

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

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

Keywords: extinction, brachiopods, species selection, imputation, Bayesian

8 **Word count:**

Table count: 1

10 **Figure count:** 6

Data archival location:

Abstract

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

1 Introduction

38 Extinction is one half of the diversification process (Raup, 1994, Stanley, 1975,
1979), second only to speciation or origination; it can also be the ultimate
40 manifestation of selection as a taxon with a beneficial trait should persist for
longer on average than a taxon without that beneficial trait (Jablonski, 2008,
42 Rabosky and McCune, 2010, Raup, 1994, Stanley, 1975). How do
trait-dependent extinction rates change over time and in relation to extinction
44 intensity?

While estimation of both trait-dependent speciation and extinction rates from
46 phylogenies of extant taxa has grown dramatically (Fitzjohn, 2010, Goldberg
et al., 2011, 2005, Maddison et al., 2007, Rabosky et al., 2013, Stadler, 2011,
48 2013, Stadler and Bokma, 2013), there are two major ways to estimate
trait-dependent extinction: analysis of phylogenies, and analysis of the fossil
50 record. These two directions, phylogenetic comparative and paleobiological, are
complementary and intertwined in the field of macroevolution (Hunt and
52 Rabosky, 2014, Jablonski, 2008, Rabosky and McCune, 2010). In the case of
extinction, analysis of the fossil record has the distinct advantage over
54 phylogenies of only extant taxa because extinction is observable; this means
that extinction rate is possible to estimate (Liow et al., 2010, Quental and
56 Marshall, 2009, Rabosky, 2010). The approach used here is thus complementary
to the analysis of trait-dependent extinction based on a phylogeny.

58 Jablonski (1986) observed that for bivalves at the end Cretaceous mass
extinction event, the effects of some biological traits on taxonomic survival
60 decreased. However, this pattern was not the case for the effect of geographic
range on survival (Jablonski, 1986, Payne and Finnegan, 2007). There are
62 multiple possible macroevolutionary mechanisms which may underlie this

pattern: the effect of geographic range on survival remains constant and those of
64 other biological traits decrease, the effect of geographic range on survival
increases and those of other biological traits stay constant, or the effects of all
66 traits decrease potentially by different degrees.

While Jablonski (1986) phrased his conclusions in terms of background versus
68 mass extinction, these states are not distinguishable in terms of extinction rate
alone; my analysis treats the time period analyzed as part of the same
70 continuum (Payne and Finnegan, 2007, Simpson and Harnik, 2009, Wang, 2003).
Additionally, in order to test the proposed macroevolutionary mechanism behind
72 the Jablonski (1986) scenario; not only do the taxon trait effects need to be
modeled, but the correlation between trait effects need to be modeled as well.

74 Here I model brachiopod taxon durations because trait based differences in
extinction risk should manifest as differences in taxon durations. Brachiopods
76 are an ideal group for this study as they are well known for having an
exceptionally complete fossil record (Foote, 2000b, Foote and Raup, 1996). I
78 focus on the brachiopod record from the post-Cambrian Paleozoic, from the
start of the Ordovician (approximately 485 My) through the end Permian
80 (approximately 252 My) as this represents the time of greatest global
brachiopod diversity (Alroy, 2010) meaning a large sample size for this analysis.

82 The analysis of taxon durations, or time from origination to extinction, falls
under the purview of survival analysis, a field of applied statistics commonly
84 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,
86 2015, Van Valen, 1973, 1979). I adopt a hierarchical modeling approach
(Gelman et al., 2013, Gelman and Hill, 2007, Klein and Moeschberger, 2003),
88 which represents both a conceptual and statistical unification of the
paleontological dynamic and cohort survival analytic approaches (Baumiller,

90 1993, Crampton et al., 2016, Ezard et al., 2012, Foote, 1988, Raup, 1975, 1978,
Simpson, 2006, Smits, 2015, Van Valen, 1973, 1979).

92 **1.1 Factors affecting brachiopod survival**

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

Geographic range is widely considered the most important biological trait for
104 estimating differences in extinction risk at nearly all times, with large
geographic range associated with low extinction risk (Finnegan et al., 2012,
106 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
108 random; a taxon with an unrestricted range is less likely to go extinct at
random than a taxon with a restricted range.

110 Epicontinental seas are a shallow-marine environment where the ocean has
spread over the continental interior or craton of a continental shelf with a depth
112 typically less than 100m. In contrast, open-ocean coastline environments have
much greater variance in depth, do not cover the continental craton, and can
114 persist during periods of low sea level (Miller and Foote, 2009). Because of this,

it is strongly expected that taxa which favor epicontinental seas would be at
116 great risk during periods of low sea levels, such as during glacial periods, when
epicontinental seas are drained. During the Paleozoic (approximately 541–252
118 My), epicontinental seas were widely spread globally but declined over the
Mesozoic (approximately 252–66 My) and have nearly disappeared during the
120 Cenozoic (approximately 66–0 My) as open-ocean coastlines became the
dominant shallow-marine setting (Johnson, 1974, Miller and Foote, 2009, Peters,
122 2008).

Miller and Foote (2009) demonstrated that during several mass extinctions taxa
124 associated with open-ocean environments tend to have a greater extinction risk
than those taxa associated with epicontinental seas. During periods of
126 background extinction, however, they found no consistent difference between
taxa favoring either environment. Miller and Foote (2009) hypothesize that
128 open-ocean taxa may have a greater extinction rate because these environments
would be more strongly affected by waterborne selective pressures which would
130 propagate more quickly than in epicontinental environments with sluggish
circulation. These two environment types represent the primary identifiable
132 environmental dichotomy observed in ancient marine systems (Miller and Foote,
2009, Peters, 2008, Sheehan, 2001). Given these findings, I predict that as
134 extinction risk increases, the extinction risk associated with open-ocean
environments should generally increase.

136 Because environmental preference is defined here as the continuum between
occurring exclusively in open-ocean environments versus epicontinental
138 environments, intermediate values are considered “generalists” in the sense that
they favor neither end member. A long-standing hypothesis is that generalists or
140 unspecialized taxa will have greater survival than specialists (Baumiller, 1993,
Liow, 2004, 2007, Nürnberg and Aberhan, 2013, 2015, Simpson, 1944, Smits,

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 end-members.

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 (Payne et al., 2014). Harnik et al. (2014) analyzed the effect of body size selectivity in Devonian brachiopods in both a phylogenetic and 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 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 results, I expect that if body size has any effect on brachiopod taxonomic survival it is very small.

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, 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 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 case, for sampling to be a confounding factor there must be consistent relationship between the quality of sampling of a taxon and its apparent duration (e.g. greater sampling, longer duration). If there is no relationship between sampling and duration then interpretation can be made clearly; while 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

major determinant of the taxon’s apparent duration. By including sampling as a
170 covariate in the model, this effect can be quantified and can be taken into
account when interpreting the estimates of the effects of the other covariates.

172 **2 Materials and Methods**

2.1 Fossil occurrence information

174 The brachiopod dataset analyzed here was sourced from the Paleobiology
Database (<http://www.paleodb.org>) which was then filtered based on
176 taxonomic (Rhynchonelliformea: Rhynchonellata, Chileata, Obolellida,
Kutorginida, Strophomenida, Spiriferida), temporal (post-Cambrian Paleozoic),
178 stratigraphic, and other occurrence information used in this analysis. Analyzed
occurrences were restricted to those with paleolatitude and paleolongitude
180 coordinates, assignment to either epicontinental or open-ocean environment,
and belonging to a genus present in the body size dataset (Payne et al., 2014).
182 Epicontinental versus open-ocean assignments for each fossil occurrence are
partially based on those from Miller and Foote (2009), with additional
184 occurrences assigned similarly (Miller and Foote, personal communication).
These filtering criteria are very similar to those from Foote and Miller (2013)
186 with an additional constraint of being present in the body size data set from
Payne et al. (2014). In total, there 1130 were genera included in the dataset.
188 Fossil occurrences were analyzed at the genus level which is common for
paleobiological, macroevolutionary and macroecological studies of marine
190 invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2012, Kiessling
and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,
192 Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).

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

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

210 The covariates of duration included in this analysis are geographic range size
(r), environmental preference (v, v^2), body size (m), and sampling (s).

212 Geographic range was calculated using an occupancy approach corrected for
incomplete sampling. First, the paleolatitude-paleolongitude coordinates for all
214 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
216 grid placed on the map. Each grid cell represents approximately 250,000 km².
The map projection and regular lattice were made using shape files from
218 <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,
2015). For each stage, the total number of occupied grid cells was calculated.

220 Then, for each temporal bin, the relative occurrence probability of the observed
taxa was calculated using the JADE method developed by Chao et al. (2015).
222 This method accounts for the fact that taxa with an occupancy of 0 cannot be
observed which means that occupancy follows a truncated Binomial distribution.
224 This correction is critical when comparing occupancies from different times with
different geographic sampling. Finally, for each genus, the mean relative
226 occurrence probability was calculated as the average of that genus' occurrence
probabilities for all stages it was sampled.

228 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
230 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
232 expressed.

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

where v is the percent of the distribution defined in equation 1 less than or
234 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.

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

The sampling probability for individual taxa was calculated using the standard
240 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
242 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

244 by definition sampled. Because taxa that were right censored only include a first
appearance, one was subtracted from the numerator and denominator instead of
246 two.

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

Prior to analysis, geographic range was logit transformed and body size was
256 natural-log transformed; both of these transformations make these variables
defined for the entire real line. Sampling probability was transformed as
258 $(s(n - 1) + 0.5)/n$ where n is the sample size as recommended by Smithson and
Verkuilen (2006); this serves to slightly shrink the range of the data so that
260 there are no values of 0 or 1. All covariates except for sampling were
standardized by subtracting the mean from all values and dividing by twice its
262 standard deviation, which follows Gelman and Hill (2007). This standardization
means that the associated regression coefficients are comparable as the expected
264 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.

266 2.2 Details of model

Hierarchical modelling is a statistical approach which explicitly takes into
268 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
 270 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
 272 (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
 274 distributions and are also estimated like the other parameters of interest (e.g.
 covariate effects) (Gelman et al., 2013). The subsequent estimates are partially
 276 pooled together, where parameters from groups with large samples or effects
 remain large while those of groups with small samples or effects are pulled
 278 towards the overall group mean. All covariate effects (regression coefficients), as
 well as the intercept term (baseline extinction risk), were allowed to vary by
 280 group (origination cohort). The covariance between covariate effects was also
 modeled.

282 Genus durations were assumed to follow a Weibull distribution which allows for
 age-dependent extinction (Klein and Moeschberger, 2003): $y \sim \text{Weibull}(\alpha, \sigma)$.
 284 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
 286 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
 288 extinction risk decreases with age. Note that the Weibull distribution is
 equivalent to the exponential distribution when $\alpha = 1$.

290 In the case of the right- and left-censored observations mentioned above, the
 probability of those observations has a different calculation (Klein and
 292 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
 294 function. In contrast, the likelihood of a left-censored observation is calculated
 $p(y|\theta) = F(y)$.

296 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
 298 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
 300 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
 302 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
 304 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
 306 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
 308 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.

310 The elements of μ were given independent normally distributed priors. The
 effects of geographic range size and the breadth of environmental preference
 312 were given informative priors reflecting the previous findings while the others
 were given weakly informative favoring no effect. The correlation matrix Ω was
 314 given an almost uniform LKJ distributed prior proposed by Lewandowski et al.
 (2009) and recommended by Stan Development Team (2014b). These priors are

316 defined

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

The log of the shape parameter α was given a weakly informative prior
318 $\log(\alpha) \sim \mathcal{N}(0, 1)$ centered at $\alpha = 1$, which corresponds to the Law of Constant
Extinction (Van Valen, 1973).

320 2.3 Imputation of sampling probability

The vector sampling s has two parts: the observed part s^o , and the unobserved
322 part s^u . Of the 1130 total observations, 539 have a duration of 3 or more and
have an observed gap statistic. The gap statistic for the remaining 591
324 observations was imputed. As stated above, the unobserved part is the imputed,
or filled in, based on the observed part s^o . Because sampling varies between 0
326 and 1, I chose to model it has a Beta regression with the other covariates X (e.g.
geographic range size) as predictors of sampling; this assumes that the sampling
328 value for all taxa come from the same distribution. Predicting sampling
probability from the other covariate information and then using it as an
330 additional covariate of duration is acceptable in the case of imputation where
the sample goal is accurate prediction (Gelman and Hill, 2007). Importantly, I
332 make no assumptions of causal structure.

The Beta regression is defined

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

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

mean parameter

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

and total count parameter

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

where α and β are the characteristic parameters of the Beta distribution.

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

elements of γ , and both δ (Eq. 2) and λ (Eq. 4) were given weakly informative

priors where

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

2.4 Posterior inference and posterior predictive checks

The joint posterior was approximated using a Markov-chain Monte Carlo

routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn

Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic

programming language Stan (Stan Development Team, 2014a). The posterior

distribution was approximated from four parallel chains run for 10,000 steps

each, split half warm-up and half sampling and thinned to every 10th sample for

350 a total of 4000 posterior samples. Chain convergence was assessed via the scale
reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$) indicate approximate
352 convergence. Convergence means that the chains are approximately stationary
and the samples are well mixed (Gelman et al., 2013).

354 Model adequacy was evaluated using a couple of posterior predictive checks.
Posterior predictive checks are a means for understanding model fit or adequacy
356 where the basic idea is that replicated data sets simulated from the fitted model
should be similar to the original data and systematic differences between the
358 simulations and observations indicate weaknesses of the model fit (Gelman
et al., 2013). For both approaches used here, 1000 posterior predictive datasets
360 were generated from 1000 unique draws from the posterior distribution of each
parameter. The two posterior predictive checks used in this analysis are a
362 comparison of a non-parameteric estimate of the survival function $S(t)$ from the
empirical dataset to the non-parameteric estimates of $S(t)$ from the 1000
364 posterior predictive datasets, and comparison of the observed genus durations to
the average posterior predictive estimate of $\log(\sigma)$ (Eq. 2). The former is to see
366 if simulated data has a similar survival pattern to the observed, while the latter
is to see if the model systematically over- or under- estimates taxon survival.

368 **3 Results**

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

the majority of observations, model fit is generally not biased.

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

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

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

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

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

398 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
400 detectable visually (Fig. 4), 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
402 β^{v^2} .

While most cohort-specific estimates are very similar to the overall cohort-level
404 estimate, there are a few notable excursions away from the overall mean (Fig.
4). There are simultaneous excursions in both β^0 and β^v for cohorts originating
406 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
408 Frasnian/Famennian boundary. These cohorts are marked by both a high
extinction intensity and an increase in expected duration for taxa favoring
410 epicontinental environments over open-ocean ones; this is consistent with the
results of Miller and Foote (2009).

412 Cohorts originating from the Silurian through the Early Devonian have a
slightly lower extinction intensity than the overall mean; these cohorts are those
414 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
416 that epicontinental favoring taxa are expected to have greater duration than
open-ocean favoring taxa. Both the Silurian and Devonian periods are notable
418 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,
420 Munnecke et al., 2010).

The cohort-specific relationships between environmental preference and $\log(\sigma)$
422 were calculated from the estimates of β^0 , β^v , and β^{v^2} (Fig. 5) and reflect how
these three parameters act in concert and not just individually (Fig. 4). Beyond
424 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
426 like the overall expected relationship and has the most sharply curved
appearance due to a high estimate β^{v^2} (Fig. 4). This cohort has the biggest

428 difference in extinction risk between environmental generalists and specialists.
 The cohorts originating during the Emsian (407-393 My) and Frasnian (382 -
 430 372 My) are tied for second in sharpness of curvature. The least sharply curved
 cohorts include those originating during Tremadocian (484-477 My), Hirnantian
 432 (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
 434 through the Early Devonian range identified earlier as having lower expected
 extinction intensity than what is expected from the group-level estimate.

436 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
 438 the elements of Ω are distinguishable from 0: the correlation of β^0 (extinction
 intensity) with both β^r and β^v (Fig. 6).

440 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
 442 negatively correlated; this means that for cohorts experiencing a lower
 extinction intensity (β^0 decreases), the magnitude of the effect of geographic
 444 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.

446 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
 448 intensity increases it is expected that epicontinental taxa become more favored
 over open-ocean environments (i.e. as β^0 increases, β^v decreases).

450 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
 452 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

454 1); the effect of geographic range might simply not vary enough relative to the
much noisier environmental preference.

456 Sampling was found to have a negative effect (positive δ) on duration: greater
sampling, shorter duration (Table 1). While this effect appears large compared
458 to the other regression coefficients, this is only because sampling was not
standardized like the other covariates.

460 To make the coefficients comparable, δ is multiplied by twice the posterior mean
of the standard deviation of sampling probability; the transformed value of δ is
462 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% probability that the effect of
464 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 than δ . Finally, μ^{v^2} has a
466 100% probability of having a greater magnitude of effect than δ .

The Weibull shape parameter α was found to be approximately 1.36 (± 0.05 SD)
468 with a 100% probability of being greater than 1. This result is not consistent
with the Law of Constant Extinction (Van Valen, 1973) and is instead
470 consistent with accelerating extinction risk with taxon age. This may indicate
that older taxa are out-competed by younger taxa, a result consistent with some
472 empirical results (Quental and Marshall, 2013, Smits, 2015, Wagner and
Estabrook, 2014) and (ironically) with a recently proposed Red Queen-like
474 model of evolution (Rosindell et al., 2015). This results, however, is not
consistent with other empirical results (Crampton et al., 2016, Finnegan et al.,
476 2008) and could potentially be caused by the minimum resolution of the fossil
record (Sepkoski, 1975). It is thus unclear if a strong biological inference can be
478 made from this result, which means that further work is necessary on the effect
of taxon age on extinction risk.

480 4 Discussion

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

I find that as intensity increases (β^0 decreases), the magnitude of the effect of
geographic range increases. I also find that as intensity increases, the effect of
490 favoring epicontinental environments over open-ocean environments is expected to
increase; this is consistent with the results of Miller and Foote (2009). There
492 is no evidence for a correlation between the effect of geographic range and
environmental preference. Additionally, the between-cohort variance in effect of
geographic range is much less than the between-cohort variance of the effect of
494 environmental preference. As stated earlier, this may underlie the lack of
correlation between these two effects.
496

Additionally, the lower between-cohort variance in the effect of geographic range
498 versus that higher between-cohort variance implies that for cohorts with a
greater than average extinction intensity, the difference in the effect of geographic
range and the group-level effect of geographic range is expected to be smaller
500 than the difference between the effect of environmental preference and the
group-level effect of environmental preference.
502

I find consistent support for the “survival of the unspecialized,” with respect to
504 epicontinental versus open-ocean environmental preference, as a time-invariant
generalization of brachiopod survival; taxa with intermediate environmental

506 preferences are expected to have lower extinction risk than taxa specializing in
 either epicontinental or open-ocean environments (Fig. 3), though the curvature
 508 of the relationship varies from rather shallow to very peaked (Fig. 5). However,
 this relationship is not symmetric about 0, as taxa favoring epicontinental
 510 environments are expected to have a greater duration than taxa favoring
 open-ocean environments. This description of environment only describes one
 512 major aspect of a taxon's environmental context, with factors such as
 bathymetry and temperature being further descriptors of a taxon's adaptive
 514 zone (Harnik, 2011, Harnik et al., 2012, Heim and Peters, 2011, Nürnberg and
 Aberhan, 2013); inclusion of these factors in future analyses would potentially
 516 improve our understanding of the “survival of the unspecialized” hypothesis
 (Simpson, 1944).

518 Hopkins et al. (2014), in their analysis of niche conservatism and substrate
 lithological preference in marine invertebrates, found that brachiopods were
 520 among the least “conservative” groups; taxa were found to easily change
 substrate preference on short time scales. While substrate preference is not the
 522 same as environmental preference (as defined here), a question does arise: are
 there three classes of environmental preference instead of two? These classes
 524 would be taxa with broad tolerance (“true” generalists), inflexible specialists
 (“true” specialists), and flexible but with a narrow tolerance. A flexible taxa is
 526 one with a narrow habitat preference at one time, but with preference that
 changes over time with changing environmental availability. My analysis
 528 assumes that traits are constant over the duration of the taxon meaning that
 this scenario is not detectable; future work should explore how environmental
 530 preference changes over lineage duration in relation to environmental availability
 to estimate if the generalists–specialists continuum is actually ternary
 532 relationship.

An alternative approach for specifically modeling survival that can take into
534 account imperfect observation than the method used here is the
Cormack-Jolly-Seber (CJS) model (Liow et al., 2008, Liow and Nichols, 2010,
536 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
538 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
540 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:
542 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
544 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
546 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
548 might increase the duration of genera through self-renewal (Raup, 1991, 1994),
which would be an example of the difference in biological pattern between
550 species and genera (Jablonski, 1987, 2007, 2008). This could lead to a trait
appearing to decrease generic level extinction risk by that trait increasing
552 species level origination rate instead of decreasing species level extinction risk.

The model used here could be improved through either increasing the number of
554 analyzed traits, expanding the hierarchical structure of the model to include
other major taxonomic groups of interest, and the inclusion of explicit
556 phylogenetic relationships between the taxa in the model as an additional
hierarchical effect. An example trait that may be of particular interest is the
558 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

560 taxa, taxa that are attached to the substrate are expected to have a greater
duration than those that are not (Alexander, 1977).

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

With significant updates, it would also be possible to compare the brachiopod
570 record with with Moden groups such as bivalvae or brachiopods (Sepkoski,
1981), though remembering that the groups may not necessarily share all
572 cohorts with the brachiopods. This particular model expansion would act as a
test of any universal cross-taxonomic patterns in the effects of emergent traits
574 on extinction such as has been proposed for geographic range (Payne and
Finnegan, 2007). Additionally, this expanded model would also act as a test of
576 the distinctness of the Sepkoski (1981) three-fauna hypothesis in terms of
trait-dependent extinction.

578 Traits like environmental preference or geographic range (Hunt et al., 2005,
Jablonski, 1987) are most likely heritable. Without phylogenetic context, this
580 analysis assumes that differences in extinction risk between taxa are
independent of the shared evolutionary history of those taxa (Felsenstein, 1985).
582 In contrast, the origination cohorts only capture shared temporal context. For
example, if taxon duration is phylogenetically heritable, then closely related
584 taxa may have more similar durations as well as more similar biological traits.
Without taking into account phylogenetic similarity the effects of these
586 biological traits would be inflated solely due to inheritance. The inclusion of

phylogenetic context as an additional individual-level hierarchical effect,
 588 independent of origination cohort, would allow for determining how much of the
 observed variability is due to shared evolutionary history versus shared temporal
 590 context versus actual differences associated with biological traits (Harnik et al.,
 2014, Smits, 2015). The combination of the phylogenetic comparative and
 592 paleontological approaches requires both sources of data which is not possible
 for this analysis because there is no phylogenetic hypothesis for all Paleozoic
 594 taxa, something that is frequently the case for marine invertebrates with a good
 fossil record. When both data sources are available has been possible, however,
 596 the analysis can more fully address the questions of interest in macroevolution
 (Fritz et al., 2013, Harnik et al., 2014, Raia et al., 2012a,b, Simpson et al., 2011,
 598 Slater et al., 2012, Slater, 2013, 2015, Smits, 2015, Tomiya, 2013).

In summary, patterns of Paleozoic brachiopod survival were analyzed using a
 600 fully Bayesian hierarchical survival modelling approach while also eschewing the
 traditional separation between background and mass extinction. I find that
 602 cohort extinction intensity is negatively correlated with both the cohort-specific
 effects of geographic range and environmental preference. These results imply
 604 that as extinction intensity increases (β^0) increases, it is expected that both
 effects will increase in magnitude. However, the change in effect of
 606 environmental preference is expected to be greater than the change in the effect
 of geographic range. Additionally, I find support for greater survival in
 608 environmental generalists over specialists in all origination cohorts analyzed;
 this is consistent with the long standing “survival of the unspecialized”
 610 hypothesis (Baumiller, 1993, Liow, 2004, 2007, Nürnberg and Aberhan, 2013,
 2015, Simpson, 1944, 1953, Smits, 2015). The results of this analysis support the
 612 conclusion that for Paleozoic brachiopods, as extinction intensity increases
 overall extinction selectivity is expected to increase as well.

614 Acknowledgements

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, R. Ree, and G. Slater
616 for helpful discussion during the conception of this study. I'd also like to thank
D. Bapst, N. Pierrehumbert and M. Villarosa Garcia for additional comments.
618 An additional thank you to A. Miller for the epicontinental versus open-ocean
assignments. Finally, thank you to the reviewers for their helpful comments that
620 improved this manuscript. This entire study would would not have been possible
without the Herculean effort of the many contributors to the Paleobiology
622 Database. In particular, I would like to thank J. Alroy, M. Aberhan, D. Bottjer,
M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland, L. Ivany, W. Kiessling,
624 B. Kröger, A. McGowan, T. Olszewski, P. Novack-Gottshall, M. Patzkowsky, M.
Uhen, L. Villier, and P. Wager. This work was supported by a NASA
626 Exobiology grant (NNX10AQ446) to A. Miller and M. Foote. I declare no
conflicts of interest. This is Paleobiology Database publication XXX.

628 References

- Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
630 to the mode of stabilization on the substrate. *Palaeogeography,*
Palaeoclimatology, Palaeoecology 21:209–226.
- Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal
632 Groups. *Science* 329:1191–1194.
- , 2014. A simple Bayesian method of inferring extinction. *Paleobiology*
634 40:584–607.
- Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of
636 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.
638 Inouye, 2015. Unveiling the species-rank abundance distribution by
generalizing the Good-Turing sample coverage theory. *Ecology* 96:1189–1201.
640
- Cooper, W. S., 1984. Expected time to extinction and the concept of
642 fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- Crampton, J. S., R. A. Cooper, P. M. Sadler, and M. Foote, 2016.
644 Greenhouseicehouse transition in the Late Ordovician marks a step change in
extinction regime in the marine plankton. *Proceedings of the National*
646 *Academy of Sciences P.* 201519092.
- Edwards, D. and U. Fanning, 1985. Evolution and environment in the late
648 Silurian–early Devonian: the rise of pteridophytes. *Philosophical Transactions*
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
650 birth and death (in macroevolutionary birth-death models). *Biology letters*
652 8:139–42.

- Felsenstein, J., 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- 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.
- 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.
- Fitzjohn, R. G., 2010. Quantitative Traits and Diversification. *Systematic Biology* Pp. 1–15.
- Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology* 14:258–271.
- , 2000a. Origination and extinction components of taxonomic diversity: general problems. *Paleobiology* 26:74–102.
- , 2000b. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.
- , 2014. Environmental controls on geographic range size in marine animal genera. *Paleobiology* 40:440–458.
- Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine animal genera. *Paleobiology* 39:171–192.
- Foote, M. and D. Raup, 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 22:121–140.
- Fritz, S. A., J. Schnitzler, J. T. Eronen, C. Hof, K. Böhning-Gaese, and C. H.

676 Graham, 2013. Diversity in time and space: wanted dead and alive. *Trends in Ecology & Evolution* 28:509–16.

678 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.

680 Gelman, A. and J. Hill, 2007. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY.

682 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.

684 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.

686 Harnik, P. G., 2011. Direct and indirect effects of biological factors on extinction risk in fossil bivalves. *Proceedings of the National Academy of Sciences of the United States of America* 108:13594–9.

690 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.

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.

696 Heim, N. A. and S. E. Peters, 2011. Regional environmental breadth predicts geographic range and longevity in fossil marine genera. *PloS one* 6:e18946.

698

- 700 Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL
<http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- 702 Hoehn, K. B., P. G. Harnik, and V. L. Roth, 2015. A framework for detecting
 natural selection on traits above the species level. *Methods in Ecology and*
 704 *Evolution* .
- Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively
 706 Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine*
Learning Research 15:1351–1381.
- 708 Hopkins, M. J., C. Simpson, and W. Kiessling, 2014. Differential niche dynamics
 among major marine invertebrate clades. *Ecology Letters* 17:314–323.
- 710 Hunt, G. and D. L. Rabosky, 2014. Phenotypic Evolution in Fossil Species:
 Pattern and Process. *Annual Review of Earth and Planetary Sciences*
 712 42:421–441.
- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
 714 comment on "On the heritability of geographic range sizes". *American*
Naturalist 166:129–135.
- 716 Jablonski, D., 1986. Background and mass extinctions: the alternation of
 macroevolutionary regimes. *Science* 231:129–133.
- 718 ———, 1987. Heritability at the species level: analysis of geographic ranges of
 cretaceous mollusks. *Science* 238:360–363.
- 720 ———, 2007. Scale and hierarchy in macroevolution. *Palaeontology* 50:87–109.
- , 2008. Species Selection: Theory and Data. *Annual Review of Ecology,*
 722 *Evolution, and Systematics* 39:501–524.
- Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and

- 724 living molluscs. *Proceedings. Biological sciences / The Royal Society*
270:401–6.
- 726 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
728 and reef evolution: Insights from oxygen isotopes in apatite. *Earth and*
Planetary Science Letters 284:599–609.
- 730 Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
732 benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology*
33:414–434.
- 734 Klein, J. P. and M. L. Moeschberger, 2003. *Survival Analysis: Techniques for*
Censored and Truncated Data. 2nd ed. Springer, New York.
- 736 Kleinbaum, D. G. and M. Klein, 2005. *Survival analysis: a self-learning text*. 2
ed. Springer, New York, NY.
- 738 Lewandowski, D., D. Kurowicka, and H. Joe, 2009. Generating random
correlation matrices based on vines and extended onion method. *Journal of*
740 *Multivariate Analysis* 100:1989–2001.
- Liow, L. H., 2004. A test of Simpson’s ”rule of the survival of the relatively
742 unspecialized” using fossil crinoids. *The American naturalist* 164:431–43.
- , 2007. Does versatility as measured by geographic range, bathymetric
744 range and morphological variability contribute to taxon longevity? *Global*
Ecology and Biogeography 16:117–128.
- 746 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
748 mammals. *Proceedings of the National Academy of Sciences* 105:6097.

- Liow, L. H. and J. D. Nichols, 2010. Estimating rates and probabilities of
750 origination and extinction using taxonomic occurrence data:
Capture-mark-recapture (CMR) approaches. *Quantitative Methods in*
752 *Paleobiology* Pp. 81–94.
- Liow, L. H., T. B. Quental, and C. R. Marshall, 2010. When can decreasing
754 diversification rates be detected with molecular phylogenies and the fossil
record? *Systematic biology* 59:646–59.
- 756 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.
- 758 Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:
the kinetics of mass extinction and origination. *Science* 326:1106–9.
- 760 Munnecke, A., M. Calner, D. A. T. Harper, and T. Servais, 2010. Ordovician
and Silurian sea-water chemistry, sea level, and climate: A synopsis.
762 *Palaeogeography, Palaeoclimatology, Palaeoecology* 296:389–413.
- Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range
764 predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*
39:360–372.
- 766 ———, 2015. Interdependence of specialization and biodiversity in Phanerozoic
marine invertebrates. *Nature communications* 6:6602.
- 768 Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental
than evolvability. *PloS one* 7:e38025.
- 770 Payne, J. L. and S. Finnegan, 2007. The effect of geographic range on
extinction risk during background and mass extinction. *Proceedings of the*
772 *National Academy of Sciences* 104:10506–11.
- Payne, J. L., N. A. Heim, M. L. Knope, and C. R. McClain, 2014. Metabolic

774 dominance of bivalves predates brachiopod diversity decline by more than 150
million years. *Proceedings of the Royal Society B* 281:20133122.

776 Peters, S. E., 2008. Environmental determinants of extinction selectivity in the
fossil record. *Nature* 454:626–9.

778 Quental, T. B. and C. R. Marshall, 2009. Extinction during evolutionary
radiations: reconciling the fossil record with molecular phylogenies. *Evolution*
780 63:3158–67.

———, 2013. How the Red Queen Drives Terrestrial Mammals to Extinction.
782 *Science* 341:290–292.

Rabosky, D. L., 2010. Extinction rates should not be estimated from molecular
784 phylogenies. *Evolution* 64:1816–24.

Rabosky, D. L. and A. R. McCune, 2010. Reinventing species selection with
786 molecular phylogenies. *Trends in ecology & evolution* 25:68–74.

Rabosky, D. L., F. Santini, J. Eastman, S. a. Smith, B. Sidlauskas, J. Chang,
788 and M. E. Alfaro, 2013. Rates of speciation and morphological evolution are
correlated across the largest vertebrate radiation. *Nature Communications*
790 4:1–8.

Raia, P., F. Carotenuto, F. Passaro, D. Fulgione, and M. Fortelius, 2012a.
792 Ecological specialization in fossil mammals explains Cope’s rule. *The
American naturalist* 179:328–37.

794 Raia, P., F. Carotenuto, F. Passaro, P. Piras, D. Fulgione, L. Werdelin,
J. Saarinen, and M. Fortelius, 2012b. Rapid action in the Palaeogene, the
796 relationship between phenotypic and taxonomic diversification in Cenozoic
mammals. *Proceedings of the Royal Society B: Biological Sciences*
798 280:20122244–20122244.

- Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen's Law.
800 Paleobiology 1:82–96.
- , 1978. Cohort Analysis of generic survivorship. Paleobiology 4:1–15.
- 802 ———, 1991. Extinction: Bad Genes or Bad Luck? Norton, New York.
- , 1994. The role of extinction in evolution. Proceedings of the National
804 Academy of Sciences 91:6758–6763.
- Rosindell, J., L. J. Harmon, and R. S. Etienne, 2015. Unifying ecology and
806 macroevolution with individual-based theory. Ecology letters 18:472–82.
- Royle, J. A. and R. M. Dorazio, 2008. Hierarchical modeling and inference in
808 ecology: the analysis of data from populations, metapopulations and
communities. Elsevier, London.
- 810 Sepkoski, J. J., 1975. Stratigraphic biases in the analysis of taxonomic
survivorship. Paleobiology 1:343–355.
- 812 ———, 1981. A factor analytic description of the Phanerozoic marine fossil
record. Paleobiology 7:36–53.
- 814 Sheehan, P., 2001. The late Ordovician mass extinction. Annual Review of
Earth and Planetary Sciences 29:331–364.
- 816 Simpson, C., 2006. Levels of selection and large-scale morphological trends.
Ph.D. thesis, University of Chicago.
- 818 Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine
bivalve extinction over the post-Paleozoic. Paleobiology 35:631–647.
- 820 Simpson, C., W. Kiessling, H. Mewis, R. C. Baron-Szabo, and J. Müller, 2011.
Evolutionary diversification of reef corals: a comparison of the molecular and
822 fossil records. Evolution 65:3274–84.

- Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University
824 Press, New York.
- , 1953. *The Major Features of Evolution*. Columbia University Press,
826 New York.
- Slater, G., L. Harmon, and M. Alfaro, 2012. Integrating fossils with molecular
828 phylogenies improves inference of trait evolution. *Evolution* .
- Slater, G. J., 2013. Phylogenetic evidence for a shift in the mode of mammalian
830 body size evolution at the Cretaceous-Palaeogene boundary. *Methods in
Ecology and Evolution* 4:734–744.
- , 2015. Iterative adaptive radiations of fossil canids show no evidence for
832 diversity-dependent trait evolution. *Proceedings of the National Academy of
Sciences* 112:4897–4902.
834
- Smithson, M. and J. Verkuilen, 2006. A better lemon squeezer?
836 Maximum-likelihood regression with beta-distributed dependent variables.
Psychological methods 11:54–71.
- Smits, P. D., 2015. Expected time-invariant effects of biological traits on
838 mammal species duration. *Proceedings of the National Academy of Sciences*
840 112:13015–13020.
- Solow, A. and W. Smith, 1997. On fossil preservation and the stratigraphic
842 ranges of taxa. *Paleobiology* 23:271–277.
- Stadler, T., 2011. Mammalian phylogeny reveals recent diversification rate
844 shifts. *Proceedings of the National Academy of Sciences* Pp. 1–6.
- , 2013. Recovering speciation and extinction dynamics based on
846 phylogenies. *Journal of evolutionary biology* 26:1203–1219.

- 848 Stadler, T. and F. Bokma, 2013. Estimating speciation and extinction rates for
phylogenies of higher taxa. *Systematic biology* 62:220–30.
- Stan Development Team, 2014a. Stan: A c++ library for probability and
850 sampling, version 2.5.0. URL <http://mc-stan.org/>.
- , 2014b. Stan Modeling Language Users Guide and Reference Manual,
852 Version 2.5.0. URL <http://mc-stan.org/>.
- Stanley, S. M., 1975. A theory of evolution above the species level. *Proc. Natl.*
854 *Acad. Sci. USA* 72:646.
- , 1979. *Macroevolution: pattern and process*. W. H. Freeman, San
856 Francisco.
- Tomiya, S., 2013. Body Size and Extinction Risk in Terrestrial Mammals Above
858 the Species Level. *The American Naturalist* 182:196–214.
- Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- 860 ———, 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.
862 Sidor, C. A. E. Strömberg, and G. P. Wilson, 2013. Bivalve network reveals
latitudinal selectivity gradient at the end-Cretaceous mass extinction.
864 *Scientific Reports* 3:1790.
- Wagner, P. J. and G. F. Estabrook, 2014. Trait-based diversification shifts
866 reflect differential extinction among fossil taxa. *Proceedings of the National*
Academy of Sciences 111:16419–16424.
- 868 Wagner, P. J. and J. D. Marcot, 2013. Modelling distributions of fossil sampling
rates over time, space and taxa: assessment and implications for
870 macroevolutionary studies. *Methods in Ecology and Evolution* 4:703–713.

- Wang, S. C., 2003. On the continuity of background and mass extinction.
872 Paleobiology 29:455–467.
- Wang, S. C. and C. R. Marshall, 2004. Improved confidence intervals for
874 estimating the position of a mass extinction boundary. Paleobiology 30:5.
- Williams, C., C. H. C. Brunton, and S. J. Carlson, 2007. Treatise on
876 invertebrate paleontology. Part H, Brachiopoda. Geological Society of
 America, Boulder, Colorado.

Table 1: Group-level estimates of the effects of biological traits on brachiopod generic survival, the standard deviation of the between-cohort effects, and estimates of both the effect of sampling δ and the Weibull shape parameter α . μ values are the location parameters of the effects, while τ values are the scale terms describing the variation between cohorts. The mean, standard deviation (SD), 10th, 50th, and 90th quantiles of the posterior 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
Standard deviation	τ^i	intercept	0.51	0.11	0.38	0.50	0.65
	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
Other	δ	sampling	0.90	0.15	0.71	0.90	1.09
	α	“time”	1.36	0.04	1.30	1.36	1.42

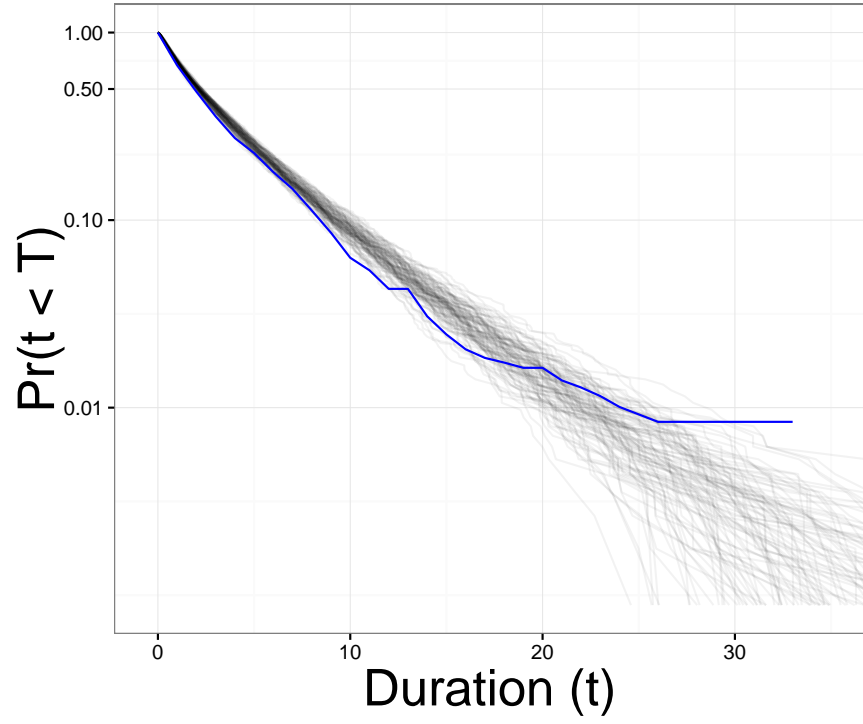


Figure 1: Comparison of the empirical estimate of $S(t)$ (highlighted) versus estimates from 1000 posterior predictive data sets (black). $S(t)$ corresponds to the probability that the age of a genus t is less than the genus' ultimate duration T . The vertical axis is log10 transformed.

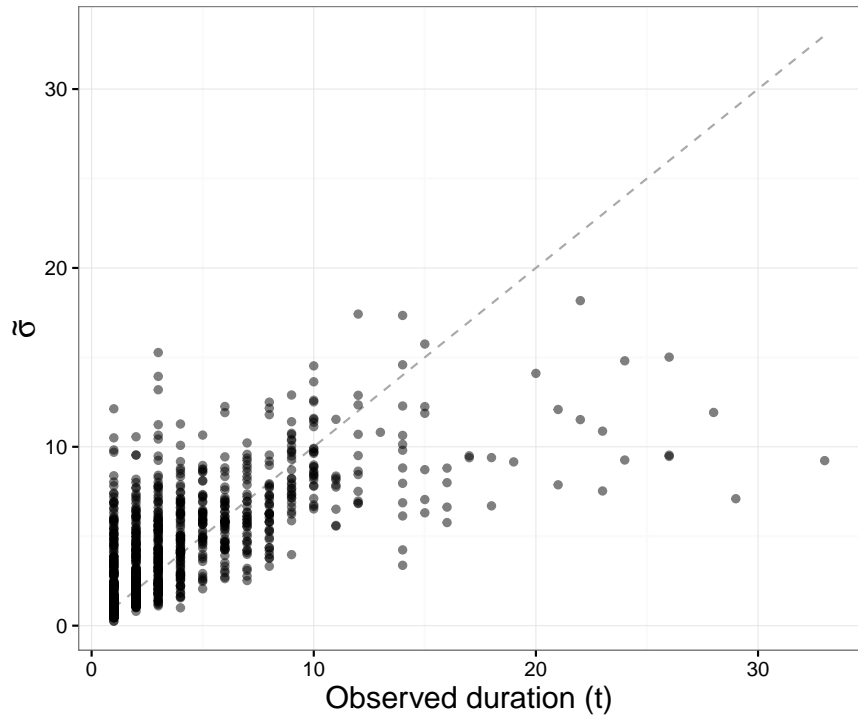


Figure 2: Comparison of all observed genus durations in number of geological stages to the average posterior predictive estimates of $\log(\sigma)$. The dashed, diagonal line corresponds to $x = y$.

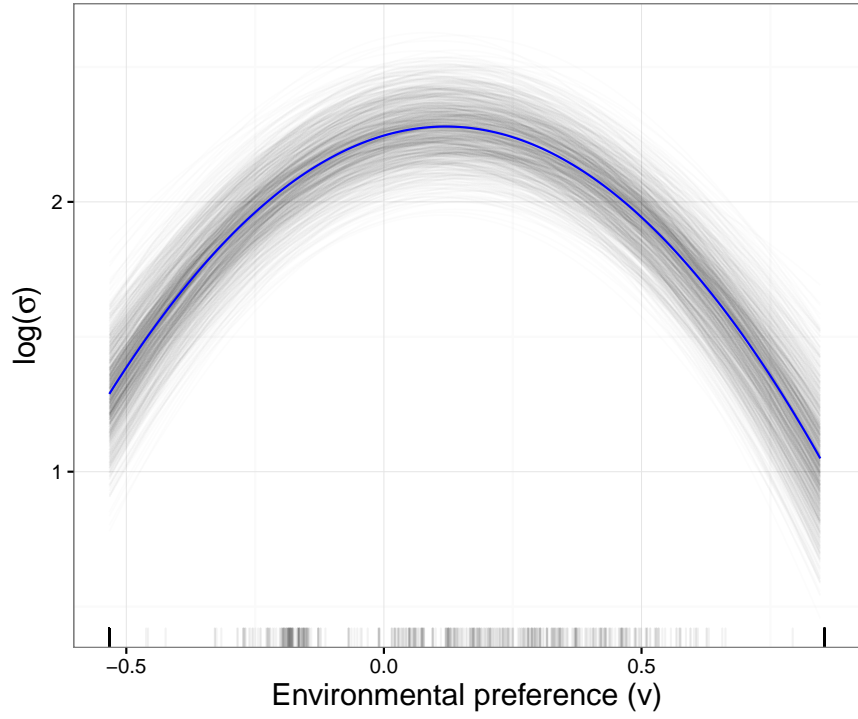


Figure 3: The overall expected relationship between environmental affinity v_i and a $\log(\sigma)$ when $r = 0$ and $m = 0$. Each grey line corresponds to a single draw from the posterior predictive distribution, while the highlighted line corresponds to the median of the posterior predictive distribution. 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.

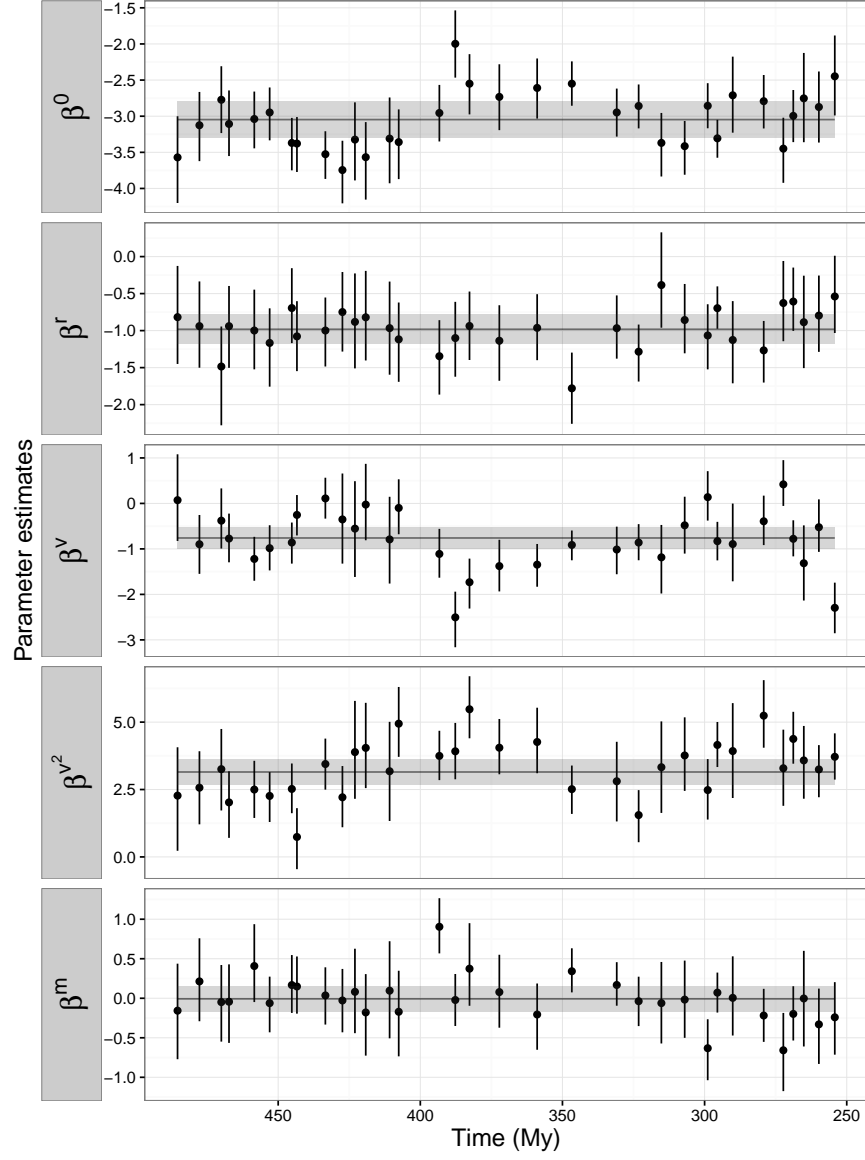


Figure 4: Comparison of cohort-specific estimates of β^0 , the effect of geographic range on extinction risk β^r , the effect of environmental preference β^v and β^{v^2} , and body size β^m . Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. Points are 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).

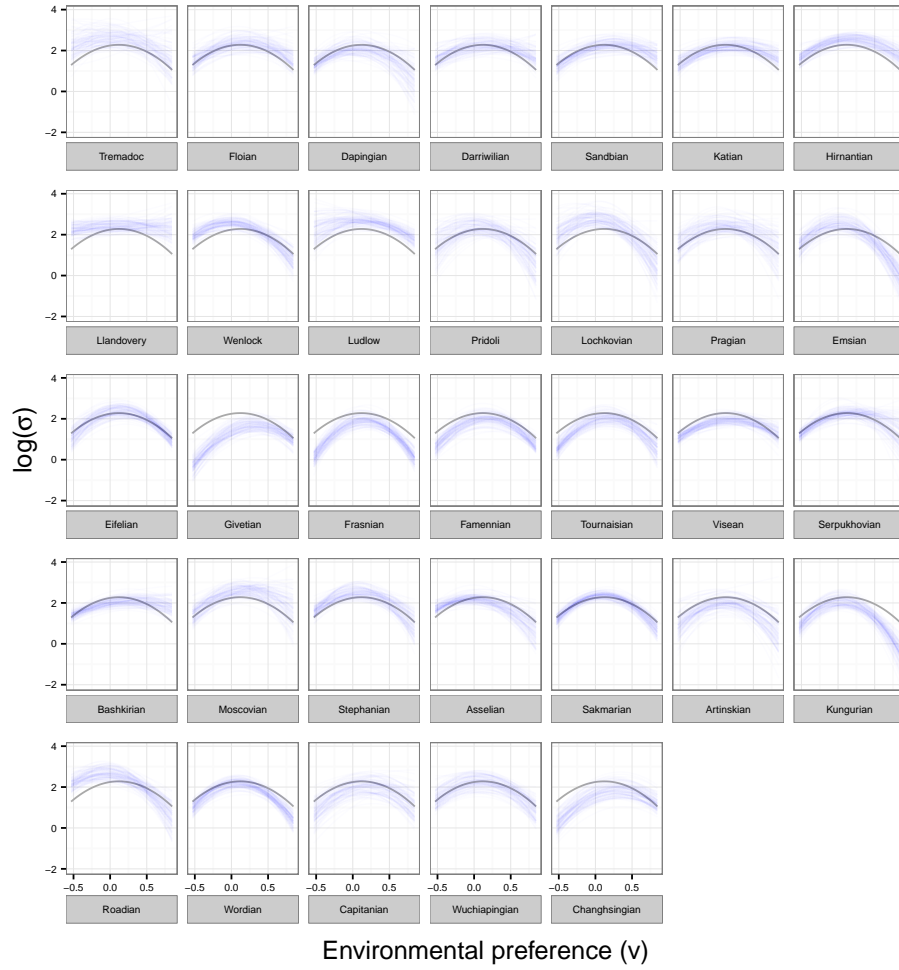


Figure 5: Comparison of origination cohort-specific (posterior predictive) estimates of the effect of environmental preference on $\log(\sigma)$ to the mean overall estimate of the effect of environmental preference. Cohort-specific estimates are from 100 posterior predictive simulations across the range 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. Facet names correspond to the name of the stage in which that cohort originated.

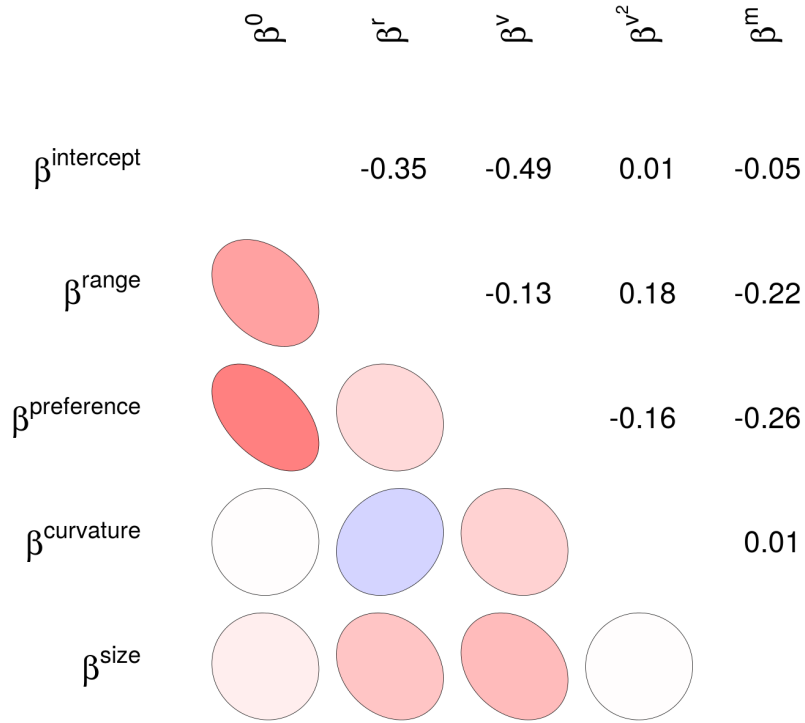


Figure 6: Mixed graphical and numerical representation of the correlation matrix Ω of variation in cohort-specific covariate estimates. These correlations are between the estimates of the cohort-level effects of covariates, along with intercept/baseline extinction risk. The median estimates of the correlations are presented numerically (upper-triangle) and as idealized ellipses representing that much correlation (lower-triangle).