H., Rees, J.A., Soltis, D.E., Williams, T. & Cranston, K.A. (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences*, **112**, 12764–12769.

Hof, A.R., Jansson, R. & Nilsson, C. (2012) The usefulness of elevation as a predictor variable in species distribution modelling. *Ecological Modelling*, **246**, 86–90.

Holt, B.G., Costa, G.C., Penone, C., Lessard, J.P., Brooks, T.M., Davidson, A.D., Blair Hedges, S., Radeloff, V.C., Rahbek, C., Rondinini, C. & Graham, C.H. (2017) Environmental variation is a major predictor of global trait turnover in mammals. *Journal of Biogeography*, **19**, 134–13.

Howe, G.T., Aitken, S.N., Neale, D.B., Jermstad, K.D., Wheeler, N.C. & Chen, T.H. (2003) From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany*, **81**, 1247–1266.

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

Hua, X. & Wiens, J.J. (2013) How Does Climate Influence Speciation? *The American Naturalist*, **182**, 1–12.

Hugall, A.F., Foster, R., Hutchinson, M. & Lee, M.S.Y. (2008) Phylogeny of Australasian agamid lizards based on nuclear and mitochondrial genes: implications for morphological evolution and biogeography. *Biological Journal of the Linnean Society*, **93**, 343–358.

Hugall, A.F., Foster, R. & Lee, M.S.Y. (2007) Calibration Choice, Rate Smoothing, and the Pattern of Tetrapod Diversification According to the Long Nuclear Gene RAG-1. *Systematic Biology*, **56**, 543–563.

Jablonski, D., Belanger, C.L., Berke, S.K., Huang, S., Krug, A.Z., Roy, K., Tomasovych, A. & Valentine, J.W. (2013) Out of the tropics, but how? Fossils, bridge species, and thermal ranges in the dynamics of the marine latitudinal diversity gradient. *Proceedings of the National Academy of Sciences*, **110**, 10487–10494.

Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the Tropics: Evolutionary Dynamics of the Latitudinal Diversity Gradient. *Science*, **314**, 102–106.

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

Jablonski, D., Huang, S., Roy, K. & Valentine, J.W. (2017) Shaping the Latitudinal Diversity Gradient: New Perspectives from a Synthesis of Paleobiology and Biogeography. *The American Naturalist*, **189**, 1–12.

Jetz, W. & Pyron, R.A. (2018) The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, **2**, 850–858.

Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Jezkova, T. & Wiens, J.J. (2018) Testing the role of climate in speciation: New methods and applications to squamate reptiles (lizards and snakes). *Molecular Ecology*, **27**, 2754–2769.

Jones, M.E., Anderson, C.L., Hipsley, C.A., Müller, J., Evans, S.E. & Schoch, R.R. (2013) Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evolutionary Biology*, **13**, 208.

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

Kaschner, K., Watson, R., Trites, A.W. & Pauly, D. (2006) Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, **316**, 285–310.

Katul, G.G., Oren, R., Manzoni, S., Higgins, C. & Parlange, M.B. (2012) Evapotranspiration: A process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system. *Reviews of Geophysics*, **50**, 1–25.

Kay, K., Voelckel, C., Yang, J.Y., Hufford, K.M., Kaska, D.D. & Hodges, S.A. (2006) *Floral characters and species diversification*. Oxford Univ. Press Oxford, Oxford.

Kearney, M., Phillips, B.L., Tracy, C.R., Christian, K.A., Betts, G. & Porter, W.P. (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography*, **31**, 423–434.

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

Kennedy, J.D., Wang, Z., Weir, J.T., Rahbek, C., Fjeldså, J. & Price, T.D. (2014) Into and out of the tropics: the generation of the latitudinal gradient among New World passerine birds. *Journal of Biogeography*, **41**, 1746–1757.

Kennett, J.P. (1982) *Marine Geology, 813 pp.* Prentice-Hall, Englewood Cliffs, NJ.

Keogh, J.S. (1998) Molecular phylogeny of elapid snakes and a consideration of their biogeographic history. *Biological Journal of the Linnean Society*, **63**, 177–203.

Kerkhoff, A.J., Moriarty, P.E. & Weiser, M.D. (2014) The latitudinal species richness gradient in New World woody angiosperms is consistent with the tropical conservatism hypothesis. *Proceedings of the National Academy of Sciences*, **111**, 8125–8130.

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

Knowles, L.L. & Alvarado-Serrano, D.F. (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology*, **19**, 3727–3745.

Kozak, K.H. & Wiens, J.J. (2010) Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology Letters*, **13**, 1378–1389.

Kozak, K.H. & Wiens, J.J. (2016) What explains patterns of species richness? The relative importance of climatic-niche evolution, morphological evolution, and ecological limits in salamanders. *Ecology and Evolution*, **6**, 5940–5949.

Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J. & Scott, J.K. (2011) CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, **3**, 53–64.

Kuhn, T.S., Mooers, A.O. & Thomas, G.H. (2011) A simple polytomy resolver for dated phylogenies. *Methods in Ecology and Evolution*, **2**, 427–436.

Lassueur, T., Joost, S. & Randin, C.F. (2006) Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecological Modelling*, **198**, 139–153.

Lee, M.S.Y., Oliver, P.M. & Hutchinson, M.N. (2009) Phylogenetic uncertainty and molecular clock calibrations: A case study of legless lizards (Pygopodidae, Gekkota). *Molecular Phylogenetics and Evolution*, **50**, 661–666.

Lewis, P.O. (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50**, 913–925.

Lewitus, E. & Morlon, H. (2016) Characterizing and Comparing Phylogenies from their Laplacian Spectrum. *Systematic Biology*, **65**, 495–507.

Lima-Ribeiro, M.S., Varela, S., González-Hernández, J., de Oliveira, G., Diniz-Filho, J.A.F. & Terribile, L.C. (2015) Ecoclimate: A database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodiversity Informatics*, **10**, 1–21.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.

MacArthur, R.H. (1969a) Patterns of communities in the tropics. *Biological Journal of the Linnean Society*, **1**, 19–30.

MacArthur, R.H. (1969b) Patterns of communities in the tropics. *Biological Journal of the Linnean Society*, **1**, 19–30.

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

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

Magallón, S. & Sanderson, M.J. (2001) Absolute diversification rates in angiosperm clades. *Evolution*, **55**, 1762–1780.

Marin, J., Donnellan, S.C., Blair Hedges, S., Doughty, P., Hutchinson, M.N., Cruaud, C. & Vidal, N. (2012) Tracing the history and biogeography of the Australian blindsnake radiation. *Journal of Biogeography*, **40**, 928–937.

Marin, J., Donnellan, S.C., Hedges, S.B., Puillandre, N., Aplin, K.P., Doughty, P., Hutchinson, M.N., Couloux, A. & Vidal, N. (2013) Hidden species diversity of Australian burrowing snakes (Ramphotyphlops). *Biological Journal of the Linnean Society*, **110**, 427–441.

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

Martínez-Meyer, E., Peterson, A.T., Servín, J.I. & Kiff, L.F. (2006) Ecological niche modelling and prioritizing areas for species reintroductions. *Oryx*, **40**, 411–8.

McMaster, G.S. & Wilhelm, W.W. (1997) Growing degree-days: one equation, two interpretations. *Agricultural and Forest Meteorology*, **87**, 291–300.

McPeek, M.A. & Brown, J.M. (2007) Clade age and not diversification rate explains species richness among animal taxa. *The American Naturalist*, **169**, E97–E106.

Mecklenburg, C.W., Mecklenburg, A.T., Sheiko, B.A. & Steinke, D. (2016) *Pacific Arctic Marine Fishes*. CAFF International Secretariat.

Mecklenburg, C.W., Møller, P.R. & Steinke, D. (2011) Biodiversity of arctic marine fishes: taxonomy and zoogeography. *Marine Biodiversity*, **41**, 109–140.

Merow, C., Smith, M.J. & Silander Jr, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Metzger, M.J., Bunce, R.G.H., Jongman, R.H.G., Sayre, R., Trabucco, A. & Zomer, R. (2013) A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*, **22**, 630–638.

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

Misof, B., Meyer, B., von Reumont, B.M., Kück, P., Misof, K. & Meusemann, K. (2013) Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. *BMC bioinformatics*, **14**, 348.

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

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

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

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

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

Mora, C., Tittensor, D.P. & Myers, R.A. (2008) The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **275**, 149–155.

Moreno-Amat, E., Mateo, R.G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.C. & García-Amorena, I. (2015) Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling*, **312**, 308–317.

Morin, X., Augspurger, C. & Chuine, I. (2007) Process-Based Modeling of Species' Distributions: What Limits Temperate Tree Species' Range Boundaries? *Ecology*, **88**, 2280–2291.

Morin, X. & Thuiller, W. (2009) Comparing Niche- and Process-Based Models to Reduce Prediction Uncertainty in Species Range Shifts under Climate Change. *Ecology*, **90**, 1301–1313.

Morlon, H., Parsons, T. & Plotkin, J. (2011) Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences*, **108**, 16327–16332.

Morlon, H., Schwilk, D.W., Bryant, J.A., Marquet, P.A., Rebelo, A.G., Tauss, C., Bohannan, B.J.M. & Green, J.L. (2010) Spatial patterns of phylogenetic diversity. *Ecology Letters*, **14**, 141–149.

Mulcahy, D.G., Noonan, B.P., Moss, T., Townsend, T.M., Reeder, T.W., Sites Jr., J.W. & Wiens, J.J. (2012) Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Molecular Phylogenetics and Evolution*, **65**, 974–991.

Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M. & Anderson, R.P. (2014) ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxentecological niche models. *Methods in Ecology and Evolution*, **5**, 1198–1205.

Near, T.J., Bolnick, D.I. & Wainwright, P.C. (2005) Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution*, **59**, 1768–1782.

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

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

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

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

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

Nosil, P. & Mooers, A.O. (2005) Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution*, **59**, 2256–2263.

O'Grady, S.M. & DeVries, A.L. (1982) Osmotic and ionic regulation in polar fishes. *Journal of Experimental Marine Biology and Ecology*, **57**, 219–228.

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

Oliver, P.M. & Sanders, K.L. (2009) Molecular evidence for Gondwanan origins of multiple lineages within a diverse Australasian gecko radiation. *Journal of Biogeography*, **36**, 2044–2055.

Olson, D., Dinerstein, E., Wikramanayake, E., Burgess, N., Powell, G., Underwood, E., D'amico, J., Itoua, I., Strand, H. & Morrison, J. (2001) Terrestrial ecoregions of the world: a new map of life on earth. *BioScience*, **51**, 933–938.

Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C.M., Myers, C.E. & Peterson, A.T. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, **263**, 10–18.

Penone, C., Weinstein, B.G., Graham, C.H., Brooks, T.M., Rondinini, C., Hedges, S.B., Davidson, A.D. & Costa, G.C. (2016) Global mammal beta diversity shows parallel assemblage structure in similar but isolated environments. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **283**, 20161028.

Peterson, A.T. & Nakazawa, Y. (2008) Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. *Global Ecology and Biogeography*, **17**, 135–144.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. Extinction Risk from Climate Change. Princeton University Press.

Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pineda, E. & Lobo, J.M. (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182–190.

Portner, H.O. (2002) Climate variations and the physiological basis of temperature dependent biogeography: systemic to molecular hierarchy of thermal tolerance in animals. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **132**, 739–761.

Pradervand, J.N., Dubuis, A., Pellissier, L., Guisan, A. & Randin, C. (2014) Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography*, **38**, 79–96.

Price, S.A., Hopkins, S.S.B., Smith, K.K. & Roth, V.L. (2012) Tempo of trophic evolution and its impact on mammalian diversification. *Proceedings of the National Academy of Sciences*, **109**, 7008–7012.

Priede, I.G. & Froese, R. (2013) Colonization of the deep sea by fishes. *Journal of Fish Biology*, **83**, 1528–1550.

Priede, I.G. (2017) *Deep-sea fishes: biology, diversity, ecology and fisheries*. Cambridge University Press.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M. & Lemmon, A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, **526**, 569–573.

Pulido-Santacruz, P. & Weir, J.T. (2016) Extinction as a driver of avian latitudinal diversity gradients. *Evolution*, **70**, 860–872.

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

Pyron, R.A. (2011) Divergence Time Estimation Using Fossils as Terminal Taxa and the Origins of Lissamphibia. *Systematic Biology*, **60**, 466–481.

Pyron, R.A. (2014) Temperate extinction in squamate reptiles and the roots of latitudinal diversity gradients. *Global Ecology and Biogeography*, **23**, 1126–1134.

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

Pyron, R.A., Burbrink, F.T. & Wiens, J.J. (2013) A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, **13**, 93.

Pyron, R.A. & Wiens, J.J. (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, **61**, 543–583.

Pyron, R.A. & Wiens, J.J. (2013) Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **280**, 20131622.

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

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

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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

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

Rabosky, D.L. (2017) How to make any method "fail": BAMM at the kangaroo court of false equivalency. *arXivorg*.

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*, pp. 1–20.

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

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

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

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

Rabosky, D.L. & Hurlbert, A.H. (2015) Species Richness at Continental Scales Is Dominated by Ecological Limits. *The American Naturalist*, **185**, 572–583.

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

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.

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**, 1958.

Rabosky, D.L., Slater, G.J. & Alfaro, M.E. (2012) Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biology*, **10**, e1001381.

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

Radosavljevic, A. & Anderson, R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, **41**, 629–643.

Ramirez-Villegas, J. & Jarvis, A. (2010) *Downscaling global circulation model outputs: the delta method decision and policy analysis Working Paper No. 1*, volume 1, pp. 1–18. International Center for Tropical Agriculture (CIAT).

Rawlings, L.H., Rabosky, D.L., Donnellan, S.C. & Hutchinson, M.N. (2008) Python phylogenetics: inference from morphology and mitochondrial DNA. *Biological Journal of the Linnean Society*, **93**, 603–619.

Ready, J., Kaschner, K., South, A.B., Eastwood, P.D., Rees, T., Rius, J., Agbayani, E., Kullander, S. & Froese, R. (2010) Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*, **221**, 467–478.

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

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.

Reeder, T.W., Townsend, T.M., Mulcahy, D.G., Noonan, B.P., Wood Jr, P.L., Sites Jr, J.W. & Wiens, J.J. (2015) Integrated Analyses Resolve Conflicts over Squamate Reptile Phylogeny and Reveal Unexpected Placements for Fossil Taxa. *PLoS ONE*, **10**, e0118199.

Reynolds, R.G., Niemiller, M.L. & Revell, L.J. (2014) Toward a Tree-of-Life for the boas and pythons: Multilocus species-level phylogeny with unprecedented taxon sampling . *Molecular Phylogenetics and Evolution*, **71**, 201–213.

Ricklefs, R.E. (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.

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

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

Rivas-Martínez, S. & Rivas-Sáenz, S. (2009) Synoptical worldwide bioclimatic classification system. `www.globalbioclimatics.org`. Accessed 15 February 2016.

Robertson, T., Döring, M., Guralnick, R.P., Bloom, D., Wieczorek, J., Braak, K., Otegui, J., Russell, L. & Desmet, P. (2014) The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. *PLoS ONE*, **9**, e102623.

Robinson, D.F. & Foulds, L.R. (1981) Comparison of phylogenetic trees. *Mathematical Biosciences*, **53**, 131–147.

Rodda, G.H., Jarnevich, C.S. & Reed, R.N. (2011) Challenges in Identifying Sites Climatically Matched to the Native Ranges of Animal Invaders. *PLoS ONE*, **6**, e14670.

Rödder, D., Schmidtlein, S., Veith, M. & Lötters, S. (2009) Alien Invasive Slider Turtle in Unpredicted Habitat: A Matter of Niche Shift or of Predictors Studied? *PLoS ONE*, **4**, e7843–9.

Rogers, A.D. (2015) Environmental Change in the Deep Ocean. *Annual Review of Environment and Resources*, **40**, 1–38.

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

Rolland, J., Condamine, F.L., Jiguet, F. & Morlon, H. (2014) Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLoS Biology*, **12**, e1001775.

Rolland, J. & Salamin, N. (2016) Niche width impacts vertebrate diversification. *Global Ecology and Biogeography*, **25**, 1252–1263.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L. & Rasnitsyn, A.P. (2012a) A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. *Systematic Biology*, **61**, 973–999.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. (2012b) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, **61**, 539–542.

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

Roy, K. & Goldberg, E.E. (2007) Origination, extinction, and dispersal: Integrative models for understanding present-day diversity gradients. *American Naturalist*, pp. S71–85. University of California, San Diego, San Diego, United States.

Sanders, K.L., Lee, M.S.Y., Leys, R., Foster, R. & Scott Keogh, J. (2008) Molecular phylogeny and divergence dates for Australasian elapids and sea snakes (hydrophiinae): evidence from seven genes for rapid evolutionary radiations. *Journal of Evolutionary Biology*, **21**, 682–695.

Sanders, K.L. & Lee, M.S.Y. (2008) Molecular evidence for a rapid late-Miocene radiation of Australasian venomous snakes (Elapidae, Colubroidea). *Molecular Phylogenetics and Evolution*, **46**, 1165–1173.

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

Sanderson, M.J., McMahon, M.M., Stamatakis, A., Zwickl, D.J. & Steel, M. (2015) Impacts of Terraces on Phylogenetic Inference. *Systematic Biology*, **64**, 709–726.

Sanderson, M.J., Purvis, A. & Henze, C. (1998) Phylogenetic supertrees: assembling the trees of life. *Trends in Ecology and Evolution*, **13**, 105–109.

Sankoff, D. (1975) Minimal Mutation Trees of Sequences. *Siam Journal on Applied Mathematics*, **28**, 35–42.

Saupe, E.E., Barve, V., Myers, C.E., Soberón, J., Barve, N., Hensz, C.M., Peterson, A.T., Owens, H.L. & Lira-Noriega, A. (2012) Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling*, **237-238**, 11–22.

Sayre, R., Comer, P., Warner, H. & Cress, J. (2009) *A new map of standardized terrestrial ecosystems of the conterminous United States:* US Geological Survey Professional Paper 1768. Technical report, US Geological Survey.

Scharf, I., Feldman, A., Novosolov, M., Pincheira-Donoso, D., Das, I., Böhm, M., Uetz, P., Torres-Carvajal, O., Bauer, A., Roll, U. & Meiri, S. (2014) Late bloomers and baby boomers: ecological drivers of longevity in squamates and the tuatara. *Global Ecology and Biogeography*, **24**, 396–405.

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

Schoener, T.W. (1968) The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna. *Ecology*, **49**, 704–726.

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

Simões, T.R., Caldwell, M.W., Tałanda, M., Bernardi, M., Palci, A., Vernygora, O., Bernardini, F., Mancini, L. & Nydam, R.L. (2018) The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. *Nature*, **557**, 706–709.

Siqueira, A.C., Oliveira-Santos, L.G.R., Cowman, P.F. & Floeter, S.R. (2016) Evolutionary processes underlying latitudinal differences in reef fish biodiversity. *Global Ecology and Biogeography*, **25**, 1466–1476.

Sistrom, M., Hutchinson, M., Bertozzi, T. & Donnellan, S.C. (2014) Evaluating evolutionary history in the face of high gene tree discordance in Australian Gehyra (Reptilia: Gekkonidae). *Heredity*, **113**, 52–63.

Skinner, A., Hugall, A.F. & Hutchinson, M.N. (2011) Lygosomine phylogeny and the origins of Australian scincid lizards. *Journal of Biogeography*, **38**, 1044–1058.

Smith, B.T., Bryson Jr., R.W., Houston, D.D. & Klicka, J. (2012) An asymmetry in niche conservatism contributes to the latitudinal species diversity gradient in New World vertebrates. *Ecology Letters*, **15**, 1318–1325.

Smith, S.A., Beaulieu, J.M. & Donoghue, M.J. (2010) An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences*, **107**, 5897–5902.

Smith, S.A., Beaulieu, J.M. & Donoghue, M.J. (2009) Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology*, **9**, 37.

Smith, S.A. & Brown, J.W. (2018) Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, **66**, 152–13.

Smith, S.A., Brown, J.W., Yang, Y., Bruenn, R., Drummond, C.P., Brockington, S.F., Walker, J.F., Last, N., Douglas, N.A. & Moore, M.J. (2018) Disparity, diversity, and duplications in the Caryophyllales. *New Phytologist*, **217**, 836–854.

Smith, S.A. & O'Meara, B.C. (2012) treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, **28**, 2689–2690.

Somero, G.N. & DeVries, A.L. (1967) Temperature tolerance of some Antarctic fishes. *Science*, **156**, 257–258.

Spalding, M.D., Fox, H.E., Allen, G.R. & Davidson, N. (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, **57**, 573–583.

Spalding, M.D., Agostini, V.N., Rice, J. & Grant, S.M. (2012) Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. *Ocean and Coastal Management*, **60**, 19–30.

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

Stadler, T., Rabosky, D.L., Ricklefs, R.E. & Bokma, F. (2014) On Age and Species Richness of Higher Taxa. *The American Naturalist*, **184**, 447–455.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stanton, J.C., Pearson, R.G., Horning, N., Ersts, P. & Reşit Akçakaya, H. (2012) Combining static and dynamic variables in species distribution models under climate change. *Methods in Ecology and Evolution*, **3**, 349–357.

Stephens, P.R. & Wiens, J.J. (2003) Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. *The American Naturalist*, **161**, 112–128.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T. & Roeckner, E. (2013) Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, **5**, 146–172.

Stuart-Smith, R.D., Bates, A.E., Lefcheck, J.S., Duffy, J.E., Baker, S.C., Thomson, R.J., Stuart-Smith, J.F., Hill, N.A., Kininmonth, S.J., Airoldi, L., Becerro, M.A., Campbell, S.J., Dawson, T.P., Navarrete, S.A., Soler, G.A., Strain, E.M.A., Willis, T.J. & Edgar, G.J. (2013) Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature*, **501**, 539–542.

Svenning, J.C., Fløjgaard, C., Marske, K.A., Nogues-Bravo, D. & Normand, S. (2011) Applications of species distribution modeling to paleobiology. *Quaternary Science Reviews*, **30**, 2930–2947.

Swenson, N.G., Erickson, D.L., Mi, X., Bourg, N.A., Forero-Montaña, J., Ge, X., Howe, R., Lake, J.K., Liu, X., Ma, K., Pei, N., Thompson, J., Uriarte, M., Wolf, A., Wright, S.J., Ye, W., Zhang, J., Zimmerman, J.K. & Kress, W.J. (2012) Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. *Ecology*, **93**, S112–S125.

Synes, N.W. & Osborne, P.E. (2011) Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecology and Biogeography*, **20**, 904–914.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B. & Harmon, L.J. (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist*, **207**, 454–467.

Thomas, G.H., Hartmann, K., Jetz, W., Joy, J.B., Mimoto, A. & Mooers, A.O. (2013) PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences. *Methods in Ecology and Evolution*, **4**, 1011–1017.

Thomson, R.C. & Shaffer, H.B. (2009) Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Systematic Biology*, **59**, 42–58.

Thornthwaite, C.W. (1948) An approach toward a rational classification of climate. *Geographical Review*, **38**, 55–94.

Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2020–2027.

Thuiller, W., Richardson, D.M., Pyšek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.

Title, P.O. & Bemmels, J.B. (2017) ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, **41**, 291–307.

Title, P.O. & Burns, K.J. (2015) Rates of climatic niche evolution are correlated with species richness in a large and ecologically diverse radiation of songbirds. *Ecology Letters*, **18**, 433–440.

Title, P.O. & Rabosky, D.L. (2017) Do Macrophylogenies Yield Stable Macroevolutionary Inferences? An Example from Squamate Reptiles. *Systematic Biology*, **66**, 843–856.

Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V. & Worm, B. (2010) Global patterns and predictors of marine biodiversity across taxa. *Nature*, **466**, 1098–1101.

Tonini, J.F.R., Beard, K.H., Ferreira, R.B., Jetz, W. & Pyron, R.A. (2016) Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, **204**, 23–31.

Tuanmu, M.N. & Jetz, W. (2015) A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, **24**, 1329–1339.

Tucker, C.M., Cadotte, M.W., Carvalho, S.B., Davies, T.J., Ferrier, S., Fritz, S.A., Grenyer, R., Helmus, M.R., Jin, L.S., Mooers, A.O., Pavoine, S., Purschke, O., Redding, D.W., Rosauer, D.F., Winter, M. & Mazel, F. (2017) A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, **92**, 698–715.

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.

Uetz, P. & Hošek, J. (2015) The reptile database. `http://www.reptile-database.org`. Accessed 15 September 2015.

Varadhan, R., University, J.H., Borchers, H.W. & Research., A.C. (2018) *dfoptim: Derivative-Free Optimization*. R package version 2018.2-1.

Vecchione, M., Falkenhaug, T., Sutton, T., Cook, A., Gislason, A., Hansen, H.Ø., Heino, M., Miller, P.I., Piatkowski, U., Porteiro, F., Søiland, H. & Bergstad, O.A. (2015) The effect of the North Atlantic Subpolar Front as a boundary in pelagic biogeography decreases with increasing depth and organism size. *Progress in Oceanography*, **138**, 105–115.

Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290–2299.

Vidal, N., Marin, J., Sassi, J., Battistuzzi, F.U., Donnellan, S.C., Fitch, A.J., Fry, B.G., Vonk, F.J., Rodriguez de la Vega, R.C., Couloux, A. & Hedges, S.B. (2012) Molecular evidence for an Asian origin of monitor lizards followed by Tertiary dispersals to Africa and Australasia. *Biology Letters*, **8**, 853–855.

Vörösmarty, C.J., Douglas, E.M., Green, P.A. & Revenga, C. (2005) Geospatial Indicators of Emerging Water Stress: An Application to Africa. *Ambio*, **34**, 230–236.

Waltari, E., Hijmans, R.J., Peterson, A.T., Nyári, Á.S., Perkins, S.L. & Guralnick, R.P. (2007) Locating Pleistocene Refugia: Comparing Phylogeographic and Ecological Niche Model Predictions. *PLoS ONE*, **2**, e563.

Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. (2016) Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. *PLoS ONE*, **11**, e0156720–17.

Wang, T., Hamann, A., Spittlehouse, D.L. & Murdock, T.Q. (2012) ClimateWNA—High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, **51**, 16–29.

Warnock, R.C., Parham, J.F., Joyce, W.G., Lyson, T.R. & Donoghue, P.C. (2014) Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **282**, 20141013.

Warren, D.L., Glor, R.E. & Turelli, M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.

Warren, D.L. & Seifert, S. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.

Warren, D.L., Wright, A.N., Seifert, S.N. & Shaffer, H.B. (2014) Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Diversity and Distributions*, **20**, 334–343.

Webb, C.O., Ackerly, D., McPeek, M.A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

Weinstein, B.G., Tinoco, B., Parra, J.L., Brown, L.M., McGuire, J.A., Stiles, F.G. & Graham, C.H. (2014) Taxonomic, Phylogenetic, and Trait Beta Diversity in South American Hummingbirds. *The American Naturalist*, **184**, 211–224.

Weir, J.T. (2014) Environmental harshness, latitude and incipient speciation. *Molecular Ecology*, **23**, 251–253.

Weir, J.T. & Schluter, D. (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**, 1574–1576.

Wieczorek, J., Bloom, D., Guralnick, R.P., Blum, S., Döring, M., Giovanni, R., Robertson, T. & Vieglais, D. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*, **7**, e29715.

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.

Wiens, J.J. & Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, **19**, 639–644.

Wiens, J.J., Hutter, C.R., Mulcahy, D.G., Noonan, B.P., Townsend, T.M., Sites Jr, J.W. & Reeder, T.W. (2012) Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biology Letters*, **8**, 1043–1046.

Willmott, C. & Feddema, J. (1992) A More Rational Climatic Moisture Index. *The Professional Geographer*, **44**, 84–88.

Wilson, A.M. & Jetz, W. (2016) Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. *PLoS Biology*, **14**, e1002415–20.

Wilson, M.F.J., O'Connell, B., Brown, C., Guinan, J.C. & Grehan, A.J. (2007) Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. *Marine Geodesy*, **30**, 3–35.

Wisz, M.S., Broennimann, O., Gronkjaer, P., Moller, P.R., Olsen, S.M., Swingedouw, D., Hedeholm, R.B., Nielsen, E.E., Guisan, A. & Pellissier, L. (2015) Arctic warming will promote Atlantic–Pacific fish interchange. *Nature Climate Change*, **5**, 261–265.

Wright, A.N., Hijmans, R.J., Schwartz, M.W. & Shaffer, H.B. (2014) Multiple sources of uncertainty affect metrics for ranking conservation risk under climate change. *Diversity and Distributions*, **21**, 111–122.

Wright, A.M., Lyons, K.M., Brandley, M.C. & Hillis, D.M. (2015) Which came first: The lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **324**, 504–516.

Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G., McGlinn, D.J., O'Meara, B.C., Moles, A.T., Reich, P.B., Royer, D.L., Soltis, D.E., Stevens, P.F., Westoby, M., Wright, I.J., Aarssen, L., Bertin, R.I., Calaminus, A., Govaerts, R., Hemmings, F., Leishman, M.R., Oleksyn, J., Soltis, P.S., Swenson, N.G., Warman, L. & Beaulieu, J.M. (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature*, **506**, 89–92.

Zheng, Y. & Wiens, J.J. (2016) Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, **94**, 537–547.

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

Zomer, R., Trabucco, A., van Straaten, O. & Bossio, D. (2006) *Carbon, land and water: A global analysis of the hydrologic dimensions of climate change mitigation through afforestation/reforestation*, volume 101. IWMI.

Zomer, R.J., Trabucco, A., Bossio, D.A. & Verchot, L.V. (2008) Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems and Environment*, **126**, 67–80.

Zwickl, D.J. (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.* Ph.D. thesis, The University of Texas at Austin.