

Species occurrence as a function of both emergent biological traits  
and environmental context

Peter D. Smits<sup>1,\*</sup>

1. University of Chicago, Chicago, Illinois 60637.

\* Corresponding author; e-mail: psmits@uchicago.edu.

*Manuscript elements:*

*Keywords:*

*Manuscript type:* Article

Prepared using the suggested L<sup>A</sup>T<sub>E</sub>X template for *Am. Nat.*

## Introduction

- 2 How do species pools change over time as species are recruited or go extinct? When are ecotypes enriched or depleted? How does global and regional environmental context affect the distribution of
- 4 species ecotypes (e.g. guilds) in a regional species pool?

A regional species pool is the set of species which form communities in a specific region; local  
6 communities are subsets of the regional pool. The composition of a regional species pool changes  
over time due to speciation, migration, extinction. Local scale processes like resource competition  
8 only affect the regional species pool if all communities are affected.

Valentine and Bambach how they presented guilds in paleobiology. Bush and Bambach presented an  
10 ecocube to describe what how marine invertebrates partition space and resources (Bambach et al.,  
2007; Bush and Bambach, 2011; Bush et al., 2007). Unique combinations represent what possible  
12 ecotypes are observable. The distribution of ecocube occupancy is then normally analyzed as raw  
counts of unique combinations or using ordination methods and the change in disparity over time is  
14 estimated (Bambach et al., 2007; Bush and Bambach, 2011; Bush et al., 2007).

One of the greatest challenges with analyzing species occurrence data is the inherent incompleteness  
16 of any sample (Foote, 2001; Foote and Sepkoski, 1999; Lloyd et al., 2011; Royle and Dorazio, 2008;  
Royle et al., 2014; Wang and Marshall, 2016). In the modern, only presences are certain as an  
18 absence can be caused by both the species being truly absent or the species never having been  
sampled (Royle and Dorazio, 2008; Royle et al., 2014). For paleontological data in the context of  
20 this study, the incomplete preservation of fossil communities combined with the incomplete  
sampling of what fossils there are means that the true times of origination or extinction may not be  
22 observed (Foote, 2001; Foote and Sepkoski, 1999; Wang et al., 2016; Wang and Marshall, 2016)

Smits (2015) found several systematic differences in mammal species durations associated with  
24 various species traits. Omnivorous taxa were found to have, on average, a greater duration than  
other dietary categories. Additionally, arboreal taxa were found to have a shorter duration than  
26 other locomotor categories.

An unresolved question from Smits (2015) is whether the greater extinction risk faced by arboreal is  
28 constant over time or if there was a change in extinction risk at the Paleogene/Neogene boundary.  
Specifically, the question is whether the extinction risk arboreal taxa increased in the Neogene,  
30 driving the loss of arboreal taxa and average extinction risk of arboreal taxa down.

There are no observed massive cross-taxonomic turnover events in the North American record,  
32 unlike the Neogene record Europe (Alroy, 1996, 2009; Alroy et al., 2000; Eronen et al., 2015; Janis,  
1993).

34 The effect of climate on diversity and the diversification process has been the focus of considerable  
research with many analyses favoring diversification being more biologically-mediated than  
36 climate-mediated (Alroy, 1996; Alroy et al., 2000; Clyde and Gingerich, 1998; Figueirido et al.,  
2012). Scale of analysis makes a big difference in interpretation of results, both temporal and  
38 geographic. For example when the mammal fossil record analyzed at small temporal and geographic  
scales a correlation between diversity and climate are observable (Clyde and Gingerich, 1998).  
40 However, when the record is analyzed at the scale of the continent and the Cenozoic there is no  
correlation with diversity and climate (Alroy et al., 2000). This result, however, does not go  
42 against the idea that there may be short periods of correlation and that this correlation change or  
reverse direction over time; instead this result means that there is no single direction of correlation  
44 between diversity and climate (Figueirido et al., 2012).

In the case of a fluctuating correlation between diversity and climate it is hard to make the  
46 argument of an actual causal link between the two without understanding the ecological differences  
in mammalian fauna over time; when this analysis is based on diversity or taxonomy alone no  
48 mechanisms are possible to infer. After all, taxonomy conflates many potential factors that could  
affect diversification into a single variable; by separating the effects of shared common ancestry (i.e.  
50 phylogeny) from species ecology the subtle differences in the diversification process can be observed  
(Smits, 2015).

52 There are many candidate climatic events that may have influenced the distribution of mammal  
ecotypes regionally, if not globally (Blois and Hadly, 2009; Janis, 1993; Zachos et al., 2008, 2001).

54 The Paleocene-Eocene Temermal Maximum is associated with species dwarfing and rearrangement  
of local communities, though regional effects are less known CITATION. The Mid-Miocene  
56 climactic optimum is associated with WHAT CITATION. The

The general cooling throughout the Cenozoic and the development of ice-caps in the Neogene. The  
58 Oligo-Miocene boundary.

One of the most stunning environmental transitions of the Cenozoic in North America was gradual  
60 “opening-up” of the landscape with the shift from closed or partially forested environments of the  
Paleogene to the savannah and grasslands environments that characterize the Neogene (Blois and  
62 Hadly, 2009; Janis, 1993; Janis et al., 2000; Strömberg, 2005).

Fourth-corner modeling an approach to explaining the patterns of either species abundance or  
64 presence/absence as a product of species traits, environmental factors, and the interaction between  
traits and environment CITATION. In modern ecological studies, what is being modeled is species  
66 occurrences at localities distributed across a region CITATION. In this study, what is being  
modeled is the pattern of species occurrence over time for most of the Cenozoic in North America  
68 (Fig. 1). These two approaches, modern and palentological, are different views of the same  
three-dimensional pattern: species at localities over time. The temporal limitations of modern  
70 ecological studies and difficulties with uneven spatial occurrences of fossils in paleontological studies  
means that these approaches are complimentary but reveal different patterns of how species are  
72 distributed in time and space.

Ultimately, the goal of this analysis are to understand when are unique ecotypes enriched or  
74 depleted in the North American mammal regional species pool and how changes in ecotypic  
diversity are related to changes in species’ environmental context.

## 76 Materials and Methods

### Taxon occurrences and species-level information

78 All fossil occurrence information was downloaded from the Paleobiology Database. Occurrences (PBDB) were restricted to all Mammalia sampled in North America between the  
80 Maastrichtian and Gelasian stages. Taxonomic, stratigraphic, and ecological metadata for each occurrence was included. The raw data is available for download at <http://goo.gl/2s1geU>.

82 This raw data was then sorted, cleaned, and manipulated programmatically prior to analysis. Species taxonomic assignments given by the PBDB were updated for accuracy and consistency. For  
84 example, species classified in the order Artiodactyla were reclassified as Cetartiodactyla. These re-assignments follow Smits (2015) and were Janis et al. (2008, 1998) and the Encyclopedia of Life  
86 WEBSITE. Additionally, Taxa who's life habit was classified as either volant (i.e. Chiroptera) or aquatic (e.g. Cetacea) were excluded from this analysis because of both differences in fossilization  
88 potential and applicability to the study of terrestrial species pools.

The life habit and dietary categories provided through the PBDB were coarsened to increase per  
90 ecotype sample size; this coarsening follows the same procedure as Smits (2015). Additionally, life habit category was further modified to break-up the vague “ground-dwelling” category;  
92 re-classifying these species by ankle posture gives more precise information about that species’ environmental context. Ground-dwelling taxa were reassigned following ? by species taxonomic  
94 context. Species ecotype is defined as the interaction between life habit and diet categories. Ecotype categories with less than 10 species having ever been in that combination were excluded, yielding a  
96 total of 18 of 24 possible ecotypes.

Table 2: Posture assignment based on taxonomy

Order	Family	Stance
	Ailuridae	plantigrade
	Allomyidae	plantigrade

Continued on next page

**Table 2 – continued from previous page**

Order	Family	Stance
	Amphicyonidae	plantigrade
	Amphilemuridae	plantigrade
	Anthracotheriidae	digitigrade
	Antilocapridae	unguligrade
	Apheliscidae	plantigrade
	Aplodontidae	plantigrade
	Apternodontidae	scansorial
	Arctocyonidae	unguligrade
	Barbourofelidae	digitigrade
	Barylambdidae	plantigrade
	Bovidae	unguligrade
	Camelidae	unguligrade
	Canidae	digitigrade
	Cervidae	unguligrade
	Cimolodontidae	scansorial
	Coryphodontidae	plantigrade
	Cricetidae	plantigrade
	Cylindrodontidae	plantigrade
	Cyriacotheriidae	plantigrade
	Dichobunidae	unguligrade
Dinocerata		unguligrade
	Dipodidae	digitigrade
	Elephantidae	digitigrade
	Entelodontidae	unguligrade
	Eomyidae	plantigrade

Continued on next page

**Table 2 – continued from previous page**

Order	Family	Stance
	Erethizontidae	plantigrade
	Erinaceidae	plantigrade
	Esthonychidae	plantigrade
	Eutypomyidae	plantigrade
	Felidae	digitigrade
	Florentiamyidae	plantigrade
	Gelocidae	unguligrade
	Geolabididae	plantigrade
	Glyptodontidae	plantigrade
	Gomphotheriidae	unguligrade
	Hapalodectidae	plantigrade
	Heteromyidae	digitigrade
	Hyaenidae	digitigrade
	Hyaenodontidae	digitigrade
	Hypertragulidae	unguligrade
	Ischyromyidae	plantigrade
	Jimomyidae	plantigrade
Lagomorpha		digitigrade
	Leptictidae	plantigrade
	Leptochoeridae	unguligrade
	Leptomerycidae	unguligrade
	Mammutidae	unguligrade
	Megalonychidae	plantigrade
	Megatheriidae	plantigrade
	Mephitidae	plantigrade

Continued on next page

**Table 2 – continued from previous page**

Order	Family	Stance
	Merycoidodontidae	digitigrade
Mesonychia		unguligrade
	Mesonychidae	digitigrade
	Micropternodontidae	plantigrade
	Mixodectidae	plantigrade
	Moschidae	unguligrade
	Muridae	plantigrade
	Mustelidae	plantigrade
	Mylagaulidae	fossorial
	Mylodontidae	plantigrade
	Nimravidae	digitigrade
	Nothrotheriidae	plantigrade
Notoungulata		unguligrade
	Oromerycidae	unguligrade
	Oxyaenidae	digitigrade
	Palaeomerycidae	unguligrade
	Palaeoryctidae	plantigrade
	Pampatheriidae	plantigrade
	Pantolambdidae	plantigrade
	Periptychidae	digitigrade
Perissodactyla		unguligrade
	Phenacodontidae	unguligrade
Primates		plantigrade
	Procyonidae	plantigrade
	Proscalopidae	plantigrade

Continued on next page

**Table 2 – continued from previous page**

Order	Family	Stance
	Protoceratidae	unguligrade
	Reithroparamyidae	plantigrade
	Sciuravidae	plantigrade
	Sciuridae	plantigrade
	Simimyidae	plantigrade
	Soricidae	plantigrade
	Suidae	digitigrade
	Talpidae	fossorial
	Tayassuidae	unguligrade
	Tenrecidae	plantigrade
	Titanoideidae	plantigrade
	Ursidae	plantigrade
	Viverravidae	plantigrade
	Zapodidae	plantigrade

<sup>98</sup> Species mass information was gathered from multiple different sources where a plurality of the body size estimates are from the PBDB. Body part measurements for many species are also available through the PBDB. Just as with Smits (2015), these measurements and corresponding regression equations were used to get mass estimates for more species. Additional mass estimates and body part measurements were sourced from numerous publications and the Neogene Old World Database; see the supplementary material to Smits (2015) for details. Mass was log-transformed and then mean-centered and rescaled by dividing by two-times its standard deviation; this insures that the magnitude of effects for both continuous and discrete covariates are comparable (Gelman, 2008; Gelman and Hill, 2007).

Table 1: 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.

All fossil occurrences from 64 to 2 million years ago (Mya) were binned into 31 2 million year (My)

<sup>108</sup> bins. This temporal length was chosen because it is approximately the resolution of the North American mammal fossil record.

## <sup>110</sup> Environmental and temporal covariates

The group-level covariates in this study are descriptors of species' environmental context, specifically global temperature estimates and Graham's floral intervals CITATION. Global temperature across most of the Cenozoic was calculated from Mg/Ca isotope record from deep sea carbonates (Cramer et al., 2011). Mg/Ca based temperature estimates are preferable to the frequently used  $\delta^{18}\text{O}$  temperature proxy (Alroy et al., 2000; Figueirido et al., 2012; Zachos et al., 2008, 2001) because Mg/Ca estimates do not conflate temperature with ice sheet volume and depth/stratification changes; this makes it preferable as an estimate of global temperature for macroevolutionary and macroecological studies (Ezard et al., 2016).

Two aspects of the Mg/Ca-based temperature curve were included in this analysis: mean and range.

<sup>120</sup> Both were calculated as the mean of all respective estimates for each 2 My temporal bins. Both mean and range were then rescaled as above: subtract mean, divide by twice the standard deviation.

<sup>122</sup> The other major set of environmental factors included in this study are Graham's Cenozoic plant phases CITATION. Graham's plant phases are holistic descriptors of the taxonomic composition of <sup>124</sup> which plants were present at a given time and their relative modernity, with younger phases

Table 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	Legendre (1986)
General	$\log(m) = 2.9677x - 5.6712$	mandible length	?
General	$\log(m) = 3.68x - 3.83$	skull length	?
Carnivores	$\log(m) = 2.97x + 1.681$	lower m1 length	?
Insectivores	$\log(m) = 1.628x + 1.726$	lower m1 area	?
Insectivores	$\log(m) = 1.714x + 0.886$	upper M1 area	?
Lagomorph	$\log(m) = 2.671x - 2.671$	lower toothrow area	Tomiya (2013)
Lagomorph	$\log(m) = 4.468x - 3.002$	lower m1 length	Tomiya (2013)
Marsupials	$\log(m) = 3.284x + 1.83$	upper M1 length	?
Marsupials	$\log(m) = 1.733x + 1.571$	upper M1 area	?
Rodentia	$\log(m) = 1.767x + 2.172$	lower m1 area	Legendre (1986)
Ungulates	$\log(m) = 1.516x + 3.757$	lower m1 area	?
Ungulates	$\log(m) = 3.076x + 2.366$	lower m2 length	?
Ungulates	$\log(m) = 1.518x + 2.792$	lower m2 area	?
Ungulates	$\log(m) = 3.113x - 1.374$	lower toothrow length	?

representing increasingly modern taxa CITATION. Graham CITATION defines four intervals from

126 the Cretaceous to the Pliocene, though only three of these intervals are included in this analysis.

Graham's plant phases CITATION was included as a series of "dummy variables" encoding the

128 three phases included in this analysis. This means that the first phase is synonymous with the

intercept and phases

### 130 Modelling species occurrence

Two different models were used in this study: a pure-presence model and a birth-death model. Both

132 models at their core are hidden Markov model where the latent aspect of the process has an

absorbing state (Allen, 2011). The difference between these two models is if the probability of a

134 species origination and survival are considered equal or different (Table 4). Something that is

important to realize is that while there are only two state "codes" in a presence-absence matrix (i.e.

136 0/1), there are in fact three states in a birth-death model: never having originated, extant, and

extinct. The last of these is the absorbing state, as once a species has gone extinct it cannot

138 re-originate (Allen, 2011); this is made obvious in the transition matrices as the probability of an

		State at $t + 1$		
		$0_{never}$	1	$0_{extinct}$
State at $t$	$0_{never}$	$1 - \theta$	$\theta$	0
	1	0	$\theta$	$1 - \theta$
	$0_{extinct}$	0	0	1

(a) Pure-presence

		State at $t + 1$		
		$0_{never}$	1	$0_{extinct}$
State at $t$	$0_{never}$	$1 - \phi$	$\phi$	0
	1	0	$\pi$	$1 - \pi$
	$0_{extinct}$	0	0	1

(b) Birth-death

Table 4: Transition matrices for the pure-presence (4a) and birth-death (4b) models. Both of these models share the core machinery of discrete-time birth-death processes but make distinct assumptions about the equality of originating and surviving (Eq. 2, and 3). Note also that while there are only two state “codes” (0, 1), there are in fact three states: never having originated  $0_{never}$ , present 1, extinct  $0_{extinct}$  (Allen, 2011).

extinct species changing states is 0 (Table 4). See below for parameter explanations (Tables 6, and

140 7).

## Data augmentation

- 142 All presence/absence observations are incomplete. The hidden Markov model at the core of this  
 analysis allows for observed absences to be used meaningfully to estimate the number of unobserved  
 144 species. Of specific concern in this analysis is the unknown “true” size of the dataset; how many  
 species could have actually been observed? While many species have been observed, the natural  
 146 incompleteness of all observations, especially in the case of paleontological data, there are obviously  
 many species which were never sampled (Royle and Dorazio, 2008; Royle et al., 2007).
- 148 Let  $N$  by the total number of observed species,  $M$  be the upper limit of possible species that could  
 have existed given a model of species presence, and  $N^*$  is the all-zero histories where  $N^* = M - N$ .  
 150 This approach assumes that  $\hat{N} \sim \text{Binomial}(M, \psi)$  where  $\hat{N}$  is the estimated “true” number of  
 species and  $\psi$  is the probability that any augmented species should actually be “present.” Because  
 152  $M$  is user defined, this approach effectively gives  $\psi$  a uniform prior over  $N$  to  $M$  (Royle and  
 Dorazio, 2008). For this study,  $M = N + \lfloor N/4 \rfloor$ .
- 154 Data imputation is the process of estimating missing data for partially observed covariates (Gelman  
 and Hill, 2007; Rubin, 1996), this is simple in a Bayesian context because data are also parameters  
 156 (Gelman et al., 2013). Augmented species also have no known mass so a mass estimate must be

imputed for each possible species (Royle and Dorazio, 2012). This procedure assumes that mass  
 158 values for augmented species are from the same distribution as observed species. The distribution of  
 observed mass values is estimated as part of the model, and new mass values are then generated  
 160 from this distribution. This approach is an example of imputing data missing completely at random  
 (Gelman and Hill, 2007; Royle and Dorazio, 2012). Because log mass values are rescaled as a part of  
 162 this study, the body mass distribution is already known ( $\mathcal{N}(0, 0.5)$ ); augmented species body mass  
 just simply drawn from this distribution.

164 In addition to body mass information, the augmented species need an ecotype classification. Because  
 these species are completely unknown, they were all classified as “augmented,” an additional  
 166 grouping indicating their unknown biology. This classification has no biological interpretation.

### Observation process

168 The type of hidden Markov model used in this study has three characteristic probabilities:  
 probability  $p$  of observing a species given that it is present, probability  $\phi$  of a species surviving from  
 170 one time to another, and probability  $\pi$  of a species first appearing (Royle and Dorazio, 2008). In  
 this formulation, the probability of a species going extinct is  $1 - \pi$ . For the pure-presence model  
 172  $\phi = \pi$ , while for the birth-death model  $\phi \neq \pi$ .

The probability of observing a species that is present  $p$  is modeled as a logistic regression was a  
 174 time-varying intercept and species mass as a covariate. The effect of species mass on  $p$  was assumed  
 linear and constant over time and given a prior reflecting a possible positive relationship; these  
 176 assumptions are reflected in the structure of the model Equation 1. The parameters associated with  
 this part of the model are described in Table 5.

$$\begin{aligned} y_{i,t} &\sim \text{Bernoulli}(p_{i,t} z_{i,t}) \\ p_{i,t} &= \text{logit}^{-1}(\alpha_0 + \alpha_1 m_i + r_t) \\ r_t &\sim \mathcal{N}(0, \sigma) \end{aligned} \tag{1}$$

Table 5: Observation parameters

Parameter	dimensions	explanation
$y$	$N \times T$	observed species presence/absence
$z$	$N \times T$	“true” species presence/absence
$p$	$T$	probability of observing a species that is present at time $t$
$m$	$N$	species log mass, rescaled
$\alpha_0$	1	average log-odds of $p$
$\alpha_1$	1	change in average log-odds of $p$ per change mass
$r$	$T$	difference from $\alpha_0$ associated with time $t$
$\sigma$	1	standard deviation of $r$

Table 6: Parameters for the model of presence in the pure-presence model

Parameter	dimensions	explanation
$z$	$N \times T$	“true” species presence/absence
$\theta$	$N \times T - 1$	probability of $z = 1$
$a$	$T - 1 \times D$	ecotype-varying intercept; mean value of log-odds of $\theta$
$m$	$N$	species log mass, rescaled
$b_1$	1	effect of species mass on log-odds of $\theta$
$b_2$	1	effect of species mass, squared, on log-odds of $\theta$
$U$	$T \times D$	matrix of group-level covariates
$\gamma$	$U \times D$	matrix of group-level regression coefficients
$\Sigma$	$D \times D$	covariance matrix of $a$
$\Omega$	$D \times D$	correlation matrix of $a$
$\tau$	$D$	vector of standard deviations for each ecotype $a_d$

## 178 Pure-presence process

For the pure-presence model there is only a single probability dealing with the presence of a species  
 180  $\theta$  (Table 4a). This probability was modeled as multi-level logistic regression with both species-level  
 and group-level covariates (Gelman et al., 2013; Gelman and Hill, 2007). The parameters associated  
 182 with pure-presence model are presented in Table 6 and the full sampling statement in Equation 2.

The species-level of the model (Eq. 2) is a logistic regression with varying-intercept that varies by  
 184 ecotype. Additionally, species mass was included as a covariate associated with two regression  
 coefficients allowing a quadratic relationship with log-odds of occurrence. This assumption is based  
 186 on the known distribution of mammal body masses where species with intermediate mass values are  
 more common than either small or large bodied species. These assumptions are also reflected in the  
 188 choice of priors for these regression coefficients.

The values of each ecotype's intercept are themselves modeled as regressions using the group-level  
 190 covariates associated with environmental context. Each of these regressions has an associated  
 variance of possible values of each ecotype's intercept (Gelman and Hill, 2007). In addition, the  
 192 covariances between ecotype intercepts, given this group-level regression, are modeled (Gelman and  
 Hill, 2007).

194 All parameters not modeled elsewhere were given weakly informative priors (Gelman et al., 2013)  
 CITATION STAN MANUAL STATISTICAL RETHINKING. Weakly informative means that  
 196 priors do not necessarily encode actual prior information but instead help regularize or weakly  
 constrain posterior estimates. These priors have a concentrated probability density around and near  
 198 zero; this has the effect of tempering our estimates and help prevent overfitting the model to the  
 data (Gelman et al., 2013) CITATION STAN MANUAL STATISTICAL RETHINKING.

$$\begin{aligned}
 y_{i,t} &\sim \text{Bernoulli}(p_{i,t} z_{i,t}) & \alpha_0 &\sim \mathcal{N}(0, 1) \\
 p_{i,t} &= \text{logit}^{-1}(\alpha_0 + \alpha_1 m_i + r_t) & \alpha_1 &\sim \mathcal{N}(1, 1) \\
 r_t &\sim \mathcal{N}(0, \sigma) & \sigma &\sim \mathcal{N}^+(1) \\
 z_{i,1} &\sim \text{Bernoulli}(\rho) & b_1 &\sim \mathcal{N}(0, 1) \\
 z_{i,t} &\sim \text{Bernoulli}(\theta_{i,t}) & b_2 &\sim \mathcal{N}(-1, 1) \\
 \theta_{i,t} &= \text{logit}^{-1}(a_{t,j[i]} + b_1 m_i + b_2 m_i^2) & \gamma &\sim \mathcal{N}(0, 1) \\
 a &\sim \text{MVN}(u\gamma, \Sigma) & \tau &\sim \mathcal{N}^+(1) \\
 \Sigma &= \text{diag}(\tau)\Omega\text{diag}(\tau) & \Omega &\sim \text{LKJ}(2)
 \end{aligned} \tag{2}$$

200 **Birth-death process**

In the birth-death model,  $\phi \neq \pi$  and so each of these probabilities are modeled separately but in a  
 202 similar manner to how  $\theta$  is modeled in the pure-presence model (Eq. 2, Table 4b). The parameters  
 associated with the birth-death presence model are presented in Table 7 and the full sampling  
 204 statement, including observation (Eq. 1), is described in Equation 3.

Table 7: Parameters for the model of presence in the pure-presence model

Parameter	dimensions	explanation
$z$	$N \times T$	“true” species presence/absence
$\phi$	$N \times T$	probability of $z_{-,t} = 1   z_{-,t-1} = 0$
$\pi$	$N \times T - 1$	probability of $z_{-,t} = 1   z_{-,t-1} = 1$
$a^\phi$	$T - 1 \times D$	ecotype-varying intercept; mean value of log-odds of $\theta$
$a^\pi$	$T - 1 \times D$	ecotype-varying intercept; mean value of log-odds of $\theta$
$m$	$N$	species log mass, rescaled
$b_1^\phi$	1	effect of species mass on log-odds of $\phi$
$b_1^\pi$	1	effect of species mass on log-odds of $\pi$
$b_2^\phi$	1	effect of species mass, squared, on log-odds of $\phi$
$b_2^\pi$	1	effect of species mass, squared, on log-odds of $\pi$
$U$	$T \times D$	matrix of group-level covariates
$\gamma^\phi$	$U \times D$	matrix of group-level regression coefficients
$\gamma^\pi$	$U \times D$	matrix of group-level regression coefficients
$\Sigma^\phi$	$D \times D$	covariance matrix of $a^\phi$
$\Sigma^\pi$	$D \times D$	covariance matrix of $a^\pi$
$\Omega^\phi$	$D \times D$	correlation matrix of $a^\phi$
$\Omega^\pi$	$D \times D$	correlation matrix of $a^\pi$
$\tau^\phi$	$D$	vector of standard deviations for each ecotype $a_d^\phi$
$\tau^\pi$	$D$	vector of standard deviations for each ecotype $a_d^\pi$

Similar to the pure-presence model, both  $\phi$  and  $\pi$  are modeled as logistic regressions with

206 varying-intercept and one covariate associated with two parameters. The possible relationships  
 between mass and both  $\phi$  and  $\pi$  are reflected in the parameterization of the model and choice of  
 208 priors (Eq. 3).

The intercepts of  $\phi$  and  $\pi$  both vary by species ecotype and those values are themselves the product  
 210 of group-level regression using environmental factors as covariates (Eq. 3); this is identical to the

pure presence model (Eq. 2).

$$\begin{aligned}
y_{i,t} &\sim \text{Bernoulli}(p_{i,t} z_{i,t}) & \Sigma^\phi &= \text{diag}(\tau^\phi) \Omega^\phi \text{diag}(\tau^\phi) \\
p_{i,t} &= \text{logit}^{-1}(\alpha_0 + \alpha_1 m_i + r_t) & \Sigma^\pi &= \text{diag}(\tau^\pi) \Omega^\pi \text{diag}(\tau^\pi) \\
r_t &\sim \mathcal{N}(0, \sigma) & \rho &\sim U(0, 1) \\
\alpha_0 &\sim \mathcal{N}(0, 1) & b_1^\phi &\sim \mathcal{N}(0, 1) \\
\alpha_1 &\sim \mathcal{N}(1, 1) & b_1^\pi &\sim \mathcal{N}(0, 1) \\
\sigma &\sim \mathcal{N}^+(1) & b_2^\phi &\sim \mathcal{N}(-1, 1) \\
z_{i,1} &\sim \text{Bernoulli}(\phi_{i,1}) & b_2^\pi &\sim \mathcal{N}(-1, 1) \\
z_{i,t} &\sim \text{Bernoulli} \left( z_{i,t-1} \pi_{i,t} + \sum_{x=1}^t (1 - z_{i,x}) \phi_{i,t} \right) & \gamma^\phi &\sim \mathcal{N}(0, 1) \\
&& \gamma^\pi &\sim \mathcal{N}(0, 1) \\
\phi_{i,t} &= \text{logit}^{-1}(a_{t,j[i]}^\phi + b_1^\phi m_i + b_2^\phi m_i^2) & \tau^\phi &\sim \mathcal{N}^+(1) \\
\pi_{i,t} &= \text{logit}^{-1}(a_{t,j[i]}^\pi + b_1^\pi m_i + b_2^\pi m_i^2) & \tau^\pi &\sim \mathcal{N}^+(1) \\
a^\phi &\sim \text{MVN}(U \gamma^\phi, \Sigma^\phi) & \Omega^\phi &\sim \text{LKJ}(2) \\
a^\pi &\sim \text{MVN}(U \gamma^\pi, \Sigma^\pi) & \Omega^\pi &\sim \text{LKJ}(2)
\end{aligned} \tag{3}$$

## 212 Posterior inference and model adequacy

Programs that implement joint posterior inference for the above models (Eqs. 2, 3) were  
214 implemented in the probabilistic programming language Stan CITATION. The models used here  
both feature latent discrete parameters in the large matrix  $z$  (Tables 5, 6, 7; Eqs. 1, 2, 3). All  
216 methods for posterior inference implemented in Stan are derivative based which causes  
complications for actually implementing the above models because integers do not have derivatives.  
218 Instead of implementing a latent discrete parameterization, the posterior probabilities of all possible  
states of the latent parameters  $z$  were estimated (i.e. marginalized).  
220 Species durations at minimum range-through from the FAD to the LAD, but the incompleteness of

all observations means that the actual time of origination or extinction is unknown. The  
222 marginalization approach used here means that the probabilities all possible histories for a species  
are calculated, from the end members of the species having existed for the entire study interval and  
224 the species having only existed between the directly observed FAD and LAD to all possible  
intermediaries CITATION STAN MANUAL.

226 The combined size of the dataset and large number of parameters in both models (Eqs. 2, 3),  
specifically the total number of latent parameters that are the matrix  $z$ , means that stochastic  
228 approximate posterior inference is computationally very slow even using HMC. Instead, an  
approximate Bayesian approach was used: variational inference. A recently developed automatic  
230 variational inference algorithm called “automatic differentiation variational inference” (ADVI) is  
implemented in Stan and was used here CITATION. ADVI assumes that the posterior is Gaussian  
232 but still yields a true Bayesian posterior; this assumption is similar to quadratic approximation of  
the likelihood function used in maximum likelihood inference CITATION. The principal limitation  
234 of assuming the joint posterior is Gaussian is that the true topology of the log-posterior isn’t  
estimated; this is a particular burden for scale parameters which are bound to be positive (e.g.  
236 standard deviation).

After fitting both models (Eqs. 2, 3) using ADVI, model adequacy and quality of fit was assessed  
238 using a series of posterior predictive checks CITATION CITATION. Because all Bayesian models  
are inherently generative, simulations of new data sets is “free” CITATION. By simulating many  
240 theoretical data sets using the observed covariate information the congruence between predictions  
made by the model and the observed empirical data can be assessed. By combining multiple  
242 posterior predictive tests of congruence between empirical and simulated values of interest, the  
holistic adequacy of the model can be analyzed CITATION.

244 An example posterior predictive check used in this study was comparing the observed average  
number of observations per species to a distribution of simulated averages; if the empirically  
246 observed value sits in the middle of the distribution than the model is adequate in reproducing the  
observed number of occurrences per species.

- 248 Posterior simulations for time series are start with the values at  $t = 1$  and then just simulating forward.
- 250 Given parameter estimates, diversity and diversification rates are estimated through posterior predictive simulations. Given the observed presence-absence matrix  $y$ , estimates of the true
- 252 presence-absence matrix  $z$  can be simulated and the distribution of possible occurrence histories can be analyzed. This is conceptually similar to marginalization where the probability of each
- 254 possible occurrence history is estimated (Fig. 2).

The posterior distribution of  $z$  gives the estimate of standing diversity  $N_t^{stand}$  for all time points as

$$N_t^{stand} = \sum_{i=1}^M z_{i,t}. \quad (4)$$

- 256 Given estimates of  $N^{stand}$  for all time points, the estimated number of originations  $O_t$  are be estimated as

$$O_t = \sum_{i=1}^M z_{i,t} = 1 | z_{i,t-1} = 0 \quad (5)$$

- 258 and number of extinctions  $E_t$  estimated as

$$E_t = \sum_{i=1}^M z_{i,t} = 0 | z_{i,t-1} = 1. \quad (6)$$

Per-captia growth  $D^{rate}$ , origination  $O^{rate}$  and extinction  $E^{rate}$  rates are then calculated as

$$\begin{aligned} O_t^{rate} &= \frac{O_t}{N_{t-1}^{stand}} \\ E_t^{rate} &= \frac{E_t}{N_{t-1}^{stand}} \\ D_t^{rate} &= O_t^{rate} - E_t^{rate}. \end{aligned} \quad (7)$$

<sup>260</sup> **Results**

Posterior results take one of two forms: direct inspection of parameter estimates, and downstream  
<sup>262</sup> estimates of diversity and diversification rates. For the former, both the pure-presence and birth-death models (Eq. 2, and 3 are inspected. For the latter, only posterior estimates from the  
<sup>264</sup> birth-death model are considered; the reason for this is explained below in the comparison of the models' posterior predictive check results.

<sup>266</sup> **Comparing parameter estimates from the pure-presence and birth-death models**

<sup>268</sup> Comparison of the posterior predictive performance of the pure-presence and birth-death models reveals a striking difference in quality of the models' fits to the data (Fig. 3a and 3b). The  
<sup>270</sup> birth-death model is clearly able to reproduce the observed average number of occurrence, in contrast to the pure-birth model which greatly underestimates the ovserved average number of  
<sup>272</sup> occurrences. The interpretation of these results is that the results of the birth-death model are more representative of the data than the pure-presence model, though further inspection of the  
<sup>274</sup> posterior parameter estimates can provide further insight into why these models give different posterior predictive results (Gelman et al., 2013). However, it is expected that downstream analyses  
<sup>276</sup> from the birth-death model will be more reliable than that from the pure-presence model.

Occurrence probabilities estimated from the pure-presence model (Fig. 4) are much more similar to  
<sup>278</sup> the origination estimates from the birth-death model (Fig. 5) than the estimates of survival probability (Fig. 6).

<sup>280</sup> In general, both occurrence probabilities estimated from the pure-presence model (Fig. 4) and origination probabilities estimated from the birth-death model (fig. 5) increase with time. Notable,  
<sup>282</sup> ecotypes with arboreal components do not follow this average; instead, occurrence and origination probabilities appear relatively flat for most of the Cenozoic.

<sup>284</sup> The dramatic differences between origination and survival probabilities indicate how different these

processes are, and may be responsible for the better posterior predictive performance of the  
286 birth-death model over the pure-presence model (Fig. 3a, and 3b). While the estimates of both time  
series have high variance, what is striking is how mean origination probability changes over time  
288 while in general survival probabilities have relatively stable means (Fig. 5, and 6).

Estimates of origination probabilities appear to have less uncertainty than for survival (Fig. 5, and  
290 6).

The pure-presence and birth-death models differ in estimated effect of mass on the probability of  
292 observing a species that is present (Fig. 7). For the pure-presence model, mass is estimated to have  
no effect on the probability of observing a species that is presence (Fig. 7a). Contrastingly, for the  
294 birth-death model mass is found to have a negative relationship with observation such that larger  
species are less likely to be observed if present than smaller species (Fig. 7b).

296 The result from the birth-death model is unexpected given that it is generally assumed that larger  
mammals are more likely to have been collected than smaller mammals CITATION. However,  
298 collection is not preservation; similarities in preservation rate indicate similarities in how gap-filled  
species records are. What this result means is that the record of large bodied species is expected on  
300 average to be more gap-filled and less consistent from time point to time point than smaller bodied  
species. Additionally, this is presence/absence data, so higher preservation and collection in terms  
302 of individual specimens at a location or a single temporal horizon does not necessarily translate to  
high preservation over time.

304 The effect of species mass on probability of occurrence as estimated from the pure-presence (Fig. 8)  
is most similar to the effect of species mass on probability of origination as estimates from the  
306 birth-death model (Fig. 9). The striking pattern observable in both sets of estimates is the higher  
probability of occurrence for species with body sizes closer to the mean than either extremes. This  
308 result is consistent with the canonically normal distribution of mammal body sizes CITATION; it is  
then expected that the most likely to occur species would be those from the middle of the  
310 distribution, and that species originating will on average be of average mass, especially considering  
species shared common ancestry CITATION.

312 In contrast, the effect of species mass on probability of survival as estimated from the birth-death  
model (Fig. 10) indicates little effect of mass on extinction; this is consistent with previous findings  
314 from the North American mammal fossil record (Smits, 2015; Tomiya, 2013). Note that all variation  
between ecotypes is due to differences in ecotype-specific survival probability and the associated  
316 effects of plant phase.

Similarities in parameters estimates between ecotypes may be due to similar response to  
318 environmental factors. Some of the obvious patterns from inspection of the individual-level  
estimates of occurrence (Fig. 4), origination (Fig. 5), and survival probabilities (Fig. 6) that are of  
320 note are the similarities between arboreal taxa and the differences between arboreal and all other  
taxa.

322 Inspection of parameter estimates for the group-level covariates

## Analysis of diversity

324 All analyses of diversification and macroevolutionary rates has been done using only the birth-death  
model; this is because of the models better posterior predictive check performance (Fig. 3a, and 3b).

326 For the first half of the Cenozoic there appears to be a very slow decline and then plateau in total  
diversity until the WHEN? (Fig. 14a).

328 When viewed through the lens of diversification rate the structure behind this pattern begins to  
take shape (Fig. 14b). For approximately the first third of the Cenozoic, diversification rate is  
330 frequently below zero species gained per species present per two million years; this is broken up by  
few inferred spikes in diversification WHEN?. During the observed period of possible stability in

332 The comparison between per capita origination and extinction rate estimates reveals how  
diversification rate is formed (Fig. 14c, 14d). Origination rate seems most closely mimic

334 diversification rate while extinction rate has a saw-toothed pattern for the Cenozoic with no  
obvious emergent structure; inferred spikes in origination rate do not correspond to any spikes in  
336 extinction rate.

Diversity partitioned by ecotype reveals a lot of the complexity behind the pattern of mammal  
338 diversity for the Cenozoic (Fig. 15).

An impressive commonality across multiple ecotype-specific diversity time series are two spikes in  
340 diversity, either up or down (Fig. 15). Spikes of increased diversity are seen this is seen in all  
arboreal ecotypes, plantigrade insectivores, scansorial insectivores, and scansorial omnivores. The  
342 converse pattern, spikes of decreased diversity, are strongly observed in the diversity history of  
digitigrade herbivores, with weaker decreases observed in the histories of digitigrade carnivores, and  
344 unguligrade herbivores.

Arboreal ecotypes are estimated to have disappeared and reappeared as short bursts over the last  
346 65 million years in North America, with arboreal carnivores and insectivores being most often rarer  
than arboreal herbivores and omnivores.

348 Fossorial ecotypes, whether herbivorous and insectivorous appear rare or possibly absent for most  
the Cenozoic, which maximum estimated diversity being obtained in the latter half of the time  
350 series (Fig. 15).

Plantigrade ecotypes appear most variable, plantigrade herbivores having a very different diversity  
352 history than plantigrade carnivores and omnivores and plantigrade insectivores.

Unguligrade and plantigrade herbivores have relatively stable standing diversities throughout the  
354 Cenozoic of North America (Fig. 15). Similarly, digitigrade carnivorous taxa appear to have a  
relatively stable diversity for the Cenozoic.

## 356 Acknowledgements

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for helpful discussion and  
358 advice. This entire study would not have been possible without the Herculean effort of the  
many contributors to the Paleobiology Database. In particular, I would like to thank J. Alroy and  
360 M. Uhen for curating most of the mammal occurrences recorded in the PBDB. This is Paleobiology  
Database publication XXX.

<sup>362</sup> **References**

- Allen, L. J. S. 2011. An introduction to stochastic processes with applications to biology. 2nd ed.  
<sup>364</sup> Chapman and Hall/CRC, Boca Raton, FL.
- Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North  
<sup>366</sup> American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology* 127:285–311.
- . 2009. Speciation and extinction in the fossil record of North American mammals. Pages  
<sup>368</sup> 302–323 *in* R. K. Butlin, J. R. Bridle, and D. Schluter, eds. *Speciation and patterns of diversity*.  
Cambridge University Press, Cambridge.
- <sup>370</sup> Alroy, J., P. L. Koch, and J. C. Zachos. 2000. Global climate change and North American  
mammalian evolution. *Paleobiology* 26:259–288.
- <sup>372</sup> Bambach, R. K., A. M. Bush, and D. H. Erwin. 2007. Autecology and the filling of ecospace: Key  
metazoan radiations. *Palaeontology* 50:1–22.
- <sup>374</sup> Blois, J. L., and E. A. Hadly. 2009. Mammalian Response to Cenozoic Climatic Change. *Annual  
Review of Earth and Planetary Sciences* 37:181–208.
- <sup>376</sup> Brown, A. M., D. I. Warton, N. R. Andrew, M. Binns, G. Cassis, and H. Gibb. 2014. The  
fourth-corner solution - using predictive models to understand how species traits interact with  
<sup>378</sup> the environment. *Methods in Ecology and Evolution* 5:344–352.
- Bush, A. M., and R. K. Bambach. 2011. Paleoenologic Megatrends in Marine Metazoa, vol. 39.
- <sup>380</sup> Bush, A. M., R. K. Bambach, and G. M. Daley. 2007. Changes in theoretical ecospace utilization in  
marine fossil assemblages between the mid-Paleozoic and late Cenozoic. *Paleobiology* 33:76–97.
- <sup>382</sup> Clyde, W. C., and P. D. Gingerich. 1998. Mammalian community response to the latest Paleocene  
thermal maximum: an isotaphonomic study in the northern Bighorn Basin, Wyoming. *Geology*  
<sup>384</sup> 26:1011–1014.
- Cramer, B. S., K. Miller, P. Barrett, and J. Wright. 2011. Late Cretaceous-Neogene trends in deep  
<sup>386</sup> ocean temperature and continental ice volume: Reconciling records of benthic foraminiferal

- geochemistry ( $\delta^{18}\text{O}$  and Mg/Ca) with sea level history. *Journal of Geophysical Research: Oceans* 116:1–23.
- Eronen, J. T., C. M. Janis, C. P. Chamberlain, and A. Mulch. 2015. Mountain uplift explains differences in Palaeogene patterns of mammalian evolution and extinction between North America and Europe. *Proceedings of the Royal Society B: Biological Sciences* 282:20150136.
- Ezard, T. H. G., A. Purvis, and H. Morlon. 2016. Environmental changes define ecological limits to species richness and reveal the mode of macroevolutionary competition. *Ecology Letters* 19:899–906.
- Figueirido, B., C. M. Janis, J. A. Pérez-Claros, M. De Renzi, and P. Palmqvist. 2012. Cenozoic climate change influences mammalian evolutionary dynamics. *Proceedings of the National Academy of Sciences* 109:722–727.
- Foote, M. 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. *Paleobiology* 27:602–630.
- Foote, M., and J. J. Sepkoski. 1999. Absolute measures of the completeness of the fossil record. *Nature* 398:415–7.
- Gelman, A. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* pages 2865–2873.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. Bayesian data analysis. 3rd ed. Chapman and Hall, Boca Raton, FL.
- Gelman, A., and J. Hill. 2007. Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- Janis, C. M. 1993. Tertiary mammal evolution in the context of changing climates, vegetation, and tectonic events. *Annual Review of Ecology and Systematics* 24:467–500.
- Janis, C. M., J. Damuth, and J. M. Theodor. 2000. Miocene ungulates and terrestrial primary

- productivity: where have all the browsers gone? *Proceedings of the National Academy of Sciences*  
412 97:7899–904.
- Janis, C. M., G. F. Gunnell, and M. D. Uhen. 2008. Evolution of Tertiary mammals of North  
414 America. Vol. 2. Small mammals, xenarthrans, and marine mammals. Cambridge University  
Press, Cambridge.
- Janis, C. M., K. M. Scott, and L. L. Jacobs. 1998. Evolution of Tertiary mammals of North  
416 America. Vol. 1. Terrestrial carnivores, ungulates, and ungulate-like mammals. Cambridge  
University Press, Cambridge.
- Legendre, S. 1986. Analysis of mammalian communities from the Late Eocene and Oligocene of  
420 Southern France. *Paleovertebra* 16:191–212.
- Lloyd, G. T., J. R. Young, and A. B. Smith. 2011. Taxonomic Structure of the Fossil Record is  
422 Shaped by Sampling Bias. *Systematic Biology* 61:80–89.
- Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis  
424 of data from populations, metapopulations and communities. Elsevier, London.
- . 2012. Parameter-expanded data augmentation for Bayesian analysis of capture-recapture  
426 models. *Journal of Ornithology* 152:521–537.
- Royle, J. A., R. M. Dorazio, and W. a. Link. 2007. Analysis of Multinomial Models With Unknown  
428 Index Using Data Augmentation. *Journal of Computational and Graphical Statistics* 16:67–85.
- Royle, J. A., J. D. Nichols, M. Kéry, E. Ranta, and M. Kery. 2014. detection is of species when  
430 Modelling occurrence and abundance imperfect 110:353–359.
- Rubin, D. B. 1996. Multiple imputation after 18+ years. *Journal of the American Statistical  
432 Association* 91:473–489.
- Smits, P. D. 2015. Expected time-invariant effects of biological traits on mammal species duration.  
434 *Proceedings of the National Academy of Sciences* 112:13015–13020.
- Strömberg, C. A. E. 2005. Decoupled taxonomic radiation and ecological expansion of open-habitat

- <sup>436</sup> grasses in the Cenozoic of North America. *Proceedings of the National Academy of Sciences of the United States of America* 102:11980–4.
- <sup>438</sup> Tomiya, S. 2013. Body Size and Extinction Risk in Terrestrial Mammals Above the Species Level. *The American Naturalist* 182:196–214.
- <sup>440</sup> Wang, S. C., P. J. Everson, H. J. Zhou, D. Park, and D. J. Chudzicki. 2016. Adaptive credible intervals on stratigraphic ranges when recovery potential is unknown. *Paleobiology* 42:240–256.
- <sup>442</sup> Wang, S. C., and C. R. Marshall. 2016. Estimating times of extinction in the fossil record. *Biology Letters* 12:20150989.
- <sup>444</sup> Warton, D. I., B. Shipley, and T. Hastie. 2015. CATS regression - a model-based approach to studying trait-based community assembly. *Methods in Ecology and Evolution* 6:389–398.
- <sup>446</sup> Zachos, J. C., G. R. Dickens, and R. E. Zeebe. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451:279–283.
- <sup>448</sup> Zachos, J. C., M. Pagani, L. Sloan, E. Thomas, and K. Billups. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:686–693.

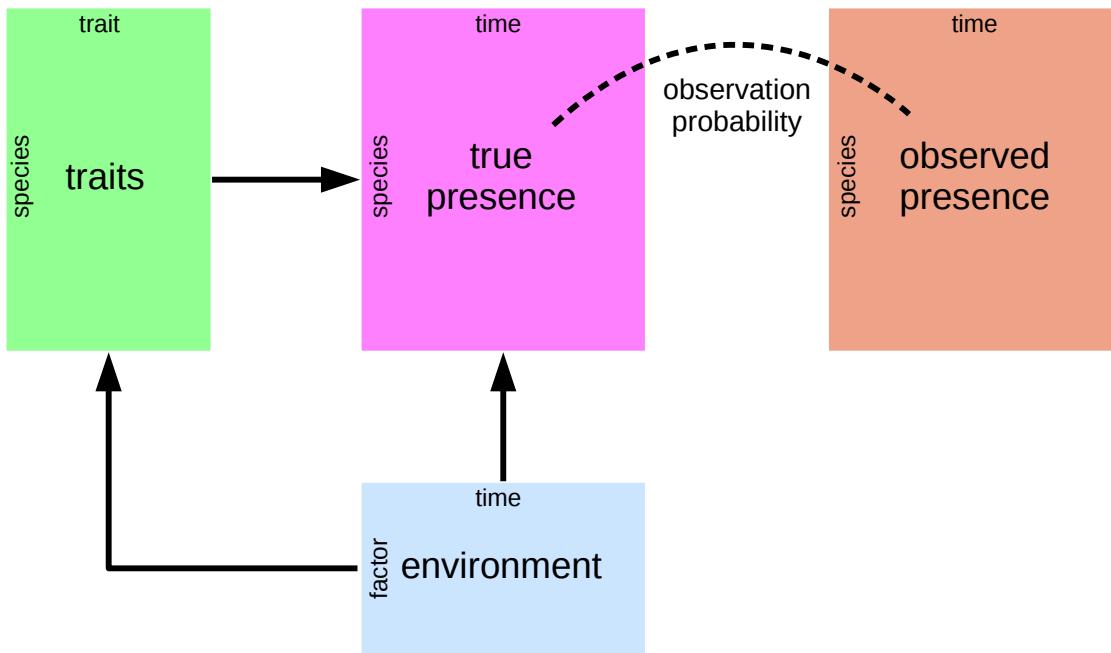


Figure 1: Conceptual diagram of the paleontological fourth corner problem. The observed presence matrix (orange) is the empirical presence/absence pattern for all species for all time points; this matrix is an incomplete observation of the “true” presence/absence pattern (purple). The estimated true presence matrix is modeled as a function of both environmental factors over time (blue) and multiple species traits (green). Additionally, the effect of environmental factors on species traits are also modeled as traits are expected to mediate the effects of a species environmental context. This diagram is based partially on material presented in Brown et al. (2014) and Warton et al. (2015).

	Time Bin							
	1	2	3	4	5	6	7	8
Observed	0	0	0	1	0	1	1	0
Certain	?	?	?	1	1	1	1	?
Potential	0	0	0	1	1	1	1	0
Potential	0	0	1	1	1	1	1	0
Potential	0	1	1	1	1	1	1	0
Potential	1	1	1	1	1	1	1	0
Potential	0	0	0	1	1	1	1	1
Potential	0	0	1	1	1	1	1	1
Potential	0	1	1	1	1	1	1	1
Potential	1	1	1	1	1	1	1	1

Figure 2: Conceptual figure of all possible occurrence histories for an observed species. The first row represents the observed presence/absence pattern for a single species at eight time points. The second row corresponds to the known aspects of the “true” occurrence history of that species. The remaining rows correspond to all possible occurrence histories that are consistent with the observed data. The process of parameter marginalization described in the text

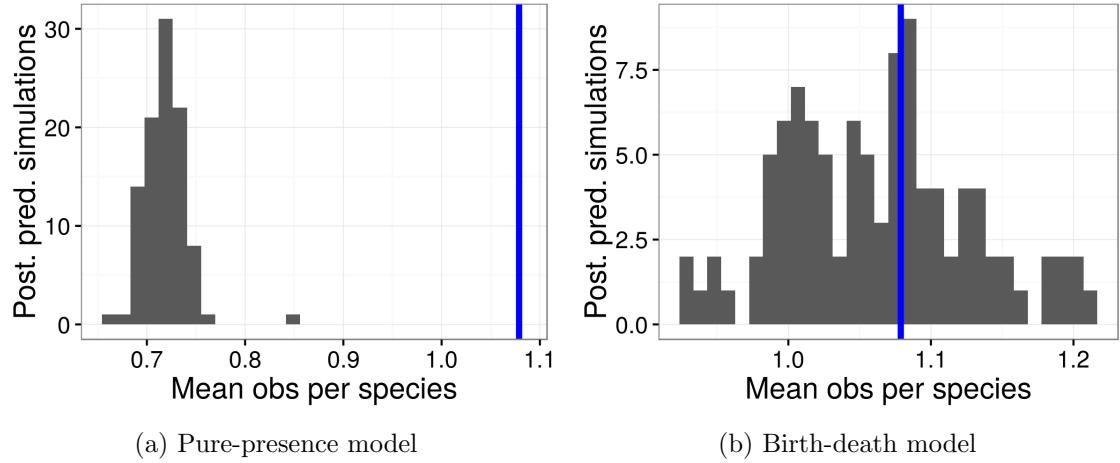


Figure 3: Comparison of the average observed number of occurrences per species (blue line) to the average number of occurrences from 100 posterior predictive datasets using the posterior estimates from the pure-presence and birth-death models.

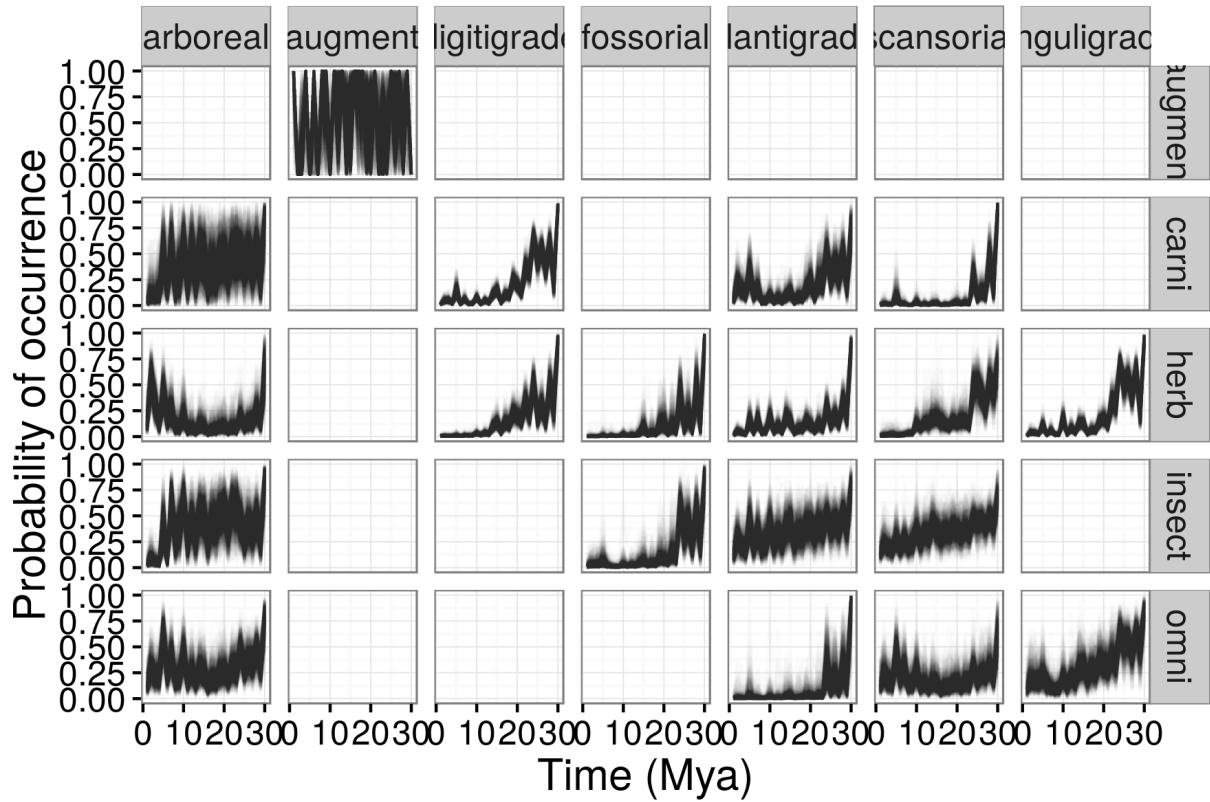


Figure 4: Probability of a mammal ecotype occurring over time as estimated from the pure-presence model. Each panel depicts 100 random samples from the model's posterior. The columns are by locomotor category and rows by dietary category; their intersections are the observed and analyzed ecotypes. Panels with no lines are ecotypes not observed in the dataset.

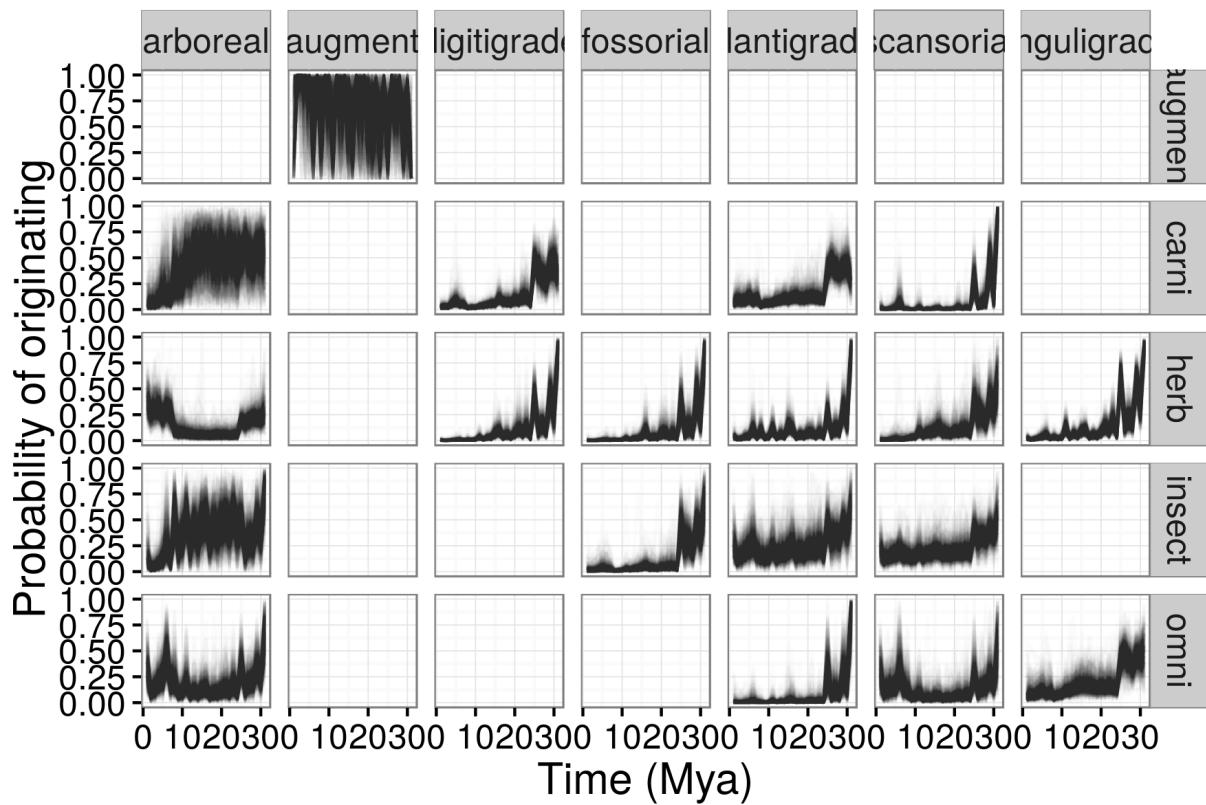


Figure 5: Probability of a mammal ecotype origination probabilities at each time point as estimated from the birth-death model. Each panel depicts 100 random samples from the model's posterior. The columns are by locomotor category and rows by dietary category; their intersections are the observed and analyzed ecotypes. Panels with no lines are ecotypes not observed in the dataset.

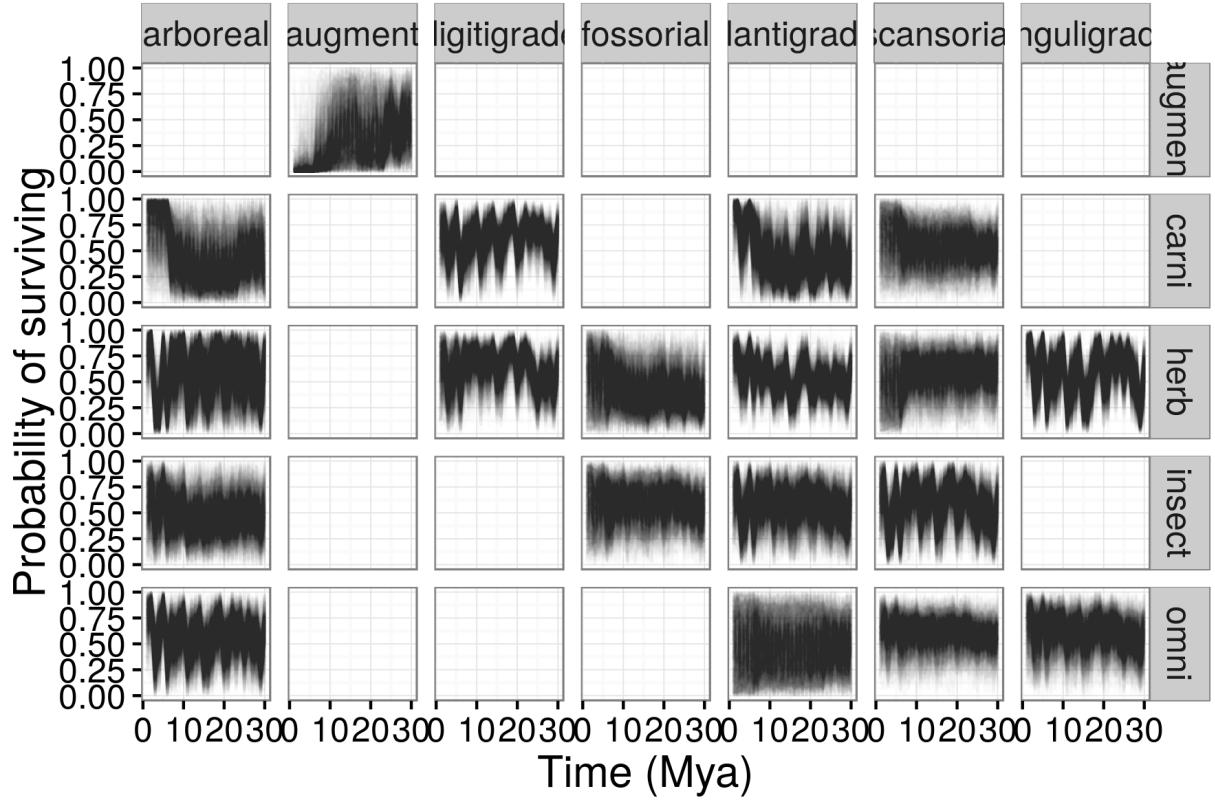


Figure 6: Probability of a mammal ecotype survival probabilities at each time point as estimated from the birth-death model. Each panel depicts 100 random samples from the model’s posterior. The columns are by locomotor category and rows by dietary category; their intersections are the observed and analyzed ecotypes. Panels with no lines are ecotypes not observed in the dataset.

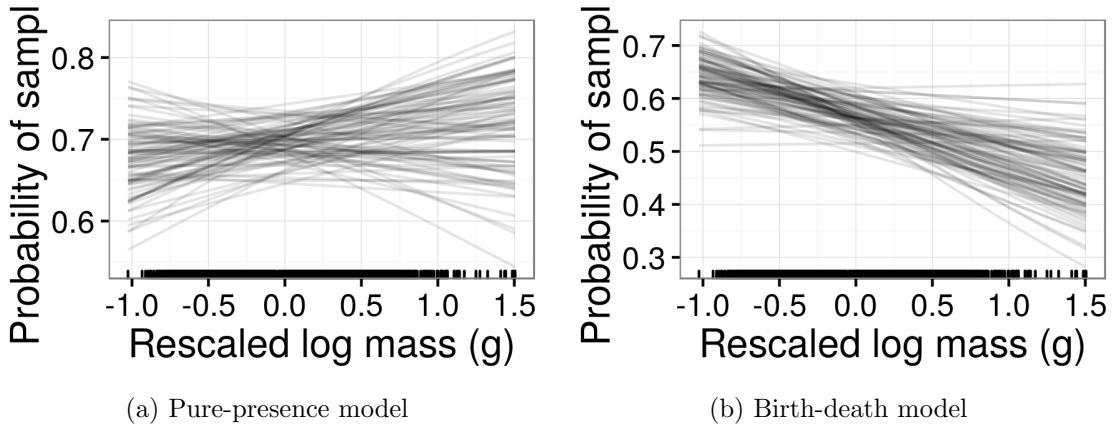


Figure 7: Estimates of the effect of species mass on probability of observing a present species ( $p$ ). Mass has been log-transformed, centered, and rescaled; this means that a mass of 0 corresponds to the mean of log-mass of all observed species and that mass is in standard deviation units. Estimates are from both the pure-presence and birth-death models.

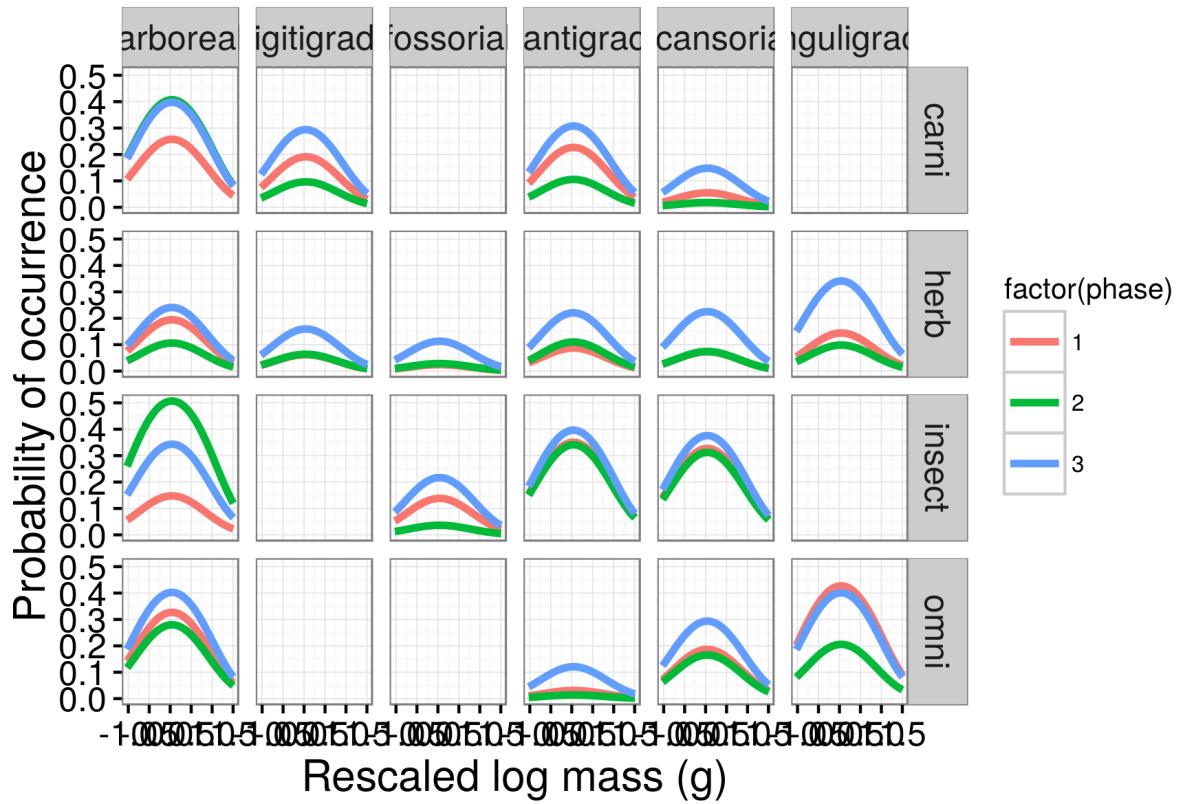


Figure 8: Mean estimate of the effect of species mass on the probability of a species occurrence for each of the three plant phases. The effect of mass is considered constant over time and that the only aspect of the model that changes with plant phase is the intercept of the relationship between mass and occurrence. The three plant phases are indicated by the color of the line. Mass has been log-transformed, centered, and rescaled; this means that a mass of 0 corresponds to the mean of log-mass of all observed species and that mass is in standard deviation units. Only the mean estimates of the effects of both mass and plant phase are plotted for clarity; these estimates are obviously made with uncertainty.

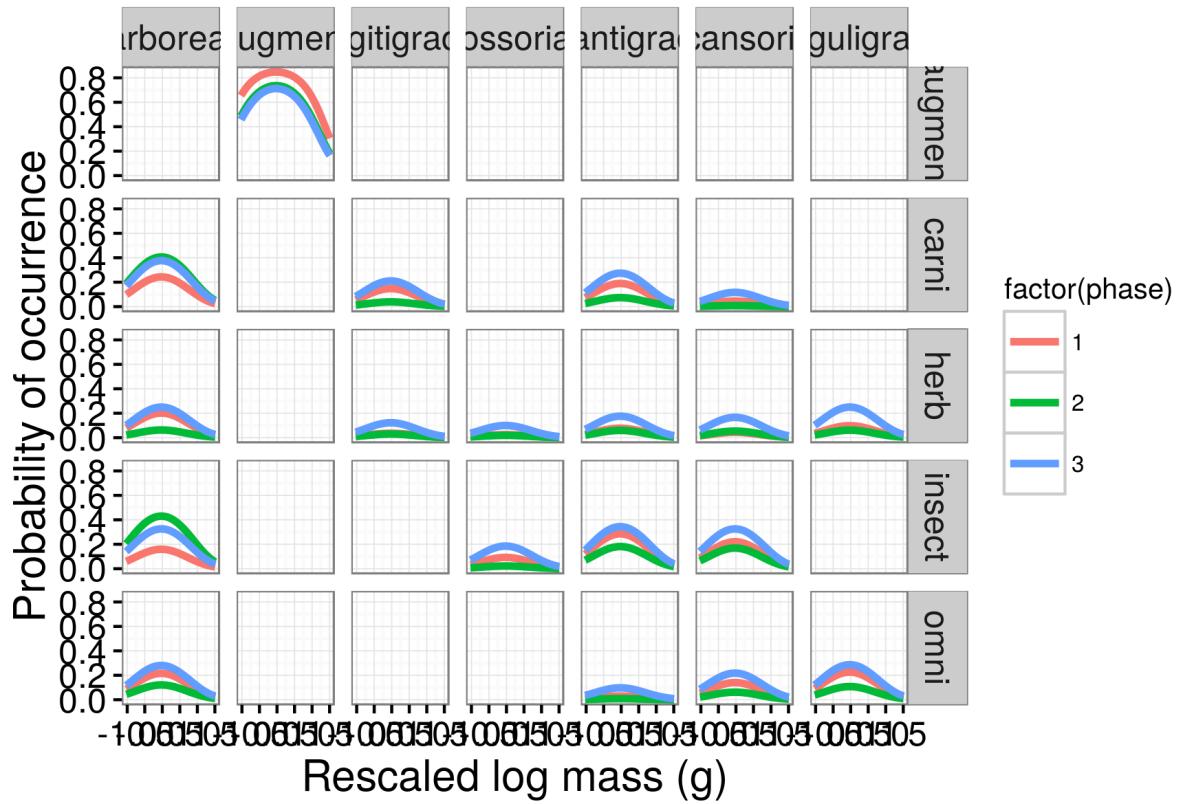


Figure 9: Mean estimate of the effect of species mass on the probability of a species originating for each of the three plant phases. The effect of mass is considered constant over time and that the only aspect of the model that changes with plant phase is the intercept of the relationship between mass and origination. The three plant phases are indicated by the color of the line. Mass has been log-transformed, centered, and rescaled; this means that a mass of 0 corresponds to the mean of log-mass of all observed species and that mass is in standard deviation units. Only the mean estimates of the effects of both mass and plant phase are plotted for clarity; these estimates are obviously made with uncertainty.

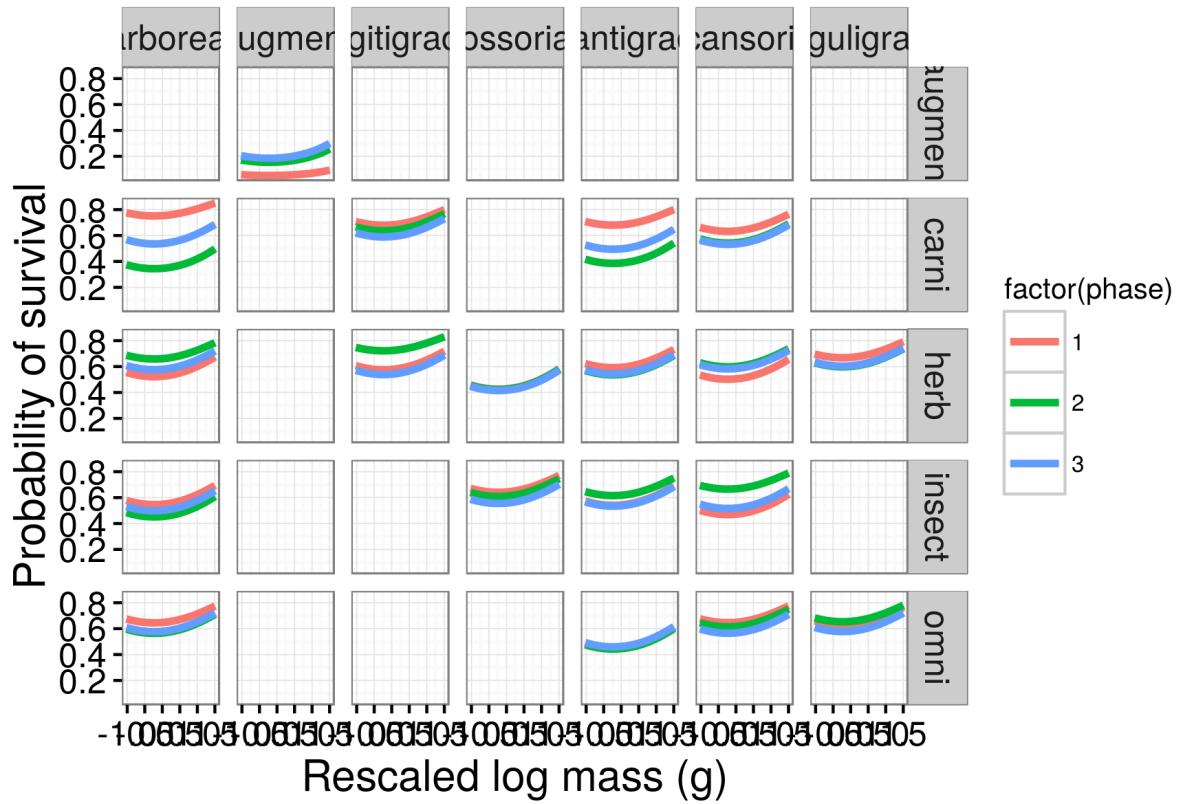


Figure 10: Mean estimate of the effect of species mass on the probability of a species survival for each of the three plant phases. The effect of mass is considered constant over time and that the only aspect of the model that changes with plant phase is the intercept of the relationship between mass and survival. The three plant phases are indicated by the color of the line. Mass has been log-transformed, centered, and rescaled; this means that a mass of 0 corresponds to the mean of log-mass of all observed species and that mass is in standard deviation units. Only the mean estimates of the effects of both mass and plant plant are plotted for clarity; these estimates are obviously made with uncertainty.

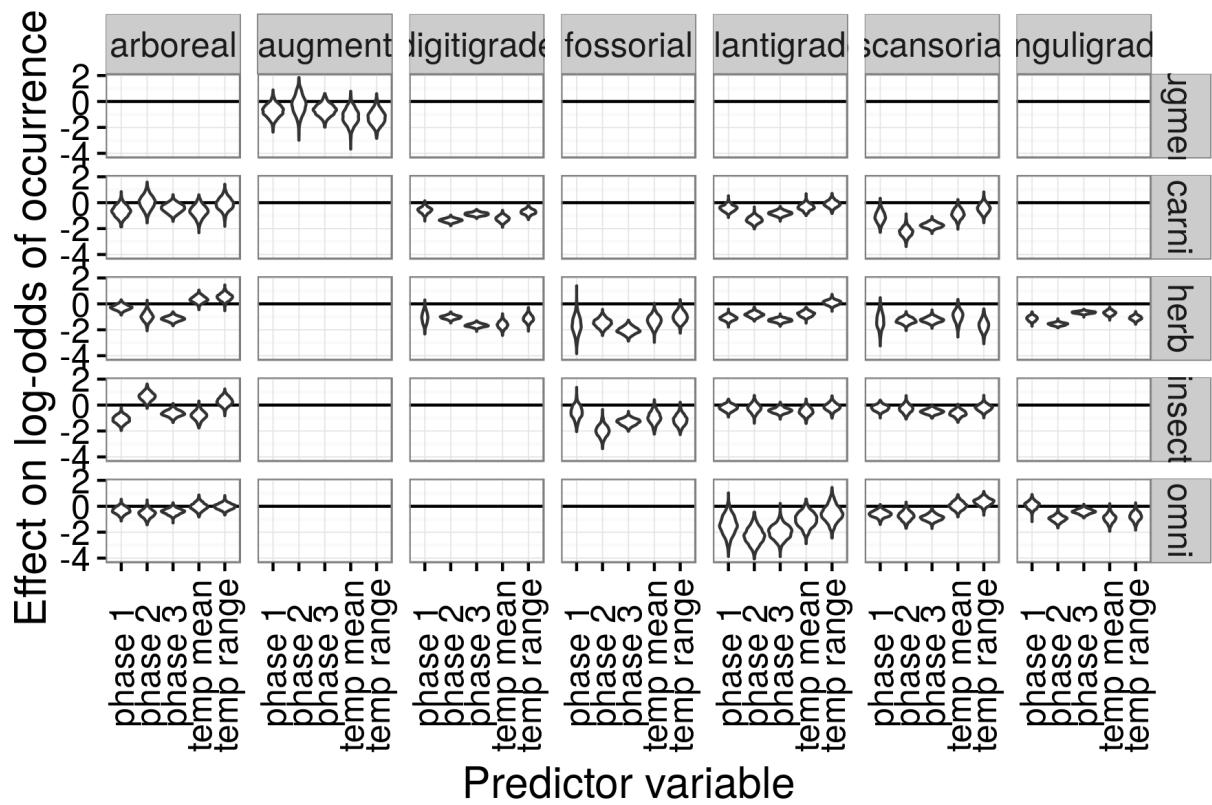


Figure 11: Estimated effects of the group-level covariates describing environmental context on log-odds of species occurrence. These estimates are from the pure-presence model.

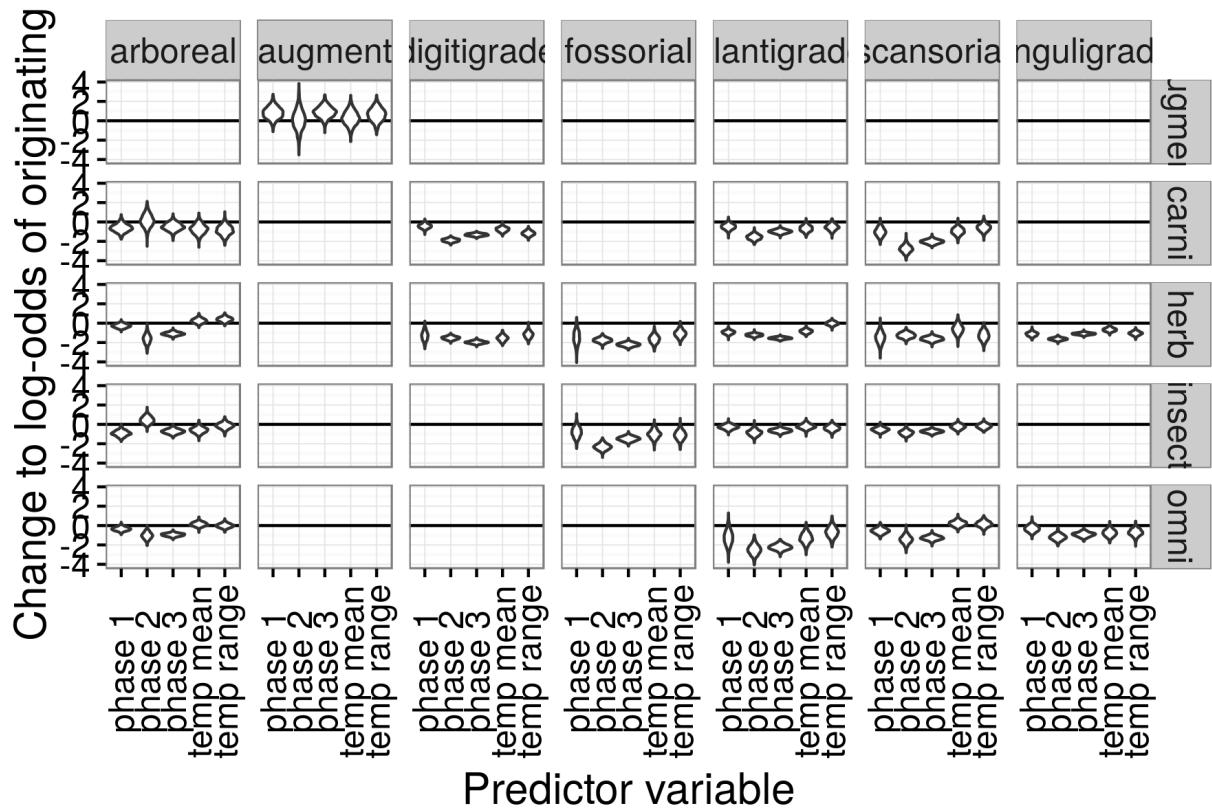


Figure 12: Estimated effects of the group-level covariates describing environmental context on log-odds of species origination. These estimates are from the birth-death model.

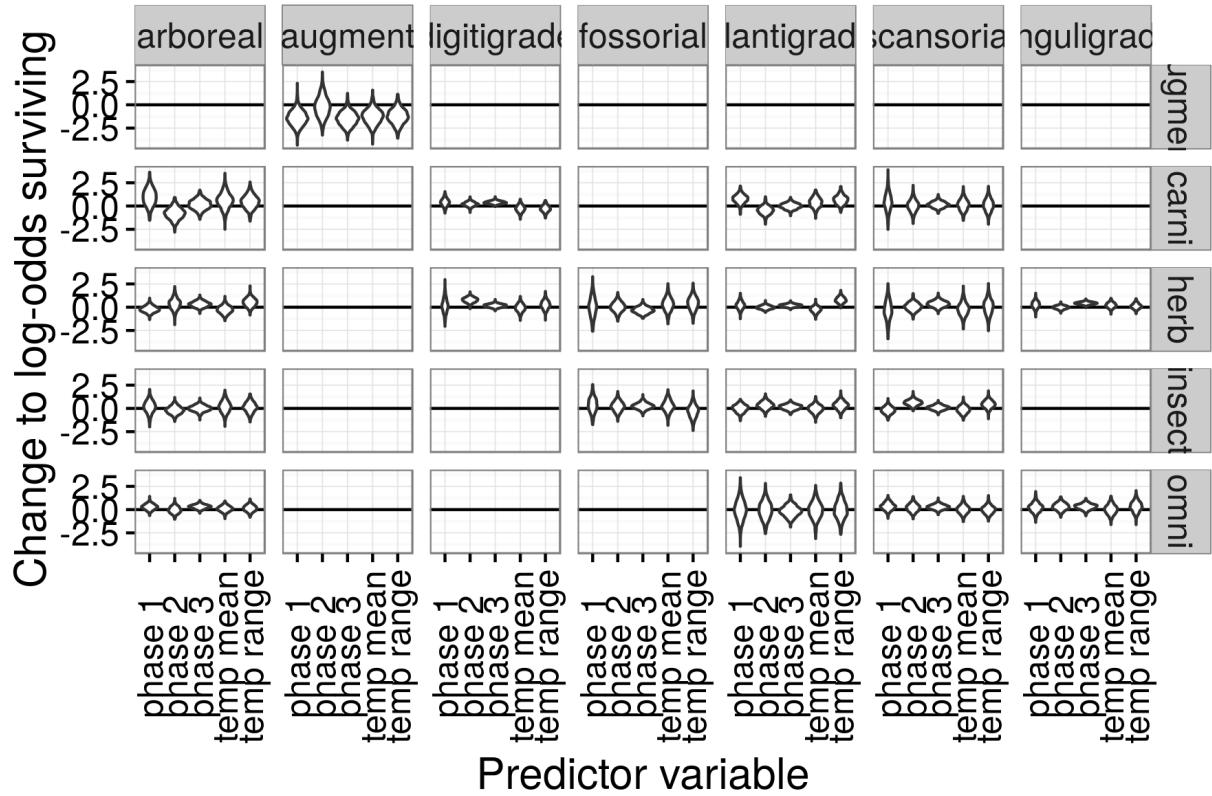


Figure 13: Estimated effects of the group-level covariates describing environmental context on log-odds of species survival. These estimates are from the birth-death model.

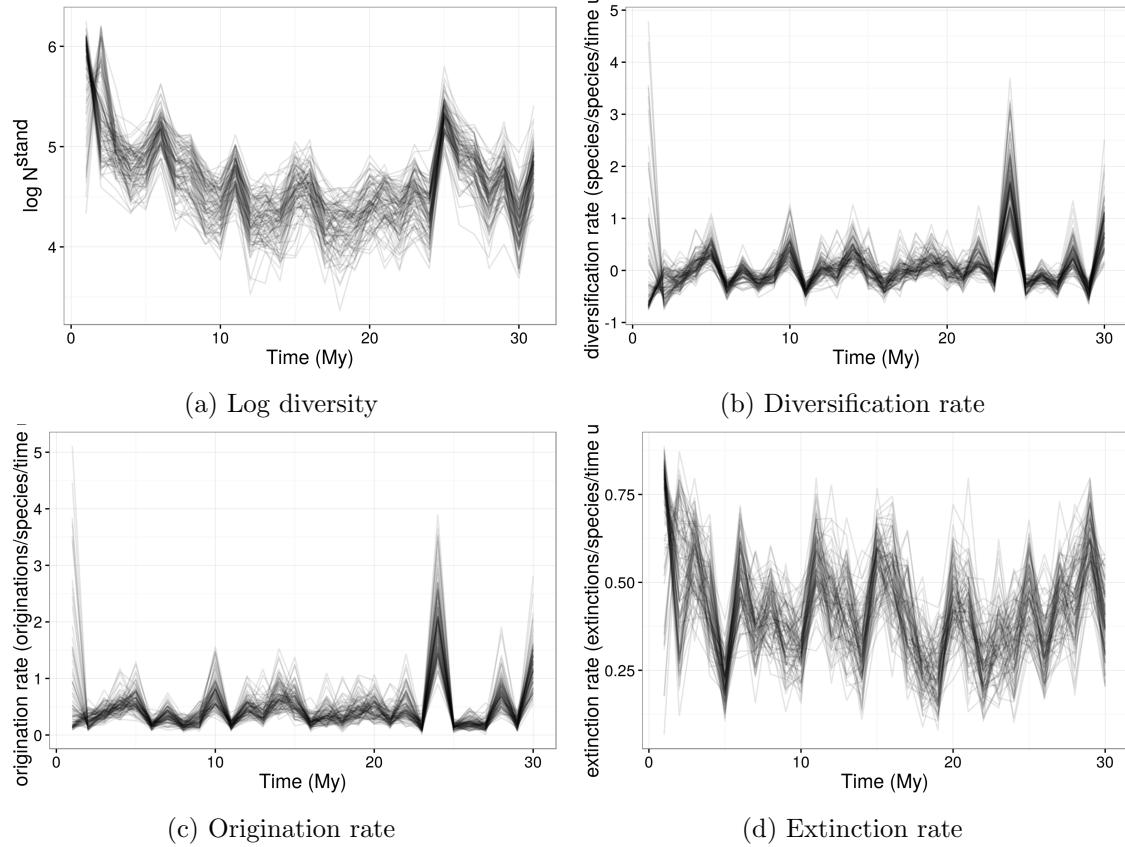


Figure 14: Posterior estimates of the time series of Cenozoic North American mammal diversity and its characteristic macroevolutionary rates; all estimates are from the birth-death model and 100 posterior draws are plotted to indicate the uncertainty in these estimates. The dramatic differences between diversity estimates at the first and second time points and the penultimate and last time points in this series are caused by well known edge effects in discrete-time birth-death models caused by  $p_{-,t=1}$  and  $p_{-,t=T}$  being partially unidentifiable (Royle and Dorazio, 2008); the hierarchical modeling strategy used here helps mitigate these effects but they are still present (Gelman et al., 2013; Royle and Dorazio, 2008). Diversification rate is in units of species gained per species present per time unit (2 My), origination rate is in units of species originating per species present per time unit, and extinction rate is in units of species becoming extinct per species present per time unit.

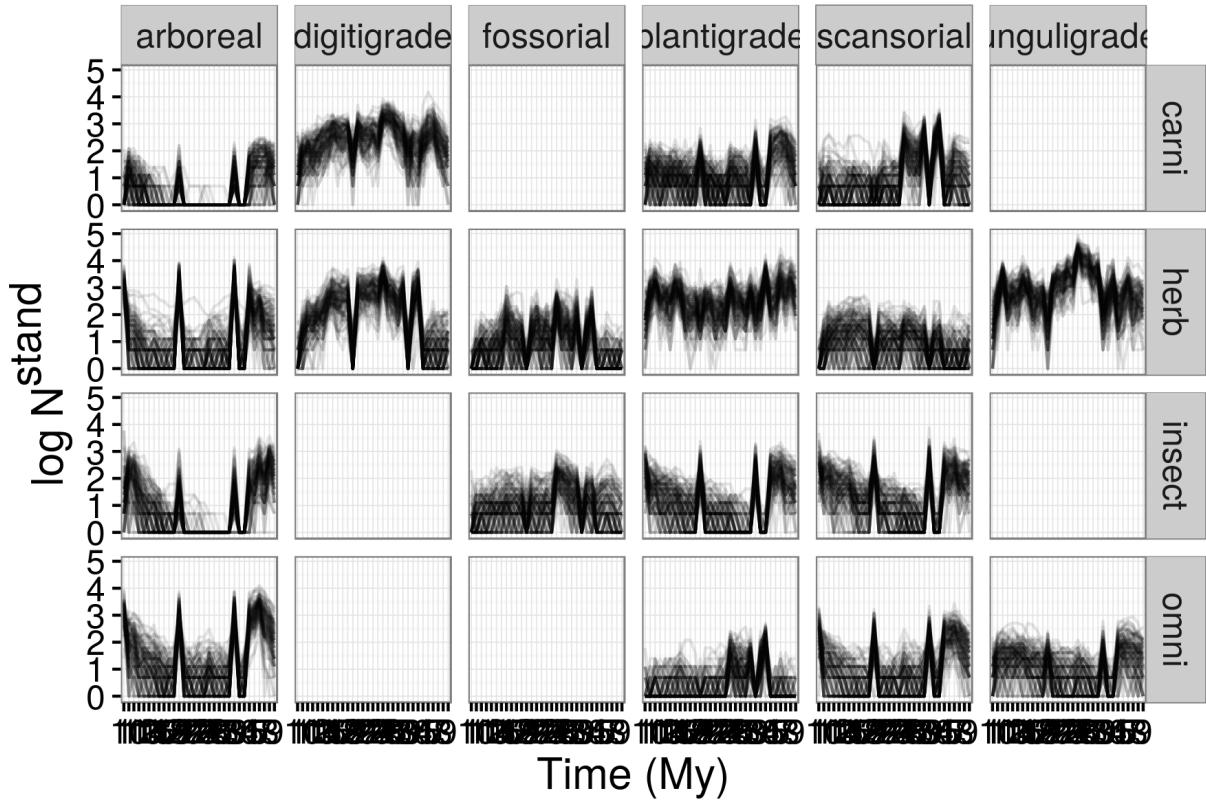


Figure 15: Posterior of standing log-diversity of North American mammals by ecotype for the Cenozoic as estimated from the birth-death model; 100 posterior draws are plotted to indicate the uncertainty in these estimates and what is technically plotted is log of diversity plus 1. The dramatic differences between diversity estimates at the first and second time points and the penultimate and last time points in this series are caused by well known edge effects in discrete-time birth-death models caused by  $p_{-,t=1}$  and  $p_{-,t=T}$  being partially unidentifiable (Royle and Dorazio, 2008); the hierarchical modeling strategy used here helps mitigate these effects but they are still present (Gelman et al., 2013; Royle and Dorazio, 2008).