

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

Peter D. Smits^{1,*}

1. University of California – Berkeley, Berkeley, California 94720;

* Corresponding author; e-mail: psmits@berkeley.edu

Manuscript elements: Figure 1, figure 2, table 1, online appendices A and B (including figure A1 and figure A2). Figure 2 is to print in color.

Keywords: species selection, geographic range, environmental preference, paleobiology, Bayesian

Manuscript type: Article.

Prepared using the suggested L^AT_EX template for *Am. Nat.*

Abstract

As average species fitness decreases, do trait-based differences in species' fitness (selection) increase or decrease? Similarly, as taxonomic extinction intensity increases, how do the effects of species traits on taxonomic survival change? Do the selective differences associated with some traits change more than others? 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. My analytical framework eschews the traditional distinction between background and mass extinction, and instead considers extinction intensity as a continuum. I find strong evidence for a non-linear relationship between environmental preference for epicontinental versus open-ocean environments and expected taxon duration, where taxa with intermediate or no environmental preference are expected to have greater durations than taxa which appear exclusively in either environmental end-member. I find evidence for a correlation between average cohort extinction risk and the effect of geographic range on extinction risk for that cohort where I estimate that as extinction intensity increases that the magnitude of the effect of geographic range on extinction risk increases, thus increasing the difference in extinction risk between small and large ranged taxa. I also find evidence a similar relationship between average cohort extinction risk and the effect of the linear aspect of environmental preference on expected duration, so that as expected cohort average extinction intensity increases it is expected that taxa which preferentially appear in epicontinental environments will have a greater expected duration than those taxa which prefer open-ocean environments. These results support the conclusion that, for Paleozoic brachiopods and the effects of geographic range and environmental preference, as extinction

intensity increases the magnitude of selection also increases.

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 (?), and trait-associated differences in fitness is the hallmark of (species) selection (?).

? 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 show unimodal 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 traits and differences in fitness (extinction **selectivity**) (?). 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 develop a statistical model describing the relationship between brachiopod taxon durations and multiple functional taxon traits in order to understand the relationship between extinction intensity and selectivity over time. Trait-dependent differences in extinction risk should be associated with differences in taxon duration (??). 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 (?). 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 (?). 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 (?). 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 (?).

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.

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 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 envi-

ronments 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 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.

Methods

The brachiopod dataset analyzed here was sourced from the Paleobiology Database (<http://www.paleodb.org>) and then filtered to a limited taxonomic selection (Rhynchonelliformea: Rhynchonellata, Chileata, Obolellida, Kutorginida, Strophomenida, Spiriferida). Temporal (post-Cambrian Paleozoic), stratigraphic, and other occurrence information used in this analysis was also downloaded. Analyzed occurrences were restricted to those with paleolatitude and paleolongitude coordinates, have been assigned 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 based on those from previous analyses ??. 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 genera were included in this analysis.

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.

The covariates of duration included in this analysis are geographic range size (r), environmental preference (v, v^2), shell length (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 km². The map projection and regular lattice were made using shape files from <http://www.naturalearthdata.com/> and the raster package for R (?). For each 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.

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

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

Prior to analysis, geographic range was logit transformed and the number of samples 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.

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 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. Taxa that existed for only a single stage were left-censored, which implies that that taxon went extinct at any point between 0 and 1 stages. In contrast to right-censored data, the likelihood of a left-censored observation is calculated from $p(y|\theta) = F(y)$. This censoring strategy improves model

fit greatly over treating these taxa as being fully observed (see Appendix).

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

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

where i indexes across all observations where $i = 1, \dots, n$ where n is the total 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 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 recommended by ?. These priors are defined

$$\begin{aligned}
\mu^0 &\sim \mathcal{N}(0, 5) \\
\mu^r &\sim \mathcal{N}(-1, 1) \\
\mu^v &\sim \mathcal{N}(0, 1) \\
\mu^{v^2} &\sim \mathcal{N}(1, 1) \\
\mu^m &\sim \mathcal{N}(0, 0.5) \\
\delta &\sim \mathcal{N}(1) \\
\tau &\sim C^+(1) \\
\Omega &\sim \text{LKJ}(2).
\end{aligned} \tag{3}$$

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 (?).

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 $N \times (D - 3)$ matrix of covariates (i.e. geographic range size, environmental preference, body size; no interactions) 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 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

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

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 as recommended by ?:

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

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 40,000 steps, split half warm-up and half sampling and thinned to every 20th 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 (?). After the model was fit to the data, 100 datasets were simulated from the posterior predic-

tive distribution of the model. These simulations were used to test for adequacy of model fit as described below.

Survival analysis is complicated by censored observations where the ultimate time of extinction for some taxa could not be fully observed during the study window. Importantly, posterior predictive simulations for these observations must be similarly censored. To accomplish this, posterior predictive simulated durations for right-censored observations were the minimum of its final observed duration and the simulated duration. For left-censored individuals, their simulated duration was pegged at a minimum of one stage with simulated values less than one stage set to one.

Model adequacy was evaluated using a series of posterior predictive checks. Posterior predictive checks are a means for understanding model fit or adequacy. The concept of model adequacy is that if our model is an adequate descriptor of the observed data, then data simulated from the posterior predictive distribution should be similar to the observed given the same covariates, etc. (?). Posterior predictive checks can take many forms but the basic idea is to compare some property of the empirical data to that property estimated from each of the simulated datasets. Additionally, for structured datasets like the one analyzed here, the fit of the model to different parts of the data can be assessed, which in turn can reveal a great deal if the model has good fit to some aspects of data but not others; it is in these scenarios when knowledge about the biology, geology, and paleoenvironment becomes important in order to explain what unmodeled processes might lead to these discrepancies between our data and the model (?).

The types of posterior predictive tests used in this analysis fall into two categories: comparison of observed mean and median genus duration to a distribution of mean and median genus duration estimates from the posterior simulations, comparison of a non-parametric estimate of the survival function from the observed data to estimates of that same survival function from the simulations.

The survival function describes the probability of a taxon persisting given that it has survived up to time t ; this is expressed $P(T > t)$ because T is the true extinction time of the species and t is

some arbitrary time of observation and we are estimating that probability that t is less than T . It is important to note, however, that the survival function does not reflect density of observations unlike e.g. histograms. Instead, this posterior predictive check reflects the model's ability to predict genus survival.

These posterior predictive tests were done for the entire data and for each of the origination cohorts.

Results

I first present the results of the multiple posterior predictive checks for the whole dataset as well as each of the origination cohorts. I next present the parameter posterior estimates and their interpretations. Importantly, the adequacy of model fit associated with each cohort-specific interpretation is presented to demonstrate which results for which I have the greatest confidence.

Comparisons between the observed distribution of durations to the distributions of 100 simulated datasets reveals the relatively good but heterogeneous fit of the model to the data (Fig. 1). The two major aspects of possible misfit that are observable are at durations of 1 and 8-9 stages. In both of these cases, the model slightly over-estimates the number of observations with duration of 1 stage, and slightly under-estimates the number of individuals with a duration of 8-9 stages. The goal of this model is estimating the expected duration of a genus given its covariate information. While the model estimates are not exact, it is possible that our model fits the “middle” of our data well but does poorly towards the tails.

Comparisons between the survival functions estimated from the empirical data and from 100 simulated datasets further expands the picture of model (mis)fit (Fig. 2). The survival curves of the 100 simulated datasets are very similar to the survival function estimated from the empirical data. The major points of potential misfit between the model and the data are underestimated percentage of taxa with duration 1 stage, and an over-estimate of probability of species surviving at least 9-16 stages. Importantly, the major divergence between the observed and estimated

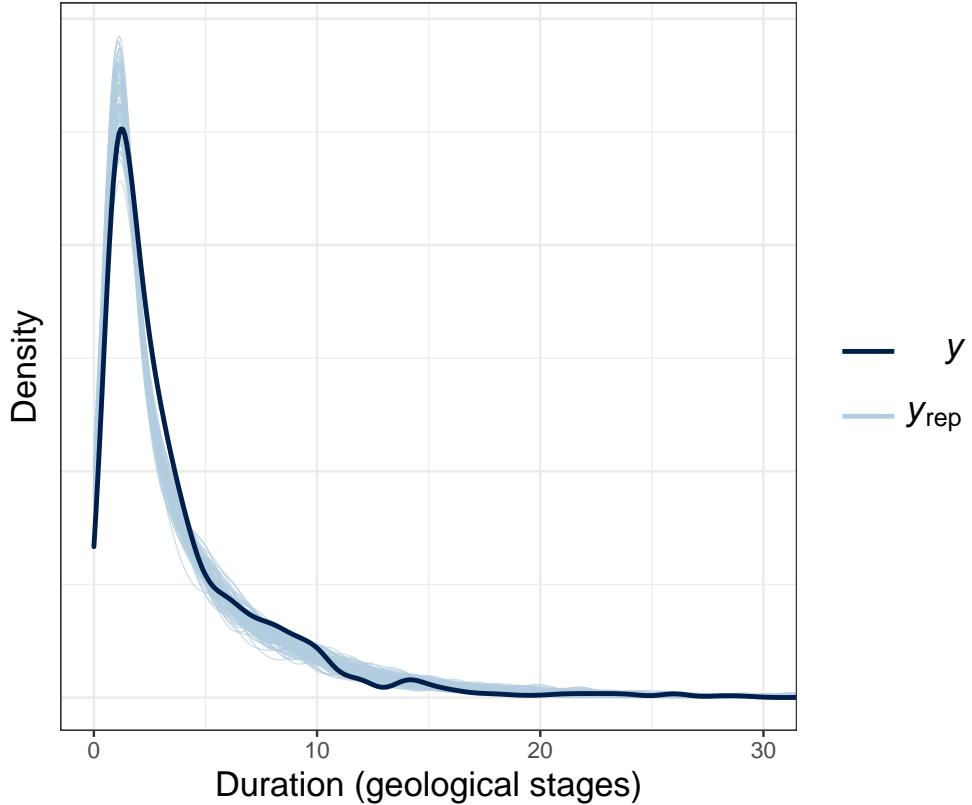


Figure 1: Comparison of the distribution of the observed data (black) to 100 simulated distributions (blue). This is a close-up view of the bulk of the distribution which shows the more subtle aspects of (mis)fit between the data and the model.

applies to taxa with a less than 15% probability of continuing to surviving. Keep in mind, also, that the survival curve as presented does not depict density of observations as in Fig. 1.

In addition to distributional comparisons, model adequacy at the total data level was assessed through comparison of the mean and median of the observed data to those from simulated data sets. While the previous posterior predictive checks have focused on the relatively good but heterogeneous fit of the model to the entire distribution of the data, the fitted model's ability to predict the mean and median of the observed data appears adequate (Fig. 3a, 3b). Because the principle goal of this model is to obtain adequate prediction of how a taxon's expected duration for a given set of ecological covariates, the seemingly adequate fit of our model to mean taxon

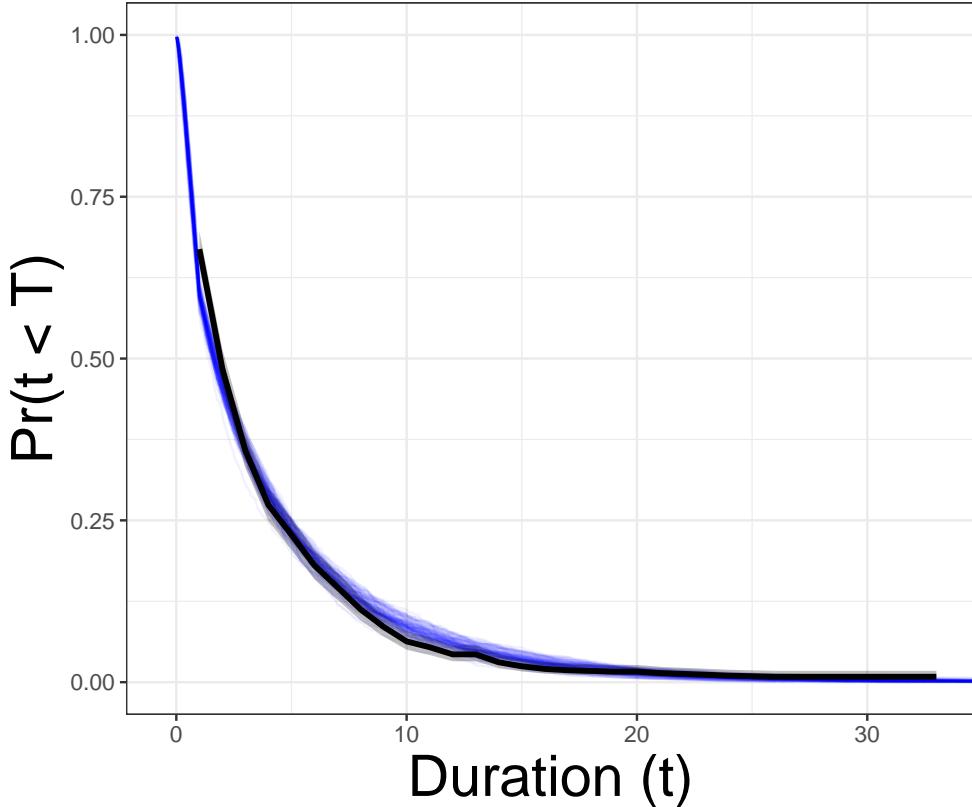


Figure 2: Comparison of the empirical estimate of $S(t)$ (highlighted) versus estimates from 100 posterior predictive data sets (black). $S(t)$ corresponds to the probability that the age of a genus t is less than the genus' ultimate duration T .

duration is reassuring (Fig. 3a). Additionally, given the skewness of the observed taxon durations (Fig. 1, the ability for the model to recapitulate the median observed taxon duration is important and points to the overall good fit of the model to the data.

When considered together, all of the above posterior predictive checks indicate approximately adequate model fit for key questions such as expected taxon duration (Fig. 3a). However, there is obviously heterogeneity in model fit because, while the model can recapitulate some aspects of the observed data (Fig. 3a, 3b), there are obvious discrepancies between the model and the data (Fig. 1, 2). By performing the same posterior predictive tests for each of the origination cohorts, it may be possible to get a better picture of the sources of model misfit.

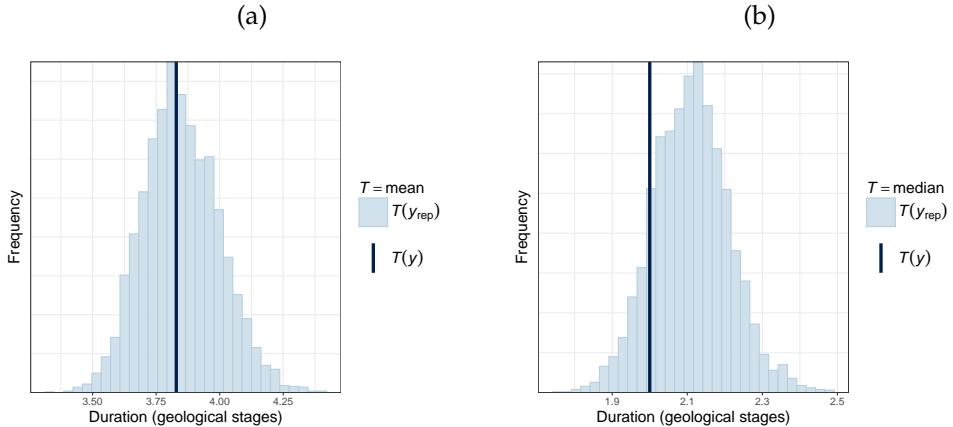


Figure 3: Comparison of the (A) observed mean genus duration (black vertical line) to a distribution of means estimated from 100 simulated datasets (blue), and (B) comparison of the observed median genus duration (black vertical line) to a distribution of medians estimated from 100 simulated datasets (blue). Model fit is evaluated by the similarity between the observed and the estimated, where good fit is demonstrated by the vertical line being “within” the simulated distribution.

When the posterior predictive tests are visualized for each of the origination cohorts, a complex picture of model fit emerges. Comparison between the empirical survival functions estimated for each cohort to those estimated from the simulated datasets reveals the degree of heterogeneity in model fit (Fig. 4, as some origination cohorts appear to be very well fit by the model (e.g. Tremadoc, Darriwilian Wenlock, Ludlow, Lochkovian, etc.) while others are more poorly fit by the model (e.g. Visean, Bashirian, Moscovian, Stephanian, Roadian, etc.). The poverty of model fit to these cohorts may indicate that these origination cohorts are undergoing a different extinction process whose aspects are unmodeled in this analysis. However, for those cohorts where our model recapitulates the empirical survival function indicates approximately adequate fit and that our model may be capturing some of the processes underlying taxon extinction.

For nearly every origination cohort, the model is able to approximately recapitulate the observed mean duration (Fig. 5). The only exceptions to this observed mean is within the tails

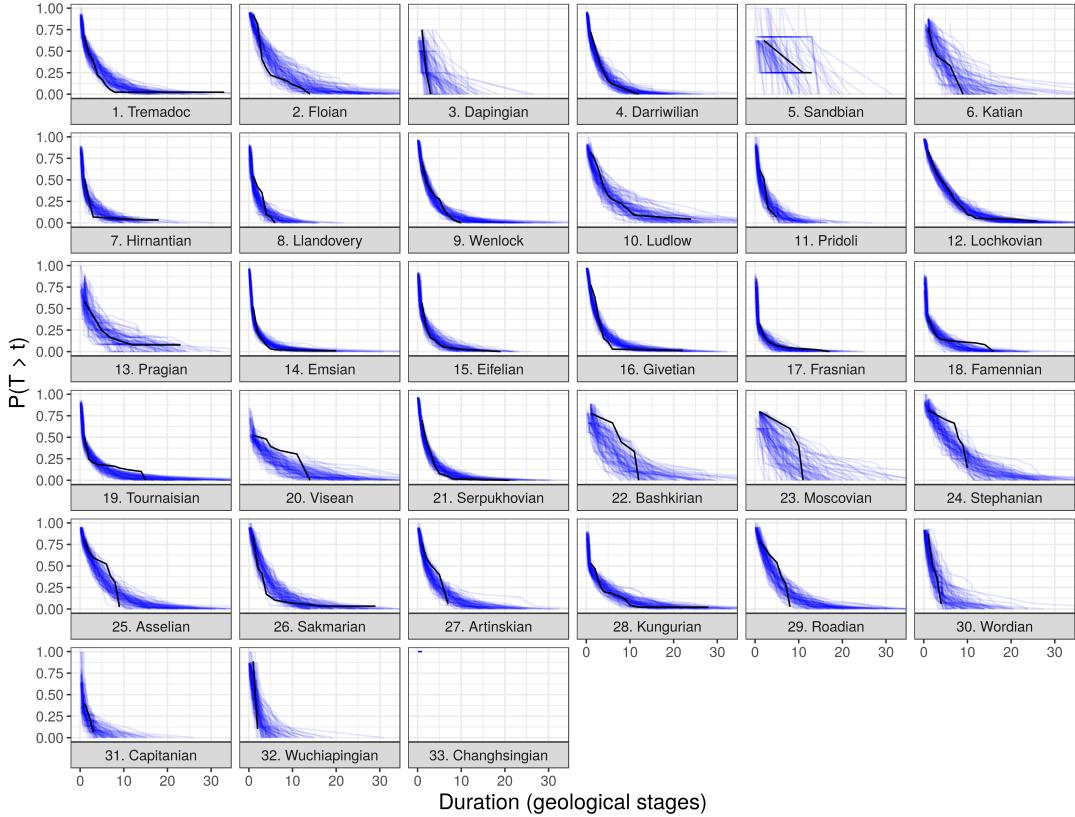


Figure 4: Comparison of the empirical estimate of $S(t)$ (highlighted) versus estimates from 100 posterior predictive data sets (black) for each of the origination cohorts. $S(t)$ corresponds to the probability that the age of a genus t is less than the genus' ultimate duration T . By comparing the fit of the model to the individual cohorts, when and where the model (mis)fits is more observable.

of the posterior predictive distributions (e.g. Moscovian, Sakmarian, Wordian, etc.). In general, however, posterior predictive distribution of the model is able to adequately estimate the mean duration of nearly all origination cohorts.

In comparison, the model has a much more heterogeneous fit to each origination cohort's median taxon duration (Fig. 6). The skewness of the distribution underlying (Fig. 1) means that for some origination cohorts, median duration might be pegged at 1 stage; this means that the posterior predictive distributions for some cohorts can be extremely skewed. The cohorts with notably inadequate poor predictions are the Hirnantian, Priodoli, Emsian, Eifelian, Givetian,

Asselian, Sakmarian, Artinskian, Kungurian, and Roadian. The remaining 13 cohorts, however, have decently adequate fit. These results indicate that our model is very good at recapitulate mean taxon duration (Fig. 3a, 5), and it is capable of esimtaing overall median duration and median duration of most origination cohorts (Fig. 3b, 6).

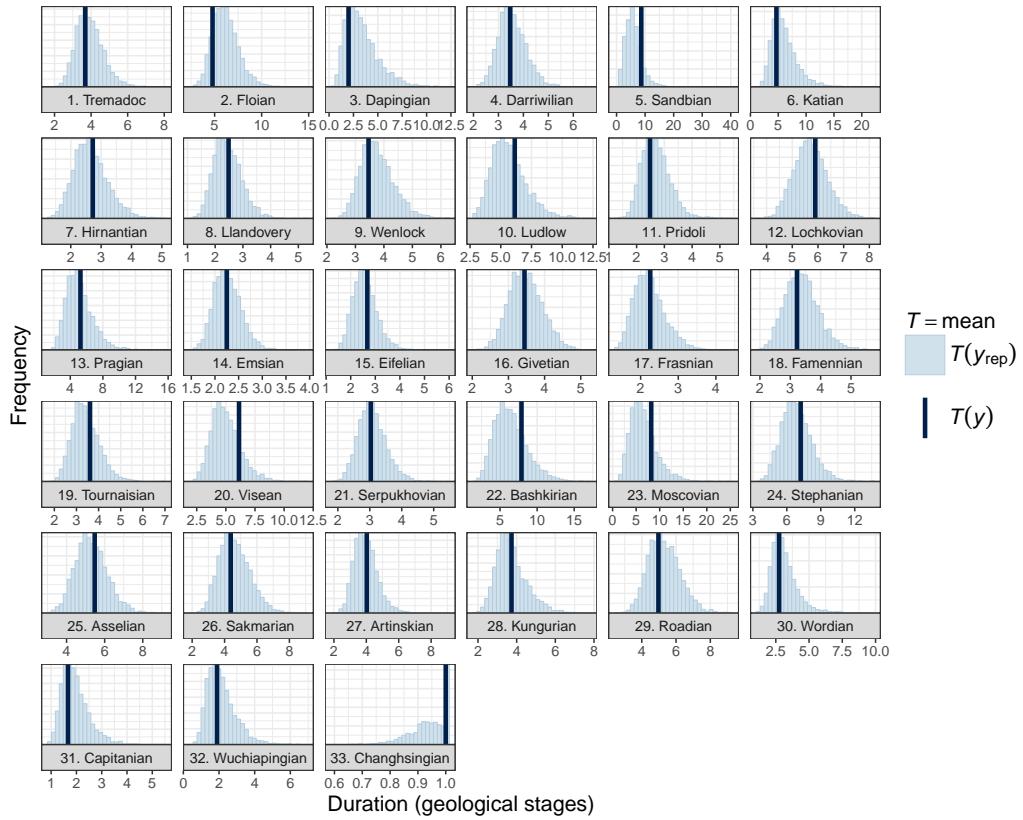


Figure 5: Comparison of the observed mean genus duration (black vertical line) to a distribution of means estimated from 100 simulated datasets (blue) for each of the origination cohorts. Model fit is evaluated by the similarity between the observed and the estimated, where good fit is demonstrated by the vertical line being “within” the simulated distribution.

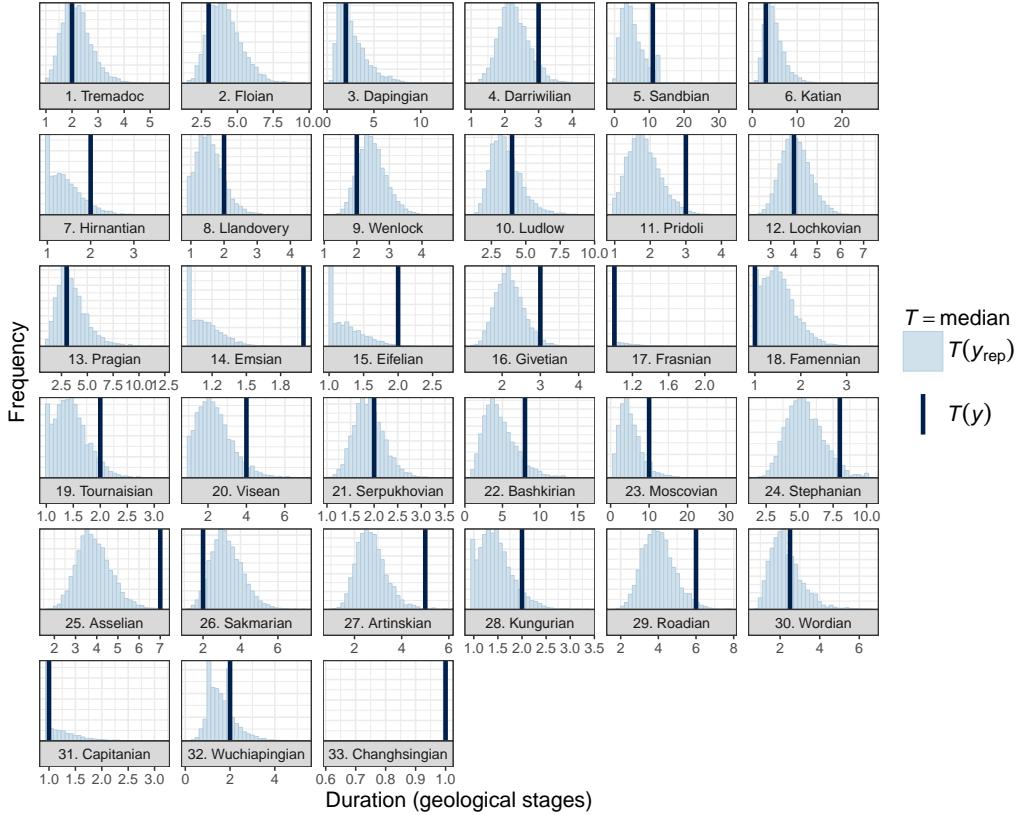


Figure 6: Comparison of the observed median genus duration (black vertical line) to a distribution of medians estimated from 100 simulated datasets (blue) for each of the origination cohorts. Model fit is evaluated by the similarity between the observed and the estimated, where good fit is demonstrated by the vertical line being “within” the simulated distribution.

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	-2.34	0.14	-2.52	-2.34	-2.16
	μ^r	geographic	1.19	0.21	0.93	1.19	1.45
		range					
	μ^v	environmental	-0.34	0.20	-0.59	-0.34	-0.07
		preference					
Standard deviation	μ^{v^2}	environmental	2.15	0.30	1.76	2.15	2.53
		preference ²					
	μ^m	body size	0.05	0.11	-0.09	0.05	0.19
	τ^i	intercept	0.51	0.11	0.38	0.50	0.65
	τ^r	geographic	0.77	0.17	0.55	0.75	1.00
Other	τ^v	range					
		environmental	0.97	0.18	0.76	0.95	1.20
	τ^{v^2}	preference					
		environmental	1.08	0.34	0.66	1.07	1.52
		preference ²					
	τ^m	body size	0.41	0.11	0.27	0.41	0.56
Other	δ	sampling	-2.77	0.15	-2.97	-2.77	-2.58
	α	"time"	1.61	0.05	1.54	1.61	1.67

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 increase (Table 1). Given the estimates of μ^r and τ^r , there is an approximately 7% ($\pm 5\%$ SD) probability that this relationships would be reversed ($\text{Pr}(\mathcal{N}(\mu^r, \tau^r) > 0)$).

Body size is estimated to have a slightly positive estimate, which means that species with greater than average size are expected to have a shorter duration than species with below average size (Table 1).

The group-level estimate of the effect of environmental preference is estimated from both μ^v and μ^{v^2} . 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, given the estimate of between-cohort variance τ^v , there is approximately 36.1% ($\pm 8\%$ SD) probability that, for any given cohort, taxa favoring open-ocean environments would have a greater expected duration than taxa favoring epicontinental environments ($\text{Pr}(\mathcal{N}(\mu^v, \tau^v) > 0)$). The estimate of μ^{v^2} indicates that the overall relationship between environmental preference and $\log(\sigma)$ is concave down (Fig. 7), with only a 3.6% ($\pm 4\%$ SD) probability that any given cohort is convex up ($\text{Pr}(\mathcal{N}(\mu^{v^2}, \tau^{v^2}) < 0)$).

The cohort-specific estimates of all the regression coefficients demonstrate a lot of between cohort variance, with no obvious long-term trends (Fig. 8). While most cohort-specific estimates are very similar to the overall cohort-level estimate, there are a few notable cohorts for which two or more individual-level parameter estimates diverge greatly from the group-level averages (Fig. 8). What's even more interesting is that for many cohort's for which two or more covariate effect estimates are different from the overall group-level mean are well fit by the model (Fig. 4, 5, 6).

In chronological order, the Wenlock (433.4-427.4 Mya) is the first origination cohort for which at least two effect estimates diverge from their group-level averages: the effect of the linear aspect of environmental preference β^v increases (shifting the parabola left-ward) while the effect of average body size β^m decreases meaning that greater than average size is associated with a longer duration than taxa with less than average body sizes (Fig. 8). Additionally, this cohort is

well fit by our model (Fig. 4, 5, 6) which allows us to have greater confidence in these results. This cohort is one of the few for which the estimated effect of body size is different from the group-level average.

During the Pragian (410.8-407.6 Mya) there are simultaneous excursions in mean taxon duration or extinction intensity β^0 and the effect of geographic range on duration β^r . This cohort has lower than expected extinction intensity β^0 and greater than expected effect of geographic range β^r . Additionally, this cohort is well fit by our model (Fig. 4, 5, 6) which increases confidence in these results.

The Emisian (407.6-393.3 Mya) origination cohort is estimated to have greater than average estimate of β^0 , lower than average estimate of β^r , and a greater than average estimate of β^m . Additionally, our model has less than adequate fit to this origination cohort as it underestimates the median duration of taxa in that cohort (Fig. 6), though the model successfully recapitulates mean duration and does a fair job of reproducing the survival function for that cohort (Fig. 4, 5).

The effects of the linear and quadratic aspects of environmental preference, β^v and β^{v^2} , for taxa originating in the Frasnian are both divergent from the overall group-level mean. For this cohort, β^v is estimated to be less than expected while β^{v^2} is estimated to be greater than expected. This result means that the curvature of the relationship increased while the apex was right-shifted from average. Additionally, this origination cohort is well fit by the model (Fig. 4, 5, 6) which increases confidence in interpreting these results.

Taxa originating during the Famennian (372.2-358.9 Mya) were estimated to have a greater than average expected duration than the group-level average (i.e. lower than average β^0) as well as lower than average estimate of β^v (i.e. right-shifted) (Fig. 8). Finally, this cohort has only moderate quality of fit as the model is able to estimate average cohort duration (Fig. 5) but has difficulty reproducing the survival function and median duration (Fig. 4, 6), meaning that these excursions from the group-level mean may be driven by the difference in processes underlying extinction of taxa in this origination cohort.

Many of the effect estimates for the Sakmarian origination cohort (295.5-290.1 Mya) are di-

vergent from the group-level averages (Fig. 8). Specifically, β^0 was estimated to be below the group-level average indicating greater average duration than average origination cohort, β^r was estimated to be above the group-level average indicating that geographic range has a weaker than normal effect on extinction risk, β^v has an above average estimate indicating a left-shifted relationship between environmental preference and duration, and finally β^m was estimated to be below the group-level average which means that taxa with greater than average body sizes are expected to have greater durations than taxa originating in that cohort with below average body size. The model does a poor job of reproducing the origination cohort survival curve (Fig. 4) while slightly under estimating mean and median durations of the origination cohort (Fig. 5).

The Kungarian origination cohort (279.2-272.3 Mya) is the final cohort for which there are at least two effect estimates are divergent from the group-level averages. Both the linear and quadratic aspects of environmental preference on expected duration for taxa in this origination cohort have divergent estimates (Fig. 8). β^v is estimated to be lower than the group-level average indicating a left-shifted relationship, while β^{v^2} is estimated to be greater than the group-level average which indicates an increasing peakedness to the relationship between environmental preference and expected duration. However, the model is able reproduce many aspects of that origination cohort's survival curve but is far from perfect (Fig. 4). Additionally, the model has a good fit to that cohort's mean duration (Fig. 5) while also underestimating median taxon duration (Fig. 6).

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. 9) and reflect how these three parameters act in concert and not just individually (Fig. 8). Because of the relationship between β^v and β^{v^2} , it is important to consider them together when drawing conclusions from the model. In many cases, the cohort-specific estimated relationship between environmental preference and duration is approximately equal to the group-level average, but 14 of the 33 analyzed origination cohorts have at least one of these three parameters being noticeably different from the group-level average.

The cohort originating in the Tremadocian stage are estimated to have a greater-than average

estimate of the linear aspect of environmental preference, β^v (Fig. 8); this means that the relationship between environmental preference and duration is slightly right-shifted than the average (Fig. 9). Additionally, the model adequately recapitulates the empirical survival function along with observed mean and median durations of that cohort (Fig. 4, 5, 6). These results indicate that this cohort has a different relationship between environmental preference and duration than the average origination-cohort.

The Pragian, Emisian, Moscovian, and Capitanian origination cohorts are all estimated have a different average durations from the group-level average based on their estimates of β^0 (Fig. 8). While the cohort-specific estimates vary from greater-than or less-than average, none of these origination cohorts are estimated to have a simultaneous divergence in the linear or quadratic aspects of environmental preference. This result means that most variation in these cohorts relationships between environmental preference and duration is driven not by changes to actual environmental preference than overall changes in average duration, i.e. extinction intensity (Fig. 9). Of these cohorts, the Pragian and Moscovian origination cohorts are estimated to have greater than average species durations (Fig. 8), though the Moscovian is extremely poorly fit by this model (Fig. 4, 5, 6). In contrast, the Emisian and Capitanian origination cohorts are estimated to have greater-than average estimates of β^0 indicating greater-than average extinction intensity, as average duration decreases as β^0 increases (Fig. 8).

The Wenlock, Visean, Serpukhovian, and Asseilian cohorts all have estimates of the linear aspect of environmental preference, β^v , that are lower than the group-level average. Additionally, this is the only aspect of the relationship between environmental preference and expected taxon duration for these cohorts than is divergent from the group-level means (Fig. 8). This change to β^v manifests as a left-shifted relationship between environmental preference and duration such that the apex of the relationship is closer to 0 which means that there is no “advantage” for favoring either an epicontinental and open-ocean environment; this is slightly different from the group-level average relationship which has an approximate 64% probability of the apex being right-shifted so that taxa which slightly favor epicontinental environments are expected to have

a greater duration than those that favor open-ocean environments (Fig. 7). Keep in mind, however, there is an approximately 96.4% probability that any cohort would have a concave down relationship between environmental preference and duration which means that taxa which favor purely open-ocean or epicontinental environments are expected to have shorter durations than taxa which can exist in either environment (Fig. 9).

The Famennian and Tournaisian origination cohorts are all estimated to have greater than average estimates of the linear aspect of environmental preference (β^v) (Fig. 8). These estimates mean that the relationship between environmental preference and duration is right-shifted such that taxa which prefer epicontinental environments over open-ocean are expected to have greater durations than average. Specifically, this means that taxa that can tolerate both open-ocean and epicontinental environments but preferentially occur in epicontinental environments are expected to have greater durations than those that occur equally in these two environments (Fig. 9).

The Frasnian origination cohort is one of two origination cohorts for which the quadratic aspect of the relationship between environmental preference and duration β^{v^2} . Additionally, this parameter is the only aspect of environmental preference (or average duration) that is divergent for this cohort (Fig. 8). Additionally, this cohort is well fit by the model so confidence in these results may represent aspects of the data generating process.

The Sakmarian and Kungarian are the only two origination cohorts for which two of β^0 , β^v , and β^{v^2} are divergent from the group-level averages (Fig. 8), though they have very different divergences. As explained earlier, the Sakmarian origination cohort is estimated to have greater-than average β^0 (increased extinction risk) and a left-shifted relationship between environmental preference and duration (lower-than average estimate of β^v), while the Kungarian cohort is characterized by a greater-than average estimate of β^{v^2} and a lower-than average estimate of β^{v^2} , which translates into a relationship between environmental preference and duration that is both a more peaked than average as well as being more left-shifted than average (Fig. 9).

Included in this model are also estimates of the correlation between the covariate estimate

time-series, as well as that of the intercept (Fig. 10). These estimates allow for testing any hypotheses about the relationship between extinction intensity (the intercept) and extinction selectivity (covariate effects).

There is an approximately 99.4% probability that cohort estimates of β^0 and β^r are negatively correlated, with median estimate of correlation being -0.57. This result means that for any cohort, we would expect that if extinction intensity increases (β^0 increases), the estimated effect of geographic range increases (β^r increases). This result is strong evidence for a relationship between intensity and selectivity with respect to geographic range size.

There is an approximate 94.2% probability that the cohort-specific estimates of β^0 and β^v are negatively correlated; this means that as extinction intensity increases it is expected that epicontinental taxa become more favored over open-ocean environments (i.e. as β^0 increases, β^v decreases). This result is strong evidence for a relationship between intensity and selectivity with respect to the linear aspect of environmental preference.

Correlations between the non-intercept estimates reflect potential similarities in selective pressures between cohorts, however there is only weak evidence of any potential cross-correlations in cohort-specific covariate effects. There is an approximate 57.6% probability that β^r and β^v are positively correlated. This lack of cross-correlation may be due in part to the 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.

Conversely, There is a 80.4% probability that estimates of the effect of geographic range (β^r) and the quadratic aspect of environmental preference (β^{v^2}) are positively correlated; this is weak evidence of a relationship between the effects of these covariates. Thus, as the effect of geographic range increases, we might expect with weak evidence that the peakedness of relationship between environmental preference and duration to increase. However, because there is only a 80.4% probability of a positive correlation, this result cannot be interpreted with authority. Instead, this result is an opportunity for future research to understand a potential relationship between

geographic range, environmental preference, and species duration.

Sampling was found to have a positive effect (negative 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.

The Weibull shape parameter α was found to be approximately 1.61 (± 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 result is consistent with recent empirical results and may be caused by newly originating species have a fundamentally lower risk of extinction compared to species which have already originated (???). This result is also consistent with a recently proposed nearly-neutral evolution where competition/selection/evolution drives whole communities to increase in average fitness over time while still maintaining constant relative fitness across the community, thus older species are more likely to go extinct because of having a fundamentally lower average fitness than newly originating species (?). This results, however, is not consistent with other empirical results from the marine fossil record (??) 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.

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 to affect extinction risk was geographic range while traits that had previously been associated with differences in duration 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 difference in survival for taxa favoring epicontinental environments over open-ocean environments is expected to increase; this is consistent with the results of ?. There is no evidence for a correlation between the effects of geographic range and environmental preference on taxon duration.

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. 7), though the curvature of the relationship varies from rather shallow to very peaked (Fig. 9). However, this relationship is not symmetric about 0, as taxa favoring epicontinental environments are expected with approximately 75% probability to have a greater duration than taxa favoring open-ocean environments. This description of environment preference 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 the extent and complexity of the “survival of the unspecialized” hypothesis as it applies to all dimensions of an adaptive zone.

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

broad tolerances and flexible taxa with narrow per-stage preference end up being treated the same way. Future work 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 Modern 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 (??????????).

Conclusion

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 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.

Acknowledgments

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, R. Ree, and G. Slater for helpful discussion during the conception of this study. I’d also like to thank D. Bapst, N. Pierrehumbert and M. Villarosa Garcia for additional comments. An additional thank you to A. Miller for the epicontinental versus open-ocean assignments. Finally, thank you to P. Harnik, L. L. Hsiang, G. Simpson, and the other anonymous reviewers for their comments at various points during this manuscripts life that improved this manuscript. This entire study would not have been possible without the Herculean effort of the many contributors to the Paleobiology Database. In particular, I would like to thank J. Alroy, M. Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N.

Heim, A. Hendy, S. Holland, L. Ivany, W. Kiessling, 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 Exobiology grant (NNX10AQ446) to A. Miller and M. Foote. I declare no conflicts of interest. This is Paleobiology Database publication XXX. The code necessary for reproducing these results is available as a Zenodo archive DOI 10.5281/zenodo.46928.

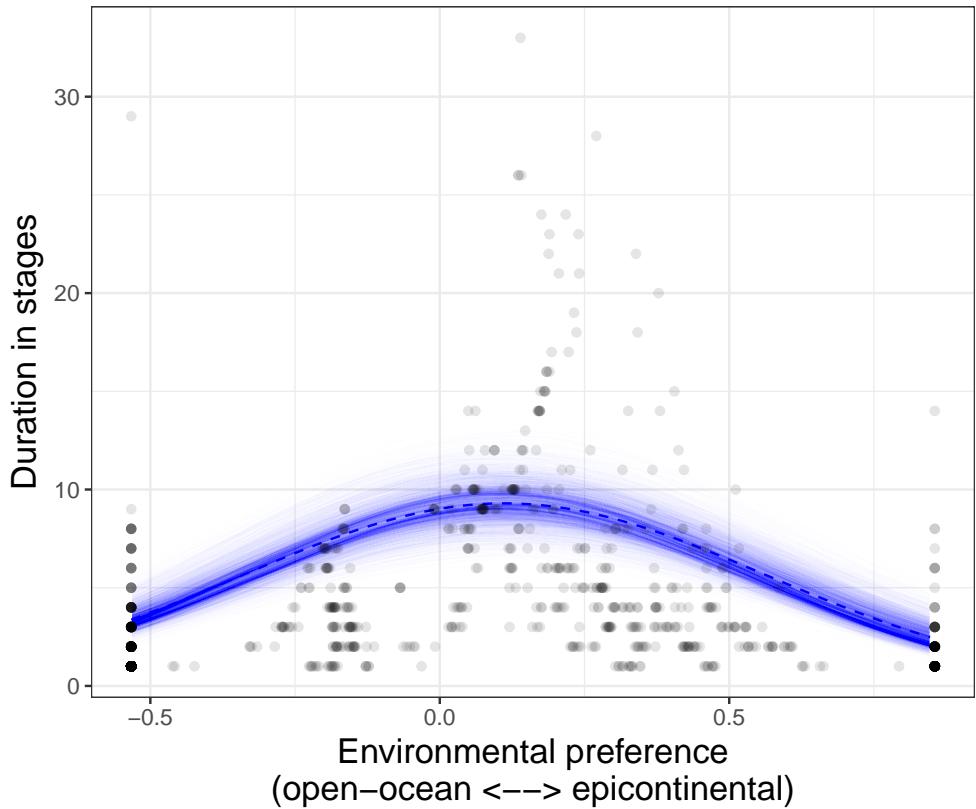


Figure 7: The overall expected relationship between environmental affinity v_i and a $\log(\sigma)$ when $r = 0$ and $m = 0$. The 1000 semi-transparent lines corresponds to a single draw from the posterior predictive distribution, while the highlighted line corresponds to the median of the posterior 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.

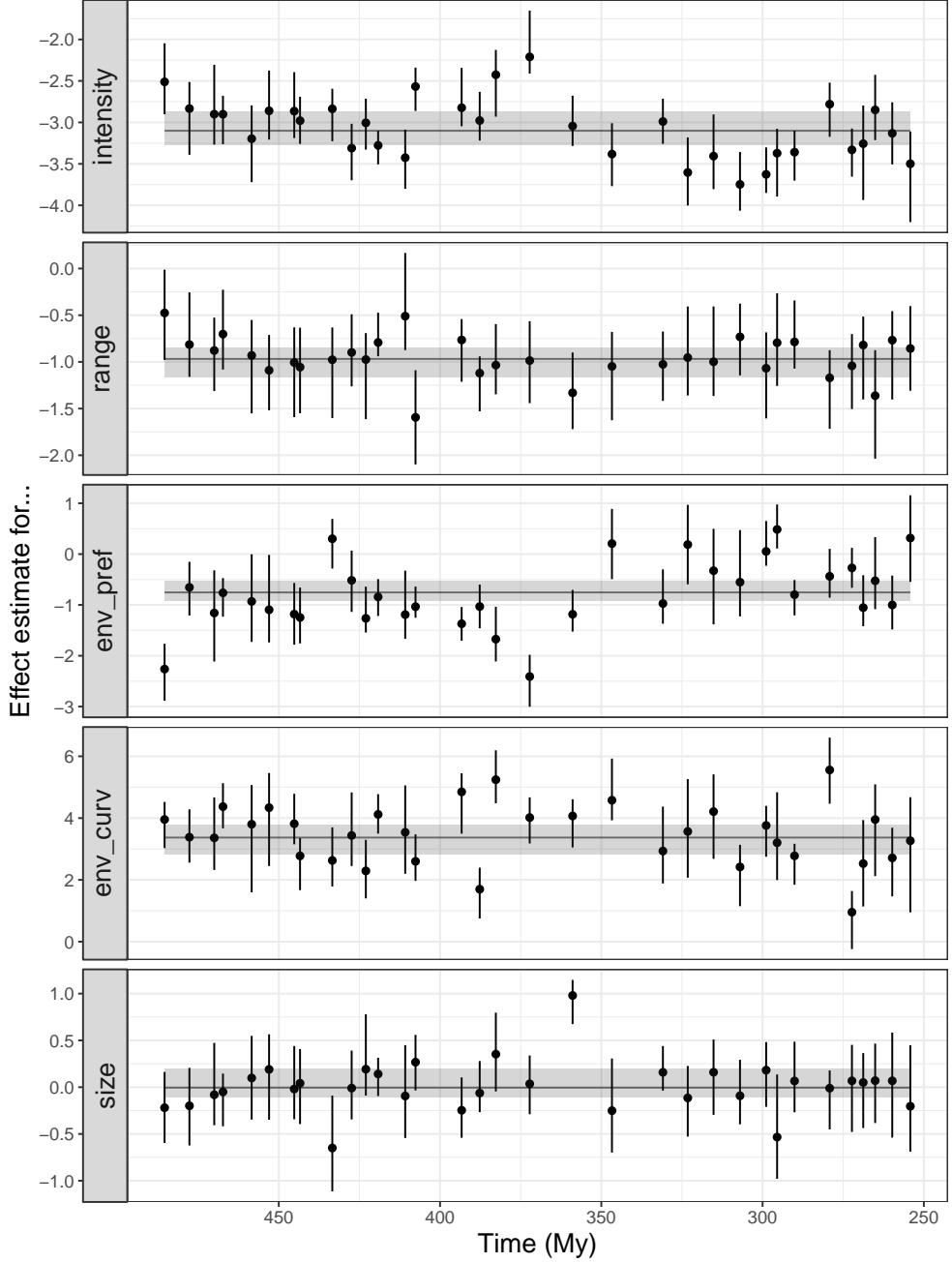


Figure 8: Comparison of cohort-specific estimates of β^0 , the effect of geographic range on extinction risk β^r , the effect of environmental preference β^v and β^{v^2} , and body size β^m . Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. Points are 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).

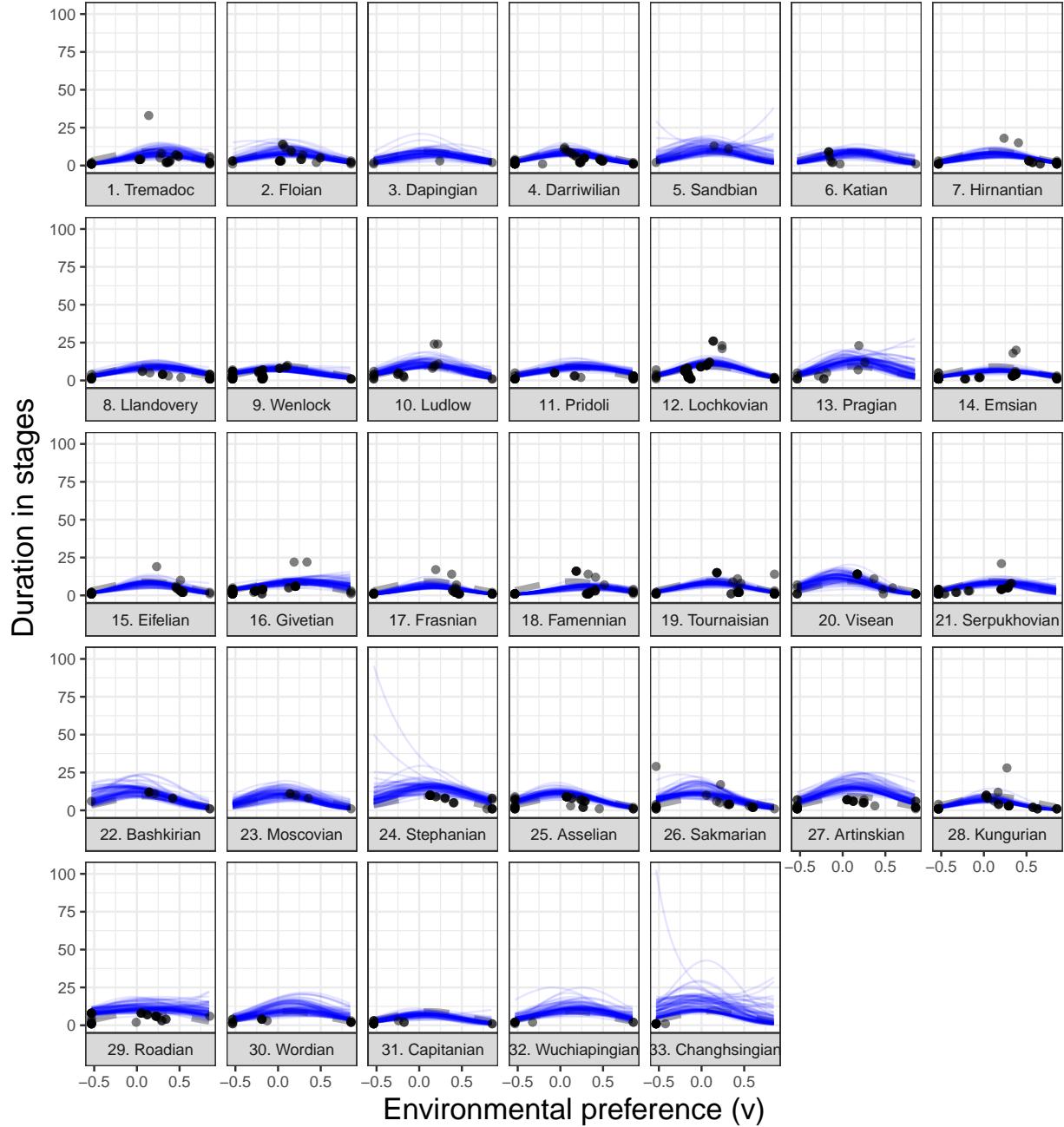


Figure 9: Comparison of origination cohort-specific (posterior predictive) estimates of the effect of environmental preference on $\log(\sigma)$ to the mean overall estimate of the effect of environmental preference. Cohort-specific estimates are from 100 posterior predictive simulations across the range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

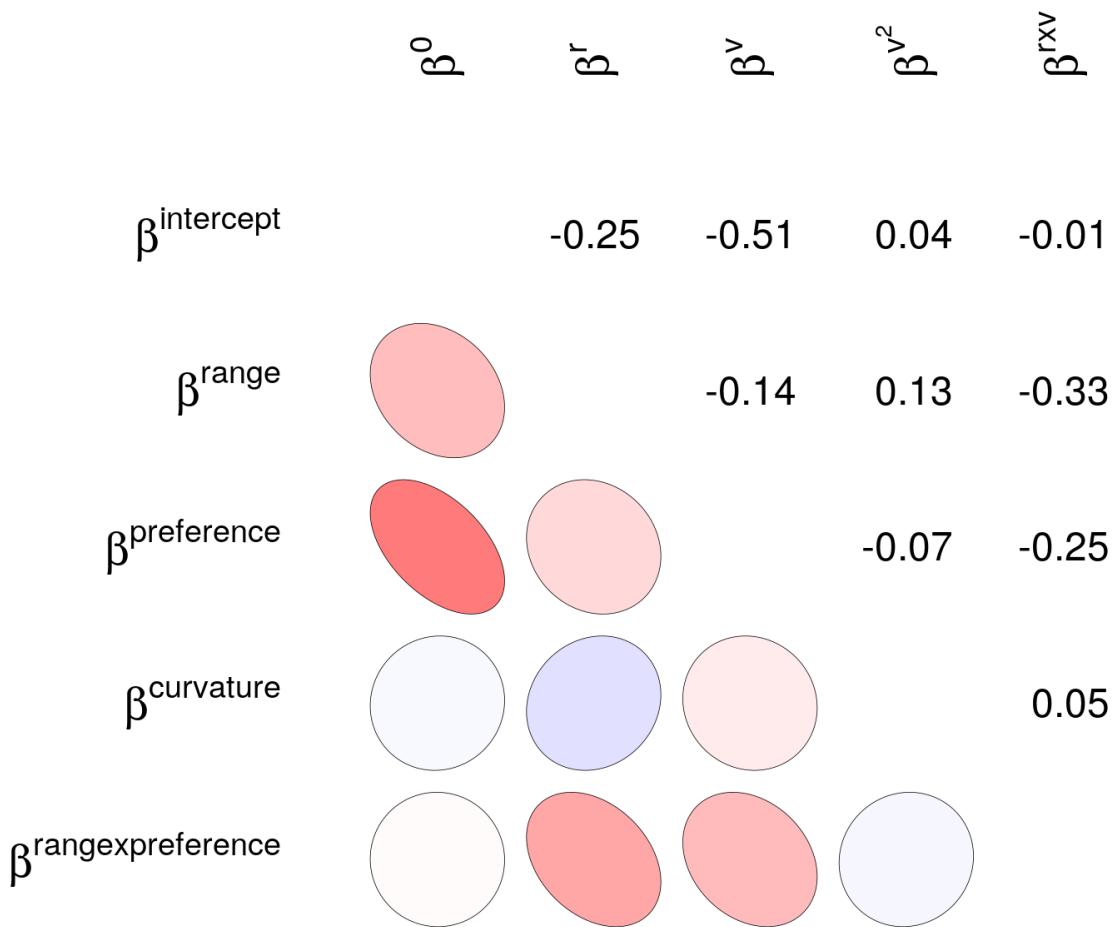


Figure 10: Mixed graphical and numerical representation of the correlation matrix Ω of variation in cohort-specific covariate estimates. These correlations are between the estimates of the cohort-level effects of covariates, along with intercept/baseline extinction risk. The median estimates of the correlations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