Death and taxa: time-invariant differences in mammal species duration

Peter D Smits *

*University of Chicago, Chicago, Illinois, United States of America, 60637

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Determining which biological traits influence differences in extinction risk is vital for understanding the differential diversification of life and for making predictions about species' vulnerability to anthropogenic impacts. Here I present a hierarchical Bayesian survival model of North American Cenozoic mammal species durations as predicted by species-level ecological factors, time of origination, and phylogenetic relationships. I find support for the "survival of the unspecialized" as a time-invariant generalization of trait-based extinction risk. Furthermore, I find that phylogenetic and temporal effects are both substantial factors associated with differences in species durations. The current biodiversity crisis is partially incongruous with previous patterns. This incongruity implies that the either the fossil record is a poor predictor of risk in the modern world, or that the current biodiversity crisis has transitioned into a state comparable to the great mass extinctions in Earth history.

paleobiology | macroevolution | macroecology | extinction

Why extinction risk varies among species remains one of the most fundamental questions in paleobiology and conservation biology [1, 2, 3, 4, 5]. To address this issue, I test for associations between extinction risk and multiple species-level traits during times of background extinction and in the modern world; which traits have time-invariant effects on species duration; and whether extinction is age-independent. I approach these questions together by using a model of species duration whose parameter estimates act as direct tests of these questions. Cenozoic mammals are an ideal focus for this study because their fossil record is well sampled and well resolved both temporally and spatially, and because individual species ecology and taxonomic position are generally understood [6, 7, 8, 4, 1, 9, 10].

The species-level traits studied here are bioprovince occupancy, body mass, and both dietary and locomotor categories. These traits are related to aspects of a species' adaptive zone such as population density, expected range size, potential prey, and dispersal ability [8, 11]. It is expected that species with larger geographic ranges have lower extinction rates than species with smaller geographic ranges [12, 13]; however, how traits more directly related to species–environment interactions play a role is more nebulous.

Body size is a complex trait related to many life history characteristics. There are three general hypotheses of how body size may effect extinction risk: 1) positive effect where an increase in body size causes an increase in extinction risk, potentially due to associated decrease in reproductive rate or similar [7, 14]; 2) negative effect where an increase in body size causes a decrease in extinction risk because of an expected positive relationship between body size and geographic range; and 3) no effect of body size on extinction risk [9].

The strongest expectation of the effect of dietary category on extinction risk is that omnivores will have the lowest extinction risk of all species. This hypothesis is based on the long standing "survival of the unspecialized" hypothesis where more generalist species (e.g. omnivores) have greater survival than specialist species (e.g. carnivores/herbivores) [15, 1]. It has also been observed that both carnivores and herbivores have greater diversification rates than omnivores, with herbivores diversifying faster than carnivores [16]. How this result translates into expectations of differences in extinction risk is currently unknown. In modern taxa, higher trophic levels (e.g. carnivores versus herbivores) have been associated with an increase in extinction risk, most likely because of human extermination of top predators [14, 17].

Similarly, there are few simple expectations of how locomotor category may effect extinction risk. During the Cenozoic there was a shift at the Paleogene/Neogene boundary from predominately closed to predominately open environments [18, 19]. Based on this observation, a simple prediction is that arboreal taxa will have the greatest extinction risk of all, with both scansorial and ground dwelling taxa having lower extinction risks.

Time-invariant factors are those that have a constant directional effect even if the magnitude varies. Because change in the magnitude of extinction risk is not necessarily the best indicator of a shift from background to mass extinction [20], it is better to look for changes in either the direction of selection, the loss of a selective pressure, or the appearance of novel selective pressures [12].

I use a hierarchical Bayesian survival model of species duration as predicted by the covariates of interest along with species' temporal and phylogenetic context. Species duration, in 2 My bins, was modeled as being drawn from either an exponential or Weibull distribution parameterized as a hierarchical regression [21]. The exponential is a special case of the Weibull where the shape parameter, α, is 1 which corresponds to the Law of Constant Extinction, which states that extinction is age-independent [2]. Origination cohorts were modeled as exchangeable draws from a common normal distribution. Phylogenetic effect was modeled assuming species duration may have evolved via Brownian motion [22]. Extended explanation of the model, choice of priors, parameter estimation, and posterior predictive check results are provided in the supplementary online text. The results from the Weibull model are detailed here because this model has a better fit to the data the exponential (Fig. 1, S1, S2).

Results

Species with greater bioprovince occupancy are found to be associated with lower extinction risk than taxa with smaller bioprovince occupancy (Mean $\beta_{occupancy} = -0.53$, Std $= 0.08$). This is consistent with prior expectations. Body size has nearly

Reserved for Publication Footnotes

zero association with expected duration (Mean $\beta_{size} = -0.05$, Std = 0.05), a similar result to some previous studies [9]. However, previous studies were performed at the generic-level and were unable to determine how body size may effect species level extinction, as the effect of either extinction or speciation cannot be distinguished [9, 7].

Some clear patterns emerge from the pairwise differences in effect of each dietary category on expected duration (Fig. 2). Consistent with expectations from the "survival of the unspecialized" hypothesis [15, 1], omnivory appears to be associated with the lowest expected extinction risk. Carnivory is associated a greater expected duration than either herbivory or insectivory, but a greater expected extinction risk then omnivory. Finally, herbivory and insectivory have approximately equal effects on expected duration. Given previous results, these results imply that carnivores have a greater origination rate than omnivores [16]. These results also imply that herbivores, which have the greatest extinction risk, must also have a very high origination rate in order to have the greatest diversification rate among these three categories [16].

For locomotor category, both scansoriality and ground dwelling life habitat are associated with a greater expected duration than arboreality (Fig. 2). Scansorial and ground dwelling life habits also have approximately equal expected effects on extinction risk. This is consistent with the expectation that arboreality will confer greater extinction risk due to the loss of associated environment with the shift from open to closed habitat at the Paleogene/Neogene boundary [18].

Of the three sources of variance present in the model, individual species variance accounts for approximately 80% of the observed, unmodeled variance (Fig. 4). Both cohort and phylogenetic effects account for the other 20% of the observed variance. This result means that extinction risk has both temporal and phylogenetic aspects, as both contribute greater than 0% of the observed variability in the data [22].

The estimates for the individual cohort effects show a weak pattern of greater extinction risk in older Cenozoic cohorts compared to younger cohorts (Fig. 3). This potential slowdown in extinction risk is consistent with previous analyses of marine invertebrates [23, 24] and mammals [25, 26]. There are two prevailing hypotheses as to the cause of this slowdown: 1) extinction risk is constant within, but varies between, clades so over time clades with low extinction rates increases in proportion of total diversity thus bringing down expected extinction risk; or 2) over time taxa increase in mean fitness and thus decrease in expected extinction risk [23]. The observed decrease in extinction risk with age, along with the variance partitioning results (Fig. 4) are consistent with both of these hypotheses with neither being more "important" than the other.

Interestingly, the shift from older cohorts with a higher extinction risk to younger cohorts with lower extinction risk is approximately at the Paleogene–Neogene boundary. Given the association with arboreality (a heritable trait) and increased extinction risk (Fig. 2), the decrease in expected extinction risk over time might relate to the preferential loss of arboreal taxa over the Cenozoic. However, because the model used here does not allow for change in time-invariant effects, I cannot identify whether this boundary is associated with a shift in the direction or magnitude of the expected effect of arboreality on extinction risk.

Discussion

When these results are compared to factors contributing to increased extinction risk in extant mammals, there are some in congruencies. As expected, large range size is always associated with lower extinction risk in the modern world [27, 28, 14, 17]. While my analysis found body size to have almost no time-invariant effect on extinction risk, this is not necessarily the case in extant mammals where increased body size is associated with increased extinction risk [14, 17]. However, this pattern is partially clade dependent [27]. As stated earlier, higher trophic levels have been found to be associated with greater extinction risk in extant mammals [17, 14]. In contrast, I found that omnivores and carnivores have a lower expected extinction risk than either insectivores or herbivores (Fig. 2). Finally, phylogeny has been found to be a good predictor of differences in extinction risk in extant mammals as certain clades are at much higher risks than others [28]. This effect seems much greater in the Recent than for the whole Cenozoic, implying that current extinction risk is more phylogenetically concentrated than in previous mammal extinctions.

Whether these incongruities are within the standard range of time-variant effects is unknown, though my comparisons do imply that current processes are different from those studied here. These incongruities imply that the fossil record may not provide a guide for conservation biology. However, this is not a model of what makes taxa vulnerable during mass extinctions and that may account for these incongruities, assuming mass extinctions are qualitatively different than background extinction [12]. These results would also be inapplicable if the current biodiversity crisis is qualitatively different from either background or mass extinction as preserved in the fossil record.

By modeling how different ecologies and historical factors effect a species' expected extinction risk, it is possible to better understand what processes may have driven the resulting pattern of selection (i.e. diversity). This analysis finds broad support for the "survival of the unspecialized" hypothesis [1, 15] as a time-invariant generalization about extinction risk. I also find that there are substantial effects of both cohort and phylogeny on extinction risk, which supports the idea that the decrease in extinction risk [23] over time has both temporal and phylogenetic components. Additionally, I found evidence of increasing extinction risk with species age, the cause of which is unknown. These results, when compared to current mammal extinction risk factors, support the hypothesis that the current biodiversity crisis is qualitatively different from "normal" extinction, possibly because the current biodiversity crisis is a mass extinction event.

Materials and Methods

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

Species dietary and locomotor category assignments were done using the assignments in the PBDB, which were reassigned into coarser categories (Supplementary Table S1). This was done to improve interpretability, increase sample size per category, and make results comparable to previous studies [11, 16].

All individual fossil occurrences were assigned to 2 My bins ranging through the entire Cenozoic. Taxon duration was measured as the number of 2 My bins from the first occurrence to the last occurrence, inclusive. This bin size was chosen because it approximately reflects the resolution of the North American Cenozoic mammal fossil record [6, 26, 10]. Species originating in the youngest cohort, 0-2 My, were excluded from analysis because every species duration would be both left and right censored, which is illogical.

Species body size estimates in grams were sourced from a large selection of primary literature and database compilations. Databases used include the PBDB, PanTHERIA [29], and the Neogene Old World Mammal database (NOW;

http://www.helsinki.fi/science/now/). Major sources of additional compiled body size estimates include [30, 31, 32, 33, 34, 9]. These were then supplemented with an additional literature search to try and fill in the remaining gaps. In many cases, species body mass was estimated using various published regression equations based on tooth or skull measurements (Supplementary Table S2). If multiple specimens were measured, I used the mean of specimen measures as the species mean. See Supplementary Table S3 for a complete list of individual measures and sources.

Biogeographic network. Species geographic extent was measured as the mean of the relative number of bioprovinces occupied by a species for each 2 My bin the species was present. Bioprovinces were identified using a network-theoretic approach that has previously been applied to paleontological data [35, 36]. This approach relies on defining a biogeographic bipartite network of taxa and localities. In this study, taxa were defined as species and localities were grid cells from a regular lattice on a global equal-area cylinder map projection. The regular lattice was defined as a 70 x 34 global grid where each cell corresponds to approximately 250000 km^2. This network is considered bipartite because taxa are connected to localities based on their occurrence but taxa are not connected to taxa nor are localities connected to localities.

A biogeographic network was constructed for each of the 2 My bins used in this study. Emergent bioprovinces were then identified using the map equation [37, 38] as has been done before [35, 39, 36]. These bioprovinces correspond to taxa and localities that are more interconnected with each other than with other nodes.

The map projection and regular lattice were made using shape files from http://www.naturalearthdata.com/ and the `raster` package for R [40]. The map equation and other network related analysis was done using the `igraph` package for R [41].

Supertree.

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

Survival model. Presented here is the model development process used to formulate the two survival models used in this study.

First, define y as a vector of length n where the ith element is the duration of species i, where $i = 1, \cdots, n$.

The simplest survival model where durations are assumed to follow an exponential distribution with a single "rate" or inverse-scale parameter λ [42]. This is written out as

$$p(y|\lambda) = \lambda \exp(-\lambda y)$$
$$y \sim \text{Exp}(\lambda). \quad \textbf{[1]}$$

The exponential distribution corresponds to situations where extinction risk is independent of age. To understand this, we need to define two functions: the survival function $S(t)$ and the hazard function $h(t)$.

$S(t)$ corresponds to the probability that a species having existed for t 2 My bins will not have gone extinct while $h(t)$ corresponds to the instantaneous extinction rate for some taxon age t [42]. For an exponential model, $S(t)$ is defined

$$S(t) = \exp(-\lambda t) \quad \textbf{[2]}$$

and $h(t)$ is defined

$$h(t) = \lambda \quad \textbf{[3]}$$

The choice of the exponential distribution corresponds directly to the Law of Constant Extinction [2] as the right side of Eq. 3 does not depend on species age t.

The current sampling statement (Eq. 1) assumes that all species share the same rate parameter with no variation. To allow for variation in λ associated with relevant covariate information like species body size, λ is reparameterized as $\lambda_i = \exp(\sum \beta^T \mathbf{X}_i)$ with i indexing a given observation and its covariates, β is a vector of regression coefficients, and $\mathbf{X}$ is a matrix of covariates. This is a standard regression formulation, where one column of $\mathbf{X}$ is all 1-s and its corresponding β coefficient is the intercept. This approach is essentially a generalized linear model (GLM) approach where instead of normally distributed errors there are exponentially distributed errors [42].

To relax the assumption of age-independent extinction of the Law of Constant Extinction we substitute the Weibull distribution for the exponential [42]. The Weibull distribution has a shape parameter α and scale parameter σ. Conceptually, σ is the inverse of λ. α modifies the impact of taxon age on extinction risk. When $\alpha > 1$ then $h(t)$ is a monotonically increasing function, but when $\alpha < 1$ then $h(t)$ is a monotonically decreasing function. When $\alpha = 1$ then the Weibull distribution is equivalent to the exponential.

The Weibull distribution and sampling statement were defined

$$p(y|\alpha, \sigma) = \frac{\alpha}{\sigma}\left(\frac{y}{\sigma}\right)^{\alpha-1} \exp\left(-\left(\frac{y}{\sigma}\right)^{\alpha}\right)$$
$$y \sim \text{Weibull}(\alpha, \sigma). \quad \textbf{[4]}$$

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

$$S(t) = \exp\left(-\left(\frac{t}{\sigma}\right)^{\alpha}\right) \quad \textbf{[5]}$$

$$h(t) = \frac{\alpha}{\sigma}\left(\frac{t}{\sigma}\right)^{\alpha-1}. \quad \textbf{[6]}$$

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

$$\sigma_i = \exp\left(\frac{-(\beta_0}{\alpha}\right) \quad \textbf{[7]}$$

where β_0 is the intercept term.

A survival model was fit in a Bayesian context where species duration were assumed to be drawn from either an exponential or Wiebull distribution (Eq. 4) with shape α and scale σ parameters. α was assumed constant, which is standard practice in survival analysis [42]. α was given a weakly informative half-Cauchy (C^+) prior. σ was reparameterized as an exponentiated regression model (Eq. 7). This was further expanded (Eq. 8) to allow for two hierarchical factors as discussed below. This is written

$$\sigma_i = \exp\left(\frac{-(h_i + \eta_{j[i]} + \sum \beta^T \mathbf{X}_i)}{\alpha}\right) \quad \textbf{[8]}$$

where equivalent statement for the exponential distribution is defined

$$\lambda_i = \exp\left(h_i + \eta_{j[i]} + \sum \beta^T \mathbf{X}_i)\right). \quad \textbf{[9]}$$

$\mathbf{X}$ is an $n \times K$ matrix of species-level covariates. Three of the covariates of interest are the logit of mean relative occupancy, and the logarithm of body size (g). The discrete covariate index variables of dietary and locomotor category were transformed into $n \times (k-1)$ matrices where each column is an indicator variable (0/1) for that species's category, k being the number of categories of the index variable (3 and 4, respectively). Only $k-1$ indicator variables are necessary as the intercept takes on the remaining value. Finally, a vector of 1-s was included in the matrix $\mathbf{X}$ whose corresponding β coefficient is the intercept, making K equal eight.

β is the vector of regression coefficients. The intercept term was given a weak normal prior, $\beta_0 \sim \mathcal{N}(0, 10)$ while all of these other coefficients were slightly more informative priors, e.g. $\beta_{mass} \sim \mathcal{N}(0, 5)$. These priors were chosen because it is expected that the effect size of each variable on duration will be small, as is generally the case with binary covariates [43].

Regression coefficients are not directly comparable without first standardizing the input variables to have equal standard deviations. This is accomplished by subtracting the mean of the covariate from all values and then dividing by the standard deviation, resulting in a variable with mean of zero and a standard deviation of one. This linear transform greatly improves the interpretability of the coefficients as expected change in mean duration given a difference of one standard deviation in the covariate [44]. Additionally, this makes the intercept directly interpretable as the estimate of mean (transformed) σ (Eq. 7). However, because the expected standard deviation for a random binary variable is 0.5, in order to make comparisons between the binary and continuous variables, the continuous inputs must instead be divided by twice their standard deviation [45].

The two hierarchical effects of interest in this study are origination cohort and shared evolutionary history, or phylogeny. Hierarchical modeling can be considered an intermediate between complete and no pooling of groups [43], where complete pooling is when the differences between groups are ignored and no pooling is where different groups are analyzed separately. By allowing for partial pooling, we are modeling the appropriate compromise between these two extremes, allowing for better and potentially more informative overall inference. This is done by having all of the groups share the same normal prior with mean 0 and a scale parameter estimated from the data, which then acts as an indicator of the amount of pooling. A scale of 0 and ∞ indicate complete and no pooling, respectively. The choice of mean 0 allows for the individual group estimates to be interpreted as deviations from the intercept. Hierarchical modeling is analogous to mixed-effects modeling [43].

Origination cohort is defined as the group of species which all originated during the same 2 My temporal bin. Because the most recent temporal bin, 0-2 My, was excluded, there are 32 different cohorts. The effect of origination cohort j was modeled with each group being a sample from a common cohort effect, η, which was considered normally distributed with mean 0, and standard deviation σ_c. The value of σ_c was then estimated from the data itself, corresponding to the amount of pooling in the individual estimates of η_j. This approach is a conceptual and statistical unification between dynamic and cohort survival analysis in paleontology [46, 47, 48, 49, 50], with σ_c acting as a measure of compromise between these two end members.

$$\eta_j \sim \mathcal{N}(0, \sigma_c)$$
$$\sigma_c \sim \mathrm{C}^+(0, 2.5)$$

The choice of the half-Cauchy prior on σ_c follows [51].

The impact of shared evolutionary history, or phylogeny, was modeling as an individual effect where each observation, i, is modeled as a multivariate normal, h, where the covariance matrix Σ is known up to a constant, σ_p^2 [52, 22]. This is written

$$h \sim \text{multivariate}\, \mathcal{N}(0, \mathbf{\Sigma})$$
$$\mathbf{\Sigma} = \sigma_p^2 \mathbf{V}_{phy}$$
$$\sigma_p \sim \mathrm{C}^+(0, 2.5).$$

$\mathbf{V}_{phy}$ is the phylogenetic covariance matrix defined as an $n \times n$ matrix where the diagonal elements are the distance from root to tip, in branch length, for each observation and the off-diagonal elements are the amount of shared history, measured in branch length, between observations i and j. σ_p was given a weakly informative half-Cauchy hyperprior. Note that because the phylogeny used here is primarily based on taxonomy, estimates of σ_p represent minimum estimates [52, 22]. Improved phylogenetic estimates of all fossil Cenozoic mammals would greatly improve this estimate.

An important part of survival analysis is the inclusion of censored observations where the failure time has not been observed [53, 42]. The most common censored observation is right censored, where the point of extinction had not yet been observed in the period of study, such as taxa that are still present in the most recent time bin (0-2 My). Left censored observations, on the other hand, correspond to observations that went extinct any time between 0 and some known point. In order to account for the minimum resolution of the fossil record encountered here, taxa that occurred in only a single time bin were left censored.

Censored data are modeled using the survival function of the distribution, $S(t)$, defined earlier for the Weibull distribution (Eq. 5) with σ defined as above (Eq. 8). $S(t)$ is the probability that an observation will survive longer than a given time t. The likelihood of uncensored observations is evaluated as normal using Equation 4 while right censored observations are evaluated at $S(t)$ and left censored observations are evaluated at $1 - S(t)$. Note, $1 - S(t)$ is equivalent to the cumulative distribution function and $S(t)$ is equivalent to the complementary cumulative distribution function [21].

The full likelihood for both uncensored and both right and left censored observations is written

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

where C is the set of uncensored observations, R is the set of right censored observations, and L is the set of left censored observations.

Estimation. Parameter posteriors were approximated using a Markov-chain Monte Carlo (MCMC) routine implemented in the Stan programming language [54]. Stan implements a Hamiltonian Monte Carlo using a No-U-Turn sampler [55]. Posterior approximation was done using four parallel MCMC chains run for 30000 steps, thinned to every thirtieth sample, split evenly between warm-up and sampling. Convergence was evaluated using the scale reduction factor, $\hat{R}$. Values of $\hat{R}$ close to 1, or less than or equal to 1.1, indicate approximate convergence. Convergence means that the chains are approximately stationary and the samples are well mixed [21].

In order to speed up the posterior approximation, a custom multivariate normal sampler was used to estimate the unknown constant term in the covariance matrix. This is necessary because inverting and solving the complete covariance matrix on every iteration is a memory intense procedure. The custom sampler limits the necessary number of operations and matrix inversions per iteration.

Model evaluation. The most basic assessment of model fit is that simulated data generated using the fitted model should be similar to the observed. This is the idea behind posterior predictive checks. Using the covariates from each of the observed durations, and randomly drawn parameter estimates from their marginal posteriors, a simulated data set y^{rep} was generated. This process was repeated 1000 times and the distribution of y^{rep} was compared with the observed [21].

An example posterior predictive check used in this study is a graphical comparison between the empirically estimated survival curve and the survival curves estimated from 1000 simulation sets. Survival curves were estimated using the Kaplan-Meier approach, which is a non-parametric estimate of $S(t)$ [42]. Other posterior predictive checks used here include comparison of the mean and quantiles of the observed durations to the distributions of the same quantities from the simulations, and inspection of the deviance residuals. For an extended explanation of how deviance residuals for a Weibull model are calculated, please see the Supplementary information.

The exponential and Weibull models were compared for out-of-sample predictive accuracy using the widely-applicable information criterion (WAIC) [56]. Because the Weibull model reduces to the exponential model when $\alpha = 0$, our interest is not in choosing between these models. Instead comparison of WAIC values is useful for better understanding the effect of model complexity on out-of-sample predictive accuracy. An explanation of how WAIC is calculated is presented in the Supplementary information following the "WAIC 2" formulation recommended by [21].

There are three different variance components in this model: sample σ_y^2, cohort σ_c^2, and phylogenetic σ_p^2. The sample variance, σ_y^2, is similar to the residual variance from a normal linear regression. Partitioning the variance between these sources allows the relative amount of unexplained variance of the sample to be compared. However, the Weibull based model used here (Eq. 4) does not include an estimate of the sample variance, σ_y^2. Partitioning the variance between these three components was approximated via a simulation approach modified from [57]. For explination of this method, please see Supplementary information.

I used variance partitioning coefficients (VPC) to estimate the relative importance of the different variance components [43]. Phylogenetic heritability, h_p^2 [52, 22], is identical to the VPC of the phylogenetic effect. Additionally, because phylogenetic effect was estimated using a principally taxonomy based tree the estimates derived here can be considered minimum estimates of the phylogenetic effect.

ACKNOWLEDGMENTS. P.D.S would like to thank M. Foote, K. Angielczyk, R. Ree, P.D. Polly for discussion; N. Pierrehumbert, E. Sander, L. Southcott for draft comments; and J. Alroy and the Fossilworks/Paleobiology Database for data accumulation, management, and availability. This is Paleobiology Database publication number XXX.

1. Simpson GG (1944) Tempo and Mode in Evolution (Columbia University Press, New York).
2. Van Valen L (1973) A new evolutionary law. Evolutionary Theory 1:1–30.
3. Raup DM (1994) The role of extinction in evolution. Proceedings of the National Academy of Sciences 91:6758–6763.
4. Quental TB, Marshall CR (2013) How the Red Queen Drives Terrestrial Mammals to Extinction. Science 341:290–292.
5. Wagner PJ, Estabrook GF (2014) Trait-based diversification shifts reflect differential extinction among fossil taxa. Proceedings of the National Academy of Sciences 111:16419–16424.
6. Alroy J (2009) in Speciation and patterns of diversity, eds Butlin RK, Bridle JR, Schluter D (Cambridge University Press, Cambridge), pp 302–323.
7. Liow LH, et al. (2008) Higher origination and extinction rates in larger mammals. Proceedings of the National Academy of Sciences 105:6097.
8. Smith FA, Brown J, Haskell J, Lyons S (2004) Similarity of mammalian body size across the taxonomic hierarchy and across space and .... The American Naturalist 163:672–691.
9. Tomiya S (2013) Body Size and Extinction Risk in Terrestrial Mammals Above the Species Level. The American Naturalist 182:196–214.
10. Marcot JD (2014) The fossil record and macroevolutionary history of North American ungulate ungulate mammals: standardizing variation in intensity and geography of sampling. Paleobiology 40:237–254.
11. Jernvall J, Fortelius M (2004) Maintenance of trophic structure in fossil mammal communities: site occupancy and taxon resilience. American Naturalist 164:614–624.
12. Jablonski D (1986) Background and mass extincitons: the alternation of macroevolutionary regimes. Science 231:129–133.
13. Roy K, Hunt G, Jablonski D, Krug AZ, Valentine JW (2009) A macroevolutionary perspective on species range limits. Proceedings of the Royal Society B: Biological Sciences 276:1485–1493.
14. Liow LH, Fortelius M, Lintulaakso K, Mannila H, Stenseth NC (2009) Lower Extinction Risk in SleeporHide Mammals. The American Naturalist 173:264–272.
15. Liow LH (2004) A test of Simpson's "rule of the survival of the relatively unspecialized" using fossil crinoids. The American naturalist 164:431–43.
16. Price SA, Hopkins SSB, Smith KK, Roth VL (2012) Tempo of trophic evolution and its impact on mammalian diversification. Proceedings of the National Academy of Sciences 109:7008–12.
17. Purvis A, Gittleman JL, Cowlishaw G, Mace GM (2000) Predicting extinction risk in declining species. Proceedings of the Royal Society B: Biological Sciences 267:1947–52.
18. Blois JL, Hadly EA (2009) Mammalian Response to Cenozoic Climatic Change. Annual Review of Earth and Planetary Sciences 37:181–208.
19. Janis CM (1993) Tertiary mammal evolution in the context of changing climates, vegetation, and tectonic events. Annual Review of Ecology and Systematics 24:467–500.
20. Wang SC (2003) On the continuity of background and mass extinction. Paleobiology 29:455–467.
21. Gelman A, et al. (2013) Bayesian data analysis (Chapman and Hall, Boca Raton, FL), 3 edition, p 675.
22. Housworth EA, Martins P, Lynch M (2004) The Phylogenetic Mixed Model. The American Naturalist 163:84–96.
23. Raup DM, Sepkoski JJ (1982) Mass Extinctions in the Marine Fossil Record. Science 215:1501–1503.
24. Foote M (2003) Origination and extinction through the Phanerozoic: a new approach. Journal of Geology 111:125–148.
25. Alroy J (2010) in Quantitative Methods in Paleobiology, eds Alroy J, Hunt G (The Paleontological Society), pp 55–80.
26. Alroy J, Koch PL, Zachos JC (2000) Global climate change and North American mammalian evolution. Paleobiology 26:259–288.
27. Fritz SA, Bininda-Emonds ORP, Purvis A (2009) Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics. Ecology Letters 12:538–49.
28. Fritz SA, Purvis A (2010) Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. Conservation Biology 24:1042–51.
29. Jones KE, et al. (2009) PanTHERIA : a species-level database of life history , ecology , and geography of extant and recently extinct mammals. Ecology 90:2648.
30. Brook BW, Bowman DMJS (2004) The uncertain blitzkrieg of Pleistocene megafauna. Journal of Biogeography 31:517–523.
31. Freudenthal M, Martín-suárez E (2013) Estimating body mass of fossil rodents. Scripta Geologica 145:1–130.
32. McKenna RT (2011) Masters (Portland State University).
33. Raia P, Carotenuto F, Passaro F, Fulgione D, Fortelius M (2012) Ecological specialization in fossil mammals explains Cope's rule. The American naturalist 179:328–37.
34. Smith FA, et al. (2004) Similarity of Mammalian Body Size across the Taxonomic Hierarchy and across Space and Time. The American Naturalist 163:672–691.
35. Sidor CA, et al. (2013) Provincialization of terrestrial faunas following the end-Permian mass extinction. Proceedings of the National Academy of Sciences 110:8129–33.
36. Vilhena DA, et al. (2013) Bivalve network reveals latitudinal selectivity gradient at the end-Cretaceous mass extinction. Scientific Reports 3:1790.
37. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105:1118–23.
38. Rosvall M, Axelsson D, Bergstrom C (2009) The map equation. The European Physical Journal Special Topics 178:13–24.
39. Vilhena DA (2013) Ph.D. thesis (University of Washington).
40. Hijmans RJ (2015) raster: Geographic data analysis and modeling R package version 2.3-24.
41. Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Systems:1695.
42. Klein JP, Moeschberger ML (2003) Survival Analysis: Techniques for Censored and Truncated Data (Springer, New York), 2nd edition.
43. Gelman A, Hill J (2007) Data Analysis using Regression and Multilevel/Hierarchical Models (Cambridge University Press, New York, NY), p 625.
44. Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution 1:103–113.
45. Gelman A (2008) Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine pp 2865–2873.
46. Foote M (1988) Survivorship analysis of Cambrian and Ordovician Trilobites. Paleobiology 14:258–271.
47. Raup DM (1978) Cohort Analysis of generic survivorship. Paleobiology 4:1–15.
48. Raup DM (1975) Taxonomic survivorship curves and Van Valen's Law. Paleobiology 1:82–96.
49. Van Valen L (1979) Taxonomic survivorship curves. Evolutionary Theory 4:129–142.
50. Baumiller TK (1993) Survivorship analysis of Paleozoic Crinoidea: effect of filter morphology on evolutionary rates. Paleobiology 19:304–321.
51. Gelman A (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1:515–533.
52. Lynch M (1991) Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.
53. Ibrahim JG, Chen MH, Sinha D (2001) Bayesian Survival Analysis (Springer, New York), p 480.
54. Stan Development Team (2014) Stan: A c++ library for probability and sampling, version 2.5.0.
55. Hoffman MD, Gelman A (2011) The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. arXiv 1111.
56. Watanabe S (2010) Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. Journal of Machine Learning Research 11:3571–3594.
57. Goldstein H, Browne W, Rasbash J (2002) Partitioning variation in multilevel models. Understanding Statistics 1:1–12.

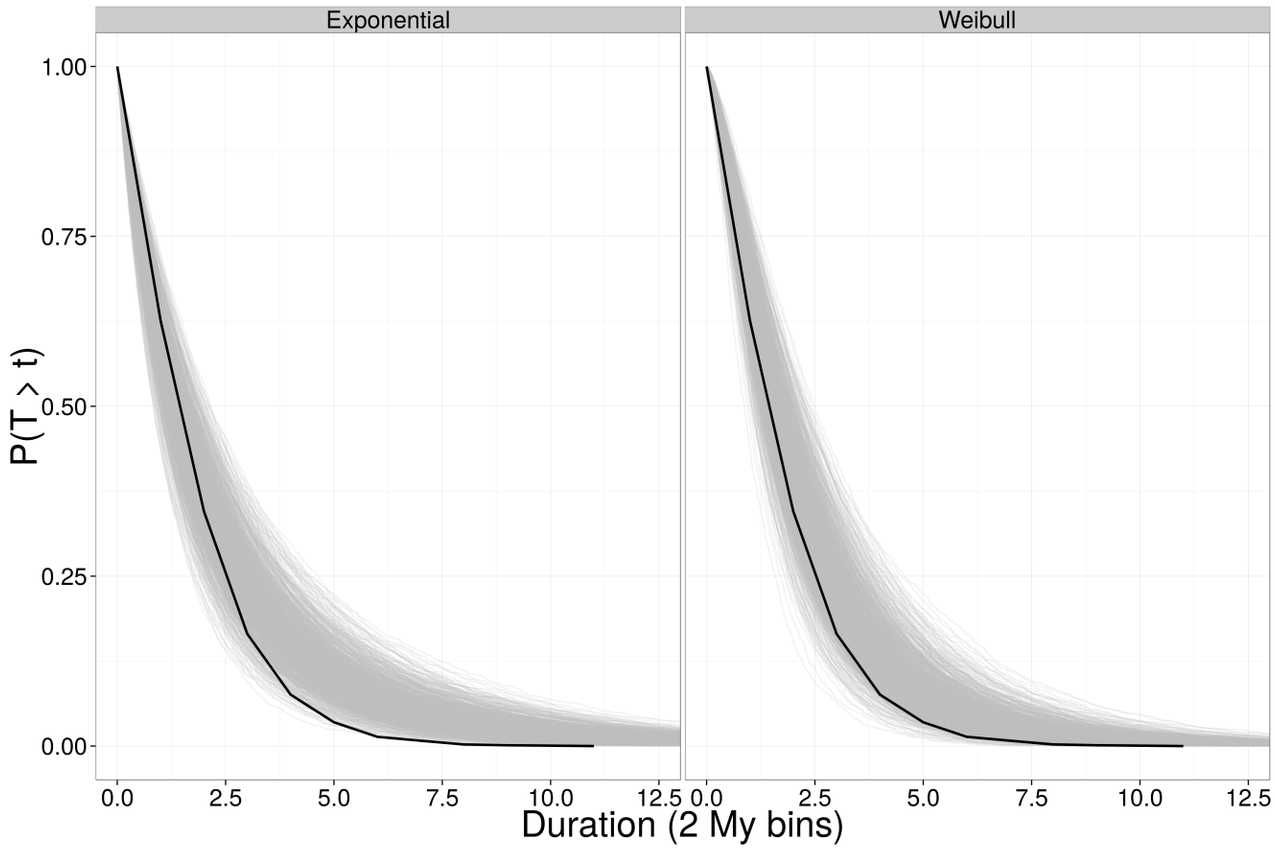
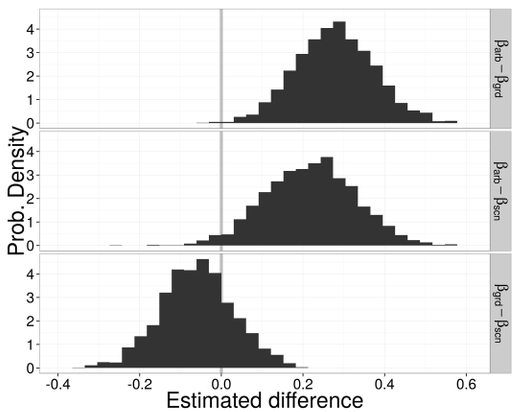


Fig. 1: Comparison of estimates of the empirical survival function (black) and survival functions from 1000 simulated data sets using the fitted model (dark grey). The survival function is the probability that a species with duration t will not have gone extinct. Simulated data sets were generated by drawing parameter values randomly from their estimated posteriors and using the observed covariate information to estimate durations for all the observed species. On the left are the results from the exponential survival model, while on the right are the results the Weibull model.

(a)



(b)

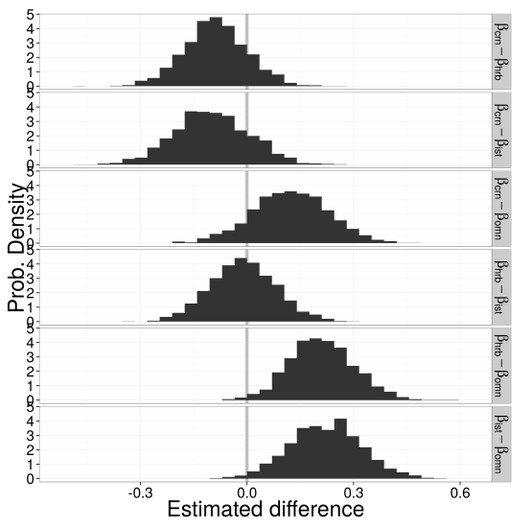


Fig. 2: Pairwise differences in effect of the locomotor (**A**) and dietary categories (**B**) on expected duration from 1000 samples from the posterior distribution. Comparisons of locomotor categories, from top to bottom (**A**), are: arboreal versus ground dwelling, arboreal versus scansorial, and ground dwelling versus scansorial. For dietary category, from top to bottom (**B**): carnivore versus herbivore, carnivore versus insectivore, carnivore versus omnivore, herbivore versus insectivore, herbivore versus omnivore, and insectivore versus omnivore. Values to the left indicate that the first category is expected to have a greater duration than the second, while values to the right indicate that the first category is expected to have a shorter duration.

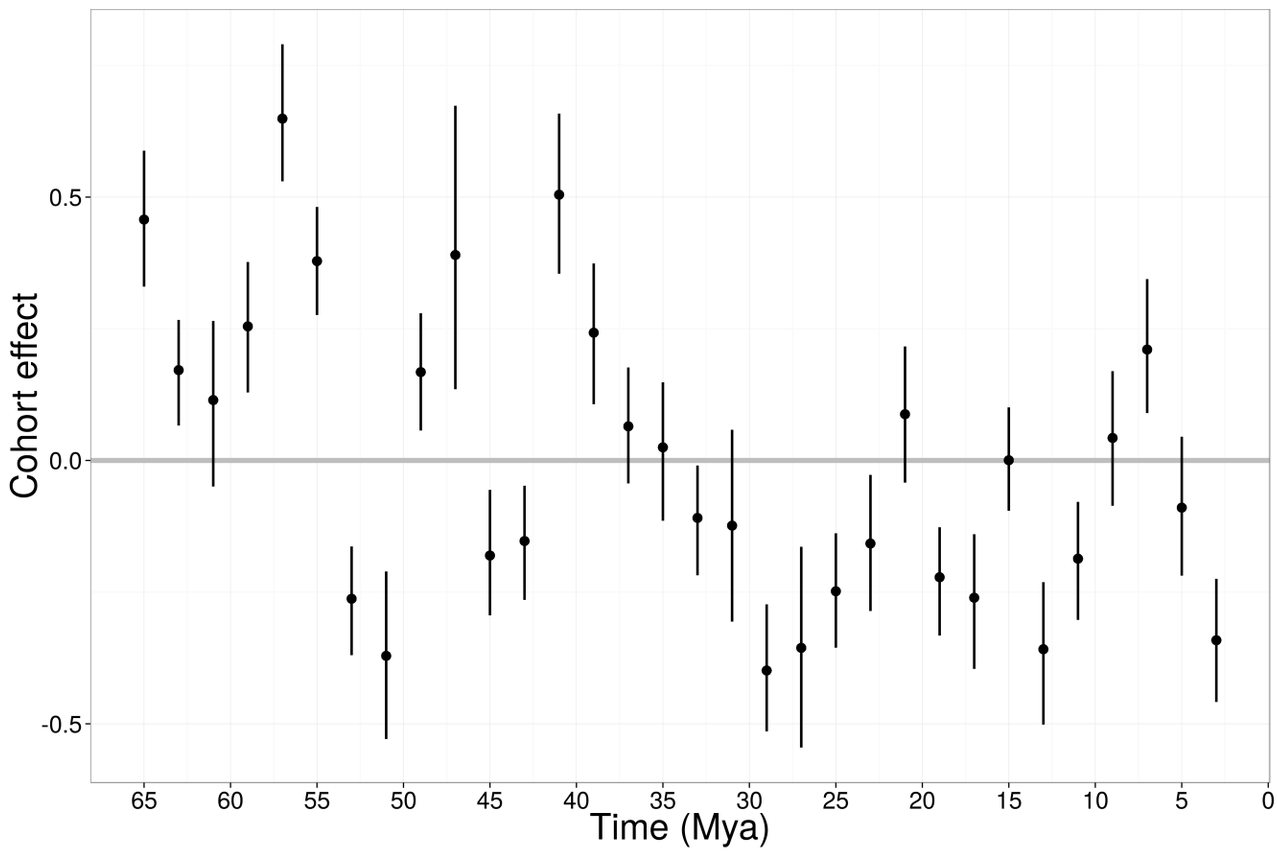


Fig. 3: Summaries of posterior estimates of individual cohort effect depicted as medians and 80% credible intervals. High values correspond to shorter species durations while lower values correspond to greater species durations compared to the mean duration. Lines are placed at the middle of the 2 My origination cohorts.

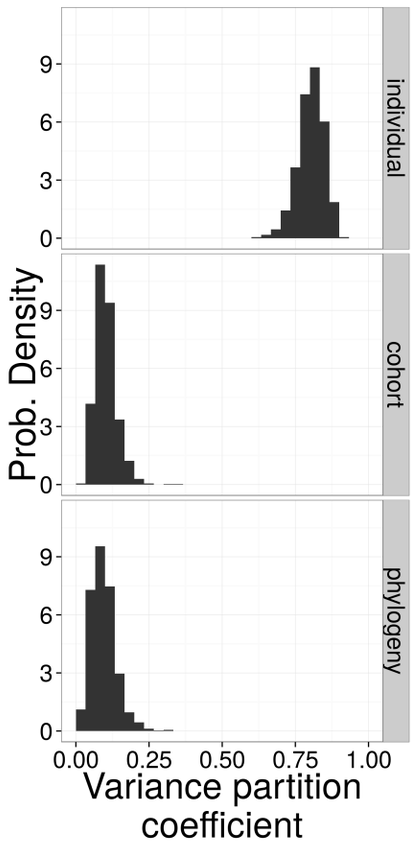


Fig. 4: Estimates of the variance partitioning coefficients for the three different sources of variance: species, cohort, and phylogeny. Higher values correspond to greater contribution to total observed variance. Each of the estimates is a distribution of 1000 approximating simulations due to the model's non-normally distributed errors.

Table 1: Marginal posterior estimates for the praameters of interested based on 1000 posterior samples. The intercept can be interpreted as the estimate for the mean observed species. The other values are the effect of a trait on the expected species duration as expressed as deviation from the mean. The categorical variables are binary index variables where an observation is of that category or not. $\hat{R}$ values of less than 1.1 indicate approximate chain convergence for the posterior samples.

	mean	sd	2.5%	25%	50%	75%	97.5%	$\hat{R}$
alpha	1.29	0.03	1.23	1.27	1.29	1.31	1.36	1.00
intercept	-0.78	0.14	-1.05	-0.87	-0.78	-0.68	-0.51	1.00
logit(occupancy)	-0.53	0.08	-0.69	-0.59	-0.53	-0.48	-0.38	1.00
log(size)	-0.05	0.05	-0.14	-0.08	-0.05	-0.01	0.05	1.00
ground dwelling	-0.28	0.10	-0.47	-0.34	-0.28	-0.21	-0.09	1.00
scansorial	-0.22	0.11	-0.43	-0.29	-0.22	-0.14	-0.00	1.00
herbivore	0.09	0.09	-0.09	0.03	0.09	0.14	0.27	1.00
insectivore	0.10	0.11	-0.11	0.03	0.10	0.17	0.31	1.00
omnivore	-0.12	0.11	-0.33	-0.19	-0.12	-0.05	0.09	1.00
sd cohort	0.33	0.06	0.23	0.29	0.33	0.37	0.48	1.00
sd phylogeny	0.11	0.05	0.03	0.07	0.10	0.14	0.23	1.03