

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. 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 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. My analysis framework eschews the traditional distinction between background and mass extinction, instead the entire time period is analyzed where these "states" are part of a continuum. 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 many parts of the Paleozoic this hypothesis is supported, there are many times where this hypothesized relationship is absent or even reversed. Importantly, I find that as extinction risk increases, the steepness of this relationship is expected to increase as well.

36 1 Introduction

How do biological traits affect extinction risk? Jablonski (1986) observed that
38 during periods of high extinction risk, the effects of biological traits on survival

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

44 Jablonski (1986) phrased his conclusions in terms of background versus mass
extinction, but this scenario is readily transferable to a continuous variation
46 framework as there is no obvious distinction in terms of extinction rate between
these two states (Wang, 2003). Additionally, the Jablonski (1986) scenario has
48 strong model structure requirements in order to test its proposed
macroevolutionary mechanism; not only do the taxon trait effects need to be
50 modeled, but the correlation between the trait 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 brachiopod taxon durations because trait based differences in extinction
risk should manifest as differences in taxon durations. Namely, a taxon with a
60 beneficial trait should survive longer, on average, than a taxon without that
beneficial trait. Conceptually, taxon survival can be considered an aspect of
62 “taxon fitness” along with expected lineage specific branching/origination rate
(Cooper, 1984, Palmer and Feldman, 2012). Brachiopods are an ideal group for
64 this study as they are well known for having an exceptionally complete fossil
records (Foote, 2000). Specifically, I focus on the brachiopod record from most

66 of the Paleozoic, specifically from the start of the Ordovician (approximately
485 Mya) through the end Permian (approximately 252 Mya) as this represents
68 the time of greatest global brachiopod diversity (Alroy, 2010).

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

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

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

84 Epicontinental seas are a shallow-marine environment where the ocean has
spread over the surface of a continental shelf with a depth typically less than
86 100m. In contrast, open-ocean coastline environments have much greater
variance in depth, do not cover the continental shelf, and can persist during
88 periods of low sea level. Because of this, it is strongly expected that taxa which
favor epicontinental seas would be at great risk during periods of low sea levels,
90 such as during glacial periods, where these seas are drained. During the
Paleozoic (approximately 541–252 My), epicontinental seas were widely spread

⁹² globally but declined over the Mesozoic (approximately 252–66 My) and
⁹³ eventually diminished disappearing during the Cenozoic (approximately 66–0
⁹⁴ My) as open-ocean coastlines became the dominant shallow-marine setting
(Johnson, 1974, Miller and Foote, 2009, Peters, 2008).

⁹⁶ Given the above information, I predict that as extinction risk increases, taxa
⁹⁷ associated with open-ocean environments should generally increase in extinction
⁹⁸ risk versus those that favor epicontinental seas. Additionally, there is a possible
⁹⁹ nonlinear relationship between environmental preference and taxon duration. A
¹⁰⁰ long standing hypothesis is that generalists or unspecialized taxa will have
¹⁰¹ greater survival than specialists (Baumiller, 1993, Liow, 2004, 2007, Nürnberg
¹⁰² and Aberhan, 2013, 2015, Simpson, 1944). In this analysis I allowed for
¹⁰³ environmental preference to possibly have a parabolic effect on taxon duration
¹⁰⁴ Body size, measured as shell length (Payne et al., 2014), was also considered as
¹⁰⁵ a potentially informative covariate. Body size is a proxy for metabolic activity
¹⁰⁶ and other correlated life history traits (Payne et al., 2014). There is no strong
¹⁰⁷ hypothesis of how body size effects extinction risk in brachiopods, meaning a
¹⁰⁸ positive, negative, or zero effect are all plausible.

I adopt a hierarchical Bayesian survival modeling approach, which represents a
¹¹⁰ conceptual and statistical unification of the paleontological dynamic and cohort
survival analytic approaches (Baumiller, 1993, Foote, 1988, Raup, 1975, 1978,
¹¹² Simpson, 2006, Van Valen, 1973, 1979). By using a Bayesian framework I am
¹¹³ able to quantify the uncertainty inherent in the estimates of the effects of
¹¹⁴ biological traits on survival, especially in cases where the covariates of interest
(i.e. biological traits) are themselves known with error.

₁₁₆ **2 Materials and Methods**

2.1 Fossil occurrence information

₁₁₈ The dataset analyzed here was sourced from the Paleobiology Database
(<http://www.paleodb.org>) which was then filtered based on taxonomic,
₁₂₀ temporal, stratigraphic, and other occurrence information that was necessary
for this analysis. These filtering criteria are very similar to those from Foote and
₁₂₂ 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
₁₂₄ for each fossil occurrence are partially based on those from Miller and Foote
(2009), with additional occurrences assigned similarly (Miller and Foote,
₁₂₆ personal communication).

Fossil occurrences were analyzed at the genus level which is common for
₁₂₈ paleobiological, macroevolution, or macroecological studies of marine
invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2013, Kiessling
₁₃₀ and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,
Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).

₁₃₂ While species diversity dynamics are of much greater interest than those of
higher taxa, the nature of the fossil record makes accurate and precise
₁₃₄ taxonomic assignments at the species level for all occurrences. In particular, the
simplicity of brachiopod external morphology and the quality of preservation
₁₃₆ 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
₁₃₈ of confidence and accuracy in the data analyzed here.

Sampled occurrences were restricted to those with paleolatitude and
₁₄₀ paleolongitude coordinates, assignment to either epicontinental or open-ocean

environment, and belonging to a genus present in the body size dataset (Payne
142 et al., 2014). Genus duration was calculated as the number of geologic stages
from first appearance to last appearance, inclusive. Durations were based on
144 geologic stages as opposed to millions of years because of the inherently discrete
nature of the fossil record; dates are not assigned to fossils themselves but
146 instead fossils are known from a geological interval which represents some
temporal range. Stages are effectively irreducible temporal intervals in which
148 taxa may occur.

Genera with a last occurrence in or after Changhsingian stage were right
150 censored at the Changhsingian. Genera with a duration of only one stage were
left censored (Appendix C). The covariates used to model genus duration were
152 geographic range size (r), environmental preference (v, v^2), and body size (m).

Geographic range was calculated using an occupancy approach. First, all
154 occurrences were projected onto an equal-area cylindrical map projection. Each
occurrence was then assigned to one of the cells from a 70×34 regular raster
156 grid placed on the map. Each grid cell represents approximately $250,000 \text{ km}^2$.
The map projection and regular lattice were made using shape files from
158 <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,
2015).

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

166 Body size data was sourced directly from Payne et al. (2014). Because those

measurements are presented without error, a measurement error model similar
168 to the one for environmental affinity could not be implemented (Appendix A).

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

2.2 Analytical approach

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

188 This partial pooling is one of the greatest advantages of hierarchical modeling.
By letting the groups “support” each other, parameter estimates then better
190 reflect our statistical uncertainty. Additionally, this partial pooling helps control
for multiple comparisons and possibly spurious results as effects with little

¹⁹² support are drawn towards the overall group mean (Gelman et al., 2013,
Gelman and Hill, 2007).

¹⁹⁴ All covariate effects (regression coefficients), as well as the intercept term
(baseline extinction risk), were allowed to vary by group (origination cohort).

¹⁹⁶ The covariance/correlation between covariate effects was also modeled. This
hierarchical structure allows inference for how covariates effects may change
¹⁹⁸ with respect to each other while simultaneously estimating the effects
themselves, propagating our uncertainty through all estimates.

²⁰⁰ Additionally, instead of relying on point estimates of environmental affinity, I
treat environmental affinity as a continuous measure of the difference between
²⁰² the taxon's environmental occurrence pattern and the background occurrence
pattern (Appendix A).

²⁰⁴ 2.3 Survival model

Genus durations were modeled as time-till-event data (Klein and Moeschberger,
²⁰⁶ 2003), with covariate information used in estimates of extinction risk as a
hierarchical regression model. Genus durations were assumed to follow either an
²⁰⁸ exponential or Weibull distribution. THe use of either of these distributions
makes assumptions about how duration may effect extinction risk (Klein and
²¹⁰ Moeschberger, 2003). The exponential distribution assumes that extinction risk
is independent of duration. In contrast, the Weibull distribution allows for age
²¹² dependent extinction via the shape parameter α , though only as a monotonic
function of duration. Importantly, the Weibull distribution is equivalent to the
²¹⁴ exponential distribution when $\alpha = 1$.

The following variables are defined: y_i is the duration of genus i in geologic
²¹⁶ stages, X is the matrix of covariates including a constant term, B_j is the vector

of regression coefficients for origination cohort j , Σ is the covariance matrix of
 218 the regression coefficients, τ is the vector of scales the standard deviations of
 the between-cohort variation in regression coefficient estimates, and Ω is the
 220 correlation matrix of the regression coefficients.

The exponential model is defined

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

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

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

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

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

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

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

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

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

²²² Similarly, the Weibull model is defined

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

The principal difference between this model and the previous (Eq. 1) is the
²²⁴ inclusion of the shape parameter α . Note that σ is approximately equivalent to
 $1/\lambda$.

²²⁶ For an explanation of how this model was developed, parameter explanations,
and choice of priors, please see Appendix B. Note that these models (Eq. 1, 2)
²²⁸ do not include how the uncertainty in environmental affinity is included nor how
censored observations are included. For an explanation of both of these aspects,
²³⁰ see Appendices A and C.

2.4 Parameter estimation

²³² The joint posterior was approximated using a Markov-chain Monte Carlo
²³³ routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
²³⁴ Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic
²³⁵ programming language Stan (Stan Development Team, 2014a). The posterior
²³⁶ distribution was approximated from four parallel chains run for 10,000 draws
²³⁷ each, split half warm-up and half sampling and thinned to every 10th sample for
²³⁸ a total of 5000 posterior samples. Chain convergence was assessed via the scale
²³⁹ reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$) indicate approximate
²⁴⁰ convergence. Convergence means that the chains are approximately stationary
and the samples are well mixed (Gelman et al., 2013).

2.5 Model evaluation

²⁴² Models were evaluated using both posterior predictive checks and an estimate of
²⁴³ out-of-sample predictive accuracy. The motivation behind posterior predictive
²⁴⁴ checks as tools for determining model adequacy is that replicated data sets
²⁴⁵ using the fitted model should be similar to the original data (Gelman et al.,
²⁴⁶ 2013). Systematic differences between the simulations and observations indicate
²⁴⁷ weaknesses of the model fit. An example of a technique that is very similar
would be inspecting the residuals from a linear regression.

²⁴⁸ The strategy behind posterior predictive checks is to draw simulated values
²⁴⁹ from the joint posterior predictive distribution, $p(y^{rep}|y)$, and then compare
²⁵⁰ those draws to the empirically observed values (Gelman et al., 2013). To
²⁵¹ accomplish this, for each replicate, a single value is drawn from the marginal
²⁵² posterior distributions of each regression coefficient from the final model as well
²⁵³ as α for the Weibull model (Eq. 1, 2). Then, given the covariate information \mathbf{X} ,

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

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

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

276 3 Results

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

²⁸⁰ Comparisons of the survival functions estimated from 1000 posterior predictive
data sets to the estimated survival function of the observed genera demonstrates
²⁸² that both the exponential and Weibull models approximately capture the
observed pattern of extinction (Fig. 1). The major difference in fit between the
²⁸⁴ two models is that the Weibull model has a slightly better fit for longer lived
taxa than the exponential model.

²⁸⁶ Additionally, the Weibull model is expected to have slightly better out-of-sample
predictive accuracy when compared to the exponential model (WAIC 4576
²⁸⁸ versus 4604, respectively). Because the difference in WAIC between these
two models is large, while results from both the exponential and Weibull models
²⁹⁰ will be presented, only those from the Weibull model will be discussed.

Estimates of the overall mean covariate effects μ can be considered
²⁹² time-invariant generalizations for brachiopod survival during the Paleozoic (Fig.
1). Consistent with prior expectations, geographic range size has a negative
²⁹⁴ effect on extinction risk, where genera with large ranges having greater
durations than genera with small ranges.

²⁹⁶ I find that while the mean estimate of the effect of body size on extinction risk
is negative, implying that increasing body size decreases extinction risk, this
²⁹⁸ estimate is within 2 standard deviations of 0 (mean $\mu_m = -0.09$, standard
deviation 0.09; Fig. 1). Because of this, I infer that body size has no
³⁰⁰ distinguishable effect on brachiopod taxonomic survival.

Interpretation of the effect of environmental preference v on duration is slightly
³⁰² more involved. Because a quadratic term is the equivalent of an interaction
term, both μ_v and μ_{v^2} have to be interpreted together because it is illogical to
³⁰⁴ change values of v without also changing values v^2 . To determine the nature of
the effect of v on duration I calculated the multiplicative effect of environmental

³⁰⁶ preference on extinction risk.

Given mean estimated extinction risk $\tilde{\sigma}$, we can define the extinction risk
³⁰⁸ multiplier of an observation with environmental preference v_i as

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

This function $f(v_i)$ has a y-intercept of $\exp(0)$ or 1 because it does not have a
³¹⁰ non-zero intercept term. Equation 3 can be either concave up or down. A
concave down $f(v_i)$ may indicate that genera of intermediate environmental
³¹² preference have greater durations than either extreme, and *vice versa* for
concave up function.

³¹⁴ The expected effect of environmental preference as a multiplier of expected
extinction risk can then be visualized (Fig. 2). This figure depicts 1000 posterior
³¹⁶ predictive estimates of Eq. 3 across all possible values of v . The number
indicates the posterior probability that the function is concave down, with
³¹⁸ generalists having lower extinction risk/greater duration than either type of
specialist. Note that the inflection point/optimum of Fig. 2 is approximately
³²⁰ $x = 0$, something that is expected given the estimate of μ_v (Fig. 1).

The matrix Σ describing the covariance between the different coefficients
³²² describes how these coefficients might vary together across the origination
cohorts. Similar to how this was modeled (Eq. 1, 2), for interpretation purposes
³²⁴ Σ can be decomposed into a vector of standard deviations $\vec{\tau}$ and a correlation
matrix Ω .

³²⁶ The estimates of the standard deviation of between-cohort coefficient estimates
 τ indicate that some effects can vary greatly between-cohorts (Fig. 1).
³²⁸ Coefficients with greater values of τ have greater between-cohort variation. The
covariate effects with the greatest between origination cohort variation are β_r ,

³³⁰ β_v , and β_{v^2} . Estimates of β_m have negligible between cohort variation, as there
is less between cohort variation than the between cohort variation in baseline
³³² extinction risk β_0 . However the amount of between cohort variation in estimates
of β_{v^2} means that it is possible for the function describing the effect of
³³⁴ environmental affinity to be upward facing for some cohorts (Eq. 3), which
corresponds to environmental generalists being shorter lived than specialists in
³³⁶ that cohort.

The correlation terms of Ω (Fig. 3a) describe the relationship between the
³³⁸ coefficients and how their estimates may vary together across cohorts. The
correlations between the intercept term β_0 and the effects of the taxon traits are
³⁴⁰ of particular interest for evaluating the Jablonski (1986) scenario (Fig. 3a first
column/last row). Keep in mind that when β_0 is low, extinction risk is low; and
³⁴² conversely, when β_0 is high, then extinction risk is high.

Marginal posterior probabilities of the correlations between the level of baseline
³⁴⁴ extinction risk β_0 and the effects of the taxon traits indicate that the correlation
between expected extinction risk and both geographic range β_r and β_{v^2} are of
³⁴⁶ particular note (Fig. 3b).

There is approximately a 98% probability that β_0 and β_r are negatively
³⁴⁸ correlated (Fig. 3b), meaning that as extinction risk increases, the
effect/importance of geographic range on genus duration increases. This means
³⁵⁰ that increases in baseline extinction rate are correlated with an increased
importance of geographic range size. There is a 93% probability that β_0 and β_{v^2}
³⁵² are negatively correlated (Fig. 3b), meaning that as extinction risk increases,
the peakedness of $f(v_i)$ increases and the relationship tends towards concave
³⁵⁴ down. Additionally, there is a 97% probability that values of β_r and β_{v^2} are
positively correlated (Mean correlation 0.51, standard deviation 0.23).

³⁵⁶ While the overall group level estimates are of particular importance when
³⁵⁷ defining time-invariant differences in extinction risk, it is also important and
³⁵⁸ useful to analyze the individual level parameter estimates in order to better
understand how parameters actually vary across cohorts.

³⁶⁰ In comparison to the overall mean extinction risk μ_0 , cohort level estimates β_0
show some amount of variation through time as expected by estimates of τ_0
³⁶² (Fig. 4a). A similar, if slightly greater, amount of variation is also observable in
cohort estimates of the effect of geographic range β_r (Fig. 4b). Again, smaller
³⁶⁴ values of β_0 correspond to lower expected extinction risk. Similarly, smaller
values of β_r correspond to greater decrease in extinction risk with increasing
³⁶⁶ geographic range

How the effect of environmental affinity varies between cohorts can be observed
³⁶⁸ by using the cohort specific coefficients estimates. Following the same procedure
used earlier (Fig. 1), but substituting cohort specific estimates of β_v and β_{v^2} for
³⁷⁰ μ_v and μ_{v^2} , the cohort specific effect of environmental preference as a multiplier
of mean extinction risk can be calculated. This was done only for the Weibull
³⁷² model, though the observed pattern should be similar for the exponential model.

As expected based on the estimates of τ_v and τ_{v^2} , there is greater variation in
³⁷⁴ the peakedness of $f(v_i)$ than there is variation between convave up and down
functions (Fig. 5). 12 of the 33 cohorts have less than 50% posterior probability
³⁷⁶ that generalists are potentially expected to be shorter lived than specialists,
though two of those cases have approximately a 50% probability of being either
³⁷⁸ concave up or down. This is congruent with the 0.72 posterior probability that
 μ_{v^2} is positive/ $f(v_i)$ is concave down.

³⁸⁰ Additionally, for some cohorts there is a quite striking pattern where the effect
of environmental preference v has a nearly-linear relationship (Fig. 5). These are

³⁸² primarily scenarios where one of the end member preferences is expected to
³⁸³ have a greater duration than either intermediate or the opposite end member
³⁸⁴ preference. Whatever curvature is present in these nearly-linear cases is due to
³⁸⁵ the definition of $f(v)$ as it is not defined for non-negative values of σ (Eq. 3). For
³⁸⁶ all stages between the Emsian through the Viséan, inclusive, intermediate
³⁸⁷ preferences are of intermediate extinction risk when compared with
³⁸⁸ epicontinental specialists (lowest risk) or open-ocean specialists (highest risk).
This time period represents most of the Devonian through the early
³⁹⁰ Carboniferous.

4 Discussion

³⁹² My results demonstrate that both the effects of geographic range and the
³⁹³ peakedness/concavity of environmental preference are both negatively
³⁹⁴ correlated with baseline extinction risk, meaning that as baseline extinction risk
increases the effect sizes of both these traits are expected to increase (Fig. 3b).
³⁹⁶ This result supports neither of the two proposed macroevolutionary mechanisms
for how biological traits should correlate with extinction risk. The observed
³⁹⁸ correlation between the two effects as well as between the effects and baseline
extinction risk instead implies that as baseline extinction risk increases, the
⁴⁰⁰ strength of the total selection gradient on biological traits (except body size)
increases. This manifests as greater differences in extinction risk for each unit
⁴⁰² difference in the biological covariates during periods of high extinction risk,
while a relatively flatter selection gradient during periods of low extinction risk.
⁴⁰⁴ For the approximately 233 My period analyzed there is an approximate 75%
posterior probability that brachiopod genera with intermediate environmental
⁴⁰⁶ preferences are expected to have a lower extinction risk than either end

members. However, the over all curvature of $f(v_i)$ is not very peaked meaning
408 that when averaged over the entire Phanerozoic this relationship may not lead
to large differences in extinction risk (Fig. 2). Note that the duration of the
410 period analyzed is approximately four times then length of the Cenozoic (e.g.
time since the extinction of the non-avian dinosaurs). This result gives weak
412 support for the universality of the hypothesis that environmental generalists
have greater survival than environmental specialists (Liow, 2004, 2007,
414 Nürnberg and Aberhan, 2013, 2015, Simpson, 1944).

The posterior variance in the estimate of overall $f(v_i)$ reflects the large between
416 cohort variance in cohort specific estimates of $f(v_i)$ (Fig. 5). Given that there is
only a 75% posterior probability that the expected overall estimate of $f(v_i)$ is
418 concave down, it is not surprising that there are some stages where the
estimated relationship is in fact the reverse of the prior expectation.

420 Additionally, some of those same stages where $f(v_i)$ does not resemble the prior
expectation of a concave down nonlinear relation are instead is highly skewed
422 and effectively linear (Fig. 5). These results demonstrate that, while the
group-level estimate may only weakly support one hypothesis, the cohort-level
424 estimates may exhibit very different characteristics. These results are also
consistent with aspects of Miller and Foote (2009) who found that the effect of
426 environmental preference on extinction risk was quite variable and without
obvious patterning during times of background extinction.

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

The proposed mechanism for the end Ordovician mass extinction is a decrease
432 in sea level and the draining of epicontinental seas due to protracted glaciation
(Johnson, 1974, Sheehan, 2001). My results are broadly consistent with this

434 scenario with both epicontinental and open-ocean specialists having a much
lower expected duration than intermediate taxa (Fig. 5). All of the stages
436 between the Darriwillian and the Llandovery, except the Hirnantian, have a
high probability (90+%) that $f(v)$ is concave down. The pattern for the
438 Darriwillian, which proceeds the supposed start of Ordovician glacial activity,
demonstrates that taxa tend to favor open-ocean environments are expected to
440 have a greater duration than either intermediate or epicontinental specialists, in
decreasing order.

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

454 Of concern is the use of genera as the unit of the study and how to exactly
interpret the effects of the biological traits. For example, if any of the traits
456 analyzed here are associated with increases in speciation rates, this might
“artificially” increase the duration of genera through self-renewal (Raup, 1991,
458 1994). This could lead to a trait appearing to decrease generic level extinction
risk by increasing species level origination rate instead of decreasing species
460 level extinction risk. However, given the nature of the brachiopod fossil record

and the difficulty of identifying individual specimens to the species level, there
462 is no simple solution to decreasing this uncertainty in the interpretations of how
the biological traits studied here actually affect extinction risk.

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

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

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

Taxon traits like environmental preference or geographic range (Hunt et al.,
480 2005, Jablonski, 1987) are most likely heritable, at least phylogenetically
(Housworth et al., 2004, Lynch, 1991). Without phylogenetic context, this
482 analysis assumes that differences in extinction risk between taxa are
independent of those taxa's shared evolutionary history (Felsenstein, 1985). In
484 contrast, the origination cohorts only capture shared temporal context. The
inclusion of phylogenetic context as an addition individual level hierarchical
486 structure independent of origination cohort would allow for determining how

much of the observed variability is due to shared evolutionary history versus
488 actual differences associated with these taxonomic traits.

In summary, patterns of Paleozoic brachiopod survival were analyzed using a
490 fully Bayesian hierarchical survival modelling approach while also eschewing the
traditional separation between background and mass extinction. I modeled both
492 the overall mean effect of biological covariates on extinction risk while also
modeling the correlation between cohort-specific estimates of covariate effects. I
494 find that as baseline extinction risk increases, the strength of the selection
gradient on biological traits (except body size) increases. This manifests as
496 greater differences in extinction risk for each unit difference in the biological
covariates during periods of high extinction risk, while a much flatter total
498 selection gradient during periods of low extinction risk. I also find some support
for “survival of the unspecialized” (Liow, 2004, 2007, Nürnberg and Aberhan,
500 2013, 2015, Simpson, 1944) as a general characterization of the effect of
environmental preference on extinction risk (Fig. 2), though there is
502 heterogeneity between origination cohorts with most periods of time conforming
to this hypothesis (Fig. 5).

504 Acknowledgements

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for
506 helpful discussion and advice. Additionally, thank you A. Miller for the
epicontinental versus open-ocean assignments. This entire study would
508 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.
510 Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland,
L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T. Olszewski, P.

⁵¹² 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
⁵¹⁴ M. Foote. This is Paleobiology Database publication XXX.

References

- 516 Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
to the mode of stabilization on the substrate. *Palaeogeography,*
518 *Palaeoclimatology, Palaeoecology* 21:209–226.
- Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal
520 Groups. *Science* 329:1191–1194.
- Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of
522 filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- Carvalho, C. M., N. G. Polson, and J. G. Scott, 2009. Handling Sparsity via the
524 Horseshoe. *in* Proceedings of the 12th International Conference on Artificial
Intelligence and Statistics, vol. 5, Pp. 73–80.
- 526 ———, 2010. The horseshoe estimator for sparse signals. *Biometrika*
97:465–480.
- 528 Cooper, W. S., 1984. Expected time to extinction and the concept of
fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- 530 Felsenstein, J., 1985. Phylogenies and the comparative method. *American
Naturalist* 125:1–15.
- 532 Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites.
Paleobiology 14:258–271.
- 534 ———, 2000. Origination and extinction components of taxonomic diversity:
Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.
- 536 ———, 2006. Substrate affinity and diversity dynamics of Paleozoic marine
animals. *Paleobiology* 32:345–366.

- 538 Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine
animal genera. *Paleobiology* 39:171–192.
- 540 Gelman, A., 2006. Prior distributions for variance parameters in hierarchical
models. *Bayesian Analysis* 1:515–533.
- 542 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,
544 FL.
- Gelman, A. and J. Hill, 2007. Data Analysis using Regression and
546 Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- Harnik, P. G., C. Simpson, and J. L. Payne, 2013. Long-term differences in
548 extinction risk among the seven forms of rarity. *Proceedings of the Royal
Society B: Biological Sciences* 282.
- 550 Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL
<http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- 552 Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively
Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine
554 Learning Research* 15:1351–1381.
- Housworth, E. A., P. Martins, and M. Lynch, 2004. The Phylogenetic Mixed
556 Model. *The American Naturalist* 163:84–96.
- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
558 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.

- 562 Jablonski, D., 1986. Background and mass extincitons: the alternation of
macroevolutionary regimes. *Science* 231:129–133.
- 564 ———, 1987. Heritability at the species level: analysis of geographic ranges of
cretaceous mollusks. *Science* 238:360–363.
- 566 Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and
living molluscs. *Proceedings. Biological sciences / The Royal Society*
568 270:401–6.
- Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- 570 Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
benthic biodiversity dynamics through Triassic Jurassic time. *Paleobiology*
572 33:414–434.
- 574 Klein, J. P. and M. L. Moeschberger, 2003. *Survival Analysis: Techniques for
Censored and Truncated Data*. 2nd ed. Springer, New York.
- 576 Liow, L. H., 2004. A test of Simpson's "rule of the survival of the relatively
unspecialized" using fossil crinoids. *The American naturalist* 164:431–43.
- 578 ———, 2007. Does versatility as measured by geographic range, bathymetric
range and morphological variability contribute to taxon longevity? *Global
Ecology and Biogeography* 16:117–128.
- 580 Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary
biology. *Evolution* 45:1065–1080.
- 582 Miller, A. I. and S. R. Connolly, 2001. Substrate affinities of higher taxa and
the Ordovician Radiation. *Paleobiology* 27:768–778.
- 584 Miller, A. I. and M. Foote, 2009. Epicontinent seas versus open-ocean settings:
the kinetics of mass extinction and origination. *Science* 326:1106–9.

- 586 Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range
predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*
588 39:360–372.
- 590 ———, 2015. Interdependence of specialization and biodiversity in Phanerozoic
marine invertebrates. *Nature communications* 6:6602.
- 592 Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental
than evolvability. *PloS one* 7:e38025.
- 594 Payne, J. L. and S. Finnegan, 2007. The effect of geographic range on
extinction risk during background and mass extinction. *Proceedings of the
National Academy of Sciences* 104:10506–11.
- 596 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
598 million years. *Proceedings of the Royal Society B* 281:20133122.
- 600 Peters, S. E., 2008. Environmental determinants of extinction selectivity in the
fossil record. *Nature* 454:626–9.
- 602 Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen's Law.
Paleobiology 1:82–96.
- 604 ———, 1978. Cohort Analysis of generic survivorship. *Paleobiology* 4:1–15.
- 606 ———, 1991. Extinction: Bad Genes or Bad Luck? Norton, New York.
- 608 ———, 1994. The role of extinction in evolution. *Proceedings of the National
Academy of Sciences* 91:6758–6763.
- Sepkoski Jr., J. J., 1981. A factor analytic description of the Phanerozoic
marine fossil record. *Paleobiology* 7:36–53.

- Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of Earth and Planetary Sciences* 29:331–364.
- Simpson, C., 2006. Levels of selection and large-scale morphological trends. Ph.D. thesis, University of Chicago.
- Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University Press, New York.
- , 1953. *The Major Features of Evolution*. Columbia University Press, New York.
- Stan Development Team, 2014a. Stan: A c++ library for probability and sampling, version 2.5.0. URL <http://mc-stan.org/>.
- , 2014b. Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0. URL <http://mc-stan.org/>.
- Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- , 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.
- 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 latitudinal selectivity gradient at the end-Cretaceous mass extinction. *Scientific Reports* 3:1790.
- Wang, S. C., 2003. On the continuity of background and mass extinction. *Paleobiology* 29:455–467.
- Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and

- ⁶³² Widely Applicable Information Criterion in Singular Learning Theory.
Journal of Machine Learning Research 11:3571–3594.

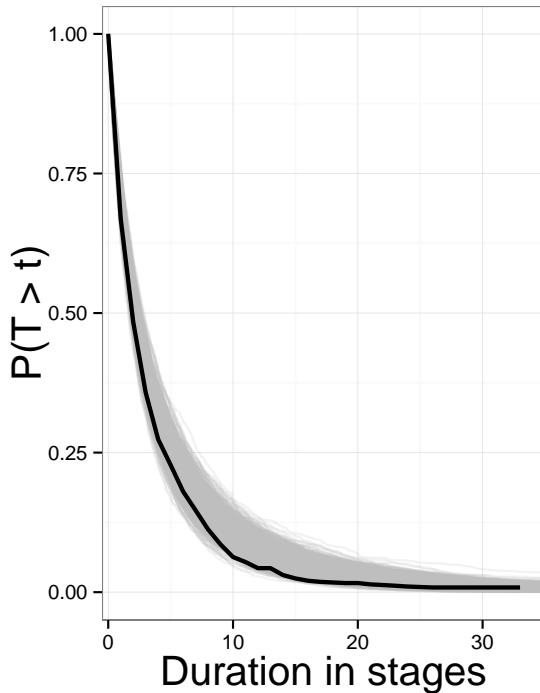


Figure 1: Comparison of empirical estimates of $S(t)$ versus estimates from 1000 posterior predictive data sets. $S(t)$ corresponds to $P(T > t)$ as it is the probability that a given genus observed at age t will continue to live. This is equivalent to the probability that t is less than the genus' ultimate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

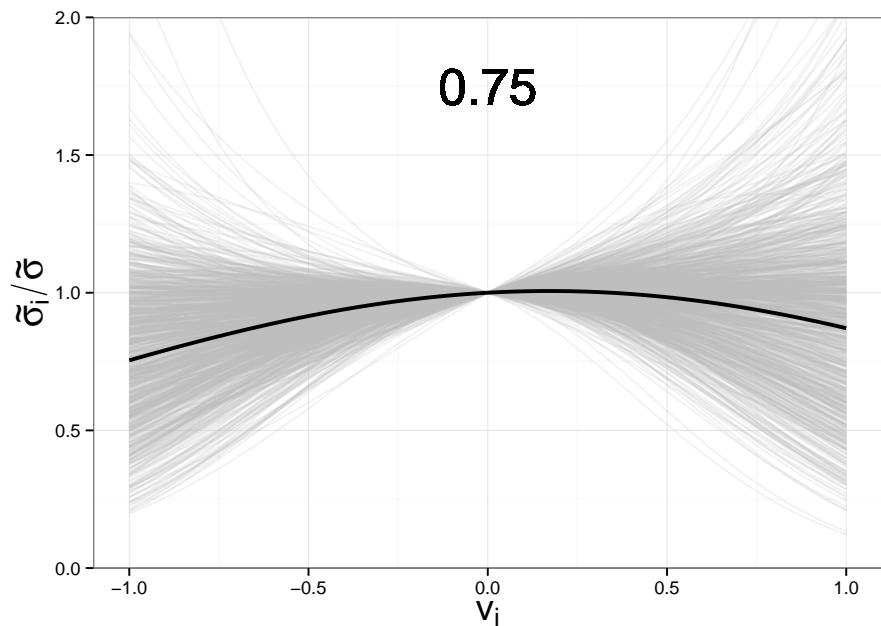


Figure 2: The overall expected relationship $f(v_i)$ between environmental affinity v_i and a multiplier of extinction risk (Eq. 3). Each grey line corresponds to a single draw from the posterior predictive distribution, while the black corresponds to the median of the posterior predictive distribution. 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.

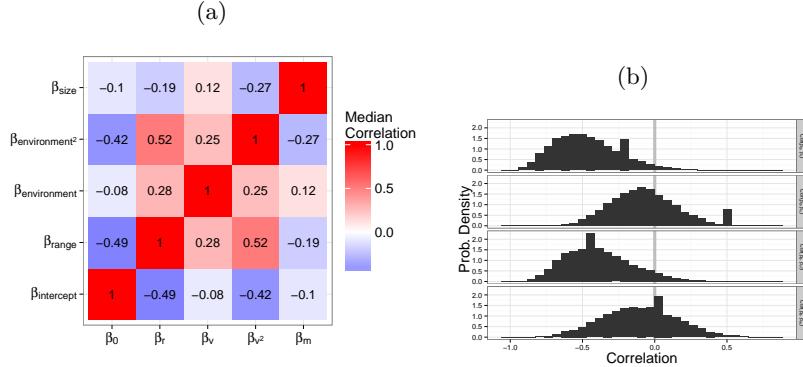


Figure 3: **A:** Heatmap for the median estimates of the terms of the correlation matrix Ω between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the 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.

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_{e^2}	1.88	0.66
τ_m	0.32	0.13

Table 1: Group-level estimates of the intercept terms the effects of biological traits on brachiopod generic survival from equations 1 and 2, presented as means and standard deviations. μ values are the location parameters of the effects, while τ values are the scale terms describing the variation between cohorts. The subscripts correspond to the following: i intercept, r geographic range, e environmental affinity, e^2 environmental affinity squared, m body size.

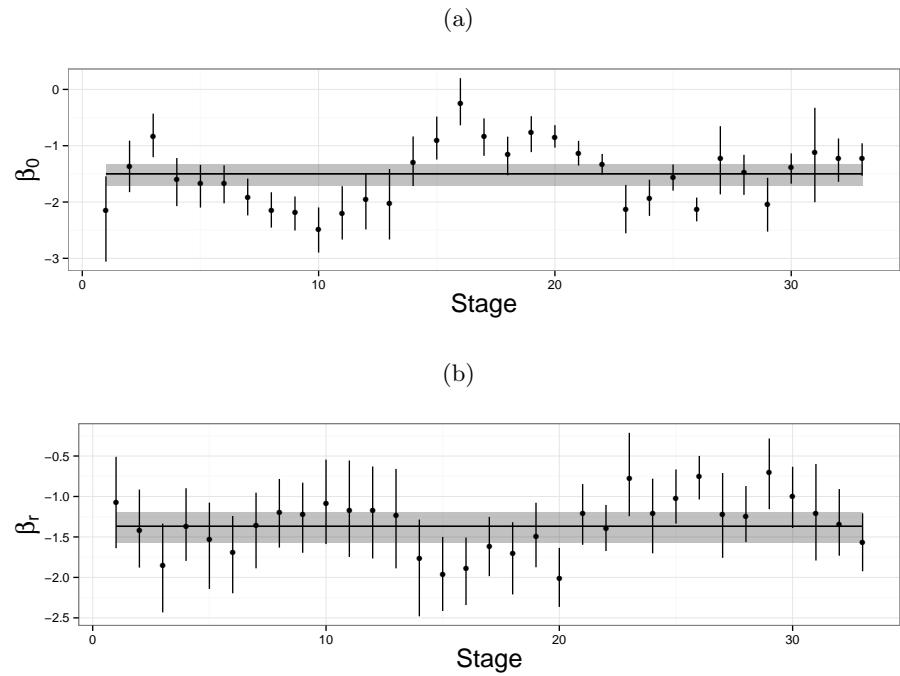


Figure 4: Comparison of cohort-specific estimates of β_0 presented along with the estimate for the overall baseline extinction risk. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction 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.

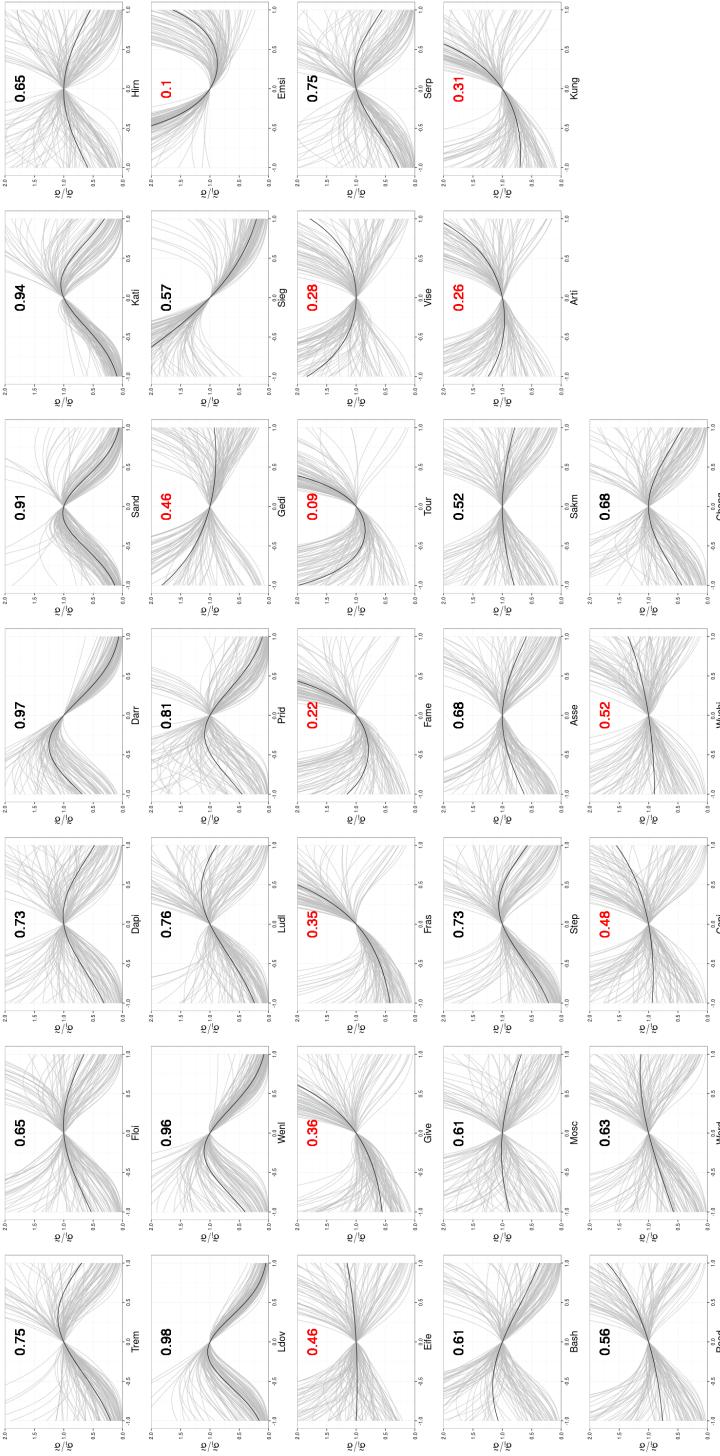


Figure 5: Comparison of the cohort-specific estimates of $f(v_i)$ (Eq. 3) for the 33 analyzed origination cohorts. The stage of origination is labeled on the x-axis of each panel. The oldest stage is in the upper left, while the youngest is in the lower left. The number in each panel corresponds to the 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.

634 **A Uncertainty in environmental preference**

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

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

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

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

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

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

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

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

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

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

While this approach is noticeably more complicated than previous ones (Foote,
 670 2006, Kiessling and Aberhan, 2007, Miller and Connolly, 2001, Simpson and
 Harnik, 2009) there are some important benefits to both using a continuous
 672 measure of affinity as well directly modeling our uncertainty. In order to show
 some of these benefits, I performed a simulation analysis of how
 674 modal/maximum *a posteriori* (MAP) estimates versus full posterior estimates.

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

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

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

680 Using this background occurrence ratio, I randomly generated the occurrence
patterns of 1000 simulated taxa. This was done at multiple sample sizes (1, 2, 3,
682 4, 5, 10, 25, 50, 100) in order to demonstrate the effects of increasing sample
size on the confidence of environmental affinity. For each simulated taxon I
684 calculated the full posterior distribution while assuming a flat Beta prior
(Beta(1,1)). Using the full posterior I calculated the MAP probability of
686 occurring in epicontinental environments. The environmental affinity was
calculated for each of the simulated taxa using both the full posterior and the
688 MAP estimate. In this toy example, environmental affinity can range between
-0.5 and 0.5.

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

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

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

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

714 B Survival model

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

$$y_i \sim \text{Exponential}(\lambda) \quad (9)$$
$$y_i \sim \text{Weibull}(\alpha, \sigma).$$

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

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

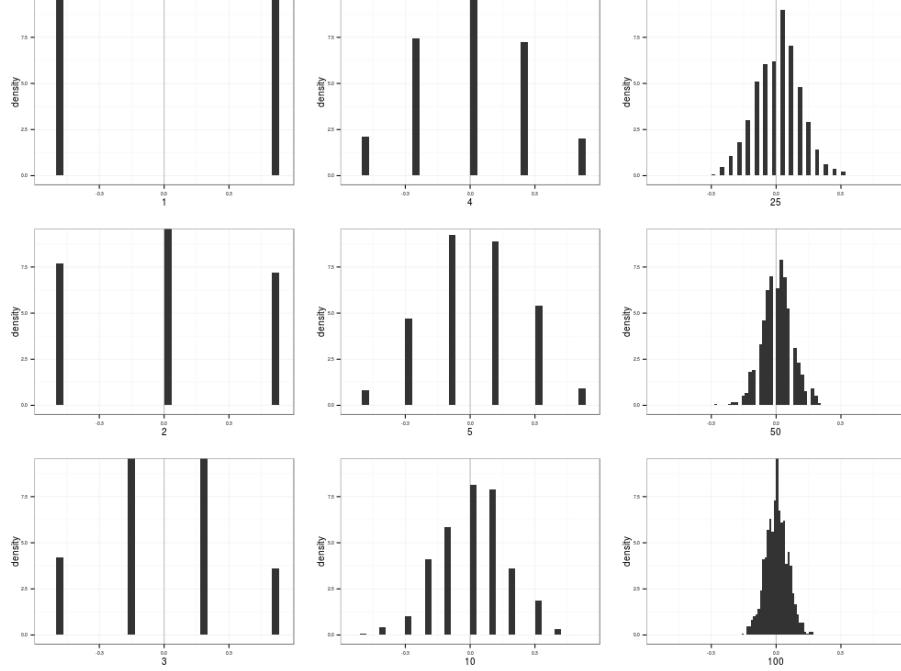


Figure 6: Histograms of the distributions of from the beta distribution defined in Eq. 8. As to be expected, as sample size increases the draws better resemble the underlying true distribution. Sample size is indicated as the label of the x-axis, increasing in column major order.

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

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

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

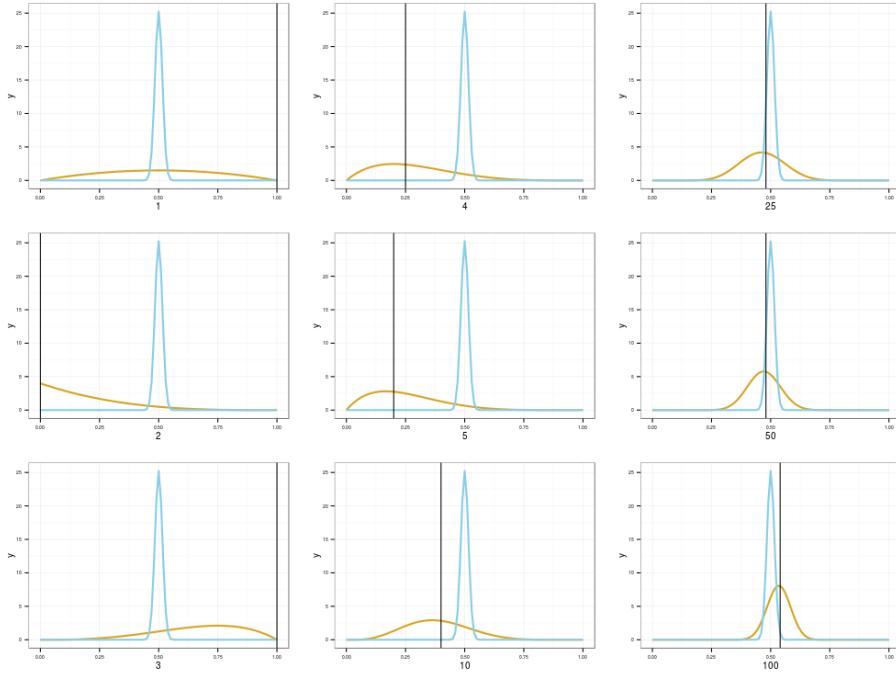


Figure 7: Comparisons of the underlying distribution (blue) to posterior estimates based on increasing sample size (gold). Each posterior estimate is represented for only a single realization of draws, each with sample size indicated as the x-axis label (increasing in column major order). Black vertical 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.

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

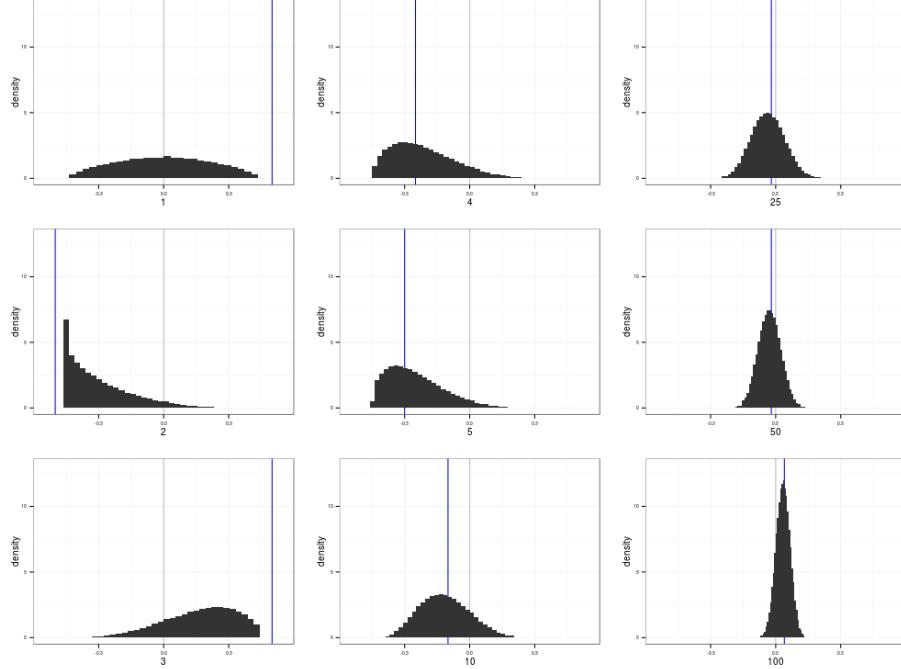


Figure 8: Histograms of the difference in the underlying occurrence distribution and the posterior distribution estimates from the previous graph (Fig. 7). The “true” value is included in all distributions of environmental affinities. Each affinity estimate is represented for only a single realization 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.

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

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

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

⁷⁴² of means μ and covariance matrix Σ . This is written as

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

⁷⁴³ where $\vec{\mu}$ is length k vector representing the overall mean of the distributions of
⁷⁴⁴ β coefficients. Σ is a $k \times k$ covariance matrix of the β coefficients.

What remains is assigning priors the elements of $\vec{\mu}$ and the covariance matrix Σ .

⁷⁴⁵ All elements of $\vec{\mu}$ except for μ_r were given horseshoe priors (Carvalho et al.,
2009, 2010) while μ_r was given an informative normal prior ($\mu_r \sim \mathcal{N}(-1, 1)$).
⁷⁴⁶ Horseshoe priors are a strong regularizing priors with effectively infinite density
at 0 and heavy, Cauchy-like tails (Carvalho et al., 2009, 2010) which allow
⁷⁴⁷ weakly inferred effects to be strongly drawn towards 0 while truly strong effects
can remain large. The horseshoe prior consists of a normal distribution with
⁷⁴⁸ scale term that is the product between a global shrinkage parameter ν and a
local shrinkage parameter ψ unique to each of the parameters of interest. These
⁷⁴⁹ parameters are themselves given half-Cauchy priors (Eq. 1 and 2).
⁷⁵⁰

The prior for Σ is a bit more complicated due to its multivariate nature.

⁷⁵¹ Following the Stan Development Team (2014b), I modeled the scale terms
separate from the correlation structure of the coefficients. This is possible
⁷⁵² because of the relationship between a covariance and a correlation matrix,
defined as

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

⁷⁵³ where $\vec{\tau}$ is a length k vector of variances and $\text{Diag}(\tau)$ is a diagonal matrix.

I used a LKJ prior distribution for correlation matrix Ω as recommended by
⁷⁵⁴ Stan Development Team (2014b). The LKJ distribution is a single parameter
multivariate distribution where values of the parameter η greater than 1

⁷⁶⁴ concentrate density at the unit correlation matrix, which corresponds to no
correlation between the β coefficients. The scale parameters, $\vec{\tau}$, are given weakly
⁷⁶⁶ informative half-Cauchy (C^+) priors following Gelman (2006).

C Censored observations

⁷⁶⁸ 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
⁷⁷⁰ two classes of censored observations encountered in this study were right and
left censored observations. Right censored genera are those that did not go
⁷⁷² 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
⁷⁷⁴ their duration.

In the context of this study, I considered all genera that had a duration of only
⁷⁷⁶ one geologic stage to be left censored as we do not have a finer degree of
resolution.

⁷⁷⁸ 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
⁷⁸⁰ will not have gone extinct while $h(t)$ corresponds to the instantaneous
extinction rate at taxon age t Klein and Moeschberger (2003). For an
⁷⁸² exponential model, $S(t)$ is defined as

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

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

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

⁷⁸⁴ $S(t)$ is equivalent to the complementary cumulative distribution function,
 $1 - F(t)$ (Klein and Moeschberger, 2003).

⁷⁸⁶ For right censored observations, instead of calculating the likelihood as normal
 (Eq. 11) the likelihood of an observation is evaluated using $S(t)$. Conceptually,
⁷⁸⁸ this approach calculates the likelihood of observing a taxon that existed for at
 least that long. For left censored data, instead the likelihood is calculated using
⁷⁹⁰ $1 - S(t)$ which corresponds to the likelihood of observing a taxon that existed
 no longer than t .

⁷⁹² The full likelihood statements incorporating fully observed, right censored, and
 left censored observations are then

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

⁷⁹⁴ where C is the set of all fully observed taxa, R the set of all right censored taxa,
 and L the set of all left-censored taxa.

⁷⁹⁶ D Widely applicable information criterion

WAIC can be considered a fully Bayesian alternative to the Akaike information
⁷⁹⁸ criterion, where WAIC acts as an approximation of leave-one-out
 cross-validation which acts as a measure of out-of-sample predictive accuracy
⁸⁰⁰ (Gelman et al., 2013). WAIC is calculated starting with the log pointwise
 posterior predictive density calculated as

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

802 where n is sample size, S is the number posterior simulation draws, and Θ
 804 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
 806 adjust for overfitting. The effective number of parameters is calculated,
 following the recommendations of Gelman et al. (2013), as

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

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

810 Given both equations 17 and 18, WAIC is then calculated

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

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