

Materials and Methods

Taxon occurrences and species-level information

All fossil occurrence information used in this analysis was downloaded from the Paleobiology Database (PBDB). The initial download restricted occurrences to Mammalia observed in North America between the Maastrichtian (72-66 Mya) and Gelasian (2.58-1.8 Mya) stages [?]. Taxonomic, stratigraphic, and ecological metadata for each occurrence and species was also downloaded. A new download for a raw, unfiltered PBDB datafile following the same criterion used here is available at <http://goo.gl/2slgeU>. The raw datafile used as a part of this study, along with all code for filtering and manipulating this download is available at <http://github.com/psmits/coping>.

After being downloaded, the raw occurrence data was then sorted, cleaned, and manipulated programmatically before analysis. Occurrences were restricted to those occurring between 64 and 2 million years ago (Mya); this age restriction was to insure that observation time series lines up with the temperature time series [?]. All taxa whose life habit was classified as either volant (i.e. Chiroptera) or aquatic (e.g. Cetacea) were excluded from this analysis because of their lack of direct applicability to the study of terrestrial species pools.

Many species taxonomic assignments as present in the raw PBDB data were updated for accuracy and consistency. Species present in the PBDB have some taxonomic information, including possible Family and Order assignments. In order to increase consistency between species and reflect more recent taxonomic assignments, each species taxonomic assignments updated as follows: 1) species family and order assignments as present in the Encyclopedia of life (<http://eol.org>) was downloaded using the *taxize* package for R; 2) for species not present in the EoL or not assigned order, their taxonomic information was further updated based on whatever family information was recorded in the PBDB or EoL; 3) for species still missing order assignments, their genus information was used to assign either an order or family, which was then used to assign an order. This procedure is similar to that used in [?] and is detailed in the code repository associated with this study.

Species ecotype is defined based on a combination of locomotor and diet categories; the goal is to classify species based on the manner with which

they interact with their environment. Mammal species records in the PBDB have life habit (i.e. locomotor category) and dietary category assignments. In order to simplify interpretation, analysis, and per-ecotype sample size these classifications were coarsened in a similar manner to [?] (Table 1). The life history category was then further edited to better reflect the diversity of mammal locomotor modes and our knowledge of species' ecologies. Ground dwelling species locomotor categories were reassigned based on the ankle posture associated with their taxonomic group, as described in Table 2 [?]. Ankle posture was assumed uniform for all species within a taxonomic group except for those species assigned a non-ground dwelling locomotor category by the PBDB, which retained their non-ground dwelling assignment. All species for which it was possible to assign a locomotor category had one assigned, including species for which post-crania are unknown but for which a taxonomic grouping is known. Ground dwelling species which were unable to be reassigned based on ankle posture were excluded from analysis. Finally, ecotype categories with less than 10 total species were excluded, yielding a total of 18 observed ecotypes out of a possible 24.

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.

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.

Table 2: Ankle posture assignment as based on taxonomy. Assignments are based on [?]. Taxonomic groups are presented alphabetically and without reference for the nestedness of families in orders.

Order	Family	Stance
	Ailuridae	plantigrade
Continued on next page		

Table 2 – continued from previous page

Order	Family	Stance
Dinocerata	Allomyidae	plantigrade
	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
	Dipodidae	digitigrade
	Elephantidae	digitigrade
	Entelodontidae	unguligrade
	Eomyidae	plantigrade
	Erethizontidae	plantigrade
	Erinaceidae	plantigrade
	Esthonychidae	plantigrade
	Eutypomyidae	plantigrade
	Felidae	digitigrade
	Florentiamyidae	plantigrade
	Gelocidae	unguligrade
	Geolabididae	plantigrade
	Glyptodontidae	plantigrade
Continued on next page		

Table 2 – continued from previous page

Order	Family	Stance
Lagomorpha	Gomphotheriidae	unguligrade
	Hapalodectidae	plantigrade
	Heteromyidae	digitigrade
	Hyaenidae	digitigrade
	Hyaenodontidae	digitigrade
	Hypertragulidae	unguligrade
	Ischyromyidae	plantigrade
	Jimomyidae	plantigrade
		digitigrade
	Leptictidae	plantigrade
	Leptochoeridae	unguligrade
	Leptomerycidae	unguligrade
	Mammutidae	unguligrade
	Megalonychidae	plantigrade
	Megatheriidae	plantigrade
Mesonychia	Mephitidae	plantigrade
	Merycoidodontidae	digitigrade
		unguligrade
	Mesonychidae	digitigrade
	Micropternodontidae	plantigrade
	Mixodectidae	plantigrade
	Moschidae	unguligrade
	Muridae	plantigrade
	Mustelidae	plantigrade
	Mylagaulidae	fossorial
Notoungulata	Mylodontidae	plantigrade
	Nimravidae	digitigrade
	Nothrotheriidae	plantigrade
		unguligrade
	Oromerycidae	unguligrade
	Oxyaenidae	digitigrade
	Palaeomerycidae	unguligrade
	Palaeoryctidae	plantigrade
	Pamphathiidae	plantigrade
	Pantolambdidae	plantigrade
		plantigrade
Continued on next page		

Table 2 – continued from previous page

Order	Family	Stance
Perissodactyla	Periptychidae	digitigrade
		unguligrade
Primates	Phenacodontidae	unguligrade
		plantigrade
	Procyonidae	plantigrade
	Proscalopidae	plantigrade
	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

Estimates of species mass used in this study were sourced from multiple databases and papers, especially those focusing on similar macroevolutionary or macroecological questions [? ? ? ? ?]; this is similar to what was done in [?]. When species mass was not available, proxy measures were used and then transformed into estimates of mass. For example, given a measurement of a mammal tooth size, it is possible and routine to estimate its mass given some regression equation. The PBDB has one or more body part measures for many species. These were used as body size proxies for many species, as was the case in [?]. Mass was log-transformed and then rescaled by first subtracting mean log-mass from all mass estimates, then dividing by two-times its standard deviation; this insures that the magnitude of effects for both continuous and discrete covariates are directly comparable [? ?].

In total, 1400 mammal species occurrence histories were included in this study after applying all of the restrictions above.

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	[?]
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 tooththrow area	[?]
Lagomorph	$\log(m) = 4.468x - 3.002$	lower m1 length	[?]
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	[?]
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 tooththrow length	[?]

All fossil occurrences from 64 to 2 million years ago (Mya) were binned into the 19 North American Land Mammal Ages (NALMA) covered by this interval CITATION. The choice of binning by NALMA reflects the belief that these represent distinct communities or periods of mammal evolution, something that is central to this study. Additionally, because of the inherently discrete nature of the fossil record it can be hard to re-bin fossils by temporal interval because of the inherent uncertainty in their ages CITATION.

Environmental and temporal covariates

The environmental covariates used in this study are collectively referred to as group-level covariates because they predict the response of a “group” of individual-level observations (i.e. species occurrences of an ecotype). Addi-

Table 4: Definitions of the start and stop times of the three plant phases used this study as defined by ?].

Plant phase	Phase code	Start	Stop
Paleocene-Eocene	Pa-Eo	66	50
Eocene-Miocene	Eo-Mi	50	16
Miocene-Pleistocene	Mi-Pl	16	2

tionally, these covariates are defined for temporal bins and not the species themselves; as such they predict the parts of each species occurrence history. The group-level covariates in this study are two global temperature estimates and the Cenozoic “plant phases” defined by ?].

Global temperature across most of the Cenozoic was calculated from Mg/Ca isotope record from deep sea carbonates [?]. Mg/Ca based temperature estimates are preferable to the frequently used $\delta^{18}\text{O}$ temperature proxy [? ? ? ?] because Mg/Ca estimates do not conflate temperature with ice sheet volume and depth/stratification changes. The former is particularly important to this analysis as the current polar ice-caps appeared and grew during the second half of the Cenozoic. These properties make Mg/Ca based temperature estimates preferable for macroevolutionary and macroecological studies [?]. Two aspects of the Mg/Ca-based temperature curve were included in this analysis: mean and range. Both were calculated as the mean of all respective estimates for each of the NALMA units. The distributions of the temperature mean and range estimates were then rescaled by subtracting their respective means from all values and then dividing by twice their respective standard deviations.

The second set of environmental factors included in this study are the Cenozoic plant phases defined in ?]. Graham’s plant phases are holistic descriptors of the taxonomic composition of 12 ecosystem types, which plants are present at a given time, and the relative modernity of those plant groups with younger phases representing increasingly modern taxa [?]. ?] defines four intervals from the Cretaceous to the Pliocene, though only three of these intervals take place during the time frame being analyzed. Graham’s plant phases was included as a series of “dummy variables” encoding the three phases included in this analysis [?]; this means that the first phase is synonymous with the intercept and subsequent phases are defined by their differences from the first

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

Table 5: Transition matrix for the birth-death model (Eq. ??). Note that while there are only two state “codes” (0, 1), there are in fact three states: never having originated 0_{never} , present 1, extinct $0_{extinct}$ [?].

phase. The temporal boundaries of these plant phases are defined in Table 4.

Modelling species occurrence

At the core of the model used in this study is hidden Markov process where the latent process has an absorbing state; this can also be referred to as a discrete-time birth-death model [?]. While there are only two state “codes” in a presence-absence matrix (i.e. 0/1), there are in fact three states in a birth-death model: not having originated yet, extant, and extinct. The last of these is the absorbing state, as once a species has gone extinct it cannot re-originate [?]. Thus, in the transition matrices the probability of an extinct species changing states is 0 (Table 5). See below for parameter explanations (Tables ??, and ??).

Basic model

I will begin defining the model used in this study by focusing on the basic machinery of the hidden Markov process at the model’s core. This basic model is similar to the Jolly-Seber capture-mark-recapture model CITATION. The basic model has three characteristic probabilities: probability p of observing a species given that it is present, probability π of a species surviving from one time to another, and probability ϕ of a species first appearing [?] (Table ??). In this formulation, the probability of a species becoming extinct is $1 - \pi$. The inclusion of species and temporal information means that all three of these probabilities are defined for every species at every time point (Table ??;

Table 6: Parameters associated with the observation process part of the hidden Markov model.

Parameter	dimensions	explanation
y	$N \times T$	observed species presence/absence
z	$N \times T$	“true” species presence/absence
p	$N \times T$	probability of observing a species at time t if it is present
ϕ	$N \times T$	probability of species originating from time t to $t + 1$ if it is not present
π	$N \times T - 1$	probability of species surviving at time t , given that it is already originated

how this is accomplished is described below. Importantly, only origination can occur during the first time step as nothing is present in order to survive.

$$\begin{aligned}
y_{i,t} &\sim \text{Bernoulli}(p_{i,t}z_{i,t}) \\
z_{i,1} &\sim \text{Bernoulli}(\phi_{i,1}) \\
z_{i,t} &\sim \text{Bernoulli}\left(z_{i,t-1}\pi_{i,t} + \sum_{x=1}^t (1 - z_{i,x})\phi_{i,t}\right)
\end{aligned} \tag{1}$$

The parameters associated with Equation 1 are described in Table 6; this formulation is identical to that described in where individual covariates are included for all parameters [?]. The product term that appears when calculating values of z not at $t = 1$ ensures that once a species goes extinct it does not re-originate. The basic model described here (Eq. ??) does not include the additional, necessary prior information.

Expanding on the basics

Expanding on the basic model involves modeling the observation, origination and survival as independent multi-level logistic regressions. Origination and survival probabilities share the same covariates and model structure, but observation probability is modeled as a function of a smaller selection of covariates.

The probability of observing a species given that it is present (logit p) is modeled as a logistic regression with a time-varying intercept with an additional zero-centered varying effect for species’ functional group, respectively. The effect of species mass was also included through a slope term.

Table 7: Parameters for the first expansions

Parameter	dimensions	explanation
u	T	time-varying intercept
f	J	effect of functional group on observation
f^ϕ	$N \times T - 1$	
f^π	$N \times T$	
o^ϕ	K	
o^π	K	
β^ϕ	1	
β^π	1	

The log-odds of a species originating (logit pi) or surviving (logit ϕ) are modeled independently but take the same form: a regression with an intercept that varies by both time and functional group, an additional taxonomic order varying-intercept term, and the slope term for species mass. Importantly, the time and functional group varying-intercept is itself modeled such that each the intercept for each functional group is a time series with it's own group-level covariates (described below).

The expanded model incorporating these regression models is written as

$$\begin{aligned}
y_{i,t} &\sim \text{Bernoulli}(p_{i,t}z_{i,t}) \\
p_{i,t} &= \text{logit}^{-1}(u_t + f_{j[i]}) \\
z_{i,1} &\sim \text{Bernoulli}(\phi_{i,1}) \\
z_{i,t} &\sim \text{Bernoulli}\left(z_{i,t-1}\pi_{i,t} + \sum_{x=1}^t (1 - z_{i,x})\phi_{i,t}\right) \\
\phi_{i,t} &= \text{logit}^{-1}(f_{j[i],t}^\phi + o_{k[i]}^\phi + \beta^\phi m_i). \\
\pi_{i,t} &= \text{logit}^{-1}(f_{j[i],t}^\pi + o_{k[i]}^\pi + \beta^\pi m_i)
\end{aligned} \tag{2}$$

How the priors for the varying-effects and coefficients in this expanded model are described below along with the complete model.

Complete model

The expanded model (Eq. 2) is still incomplete as it is missing the group-level covariates such as global temperature, and it is missing all of the necessary

priors.

Here I describe how the effects of mammal functional group on origination and survival are modeled. f^ϕ and f^π are modeled as the responses from a multivariate normal distribution where the elements of the mean vector are independent regressions. The time-series structure of these regressions is represented as a

$$\begin{aligned}
f^\phi &\sim \text{MVN}(\mu^\phi, \Sigma^\phi) \\
f^\pi &\sim \text{MVN}(\mu^\pi, \Sigma^\pi) \\
\mu_{j,t}^\phi &= \alpha_{j,t}^\phi + U * \gamma_j^\phi \\
\mu_{j,t}^\pi &= \alpha_{j,t}^\pi + U * \gamma_j^\pi \\
\alpha_{j,t}^\phi &\sim \begin{cases} \mathcal{N}(0, \sigma_j^\phi) & \text{if } t = 1 \\ \mathcal{N}(\alpha_{j,t-1}^\phi, \sigma_j^\phi) & \text{if } t > 1 \end{cases} \\
\alpha_{j,t}^\pi &\sim \begin{cases} \mathcal{N}(0, \sigma_j^\pi) & \text{if } t = 1 \\ \mathcal{N}(\alpha_{j,t-1}^\pi, \sigma_j^\pi) & \text{if } t > 1 \end{cases}
\end{aligned} \tag{3}$$

Posterior inference and model adequacy

Computer programs that implement joint posterior inference for the above model (Eqs. ??) were written in the probabilistic programming language Stan [?]. Both models feature a large matrix of latent discrete parameters z (Tables ??, ??; Eqs. ??, ??). All methods for posterior inference implemented in Stan are derivative-based; this causes complications for actually implementing the above models, because integers do not have derivatives. Instead of implementing a latent discrete parameterization, the log posterior probabilities of all possible states of the latent parameters z were calculated and summed (i.e. marginalized).

Species durations at minimum range through from a species first appearance to their last appearance in the fossil record, but the incompleteness of all observations means that the actual times of origination and extinction are unknown. The marginalization approach used here means that the probabilities of all possible histories for a species are calculated, from the end members of the species having existed for the entire study interval and the species having

only existed between the directly observed first and last appearances to all possible intermediaries (Fig 1) [?]. This process is identical, language-wise, to assuming range-through and then estimating the possibility of all possible range extension due to incomplete sampling.

	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 1: 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. By marginalizing over all possible occurrence histories, the probability of each potential history is estimated. The process of parameter marginalization is described in the text.

The combined size of the dataset and large number of parameters in both models (Eqs. ??, ??), specifically the total number of latent parameters that are the matrix z , means that stochastic approximate posterior inference is computationally very slow even using NUTS based HMC as implemented in Stan [?]. Instead, an approximate Bayesian approach was used: variational inference. A recently developed automatic variational inference algorithm called “automatic differentiation variational inference” (ADVI) is implemented in Stan and was used here [? ?]. ADVI assumes that the posterior is Gaus-

sian but still yields a true Bayesian posterior; this assumption is similar to quadratic approximation of the likelihood function commonly used in maximum likelihood based inference [?]. The principal limitation 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 bounded to be positive (e.g. standard deviation).

Of additional concern for posterior inference is the partial identifiability of observation parameters $p_{t=1}$ and $p_{t=T}$ [?]. This issue means that the estimates of sampling probabilities at the “edges” of the time series cannot fully be estimated because there are no known “gaps” in species occurrence histories that are guaranteed to be filled. Instead, the values of the first and final columns of the “true” presence-absence matrix z for those observations that do not already have presences in the observed presence-absence matrix y cannot be estimated [?]. The hierarchical modeling approach used here helps mitigate this problem by pulling the values of $p_{t=1}$ and $p_{t=T}$ towards the overall mean of p [?], and in fact this approach might be more analytically sound than the more ad-hoc approaches that are occasionally used to overcome this hurdle [?]. Additionally, because $p_{t=1}$ and $p_{t=T}$ are only partially identifiable, estimates of occurrence θ and origination ϕ at $t = 1$ and estimates of θ , ϕ and survival π at $t = T$ may suffer from similar edge effects. Again, the hierarchical modeling approach used here may help correct for this reality by drawing these estimates towards the overall means of those parameters.

After fitting both models (Eqs. ??, ??) using ADVI, model adequacy and quality of fit were assessed using a posterior predictive check [?]. By simulating 100 theoretical data sets from the posterior estimates of the model parameters and the observed covariate information the congruence between predictions made by the model and the observed empirical data can be assessed. These datasets are simulated by starting with the observed states of the presence-absence matrix at $t = 1$; from there, the time series roll forward as stochastic processes with covariate information given from the empirical observations. Importantly, this is fundamentally different from observing the posterior estimates of the “true” presence-absence matrix z . The posterior predictive check used in this study is to compare the observed average number of observations per species to a distribution of simulated averages; if the empirically observed value sits in the middle of the distribution then the model can be considered adequate in reproducing the observed number of occurrences per species.

The ADVI assumption of a purely Gaussian posterior limits the utility and accuracy of the posterior predictive checks because parameter estimates do not reflect the true posterior distribution and are instead just an approximation [?]. Because of this, posterior predictive estimates are themselves only approximate checks of model adequacy. The posterior predictive check that is used in this study focuses on mean occurrence and not to any scale parameters that might be most affected by the ADVI assumptions.

Given parameter estimates, diversity and diversification rates are estimated through posterior predictive simulations. Given the observed presence-absence matrix y , estimates of the true 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 possible occurrence history is estimated (Fig. 1), but now these occurrence histories are generated relative to their estimated probabilities.

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

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

Given estimates of N_t^{stand} for all time points, the estimated number of originations O_t is estimated as

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

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

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