Materials and Methods

Taxon occurrence information

All fossil occurrence information information was downloaded from the Paleobiology Database. Occurrences (PBDB) were restricted to all Mammalia sampled in North America between the 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/2slgeU`.

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 example, species classified in the order Artidodactyla were reclassified as Cetartiodactyla. These re-assignments follow **?**] and were based on CITATIONS. 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 potential and applicability to the study of terrestrial species pools.

The life habit and dietary categories provided through the PBDB where coarsened to increase per ecotype sample size; this coarsening follows the same procedure as **?**]. Additionally, life habit category was further modified to break-up the vague "ground-dwelling" category; 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 context. Species ecotype is defined as the interaction between life habit and diet categories. Ecotype categories with less than 10 species havig ever been that combination were excluded, yielding a total of 18 of 21 possible ecotypes.

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 **?**], 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 CITATIONS, the Neogene Old World Database as in **?**]. 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 and follows CITATION.

All fossil occurrences from 66 to 2 million years ago (Mya) were binned into 31 2 million year (My) bins. This temporal length was chosen because it is approximately the resolution of the North American mammal fossil record.

Environmental and temporal covariates

Modelling species occurrence

Two different models were used in this study: a pure-presence model and a birth-death model. Both models at their core are hidden Markov model with an absorbing state. The difference between these two models is if the probability of a species origination and survival are considered equal or different.

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 ϕ of a species surviving from one time to another, and probability π of a species first appearing CITATION. In this formulation, the probability of a species going extinct is $1 - \pi$. For the pure-presence model $\phi = \pi$, 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 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 assumptions are reflected in the structure of the model Equation 1. The parameters associated with this part of the model are described in Table 1.

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

For the pure-presence model there is only a single probability dealing with the presence of a species θ. This probability was modeled as multi-level logistic

Parameter	dimensions	explanation
y	$N \times T$	
z	$N \times T$	
p	T	probability of observing a species that is present at time t
α_0	1	average log-odds of p
α_1	1	change in average log-odds of p per change mass
r	T	difference from α_0 associated with time t
σ	1	standard deviation of r

Table 1: Observation parameters

regression with both species-level and group-level covariates CITATION. The parameters associated with pure-presence model are presented in Table 2 and the sampling statement in Equation 2.

The species-level of the model (Eq. 2) is a varying-intercept model where the intercept varies by 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 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 choice of priors for these regression coefficients.

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

Following recommendations from CITATION, CITATION, CITATION all parameters not modeled elsewhere were given weakly informative priors. Weakly informative means that 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 zero; this has the effect of tempering our estimates and help prevent overfitting the model to the data CITATION.

Table 2: Parameters for the model of presence in the pure-presence model

Parameter	dimensions	explanation
z	$N \times T$	presence of species at any time
θ	$N \times T - 1$	probability of $z = 1$
a	$T - 1 \times D$	ecotype-varying intercept; mean value of log-odds of θ
b_1	1	effect of species mass on log-odds of θ
b_2	1	effect of species mass, squared, on log-odds of θ
γ	$U \times D$	matrix of group-level regression coefficients
Σ	$D \times D$	covariance matrix of a
Ω	$D \times D$	correlation matrix of a
τ	D	vector of standard deviations for each ecotype a_d

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

In the birth-death model, $\phi \neq \pi$ and so each of these probabilities are modeled separately but in a similar manner to how θ is modeled in the pure-presence model (Eq. 2).

Posterior inference and model adequacy