Supplemental information for "Death and taxa"

Peter D Smits

1 Supertree inference

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.

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 [5]. Finally, this taxonomic information was further updated using a published taxonomy of fossil mammals [11, 12].

This taxonomy serves as an initial phylogenetic hypothesis which was then combined with a selection of species-level phylogenies [3, 17] in order to better constrain a minimum estimate of the actual phylogenetic relationships of the species. The supertree was inferred via matrix representation parsimony implemented in the `phytools` package for R [18]. 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 in order to minize stratigraphic gaps. 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 [1]. The minimum length is necessary to avoid zero-length branches which cause the phylogenetic covariance matrix not to be positive definite, which is important for computation (see below). While other time scaling approaches are possible [2, 10] this method was chosen for its simplicity and not requiring additional information about diversification rates which are the interest of this study.

2 Modeling censored observations

Censored data are modeled using the survival function of the distribution, $S(t)$, defined earlier for the Weibull distribution (Eq. 5, 6) with σ defined as above (Eq. 8, 9). $S(t)$ 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 [7].

The final sampling statement/likelihood for both uncensored and both right and left censored observations is then 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.

3 Deviance residuals

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 cumulative hazard function [13]. Given $S(t)$, 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.

4 Variance partitioning

I calculated VPC using a resampling approach based on [8]. 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 υ_c, the Weibull variance (Eq. 1) 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 υ_p, the Weibull variance (Eq. 1) of each element of σ^{p*} with α drawn from the posterior estimate.

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

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

The simulated values of h were drawn from a univariate normal distribution because each simulated value is in isolation, so there 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

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

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.

5 Widely applicable information criterion

WAIC can be considered fully Bayesian alternative to the Akaike information criterion, where WAIC acts as an approximation of leave-one-out cross-validation which acts as a measure of out-of-sample predictive accuracy [7]. The following explanation uses the "WAIC 2" formulation recommended by [7].

WAIC is calculated starting with the log pointwise posterior predictive density calculated as

$$\text{lppd} = \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\Theta^S)\right), \tag{2}$$

where n is sample size, S is the number posterior simulation draws, and Θ represents all of the estimated parameters of the model. This is similar to calculating the likelihood of each observation given the entire posterior.

A correction for the effective number of parameters is then added to lppd to adjust for overfitting. The effective number of parameters is calculated, following derivation and recommendations of [7], as

$$p_{\text{WAIC}} = \sum_{i=1}^{n} V_{s=1}^{S}(\log p(y_i|\Theta^S)). \tag{3}$$

where V is the sample posterior variance of the log predictive density for each data point.

Given both equations 2 and 3, WAIC is then calculated

$$\text{WAIC} = \text{lppd} - p_{\text{WAIC}}. \tag{4}$$

When comparing two or more models, lower WAIC values indicate better out-of-sample predictive accuracy. Importantly, WAIC is just one way of comparing models. When combined with posterior predictive checks it is possible to get a more complete understanding of model fit.

6 Results from posterior predictive checks

With all marginal posterior estimates having converged ($\hat{R} < 1.1$) it is possible to examine the quality of model fit (Table 1). If the model is an adequate descriptor of the observed data, then relatively confident inference can be made [7].

Visual examination of the deviance residuals from twelve different sets of posterior predictive simulations indicates a systematic weakness estimating durations greater than 3 2-My bins (Fig. S1). However, posterior predictive estimates of the 25th, 50th, and 75th quantiles of the observed indicate adequate fit. (Fig. S2). Importantly, this indicates that the model has approximate fit for 50+% of the data. Because, the inferred model can be inferred to be approximately adequate at capturing the observed variation.

The Wiebull model (6140.37) also had a much lower WAIC score than the Exponential model (16697.35). This large a difference indicates that the Weibull model probably has the lower out-of-sample predictive accuracy of the two.

7 Concerns surrounding estimates of α

The estimate of the Weibull shape parameter, α, is greater than 1 meaning that extinction risk is expected to increase with taxon age (Table 1). As the value of α is between 1 and 1.5, extinction risk for a given species only gradually increases with age (Fig. S3). There are three possible explanations for this result: 1) older taxa being out competed by younger taxa [22]; or 2) this is an artifact of the minimum resolution of the fossil record [19].

An additional concern is that there may be an upward bias in estimates of α at this sample size, similar to that for scale parameters [7]. The plausibility of third possibility in this example can be explored in simulation. I simulated from

10, 100, 1000, and 10000 samples from a Weibull($alpha = 1.3$, $\sigma = 1$) 100 times each. For each of these simulated datasets, I then estimated the values of α and σ in a simple maximum likelihood context in order to just get the model estimate. The modal estimates of both parameters for the simulated datasets were then compared to the known values (Fig. S4). The results from these simulations demonstrate that the estimates of α in the above analyses (Table 1) should not be particularity biased based on my sample size of approximately 2000 species.

The model used in this analysis, however, is unable to distinguish between the remaining two hypotheses [19, 22]. Further work on how to better constrain estimates α is necessary. A possibly is somehow incorporating these hypotheses as prior information.

8 Supplementary figures

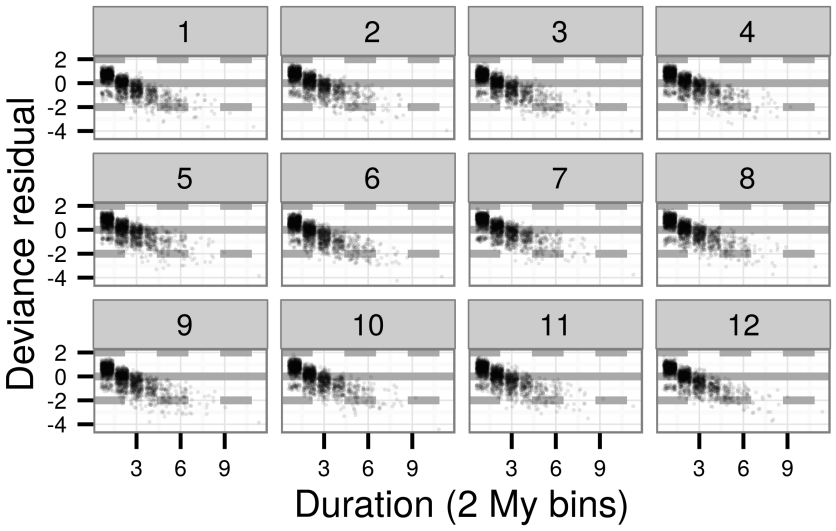

Figure S1: Deviance residuals from the fitted survival model compared to observed durations. Each graph depicts the residuals from single draws from the posterior distributions of all estimated parameters. Positive values indicate an underestimate of the observed duration, while negative values indicate an overestimate of the observed duration. A small amount of noise is added to each point to increase clarity. Twelve different examples are provided here to indicate consistency across multiple realizations.

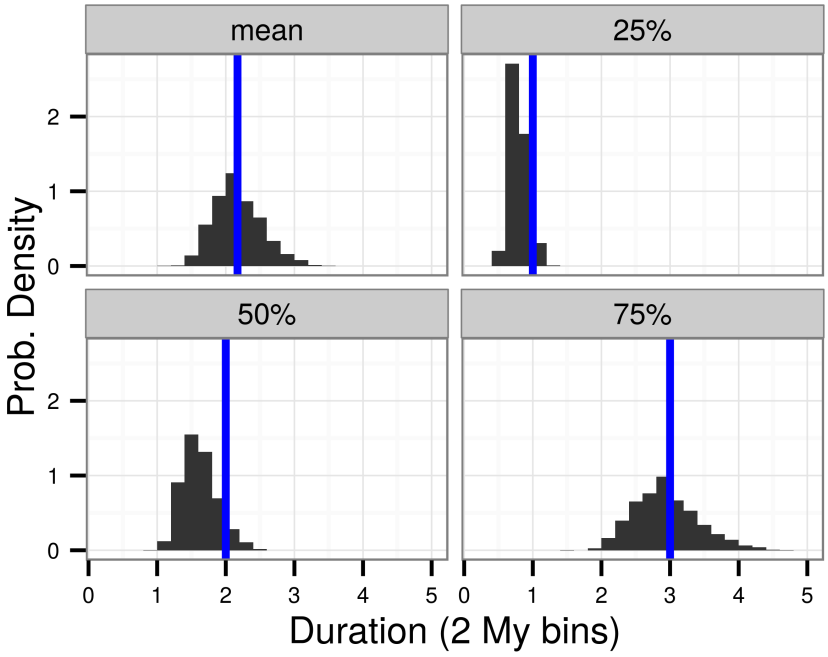

Figure S2: The results of additional posterior predictive checks for four summaries of the observed durations, as labeled. Blue vertical lines indicate the observed value. None of the observed values are significantly different from the posterior predictive distributions.

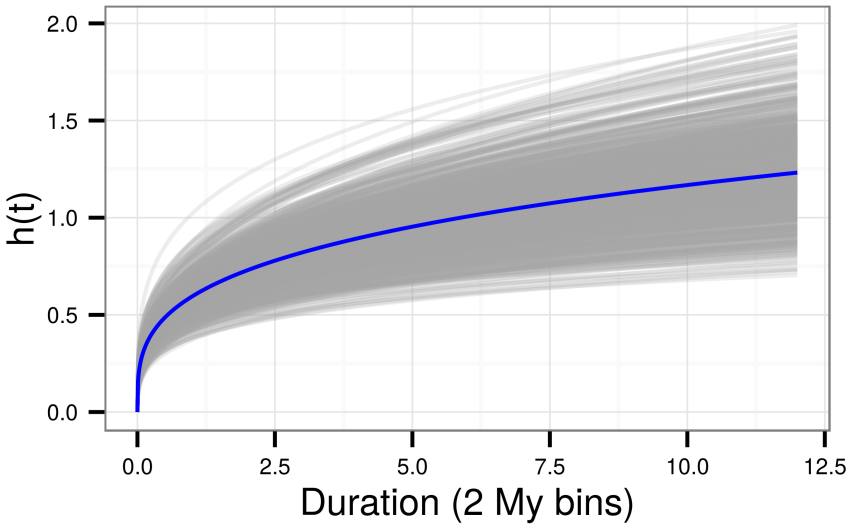

Figure S3: 1000 estimates of the hazard function ($h(t)$ for the observed species mean (grey), along with the median estimated hazard function (blue). $h(t)$ is an estimate of the rate at which a species of age t is expected to go extinct. Hazard functions were estimated from random draws from the estimated posterior distributions and evaluated with all covariate information set to 0, which corresponds to the expected duration of the mean species.

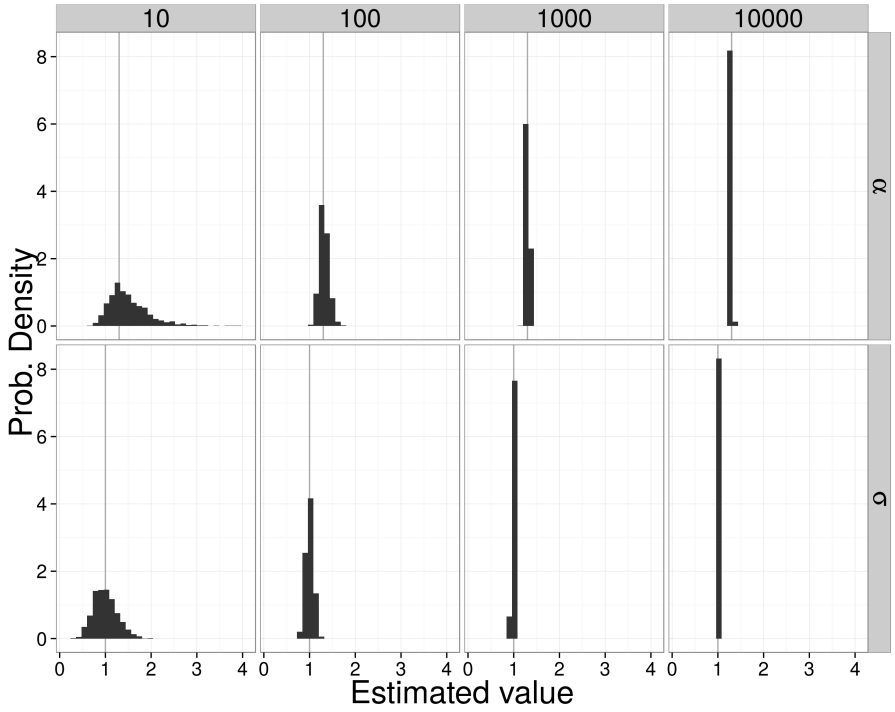

Figure S4: Comparison of maximum likelihood estimates of shape (α) and scale (σ) parameters from 1000 simulated data sets from 4 different sample sizes. Vertical lines are the actual parameter value used to generate the data. When sample size is approximately 100 or greater, estimates are not overly biased.

9 Supplementary tables

Table S1: 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.

Table S2: Regression equations used in this study for estimating body size. Equations are presented with reference to taxonomic grouping, part name, and reference.

Group	Equation	log(Measurement)	Source
General	$\log(m) = 1.827x + 1.81$	lower m1 area	[14]
General	$\log(m) = 2.9677x - 5.6712$	mandible length	[6]
General	$\log(m) = 3.68x - 3.83$	skull length	[15]
Carnivores	$\log(m) = 2.97x + 1.681$	lower m1 length	[21]
Insectivores	$\log(m) = 1.628x + 1.726$	lower m1 area	[4]
Insectivores	$\log(m) = 1.714x + 0.886$	upper M1 area	[4]
Lagomorph	$\log(m) = 2.671x - 2.671$	lower toothrow area	[20]
Lagomorph	$\log(m) = 4.468x - 3.002$	lower m1 length	[20]
Marsupials	$\log(m) = 3.284x + 1.83$	upper M1 length	[9]
Marsupials	$\log(m) = 1.733x + 1.571$	upper M1 area	[9]
Rodentia	$\log(m) = 1.767x + 2.172$	lower m1 area	[14]
Ungulates	$\log(m) = 1.516x + 3.757$	lower m1 area	[16]
Ungulates	$\log(m) = 3.076x + 2.366$	lower m2 length	[16]
Ungulates	$\log(m) = 1.518x + 2.792$	lower m2 area	[16]
Ungulates	$\log(m) = 3.113x - 1.374$	lower toothrow length	[16]

References

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

[2] 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.

[3] 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.

[4] Jonathan I Bloch, Kenneth D Rose, and Philip D Gingerich. New species of Batodonoides (Lipotyphla, Geolabididae) from the Early Eocene of Wyoming: smallest known mammal? *Journal of Mammalogy*, 79(3):804–827, 1998.

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

[6] John R Foster. Preliminary body mass estimates for mammalian genera of the Morrison Formation (Upper Jurassic, North America). *PaleoBios*, 28:114–122, 2009.

[7] 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.

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

[9] Cynthia L Gordon. A First Look at Estimating Body Size in Dentally Conservative Marsupials. *Journal of Mammalian Evolution*, page 21, 2003.

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

[11] 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.

[12] 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.

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

[14] Serge Legendre. Analysis of mammalian communities from the Late Eocene and Oligocene of Southern France. *Paleovertebrata*, 16(4):191–212, 1986.

[15] Zhe-Xi Luo, Alfred W Crompton, and Ai-Lin Sun. A New Mammaliaform from the Early Jurassic and Evolution of Mammalian Characteristics. *Science*, 292:1535–1540, 2001.

[16] M. Mendoza, C. M. Janis, and P. Palmqvist. Estimating the body mass of extinct ungulates: a study on the use of multiple regression. *Journal of Zoology*, 270:90–101, May 2006.

[17] 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.

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

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

[20] Susumu Tomiya. Body Size and Extinction Risk in Terrestrial Mammals Above the Species Level. *The American Naturalist*, 182:196–214, September 2013.

[21] Blair Van Valkenburgh. Skeletal and dental predictors of body mass in carnivores. In John Damuth and Bruce J Macfadden, editors, *Body size in mammalian palebiology: estimation and biological implications*, pages 181–205. Cambridge University Press, Cambridge, 1990.

[22] Peter J Wagner and George F Estabrook. Trait-based diversification shifts reflect differential extinction among fossil taxa. *Proceedings of the National Academy of Sciences*, 111:16419–16424, October 2014.