Occupancy history influences extinction risk of fossil marine microplankton groups

Isaiah E. Smith*, Ádám T. Kocsis, Wolfgang Kiessling

Department of Geography and Geosciences, GeoZentrum Nordbayern, Friedrich-Alexander-

Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

* Corresponding author

Correspondence email: isaiah.em.smith@gmail.com

**Abstract**

Geographic range has long been acknowledged as an important determinant of extinction risk. The trajectory of geographic range through time, however, has not received as much scientific attention. Here, we test the role of change in geographic range – assessed by a measure of proportional occupancy of grid cells – in determining the extinction risk in four major microplankton groups over the last 66 million years: foraminifera, calcareous nannofossils, radiolarians, and diatoms. Logistic regression was used to assess the importance of standing occupancy and occupancy change in the extinction risk of species. We find that while standing occupancy is a major determinant of extinction risk in all microplankton groups, the change in occupancy accounts for an average of 41% of the explanatory power shared by the two analyzed variables, with a maximum value of 77%. We also find that as temporal resolution decreases, the predictive ability of these variables increases. Our results highlight the importance of incorporating both geographic range and its change through time into extinction models. The ability of occupancy trajectory to help predict extinction risk underlines the necessity of paleontological data in modern conservation efforts.

## 1. Introduction

There is a rich literature documenting the effect of smaller geographic range size increasing risk in contemporary and ancient extinctions (e.g. Foote et al., 2016; Foote et al., 2007; McKinney, 1997; Payne and Finnegan, 2007; Purvis et al., 2000; Staude et al., 2020). The International Union for the Conservation of Nature (IUCN) uses geographic range size as one of the five key criteria by which the risk status of a species is assessed in the "Red List of Threatened Species" (Mace et al., 2008). The temporal trajectory of geographic range as a predictor of global extinction has been previously explored in the paleontological literature

48 (Liow et al., 2010; Foote et al., 2007, Tietje and Kiessling, 2013; Kiessling and Kocsis, 2016;

49 Saulsbury et al., 2023). Here, we further build upon this topic and explore how it applies to

50 marine microplankton.

51     Based on a data set of Cenozoic marine invertebrates from the Paleobiology Database

52 (https://paleobiodb.org/), Kiessling and Kocsis (2016), suggested that the trajectory of

53 geographic range has the potential to inform extinction risk. However, the coarse stratigraphic

54 resolution of the macroinvertebrate record (geological stages, about 5 million years in duration)

55 puts constraints on the fidelity of any approach that depends on the spatiotemporal distribution of

56 species. Due to their sheer abundance, unicellular groups are less affected by such issues and can

57 be used for finely resolved studies of assemblage changes (e.g., Strack et al. 2024) and

58 biogeography (e.g., Swain et al., 2024). We assess here the importance of geographic range

59 (expressed as proportional grid occupancy) and its temporal trajectory on extinction risk of

60 marine planktonic organisms. By using a temporally finely-resolved dataset of fossil plankton,

61 we can better assess the degree to which the trajectory of geographic occupancy influences

62 extinction risk in marine life.

63

64 **2. Data and Methods**

65 **2.1 Sourcing and cleaning of raw data**

66     We downloaded occurrence records from the Neptune Sandbox Berlin ("NSB"; Lazarus,

67 1994; Renaudie et al., 2020; data downloaded 30 August 2023) using the R package

68 "NSBcompanion" (Renaudie, 2019; version 2.2) as well as the Triton database (Fenton et al.

69 2021, version 2). Four taxonomic groups were downloaded: planktonic foraminifera, calcareous

70 nannofossils, radiolarians, and diatoms. Each of the four data sets were downloaded with the

71 taxonomy resolved using the IODP Taxonomic Name List Project (Renaudie et al., 2020), a

72   built-in option that we specified prior to downloading. Additionally, questionably-identified taxa

73   were excluded from the download. Open-nomenclature taxa and possibly problematic or

74   reworked occurrences were also excluded using the built-in NSB download options. The NSB

75   holds taxon occurrences stretching back to the late Jurassic, but we limit our analysis to the

76   Cenozoic record (i.e. the last 66 million years) to ensure a consistent age range for all for

77   taxonomic groups, since both the diatom and the radiolarian NSB records only exist for the

78   Cenozoic.

79        The NSB database includes the estimated age as well as the modern and estimated paleo-

80   coordinates (longitude and latitude) of each fossil occurrence calculated internally based on the

81   plate tectonic reconstruction by Boyden et al. (2011, Fig. S1). Each dataset (foraminifera,

82   calcareous nannofossils, radiolarians, diatoms) was cleaned to remove any occurrences that were

83   missing age, paleo-coordinate, and/or relevant taxonomic information. The counts of unique

84   species and the number of occurrence records in each data set (post-cleaning) are provided in

85   Table 1. All cleaning and subsequent analyses were carried out in R 4.4.3 (R Core Team, 2022).

86

87

88

89

90

91

92

93

94

95

96

| Data Set Summaries | | | | |
|---|---|---|---|---|
| | Foraminifera | Calcareous Nannofossils | Radiolarians | Diatoms |
| Number of unique species | 552 | 737 | 840 | 966 |
| Number of occurrences | 96426 | 150346 | 90201 | 91194 |
| Percent of species that are extant and have records in more than 1 time bin | 6.1 | 1.5 | 11.3 | 14.6 |
| Number of unique species-bin pairings with "gap-omission" approach, excluding NA occurrences for occupancy and occupancy change. Bin size = 1 million years. | 2951 | 3817 | 4512 | 4225 |

97  **Table 1.** The number of unique species, the number of total occurrence records, the proportion of
98  species that are both extant and occur in more than one bin, and the number of species-bin
99  pairings (post-cleaning).
100

101      We assigned occurrences from each data set to time bins of either 0.1, 0.2, 0.5, or 1.0

102  million years, noting that 0.1 million years is currently the lower limit for global correlation. For

103  each time bin size, the first bin stretched from the present (0 Ma) to either 0.1, 0.2, 0.5, or 1.0

104  million years into the past. Each subsequent bin encompassed the following increment stretching

105  progressively further into the past.

106      We assessed stratigraphic ranges as defined by the oldest and youngest fossil

107  occurrences. Due to reworking and other processes, the documented raw ranges may not reflect

108  the true durations of species. Therefore, we also applied the recommended "Pacman profiling"

109  (Lazarus et al., 2012), a stratigraphic outlier correction, to reduce the impact of outliers and

110  reworking on the data. The degree of Pacman trimming on the NSB data was determined via a

111    calibration process that used speciation and extinction ages of a given subset of each taxonomic

112    group. Based on this subset, the degree of trimming necessary to restore the "true" temporal

113    ranges of species could be estimated. Calibration ages were sourced from the Triton database

114    (Fenton et al. 2021) for foraminifera, Nigrini et al. (2006, obtained from Lazarus et al., 2012) for

115    radiolarians, the "Barron Diatom Catalog" (Lazarus et al., 2014) for diatoms, and a custom

116    species list constructed from Mikrotax (www.mikrotax.org; Huber et al., 2017) for calcareous

117    nannofossils. Potential trim values ranging from 0% to 16% of the raw ranges, at 1% intervals,

118    were analyzed. Pacman calibration was carried out on datasets after they had been trimmed to the

119    last 66 million years. Trim values were selected such that they minimized the average absolute

120    difference between the actual and the represented speciation or extinction ages of the species

121    present in the calibration set. The best-performing trim values were implemented in this study,

122    although the key results presented here do not change in the absence of Pacman profiling. Those

123    trim values were as follows: foraminifera (top: 15%, bottom 3%), calcareous nannofossils (top:

124    15%, bottom: 5%), diatoms (top: 11%, bottom: 4%), radiolarians (top: 10%, bottom: 7%). Per

125    capita extinction rates were calculated using the formula from Foote (1999), without normalizing

126    for interval length.

127

128    **2.2 Analysis of completeness**

129         In order to quantify the degree to which sampling completeness affected downstream

130    analyses, we employed two separate completeness metrics: the three-timer completeness metric

131    (Alroy, 2008) and the simple completeness metric ("SCM"; Benton, 1985). The three-timer

132    completeness metric is the ratio of "three-timer" taxa (those which occur in bin $i$-1, bin $i$, and bin

133    $i$+1) to all taxa which occur in both bin $i$-1 and bin $i$+1 (irrespective of their presence in bin $i$,

134    "part-timer" taxa). The three-timer metric was calculated from the three-timer and part-timer

135  counts returned by the "divDyn" R extension package (Kocsis et al., 2019; version 0.8.3). The

136  simple completeness metric is the ratio of time bins with confirmed taxon occurrences to the

137  inferred (by recorded observations before and after a focal time interval) number of time bins

138  occupied by that taxon.

139

140  **2.3 Calculating occupancy**

141      For each data set, paleo-coordinates of samples were assigned to equal-area geographic

142  cells using the R package "icosa" (Kocsis, 2020; version 0.11.1) for the calculation of

143  proportional grid occupancy. Proportional grid occupancy is a recognized metric for assessing

144  geographic range in the fossil record, where contemporaneous sampling is impossible and

145  incomplete preservation is common (Foote et al., 2007; Darroch et al., 2022). Several cell-sizes

146  were analyzed ranging in edge length between 3.33° to 2°. There was little variation in results

147  within this range, so the highest resolution (4002 cells with 2° edge length, mean area of 1.3 x

148  $10^5$ km$^2$) was selected for this study. The present-day distribution of samples can be seen in

149  supplementary Fig. S1.

150  As counts of occupied cells tend to be biased by sampling (Kiessling, 2005), we calculated

151  proportional occupancy of each species in every time bin. Proportional occupancy is the number

152  of geographic cells occupied by the species divided by the total number of sampled cells in a

153  given time bin. For simplicity, we refer to what is actually proportional occupancy as *occupancy*

154  from here forward. Furthermore, the number of unique Longhurst (2007) biogeographic

155  planktonic provinces that were occupied by each species in each time bin was calculated, and the

156  Pearson correlation of this value with the raw number of occupied geographic cells was

157  calculated. Autocorrelation was accounted for by differencing temporally consecutive values

158  prior to calculating correlation values.

**2.4 Change in occupancy**

159

160  In addition to standing *occupancy*, the change in *occupancy* between consecutive time

161 bins was calculated by taking the natural log of the ratio of *occupancy* in time bin *i* to *occupancy*

162 in time bin *i*-1. The log transformation serves to standardize the magnitude of change and

163 produces positive values for increases in *occupancy* (range expansions) and negative values for

164 decreases in *occupancy* (range contractions). The correlation of the first differences between

165 occupancy and occupancy change was computed to determine if the data was affected by

166 multicollinearity.

167  Initially, instances where *occupancy* values in bin *i* or *i*-1 were 0 (no occurrences) were

168 coded as missing data for *occupancy change* and removed from the final data set. While

169 removing these records prevents the inclusion of undefined *occupancy change* values in the final

170 data set, it greatly reduces the number of occurrences for a given taxon, especially for species

171 whose sampling is fragmentary. This effect is magnified by the fact that for each time bin with

172 zero occurrences of a given taxon (a "gap" in that taxon's fossil record), two data points are

173 removed from the final data set for that taxon. This overall loss of data becomes more

174 pronounced with smaller bin sizes.

175 To combat this effect, we employed a "gap-omission" approach, whereby the change in

176 *occupancy* was calculated based on the previous occurrence of the taxon (regardless of when that

177 was) rather than the previous time bin, *per se*. Thus, *i* and *i*-1 do not necessarily correspond to

178 sequential time bins in this approach, but rather consecutive positive sampling intervals for each

179 given taxon. With this approach, consecutive taxon occurrences are included even when

180 separated by "gaps", thus retaining more data to the final data set. Although both approaches

181 yield the same basic results (see Tables S1 and S2), we used the "gap-omission" approach for the

182 sake of retaining a larger data set.

## 2.5 Binomial logistic modeling

183 

184       For every species, a record of each time bin in which that species occurred was included

185 in the final dataset as a single row. Each unique species-bin pairing (row) is characterized with

186 the *occupancy* and a binary extinction indicator in the focal time bin as well as the change in

187 *occupancy* from the previous time bin. An extinction indicator value of 1 was assigned if an

188 occurrence was the last time bin in which a species occurred for the entire data set (the species

189 went extinct or permanently disappeared from the fossil record during this interval). An

190 extinction value of 0 was assigned for all other occurrence records (the species did not go extinct

191 during this interval). Species that are still extant, or those which only went extinct during the

192 most recent time bin (which spans up to the present), would by default be assigned an extinction

193 value of 1 in the most recent time bin. To avoid this edge effect, all occurrences from the most

194 recent time bin were removed prior to model fitting.

195       Binomial logistic models were constructed to examine the dependency of extinction on

196 *occupancy* and *occupancy change*. Both variables were examined with respect to the per-interval

197 probability of extinction. The saturated generalized linear model structure of "glm(*extinction ~*

198 *occupancy * occupancy_change* , family = binomial(link = 'logit'))," was used. The stepAIC()

199 function in the R package "MASS" (Ripley et al., 2013; version 7.3) was used to select the best-

200 fitting model containing some combination of these variables and their interaction term.

201 

## 2.6 Model performance and predictor importance

202 

203       We calculated the adjusted amount of deviance ($D^2$ of Guisan and Zimmermann, 2000)

204 accounted for by each computed logistic model. Deviance in a generalized linear model is

205 analogous to variance of ordinary linear regression. In each of the 16 data sets (four groups with

206 four time resolutions each), the Lindeman, Merenda, and Gold (1980, "LMG") indices of

207 correlated input relative importance (henceforth referred to as "relative importance") were

208 calculated for the *occupancy* and *occupancy change* terms with respect to predicting the

209 *extinction* term. This statistical approach was used to represent the explanatory power of each

210 model term with respect to one another, an insight that is not directly apparent with simple model

211 coefficients.

212

213 **2.7 Extinction probabilities of extant species**

214   The World Register of Marine Species (https://www.marinespecies.org/), with the

215 assistance of the R package "taxize" (Chamberlain and Szoecs, 2013; version 0.9.100), was used

216 to identify extant species. These data on extant taxa were downloaded on 25 September 2023.

217   In order to predict the extinction probabilities of extant species, the data sets were

218 reanalyzed and re-fit to models using only the extinct species. Although this technique reduced

219 the overall amount of data used to fit the model, it allowed for the prediction of extinction

220 probabilities of extant species without circularity. Other than removing extant species, all other

221 processes were carried out in the same way as described above.

222   After selecting the best model for each plankton group, that model was used to predict the

223 extinction probability of extant species. Using the fitted models along with the *occupancy* and

224 *occupancy change* values for each extant species in the present bin (that which ends at the

225 present, 0 Ma), a probability of the binary response variable occurring as a 1 (extinction) can be

226 calculated. This represents the probability that the species will *not* appear again during the next

227 time bin of the same length (that which begins at the present, 0 Ma) or in other future time bins.

228 Extinction predictions were made on extant species subsets without upper Pacman trimming, and

229 the average probability of extinction for all extant species was calculated in each data set.

230

**2.8 Robustness testing**

Further analyses tested the robustness of our results, specifically for the proportional

occupancy's utility as a metric of geographic range. The same analyses at a bin size of 1 million

years were carried out using latitudinal range and change of latitudinal range instead of

proportional occupancy and its change. Additionally, the same analyses were carried out using

proportional occupancy of Longhurst (2007) provinces and the change in proportional occupancy

of Longhurst provinces for data sorted into 1 million-year bins. Mixed-effect models, in which

each taxon was considered as a random-effect, were also constructed to check if species identity

substantially impacted the basic model results.

Although containing only records of planktonic foraminifers (many of which were

sourced from the NSB), the Triton database includes information on the original purpose of each

study from which records were sourced, as well as the age of speciation and extinction for each

species. With this additional information, the Triton dataset can be used to confirm the suitability

of methods used with the Neptune dataset with a different collection of fossil occurrences. Given

that some studies may not record every present taxon if it is not a zonal marker or thought to be

particularly informative, the Triton dataset was subset to include only studies whose purpose was

noted as "community analysis" (Fenton et al. 2021). Because studies whose purpose was to

analyze community structure would likely document all present species, by using this subset,

studies that potentially excluded some species were removed from the final data set.

Additionally, each included species history was subset to exclude any occurrences that occurred

before or after the speciation and extinction ages noted in the Triton dataset, respectively,

reducing the potential impact of reworked fossils in the analysis. Because each species in Triton

was trimmed in this manner, these data did not undergo Pacman profiling as the NSB data did.

254 After these additional data-cleaning actions were taken, the Triton dataset had 197,832 usable

255 occurrence records and was analyzed in the same way as the NSB data.

256

257 **3. Results**

258      The utilized datasets show all four groups with elevated extinction rates coming out of

259 the end-Cretaceous mass extinction and returning to relative stasis approximately 5-15 Ma after

260 the event (Fig. 1). All groups underwent decreases in diversity, corresponding with the

261 Paleocene-Eocene Thermal Maximum (PETM, Fig. S2). Radiolarians and diatoms show spikes

262 in extinction ($> 0.4$) during the Eocene, including at the Eocene-Oligocene transition as well as

263 at approximately 44 and 41 Ma. The extinction rates (Fig. 1) and diversity patterns (Fig. S2) of

264 each plankton group match those of previous analyses of Neptune data (Jamson et al., 2022), and

265 various biotic events in the Cenozoic, including the Eocene-Oligocene and the Oligocene-

266 Miocene transitions, can be detected.

267

268

269
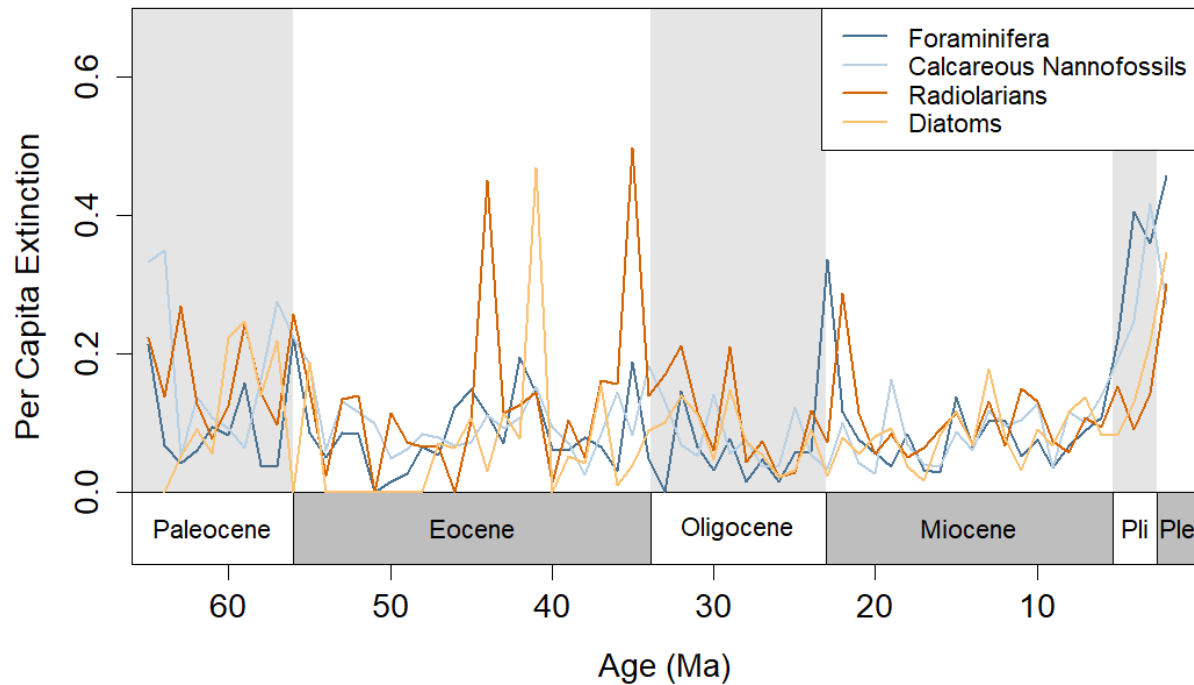
**Figure 1.** Per capita extinction rates calculated using the formula in Foote (1999) for each of the four taxonomic groups during the Cenozoic, calculated with 1-million-year bins. The timescale of Gradstein et al. (2012) was used here to match age assignments in the NSB.

*Occupancy* is statistically significant (i.e., statistically non-zero, *p = 0.05*) in all of the

analyzed combinations of taxonomic group and bin size. *Occupancy change* is significant in all

models except for calcareous nannofossils with a bin size of 0.2 Ma (Table 2). The term for the

interaction between occupancy and occupancy change is significant in all but three models:

foraminifera with bin size of 0.2 Ma, foraminifera with a bin size of 0.5 Ma, and diatoms with a

bin size of 1.0 Ma (Table 2).

| Occupancy Importance | Occupancy Change Importance | Interaction Importance | D² | Bin Size (Ma) | Group |
|---|---|---|---|---|---|
| 0.062 ± 0.010 | 0.052 ± 0.008 | 0.040 ± 0.007 | 0.155 | 1 | F |
| 0.050 ± 0.008 | 0.028 ± 0.006 | -- | 0.093 | 0.5 | F |
| 0.049 ± 0.007 | 0.014 ± 0.004 | -- | 0.062 | 0.2 | F |
| 0.036 ± 0.006 | 0.003 ± 0.001 | 0.002 ± 0.001 | 0.041 | 0.1 | F |
| 0.037 ± 0.006 | 0.040 ± 0.006 | 0.049 ± 0.007 | 0.126 | 1 | N |
| 0.045 ± 0.006 | 0.029 ± 0.005 | 0.029 ± 0.004 | 0.103 | 0.5 | N |
| 0.052 ± 0.005 | -- | 0.009 ± 0.002 | 0.072 | 0.2 | N |
| 0.050 ± 0.004 | 0.006 ± 0.002 | 0.002 ± 0.001 | 0.058 | 0.1 | N |
| 0.015 ± 0.003 | 0.032 ± 0.005 | 0.026 ± 0.004 | 0.073 | 1 | R |
| 0.019 ± 0.003 | 0.018 ± 0.003 | 0.013 ± 0.003 | 0.050 | 0.5 | R |
| 0.016 ± 0.003 | 0.014 ± 0.003 | 0.003 ± 0.001 | 0.033 | 0.2 | R |
| 0.013 ± 0.003 | 0.004 ± 0.002 | 0.001 ± 0.0004 | 0.019 | 0.1 | R |
| 0.017 ± 0.004 | 0.020 ± 0.005 | -- | 0.037 | 1 | D |
| 0.012 ± 0.003 | 0.011 ± 0.003 | 0.002 ± 0.001 | 0.025 | 0.5 | D |
| 0.007 ± 0.002 | 0.010 ± 0.003 | 0.003 ± 0.001 | 0.020 | 0.2 | D |
| 0.002 ± 0.001 | 0.007 ± 0.002 | 0.004 ± 0.001 | 0.014 | 0.1 | D |

**Table 2.** Relative importance values shown with standard error for the *occupancy*, *occupancy change*, and interaction term for each group at each bin size. The D² values for each model are also reported. Darker blue corresponds to higher values. F = foraminifera, N = calcareous nannofossils, R = radiolarians, D = diatoms.

The signs of the *occupancy* and the *occupancy change* coefficients are always negative (Table 3), meaning that smaller instantaneous geographic range sizes and more-negative changes in geographic range size both correspond to larger extinction probabilities (Fig. 2). The $D^2$ values of each analyzed model combination set are shown in Table 2. The maximum $D^2$ for any model is 0.155, occurring for the saturated foraminifera model at a bin size of 1.0 Ma . The relative importance of the *occupancy change* term averages 0.019, or about 41% of the total explanatory power shared by *occupancy* and *occupancy change*, with higher values being achieved with larger bin sizes.

| Occupancy | Occupancy Change | Interaction | Bin Size (Ma) | Group |
|---|---|---|---|---|
| -4.975 | -0.624 | -7.860 | 1 | F |
| -4.419 | -0.700 | -- | 0.5 | F |
| -4.863 | -0.463 | -- | 0.2 | F |
| -4.406 | -0.389 | 2.299 | 0.1 | F |
| -3.275 | -0.393 | -11.363 | 1 | N |
| -4.080 | -0.343 | -9.593 | 0.5 | N |
| -4.576 | -- | -4.313 | 0.2 | N |
| -4.516 | -0.522 | 1.970 | 0.1 | N |
| -2.063 | -0.711 | -4.907 | 1 | R |
| -2.484 | -0.598 | -2.858 | 0.5 | R |
| -2.138 | -0.922 | 2.021 | 0.2 | R |
| -1.902 | -0.513 | 1.443 | 0.1 | R |
| -2.095 | -0.627 | -- | 1 | D |
| -2.028 | -0.734 | 2.096 | 0.5 | D |
| -1.494 | -0.830 | 2.139 | 0.2 | D |
| -0.918 | -0.799 | 2.036 | 0.1 | D |

**Table 3.** Model coefficients for the occupancy, occupancy change, and interaction term for each group and bin size. Darker blue corresponds to more positive values, and darker red corresponds to more negative values. F = foraminifera, N = calcareous nannofossils, R = radiolarians, D = diatoms.
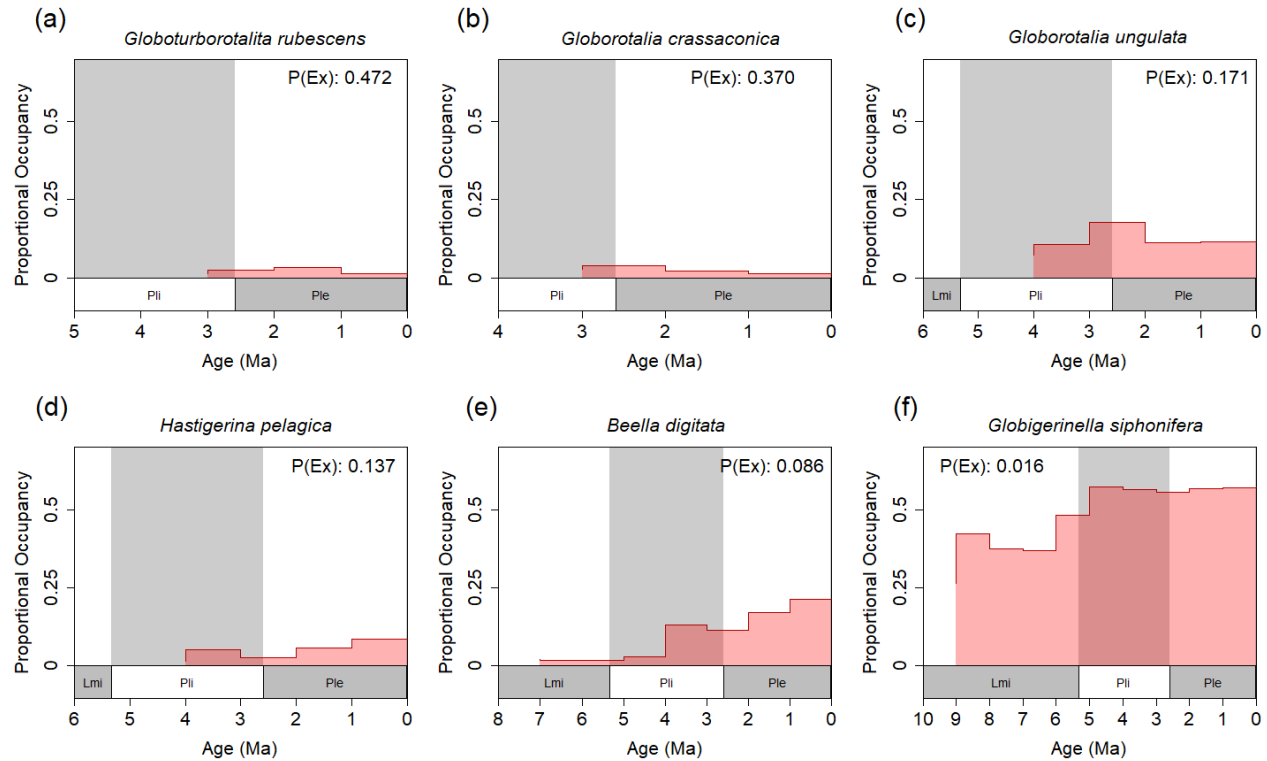
**Figure 2.** A selection of proportional occupancy through time plots for 6 species of foraminifers sourced from the NSB with one-million-year bin size. In each panel, the current extinction probability of that species—predicted for the next 1 million years using that species' historical geospatial records— is shown. Panels A through F are ordered according to decreasing current extinction probability. Notice the association of relatively small standing occupancy values and relatively large occupancy decreases with increased probability of extinction. "Lmi" = Late Miocene, "Pli" = Pliocene, "Ple" = Pleistocene.

$D^2$ values and relative importance values of the *occupancy change* term increase systematically with bin size, and calcareous organisms tend to have higher values than siliceous organisms for both metrics (Table 2), indicating better overall explanatory power in calcareous organisms. The maximum relative importance value of *occupancy* is 0.062, and the maximum relative importance value of *occupancy change* is 0.052, both of which were reported for the foraminifera data set with a time bin of 1.0 Ma (the $D^2$ of this particular saturated model is 0.155). Not surprisingly, sampling completeness increases with larger temporal grain. Foraminifera and calcareous nannofossils have consistently higher three-timer sampling

328    completeness than the siliceous groups across all bin sizes (Table 4, see Table S3 for SCM

329    completeness).

330
331

| Three-Timer Completeness | | | | |
|---|---|---|---|---|
| Bin Size (million years) | **0.1** | **0.2** | **0.5** | **1.0** |
| Foraminifera | $0.869 \pm 0.005$ | $0.921 \pm 0.005$ | $0.957 \pm 0.006$ | $0.949 \pm 0.009$ |
| Calcareous Nannofossils | $0.891 \pm 0.004$ | $0.918 \pm 0.005$ | $0.942 \pm 0.006$ | $0.950 \pm 0.008$ |
| Radiolarians | $0.857 \pm 0.005$ | $0.879 \pm 0.006$ | $0.922 \pm 0.007$ | $0.947 \pm 0.008$ |
| Diatoms | $0.830 \pm 0.006$ | $0.867 \pm 0.007$ | $0.898 \pm 0.008$ | $0.909 \pm 0.010$ |

332    **Table 4.** Three-timer completeness scores (Alroy 2008) calculated for each full data set at each
333    of the four examined bin sizes. Shown with 95% confidence intervals.
334
335

336         When fit to extinct-only data subsets, AIC-selected models still retain both the *occupancy*

337    and the *occupancy change* term for all groups at all bin sizes. Extant diatoms show a

338    significantly higher probability of extinction than the two calcareous groups, and extinct

339    radiolarians have a significantly lower probability of extinction than calcareous nannofossils and

340    diatoms (Table 5).

341

342

343

344

345

346

347

| | Species | |
|---|---|---|
| | **Extant** | **Extinct** |
| **Foraminifera** | 0.130 ± 0.049 | 0.235 ± 0.033 |
| **Calcareous Nannofossils** | 0.070 ± 0.044 | 0.262 ± 0.016 |
| **Radiolarians** | 0.177 ± 0.014 | 0.208 ± 0.012 |
| **Diatoms** | 0.191 ± 0.010 | 0.269 ± 0.020 |

348  **Table 5.** Mean of all extant-organism extinction probabilities (left) and
349  extinct-organism extinction probabilities (right) produced for each data
350  set. Means are shown with 95% confidence interval.
351

352      The raw number of cells occupied by each species during each time bin shows a strong

353  positive correlation with the number of occupied Longhurst provinces for each taxonomic group

354  at each bin size, even when autocorrelation is removed (Table S4). The same analyses conducted

355  on proportional occupancy of Longhurst provinces and the change in proportional occupancy of

356  Longhurst provinces, as well as on the latitudinal range of species and the change in latitudinal

357  range of species, yield best-fit models that retain both variables and their interaction term (Tables

358  S5 and S6). When the logistic modeling is applied to the Triton data, both the *occupancy* and the

359  *occupancy change* terms are retained across the three largest bin sizes (Table S9). The maximum

360  $D^2$ value of 0.072 is achieved with a bin size of 1.0 Ma, and the LMG values of *occupancy* and

361  *occupancy change* are 0.050 and 0.016, respectively. Tests for multicollinearity show only minor

362  correlation (mean = 0.23, maximum = 0.32) between the examined variables (Table S7).

363  Although several datasets failed to reach convergence, the mixed-effect models show similar

364  results to those reported as part of the main analysis (Table S8).

365

366

**4. Discussion**

**4.1 Geographic range as a driver of microplankton diversity**

While absolute geographic range size is an informative predictor of extinction risk, various other factors relating to geographic range also play an important role in global biodiversity patterns. Powell and Glazier (2017) found that in the same four groups of microplankton analyzed in this study, latitudinal diversity gradients are produced by asymmetric shifts in geographic range, rather than variations in diversification rate with respect to latitude. On the contrary, Raja and Kiessling (2021) found that the extratropics had higher average origination rates than the tropics. Supporting Powell and Glazier's (2017) finding that asymmetric shifts in geographic range are key drivers of latitudinal diversity gradients, Raja and Kiessling (2021) showed that dispersal was more likely to occur from the extratropics towards the tropics. Both studies suggest that latitudinal diversity gradients, and thus the geographic distribution of a species, are closely linked to paleoclimate regimes. Indeed, changes in global circulation patterns, water column stratification, and temperature are all among the major influences on global plankton diversity (Lowery et al., 2020)

Although the ranges of marine microplankton have been known to shift in response to climate (Ying et al., 2024, Chaabane et al., 2024), this is not always the case, as species ranges sometimes fail to keep up with shifting temperature zones (Trubovitz et al., 2020). Trubovitz et al. (2023) found that radiolarian abundance is not a significant predictor of extinction risk and that external drivers (such as climate) are more likely to predict extinctions. Thus, while some species do migrate in response to climate change, larger geographic ranges may provide a geographic cushion to species that don't: as local temperatures change, more widespread species undergo a more drawn-out series of local extirpations before global extinction occurs. This agrees with the well-established phenomenon, and that which we also report here, that larger

391  instantaneous geographic occupancy reduces a species' risk of extinction (Foote et al., 2016;

392  Foote et al., 2007; McKinney, 1997; Payne and Finnegan, 2007; Purvis et al., 2000; Staude et al.,

393  2020).

394       Additionally, the trajectory of a species' geographic range through time might indirectly

395  reflect shifts in regional or global climate. As paleoclimate zones shift, geographic cells may

396  become inhospitable to a species, and the species may undergo extirpation in that geographic

397  cell. A more rapid change in a species' occupancy through time may reflect a more rapid change

398  in paleoclimate and hospitable regions of Earth. Continued reduction in occupancy over time can

399  thus provide insight into the effects of long-term climatic, geographic, or biological trends on the

400  extinction probability of marine microplankton.

401

402  **4.2 History of occupancy / legacy effects**

403       The trajectory of various ecological variables through time has been shown to impact the

404  current and future direction of species diversity trends. These legacies may include past climatic

405  events or geographic range shifts influencing modern distributions or extinctions of species

406  (Svenning et al., 2015). The interaction of historic information with current information can

407  provide insight about ecological processes that neither historic nor current information could

408  provide on its own.

409       The historic trajectory of climate change impacts the probability of extinction occurring

410  with a short-term change in climate. A warming event occurring after a long-term warming trend

411  leads to greater extinction rates (Mathes et al., 2021) than a warming event occurring after a

412  long-term cooling trend. Understanding the historical conditions leading up to a study period of

413  interest may thus be essential to understanding the key drivers as to what goes extinct versus

414  what survives.

415     Although the effect of climate and geographic range legacies on instantaneous

416     geographic range is well-studied (Svenning et al., 2015), the effect of geographic range legacies

417     on instantaneous extinction probability has not received as much attention. Of course,

418     populations of species cannot "look" backwards, but are instead influenced by the current

419     conditions present in an environment. The predictive capability of the *occupancy change* term

420     may thus be an indicator of continued unfavorable conditions (perhaps spanning millions of

421     years) acting on a population at a given time. Kiessling and Kocsis (2016) found that the legacy

422     of geographic range (represented as its change to the present from the previous bin) is an

423     informative predictor for extinction risk in marine macroinvertebrates. Our results build upon

424     those of Kiessling and Kocsis (2016), demonstrating both that these findings hold true for marine

425     microplankton, as well as that temporal scale (bin size) is a key variable in detecting the

426     importance of geographic legacy effects.

427

428     **4.3 Scale dependency of extinction drivers:**

429     Although previous studies have analyzed various drivers of extinction through geologic

430     time, relatively little research has gone into understanding the scale-dependency of these

431     extinction drivers. Scale-dependency in extinction studies manifests in various variables, such as

432     area (Fagan et al., 2005; Guardiola et al., 2013) or taxon age (Henao Diaz et al., 2019).

433     Analyzing data at different temporal scales is also imperative to detect true ecological signals

434     (Hewitt et al., 2010). We find that as temporal resolution decreases (bin size increases), the

435     relative importance of both the *occupancy* and *occupancy change* variables increases (Table 2).

436     This could result from there being more records in a single temporal bin as bin size

437     increases, thus increasing statistical power. With larger bin sizes, it is easier to detect biological

438     signals that may otherwise be lost in the noise of fragmentary data. We show here that the

439    seemingly-arbitrary selection of temporal bin size can have major impacts on conclusions drawn

440    about microplankton diversification, and that coarser resolutions may more reliably indicate

441    actual macroevolutionary trends.

442

443    **4.4 Calcareous vs siliceous microfossils**

444          In general, the explanatory power of each of the model terms is smaller in the siliceous

445    groups than in the calcareous groups. Although *occupancy* and *occupancy change* were found to

446    be informative across all groups, the signals are weaker in diatoms and radiolarians (Table 2,

447    Table 3). This discrepancy likely results from minor variations in sampling, as evidenced by

448    lower 3-timer completeness values for the two siliceous groups. The difference may also be a

449    result of variable taphonomic pathways between the calcareous and siliceous organisms

450    (Boltovskoy 1994). Nonetheless, both *occupancy* and *occupancy change* are important predictors

451    of extinction regardless of the group, and these findings further underscore the importance of

452    accounting for sampling when analyzing paleontological data.

453

454    **4.5 Robustness testing**

455          There is a strong correlation between the number of occupied Longhurst provinces and

456    the number of individual occupied geographic cells for each species-bin pairing. This

457    demonstrates that, although the locations of the various drilling expeditions that sourced much of

458    the data in the Neptune database are not entirely random, when taken together they still account

459    for a diverse spread of planktonic biogeographic regions around the globe. This supports the idea

460    that the collection of data contained in the Neptune database is comprehensive enough to study

461    large-scale biogeographic trends. Additionally, AIC-selected models contained both the

462    *occupancy* and *occupancy change* terms even when geographic range was measured as

463    latitudinal expanse or as a proportion of occupied Longhurst provinces (Tables S5 and S6). This

464    suggests that the significance of proportional occupancy change in predicting extinction is not

465    merely an artifact of data processing.

466      The AIC-selected model for each bin size in the Triton data set always retains both

467    *occupancy* and *occupancy change* as significant except with a bin size of 0.1 Ma (Table S9).

468    Although the Triton dataset has substantially more occurrence records after preprocessing, it has

469    consistently lower diversity compared to the other taxonomic groups from the NSB (Fig. S2).

470    This could indicate a greater propensity for "lumping" in the Triton dataset than in the NSB,

471    which in turn could change how spatio-temporal signals manifest. The similar results obtained

472    from the Triton dataset further confirm the suitability of these methods with an alternative data

473    set and reaffirm the importance of *occupancy* and *occupancy change* when modeling extinction.

474      Taken together, our findings suggest that the change in geographic occupancy is an

475    important metric for predicting extinction across marine life. Kiessling and Kocsis (2016) looked

476    exclusively at skeletal macroinvertebrates, whereas we here analyze several protist lineages of

477    marine plankton. The broad taxonomic scope of these findings emphasizes the fundamental

478    importance of the trajectory in geographic range as a biological metric, which can be a key

479    aspect of taxon dynamics through time. Although the explanatory power of the model may seem

480    low (up to 15.5%), it is an important factor given the many other variables that influence

481    extinction risk (McKinney 1997)

482

483    **4.6 Future Perspectives**

484      Although modern studies can track geographic occupancy change over the course of

485    decades (if there is a history of consistent data collection), estimates of marine species durations

486    average between 5-10 million years (Foote and Raup, 1996; Raup, 1991), much longer than

487  human-collected records can encompass. To fully understand the change of occupancy through a

488  species' duration, records extending beyond those which could have been manually recorded by

489  conservation biologists are needed. Although some modern conservation practitioners have been

490  hesitant to fully embrace long-term paleontological data, this study provides yet another

491  argument for the incorporation of historical perspectives and fossil evidence in conservation

492  efforts (Dietl et al., 2019; Kiessling et al., 2019; Smith et al., 2018).

493       While for simplicity's sake, this study only looked at the interaction of *occupancy* and the

494  first degree of *occupancy change* (bin number *i* to *i*-1), future iterations could incorporate entire

495  occupancy histories into model fitting using even more advanced techniques. This may help the

496  model overcome variations in sampling intensity or localized paleoenvironmental events, and let

497  the models provide information not only on decline, but information on *continued* decline —

498  another hallmark of increased extinction risk.

499

500  **5.  Conclusion**

501       In providing evidence that the geological history of species distributions plays a

502  significant role in species extinction risk, our study demonstrates the importance of

503  paleontological data for assessing modern species extinction risk. These findings provide

504  empirical support for the connection between continued range reduction and ultimate global

505  extinction in marine microplankton. We also demonstrate here the importance of temporal grain

506  in detecting biological signal in fragmentary fossil data.

507

508  **Data availability statement**

509  All data and code is currently accessible at the reserved stable repository DOI:
510  10.5281/zenodo.15174296 (Smith, 2025).
511

## Author contributions

WK and AK developed the conceptual framework. IS constructed the analytical pipeline and carried out analyses and drafted manuscript. All authors contributed to the development of the manuscript.

## Competing interests

The authors declare that they have no conflict of interest.

## Acknowledgements

The study was supported by the Deutsche Forschungsgemeinschaft (Ko 5382/2-1) and was embedded in the Research Unit TERSANE (FOR 2332). The authors would like to thank Johan Renaudie for ongoing assistance with the Neptune Sandbox Berlin, as well as for providing insightful comments and suggestions.

## References

Alroy, J. (2008). Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences*, *105*(supplement_1), 11536-11542.

Benton, M. J. (1985). Mass extinction among non-marine tetrapods. *Nature*, *316*(6031), 811-814.

Boltovskoy, D. (1994). The sedimentary record of pelagic biogeography. *Progress in Oceanography*, *34*(2-3), 135-160.

Boyden, J. A., Müller, R. D., Gurnis, M., Torsvik, T. H., Clark, J. A., Turner, M., Ivey-Law, H., Watson, R. J., & Cannon, J. S. (2011). Next-generation plate-tectonic reconstructions using GPlates, *in* Baru, C., and Keller, G. R., eds., Geoinformatics: Cyberinfrastructure for the Solid Earth Sciences: Cambridge, Cambridge University Press, p. 95-114.

Chaabane, S., de Garidel-Thoron, T., Meilland, J., Sulpis, O., Chalk, T. B., Brummer, G. J. A., Mortyn, P.G., Giraud, X., Howa, H., Casajus, N., Kuroyanagi, A., Beaugrand, G., & Schiebel, R. (2024). Migrating is not enough for modern planktonic foraminifera in a changing ocean. *Nature*, 1-7.

Chamberlain, S., Szocs, E. (2013). "taxize - taxonomic search and retrieval in R." *F1000Research*. https://f1000research.com/articles/2-191/v2.

Darroch, S. A., Saupe, E. E., Casey, M. M., & Jorge, M. L. (2022). Integrating geographic ranges across temporal scales. *Trends in Ecology & Evolution*.

Dietl, G. P., Smith, J. A., & Durham, S. R. (2019). Discounting the past: the undervaluing of paleontological data in conservation science. *Frontiers in Ecology and Evolution*, 108.

Fagan, W. F., Aumann, C., Kennedy, C. M., & Unmack, P. J. (2005). Rarity, fragmentation, and the scale dependence of extinction risk in desert fishes. *Ecology*, *86*(1), 34-41.

Fenton, I. S., Woodhouse, A., Aze, T., Lazarus, D., Renaudie, J., Dunhill, A. M., Young, J. R., & Saupe, E. E. (2021). Triton, a new species-level database of Cenozoic planktonic foraminiferal occurrences. *Scientific Data*, *8*(1), 160.

553 Foote, M., & Raup, D. M. (1996). Fossil preservation and the stratigraphic ranges of
554     taxa. *Paleobiology*, *22*(2), 121-140.
555 Foote, M., Crampton, J. S., Beu, A. G., Marshall, B. A., Cooper, R. A., Maxwell, P. A., &
556     Matcham, I. (2007). Rise and fall of species occupancy in Cenozoic fossil
557     mollusks. *Science*, *318*(5853), 1131-1134.
558 Foote, M., Ritterbush, K. A., & Miller, A. I. (2016). Geographic ranges of genera and their
559     constituent species: structure, evolutionary dynamics, and extinction resistance.
560     *Paleobiology*, *42*(2), 269-288.
561 Foote, M. (1999). Morphological diversity in the evolutionary radiation of Paleozoic and post-
562     Paleozoic crinoids. *Paleobiology*, *25*(S2), 1-115.
563 Gradstein, F.M., Ogg, J.G., Schmitz, M.D., and Ogg, G.M. (2012). The Geologic Time Scale.
564     Elsevier, Amsterdam.
565 Guardiola, M., Pino, J., & Rodà, F. (2013). Patch history and spatial scale modulate local plant
566     extinction and extinction debt in habitat patches. *Diversity and Distributions*, *19*(7), 825-
567     833.
568 Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology.
569     *Ecological modelling*, *135*(2-3), 147-186.
570 Henao Diaz, L. F., Harmon, L. J., Sugawara, M. T., Miller, E. T., & Pennell, M. W. (2019).
571     Macroevolutionary diversification rates show time dependency. *Proceedings of the*
572     *National Academy of Sciences*, *116*(15), 7403-7408.
573 Hewitt, J. E., Thrush, S. F., & Lundquist, C. (2010). Scale-dependence in ecological systems.
574     *Encyclopedia of Life Sciences (ELS)*, 1-7.
575 Huber, B. T., Petrizzo, M. R., Young, J. R., Falzoni, F., Gilardoni, S. E., Bown, P. R., & Wade,
576     B. S. (2016). Pforams@ microtax. *Micropaleontology*, *62*(6), 429-438. Jamson, K. M.,
577     Moon, B. C., & Fraass, A. J. (2022). Diversity dynamics of microfossils from the
578     Cretaceous to the Neogene show mixed responses to events. *Palaeontology*, *65*(4),
579     e12615.
580 Jamson, K. M., Moon, B. C., & Fraass, A. J. (2022). Diversity dynamics of microfossils from the
581     Cretaceous to the Neogene show mixed responses to events. *Palaeontology*, *65*(4),
582     e12615.
583 Kiessling, W., & Kocsis, Á. T. (2016). Adding fossil occupancy trajectories to the assessment of
584     modern extinction risk. *Biology letters*, *12*(10), 20150813.
585 Kiessling, W., Raja, N. B., Roden, V. J., Turvey, S. T., & Saupe, E. E. (2019). Addressing
586     priority questions of conservation science with palaeontological data. *Philosophical*
587     *Transactions of the Royal Society B*, *374*(1788), 20190222.
588 Kiessling, W. (2005). Habitat effects and sampling bias on Phanerozoic reef
589     distribution. *Facies*, *51*(1-4), 24-32.
590 Kocsis, Á. (2020). icosa: global triangular and penta-hexagonal grids based on tessellated
591     icosahedra, R package version 0.10. 0. In.
592 Kocsis, Á. T., Reddin, C. J., Alroy, J., & Kiessling, W. (2019). The R package divDyn for
593     quantifying diversity dynamics using fossil sampling data. *Methods in Ecology and*
594     *Evolution*, *10*(5), 735-743.
595 Lazarus, D., Barron, J., Renaudie, J., Diver, P., & Türke, A. (2014). Cenozoic planktonic marine
596     diatom diversity and correlation to climate change. *PLoS One*, *9*(1), e84857.
597 Lazarus, D., Weinkauf, M., & Diver, P. (2012). Pacman profiling: a simple procedure to identify
598     stratigraphic outliers in high-density deep-sea microfossil data. *Paleobiology*, *38*(1), 144-
599     161.

Lazarus, D. (1994). Neptune: a marine micropaleontology database. *Mathematical Geology*, *26*, 817-832.

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). Introduction to bivariate and multivariate analysis. *Scott, Foresman & Company Glenview, IL*.

Liow, L. H., Skaug, H. J., Ergon, T., & Schweder, T. (2010). Global occurrence trajectories of microfossils: environmental volatility and the rise and fall of individual species. *Paleobiology*, *36*(2), 224-252.

Longhurst, A. R. (2007). *Ecological geography of the sea*. Elsevier.

Lowery, C. M., Bown, P. R., Fraass, A. J., & Hull, P. M. (2020). Ecological response of plankton to environmental change: thresholds for extinction. *Annual Review of Earth and Planetary Sciences*, *48*(1), 403-429.

Mace, G. M., Collar, N. J., Gaston, K. J., Hilton-Taylor, C., Akçakaya, H. R., Leader-Williams, N., Milner-Gulland, E. J., & Stuart, S. N. (2008). Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation biology*, *22*(6), 1424-1442.

Mathes, G. H., van Dijk, J., Kiessling, W., & Steinbauer, M. J. (2021). Extinction risk controlled by interaction of long-term and short-term climate change. *Nature Ecology & Evolution*, *5*(3), 304-310.

McKinney, M. L. (1997). Extinction vulnerability and selectivity: combining ecological and paleontological views. *Annual review of ecology and systematics*, *28*(1), 495-516.

Nigrini, C., Sanfilippo, A., & Moore Jr., T. C. (2006). Cenozoic radiolarian biostratigraphy: a magnetobiostratigraphic chronology of Cenozoic sequences from ODP Sites 1218, 1219, and 1220, Equatorial Pacific. Proceedings of the Ocean Drilling Program, Scientific Results 199:1–76.

Payne, J. L., & Finnegan, S. (2007). The effect of geographic range on extinction risk during background and mass extinction. *Proceedings of the National Academy of Sciences*, *104*(25), 10506-10511.

Powell, M. G., & Glazier, D. S. (2017). Asymmetric geographic range expansion explains the latitudinal diversity gradients of four major taxa of marine plankton. *Paleobiology*, *43*(2), 196-208.

Purvis, A., Gittleman, J. L., Cowlishaw, G., & Mace, G. M. (2000). Predicting extinction risk in declining species. *Proceedings of the royal society of London. Series B: Biological Sciences*, *267*(1456), 1947-1952.

Raja, N. B., & Kiessling, W. (2021). Out of the extratropics: The evolution of the latitudinal diversity gradient of Cenozoic marine plankton. *Proceedings of the Royal Society B*, *288*(1950), 20210545.

Raup, D. M. (1991). A kill curve for Phanerozoic marine species. *Paleobiology*, *17*(1), 37-48.

R Core Team (2022). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Renaudie, J., Lazarus, D. B., & Diver, P. (2020). NSB (Neptune Sandbox Berlin): An expanded and improved database of marine planktonic microfossil data and deep-sea stratigraphy. *Palaeontologia Electronica*, *23*(1), 1-28.

Renaudie, J. (2019). plannapus/NSBcompanion: NSBcompanion 2.1 (v2.1.1). Zenodo. https://doi.org/10.5281/zenodo.3408190.

644    Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B.
645        (2013). Package 'mass'. *Cran r*, *538*(113-120), 822.
646    Saulsbury, J. G., Parins-Fukuchi, C. T., Wilson, C. J., Reitan, T., & Liow, L. H. (2023). Age-
647        dependent extinction and the neutral theory of biodiversity. *Proceedings of the National*
648        *Academy of Sciences*, *121*(1), e2307629121.
649    Smith, I. (2025). Preprint -- Occupancy Trajectory Microplankton. DOI:
650        10.5281/zenodo.15174296. Zenodo.
651    Smith, J. A., Durham, S. R., & Dietl, G. P. (2018). Conceptions of long-term data among marine
652        conservation biologists and what conservation paleobiologists need to know. *Marine*
653        *conservation paleobiology*, 23-54.
654    Staude, I. R., Navarro, L. M., & Pereira, H. M. (2020). Range size predicts the risk of local
655        extinction from habitat loss. *Global Ecology and Biogeography*, *29*(1), 16-25.
656    Strack, T., Jonkers, L., C. Rillo, M., Baumann, K. H., Hillebrand, H., & Kucera, M. (2024).
657        Coherent response of zoo-and phytoplankton assemblages to global warming since the
658        Last Glacial Maximum. *Global Ecology and Biogeography*, *33*(6), e13841.
659    Svenning, J. C., Eiserhardt, W. L., Normand, S., Ordonez, A., & Sandel, B. (2015). The
660        influence of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annual*
661        *Review of Ecology, Evolution, and Systematics*, *46*(1), 551-572.
662    Swain, A., Woodhouse, A., Fagan, W. F., Fraass, A. J., & Lowery, C. M. (2024). Biogeographic
663        response of marine plankton to Cenozoic environmental changes. *Nature*, 1-8.
664    Tietje, M., & Kiessling, W. (2013). Predicting extinction from fossil trajectories of geographical
665        ranges in benthic marine molluscs. *Journal of Biogeography*, *40*(4), 790-799.
666    Trubovitz, S., Lazarus, D., Renaudie, J., & Noble, P. J. (2020). Marine plankton show threshold
667        extinction response to Neogene climate change. *Nature communications*, *11*(1), 5069.
668    Trubovitz, S., Renaudie, J., Lazarus, D., & Noble, P. J. (2023). Abundance does not predict
669        extinction risk in the fossil record of marine plankton. *Communications Biology*, *6*(1),
670        554.
671    Ying, R., Monteiro, F. M., Wilson, J. D., Ödalen, M., & Schmidt, D. N. (2024). Past
672        foraminiferal acclimatization capacity is limited during future warming. *Nature*, 1-5.
673