

Title: How macroecology affects macroevolution: the interplay
2 between extinction intensity and trait-dependent extinction in
brachiopods.

4 **Running title:** Trait-dependent extinction in brachiopods

5 **Author:** Peter D Smits, psmits@uchicago.edu, Committee on Evolutionary
6 Biology, University of Chicago, IL, USA.

Keywords: species selection, paleobiology, Bayesian

8 **Word count:** \approx 7000

Table count: 1

10 **Figure count:** 7

Data archival location: Zenodo DOI 10.5281/zenodo.46928.

Abstract

As extinction intensity increases, how do the effects of traits on
 14 taxonomic survival change? Does the extinction rate associated with
 certain traits increase while that of others decreases? Using a hierarchical
 16 Bayesian approach, I develop a model of how the effects of biological
 traits on extinction risk can vary with respect to extinction intensity,
 18 origination cohort (i.e. time of origination), and in relation to each other.
 The emergent traits I analyze in relation to their patterns of
 20 Paleozoic brachiopod genus durations are geographic range, affinity for
 epicontinental seas versus open ocean environments, and body size.
 22 Additionally, I estimate the effects of environmental generalization versus
 specialization on taxonomic survival by allowing environmental preference
 24 to have a nonlinear effect on duration. My analytical framework eschews
 the traditional distinction between background and mass extinction, and
 26 instead considers extinction intensity as a continuum. I find that the
 cohort-specific effects of geographic range and environmental preference
 28 are negatively correlated with baseline extinction intensity. Additionally, I
 find support for greater survival of environmental generalists versus
 specialists in all origination cohorts. These results support the conclusion
 30 that for Paleozoic brachiopods, as extinction intensity increases overall
 extinction selectivity increases.
 32

1 Introduction

34 Extinction is one half of the diversification process (Raup, 1994, Stanley, 1975,
 1979), second only to speciation or origination in shaping changes to diversity; it
 36 can also be the ultimate manifestation of selection as a taxon with a beneficial
 trait should persist for longer on average than a taxon without that beneficial
 38 trait (Jablonski, 2008, Rabosky and McCune, 2010, Raup, 1994, Stanley, 1975).

Species duration is a measure of species fitness CITATION, and trait-associated
40 differences in fitness is the hallmark of (species) selection CITATION.

Jablonski (1986) observed that for bivalves at the end Cretaceous mass
42 extinction event, previous trait-associated differences in survival no longer
mattered except for the case of geographic range. Based on this evidence,
44 Jablonski (1986) proposed the idea of "macroevolutionary modes" and that
mass extinction and background extinction are fundamentally different
46 processes. However, based on estimates of extinction rates over time, there is no
evidence of there being two or more "types" of extinction (Wang, 2003). Instead,
48 extinction rates for marine invertebrates form a unimodal distribution where
estimates of extinction rate/intensity show continuous variation.

50 The apparent disconnect between the theory of macroevolutionary modes and
the observation of continuous variation in extinction rates implies the possibility
52 of a relationship between the strength of selection (extinction **intensity**) and
the association between of traits and differences in fitness (extinction
54 **selectivity**) CITATION PAYNE. As extinction intensity increases, what
happens to extinction selectivity? How do trait-associated differences in fitness
56 change as average extinction rate changes over time?

Here I model brachiopod taxon durations as a function of multiple functional
58 taxon traits because trait-dependent differences in extinction risk should be
associated with differences in taxon duration CITATION. Brachiopods are an
60 ideal group for this study as they have an exceptionally complete fossil record
(Foote, 2000b, Foote and Raup, 1996). I focus on the brachiopod record from
62 the post-Cambrian Paleozoic, from the start of the Ordovician (approximately
485 My) through the end Permian (approximately 252 My) as this represents
64 the time of greatest global brachiopod diversity (Alroy, 2010) which results in a
large sample size.

66 The analysis of taxon durations, or time from origination to extinction, falls
under the purview of survival analysis, a field of applied statistics commonly
68 used in health care and engineering (Klein and Moeschberger, 2003) but has a
long history in paleontology (Crampton et al., 2016, Simpson, 1944, 1953, Smits,
70 2015, Van Valen, 1973, 1979). I adopt a hierarchical Bayesian modeling
approach (Gelman et al., 2013, Gelman and Hill, 2007) in order to unify the
72 previously distinct dynamic and cohort paleontological survival approaches
(Baumiller, 1993, Crampton et al., 2016, Ezard et al., 2012, Foote, 1988, Raup,
74 1975, 1978, Simpson, 2006, Van Valen, 1973, 1979).

To analyze a potential association between extinction intensity and extinction
76 selectivity, average extinction rate and the trait-based differences in extinction
rate need to be estimated. In particular, how these values vary over time needs
78 to be estimated; this is a data intensive and parameter rich requirement. In
general, there are two approaches to this problem, which I will call phylogenetic
80 comparative (PCM) or paleobiological. PCM approaches uses branch length
information from a time-scaled phylogeny in order to estimate the speciation
82 and extinction rates from a birth-death model that best fits the phylogeny's
topology (Fitzjohn, 2010, Goldberg et al., 2011, 2005, Maddison et al., 2007,
84 Rabosky et al., 2013, Stadler, 2013, Stadler and Bokma, 2013, ?) MORE
CITATION. In contrast, the paleobiological approach to estimating extinction
86 rates, as well as trait-dependent differences in those rates, is a type of
time-series analysis where the occurrence and sampling histories of fossil taxa
88 are used to estimated the underlying rates of a birth-death process (with
incomplete observation) that fits the observed data.

90 While the capabilities of the PCM approach have increased dramatically from
their first inception, being capable of estimating time-varying rates as well as
92 possible trait-dependent or clade-dependent differences in those rates

CITATION, there are known analytical limitations that hamper analysis of
94 extinction rates and more specifically trait-dependent differences in extinction
in non-ultrametric trees (i.e. those with fossil tips), estimates which are critical
96 to answering the question of how extinction intensity relates to extinction
selectivity. Additionally, many fossil invertebrate groups do not have an inferred
98 phylogeny that covers a significant portion of history of a clade CITATION; this
is the case for post-Cambrian Paleozoic brachiopods. What phylogenies do exist
100 for this group exist at relatively small taxonomic and temporal scales
CITATION, and are unsuited for questions which require large datasets in order
102 to fit parameter rich models. This reality means that a paleobiological approach
to this question is more appropriate than a PCM one.

104 Previous work that has dealt with extinction intensity versus selectivity
approached this problem by independently estimating the extinction intensity
106 and selectivity at different time points or for different origination cohorts and
then comparing those estimates CITATION. I find this approach problematic
108 for a few reasons. First, by treating each time point or cohort as independent
the estimates from each model are not relative to each other and they do not
110 represent the greatest compromise across all available data CITATIONS.
Second, by treating each time point or cohort as independent any and all
112 post-hoc analyses are at risk of false positive results because of multiple
comparisons CITATIONS. Third, post-hoc analysis of correlations between
114 maximum likelihood estimates is much less satisfying or coherent than including
those correlations as estimates of the initial model, something that is not
116 possible when each time point or cohort is considered independent CITATION.

The hierarchical Bayesian approach used here involves developing a single model
118 to estimate all of time points or cohorts simultaneously while also estimating
the group averages for all parameter CITATION. By averaging across groups,

¹²⁰ the analytical advantages of specific prior choices can induce the partial pooling
¹²¹ and regularization of parameter estimates which smooths over groups with small
¹²² sample sizes or weakly estimated effects. Additionally, individual estimates are
¹²³ much more comparable, not only between time points or cohorts, but also
¹²⁴ relative to the groups average CITATION. This approach is conceptually and
¹²⁵ analytically similar to mixed-effects modeling but the Bayesian framework
¹²⁶ allows the use of strongly regularizing priors in order to better constrain
parameter estimates.

¹²⁸ 1.1 Factors affecting brachiopod survival

Conceptually, taxon survival can be considered an aspect of “taxon fitness”
¹³⁰ (Cooper, 1984, Palmer and Feldman, 2012). Traits associated with taxon
survival are thus examples of species (or higher-level) selection, as differences in
¹³² survival are analogous to differences in fitness. The traits analyzed here are all
examples of emergent and aggregate traits (Jablonski, 2008, Rabosky and
¹³⁴ McCune, 2010); specifically I analyze genus-level traits. Emergent traits are
those which are not measurable from individuals of the lower level (e.g.
¹³⁶ organisms within species) such as geographic range, or even fossil sampling rate.
Aggregate traits, like body size or environmental preference, are the average of a
¹³⁸ shared trait across all members of a lower level.

Geographic range is widely considered the most important biological trait for
¹⁴⁰ estimating differences in extinction risk at nearly all times, with large
geographic range associated with low extinction risk (Finnegan et al., 2012,
¹⁴² Harnik et al., 2012, Jablonski, 1986, 1987, 2008, Jablonski and Roy, 2003, Payne
and Finnegan, 2007). This stands to reason even if extinction is completely at
¹⁴⁴ random; a taxon with an unrestricted range is less likely to go extinct at

random than a taxon with a restricted range.

¹⁴⁶ Epicontinental seas are a shallow-marine environment where the ocean has spread over the continental interior or craton with a depth typically less than
¹⁴⁸ 100m. In contrast, open-ocean coastline environments have much greater variance in depth, do not cover the continental craton, and can persist during
¹⁵⁰ periods of low sea level (Miller and Foote, 2009). Because of this, a simple hypothesis that taxa which favor epicontinental seas would be at great risk
¹⁵² during periods of low sea levels, such as during glacial periods, when epicontinental seas are drained. During the Paleozoic (approximately 541–252
¹⁵⁴ My), epicontinental seas were widely spread globally but declined over the Mesozoic (approximately 252–66 My) and have nearly disappeared during the
¹⁵⁶ Cenozoic (approximately 66–0 My) as open-ocean coastlines became the dominant shallow-marine setting (Johnson, 1974, Miller and Foote, 2009, Peters,
¹⁵⁸ 2008, Sheehan, 2001). Taxa in epicontinental environments could also have a greater extinction susceptibility than taxa in open-ocean environments due to
¹⁶⁰ anoxic events due to enhanced water stratification or poor water circulation (Peters, 2007).

¹⁶² Miller and Foote (2009) demonstrated that during several mass extinctions taxa associated with open-ocean environments tend to have a greater extinction risk
¹⁶⁴ than those taxa associated with epicontinental seas. During periods of background extinction, however, they found no consistent difference between
¹⁶⁶ taxa favoring either environment. Miller and Foote (2009) hypothesize that open-ocean taxa may have a greater extinction rate because these environments
¹⁶⁸ would be more strongly affected by waterborne hazards such as fallout from impacts or volcanic events which would propagate more quickly than in
¹⁷⁰ epicontinental environments with sluggish circulation. These two environment types represent the primary identifiable environmental dichotomy observed in

¹⁷² ancient marine systems (Miller and Foote, 2009, Sheehan, 2001). Given these
¹⁷⁴ findings, I would hypothesize that as extinction risk increases, the extinction
risk associated with open-ocean environments should generally increase.

¹⁷⁶ Because environmental preference is defined here as the continuum between
occurring exclusively in open-ocean environments versus epicontinental
¹⁷⁸ environments, intermediate values are considered “generalists” in the sense that
they favor neither end member. A long-standing hypothesis is that generalists or
¹⁸⁰ unspecialized taxa will have greater survival than specialists (Baumiller, 1993,
Liow, 2004, 2007, Nürnberg and Aberhan, 2013, 2015, Simpson, 1944, Smits,
¹⁸² 2015). Because of this, the effect of environmental preference was modeled as a
quadratic function where a concave down relationship between preference and
expected duration indicates that generalists are favored over specialists
¹⁸⁴ end-members.

¹⁸⁶ Body size, measured as shell length, is also considered as a trait that may
potentially influence extinction risk (Harnik, 2011, Payne et al., 2014). Body
size is a proxy for metabolic activity and other correlated life history traits
¹⁸⁸ (Payne et al., 2014). Harnik et al. (2014) analyzed the effect of body size
selectivity in Devonian brachiopods in both a phylogenetic and
¹⁹⁰ non-phylogenetic context; finding that body size was not found to be associated
with differences in taxonomic duration. It has also been found that, at least in
¹⁹² the case of some bivalve subclades, body size can be as important a factor as
geographic range size in determining extinction risk (Harnik, 2011). Given these
¹⁹⁴ results, I expect that if body size has any effect on brachiopod taxonomic
survival it is very small.

¹⁹⁶ It is well known that, given the incompleteness of the fossil record, the observed
duration of a taxon is an underestimate of that taxon’s true duration (Alroy,
¹⁹⁸ 2014, Foote and Raup, 1996, Liow and Nichols, 2010, Solow and Smith, 1997,

Wagner and Marcot, 2013, Wang and Marshall, 2004). Because of this, the
200 concern is that a taxon's observed duration may reflect its relative chance of
being sampled and not any of the effects of the covariates of interest. In this
202 case, for sampling to be a confounding factor there must be consistent
relationship between the quality of sampling of a taxon and its apparent
204 duration (e.g. greater sampling, longer duration). If there is no relationship
between sampling and duration then interpretation can be made clearly; while
206 observed durations are obviously truncated true durations, a lack of a
relationship would indicate that the amount and form of this truncation is not a
208 major determinant of the taxon's apparent duration. By including sampling as a
covariate in the model, this effect can be quantified and can be taken into
210 account when interpreting the estimates of the effects of the other covariates.

2 Materials and Methods

212 2.1 Fossil occurrence information

The brachiopod dataset analyzed here was sourced from the Paleobiology
214 Database (<http://www.paleodb.org>) which was then filtered based on
taxonomic (Rhyconelliformea: Rhynchonellata, Chileata, Obolellida,
216 Kutorginida, Strophomenida, Spiriferida), temporal (post-Cambrian Paleozoic),
stratigraphic, and other occurrence information used in this analysis. Analyzed
218 occurrences were restricted to those with paleolatitude and paleolongitude
coordinates, assignment to either epicontinental or open-ocean environment,
220 and belonging to a genus present in the body size dataset (Payne et al., 2014).
Epicontinental versus open-ocean assignments for each fossil occurrence are
222 partially based on those from Miller and Foote (2009), with additional

occurrences assigned similarly (Miller and Foote, personal communication).

²²⁴ These filtering criteria are very similar to those from Foote and Miller (2013)
with an additional constraint of being present in the body size data set from
²²⁶ Payne et al. (2014). In total, there 1130 were genera included in the dataset.

Fossil occurrences were analyzed at the genus level which is common for
²²⁸ paleobiological, macroevolutionary and macroecological studies of marine
invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2012, Kiessling
²³⁰ and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,
Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).

²³² While species diversity dynamics are frequently of much greater interest than
those of higher taxa (though see Foote 2014, Hoehn et al. 2015), the nature of
²³⁴ the fossil record makes accurate, precise, and consistent taxonomic assignments
at the species level difficult for all occurrences. As such, the choice to analyze
²³⁶ genera as opposed to species was in order to assure a minimum level of
confidence and accuracy in the data analyzed here.

²³⁸ Genus duration was calculated as the number of geologic stages from first
appearance to last appearance, inclusive. Durations were based on geologic
²⁴⁰ stages as opposed to millions of years because of the inherently discrete nature
of the fossil record; dates are not assigned to individual fossils themselves but
²⁴² instead fossils are assigned to a geological interval which represents some
temporal range. In this analysis, stages are effectively irreducible temporal
²⁴⁴ intervals in which taxa may occur. Genera with a last occurrence in or after
Changhsingian stage (e.g. the final stage of the study interval) were right
²⁴⁶ censored at the Changhsingian; genera with a duration of only one stage were
left censored (Klein and Moeschberger, 2003). How the likelihood of censored
²⁴⁸ observations is calculated is detailed in section 2.2.

The covariates of duration included in this analysis are geographic range size

²⁵⁰ (r), environmental preference (v, v^2), the statistical interactions between
geographic range size and environmental preference ($r \times v, r \times v^2$), body size
²⁵² (m), and sampling (s). The interaction terms are important to include because
of the known multidirectional relationship between geographic range,
²⁵⁴ environmental preference, and duration CITATION.

²⁵⁶ Geographic range was calculated as relative occupancy corrected for incomplete
sampling. First, the paleolatitude-paleolongitude coordinates for all occurrences
were projected onto an equal-area cylindrical map projection. Each occurrence
²⁵⁸ 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. The map
²⁶⁰ projection and regular lattice were made using shape files from
<http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,
²⁶² 2015). For each stage, the total number of occupied grid cells was calculated.
Then, for each temporal bin, the relative occurrence probability of the observed
²⁶⁴ taxa was calculated using the JADE method developed by Chao et al. (2015).
This method accounts for the fact that taxa with an occupancy of 0 cannot be
²⁶⁶ observed which means that occupancy follows a truncated Binomial distribution.
This correction is critical when comparing occupancies from different times with
²⁶⁸ different geographic sampling. Finally, for each genus, the mean relative
occurrence probability was calculated as the average of that genus' occurrence
²⁷⁰ probabilities for all stages it was sampled to yield relative occupancy, my proxy
for geographic range.

²⁷² Environmental preference was defined as probability of observing the ratio of
epicontinental occurrences to total occurrences ($\theta_i = e_i/E_i$) or greater given the
²⁷⁴ background occurrence probability θ'_i as estimated from all other taxa occurring
at the same time (e'_i/E'_i). This measure of environmental preference is

²⁷⁶ expressed.

$$\begin{aligned} p(\theta'_i | e'_i, E'_i) &\propto \text{Beta}(e'_i, E'_i - e'_i) \text{Beta}(1, 1) \\ &= \text{Beta}(e'_i + 1, E'_i - e'_i + 1), \end{aligned} \tag{1}$$

²⁷⁸ where v is the percent of the distribution defined in equation 1 less than or equal to θ_i . The Beta distribution is used here because it is a continuous distribution bounded at 0 and 1, which is idea for modeling percentages.

²⁸⁰ Body size, measured as shell length, was sourced directly from Payne et al. (2014). These measurements were made from brachiopod taxa figured in the ²⁸² *Treatise on Invertebrate Paleontology* (Williams et al., 2007).

²⁸⁴ The sampling probability for individual taxa was calculated using the standard gap statistic (Foote, 2000a, Foote and Raup, 1996). The gap statistic is calculated as the number of stages in which the taxon was sampled minus two ²⁸⁶ divided by the duration of the taxon minus two. Subtracting two from both the numerator and denominator is because the first and last appearance stages are ²⁸⁸ by definition sampled. Because taxa that were right censored only include a first appearance, one was subtracted from the numerator and denominator instead of ²⁹⁰ two.

²⁹² The minimum duration for which a gap statistic can be calculated is three stages, so I chose the impute the gap statistic for all observations with a ²⁹⁴ duration less than 3. Imputation is the “filling in” of missing observations based on the observed values (Gelman and Hill, 2007, Rubin, 1996). This is fairly ²⁹⁶ straight forward in a Bayesian framework because both covariates and parameters are considered random variables, meaning that the missing values of ²⁹⁸ sampling can be modeled as coming from some probability distribution. The model for imputing sampling probability is described in section 2.3.

Prior to analysis, geographic range was logit transformed and body size was

³⁰⁰ natural-log transformed; both of these transformations make these variables
³⁰¹ defined for the entire real line. Sampling probability was transformed as
³⁰² $(s(n - 1) + 0.5)/n$ where n is the sample size as recommended by Smithson and
³⁰³ Verkuilen (2006); this serves to slightly shrink the range of the data so that
³⁰⁴ there are no values of 0 or 1. All covariates except for sampling were
³⁰⁵ standardized by subtracting the mean from all values and dividing by twice its
³⁰⁶ standard deviation, which follows Gelman and Hill (2007). This standardization
³⁰⁷ means that the associated regression coefficients are comparable as the expected
³⁰⁸ change per 1-unit change in the rescaled covariates. Finally, D is defined as the
total number of covariates, excluding sampling, plus one for the intercept term.

³¹⁰ 2.2 Details of model

Hierarchical modelling is a statistical approach which explicitly takes into
³¹¹ account the structure of the observed data in order to model both the within
and between group variance (Gelman et al., 2013, Gelman and Hill, 2007). The
³¹² units of study (e.g. genera) each belong to a single group (e.g. origination
cohort). Each group is considered a draw from a shared probability distribution
³¹³ (e.g. prior) of all cohorts, observed and unobserved. The group-level parameters,
or the hyperparameters of this shared prior, are themselves given (hyper)prior
³¹⁴ distributions and are also estimated like the other parameters of interest (e.g.
covariate effects) (Gelman et al., 2013). The subsequent estimates are partially
³¹⁵ pooled together, where parameters from groups with large samples or effects
remain large while those of groups with small samples or effects are pulled
³¹⁶ towards the overall group mean. All covariate effects (regression coefficients), as
well as the intercept term (baseline extinction risk), were allowed to vary by
³¹⁷ group (origination cohort). The covariance between covariate effects was also
modeled.

326 Genus durations were assumed to follow a Weibull distribution which allows for
 age-dependent extinction (Klein and Moeschberger, 2003): $y \sim \text{Weibull}(\alpha, \sigma)$.
 328 The Weibull distribution has two parameters: scale σ , and shape α . When
 $\alpha = 1$, σ is equal to the expected duration of any taxon. α is a measure of the
 330 effect of age on extinction risk where values greater than 1 indicate that
 extinction risk increases with age, and values less than 1 indicate that
 332 extinction risk decreases with age. Note that the Weibull distribution is
 equivalent to the exponential distribution when $\alpha = 1$.
 334 In the case of the right- and left-censored observations mentioned above, the
 probability of those observations has a different calculation (Klein and
 336 Moeschberger, 2003). For right-censored observations, the likelihood is
 calculated $p(y|\theta) = 1 - F(y) = S(y)$ where $F(y)$ is the cumulative distribution
 338 function. In contrast, the likelihood of a left-censored observation is calculated
 $p(y|\theta) = F(y)$.
 340 The scale parameter σ was modeled as a regression following Kleinbaum and
 Klein (2005) with both varying intercept and varying slopes and the effect of
 342 sampling; this is expressed

$$\sigma_i = \exp\left(\frac{-\mathbf{X}_i B_{j[i]} + \delta s_i}{\alpha}\right) \quad (2)$$

where i indexes across all observations where $i = 1, \dots, n$ where n is the total
 344 number of observations, $j[i]$ is the cohort membership of the i th observation
 where $j = 1, \dots, J$ where J is the total number of cohorts, X is a $N \times D$ matrix
 346 of covariates along with a column of 1's for the intercept term, B is a $J \times D$
 matrix of cohort-specific regression coefficients, and δ is the regression
 348 coefficient for the effect of sampling s . δ does not vary by cohort.

Each of the rows of matrix B are modeled as realizations from a multivariate

³⁵⁰ normal distribution with length D location vector μ and $J \times J$ covariance matrix Σ : $B_j \sim \text{MVN}(\mu, \Sigma)$. The covariance matrix was then decomposed into a length J vector of scales τ and a $J \times J$ correlation matrix Ω , defined $\Sigma = \text{diag}(\tau)\Omega\text{diag}(\tau)$ where “diag” indicates a diagonal matrix.

³⁵⁴ The elements of μ were given independent normally distributed priors. The effects of geographic range size and the breadth of environmental preference ³⁵⁶ were given informative priors reflecting the previous findings while the others were given weakly informative favoring no effect. The correlation matrix Ω was ³⁵⁸ given an LKJ distributed prior (Lewandowski et al., 2009) that slightly favors an identity matrix as recommended by ?. These priors are defined

$$\begin{aligned}\mu^0 &\sim \mathcal{N}(0, 5) \\ \mu^r &\sim \mathcal{N}(-1, 1) \\ \mu^v &\sim \mathcal{N}(0, 1) \\ \mu^{v^2} &\sim \mathcal{N}(1, 1) \\ \mu^{r \times v} &\sim \mathcal{N}(0, 1) \\ \mu^{r \times v^2} &\sim \mathcal{N}(1, 1) \\ \mu^m &\sim \mathcal{N}(0, 1) \\ \tau &\sim \text{C}^+(1) \\ \Omega &\sim \text{LKJ}(2).\end{aligned}\tag{3}$$

³⁶⁰ The log of the shape parameter α was given a weakly informative prior $\log(\alpha) \sim \mathcal{N}(0, 1)$ centered at $\alpha = 1$, which corresponds to the Law of Constant ³⁶² Extinction (Van Valen, 1973).

2.3 Imputation of sampling probability

364 The vector sampling s has two parts: the observed part s^o , and the unobserved
part s^u . Of the 1130 total observations, 539 have a duration of 3 or more and
366 have an observed gap statistic. The gap statistic for the remaining 591
observations was imputed. As stated above, the unobserved part is the imputed,
368 or filled in, based on the observed part s^o . Because sampling varies between 0
and 1, I chose to model it as a Beta regression with matrix W being a
370 $N \times (D - 3)$ matrix of covariates (i.e. geographic range size, environmental
preference, body size; no interactions) as predictors of sampling; this assumes
372 that the sampling value for all taxa come from the same distribution.
Importantly, I make no assumptions of causal structure.

374 Predicting sampling probability using the other covariate that are then included
in the model of duration is acceptable and appropriate in the case of imputation
376 where the sample goal is accurate prediction (Gelman and Hill, 2007, Rubin,
1996). Not including these covariates can lead to biased estimates of the
378 imputed variable; if the covariates themselves are related, not including them
will bias this correlation towards zero which then leads to improper imputation
380 and inference (Rubin, 1996).

The Beta regression is defined

$$s^o \sim \text{Beta}(\phi = \text{logit}^{-1}(X^o\gamma), \lambda), \quad (4)$$

382 where γ is a length D vector of regression coefficients, and X defined as above.

The Beta distribution used in the regression is reparameterized in terms of a
384 mean parameter

$$\phi = \frac{\alpha}{\alpha + \beta} \quad (5)$$

and total count parameter

$$\lambda = \alpha + \beta \quad (6)$$

³⁸⁶ where α and β are the characteristic parameters of the Beta distribution
(Gelman et al., 2013).

³⁸⁸ The next step is to then estimate $s^u|s^o, X^o, X^u, \gamma$, the posterior distribution of
which are folded back into s and used as a covariate of duration (Eq. 2). All the
³⁹⁰ elements of γ , and both δ (Eq. 2) and λ (Eq. 4) were given weakly informative
priors where

$$\begin{aligned} \gamma &\sim \mathcal{N}(0, 1) \\ \delta &\sim \mathcal{N}(0, 1) \\ \lambda &\sim \text{Pareto}(0.1, 1.5). \end{aligned} \quad (7)$$

³⁹² 2.4 Posterior inference and posterior predictive checks

The joint posterior was approximated using a Markov-chain Monte Carlo
³⁹⁴ routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic
³⁹⁶ programming language Stan (Stan Development Team, 2014). The posterior
distribution was approximated from four parallel chains run for 10,000 steps
³⁹⁸ each, split half warm-up and half sampling and thinned to every 10th sample for
a total of 4000 posterior samples. Chain convergence was assessed via the scale
⁴⁰⁰ reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$) indicate approximate
convergence. Convergence means that the chains are approximately stationary
⁴⁰² and the samples are well mixed (Gelman et al., 2013).

Model adequacy was evaluated using a couple of posterior predictive checks.
⁴⁰⁴ Posterior predictive checks are a means for understanding model fit or adequacy
where the basic idea is that replicated data sets simulated from the fitted model

406 should be similar to the original data and systematic differences between the
407 simulations and observations indicate weaknesses of the model fit (Gelman
408 et al., 2013). For both approaches used here, each posterior predictive datasets
409 were generated from a unique draw from the posterior distribution of each
410 parameter. The two posterior predictive checks used in this analysis are a
411 comparison of a non-parametric estimate of the survival function $S(t)$ from the
412 empirical dataset to the non-parametric estimates of $S(t)$ from the 100 posterior
413 predictive datasets, and comparison of the observed genus durations to the
414 average posterior predictive estimate of $\log(\sigma)$ (Eq. 2). The former is to see if
415 simulated data has a similar survival pattern to the observed, while the latter is
416 to see if the model systematically over- or under- estimates taxon survival.

3 Results

418 Comparison of the posterior predictive estimates of $S(t)$ to the empirical
419 estimate reveal few obvious biases except for the case of values from the far
420 right tail of observed durations (Fig. 1). This result is reinforced by the
421 additional posterior predictive comparison where most estimates are not
422 systematically biased except for a consistent under-estimate of $\log(\sigma)$ for older
423 taxa (Fig. 2). The results of both posterior predictive checks indicate that, for
424 the majority of observations, model fit is generally not biased.

The cohort-level estimate of the effect of geographic range size indicates that as
426 a taxon's geographic range increases, that taxon's duration is expected to
427 increase (Table 1). Given the estimates of μ^r and τ^r , there is a less than 3.7%
428 ($\pm 0.04\%$ SD) probability that this relationships would be reversed
 $(\text{Pr}N(\mu^r, \tau^r) > 0)$). The between-cohort variance τ^r is the lowest of all the
430 regression coefficients (Table 1).

Body size is estimated to have no effect on taxon duration, with the estimate
432 being nearly 0 (Table 1). The variance between the cohort-specific estimates of
the effect of body size τ^m is estimated to be greater than the variance of
434 between-cohort estimates of the effect of geographic range size τ^r .

The group-level estimate of the effect of environmental preference is estimated
436 from both μ^v and μ^{v^2} .

The estimate of μ^v indicates that epicontinental favoring taxa are expected to
438 have a greater duration than open-ocean favoring taxa (Table 1). Additionally,
given the estimate of between-cohort variance τ^v , there is approximately 18%
440 ($\pm 7\%$ SD) probability that, for any given cohort, taxa favoring open-ocean
environments would have a greater expected duration than taxa favoring
442 epicontinental environments ($\Pr(\mathcal{N}(\mu^v, \tau^v) > 0)$).

The estimate of μ^{v^2} indicates that the overall relationship between
444 environmental preference and $\log(\sigma)$ is concave down (Fig. ??), with only a 2.7%
($\pm 3\%$ SD) probability that any given cohort is convex up ($\Pr(\mathcal{N}(\mu^{v^2}, \tau^{v^2}) < 0)$).

446 The cohort-specific estimates of all the regression coefficients demonstrate a lot
of between cohort variance, with no obvious trends. As indicated in Table 1 and
448 detectable visually (Fig. 6), the between-cohort estimates for β^0 , β^r , and β^m all
have much lower variance than the between-cohort estimates of both β^v and
450 β^{v^2} .

While most cohort-specific estimates are very similar to the overall cohort-level
452 estimate, there are a few notable excursions away from the overall mean (Fig.
6). There are simultaneous excursions in both β^0 and β^v for cohorts originating
454 in the Givetian (387-382 My) and Frasnian (382-372 My) stages; both of which
directly precede the late Devonian mass extinction event at the
456 Frasnian/Famennian boundary. These cohorts are marked by both a high

extinction intensity and an increase in expected duration for taxa favoring
458 epicontinental environments over open-ocean ones; this is consistent with the
results of Miller and Foote (2009).

460 Cohorts originating from the Silurian through the Early Devonian have a
slightly lower extinction intensity than the overall mean; these cohorts are those
462 originating in the Llandovery (443-443 My) through the Emsian (407-393 My).
This is also a time period is also when there is the lowest overall probability
464 that epicontinental favoring taxa are expected to have greater duration than
open-ocean favoring taxa. Both the Silurian and Devonian periods are notable
466 for having been periods with a mostly “hothouse” climate, with no polar icecaps
and a high sea-level (Edwards and Fanning, 1985, Joachimski et al., 2009,
468 Munnecke et al., 2010).

The cohort-specific relationships between environmental preference and $\log(\sigma)$
470 were calculated from the estimates of β^0 , β^v , and β^{v^2} (Fig. ??) and reflect how
these three parameters act in concert and not just individually (Fig. 6). Beyond
472 results already discussed above in the context of the parameters individually, it
is notable that the cohort originating in the Kungurian (279-272 My) is least
474 like the overall expected relationship and has the most sharply curved
appearance due to a high estimate β^{v^2} (Fig. 6). This cohort has the biggest
476 difference in extinction risk between environmental generalists and specialists.

The cohorts originating during the Emsian (407-393 My) and Frasnian (382 -
478 372 My) are tied for second in sharpness of curvature. The least sharply curved
cohorts include those originating during Tremadocian (484-477 My), Hirnantian
480 (445-443 My), Llandovery (443-433 My), and Ludlow (427-423 My). Except for
the Tremadocian cohort, most of these cohorts originate during the Silurian
482 through the Early Devonian range identified earlier as having lower expected
extinction intensity than what is expected from the group-level estimate.

- 484 The correlations of the cohort-specific estimates of the regression coefficients are
 estimated as the off-diagonal elements of the correlation matrix Ω . Only two of
 486 the elements of Ω are distinguishable from 0: the correlation of β^0 (extinction
 intensity) with both β^r and β^v (Fig. 10).
- 488 There is an approximate 90% probability that the cohort-specific estimates of
 baseline extinction intensity β^0 and the effect of geographic range β^r are
 490 negatively correlated; this means that for cohorts experiencing a lower
 extinction intensity (β^0 decreases), the magnitude of the effect of geographic
 492 range is expected to decrease as well, and *vice versa*; this is in contrast to the
 observation made by Jablonski (1986) with regards to late Cretaceous bivalves.
- 494 Similarly, there is an approximate 97.4% probability that the cohort-specific
 estimates of β^0 and β^v are negatively correlated; this means that as extinction
 496 intensity increases it is expected that epicontinental taxa become more favored
 over open-ocean environments (i.e. as β^0 increases, β^v decreases).
- 498 There is only an approximate 30% probability that β^r and β^v are positively
 correlated. This lack of cross-correlation may be due in part to the much higher
 500 between-cohort variance of the effect of environmental preference τ^v than the
 very small between-cohort variance in the effect of geographic range τ^r (Table
 502 1); the effect of geographic range might simply not vary enough relative to the
 much noisier environmental preference.
- 504 Comparison of observed values of sampling, as measured by the gap statistic, to
 random draws from the posterior estimates of the imputed sampling values
 506 indicate that they are very similar (Fig. 11. This result is very encouraging as
 this is the ultimate goal of multiple imputation: to fill in missing data with
 508 values similar to the observed while taking into account the randomness of that
 variable (Gelman and Hill, 2007, Rubin, 1996). The estimates of δ are based on

510 the set of observed values and the entire posterior of imputed values.

Sampling was found to have a negative effect (positive δ) on duration: greater
512 sampling, shorter duration (Table 1). While potentially counter intuitive, this
result is most likely due to some long lived taxa only be sampled in the stages of
514 the first and last appearance. Also, longer lived taxa have more opportunities to
not be sampled than shorter lived taxa. These two factors will lead to this result.

516 While the effect of sampling appears large compared to the other regression
coefficients, this is only because sampling was not standardized like the other
518 covariates. To make the coefficients comparable, δ is multiplied by twice the
posterior mean of the standard deviation of sampling probability; the
520 transformed value of δ is then 0.642 (± 0.1 SD). This effect is relatively small
compared to the other covariate effects (Table 1). There is then a 98.6%
522 probability that the effect of geographic range μ^r has a greater magnitude than
 δ . Similarly, μ^v has a 71.8% probability of having a greater magnitude of effect
524 than δ . Finally, μ^{v^2} has a 100% probability of having a greater magnitude of
effect than δ .

526 The Weibull shape parameter α was found to be approximately 1.36 (± 0.05 SD)
with a 100% probability of being greater than 1. This result is not consistent
528 with the Law of Constant Extinction (Van Valen, 1973) and is instead
consistent with accelerating extinction risk with taxon age. This may indicate
530 that older taxa are out-competed by younger taxa, a result consistent with some
empirical results (Quental and Marshall, 2013, Smits, 2015, Wagner and
532 Estabrook, 2014) and (ironically) with a recently proposed Red Queen-like
model of evolution (Rosindell et al., 2015). This results, however, is not
534 consistent with other empirical results (Crampton et al., 2016, Finnegan et al.,
2008) and could potentially be caused by the minimum resolution of the fossil
536 record (Sepkoski, 1975). It is thus unclear if a strong biological inference can be

made from this result, which means that further work is necessary on the effect
538 of taxon age on extinction risk.

4 Discussion

540 The generating observation behind this study was that for bivalves at the end
Cretaceous mass extinction event, the only biological trait that was found the
542 affect extinction risk was geographic range while traits that had previously been
beneficial had no effect (Jablonski, 1986). This observation raises two linked
544 questions: how does the effect of geographic range change with changing
extinction intensity, and how does the effect of other biological traits change
546 with changing extinction intensity?

I find that as intensity increases (β^0 decreases), the magnitude of the effect of
548 geographic range increases. I also find that as intensity increases, the effect of
favoring epicontinental environments of open-ocean environments is expected to
550 be increase; this is consistent with the results of Miller and Foote (2009). There
is no evidence for a correlation between the effect of geographic range and
552 environmental preference. Additionally, the between-cohort variance in effect of
geographic range is much less than the between-cohort variance of the effect of
554 environmental preference which may underlie the lack of correlation between
these two effects.

556 Additionally, the lower between-cohort variance in the effect of geographic range
versus that higher between-cohort variance implies that for cohorts with a
558 greater than average extinction intensity, the difference in the effect geographic
range and the group-level effect of geographic range is expected to be smaller
560 than the difference between the effect of environmental preference and the
group-level effect of environmental preference.

I find consistent support for the “survival of the unspecialized,” with respect to epicontinental versus open-ocean environmental preference, as a time-invariant generalization of brachiopod survival; taxa with intermediate environmental preferences are expected to have lower extinction risk than taxa specializing in either epicontinental or open-ocean environments (Fig. ??), though the curvature of the relationship varies from rather shallow to very peaked (Fig. ??). However, this relationship is not symmetric about 0, as taxa favoring epicontinental environments are expected to have a greater duration than taxa favoring open-ocean environments. This description of environment only describes one major aspect of a taxon’s environmental context, with factors such as bathymetry and temperature being further descriptors of a taxon’s adaptive zone (Harnik, 2011, Harnik et al., 2012, Heim and Peters, 2011, Nürnberg and Aberhan, 2013); inclusion of these factors in future analyses would potentially improve our understanding of the “survival of the unspecialized” hypothesis (Simpson, 1944).

Hopkins et al. (2014), in their analysis of niche conservatism and substrate lithological preference in marine invertebrates, found that brachiopods were among the least “conservative” groups; taxa were found to easily change substrate preference on short time scales. While substrate preference is not the same as environmental preference (as defined here), a question does arise: are there three classes of environmental preference instead of two? These classes would be taxa with broad tolerance (“true” generalists), inflexible specialists (“true” specialists), and flexible but with a narrow tolerance. A flexible taxon is one with a narrow habitat preference at one time, but with preference that changes over time with changing environmental availability. My analysis assumes that traits are constant over the duration of the taxon meaning that this scenario is not detectable; taxa with broad tolerances and flexible taxa with

narrow per-stage preference end up being treated the same way. Future work
590 should explore how environmental preference changes over lineage duration in
relation to environmental availability to estimate if the generalists–specialists
592 continuum is actually ternary relationship.

An alternative approach for specifically modeling survival that can take into
594 account imperfect observation than the method used here is the
Cormack-Jolly-Seber (CJS) model (Liow et al., 2008, Liow and Nichols, 2010,
596 Royle and Dorazio, 2008, Tomiya, 2013). This model is a type of hidden Markov
model with an absorbing state (i.e. extinction). In this model, survival is defined
598 as the probability of surviving from time t to time $t + 1$. Additionally, the effect
of preservation and sighting is estimated as probability of observing a taxon that
600 is present; this can extend the duration of a taxon beyond its last occurrence.
This approach is a fundamentally different from the method used in my analysis:
602 I am estimating the biasing effect of sampling probability on taxon duration to
then compare with effects of other covariates, while the CJS model estimates the
604 pre-sampling fossil record and then estimates per-time unit survival probability.

The use of genera as the unit of the study and how to exactly interpret the
606 effects of the biological traits is an important question. For example, if any of
the traits analyzed here are associated with increases in speciation rates, this
608 might increase the duration of genera through self-renewal (Raup, 1991, 1994),
which would be an example of the difference in biological pattern between
610 species and genera (Jablonski, 1987, 2007, 2008). This could lead to a trait
appearing to decrease generic level extinction risk by that trait increasing
612 species level origination rate instead of decreasing species level extinction risk.

The model used here could be improved through either increasing the number of
614 analyzed traits, expanding the hierarchical structure of the model to include
other major taxonomic groups of interest, and the inclusion of explicit

616 phylogenetic relationships between the taxa in the model as an additional
hierarchical effect. An example trait that may be of particular interest is the
618 affixing strategy or method of interaction with the substrate of the taxon, which
has been found to be related to brachiopod survival where, for cosmopolitan
620 taxa, taxa that are attached to the substrate are expected to have a greater
duration than those that are not (Alexander, 1977).

622 It is theoretically possible to expand this model to allow for comparisons both
within and between major taxonomic groups which would better constrain the
624 brachiopod estimates while also allowing for estimation of similarities and
differences in cross-taxonomic patterns. The major issue surrounding this
626 particular expansion involves finding a similarly well sampled taxonomic group
that is present during the Paleozoic. Potential groups include Crinoidea,
628 Ostracoda, and other members of the “Paleozoic fauna” (Sepkoski, 1981).

With significant updates, it would also be possible to compare the brachiopod
630 record with with Modern groups such as bivalves or gastropods (Sepkoski, 1981),
though remembering that the groups may not necessarily share all cohorts with
632 the brachiopods. This particular model expansion would act as a test of any
universal cross-taxonomic patterns in the effects of emergent traits on extinction
634 such as has been proposed for geographic range (Payne and Finnegan, 2007).

Additionally, this expanded model would also act as a test of the distinctness of
636 the Sepkoski (1981) three-fauna hypothesis in terms of trait-dependent
extinction.

638 Traits like environmental preference or geographic range (Hunt et al., 2005,
Jablonski, 1987) are most likely heritable. Without phylogenetic context, this
640 analysis assumes that differences in extinction risk between taxa are
independent of the shared evolutionary history of those taxa (Felsenstein, 1985).
642 In contrast, the origination cohorts only capture shared temporal context. For

example, if taxon duration is phylogenetically heritable, then closely related
644 taxa may have more similar durations as well as more similar biological traits.
Without taking into account phylogenetic similarity the effects of these
646 biological traits would be inflated solely due to inheritance. The inclusion of
phylogenetic context as an additional individual-level hierarchical effect,
648 independent of origination cohort, would allow for determining how much of the
observed variability is due to shared evolutionary history versus shared temporal
650 context versus actual differences associated with biological traits (Smits, 2015).

The combination and integration of the phylogenetic comparative and
652 paleontological approaches requires both sources of data, something which is not
possible for this analysis because there is no phylogenetic hypothesis for all
654 Paleozoic taxa, something that is frequently the case for marine invertebrates
with a good fossil record. When both data sources are available has been
656 possible, however, the analysis can more fully address the questions of interest
in macroevolution (Fritz et al., 2013, Harnik et al., 2014, Raia et al., 2012, 2013,
658 Simpson et al., 2011, Slater et al., 2012, Slater, 2013, 2015, Smits, 2015, Tomiya,
2013).

660 In summary, patterns of Paleozoic brachiopod survival were analyzed using a
fully Bayesian hierarchical survival modelling approach while also eschewing the
662 traditional separation between background and mass extinction. I find that
cohort extinction intensity is negatively correlated with both the cohort-specific
664 effects of geographic range and environmental preference. These results imply
that as extinction intensity increases (β^0) increases, it is expected that both
666 effects will increase in magnitude. However, the change in effect of
environmental preference is expected to be greater than the change in the effect
668 of geographic range. Additionally, I find support for greater survival in
environmental generalists over specialists in all origination cohorts analyzed;

670 this is consistent with the long standing “survival of the unspecialized”
hypothesis (Baumiller, 1993, Liow, 2004, 2007, Nürnberg and Aberhan, 2013,
672 2015, Simpson, 1944, 1953, Smits, 2015). The results of this analysis support the
conclusion that for Paleozoic brachiopods, as extinction intensity increases
674 overall extinction selectivity is expected to increase as well.

Acknowledgements

676 I would like to thank K. Angielczyk, M. Foote, P. D. Polly, R. Ree, and G.
Slater for helpful discussion during the conception of this study. I'd also like to
678 thank D. Bapst, N. Pierrehumbert and M. Villarosa Garcia for additional
comments. An additional thank you to A. Miller for the epicontinental versus
680 open-ocean assignments. Finally, thank you to the reviewers for their helpful
comments that improved this manuscript. This entire study would not
682 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.
684 Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland,
L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T. Olszewski, P.
686 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
688 M. Foote. I declare no conflicts of interest. This is Paleobiology Database
publication XXX. The code necessary for reproducing these results is available
690 as a Zenodo archive DOI 10.5281/zenodo.46928.

References

- 692 Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
to the mode of stabilization on the substrate. *Palaeogeography,*
694 *Palaeoclimatology, Palaeoecology* 21:209–226.
- Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal
696 Groups. *Science* 329:1191–1194.
- , 2014. A simple Bayesian method of inferring extinction. *Paleobiology*
698 40:584–607.
- Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of
700 filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- Chao, A., T. C. Hsieh, R. L. Chazdon, R. K. Colwell, N. J. Gotelli, and B. D.
702 Inouye, 2015. Unveiling the species-rank abundance distribution by
generalizing the Good-Turing sample coverage theory. *Ecology* 96:1189–1201.
- 704 Cooper, W. S., 1984. Expected time to extinction and the concept of
fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- 706 Crampton, J. S., R. A. Cooper, P. M. Sadler, and M. Foote, 2016.
Greenhouse–icehouse transition in the Late Ordovician marks a step change
708 in extinction regime in the marine plankton. *Proceedings of the National
Academy of Sciences* 113:1498–1503.
- 710 Edwards, D. and U. Fanning, 1985. Evolution and environment in the late
Silurian–early Devonian: the rise of pteridophytes. *Philosophical Transactions
712 of the Royal Society B: Biological Sciences* 309:147–165.
- Ezard, T. H. G., P. N. Pearson, T. Aze, and A. Purvis, 2012. The meaning of
714 birth and death (in macroevolutionary birth-death models). *Biology Letters*
8:139–42.

- ⁷¹⁶ Felsenstein, J., 1985. Phylogenies and the comparative method. *The American Naturalist* 125:1–15.
- ⁷¹⁸ Finnegan, S., N. A. Heim, S. E. Peters, and W. W. Fischer, 2012. Climate change and the selective signature of the Late Ordovician mass extinction. *Proceedings of the National Academy of Sciences* 109:6829–6834.
- ⁷²⁰ Finnegan, S., J. L. Payne, and S. C. Wang, 2008. The Red Queen revisited: reevaluating the age selectivity of Phanerozoic marine genus extinctions. *Paleobiology* 34:318–341.
- ⁷²² Fitzjohn, R. G., 2010. Quantitative Traits and Diversification. *Systematic Biology* 59:619–633.
- ⁷²⁴ Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology* 14:258–271.
- ⁷²⁶ ———, 2000a. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology* 26:74–102.
- ⁷²⁸ ———, 2000b. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.
- ⁷³⁰ ———, 2014. Environmental controls on geographic range size in marine animal genera. *Paleobiology* 40:440–458.
- ⁷³² Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine animal genera. *Paleobiology* 39:171–192.
- ⁷³⁴ Foote, M. and D. Raup, 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 22:121–140.
- ⁷³⁶ Fritz, S. A., J. Schnitzler, J. T. Eronen, C. Hof, K. Böhning-Gaese, and C. H.

- Graham, 2013. Diversity in time and space: wanted dead and alive. *Trends in Ecology & Evolution* 28:509–16.
- 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, FL.
- Gelman, A. and J. Hill, 2007. Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- Goldberg, E. E., L. T. Lancaster, and R. H. Ree, 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology* 60:451–65.
- Goldberg, E. E., K. Roy, R. Lande, and D. Jablonski, 2005. Diversity, endemism, and age distributions in macroevolutionary sources and sinks. *The American Naturalist* 165:623–33.
- Harnik, P. G., 2011. Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proceedings of the National Academy of Sciences* 108:13594–13599.
- Harnik, P. G., P. C. Fitzgerald, J. L. Payne, and S. J. Carlson, 2014. Phylogenetic signal in extinction selectivity in Devonian terebratulide brachiopods. *Paleobiology* 40:675–692.
- Harnik, P. G., C. Simpson, and J. L. Payne, 2012. Long-term differences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society B: Biological Sciences* 279:4969–4976.
- Heim, N. A. and S. E. Peters, 2011. Regional environmental breadth predicts geographic range and longevity in fossil marine genera. *PLoS one* 6:e18946.

- Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL
764 <http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- Hoehn, K. B., P. G. Harnik, and V. L. Roth, 2015. A framework for detecting
766 natural selection on traits above the species level. Methods in Ecology and
768 Evolution Pp. doi: 10.1111/2041-210X.12461.
- Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively
768 Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine
770 Learning Research 15:1351–1381.
- Hopkins, M. J., C. Simpson, and W. Kiessling, 2014. Differential niche dynamics
772 among major marine invertebrate clades. Ecology Letters 17:314–323.
- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
774 comment on "On the heritability of geographic range sizes". The American
776 Naturalist 166:129–135.
- Jablonski, D., 1986. Background and mass extinctions: the alternation of
776 macroevolutionary regimes. Science 231:129–133.
- 778 ———, 1987. Heritability at the species level: analysis of geographic ranges of
780 cretaceous mollusks. Science 238:360–363.
- , 2007. Scale and hierarchy in macroevolution. Palaeontology 50:87–109.
- , 2008. Species Selection: Theory and Data. Annual Review of Ecology,
782 Evolution, and Systematics 39:501–524.
- Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and
784 living molluscs. Proceedings of the Royal Society B: Biological Sciences
786 270:401–406.
- Joachimski, M. M., S. Breisig, W. Buggisch, J. A. Talent, R. Mawson,

- M. Gereke, J. R. Morrow, J. Day, and K. Weddige, 2009. Devonian climate
788 and reef evolution: Insights from oxygen isotopes in apatite. *Earth and*
Planetary Science Letters 284:599–609.
- 790 Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
792 benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology*
33:414–434.
- 794 Klein, J. P. and M. L. Moeschberger, 2003. *Survival Analysis: Techniques for*
Censored and Truncated Data. 2nd ed. Springer, New York.
- 796 Kleinbaum, D. G. and M. Klein, 2005. *Survival analysis: a self-learning text*. 2
ed. Springer, New York, NY.
- 798 Lewandowski, D., D. Kurowicka, and H. Joe, 2009. Generating random
correlation matrices based on vines and extended onion method. *Journal of*
800 *Multivariate Analysis* 100:1989–2001.
- Liow, L. H., 2004. A test of Simpson’s “rule of the survival of the relatively
802 unspecialized” using fossil crinoids. *The American Naturalist* 164:431–43.
- , 2007. Does versatility as measured by geographic range, bathymetric
804 range and morphological variability contribute to taxon longevity? *Global*
Ecology and Biogeography 16:117–128.
- 806 Liow, L. H., M. Fortelius, E. Bingham, K. Lintulaakso, H. Mannila, L. Flynn,
and N. C. Stenseth, 2008. Higher origination and extinction rates in larger
808 mammals. *Proceedings of the National Academy of Sciences* 105:6097–6102.
- Liow, L. H. and J. D. Nichols, 2010. Estimating rates and probabilities of
810 origination and extinction using taxonomic occurrence data:
Capture-mark-recapture (CMR) approaches. Pp. 81–94, *in* J. Alroy and

- 812 G. Hunt, eds. Quantitative Methods in Paleobiology. The Paleontological Society.
- 814 Maddison, W. P., P. E. Midford, and S. P. Otto, 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* 56:701.
- 816 Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings: the kinetics of mass extinction and origination. *Science* 326:1106–9.
- 818 Munnecke, A., M. Calner, D. A. T. Harper, and T. Servais, 2010. Ordovician and Silurian sea-water chemistry, sea level, and climate: A synopsis. *Palaeogeography, Palaeoclimatology, Palaeoecology* 296:389–413.
- Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology* 39:360–372.
- 822 ———, 2015. Interdependence of specialization and biodiversity in Phanerozoic marine invertebrates. *Nature Communications* 6:6602.
- 826 Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental than evolvability. *PloS one* 7:e38025.
- 828 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.
- 830 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 million years. *Proceedings of the Royal Society B: Biological Sciences* 281:20133122.
- 834 Peters, S. E., 2007. The problem with the Paleozoic. *Paleobiology* 33:165–181.

- 836 ———, 2008. Environmental determinants of extinction selectivity in the fossil record. *Nature* 454:626–629.
- 838 Quental, T. B. and C. R. Marshall, 2013. How the Red Queen Drives Terrestrial Mammals to Extinction. *Science* 341:290–292.
- 840 Rabosky, D. L. and A. R. McCune, 2010. Reinventing species selection with molecular phylogenies. *Trends in Ecology & Evolution* 25:68–74.
- 842 Rabosky, D. L., F. Santini, J. Eastman, S. a. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro, 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* 4:1–8.
- 846 Raia, P., F. Carotenuto, F. Passaro, D. Fulgione, and M. Fortelius, 2012. Ecological specialization in fossil mammals explains Cope’s rule. *The American Naturalist* 179:328–37.
- 850 Raia, P., F. Carotenuto, F. Passaro, P. Piras, D. Fulgione, L. Werdelin, J. Saarinen, and M. Fortelius, 2013. Rapid action in the Palaeogene, the relationship between phenotypic and taxonomic diversification in Coenozoic mammals. *Proceedings of the Royal Society B: Biological Sciences* 280:20122244.
- 854 Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen’s Law. *Paleobiology* 1:82–96.
- 856 ———, 1978. Cohort Analysis of generic survivorship. *Paleobiology* 4:1–15.
- , 1991. *Extinction: Bad Genes or Bad Luck?* Norton, New York.
- 858 ———, 1994. The role of extinction in evolution. *Proceedings of the National Academy of Sciences* 91:6758–6763.

- 860 Rosindell, J., L. J. Harmon, and R. S. Etienne, 2015. Unifying ecology and
macroevolution with individual-based theory. *Ecology Letters* 18:472–482.
- 862 Royle, J. A. and R. M. Dorazio, 2008. Hierarchical modeling and inference in
ecology: the analysis of data from populations, metapopulations and
864 communities. Elsevier, London.
- Rubin, D. B., 1996. Multiple imputation after 18+ years. *Journal of the
866 American Statistical Association* 91:473–489.
- Sepkoski, J. J., 1975. Stratigraphic biases in the analysis of taxonomic
868 survivorship. *Paleobiology* 1:343–355.
- , 1981. A factor analytic description of the Phanerozoic marine fossil
870 record. *Paleobiology* 7:36–53.
- Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of
872 Earth and Planetary Sciences* 29:331–364.
- Simpson, C., 2006. Levels of selection and large-scale morphological trends.
874 Ph.D. thesis, University of Chicago.
- Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine
876 bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- Simpson, C., W. Kiessling, H. Mewis, R. C. Baron-Szabo, and J. Müller, 2011.
878 Evolutionary diversification of reef corals: a comparison of the molecular and
fossil records. *Evolution* 65:3274–3284.
- 880 Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University
Press, New York.
- 882 ———, 1953. *The Major Features of Evolution*. Columbia University Press,
New York.

- 884 Slater, G., L. Harmon, and M. Alfaro, 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* 66:3931–3944.
- 886 Slater, G. J., 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution* 4:734–744.
- 888 ———, 2015. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proceedings of the National Academy of Sciences* 112:4897–4902.
- 892 Smithson, M. and J. Verkuilen, 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11:54–71.
- 894 Smits, P. D., 2015. Expected time-invariant effects of biological traits on mammal species duration. *Proceedings of the National Academy of Sciences* 112:13015–13020.
- 898 Solow, A. and W. Smith, 1997. On fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 23:271–277.
- 900 Stadler, T., 2013. Recovering speciation and extinction dynamics based on phylogenies. *Journal of Evolutionary Biology* 26:1203–1219.
- 902 Stadler, T. and F. Bokma, 2013. Estimating speciation and extinction rates for phylogenies of higher taxa. *Systematic biology* 62:220–30.
- 904 Stan Development Team, 2014. Stan: A c++ library for probability and sampling, version 2.5.0. URL <http://mc-stan.org/>.
- 906 Stanley, S. M., 1975. A theory of evolution above the species level. *Proceedings of the National Academy of Sciences* 72:646–650.

- 908 ———, 1979. Macroevolution: pattern and process. W. H. Freeman, San Francisco.
- 910 Tomiya, S., 2013. Body Size and Extinction Risk in Terrestrial Mammals Above the Species Level. *The American Naturalist* 182:196–214.
- 912 Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- , 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.
- 914 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.
- 918 Wagner, P. J. and G. F. Estabrook, 2014. Trait-based diversification shifts reflect differential extinction among fossil taxa. *Proceedings of the National Academy of Sciences* 111:16419–16424.
- 922 Wagner, P. J. and J. D. Marcot, 2013. Modelling distributions of fossil sampling rates over time, space and taxa: assessment and implications for macroevolutionary studies. *Methods in Ecology and Evolution* 4:703–713.
- 924 Wang, S. C., 2003. On the continuity of background and mass extinction. *Paleobiology* 29:455–467.
- 926 Wang, S. C. and C. R. Marshall, 2004. Improved confidence intervals for estimating the position of a mass extinction boundary. *Paleobiology* 30:5–18.
- 928 Williams, C., C. H. C. Brunton, and S. J. Carlson, 2007. Treatise on invertebrate paleontology. Part H, Brachiopoda. Geological Society of America, Boulder, Colorado.

Table 1: Estimates of various parameters in the model used here. These include group-level estimates of the effects of biological traits on brachiopod generic survival, the standard deviation of the between-cohort effects, as well as the estimates of both the effect of sampling δ and the Weibull shape parameter α . The mean, standard deviation (SD), 10th, 50th, and 90th quantiles of the marginal posteriors are presented.

type	parameter	effect of	mean	SD	10%	50%	90%
Mean	μ^i	intercept	-3.05	0.20	-3.30	-3.05	-2.80
	μ^r	geographic range	-0.98	0.16	-1.18	-0.98	-0.79
	μ^v	environmental preference	-0.76	0.19	-0.99	-0.76	-0.52
	μ^{v^2}	environmental preference ²	3.15	0.36	2.69	3.15	3.62
	μ^m	body size	-0.01	0.13	-0.17	-0.01	0.15
	τ^i	intercept	0.51	0.11	0.38	0.50	0.65
Standard deviation	τ^r	geographic range	0.50	0.16	0.30	0.49	0.71
	τ^v	environmental preference	0.84	0.17	0.63	0.82	1.05
	τ^{v^2}	environmental preference ²	1.51	0.36	1.08	1.48	1.97
	τ^m	body size	0.47	0.13	0.32	0.46	0.64
	δ	sampling	0.90	0.15	0.71	0.90	1.09
	α	“time”	1.36	0.04	1.30	1.36	1.42
Other							

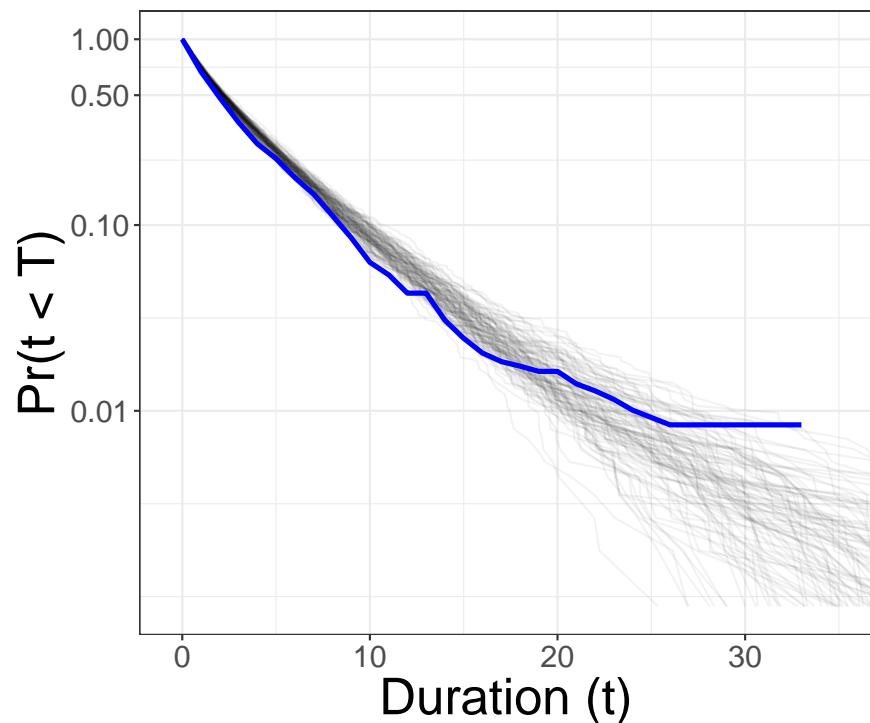


Figure 1: Comparison of the empirical estimate of $S(t)$ (highlighted) versus estimates from 100 posterior predictive data sets (black). $S(t)$ corresponds to the probability that the age of a genus t is less than the genus' ultimate duration T . The vertical axis is log10 transformed.

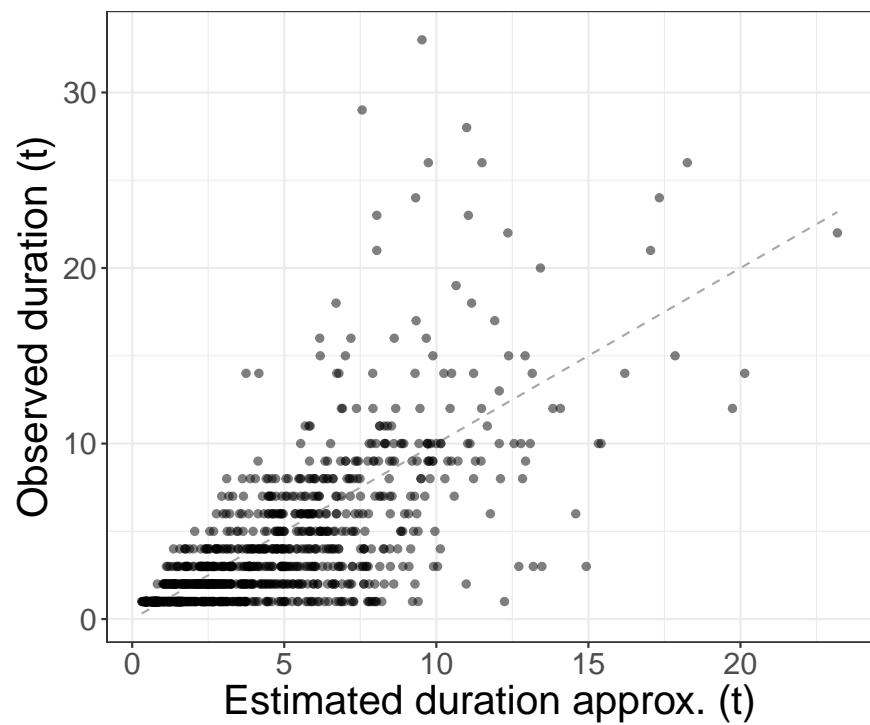


Figure 2: Comparison of all observed genus durations in number of geological stages to the average posterior predictive estimates of $\log(\sigma)$. The dashed, diagonal line corresponds to $x = y$.

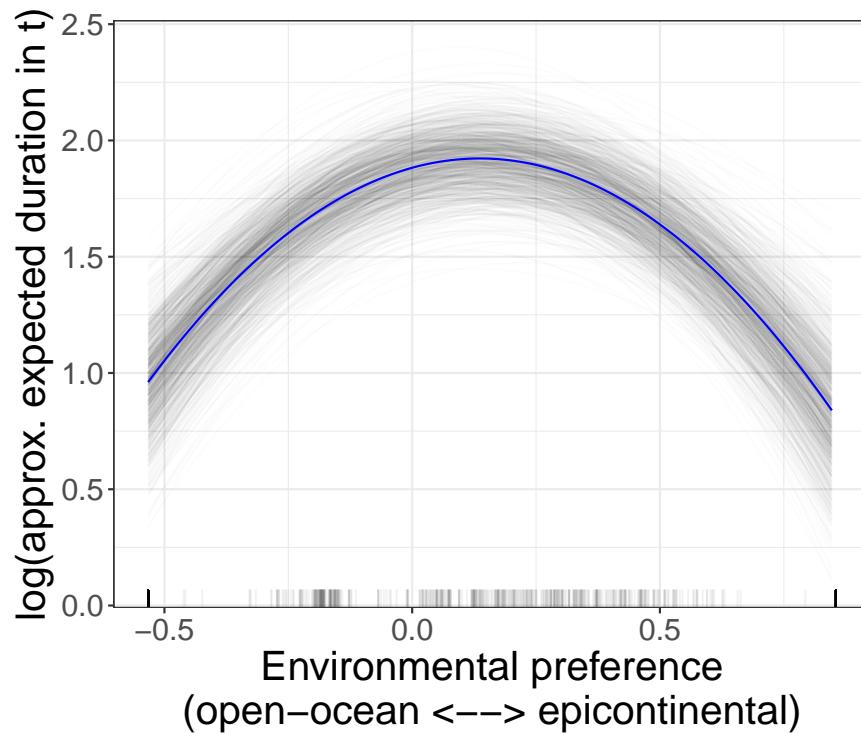


Figure 3: The overall expected relationship between environmental affinity v_i and a $\log(\sigma)$ when $r = 0$ and $m = 0$. The 1000 semi-transparent lines corresponds to a single draw from the posterior predictive distribution, while the highlighted line corresponds to the median of the posterior predictive distribution. The overall relationship is concave down with an optimum greater than 0, which means that taxa favoring epicontinental environments are expected to have longer durations than those favoring open-ocean environments. The tick marks along the bottom of the plot correspond to the (rescaled) observed values of environmental preference.

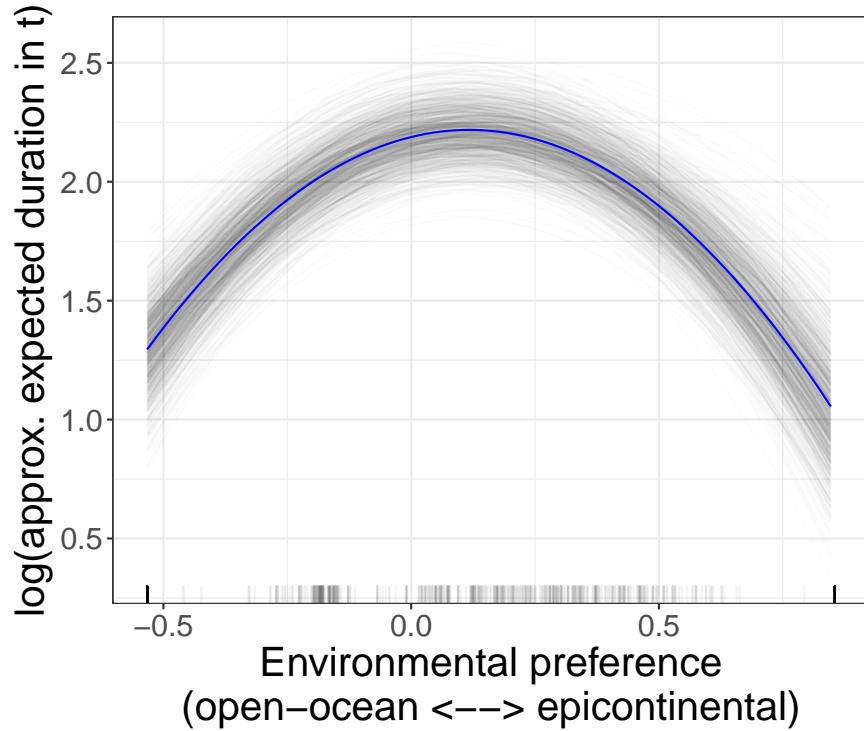


Figure 4: The overall expected relationship between environmental affinity v_i and a $\log(\sigma)$ when $r = 0$ and $m = 0$. The 1000 semi-transparent lines corresponds to a single draw from the posterior predictive distribution, while the highlighted line corresponds to the median of the posterior predictive distribution. The overall relationship is concave down with an optimum greater than 0, which means that taxa favoring epicontinental environments are expected to have longer durations than those favoring open-ocean environments. The tick marks along the bottom of the plot correspond to the (rescaled) observed values of environmental preference.

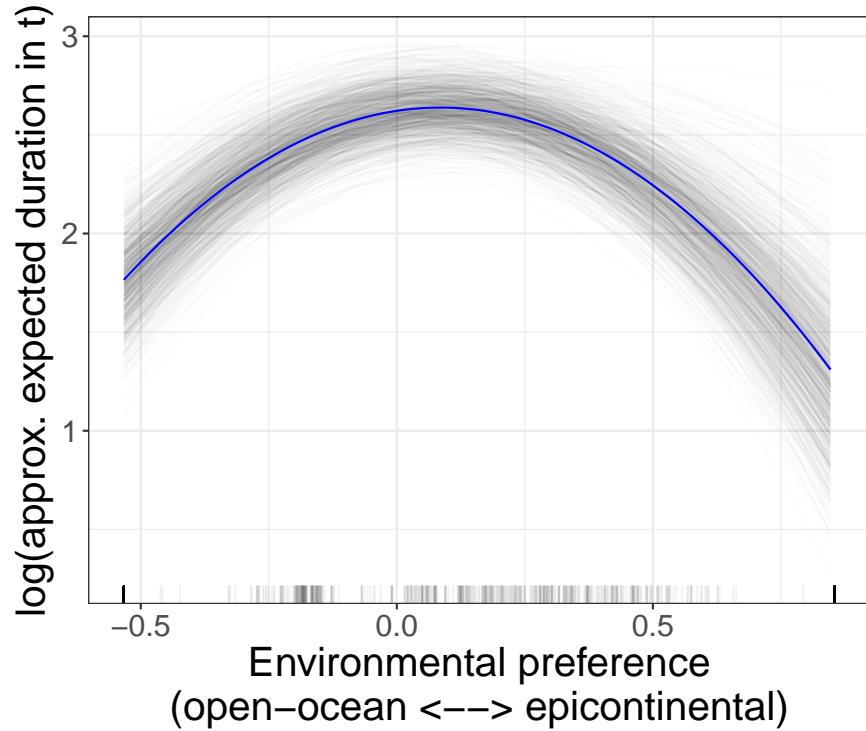


Figure 5: The overall expected relationship between environmental affinity v_i and a $\log(\sigma)$ when $r = 0$ and $m = 0$. The 1000 semi-transparent lines corresponds to a single draw from the posterior predictive distribution, while the highlighted line corresponds to the median of the posterior predictive distribution. The overall relationship is concave down with an optimum greater than 0, which means that taxa favoring epicontinental environments are expected to have longer durations than those favoring open-ocean environments. The tick marks along the bottom of the plot correspond to the (rescaled) observed values of environmental preference.

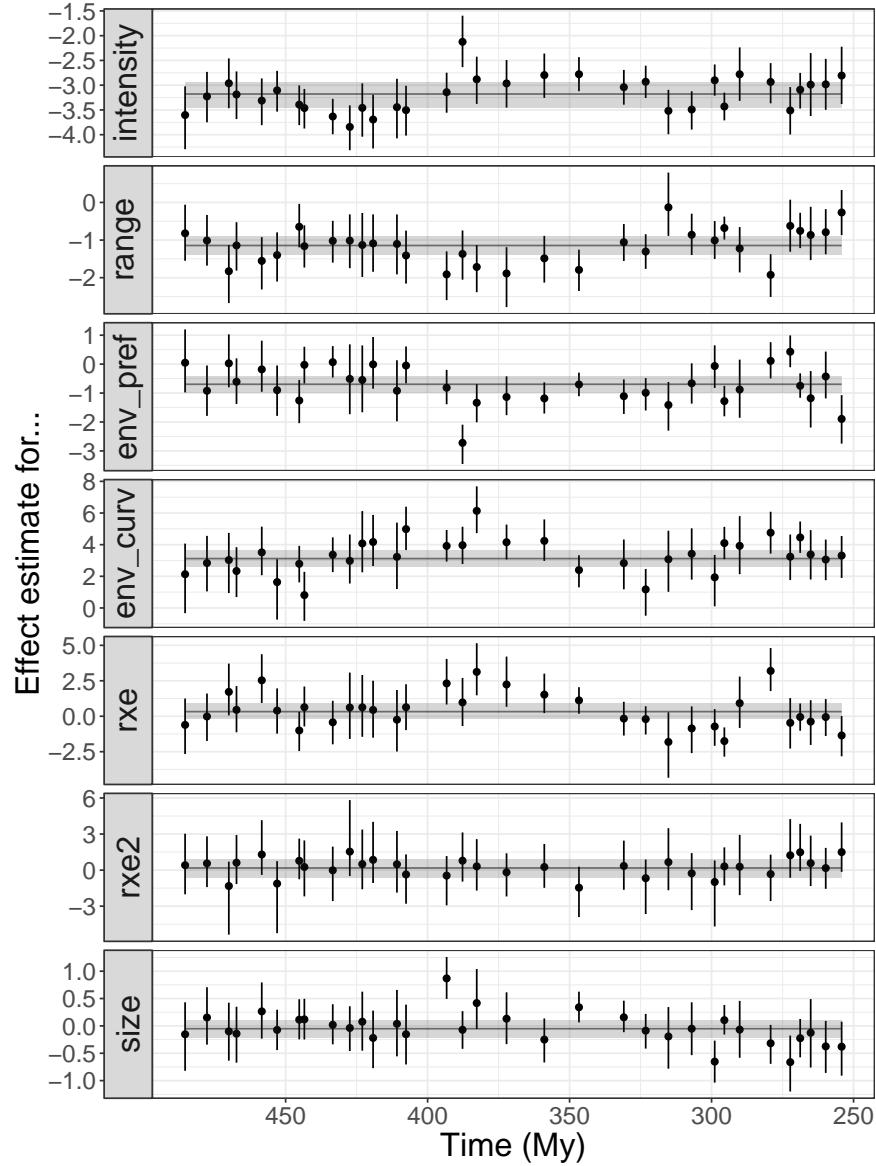


Figure 6: Comparison of cohort-specific estimates of β^0 , the effect of geographic range on extinction risk β^r , the effect of environmental preference β^v and β^{v^2} , and body size β^m . Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. Points are plotted at the midpoint of the cohorts stage of origination in millions of years before present (My). Black, horizontal lines are the overall estimates of covariate effects along with 80% credible intervals (shaded).

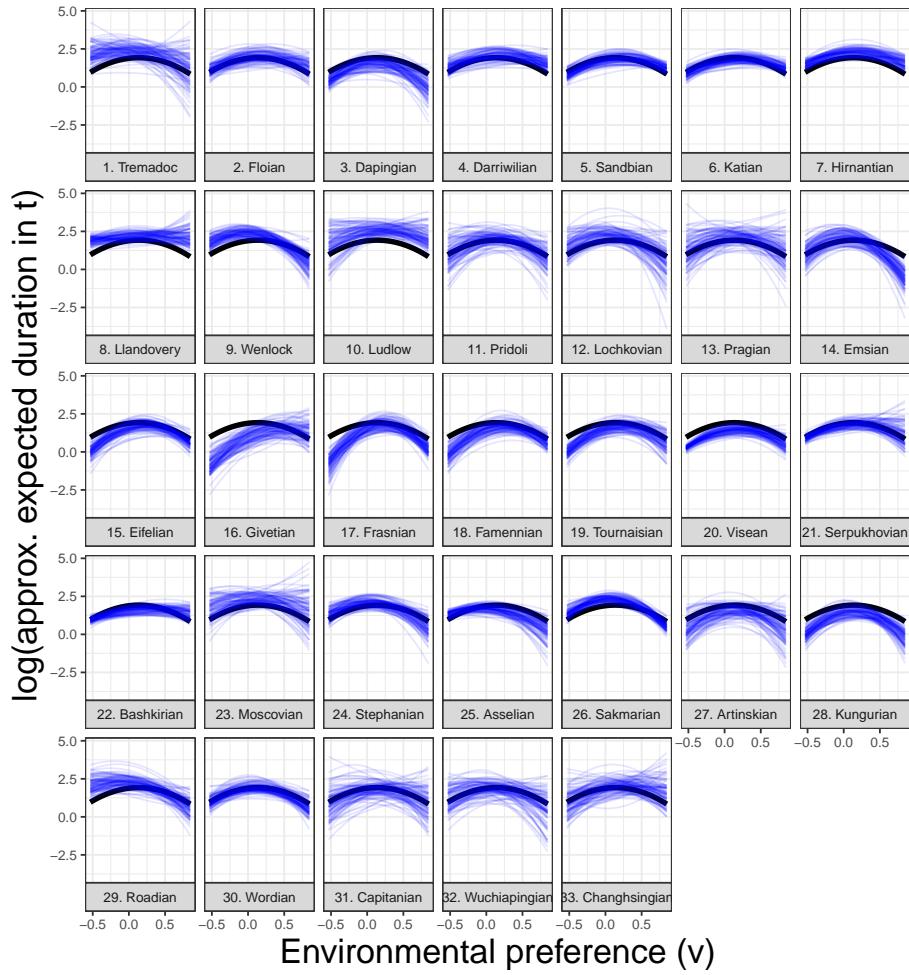


Figure 7: Comparison of origination cohort-specific (posterior predictive) estimates of the effect of environmental preference on $\log(\sigma)$ to the mean overall estimate of the effect of environmental preference. Cohort-specific estimates are from 100 posterior predictive simulations across the range of (transformed and rescaled) observed values of environmental preference. The oldest cohort is in the top-right and younger cohorts proceed left to right, with the youngest cohort being the right-most facet of the last row. Panel names correspond to the name of the stage in which that cohort originated.

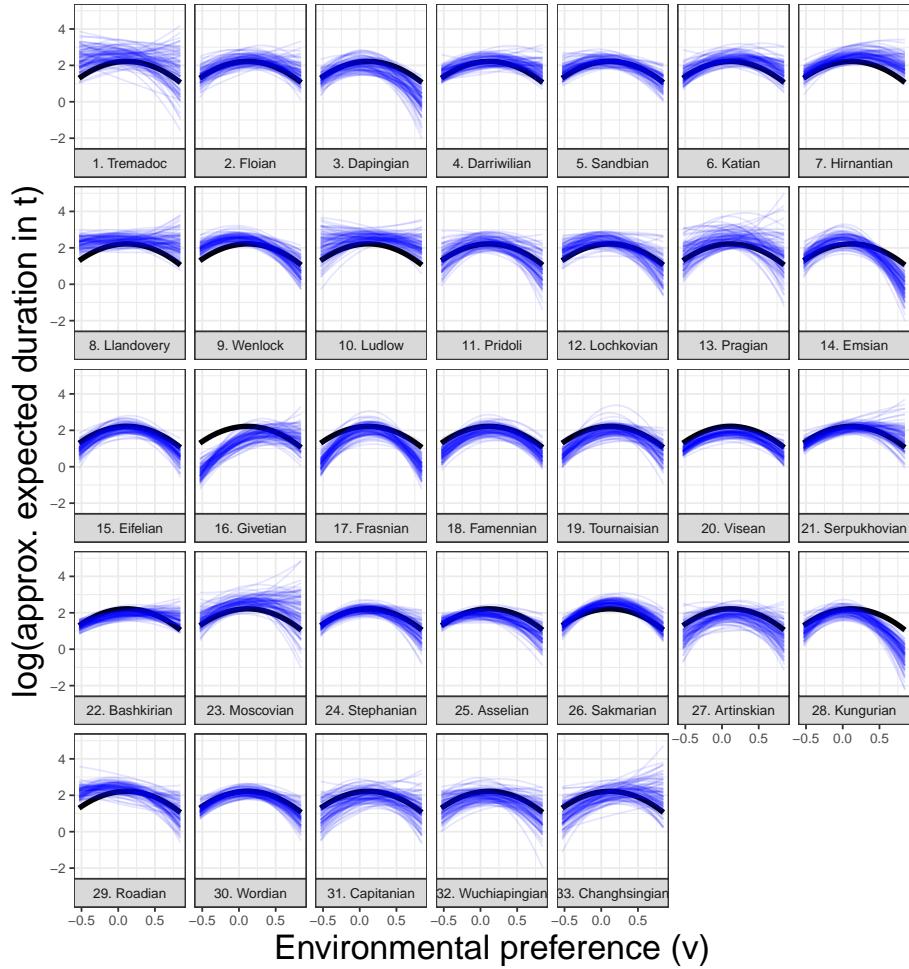


Figure 8: Comparison of origination cohort-specific (posterior predictive) estimates of the effect of environmental preference on $\log(\sigma)$ to the mean overall estimate of the effect of environmental preference. Cohort-specific estimates are from 100 posterior predictive simulations across the range of (transformed and rescaled) observed values of environmental preference. The oldest cohort is in the top-right and younger cohorts proceed left to right, with the youngest cohort being the right-most facet of the last row. Panel names correspond to the name of the stage in which that cohort originated.

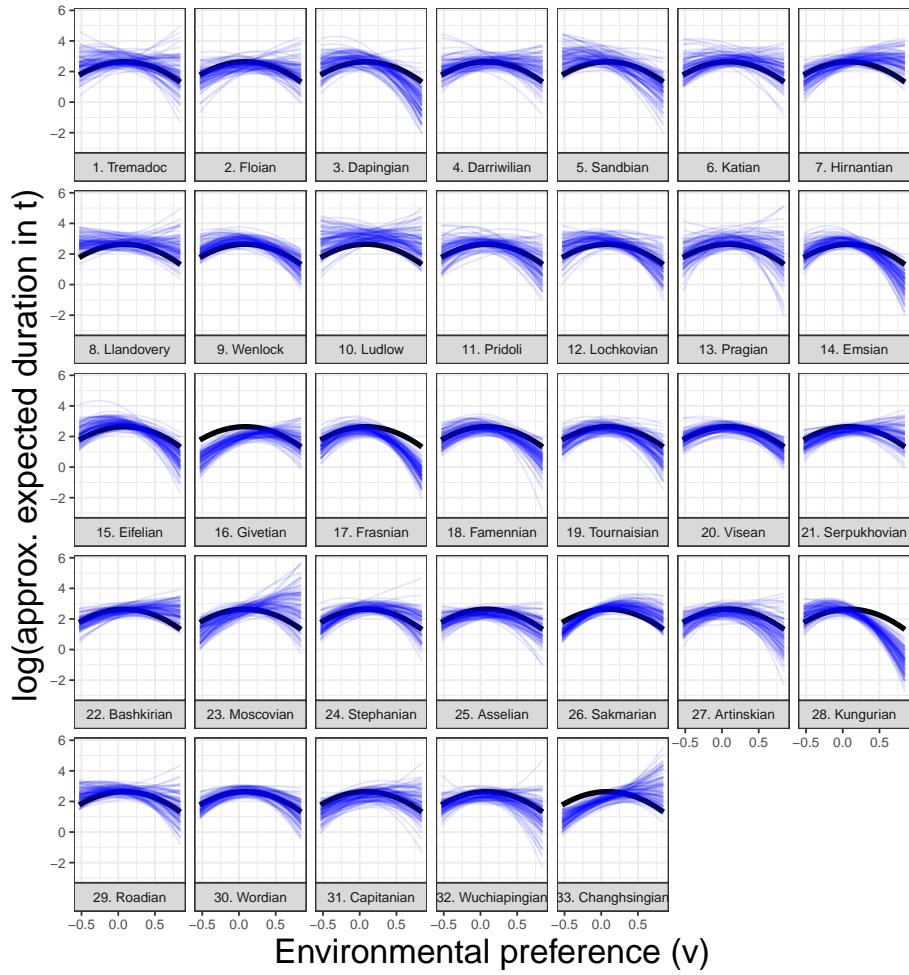


Figure 9: Comparison of origination cohort-specific (posterior predictive) estimates of the effect of environmental preference on $\log(\sigma)$ to the mean overall estimate of the effect of environmental preference. Cohort-specific estimates are from 100 posterior predictive simulations across the range of (transformed and rescaled) observed values of environmental preference. The oldest cohort is in the top-right and younger cohorts proceed left to right, with the youngest cohort being the right-most facet of the last row. Panel names correspond to the name of the stage in which that cohort originated.

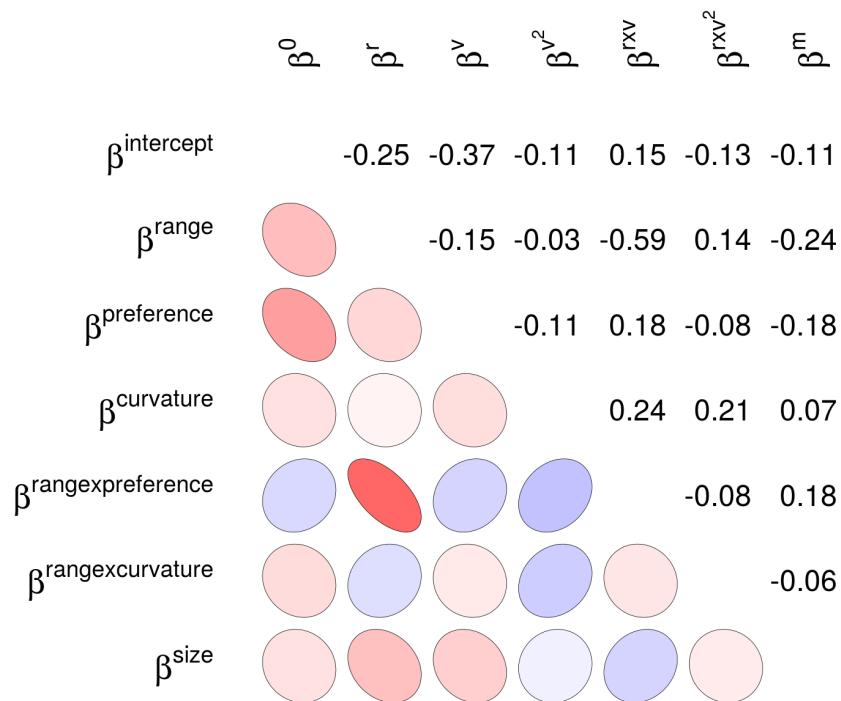


Figure 10: Mixed graphical and numerical representation of the correlation matrix Ω of variation in cohort-specific covariate estimates. These correlations are between the estimates of the cohort-level effects of covariates, along with intercept/baseline extinction risk. The median estimates of the correlations are presented numerically (upper-triangle) and as idealized ellipses representing that much correlation (lower-triangle). The darkness of the ellipse corresponds to the magnitude of the correlation.

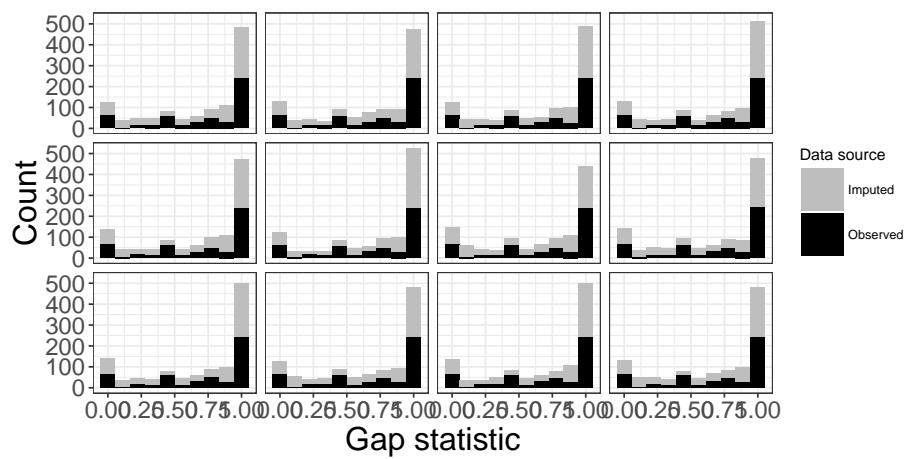


Figure 11: Histograms of the distribution of gap statistic values from both the observed values and the imputed values from 12 unique posterior realizations. For each panel the observed values are identical but the imputed values are from a single set of their posterior estimates.