

THE UNIVERSITY OF CHICAGO

REMODELING THE FOSSIL RECORD: ANALYSIS OF EMERGENT EVOLUTIONARY  
AND ECOLOGICAL PATTERNS

A DISSERTATION SUBMITTED TO  
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES DIVISION  
IN CANDIDACY FOR THE DEGREE OF  
PH.D.

DEPARTMENT OF EVOLUTIONARY BIOLOGY

BY  
PETER DAVID SMITS

CHICAGO, ILLINOIS  
GRADUATION DATE

Copyright © 2017 by Peter David Smits

All Rights Reserved

Dedication Text

Epigraph Text

## TABLE OF CONTENTS

LIST OF FIGURES . . . . .	vii
LIST OF TABLES . . . . .	viii
ACKNOWLEDGMENTS . . . . .	ix
ABSTRACT . . . . .	x
1 INTRODUCTION . . . . .	1
1.1 Emergent patterns, macroevolution, and macroecology . . . . .	1
1.2 Structured data and modelling emergent patterns . . . . .	4
1.3 Study summaries . . . . .	8
2 EXPECTED TIME-INVARIANT DIFFERENCES IN MAMMAL SPECIES DURATION . . . . .	10
2.1 Introduction . . . . .	10
2.2 Results . . . . .	13
2.3 Discussion . . . . .	15
2.4 Materials and Methods . . . . .	17
2.4.1 Species occurrence and covariate information . . . . .	17
2.4.2 Survival model . . . . .	19
2.4.3 Estimation . . . . .	24
2.4.4 Posterior evaluation . . . . .	25
2.5 Supplemental information for “Death and taxa” . . . . .	26
2.5.1 Supertree inference . . . . .	26
2.5.2 Modeling censored observations . . . . .	27
2.5.3 Deviance residuals . . . . .	28
2.5.4 Variance partitioning . . . . .	29
2.5.5 Widely applicable information criterion . . . . .	30
2.5.6 Results from posterior predictive checks . . . . .	31
2.5.7 Data quality concerns . . . . .	32
2.5.8 Concerns surrounding estimates of $\alpha$ . . . . .	32
3 HOW MACROECOLOGY AFFECTS MACROEVOLUTION: THE INTERPLAY BETWEEN EXTINCTION INTENSITY AND TRAIT-DEPENDENT EXTINCTION IN BRACHIOPODS . . . . .	42
3.1 Introduction . . . . .	43
3.1.1 Factors affecting brachiopod survival . . . . .	44
3.2 Materials and Methods . . . . .	47
3.2.1 Fossil occurrence information . . . . .	47
3.2.2 Details of model . . . . .	50
3.2.3 Imputation of sampling probability . . . . .	53
3.2.4 Posterior inference and posterior predictive checks . . . . .	54

3.3 Results . . . . .	55
3.4 Discussion . . . . .	59
4 TAXON OCCURRENCE AS A FUNCTION OF BOTH EMERGENT BIOLOGICAL TRAITS AND ITS ENVIRONMENTAL CONTEXT . . . . .	73
5 CONCLUSION . . . . .	74
REFERENCES . . . . .	75

## LIST OF FIGURES

2.1	Posterior predictive checks . . . . .	34
2.2	Effect of mammal ecotypes on survival . . . . .	35
2.3	Effect of cohort on mammal survival . . . . .	36
2.4	Partitioned variance for mammal survival . . . . .	36
2.5	Deviance residuals of fitted model . . . . .	37
2.6	Additional posterior predictive checks . . . . .	38
2.7	Estimate of hazard function for mammal survival . . . . .	39
2.8	Simulation of sample size and estimates of $\alpha$ . . . . .	40
3.1	Posterior predictive check of survival . . . . .	66
3.2	Posterior predictive check of congruence . . . . .	67
3.3	Estimated relationship between environment and survival . . . . .	68
3.4	Cohort specific relationship between environment and survival . . . . .	69
3.5	Effect of environmental covariates on survival . . . . .	70
3.6	Correlation matrix of effects of environmental covariates . . . . .	71
3.7	Example imputed gap statistic distributions . . . . .	72

## LIST OF TABLES

2.1	Posterior estimates of covariates on mammal survival . . . . .	34
2.2	Cypher for ecotype assignments . . . . .	37
2.3	Equations used to estimate mammal mass . . . . .	41
3.1	Parameter estimates for brachiopod survival model . . . . .	65

## **ACKNOWLEDGMENTS**

## **ABSTRACT**

# CHAPTER 1

## INTRODUCTION

Species traits are the bridge between evolution and ecology [104? ]. A trait is an identifiable property of an organism, such as individual body size, while a species trait is some identifiable property of the entire species, such as the average body size or geographic range of a species [104]. A class of species traits called functional traits are those traits which clearly describe a species means of interacting with their environment such as leaf surface area or trophic role [104]. In macroevolutionary studies, analyses are typically focused on a patterns associated with a single single trait or are instead traitless a analysis of diversity and the diversification process [163, 162, 123, 134, 172, 67, 68, 91, 117]. Macroecological studies are frequently concerned with describing the distribution of species and individuals over space or time, such as shifts in community composition along some gradient or axis [174, 173, 20, 19, 21, 29, 43, 80, 81]. My desire with this dissertation is present the types of analyses and results that are possible through a synthesis of both macroevolution and macroecology; my approach is to develop inference devices (i.e. statistical models) to better understand the interactions between the effects of multiple species traits, as well as those of a species' temporal and environmental context, on diversity and differential diversification both in space and time.

### 1.1 Emergent patterns, macroevolution, and macroecology

An emergent pattern is one that is not observable or predictable from its individual constituent parts. Emergence is ubiquitous in biological systems: cells form a tissue with a complex function, species extinction requires all individual members to die for possibly unrelated reasons, and a species global geographic range is the product of many individual ranges. The

history of a species, or set of species, over time is inherently an emergent pattern as the temporal history of a species is not knowable from an instantaneous sample. Macroevolution and macroecology are the studies of emergent patterns in evolutionary and ecological data, respectively [19, 20, 182, 181]. Traditionally, macroevolution is the study of patterns over time while macroecology is concerned with patterns over space, but I find this division overly reductive.

Both macroevolution and macroecology are disciplines concerned with emergent patterns; they both implicitly and explicitly accept a hierarchical perspective on biology as emergence is not possible without different levels of organization, however they are defined. Even if an analysis is concerned with only a single species, the concept of a species is itself an emergent property or label for a collection of populations and individuals which all share a common evolutionary history. While it may be argued that the species label or identity is a non-biological construct or is simply heuristic for understanding the complexity of populations and reproductive isolation, we are still concerned with patterns associated with that construct as well as its intrinsic properties (e.g. extinction, conservation, ecosystem services) [24, 74].

Macroevolution is much more than evolution above the species level [38]; this too is overly reductive and assigns too much meaning to a single level of organization rather than embracing the multitudes of possible levels of organisation. Instead, I propose defining macroevolution, *qua* field of study, as the study of emergent evolutionary patterns; this means patterns of speciation/extinction (diversity) as well as trait evolution (disparity), both of which are only observable when considering more than one taxon or when considering the temporal history of one or more taxa. In complement, macroecology is then the study of emergent ecological patterns, which means patterns in spatial distribution or community composition which are observable only when considering more than one taxon or when considering the temporal history of one or more taxa [19, 20, 174]. Because taxa inherently respond differently and

individually to environmental changes, both biotic and abiotic, macroecological patterns are those due to the similarity in response across individuals [15].

Species selection is enshrined as one of the most important patterns in macroevolution [181, 182, 190, 74, 133, 165, 119]. (**author?**) [133] portray species selection as resultant phenomenon of the heritability of speciation and extinction rates, a powerful and useful expansion of which phenomenon fall under this category by divorcing levels of organization from the definition and thus avoiding the unproductive species selection versus species sorting debate which has been the cause for a considerable amount of confusion and rhetorical fights [189, 190, 98, 119], not the least of which is the adoption of “sorting” as an important term for understanding community assembly CITATION. However, I think the definition presented by (**author?**) [133] misses the mark as an operational definition to inspire and guide future study. Species inherit more than just speciation and extinction rates; they also inherit traits which themselves may be linked to differences in speciation or extinction rates due to their effects on species fitness. Species fitness is a concept that is rarely discussed and difficult to define yet is vitally important to understanding species selection as process [23, 116]. Here I adopt the inclusive definition of fitness presented in (**author?**) [23]: species fitness is, minimally, the expected duration of that species. Differences in expected species duration that are associated with species traits are then the product of (species) selection.

A generative definition of species fitness (i.e. relating to speciation rate) is more difficult to develop for a variety of reasons, not the least of which being that the why and how speciation rates can vary both across species and time is not well understood [131, 132, 24]. At a minimum, given the earlier definition of fitness wrt duration, species fitness wrt speciation would require the association of differences speciation rate with one or more species traits; this difference in fitness is then the raw material for selection.

Extinction is a property of, or a phenomenon affecting, species as it requires the death of

all organisms within a species which do not have to all occur for the same reason [165]. Extinction is a fundamentally emergent phenomenon that is the ultimate manifestation of selection; it is also central to macroevolutionary studies and the definition of (species) fitness used here [23]. Extinction is featured centrally in one of the few “laws” in macroevolution and paleobiology: the Law of Constant Extinction [184, 97]. This law states that a species risk of going extinct is independent of that species age [184, 97], a conclusion reached via analysis of patterns of (higher) taxon survival patterns. The Red Queen hypothesis was proposed as a process that would result in the observation underlying the Law of Constant Extinction [184], though it has obviously grown to have a life of its own [97].

The functional composition of a community or species pool is a property of that unit; observing a single species at a locality does not reveal the functional composition of the community in which it interacts. The composition of a community or species pool in terms of functional groups is a community ecology exercise. Comparing the distribution of functional groups across communities or species pools is where community ecology and macroecology intersect [104, 20? ]. In paleobiology, a successful means of classifying marine invertebrate functional groups has been a three dimensional classification scheme called an “ecocube” which uses consistently identifiable functional traits to label both possible and observed functional groups [21, 9]. It is this strategy that inspires the third study presented in this dissertation.

## 1.2 Structured data and modelling emergent patterns

An inference device is a theoretical tool for improving our knowledge by processing new information and observations [103] CITATION; this device has initial conditions describing what we know (e.g. nothing), mechanisms for updating this knowledge to reflect new information, and can then produce an updated “picture” that better represents our current

knowledge as well as the uncertainty surrounding this knowledge. Each inference device has a specific and narrow purpose and functionality [103] CITATION; unless the mechanisms are similar, a device for processing the rate of imperfections in the manufacturing of widgets cannot process the queuing times of callers to a help line.

We can think of the well known battery of statistical tests CITATION BIOMETRY as re-usable inference devices with very narrow utility; these are unmodifiable tools for handling very specific questions and data. All (Bayesian statistical) models act as inference devices because they fulfill the requirements described above: initial conditions, updating mechanism, and output as updated knowledge along with the uncertainty surrounding that knowledge [103] JAYNES. By developing a new model for each new question there is a precision of translation; the model actually reflects the questions at hand, something that is preferable to forcing questions and data to fit into pre-made inference devices (e.g. models, tests) that do not update knowledge in a means relevant to the actual question(s) of interest.

Structure occurs naturally in the collection of data. For example, imagine a drug trial that takes place across multiple hospitals. It is possible to consider the results from all hospitals in aggregate by ignoring the hospital labels; alternatively, these results can be considered individually by hospital. In a biological example, imagine the study of individuals within a single species that are collected from multiple locations. For many reasons, we might expect that individuals from the same location are more similar to each other than to individuals from other locations. The goal in the analysis of structured data through hierarchical or multi-level models is to leverage this structure into the analysis in order to improve estimation by having groupings share information about associated parameter estimates [51, 103].

Two of the most important analytical approaches at the core of macroevolutionary study are the birth-death process for diversification [143, 139, 111, 113, 110, 112, 179], and the random walk heuristic for trait evolution [142, 31, 16, 53, 146, 147, 148, 159, 68, 67]. Discrete-state

Markov processes also feature when modeling the evolution of discrete traits and biogeography CITATIONS. All three of the studies covered by this dissertation make use of some variant to the birth-death process. The first two studies are analyses of extinction, which is a pure-death processes. The third study utilizes a discrete-time birth-death process to model species presence in a species pool.

Similar analytical cores are harder to identify for macroecology as a whole, so instead I will focus on species distribution models (SDMs) as a powerful framework for understanding the distribution of species, both due to environmental factors [28] and species traits [160]. SDMs are a class of models which attempt to operationalize the multitude of processes which result in the distribution of one or more species in both space and/or time. In effect, SDMs are a means of operationalizing the concept of a species’ “realized niche” in order to understand the limits on a species distribution [28]. Typically, SDMs are used to analyze the relationship between species presence at a locality and the environmental factors which characterize that locality. From this analysis, the possible distribution of a species in space can be then estimated and compared to the observed distribution of that species [28, 8, 122]. The maximum entropy theory of community assembly, and its related model, view community assembly as an ecological sorting process where traits mediate the effects of environmental filters [160, 196]; also called “community assembly via trait selection” (CATS regression). By analyzing the composition of species at localities based on their traits, the strength and relative importance of which traits most strongly structure community composition can be elicited.

Both of these approaches to analyzing species distributions can be united in a single fourth-corner model [196, 18]. The fourth-corner problem is an old problem in community ecology originating in the multivariate analysis literature: assuming species distribution is the result of functional traits interacting with environmental factors, how do we estimate which interactions

are important and their relative strengths [88, 26]? By phrasing the fourth-corner problem as a model based framework, results are much more easily interpretable and actually provide estimates of the effects of species traits and environmental factors instead of the simple significance provided from the older Monte Carlo based methodology [18, 76, 125, 124].

The third study in this dissertation makes extensive use of this framework by casting the fourth-corner problem into an additional dimension: time. By combining the fourth-corner framework outlined above [196, 18] with the birth-death process used for modeling diversification into a single unified model of species occurrence through time as a function of both species traits and changing environmental context I've developed a powerful analytical bridge between macroecology and macroevolution.

I emphasize model-based approaches to analysis as well as question or study specific models because a common language is necessary for clear, coherent, and translatable results that actually relate to the question(s) at hand. Some of the greatest limits to paleobiological, macroevolutionary, and macroecological study are a lack of strong, mechanistic predictive theories that can be expressed mathematically. Some of the greatest strides in advancing discussions of macroevolutionary and macroecological theory disputes have come from translating verbal theory into mathematical and statistical models [143, 111, 31, 68? , 160]. The reality of how complex biological process is rarely integrated into paleobiological analyses of macroevolutionary and macroecological patterns. A move to a model-forward approach to paleobiology, heavily steeped in evolutionary and ecological theory, would be beneficial for advancement of theories in macroevolution and macroecology.

Paleobiologists historically believe that neontologists ignore their approaches and insights into macroevolution and macroecology study and theory CITATION, but without a concerted effort to engage within the same theoretical framework and language when developing scientific questions and the related analytical tools (i.e. statistical models) then this worry

and resentment is all but preordained. Because the systems paleobiologists study are unknown to, have no direct impact on, or are contextualized wrt the systems studied by neontologists, a push towards unification and synthesis most likely has to begin with the paleobiological community; luckily, it appears that the neontological community is receptive paleobiological insight CITATION. The simplest and fastest way to begin not just a dialogue but a unification is by translating verbal macroevolutionary and macroecological theories from paleobiology into statistical models that are readable by all researchers, both paleontological and neontological.

### 1.3 Study summaries

Each of the chapters that make up this dissertation feature structured data and a hierarchical model describing this structure. The first two chapters have a decidedly macroevolutionary bent while the third has a macroecological bent. However, all three are informed by both evolution and ecology; after all, the focus is of all studies is the relationship between functional traits to emergent patterns.

The first chapter is an analysis of North American mammal species duration wrt multiple species traits, origination cohort, and shared evolutionary history (i.e. phylogeny). All of these factors are emergent properties of a species, with origination cohort and shared evolutionary history also being data structuring factors. Additionally, the possibility of species age affecting extinction risk is also considered as a possible test of the Law of Constant Extinction [184]. This analysis is also concerned with the long standing “survival of the unspecialized” hypothesis wrt species duration/extinction risk [168].

The second chapter is an analysis of post-Cambrian brachiopod genus duration wrt multiple species traits and origination cohort. As with the first chapter, all of these factors are emergent properties of species, and origination cohort is used as a structuring factor. Similarly, the

possibility of age-dependent extinction is allowed. What distinguishes this study from the first is the additional focus on the relationship between the selective importance of these emergent factors and the higher-level average fitness of an origination cohort.

The third chapter is an analysis of the functional composition of the North American mammal regional species pool over time with the goal of understanding when are specific ecotypes enriched or depleted and how these patterns may be associated with species' environmental context. The structuring factors in this analysis are both the mammal ecotypes and the temporal units of the discrete time-series describing species presence and absence in the regional species pool. This analysis presents a multi-level model where the properties of the structuring factor (e.g. environmental context) are included as covariates affecting the emergent ecotype-specific pattern.

## CHAPTER 2

# EXPECTED TIME-INVARIANT DIFFERENCES IN MAMMAL SPECIES DURATION

Determining which biological traits influence differences in extinction risk is vital for understanding the differential diversification of life and for making predictions about species' vulnerability to anthropogenic impacts. Here I present a hierarchical Bayesian survival model of North American Cenozoic mammal species durations in relation to species-level ecological factors, time of origination, and phylogenetic relationships. I find support for the "survival of the unspecialized" as a time-invariant generalization of trait-based extinction risk. Furthermore, I find that phylogenetic and temporal effects are both substantial factors associated with differences in species durations. Finally, I find that the estimated effects of these factors are partially incongruous with how these factors are correlated with extinction risk of the extant species. This parallels previous observations that background extinction is a poor predictor of mass extinction events and suggests that attention should be focused on mass extinctions to gain insight into modern species loss.

### 2.1 Introduction

Why extinction risk varies among species remains one of the most fundamental questions in paleobiology and conservation biology [168, 184, 141, 129, 192]. To address this issue, I test for similarities in associations between extinction risk and multiple species-level traits during times of background extinction and in the modern world; which traits have time-invariant effects on species duration; and whether extinction is age-independent. I approach these questions together by using a model of species duration whose parameter estimates

act as direct tests of these questions. Cenozoic mammals are an ideal focus for this study because their fossil record is well sampled and well resolved both temporally and spatially, and because individual species ecology and taxonomic position are generally understood [168, 129, 3, 91, 173, 183, 102].

Time-invariant factors are those that have a constant directional effect even if their magnitude varies. Because change in the magnitude of extinction risk is not necessarily the best indicator of a shift from background to mass extinction [194], it is better to look for changes in either the direction of selection, the loss of a selective pressure, or the appearance of novel selective pressures [72].

The species-level traits studied here are bioprovince occupancy, body mass, and dietary and locomotor categories. These traits are related to aspects of a species' adaptive zone such as population density, expected range size, potential prey, and dispersal ability [173, 81] and are a combination of aggregate and emergent traits [74]. It is expected that species with larger geographic ranges have lower extinction rates than species with smaller geographic ranges [72, 152]; however, how traits more directly related to species–environment interactions may affect species extinction risk is more nebulous.

Body size is a complex trait related to many life history characteristics. There are three general hypotheses of how body size may effect extinction risk: 1) positive effect where an increase in body size causes an increase in extinction risk, potentially due to associated decrease in reproductive rate or other similar life history traits [91, 94]; 2) negative effect where an increase in body size causes a decrease in extinction risk because of an expected positive relationship between body size and geographic range; and 3) no effect of body size on extinction risk [183].

The strongest expectation for the effects of dietary category on extinction risk is that

omnivores will have the lowest extinction risk of all species. This expectation is based on the long standing “survival of the unspecialized” hypothesis where more generalist species (e.g. omnivores) have greater survival than specialist species (e.g. carnivores/herbivores) [168, 92]. It has also been observed that both carnivores and herbivores have greater diversification rates than omnivores, with herbivores diversifying faster than carnivores [126]. How this result translates into differences in extinction risk is currently unknown [130]. In modern taxa, higher trophic levels (e.g. carnivores versus herbivores) have been associated with greater extinction risk, most likely because of human extermination of top predators [94, 127].

Similarly, there are few expectations of how locomotor category may effect extinction risk. During the Cenozoic, there was a shift at the Paleogene/Neogene boundary from predominately closed to predominately open environments [15, 77]. Based on this observation, a prediction is that arboreal taxa will have the greatest extinction risk of all, with both scansorial and ground dwelling taxa having lower extinction risks.

I use a hierarchical Bayesian survival model of species duration as predicted by the covariates of interest along with species’ temporal and phylogenetic context. Species duration, in 2 My bins, was modeled as realizations from either an exponential or Weibull distribution-based hierarchical model [51]. The exponential distribution corresponds to the Law of Constant Extinction, which states that extinction is age-independent [184]. Note that the exponential is a special case of the Weibull when its shape parameter,  $\alpha$ , is 1. The Weibull distribution allows for extinction to be taxon-age dependent, where values of  $\alpha$  greater than 1 corresponds to increasing risk with age and values less than 1 corresponds to decreasing risk with age. Origination cohort and phylogenetic position were modeled as independent effects. Phylogenetic effect was modeled assuming species duration may have evolved via a Brownian motion-like process [100, 66]. The results from the Weibull model are detailed here because this model has a better fit to the data the exponential (Weibull WAIC 6140.37,

exponential WAIC 16697.35; Fig. 2.1, S1, S2).

## 2.2 Results

A summary of the posterior distributions for the most relevant parameter estimates is presented in Table 2.1. All posterior inference is based on these estimates. For the results from the posterior predictive checks and discussion of the estimation of  $\alpha$ , please see the accompanying Supplemental information. Additionally, See the Supplemental information for discussion surrounding use of Paleobiology Database and accompanying data quality concerns.

Species with greater bioprovence occupancy are found to be associated with lower extinction risk than taxa with smaller bioprovence occupancy ( $\beta_{occupancy}$ mean = -0.53, std = 0.08). This is consistent with previous findings. Body size has nearly zero association with expected duration ( $\beta_{size}$ mean = -0.05, std = 0.05), a similar result to some previous studies [183]. However, previous studies were performed at the generic level and were unable to determine how body size may effect species-level extinction, as the effect of either extinction or speciation cannot be distinguished [91, 183].

Some clear patterns emerge from the pairwise differences in effect of each dietary category on expected duration (Fig. 2.2). Consistent with expectations from the “survival of the unspecialized” hypothesis [92, 168], omnivory appears to be associated with the lowest expected extinction risk. Carnivory is associated a greater expected duration than either herbivory or insectivory, but a greater expected extinction risk than omnivory. Finally, herbivory and insectivory have approximately equal effects on expected duration. Given previous results, these results imply that carnivores have a greater origination rate than omnivores [126]. These results also imply that herbivores, which have the greatest extinction

risk, must also have a very high origination rate in order to have the greatest diversification rate among these three categories [126].

For locomotor category, both scansoriality and ground dwelling life habitat are associated with a greater expected duration than arboreality (Fig. 2.2). Scansorial and ground dwelling life habits also have approximately equal expected effects on extinction risk. This is consistent with the expectation that arboreality will confer greater extinction risk due to the loss of associated environment with the shift from open to closed habitat at the Paleogene/Neogene boundary [15]. However, there are two possible processes which could lead to the observed pattern: arboreality confers an intrinsic difference in extinction risk or it might not be that arboreal taxa have an intrinsically higher risk but were instead “hit harder” by the environmental shift than other taxa. This analysis cannot distinguish between these two processes. Note that, while this is a study of North American Cenozoic mammals, for European Cenozoic mammals this transitionary period corresponds to the Vallesian which was a sudden shift in species demography away from arboreality [1, 108].

Of the three sources of variance present in the model, individual species variance accounts for approximately 80% of the observed, unmodeled variance (Fig. 2.4). Note that the individual variance was approximated using an simulation approach [56] because the Weibull distribution does not have a variance term. Both cohort and phylogenetic effects account for the other 20% of the observed variance. This result means that extinction risk has both temporal and phylogenetic aspects, as both contribute greater than 0% of the observed variability in the data [66].

The estimates for the individual cohort effects show a weak pattern of greater extinction risk in older Cenozoic cohorts compared to younger cohorts (Fig. 2.3). This potential slowdown in extinction risk is consistent with previous analyses of marine invertebrates [144, 36] and mammals [4, 7]. There are two prevailing hypotheses as to the cause of this slowdown:

1) extinction risk is constant within, but varies between, clades so over time clades with low extinction rates increases in proportion of total diversity thus bringing down expected extinction risk; or 2) over time taxa increase in mean fitness and thus decrease in expected extinction risk [144]. The observed decrease in extinction risk with age, along with the variance partitioning results (Fig. 2.4) are consistent with both of these hypotheses with neither being more “important” than the other.

Interestingly, the shift from older cohorts with a higher extinction risk to younger cohorts with lower extinction risk is approximately at the Paleogene–Neogene boundary. Given the association with arboreality and increased extinction risk (Fig. 2.2), the decrease in expected extinction risk over time might relate to the preferential loss of arboreal taxa over the Cenozoic. However, because the model used here does not allow for time-varying effects, I cannot identify whether this boundary is associated with a shift in the direction or magnitude of the expected effect of arboreality on extinction risk.

## 2.3 Discussion

My results indicate that Cenozoic North American mammal “generalists” are expected to have a lower extinction risk than “specialists.” This implies that the diversification of specialized taxa would have required either a driven trend away from generality [? ] or an increase in speciation rate relative to extinction rate [181]. This requires that specialist traits should somehow increase or be associated with increases in speciation rate, perhaps via niche partitioning or changes in habitat heterogeneity. For example, descendant species of omnivores many divide available prey items more finely or arboreal taxa may increase in both extinction and speciation rates via increases in habitat heterogeneity. Possible evidence to support this hypothesis would be to demonstrate differences in speciation rate associated

with those traits analyzed here or other similar traits.

When these results are compared to factors contributing to increased extinction risk in extant mammals, there are some incongruencies. As expected, large range size is consistently associated with lower extinction risk in the modern world [94, 127, 46, 47]. While my analysis found body size to have almost no time-invariant effect on extinction risk, in extant mammals this is not necessarily the case as increased body size is associated with increased extinction risk [94, 127]. However, this pattern is partially clade dependent [46]. As stated earlier, higher trophic levels have been found to be associated with greater extinction risk in extant mammals [94, 127]. In contrast, I found that omnivores and carnivores have a lower expected extinction risk than either insectivores or herbivores (Fig. 2.2). Finally, phylogeny has been found to be a good predictor of differences in extinction risk in extant mammals as certain clades are at much higher risks than others [47]. This effect seems much greater in the Recent than for the whole Cenozoic, implying that current extinction risk is more phylogenetically concentrated than during times of background extinction levels during the Cenozoic.

Whether these incongruities are within the standard range of time-variant effects is unknown, though my comparisons do imply that current processes are different from those studied here. However, this is not a model of what makes taxa vulnerable during mass extinctions and that may account for these incongruities, assuming mass extinctions are qualitatively different than background extinction [72]. These results would also be inapplicable if the current biodiversity crisis is qualitatively different from either background or mass extinction as preserved in the fossil record.

By modeling how different ecologies and historical factors effect a species' expected extinction risk, it is possible to better understand what processes may have driven the resulting pattern of selection (i.e. diversity) while also providing a baseline for evaluating the current biodiversity crisis. This analysis finds support for the “survival of the unspecialized” hypothesis [168, 92]

as a time-invariant generalization about extinction risk. I also find that there are substantial effects of both cohort and phylogeny on extinction risk, which supports the idea that the decrease in extinction risk [144] over time has both temporal and phylogenetic components. Additionally, I found evidence of increasing extinction risk with species age, the cause of which is unknown. These results show that, like prior mass extinction events in the fossil record, the current biodiversity crisis is qualitatively different from the previous period of background extinction in the fossil record [72].

## 2.4 Materials and Methods

### 2.4.1 *Species occurrence and covariate information*

Fossil occurrence information was downloaded from the Paleobiology Database (PBDB; <http://paleodb.org/>). Occurrence, taxonomic, stratigraphic, and biological information was downloaded for all North American mammals. This data set was filtered so that only occurrences identified to the species-level, excluding all “sp.”-s. All aquatic and volant taxa were also excluded. Additionally, all occurrences without latitude and longitude information were excluded from the sample.

Species dietary and locomotor category assignments were done using the assignments in the PBDB, which were reassigned into coarser categories (Supplementary Table S1). This was done to improve interpretability, increase sample size per category, and make results comparable to previous studies [81, 126].

All individual fossil occurrences were assigned to 2 My bins ranging through the entire Cenozoic. Taxon duration was measured as the number of 2 My bins from the first occurrence to the last occurrence, inclusive. This bin size was chosen because it approximately reflects

the resolution of the North American Cenozoic mammal fossil record [3, 7, 102]. Species originating in the youngest cohort, 0-2 My, were excluded from analysis because every species duration would be both left and right censored, which is illogical.

Species body size estimates in grams were sourced from a large selection of primary literature and database compilations. Databases used include the PBDB, PanTHERIA [84], and the Neogene Old World Mammal database (NOW; <http://www.helsinki.fi/science/now/>). Major sources of additional compiled body size estimates include [183, 17, 45, 105, 135, 175]. These were then supplemented with an additional literature search to try and fill in the remaining gaps. In many cases, species body mass was estimated using various published regression equations based on tooth or skull measurements (Supplementary Table S2). If multiple specimens were measured, I used the mean of specimen measures as the species mean. See Dataset S1 for a complete list of mass estimates and sources.

## Biogeographic network

Species geographic extent was measured as the mean of the relative number of bioprovinces occupied by a species for each 2 My bin the species was present. Bioprovinces were identified using a network-theoretic approach that has previously been applied to paleontological data [161, 188]. This approach relies on defining a biogeographic bipartite network of taxa and localities. In this study, taxa were defined as species and localities were grid cells from a regular lattice on a global equal-area cylinder map projection. The regular lattice was defined as a 70 x 34 global grid where each cell corresponds to approximately 250000 km<sup>2</sup>. An advantage of this approach is that this approach reduces to occupancy when all localities are independent and to a single bioprovince when all localities are identical.

A biogeographic network was constructed for each of the 2 My bins used in this study.

Emergent bioprovinces were then identified using the map equation [151, 150] as has been done before [161, 187, 188]. These bioprovinces correspond to taxa and localities that are more interconnected with each other than with other nodes.

The map projection and regular lattice were made using shape files from <http://www.naturalearthdata.com/> and the **raster** package for R [? ]. Bioprovince identification was done using the map equation as implemented in the **igraph** package for R [? ].

## Supertree

As there is no single, combined formal phylogenetic hypothesis of all Cenozoic fossils mammals from North America, it was necessary to construct a semi-formal supertree. This was done by combining taxonomic information for all the observed species and a few published phylogenies using matrix representation parsimony [13]. For further explanation of the methodology used to construct this supertree, please see the Supplementary information.

### 2.4.2 Survival model

Presented here is the model development process used to formulate the two survival models used in this study. First, define  $y$  as a vector of length  $n$  where the  $i$ th element is the duration of species  $i$ , where  $i = 1, \dots, n$ .

The simplest survival model where durations are assumed to follow an exponential distribution with a single “rate” or inverse-scale parameter  $\lambda$  [86]. This is written

$$p(y|\lambda) = \lambda \exp(-\lambda y)$$

$$y \sim \text{Exp}(\lambda). \quad (2.1)$$

The exponential distribution corresponds to situations where extinction risk is independent of age. To understand this, we need to define two functions: the survival function  $S(t)$  and the hazard function  $h(t)$ .  $S(t)$  is the probability that a species having existed for  $t$  2 My bins will not have gone extinct while  $h(t)$  corresponds to the instantaneous extinction rate for some taxon age  $t$  [86]. For an exponential model,  $S(t)$  is

$$S(t) = \exp(-\lambda t) \quad (2.2)$$

and  $h(t)$  is defined

$$h(t) = \lambda \quad (2.3)$$

The choice of the exponential distribution corresponds directly to the Law of Constant Extinction [184] as the right side of Eq. 2.3 does not depend on species age  $t$ .

The current sampling statement (Eq. 2.1) assumes that all species share the same rate parameter with no variation. To allow for variation in  $\lambda$  associated with relevant covariate information like species body size,  $\lambda$  is reparameterized as  $\lambda_i = \exp(\sum \beta^T \mathbf{X}_i)$  with  $i$  indexing a given observation and its covariates,  $\beta$  is a vector of regression coefficients, and  $\mathbf{X}$  is a matrix of covariates. This is a standard regression approach, where one column of  $\mathbf{X}$  is all 1-s and its corresponding  $\beta$  coefficient is the intercept.

$\mathbf{X}$  is an  $n \times K$  matrix of species-level covariates. Three of the covariates of interest are the logit of mean relative occupancy, and the logarithm of body size (g). The discrete covariate index variables of dietary and locomotor category were transformed into  $n \times (k - 1)$  matrices where each column is an indicator variable (0/1) for that species's category,  $k$  being the number of categories of the index variable (3 and 4, respectively). Only  $k - 1$  indicator variables are necessary as the intercept takes on the remaining value. For example, the difference in effect of arboreality versus scansoriality on extinction risk, given that arboreality

is the reference category, is the coefficient for the scansorial indicator variable as that is the difference between between the effect of arboreality (the intercept  $\beta_0$ ) and scansoriality (the intercept + scansorial effect  $\beta_s$ ); Fig. 2.2). Finally, a vector of 1-s was included in the matrix  $\mathbf{X}$  whose corresponding  $\beta$  coefficient is the intercept, making  $K$  equal eight.

$\beta$  is the vector of regression coefficients. The intercept term was given a weak normal prior,  $\beta_0 \sim \mathcal{N}(0, 10)$  while all of these other coefficients were slightly more informative priors, e.g.  $\beta_{mass} \sim \mathcal{N}(0, 5)$ . These priors were chosen because it is expected that the effect size of each variable on duration will be small, as is generally the case with binary covariates [52].

Regression coefficients are not directly comparable without first standardizing the input variables to have equal standard deviations. This is accomplished by subtracting the mean of the covariate from all values and then dividing by the standard deviation, resulting in a variable with mean of zero and a standard deviation of one. This linear transform greatly improves the interpretability of the coefficients as expected change in mean duration given a difference of one standard deviation in the covariate [155]. Additionally, this makes the intercept directly interpretable as the estimate of mean (transformed)  $\sigma$  (Eq. 2.7). However, because the expected standard deviation for a random binary variable is 0.5, in order to make comparisons between the binary and continuous variables, the continuous inputs were divided by twice their standard deviation [50].

Origination cohort is defined as the group of species which all originated during the same 2 My temporal bin. Because the most recent temporal bin, 0-2 My, was excluded, there are 32 total cohorts. The effect of origination cohort  $j$  was modeled with each group being a sample from a common cohort effect,  $\eta$ , which was considered normally distributed with mean 0, and standard deviation  $\sigma_c$ . The value of  $\sigma_c$  was then estimated from the data itself, corresponding to the amount of pooling in the individual estimates of  $\eta_j$ . This approach is a conceptual and statistical unification between dynamic and cohort survival analysis in

paleontology [35, 138, 137, 185, 12], with  $\sigma_c$  acting as a measure of compromise between these two end members. The choice of the half-Cauchy prior for  $\sigma_c$  follows [49].

$$\eta_j \sim \mathcal{N}(0, \sigma_c)$$

$$\sigma_c \sim C^+(0, 2.5)$$

The impact of shared evolutionary history, or phylogeny, was modeled as an individual effect where each observation,  $i$ , is modeled as a multivariate normal,  $h$ , where the covariance matrix  $\Sigma$  is known up to a constant,  $\sigma_p^2$  [100, 66]. This is written

$$h \sim MVN(0, \Sigma)$$

$$\Sigma = \sigma_p^2 \mathbf{V}_{phy}$$

$$\sigma_p \sim C^+(0, 2.5).$$

$\mathbf{V}_{phy}$  is the phylogenetic covariance matrix defined as an  $n \times n$  matrix where the diagonal elements are the distance from root to tip, in branch length, for each observation and the off-diagonal elements are the amount of shared history, measured in branch length, between observations  $i$  and  $j$ .  $\sigma_p$  was given a weakly informative half-Cauchy hyperprior. Note that because the phylogeny used here is primarily based on taxonomy, estimates of  $\sigma_p$  represent minimum estimates [100, 66]. Improved phylogenetic estimates of all fossil Cenozoic mammals would greatly improve this estimate.

To relax the assumption of age-independent extinction of the Law of Constant Extinction, the Weibull distribution is substituted for the exponential [86]. The Weibull distribution has a shape parameter  $\alpha$  and scale parameter  $\sigma$ . Conceptually,  $\sigma$  is the inverse of  $\lambda$ .  $\alpha$  modifies the impact of taxon age on extinction risk. When  $\alpha > 1$  then  $h(t)$  is a monotonically increasing

function, but when  $\alpha < 1$  then  $h(t)$  is a monotonically decreasing function. When  $\alpha = 1$  then the Weibull distribution is equivalent to the exponential.

The Weibull distribution and sampling statement were defined

$$p(y|\alpha, \sigma) = \frac{\alpha}{\sigma} \left(\frac{y}{\sigma}\right)^{\alpha-1} \exp\left(-\left(\frac{y}{\sigma}\right)^\alpha\right)$$

$$y \sim \text{Weibull}(\alpha, \sigma). \quad (2.4)$$

The corresponding  $S(t)$  and  $h(t)$  functions are defined

$$S(t) = \exp\left(-\left(\frac{t}{\sigma}\right)^\alpha\right) \quad (2.5)$$

$$h(t) = \frac{\alpha}{\sigma} \left(\frac{t}{\sigma}\right)^{\alpha-1}. \quad (2.6)$$

To allow for  $\sigma$  to vary with a given observation's covariate information it is reparameterized in a similar fashion to  $\lambda$  with a few key differences. Because  $\sigma = 1/\lambda$  in order to preserve the interpretation of  $\beta$ , while taking  $\alpha$  into account,  $\sigma$  is reparameterized as

$$\sigma_i = \exp\left(\frac{-\beta}{\alpha}\right). \quad (2.7)$$

Given the above, the survival model was then fit in a Bayesian context using both exponential and Weibull distributions. The Weibull's  $\alpha$  parameter was assumed constant across species, which is standard practice in survival analysis [86].  $\alpha$  was given a weakly informative half-Cauchy ( $C^+$ ) prior.  $\sigma$  was reparameterized as an exponentiated regression model (Eq. 2.7). This was further expanded (Eq. 2.8) to allow for two hierarchical factors as discussed above. This is written

$$\sigma_i = \exp\left(\frac{-(h_i + \eta_j[i] + \sum \beta^T \mathbf{X}_i)}{\alpha}\right) \quad (2.8)$$

where equivalent statement for the exponential distribution is defined

$$\lambda_i = \exp\left(h_i + \eta_j[i] + \sum \beta^T \mathbf{X}_i\right). \quad (2.9)$$

An important part of survival analysis is the inclusion of censored observations where the failure time has not been observed [70, 86]. The most common censored observation is right censored, where the point of extinction had not yet been observed in the period of study, such as taxa that are still present in the most recent time bin (0-2 My). Left censored observations, on the other hand, correspond to observations that went extinct any time between 0 and some known point. To account for this uncertainty, the probability of a left censored observation is found by integrating over all possible durations between 0 and 1 time bin. For an explanation of how censored observations are included in the sampling statement, please see the Supplementary information.

#### 2.4.3 Estimation

Parameter posteriors were approximated using a Markov-chain Monte Carlo (MCMC) routine implemented in the Stan programming language [? ]. Stan implements a version of Hamiltonian Monte Carlo called the No-U-Turn sampler [? ]. Posterior approximation was done using four parallel MCMC chains run for 30000 steps, thinned to every thirtieth sample, split evenly between warm-up and sampling. Convergence was evaluated using the scale reduction factor,  $\hat{R}$ . Values of  $\hat{R}$  close to 1, or less than or equal to 1.1, indicate approximate convergence. Convergence means that the chains are approximately stationary and the samples are well mixed [51].

#### 2.4.4 Posterior evaluation

The most basic assessment of model fit is that simulated data generated given the model should be similar to the observed. This is the idea behind posterior predictive checks. Using the covariates from each of the observed durations, and randomly drawn parameter estimates from their marginal posteriors, a simulated data set  $y^{rep}$  was generated. This process was repeated 1000 times and the distribution of  $y^{rep}$  was compared with the observed [51]. For results from the posterior predictive tests used in this study, please see the Supplementary information.

The exponential and Weibull models were compared for out-of-sample predictive accuracy using the widely-applicable information criterion (WAIC) [197]. Because the Weibull model reduces to the exponential model when  $\alpha = 0$ , our interest is not in choosing between these models. Instead comparison of WAIC values is useful for better understanding the effect of model complexity on out-of-sample predictive accuracy. An explanation of how WAIC is calculated is presented in the Supplementary information following the recommended “WAIC 2” formulation [51].

There are three different variance components in this model: sample component, cohort  $\sigma_c^2$ , and phylogenetic  $\sigma_p^2$ . Partitioning the variance between these sources allows the relative amount of unexplained variance of the sample to be compared. The sample component is similar to the residual variance from a normal linear regression. However, the Weibull based model used here (Eq. 2.4) does not include an estimate of the variance similar to the squared scale term of the a Normal distribution. Instead, the sample component was approximated via a simulation approach modified from [56]. For explanation of this method, please see Supplementary information.

I used variance partitioning coefficients (VPC) to estimate the relative importance of the

different variance components [52]. Phylogenetic heritability,  $h_p^2$  [100, 66], is identical to the VPC of the phylogenetic effect. Phylogenetic heritability is a measure of how shared evolutionary history impacts differences in individual species trait values (e.g. duration). This is a broad sense “heritability” as it combines both genetic inheritance and other, non-genetic shared history factors. Importantly, because phylogenetic effect was estimated using a principally taxonomy based tree the estimates derived here can be considered minimum estimates of the phylogenetic effect.

## 2.5 Supplemental information for “Death and taxa”

### 2.5.1 Supertree inference

As there is no single, combined formal phylogenetic hypothesis of all Cenozoic fossils mammals from North America, it was necessary to construct a semi-formal supertree. This was done by combining taxonomic information for all the observed species and a few published phylogenies.

The initial taxonomic classification of the observed species was based on the associated taxonomic information from the PBDB. This information was then updated using the Encyclopedia of Life (<http://eol.org/>) which collects and collates taxonomic information in a single database. This was done programatically using the `taxize` package for R [? ]. Finally, this taxonomic information was further updated using a published taxonomy of fossil mammals [79, 78].

This taxonomy serves as an initial phylogenetic hypothesis which was then combined with a selection of species-level phylogenies [13, 135] in order to better constrain a minimum estimate of the actual phylogenetic relationships of the species. The supertree was inferred via matrix representation parsimony implemented in the `phytools` package for R [? ]. Of

the two most parsimonious trees found, I used only one for analysis.

Polytomies were resolved in order of species first appearance in order to minimize stratigraphic gaps. The resulting tree was then time scaled using the `paleotree` package via the “minimum branch length” approach with a minimum length of 0.1 My [10]. The minimum length is necessary to avoid zero-length branches which cause the phylogenetic covariance matrix not to be positive definite, which is important for computation (see below). While other time scaling approaches are possible [61, 11] this method was chosen for its simplicity and not requiring additional information about diversification rates which are the interest of this study.

### *2.5.2 Modeling censored observations*

Censored data are modeled using the survival function of the distribution,  $S(t)$ , defined earlier for the Weibull distribution (Eq. 5, 6) with  $\sigma$  defined as above (Eq. 8, 9).  $S(t)$  is the probability that an observation will survive longer than a given time  $t$ .

The likelihood of uncensored observations is evaluated as normal using equation 4 while right censored observations are evaluated at  $S(t)$  and left censored observations are evaluated at  $1 - S(t)$ . Note,  $1 - S(t)$  is equivalent to the cumulative distribution function and  $S(t)$  is equivalent to the complementary cumulative distribution function [51].

The final sampling statement/likelihood for both uncensored and both right and left censored observations is then written

$$L \propto \prod_{i \in C} \text{Weibull}(y_i | \alpha, \sigma) \prod_{j \in R} S(y_j | \alpha, \sigma) \prod_{k \in L} (1 - S(y_k | \alpha, \sigma)),$$

where  $C$  is the set of uncensored observations,  $R$  is the set of right censored observations,

and  $L$  is the set of left censored observations.

### 2.5.3 Deviance residuals

In standard linear regression, residuals are defined as  $r_i = y_i - y_i^{est}$ . For the model used here, this definition is inadequate. The equivalent values for survival analysis are deviance residuals. To define how deviance residuals are calculated, we first define the cumulative hazard function [86]. Given  $S(t)$ , we define the cumulative hazard function as

$$\Lambda(t) = -\log(S(t)).$$

Next, we define martingale residuals  $m$  as

$$m_i = I_i - \Lambda(t_i).$$

$I$  is the inclusion vector of length  $n$ , where  $I_i = 1$  means the observation is completely observed and  $I_i = 0$  means the observation is censored. Martingale residuals have a mean of 0, range between 1 and  $-\infty$ , and can be viewed as the difference between the observed number of deaths between 0 and  $t_i$  and the expected number of deaths based on the model. However, martingale residuals are asymmetrically distributed, and can not be interpreted in the same manner as standard residuals.

The solution to this is to use the deviance residuals,  $D$ . This is defined as a function of martingale residuals and takes the form

$$D_i = \text{sign}(m_i) \sqrt{-2[m_i + I_i \log(I_i - m_i)]}.$$

Deviance residuals have a mean of 0 and a standard deviation of 1 by definition.

#### 2.5.4 Variance partitioning

I calculated VPC using a resampling approach based on [56]. The procedure is as follows:

1. Simulate  $w$  (50,000) values of  $\eta$ ;  $\eta \sim \mathcal{N}(0, \sigma_c)$ .
2. For a given value of  $\beta^T \mathbf{X}$ , calculate  $\sigma^{c*}$  (Eq. 7) for all  $w$  simulations, holding  $h$  constant at 0.
3. Calculate  $v_c$ , the Weibull variance (Eq. 2.10) of each element of  $\sigma^{c*}$  with  $\alpha$  drawn from the posterior estimate.
4. Simulate  $w$  values of  $h$ ;  $h \sim \mathcal{N}(0, \sigma_p)$ .
5. For a given value of  $\beta^T \mathbf{X}$ , calculate  $\sigma^{p*}$  (Eq. 7) for all  $w$  simulations, holding  $\eta$  constant at 0.
6. Calculate  $v_p$ , the Weibull variance (Eq. 2.10) of each element of  $\sigma^{p*}$  with  $\alpha$  drawn from the posterior estimate.
7.  $\sigma_{y*}^2 = \frac{1}{2} \left( \left( \frac{1}{w} \sum_i^w v_{pi} \right) + \left( \frac{1}{w} \sum_j^w v_{cj} \right) \right)$ .
8.  $\sigma_{c*}^2 = var(v_c)$  and  $\sigma_{p*}^2 = var(v_p)$ .

The simulated values of  $h$  were drawn from a univariate normal distribution because each simulated value is in isolation, so there is no concern of phylogenetic autocorrelation. The chosen value for  $\beta^T \mathbf{X}$  was a draw from the posterior estimate of the intercept. Because input variables were standardized prior to model fitting, the intercept corresponds to the estimated effect on survival of the sample mean.

Weibull variance is calculated as

$$var(x) = \sigma^2 \left( \Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2 \right), \quad (2.10)$$

where  $\Gamma$  is the gamma function.

The variance partitioning coefficients are then calculated, for example, as  $VPC_{phylo} = \frac{\sigma_{p*}^2}{\sigma_{y*}^2 + \sigma_{c*}^2 + \sigma_{p*}^2}$  and similarly for the other components.

### 2.5.5 Widely applicable information criterion

WAIC can be considered fully Bayesian alternative to the Akaike information criterion, where WAIC acts as an approximation of leave-one-out cross-validation which acts as a measure of out-of-sample predictive accuracy [51]. The following explanation uses the “WAIC 2” formulation recommended by [51].

WAIC is calculated starting with the log pointwise posterior predictive density calculated as

$$\text{lppd} = \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \Theta^S) \right), \quad (2.11)$$

where  $n$  is sample size,  $S$  is the number posterior simulation draws, and  $\Theta$  represents all of the estimated parameters of the model. This is similar to calculating the likelihood of each observation given the entire posterior.

A correction for the effective number of parameters is then added to lppd to adjust for overfitting. The effective number of parameters is calculated, following derivation and

recommendations of [51], as

$$p_{\text{WAIC}} = \sum_{i=1}^n V_{s=1}^S (\log p(y_i | \Theta^S)). \quad (2.12)$$

where  $V$  is the sample posterior variance of the log predictive density for each data point.

Given both equations 2.11 and 2.12, WAIC is then calculated

$$\text{WAIC} = \text{lppd} - p_{\text{WAIC}}. \quad (2.13)$$

When comparing two or more models, lower WAIC values indicate better out-of-sample predictive accuracy. Importantly, WAIC is just one way of comparing models. When combined with posterior predictive checks it is possible to get a more complete understanding of model fit.

### 2.5.6 Results from posterior predictive checks

With all marginal posterior estimates having converged ( $\hat{R} < 1.1$ ) it is possible to examine the quality of model fit (Table 1). If the model is an adequate descriptor of the observed data, then relatively confident inference can be made [51].

Visual examination of the deviance residuals from twelve different sets of posterior predictive simulations indicates a systematic weakness estimating durations greater than 3 2-My bins (Fig. 2.5). However, the comparison of posterior predictive estimates of the 25th, 50th, and 75th quantiles to the observed indicate adequate fit. (Fig. 2.6). Importantly, this indicates that the model has approximate fit for 50+% of the data. Because, the inferred model can be inferred to be approximately adequate at capturing the observed variation.

The Weibull model (6140.37) also had a much lower WAIC score than the Exponential model (16697.35). This large a difference indicates that the Weibull model probably has the lower out-of-sample predictive accuracy of the two.

### *2.5.7 Data quality concerns*

A concern with using the PBDB as a primary data source, though this concerns are general to most paleontological data, is that the results are an artifact of taxonomy or the database itself [191]. However, to obtain the results obtained in this analysis there would need to be a systematic error in assignments of all of diet, locomotor, and taxonomic categories for a large portion of the close to 2000 sampled species. It is important to note that species included have to have body size information, much of which is found from other sources (see Dataset S1). this means that, for many taxa, that species name has to appear in occur in more than one place. this is a strong filter for misspellings and potentially invalid taxa. Additionally, given that most mammal fossils are teeth which allows for relatively accurate dietary category assignment.

A possible concern, however, is that omnivorous taxa have feature poor morphology and thus longer durations may reflect a single anagenetic lineage as opposed to a single “species.” However it is possible to consider that, from a population genetic perspective, it can be argued that a single unbranching lineage is still a single biological “unit.”

### *2.5.8 Concerns surrounding estimates of $\alpha$*

The estimate of the Weibull shape parameter,  $\alpha$ , is greater than 1 meaning that extinction risk is expected to increase with taxon age (Table 1). As the value of  $\alpha$  is between 1 and

1.5, extinction risk for a given species only gradually increases with age (Fig. 2.7). There are three possible explanations for this result: 1) older taxa being out competed by younger taxa [192]; or 2) this is an artifact of the minimum resolution of the fossil record [156].

An additional concern is that there may be an upward bias in estimates of  $\alpha$  at this sample size, similar to that for scale parameters [51]. The plausibility of third possibility in this example can be explored in simulation. I simulated from 10, 100, 1000, and 10000 samples from a Weibull( $\alpha = 1.3$ ,  $\sigma = 1$ ) 100 times each. For each of these simulated datasets, I then estimated the values of  $\alpha$  and  $\sigma$  in a simple maximum likelihood context in order to just get the model estimate. The modal estimates of both parameters for the simulated datasets were then compared to the known values (Fig. 2.8). The results from these simulations demonstrate that the estimates of  $\alpha$  in the above analyses (Table 1) should not be particularly biased based on my sample size of approximately 2000 species.

The model used in this analysis, however, is unable to distinguish between the remaining two hypotheses [156, 192]. Further work on how to better constrain estimates  $\alpha$  is necessary. A possibly is somehow incorporating these hypotheses as prior information.

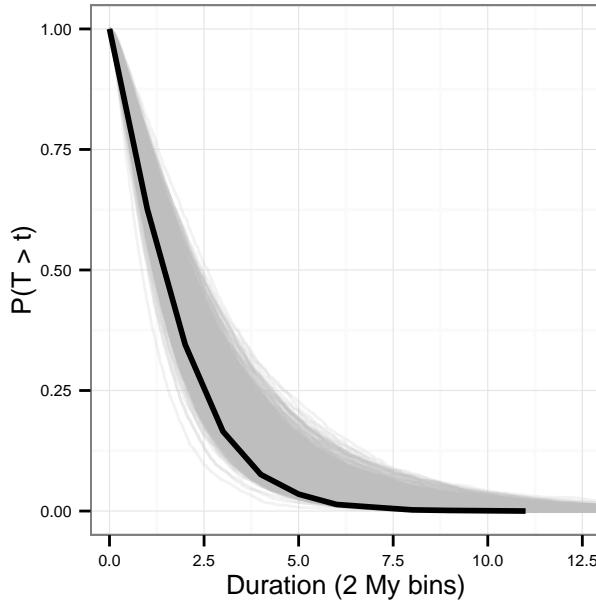


Figure 2.1: Weibull-based model estimates (grey) from 1000 posterior predictive data sets of the empirical survival function (black). The survival function is the probability that a species with duration  $t$  will not have gone extinct. Simulated data sets were generated by drawing parameter values randomly from their estimated posteriors and using the observed covariate information to estimate durations for all the observed species.

Table 2.1: Marginal posterior estimates for the parameters of interest based on 1000 posterior samples. The intercept  $\beta_0$  can also be interpreted as the estimate for the mean observed species. The remaining  $\beta$  values can be interpreted as the effect of a trait on the expected species duration as expressed as deviation from the mean. The categorical variables are binary index variables where an observation is of that category or not. See Equation 2.6 for an explanation of the effect of  $\alpha$  on extinction risk.  $\hat{R}$  values of less than 1.1 indicate approximate chain convergence for the posterior samples.

parameter	effect	mean	sd	2.5%	25%	50%	75%	97.5%	$\hat{R}$
$\alpha$	“age”	1.29	0.03	1.23	1.27	1.29	1.31	1.36	1.00
$\beta_0$	arboreal/carnivore	-0.78	0.14	-1.05	-0.87	-0.78	-0.68	-0.51	1.00
$\beta_o$	occupancy	-0.53	0.08	-0.69	-0.59	-0.53	-0.48	-0.38	1.00
$\beta_{size}$	body size	-0.05	0.05	-0.14	-0.08	-0.05	-0.01	0.05	1.00
$\beta_g$	ground dwelling	-0.28	0.10	-0.47	-0.34	-0.28	-0.21	-0.09	1.00
$\beta_s$	scansorial	-0.22	0.11	-0.43	-0.29	-0.22	-0.14	-0.00	1.00
$\beta_h$	herbivore	0.09	0.09	-0.09	0.03	0.09	0.14	0.27	1.00
$\beta_i$	insectivore	0.10	0.11	-0.11	0.03	0.10	0.17	0.31	1.00
$\beta_o$	omnivore	-0.12	0.11	-0.33	-0.19	-0.12	-0.05	0.09	1.00
$\sigma_c$	sd cohort	0.33	0.06	0.23	0.29	0.33	0.37	0.48	1.00
$\sigma_p$	sd phylogeny	0.11	0.05	0.03	0.07	0.10	0.14	0.23	1.03

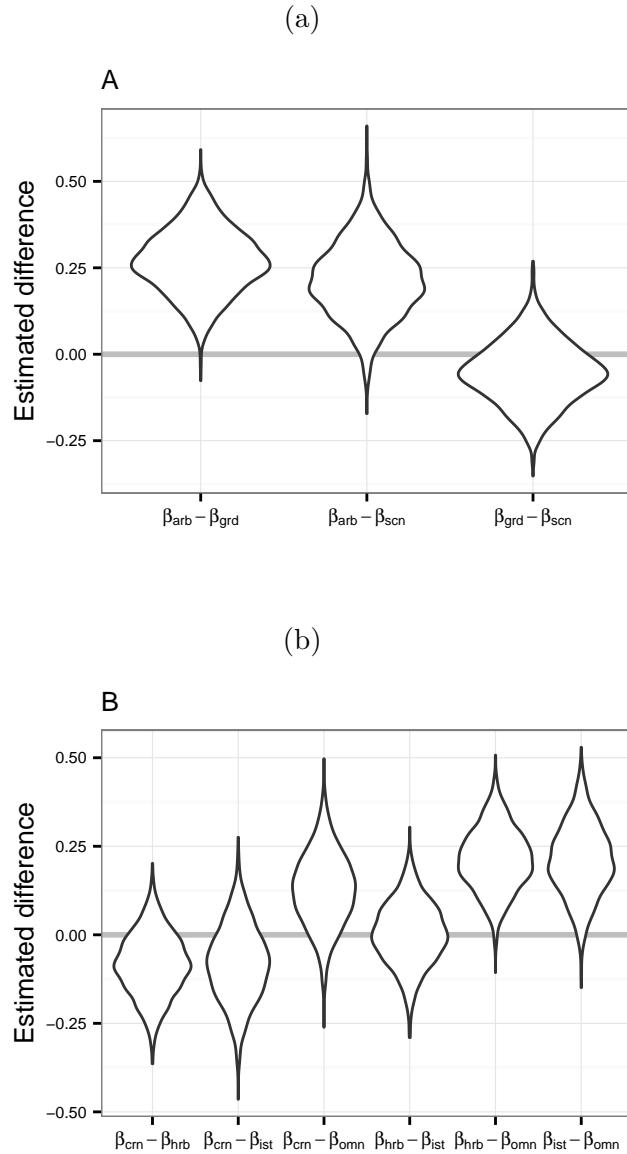


Figure 2.2: Pairwise differences in effect of the locomotor (**A**) and dietary categories (**B**) on expected duration from 1000 samples from the posterior distribution. Comparisons of locomotor categories, from top to bottom (**A**), are: arboreal ( $\beta_{arb} = \beta_0$ ) versus ground dwelling ( $\beta_{grd} = \beta_0 + \beta_g$ ), arboreal versus scansorial ( $\beta_{scn} = \beta_0 + \beta_s$ ), and ground dwelling versus scansorial. For dietary category, from top to bottom (**B**): carnivore ( $\beta_{crn} = \beta_0$ ) versus herbivore ( $\beta_{hrb} = \beta_0 + \beta_h$ ), carnivore versus insectivore ( $\beta_{ist} = \beta_0 + \beta_i$ ), carnivore versus omnivore ( $\beta_{omn} = \beta_0 + \beta_o$ ), herbivore versus insectivore, herbivore versus omnivore, and insectivore versus omnivore. Negative values indicate that the first category is expected to have a greater duration than the second, while positive values indicate that the first category is expected to have a shorter duration.

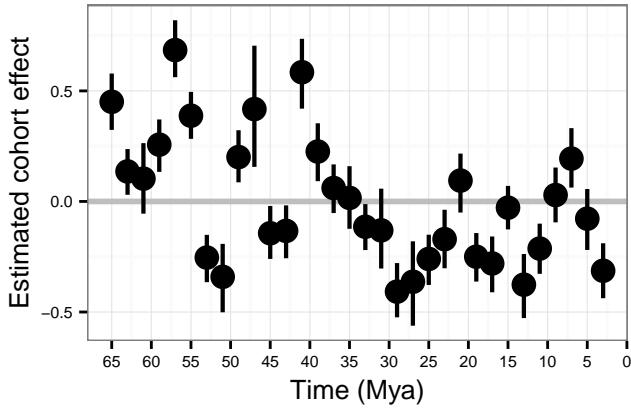


Figure 2.3: Summaries of posterior estimates of individual cohort effect depicted as medians and 80% credible intervals. High values correspond to shorter species durations while lower values correspond to greater species durations compared to the mean duration. Lines are placed at the middle of the 2 My origination cohorts.

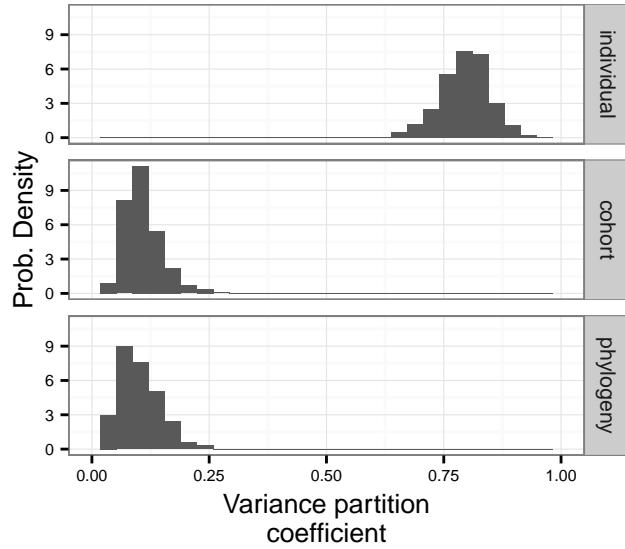


Figure 2.4: Estimates of the variance partitioning coefficients for the three different sources of variance: species, cohort, and phylogeny. Higher values correspond to greater contribution to total observed variance. Each of the estimates is a distribution of 1000 approximating simulations due to the model's non-normally distributed errors.

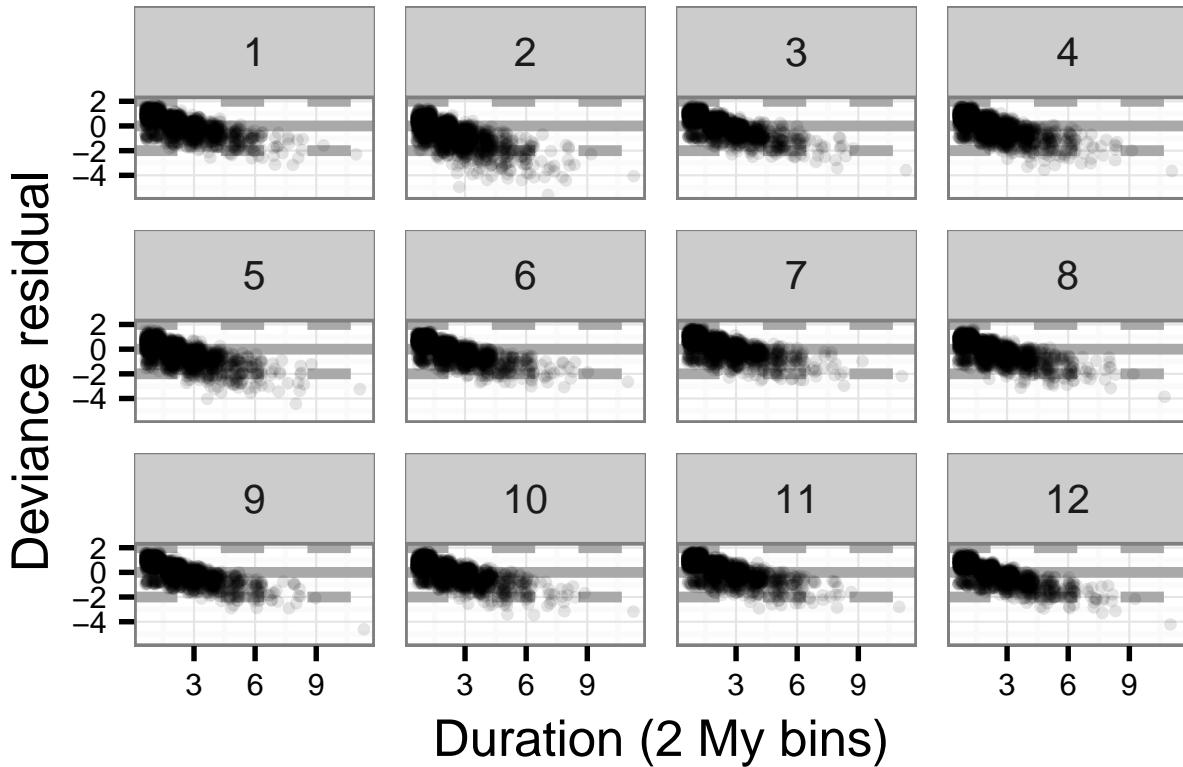


Figure 2.5: Deviance residuals from the fitted survival model compared to observed durations. Each graph depicts the residuals from single draws from the posterior distributions of all estimated parameters. Positive values indicate an underestimate of the observed duration, while negative values indicate an overestimate of the observed duration. A small amount of noise is added to each point to increase clarity. Twelve different examples are provided here to indicate consistency across multiple realizations.

Table 2.2: Species trait assignments in this study are a coarser version of the information available in the PBDB. Information was coarsened to improve per category sample size and uniformity and followed this table.

This study		PBDB categories
Diet	Carnivore	Carnivore
	Herbivore	Browser, folivore, granivore, grazer, herbivore.
	Insectivore	Insectivore.
	Omnivore	Frugivore, omnivore.
Locomotor	Arboreal	Arboreal.
	Ground dwelling	Fossorial, ground dwelling, semifossorial, saltatorial.
	Scansorial	Scansorial.

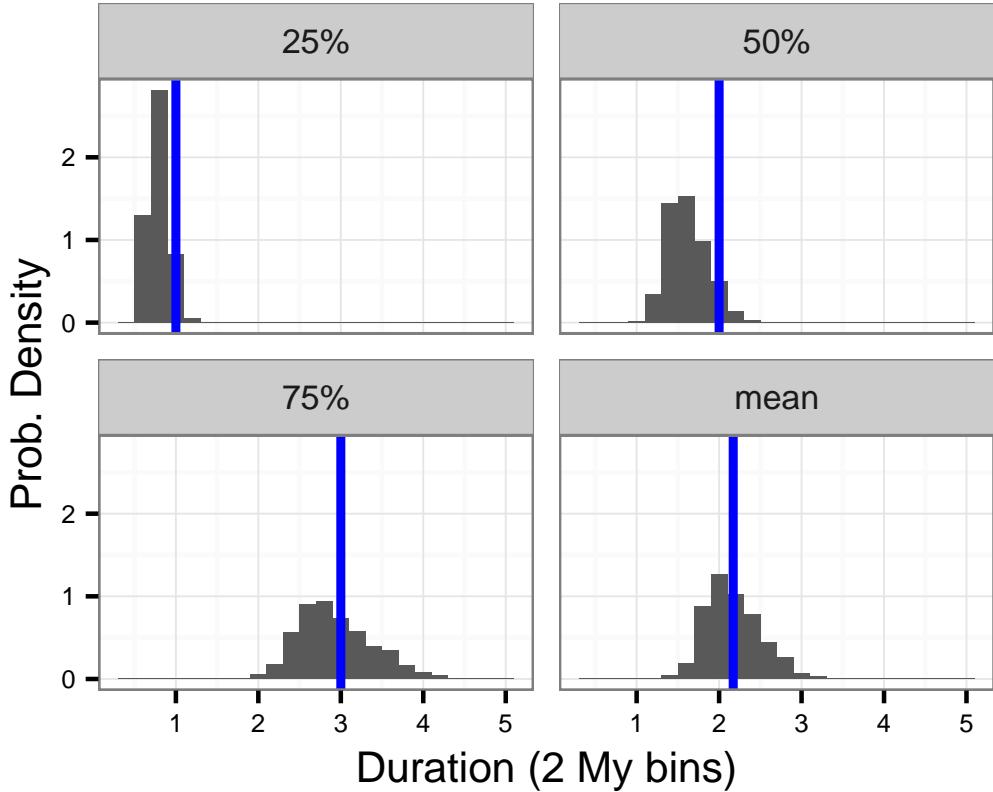


Figure 2.6: The results of additional posterior predictive checks for four summaries of the observed durations, as labeled. Blue vertical lines indicate the observed value. None of the observed values are significantly different from the posterior predictive distributions.

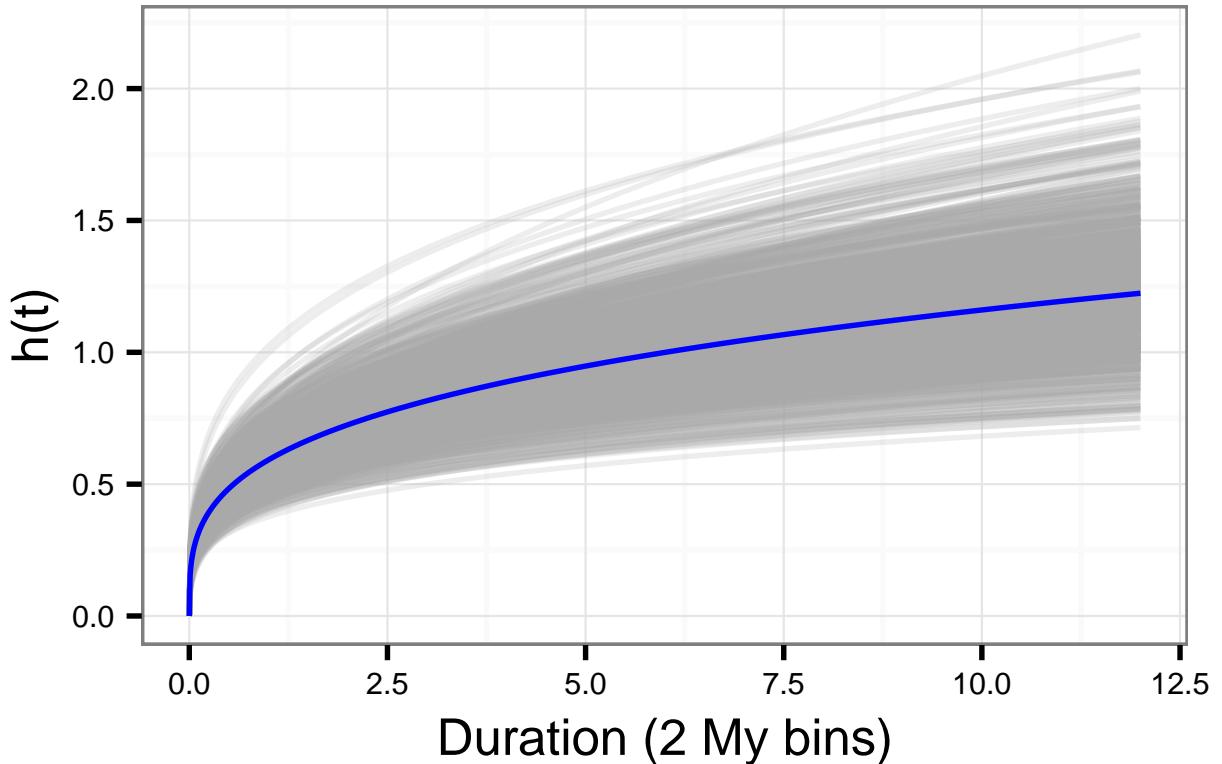


Figure 2.7: 1000 estimates of the hazard function ( $h(t)$ ) for the observed species mean (grey), along with the median estimated hazard function (blue).  $h(t)$  is an estimate of the rate at which a species of age  $t$  is expected to go extinct. Hazard functions were estimated from random draws from the estimated posterior distributions and evaluated with all covariate information set to 0, which corresponds to the expected duration of the mean species.

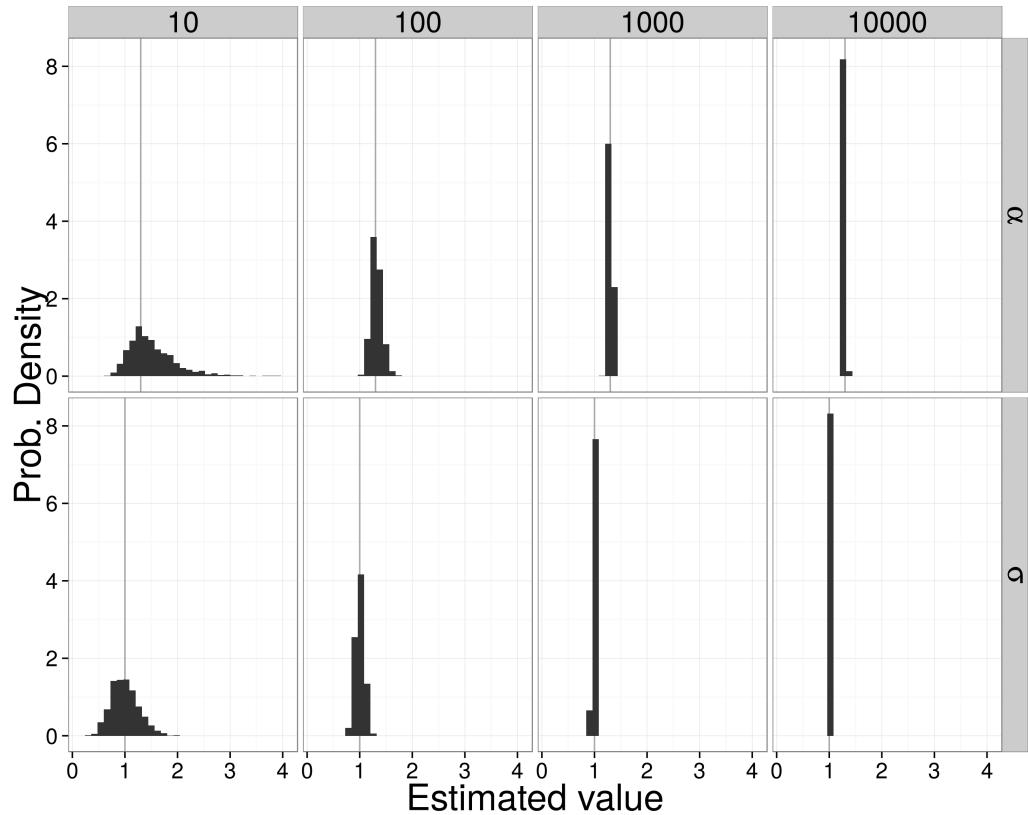


Figure 2.8: Comparison of maximum likelihood estimates of shape ( $\alpha$ ) and scale ( $\sigma$ ) parameters from 1000 simulated data sets from 4 different sample sizes. Vertical lines are the actual parameter value used to generate the data. When sample size is approximately 100 or greater, estimates are not overly biased.

Table 2.3: Regression equations used in this study for estimating body size. Equations are presented with reference to taxonomic grouping, part name, and reference.

Group	Equation	log(Measurement)	Source
General	$\log(m) = 1.827x + 1.81$	lower m1 area	[89]
General	$\log(m) = 2.9677x - 5.6712$	mandible length	[44]
General	$\log(m) = 3.68x - 3.83$	skull length	[99]
Carnivores	$\log(m) = 2.97x + 1.681$	lower m1 length	[186]
Insectivores	$\log(m) = 1.628x + 1.726$	lower m1 area	[14]
Insectivores	$\log(m) = 1.714x + 0.886$	upper M1 area	[14]
Lagomorph	$\log(m) = 2.671x - 2.671$	lower toothrow area	[183]
Lagomorph	$\log(m) = 4.468x - 3.002$	lower m1 length	[183]
Marsupials	$\log(m) = 3.284x + 1.83$	upper M1 length	[57]
Marsupials	$\log(m) = 1.733x + 1.571$	upper M1 area	[57]
Rodentia	$\log(m) = 1.767x + 2.172$	lower m1 area	[89]
Ungulates	$\log(m) = 1.516x + 3.757$	lower m1 area	[106]
Ungulates	$\log(m) = 3.076x + 2.366$	lower m2 length	[106]
Ungulates	$\log(m) = 1.518x + 2.792$	lower m2 area	[106]
Ungulates	$\log(m) = 3.113x - 1.374$	lower toothrow length	[106]

## CHAPTER 3

### HOW MACROECOLOGY AFFECTS MACROEVOLUTION: THE INTERPLAY BETWEEN EXTINCTION INTENSITY AND TRAIT-DEPENDENT EXTINCTION IN BRACHIOPODS

As extinction intensity increases, how do the effects of traits on taxonomic survival change? Does the extinction rate associated with certain traits increase while that of others decreases? Using a hierarchical Bayesian approach, I develop a model of how the effects of biological traits on extinction risk can vary with respect to extinction intensity, origination cohort (i.e. time of origination), and in relation to each other. The emergent traits I analyze in relation to their patterns of Paleozoic brachiopod genus durations are geographic range, affinity for epicontinental seas versus open ocean environments, and body size. Additionally, I estimate the effects of environmental generalization versus specialization on taxonomic survival by allowing environmental preference to have a nonlinear effect on duration. My analytical framework eschews the traditional distinction between background and mass extinction, and instead considers extinction intensity as a continuum. I find that the cohort-specific effects of geographic range and environmental preference are negatively correlated with baseline extinction intensity. Additionally, I find support for greater survival of environmental generalists versus specialists in all origination cohorts. These results support the conclusion that for Paleozoic brachiopods, as extinction intensity increases overall extinction selectivity increases.

### 3.1 Introduction

Extinction is one half of the diversification process [141, 182, 181], second only to speciation or origination; it can also be the ultimate manifestation of selection as a taxon with a beneficial trait should persist for longer on average than a taxon without that beneficial trait [133, 74, 141, 181].

While estimation of both trait-dependent speciation and extinction rates from phylogenies of extant taxa has grown dramatically [101, 34, 54, 55, 134, 179, 178, 180], there are two major ways to estimate trait-dependent extinction: analysis of phylogenies, and analysis of the fossil record. These two directions, phylogenetic comparative and paleobiological, are complementary and intertwined in the field of macroevolution [133, 74, 69]. In the case of extinction, analysis of the fossil record has the distinct advantage over phylogenies of only extant taxa because extinction is observable; this means that extinction rate is possible to estimate [130, 128, 96]. The approach used here is thus complementary to the analysis of trait-dependent extinction based on a phylogeny.

(**author?**) [72] observed that for bivalves at the end Cretaceous mass extinction event, the effects of some biological traits on taxonomic survival decreased. However, this pattern was not the case for the effect of geographic range on survival [72, 117]. There are multiple possible macroevolutionary mechanisms which may underlie this pattern: the effect of geographic range on survival remains constant and those of other biological traits decrease, the effect of geographic range on survival increases and those of other biological traits stay constant, or the effects of all traits decrease potentially by different degrees.

While (**author?**) [72] phrased his conclusions in terms of background versus mass extinction, these states are not distinguishable in terms of extinction rate alone; my analysis treats the time period analyzed as part of the same continuum [194, 117, 166]. Additionally, in order

to test the proposed macroevolutionary mechanism behind the (**author?**) [72] scenario; not only do the taxon trait effects needs to be modeled, but the correlation between trait effects need to be modeled as well.

Here I model brachiopod taxon durations because trait based differences in extinction risk should manifest as differences in taxon durations. Brachiopods are an ideal group for this study as they are well known for having an exceptionally complete fossil record [40, 42]. I focus on the brachiopod record from the post-Cambrian Paleozoic, from the start of the Ordovician (approximately 485 My) through the end Permian (approximately 252 My) as this represents the time of greatest global brachiopod diversity [5] meaning a large sample size for this analysis.

The analysis of taxon durations, or time from origination to extinction, falls under the purview of survival analysis, a field of applied statistics commonly used in health care and engineering [86] but has a long history in paleontology [168, 169, 184, 185, 25]. I adopt a hierarchical modeling approach [52, 51], which represents both a conceptual and statistical unification of the paleontological dynamic and cohort survival analytic approaches [184, 185, 138, 137, 35, 12, 164, 25, 30].

### *3.1.1 Factors affecting brachiopod survival*

Conceptually, taxon survival can be considered an aspect of “taxon fitness” [23, 116]. Traits associated with taxon survival are thus examples of species (or higher-level) selection, as differences in survival are analogous to differences in fitness. The traits analyzed here are all examples of emergent and aggregate traits [74, 133]; specifically I analyze genus-level traits. Emergent traits are those which are not measurable at a lower level (e.g. species versus individual organism) such as geographic range, or even fossil sampling rate. Aggregate

traits, like body size or environmental preference, are the average of a shared trait across all members of a lower level.

Geographic range is widely considered the most important biological trait for estimating differences in extinction risk at nearly all times, with large geographic range associated with low extinction risk [72, 71, 75, 117, 74, 58, 32]. This stands to reason even if extinction is completely at random; a taxon with an unrestricted range is less likely to go extinct at random than a taxon with a restricted range.

Epicontinental seas are a shallow-marine environment where the ocean has spread over the continental interior or craton with a depth typically less than 100m. In contrast, open-ocean coastline environments have much greater variance in depth, do not cover the continental craton, and can persist during periods of low sea level [107]. Because of this, a simple hypothesis that taxa which favor epicontinental seas would be at great risk during periods of low sea levels, such as during glacial periods, when epicontinental seas are drained. During the Paleozoic (approximately 541–252 My), epicontinental seas were widely spread globally but declined over the Mesozoic (approximately 252–66 My) and have nearly disappeared during the Cenozoic (approximately 66–0 My) as open-ocean coastlines became the dominant shallow-marine setting [158, 121, 107, 83]. Taxa in epicontinental environments could also have a greater extinction susceptibility than taxa in open-ocean environments due to anoxic events due to enhanced water stratification or poor water circulation [120].

(author?) [107] demonstrated that during several mass extinctions taxa associated with open-ocean environments tend to have a greater extinction risk than those taxa associated with epicontinental seas. During periods of background extinction, however, they found no consistent difference between taxa favoring either environment. (author?) [107] hypothesize that open-ocean taxa may have a greater extinction rate because these environments would be more strongly affected by waterborne hazards such as fallout from impacts or volcanic

events which would propagate more quickly than in epicontinental environments with sluggish circulation. These two environment types represent the primary identifiable environmental dichotomy observed in ancient marine systems [107, 158]. Given these findings, I would hypothesize that as extinction risk increases, the extinction risk associated with open-ocean environments should generally increase.

Because environmental preference is defined here as the continuum between occurring exclusively in open-ocean environments versus epicontinental environments, intermediate values are considered “generalists” in the sense that they favor neither end member. A long-standing hypothesis is that generalists or unspecialized taxa will have greater survival than specialists [168, 92, 93, 114, 115, 12? ]. Because of this, the effect of environmental preference was modeled as a quadratic function where a concave down relationship between preference and expected duration indicates that generalists are favored over specialists end-members.

Body size, measured as shell length, is also considered as a trait that may potentially influence extinction risk [118, 59]. Body size is a proxy for metabolic activity and other correlated life history traits [118]. (**author?**) [60] analyzed the effect of body size selectivity in Devonian brachiopods in both a phylogenetic and non-phylogenetic context; finding that body size was not found to be associated with differences in taxonomic duration. It has also been found that, at least in the case of some bivalve subclades, body size can be as important a factor as geographic range size in determining extinction risk [59]. Given these results, I expect that if body size has any effect on brachiopod taxonomic survival it is very small.

It is well known that, given the incompleteness of the fossil record, the observed duration of a taxon is an underestimate of that taxon’s true duration [177, 193, 195, 95, 6, 40]. Because of this, the concern is that a taxon’s observed duration may reflect its relative chance of being sampled and not any of the effects of the covariates of interest. In this case, for sampling to be a confounding factor there must be consistent relationship between the quality of sampling

of a taxon and its apparent duration (e.g. greater sampling, longer duration). If there is no relationship between sampling and duration then interpretation can be made clearly; while observed durations are obviously truncated true durations, a lack of a relationship would indicate that the amount and form of this truncation is not a major determinant of the taxon's apparent duration. By including sampling as a covariate in the model, this effect can be quantified and can be taken into account when interpreting the estimates of the effects of the other covariates.

## 3.2 Materials and Methods

### 3.2.1 *Fossil occurrence information*

The brachiopod dataset analyzed here was sourced from the Paleobiology Database (<http://www.paleodb.org>) which was then filtered based on taxonomic (Rhynchonelliformea: Rhynchonellata, Chileata, Obolellida, Kutorginida, Strophomenida, Spiriferida), temporal (post-Cambrian Paleozoic), stratigraphic, and other occurrence information used in this analysis. Analyzed occurrences were restricted to those with paleolatitude and paleolongitude coordinates, assignment to either epicontinental or open-ocean environment, and belonging to a genus present in the body size dataset [118]. Epicontinental versus open-ocean assignments for each fossil occurrence are partially based on those from **(author?)** [107], with additional occurrences assigned similarly (Miller and Foote, personal communication). These filtering criteria are very similar to those from **(author?)** [39] with an additional constraint of being present in the body size data set from **(author?)** [118]. In total, there 1130 genera were included in the dataset.

Fossil occurrences were analyzed at the genus level which is common for paleobiological, macroevolutionary and macroecological studies of marine invertebrates [5, 39, 58, 85, 107,

114, 115, 117, 166, 188]. While species diversity dynamics are frequently of much greater interest than those of higher taxa (though see (**author?**) 37, 63), the nature of the fossil record makes accurate, precise, and consistent taxonomic assignments at the species level difficult for all occurrences. As such, the choice to analyze genera as opposed to species was in order to assure a minimum level of confidence and accuracy in the data analyzed here.

Genus duration was calculated as the number of geologic stages from first appearance to last appearance, inclusive. Durations were based on geologic stages as opposed to millions of years because of the inherently discrete nature of the fossil record; dates are not assigned to individual fossils themselves but instead fossils are assigned to a geological interval which represents some temporal range. In this analysis, stages are effectively irreducible temporal intervals in which taxa may occur. Genera with a last occurrence in or after Changhsingian stage (e.g. the final stage of the study interval) were right censored at the Changhsingian; genera with a duration of only one stage were left censored [86]. How the likelihood of censored observations is calculated is detailed in section 3.2.2.

The covariates of duration included in this analysis are geographic range size ( $r$ ), environmental preference ( $v, v^2$ ), body size ( $m$ ), and sampling ( $s$ ).

Geographic range was calculated as relative occupancy corrected for incomplete sampling. First, the paleolatitude-paleolongitude coordinates for all occurrences were projected onto an equal-area cylindrical map projection. Each occurrence was then assigned to one of the cells from a  $70 \times 34$  regular raster grid placed on the map. Each grid cell represents approximately 250,000 km<sup>2</sup>. The map projection and regular lattice were made using shape files from <http://www.naturalearthdata.com/> and the **raster** package for R [? ]. For each stage, the total number of occupied grid cells was calculated. Then, for each temporal bin, the relative occurrence probability of the observed taxa was calculated using the JADE method developed by (**author?**) [22]. This method accounts for the fact that taxa with an occupancy of 0

cannot be observed which means that occupancy follows a truncated Binomial distribution. This correction is critical when comparing occupancies from different times with different geographic sampling. Finally, for each genus, the mean relative occurrence probability was calculated as the average of that genus' occurrence probabilities for all stages it was sampled to yield relative occupancy, my proxy for geographic range.

Environmental preference was defined as probability of observing the ratio of epicontinental occurrences to total occurrences ( $\theta_i = e_i/E_i$ ) or greater given the background occurrence probability  $\theta'_i$  as estimated from all other taxa occurring at the same time ( $e'_i/E'_i$ ). This measure of environmental preference is expressed.

$$\begin{aligned} p(\theta'_i | e'_i, E'_i) &\propto \text{Beta}(e'_i, E'_i - e'_i)\text{Beta}(1, 1) \\ &= \text{Beta}(e'_i + 1, E'_i - e'_i + 1), \end{aligned} \tag{3.1}$$

where  $v$  is the percent of the distribution defined in equation 3.1 less than or equal to  $\theta_i$ . The Beta distribution is used here because it is a continuous distribution bounded at 0 and 1, which is idea for modeling percentages.

Body size, measured as shell length, was sourced directly from (**author?**) [118]. These measurements were made from brachiopod taxa figured in the *Treatise on Invertebrate Paleontology* [198].

The sampling probability for individual taxa was calculated using the standard gap statistic [40, 41]. The gap statistic is calculated as the number of stages in which the taxon was sampled minus two divided by the duration of the taxon minus two. Subtracting two from both the numerator and denominator is because the first and last appearance stages are by definition sampled. Because taxa that were right censored only include a first appearance, one was subtracted from the numerator and denominator instead of two.

The minimum duration for which a gap statistic can be calculated is three stages, so I chose to impute the gap statistic for all observations with a duration less than 3. Imputation is the “filling in” of missing observations based on the observed values [154, 52]. This is fairly straight forward in a Bayesian framework because both covariates and parameters are considered random variables, meaning that the missing values of sampling can be modeled as coming from some probability distribution. The model for imputing sampling probability is described in section 3.2.3.

Prior to analysis, geographic range was logit transformed and body size was natural-log transformed; both of these transformations make these variables defined for the entire real line. Sampling probability was transformed as  $(s(n - 1) + 0.5)/n$  where  $n$  is the sample size as recommended by (**author?**) [176]; this serves to slightly shrink the range of the data so that there are no values of 0 or 1. All covariates except for sampling were standardized by subtracting the mean from all values and dividing by twice its standard deviation, which follows (**author?**) [52]. This standardization means that the associated regression coefficients are comparable as the expected change per 1-unit change in the rescaled covariates. Finally,  $D$  is defined as the total number of covariates, excluding sampling, plus one for the intercept term.

### 3.2.2 Details of model

Hierarchical modelling is a statistical approach which explicitly takes into account the structure of the observed data in order to model both the within and between group variance [51, 52]. The units of study (e.g. genera) each belong to a single group (e.g. origination cohort). Each group is considered a draw from a shared probability distribution (e.g. prior) of all cohorts, observed and unobserved. The group-level parameters, or the hyperparameters of this shared prior, are themselves given (hyper)prior distributions and are also estimated

like the other parameters of interest (e.g. covariate effects) [51]. The subsequent estimates are partially pooled together, where parameters from groups with large samples or effects remain large while those of groups with small samples or effects are pulled towards the overall group mean. All covariate effects (regression coefficients), as well as the intercept term (baseline extinction risk), were allowed to vary by group (origination cohort). The covariance between covariate effects was also modeled.

Genus durations were assumed to follow a Weibull distribution which allows for age-dependent extinction [86]:  $y \sim \text{Weibull}(\alpha, \sigma)$ . The Weibull distribution has two parameters: scale  $\sigma$ , and shape  $\alpha$ . When  $\alpha = 1$ ,  $\sigma$  is equal to the expected duration of any taxon.  $\alpha$  is a measure of the effect of age on extinction risk where values greater than 1 indicate that extinction risk increases with age, and values less than 1 indicate that extinction risk decreases with age. Note that the Weibull distribution is equivalent to the exponential distribution when  $\alpha = 1$ .

In the case of the right- and left-censored observations mentioned above, the probability of those observations has a different calculation [86]. For right-censored observations, the likelihood is calculated  $p(y|\theta) = 1 - F(y) = S(y)$  where  $F(y)$  is the cumulative distribution function. In contrast, the likelihood of a left-censored observation is calculated  $p(y|\theta) = F(y)$ .

The scale parameter  $\sigma$  was modeled as a regression following (**author?**) [87] with both varying intercept and varying slopes and the effect of sampling; this is expressed

$$\sigma_i = \exp\left(\frac{-\mathbf{X}_i B_{j[i]} + \delta s_i}{\alpha}\right) \quad (3.2)$$

where  $i$  indexes across all observations where  $i = 1, \dots, n$  where  $n$  is the total number of observations,  $j[i]$  is the cohort membership of the  $i$ th observation where  $j = 1, \dots, J$  where  $J$  is the total number of cohorts,  $X$  is a  $N \times D$  matrix of covariates along with a column of 1's for the intercept term,  $B$  is a  $J \times D$  matrix of cohort-specific regression coefficients, and

$\delta$  is the regression coefficient for the effect of sampling  $s$ .  $\delta$  does not vary by cohort.

Each of the rows of matrix  $B$  are modeled as realizations from a multivariate normal distribution with length  $D$  location vector  $\mu$  and  $J \times J$  covariance matrix  $\Sigma$ :  $B_j \sim \text{MVN}(\mu, \Sigma)$ . The covariance matrix was then decomposed into a length  $J$  vector of scales  $\tau$  and a  $J \times J$  correlation matrix  $\Omega$ , defined  $\Sigma = \text{diag}(\tau)\Omega\text{diag}(\tau)$  where “diag” indicates a diagonal matrix.

The elements of  $\mu$  were given independent normally distributed priors. The effects of geographic range size and the breadth of environmental preference were given informative priors reflecting the previous findings while the others were given weakly informative favoring no effect. The correlation matrix  $\Omega$  was given an LKJ distributed prior [90] that slightly favors an identity matrix as recommended by [?]. These priors are defined

$$\begin{aligned}\mu^0 &\sim \mathcal{N}(0, 5) \\ \mu^r &\sim \mathcal{N}(-1, 1) \\ \mu^v &\sim \mathcal{N}(0, 1) \\ \mu^{v^2} &\sim \mathcal{N}(1, 1) \\ \mu^m &\sim \mathcal{N}(0, 1) \\ \tau &\sim \text{C}^+(1) \\ \Omega &\sim \text{LKJ}(2).\end{aligned}\tag{3.3}$$

The log of the shape parameter  $\alpha$  was given a weakly informative prior  $\log(\alpha) \sim \mathcal{N}(0, 1)$  centered at  $\alpha = 1$ , which corresponds to the Law of Constant Extinction [184].

### 3.2.3 Imputation of sampling probability

The vector sampling  $s$  has two parts: the observed part  $s^o$ , and the unobserved part  $s^u$ . Of the 1130 total observations, 539 have a duration of 3 or more and have an observed gap statistic. The gap statistic for the remaining 591 observations was imputed. As stated above, the unobserved part is the imputed, or filled in, based on the observed part  $s^o$ . Because sampling varies between 0 and 1, I chose to model it as a Beta regression with matrix  $W$  being a  $N \times (D - 1)$  matrix of covariates (i.e. geographic range size, environmental preference, body size) as predictors of sampling; this assumes that the sampling value for all taxa come from the same distribution. Importantly, I make no assumptions of causal structure.

Predicting sampling probability using the other covariate that are then included in the model of duration is acceptable and appropriate in the case of imputation where the sample goal is accurate prediction [154, 52]. Not including these covariates can lead to biased estimates of the imputed variable; if the covariates themselves are related, not including them will bias this correlation towards zero which then leads to improper imputation and inference [154].

The Beta regression is defined

$$s^o \sim \text{Beta}(\phi = \text{logit}^{-1}(X^o\gamma), \lambda), \quad (3.4)$$

where  $\gamma$  is a length  $D$  vector of regression coefficients, and  $X$  defined as above. The Beta distribution used in the regression is reparameterized in terms of a mean parameter

$$\phi = \frac{\alpha}{\alpha + \beta} \quad (3.5)$$

and total count parameter

$$\lambda = \alpha + \beta \quad (3.6)$$

where  $\alpha$  and  $\beta$  are the characteristic parameters of the Beta distribution [51].

The next step is to then estimate  $s^u|s^o, X^o, X^u, \gamma$ , the posterior distribution of which are folded back into  $s$  and used as a covariate of duration (Eq. 3.2). All the elements of  $\gamma$ , and both  $\delta$  (Eq. 3.2) and  $\lambda$  (Eq. 3.4) were given weakly informative priors where

$$\begin{aligned}\gamma &\sim \mathcal{N}(0, 1) \\ \delta &\sim \mathcal{N}(0, 1) \\ \lambda &\sim \text{Pareto}(0.1, 1.5).\end{aligned}\tag{3.7}$$

### 3.2.4 Posterior inference and posterior predictive checks

The joint posterior was approximated using a Markov-chain Monte Carlo routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn Sampler [64] as implemented in the probabilistic programming language Stan [? ]. The posterior distribution was approximated from four parallel chains run for 10,000 steps each, split half warm-up and half sampling and thinned to every 10th sample for a total of 4000 posterior samples. Chain convergence was assessed via the scale reduction factor  $\hat{R}$  where values close to 1 ( $\hat{R} < 1.1$ ) indicate approximate convergence. Convergence means that the chains are approximately stationary and the samples are well mixed [51].

Model adequacy was evaluated using a couple of posterior predictive checks. Posterior predictive checks are a means for understanding model fit or adequacy where the basic idea is that replicated data sets simulated from the fitted model should be similar to the original data and systematic differences between the simulations and observations indicate weaknesses of the model fit [51]. For both approaches used here, each posterior predictive datasets were generated from a unique draw from the posterior distribution of each parameter. The two

posterior predictive checks used in this analysis are a comparison of a non-parametric estimate of the survival function  $S(t)$  from the empirical dataset to the non-parametric estimates of  $S(t)$  from the 100 posterior predictive datasets, and comparison of the observed genus durations to the average posterior predictive estimate of  $\log(\sigma)$  (Eq. 3.2). The former is to see if simulated data has a similar survival pattern to the observed, while the latter is to see if the model systematically over- or under- estimates taxon survival.

### 3.3 Results

Comparison of the posterior predictive estimates of  $S(t)$  to the empirical estimate reveal few obvious biases except for the case of values from the far right tail of observed durations (Fig. 3.1). This result is reinforced by the additional posterior predictive comparison where most estimates are not systematically biased except for a consistent under-estimate of  $\log(\sigma)$  for older taxa (Fig. 3.2). The results of both posterior predictive checks indicate that, for the majority of observations, model fit is generally not biased.

The cohort-level estimate of the effect of geographic range size indicates that as a taxon's geographic range increases, that taxon's duration is expected to increase (Table 3.1). Given the estimates of  $\mu^r$  and  $\tau^r$ , there is a less than 3.7% ( $\pm 0.04\%$  SD) probability that this relationships would be reversed ( $\Pr(\mathcal{N}(\mu^r, \tau^r) > 0)$ ). The between-cohort variance  $\tau^r$  is the lowest of all the regression coefficients (Table 3.1).

Body size is estimated to have no effect on taxon duration, with the estimate being nearly 0 (Table 3.1). The variance between the cohort-specific estimates of the effect of body size  $\tau^m$  is estimated to be greater than the variance of between-cohort estimates of the effect of geographic range size  $\tau^r$ .

The group-level estimate of the effect of environmental preference is estimated from both  $\mu^v$  and  $\mu^{v^2}$ .

The estimate of  $\mu^v$  indicates that epicontinental favoring taxa are expected to have a greater duration than open-ocean favoring taxa (Table 3.1). Additionally, given the estimate of between-cohort variance  $\tau^v$ , there is approximately 18% ( $\pm 7\%$  SD) probability that, for any given cohort, taxa favoring open-ocean environments would have a greater expected duration than taxa favoring epicontinental environments ( $\Pr(\mathcal{N}(\mu^v, \tau^v) > 0)$ ).

The estimate of  $\mu^{v^2}$  indicates that the overall relationship between environmental preference and  $\log(\sigma)$  is concave down (Fig. 3.3), with only a 2.7% ( $\pm 3\%$  SD) probability that any given cohort is convex up ( $\Pr(\mathcal{N}(\mu^{v^2}, \tau^{v^2}) < 0)$ ).

The cohort-specific estimates of all the regression coefficients demonstrate a lot of between cohort variance, with no obvious trends. As indicated in Table 3.1 and detectable visually (Fig. 3.4), the between-cohort estimates for  $\beta^0$ ,  $\beta^r$ , and  $\beta^m$  all have much lower variance than the between-cohort estimates of both  $\beta^v$  and  $\beta^{v^2}$ .

While most cohort-specific estimates are very similar to the overall cohort-level estimate, there are a few notable excursions away from the overall mean (Fig. 3.4). There are simultaneous excursions in both  $\beta^0$  and  $\beta^v$  for cohorts originating in the Givetian (387-382 My) and Frasnian (382-372 My) stages; both of which directly precede the late Devonian mass extinction event at the Frasnian/Famennian boundary. These cohorts are marked by both a high extinction intensity and an increase in expected duration for taxa favoring epicontinental environments over open-ocean ones; this is consistent with the results of (author?) [107].

Cohorts originating from the Silurian through the Early Devonian have a slightly lower extinction intensity than the overall mean; these cohorts are those originating in the Llandovery (443-443 My) through the Emsian (407-393 My). This is also a time period is also when

there is the lowest overall probability that epicontinental favoring taxa are expected to have greater duration than open-ocean favoring taxa. Both the Silurian and Devonian periods are notable for having been periods with a mostly “hothouse” climate, with no polar icecaps and a high sea-level [27, 82, 109].

The cohort-specific relationships between environmental preference and  $\log(\sigma)$  were calculated from the estimates of  $\beta^0$ ,  $\beta^v$ , and  $\beta^{v^2}$  (Fig. 3.5) and reflect how these three parameters act in concert and not just individually (Fig. 3.4). Beyond results already discussed above in the context of the parameters individually, it is notable that the cohort originating in the Kungurian (279-272 My) is least like the overall expected relationship and has the most sharply curved appearance due to a high estimate  $\beta^{v^2}$  (Fig. 3.4). This cohort has the biggest difference in extinction risk between environmental generalists and specialists. The cohorts originating during the Emsian (407-393 My) and Frasnian (382 - 372 My) are tied for second in sharpness of curvature. The least sharply curved cohorts include those originating during Tremadocian (484-477 My), Hirnantian (445-443 My), Llandovery (443-433 My), and Ludlow (427-423 My). Except for the Tremadocian cohort, most of these cohorts originate during the Silurian through the Early Devonian range identified earlier as having lower expected extinction intensity than what is expected from the group-level estimate.

The correlations of the cohort-specific estimates of the regression coefficients are estimated as the off-diagonal elements of the correlation matrix  $\Omega$ . Only two of the elements of  $\Omega$  are distinguishable from 0: the correlation of  $\beta^0$  (extinction intensity) with both  $\beta^r$  and  $\beta^v$  (Fig. 3.6).

There is an approximate 90% probability that the cohort-specific estimates of baseline extinction intensity  $\beta^0$  and the effect of geographic range  $\beta^r$  are negatively correlated; this means that for cohorts experiencing a lower extinction intensity ( $\beta^0$  decreases), the magnitude of the effect of geographic range is expected to decrease as well, and *vice versa*; this is in

contrast to the observation made by (**author?**) [72] with regards to late Cretaceous bivalves.

Similarly, there is an approximate 97.4% probability that the cohort-specific estimates of  $\beta^0$  and  $\beta^v$  are negatively correlated; this means that as extinction intensity increases it is expected that epicontinental taxa become more favored over open-ocean environments (i.e. as  $\beta^0$  increases,  $\beta^v$  decreases).

There is only an approximate 30% probability that  $\beta^r$  and  $\beta^v$  are positively correlated. This lack of cross-correlation may be due in part to the much higher between-cohort variance of the effect of environmental preference  $\tau^v$  than the very small between-cohort variance in the effect of geographic range  $\tau^r$  (Table 3.1); the effect of geographic range might simply not vary enough relative to the much noisier environmental preference.

Comparison of observed values of sampling, as measured by the gap statistic, to random draws from the posterior estimates of the imputed sampling values indicate that they are very similar (Fig. 3.7). This result is very encouraging as this is the ultimate goal of multiple imputation: to fill in missing data with values similar to the observed while taking into account the randomness of that variable [154, 52]. The estimates of  $\delta$  are based on the set of observed values and the entire posterior of imputed values.

Sampling was found to have a negative effect (positive  $\delta$ ) on duration: greater sampling, shorter duration (Table 3.1). While potentially counter intuitive, this result is most likely due to some long lived taxa only be sampled in the stages of the first and last appearance. Also, longer lived taxa have more opportunities to not be sampled than shorter lived taxa. These two factors will lead to this result.

While the effect of sampling appears large compared to the other regression coefficients, this is only because sampling was not standardized like the other covariates. To make the coefficients comparable,  $\delta$  is multiplied by twice the posterior mean of the standard deviation

of sampling probability; the transformed value of  $\delta$  is then 0.642 ( $\pm 0.1$  SD). This effect is relatively small compared to the other covariate effects (Table 3.1). There is then a 98.6% probability that the effect of geographic range  $\mu^r$  has a greater magnitude than  $\delta$ . Similarly,  $\mu^v$  has a 71.8% probability of having a greater magnitude of effect than  $\delta$ . Finally,  $\mu^{v^2}$  has a 100% probability of having a greater magnitude of effect than  $\delta$ .

The Weibull shape parameter  $\alpha$  was found to be approximately 1.36 ( $\pm 0.05$  SD) with a 100% probability of being greater than 1. This result is not consistent with the Law of Constant Extinction [184] and is instead consistent with accelerating extinction risk with taxon age. This may indicate that older taxa are out-competed by younger taxa, a result consistent with some empirical results [192, 129? ] and (ironically) with a recently proposed Red Queen-like model of evolution [149]. This results, however, is not consistent with other empirical results [33, 25] and could potentially be caused by the minimum resolution of the fossil record [156]. It is thus unclear if a strong biological inference can be made from this result, which means that further work is necessary on the effect of taxon age on extinction risk.

### 3.4 Discussion

The generating observation behind this study was that for bivalves at the end Cretaceous mass extinction event, the only biological trait that was found the affect extinction risk was geographic range while traits that had previously been beneficial had no effect [72]. This observation raises two linked questions: how does the effect of geographic range change with changing extinction intensity, and how does the effect of other biological traits change with changing extinction intensity?

I find that as intensity increases ( $\beta^0$  decreases), the magnitude of the effect of geographic range increases. I also find that as intensity increases, the effect of favoring epicontinental

environments of open-ocean environments is expected to be increase; this is consistent with the results of (**author?**) [107]. There is no evidence for a correlation between the effect of geographic range and environmental preference. Additionally, the between-cohort variance in effect of geographic range is much less than the between-cohort variance of the effect of environmental preference which may underlie the lack of correlation between these two effects.

Additionally, the lower between-cohort variance in the effect of geographic range versus that higher between-cohort variance implies that for cohorts with a greater than average extinction intensity, the difference in the effect geographic range and the group-level effect of geographic range is expected to be smaller than the difference between the effect of environmental preference and the group-level effect of environmental preference.

I find consistent support for the “survival of the unspecialized,” with respect to epicontinental versus open-ocean environmental preference, as a time-invariant generalization of brachiopod survival; taxa with intermediate environmental preferences are expected to have lower extinction risk than taxa specializing in either epicontinental or open-ocean environments (Fig. 3.3), though the curvature of the relationship varies from rather shallow to very peaked (Fig. 3.5). However, this relationship is not symmetric about 0, as taxa favoring epicontinental environments are expected to have a greater duration than taxa favoring open-ocean environments. This description of environment only describes one major aspect of a taxon’s environmental context, with factors such as bathymetry and temperature being further descriptors of a taxon’s adaptive zone [114, 58, 59, 62]; inclusion of these factors in future analyses would potentially improve our understanding of the “survival of the unspecialized” hypothesis [168].

(**author?**) [65], in their analysis of niche conservatism and substrate lithological preference in marine invertebrates, found that brachiopods were among the least “conservative” groups; taxa were found to easily change substrate preference on short time scales. While substrate

preference is not the same as environmental preference (as defined here), a question does arise: are there three classes of environmental preference instead of two? These classes would be taxa with broad tolerance (“true” generalists), inflexible specialists (“true” specialists), and flexible but with a narrow tolerance. A flexible taxon is one with a narrow habitat preference at one time, but with preference that changes over time with changing environmental availability. My analysis assumes that traits are constant over the duration of the taxon meaning that this scenario is not detectable; taxa with broad tolerances and flexible taxa with narrow per-stage preference end up being treated the same way. Future work should explore how environmental preference changes over lineage duration in relation to environmental availability to estimate if the generalists–specialists continuum is actually ternary relationship.

An alternative approach for specifically modeling survival that can take into account imperfect observation than the method used here is the Cormack-Jolly-Seber (CJS) model [153, 91, 183, 95]. This model is a type of hidden Markov model with an absorbing state (i.e. extinction). In this model, survival is defined as the probability of surviving from time  $t$  to time  $t + 1$ . Additionally, the effect of preservation and sighting is estimated as probability of observing a taxon that is present; this can extend the duration of a taxon beyond its last occurrence. This approach is a fundamentally different from the method used in my analysis: I am estimating the biasing effect of sampling probability on taxon duration to then compare with effects of other covariates, while the CJS model estimates the pre-sampling fossil record and then estimates per-time unit survival probability.

The use of genera as the unit of the study and how to exactly interpret the effects of the biological traits is an important question. For example, if any of the traits analyzed here are associated with increases in speciation rates, this might increase the duration of genera through self-renewal [140, 141], which would be an example of the difference in biological pattern between species and genera [71, 73, 74]. This could lead to a trait appearing to

decrease generic level extinction risk by that trait increasing species level origination rate instead of decreasing species level extinction risk.

The model used here could be improved through either increasing the number of analyzed traits, expanding the hierarchical structure of the model to include other major taxonomic groups of interest, and the inclusion of explicit phylogenetic relationships between the taxa in the model as an additional hierarchical effect. An example trait that may be of particular interest is the affixing strategy or method of interaction with the substrate of the taxon, which has been found to be related to brachiopod survival where, for cosmopolitan taxa, taxa that are attached to the substrate are expected to have a greater duration than those that are not [2].

It is theoretically possible to expand this model to allow for comparisons both within and between major taxonomic groups which would better constrain the brachiopod estimates while also allowing for estimation of similarities and differences in cross-taxonomic patterns. The major issue surrounding this particular expansion involves finding a similarly well sampled taxonomic group that is present during the Paleozoic. Potential groups include Crinoidea, Ostracoda, and other members of the “Paleozoic fauna” [157].

With significant updates, it would also be possible to compare the brachiopod record with with Modern groups such as bivalves or gastropods [157], though remembering that the groups may not necessarily share all cohorts with the brachiopods. This particular model expansion would act as a test of any universal cross-taxonomic patterns in the effects of emergent traits on extinction such as has been proposed for geographic range [117]. Additionally, this expanded model would also act as a test of the distinctness of the (**author?**) [157] three-fauna hypothesis in terms of trait-dependent extinction.

Traits like environmental preference or geographic range [71? ] are most likely heritable.

Without phylogenetic context, this analysis assumes that differences in extinction risk between taxa are independent of the shared evolutionary history of those taxa [31]. In contrast, the origination cohorts only capture shared temporal context. For example, if taxon duration is phylogenetically heritable, then closely related taxa may have more similar durations as well as more similar biological traits. Without taking into account phylogenetic similarity the effects of these biological traits would be inflated solely due to inheritance. The inclusion of phylogenetic context as an additional individual-level hierarchical effect, independent of origination cohort, would allow for determining how much of the observed variability is due to shared evolutionary history versus shared temporal context versus actual differences associated with biological traits [? ].

The combination and integration of the phylogenetic comparative and paleontological approaches requires both sources of data, something which is not possible for this analysis because there is no phylogenetic hypothesis for all Paleozoic taxa, something that is frequently the case for marine invertebrates with a good fossil record. When both data sources are available has been possible, however, the analysis can more fully address the questions of interest in macroevolution [? 171? , 167, 183, 170, 136, 135, 60, 48].

In summary, patterns of Paleozoic brachiopod survival were analyzed using a fully Bayesian hierarchical survival modelling approach while also eschewing the traditional separation between background and mass extinction. I find that cohort extinction intensity is negatively correlated with both the cohort-specific effects of geographic range and environmental preference. These results imply that as extinction intensity increases ( $\beta^0$ ) increases, it is expected that both effects will increase in magnitude. However, the change in effect of environmental preference is expected to be greater than the change in the effect of geographic range. Additionally, I find support for greater survival in environmental generalists over specialists in all origination cohorts analyzed; this is consistent with the long standing

“survival of the unspecialized” hypothesis [92, 93, 168, 169? , 115, 114, 12]. The results of this analysis support the conclusion that for Paleozoic brachiopods, as extinction intensity increases overall extinction selectivity is expected to increase as well.

Table 3.1: Estimates of various parameters in the model used here. These include group-level estimates of the effects of biological traits on brachiopod generic survival, the standard deviation of the between-cohort effects, as well as the estimates of both the effect of sampling  $\delta$  and the Weibull shape parameter  $\alpha$ . The mean, standard deviation (SD), 10th, 50th, and 90th quantiles of the marginal posteriors are presented.

type	parameter	effect of	mean	SD	10%	50%	90%
Mean	$\mu^i$	intercept	-3.05	0.20	-3.30	-3.05	-2.80
	$\mu^r$	geographic range	-0.98	0.16	-1.18	-0.98	-0.79
	$\mu^v$	environmental preference	-0.76	0.19	-0.99	-0.76	-0.52
	$\mu^{v^2}$	environmental preference <sup>2</sup>	3.15	0.36	2.69	3.15	3.62
Standard deviation	$\mu^m$	body size	-0.01	0.13	-0.17	-0.01	0.15
	$\tau^i$	intercept	0.51	0.11	0.38	0.50	0.65
	$\tau^r$	geographic range	0.50	0.16	0.30	0.49	0.71
	$\tau^v$	environmental preference	0.84	0.17	0.63	0.82	1.05
Other	$\tau^{v^2}$	environmental preference <sup>2</sup>	1.51	0.36	1.08	1.48	1.97
	$\tau^m$	body size	0.47	0.13	0.32	0.46	0.64
	$\delta$	sampling	0.90	0.15	0.71	0.90	1.09
	$\alpha$	“time”	1.36	0.04	1.30	1.36	1.42

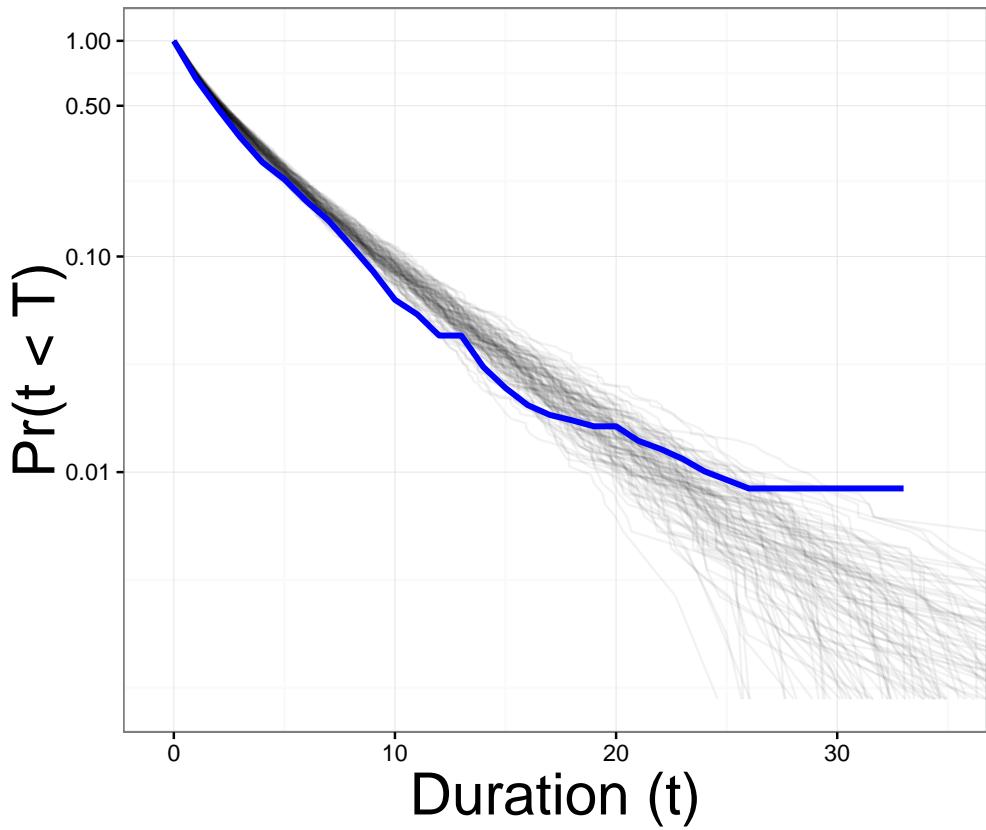


Figure 3.1: Comparison of the empirical estimate of  $S(t)$  (highlighted) versus estimates from 100 posterior predictive data sets (black).  $S(t)$  corresponds to the probability that the age of a genus  $t$  is less than the genus' ultimate duration  $T$ . The vertical axis is log10 transformed.

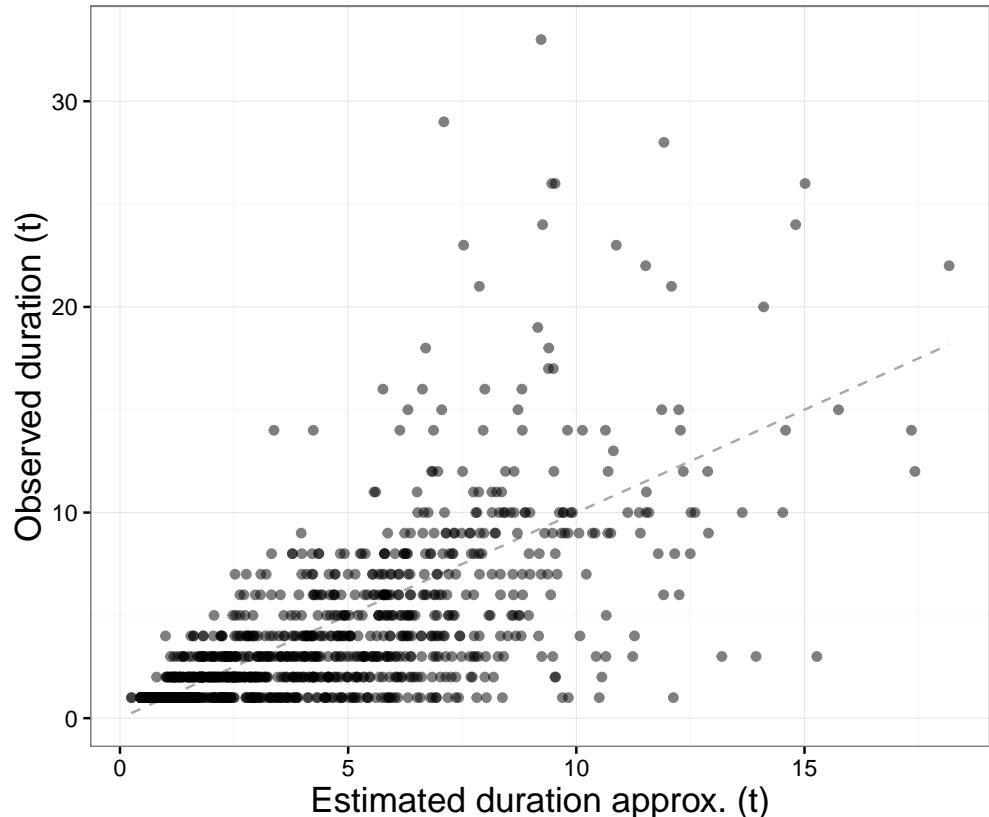


Figure 3.2: Comparison of all observed genus durations in number of geological stages to the average posterior predictive estimates of  $\log(\sigma)$ . The dashed, diagonal line corresponds to  $x = y$ .

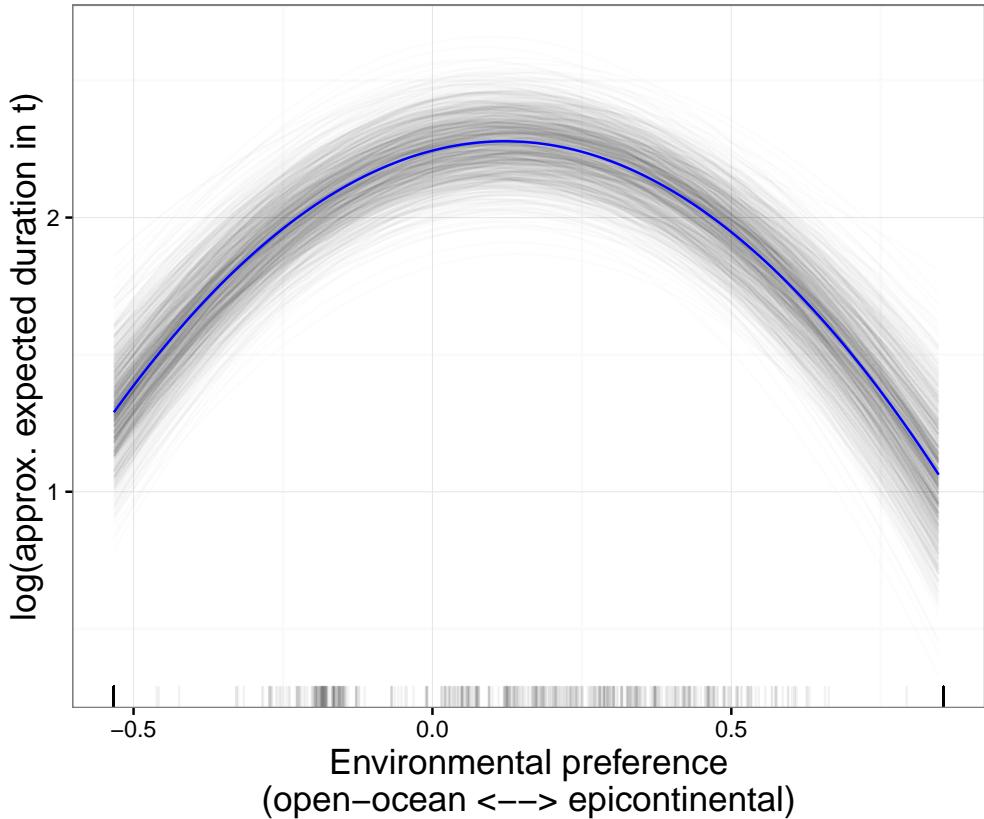


Figure 3.3: The overall expected relationship between environmental affinity  $v_i$  and  $\log(\sigma)$  when  $r = 0$  and  $m = 0$ . The 1000 semi-transparent lines corresponds to a single draw from the posterior predictive distribution, while the highlighted line corresponds to the median of the posterior predictive distribution. The overall relationship is concave down with an optimum greater than 0, which means that taxa favoring epicontinental environments are expected to have longer durations than those favoring open-ocean environments. The tick marks along the bottom of the plot correspond to the (rescaled) observed values of environmental preference.

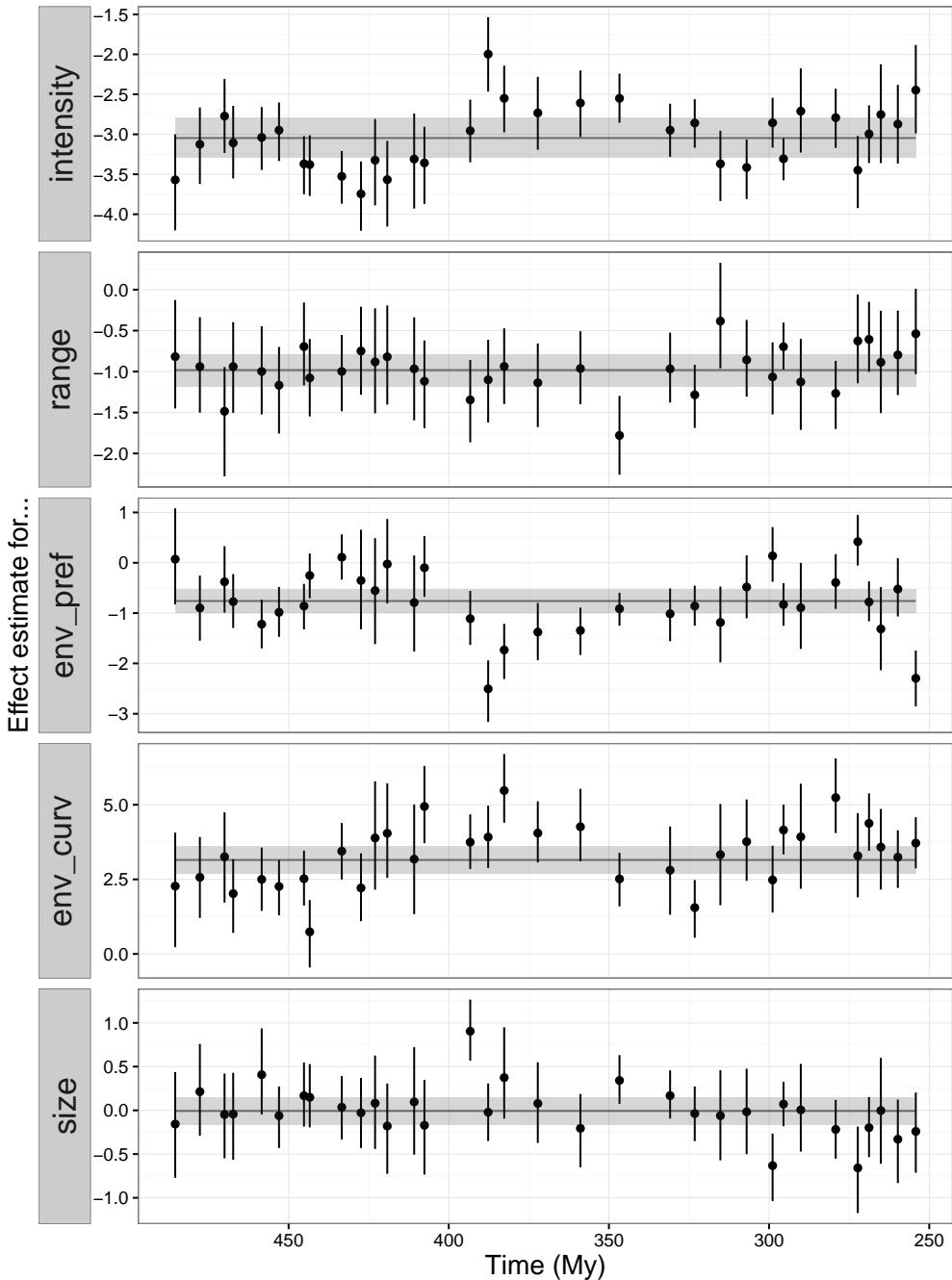


Figure 3.4: Comparison of cohort-specific estimates of  $\beta^0$ , the effect of geographic range on extinction risk  $\beta^r$ , the effect of environmental preference  $\beta^v$  and  $\beta^{v^2}$ , and body size  $\beta^m$ . Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. Points are plotted at the midpoint of the cohorts stage of origination in millions of years before present (My). Black, horizontal lines are the overall estimates of covariate effects along with 80% credible intervals (shaded).

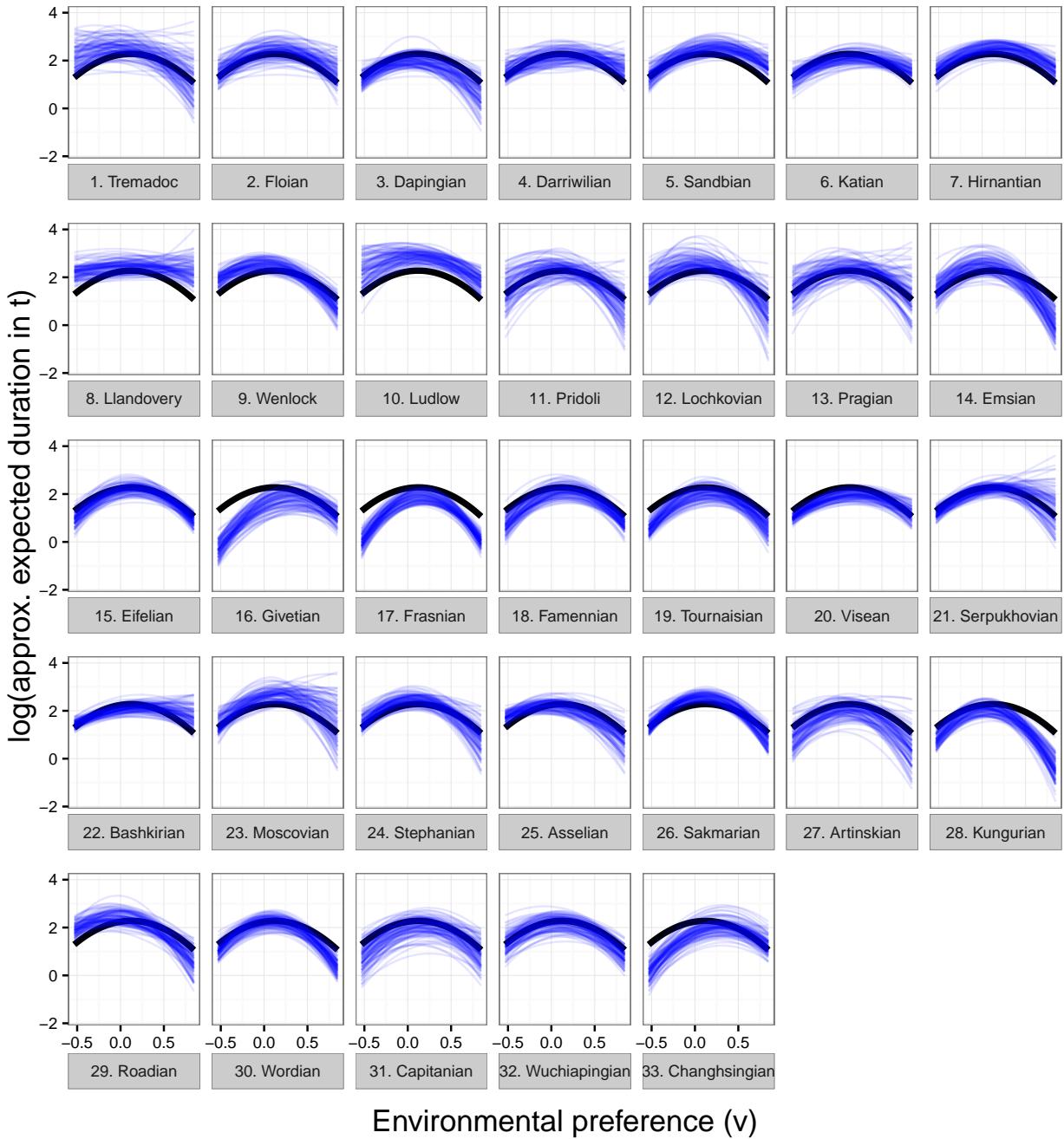


Figure 3.5: Comparison of origination cohort-specific (posterior predictive) estimates of the effect of environmental preference on  $\log(\sigma)$  to the mean overall estimate of the effect of environmental preference. Cohort-specific estimates are from 100 posterior predictive simulations across the range of (transformed and rescaled) observed values of environmental preference. The oldest cohort is in the top-right and younger cohorts proceed left to right, with the youngest cohort being the right-most facet of the last row. Panel names correspond to the name of the stage in which that cohort originated.

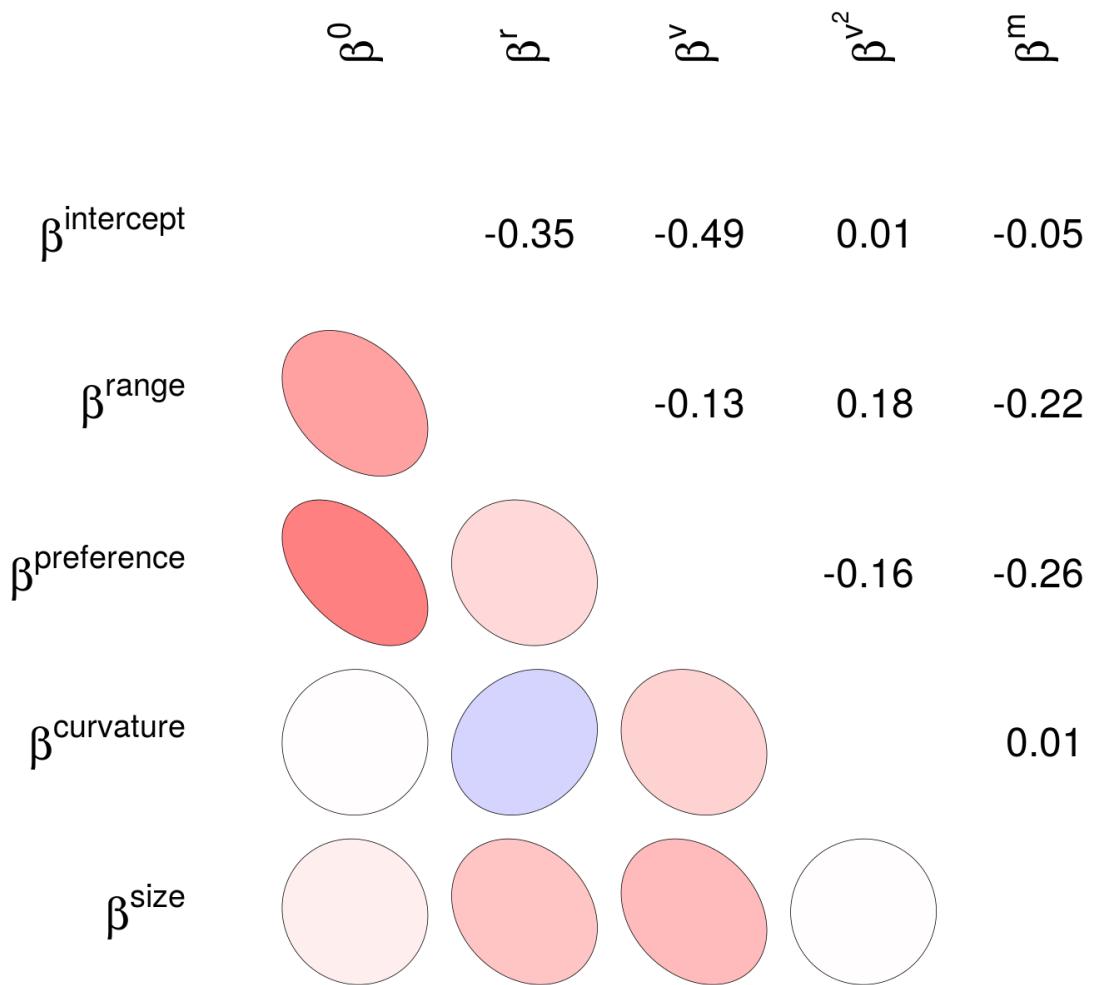


Figure 3.6: Mixed graphical and numerical representation of the correlation matrix  $\Omega$  of variation in cohort-specific covariate estimates. These correlations are between the estimates of the cohort-level effects of covariates, along with intercept/baseline extinction risk. The median estimates of the correlations are presented numerically (upper-triangle) and as idealized ellipses representing that much correlation (lower-triangle). The darkness of the ellipse corresponds to the magnitude of the correlation.

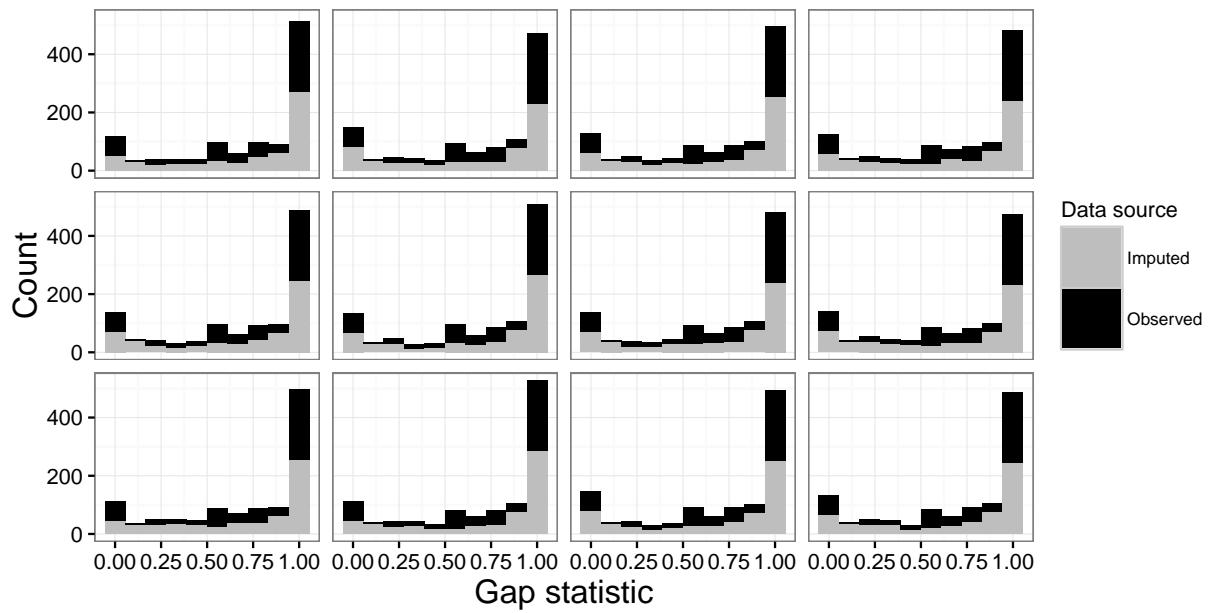


Figure 3.7: Histograms of the distribution of gap statistic values from both the observed values and the imputed values from 12 unique posterior realizations. For each panel the observed values are identical but the imputed values are from a single set of their posterior estimates.

## **CHAPTER 4**

### **TAXON OCCURRENCE AS A FUNCTION OF BOTH EMERGENT BIOLOGICAL TRAITS AND ITS ENVIRONMENTAL CONTEXT**

Place holder text.

## **CHAPTER 5**

### **CONCLUSION**

## REFERENCES

- [1] Jordi Agustí, Lluís Cabrera, and Miguel Garcés. The Vallesian Mammal Turnover: A Late Miocene record of decoupled land-ocean evolution. *Geobios*, 46(1-2):151–157, jan 2013.
- [2] Richard R Alexander. Generic longevity of articulate brachiopods in relation to the mode of stabilization on the substrate. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 21:209–226, 1977.
- [3] John Alroy. Speciation and extinction in the fossil record of North American mammals. In Roger K Butlin, Jon R Bridle, and Dolph Schlüter, editors, *Speciation and patterns of diversity*, pages 302–323. Cambridge University Press, Cambridge, 2009.
- [4] John Alroy. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. In John Alroy and Gene Hunt, editors, *Quantitative Methods in Paleobiology*, pages 55–80. The Paleontological Society, 2010.
- [5] John Alroy. The Shifting Balance of Diversity Among Major Marine Animal Groups. *Science*, 329(5996):1191–1194, 2010.
- [6] John Alroy. A simple Bayesian method of inferring extinction. *Paleobiology*, 40(4):584–607, jul 2014.
- [7] John Alroy, Paul L Koch, and James C Zachos. Global climate change and North American mammalian evolution. *Paleobiology*, 26(1981):259–288, 2000.
- [8] Mike Austin. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, 200:1–19, 2007.
- [9] Richard K. Bambach, Andrew M. Bush, and Douglas H. Erwin. Autecology and the filling of ecospace: Key metazoan radiations. *Palaeontology*, 50(1):1–22, 2007.
- [10] David W Bapst. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, 3:803–807, 2012.
- [11] David W. Bapst. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods in Ecology and Evolution*, 4(8):724–733, aug 2013.
- [12] Tomasz K Baumiller. Survivorship analysis of Paleozoic Crinoidea: effect of filter morphology on evolutionary rates. *Paleobiology*, 19(3):304–321, 1993.
- [13] Olaf R P Bininda-Emonds, Marcel Cardillo, Kate E Jones, Ross D E Macphee, Robin M D Beck, Richard Grenyer, Samantha A Price, Rutger A Vos, John L Gittleman, and Andy Purvis. The delayed rise of present-day mammals. *Nature*, 446(7135):507–512, 2007.

- [14] Jonathan I Bloch, Kenneth D Rose, and Philip D Gingerich. New species of Batodonoides (Lipotyphla, Geolabididae) from the Early Eocene of Wyoming: smallest known mammal? *Journal of Mammalogy*, 79(3):804–827, 1998.
- [15] Jessica L Blois and Elizabeth A Hadly. Mammalian Response to Cenozoic Climatic Change. *Annual Review of Earth and Planetary Sciences*, 37(1):181–208, may 2009.
- [16] Fred L Bookstein. Random walk and the existence of evolutionary rates. *Paleobiology*, 13(4):446–464, 1987.
- [17] Barry W. Brook and David M. J. S. Bowman. The uncertain blitzkrieg of Pleistocene megafauna. *Journal of Biogeography*, 31(4):517–523, apr 2004.
- [18] Alexandra M. Brown, David I. Warton, Nigel R. Andrew, Matthew Binns, Gerasimos Cassis, and Heloise Gibb. The fourth-corner solution - using predictive models to understand how species traits interact with the environment. *Methods in Ecology and Evolution*, 5(4):344–352, apr 2014.
- [19] James H Brown and Brian A Maurer. Macroecology: the division of food and space among species on continents. *Science*, 243(4895):1145–1150, 1989.
- [20] James J Brown. *Macroecology*. University of Chicago Press, Chicago, 1995.
- [21] Andrew M Bush, Richard K Bambach, and Gwen M Daley. Changes in theoretical ecospace utilization in marine fossil assemblages between the mid-Paleozoic and late Cenozoic. *Paleobiology*, 33(1):76–97, 2007.
- [22] Anne Chao, T. C. Hsieh, Robin L. Chazdon, Robert K. Colwell, Nicholas J. Gotelli, and B. D. Inouye. Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology*, 96(5):1189–1201, 2015.
- [23] William S Cooper. Expected time to extinction and the concept of fundamental fitness. *Journal of Theoretical Biology*, 107:603–629, 1984.
- [24] Jerry A Coyne and H Allen Orr. *Speciation*. Sinauer Associates, Sunderland, MA, 2004.
- [25] James S. Crampton, Roger A. Cooper, Peter M. Sadler, and Michael Foote. Greenhouse–icehouse transition in the Late Ordovician marks a step change in extinction regime in the marine plankton. *Proceedings of the National Academy of Sciences*, 113(6):1498–1503, 2016.
- [26] Stéphane Dray and Pierre Legendre. Testing the species traits-environment relationships: the fourth-corner problem revisited. *Ecology*, 89(12):3400–3412, 2008.
- [27] Dianne Edwards and Una Fanning. Evolution and environment in the late Silurian–early Devonian: the rise of pteridophytes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 309:147–165, 1985.

- [28] Jane Elith and John R Leathwick. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40:677–697, 2009.
- [29] Jussi T Eronen, Alistair R Evans, Mikael Fortelius, and Jukka Jernvall. The impact of regional climate on the evolution of mammals: a case study using fossil horses. *Evolution*, 64(2):398–408, 2009.
- [30] Thomas H G Ezard, Paul N Pearson, Tracy Aze, and Andy Purvis. The meaning of birth and death (in macroevolutionary birth-death models). *Biology Letters*, 8(1):139–42, mar 2012.
- [31] J Felsenstein. Phylogenies and the comparative method. *The American Naturalist*, 125(1):1–15, 1985.
- [32] S. Finnegan, N. A. Heim, S. E. Peters, and W. W. Fischer. Climate change and the selective signature of the Late Ordovician mass extinction. *Proceedings of the National Academy of Sciences*, 109:6829–6834, 2012.
- [33] Seth Finnegan, Jonathan L. Payne, and Steve C. Wang. The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions. *Paleobiology*, 34(3):318–341, sep 2008.
- [34] R G Fitzjohn. Quantitative Traits and Diversification. *Systematic Biology*, 59(6):619–633, 2010.
- [35] Michael Foote. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology*, 14(3):258–271, 1988.
- [36] Michael Foote. Origination and extinction through the Phanerozoic: a new approach. *Journal of Geology*, 111:125–148, 2003.
- [37] Michael Foote. Environmental controls on geographic range size in marine animal genera. *Paleobiology*, 40(3):440–458, 2014.
- [38] Michael Foote and Arnold I Miller. *Principles of Paleontology*. Freeman, New York, third edition, 2007.
- [39] Michael Foote and Arnold I. Miller. Determinants of early survival in marine animal genera. *Paleobiology*, 39(2):171–192, mar 2013.
- [40] Michael Foote and DM Raup. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, 22(2):121–140, 1996.
- [41] Mike Foote. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology*, 26(sp4):74–102, dec 2000.

- [42] Mike Foote. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology*, 26(4):578–605, dec 2000.
- [43] Mikael Fortelius, Jussi Eronen, Jukka Jernvall, Liping Liu, Diana Pushkina, Juhani Rinne, Alexey Tesakov, Inesa Vislobokova, Zhaoqun Zhang, and Liping Zhou. Fossil mammals resolve regional patterns of Eurasian climate change over 20 million years. *Evolutionary Ecology Research*, 4:1005–1016, 2002.
- [44] John R Foster. Preliminary body mass estimates for mammalian genera of the Morrison Formation (Upper Jurassic, North America). *PaleoBios*, 28:114–122, 2009.
- [45] Matthijs Freudenthal and Elvira Martín-Suárez. Estimating body mass of fossil rodents. *Scripta Geologica*, 145:1–130, 2013.
- [46] Susanne A Fritz, Olaf R P Bininda-Emonds, and Andy Purvis. Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. *Ecology Letters*, 12(6):538–49, jun 2009.
- [47] Susanne A Fritz and Andy Purvis. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4):1042–51, aug 2010.
- [48] Susanne A Fritz, Jan Schnitzler, Jussi T Eronen, Christian Hof, Katrin Böhning-Gaese, and Catherine H Graham. Diversity in time and space: wanted dead and alive. *Trends in Ecology & Evolution*, 28(9):509–16, sep 2013.
- [49] Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.
- [50] Andrew Gelman. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, pages 2865–2873, 2008.
- [51] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall, Boca Raton, FL, 3 edition, 2013.
- [52] Andrew Gelman and Jennifer Hill. *Data Analysis using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, New York, NY, 2007.
- [53] PD Gingerich. Quantification and comparison of evolutionary rates. *American Journal of Science*, 293:453–478, 1993.
- [54] Emma E Goldberg, Lesley T Lancaster, and Richard H Ree. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, 60(4):451–65, jul 2011.

- [55] Emma E Goldberg, Kaustuv Roy, Russell Lande, and David Jablonski. Diversity, endemism, and age distributions in macroevolutionary sources and sinks. *The American Naturalist*, 165(6):623–33, jun 2005.
- [56] Harvey Goldstein, William Browne, and Jon Rasbash. Partitioning variation in multi-level models. *Understanding Statistics*, 1(4):1–12, 2002.
- [57] Cynthia L Gordon. A First Look at Estimating Body Size in Dentally Conservative Marsupials. *Journal of Mammalian Evolution*, page 21, 2003.
- [58] P. G. Harnik, C. Simpson, and J. L. Payne. Long-term differences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749):4969–4976, oct 2012.
- [59] Paul G Harnik. Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proceedings of the National Academy of Sciences*, 108(33):13594–13599, aug 2011.
- [60] Paul G. Harnik, Paul C. Fitzgerald, Jonathan L. Payne, and Sandra J. Carlson. Phylogenetic signal in extinction selectivity in Devonian terebratulide brachiopods. *Paleobiology*, 40(4):675–692, jun 2014.
- [61] Matthew M Hedman. Constraints on clade ages from fossil outgroups. *Paleobiology*, 36(1):16–31, 2010.
- [62] Noel A Heim and Shanan E Peters. Regional environmental breadth predicts geographic range and longevity in fossil marine genera. *PLoS one*, 6(5):e18946, jan 2011.
- [63] Kenneth B. Hoehn, Paul G. Harnik, and V. Louise Roth. A framework for detecting natural selection on traits above the species level. *Methods in Ecology and Evolution*, pages doi: 10.1111/2041–210X.12461, 2015.
- [64] Matthew D Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1351–1381, 2014.
- [65] Melanie J. Hopkins, Carl Simpson, and Wolfgang Kiessling. Differential niche dynamics among major marine invertebrate clades. *Ecology Letters*, 17(3):314–323, 2014.
- [66] Elizabeth A Housworth, P Martins, and Michael Lynch. The Phylogenetic Mixed Model. *The American Naturalist*, 163(1):84–96, 2004.
- [67] G Hunt. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. *Proceedings of the National Academy of Sciences*, 104:18404–18408, 2007.
- [68] Gene Hunt. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology*, 32(4):578–601, 2006.

- [69] Gene Hunt and Daniel L Rabosky. Phenotypic Evolution in Fossil Species: Pattern and Process. *Annual Review of Earth and Planetary Sciences*, 42(1):421–441, 2014.
- [70] Joseph G Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. *Bayesian Survival Analysis*. Springer, New York, 2001.
- [71] D Jablonski. Heritability at the species level: analysis of geographic ranges of cretaceous mollusks. *Science*, 238(4825):360–363, oct 1987.
- [72] David Jablonski. Background and mass extincitons: the alternation of macroevolutionary regimes. *Science*, 231(4734):129–133, 1986.
- [73] David Jablonski. Scale and hierarchy in macroevolution. *Palaeontology*, 50(September 2006):87–109, 2007.
- [74] David Jablonski. Species Selection: Theory and Data. *Annual Review of Ecology, Evolution, and Systematics*, 39(1):501–524, dec 2008.
- [75] David Jablonski and Kaustuv Roy. Geographical range and speciation in fossil and living molluscs. *Proceedings of the Royal Society B: Biological Sciences*, 270(1513):401–406, feb 2003.
- [76] Tahira Jamil, Wim A. Ozinga, Michael Kleyer, and Cajo J F Ter Braak. Selecting traits that explain species-environment relationships: A generalized linear mixed model approach. *Journal of Vegetation Science*, 24(6):988–1000, 2013.
- [77] Christine M Janis. Tertiary mammal evolution in the context of changing climates, vegetation, and tectonic events. *Annual Review of Ecology and Systematics*, 24:467–500, 1993.
- [78] Christine M Janis, Gregg F Gunnell, and Mark D Uhen. *Evolution of Tertiary mammals of North America. Vol. 2. Small mammals, xenarthrans, and marine mammals*. Cambridge University Press, Cambridge, 2008.
- [79] Christine M Janis, K M Scott, and L L Jacobs. *Evolution of Tertiary mammals of North America. Vol. 1. Terrestrial carnivores, ungulates, and ungulatelike mammals*. Cambridge University Press, Cambridge, 1998.
- [80] Jukka Jernvall and Mikael Fortelius. Common mammals drive the evolutionary increase of hypsodonty in the Neogene. *Nature*, 417(6888):538–40, may 2002.
- [81] Jukka Jernvall and Mikael Fortelius. Maintenance of trophic structure in fossil mammal communities: site occupancy and taxon resilience. *The American Naturalist*, 164(5):614–624, nov 2004.
- [82] M. M. Joachimski, S. Breisig, W. Buggisch, J. A. Talent, R. Mawson, M. Gereke, J. R. Morrow, J. Day, and K. Weddige. Devonian climate and reef evolution: Insights from oxygen isotopes in apatite. *Earth and Planetary Science Letters*, 284(3-4):599–609, 2009.

- [83] J G Johnson. Extinction of Perched Faunas. *Geology*, 2:479–482, 1974.
- [84] Kate E Jones, Jon Bielby, Marcel Cardillo, Susanne A Fritz, Justin O'Dell, C David L Orme, Kamran Safi, Wes Sechrest, E Boakes, C Carbone, C Connolly, M J Cutts, J K Foster, R Grenyer, M Habib, C A Plaster, S A Price, E A Rigby, J Rist, Amber Teacher, Olaf R P Bininda-Emonds, John L Gittleman, Georgina M Mace, and Andy Purvis. PanTHERIA : a species-level database of life history , ecology , and geography of extant and recently extinct mammals. *Ecology*, 90(9):2648, 2009.
- [85] Wolfgang Kiessling and Martin Aberhan. Environmental determinants of marine benthic biodiversity dynamics through Triassic Jurassic time. *Paleobiology*, 33(3):414–434, 2007.
- [86] John P Klein and Melvin L Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, 2nd edition, 2003.
- [87] D G Kleinbaum and M Klein. *Survival analysis: a self-learning text*. Springer, New York, NY, 2 edition, 2005.
- [88] Pierre Legendre, René Galzin, and Mireille L Harmelin-Vivien. Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology*, 78(2):547–562, 1997.
- [89] Serge Legendre. Analysis of mammalian communities from the Late Eocene and Oligocene of Southern France. *Paleovertebrata*, 16(4):191–212, 1986.
- [90] Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001, oct 2009.
- [91] L H Liow, Mikael Fortelius, E Bingham, K Lintulaakso, H Mannila, L Flynn, and N C Stenseth. Higher origination and extinction rates in larger mammals. *Proceedings of the National Academy of Sciences*, 105(16):6097–6102, 2008.
- [92] Lee Hsiang Liow. A test of Simpson's "rule of the survival of the relatively unspecialized" using fossil crinoids. *The American Naturalist*, 164(4):431–43, oct 2004.
- [93] Lee Hsiang Liow. Does versatility as measured by geographic range, bathymetric range and morphological variability contribute to taxon longevity? *Global Ecology and Biogeography*, 16(1):117–128, 2007.
- [94] Lee Hsiang Liow, Mikael Fortelius, Kari Lintulaakso, Heikki Mannila, and Nils Chr Stenseth. Lower Extinction Risk in SleeporHide Mammals. *The American Naturalist*, 173(2):264–272, 2009.
- [95] Lee Hsiang Liow and James D Nichols. Estimating rates and probabilities of origination and extinction using taxonomic occurrence data: Capture-mark-recapture (CMR) approaches. In John Alroy and Gene Hunt, editors, *Quantitative Methods in Paleobiology*, pages 81–94. The Paleontological Society, 2010.

- [96] Lee Hsiang Liow, Tiago B Quental, and Charles R Marshall. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? *Systematic biology*, 59(6):646–59, dec 2010.
- [97] Lee Hsiang Liow, Leigh Van Valen, and Nils Chr Stenseth. Red Queen: from populations to taxa and communities. *Trends in Ecology & Evolution*, 26(7):349–358, jul 2011.
- [98] Elisabeth A Lloyd and Stephen J Gould. Species selection on variability. *Proceedings of the National Academy of Sciences*, 90:595–599, 1993.
- [99] Zhe-Xi Luo, Alfred W Crompton, and Ai-Lin Sun. A New Mammaliaform from the Early Jurassic and Evolution of Mammalian Characteristics. *Science*, 292:1535–1540, 2001.
- [100] Michael Lynch. Methods for the analysis of comparative data in evolutionary biology. *Evolution*, 45(5):1065–1080, 1991.
- [101] W P Maddison, P E Midford, and S P Otto. Estimating a binary character’s effect on speciation and extinction. *Systematic Biology*, 56(5):701, 2007.
- [102] Jonathan D. Marcot. The fossil record and macroevolutionary history of North American ungulate ungulate mammals: standardizing variation in intensity and geography of sampling. *Paleobiology*, 40(2):237–254, feb 2014.
- [103] Richard McElreath. *Statistical rethinking: a Bayesian course with examples in R and Stan*. CRC Press, Boca Raton, FL, 2016.
- [104] Brian J McGill, Brian J Enquist, Evan Weiher, and Mark Westoby. Rebuilding community ecology from functional traits. *TRENDS in Ecology and Evolution*, 21(4):178–185, 2006.
- [105] Ryan Thomas McKenna. *Potential for Speciation in Mammals Following Vast, Late Miocene Volcanic Interruptions in the Pacific Northwest*. Masters, Portland State University, 2011.
- [106] M. Mendoza, C. M. Janis, and P. Palmqvist. Estimating the body mass of extinct ungulates: a study on the use of multiple regression. *Journal of Zoology*, 270:90–101, may 2006.
- [107] Arnold I Miller and Michael Foote. Epicontinental seas versus open-ocean settings: the kinetics of mass extinction and origination. *Science*, 326(5956):1106–9, nov 2009.
- [108] Isaac Casanovas-Vilar Salvador Moyà-Solà, Jordi Agustí, and Meike Kohler. 9 The geography of a faunal turnover : tracking the Vallesian Crisis. In Ashraf M T Elewa, editor, *Migration of Organisms: Climate, geography, ecology*, pages 247–300. Springer, Berlin, 2005.

- [109] Axel Munnecke, Mikael Calner, David A T Harper, and Thomas Servais. Ordovician and Silurian sea-water chemistry, sea level, and climate: A synopsis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 296(3-4):389–413, 2010.
- [110] S Nee. Inferring speciation rates from phylogenies. *Evolution*, 55(4):661–8, apr 2001.
- [111] S Nee, a O Mooers, and P H Harvey. Tempo and mode of evolution revealed from molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America*, 89(17):8322–6, sep 1992.
- [112] Sean Nee. Birth-Death Models in Macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):1–17, dec 2006.
- [113] Sean Nee, RM May, and PH Harvey. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 344:305–311, 1994.
- [114] Sabine Nürnberg and Martin Aberhan. Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*, 39(3):360–372, apr 2013.
- [115] Sabine Nürnberg and Martin Aberhan. Interdependence of specialization and biodiversity in Phanerozoic marine invertebrates. *Nature Communications*, 6:6602, jan 2015.
- [116] Michael E Palmer and Marcus W Feldman. Survivability is more fundamental than evolvability. *PloS one*, 7(6):e38025, jan 2012.
- [117] Jonathan L Payne and Seth Finnegan. The effect of geographic range on extinction risk during background and mass extinction. *Proceedings of the National Academy of Sciences*, 104:10506–11, jun 2007.
- [118] Jonathan L Payne, Noel A Heim, Matthew L Knope, and Craig R McClain. Metabolic dominance of bivalves predates brachiopod diversity decline by more than 150 million years. *Proceedings of the Royal Society B: Biological Sciences*, 281:20133122, 2014.
- [119] Matthew W Pennell, Luke J Harmon, and Josef C Uyeda. Is there room for punctuated equilibrium in macroevolution? *Trends in ecology & evolution*, 29(1):23–32, jan 2014.
- [120] Shanan E. Peters. The problem with the Paleozoic. *Paleobiology*, 33(2):165–181, 2007.
- [121] Shanan E Peters. Environmental determinants of extinction selectivity in the fossil record. *Nature*, 454(7204):626–629, jul 2008.
- [122] Steven J Phillips, Robert P Anderson, and Robert E Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190:231–259, 2006.
- [123] Mathias M Pires, Daniele Silvestro, and Tiago B Quental. Continental faunal exchange and the asymmetrical radiation of carnivores. *Proceedings of the Royal Society B: Biological Sciences*, 282:20151952, 2015.

- [124] Laura J. Pollock, Michael J. Bayly, and Peter a. Vesk. The Roles of Ecological and Evolutionary Processes in Plant Community Assembly: The Environment, Hybridization, and Introgression Influence Co-occurrence of Eucalyptus. *The American Naturalist*, pages 000–000, mar 2015.
- [125] Laura J. Pollock, William K. Morris, and Peter A. Vesk. The role of functional traits in species distributions revealed through a hierarchical model. *Ecography*, 35(8):716–725, 2012.
- [126] Samantha A Price, Samantha S B Hopkins, Kathleen K Smith, and V Louise Roth. Tempo of trophic evolution and its impact on mammalian diversification. *Proceedings of the National Academy of Sciences*, 109(18):7008–12, may 2012.
- [127] A Purvis, J L Gittleman, G Cowlishaw, and G M Mace. Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, 267(1456):1947–52, oct 2000.
- [128] Tiago B Quental and Charles R Marshall. Extinction during evolutionary radiations: reconciling the fossil record with molecular phylogenies. *Evolution*, 63(12):3158–3167, dec 2009.
- [129] Tiago B Quental and Charles R Marshall. How the Red Queen Drives Terrestrial Mammals to Extinction. *Science*, 341(6143):290–292, sep 2013.
- [130] Daniel L Rabosky. Extinction rates should not be estimated from molecular phylogenies. *Evolution*, 64(6):1816–1824, jun 2010.
- [131] Daniel L. Rabosky. Reproductive isolation and the causes of speciation rate variation in nature. *Biological Journal of the Linnean Society*, 2015.
- [132] Daniel L Rabosky and Daniel R Matute. Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. *Proceedings of the National Academy of Sciences*, 110(38):15354–15359, 2013.
- [133] Daniel L Rabosky and Amy R McCune. Reinventing species selection with molecular phylogenies. *Trends in Ecology & Evolution*, 25(2):68–74, feb 2010.
- [134] Daniel L. Rabosky, Francesco Santini, Jonathan Eastman, Stephen a. Smith, Brian Sidlauskas, Jonathan Chang, and Michael E. Alfaro. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, 4:1–8, jun 2013.
- [135] P Raia, F Carotenuto, F Passaro, D Fulgione, and M Fortelius. Ecological specialization in fossil mammals explains Cope’s rule. *The American Naturalist*, 179(3):328–37, mar 2012.

- [136] P. Raia, F. Carotenuto, F. Passaro, P. Piras, D. Fulgione, L. Werdelin, J. Saarinen, and Mikael Fortelius. Rapid action in the Palaeogene, the relationship between phenotypic and taxonomic diversification in Coenozoic mammals. *Proceedings of the Royal Society B: Biological Sciences*, 280:20122244, nov 2013.
- [137] David M Raup. Taxonomic survivorship curves and Van Valen's Law. *Paleobiology*, 1(1):82–96, jan 1975.
- [138] David M Raup. Cohort Analysis of generic survivorship. *Paleobiology*, 4(1):1–15, 1978.
- [139] David M Raup. Mathematical models of cladogenesis. *Paleobiology*, 11(1):42–52, 1985.
- [140] David M Raup. *Extinction: Bad Genes or Bad Luck?* Norton, New York, 1991.
- [141] David M Raup. The role of extinction in evolution. *Proceedings of the National Academy of Sciences*, 91(July):6758–6763, 1994.
- [142] David M Raup and Stephen Jay Gould. Stochastic simulation and evolution of morphology – towards a nomothetic paleontology. *Systematic Zoology*, 23(3):305–322, 1974.
- [143] David M Raup, Stephen Jay Gould, Thomas J M Schopf, and Daniel S Simberloff. Stochastic models of phylogeny and the evolution of diversity. *The Journal of Geology*, 81(5):525–542, 1973.
- [144] David M Raup and J John Sepkoski. Mass Extinctions in the Marine Fossil Record. *Science*, 215(4539):1501–1503, 1982.
- [145] Ian W Renner and David I Warton. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics*, 69:274–281, 2013.
- [146] Peter D Roopnarine. The description and classification of evolutionary mode: a computational approach. *Paleobiology*, 27(3):446–465, 2001.
- [147] Peter D Roopnarine. Analysis of rates of morphologic evolution. *Annual Review of Ecology, Evolution, and Systematics*, 34:605–632, 2003.
- [148] Peter D Roopnarine, Gabe Byars, and Paul Fitzgerald. Anagenetic evolution, stratophentic patterns, and random walk models. *Paleobiology*, 25(1):41–57, 1999.
- [149] James Rosindell, Luke J Harmon, and Rampal S Etienne. Unifying ecology and macroevolution with individual-based theory. *Ecology Letters*, 18(5):472–482, may 2015.
- [150] Martin Rosvall, D Axelsson, and CT Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(14):13–24, 2009.

- [151] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–23, jan 2008.
- [152] Kaustuv Roy, Gene Hunt, David Jablonski, Andrew Z Krug, and James W Valentine. A macroevolutionary perspective on species range limits. *Proceedings of the Royal Society B: Biological Sciences*, 276:1485–1493, apr 2009.
- [153] J. Andrew Royle and Robert M Dorazio. *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Elsevier, London, 2008.
- [154] Donald B Rubin. Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473–489, 1996.
- [155] Holger Schielzeth. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2):103–113, feb 2010.
- [156] J John Sepkoski. Stratigraphic biases in the analysis of taxonomic survivorship. *Paleobiology*, 1(4):343–355, 1975.
- [157] J John Sepkoski. A factor analytic description of the Phanerozoic marine fossil record. *Paleobiology*, 7(1):36–53, 1981.
- [158] PM Sheehan. The late Ordovician mass extinction. *Annual Review of Earth and Planetary Sciences*, 29:331–364, 2001.
- [159] H David Sheets and Charles E Mitchell. Uncorrelated change produces the apparent dependence of evolutionary rate on interval. *Paleobiology*, 27(3):429–445, 2001.
- [160] Bill Shipley, Denis Vile, and Eric Garnier. From plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science*, 314:812–814, 2006.
- [161] Christian A Sidor, Daril A Vilhena, Kenneth D Angielczyk, Adam K Huttenlocker, Sterling J Nesbitt, Brandon R Peecook, J Sébastien Steyer, Roger M H Smith, and Linda A Tsuji. Provincialization of terrestrial faunas following the end-Permian mass extinction. *Proceedings of the National Academy of Sciences*, 110(20):8129–33, may 2013.
- [162] Daniele Silvestro, Alexandre Antonelli, Nicolas Salamin, and Tiago B Quental. The role of clade competition in the diversification of North American canids. *Proceedings of the National Academy of Sciences of the United States of America*, 112(28):8684–9, jul 2015.
- [163] Daniele Silvestro, Jan Schnitzler, Lee Hsiang Liow, Alexandre Antonelli, and Nicolas Salamin. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Systematic biology*, 63(3):349–67, may 2014.

- [164] Carl Simpson. *Levels of selection and large-scale morphological trends*. PhD thesis, University of Chicago, 2006.
- [165] Carl Simpson. The case for species selection. *bioRxiv*, 2016.
- [166] Carl Simpson and Paul G. Harnik. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology*, 35(4):631–647, dec 2009.
- [167] Carl Simpson, Wolfgang Kiessling, Heike Mewis, Rosemarie C Baron-Szabo, and Johannes Müller. Evolutionary diversification of reef corals: a comparison of the molecular and fossil records. *Evolution*, 65(11):3274–3284, nov 2011.
- [168] George Gaylord Simpson. *Tempo and Mode in Evolution*. Columbia University Press, New York, 1944.
- [169] George Gaylord Simpson. *The Major Features of Evolution*. Columbia University Press, New York, 1953.
- [170] GJ Slater, LJ Harmon, and ME Alfaro. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, 66(12):3931–3944, 2012.
- [171] Graham J. Slater. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution*, 4(8):734–744, aug 2013.
- [172] Graham J. Slater. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proceedings of the National Academy of Sciences*, 112(16):4897–4902, 2015.
- [173] F A Smith, J Brown, J Haskell, and S Lyons. Similarity of mammalian body size across the taxonomic hierarchy and across space and time. *The American Naturalist*, 163:672–691, 2004.
- [174] F. A. Smith, S. K. Lyons, S.K. Morgan Ernest, and J. H. Brown. Macroecology: more than the division of food and space among species on continents. *Progress in Physical Geography*, 32(2):115–138, apr 2008.
- [175] Felisa A Smith, James H Brown, John P Haskell, S Kathleen Lyons, John Alroy, Eric L Charnov, Tamar Dayan, Brian J Enquist, S K Morgan Ernest, Elizabeth A Hadly, Kate E Jones, Dawn M Kaufman, Pablo A Marquet, Brian A Maurer, Karl J Niklas, Warren P Porter, Bruce Tiffney, and Michael R Willig. Similarity of Mammalian Body Size across the Taxonomic Hierarchy and across Space and Time. *The American Naturalist*, 163(5):672–691, 2004.
- [176] Michael Smithson and Jay Verkuilen. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1):54–71, 2006.

- [177] AR Solow and Woollcott Smith. On fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, 23(3):271–277, 1997.
- [178] T Stadler. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences*, 108(15):6187–6192, 2011.
- [179] T Stadler. Recovering speciation and extinction dynamics based on phylogenies. *Journal of Evolutionary Biology*, 26:1203–1219, may 2013.
- [180] Tanja Stadler and Folmer Bokma. Estimating speciation and extinction rates for phylogenies of higher taxa. *Systematic biology*, 62(2):220–30, mar 2013.
- [181] S M Stanley. A theory of evolution above the species level. *Proceedings of the National Academy of Sciences*, 72(2):646–650, 1975.
- [182] S M Stanley. *Macroevolution: pattern and process*. W. H. Freeman, San Francisco, 1979.
- [183] Susumu Tomiya. Body Size and Extinction Risk in Terrestrial Mammals Above the Species Level. *The American Naturalist*, 182:196–214, sep 2013.
- [184] Leigh Van Valen. A new evolutionary law. *Evolutionary Theory*, 1:1–30, 1973.
- [185] Leigh Van Valen. Taxonomic survivorship curves. *Evolutionary Theory*, 4:129–142, 1979.
- [186] Blair Van Valkenburgh. Skeletal and dental predictors of body mass in carnivores. In John Damuth and Bruce J Macfadden, editors, *Body size in mammalian paleobiology: estimation and biological implications*, pages 181–205. Cambridge University Press, Cambridge, 1990.
- [187] Daril A Vilhena. *Boundaries and dynamics of biomes*. PhD thesis, University of Washington, 2013.
- [188] Daril A Vilhena, Elisha B Harris, Carl T Bergstrom, Max E Maliska, Peter D Ward, Christian A Sidor, Caroline A E Strömberg, and Gregory P Wilson. Bivalve network reveals latitudinal selectivity gradient at the end-Cretaceous mass extinction. *Scientific Reports*, 3:1790, may 2013.
- [189] Elisabeth S Vrba. What is species selection? *Systematic Zoology*, 33(3):318–328, 1984.
- [190] Elisabeth S Vrba and Stephen Jay Gould. The hierarchical expansion of sorting and selection: sorting and selection cannot be equated. *Paleobiology*, 12(2):217–228, 1986.
- [191] P J Wagner, M Aberhan, A Hendy, and W Kiessling. The effects of taxonomic standardization on sampling-standardized estimates of historical diversity. *Proceedings of the Royal Society B: Biological Sciences*, 274(1608):439, 2007.

- [192] Peter J Wagner and George F Estabrook. Trait-based diversification shifts reflect differential extinction among fossil taxa. *Proceedings of the National Academy of Sciences*, 111:16419–16424, oct 2014.
- [193] Peter J. Wagner and Jonathan D. Marcot. Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. *Methods in Ecology and Evolution*, 4(8):703–713, aug 2013.
- [194] Steve C. Wang. On the continuity of background and mass extinction. *Paleobiology*, 29(4):455–467, dec 2003.
- [195] Steve C Wang and C R Marshall. Improved confidence intervals for estimating the position of a mass extinction boundary. *Paleobiology*, 30(1):5–18, 2004.
- [196] David I. Warton, Bill Shipley, and Trevor Hastie. CATS regression - a model-based approach to studying trait-based community assembly. *Methods in Ecology and Evolution*, 6(4):389–398, apr 2015.
- [197] Sumio Watanabe. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, 11:3571–3594, 2010.
- [198] C Williams, C H C Brunton, and S J Carlson. *Treatise on invertebrate paleontology. Part H, Brachiopoda*. Geological Society of America, Boulder, Colorado, 2007.