Title: How macroecology affects macroevolution: the interplay

between extinction intensity and trait-dependent extinction in

brachiopods.

Running title: Trait-dependent extinction in brachiopods

Author: Peter D Smits, psmits@uchicago.edu, Committee on Evolutionary

Biology, University of Chicago, IL, USA.

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 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. 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. 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

Extinction is one half of the diversification process (**???**), second only to speciation or origination in shaping changes to diversity; it 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 trait (**????**). Species duration is a measure of species fitness CITATION, and trait-associated

1

differences in fitness is the hallmark of (species) selection CITATION.

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

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

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

The analysis of taxon durations, or time from origination to extinction, falls

under the purview of survival analysis, a field of applied statistics commonly used in health care and engineering (**?**) but has a long history in paleontology (**??????**). I adopt a hierarchical Bayesian modeling approach (**??**) in order to unify the previously distinct dynamic and cohort paleontological survival approaches (**?????????**).

In order to test for an association between extinction intensity and extinction selectivity, both extinction rate and trait-based differences in extinction rate need to be estimated. Previous work has approached this problem by estimating the extinction intensity and selectivity at different time points or for different origination cohorts independently and then comparing those estimates CITATION. I find this approach problematic 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 represent the greatest compromise across all available data CITATIONS. Second, by treating each time point or cohort as independent any and all post-hoc analyses are at risk of false positive results because of multiple comparisons CITATIONS. Third, post-hoc analysis of correlations between maximum likelihood estimates is much less satisfying or coherent than including those correlations as estimates of the initial model, something that is not possible when each time point or cohort is considered independent CITATION.

The hierarchical Bayesian approach used here involves developing a single model to estimate all of time points or cohorts simultaneously while also estimating the group averages for all parameter CITATION. By averaging across groups, the analytical advantages of specific prior choices can induce the partial pooling and regularization of parameter estimates which smooths over groups with small sample sizes or weakly estimated effects. Additionally, individual estimates are much more comparable, not only between time points or cohorts, but also

3

relative to the groups average CITATION. This approach is conceptually and analytically similar to mixed-effects modeling but the Bayesian framework allows the use of strongly regularizing priors in order to better constrain parameter estimates.

While estimation of both trait-dependent speciation and extinction rates from phylogenies of extant taxa has grown dramatically (????????), there are two major ways to estimate trait-dependent extinction: analysis of phylogenies, and analysis of the fossil record. These two directions, phylogenetic comparative and paleobiological, are complementary and intertwined in the field of macroevolution (???). In the case of extinction, analysis of the fossil record has the distinct advantage over phylogenies of only extant taxa because extinction is observable; this means that extinction rate is possible to estimate (???). The approach used here is thus complementary to the analysis of trait-dependent extinction based on a phylogeny.

1.1 Factors affecting brachiopod survival

Conceptually, taxon survival can be considered an aspect of "taxon fitness" (??). Traits associated with taxon survival are thus examples of species (or higher-level) selection, as differences in survival are analogous to differences in fitness. The traits analyzed here are all examples of emergent and aggregate traits (??); specifically I analyze genus-level traits. Emergent traits are those which are not measurable at a lower level (e.g. species versus individual organism) such 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 members of a lower level.

Geographic range is widely considered the most important biological trait for

4

estimating differences in extinction risk at nearly all times, with large geographic range associated with low extinction risk (**???????**). This stands to reason even if extinction is completely at random; a taxon with an unrestricted range is less likely to go extinct at random than a taxon with a restricted range.

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

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

dichotomy observed in ancient marine systems (**??**). Given these findings, I would hypothesize that as extinction risk increases, the extinction risk associated with open-ocean environments should generally increase.

Because environmental preference is defined here as the continuum between occurring exclusively in open-ocean environments versus epicontinental environments, intermediate values are considered "generalists" in the sense that they favor neither end member. A long-standing hypothesis is that generalists or unspecialized taxa will have greater survival than specialists (**???????**). 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 (**??**). Body size is a proxy for metabolic activity and other correlated life history traits (**?**). **?** 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 (**?**). 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 (**??????**). 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

6

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 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.

2   Materials and Methods

2.1   Fossil occurrence information

The brachiopod dataset analyzed here was sourced from the Paleobiology

Database (http://www.paleodb.org) which was then filtered based on

taxonomic (Rhychonelliformea: Rhynchonellata, Chileata, Obolellida,

Kutorginida, Strophomenida, Spiriferida), temporal (post-Cambrian Paleozoic),

stratigraphic, and other occurrence information used in this analysis. Analyzed

occurrences were restricted to those with paleolatitude and paleolongitude

coordinates, assignment to either epicontinental or open-ocean environment,

and belonging to a genus present in the body size dataset (?). Epicontinental

versus open-ocean assignments for each fossil occurrence are partially based on

those from ?, with additional occurrences assigned similarly (Miller and Foote,

personal communication). These filtering criteria are very similar to those from

? with an additional constraint of being present in the body size data set from

?. In total, there 1130 were genera included in the dataset.

Fossil occurrences were analyzed at the genus level which is common for

paleobiological, macroevolutionary and macroecological studies of marine

invertebrates (**??????????**). While species diversity dynamics are frequently of

much greater interest than those of higher taxa (though see **??**), the nature of

the fossil record makes accurate, precise, and consistent taxonomic assignments

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

confidence and accuracy in the data analyzed here.

Genus duration was calculated as the number of geologic stages from first

appearance to last appearance, inclusive. Durations were based on geologic

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

instead fossils are assigned to a geological interval which represents some

temporal range. In this analysis, stages are effectively irreducible temporal

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

censored at the Changhsingian; genera with a duration of only one stage were

left censored (**?**). How the likelihood of censored observations is calculated is

detailed in section 2.2.

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

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

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

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 km^2. The map

projection and regular lattice were made using shape files from

http://www.naturalearthdata.com/ and the `raster` package for R (**?**). For each

8

stage, the total number of occupied grid cells was calculated. Then, for each temporal bin, the relative occurrence probability of the observed taxa was calculated using the JADE method developed by **?**. 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. This correction is critical when comparing occupancies from different times with different geographic sampling. Finally, for each genus, the mean relative occurrence probability was calculated as the average of that genus' occurrence probabilities for all stages it was sampled to yield relative occupancy, my proxy for geographic range.

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 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 expressed.

$$
\begin{aligned}
p\left(\theta_i'|e_i', E_i'\right) &\propto \text{Beta}(e_i', E_i' - e_i')\text{Beta}(1,1) \\
&= \text{Beta}(e_i' + 1, E_i' - e_i' + 1),
\end{aligned}
\tag{1}
$$

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

Body size, measured as shell length, was sourced directly from **?**. These measurements were made from brachiopod taxa figured in the *Treatise on Invertebrate Paleontology* (**?**).

The sampling probability for individual taxa was calculated using the standard gap statistic (**??**). The gap statistic is calculated as the number of stages in which the taxon was sampled minus two divided by the duration of the taxon minus two. Subtracting two from both the numerator and denominator is

9

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

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 duration less than 3. Imputation is the "filling in" of missing observations based on the observed values (??). This is fairly straight forward in a Bayesian framework because both covariates and parameters are considered random variables, meaning that the missing values of 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 natural-log transformed; both of these transformations make these variables defined for the entire real line. Sampling probability was transformed as $(s(n-1) + 0.5)/n$ where n is the sample size as recommended by ?; this serves to slightly shrink the range of the data so that 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 standard deviation, which follows ?. This standardization means that the associated regression coefficients are comparable as the expected 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.

2.2 Details of model

Hierarchical modelling is a statistical approach which explicitly takes into account the structure of the observed data in order to model both the within

10

and between group variance (**??**). The 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 (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 distributions and are also estimated like

the other parameters of interest (e.g. covariate effects) (**?**). The subsequent

estimates are partially pooled together, where parameters from groups with

large samples or effects remain large while those of groups with small samples or

effects are pulled towards the overall group mean. All covariate effects

(regression coefficients), as well as the intercept term (baseline extinction risk),

were allowed to vary by group (origination cohort). The covariance between

covariate effects was also modeled.

Genus durations were assumed to follow a Weibull distribution which allows for

age-dependent extinction (**?**): $y \sim \text{Weibull}(\alpha, \sigma)$. 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 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 extinction risk decreases with age. Note that the

Weibull distribution is equivalent to the exponential distribution when $\alpha = 1$.

In the case of the right- and left-censored observations mentioned above, the

probability of those observations has a different calculation (**?**). For

right-censored observations, the likelihood is calculated $p(y|\theta) = 1 - F(y) = S(y)$

where $F(y)$ is the cumulative distribution function. In contrast, the likelihood of

a left-censored observation is calculated $p(y|\theta) = F(y)$.

The scale parameter σ was modeled as a regression following **?** with both

varying intercept and varying slopes and the effect of sampling; this is expressed

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

where i indexes across all observations where $i = 1, \ldots, n$ where n is the total number of observations, $j[i]$ is the cohort membership of the ith observation where $j = 1, \ldots, J$ where J is the total number of cohorts, X is a $N \times D$ matrix 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 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 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.

The elements of μ were given independent normally distributed priors. The effects of geographic range size and the breadth of environmental preference were given informative priors reflecting the previous findings while the others were given weakly informative favoring no effect. The correlation matrix Ω was given an LKJ distributed prior (?) that slightly favors an identity matrix as

12

recommended by **?**. These priors are defined

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

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 Extinction (**?**).

2.3 Imputation of sampling probability

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 have an observed gap statistic. The gap statistic for the remaining 591 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 and 1, I chose to model it as a Beta regression with matrix W being a $N \times (D - 1)$ matrix of covariates (i.e. geographic range size, environmental preference, body size) as predictors of sampling; this assumes that the sampling value for all taxa come from the same distribution. Importantly, I make no assumptions of causal structure.

Predicting sampling probability using the other covariate that are then included

13

in the model of duration is acceptable and appropriate in the case of imputation where the sample goal is accurate prediction (**??**). Not including these covariates can lead to biased estimates of the imputed variable; if the covariates themselves are related, not including them will bias this correlation towards zero which then leads to improper imputation and inference (**?**).

The Beta regression is defined

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

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

$$\phi = \frac{\alpha}{\alpha + \beta} \tag{5}$$

and total count parameter

$$\lambda = \alpha + \beta \tag{6}$$

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

$$\gamma \sim \mathcal{N}(0, 1)$$
$$\delta \sim \mathcal{N}(0, 1) \tag{7}$$
$$\lambda \sim \text{Pareto}(0.1, 1.5).$$

14

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 (**?**) as implemented in the probabilistic programming language Stan (**?**). 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 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 convergence. Convergence means that the chains are approximately stationary and the samples are well mixed (**?**).

Model adequacy was evaluated using a couple of posterior predictive checks. 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 should be similar to the original data and systematic differences between the simulations and observations indicate weaknesses of the model fit (**?**). For both approaches used here, each posterior predictive datasets were generated from a unique draw from the posterior distribution of each parameter. The two posterior predictive checks used in this analysis are a comparison of a non-parametric estimate of the survival function $S(t)$ from the empirical dataset to the non-parametric estimates of $S(t)$ from the 100 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 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.

15

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 ($\mathrm{Pr}\mathcal{N}(\mu^r, \tau^r) > 0)$). 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

16

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 $(\Pr(\mathcal{N}(\mu^{v^2}, \tau^{v^2}) < 0)$.

The cohort-specific estimates of all the regression coefficients demonstrate a lot

398 of between cohort variance, with no obvious trends. As indicated in Table 1 and detectable visually (Fig. 4), the between-cohort estimates for β^0, β^r, and β^m all

400 have much lower variance than the between-cohort estimates of both β^v and β^{v^2}.

402 While most cohort-specific estimates are very similar to the overall cohort-level estimate, there are a few notable excursions away from the overall mean (Fig.

404 4). There are simultaneous excursions in both β^0 and β^v for cohorts originating in the Givetian (387-382 My) and Frasnian (382-372 My) stages; both of which

406 directly precede the late Devonian mass extinction event at the Frasnian/Famennian boundary. These cohorts are marked by both a high

408 extinction intensity and an increase in expected duration for taxa favoring epicontinental environments over open-ocean ones; this is consistent with the

410 results of **?**.

Cohorts originating from the Silurian through the Early Devonian have a

412 slightly lower extinction intensity than the overall mean; these cohorts are those originating in the Llandovery (443-443 My) through the Emsian (407-393 My).

414 This is also a time period is also when there is the lowest overall probability that epicontinental favoring taxa are expected to have greater duration than

416 open-ocean favoring taxa. Both the Silurian and Devonian periods are notable for having been periods with a mostly "hothouse" climate, with no polar icecaps

418 and a high sea-level (**???**).

17

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

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

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 negatively correlated; this means that for cohorts experiencing a lower extinction intensity (β^0 decreases), the magnitude of the effect of geographic range is expected to decrease as well, and *vice versa*; this is in contrast to the observation made by ? with regards to late Cretaceous bivalves.

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

18

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

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

Comparison of observed values of sampling, as measured by the gap statistic, to random draws from the posterior estimates of the imputed sampling values indicate that they are very similar (Fig. 7. This result is very encouraging as this is the ultimate goal of multiple imputation: to fill in missing data with values similar to the observed while taking into account the randomness of that variable (??). The estimates of δ are based on 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 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 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.

While the effect of sampling appears large compared to the other regression coefficients, this is only because sampling was not standardized like the other covariates. 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 then 0.642 ($\pm$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 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 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) with a 100% probability of being greater than 1. This result is not consistent with the Law of Constant Extinction (**?**) and is instead 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 empirical results (**???**) and (ironically) with a recently proposed Red Queen-like model of evolution (**?**). This results, however, is not consistent with other empirical results (**??**) and could potentially be caused by the minimum resolution of the fossil record (**?**). 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 of taxon age on extinction risk.

4   Discussion

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 affect extinction risk was geographic range while traits that had previously been beneficial had no effect (**?**). This observation raises two linked questions: how does the effect of geographic range change with changing 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

20

favoring epicontinental environments of open-ocean environments is expected to be increase; this is consistent with the results of **?**. There 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 then the between-cohort variance of the effect of environmental preference which may underlie the lack of correlation between these two effects.

Additionally, the lower between-cohort variance in the effect of geographic range versus that higher between-cohort variance implies that for cohorts with a 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 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. 3), though the curvature of the relationship varies from rather shallow to very peaked (Fig. 5). 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 (**????**); inclusion of these factors in future analyses would potentially improve our understanding of the "survival of the unspecialized" hypothesis (**?**).

?, in their analysis of niche conservatism and substrate lithological preference in marine invertebrates, found that brachiopods were among the least

21

"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 should explore how environmental preference changes over lineage duration in relation to

environmental availability to estimate if the generalists–specialists continuum is actually ternary relationship.

An alternative approach for specifically modeling survival that can take into account imperfect observation than the method used here is the

Cormack-Jolly-Seber (CJS) model (**????**). This model is a type of hidden Markov model with an absorbing state (i.e. extinction). In this model, survival

is defined 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 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: 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 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

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 might increase the duration of genera through self-renewal (??), which would be an example of the difference in biological pattern between species and genera (???). This could lead to a trait appearing to decrease generic level extinction risk by that trait increasing species level origination rate instead of decreasing species level extinction risk.

The model used here could be improved through either increasing the number of analyzed traits, expanding the hierarchical structure of the model to include other major taxonomic groups of interest, and the inclusion of explicit phylogenetic relationships between the taxa in the model as an additional hierarchical effect. An example trait that may be of particular interest is the 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 taxa, taxa that are attached to the substrate are expected to have a greater duration than those that are not (?).

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

With significant updates, it would also be possible to compare the brachiopod record with with Moden groups such as bivalves or gastropods (?), though remembering that the groups may not necessarily share all 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 on extinction such as has been proposed for geographic range (**?**). Additionally, this expanded model would also act as a test of the distinctness of the **?** three-fauna hypothesis in terms of trait-dependent extinction.

Traits like environmental preference or geographic range (**??**) are most likely heritable. Without phylogenetic context, this analysis assumes that differences in extinction risk between taxa are independent of the shared evolutionary history of those taxa (**?**). In contrast, the origination cohorts only capture shared temporal context. For example, if taxon duration is phylogenetically heritable, then closely related taxa may have more similar durations as well as more similar biological traits. Without taking into account phylogenetic similarity the effects of these biological traits would be inflated solely due to inheritance. The inclusion of phylogenetic context as an additional individual-level hierarchical effect, independent of origination cohort, would allow for determining how much of the observed variability is due to shared evolutionary history versus shared temporal context versus actual differences associated with biological traits (**?**).

The combination and integration of the phylogenetic comparative and paleontological approaches requires both sources of data, something which is not possible for this analysis because there is no phylogenetic hypothesis for all Paleozoic 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, the analysis can more fully address the questions of interest in macroevolution (**?????????**).

In summary, patterns of Paleozoic brachiopod survival were analyzed using a fully Bayesian hierarchical survival modelling approach while also eschewing the traditional separation between background and mass extinction. I find that

24

cohort extinction intensity is negatively correlated with both the cohort-specific

effects of geographic range and environmental preference. These results imply

that as extinction intensity increases (β^0) increases, it is expected that both

effects will increase in magnitude. However, the change in effect of

environmental preference is expected to be greater than the change in the effect

of geographic range. Additionally, I find support for greater survival in

environmental generalists over specialists in all origination cohorts analyzed;

this is consistent with the long standing "survival of the unspecialized"

hypothesis (**????????**). The results of this analysis support the conclusion that

for Paleozoic brachiopods, as extinction intensity increases overall extinction

selectivity is expected to increase as well.

Acknowledgements

⁶³⁰ publication XXX. The code necessary for reproducing these results is avaliable as a Zenodo archive DOI 10.5281/zenodo.46928.

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
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 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: 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. Panel 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). The darkness of the ellipse corresponds to the magnitude of the correlation.

Figure 7: 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.