Marine biodiversity and geographic distributions are independent on large scales

Gawain T. Antell[1,4*], Wolfgang Kiessling[2], Martin Aberhan[3], Erin E. Saupe[1*]

[1]Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, UK

[2]Section Paleobiology, Friedrich-Alexander Universität Erlangen-Nürnberg, Loewenichstraße 28, 91054 Erlangen, DE

[3]Museum für Naturkunde, Invalidenstraße 43, 10115 Berlin, DE

[4]Lead Contact

***Correspondence:** gawainantell@gmail.com and erin.saupe@earth.ox.ac.uk

SUMMARY

Fundamental ecological and evolutionary theories such as community saturation and diversity-dependent diversification assume that biotic competition restricts resource use, and thence limits realized niche breadth and geographic range size [1-3]. This principle is called competitive exclusion. The corollary, ecological release, posits that after competitors disappear from a region, species that were previously excluded will invade. Hundreds of field experiments have demonstrated ecological release in living populations, but few of these studies were conducted in marine environments, and almost no work extended beyond 10 years and 1,000 km^2 [4, 5]. In limited investigation of marine taxa at larger spatio-temporal scales, macroecologists and paleobiologists have observed little evidence of competitive exclusion [6-9]. Here, we quantified spatial trends in the rich and densely-sampled fossil history of brachiopods and bivalves, while accounting for inconsistent sampling coverage through time using a new method of spatial standardization. The number of potential competitors in a region did not explain the geographic

distribution of constituent species or genera. Furthermore, although ecological release predicts species to expand after extinction events, survivors of intervals with net species loss expanded as rarely as species in other intervals. Regression model estimates indicated different spatial responses of brachiopods and bivalves, and of habitat specialists and generalists, but no effect from changes in number of potential competitors. Biotic competition may control the distribution of populations, but on larger spatio-temporal scales, non-competitive factors may have driven biogeographic patterns of brachiopods and bivalves.

RESULTS AND DISCUSSION

We performed three sets of analyses to test species' geographic responses to inferred competitive pressure. The first approach tested the degree to which mean geographic range size of species in regional marine assemblages varied inversely with number of species in those regions. This analysis investigated trends across all taxa of bivalves and brachiopods that might have interacted in a given region, an "assemblage approach" that produced continuous time series of central tendencies. We also analyzed data in a "taxon-tracking approach:" each datum represented a single species that was traced through successive stages. We tested whether species that survived intervals of net species loss expanded in distribution compared to species from intervals of background extinction rates, and we modeled each species' change in geographic distribution through time as a function of environmental affinity, clade membership, and change in the number of species in a region.

All analyses focused on marine bivalves and brachiopods—diverse and well-sampled clades that have figured prominently in discussion of how competition influences macroevolution [e.g. 10]. Each of the three analyses was applied to rasterized occurrences of

brachiopod and bivalve species compiled from the global Paleobiology Database [1]. We repeated the assemblage-based time series analysis on an independent, finer-resolution dataset of bivalve species in the northwest Atlantic. Members of Bivalvia and Brachiopoda overlap in body size, trophic position, and preferences for substrates and habitats. Therefore, species from these groups are candidate competitors, especially for substrate availability, even though the clades differ physiologically and diverged before the Early Cambrian [11, 12]. Our study does not assume that all co-existing species competed simultaneously for space or other resources, or that competition was stronger between than within clades. Rather, because our global dataset includes species that occupied a variety of habitats, we expect that subsets of species competed in the living community, regardless of whether these competitors can be identified individually.

We converted occurrences to presence or absence values for grid cells of standard area. Our novel subsampling procedure controlled both total sampling area and dispersion of sampling sites through time (Figures S1 and S2; Methods S1). Although spatial subsampling compresses the range of observed values in a similar way to classical rarefaction, power analysis indicated a low likelihood of type 2 error (Figure S3). We calculated geographic range size both as the number of equal-area grid cells in which a species was observed ("occupancy"), and as the length of the minimum spanning tree connecting occupied grid cell centroids [13]. Occupancy directly reflects area on a map, whereas tree length is a continuous variable that accounts for dispersion.

We parameterized competitive pressure in terms of the number of species in an assemblage, to consider the cumulative pressure from many species on any focal species. Taxon count has been treated as a proxy of competitive pressure in recent theoretical work [14] and in previous empirical work on brachiopods and bivalves [10, 15] and other taxa [16]. Co-

occurrence frequencies and temporal replacement, alternative metrics for competitive pressure often used for fossil data, have power only to detect extreme, pairwise cases of exclusion [6-8]. Common metrics used in modern ecology, such as changes in abundance or experimental removal of competitor populations (examples in [5]), are difficult or impossible to apply to the fossil record.

The reliability of species count as a metric of competitive pressure could depend on the degree of niche overlap in a study ecosystem, a property that may have differed between Paleozoic and post-Paleozoic marine fauna [17]. However, analyses herein considered many time bins and spatial regions so as to determine dominant patterns across the Phanerozoic; we also replicated analysis on fine-scale spatio-temporal data in the short and ecologically complex interval of the Neogene–Quaternary. Ultimately, the particular biotic parameter to which species respond most strongly (e.g. abundance of the closest competitor *versus* total number of competitors) remains unknown.

Time Series Analysis

We calculated the correlation between replicate time series of species count and geographic range size in an assemblage drawn at random from each time bin. If ecological release occurred on the spatio-temporal scale of analysis, one would expect a negative correlation between time series of species count and geographic range size. Instead, the estimated relationship was too close to null to consider the signal to be either positive or negative (Figure 1; Table S3). We plotted cross-correlation functions at lags of one to five time bins to test for delayed ecological responses, but the patterns were congruent with the simultaneous comparisons (i.e. lag zero). Results from the finer-grain northwest Atlantic dataset also concurred (Figure 2; Table S3), as did results from sensitivity analyses—for instance where

global data were subsampled more dispersedly or within smaller regions, treated with classical or coverage-based richness rarefaction, or analyzed at the level of genera (Tables S2 and S3; Methods S1 section 2.3).

Biotic Crises *versus* Background Intervals

We tracked individual species across stage boundaries to test whether taxa expanded their distributions after biotic crises in comparison to periods of background extinction dynamics. Geographic behavior varied considerably both among species within each interval and in the mean species' trend among intervals; however, the central tendency of species' geographic range size changes was stasis (Table S4). Expansion and contraction in species' geographic ranges were distributed evenly across intervals of increases and decreases in species count, regardless of geographic range metric (Figure 3). This result is incongruent with a hypothesis of ecological release, which would have predicted species that survived the extinction of competitors to expand into newly vacated geographic space.

Mixed-Effects Models

Change in species count did not predict change in geographic range size of species tracked through time. The three highest-ranking linear mixed-effects models for minimum spanning tree length change did not include species count as a predictor (Tables S5). In all models that did use species count as a predictor variable, including highly-ranked models for occupancy, the confidence interval estimate included zero (Figure 4, Tables S6 and S7). Thus, regression analysis agreed with the other two analyses in that it did not provide evidence for ecological release.

In contrast to change in species count, a strongly influential predictor of species' distribution dynamics was the point in time in a lineage (e.g. mid existence) at which geographic

ranges were measured. Evidence for the strength of this variable includes the fact that it was the only predictor to appear in every non-trivial model (Table S5). Species expanded after appearing in the fossil record and contracted prior to disappearing from the record (Figure 4, Tables S6 and S7). Hence, our results corroborate previous descriptions of waxing and waning through time [18-20]. Additionally, regression models suggested that brachiopod species expanded in geographic range more than did bivalves, provided suitable habitat; generalists for substrate type expanded more than did specialists; and, surprisingly, water depth specialists expanded more than did generalists. However, the clarity of these results varied with the metric of geographic range size (Figure 4, Tables S5–S7).

Debate has raged for centuries [12] on whether bivalves displaced brachiopods competitively [15, 21] or diversified independently [10, 11, 22]. Our study did not intend to answer this question explicitly. However, results from the mixed-effects models provide information that bears on the issue tangentially. Unexpectedly, the model predicted brachiopods, not bivalves, to expand more in spatial distribution when suitable habitat is present. This result seems particularly anomalous with a hypothesis of competitive displacement. Moreover, many species of both clades were associated with specific habitats, and the mixed-effects model predicted differential expansion for each environmental affinity. Together, these results are consistent with hypotheses that abiotic environmental conditions mediated shifts in dominance among benthic macrofauna. For instance, water chemistry perturbations in offshore shelf settings during the Permo-Triassic transition might have preferentially extirpated brachiopod species and prompted different diversity dynamics in the two clades [10, 23]. Alternatively or additionally, brachiopods might have specialized on substrates during the post-Paleozoic more than did

bivalves. Siliciclastic habitat specialists (e.g. sandy-bottom dwellers) seem to have experienced a dampening of geographic range expansion (Figure 4) and origination rates [22], for instance.

Conclusions

Evolutionary theorists beginning with Darwin have assumed that biotic competition at the finest spatio-temporal scales could impose a stabilizing feedback on diversity at the largest scales [24]. Geographic range size has been a foremost candidate for facilitating this process [25]. Mechanistically, death of competitors during mass extinction could release surviving species from refugia [26, 27]. As large geographic ranges increase the probability of survival [20, 28], albeit to a lesser degree in the midst of a mass extinction event [29], species richness could rebound in the interval after an extinction pulse (assuming no reduction in origination rates). With many species in competition, geographic distributions might shrink. Since restricted spatial distributions come with a high risk of extinction, richness might drop back to an equilibrium level (assuming no increase in origination rates). The scale of our analysis is appropriate to search for such a stabilizing feedback between the number of species and the size of their geographic ranges, since the median temporal resolution of our datasets (2–5 million years) matches the timespan on which diversity has responded to certain mass extinction events [26, 30, 31]. A mechanism for diversity-dependence based on ecological release, however, appears unfounded in marine bivalves and brachiopods. An alternative explanation for diversity-dependence, if present, is that strong incumbency could reduce the persistence or spread of nascent species [2, 17, 32].

Our assessment of geographic occurrences across all oceans through 485 million years indicates that competitive exclusion and ecological release have not dominated as macroecological patterns in the marine benthos at the level of species. This result contrasts with

evidence from population-level studies [4, 5]. One explanation for these differing patterns is that fine-scale spatio-temporal heterogeneity might induce asynchronous dynamics among populations of a species. Populations at one edge of a geographic range might contract while those elsewhere expand, such that the cumulative evolutionary pressure over all populations would be null [33-35]. Environmental filtering and other abiotic processes could emerge as the dominant determinants of species' distribution areas at coarse scales [6, 36-39]. Brachiopod and bivalve species interact with members of many other clades [15], and antagonistic interactions with predators or parasites also could affect macroevolution in our focal clades [40]. The scale-dependent effect of within-guild biotic interactions investigated here could extend to these other negative interactions, as discussed elsewhere in relation to the Red Queen model [41, 42]. Studies at the meso-scale of 10^3–10^5 years, in regional habitat zones, are a clear next step to pinpoint the scale at which abiotic processes dominate species' geographic patterns to the exclusion of biotic controls.

Characteristics of species, communities, or environments could alter the effect size of competitive interactions. The conclusion that factors other than competition control most species' geographic range sizes might be generalizable to other marine taxa [6, 8], for which geographic shifts are more predictably tied to environmental conditions such as temperature than are terrestrial taxa [43]. Terrestrial organisms that disperse long distances, such as many plants [39, 44], may share macroecological patterns with marine taxa. Meanwhile, in terrestrial organisms with limited dispersal capacity, such as many animals, competition might have controlled large-scale spatio-temporal distributions to a larger degree [16, 26, 27, 32, 44-46]. Biotic interactions also might exert more influence in communities with simple interaction networks, or in environments subject to widespread, homogenizing disturbances.

ACKNOWLEDGEMENTS

We thank Roger Close, Roger Benson, Neil Brocklehurst, and Adam Kocsis for methodological insights and Mark Patzkowsky, Tim Coulson, Corinne Myers, and Bruce Lieberman for further discussions. Andrew Zaffos and an anonymous reviewer improved the article. Contributors to the Paleobiology Database and Integrated Digitized Biocollections database enabled the work. Leverhulme Grant DGR01020 (E.E.S.), DFG Grants Ki 806/15-1 (W.K.) and AB 109/10 (M.A.), and the Clarendon Foundation (G.T.A.) supported the authors. This is Paleobiology Database publication 349.

AUTHOR CONTRIBUTIONS

E.E.S. formulated the idea. G.T.A. and M.A. curated the data. G.T.A. and E.E.S. designed the analysis. G.T.A. and W.K. programmed the code. G.T.A. generated the results and drafted the manuscript, which all authors critically edited and approved.

DECLARATION OF INTERESTS

The authors declare no competing interests.

MAIN TEXT FIGURES

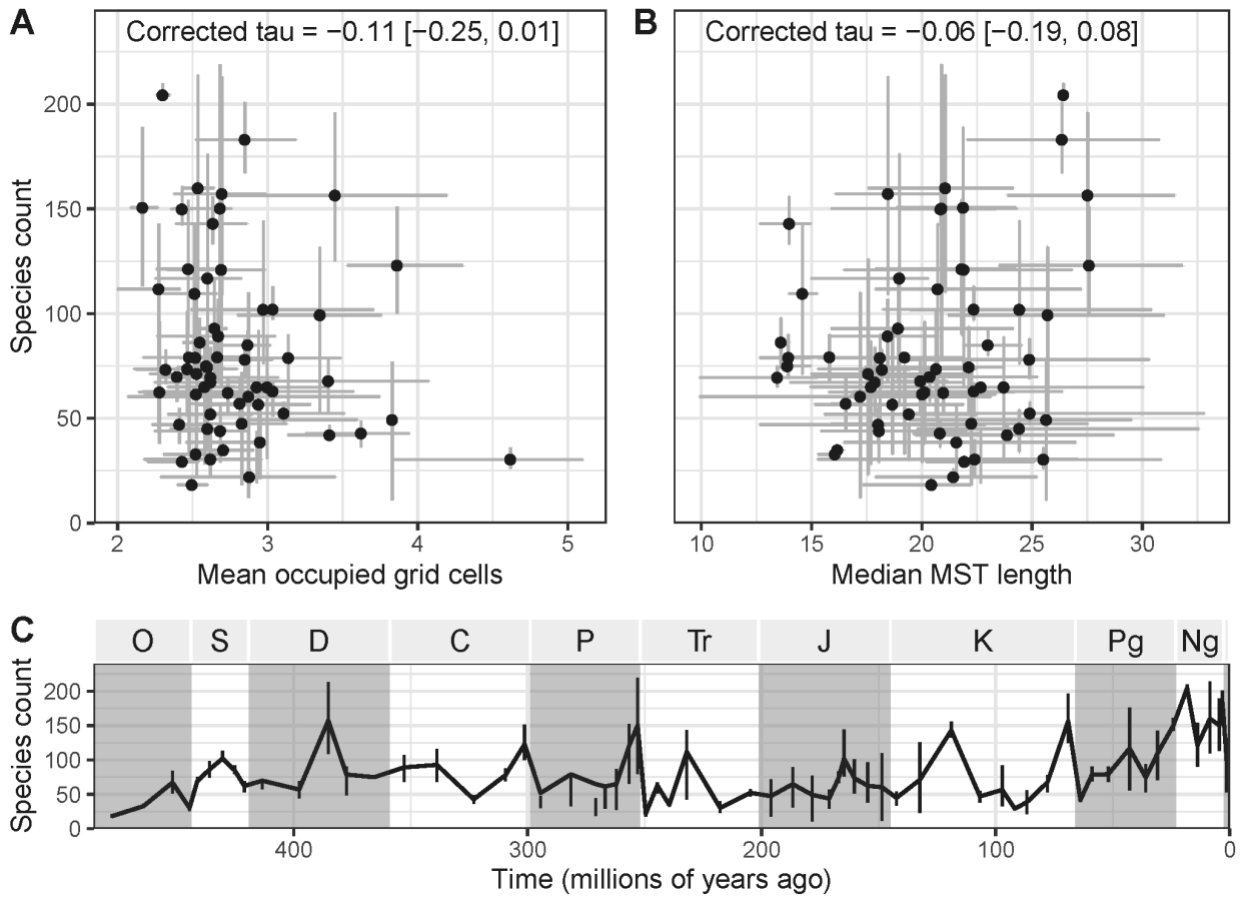


Figure 1. Regional Species Count Versus Geographic Range Size Through Time.

Scatterplots indicate the relationship between species count and (**A**) mean number of occupied grid cells or (**B**) median square-root length of tree connecting occupied cells (MST length) from 63 time bins (Table S1). Error bars mark the interquartile range. Correlation coefficients are corrected for autoregression in the predictor time series. The slight tendency toward negative values is attributable to the sampling bias of equal magnitude (Table S3; Methods S1). (**C**) Mean species count values among regional subsamples of global data are plotted across the 485-million-year study interval. Geological periods: O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; N, Neogene.

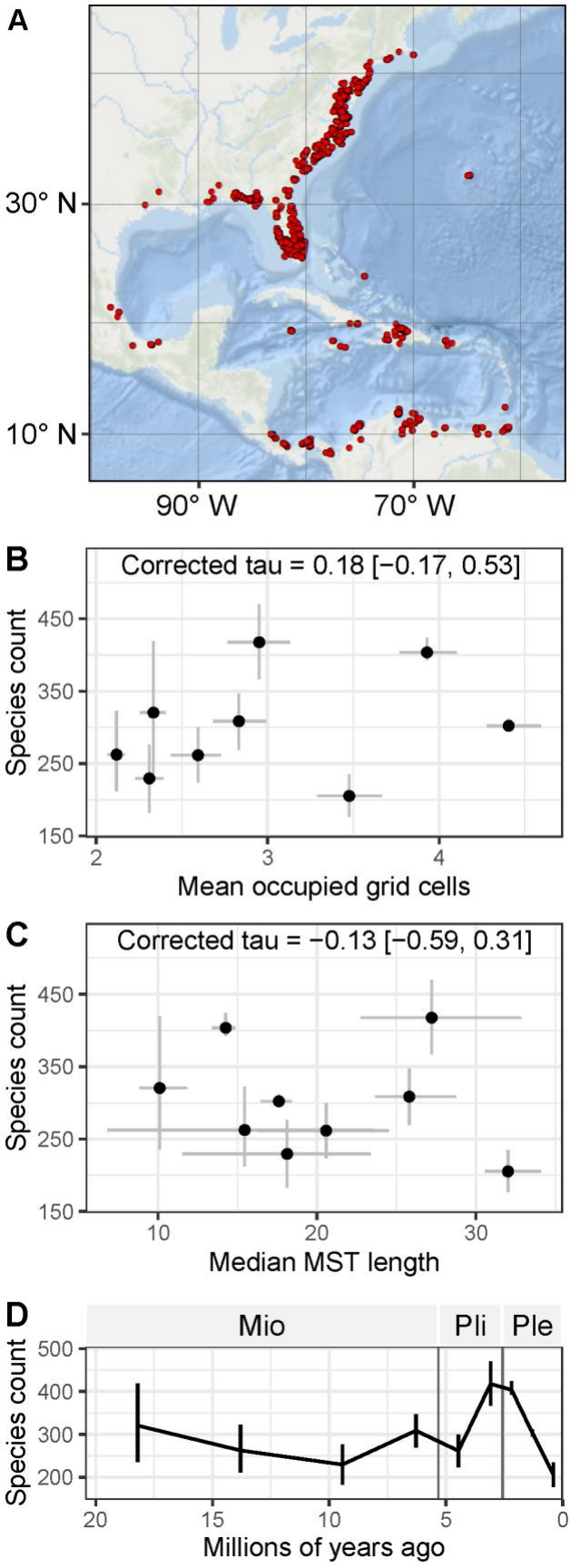


Figure 2. Species Count and Geographic Range Size from the Neogene–Quaternary Atlantic

(**A**) Northwest Atlantic bivalve specimen localities are plotted at modern-day coordinates. Scatterplots between species count and (**B**) mean number of occupied grid cells or (**C**) median square-root length of tree connecting occupied cells (MST length) are plotted as in Figure 1. Correlation coefficients for every pair of time series are listed in Table S3. Data ranged from the Burdigalian (Early Miocene) through the Pleistocene (Table S1). Geological epochs: Mio, Miocene; Pli, Pliocene; Ple, Pleistocene.

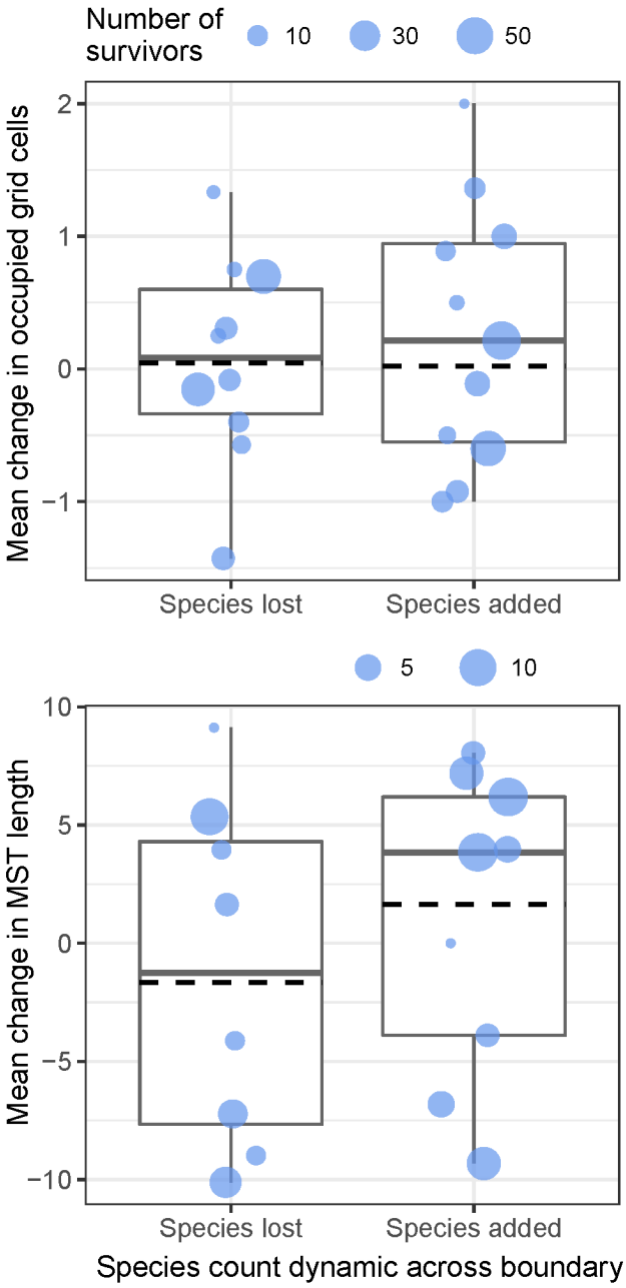


Figure 3. Occupancy Change Across Intervals of Net Species Loss or Gain Point size is proportional to the number of species sampled on both sides of a given boundary. Phanerozoic intervals were included only when successive bins contained ≥ 10 cells of the same habitat type inhabited by multiple surviving species (Table S4). (**A**) Species' occurrences were similarly dispersed (weighted Wilcoxon *p*-value = 0.79, *t*-value = -0.27, df = 19) between intervals when species count in a region decreased (n = 10) and increased (n = 11). The unweighted means (solid horizontal lines) aligned closely with n-weighted means (dashed lines, expansion of 0.05 grid cells per species for intervals of species count decrease *versus* expansion of 0.02 grid cells per species for intervals of species count increase). (**B**) Results were congruent when geographic distributions were measured using minimum spanning tree length: geographic trends were equivalent (*p*-value = 0.31, *t*-value = 1.1, df = 15) between intervals of species count decrease (average reduction of 1.66 km per species, n = 8) and increase (average expansion of 1.65 km per species, n = 9). When newly-originated and nearly-extinct species were evaluated along with mid-duration species, strong outliers appeared but the general trends were still consistent with those of mid-duration species alone (Table S4).

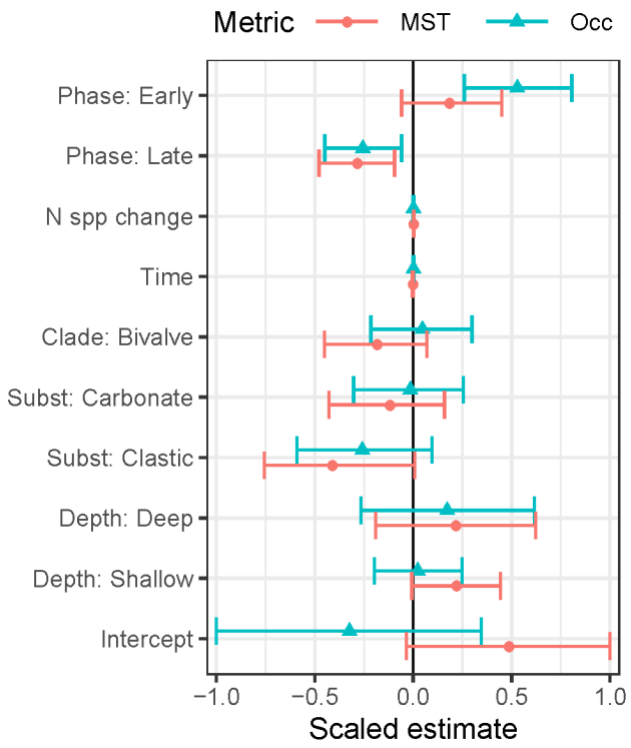


Figure 4. Predicted Change in Geographic Range Size for Boundary-Crossing Species

A mixed-effects model estimated a null effect from change in species count, contrary to expectations of ecological release. Estimates were relative to mid-existence (*versus* early or late) brachiopod species (*versus* bivalve) with no preference for habitat (*versus* preference for substrate or water depth). Substrate preference (subst.) was either carbonate (e.g. reef) or siliciclastic (clastic, e.g. sand). Error bars = 95% confidence intervals from profiling. Data for each species was standardized within subsamples of grid cells that contained suitable substrate and water depth. Color indicates whether a model was based on occupancy (Occ, n = 708) or square-root minimum spanning tree length (MST, n = 239); coefficients for the two models were scaled for comparison. Unscaled estimates are given in Table S7. Models based on subsets of predictor variables are described in Tables S5 and S6.

STAR METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Gawain Antell (gawainantell@gmail.com). This study did not generate reagents.

METHOD DETAILS

Global Dataset

Global bivalve and brachiopod occurrences across the Phanerozoic were downloaded from the Paleobiology Database (PBDB; paleobiodb.org) in July 2018, according to the following criteria: "taxa to include" as "Brachiopoda, Bivalvia"; "taxonomic resolution" as "genus"; "identification" as "latest"; "modifiers" as excluding the specific modifiers "aff.," "cf.," "sensu lato," "informal," "ex gr.," "?" or empty from genus and species names; "interval" as Cambrian–Holocene, with "time rule" as "overlap"; and "environment" as excluding "terrestrial." We chose not to restrict records to "accepted names only" in the download step, because this filter would have removed a large number of records that contained valid coordinates despite lacking minor fields associated with taxonomy (e.g. the year of species description). Furthermore, when this filter was applied in a sensitivity test, results were equivalent (Methods S1 section 2.3.1). We excluded records that (i) lacked paleocoordinates, (ii) contained metadata that indicated a geographic resolution greater than local area, (iii) were identified as "sp," "spp.," or a single letter, (iv) were flagged as "nomen dubium," "nomen vanum," "nomen nudum," or "indet.," (v) were identified at a rank above species level, or (vi) lacked taxonomic rank information above the genus level.

The frequency of synonyms could be high in the PBDB, because taxonomists have revised nomenclature after many of the names were entered. Therefore, we took precautions to validate taxonomic identifications. In particular, vetting included (i) cross-referencing entries against published lists [22, 47] and the World Register of Marine Species (marinespecies.org), (ii) hand-checking the original publication record for all species with durations longer than 50 Ma, and (iii) searching for synonyms based on spelling errors (e.g. *Saccostrea cuculata* instead of *S. cucullata*) and suffix variants (e.g. *-us* instead of *-um*).

We treated data at the level of species for the main text, because we were interested in the potential effects of inter- and intra-specific interactions [48]. Limitations of the fossil record include the potential to overlook cryptic species, because taxonomic definitions for fossils depend on morphology alone. This issue, however, likely affects both brachiopods and bivalves, from all time periods, and therefore is unlikely to result in directional bias. Nevertheless, we also repeated analyses at the level of genera (Table S3). To minimize the contributions of waste-bin taxa, only genera with species-level identifications were used; that is, we aggregated the species-level records to their encompassing genera.

We binned data at the finest temporal resolution (stages or clustered stages; Table S1) that retained sufficient occurrences in every time bin to calculate the variables of interest. Boundary dates followed the International Chronostratigraphic Chart from the International Commission on Stratigraphy [49] (stratigraphy.org). Spatial sampling in the Cambrian was too sparse to achieve sub-epoch resolution, so the Phanerozoic dataset was truncated to Ordovician–Pleistocene records. Time bins were < 20 Ma in duration (Table S1); the median was 5.1 Ma (interquartile range 3.6–7.7 Ma). A specimen record was included in a time bin if the name or age of both its maximum and minimum occurrence estimate fell within the start and end dates for

the bin. Dates were inclusive of rounding error to the nearest 0.01 Ma for boundaries younger than 10 Ma, 0.1 Ma for boundaries 50–150 Ma, and 1 Ma for boundaries older than 150 Ma. Where age estimates did not constrain records to single time bins, we matched the "early interval" and "late interval" names to time bin names where possible.

The PBDB internally derives paleo-latitude and -longitude under plate tectonic model described by Wright and others [50]. We used these coordinates to rasterize occurrences in each of 55,800 global grid cells. To ensure that all cells contained the same area, an initial 1° grid was transformed to Eckert VI projection, using the **raster** [51], **rgdal** [52], and **sp** [53] packages in R. The centroid coordinates of each of these equal-area cells were used for all subsequent calculations involving the spatial location of records (e.g. minimum spanning tree length). The transformation to ensure equal-area among cells meant that the shape of the cells was distorted (i.e. the sides differed in length), but average width of a cell was $1x10^2$ km. This spatial grain of analysis was chosen to smooth differences between densely and sparsely sampled localities while still capturing interspecific differences in regional occupancy [13]. Finally, we eliminated occurrence records for any species observed in exactly one grid cell from a single time period (please read Methods S1 for sensitivity to space-time singletons); this step reduced the influence of Lagerstätten and the incidence of invalid species identifications.

The rasterized occurrence records are available as Dataset S1. The dataset contains $1.7x10^4$ species ($8.1x10^3$ brachiopods and $8.9x10^3$ bivalves) and $5.8x10^4$ unique species-location-time points ($2.7x10^4$ for brachiopods and $3.1x10^4$ for bivalves).

Northwest Atlantic Dataset

The northwest Atlantic contains a dense record of fossil mollusks. This system is also of interest to test for ecological release because the Gulf Coast is the only region where taxa

"bloomed" after the end-Cretaceous mass extinction [54]. We downloaded occurrences from the Integrated Digitized Biocollections database in October 2018, according to the following criteria: "geopoint" entered; "specificepithet" entered; "basisofrecord" as "FossilSpecimen;" and "class" as "Bivalvia." After download, we excluded records where the species epithet was entered as "sp.," the taxon rank was entered as "doubtful," or where the radius of uncertainty exceeded 50 km. We cross-referenced the species list against the World Register of Marine Species database to identify freshwater species, which were removed. As in the PBDB dataset, we synonymized species names with spelling errors or suffix variants. We also checked that no set of coordinates fell exactly at the origin (i.e. 0° latitude and longitude), which could indicate a default entry for an unknown location. We removed records with an era assigned to the Mesozoic, Paleozoic, or Proterozoic. The iDigBio records were combined with the subset of global PBDB records from Miocene–Late Pleistocene mollusks and subset to the northwest Atlantic.

The extent of the study area (6–42° North latitude, 50–100° West longitude; Figure 2D) corresponds to the Gulf of Mexico and the western part of the Central North Atlantic biozone defined by Sutton and others [55]. The northern limit of this area matches the meso-pelagic ecoregion boundary quantified elsewhere [56]. The southern limit is set by the Isthmus of Panama and coastline of South America.

PBDB and iDigBio contain almost entirely different sets of specimens, because the PBDB draws primarily on published records, whereas iDigBio draws primarily on unpublished museum specimens. There are rare cases where specimens could have been entered into both databases (for instance, some digitized holotypes). This duplication in individual occurrence records would not affect analysis, however, because only unique space-time-taxa combinations are retained during rasterization.

Time bins for the northwest Atlantic dataset (Table S1) were shorter than those used in the global dataset. All bins were < 5 Ma in duration, with a median of 1.8 Ma (interquartile range 1.0–4.4 Ma). Sampling in the Aquitanian was so sparse in comparison to the other intervals that the dataset was truncated to Burdigalian–Late Pleistocene records only. Many iDigBio records lacked precise age information, either as a numeric estimate or geologic stage name. We therefore researched the chronostratigraphy of sedimentary units in the northwest Atlantic to determine the minimum and maximum age estimates for 485 formations and members, expanding the correlation chart of Saupe and others [57]. The compilation (Dataset S2) includes referenced opinions of age estimates from isotopic dating, sequence stratigraphy matches with sea level curves, and stratigraphic correlations and relationships. Age estimates for each occurrence record that lacked sufficient temporal information were inferred from the geologic unit with which it was associated. We searched for geologic context in the "group," "formation," "member," "verbatimLocality," and "lithostratigraphicTerms" fields of iDigBio records; and the "stratgroup," "formation," "member," "collection_name," "collection_aka," and "localsection" fields of PBDB records. We hand-checked any records that included mention of multiple geologic units in order to determine a unique assignment.

The age range of deposition in many geologic units spanned multiple time bins. Therefore, we assigned a record to a time bin if more than two-thirds of its estimated range fell within the bin. Records that spanned more than 3 bins were excluded altogether. This two-thirds overlap rule assigned 70% of geologic units to time bins; a stricter binning rule would have failed to maintain a sufficient amount for analysis. Some records could not be assigned to a time bin using the methods above, but had been identified with a stage or North American Land Mammal Age (e.g. Hemingfordian). These temporal units were converted to numeric age ranges

based on the literature [49, 58, 59]. It is probable that the two-thirds overlap rule assigned some occurrences to incorrect time bins. We account for this error later, in time series analysis, which controlled for similarity between time bins that occurred close together regardless of the source of similarity (e.g. founder effects, environmental stasis, or methodological artifacts).

The iDigBio platform provides coordinates at their modern-day positions. Therefore, we reconstructed the paleocoordinates for all records (iDigBio and PBDB) in the northwest Atlantic dataset. Modern coordinates were partitioned among plates and rotated according to PaleoAtlas models from EarthByte [60] (version 2, accessed October 2018) within the GPlates computing environment [61] (version 2.1.0, released August 2018). The coordinates of all records within a time bin were reconstructed at the median age of the bin. We rasterized occurrences into equal-area grid cells at a resolution that was twice as fine as that of the global dataset: we used a 0.5° grid, resulting in equal-area cells of 5x10 km width on average, and removed species that occurred in a single cell in only one time bin.

The rasterized occurrence records for the northwest Atlantic are available as Dataset S3. The dataset contains 1.9×10^3 species and 6.1×10^4 unique species-location-time points.

Spatial Subsampling

To account for the heterogeneity of preserved fossil occurrences, we devised a subsampling method that controls jointly the spatial extent and dispersion of sample localities. Each iteration picks a random occurrence point (a "seed cell") as a center around which to circumscribe a buffer region with a predetermined radius; the procedure then draws a given number of equal-area grid cells within the circular buffer (Figure S1). Selected cells form a subsample, within which geographic range size and species count are calculated. We assessed the performance of the subsampling routine by calculating a proxy of spatial coverage, which we

call sampling aggregation; we evaluated the explanatory power of this variable in time series analysis. We constructed buffers with the **raster** package [51] and manipulated vector geometries with the **sp** package [53] in R. Spanning tree distances were then calculated with the **vegan** package [62].

The spatial subsampling procedure is modified for the taxon-tracking approach (Figure S2 and Methods S1 section 2.2). Rather than select a seed cell at random, the taxon-tracking procedure selects a grid cell that is occupied by a focal taxon. Moreover, only cells with lithology and water depth suitable for a given taxon are considered when subsampling that species' geographic range. For instance, focusing on a species with a documented preference for shallow-water settings but no preference for substrate, the set of cells available for subsampling would consist of all cells identified with shallow-water. Species were included in taxon-tracking analysis only if their environmental preferences could be determined.

The spatial standardization procedure can be amended in many ways, for instance by selecting grid cells according to either unweighted or proximity-weighted probabilities. In the unweighted cell selection method, occupied cells were selected at random within each radius-buffered area. In the proximity-weighted method, cells were selected with scaled probability inversely proportional to their squared distance from the center of the buffer. The proximity-weighted sampling method is more computationally intensive since the weights must be calculated for every occupied cell in all iterations of the subsampling procedure. However, this method clusters subsampled cells more compactly (i.e. reduces spatial dispersion) while retaining a random component to subsampling. In cases where sampling is dense and more even across space and time, the proximity-weighted method might reduce the grain of study and hence lead to more precise estimates of assemblage-level characters, such as mean species' occupancy.

We tested the sensitivity of analyses to each parameter of the subsampling routine, including the radius and number of cells selected (Methods S1 section 2.3). For the analysis presented in the main text, we set a 1,500 km radius to confine each subsample in the global data, which agrees well with previous quantitative thresholds [13] and the common approach of constraining the dispersion of a sample to one continent or ocean basin. Consequently, the circle that constrained a random sample covered $7x10^6$ km^2, an area approximately equivalent to the size of Australia. Within this region, each assemblage-based subsample iteration drew 12 cells without replacement; each taxon-tracking subsample iteration drew 10 cells of suitable habitat. These sample sizes fall within the range that others have set when sampling grid cells [31]. For the northwest Atlantic dataset, occurrences were restricted to a single ocean basin in extent so we did not use a radial constraint during subsampling. We did, however, still standardize spatial area with the cell selection step of subsampling. The high density of occurrences allowed us to include 15 cells in the subsamples from every time bin. Each subsample procedure was replicated 500 times in each time bin so that all available data contributed to analysis, even though any individual subsample may have captured only some of the fossil species present in the global record at the time.

We did not rarefy PBDB "collection" or "reference" groups as done in earlier studies (for review, read [13]). This was because our spatial subsampling procedure already rarefied the number of grid cells in each sample, and it has been found that there is a 1:1 correlation between reference counts and grid cell counts in global PBDB datasets of marine invertebrates [1]. Indeed, the spatial standardization procedure rarefies species count as a consequence of rarefying sampling area. Additional rarefaction on species count would compress the distribution of observed values further and could cause type 2 error (Methods S1 section 2.3).

Methods S1 section 1 discusses further considerations of spatial standardization, such as discussion of proportional occupancy metrics. Methods S1 sections 2.1 and 2.2 give step-by-step descriptions of assemblage-based and taxon-tracking subsampling procedures. Figure S3 provides power analysis of the data compression caused by the subsampling procedure.

Environmental Preferences

The position and extent of different habitat types (e.g. shallow water carbonate settings) have changed through time. Many species have specific habitat preferences, in which case individuals may be constrained in the areas where they can occur. Since environmental preferences are defined at the level of species, it was not straightforward to account for habitat within an assemblage framework. Therefore, we accounted for suitable habitat only in the taxon-tracking framework. Environmental data were unavailable and unstandardized for most iDigBio specimen records, so we could determine environmental affinities for taxa in the global (PBDB) dataset but not in the northwest Atlantic dataset.

We considered environmental preferences along two axes: substrate (carbonate, siliciclastic, or generalist for both) and bathymetry (deep water, shallow water, or generalist for both). We did not code preferences for latitude or sediment grain size. Carbonate substrates already reflect the extent of tropical biomes to a large degree, and perhaps more accurately than would latitudinal coordinates. Similarly, variation in substrate type and bathymetry also capture some of the differences between fine- and coarse-grained habitats. Determining grain size affinity would improve the quality of niche characterizations only moderately, while at the same time reducing sample size substantially—grain size was not recorded in many PBDB records.

We coded binary environmental conditions based on associated data fields in the PBDB records (prior to rasterization), according to the definitions provided by Nürnberg and Aberhan

[63]. In particular, “carbonate” included any instance of “carbonate,” “limestone,” “reef rocks,” “bafflestone,” “bindstone,” “dolomite,” “framestone,” “grainstone,” “lime mudstone,” “packstone,” “rudstone,” “floatstone,” or “wackestone” in the PBDB primary lithology field. In contrast, “siliciclastic” included any instance of “shale,” “siliciclastic,” “volcaniclastic,” “claystone,” “conglomerate,” “mudstone,” “phyllite,” “quartzite,” “sandstone,” “siltstone,” “slate,” or “schist” in primary lithology. For water depth, “deep” was assigned for any environment field in the set of “basinal (carbonate),” “basinal (siliceous),” “basinal (siliciclastic),” “deep-water indet.,” “deep subtidal indet.,” “deep subtidal ramp,” “deep subtidal shelf,” “offshore,” “offshore indet.,” “offshore shelf,” “slope,” “submarine fan,” “offshore ramp,” “basin reef,” and “slope/ramp reef.” Otherwise, “shallow” was assigned for “coastal indet.,” “delta front,” “delta plain,” “deltaic indet.,” “estuary/bay,” “foreshore,” “interdistributary bay,” “lagoonal,” “lagoonal/restricted shallow subtidal,” “marginal marine indet.,” “open shallow subtidal,” “fluvial-deltaic indet.,” “paralic indet.,” “peritidal,” “prodelta,” “sand shoal,” “shallow subtidal indet.,” “shoreface,” “transition zone/lower shoreface,” “intrashelf/intraplatform reef,” “reef, buildup or bioherm,” “perireef or subreef,” or “platform/shelf-margin reef” in the environment field.

The environment in a grid cell was classified as a single condition if ≥ 80% of the paleoenvironmental records in that cell were indicative of that condition; otherwise, the cell was inferred to contain both conditions (e.g. mixed carbonate and siliciclastic substrate). Where both conditions occurred in the reference raster cell, however, we made no inference from that cell about the species’ habitat preferences.

Some occupied grid cells lacked sufficient information to assign an environmental condition. We referenced these cells against an environmental raster derived from an

independent PBDB download of all brachiopod and bivalve occurrences (not restricted to species-level identification, but excluding fossils from terrestrial environments). If the grid cell in question was characterized by a single environmental condition in the reference raster, we assigned that condition to the species' occurrence records from the corresponding cell in the vetted dataset. Cells where environmental conditions still could not be inferred along one or both axes were not considered: occurrences that fell in these cells were excluded.

After grading the environmental state for every occurrence record, we inferred the environmental preferences of each species. For each axis, a species with ≥ 80% of occurrence records classed as only one category was considered a specialist for that environmental condition. A species with > 20% of occurrence records in both categories was treated as a generalist along that axis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Time Series Analysis

The time series analysis aimed to establish the relationship between species' distributions and species count, and to evaluate the success of spatial sampling standardization. The hypothesis of ecological release predicts an inverse correlation. For each subsampled dataset, we calculated the correlations between predictor time series (species count or sampling aggregation) and response time series (occupancy or MST). We used Kendall's *tau* to estimate correlation strength because this non-parametric metric can be modified to account for ties, which occurred in time series of median occupancy.

Counts were similar in adjacent time bins, and so we corrected the time series analysis for this autocorrelation. We inspected autocorrelation and partial autocorrelation plots for a

representative iteration of each variable's time series and tested the assumption of stationarity with an augmented Dickey-Fuller test from the **fUnitRoots** package [64] in R. Based on these explorations, we chose an autoregressive (AR) error structure of order 1 to model the predictor time series (either species count or sampling aggregation) in each iteration, using the "arima" function from the base **stats** package in R [65]. The first-order coefficient estimated for species count was 0.33 and for sampling aggregation was 0.38. We calculated tau based on the predictor time series' AR(1) model residuals instead of the original series to prevent induced correlation [66]. The mean and 95% confidence interval of Kendall's *tau* coefficient are reported for each pair-wise correlation. The bounds of the confidence interval are the 2.5% and 97.5% quantiles of values obtained across bootstrap iterations. We used the "prewhiten" function in the **TSA** package [67] to check for correlations at positive time lags of 1–5 steps.

Range-Size Shifts

For every species that was observed in two or more consecutive time bins, we quantified the change in the survivor's range size between the earlier and later time bin. Some species were observed in three or more consecutive bins; in these cases, we calculated the changes associated with each of the stage boundaries that the species crossed. We compared mean change in range size among survivors ("range-size shift") of two sets of time intervals: intervals when species count increased and intervals when species count decreased.

Unbalanced numbers of newly-originated and nearly-extinct species between extinction intervals and background intervals [27, 31] could bias our analysis, because taxa at these times often experience predictable changes in geographic range size unrelated to potential effects of ecological release [18, 19]. The typical trajectory of geographic distribution over the duration of a lineage is arched like a rainbow (mathematically, a convolution of two top hat functions). That

is to say, lineages tend to expand after appearing in the fossil record and contract prior to disappearing from the record. Moreover, species in the middle of their existence, at peak occupancy, are hypothesized to experience the most competitive pressure [68]. We therefore calculated range-size shift (i) on all species, and (ii) only for species in the middle interval of their temporal durations. In the latter case, if a species was sampled across three consecutive boundaries (i.e. occurred in four consecutive time bins), the change in range size over the middle interval (but not either end interval) would contribute to the calculation of range-size shifts. Survivors of exactly one interval behaved equivalently to those that were in the middle of their duration, so these temporal doubletons were included in all range-size shift measurements.

Phanerozoic intervals were included only when successive bins contained ≥ 10 raster cells of the same habitat type inhabited by multiple surviving species. Species from the Pleistocene were not assessed because no modern occurrence data were included. To classify each interval as associated with either a net gain or loss in species number, we calculated whether the median change in species count experienced by survivors across a given boundary was positive or negative (Table S4). More species crossed certain stage boundaries than others (Table S4), which resulted in unequal variances among intervals. We therefore weighted each range-size shift by the number of survivors that contributed to that shift and compared the difference in the two groups with a two-sided Mann-Whitney (Wilcoxon rank-sum) test.

Mixed-Effects Models

We explored the nature of the relationship (i.e. estimated degrees of freedom) between change in species count and change in range size with a generalized additive mixed-effects model (GAMM). The additive component of the model fit a penalized thin plate regression spline on species count, implemented with the **mgcv** [69] and **gamm4** [70] packages in R. We

inspected this spline to determine whether the smoothing term was justified, relative to a linear model. In addition, the identity of the time interval associated with each survivor was included as a random effect in the GAMM. We included this random effect because species from the same time bin were expected to pseudoreplicate each other due to shared environmental conditions and spatial sampling structure. Moreover, this random effect obviated the need to correct for multiple testing, because a single coefficient was calculated for each fixed effect over the entire time series, instead of for every time bin as in Finnegan and others [71].

The estimates for the spline in the GAMM (degrees of freedom = 0.23) gave no justification to parameterize species count with a higher-order polynomial than a linear regression. Therefore, we performed model selection on linear mixed-effects models (LMM) rather than a more complicated regression structure. As in the GAMM, time interval was included as a random effect in the LMM. We considered the influence of the following six fixed effects: change in species count (integer), clade (categorical: Brachiopoda or Bivalvia), point in lineage duration (categorical: Early, Middle, or Late), time bin age (continuous and positive), bathymetric preference (categorical: carbonate, siliciclastic, or both), and lithological preference (categorical: shallow, deep, or both). We then built a global linear mixed-effects model (all predictor variables included; Figure 4 and Table S7) and compared the 64 possible models, including the model with no fixed effects added (that is, a horizontal line). We implemented mixed-effects models with functions in the **lme4** package [72, 73]. Following default settings, we fit models based on restricted maximum likelihood for estimation purposes and based on maximum likelihood for testing purposes.

We estimated Akaike's Information Criterion (AIC) instead of corrected AIC, given the large number of species in the dataset, to determine the subset of models that received non-trivial

weight (*Delta* AIC < 10; for derivation and discussion of weights, see Burnham and Anderson [74]. Models are listed in order of relative performance in Table S5. For fixed effects of these supported models, we recorded point estimates and the 95% confidence interval bounds from the likelihood profile (Table S6). Log-likelihood estimates can be interpreted on the original scale of the variable. We chose to report log-likelihood instead of marginal likelihood estimates because (i) interpretation of coefficients is more straightforward and (ii) statistical theory to calculate marginal likelihood of parameters over the random effect(s) in mixed-effects models is poorly developed.

DATA AND CODE AVAILABILITY

All generated data and custom software developed for this project are archived at *https://doi.org/10.5281/zenodo.3491853* and maintained at *https://github.com/GawainAntell/EcoRelease*.

REFERENCES

1. Alroy, J., Aberhan, M., Bottjer, D.J., Foote, M., Fürsich, F.T., Harries, P.J., Hendy, A.J.W., Holland, S.M., Ivany, L.C., Kiessling, W., et al. (2008). Phanerozoic trends in the global diversity of marine invertebrates. Science *321*, 97-100.

2. Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J., and Alfaro, M.E. (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. *4*, 1958.

3. Aguilée, R., Gascuel, F., Lambert, A., and Ferriere, R. (2018). Clade diversification dynamics and the biotic and abiotic controls of speciation and extinction rates. Nat. Commun. *9*, 3013.

4. Schoener, T.W. (1983). Field experiments on interspecific competition. Am. Nat. *122*, 240-285.

5. Sexton, J.P., McIntyre, P.J., Angert, A.L., and Rice, K.J. (2009). Evolution and ecology of species range limits. Annu. Rev. Ecol., Evol. Syst. *40*, 415-436.

6. Myers, C.E., and Lieberman, B.S. (2011). Sharks that pass in the night: using Geographical Information Systems to investigate competition in the Cretaceous Western Interior Seaway. Proc. R. Soc. Lond. B *278*, 681-689.

7. Klompmaker, A.A., and Finnegan, S. (2018). Extreme rarity of competitive exclusion in modern and fossil marine benthic ecosystems. Geology *46*, 723-726.

8. Benton, M.J. (1996). Testing the roles of competition and expansion in tetrapod evolution. Proc. R. Soc. Lond. B *263*, 641-646.

9. Rillo, M.C., Sugawara, M.T., Cabella, B., Jonkers, L., Baranowski, U.K., Kučera, M., and Ezard, T.H. (2019). On the mismatch in the strength of competition among fossil and modern species of planktonic Foraminifera. Global Ecol. Biogeogr.

10. Gould, S.J., and Calloway, C.B. (1980). Clams and brachiopods—ships that pass in the night. Paleobiology *6*, 383-396.

11. Payne, J.L., Heim, N.A., Knope, M.L., and McClain, C.R. (2014). Metabolic dominance of bivalves predates brachiopod diversity decline by more than 150 million years. Proc. R. Soc. Lond. B *281*, 20133122.

12. Carlson, S.J. (2016). The evolution of Brachiopoda. Annu. Rev. Earth Planet. Sci. *44*, 409-438.

13. Close, R.A., Benson, R.B.J., Upchurch, P., and Butler, R.J. (2017). Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification. Nat. Commun. *8*, 15381.

14. Herrera-Alsina, L., Pigot, A.L., Hildenbrandt, H., and Etienne, R.S. (2018). The influence of ecological and geographic limits on the evolution of species distributions and diversity. Evolution *72*, 1978-1991.

15. Miller, A.I., and Sepkoski, J.J. (1988). Modeling bivalve diversification: the effect of interaction on a macroevolutionary system. Paleobiology *14*, 364-369.

16. Laube, I., Graham, C.H., and Böhning-Gaese, K. (2013). Intra-generic species richness and dispersal ability interact to determine geographic ranges of birds. Global Ecol. Biogeogr. *22*, 223-232.

17. Wagner, P.J., Kosnik, M.A., and Lidgard, S. (2006). Abundance distributions imply elevated complexity of post-Paleozoic marine ecosystems. Science *314*, 1289-1292.

18. Foote, M., Crampton, J.S., Beu, A.G., Marshall, B.A., Cooper, R.A., Maxwell, P.A., and Matcham, I. (2007). Rise and fall of species occupancy in Cenozoic fossil mollusks. Science *318*, 1131-1134.

19. Liow, L.H., and Stenseth, N.C. (2007). The rise and fall of species: implications for macroevolutionary and macroecological studies. Proc. R. Soc. Lond. B *274*, 2745-2752.

20. Harnik, P.G., Simpson, C., and Payne, J.L. (2012). Long-term differences in extinction risk among the seven forms of rarity. Proc. R. Soc. Lond. B *279*, 4969-4976.

21. Liow, L.H., Reitan, T., and Harnik, P.G. (2015). Ecological interactions on macroevolutionary time scales: clams and brachiopods are more than ships that pass in the night. Ecol. Lett. *18*, 1030-1039.

22. Kiessling, W., and Aberhan, M. (2007). Environmental determinants of marine benthic biodiversity dynamics through Triassic–Jurassic time. Paleobiology *33*, 414-434.

23. Clapham, M.E., and Bottjer, D.J. (2007). Permian marine paleoecology and its implications for large-scale decoupling of brachiopod and bivalve abundance and diversity during the Lopingian (Late Permian). Palaeogeogr., Palaeoclimatol., Palaeoecol. *249*, 283-301.

24. Reznick, D.N., and Ricklefs, R.E. (2009). Darwin's bridge between microevolution and macroevolution. Nature *457*, 837-842.

25. Rabosky, D.L. (2013). Diversity-dependence, ecological speciation, and the role of competition in macroevolution. Annual Review of Ecology, Evolution, and Systematics *44*, 481-502.

26. Sidor, C.A., Vilhena, D.A., Angielczyk, K.D., Huttenlocker, A.K., Nesbitt, S.J., Peecook, B.R., Steyer, J.S., Smith, R.M., and Tsuji, L.A. (2013). Provincialization of terrestrial faunas following the end-Permian mass extinction. Proc. Natl. Acad. Sci. U.S.A. *110*, 8129-8133.

27. Button, D.J., Lloyd, G.T., Ezcurra, M.D., and Butler, R.J. (2017). Mass extinctions drove increased global faunal cosmopolitanism on the supercontinent Pangaea. Nat. Commun. *8*, 733.

28. Saupe, E.E., Qiao, H., Hendricks, J.R., Portell, R.W., Hunter, S.J., Soberón, J., and Lieberman, B.S. (2015). Niche breadth and geographic range size as determinants of species survival on geological time scales. Global Ecol. Biogeogr. *24*, 1159-1169.

29. Kiessling, W., and Aberhan, M. (2007). Geographical distribution and extinction risk: lessons from Triassic–Jurassic marine benthic organisms. J. Biogeogr. *34*, 1473-1489.

30. Smith, J.T., and Jackson, J.B.C. (2009). Ecology of extreme faunal turnover of tropical American scallops. Paleobiology *35*, 77-93.

31. Darroch, S.A.F., and Wagner, P.J. (2015). Response of beta diversity to pulses of Ordovician-Silurian mass extinction. Ecology *96*, 532-549.

32. Ranjard, L., Welch, D., Paturel, M., and Guindon, S. (2014). Modelling competition and dispersal in a statistical phylogeographic framework. Syst. Biol. *63*, 743-752.

33. Ricklefs, R.E. (2004). A comprehensive framework for global patterns in biodiversity. Ecol. Lett. *7*, 1-15.

34. Eldredge, N., Thompson, J.N., Brakefield, P.M., Gavrilets, S., Jablonski, D., Jackson, J.B., Lenski, R.E., Lieberman, B.S., McPeek, M.A., and Miller, W. (2005). The dynamics of evolutionary stasis. Paleobiology *31*, 133-145.

35. Schindler, D.E., Armstrong, J.B., and Reed, T.E. (2015). The portfolio concept in ecology and evolution. Front. Ecol. Environ. *13*, 257-263.

36. Willis, K.J., and Whittaker, R.J. (2002). Species diversity—scale matters. Science *295*, 1245-1248.

37. Wiens, J.J. (2011). The niche, biogeography and species interactions. Philos. Trans. R. Soc. Lond. B *366*, 2336-2350.

38. Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., and Araújo, M.B. (2011). Ecological Niches and Geographic Distributions, (Princeton, NJ: Princeton University Press).

39. Götzenberger, L., de Bello, F., Bråthen, K.A., Davison, J., Dubuis, A., Guisan, A., Lepš, J., Lindborg, R., Moora, M., Pärtel, M., et al. (2012). Ecological assembly rules in plant communities—approaches, patterns and prospects. Biol. Rev. Camb. Philos. Soc. *87*, 111-127.

40. Stanley, S.M. (2008). Predation defeats competition on the seafloor. Paleobiology *34*, 1-21.

41. Benton, M.J. (2009). The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. science *323*, 728-732.

42. Strotz, L.C., Simoes, M., Girard, M.G., Breitkreuz, L., Kimmig, J., and Lieberman, B.S. (2018). Getting somewhere with the Red Queen: chasing a biologically modern definition of the hypothesis. Biol. Lett. *14*, 20170734.

43. Sunday, J.M., Bates, A.E., and Dulvy, N.K. (2012). Thermal tolerance and the global redistribution of animals. Nature Climate Change *2*, 686.

44. Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., et al. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol. Rev. Camb. Philos. Soc. *88*, 15-30.

45. Dunhill, A.M., and Wills, M.A. (2015). Geographic range did not confer resilience to extinction in terrestrial vertebrates at the end-Triassic crisis. Nat. Commun. *6*, 7980.

46. Pigot, A.L., and Tobias, J.A. (2013). Species interactions constrain geographic range expansion over evolutionary time. Ecol. Lett. *16*, 330-338.

47. (1997-2002). Part H: Brachiopoda. In Treatise on Invertebrate Paleontology, Volume 1-4, Part H: Brachiopoda, Revised Edition, R.L. Kaesler, ed. (Boulder, Colorado: Geological Society of America).

48. Hendricks, J.R., Saupe, E.E., Myers, C.E., Hermsen, E.J., and Allmon, W.D. (2014). The generification of the fossil record. Paleobiology *40*, 511-528.

49. Cohen, K.M., Harper, D.A.T., and Gibbard, P.L. (2017). International Chronostratigraphic Chart 2017/02. 2017-02 Edition. (International Commission on Stratigraphy, IUGS).

50. Wright, N., Zahirovic, S., Müller, R., and Seton, M. (2013). Towards community-driven paleogeographic reconstructions: integrating open-access paleogeographic and paleobiology data with plate tectonics. Biogeosciences *10*, 1529-1541.

51. Hijmans, R.J. (2019). raster: Geographic data analysis and modeling. R package version 2.8-19 Edition.

52. Bivand, R., Keitt, T., and Rowlingson, B. (2018). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.3-6 Edition.

53. Bivand, R.S., Pebesma, E., and Gómez-Rubio, V. (2013). Applied Spatial Data Analysis with R, 2 Edition, (New York, NY: Springer Science & Business Media).

54. Jablonski, D. (1998). Geographic variation in the molluscan recovery from the end-Cretaceous extinction. Science *279*, 1327-1330.

55. Sutton, T.T., Clark, M.R., Dunn, D.C., Halpin, P.N., Rogers, A.D., Guinotte, J., Bograd, S.J., Angel, M.V., Perez, J.A.A., Wishner, K., et al. (2017). A global biogeographic classification of the mesopelagic zone. Deep Sea Research Part I: Oceanographic Research Papers *126*, 85-102.

56. Backus, R., Craddock, J., Haedrich, R., and Robison, B. (1977). Atlantic mesopelagic zoogeography. Fishes of the western North Atlantic *7*.

57. Saupe, E.E., Hendricks, J.R., Portell, R.W., Dowsett, H.J., Haywood, A., Hunter, S.J., and Lieberman, B.S. (2014). Macroevolutionary consequences of profound climate change on niche evolution in marine molluscs over the past three million years. Proc. R. Soc. Lond. B *281*, 20141995.

58. Bell, C.J., Lundelius Jr, E.L., Barnosky, A., Graham, R., Lindsay, E., Ruez Jr, D., Semken Jr, H., Webb, S., and Zakrzewski, R. (2004). The Blancan, Irvingtonian, and Rancholabrean mammal ages. Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology. Columbia University Press, New York, 232-314.

59. Tedford, R.H., Albright III, L.B., Barnosky, A.D., Ferrusquia-Villafranca, I., Hunt Jr, R.M., Storer, J.E., Swisher III, C.C., Voorhies, M.R., Webb, S.D., and Whistler, D.P. (2004). Mammalian biochronology of the Arikareean through Hemphillian interval (late Oligocene through early Pliocene epochs). Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology. Columbia University Press, New York, 169-231.

60. Scotese, C.R. (2016). Tutorial: PALEOMAP PaleoAtlas for GPlates and the PaleoData Plotter Program.

61. Müller, R.D., Cannon, J., Qin, X., Watson, R.J., Gurnis, M., Williams, S., Pfaffelmoser, T., Seton, M., Russell, S.H., and Zahirovic, S. (2018). GPlates: Building a virtual Earth through deep time. Geochemistry, Geophysics, Geosystems *19*, 2243-2261.

62. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2019). vegan: Community ecology package. R package version 2.5-4 Edition.

63. Nürnberg, S., and Aberhan, M. (2013). Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. Paleobiology *39*, 360-372.

64. Wuertz, D., Setz, T., and Chalabi, Y. (2017). fUnitRoots: Rmetrics – modelling trends and unit roots. R package version 3042.79 Edition.

65. R Core Team (2018). R: A Language and Environment for Statistical Computing, (Vienna, Austria: R Foundation for Statistical Computing).

66. Probst, W.N., Stelzenmüller, V., and Fock, H.O. (2012). Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: an exemplary analysis of North Sea fish population indicators. ICES J. Mar. Sci. *69*, 670-681.

67. Chan, K.-S., and Ripley, B. (2018). TSA: Time series analysis. R package version 1.2 Edition.

68. Žliobaitė, I., Fortelius, M., and Stenseth, N.C. (2017). Reconciling taxon senescence with the Red Queen's hypothesis. Nature *552*, 92.

69. Wood, S.N. (2003). Thin plate regression splines. J. R. Stat. Soc. B *65*, 95-114.

70. Wood, S., and Scheipl, F. (2017). gamm4: Generalized additive mixed models using 'mgcv' and 'lme4'. R package version 0.2-5 Edition.

71. Finnegan, S., Anderson, S.C., Harnik, P.G., Simpson, C., Tittensor, D.P., Byrnes, J.E., Finkel, Z.V., Lindberg, D.R., Liow, L.H., Lockwood, R., et al. (2015). Paleontological baselines for evaluating extinction risk in the modern oceans. Science *348*, 567-570.

72. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1-48.

73. Venables, W., and Ripley, B. (2002). Random and mixed effects. In Modern applied statistics with S. (Springer), pp. 271-300.

74. Burnham, K.P., and Anderson, D.R. (2010). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edition, (New York, NY: Springer Science & Business Media).

75. Hsieh, T., Ma, K., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (H ill numbers). Methods Ecol. Evol. *7*, 1451-1456.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Paleobiology Database	https://paleobiodb.org	Multiple, accessed July 2018
Integrated Digitized Biocollections database	https://www.idigbio.org	Multiple, accessed October 2018
PaleoAtlas tectonic plate models from EarthByte	https://www.earthbyte.org/paleomap-paleoatlas-for-gplates/, [60]	v2, accessed October 2018
Deposited data	This paper	https://github.com/GawainAntell/EcoRelease
Software and Algorithms		
R computing environment	www.R-project.org, [65]	v3.5.2
GPlates computing environment	www.gplates.org, [61]	v2.1.0
Deposited code	This paper	https://github.com/GawainAntell/EcoRelease
iNEXT package in R	[75]	v2.0.19
lme4 package in R	[72]	v1.1-21
raster package in R	[51]	v2.8-19
rgdal package in R	[52]	v.1.3-6
sp package in R	[53]	v1.3-1
vegan package in R	[62]	v2.5-4

Supplementary Figures and Tables

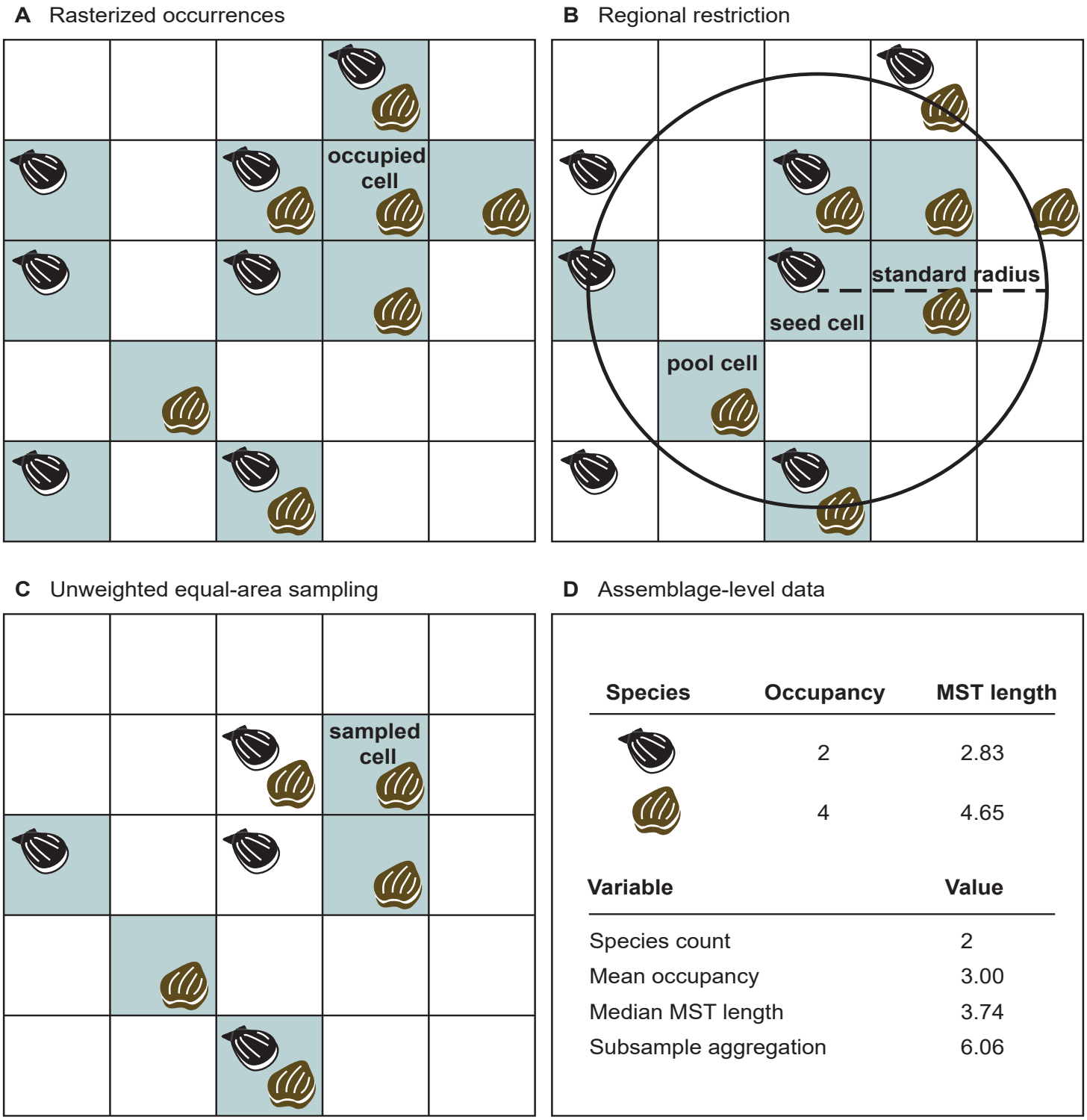
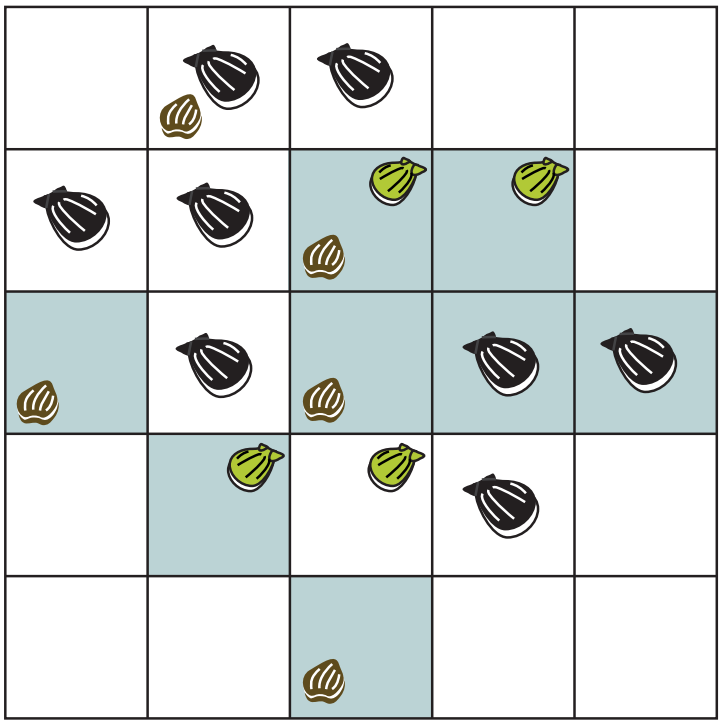
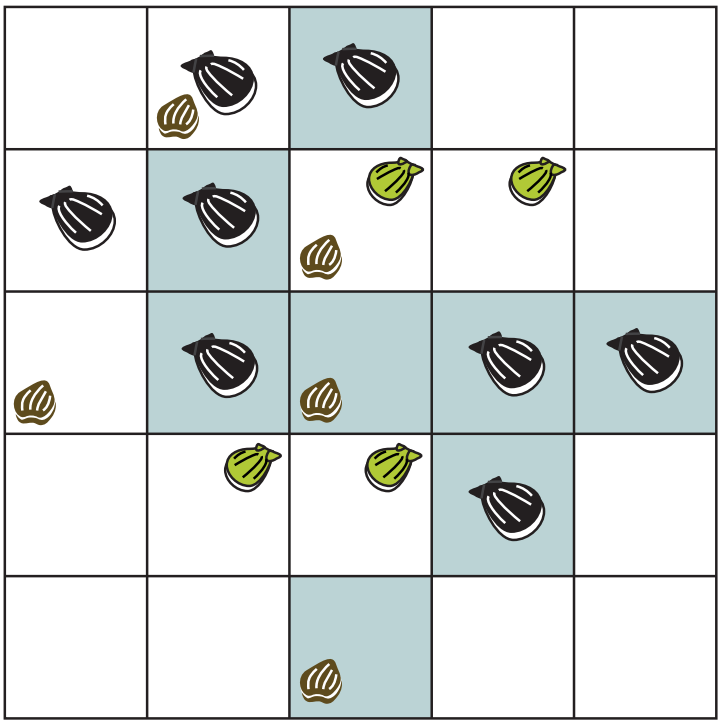


Figure S1. Simplified Diagram of Assemblage Subsampling. Related to STAR Methods and Methods S1. (A) Occurrence coordinates were rasterized. Each silhouette represents a species' presence in a grid cell. Every occupied cell (shaded) is a potential "seed cell". **(B)** A circular buffer of given radius is circumscribed around the centroid of a seed cell. Occupied cells with centroids lying inside the buffer (shaded) comprise the "subsampling pool". **(C)** A given number of cells (here, 5) are drawn from the pool to form a subsample (shaded). **(D)** For each species, range size is measured as occupancy and summed length of the minimum spanning tree that connects the centroids of occupied cells (MST). The lower table in **D** reports all assemblage-based variables used in subsequent analysis: species count, mean occupancy, median MST length, and subsample aggregation.

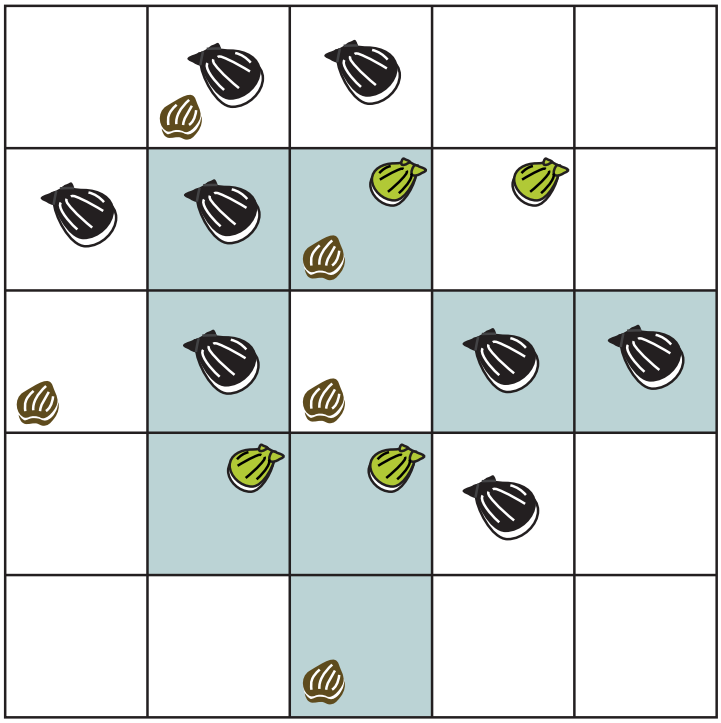
A First iteration



B Second iteration



C Third iteration



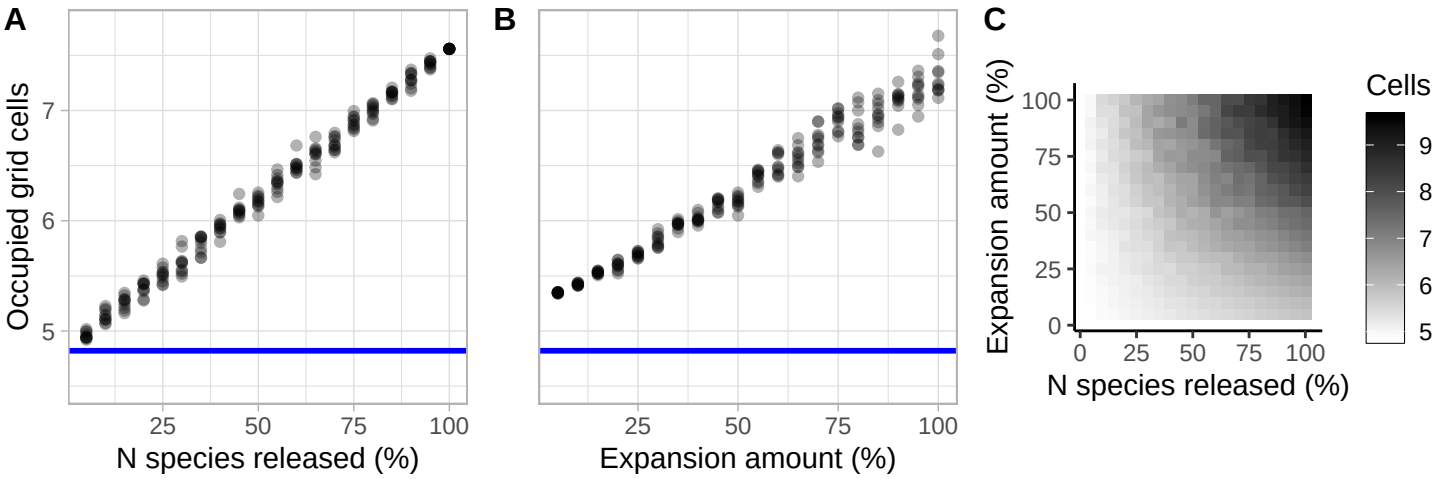
D Focal species data

Iteration	Occupancy	MST length	Species count
1	2	1.00	3
2	6	6.41	2
3	4	4.00	3

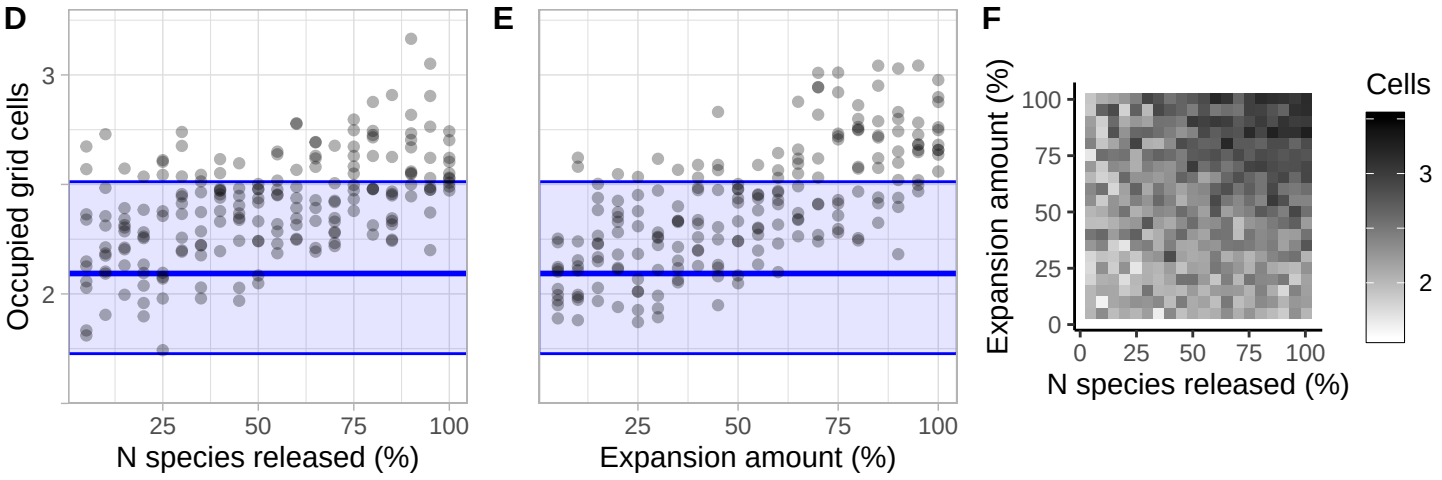
Variable	Median
Occupancy	4
MST length	4.00
Species count	3

Figure S2. Simplified Diagram of Taxon-Tracking Subsampling. Related to STAR Methods and Methods S1. (A–C) Three different realizations of the subsampling procedure illustrated for identical occurrence grids of a focal species (black pectinid). Shading indicates cells selected in a given subsample (for simplicity, each subsample in this diagram contained 8 cells). **(D)** Two metrics of range area were calculated for each iteration: occupancy and summed length of the minimum spanning tree that connects the centroids of occupied cells (MST). The range size of the species, as used in subsequent analysis, was defined as the median metric value among iterations.

Simulated geographic expansion



Observed geographic expansion after subsampling



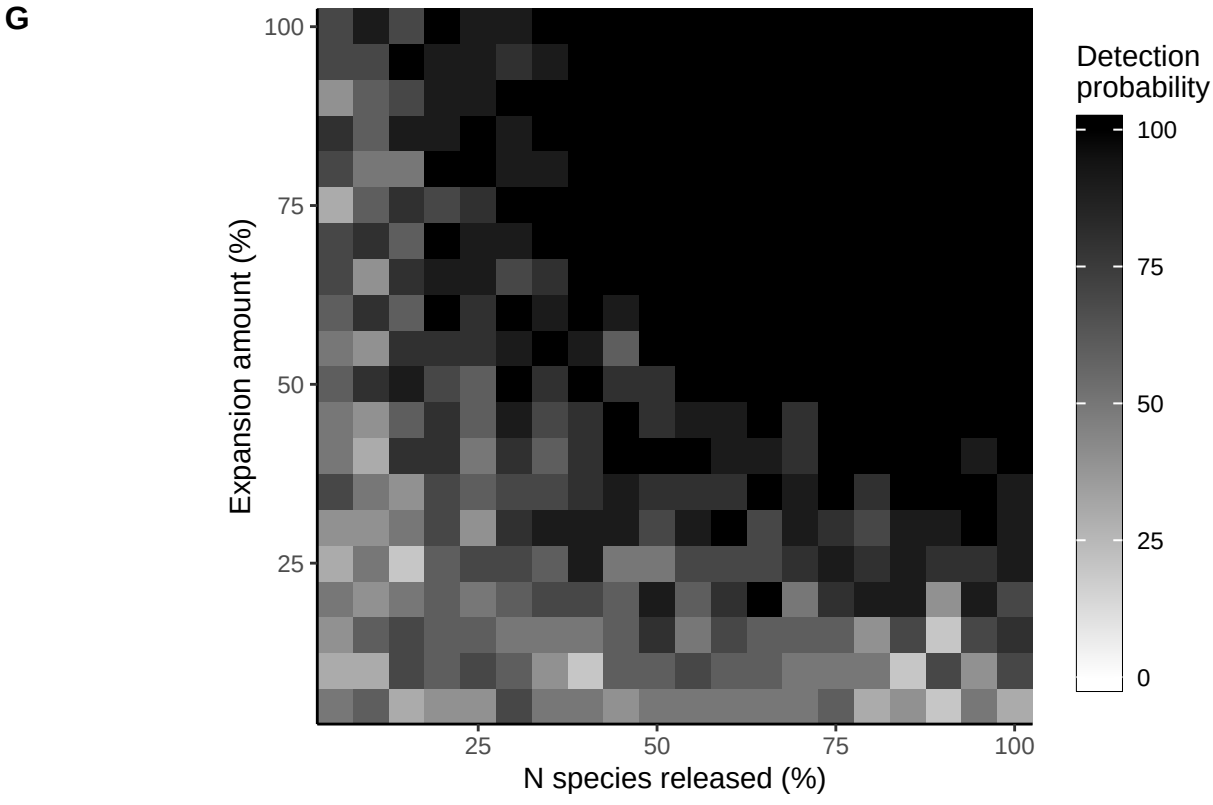


Figure S3. Power Analysis from Simulated Geographic Expansion. Related to STAR Methods and Methods S1. Our spatial standardization procedure truncates geographic distributions, and this compression raises the possibility of type 2 error in our analyses. Therefore, we simulated ecological release at a range of magnitudes and frequencies to calculate the probability of observing release in a subsample of the simulated data. We modeled range expansion based on the Middle–Late Pleistocene record of bivalves in the northwest Atlantic, where we expect that the compression from the subsampling procedure would be particularly severe. To control other biases on fossil range size, such as outcrop area, we used the same Pleistocene area before and after simulated release as the setting in which to calculate occupancy. We replicated each simulation 10 times, since measured range sizes for a given species in a given time interval vary from one iteration to the next. There were two experimental parameters in the simulation framework: the percentage of species that experience ecological release, and the percentage expansion of geographic range size that occurred for released species. In the first set of simulation experiments, no subsampling procedure was applied to occurrences. Species expanded as a linear function of the frequency and magnitude of ecological release, as intended **(A–C)**. In a second set of experiments, species' occupancy was subsampled (by selecting 15 cells from the study area) before and after ecological release. When subsampled, a signal of ecological release was still recovered, although mean occupancy was lower and varied more among iterations **(D–F)**. In **A** and **D**, the amount of release was constant at 50% expansion while the frequency of release varied from 0–100%. In **B** and **E**, the frequency of release was constant at 50% of species while the amount of release varied from 0–100%. The blue lines mark mean occupancy prior to range expansion: 4.8 grid cells without subsampling **(A, B)**, and 2.1 grid cells with subsampling (**D, E**; 95% interval among subsamples in light blue shading, 1.7–2.5 grid cells). Any point that falls above the blue line represents successful detection of release. In **C** and **F**, we simulated occupancy increases while varying the frequency and magnitude of release at the same time. In **G**, the probability of observing release in subsampled data is plotted jointly as a function of the frequency and magnitude of range expansion. Detection is calculated as the percentage of iterations ($n = 10$) in which mean occupancy after simulated release exceeded the initial value (2.1 grid cells) for a given frequency and magnitude of release. For further discussion of the power analysis, please read Methods S1 section 3.

Name	Onset	Terminus	Duration	Occurrences
A. Global Phanerozoic				
Tremadocian, Floian	485.4	470.0	15.4	196
Dapingian, Darriwilian	470.0	458.4	11.6	347
Sandbian, Katian	458.4	445.2	13.2	1226
Hirnantian	445.2	443.8	1.4	394
Rhuddanian, Aeronian	443.8	438.5	5.3	411
Telychian	438.5	433.4	5.1	337
Sheinwoodian, Homerian	433.4	427.4	6.0	1072
Gorstian, Ludfordian	427.4	423.0	4.4	924
Pridoli	423.0	419.2	3.8	313
Lochkovian, Pragian	419.2	407.6	11.6	1141
Emsian, Eifelian	407.6	387.7	19.9	914
Givetian	387.7	382.7	5.0	1293
Frasnian	382.7	372.2	10.5	937
Famennian	372.2	358.9	13.3	371
Tournaisian	358.9	346.7	12.2	641
Visean	346.7	330.9	15.8	607
Serpukhovian, Bashkirian	330.9	315.2	15.7	344
Moscovian, Kasimovian	315.2	303.7	11.5	561
Gzhelian	303.7	298.9	4.8	613
Asselian, Sakmarian	298.9	290.1	8.8	1716
Artinskian, Kungurian	290.1	272.9	17.2	2937
Roadian	272.9	268.8	4.1	1596
Wordian	268.8	265.1	3.7	1621
Capitanian	265.1	259.1	6.0	1342
Wuchiapingian	259.1	254.1	5.0	2451
Changhsingian	254.1	251.9	2.2	2411
Induan, Olenekian	251.9	247.2	4.7	430

Anisian	247.2	242.0	5.2	782
Ladinian	242.0	237.0	5.0	284
Carnian	237.0	227.0	10.0	503
Norian	227.0	208.5	18.5	689
Rhaetian	208.5	201.3	7.2	481
Hettangian, Sinemurian	201.3	190.8	10.5	581
Pliensbachian	190.8	182.7	8.1	1296
Toarcian	182.7	174.1	8.6	866
Aalenian, Bajocian	174.1	168.3	5.8	588
Bathonian	168.3	166.1	2.2	659
Callovian	166.1	163.5	2.6	719
Oxfordian	163.5	157.3	6.2	437
Kimmeridgian	157.3	152.1	5.2	304
Tithonian	152.1	145.0	7.1	486
Berriasian	145.0	139.8	5.2	317
Valanginian, Hauterivian, Barremian	139.8	125.0	14.8	656
Aptian	125.0	113.0	12.0	509
Albian	113.0	100.5	12.5	522
Cenomanian	100.5	93.9	6.6	590
Turonian	93.9	89.8	4.1	187
Coniacian, Santonian	89.8	83.6	6.2	709
Campanian	83.6	72.1	11.5	422
Maastrichtian	72.1	66.0	6.1	3197
Danian	66.0	61.6	4.4	363
Selandian, Thanetian	61.6	56.0	5.6	272
Ypresian	56.0	47.8	8.2	661
Lutetian, Bartonian	47.8	37.8	10.0	1893
Priabonian	37.8	33.9	3.9	993
Rupelian	33.9	27.8	6.1	773

Chattian, Aquitanian	27.8	20.4	7.4	632
Burdigalian	20.4	16.0	4.5	817
Langhian, Serravallian	16.0	11.6	4.3	2056
Tortonian, Messinian	11.6	5.3	6.3	1760
Zanclean	5.3	3.6	1.7	934
Piacenzian	3.6	2.6	1.0	1333
Pleistocene	2.6	0.0	2.6	2204
B. Northwest Atlantic				
Burdigalian	20.44	15.97	4.47	1019
Middle Miocene	15.97	11.63	4.34	639
Tortonian	11.63	7.25	4.38	968
Messinian	7.25	5.33	1.91	923
Zanclean	5.33	3.60	1.73	1447
Piacenzian	3.60	2.58	1.02	2557
Gelasian	2.58	1.80	0.78	1423
Calabrian	1.80	0.78	1.02	1039
Middle–Late Pleistocene	0.78	0.01	0.77	2020

Table S1. Global Phanerozoic and Neogene–Quaternary Time Bins. Related to Figures 1 and 2. (A) The global Phanerozoic dataset was divided into 63 time bins. A species' occurrence record was included in a time bin if the name or age of both its maximum and minimum occurrence estimates fell within the onset and terminus ages (in Ma) for the bin. During binning, ages were rounded to the nearest 0.01 Ma for boundaries younger than 10 Ma, 0.1 Ma for boundaries 50–150 Ma, or 1 Ma for boundaries older than 150 Ma. The number of unique occurrences for each species is tallied for each time bin. **(B)** The northwest Atlantic dataset was divided into 9 time bins. We determined minimum and maximum age estimates for 485 geologic units in the northwest Atlantic to infer the age of fossils with no temporal metadata. We assigned a record to a time bin if more than two-thirds of its estimated range fell within the bin. Records that spanned more than 3 bins were excluded altogether.

Methods S1 Section	Table	Cells	Radius (km)	Iterations	Cell selection	Rarefaction	Assemblage singletons	Taxonomic rank	Other
2.3.1	None	**10**	1500	500	Even	None	Yes	Species	**"Accepted names" only**
2.3.2	None	**10**	1500	500	Even	None	Yes	Species	**No space-time singletons**
2.3.3	S3.a	12	1500	500	Even	None	Yes	Species	Baseline method
2.3.3	None	**6**	1500	500	Even	None	Yes	Species	
2.3.4	S3.b	**6**	**750**	500	Even	None	Yes	Species	
2.3.5	None	12	1500	**1000**	Even	None	Yes	Species	
2.3.6	S3.c	12	1500	500	**Weighted**	None	Yes	Species	Main text results
2.3.7	S3.d	12	1500	500	Even	**Coverage = 0.4**	Yes	Species	
2.3.7	S3.e	12	1500	500	Even	**Coverage = 0.7**	Yes	Species	
2.3.7	S3.f	12	1500	500	Even	**Coverage = 0.7**	**No**	Species	
2.3.7	None	12	1500	500	Even	**Count = 50**	Yes	Species	
2.3.8	S3.g	12	1500	500	Even	None	Yes	**Genus**	
-	S3.h	**15**	**Atlantic**	500	Even	None	Yes	Species	Main text results

Table S2. Parameter Combinations Used in Each Sensitivity Analysis of the Assemblage Subsampling Method. Related to STAR Methods and Methods S1. Boldface indicates values that differ from those of the baseline method. All results are described in the Methods S1 text, although only key results are listed in Table S3.

	Correlation with predictor time series			
	Species count		Sampling aggregation	
Response series	Mean	95%	Mean	95%
A. Baseline				
occ with singletons, min	-0.15	(-0.18, -0.10)	-0.15	(-0.18, -0.06)
occ with singletons, med	-0.05	(-0.23, 0.14)	-0.18	(-0.36, 0.02)
occ with singletons, max	0.06	(-0.07, 0.19)	-0.19	(-0.32, -0.02)
occ with singletons, mean	-0.09	(-0.23, 0.06)	-0.23	(-0.35, -0.08)
occ no singletons, min	-0.15	(-0.21, -0.04)	0.01	(-0.16, 0.17)
occ no singletons, med	-0.15	(-0.33, 0.03)	-0.16	(-0.32, 0.03)
occ no singletons, max	0.06	(-0.07, 0.19)	-0.19	(-0.32, -0.02)
occ no singletons, mean	-0.12	(-0.25, 0.02)	-0.18	(-0.33, -0.03)
MST, min	-0.21	(-0.39, -0.03)	0.13	(-0.05, 0.30)
MST, med	-0.05	(-0.18, 0.11)	0.06	(-0.08, 0.21)
MST, max	0.15	(0.01, 0.29)	0.23	(0.11, 0.37)
MST, mean	-0.04	(-0.17, 0.11)	0.13	(-0.01, 0.26)
species count	1.00	(1.00, 1.00)	0.04	(-0.09, 0.18)
B. 6 cells, 750 km radius				
occ no singletons, mean	-0.08	(-0.23, 0.10)	-0.18	(-0.33, -0.03)
MST, median	-0.05	(-0.20, 0.11)	0.17	(0.01, 0.33)
species count	1.00	(1.00, 1.00)	0.04	(-0.12, 0.19)
C. Weighted cell selection				
occ no singletons, mean	-0.11	(-0.25, 0.01)	-0.17	(-0.32, -0.03)
MST, median	-0.06	(-0.19, 0.08)	0.09	(-0.07, 0.24)
species count	1.00	(1.00, 1.00)	-0.01	(-0.14, 0.12)

D. SQS, coverage of 0.4				
occ no singletons, mean	-0.12	(-0.26, 0.02)	-0.19	(-0.34, -0.04)
MST, median	-0.06	(-0.20, 0.08)	0.07	(-0.09, 0.21)
species count	1.00	(1.00, 1.00)	0.04	(-0.09, 0.18)
E. SQS, coverage of 0.7				
occ no singletons, mean	-0.11	(-0.25, 0.01)	-0.19	(-0.34, -0.04)
MST, median	-0.05	(-0.20, 0.09)	0.06	(-0.09, 0.21)
species count	1.00	(1.00, 1.00)	0.02	(-0.12, 0.15)
F. SQS, coverage of 0.7, no assemblage singletons				
occ no singletons, mean	0.11	(-0.03, 0.25)	-0.18	(-0.33, -0.05)
MST, median	-0.08	(-0.23, 0.05)	0.07	(-0.09, 0.21)
species count	1.00	(1.00, 1.00)	-0.09	(-0.23, 0.05)
G. Genus-level				
occ no singletons, mean	-0.10	(-0.25, 0.03)	-0.18	(-0.31, -0.04)
MST, median	-0.05	(-0.19, 0.10)	0.04	(-0.11, 0.19)
species count	1.00	(1.00, 1.00)	0.03	(-0.12, 0.17)
H. Northwest Atlantic				
occ no singletons, mean	0.18	(-0.17, 0.53)	-0.51	(-0.67, -0.33)
MST, median	-0.13	(-0.59, 0.31)	0.06	(-0.33, 0.50)
species count	1.00	(1.00, 1.00)	-0.18	(-0.56, 0.21)

Table S3. Time Series Corrected Correlation Coefficients. Related to Figures 1 and 2 and Methods S1. Kendall's τ correlation for every pair of predictor and response time series, for parameter combinations listed in Table S2. The metrics of range size were grid cell occupancy (oc) and summed length of minimum spanning tree (MST).

Early bin	Species count	Range shift	Shift SD	Survivor n
A. Occupied grid cells, mid-existence species				
Rhuddanian	Rise	0.50	0.58	4
Telychian	Rise	1.00	1.46	18
Emsian	Rise	1.36	1.57	11
Moscovian	Rise	0.89	2.62	9
Gzhelian	Rise	-0.92	2.06	13
Asselian	Rise	-0.60	1.34	45
Artinskian	Fall	1.33	2.31	3
Wordian	Fall	0.31	0.95	13
Wuchiapingian	Rise	0.21	0.73	56
Changhsingian	Fall	-0.57	1.81	7
Anisian	Fall	-1.43	1.79	14
Hettangian	Fall	-0.40	2.12	10
Pliensbachian	Fall	-0.15	0.84	39
Toarcian	Fall	0.25	1.50	4
Aalenian	Rise	2.00	2.83	2
Bathonian	Fall	0.75	0.96	4
Callovian	Rise	-0.50	1.87	6
Coniacian	Rise	-1.00	0.89	11
Ypresian	Rise	-0.11	1.88	18
Langhian	Fall	-0.08	1.08	12
Piacenzian	Fall	0.70	1.85	43
B. Minimum spanning tree length, mid-existence species				
Rhuddanian	Rise	0.00	0.00	2
Telychian	Rise	7.19	8.18	8
Emsian	Rise	8.06	3.89	4
Moscovian	Rise	3.98	11.43	5
Gzhelian	Rise	-9.32	15.26	8

Asselian	Fall	-8.98	6.24	3
Wordian	Fall	1.64	4.03	4
Wuchiapingian	Rise	3.85	8.98	11
Changhsingian	Fall	-4.13	3.24	3
Anisian	Fall	-7.22	6.33	6
Pliensbachian	Rise	-6.81	7.64	5
Bathonian	Fall	3.95	7.85	3
Callovian	Rise	-3.90	13.39	4
Coniacian	Fall	-10.12	14.29	7
Ypresian	Rise	6.19	11.54	11
Langhian	Fall	9.13	20.72	2
Piacenzian	Fall	5.36	17.18	10
	C. Occupied grid cells, all species			
Sandbian	Fall	4.50	0.71	2
Rhuddanian	Rise	0.82	1.33	11
Telychian	Rise	0.71	1.46	28
Lochkovian	Rise	2.50	0.71	2
Emsian	Rise	1.55	1.84	22
Tournaisian	Rise	1.00	0.00	2
Moscovian	Rise	2.12	2.40	24
Gzhelian	Rise	-1.09	2.01	35
Asselian	Rise	-0.47	1.39	109
Artinskian	Fall	1.00	1.55	6
Wordian	Fall	0.28	1.56	36
Wuchiapingian	Rise	-0.10	0.89	143
Changhsingian	Fall	-0.60	1.58	10
Induan	Rise	-0.50	0.71	2
Anisian	Fall	-1.12	1.86	16
Rhaetian	Rise	-1.00	2.83	2
Hettangian	Fall	-0.06	1.89	18

Pliensbachian	Rise	-0.29	0.95	56
Toarcian	Fall	0.40	1.34	5
Aalenian	Rise	1.40	1.67	5
Bathonian	Fall	0.38	0.92	8
Callovian	Rise	-1.09	1.51	11
Coniacian	Rise	-0.67	0.84	18
Ypresian	Rise	-0.12	1.74	25
Langhian	Fall	0.00	1.04	14
Piacenzian	Fall	0.70	2.15	98
D. Minimum spanning tree length, all species				
Rhuddanian	Rise	10.01	12.65	4
Telychian	Rise	0.01	15.64	13
Emsian	Rise	10.10	4.35	12
Moscovian	Rise	7.45	9.72	17
Gzhelian	Rise	-10.50	12.89	21
Asselian	Rise	-2.81	12.04	12
Artinskian	Fall	5.65	1.33	2
Wordian	Fall	-0.78	11.16	14
Wuchiapingian	Rise	-1.13	10.05	36
Changhsingian	Fall	-5.09	3.28	4
Anisian	Fall	-7.22	6.33	6
Hettangian	Fall	14.46	7.68	5
Pliensbachian	Rise	-9.11	5.77	11
Bathonian	Fall	2.24	5.98	6
Callovian	Fall	-7.61	12.91	6
Coniacian	Fall	-9.39	12.04	12
Ypresian	Rise	5.54	12.16	15
Langhian	Fall	9.13	20.72	2
Piacenzian	Rise	4.44	12.82	37

Table S4. Range Size Shifts for Each Phanerozoic Interval. Related to Figure 3. Species count dynamic and geographic range size shift as measured in occupancy **(A, C)** or minimum spanning tree length (square-root km) **(B, D)**. In **A and B**, newly originated and nearly extinct species are excluded. To calculate geographic range size trends for an interval, multiple species had to occur in successive time bins that contained at least 10 cells of suitable habitat. For minimum spanning tree calculations, a species had to occur more than once in each bin. The shift is the average and standard deviation (SD) of change in number of grid cells among survivors. Each time interval was characterized by either a rise or fall in the number of species, depending on the median change in species count associated with survivors. Each interval is referenced by the oldest geologic stage included. Weighted Wilcoxon test results for mid-existence species **(A and B)** are reported in Figure 3. When species in the first and last intervals of appearance were included **(C and D)**, two taxa that originated in the Sandbian (Upper Ordovician) were strong outliers, expanding by 4.5 grid cells on average. These points contributed to an n-weighted occupancy change of 0.35 and -0.04 grid cells in periods of species count decrease ($n = 10$) and increase ($n = 16$), respectively (p-value = 0.06, t-value= -1.9, df = 24). For measurements in MST length, less than two species were sampled in at least 10 grid cells both before and after the Sandbian, so range shift in this interval could not be assessed. The weighted mean change between sampled periods of species count decrease ($n = 9$) and increase ($n = 10$) was -2.06 and 0.79 km, respectively (p-value = 0.42, t-value = 0.83, df = 17).

Rank	Predictors	N predictors	ΔAIC	Weight (%)
	A. Occupied grid cells			
1	delta + point + time	7	0.00	23.64
2	point + delta	6	0.96	14.63
3	delta + point + clade + time	8	2.00	8.70
4	delta + point + lith + time	9	2.15	8.07
5	delta + point + lith	8	2.20	7.87
6	delta + point + clade	7	2.74	6.01
7	delta + point + bath + time	9	3.43	4.25
8	delta + point + clade + lith + time	10	4.08	3.07
9	delta + point + clade + lith	9	4.18	2.92
10	delta + point + bath	8	4.33	2.71
11	time + point	6	4.49	2.50
12	point + lith + time	8	5.13	1.82
13	delta + point + clade + bath + time	10	5.42	1.57
14	lith + point	7	5.50	1.51
15	delta + point + lith + bath + time	11	5.54	1.48
16	delta + point + lith + bath	10	5.56	1.47
17	delta + point + clade + bath	9	6.17	1.08
18	point	5	6.18	1.08
19	point + clade + time	7	6.47	0.93
20	point + clade + lith + time	9	7.04	0.70
21	delta + point + clade + lith + bath + time	12	7.44	0.57
22	point + clade + lith	8	7.49	0.56
23	delta + point + clade + lith + bath	11	7.55	0.54
24	clade + point	6	7.73	0.50
25	point + bath + time	8	7.83	0.47
26	point + lith + bath + time	10	8.47	0.34
27	point + lith + bath	9	8.85	0.28
28	bath + point	7	9.52	0.20
29	point + clade + bath + time	9	9.83	0.17

B. Minimum spanning tree length

1	point + delta	6	0.00	13.03
2	delta + point + bath	8	0.50	10.15
3	delta + point + time	7	1.10	7.52
4	delta + point + clade + time	8	1.21	7.12
5	delta + point + clade	7	1.37	6.57
6	delta + point + lith + bath	10	1.62	5.80
7	delta + point + clade + bath	9	1.75	5.43
8	delta + point + clade + lith + bath + time	12	1.78	5.35
9	delta + point + lith + bath + time	11	1.86	5.14
10	delta + point + bath + time	9	1.94	4.94
11	delta + point + clade + bath + time	10	2.12	4.52
12	delta + point + clade + lith + time	10	2.16	4.43
13	delta + point + lith + time	9	2.27	4.19
14	delta + point + lith	8	2.38	3.97
15	delta + point + clade + lith + bath	11	3.12	2.74
16	delta + point + clade + lith	9	3.96	1.80
17	point + lith + bath	9	5.61	0.79
18	bath + point	7	5.82	0.71
19	point + clade + bath	8	6.05	0.63
20	point + clade + lith + bath + time	11	6.14	0.61
21	point + clade + lith + bath	10	6.60	0.48
22	point + lith + bath + time	10	6.66	0.47
23	point	5	6.92	0.41
24	point + clade + bath + time	9	7.08	0.38
25	clade + point	6	7.27	0.34
26	point + bath + time	8	7.75	0.27
27	point + clade + time	7	7.88	0.25
28	lith + point	7	8.37	0.20
29	point + clade + lith + time	9	8.60	0.18
30	time + point	6	8.71	0.17

31	point + lith + time	8	9.22	0.13
32	point + clade + lith	8	9.41	0.12
33	delta	4	9.90	0.09
34	delta + lith + bath	8	9.95	0.09

Table S5. Relative Performance of Non-trivial Linear Mixed-effects Models. Related to Figure 4 and STAR Methods. The response variable was **(A)** the change in number of occupied grid cells, including singletons, or **(B)** minimum spanning tree length (MST; square-root km). For **A**, models were built on 708 changes in range size recorded for survivors from 26 Phanerozoic boundaries. For **B**, models were built on 239 changes in range size recorded for survivors from the 23 intervals for which tree length could be calculated. Abbreviations: delta, change in species count; time, age of boundary (Ma); point, categorical variable with the values first, middle, and last to indicate newly originated, mid duration, or nearly extinct; clade, membership in Brachiopoda or Mollusca; bath, bathymetric preference for deep, shallow, or both water depths; lith, lithological preference for carbonate, siliciclastic, or both lithologies. When geographic range size was measured in terms of number of grid cells occupied, lithological preferences appeared more important than did bathymetric preferences **(A)**. However, when range size was measured in terms of MST length, preference for water depth replaced preference for lithology as an important predictor in the best-supported models **(B)**. Weights are the ratio of ΔAIC from a given model to the sum of ΔAIC values across all candidate models, and are interpreted as the probability that a given model is the "best" (minimizes the Kullback–Leibler discrepancy) of the candidate models, given the data [S1].

Rank	Predictor	Estimate	95%
	A. Occupied grid cells		
1	(Intercept)	-0.489	(-1.278, 0.276)
	delta	0.002	(0.001, 0.004)
	point First	0.732	(0.362, 1.110)
	point Last	-0.355	(-0.624, -0.082)
	time	0.002	(-0.000, 0.005)
2	(Intercept)	0.099	(-0.275, 0.492)
	point First	0.757	(0.385, 1.135)
	point Last	-0.331	(-0.600, -0.061)
	delta	0.002	(0.001, 0.004)
3	(Intercept)	-0.498	(-1.363, 0.351)
	delta	0.002	(0.001, 0.004)
	point First	0.731	(0.359, 1.111)
	point Last	-0.355	(-0.624, -0.082)
	clade Mollusca	0.007	(-0.332, 0.343)
	time	0.003	(-0.000, 0.006)
4	(Intercept)	-0.361	(-1.208, 0.470)
	delta	0.002	(0.000, 0.004)
	point First	0.742	(0.370, 1.121)
	point Last	-0.349	(-0.618, -0.077)
	lith Carbonate	-0.017	(-0.399, 0.345)
	lith Clastic	-0.338	(-0.793, 0.142)
	time	0.002	(-0.001, 0.005)
5	(Intercept)	0.174	(-0.227, 0.588)
	delta	0.002	(0.000, 0.004)
	point First	0.765	(0.394, 1.144)
	point Last	-0.327	(-0.596, -0.058)
	lith Carbonate	-0.046	(-0.424, 0.320)
	lith Clastic	-0.394	(-0.847, 0.069)

B. Minimum spanning tree length

1	(Intercept)	0.156	(-3.034, 3.441)
	point First	3.904	(-0.546, 8.440)
	point Last	-5.336	(-8.808, -1.816)
	delta	0.038	(0.013, 0.062)
2	(Intercept)	-0.828	(-4.114, 2.522)
	delta	0.034	(0.009, 0.058)
	point First	3.259	(-1.188, 7.838)
	point Last	-5.690	(-9.166, -2.193)
	bath Deep	4.007	(-3.224, 11.229)
	bath Shallow	3.247	(-0.636, 7.205)
3	(Intercept)	2.725	(-3.535, 9.124)
	delta	0.039	(0.015, 0.064)
	point First	4.040	(-0.365, 8.660)
	point Last	-5.375	(-8.831, -1.859)
	time	-0.011	(-0.034, 0.012)
4	(Intercept)	6.079	(-1.760, 13.928)
	delta	0.037	(0.013, 0.062)
	point First	4.123	(-0.262, 8.721)
	point Last	-5.190	(-8.646, -1.689)
	clade Mollusca	-3.120	(-7.595, 1.341)
	time	-0.019	(-0.044, 0.007)
5	(Intercept)	0.979	(-2.784, 4.817)
	delta	0.036	(0.011, 0.061)
	point First	3.856	(-0.581, 8.387)
	point Last	-5.238	(-8.713, -1.715)
	clade Mollusca	-1.686	(-5.713, 2.391)

Table S6. Coefficient Estimates and Confidence Intervals. Related to Figure 4 and STAR Methods. The log-likelihood estimates and 95% confidence intervals were extracted for each predictor in the 5 models with largest weights (Table S5), with change in geographic range size as response, measured either as **(A)** number of occupied grid cells (including singletons) or **(B)** minimum spanning tree length (MST; square-root km). Abbreviations and sample sizes as in Table S5. The cumulative weight of the top 5 occupancy models was 62%, and for MST models was 44% (Table S5). The clarity of regression predictions varied depending on whether geographic range size was measured in terms of occupied grid cells or MST length. When range size was measured in terms of grid cells, expansion of newly-originated species was approximately twice as large as contraction of nearly-extinct species **(A)**, an asymmetry found previously [S2, S3]. In contrast, the magnitude of MST expansion early in a species' duration was smaller than the magnitude of later contraction **(Table S6)**. Additionally, specialists in either of the two categories of substrate lithology (carbonate or siliciclastic) tended to shrink in grid cell area compared to generalists when range size was measured in terms of grid cells **(A)**. Surprisingly, specialists in either shallow or deep water tended to expand in MST length compared to bathymetric generalists **(B)**. The confidence intervals related to water depth preferences are wide, however, since there were few specialists for which the number of grid cells with suitable habitat was large enough to perform subsampling (89 shallow-water species but only 18 deep-water species, compared to 131 generalists).

Response	Predictor	Estimate	2.5 %	97.5 %
Occ	Phase: Early	0.733	0.359	1.117
Occ	Phase: Late	-0.354	-0.623	-0.081
Occ	N spp. change	0.002	0.000	0.004
Occ	Time	0.002	-0.001	0.006
Occ	Clade: Bivalve	0.065	-0.298	0.413
Occ	Subst: Carbonate	-0.020	-0.420	0.353
Occ	Subst: Clastic	-0.358	-0.819	0.133
Occ	Depth: Deep	0.239	-0.368	0.8ine 52
Occ	Depth: Shallow	0.032	-0.273	0.344
Occ	(Intercept)	-0.448	-1.385	0.480
MST	Phase: Early	3.317	-1.065	8.121
MST	Phase: Late	-5.145	-8.621	-1.705
MST	N spp. change	0.032	0.007	0.057
MST	Time	-0.027	-0.056	0.002
MST	Clade: Bivalve	-3.297	-8.124	1.248
MST	Subst: Carbonate	-2.138	-7.729	2.871
MST	Subst: Clastic	-7.405	-13.653	0.147
MST	Depth: Deep	3.891	-3.446	11.223
MST	Depth: Shallow	3.992	-0.118	8.015
MST	(Intercept)	8.785	-0.633	18.047

Table S7. Log-likelihood Estimates and Confidence Intervals for Each Predictor in the Linear Mixed-effects Model with All Regressor Variables Included. Related to Figure 4. Scaled estimates are plotted in Figure 4. The 95% confidence intervals were derived from profiling. Estimates are relative to mid-existence (vs. early or late) brachiopod species (vs. bivalve) with no preference for habitat (vs. preference for substrate or water depth). Substrate preference (subst.) is either carbonate (e.g. reef) or siliciclastic (clastic, e.g. sand). Separate models were built for geographic distributions measured in occupancy (Occ, n = 708) or square-root minimum spanning tree length (MST, n = 239).

Supplemental References

S1. Burnham, K.P. and Anderson, D.R. (2010). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd edition (New York, NY: Springer).

S2. Liow, L.H. and Stenseth, N.C. (2007). The rise and fall of species: implications for macroevolutionary and macroecological studies. Proc. R. Soc. Lond. B *274*, 2745–2752.

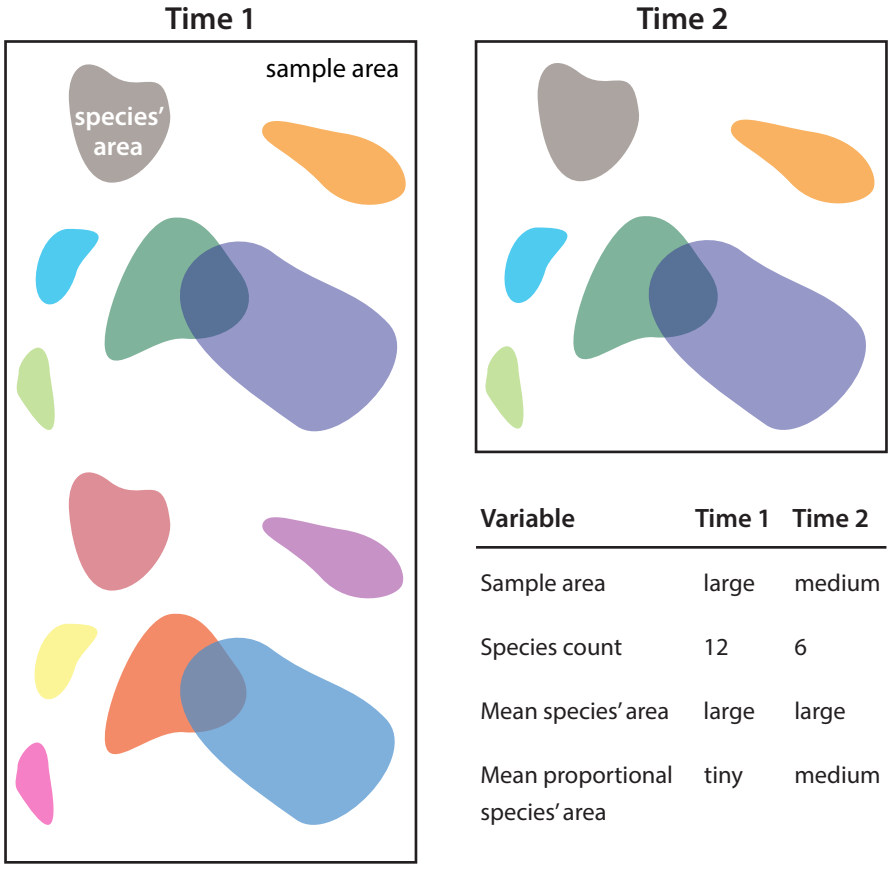
S3. Harnik, P.G., Simpson, C., and Payne, J.L. (2012). Long-term differences in extinction risk among the seven forms of rarity. Proc. R. Soc. Lond. B *279*, 4969–4976.

Methods S1

Theoretical background for, and practical development of, a procedure to standardize spatial sampling coverage. Related to STAR Methods, Figures 1 and S1–3, and Table S2–3.

1 Theoretical concerns around spatial coverage

Statistical comparisons of spatio-temporal occurrences assume that sampling universes are equivalent across space and through time. Fossil data violate this assumption consistently, because preservation potential, field sampling intensity, and amount of preserved rock area structure the record of species' distributions and diversity [S4–S10]. Furthermore, biogeographic responses vary by geographic region [S11, S12], and diversity patterns vary with the spatial scale of analysis, from local sites (α diversity) to a whole study region or biosphere (γ diversity). Turnover, the factor by which γ exceeds mean α, is termed β [S13]. β diversity increases as either the total area of sampling or the dispersion of sampling locations increases [S14]. This biological link between diversity and spatial coverage, referred to as the species-area effect, pervades all spatial grains throughout the Phanerozoic [S15, S16]. Because our question concerns the relationship between range size and diversity through time, spatial coverage standardization in terms of both total sampling area and dispersion of sampling sites was exigent to our analysis.



Variable	Time 1	Time 2
Sample area	large	medium
Species count	12	6
Mean species' area	large	large
Mean proportional species' area	tiny	medium

A simple way to account for unequal sample area is to divide a species' distribution area by the total sample area (e.g. [S3, S17, S18]). This standardized measure, sometimes called proportional occupancy, could account adequately for unequal sampling area if species' ranges were compared contemporaneously, or if species composition were homogeneous among sites

(no turnover). In practice, paleontologists often track geographic ranges through time, and species turn over among sites. A proportional measure of occupancy may not be desirable in these cases [S19]. In the context of our biological hypothesis, converting range area to a proportion of total sampling area could induce a negative correlation between time series of range size and species count as an artifact of sampling. A large sampling area will tend to contain more species than a small area. Meanwhile, if the frequency distribution of species' geographic range sizes is identical between the two regions, but these sizes are converted to proportions of the total sampling area, then the large study area will appear to contain species with small geographic ranges. Furthermore, in the following section we prove the statement that a difference in proportional occupancy of widespread species will be greater than that of restricted species when a study area increases beyond the extent of a species' geographic range.

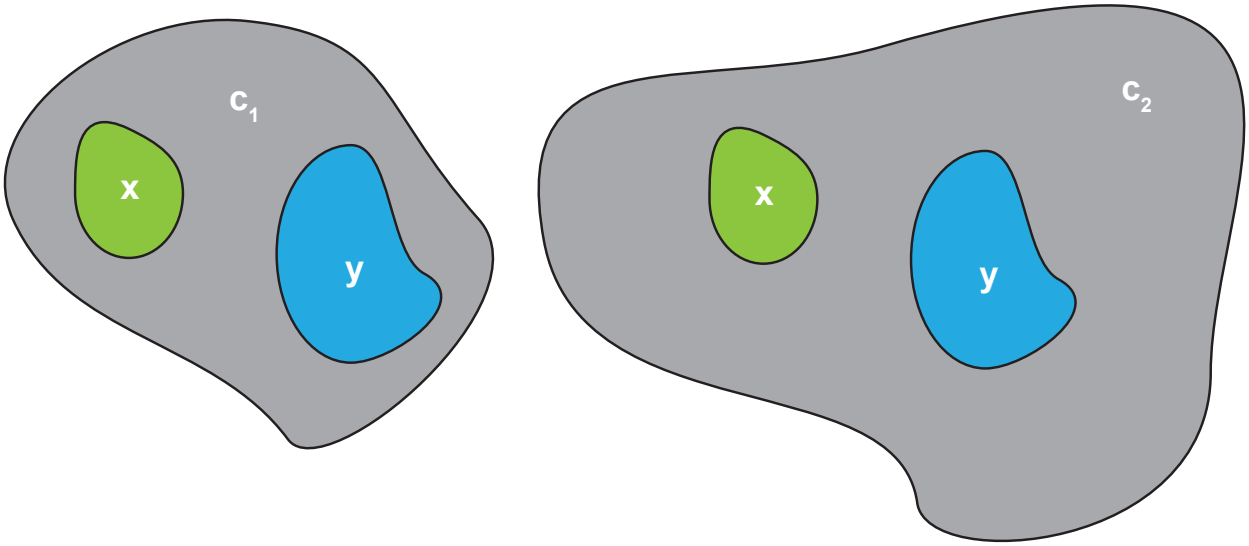
Even with constant sample area through time, variation in spatial dispersion of sampling locations could cause a negative correlation between range area and sampling amount [S20]. When localities are rarefied, each site could reflect a distinct ecological community, resulting in estimates of small distribution areas and large counts of species. In contrast, when localities are aggregated, species' distributions might be sampled more completely, but many more species would be shared among sample sites. In this latter scenario, one would observe large distribution areas and small counts of species. As with proportional occupancy, the negative relationship between geographic range size and species count that could result from variation in the aggregation of sampling locations might inflate the type 1 error rate of our analysis.

1.1 Quantitative notes on proportional occupancy

The previous section reviews proportional occupancy as a common approach to address variation in spatial sampling through time. Problems arise, however, when the metric is used to compare range sizes across time intervals, rather than within an interval. We explore these problems mathematically in this section. To begin, we specify two criteria that a range size metric must meet in order to be suitable for analysis that is simultaneously spatial and temporal: (i) the metric should be equal among all species that do not change in distribution, regardless of the size of the species' range, and (ii) the effect of sampling area should be minimal.

When total sampling area increases, the difference in proportional occupancy of widespread species will be greater than that of restricted species, a violation of the first criterion for a fair metric. Calculating changes in values as a relative difference (rather than absolute difference) avoids this situation. However, calculation of a relative change in proportions exaggerates the influence of sampling area, which violates the second criterion. The proof follows.

Statement. Let species X and Y have distributions with area x and y, respectively, such that $x < y$. Let both species occur exclusively within the preserved rock area in each of two time periods. Let the outcrop areas be c_1 and c_2, such that $c_1 < c_2$.



Then:

1. The difference in species' proportional occupancy of areas c_1 and c_2 will be smaller for species X than for species Y. We call this difference ΔA_i for a species' area i.
2. The relative change in species' proportional occupancy of areas c_1 and c_2 will be equal for species X and Y. We call this difference δA_i.
3. The magnitude of the change in proportional occupancy of areas c_1 and c_2 for either species will be larger when this change is calculated as a proportion than as a difference.

Proof.

Proof. We prove each of the 3 parts of the above statement.

1. For any species with area i, the proportional occupancy of area c is i/c. Using the definition of ΔA_i above, the difference in proportional occupancy of species X and Y in areas c_1 and c_2 can be written

$$\Delta A_i = \frac{i}{c_1} - \frac{i}{c_2} = \frac{ic_2 - ic_1}{c_1 c_2} = \frac{i(c_2 - c_1)}{c_1 c_2} \quad \forall i \in [x, y].$$

   Let k be a constant defined as $k = (c_2 - c_1)/(c_1 c_2)$. We can write $\Delta A_x = xk$ and $\Delta A_y = ay$. It is given that $0 < c_1 < c_2$, so $k > 0$. It is also given that $0 < x < y$. Thus, $\Delta A_x < \Delta A_y$.

2. The relative change in proportional occupancy of species X and Y in areas c_1 and c_2 can be written

$$\delta A_i = \frac{i}{c_1} \div \frac{i}{c_2} = \frac{ic_2}{ic_1} = \frac{c_2}{c_1} \quad \forall i \in [x, y].$$

   This is equivalent to stating that $\delta A_x = \delta A_y$.

3. Using the expressions for ΔA_i and δA_i in parts 1 and 2 above,

$$\Delta A_i = \frac{i(c_2 - c_1)}{c_1 c_2} \text{ and } \delta A_i = \frac{c_2}{c_1}$$

It is given that $0 < i < c_1 < c_2 \quad \forall i \in [x, y]$. Therefore,

$$c_1 c_2(\Delta A_i) = i(c_2 - c_1) = ic_2 - ic_1 < ic_2 < c_2 c_2 = c_1 c_2(\delta A_i)$$

It follows that $\delta A_i > \Delta A_i$ for any species with distribution area i smaller than c_1.

□

Any study that aims to quantify spatial patterns through time fairly must standardize the area of sampling. We prove in this section that transforming geographic distributions to a proportion of sampling area is not a fair way to mitigate the heterogeneity of fossil occurrences. Therefore, we developed a randomized subsampling routine that standardizes for spatial coverage without invoking proportions. The procedure draws subsamples within the confinement of a single region; the size of species' distributions in a subsample does not depend on the size of the sampling universe outside this region (Figure S1). As a result of the subsampling approach, a species' distribution that does not change from one time to another will be captured as stable, satisfying the first criterion above. Furthermore, the procedure ensures that subsamples are equal in area, which is equivalent to forcing the denominator of a proportion to be a constant through time. This satisfies the second criterion. The subsampling routine is a type of count-based rarefaction, and therefore its fairness comes with the trade-off of compression (i.e. lessened effect size). Nevertheless, we demonstrate that analysis can detect the direction of a signal accurately (section 3 below).

2 Development of a spatial subsampling method

Theoretical problems related to variable spatial sampling are considered above. To analyze the empirical data presented in the main text, we needed to develop a practical means of standardizing spatial sampling coverage. The final approach is described in the STAR Methods and illustrated in Figures S1 and S2. In brief, we standardized the dispersion of sampling by restricting analysis to a single region at a time, and then standardized the area of sampling by selecting a given number of equal-area grid cells within the region. In the next two sections, we list step-by-step instructions to spatially subsample data in the assemblage and taxon-tracking frameworks, respectively. Then, we justify the choice of values for each parameter in the subsampling method (section 2.3), e.g. the number of grid cells selected.

2.1 Assemblage subsampling routine

1. **Regional restriction** (Figure S1 A and B). Select one grid cell populated with a fossil (i.e. an "occupied cell"). Draw a circular buffer with a radius of 1500 km from the centroid of this "seed cell". Define the subsampling pool as the contemporaneous occupied cells (including seed cell) that have centroids within the buffered area. If the sample size of the pool meets or exceeds 12 cells, save the position of the seed cell. Repeat this step for all occupied cells in each time bin.

2. a. **Unweighted equal-area sampling** (Figure S1 C). Randomly select one seed cell saved from step 1. Draw randomly 12 cells without replacement from the subsampling pool associated with the seed cell. Repeat this step once in every time bin.

   b. **Proximity-weighted equal-area sampling.** Randomly select one seed cell saved from step 1. Compute the great-circle distance matrix between the centroid of each pool cell and the centroid of the seed cell. Define the probability of sampling a pool cell as the inverse of its squared distance from the seed cell. Include the seed cell in the subsample in addition to 11 pool cells drawn with weighted probabilities (without replacement).

3. **Assemblage-level data** (Figure S1 D). For each subsample, calculate the square-root of the length of the MST connecting all subsampled cells ("sampling aggregation"). For each species that occurs more than once, count the number of occupied cells and calculate the square-root of the length of the MST connecting occupied cell centroids. Save the sampling aggregation, species count, mean occupancy among species, and median MST length among species.

4. **Iterate.** Repeat steps 2–3 for a total of 500 replicates. Each time series of subsamples represents one subsampling iteration.

Due to the stochastic nature of the subsampling procedure, it is possible that an assemblage could consist entirely of species represented by a single occurrence. In such a case it would not be possible to calculate any MST lengths. Therefore, we excluded any replicate time series that included this type of assemblage. There were only two such series in the realized subsampling runs for the analysis presented in the main text.

Note also that if rarefaction methods are used to calculate species count (section 2.3.7 below), additional components exist for steps 2 and 3. In step 2 (A or B), check that the subsample contains the minimum number of records for classical rarefaction or the coverage level required for shareholder quorum subsampling. If the subsample is not large enough, take a new subsample. In step 3, use the distribution of presences (occupancy counts) to subsample the number of species to a desired coverage level or count.

2.2 Taxon-tracking subsampling routine

The below procedure applies to a focal species in a time bin. We iterated the procedure across species and time bins to generate a taxon-tracking dataset.

1. **Suitable habitat.** Cull the raster of occupied cells in the time bin such that only cells with habitat suitable for the focal species are included (if the species is a generalist along both axes, no action is necessary).

2. **Regional restriction.** This step is the same as in the assemblage-based procedure (Figure S1 A and B). Select one occupied cell. Draw a circular buffer with a radius of 1500 km from the centroid of this "seed cell". Define the subsampling pool as the contemporaneous occupied cells (including seed cell) that have centroids within the buffer area. If the sample size of the pool meets or exceeds 10 cells, save the position of the seed cell. Repeat this step for all occupied cells in the time bin.

3. **Unweighted equal-area sampling** (Figure S1 C). Randomly select one seed cell saved from step 1, and then randomly draw 9 additional cells without replacement from the pool cells associated with the seed cell. Repeat this step 500 times in every time bin.

4. **Species-level data** (Figure S2). In each time bin, find the subsamples where the focal species occurs. In each of those subsamples, calculate:

   - the number of cells occupied by the focal species (once by including and once by excluding single occurrences)
   - the square-root of the length of the MST connecting centroids of cells occupied by the focal species
   - the number of species in the subsample (including the focal species)

   Save the median value of each of these three metrics across subsamples associated with the focal species.

2.3 Sensitivity to subsampling parameters

The spatial standardization procedure contains a variety of user-defined parameters (e.g. number of iterations). We varied each such parameter in isolation, as summarized in Table S2, while developing the subsampling method. For instance, we experimented with stricter taxonomic vetting, different grid cell quotas, classical or coverage-based species richness rarefaction, and aggregation by genus names. We analyzed time series from each variation of the subsampled dataset.

As a baseline against which to compare results from sensitivity analyses, we report correlations from subsamples generated under the following parameters: 1500 km radial constraint, 12 grid cells, 500 iterations, and unweighted cell selection (Table S3 A). The area of 12 grid cells totals 1.1×10^5 km^2, approximately the land area of Honduras. The baseline considered many metrics to summarize range size among species: minimum, median, maximum, and average of occupancies or MST lengths; for occupancy, values are reported both when single-occurring species ("singletons") are included or excluded. The parameter combination from which correlations were derived for the results presented in the main text was identical to

the baseline, except with proximity-weighted instead of unweighted cell selection (Table S3 C).

2.3.1 Taxonomic data quality

The first sensitivity test we ran concerned the size and quality of the original PBDB dataset. In particular, we downloaded records in the same way specified in the "Global Dataset" section of STAR methods, except that records were filtered to those with "accepted names" at the level of species. This filter could remove some synonyms and misspellings, but it also severely restricts the number of records retrieved. In many of these excluded records, an enterer may have correctly specified the species' name, locality, stratigraphic data, and reference, but not performed the additional verification steps for the PBDB to recognize the taxonomy as "accepted" (e.g. specifying the year and author of the species' name). These valid records are retained in a download that does not distinguish between accepted and unaccepted taxonomic status. In light of the high potential for synonyms and misspellings when including "unaccepted" names, we still implemented manual vetting procedures to validate taxonomic names (as described in STAR methods).

The accepted-only data were too few in some time bins to reach a spatial coverage of 12 grid cells, so we subsampled only 10 cells at a time. The radius of constraint was 1500 km. The subsampling procedure was replicated 500 times. The same parameters were applied when subsampling the full dataset.

The magnitude and direction of correlations from the restricted dataset were equivalent to those generated from the larger dataset: species count correlated only very weakly with every range metric. Sampling aggregation correlated negatively with occupancy but positively with MST, although weakly in both cases. Because the larger dataset provided equivalent results to the more restricted version, we used the larger dataset for all further sensitivity analyses and the main text results. The additional records allowed for finer temporal resolution; the resultant binning is given in Table S1.

2.3.2 Singletons in space-time

We tested how the correlation between species count and range size might change if we included *versus* excluded species that occurred in exactly one grid cell of one time bin ("space-time singletons"). Species names that were misspelled or otherwise wrong would likely occur in the dataset as space-time singletons, and thus removing these species is expected to improve the quality of the dataset. Space-time singletons accounted for $1.3\text{x}10^4$ of $3.0\text{x}10^4$ taxa.

Including singleton taxa to test sensitivity to data quality did not change the correlation between species count and geographic distribution. Similarly, species count and range size series correlated to the same degree with spatial sample coverage whether or not space-time singletons were included.

We removed space-time singletons for all further sensitivity analyses and the main text results. There are three motivations for this decision. Firstly, we prioritized data quality above data abundance. Secondly, space-time singletons would be eliminated *a priori* during taxon-tracking analyses, which consider only species that occur in consecutive time bins. Finally, space-time singletons could exacerbate type 1 error in sensitivity analyses that performed coverage-based rarefaction, as discussed in section 2.3.7 below.

2.3.3 Number of cells

The next parameter we tested was the size of a subsample (i.e. number of cells selected). Increasing subsample size might reduce variability among subsampling iterations. After removing a filter on taxonomic status as "accepted", sufficient occurrence records were available to increase the number of subsampled cells from 10 to 12. We tested the influence of this increase in cells whilst keeping all other subsampling parameters (e.g. radius) constant. We also experimentally reduced the number of subsampled cells by half, from 12 to 6.

The correlation between species count and range size was weaker (less negative, i.e. closer to zero) for subsamples of 12 cells compared to those with 10 cells. Spatial area was better controlled for with the larger sample size, as reflected in smaller correlations between sampling aggregation and either species count or geographic distribution time series. Hence, the slightly negative correlation between species count and range size in subsamples of only 10 cells could be attributed primarily to bias in spatial coverage.

Results from subsamples of 6 cells support the interpretation that the strength of spatial bias decreases as more cells are selected. In particular, with only 6 cells, the correlation between spatial aggregation and either species count or any of the distribution metric time series was as great as, or larger than, the same correlation based on subsamples with 10 or 12 cells.

The width of the confidence intervals for correlation coefficients did not change in a uniform way or to a large degree when either more or fewer cells were selected. This might indicate that the variability in subsamples depends on which geographic province they capture, rather than on which cells are chosen from within the radius-constrained area. Alternatively, a change in the number of cells from 6 to 12 might not be large enough to induce a measurable effect on variance.

We included 12 cells in subsamples for all further sensitivity analyses and the main text results to minimize the influence of spatial bias.

2.3.4 Radius of regional constraint

While testing sensitivity to the area (number of cells) subsampled within each region, we also varied the region area (radius of constraint). Specifically, we reduced the radius of the buffer (Figure S1) from 1500 km to 750 km, while selecting 6 grid cells within the region. The magnitude of correlation between sampling aggregation and range size increased for MST

length, while the magnitude of other correlations remained constant. We set the radius at 1500 km for all further sensitivity analyses and the main text results to increase the number of cells that could be selected and minimize the influence of spatial bias.

2.3.5 Number of subsampling iterations

We noted above that the variability obtained among subsamples did not seem to depend on the number of cells included in a subsample. However, substantive variation among iterations could exist due to ecosystem differences among provinces, of which there have been up to 12 at times in the Phanerozoic [S21]. It is difficult to estimate *a priori* how many iterations are necessary to sample this regional variation thoroughly, especially because there are different numbers of possible starting points for subsampling (i.e. number of available "seed cells") among time bins. Therefore, we experimented with an increase in the number of iterations, from 500 to 1000 replicates. Doubling the number of iterations, however, did not change the mean or confidence interval for correlations. Therefore, it is likely that 500 iterations is a sufficient amount to sample most or all combinations of regions and site (cell) configurations. For this reason, we continued to replicate subsampling over 500 iterations for further sensitivity analyses and the main text results.

2.3.6 Proximity-weighted cell selection

The spatial coverage standardization routine for assemblage data (section 2.1 above) contained the option to select grid cells with scaled probability inversely proportional to the square of their distance from the regional center. In contrast, occupied cells were selected at random within each radius-buffered area for the unweighted cell selection method used in the above sensitivity analyses. The weighted method was designed to pick cells in a tighter cluster than the unweighted method (i.e. reduce spatial dispersion) whilst retaining a random component to subsampling. In cases where sampling is dense and homogeneous across space and time, the proximity-weighted method might reduce the grain of study and hence lead to more precise estimates of assemblage-level traits, such as mean species' occupancy.

We calculated time series correlations from weighted subsamples, for 500 iterations of 12 cell subsamples (Table S3 C). Weighting the selection probability reduced the correlation between species count metrics and sample aggregation, as predicted. However, weighting did not reduce the correlation between range size metrics and sample aggregation. The correlation between range size and species count time series was equivalent whether cell selection was weighted or unweighted.

2.3.7 Species count rarefaction

The number of species observed in a region depends strongly on fossil preservation and collection effort. The spatial aspects of sampling (area and extent) are discussed above (section 1). In this section, we consider the possibility that even after standardizing spatial

coverage, heterogeneous sampling effort (e.g. intensity of fossil collecting) could affect the number of species recovered. To standardize this variability, we implemented two common interpolation-based methods of (non-spatial) richness rarefaction: classical rarefaction and coverage-based rarefaction. Classical rarefaction standardizes sample size. Coverage-based rarefaction, also called shareholder quorum subsampling (SQS) or coverage-based richness interpolation, standardizes the estimated amount of the underlying frequency distribution represented in a sample [S22, S23]. Thus, sample size does not influence SQS estimates. Species count curves produced by SQS tend to be more stable than those from classical rarefaction or extrapolation methods [S24].

Coverage-based rarefaction. In the first rarefaction approach, we calculated SQS estimates of species richness within subsamples, using the **iNEXT** package [S25]. The implementation of **iNEXT** finds the same species count estimate as John Alroy's exact Perl scripts from version 4.0 onwards [S24]. Since our spatial subsampling procedure rarefied the number of grid cells in each sample, and since grid cell counts correlate directly with reference counts in global PBDB datasets of marine invertebrates [S26], we did not rarefy by PBDB "collections" or "references". Neither did we transform species count estimates, because correlations considered only relative rank of counts, rather than absolute magnitude.

Close *et al.* [S24] suggested that rarefied richness is sensitive to coverage level and thus recommend using the highest possible quorum value. After increasing the number of cells in a subsample from 10 to 12 in the global dataset (section 2.3.3 above), it was possible to increase coverage from 40% to 70% of the incidence frequency distribution in every time bin. To test the influence of coverage level, we spatially subsampled the global data while applying either 40% or 70% coverage standardization. As in the non-rarefied data, the correlation between range size and species count (as estimated by coverage-based rarefaction) was negative (Table S3 D–E). The magnitude, however, was not greater than that estimated for the correlation between sampling aggregation and geographic range size. In other words, the confounding effect of spatial coverage in samples could explain the slightly negative relationship between the variables of interest, and hence no support was found for ecological release or competitive exclusion when species count was standardized by frequency distribution coverage. Correlations were similar regardless of whether 40% or 70% coverage was used, although this similarity likely occurred because we had previously removed space-time singletons, which could influence coverage estimates.

Singleton taxa in assemblages. Frequency distribution coverage (derived mathematically from an expression named Good's u) is based in part on the relative frequency of taxa that occur only once [S22, S27]. Specifically, singletons result in a smaller estimate of Good's u, which would lead to more sampling of the coverage distribution to achieve the desired quorum level, and ultimately a higher estimate of richness. Hence, singletons could upwardly bias coverage-based richness estimates except in the unlikely case that the number of singletons were evenly distributed through time and space. Further confounding analysis, assemblage singletons cannot influence mean occupancy or median MST, because our definitions of both of those range size metrics excluded singletons. To evaluate the effect

of this asymmetry, we modified the subsampling protocol to remove single-occurring species from each assemblage before performing SQS. This added protocol reduced the correlation between sample aggregation and the SQS estimate of species count to the smallest value found in any of the sensitivity analyses. Indeed, time series correlations recovered a weakly negative relationship between species count and sampling aggregation (Table S3 F), whereas otherwise this relationship was positive (Table S3 D–E). All else equal, this inverse pattern would bias the correlation between species count and mean occupancy towards type 2 error, because the relationship between occupancy and sampling aggregation is also consistently negative.

Classical rarefaction. The classical form of rarefaction standardizes the number of sampled species units (i.e. presence counts in our rasterized dataset). Here, we selected 50 presences (species-cell combinations) in a spatial subsample and estimated the expected number of species in that selection (there were 12 cells in each subsample, within a 1500 km radius, and selected with unweighted probability). The time series was truncated to the Middle Ordovician onward because fossils in the Lower Ordovician were too sparse to obtain 50 occurrences in any given subsample. The results from classically rarefied assemblages were similar to those derived from analyses where SQS was used to rarefy at 70% coverage including singleton taxa in assemblages (Table S3 E).

2.3.8 Genus-level data

Although we were interested in processes that occur at the level of species, we repeated the analysis at the level of genera as well. The latter approach increased the number of occurrences from which to reconstruct each geographic distribution and provides comparability with previous paleontological studies (e.g. [S3, S17, S18]). We did not elevate subgenera to the rank of genera. Changing the taxonomic rank of analysis affects both taxon count and geographic range size: taxon count in any time period will decrease when assessed at higher levels, whereas geographic range sizes will increase. The empirical evidence from data aggregated by genus name suggests that the results presented in the main text are insensitive to taxonomic level. That is, the τ correlations were equivalent between analyses run on species (Table S3 A) and those run on genera (Table S3 G).

3 Power analysis

3.1 Justification

There are many reasons why a species might have had a larger geographic range than observed in the dataset for analysis: (i) fossils did not preserve in certain areas, (ii) fossiliferous sediment was eroded or overlain in certain areas, (iii) paleontologists did not collect fossils in certain areas, or (iv) an iteration of the spatial subsampling procedure did not happen to select all

occurrences of collected fossils for the species. Biases 1–3 exist for almost all quantitative paleobiology studies and have received intensive study already (e.g. [S6, S7, S9, S28]). The fourth source of underestimation in range size is specific to our new analytical approach.

Our spatial standardization procedure reduces the size of measured geographic distributions by truncating the area of observation. This compression could reduce the effect size of an underlying biological signal of ecological release. In addition, the subsampling procedure is probabilistic, and so measured range sizes for a given species in a given time interval vary from one iteration to the next. In light of these limitations, the possibility of type 2 error in our analyses merits scrutiny.

3.2 Simulation of ecological release

We simulated ecological release at a range of magnitudes and frequencies to calculate the probability of observing release in a subsample of the simulated data. We modeled range expansion based on the Middle–Late Pleistocene record of bivalves in the northwest Atlantic. This was the interval with the densest occurrences in our data (70 occupied grid cells), and so we expect that the compression from the subsampling procedure would be particularly severe. To control other biases on fossil range size, such as outcrop area, we used the same Pleistocene area before and after simulated release as the setting in which to calculate occupancy.

There were two experimental parameters in the simulation framework: the frequency with which species experience ecological release, and the percentage expansion of geographic range size that occurred for a species. We varied the frequency of release from 0–100% of species and the magnitude of release from 0–100% increase in occupancy for any given species. For each set of parameter values, we performed 10 simulations (4000 simulations in total). Considering a simulation where 50% of species expanded by 25%, for instance, we selected half of the observed species ($n = 423$) at random. For each of these "released species," we calculated an integer number of cells equivalent to 25% of current occupancy, and we generated that number of new occurrences for the species. These occurrences were placed in grid cells that were already occupied by other species, so as not to change the pool of cells available for spatial subsampling.

In the first set of simulation experiments, no subsampling procedure was applied to occurrences. In a second set of experiments, species' occupancy was subsampled before and after ecological release. From the latter simulations, we calculated the probability of observing release for a given magnitude and frequency of expansion. Subsampling consisted of selecting 15 occupied grid cells from which to calculate per-species range size, the same procedure as was applied to the empirical northwest Atlantic dataset.

3.3 Power to detect release

When no subsampling procedure was used, species expanded as a linear function of the frequency and magnitude of ecological release, as intended (Figure S3 A–C). When grid

cells were rarefied prior to calculating occupancy, the same trend emerged, although mean occupancy was lower and varied more among iterations (Figure S3 D–F).

A signal of release appeared in every simulation where at least two-thirds of species expanded in occupancy by at least 50% from their initial value (Figure S3 G). The same result held when at least half of species expanded by at least two-thirds of their initial occupancy. The probability of detecting release decreased at lower frequencies or magnitudes of expansion. Nevertheless, even when only half of species expanded by 25%, the probability of detecting this expansion reached 60% or more.

The release simulations provide an estimate of statistical power for detecting release in species that cross a single time bin. The three prongs of our empirical analysis, however, quantify patterns across many species and time bins. Therefore, even if the subsampling procedure produces a false negative of release for certain species that cross a stage boundary, a positive signal would still emerge in the final analysis. Furthermore, when characterizing the relationship between time series of geographic range size and species count, we employ a non-parametric correlation coefficient (Kendall's τ). Correlation strength then depends only on the direction of range shifts relative to species counts; the magnitude of range expansion during any intervals of ecological release does not affect correlation strength. In other words, although the spatial subsampling procedure might reduce the magnitude of observed geographic expansion in individual species relative to the true change, type 2 error in our analyses is unlikely. The high probability of detecting a signal of release in our simulations gives confidence in the empirical results.

Supplemental References

S1. Burnham, K.P. and Anderson, D.R. (2010). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd edition (New York, NY: Springer).

S2. Liow, L.H. and Stenseth, N.C. (2007). The rise and fall of species: implications for macroevolutionary and macroecological studies. Proc. R. Soc. Lond. B *274*, 2745–2752.

S3. Harnik, P.G., Simpson, C., and Payne, J.L. (2012). Long-term differences in extinction risk among the seven forms of rarity. Proc. R. Soc. Lond. B *279*, 4969–4976.

S4. Sheehan, P.M. (1977). A reflection of labor by systematists? Paleobiology *3*, 325–328.

S5. Signor, P.W., Lipps, J.H., Silver, L., and Schultz, P. (1982). Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. Geological implications of impacts of large asteroids and comets on the Earth *190*, 291–296.

S6. Behrensmeyer, A.K. and Kidwell, S.M. (1985). Taphonomy's contributions to paleobiology. Paleobiology *11*, 105–119.

S7. Kiessling, W. (2005). Habitat effects and sampling bias on Phanerozoic reef distribution. Facies *51*, 24–32.

S8. Kowalewski, M. and Bambach, R.K. (2008). The limits of paleontological resolution (Springer), pp. 1–48.

S9. Patzkowsky, M.E. and Holland, S.M. (2012). Stratigraphic paleobiology: understanding the distribution of fossil taxa in time and space (University of Chicago Press).

S10. Vilhena, D.A. and Smith, A.B. (2013). Spatial bias in the marine fossil record. PLoS One *8*, e74470.

S11. Jablonski, D. (1998). Geographic variation in the molluscan recovery from the end-Cretaceous extinction. Science *279*, 1327–1330.

S12. Button, D.J., Lloyd, G.T., Ezcurra, M.D., and Butler, R.J. (2017). Mass extinctions drove increased global faunal cosmopolitanism on the supercontinent Pangaea. Nat. Commun. *8*, 733.

S13. Tuomisto, H. (2010). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. Ecography *33*, 2–22.

S14. Connor, E.F. and McCoy, E.D. (1979). The statistics and biology of the species-area relationship. Am. Nat. *113*, 791–833.

S15. Sepkoski, J.J. (1976). Species diversity in the Phanerozoic: species-area effects. Paleobiology *2*, 298–303.

S16. Barnosky, A.D., Carrasco, M.A., and Davis, E.B. (2005). The impact of the species–area relationship on estimates of paleodiversity. PLoS Biol. *3*, e266.

S17. Foote, M., Crampton, J.S., Beu, A.G., Marshall, B.A., Cooper, R.A., Maxwell, P.A., and Matcham, I. (2007). Rise and fall of species occupancy in Cenozoic fossil mollusks. Science *318*, 1131–1134.

S18. Finnegan, S., Anderson, S.C., Harnik, P.G., Simpson, C., Tittensor, D.P., Byrnes, J.E., Finkel, Z.V., Lindberg, D.R., Liow, L.H., Lockwood, R., et al. (2015). Paleontological baselines for evaluating extinction risk in the modern oceans. Science *348*, 567–570.

S19. Kiessling, W. and Kocsis, A.T. (2016). Adding fossil occupancy trajectories to the assessment of modern extinction risk. Biology Letters *12*, 20150813.

S20. Close, R.A., Benson, R.B., Upchurch, P., and Butler, R.J. (2017). Controlling for the species-area effect supports constrained long-term Mesozoic terrestrial vertebrate diversification. Nat. Commun. *8*, 15381.

S21. Kocsis, A.T., Reddin, C.J., and Kiessling, W. (2018). The biogeographical imprint of mass extinctions. Proc. R. Soc. Lond. B *285*, 20180232.

S22. Alroy, J. (2010). The shifting balance of diversity among major marine animal groups. Science *329*, 1191–1194.

S23. Chao, A. and Jost, L. (2012). Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology *93*, 2533–2547.

S24. Close, R.A., Evers, S.W., Alroy, J., and Butler, R.J. (2018). How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves. Methods Ecol. Evol. *9*, 1386–1400.

S25. Hsieh, T., Ma, K., and Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. *7*, 1451–1456.

S26. Alroy, J., Aberhan, M., Bottjer, D.J., Foote, M., Fürsich, F.T., Harries, P.J., Hendy, A.J., Holland, S.M., Ivany, L.C., Kiessling, W., et al. (2008). Phanerozoic trends in the global diversity of marine invertebrates. Science *321*, 97–100.

S27. Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika *40*, 237–264.

S28. Darroch, S.A. and Saupe, E.E. (2018). Reconstructing geographic range-size dynamics from fossil data. Paleobiology *44*, 25–39.