

S1 Supplement to Materials and Methods

S1.1 Data Specifications

S1.1.1 Binning fossil occurrences

The estimated age of each occurrence is based on the core-specific age-model that observation is from and can be overly precise. To alleviate this overprecision, we coarsened our temporal information in an effort to limit the effects of between-core heterogeneity in age. The occurrence histories of each species was then summarized as a series of binary codes indicating the presence or last occurrence of that species. For every occurrence of a species, except the last, that species existence and survival is recorded as a 0. The last occurrence of that species is considered the bin in which the taxon has gone extinct – and is recorded as 1. This protocol means that we are reading the fossil record “as written,” a practice that is potentially dangerous as it is an overconfident statement of preservation and may be shortening the actual durations of the studied species [1–10]. However, this practice is common with marine microfossil data due to their exceptional preservation rate [11–14]. In fact, with marine microfossils collected from cores a bigger problem may be over extending the duration of a species due to mixing and smearing within the cores [15–18].

S1.1.2 Covariate transformation and standardization

Prior to analysis, geographic range was then log-plus-one transformed and standardized by mean-centering the data and then dividing by the standard deviation of the distribution of geographic ranges. This standardization means that a regression coefficient associated with each covariate describes the change in extinction probability per change in standard deviation of that covariate, that coefficients associated with similarly standardized covariates will be directly comparable in magnitude, and that the intercept term corresponds to the expected value of the outcome at when geographic range is its average value [19]. Change in geographic range between observations was measured from the standardized geographic range values and was not standardized separately.

Temperature was also transformed and standardized the in the same manner as geographic range. The change in temperature between an observation and its previous observation was measured from the standardized temperature values and was not standardized separately.

S1.2 Model Specifications

In survival analysis, the hazard function describes the instantaneous rate of extinction of a species given its age and covariate information. The hazard function is defined as the conditional probability of a species going extinct by the end of the t -th interval given that it survived up until t and the relevant covariate information X for all k 1 My intervals [20]. For

the discrete time intervals $T = 1, \dots, k$, extinction is defined as $T = t$. The discrete time hazard function is defined as

$$\lambda(t|X) = P(T = t|T \geq t, X), \quad t = 1, \dots, k. \quad (\text{S1})$$

The hazard function (Eq. S1) is easily reparameterized as a logistic regression by defining that $\lambda(t|X) = h(\Theta)$ where $h(\cdot)$ is a logit inverse-link function and Θ is the probability of a taxon going extinction during interval t [20]. $h(\Theta)$ is then modeled as with any regression. In this case, we opted for a hierarchical/mixed-effects model with multiple non-nested varying intercepts and slopes [19].

Our covariates matrix X is a $N \times D$ matrix where N is the total number of observations and D is the total number of covariates. The first column of X is entirely 1's as it corresponds to the intercept term in the regression model. The next two columns of X are two aspects of geographic range as continuous covariates: geographic range r during interval t , and the difference d between the geographic range at $t - 1$ and t . Change in geographic range was calculated from the transformed and standardized geographic range values; this means that change in geographic range is in units of changes in standard deviations. The final two columns are two aspects of global temperature: mean temperature during interval t , and the lag of mean temperature (i.e. mean temperature during interval $t - 1$.) As with change to geographic range, the lag of temperature is based on the transformed and standardized temperature estimates.

The matrix of time and phylum varying regression coefficients describing the effects of the covariates on a species' risk of extinction is called B – a w by p matrix, where w is the number of time temporal intervals and p is the number of phyla. The elements of this matrix, the regression coefficients, are themselves modeled as being multivariate normally distributed with vector of means α describing the average intercept and regression coefficient estimates of each phylum p . These phylum averages are themselves modeled as multivariate normally distributed with mean vector μ describing the overall average regression coefficients, including the intercept. μ has length D and is ordered intercept, range coefficient, change in range coefficient, temperature coefficient, temperature lag coefficient.

The effect of species age on the log-odds of species extinction is modeled as a non-nested random intercept A [20]. This term describes how the log-odds of extinction varies along a species duration, and how this effect can differ between the phyla. A is a l by p matrix, where l is the age at observation of a species and p is its phylum. A is modeled as following a multivariate normal distribution with phylum means being the vector δ and covariance matrix Σ_A . The covariation between the elements of vector δ are modeled as a multivariate normal distribution with a mean vector of all 0s and covariance matrix Σ_δ .

To complete the generative model, we need to assign final priors to the “top-level” parameters. In general, we favored weakly informative priors which help regularize our estimates. In the case of a regression coefficient, this means a Normal distribution with mean 0 and a standard deviation of 3. For our scale parameters (e.g. standard deviations), we used half-Cauchy distributed priors with heavy tails but the majority of probability density near 0.

Our top-level intercept was given a more diffuse prior than our regression coefficients, which reflects our greater degree of uncertainty about its value. Our top-level regression coefficient for the effect of geographic range was given an informative prior reflecting the overwhelming amount of evidence that species with a larger than average geographic range have a lower risk of extinction than species with an average or less than average geographic range. In the context of this analysis, this means that we are again using a weakly informative prior but instead of centering the density around -1 (i.e. larger than average geographic range decreases extinction risk).

Instead of assigning a prior distribution for each of the covariance matrices in the model, we instead decomposed the covariance matrices (e.g. Σ_B) which allows us to assign independent priors for the scale and correlation aspects of covariance. The scale parameters were assigned half-Cauchy priors as described above in the context of all other scale parameters. The correlation matrices were assigned LKJ priors each with shape parameter set to 1. This choice of shape parameter produces a uniform distribution over possible correlation matrices. These priors are also slightly more interpretable than other common prior distributions for covariance matrices such as the inverse-Wishart distribution. This approach to assigning priors to a covariance matrix is recommended by the Stan Manual [21].

In total, our model can be expressed as:

$$\begin{aligned}
t_i &\sim \text{Bernoulli}(\Theta) \\
\Theta_i &= \text{logit}^{-1}(X_i B_{w[i],p[i]} + A_{l[i],p[i]}) \\
B_{w,p} &\sim \text{MVN}(\alpha_p, \Sigma_B) \\
\alpha_p &\sim \text{MVN}(\mu, \Sigma_\alpha) \\
A_{l,p} &\sim \text{MVN}(\delta_p, \Sigma_A) \\
\delta_p &\sim \text{N}(0, \sigma_\delta) \\
\mu_d &\sim \begin{cases} \text{N}(-2, 5) & \text{if } d = \text{intercept} \\ \text{N}(-1, 1) & \text{if } d = \text{geo. range} \\ \text{N}(0, 1) & \text{else} \end{cases} \\
\delta &\sim \text{N}(0, 1) \\
\Sigma_B &= \text{diag}(\tau_B) \Omega_B \text{diag}(\tau_B) \\
\Sigma_\alpha &= \text{diag}(\tau_\alpha) \Omega_\alpha \text{diag}(\tau_\alpha) \\
\Sigma_A &= \text{diag}(\tau_A) \Omega_A \text{diag}(\tau_A) \\
\tau_B &\sim C^+(1) \\
\tau_\alpha &\sim C^+(1) \\
\tau_A &\sim C^+(1) \\
\Omega_B &\sim \text{LKJ}(1) \\
\Omega_\alpha &\sim \text{LKJ}(1) \\
\Omega_A &\sim \text{LKJ}(1)
\end{aligned} \tag{S2}$$

with i indexing the observation and bracket subscripts referencing the class of the i th

observation where $w[i]$ is the time of the i -th observation, $p[i]$ is the phylum of the i -th observation, and $d[i]$ is the age of the i -th observation.

S1.3 Model Parameter Estimation

We implemented our model (Eq. S2 using the `rstanarm` package for the R programming language [21]. This package provides an interface to the Stan probabilistic programming language for writing hierarchical/mixed-effects models in native R. Posterior estimates were obtained through Hamiltonian Monte Carlo, using 2000 steps divided equally between warm-up and sampling. In order to prevent divergent transitions adapt delta was increased to 0.999999; all other HMC/NUTS sampling parameters were kept at the defaults for `rstanarm` 2.18.2 [22].

To implement our VP model in `rstanarm`, where “data” is a `data.frame` object of all necessary data (response, covariates), is coded as:

```
form <- event ~ range + range_diff + temp + temp_lag +
  (1 + range + range_diff + temp + temp_lag | mybin/phylum) +
  (1 | age/phylum),
stan_glmer(formula = form,
  data = data,
  family = 'binomial',
  prior = normal(c(-1, 0, 0, 0), rep(1, 4), autoscale = FALSE),
  prior_intercept = normal(-2, 5, autoscale = FALSE),
  prior_aux = cauchy(0, 1, autoscale = FALSE),
  chains = 4,
  thin = 4,
  adapt_delta = 0.999999)
```

Similarly, our VP model can be implemented using the `brms` Stan interface [23, 24] as:

```
priors <- c(set_prior('normal(-2, 5)', class = 'Intercept'),
  set_prior('normal(0, 1)', class = 'b'),
  set_prior('normal(-1, 1)', class = 'b', coef = 'range'),
  set_prior('cauchy(0, 1)', class = 'sd'),
  set_prior('lkj(1)', class = 'cor'))
form <- bf(event ~ range + range_diff + temp + temp_lag +
  (1 + range + range_diff + temp + temp_lag | mybin/phylum) +
  (1 | age/phylum))
brmfit <- brm(formula = form,
  data = data,
  family = bernoulli(),
  prior = priors,
  chains = 4,
  thin = 4,
  control = list(adapt_delta = 0.999999))
```

Posterior convergence was determined using the general and HMC-specific diagnostic criteria: scale reduction factor (\hat{R} ; target < 1.1), effective sample size (eff; target value eff/steps < 0.0001), number of samples that saturated the maximum trajectory length for avoiding infinite loops (treedepth; target value 0), sample divergence, and the energy Bayesian Fraction of Mission Information (E-BFMI; target value > 0.2). For further explanation of these diagnostic criteria, see the Stan Manual [21].

Supplementary References

- [1] J. Alroy. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. *Quantitative methods in paleobiology. Paleontological Society Papers*. 16 (2010), 55–80.
- [2] J. Alroy. New methods for quantifying macroevolutionary patterns and processes. *Paleobiology* 26 (4) (2000), 707–733.
- [3] J. Alroy. Accurate and precise estimates of origination and extinction rates. *Paleobiology* 40 (03) (2014), 374–397.
- [4] M. Foote. Estimating Taxonomic Durations and Preservation Probability. *Paleobiology* 23 (3) (1997), 278–300.
- [5] M. Foote and J. J. Sepkoski. Absolute measures of the completeness of the fossil record. *Nature* 398 (6726) (1999), 415–417.
- [6] M. Foote. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. *Paleobiology* 27 (4) (2001), 602–630.
- [7] M. Foote and D. M. Raup. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 22 (2) (1996), 121–140.
- [8] G. T. Lloyd, P. N. Pearson, J. R. Young, and A. B. Smith. Sampling bias and the fossil record of planktonic foraminifera on land and in the deep sea. *Paleobiology* 38 (4) (2012), 569–584.
- [9] C. R. Marshall. Distinguishing between sudden and gradual extinctions in the fossil record: predicting the position of the Cretaceous-Tertiary iridium anomaly using the ammonite fossil record on Seymour Island, Antarctica. *Geology* 23 (8) (1995), 731–734.
- [10] S. C. Wang and C. R. Marshall. Estimating times of extinction in the fossil record. *Biology Letters* 12 (4) (2016), 20150989.
- [11] T. H. G. Ezard, G. H. Thomas, and A. Purvis. Inclusion of a near-complete fossil record reveals speciation-related molecular evolution. *Methods in Ecology and Evolution* 4 (8) (2013), 745–753.
- [12] T. H. G. Ezard and A. Purvis. Environmental changes define ecological limits to species richness and reveal the mode of macroevolutionary competition. *Ecology Letters* (2016), 1–8.
- [13] T. H. G. Ezard, T. Aze, P. N. Pearson, and A. Purvis. Interplay Between Changing Climate and Species’ Ecology Drives Macroevolutionary Dynamics. *Science* 332 (6027) (2011), 349–351.
- [14] L. H. Liow, H. J. Skaug, T. Ergon, and T. Schweder. Global occurrence trajectories of microfossils: environmental volatility and the rise and fall of individual species. *Paleobiology* 36 (2) (2010), 224–252.

- [15] F. Mekik and R. Anderson. Is the core top modern? Observations from the eastern equatorial Pacific. *Quaternary Science Reviews* 186 (2018), 156–168.
- [16] W. Broecker, K. Matsumoto, E. Clark, I. Hajdas, and G. Bonani. Radiocarbon age differences between coexisting foraminiferal species. *Paleoceanography* 14 (4) (1999), 431–436.
- [17] F. Mekik. Radiocarbon dating of planktonic foraminifer shells: A cautionary tale. *Paleoceanography* 29 (1) (2014), 13–29.
- [18] T.-H. Peng and W. S. Broecker. The impacts of bioturbation on the age difference between benthic and planktonic foraminifera in deep sea sediments. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 35 (1984), 346–352.
- [19] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press, 2006.
- [20] G. Tutz and M. Schmid. *Modeling discrete time-to-event data*. Springer International Publishing, 2016.
- [21] Stan Development Team. Stan Modeling Language Users Guide and Reference Manual. 2017.
- [22] J. Gabry and B. Goodrich. *rstanarm: Bayesian Applied Regression Modeling via Stan*. R package version 2.18.2. 2018.
- [23] P.-C. Bürkner. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80 (1) (2017), 1–28.
- [24] P.-C. Bürkner. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10 (1) (2018), 395–411.