Death and taxa: biological, temporal, and historical effects on mammal species duration

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

Abstract

1 Introduction

Extinction is the second most important aspect of diversification, after only speciation. Why species go extinct at different rates remains one of the most fundamental questions in paleobiology [25, 30, 36, 38, 44, 45, 51, 55, 57]. Determining which and how biological traits influence extinction risk is vital for understanding the differential diversification of life during the Phanerozoic and making predictions about species' vulnerability to anthropogenic impacts. Yet investigations into the factors that affect extinction rate have proved inconclusive and often contradictory results.

The fossil record of North American mammals is well resolved and represents an ideal data set to model long term evolutionary patterns [1–4, 8, 33, 38, 51, 54].

Van Valen proposed the Law of Constant Extinction [55] which states that the instantaneous rate of extinction for a taxon is independent of it's age.

Background extinction, or extinction occurring **not** at mass extinctions, is generally considered complicated with many more potential factors influencing the instantaneous extinction risk of any given species [18, 23, 30, 35, 36, 58]. Factors such as geographic range have well known effects on survival [24, 36] because the effect size is large and thus easy to identify for a given sample size. The between extinction risk and other traits, specifically organismal ones, are less well known because they most likely have much smaller effect sizes which makes identification and inference difficult.

Here I take a Bayesian hierarchical survival model approach for understanding the effects of organismal and species traits, time of origination, and shared evolutionary history on species

	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 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.

duration. This is a direct extension of the more traditional dynamic and cohort survival analyses from paleontology [7, 11, 42, 43, 50, 51, 55, 56].

The organismal and species traits studied here are biogeographic occupancy, body mass, and both dietary and locomotor categories. This is a mixture of well established factors that potentially influencing extinction risk (i.e. occupancy and body mass) and less well understood ones (i.e. dietary and locomotor categories). Dietary category is of particular interest as it has been shown to influence diversification rate [37] though because of the limitations of many phylogenetic comparative methods it is unknown if this is due to differences in speciation, extinction, or both [40].

2 Methods

2.1 Species 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.

Species dietary and locomotor category assignments were done using the assignments in initial the PBDB which were then reassigned into coarser categories (Table 1). This was done to improve interpretability, increase sample size per category, and make these results comparable to previous studies [28, 37].

Fossil occurrences were assigned to 2 My bins ranging through the entire Cenozoic. Taxon duration was measured as the number of bins from the first bin of occurrence to the last bin

of occurrence, inclusive.

Species body size estimates were sourced from a large selection of primary literature and compilations, principally the PBDB, PanTHERIA [29], the Neogene Old World Mammal database (Now; `http://www.helsinki.fi/science/now/`), and other large scale data collection efforts [9, 12, 34, 41, 52]. In many cases, species body mass was estimated from anatomical dimensions such as tooth size. These estimates were made using a variety of published regression equations. See Appendix: Data for a complete list of individual sources and equations.

2.1.1 Bioprovince occupancy

For each 2 My time bin, a bipartite biogeographic network was created between species occurrences and spatial units. Spatial units were defined was 2x2 latitude–longitude grid cells from an azimuthal equal-area map projection. In these bipartite networks, taxa can only be linked to localities and *vice versa*. Taxa are not linked to each other, nor are localities linked. Emergent bioprovinces within the biogeographic occurrence network were identified using the map equation [46, 47]. A bioprovince is a set of species–locality connections that are more interconnected within the group than without. This was done for each bin's biogeographic network using the `igraph` package for R [10, 39]. The relative number of bioprovinces occupied per time bin was then determined for each species.

2.1.2 Semi-formal supertree

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

The taxonomic information from the PBDB served as the basis for additional revision. The taxonomy of many species was updated using the Encyclopedia of Life (`http://eol.org/`), which collects and collates taxonomic information in a single database. This was done programatically using the `taxize` package for R [49]. This was additionally correct using various published phylogenies and taxonomies of fossil mammals [26, 27, 41], producing a tree that was a series of nested polytomies.

Polytomies were resolved in order of species first appearance. The resulting tree was then time scaled using the `paleotree` package via the "minimum branch length" approach with a minimum length of 0.1 My [5]. The minimum length is necessary to avoid zero-length branches which cause the phylogenetic covariance matrix not be positive definite, which is an important convenience for computation (see below). While other time scaling approaches are possible [6, 19] this method was chosen for it's simplicity and not requiring additional information about diversification rates which are of interest in this study.

2.2 Survival model

The statistical model described here was the final model at the end of a continuous model development framework where the sampling and prior distributions were iteratively modified to best reflect theory, knowledge of the data, the inclusion of important covariates, and to fit the data. This follows the approach described in Gelman and Hill [15] and Gelman et al. [16]. A survival model was fit in a Bayesian context where species duration were assumed to be drawn from a Wiebull distribution (Eq. 1) with shape α and scale σ parameters.

$$p(y_i|\alpha, \sigma) = \text{Weibull}(y_i|\alpha, \sigma)$$
$$= \frac{\alpha}{\sigma} \left(\frac{y_i}{\sigma}\right)^{\alpha-1} \exp\left(-\left(\frac{y_i}{\sigma}\right)^{\alpha}\right) \tag{1}$$

As is standard practice in survival analysis [31], α was assumed constant and was given a weakly informative half-Cauchy prior. σ was reparameterized as an exponentiated regression model (Eq. 2).

$$\sigma = \exp\left(\frac{-(h_i + \eta_{j[i]} + \sum \beta^T \mathbf{X}_i)}{\alpha}\right) \tag{2}$$

K species level covariates were included as a $n \times K$ matrix, $\mathbf{X}$. The covariates of interest included the logit of mean relative occupancy and the logarithm of body size (g). The discrete covariate index variables 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. Only $k-1$ indicator variables are necessary as the intercept takes on the remaining value. Finally, a vector of 1-s were included in the matrix $\mathbf{X}$ whose corresponding β coefficient is the intercept, making K equal eight.

β is a vector of regression coefficients, where each element was given a unique, weakly informative Normally distributed prior. 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.

Regression coefficients are not directly comparable without first standardizing the input variables to have equal standard deviations. 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 [48]. However, because the expected standard deviation for a binary variable is 0.5, in order to make comparisons between the binary and continuous variables, the continuous inputs were divided by twice their standard deviation [14]. The above model was fit with both unstandardardized and standardized inputs for illustrative purposes.

4

2.2.1 Hierarchical effects

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 [15], where complete pooling is when the differences between groups are ignored while no pooling is where different groups are analyzed separately. By allowing for partial pooling, we are modeling a compromise between these two extremes which allows for better inference. This is done by having all of the groups sharing the same prior with the 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. Hierarchical modeling is analogous to mixed-effects modeling [15].

Origination cohort is defined as the group of species which originated during the same 2 My temporal bin. The most recent temporal bin, 0-2 Mya, was excluded, leaving 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 unification between dynamic and cohort survival analysis in paleontology [7, 11, 42, 43, 56], with σ_c acting as a measure of relative importance of these two end members.

$$\eta_j \sim \mathcal{N}(0, \sigma_c)$$
$$\sigma_c \sim \text{halfCauchy}(0, 2.5)$$

The choice of prior for σ_c follows Gelman [13]

The impact of shared evolutionary history, or phylogeny, was modeling as an individual effect where each observation, i, is distributed as a multivariate normal, h, where the covariance matrix was assumed known up to a constant variance, σ_p^2 [21, 32]. More fully, this is written

$$h \sim \text{Multi}\mathcal{N}(0, \mathbf{\Sigma})$$
$$\mathbf{\Sigma} = \sigma_p^2 \mathbf{V}_{phy}$$
$$\sigma_p \sim \text{halfCauchy}(0, 2.5),$$

with $\mathbf{V}_{phy}$ being 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.

Both η and h were centered at 0 so that these effects can be interpreted as differences from the mean of the observed.

5

2.2.2 Censored observations

An important part of survival analysis is the inclusion of "censored" observations [22, 31] or observations where the failure time has not been observed. The most common censored observation is right censored, where the point of extinction had not yet been observed in the period of study. In this case, this means taxa that are still extant. Left censored observations, on the other hand, correspond to observations that went extinct any time between 0 and some known point. In this study, taxa occurring in only a single time bin were left censored. Because of the minimum resolution of the record, we cannot observe if these taxa went extinct in less than that single bin or not.

To model censored data, there are a few functions which must first be defined following Klein and Moeschberger [31]. The survival function, $S(t)$, for a Weibull model is

$$S(t|\alpha, \sigma) = \exp\left(-\left(\frac{t}{\sigma}\right)^{\alpha}\right), \tag{3}$$

where σ is defined as above (Eq. 2). $S(t)$ is the probability that an observation will survive longer than a give time, t. The likelihood of uncensored observations is evaluated as as normal using Equation 1 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 cummulative density function and $S(t)$ is equivalent to the complementary cummulative density function [16].

Fully, the likelihood for both uncensored and both right and left censored observations is written

$$L \propto \prod_{i \in C} \text{Weibull}(y_i|\alpha, \sigma) \prod_{j \in R} S(y_j|\alpha, \sigma) \prod_{k \in L} (1 - S(y_k|\alpha, \sigma)),$$
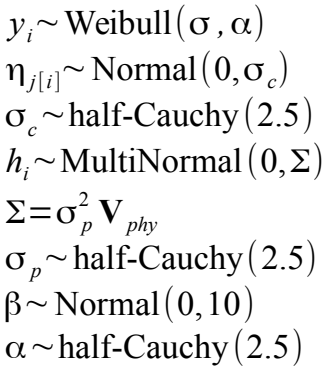
where C is the set of uncensored observations, R is the set of right censored observations, and L is the set of left censored observations.

A summary of the entire model, save for calculations for censored observations, along with the exact priors for every estimated parameter is presented in Figure 1.

2.2.3 Estimation

The parameter posteriors were approximated using a Markov-chain Monte Carlo (MCMC) routine implemented in the Stan programming language [53]. Stan implements a Hamiltonian Monte Carlo using a No-U-Turn sampler [20]. Posterior approximation was done using four parallel MCMC chains. Chain 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 [16].

Both models with and without phylogenetic effects were estimated. Because inverting a large matrix is a memory intense procedure and because the phylogenetic covariance matrix is

$y_i \sim \text{Weibull}(\sigma, \alpha)$

$\eta_{j[i]} \sim \text{Normal}(0, \sigma_c)$

$\sigma_c \sim \text{half-Cauchy}(2.5)$

$h_i \sim \text{MultiNormal}(0, \Sigma)$

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

$\sigma_p \sim \text{half-Cauchy}(2.5)$

$\beta \sim \text{Normal}(0, 10)$

$\alpha \sim \text{half-Cauchy}(2.5)$

$$\exp\left(\frac{-(\eta_{j[i]} + h_i + \sum \beta^T X_i)}{\alpha}\right)$$

$\text{MultiNormal}(0, \sigma_p^2 \mathbf{V}_{phy})$

Figure 1: Graphical depiction of full survival model, save censoring, used here. Exact values for the hyperparameters are presented to the right. The observed duration of the ith observation is indicated towards the bottom left as y_i, which is assumed to follow a Weibull distribution.

only assumed known up to a constant, every iteration of the MCMC would involve solving a very large matrix which is not ideal. In order to speed up the MCMC routine, this aspect of the model had to be reparameterized for efficiency purposes. Because of the size of the covariance matrix a custom multivariate sampler was used (see Appendix: Code).

For the model without phylogenetic effect the four MCMC chains ran for 1000 steps, with the first 500 used as warm-up and the last 500 as samples from the posterior. Because of the added complexity of estimating the phylogenetic effect, instead all four chains were run 20000 steps thinned to every twentieth sample split evenly between warm-up and sampling.

2.3   Posterior predictive checks

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 predictors 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 repeated 1000 times and the distribution of y^{rep} was compared with the observed y [16]. This was done both graphically and numerically.

An example posterior predictive check used in this study is a graphical comparison of the Kaplan-Meier (K-M) survival curve estimated from the observed data with the survival curves from 1000 simulations. K-M survival curves are non-parametric estimates of the function $S(t)$ or the probability of a species going extinct given that it has lived to time t [31]. Other posterior predictive checks included 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, defined below.

In ordinary linear regression, the residuals are defined as $r = y - y_{est}$. For hierarchical models, this definition is inadequate. For survival analysis, the equivalent values are deviance residuals. To define how deviance residuals are calculated, we first define the cummulative hazard function [31]. Given $S(t)$ (Eq. 3), we define the cumulative hazard function as

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

Next, we define martingale residuals, m, which are defined in relation to the inclusion vector I

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

I is as a vector of length n, where $I_i = 1$ means the observation is completely observed and $I_i = 0$ means the observation is censored.

Martingale residuals have a mean of 0 and ranges between 1 and $-\infty$ and can be viewed as the difference between the observed number of deaths between 0 and t_i and the expected number of deaths based on the model. However, the range of the martingale residuals is

difficult to interpret, can be asymmetrically distributed, and are not equivalent to standard residuals.

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

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

2.4 Variance partitioning

There are three different variance components in this model (Fig. 1): sample σ_y^2, cohort σ_c^2, and phylogenetic σ_p^2. However, the model used here (Eq. 1) does not include an estimate of overall variance. To partition the variance between these three components, instead an approximation was done via a simulation approach modified from Goldstein et al. [17].

The procedure is as follows:

1. Simulate w (50,000) values of η; $\eta \sim \mathcal{N}(0, \sigma_c)$.

2. For a given value of $\beta^T\mathbf{X}$, calculate σ^{c*} (Eq. 2) for all w simulations, holding h constant at 0.

3. Calculate v_c, the Weibull variance (Eq. 4) of each element of σ^{c*} with α drawn from the posterior estimate.

4. Simulate w values of h; $h \sim \mathcal{N}(0, \sigma_p)$

5. For a given value of $\beta^T\mathbf{X}$, calculate σ^{p*} (Eq. 2) for all w simulations, holding η constant at 0.

6. Calculate v_p, the Weibull variance (Eq. 4) of each element of σ^{p*} with α drawn from the posterior estimate.

7. $\sigma_{y*}^2 = \frac{1}{2}\left(\left(\frac{1}{w}\sum_i^w v_{pi}\right) + \left(\frac{1}{w}\sum_j^w v_{cj}\right)\right)$

8. $\sigma_{c*}^2 = var(v_c)$ and $\sigma_{p*}^2 = var(v_p)$

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

Weibull variance is calculated as

$$var(x) = \sigma^2\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2\right), \tag{4}$$

where Γ is the gamma function.

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

I used ratios of the variances and variance partitioning coefficients (VPC) to measure the amount of partial pooling. The former, when compared to the sample size of the hierarchical effect such as the number of cohorts, is an estimate of the relative amount of pooling, while the latter is a measure of the relative importance of the different variance components [15]. Phylogenetic heritability, h_p^2 [21], is effectively 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 tree effect.

3   Results

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat
intercept	-0.82	0.18	-1.17	-0.94	-0.82	-0.70	-0.48	1.00
logit(occupancy)	-0.66	0.08	-0.81	-0.71	-0.65	-0.60	-0.50	1.00
log(size)	-0.05	0.05	-0.15	-0.08	-0.05	-0.01	0.05	1.00
ground dwelling	-0.12	0.12	-0.34	-0.20	-0.12	-0.04	0.12	1.00
scansorial	-0.13	0.13	-0.38	-0.22	-0.13	-0.05	0.12	1.00
herbivore	0.11	0.12	-0.13	0.03	0.12	0.19	0.35	1.00
insectivore	0.09	0.13	-0.17	0.00	0.09	0.19	0.36	1.00
omnivore	-0.12	0.13	-0.37	-0.20	-0.12	-0.04	0.14	1.00
sd cohort	0.33	0.07	0.22	0.29	0.33	0.37	0.48	1.00
sd phylogeny	0.21	0.10	0.07	0.14	0.19	0.26	0.46	1.05
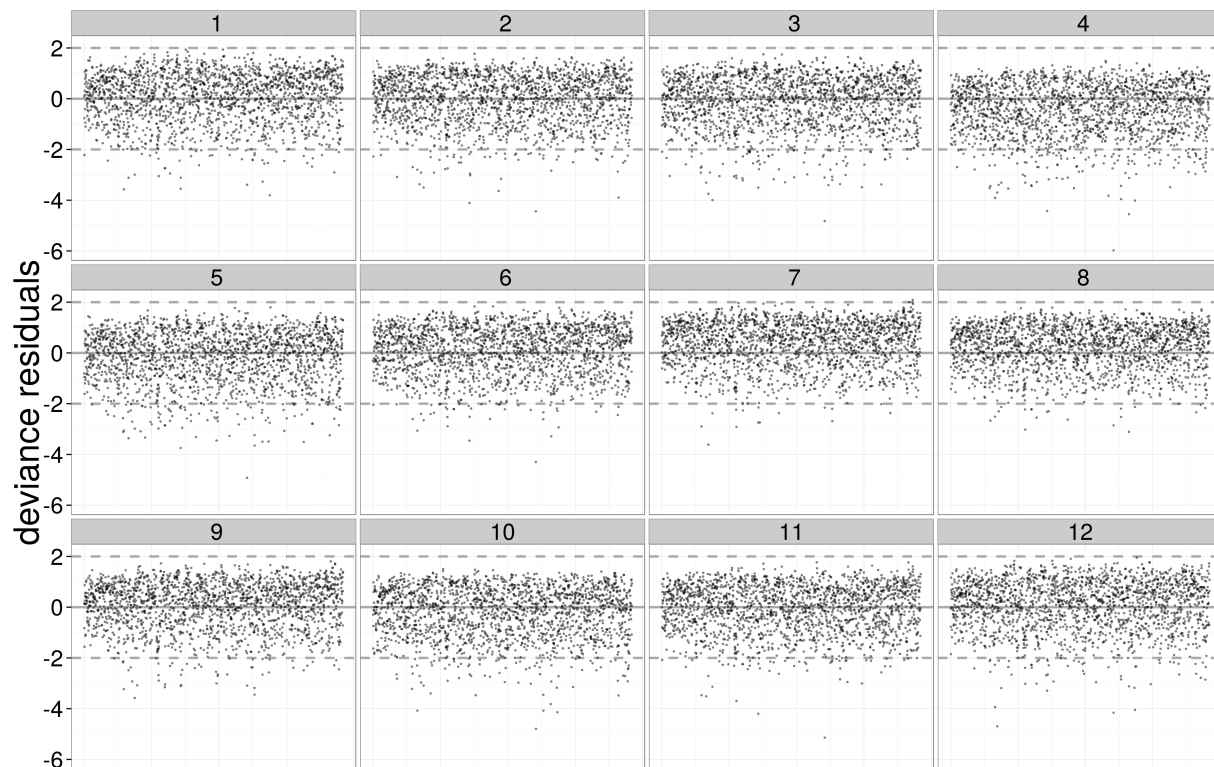
Figure 2: Deviance residuals from the fitted survival model. Each graph depicts the residuals from single draws from the posteriors distributions of all estimated parameters. Positive values indicate an under estimate of the observed duration, while negative values indicate an over estimate of the observed duration. Twelve difference examples are provided here to indicate the lack of individual observation based biases.

4   Discussion

References

[1] J. Alroy. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 127: 285–311, 1996.

[2] J. Alroy. Cope's rule and the dynamics of body mass evolution in North American fossil mammals. *Science*, 280:731–734, 1998. doi: 10.1126/science.280.5364.731.

[3] J. Alroy. Speciation and extinction in the fossil record of North American mammals. In
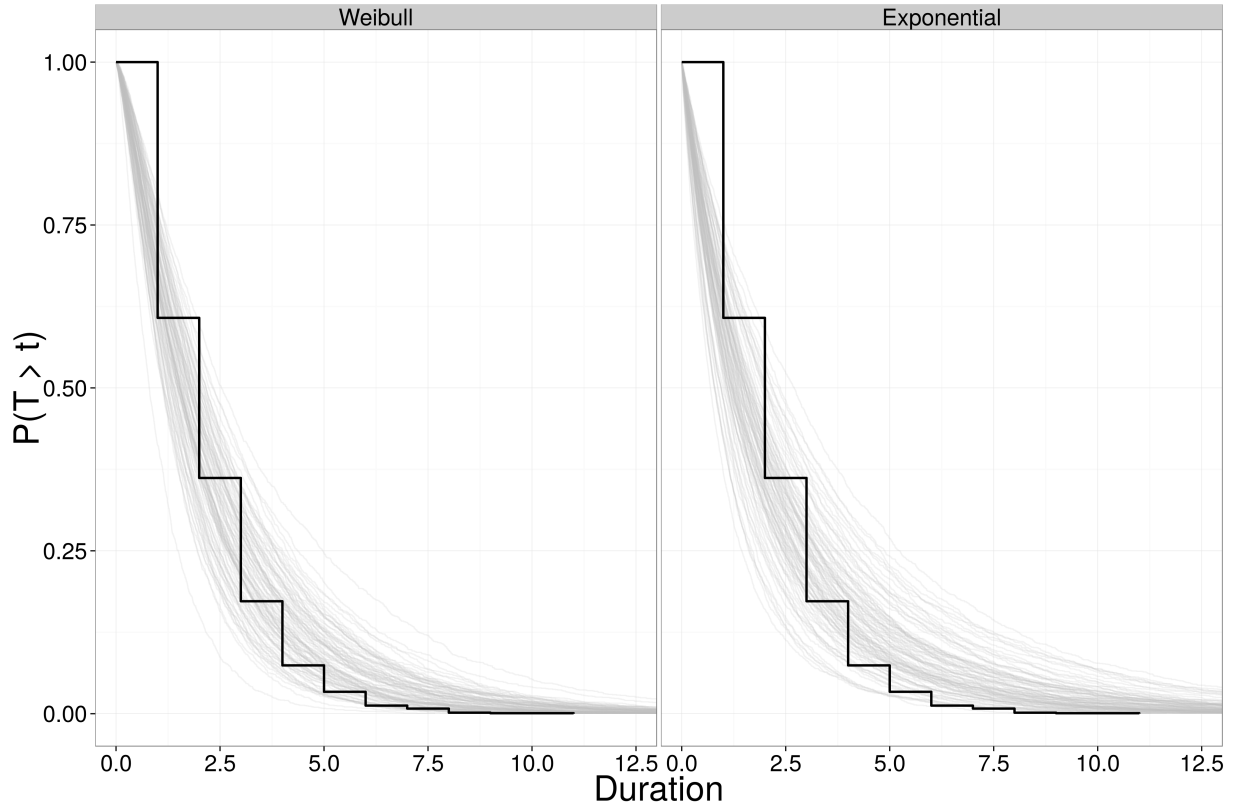
Figure 3: Comparison between K-M estimate of survival function (black) from the observed versus K-M estimates from 100 simulated data sets using the fitted model (dark grey). Simulated data sets were generated by drawing parameter values randomly from their estimated posteriors and using the observed covariate information to estimate durations for all the observed species. On the left are the results from the full survival model (Fig. 1), while on the right are the results from a simplified model where duration follows an exponential distribution and there is no phylogenetic effect.

R. K. Butlin, J. R. Bridle, and D. Schluter, editors, *Speciation and patterns of diversity*, pages 302–323. Cambridge University Press, Cambridge, 2009.

[4] J. Alroy, P. L. Koch, and J. C. Zachos. Global climate change and North American mammalian evolution. *Paleobiology*, 26(1981):259–288, 2000.

[5] D. W. Bapst. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, 3:803–807, 2012. doi: 10.1111/j. 2041-210X.2012.00223.x. URL http://doi.wiley.com/10.1111/j.2041-210X.2012. 00223.xpapers2://publication/doi/10.1111/j.2041-210X.2012.00223.x.

[6] D. W. Bapst. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods in Ecology and Evolution*, 4(8):724–733, Aug. 2013. ISSN 2041210X. doi: 10.1111/2041-210X.12081. URL http://doi.wiley.com/10.1111/2041-210X.12081.
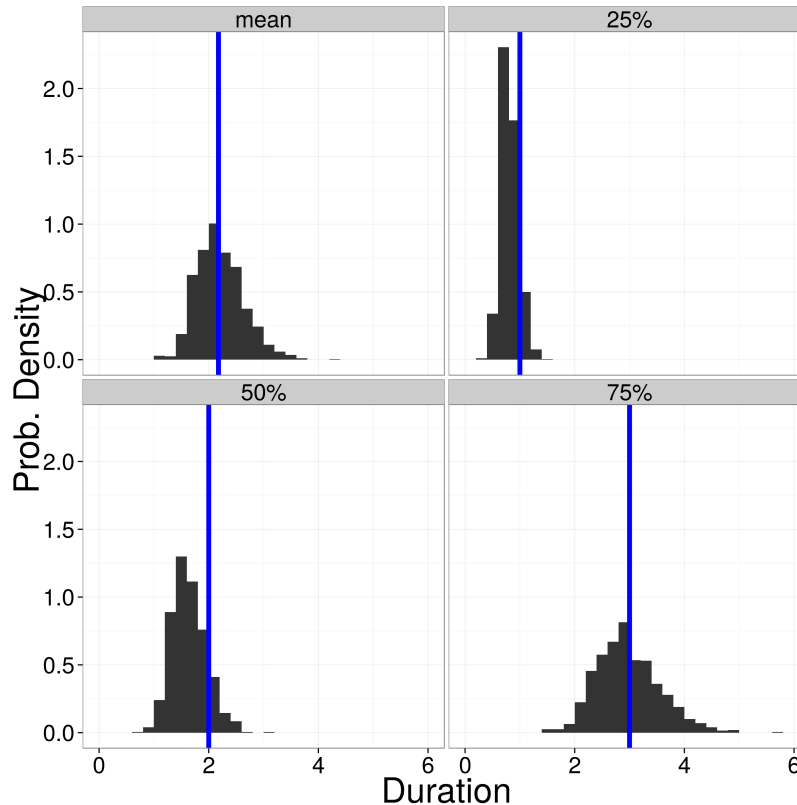
Figure 4: The results of additional posterior predictive checks for four summaries of the observed durations, as labeled. Blue vertical indicate the observed value. None of the observed are significantly different from the posterior predictive distributions.

[7] T. K. Baumiller. Survivorship analysis of Paleozoic Crinoidea: effect of filter morphology on evolutionary rates. *Paleobiology*, 19(3):304–321, 1993.

[8] J. L. Blois and E. A. Hadly. Mammalian Response to Cenozoic Climatic Change. *Annual Review of Earth and Planetary Sciences*, 37(1):181–208, May 2009. ISSN 0084-6597. doi: 10.1146/annurev.earth.031208.100055.

[9] B. W. Brook and D. M. J. S. Bowman. The uncertain blitzkrieg of Pleistocene megafauna. *Journal of Biogeography*, 31(4):517–523, Apr. 2004. ISSN 03050270. doi: 10.1046/j. 1365-2699.2003.01028.x. URL http://doi.wiley.com/10.1046/j.1365-2699.2003. 01028.x.

[10] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL http://igraph.org.

[11] M. Foote. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology*, 14 (3):258–271, 1988.
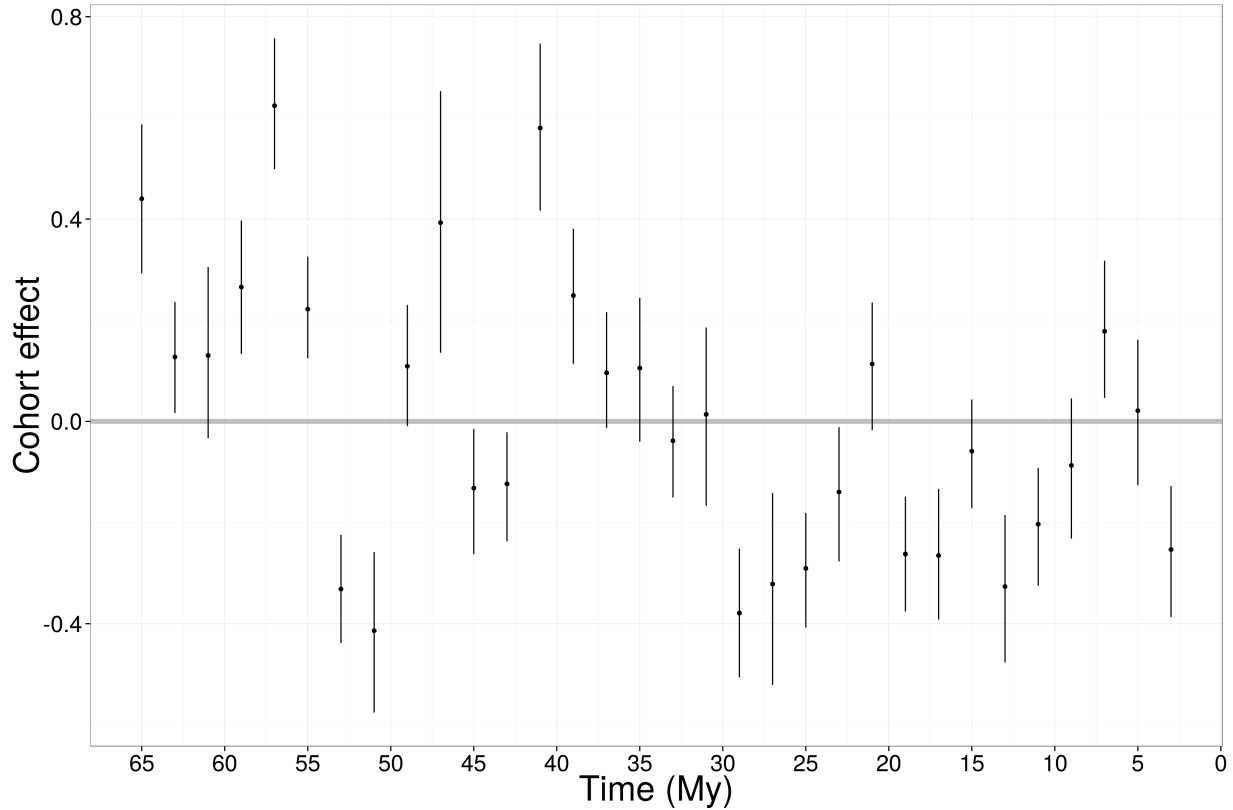
Figure 5: Summaries of estimated individual cohort effect posteriors. Depicted are medians and 80% credible intervals of the estimated posterior distributions. High values correspond to shorter species durations while lower values correspond to greater species durations compared to the mean duration. Lines are placed at the middle of the 2 My origination cohorts.

[12] M. Freudenthal and E. Martín-suárez. Estimating body mass of fossil rodents. *Scripta Geologica*, 145:1–130, 2013.

[13] A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.

[14] A. Gelman. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, pages 2865–2873, 2008. doi: 10.1002/sim.

[15] A. Gelman and J. Hill. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY, 2007.

[16] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall, Boca Raton, FL, 3 edition, 2013.

[17] H. Goldstein, W. Browne, and J. Rasbash. Partitioning variation in multilevel models. *Understanding Statistics*, 1(4):1–12, 2002. URL http://www.tandfonline.com/doi/abs/10.1207/S15328031US0104_02.
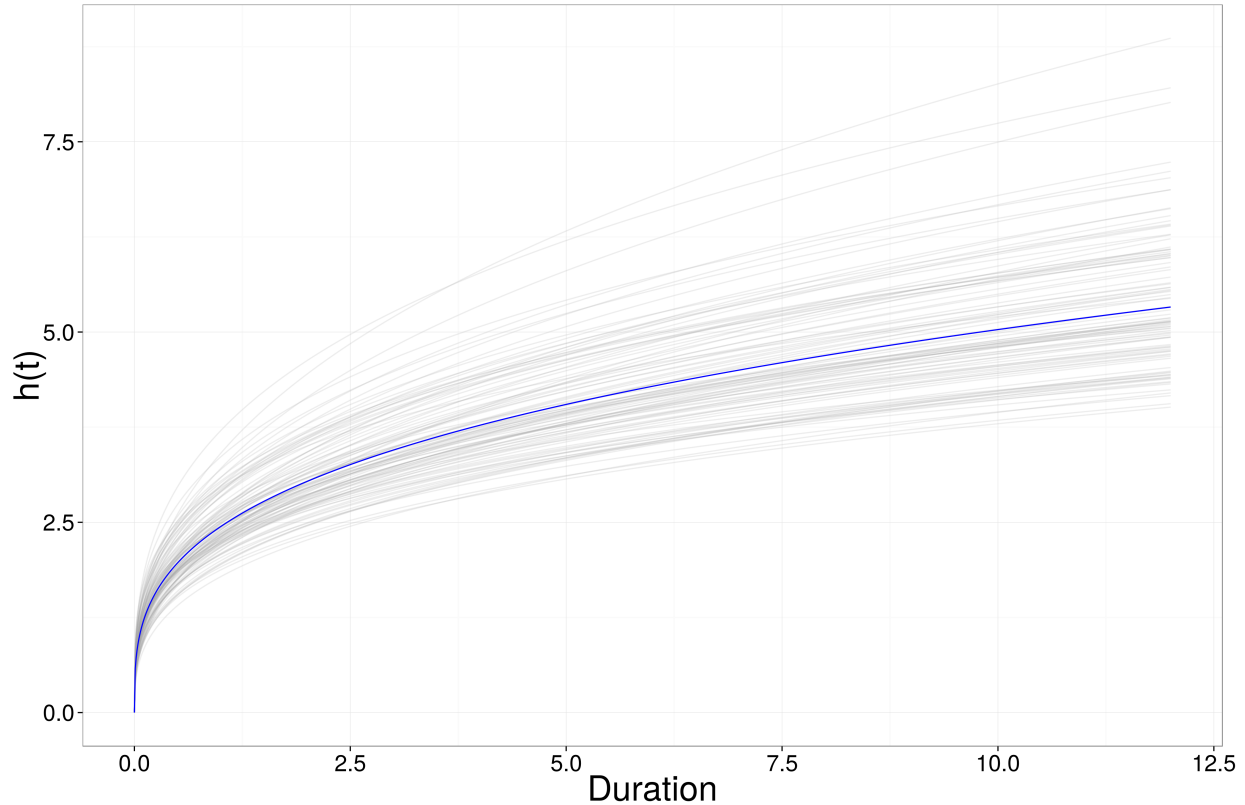
Figure 6: 100 estimates of the hazard function ($h(t)$ for the observed species mean (grey), along with the median estimated hazard function. $h(t)$ is an estimate of the number of extinctions that occur given an amount of time, t. Hazard functions were estimated from random draws from the estimated posterior distributions and evaluated with all covariate information set to 0, which corresponds to the expected duration of the mean species.

[18] P. G. Harnik, C. Simpson, and J. L. Payne. Long-term differences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society B: Biological Sciences*, (October), Oct. 2013. ISSN 0962-8452. doi: 10.1098/rspb.2012.1902. URL http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2012.1902.

[19] M. M. Hedman. Constraints on clade ages from fossil outgroups. *Paleobiology*, 36(1): 16–31, 2010.

[20] M. D. Hoffman and A. Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv*, 1111(4246), 2011. URL http://arxiv.org/abs/1111.4246.

[21] E. A. Housworth, P. Martins, and M. Lynch. The Phylogenetic Mixed Model. *The American Naturalist*, 163(1):84–96, 2004.

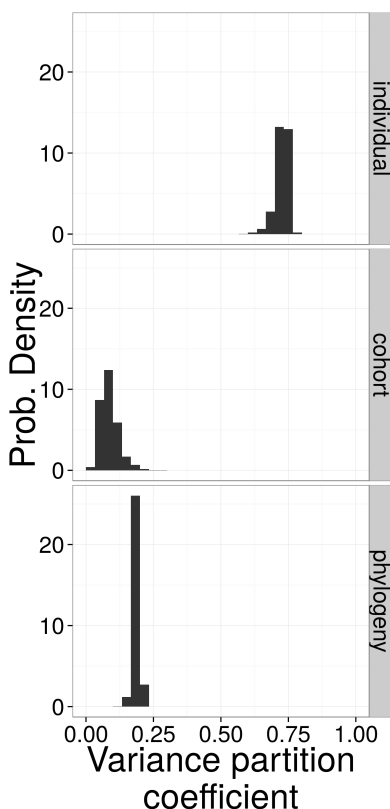[22] J. G. Ibrahim, M.-H. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, New

Figure 7: 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 variance. Each of the estimates is a distribution of 1000 approximating simulations due to the models non-normality.

York, 2001.

[23] D. Jablonski. Background and mass extincitons: the alternation of macroevolutionary regimes. *Science*, 231(4734):129–133, 1986.

[24] D. Jablonski. Heritability at the species level: analysis of geographic ranges of cretaceous mollusks. *Science*, 238(4825):360–363, Oct. 1987. ISSN 0036-8075. doi: 10.1126/science. 238.4825.360. URL `http://www.ncbi.nlm.nih.gov/pubmed/17837117`.

[25] D. Jablonski. Mass extinctions and macroevolution. *Paleobiology*, 31(sp5):192–210, 2005. URL `papers2://publication/uuid/F296080D-6609-4C57-9514-C74E5E726340`.

[26] C. M. Janis, K. M. Scott, and L. L. Jacobs. *Evolution of Tertiary mammals of North America. Vol. 1. Terrestrial carnivores, ungulates, and ungulatelike mammals.* Cambridge University Press, Cambridge, 1998.

[27] C. M. Janis, G. F. Gunnell, and M. D. Uhen. *Evolution of Tertiary mammals of*

North America. Vol. 2. Small mammals, xenarthrans, and marine mammals. Cambridge University Press, Cambridge, 2008.

[28] J. Jernvall and M. Fortelius. Maintenance of trophic structure in fossil mammal communities: site occupancy and taxon resilience. *American Naturalist*, 164(5):614–624, Nov. 2004. ISSN 1537-5323. doi: 10.1086/424967.

[29] K. E. Jones, J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, C. D. L. Orme, K. Safi, W. Sechrest, E. Boakes, C. Carbone, C. Connolly, M. J. Cutts, J. K. Foster, R. Grenyer, M. Habib, C. A. Plaster, S. A. Price, E. A. Rigby, J. Rist, A. Teacher, O. R. P. Bininda-Emonds, J. L. Gittleman, G. M. Mace, and A. Purvis. PanTHERIA : a species-level database of life history , ecology , and geography of extant and recently extinct mammals. *Ecology*, 90(9):2648, 2009.

[30] J. A. Kitchell, D. L. Clark, and A. M. Gombos. Biological selectivity of extinction: a link between background and mass extinction. *Palaios*, 1(5):504–511, 1986.

[31] J. P. Klein and M. L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data.* Springer, New York, 2nd edition, 2003.

[32] M. Lynch. Methods for the analysis of comparative data in evolutionary biology. *Evolution*, 45(5):1065–1080, 1991.

[33] J. D. Marcot. The fossil record and macroevolutionary history of North American ungulate ungulate mammals: standardizing variation in intensity and geography of sampling. *Paleobiology*, 40(2):237–254, Feb. 2014. ISSN 0094-8373. doi: 10.1666/13052. URL http://www.bioone.org/doi/abs/10.1666/13052.

[34] R. T. McKenna. *Potential for Speciation in Mammals Following Vast , Late Miocene Volcanic Interruptions in the Pacific Northwest.* Masters, Portland State University, 2011.

[35] S. Nürnberg and M. Aberhan. Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*, 39(3):360–372, Apr. 2013. ISSN 0094-8373. doi: 10.1666/12047. URL http://www.bioone.org/doi/abs/10.1666/12047.

[36] J. L. Payne and S. Finnegan. The effect of geographic range on extinction risk during background and mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 104(25):10506–11, June 2007. ISSN 0027-8424. doi: 10.1073/pnas.0701257104. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1890565&tool=pmcentrez&rendertype=abstract.

[37] S. A. Price, S. S. B. Hopkins, K. K. Smith, and V. L. Roth. Tempo of trophic evolution and its impact on mammalian diversification. *Proceedings of the National Academy of Sciences of the United States of America*, 109(18):7008–12, May 2012. ISSN 1091-6490. doi: 10.1073/pnas.1117133109. URL http://www.ncbi.nlm.nih.gov/pubmed/22509033.

[38] T. B. Quental and C. R. Marshall. How the Red Queen Drives Terrestrial Mammals to

Extinction. *Science*, 341(6143):290–292, Sept. 2013. ISSN 0022-1376. doi: 10.1086/627905. URL `http://www.journals.uchicago.edu/doi/abs/10.1086/627905`.

[39] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.R-project.org/`.

[40] D. L. Rabosky. Extinction rates should not be estimated from molecular phylogenies. *Evolution*, 64(6):1816–24, June 2010. ISSN 1558-5646. doi: 10.1111/j.1558-5646.2009. 00926.x. URL `http://www.ncbi.nlm.nih.gov/pubmed/20030708`.

[41] P. Raia, F. Carotenuto, F. Passaro, D. Fulgione, and M. Fortelius. Ecological specialization in fossil mammals explains Cope's rule. *The American naturalist*, 179(3):328–37, Mar. 2012. ISSN 1537-5323. doi: 10.1086/664081. URL `http://www.ncbi.nlm.nih.gov/pubmed/22322221`.

[42] D. M. Raup. Taxonomic survivorship curves and Van Valen's Law. *Paleobiology*, 1 (1):82–96, Jan. 1975. ISSN 0036-8075. doi: 10.1126/science.49.1254.50. URL `http://www.ncbi.nlm.nih.gov/pubmed/17777225`.

[43] D. M. Raup. Cohort Analysis of generic survivorship. *Paleobiology*, 4(1):1–15, 1978.

[44] D. M. Raup. *Extinction: Bad Genes or Bad Luck?* Norton, New York, 1991.

[45] D. M. Raup. The role of extinction in evolution. *Proceedings of the National Academy of Sciences*, 91(July):6758–6763, 1994.

[46] M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–23, Jan. 2008. ISSN 1091-6490. doi: 10.1073/pnas. 0706851105. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2234100&tool=pmcentrez&rendertype=abstract`.

[47] M. Rosvall, D. Axelsson, and C. Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(14):13–24, 2009. URL `http://www.springerlink.com/index/H8193132U6432363.pdf`.

[48] H. Schielzeth. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2):103–113, Feb. 2010. ISSN 2041210X. doi: 10. 1111/j.2041-210X.2010.00012.x. URL `http://doi.wiley.com/10.1111/j.2041-210X.2010.00012.x`.

[49] Scott Chamberlain and Eduard Szocs. taxize - taxonomic search and retrieval in r. *F1000Research*, 2013. URL `http://f1000research.com/articles/2-191/v2`.

[50] J. J. Sepkoski. Stratigraphic biases in the analysis of taxonomic survivorship. *Paleobiology*, 1(4):343–355, 1975.

[51] G. G. Simpson. *Tempo and Mode in Evolution*. Columbia University Press, New York, 1944.

[52] F. A. Smith, J. H. Brown, J. P. Haskell, S. K. Lyons, J. Alroy, E. L. Charnov, T. Dayan, B. J. Enquist, S. K. M. Ernest, E. A. Hadly, K. E. Jones, D. M. Kaufman, P. A. Marquet, B. A. Maurer, K. J. Niklas, W. P. Porter, B. Tiffney, and M. R. Willig. Similarity of Mammalian Body Size across the Taxonomic Hierarchy and across Space and Time. *The American Naturalist*, 163(5):672–691, 2004.

[53] Stan Development Team. Stan: A c++ library for probability and sampling, version 2.5.0, 2014. URL `http://mc-stan.org/`.

[54] S. Tomiya. Body Size and Extinction Risk in Terrestrial Mammals Above the Species Level. *The American Naturalist*, pages E000–E000, Sept. 2013. ISSN 00030147. doi: 10.1086/673489. URL `http://www.jstor.org/stable/info/10.1086/673489`.

[55] L. Van Valen. A new evolutionary law. *Evolutionary Theory*, 1:1–30, 1973. URL `http://ci.nii.ac.jp/naid/10011264287/`.

[56] L. Van Valen. Taxonomic survivorship curves. *Evolutionary Theory*, 4:129–142, 1979.

[57] P. J. Wagner and G. F. Estabrook. Trait-based diversification shifts reflect differential extinction among fossil taxa. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1–6, Oct. 2014. ISSN 1091-6490. doi: 10.1073/pnas. 1406304111. URL `http://www.ncbi.nlm.nih.gov/pubmed/25331898`.

[58] S. C. Wang. On the continuity of background and mass extinction. *Paleobiology*, 29(4):455–467, Dec. 2003. ISSN 0094-8373. doi: 10.1666/0094-8373(2003)029⟨0455: OTCOBA⟩2.0.CO;2. URL `http://www.bioone.org/doi/abs/10.1666/0094-8373%282003%29029%3C0455%3AOTCOBA%3E2.0.CO%3B2`.