

Title: How do biological traits affect brachiopod taxonomic
2 survival? A hierarchical Bayesian approach.

Running title: How do biological traits affect taxonomic survival?

4 Author: Peter D Smits, psmits@uchicago.edu, Committee on Evolutionary
Biology, University of Chicago

6 Keywords: extinction, macroevolution, macroecology, Paleozoic, selection

Word count: ?

8 Table count: 0.

Figure count: 5 main text, 3 supplement.

10 Data archival location: If accepted, all data and code necessary to duplicate
this analysis will be made available on DRYAD.

12

Abstract

While the effect of geographic range on extinction risk is well documented, the effects of other traits are less well known. Here, I analyze patterns of Paleozoic brachiopod genus durations and their relationship to geographic range, affinity for epicontinental seas versus open ocean environments, and body size. Additionally, I allow for environmental affinity to have a nonlinear effect on duration. Using a hierarchical Bayesian modeling approach, I also model the possible interaction between the effects of the biological traits and a taxon's time of origination. I find evidence that as extinction risk increases, the expected strength of the selection gradient on biological traits (except body size) increases. This manifests as greater expected differences in extinction risk for each unit change in geographic range and environmental preference during periods of high extinction risk, as opposed to a much flatter expected selection gradient during periods of low extinction risk. I find weak evidence for a universally nonlinear relationship between environmental preference and extinction risk such that "generalists" have a lower expected extinction risk than either "specialists". While for the majority of the Paleozoic this hypothesis is strongly supported, there are many times where this hypothesized relationship is absent or even reversed. Importantly, I find that as extinction risk increases, the peakedness of this relationship is expected to increase as well. These results demonstrate the importance of directly modeling the structure inherent in the observed data as a means to better understand which processes may have been driving the observed patterns of diversification.

1 Introduction

38 How do biological traits affect extinction risk? Jablonski (1986) observed that
during periods of high extinction risk, the effects of biological traits on survival
40 decreased in size. However, this pattern was weakest/absent in the effect of
geographic range on survival (Jablonski, 1986). Biological traits are defined here
42 as descriptors of a taxon's adaptive zone, which is the set of biotic–biotic and
biotic–abiotic interactions that a taxon can experience (Simpson, 1944). In
44 effect, these are descriptors of a taxon's broad-sense ecology.

Jablonski (1986) phrased his conclusions in terms of background versus mass
46 extinction, but this scenario is readily transferable to a continuous variation
framework as there is no obvious distinction in terms of extinction rate between
48 these two states (Wang, 2003). Additionally, the Jablonski (1986) scenario has
strong model structure requirements in order to test its proposed
50 macroevolutionary mechanism; not only do the taxon trait effects need to be
modeled, but the relationships between these effects need to be modeled as well.

52 There are two end-member macroevolutionary mechanisms which may underlie
the pattern observed by Jablonski (1986): the effect of geographic range on
54 predictive survival remains constant and those of other biological traits decrease,
and the effect of geographic range in predicting survival increases and those of
56 other biological traits stay constant. Reality, of course, may fall somewhere
along the continuum between these two opposites.

58 I model taxon durations because trait based differences in extinction risk should
manifest as differences in taxon durations. Namely, a taxon with a beneficial
60 trait should survive longer, on average, than a taxon without that beneficial
trait. Conceptually, taxon survival can be considered an aspect of “taxon fitness”
62 along with expected lineage specific branching/orignation rate (Cooper, 1984,

Palmer and Feldman, 2012). The analysis of taxon durations, or time from
64 origination to extinction, falls under the purview of survival analysis, a field of
applied statistics commonly used in health care (Klein and Moeschberger, 2003)
66 but has a long history in paleontology (Simpson, 1944, 1953, Van Valen, 1973,
1979).

68 Geographic range is widely considered the most important taxon trait for
estimating differences in extinction risk at nearly all times with large geographic
70 range associated with low extinction risk (Jablonski, 1986, 1987, Jablonski and
Roy, 2003, Payne and Finnegan, 2007). I expect this to hold true nearly always.

72 Miller and Foote (2009) demonstrated that during several mass extinctions taxa
associated with open-ocean environments tend to have a greater extinction risk
74 than those taxa associated with epicontinental seas. During periods of
background extinction, however, they found no consistent difference between
76 taxa favoring either environment. These two environment types represent the
primary environmental dichotomy observed in ancient marine systems (Miller
78 and Foote, 2009, Peters, 2008, Sheehan, 2001).

Epicontinental seas are a shallow-marine environment where the ocean has
80 spread over the surface of a continental shelf with a depth typically less than
100m. In contrast, open-ocean coastline environments have much greater
82 variance in depth, do not cover the continental shelf, and can persist during
periods of low sea level. Because of this, it is strongly expected that taxa which
84 favor epicontinental seas would be at great risk during periods of low sea levels,
such as during glacial periods, where these seas are drained. During the
86 Paleozoic, epicontinental seas were widely spread globally but declined over the
Mesozoic and eventually diminished without completely disappearing during the
88 Cenozoic as open-ocean coastlines became the dominant shallow-marine setting
(Johnson, 1974, Miller and Foote, 2009, Peters, 2008).

90 Given the above information, I predict that as extinction risk increases, taxa
associated with open-ocean environments should generally increase in extinction
92 risk versus those that favor epicontinental seas. Additionally, there is a possible
nonlinear relationship between environmental preference and taxon duration. A
94 long standing hypothesis is that generalists or unspecialized taxa will have
greater survival than specialists (Baumiller, 1993, Liow, 2004, 2007, Nürnberg
96 and Aberhan, 2013, 2015, Simpson, 1944). In this analysis I allowed for
environmental preference to possibly have a parabolic effect on taxon duration
98 Body size, measured as shell length (Payne et al., 2014), was also considered as
a potentially informative covariate. Body size is a proxy for metabolic activity
100 and other correlated life history traits (Payne et al., 2014). There is no strong
hypothesis of how body size effects extinction risk in brachiopods, meaning a
102 positive, negative, or zero effect are all plausible.

I adopt a hierarchical Bayesian survival modeling approach, which represents a
104 conceptual and statistical unification of the paleontological dynamic and cohort
survival analytic approaches (Baumiller, 1993, Foote, 1988, Raup, 1975, 1978,
106 Simpson, 2006, Van Valen, 1973, 1979). By using a Bayesian framework I am
able to quantify the uncertainty inherent in the estimates of the effects of
108 biological traits on survival, especially in cases where the covariates of interest
(i.e. biological traits) are themselves known with error.

110 **2 Materials and Methods**

2.1 Fossil occurrence information

112 The dataset analyzed here was sourced from the Paleobiology Database
(<http://www.paleodb.org>) which was then filtered based on taxonomic,

¹¹⁴ temporal, stratigraphic, and other occurrence information that was necessary
for this analysis. These filtering criteria are very similar to those from Foote and
¹¹⁶ Miller (2013) with an additional constraint of being present in the body size
data set from Payne et al. (2014). Epicontinental versus open-ocean assignments
¹¹⁸ for each fossil occurrence are partially based on those from Miller and Foote
(2009), with additional occurrences assigned similarly (Miller and Foote,
¹²⁰ personal communication).

Fossil occurrences were analyzed at the genus level which is common for
¹²² paleobiological, macroevolution, or macroecological studies of marine
invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2013, Kiessling
¹²⁴ and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,
Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).

¹²⁶ While species diversity dynamics are of much greater interest than those of
higher taxa, the nature of the fossil record makes accurate and precise
¹²⁸ taxonomic assignments at the species level for all occurrences. In particular, the
simplicity of brachiopod external morphology and the quality of preservation
¹³⁰ makes it very difficult for assignments below the genus level. As such, the choice
to analyze genera as opposed to species was in order to assure a minimum level
¹³² of confidence and accuracy in the data analyzed here.

Sampled occurrences were restricted to those with paleolatitude and
¹³⁴ paleolongitude coordinates, assignment to either epicontinental or open-ocean
environment, and belonging to a genus present in the body size dataset (Payne
¹³⁶ et al., 2014). Genus duration was calculated as the number of geologic stages
from first appearance to last appearance, inclusive. Genera with a last
¹³⁸ occurrence in or after Changhsingian stage were right censored at the
Changhsingian. Genera with a duration of only one stage were left censored
¹⁴⁰ (Appendix C). The covariates used to model genus duration were geographic

range size (r), environmental preference (v, v^2), and body size (m).

¹⁴² Geographic range was calculated using an occupancy approach. First, all occurrences were projected onto an equal-area cylindrical map projection. Each ¹⁴⁴ occurrence was then assigned to one of the cells from a 70×34 regular raster grid placed on the map. Each grid cell represents approximately 250,000 km². ¹⁴⁶ The map projection and regular lattice were made using shape files from <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans, ¹⁴⁸ 2015).

For each stage, the total number of occupied grid cells, or cells in which a fossil ¹⁵⁰ occurs, was calculated. Then, for each genus, the number of grid cells occupied by that genus was calculated. Dividing the genus occupancy by the total ¹⁵² occupancy gives the relative occupancy of that genus. Mean relative genus occupancy was then calculated as the mean of the per stage relative occupancies ¹⁵⁴ of that genus.

Body size data was sourced directly from Payne et al. (2014). Because those ¹⁵⁶ measurements are presented without error, a measurement error model similar to the one for environmental affinity could not be implemented (Appendix A).

¹⁵⁸ Prior to analysis, geographic range and body size were transformed and standardized in order to improve interpretability of the results. Geographic ¹⁶⁰ range, which can only vary between 0 and 1, was logit transformed. Body size, which is defined for all positive real values, was natural log transformed. These ¹⁶² covariates were then standardized by mean centering and dividing by two times their standard deviation following Gelman and Hill (2007).

¹⁶⁴ **2.2 Analytical approach**

Hierarchical modelling, sometimes called “mixed-effects modeling,” is a
¹⁶⁶ statistical approach which explicitly takes into account the structure of the
observed data in order to model both the within and between group variance
¹⁶⁸ (Gelman et al., 2013, Gelman and Hill, 2007). The units of study (e.g. genera)
each belong to a single grouping (e.g. origination cohort). These groups are
¹⁷⁰ considered draws from a shared probability distribution (e.g. all cohorts,
observed and unobserved). The group-level parameters are then estimated
¹⁷² simultaneously as the other parameters of interest (e.g. covariate effects)
(Gelman et al., 2013). The subsequent estimates are partially pooled together,
¹⁷⁴ where parameters from groups with large samples or effects remain large while
those of groups with small samples or effects are pulled towards the overall
¹⁷⁶ group mean.

This partial pooling is one of the greatest advantages of hierarchical modeling.
¹⁷⁸ By letting the groups “support” each other, parameter estimates then better
reflect our statistical uncertainty. Additionally, this partial pooling helps control
¹⁸⁰ for multiple comparisons and possibly spurious results as effects with little
support are drawn towards the overall group mean (Gelman et al., 2013,
¹⁸² Gelman and Hill, 2007).

All covariate effects (regression coefficients), as well as the intercept term
¹⁸⁴ (baseline extinction risk), were allowed to vary by group (origination cohort).
The covariance/correlation between covariate effects was also modeled. This
¹⁸⁶ hierarchical structure allows inference for how covariates effects may change
with respect to each other while simultaneously estimating the effects
¹⁸⁸ themselves, propagating our uncertainty through all estimates.

Additionally, instead of relying on point estimates of environmental affinity, I

¹⁹⁰ treat environmental affinity as a continuous measure of the difference between
the taxon's environmental occurrence pattern and the background occurrence
¹⁹² pattern (Appendix A).

2.3 Survival model

¹⁹⁴ Genus durations were modeled as time-till-event data (Klein and Moeschberger,
2003), with covariate information used in estimates of extinction risk as a
¹⁹⁶ hierarchical regression model. Genus durations were assumed to follow either an
exponential or Weibull distribution. THe use of either of these distributions
¹⁹⁸ makes assumptions about how duration may effect extinction risk (Klein and
Moeschberger, 2003). The exponential distribution assumes that extinction risk
²⁰⁰ is independent of duration. In contrast, the Weibull distribution allows for age
dependent extinction via the shape parameter α , though only as a monotonic
²⁰² function of duration. Importantly, the Weibull distribution is equivalent to the
exponential distribution when $\alpha = 1$.

²⁰⁴ The following variables are defined: y_i is the duration of genus i in geologic
stages, X is the matrix of covariates including a constant term, B_j is the vector
²⁰⁶ of regression coefficients for origination cohort j , Σ is the covariance matrix of
the regression coefficients, τ is the vector of scales the standard deviations of
²⁰⁸ the between-cohort variation in regression coefficient estimates, and Ω is the
correlation matrix of the regression coefficients.

²¹⁰ The exponential model is defined

$$y_i \sim \text{Exponential}(\lambda)$$

$$\lambda_i = \exp(\mathbf{X}_i B_{j[i]})$$

$$B \sim \text{MVN}(\vec{\mu}, \Sigma)$$

$$\Sigma = \text{Diag}(\vec{\tau}) \Omega \text{Diag}(\vec{\tau})$$

$$\mu_k \sim \begin{cases} \mathcal{N}(0, \psi_k \nu) & \text{if } k \neq r, \text{ or} \\ \mathcal{N}(-1, 1) & \text{if } k = r \end{cases} \quad (1)$$

$$\tau_k \sim C^+(1)$$

$$\psi_k \sim C^+(1) \text{ if } k \neq r$$

$$\nu \sim C^+(1)$$

$$\Omega \sim \text{LKJ}(2).$$

Similarly, the Weibull model is defined

$$\begin{aligned}
y_i &\sim \text{Weibull}(\alpha, \sigma) \\
\sigma_i &= \exp\left(\frac{-(\mathbf{X}_i B_{j[i]})}{\alpha}\right) \\
B &\sim \text{MVN}(\vec{\mu}, \Sigma) \\
\Sigma &= \text{Diag}(\vec{\tau}) \Omega \text{Diag}(\vec{\tau}) \\
\alpha &\sim C^+(2) \\
\mu_k &\sim \begin{cases} \mathcal{N}(0, \psi_k \nu) & \text{if } k \neq r, \text{ or} \\ \mathcal{N}(-1, 1) & \text{if } k = r \end{cases} \\
\tau_k &\sim C^+(1) \\
\psi_k &\sim C^+(1) \text{ if } k \neq r \\
\nu &\sim C^+(1) \\
\Omega &\sim \text{LKJ}(2).
\end{aligned} \tag{2}$$

²¹² The principal difference between this model and the previous (Eq. 1) is the
²¹³ inclusion of the shape parameter α . Note that σ is approximately equivalent to
²¹⁴ $1/\lambda$.

²¹⁵ For an explanation of how this model was developed, parameter explanations,
²¹⁶ and choice of priors, please see Appendix B. Note that these models (Eq. 1, 2)
²¹⁷ do not include how the uncertainty in environmental affinity is included nor how
²¹⁸ censored observations are included. For an explanation of both of these aspects,
see Appendices A and C.

220 **2.4 Parameter estimation**

The joint posterior was approximated using a Markov-chain Monte Carlo
222 routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic
224 programming language Stan (Stan Development Team, 2014a). The posterior
distribution was approximated from four parallel chains run for 10,000 draws
226 each, split half warm-up and half sampling and thinned to every 10th sample for
a total of 5000 posterior samples. Chain convergence was assessed via the scale
228 reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$) indicate approximate
convergence. Convergence means that the chains are approximately stationary
230 and the samples are well mixed (Gelman et al., 2013).

2.5 Model evaluation

232 Models were evaluated using both posterior predictive checks and an estimate of
out-of-sample predictive accuracy. The motivation behind posterior predictive
234 checks as tools for determining model adequacy is that replicated data sets
using the fitted model should be similar to the original data (Gelman et al.,
236 2013). Systematic differences between the simulations and observations indicate
weaknesses of the model fit. An example of a technique that is very similar
238 would be inspecting the residuals from a linear regression.

The strategy behind posterior predictive checks is to draw simulated values
240 from the joint posterior predictive distribution, $p(y^{rep}|y)$, and then compare
those draws to the empirically observed values (Gelman et al., 2013). To
242 accomplish this, for each replicate, a single value is drawn from the marginal
posterior distributions of each regression coefficient from the final model as well
244 as α for the Weibull model (Eq. 1, 2). Then, given the covariate information \mathbf{X} ,

a new set of n genus durations are generated giving a single replicated data set
246 y^{rep} . This is repeated 1000 times in order to provide a distribution of possible
values that could have been observed given the model.

248 In order to compare the fitted model to the observed data, various graphical
comparisons or test quantities need to be defined. The principal comparison
250 used here is a comparison between non-parameteric approximation of the
survival function $S(t)$ as estimated from both the observed data and each of the
252 replicated data sets. The purpose of this comparison is to determine if the
model approximates the same survival/extinction pattern as the original data.

254 The exponential and Weibull models were compared for out-of-sample predictive
accuracy using the widely-applicable information criterion (WAIC) (Watanabe,
256 2010). Out-of-sample predictive accuracy is a measure of the expected fit of the
model to new data. However, because the Weibull model reduces to the
258 exponential model when $\alpha = 1$ my interest is not in choosing between these
models. Instead, comparisons of WAIC values are useful for better
260 understanding the effect of model complexity on out-of-sample predictive
accuracy. The calculation of WAIC used here corresponds to the “WAIC 2”
262 formulation recommended by Gelman et al. (2013). For an explanation of how
WAIC is calculated, see Appendix D. Lower values of WAIC indicate greater
264 expected out-of-sample predictive accuracy than higher values.

3 Results

266 As stated above, posterior approximations for both the exponential and Weibull
models achieved approximate stationarity after 10,000 steps, as all parameter
268 estimates have an $\hat{R} < 1.1$.

Comparisons of the survival functions estimated from 1000 posterior predictive
270 data sets to the estimated survival function of the observed genera demonstrates
that both the exponential and Weibull models approximately capture the
272 observed pattern of extinction (Fig. 1). The major difference in fit between the
two models is that the Weibull model has a slightly better fit for longer lived
274 taxa than the exponential model.

Additionally, the Weibull model is expected to have slightly better out-of-sample
276 predictive accuracy when compared to the exponential model (WAIC 4576
versus 4604, respectively). 1). Because the difference in WAIC between these
278 two models is large, while results from both the exponential and Weibull models
will be presented, only those from the Weibull model will be discussed.

280 Estimates of the overall mean covariate effects μ can be considered
time-invariant generalizations for brachiopod survival during the Paleozoic (Fig.
282 1). Consistent with prior expectations, geographic range size has a negative
effect on extinction risk, where genera with large ranges having greater
284 durations than genera with small ranges.

I find that while the mean estimate of the effect of body size on extinction risk
286 is negative, implying that increasing body size decreases extinction risk, this
estimate is within 2 standard deviations of 0 (mean $\mu_m = -0.09$, standard
288 deviation 0.09; Fig. 1). Because of this, I infer that body size has no
distinguishable effect on brachiopod taxonomic survival.

290 Interpretation of the effect of environmental preference v on duration is slightly
more involved. Because a quadratic term is the equivalent of an interaction
292 term, both μ_v and μ_{v^2} have to be interpreted together because it is illogical to
change values of v without also changing values v^2 . To determine the nature of
294 the effect of v on duration I calculated the multiplicative effect of environmental

preference on extinction risk.

²⁹⁶ Given mean estimated extinction risk $\tilde{\sigma}$, we can define the extinction risk multiplier of an observation with environmental preference v_i as

$$\frac{\tilde{\sigma}_i}{\tilde{\sigma}} = f(v_i) = \exp\left(\frac{-(\mu_v v_i + \mu_{v^2} v^2)}{\alpha}\right). \quad (3)$$

²⁹⁸ This function $f(v_i)$ has a y-intercept of $\exp(0)$ or 1 because it does not have a non-zero intercept term. Equation 3 can be either concave up or down. A ³⁰⁰ concave down $f(v_i)$ may indicate that genera of intermediate environmental preference have greater durations than either extreme, and *vice versa* for ³⁰² concave up function.

The expected effect of environmental preference as a multiplier of expected ³⁰⁴ extinction risk can then be visualized (Fig. 2). This figure depicts 1000 posterior predictive estimates of Eq. 3 across all possible values of v . The number ³⁰⁶ indicates the posterior probability that the function is concave down, with generalists having lower extinction risk/greater duration than either type of ³⁰⁸ specialist. Note that the inflection point/optimum of Fig. 2 is approximately $x = 0$, something that is expected given the estimate of μ_v (Fig. 1).

³¹⁰ The matrix Σ describing the covariance between the different coefficients describes how these coefficients might vary together across the origination ³¹² cohorts. Similar to how this was modeled (Eq. 1, 2), for interpretation purposes Σ can be decomposed into a vector of standard deviations $\vec{\tau}$ and a correlation ³¹⁴ matrix Ω .

The estimates of the standard deviation of between-cohort coefficient estimates ³¹⁶ τ indicate that some effects can vary greatly between-cohorts (Fig. 1). Coefficients with greater values of τ have greater between-cohort variation. The ³¹⁸ covariate effects with the greatest between origination cohort variation are β_r ,

β_v , and β_{v^2} . Estimates of β_m have negligible between cohort variation, as there
320 is less between cohort variation than the between cohort variation in baseline
extinction risk β_0 . However the amount of between cohort variation in estimates
322 of β_{v^2} means that it is possible for the function describing the effect of
environmental affinity to be upward facing for some cohorts (Eq. 3), which
324 corresponds to environmental generalists being shorter lived than specialists in
that cohort.

326 The correlation terms of Ω (Fig. 3a) describe the relationship between the
coefficients and how their estimates may vary together across cohorts. The
328 correlations between the intercept term β_0 and the effects of the taxon traits are
of particular interest for evaluating the Jablonski (1986) scenario (Fig. 3a first
330 column/last row). Keep in mind that when β_0 is low, extinction risk is low; and
conversely, when β_0 is high, then extinction risk is high.

332 Marginal posterior probabilities of the correlations between the level of baseline
extinction risk β_0 and the effects of the taxon traits indicate that the correlation
334 between expected extinction risk and both geographic range β_r and β_{v^2} are of
particular note (Fig. 3b).

336 There is approximately a 98% probability that β_0 and β_r are negatively
correlated (Fig. 3b), meaning that as extinction risk increases, the
338 effect/importance of geographic range on genus duration increases. This means
that increases in baseline extinction rate are correlated with an increased
340 importance of geographic range size. There is a 93% probability that β_0 and β_{v^2}
are negatively correlated (Fig. 3b), meaning that as extinction risk increases,
342 the peakedness of $f(v_i)$ increases and the relationship tends towards concave
down. Additionally, there is a 97% probability that values of β_r and β_{v^2} are
344 positively correlated (Mean correlation 0.51, standard deviation 0.23).

While the overall group level estimates are of particular importance when
346 defining time-invariant differences in extinction risk, it is also important and
useful to analyze the individual level parameter estimates in order to better
348 understand how parameters actually vary across cohorts.

In comparison to the overall mean extinction risk μ_0 , cohort level estimates β_0
350 show some amount of variation through time as expected by estimates of τ_0
(Fig. 4a). A similar, if slightly greater, amount of variation is also observable in
352 cohort estimates of the effect of geographic range β_r (Fig. 4b). Again, smaller
values of β_0 correspond to lower expected extinction risk. Similarly, smaller
354 values of β_r correspond to greater decrease in extinction risk with increasing
geographic range

356 How the effect of environmental affinity varies between cohorts can be observed
by using the cohort specific coefficients estimates. Following the same procedure
358 used earlier (Fig. 1), but substituting cohort specific estimates of β_v and β_{v^2} for
 μ_v and μ_{v^2} , the cohort specific effect of environmental preference as a multiplier
360 of mean extinction risk can be calculated. This was done only for the Weibull
model, though the observed pattern should be similar for the exponential model.

362 As expected based on the estimates of τ_v and τ_{v^2} , there is greater variation in
the peakedness of $f(v_i)$ than there is variation between convave up and down
364 functions (Fig. 5). 12 of the 33 cohorts have less than 50% posterior probability
that generalists are potentially expected to be shorter lived than specialists,
366 though two of those cases have approximately a 50% probability of being either
concave up or down. This is congruent with the 0.72 posterior probability that
368 μ_{v^2} is positive/ $f(v_i)$ is concave down.

Additionally, for some cohorts there is a quite striking pattern where the effect
370 of environmental preference v has a nearly-linear relationship (Fig. 5). These are

primarily scenarios where one of the end member preferences is expected to
 372 have a greater duration than either intermediate or the opposite end member
 preference. Whatever curvature is present in these nearly-linear cases is due to
 374 the defintion of $f(v)$ as it is not defined for non-negative values of σ (Eq. 3). For
 all stages between the Emsian through the Viséan, inclusive, intermediate
 376 preferences are of intermediate extinction risk when compared with
 epicontinental specialists (lowest risk) or open-ocean specialists (highest risk).
 378 This time period represents most of the Devonian through the early
 Carboniferous.

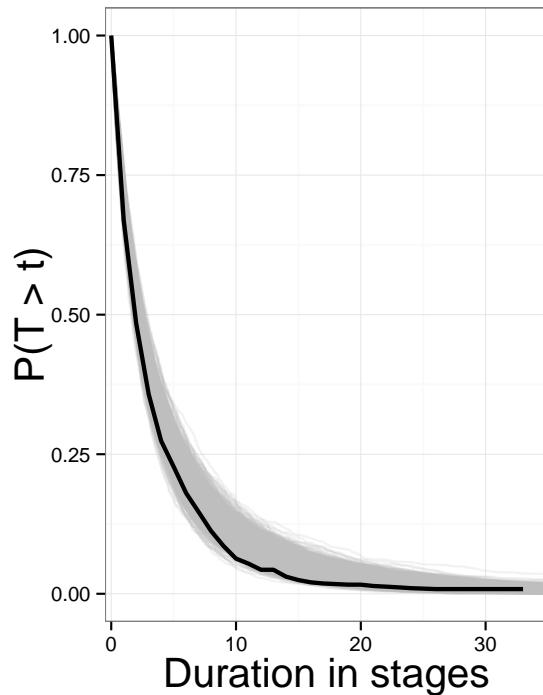


Figure 1: Comparison of empirical estimates of $S(t)$ versus estimates from 1000 posterior predictive data sets. $S(t)$ corresponds to $P(T > t)$ as it is the probability that a given genus observed at age t will continue to live. This is equivalent to the probability that t is less than the genus' ultimate duration T . Note that the Weibull (left) model has noticeably better fit to the data than the exponential (right).

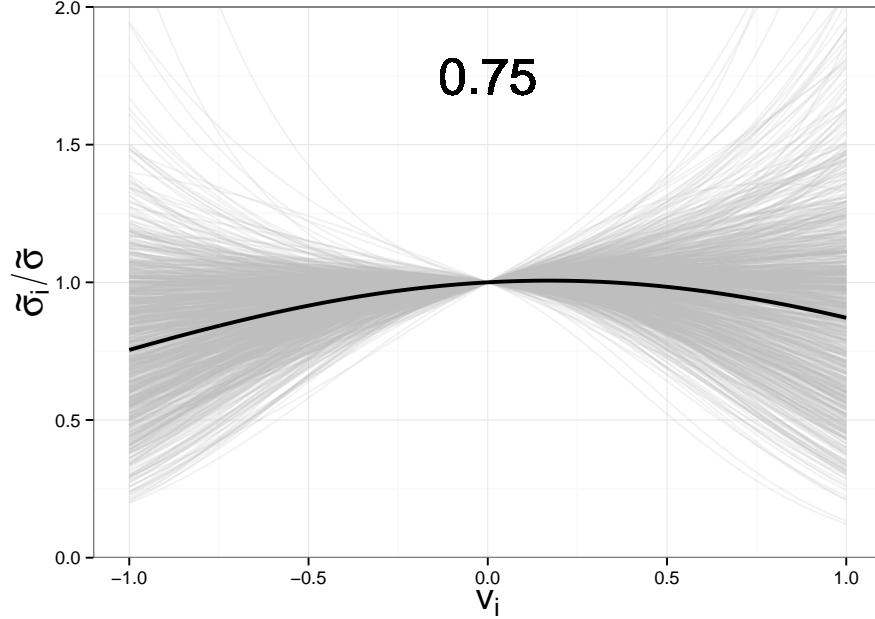


Figure 2: The overall expected relationship $f(v_i)$ between environmental affinity v_i and a multiplier of extinction risk (Eq. 3). Each grey line corresponds to a single draw from the posterior predictive distribution, while the black corresponds to the median of the posterior predictive distribution. The overall shape of $f(v_i)$ is concave down with an optimum of close 0, which corresponds to affinity approximately equal to the expectation based on background environmental occurrence rates.

380 4 Discussion

My results demonstrate that both the effects of geographic range and the
 382 peakedness/concavity of environmental preference are both negatively
 correlated with baseline extinction risk, meaning that as baseline extinction risk
 384 increases the effect sizes of both these traits are expected to increase (Fig. 3b).
 This result supports neither of the two proposed macroevolutionary mechanisms
 386 for how biological traits should correlate with extinction risk. The observed
 correlation between the two effects as well as between the effects and baseline
 388 extinction risk instead implies that as baseline extinction risk increases, the

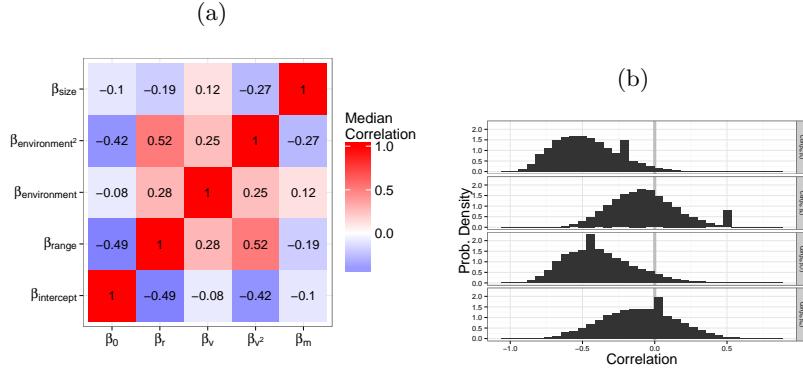


Figure 3: **A:** Heatmap for the median estimates of the terms of the correlation matrix Ω between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the effects of covariates, along with intercept/baseline extinction risk. **B:** Marginal posterior distributions of the correlations between intercept terms/baseline extinction risk and the effects of each of the covariates. These are presented for both the exponential (left) and Weibull (right) models.

strength of the total selection gradient on biological traits (except body size)
 390 increases. This manifests as greater differences in extinction risk for each unit
 difference in the biological covariates during periods of high extinction risk,
 392 while a relatively flatter selection gradient during periods of low extinction risk.

There are two mass extinction events that are captured within the time frame
 394 considered here: the Ordovician-Silurian and the Frasnian-Famennian. The
 cohorts bracketing these events are worth considering in more detail.

396 The proposed mechanism for the end Ordovician mass extinction is a decrease
 in sea level and the draining of epicontinental seas due to protracted glaciation
 398 (Johnson, 1974, Sheehan, 2001). My results are broadly consistent with this
 scenario with both epicontinental and open-ocean specialists having a much
 400 lower expected duration than intermediate taxa (Fig. 5). All of the stages
 between the Darriwillian and the Llandovery, except the Hirnantian, have a

parameter	mean	standard deviation
μ_i	-1.51	0.15
μ_r	-1.38	0.14
μ_e	-0.08	0.18
μ_{e2}	0.25	0.43
μ_m	-0.09	0.09
τ_i	0.63	0.11
τ_r	0.48	0.12
τ_e	1.07	0.23
τ_{e2}	1.88	0.66
τ_m	0.32	0.13

Table 1: j+Caption text+j,

402 high probability (90+) that $f(v)$ is concave down. The pattern for the
 Darriwillian, which proceeds the supposed start of Ordovician glacial activity,
 404 demonstrates that taxa tend to favor open-ocean environments are expected to
 have a greater duration than either intermediate or epicontinental specialists, in
 406 decreasing order.

For nearly the entire Devonian estimates of $f(v)$ indicate that one of the
 408 environmental end members is favored over the other end member of
 intermediate preference (Fig. 5). This is consistent with the predictions of Miller
 410 and Foote (2009). For almost the entirely the Givetian though the end of the
 Devonian and into the Viséan, I find that epicontinental favoring taxa are
 412 expected to have a greater duration than either intermediate or open-ocean
 specialists. Additionally, for nearly the entire Devonian and through to the
 414 Visean, the cohort-specific estimates of $f(v)$ are concave-up. This is the opposite
 pattern than what is expected (Fig. 2). This result, however, seems to reflect
 416 the intensity of the seemingly nearly-linear difference in expected duration
 across the range of v) as opposed to an inversion of the weakly expected
 418 curvilinear pattern.

There is an approximate 72% posterior probability that taxa with intermediate

⁴²⁰ environmental preferences are possibly expected to have a lesser extinction risk
than either end members, the over all curvature of $f(v_i)$ is not very peaked,
⁴²² meaning that this relationship does not lead to very strong differences in
extinction risk (Fig. 2). This result gives weak support for the hypothesis that,
⁴²⁴ in general, environmental generalists survive for longer than environmental
specialists (Liow, 2004, 2007, Nürnberg and Aberhan, 2013, 2015, Simpson,
⁴²⁶ 1944).

The variance in estimate of the overall $f(v_i)$ reflects the large between cohort
⁴²⁸ variance in cohort specific estimates of $f(v_i)$ (Fig. 5). Given that there is only a
72% posterior probability that the expected overall estimate of $f(v_i)$ is concave
⁴³⁰ down, it is not surprising that there are some stages where the theorized
relationship is in fact reversed. Additionally, as discussed earlier, some of those
⁴³² same stages where $f(v_i)$ does not resemble the theorized nonlinear relation with
the optimum in the middle, but are instead highly skewed or effectively linear
⁴³⁴ (Fig. 5).

These results do not necessarily refute “survival of the unspecialized” as a
⁴³⁶ time-invariant generalization, but instead demonstrate how, while the expected
group-level estimate of $f(v_i)$ might favor one hypothesis, there is still enough
⁴³⁸ variability between cohorts so that in some realizations this pattern may not
hold or can even be reversed. These results are also consistent with aspects of
⁴⁴⁰ Miller and Foote (2009) who found that the effect of environmental preference
on extinction risk was quite variable and without obvious patterning during
⁴⁴² times of background extinction.

This model can be improved through either increasing the number of analyzed
⁴⁴⁴ taxon traits, expanding the hierarchical structure of the model to include other
major taxonomic groups of interest, and the inclusion of explicit phylogenetic
⁴⁴⁶ relationships between the taxa in the model as an additional hierarchical effect.

An example taxon trait that may be of particular interest is the affixing
448 strategy or method of interaction with the substrate of the taxon. This trait has
been found to be related to brachiopod survival (Alexander, 1977) so its
450 inclusion may be of particular interest.

It is theoretically possible to expand this model to allow for comparisons within
452 and between major taxonomic groups. This approach would better constrain the
brachiopod estimates while also allowing for estimation of similarities and
454 differences in cross-taxonomic patterns. The major issue surrounding this
particular expansion involves finding a similarly well sampled taxonomic group
456 that is present during the Paleozoic. Example groups include Crinoidea,
Ostracoda, and other “Paleozoic” groups (Sepkoski Jr., 1981).

458 Taxon traits like environmental preference or geographic range (Hunt et al.,
2005, Jablonski, 1987) are most likely heritable, at least phylogenetically
460 (Housworth et al., 2004, Lynch, 1991). Without phylogenetic context, this
analysis assumes that differences in extinction risk between taxa are
462 independent of those taxa's shared evolutionary history (Felsenstein, 1985). In
contrast, the origination cohorts only capture shared temporal context. The
464 inclusion of phylogenetic context as an addition individual level hierarchical
structure independent of origination cohort would allow for determining how
466 much of the observed variability is due to shared evolutionary history versus
actual differences associated with these taxonomic traits.

468 In summary, patterns of Paleozoic brachiopod survival were analyzed using a
fully Bayesian hierarchical survival modelling approach. Using a varying-slopes,
470 varying-intercepts approach I am able to model both the overall mean effect of
biological covariates on extinction risk while also modeling the correlation
472 between origination cohort-specific estimates of covariate effects. I find that as
baseline extinction risk increases, the strength of the selection gradient on

⁴⁷⁴ biological traits (except body size) increases. This manifests as greater
⁴⁷⁶ differences in extinction risk for each unit difference in the biological covariates
⁴⁷⁸ during periods of high extinction risk, while a much flatter total selection
gradient during periods of low extinction risk. I also find some support for
⁴⁸⁰ “survival of the unspecialized” (Liow, 2004, 2007, Nürnberg and Aberhan, 2013,
2015, Simpson, 1944) as a general characterization of the effect of environmental
⁴⁸² preference on extinction risk (Fig. 2), though there is heterogeneity between
origination cohorts (Fig. 5). Generally, this study demonstrates the advantages
⁴⁸⁴ of a hierarchical Bayesian framework for taking into account the structured
nature of the data. Future studies of structured data should adopt similar
strategies in order to best model our knowledge instead of ignoring that
structure which can lead to poor and/or incorrect inference.

⁴⁸⁶ Acknowledgements

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for
⁴⁸⁸ helpful discussion and advice. Additionally, thank you A. Miller for the
epicontinental versus open-ocean assignments. This entire study would
⁴⁹⁰ not have been possible without the Herculean effort of the many contributors to
the Paleobiology Database. In particular, I would like to thank J. Alroy, M.
⁴⁹² Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland,
L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T. Olszewski, P.
⁴⁹⁴ Novack-Gottshall, M. Patzkowsky, M. Uhen, L. Villier, and P. Wager. This work
was supported by a NASA Exobiology grant (NNX10AQ446) to A. Miller and
⁴⁹⁶ M. Foote. This is Paleobiology Database publication XXX.

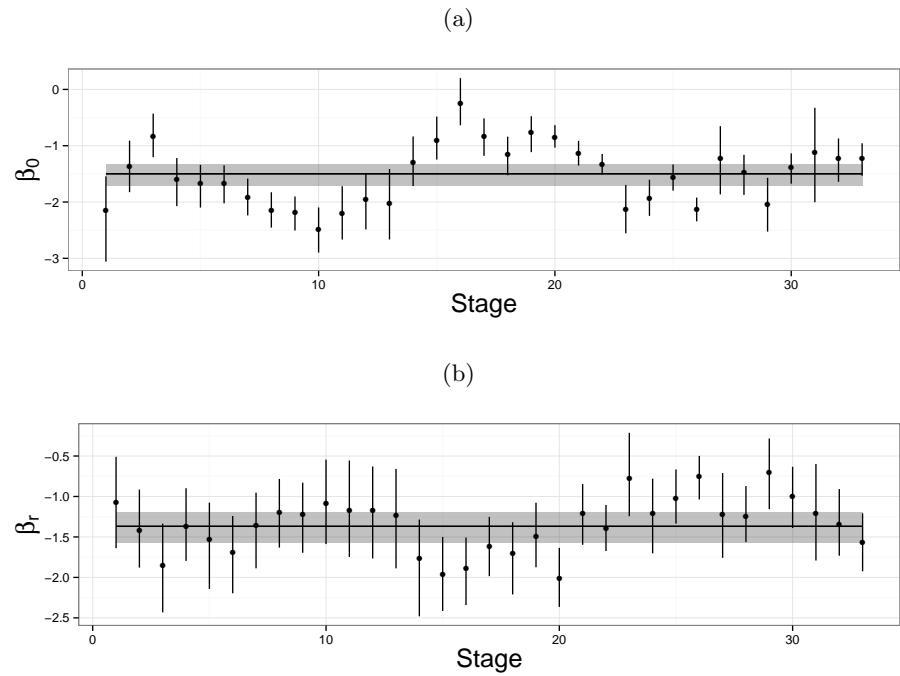


Figure 4: Comparison of cohort-specific estimates of β_0 presented along with the estimate for the overall baseline extinction risk. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction risk along with 80% credible intervals. Results are presented for the exponential (top) and Weibull (bottom) models. Comparison of cohort-specific estimates of the effect of geographic range on extinction risk β_r presented along with the estimate for the overall effect of geographic range. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction risk along with 80% credible intervals. Results are presented for the exponential (top) and Weibull (bottom) models.

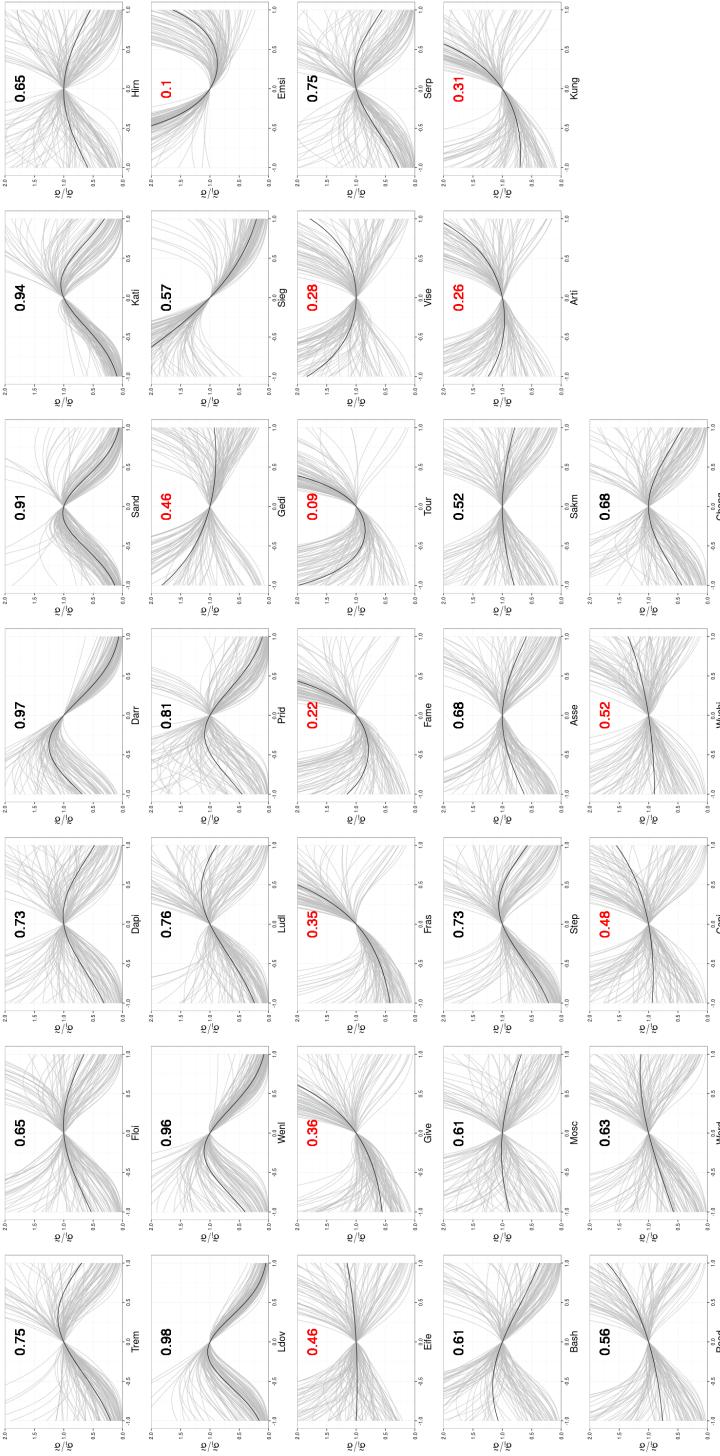


Figure 5: Comparison of the cohort-specific estimates of $f(v_i)$ (Eq. 3) for the 33 analyzed origination cohorts. The stage of origination is labeled on the x-axis of each panel. The oldest stage is in the upper left, while the youngest is in the lower left. The number in each panel corresponds to the posterior probability that $f(v_i)$ is concave down. Those that are highlighted in red have less than 51% posterior predictive probability that $f(v_i)$ is concave down.

References

- 498 Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
to the mode of stabilization on the substrate. *Palaeogeography,*
500 *Palaeoclimatology, Palaeoecology* 21:209–226.
- 502 Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal
Groups. *Science* 329:1191–1194.
- 504 Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of
filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- 506 Carvalho, C. M., N. G. Polson, and J. G. Scott, 2009. Handling Sparsity via the
Horseshoe. *in* Proceedings of the 12th International Conference on Artificial
Intelligence and Statistics, vol. 5, Pp. 73–80.
- 508 ———, 2010. The horseshoe estimator for sparse signals. *Biometrika*
97:465–480.
- 510 Cooper, W. S., 1984. Expected time to extinction and the concept of
fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- 512 Felsenstein, J., 1985. Phylogenies and the comparative method. *American
Naturalist* 125:1–15.
- 514 Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites.
Paleobiology 14:258–271.
- 516 ———, 2006. Substrate affinity and diversity dynamics of Paleozoic marine
animals. *Paleobiology* 32:345–366.
- 518 Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine
animal genera. *Paleobiology* 39:171–192.

- 520 Gelman, A., 2006. Prior distributions for variance parameters in hierarchical
models. *Bayesian Analysis* 1:515–533.
- 522 Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B.
Rubin, 2013. Bayesian data analysis. 3 ed. Chapman and Hall, Boca Raton,
524 FL.
- Gelman, A. and J. Hill, 2007. Data Analysis using Regression and
526 Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- Harnik, P. G., C. Simpson, and J. L. Payne, 2013. Long-term differences in
528 extinction risk among the seven forms of rarity. *Proceedings of the Royal
Society B: Biological Sciences* 282.
- 530 Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL
<http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- 532 Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively
Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine
534 Learning Research* 15:1351–1381.
- Housworth, E. A., P. Martins, and M. Lynch, 2004. The Phylogenetic Mixed
536 Model. *The American Naturalist* 163:84–96.
- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
538 comment on "On the heritability of geographic range sizes". *American
Naturalist* 166:129–135.
- Ibrahim, J. G., M.-H. Chen, and D. Sinha, 2001. Bayesian Survival Analysis.
Springer, New York.
- Jablonski, D., 1986. Background and mass extincitons: the alternation of
542 macroevolutionary regimes. *Science* 231:129–133.

- 544 ———, 1987. Heritability at the species level: analysis of geographic ranges of
cretaceous mollusks. *Science* 238:360–363.
- 546 Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and
living molluscs. *Proceedings. Biological sciences / The Royal Society*
548 270:401–6.
- Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- 550 Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
benthic biodiversity dynamics through Triassic Jurassic time. *Paleobiology*
552 33:414–434.
- Klein, J. P. and M. L. Moeschberger, 2003. *Survival Analysis: Techniques for
554 Censored and Truncated Data.* 2nd ed. Springer, New York.
- Liow, L. H., 2004. A test of Simpson's "rule of the survival of the relatively
556 unspecialized" using fossil crinoids. *The American naturalist* 164:431–43.
- , 2007. Does versatility as measured by geographic range, bathymetric
558 range and morphological variability contribute to taxon longevity? *Global
Ecology and Biogeography* 16:117–128.
- Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary
560 biology. *Evolution* 45:1065–1080.
- Miller, A. I. and S. R. Connolly, 2001. Substrate affinities of higher taxa and
562 the Ordovician Radiation. *Paleobiology* 27:768–778.
- Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:
564 the kinetics of mass extinction and origination. *Science* 326:1106–9.
- Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range

- predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*
568 39:360–372.
- , 2015. Interdependence of specialization and biodiversity in Phanerozoic
570 marine invertebrates. *Nature communications* 6:6602.
- Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental
572 than evolvability. *PloS one* 7:e38025.
- Payne, J. L. and S. Finnegan, 2007. The effect of geographic range on
574 extinction risk during background and mass extinction. *Proceedings of the
National Academy of Sciences* 104:10506–11.
- 576 Payne, J. L., N. A. Heim, M. L. Knope, and C. R. McClain, 2014. Metabolic
dominance of bivalves predates brachiopod diversity decline by more than 150
578 million years. *Proceedings of the Royal Society B* 281:20133122.
- Peters, S. E., 2008. Environmental determinants of extinction selectivity in the
580 fossil record. *Nature* 454:626–9.
- Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen's Law.
582 *Paleobiology* 1:82–96.
- , 1978. Cohort Analysis of generic survivorship. *Paleobiology* 4:1–15.
- 584 Sepkoski Jr., J. J., 1981. A factor analytic description of the Phanerozoic
marine fossil record. *Paleobiology* 7:36–53.
- 586 Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of
Earth and Planetary Sciences* 29:331–364.
- 588 Simpson, C., 2006. Levels of selection and large-scale morphological trends.
Ph.D. thesis, University of Chicago.

- 590 Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine
bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- 592 Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University
Press, New York.
- 594 ———, 1953. *The Major Features of Evolution*. Columbia University Press,
New York.
- 596 Stan Development Team, 2014a. Stan: A c++ library for probability and
sampling, version 2.5.0. URL <http://mc-stan.org/>.
- 598 ———, 2014b. *Stan Modeling Language Users Guide and Reference Manual*,
Version 2.5.0. URL <http://mc-stan.org/>.
- 600 Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- , 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.
- 602 Vilhena, D. A., E. B. Harris, C. T. Bergstrom, M. E. Maliska, P. D. Ward, C. A.
Sidor, C. A. E. Strömberg, and G. P. Wilson, 2013. Bivalve network reveals
- 604 latitudinal selectivity gradient at the end-Cretaceous mass extinction.
Scientific Reports 3:1790.
- 606 Wang, S. C., 2003. On the continuity of background and mass extinction.
Paleobiology 29:455–467.
- 608 Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and
Widely Applicable Information Criterion in Singular Learning Theory.
- 610 Journal of Machine Learning Research 11:3571–3594.

A Uncertainty in environmental preference

612 The calculation and inclusion of environmental affinity in the survival model is a
statistical procedure that takes into account our uncertainty based on where
614 fossils tend to occur. Because we cannot directly observe if a fossil taxon had
occurrences restricted to only a single environment, instead we can only
616 estimate its affinity with uncertainty. One advantage of using a Bayesian
analytical approach is that both parameters and data are considered random
618 samples from some underlying distribution, which means that it is possible to
model the uncertainty in our covariates of interest (Gelman et al., 2013). My
620 approach is conceptually similar to Simpson and Harnik (2009) but instead of
obtaining a single point estimate, an entire posterior distribution is estimated.

622 The first step is to determine the probability θ at which genus i occurs in an
epicontinental settings based on its own pattern of occurrences. Define e_i as the
624 number of occurrences of genus i in an epicontinental sea and o_i as the number of
occurrences of genus i not in an epicontinental sea (e.g. open ocean). Because
626 the value of interest is the probability of occurring in an epicontinental
environment, given the observed fossil record, I assume that probability follows
628 a Bernoulli distribution. We can then define our sampling statement as

$$e_i \sim \text{Bernoulli}(e_i + o_i, \theta_i). \quad (4)$$

I used a flat prior for θ_i defined as $\theta_i \sim \text{Beta}(1, 1)$. Because the beta
630 distribution is the conjugate prior for the Bernoulli distribution, the posterior is
easy to compute in closed form. The posterior probability of θ is then

$$\theta_i \sim \text{Beta}(e_i + 1, o_i + 1) \quad (5)$$

632 It is extremely important, however, to take into account the overall
 environmental occurrence probability of all other genera present at the same
 634 time as genus i . This is incorporated as an additional probability Θ . Define E_i
 as the total number of other fossil occurrences (exceptfor genus i) in
 636 epicontinental seas during stages where i occurs and O_i as the number of other
 fossil occurrences not on epicontinental seas. We can then define the sampling
 638 statement as

$$E_i \sim \text{Bernoulli}(E_i + O_i, \Theta_i). \quad (6)$$

Again, I used a flat prior of Θ_i defined as $\Theta_i \sim \text{Beta}(1, 1)$. The posterior of Θ is
 640 then simply defined as

$$\Theta_i \sim \text{Beta}(E_i + 1, O_i + 1) \quad (7)$$

I then define the environmental affinity of genus i as $v_i = \theta_i - \Theta_i$. v_i is a value
 642 that can range between -1 and 1, where negative values indicate that genus i
 tends to occur more frequently in open ocean environments than background
 644 while positive values indicate that genus i tends to occur in epicontiental
 environments.

646 While this approach is noticeably more complicated than previous ones (Foote,
 2006, Kiessling and Aberhan, 2007, Miller and Connolly, 2001, Simpson and
 648 Harnik, 2009) there are some important benefits to both using a continuous
 measure of affinity as well directly modeling our uncertainty. In order to show
 650 some of these benefits, I performed a simulation analysis of how
 modal/maximum *a posteriori* (MAP) estimates versus full posterior estimates.

652 In this simulation, I first defined the “background” epicontinental occurrence θ_b
 as 0.50 with a small amount of noise. This was represented as a beta distribution

$$\Theta_b = \text{Beta}(\alpha = 2500, \beta = 2500). \quad (8)$$

654 This choice of parameters for the distribution reflects the average number of
background occurrences for either epicontinental or open ocean environments
656 per genus.

Using this background occurrence ratio, I randomly generated the occurrence
658 patterns of 1000 simulated taxa. This was done at multiple sample sizes (1, 2, 3,
4, 5, 10, 25, 50, 100) in order to demonstrate the effects of increasing sample
660 size on the confidence of environmental affinity. For each simulated taxon I
calculated the full posterior distribution while assuming a flat Beta prior
662 ($\text{Beta}(1, 1)$). Using the full posterior I calculated the MAP probability of
occurring in epicontinental environments. The environmental affinity was
664 calculated for each of the simulated taxa using both the full posterior and the
MAP estimate. In this toy example, environmental affinity can range between
666 -0.5 and 0.5.

As should be expected, as sample size increases the distribution of MAP
668 estimates converge on the true value (Fig. 6). For taxa with less than 10
occurrences, the MAP estimate is biased towards extreme values. Note that the
670 mode of the beta distribution is not defined for situations where there were 0
draws of one of the environmental conditions. Instead, the vertical line is based
672 entirely on the observed occurrences which are technically the modal estimates
because they are the most frequently occurring/highest density.

674 In contrast, we can compare the true occurrence probability distribution versus
the posterior estimate for a given sample (Fig. 7). When sample sizes are low,
676 posterior estimates are flat and represent a compromise between the likelihood
and the flat prior (Eq. 5). Because of this, estimates from small sizes are less

678 likely to be overly biased towards the extremes. This is further emphasized by
 inspection of the estimates of environmental affinity for the simulated taxa (Fig.
 680 8). Posterior estimates from simulated taxa with small sample size have a much
 broader distribution that both allows for the extreme observation but still
 682 captures the “true” value (0).

By defining environmental preference as the difference in full posterior estimates
 684 of occurrence probability, it is possible to include taxa with low sample sizes
 that are normally discarded (Foote, 2006, Kiessling and Aberhan, 2007, Miller
 686 and Connolly, 2001, Simpson and Harnik, 2009). Additionally, 55+% of
 observed Paleozoic brachiopod genera have less than 10 occurrences which is the
 688 range of sample sizes where MAP (or ML) estimates would be potentially most
 biased. This is preferable to finding the difference between the MAP estimates
 690 (blue line; Fig. 8).

B Survival model

692 The simplest model of genus duration includes no covariate or structural
 information. Define y_i as the duration in stages of genus i , where $i = 1, \dots, n$
 694 and n is the number of observed genera. These two models are then simply
 defined as

$$y_i \sim \text{Exponential}(\lambda) \quad (9)$$

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

696 λ, α , and σ are all defined for all positive reals. Note that λ is a “rate” or
 inverse-scale while σ is a scale parameter, meaning that $\frac{1}{\lambda} = \sigma$.

698 These simple models can then be expanded to include covariate information as
 predictors by reparameterizing λ or σ as a regression (Klein and Moeschberger,

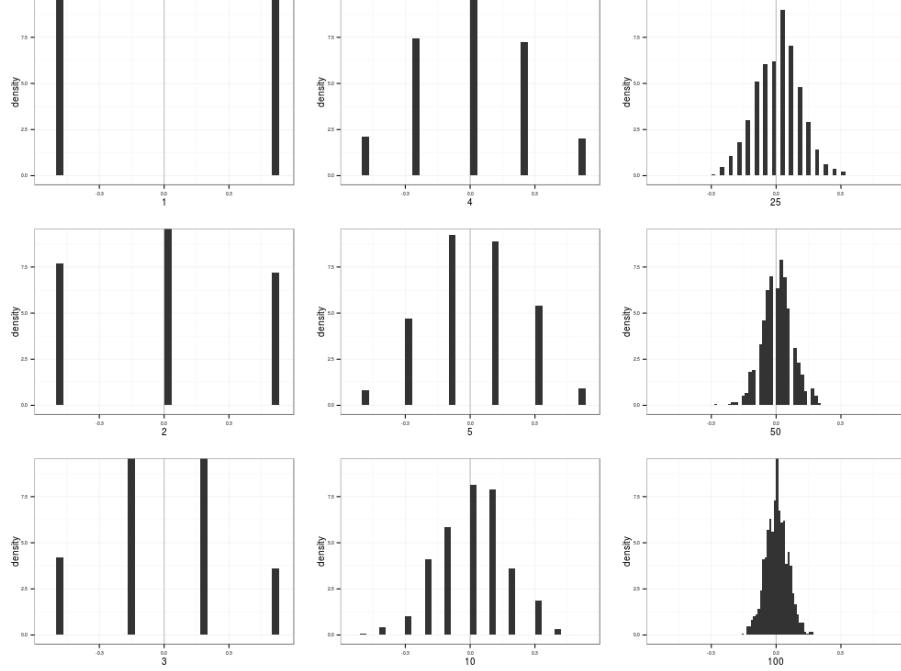


Figure 6: Histograms of the distributions of from the beta distribution defined in Eq. 8. As to be expected, as sample size increases the draws better resemble the underlying true distribution. Sample size is indicated as the label of the x-axis, increasing in column major order.

700 2003). Each of the covariates of interest is given its own regression coefficient
 (e.g. β_r) along with an intercept term β_0 . There are some additional
 702 complications to the parameterization of σ associated with the inclusion of α as
 well as for interpretability (Klein and Moeschberger, 2003). Both of these are
 704 then written as

$$\begin{aligned}\lambda_i &= \exp(\beta_0 + \beta_r r_i + \beta_v v_i + \beta_{v^2} v_i^2 + \beta_m m_i) \\ \sigma_i &= \exp\left(\frac{-(\beta_0 + \beta_r r_i + \beta_v v_i + \beta_{v^2} v_i^2 + \beta_m m_i)}{\alpha}\right).\end{aligned}\quad (10)$$

The quadratic term for environmental affinity v is to allow for the possible
 706 nonlinear relationship between environmental affinity and extinction risk.

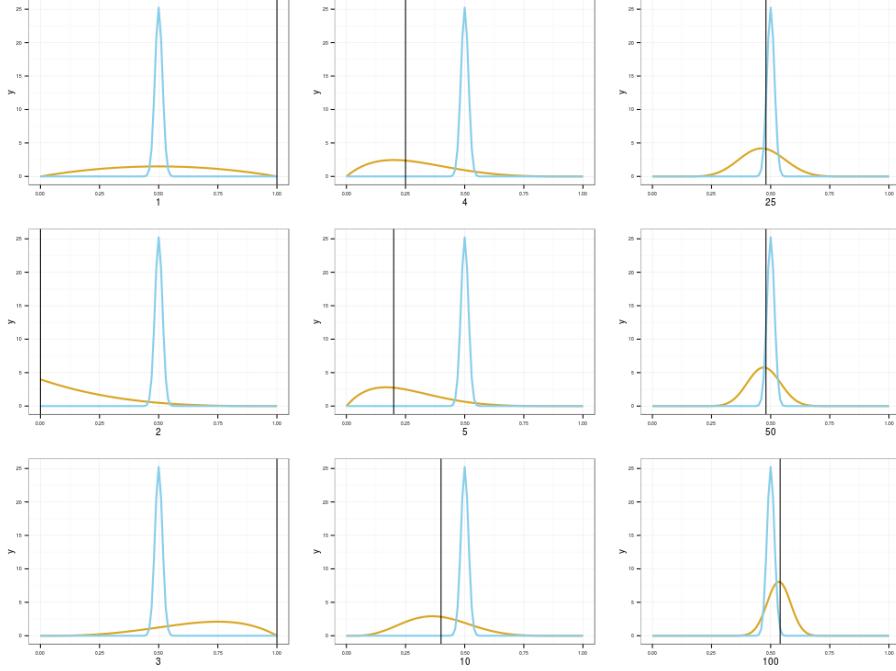


Figure 7: Comparisons of the underlying distribution (blue) to posterior estimates based on increasing sample size (gold). Each posterior estimate is represented for only a single realization of draws, each with sample size indicated as the x-axis label (increasing in column major order). Black vertical lines correspond to the MAP estimate of the simulated taxon’s affinity. This stands in contrast to the posterior distribution of expected affinity in gold.

The models which incorporate both equations 9 and 10 can then be further
 708 expanded to allow all of the β coefficients, including β_0 , to vary with origination
 cohort while also modeling their covariance and correlation. This is called a
 710 varying-intercepts, varying-slopes model (Gelman and Hill, 2007). It is much
 easier to represent and explain how this is parameterized using matrix notation.
 712 First, define \mathbf{B} as $k \times J$ matrix of the k coefficients including the intercept term
 (k = 5) for each of the J cohorts. Second, define \mathbf{X} as a $n \times k$ matrix where each
 714 column is one of the covariates of interest. Importantly, \mathbf{X} includes a columns of
 all 1s which correspond to the constant term β_0 . Third, define $j[i]$ as the

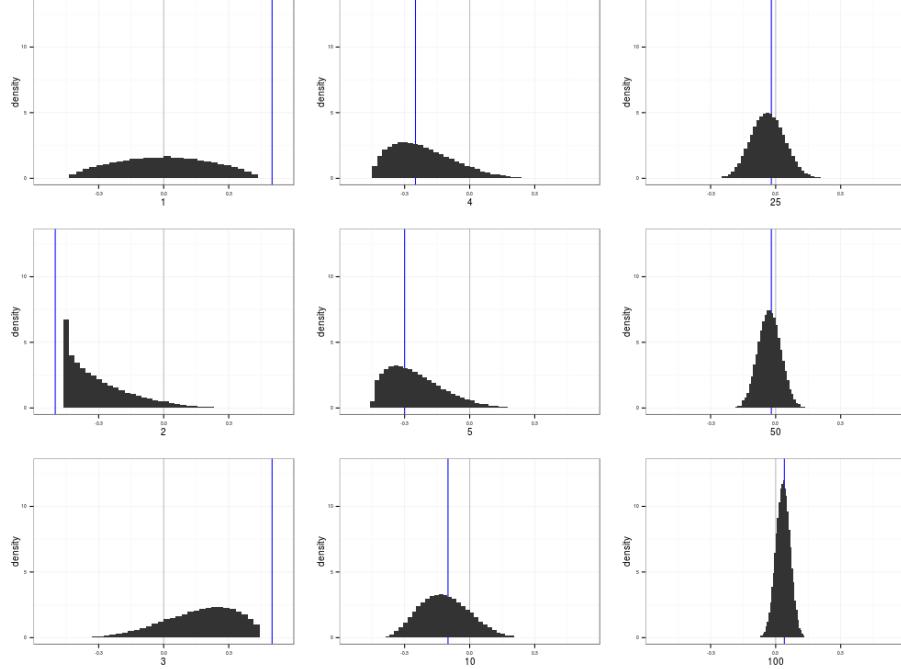


Figure 8: Histograms of the difference in the underlying occurrence distribution and the posterior distribution estimates from the previous graph (Fig. 7). The “true” value is included in all distributions of environmental affinities. Each affinity estimate is represented for only a single realization of draws, each with sample size indicated as the x-axis label (increasing in column major order). Blue vertical lines correspond to the difference in MAP estimates between the underlying distribution and the simulated taxon’s draws. This stands in contrast to the distribution of the differences between the simulated taxon and background.

716 origination cohort of genus i , where $j = 1, \dots, J$ and J is the total number of observed cohorts. We then rewrite λ and σ of equation 10 in matrix notation as

$$\begin{aligned} \lambda_i &= \exp(\mathbf{X}_i B_{j[i]}) \\ \sigma_i &= \exp\left(\frac{-(\mathbf{X}_i B_{j[i]})}{\alpha}\right). \end{aligned} \tag{11}$$

718 Because B is a matrix, I use a multivariate normal prior with unknown vector

of means μ and covariance matrix Σ . This is written as

$$B \sim \text{MVN}(\vec{\mu}, \Sigma) \quad (12)$$

720 where $\vec{\mu}$ is length k vector representing the overall mean of the distributions of
 β coefficients. Σ is a $k \times k$ covariance matrix of the β coefficients.

722 What remains is assigning priors the elements of $\vec{\mu}$ and the covariance matrix Σ .
 All elements of $\vec{\mu}$ except for μ_r were given horseshoe priors (Carvalho et al.,
 724 2009, 2010) while μ_r was given an informative normal prior ($\mu_r \sim \mathcal{N}(-1, 1)$).
 Horseshoe priors are a strong regularizing priors with effectively infinite density
 726 at 0 and heavy, Cauchy-like tails (Carvalho et al., 2009, 2010) which allow
 weakly inferred effects to be strongly drawn towards 0 while truly strong effects
 728 can remain large. The horseshoe prior consists of a normal distribution with
 scale term that is the product between a global shrinkage parameter ν and a
 locak shrinkage parameter ψ unique to each of the parameters of interest. These
 730 parameters are themselves given half-Cauchy priors (Eq. 1 and 2).

732 The prior for Σ is a bit more complicated due to its multivariate nature.
 Following the Stan Development Team (2014b), I modeled the scale terms
 734 separate from the correlation structure of the coefficients. This is possible
 because of the relationship between a covariance and a correlation matrix,
 736 defined as

$$\Sigma_B = \text{Diag}(\vec{\tau})\Omega\text{Diag}(\vec{\tau}) \quad (13)$$

where $\vec{\tau}$ is a length k vector of variances and $\text{Diag}(\tau)$ is a diagonal matrix.

738 I used a LKJ prior distribution for correlation matrix Ω as recommended by
 Stan Development Team (2014b). The LKJ distribution is a single parameter
 740 multivariate distribution where values of the parameter η greater than 1

concentrate density at the unit correlation matrix, which corresponds to no
correlation between the β coefficients. The scale parameters, $\vec{\tau}$, are given weakly
informative half-Cauchy (C^+) priors following Gelman (2006).

744 C Censored observations

A key aspect of survival analysis is the inclusion of censored, or incompletely
observed, data points (Ibrahim et al., 2001, Klein and Moeschberger, 2003). The
two classes of censored observations encountered in this study were right and
left censored observations. Right censored genera are those that did not go
extinct during the window of observation, or genera that are still extant. Left
censored observations are those taxa for which we know only an upper limit on
their duration.

In the context of this study, I considered all genera that had a duration of only
one geologic stage to be left censored as we do not have a finer degree of
resolution.

The key function for modeling censored observations is the survival function, or
 $S(t)$. $S(t)$ corresponds to the probability that a genus having existed for t stages
will not have gone extinct while $h(t)$ corresponds to the instantaneous
extinction rate at taxon age t Klein and Moeschberger (2003). For an
exponential model, $S(t)$ is defined as

$$S(t) = \exp(-\lambda t), \quad (14)$$

and for the Weibull distribution $S(t)$ is defined as

$$S(t) = \exp\left(-\left(\frac{t}{\sigma}\right)^\alpha\right). \quad (15)$$

$S(t)$ is equivalent to the complementary cumulative distribution function,
⁷⁶² $1 - F(t)$ (Klein and Moeschberger, 2003).

For right censored observations, instead of calculating the likelihood as normal
⁷⁶⁴ (Eq. 11) the likelihood of an observation is evaluated using $S(t)$. Conceptually,
 this approach calculates the likelihood of observing a taxon that existed for at
⁷⁶⁶ least that long. For left censored data, instead the likelihood is calculated using
 $1 - S(t)$ which corresponds to the likelihood of observing a taxon that existed
⁷⁶⁸ no longer than t .

The full likelihood statements incorporating fully observed, right censored, and
⁷⁷⁰ left censored observations are then

$$\begin{aligned} \mathcal{L} &\propto \prod_{i \in C} \text{Exponential}(y_i | \lambda) \prod_{j \in R} S(y_j | \lambda) \prod_{k \in L} (1 - S(y_k | \lambda)) \\ \mathcal{L} &\propto \prod_{i \in C} \text{Weibull}(y_i | \alpha, \sigma) \prod_{j \in R} S(y_j | \alpha, \sigma) \prod_{k \in L} (1 - S(y_k | \alpha, \sigma)) \end{aligned} \quad (16)$$

where C is the set of all fully observed taxa, R the set of all right censored taxa,
⁷⁷² and L the set of all left-censored taxa.

D Widely applicable information criterion

⁷⁷⁴ WAIC can be considered a fully Bayesian alternative to the Akaike information
 criterion, where WAIC acts as an approximation of leave-one-out
⁷⁷⁶ cross-validation which acts as a measure of out-of-sample predictive accuracy
 (Gelman et al., 2013). WAIC is calculated starting with the log pointwise
⁷⁷⁸ posterior predictive density calculated as

$$\text{lppd} = \sum_{i=1}^n \log \left(\frac{1}{S} \sum_{s=1}^S p(y_i | \Theta^S) \right), \quad (17)$$

where n is sample size, S is the number posterior simulation draws, and Θ
 780 represents all of the estimated parameters of the model. This is similar to
 calculating the likelihood of each observation given the entire posterior. A
 782 correction for the effective number of parameters is then added to lppd to
 adjust for overfitting. The effective number of parameters is calculated,
 784 following the recommendations of Gelman et al. (2013), as

$$p_{\text{WAIC}} = \sum_{i=1}^n V_{s=1}^S (\log p(y_i | \Theta^S)). \quad (18)$$

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

Given both equations 17 and 18, WAIC is then calculated

$$\text{WAIC} = \text{lppd} - p_{\text{WAIC}}. \quad (19)$$

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