Extinction intensity, selectivity and their combined macroevolutionary influence in the fossil record

Jonathan L. Payne¹, Andrew M. Bush², Ellen T. Chang³, Noel A. Heim¹, Matthew L. Knope⁴ and Sara B. Pruss⁵

¹Department of Geological Sciences, Stanford University, Standford, CA, USA
²Department of Ecology and Evolutionary Biology and Center for Integrative Geosciences, University of Connecticut, Storrs, CT, USA
³Division of Epidemiology, Department of Health Research and Policy, Stanford University School of Medicine, Standford, CA, USA
⁴Department of Biology, University of Hawaii, Hilo, HI, USA
⁵Department of Geosciences, Smith College, Northampton, MA, USA

Received: 10 March 2016
Accepted: 16 September 2016

The macroevolutionary effects of extinction derive from both intensity of taxonomic losses and selectivity of losses with respect to ecology, physiology and/or higher taxonomy. Increasingly, palaeontologists are using logistic regression to quantify extinction selectivity because the selectivity metric is independent of extinction intensity and multiple predictor variables can be assessed simultaneously. We illustrate the use of logistic regression with an analysis of physiological buffering capacity and extinction risk in the Phanerozoic marine fossil record. We propose the geometric mean of extinction intensity and selectivity as a metric for the influence of extinction events. The end-Permian mass extinction had the largest influence on the physiological composition of the fauna owing to its combination of high intensity and strong selectivity. In addition to providing a quantitative measure of influence to compare among past events, this approach provides an avenue for quantifying the risk posed by the emerging biodiversity crisis that goes beyond a simple projection of taxonomic losses.

1. Background
Palaeontological study of extinction has long focused on intensity, with mass extinctions defined by an unusual magnitude of taxonomic loss [1]. However, the macroevolutionary importance of mass extinction derives as much from the ecological and phylogenetic selectivity of taxonomic losses as from total magnitude [2–4]. Moreover, debate continues as to whether mass extinction is a distinct mode of taxonomic loss versus an intensified version of background extinction [5–7]. Consequently, quantitative measures of extinction selectivity that are independent of magnitude are critical for identifying proximal causes of extinctions and for quantifying their overall influence.

2. Measuring selectivity
Most previous analyses of extinction selectivity in the fossil record fall into one of three categories: (i) comparisons of extinction intensity among groups during an interval using raw data [8,9] or the G-statistic [10]; (ii) statistical tests comparing the mean and/or distribution of values of predictor variables between victims and survivors (e.g. t-tests, Kolmogorov–Smirnov tests) [11,12]; and (iii) regression analyses [13–17].
Simply comparing extinction intensity between categories does not effectively separate selectivity from intensity. Consider two extinction events. Extinction affected 1% of benthic genera and 2% of pelagic genera in the first event versus 12% and 18%, respectively, in the second. The second event is more intense and shows a greater absolute difference in intensity between groups. Despite the larger absolute difference in intensity for the second event, standard relative measures of selectivity, such as the odds ratio \([18]\), would be greater for the first. However, the greater difference in proportional losses in the second event reflects increased extinction intensity rather than increased selectivity (figure 1). Comparison between background and mass extinction is prone to this conflation of change in intensity with change in selectivity. Here, for simplicity, we leave aside the additional, important problems of correcting raw stratigraphic ranges and extinction rates for sampling effects \([19,20]\).

Statistical tests comparing mean values or distributions of traits between victims and survivors have similar limitations. For example, two events could show the same difference in mean body size between victims and survivors (i.e. the measure of selectivity in a \(t\)-test), but an event with the sizes of victims and survivors more tightly clustered around their respective means would reflect a steeper gradient in survivorship along the body size axis, suggesting a greater degree of selectivity. More generally, there is no single metric based on the \(t\)-test that can be used to compare selectivity among events because the change in probability of extinction with size is related to the difference in means for victims and survivors as well as the dispersion within each group. Additionally, interpretations of \(t\)-tests and other tests comparing distributions (e.g. Kolmogorov–Smirnov, Mann–Whitney) often focus on \(p\)-values rather than effect sizes. When following this practice, it is difficult to assess whether intervals exhibiting significant differences between victims and survivors reflect greater actual differences than non-significant intervals, or just larger sample sizes. This issue is particularly important when comparing background intervals and mass extinctions, because there are often many fewer victims in the former than the latter, limiting statistical power.

Regression analysis, like the G-statistic, is preferable to the other methods described above because it provides explicit estimates of effect size, uncertainty and statistical significance. Regression analysis alone can be applied to the simultaneous analysis of multiple predictor variables and interactions among predictors. Logistic regression is used increasingly in palaeontological analysis of extinction selectivity \([13–17]\) because outcomes are binary (each genus either goes extinct or survives) and predictors may be categorical, ordered or continuous. As illustrated in figure 1, the odds ratio (the measure of selectivity) is independent of extinction intensity and, therefore, is ideal for comparing selectivity between background and mass extinction events without confounding intensity and selectivity. Several other measures of selectivity are also independent of intensity \([10,21]\), but we focus on logistic regression coefficients for the examples below.

Multivariate adjustment via multiple regression is of particular importance owing to the potential influence of confounding variables (cf. \([22]\)). Consider the following example: extinction risk is inversely associated with geographical range within both benthic and pelagic genera. However, because pelagic life habits are typically associated with both wide geographical range and high extinction probability, geographical range and extinction risk are (false) positively associated in

![Figure 1. Hypothetical examples illustrating the difference between measuring selectivity in terms of absolute difference in risk (\(a,c\)) versus a relative difference in risk (log odds ratio) \((b,d)\).](http://rsbl.royalsocietypublishing.org/)

(a) Extinction intensity for two states of an ecological predictor (here geographical range) for four hypothetical extinction events of differing intensity. (b) Selectivity of these four events, measured as log odds ratio, is equivalent. Error bars indicate 95% confidence intervals on estimated coefficients. (c) Differences in extinction intensity across two states of one ecological predictor (tiering: benthic versus pelagic) as a function of geographical range (narrow versus wide). Point sizes reflect sample sizes. Open versus filled symbols as in (a). (d) Selectivity of (c), illustrating the confounding effect of tiering on geographical range when range is analysed without adjusting for tiering. Although wide-ranging genera go extinct preferentially overall, narrow-ranging genera go extinct preferentially within each tiering category. Multiple regression including tiering removes this confounding effect.)
the aggregate dataset (figure 1c,d)—an effect known as Simpson’s paradox [23]. Unlike univariate approaches which permit evaluation of only one predictor variable at a time, multivariate approaches enable simultaneous estimation of the mutually adjusted effects of several predictors that could otherwise act as confounders. In this case, it clearly identifies the inverse association between geographical range and extinction risk after adjusting for the association of pelagic habitat (a confounder) with both wide geographical range and extinction probability (figure 1d). Similarly, differences in rates of taxonomic turnover among higher taxa can confound the comparison of extinction intensity among stages except when using regression or the G-statistic [23].

3. Example: physiological buffering

To illustrate these principles, we present a new analysis of extinction selectivity with respect to physiological buffering capacity (i.e. the sophistication of the respiratory and circulatory systems) across the fossil record of 32,894 genera of marine animals. We follow the physiological classification scheme of Knoll et al. [8] (table 1) and assign stratigraphic ranges following Heim et al. [24]. Data are available within the Stanford Digital Repository (see Data accessibility statement for details).

Consistent with previous studies [8,9,22,25], we find that some, but not all, proposed ocean acidification and anoxic events are associated with the preferential loss of poorly buffered genera (figure 2a). The end-Guadalupian (260 Ma), end-Permian (252 Ma) and Pliensbachian/Toarcian (183 Ma) events emerge as moderately to strongly selective, whereas the Frasnian/Famennian (372 Ma), end-Triassic (201 Ma),
Albian (120 Ma), Cenomanian/Turonian (93 Ma) and Palaeocene/Eocene (56 Ma) do not. In addition, many stages not previously associated with acidification or anoxic events were also selective with respect to buffering capacity. The lack of strong change in the composition of the fauna across most stage boundaries [26], despite selective extinction, likely reflects the offsetting effects of higher origination of poorly buffered genera during most stages. Because pelagic genera tend to be moderately or well buffered and to have high rates of genus extinction, we repeated the analysis with multivariate adjustment for tiering (pelagic versus benthic), assuming additive effects (i.e. no interaction term). Stratification by stage allows for heterogeneity in the effect of tiering (or buffering capacity) across time. Adjusting for tiering increases the physiological selectivity of extinction during the Frasnian/Famennian and many Mesozoic stages, but not particularly during Cretaceous and Cenozoic acidification and anoxic events (figure 2d). Adjustment for tiering does not reduce selectivity for stages identified as physiologically selective in the univariate analysis.

A key topic for further investigation is the influence of sampling biases such as the ‘Pull of the Recent’ and Lagerstätten effects on selectivity and intensity measures. Approaches to correcting intensity metrics for sampling have been widely discussed (e.g. [19,20]). This issue is of particular importance because differences in completeness among higher taxa [27] covary with physiology (cf. table 1) and other ecological traits, such as geographical range size.

4. Quantifying extinction influence

The overall influence of an extinction event on the taxonomic and ecological composition of the global biota derives from the combination of its intensity and selectivity. Selectivity coefficients independent of extinction intensity open the opportunity to generate metrics that combine the intensity and selectivity of extinction. To quantify overall influence, we plot the geometric mean (figures 2c,f)—a simple measure appropriate for averaging variables with different numerical ranges—of selectivity (figure 2a,d; logistic regression coefficient) and extinction intensity (figure 2b,e; per cent genus extinction). The end-Permian extinction was the most influential on the physiological composition of the marine fauna (figure 2c,f), whether or not physiological selectivity is adjusted for tiering. After adjusting for tiering, the Frasnian/Famennian, end-Triassic and Pliensbachian/Toarcian events also exhibit a relatively strong influence (figure 2f). The extent to which reduced selectivity in the Cenozoic reflects a change in extinction dynamics, taxonomic composition or the ‘Pull of the Recent’ [28] cannot be determined from these data.

5. Conclusion

Phanerozoic analysis of extinction selectivity with respect to physiological buffering capacity suggests that the preferential loss of poorly buffered genera extends beyond recognized ocean acidification and anoxic events to many other stages associated with no known trigger mechanism. Adjusting for the influence of a benthic versus pelagic habitat through multiple logistic regression indicates that physiological selectivity has been common in the history of marine animal life, but particularly strong at certain times. Quantifying the influence of extinction events as the geometric mean of intensity and selectivity provides a useful metric for comparing overall influence across past extinction events, highlighting the profound influence of the end-Permian extinction relative to all other events. In addition, a quantitative influence scale has potential as a tool for assessing the potential magnitude of biological disruption associated with the emerging biodiversity crisis under a range of potential intensity and selectivity scenarios.

References