Supplementary information

Peter D Smits

February 24, 2015

1 Supplementary methods

1.1 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.

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 results comparable to previous studies [74, 80].

All individual 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 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 [51, 52, 78]. The youngest cohort, 0-2 My, was excluded from analysis.

Table 1: Species trait assignments in this study are a coarser version of the information available in the PBDB. Information was coarsened to improve per category sample size and uniformity and followed this table.

This study		PBDB categories
Diet	Carnivore	Carnivore
	Herbivore	Browser, folivore, granivore, grazer, herbivore.
	Insectivore	Insectivore.
	Omnivore	Frugivore, omnivore.
Locomotor	Arboreal	Arboreal.
	Ground dwelling	Fossorial, ground dwelling, semifossorial, saltatorial.
	Scansorial	Scansorial.

1.1.1 Body size estimates and estimation

Species body size estimates were sourced from a large selection of primary literature and data base compilations.

Databases used include the PBDB, PanTHERIA [75], and the Neogene Old World Mammal database (NOW; http://www.helsinki.fi/science/now/). Major sources of additional compiled body size estimates include [57, 61, 79, 81, 89, 91]. 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. See Supplementary Tables XXX and XXX for a complete list of individual sources and equations, respectively.

1.1.2 Biogeographic network

Species geographic extent was characterized using a network-theoretic approach that has previously been applied to paleontological data [88, 95]. 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 layed over 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 25 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 [86, 85] as has been done before [88, 94, 95]. 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 [68]. The map equation and other network related analysis was done using the `igraph` package for R [59].

1.1.3 Supertree

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

The initial taxonomic classification of the observed species was based on the associated taxonomic information from the PBDB. This information was then 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 [58]. Finally, this taxonomic information was further updated using a published taxonomy of fossil mammals [72, 73].

This taxonomy serves as in initial phylogenetic hypothesis which was then combined with a selection of species-level phylogenies [56, 81] in order to better constraint a minimum estimate of the actual phylogenetic relationships of the species. The supertree was inferred via matrix representation parsimony was implemented in the `phytools` package for R [84]. While four most parsimonious trees were found, I selected a single of these for use in analysis.

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 [53]. The minimum length is necessary to avoid zero-length branches which cause the phylogenetic covariance matrix not be positive definite, which is important for computation (see below). While other time scaling approaches are possible [54, 67] this method was chosen for it's simplicity and not requiring additional information about diversification rates which are the interest of this study.

1.2 Survival model

Presented here is the model development process used here 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 durations 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 [76]. This is written out

$$p(y|\lambda) = \lambda \exp(-\lambda y)$$
$$y \sim \text{Exp}(\lambda). \tag{1}$$

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 My will not have gone extinct while $h(t)$ corresponds to the instantaneous extinction rate for some taxon age t [76]. For an exponential model, $S(t)$ is defined

$$S(t) = \exp(-\lambda t) \tag{2}$$

and $h(t)$ is defined

$$h(t) = \lambda \tag{3}$$

The choice of the exponential distribution corresponds directly to the Law of Constant Extinction [92] as the right side of Eq. 3 does not depend on species age t.

The current sampling statement (Eq. 1) currently 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 [76].

To relax the assumption of age-independent extinction of the Law of Constant Extinction we substitute the Weibull distribution for the exponential [76]. 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 because equivalent to the exponential.

The Weibull distribution and sampling statement wre defined

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

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

$$S(t) = \exp\left(-\left(\frac{t}{\sigma}\right)^{\alpha}\right) \tag{5}$$

$$h(t) = \frac{\alpha}{\sigma} \left(\frac{t}{\sigma}\right)^{\alpha-1}. \tag{6}$$

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

$$\sigma_i = \exp\left(\frac{-(\sum \beta^T \mathbf{X}_i)}{\alpha}\right). \tag{7}$$

The 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 the fit to the data. This follows the approach described in [65] and [64]. A survival model was fit in a Bayesian context where species duration were assumed to be drawn from a Wiebull distribution (Eq. 4) with shape α and scale σ parameters. α was assumed constant, which is standard practice in survival analysis [76]. 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

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

4

where equivalent statement for the exponential distribution is defined

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

K species-level covariates were included as a $n \times K$ matrix, $\mathbf{X}$. Two 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 a vector of regression coefficients, where each element is 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 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 [87]. Additionally, this makes the intercept directly interpretable as the estimate of mean (transformed) σ (Eq. 7). 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 must instead by divided by twice their standard deviation [63].

1.2.1 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 of groups [65], 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 groups estimates to be interpreted as deviations from the intercept. Hierarchical modeling is analogous to mixed-effects modeling [65].

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

Mya 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 [60, 83, 82, 93, 55], 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 [62]

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 Σ is known up to a constant, σ_p^2 [77, 70]. This is written

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

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

1.2.2 Censored observations

An important part of survival analysis is the inclusion of censored observations where the failure time has not been observed [71, 76]. 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 extant. 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 is modeled using the survival function of the distribution, $S(t)$, defined earlier for the Weibull distribution (Eq. 5) with σ is 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 density function and $S(t)$ is equivalent to the complementary cumulative density function [64].

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.

1.2.3 Estimation

Parameter posteriors were approximated using a Markov-chain Monte Carlo (MCMC) routine implemented in the Stan programming language [90]. Stan implements a Hamiltonian Monte Carlo using a No-U-Turn sampler [69]. Posterior approximation was done using four parallel MCMC chains where 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 [64].

In order to speed up the poisterior 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.

Four different MCMC chains were run for 20000 steps and were thinned to every twentieth sample split evenly between warm-up and sampling.

1.3 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 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 [64].

An example posterior predictive check used in this study is a graphical comparison between the Kaplan-Meier (K-M) survival curve estimated from the observed data and the K-M survival curves estimated from 1000 simulation sets. K-M survival curves are non-parametric estimates of $S(t)$ [76]. 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, defined below.

In standard linear regression, residuals are defined as $r_i = y_i - y_i^{est}$. For the model used here, this definition is inadequate. The equivalent values for

survival analysis are deviance residuals. To define how deviance residuals are calculated, we first define the cummulative hazard function [76]. Given $S(t)$ (Eq. 5), we define the cumulative hazard function as

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

Next, we define martingale residuals m as

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

I is the inclusion 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, range 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, martingale residuals are asymmetrically distributed, and can not be interpreted in the same manner as 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

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

Deviance residuals have a mean of 0 and a standard deviation of 1 by definition.

1.4 Variance partitioning

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 [66].

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. 7) for all w simulations, holding h constant at 0.

3. Calculate v_c, the Weibull variance (Eq. 10) 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. 7) for all w simulations, holding η constant at 0.

6. Calculate v_p, the Weibull variance (Eq. 10) of each element of σ^{p*} with α drawn from the posterior estimate.

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

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

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

$$var(x) = \sigma^2\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2\right), \tag{10}$$

where Γ is the gamma function.

The variance partitioning coefficients are then calculated, for example, as $VPC_{phylo} = \frac{\sigma^2_{p*}}{\sigma^2_{y*} + \sigma^2_{c*} + \sigma^2_{p*}}$ and similarly for the other components.

I used variance partitioning coefficients (VPC) to estimate of the relative size importance of the different variance components [65]. Phylogenetic heritability, h^2_p [77, 70], 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.

2 Model Adequacy

Deviance residuals.
    Point estimates.
    K-M estimates of S(t).

References

[51] John Alroy. Speciation and extinction in the fossil record of North American mammals. In Roger K Butlin, Jon R Bridle, and Dolph Schluter, editors, *Speciation and patterns of diversity*, pages 302–323. Cambridge University Press, Cambridge, 2009.

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

[53] David W Bapst. paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods in Ecology and Evolution*, 3:803–807, 2012.

[54] David W. Bapst. A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods in Ecology and Evolution*, 4(8):724–733, August 2013.

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

[56] Olaf R P Bininda-Emonds, Marcel Cardillo, Kate E Jones, Ross D E Macphee, Robin M D Beck, Richard Grenyer, Samantha A Price, Rutger A Vos, John L Gittleman, and Andy Purvis. The delayed rise of present-day mammals. *Nature*, 446(7135):507–512, 2007.

[57] Barry W. Brook and David M. J. S. Bowman. The uncertain blitzkrieg of Pleistocene megafauna. *Journal of Biogeography*, 31(4):517–523, April 2004.

[58] Scott Chamberlain and Eduard Szocs. taxize - taxonomic search and retrieval in r. *F1000Research*, 2013.

[59] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[60] Michael Foote. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology*, 14(3):258–271, 1988.

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

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

[63] Andrew Gelman. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, pages 2865–2873, 2008.

[64] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall, Boca Raton, FL, 3 edition, 2013.

[65] Andrew Gelman and Jennifer Hill. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY, 2007.

[66] Harvey Goldstein, William Browne, and Jon Rasbash. Partitioning variation in multilevel models. *Understanding Statistics*, 1(4):1–12, 2002.

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

[68] Robert J. Hijmans. *raster: Geographic data analysis and modeling*, 2015. R package version 2.3-24.

[69] Matthew D. Hoffman and Andrew Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv*, 1111(4246), 2011.

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

[71] Joseph G Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. *Bayesian Survival Analysis*. Springer, New York, 2001.

[72] Christine M Janis, Gregg F Gunnell, and Mark D Uhen. *Evolution of Tertiary mammals of North America. Vol. 2. Small mammals, xenarthrans, and marine mammals*. Cambridge University Press, Cambridge, 2008.

[73] Christine 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.

[74] Jukka Jernvall and Mikael Fortelius. Maintenance of trophic structure in fossil mammal communities: site occupancy and taxon resilience. *American Naturalist*, 164(5):614–624, November 2004.

[75] Kate E Jones, Jon Bielby, Marcel Cardillo, Susanne A Fritz, Justin O'Dell, C David L Orme, Kamran Safi, Wes 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, Amber Teacher, Olaf R P Bininda-Emonds, John L Gittleman, Georgina M Mace, and Andy Purvis. PanTHERIA : a species-level database of life history , ecology , and geography of extant and recently extinct mammals. *Ecology*, 90(9):2648, 2009.

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

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

[78] Jonathan 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, February 2014.

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

[80] Samantha A Price, Samantha S B Hopkins, Kathleen K Smith, and V Louise 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.

[81] 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, March 2012.

[82] David M Raup. Taxonomic survivorship curves and Van Valen's Law. *Paleobiology*, 1(1):82–96, January 1975.

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

[84] Liam J. Revell. phytools: An r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3:217–223, 2012.

[85] Martin Rosvall, D Axelsson, and CT Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(14):13–24, 2009.

[86] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–23, January 2008.

[87] Holger Schielzeth. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2):103–113, February 2010.

[88] Christian A Sidor, Daril A Vilhena, Kenneth D Angielczyk, Adam K Huttenlocker, Sterling J Nesbitt, Brandon R Peecook, J Sébastien Steyer, Roger M H Smith, and Linda A Tsuji. Provincialization of terrestrial faunas following the end-Permian mass extinction. *Proceedings of the National Academy of Sciences*, 110(20):8129–33, May 2013.

[89] Felisa A Smith, James H Brown, John P Haskell, S Kathleen Lyons, John Alroy, Eric L Charnov, Tamar Dayan, Brian J Enquist, S K Morgan Ernest, Elizabeth A Hadly, Kate E Jones, Dawn M Kaufman, Pablo A Marquet, Brian A Maurer, Karl J Niklas, Warren P Porter, Bruce Tiffney, and Michael R Willig. Similarity of Mammalian Body Size across the Taxonomic Hierarchy and across Space and Time. *The American Naturalist*, 163(5):672–691, 2004.

[90] Stan Development Team. Stan: A c++ library for probability and sampling, version 2.5.0, 2014.

[91] Susumu Tomiya. Body Size and Extinction Risk in Terrestrial Mammals Above the Species Level. *The American Naturalist*, pages E000–E000, September 2013.

[92] Leigh Van Valen. A new evolutionary law. *Evolutionary Theory*, 1:1–30, 1973.

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

[94] Daril A Vilhena. *Boundaries and dynamics of biomes.* PhD thesis, University of Washington, 2013.

[95] Daril A Vilhena, Elisha B Harris, Carl T Bergstrom, Max E Maliska, Peter D Ward, Christian A Sidor, Caroline A E Strömberg, and Gregory P Wilson. Bivalve network reveals latitudinal selectivity gradient at the end-Cretaceous mass extinction. *Scientific reports*, 3:1790, May 2013.