Supplemental Material (not copyedited or formatted) for: Simon A. F. Darroch, Michelle M. Casey, Gawain T. Antell, Amy Sweeney, Erin E. Saupe. 2020.
"High Preservation Potential of Paleogeographic Range Size Distributions in Deep Time." The American Naturalist 196(4). DOI: https://doi.org/10.1086/710176.

Supplement to "High preservation potential of paleogeographic range size distributions in deep time"

Authors: Simon A.F. Darroch[1*], Michelle M. Casey[2], Gawain T. Antell[3], Amy Sweeney[1], and Erin E. Saupe[3]

Affiliations:

[1]Vanderbilt University, Department of Earth and Environmental Sciences, 2301 Vanderbilt Place, Nashville TN 37235-1805, USA.

[2]Towson University, Department of Physics, Astronomy and Geosciences, Smith Hall, Towson, MD 21252, USA.

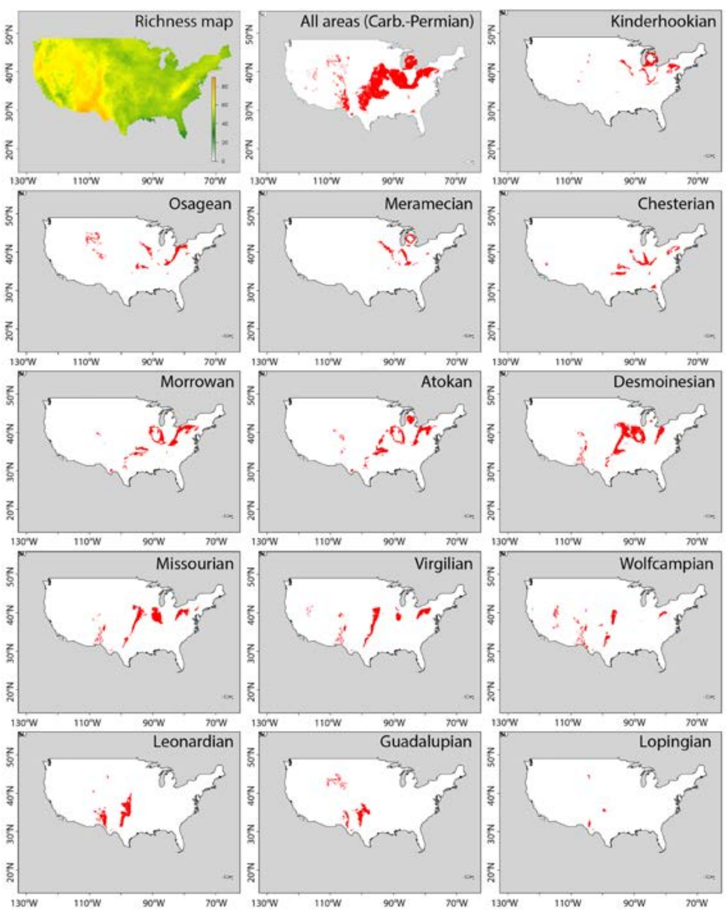
[3] University of Oxford, Department of Earth Sciences, South Parks Road, Oxford, OX1 3AN, UK.

* correspondence to simon.a.darroch@vanderbilt.edu

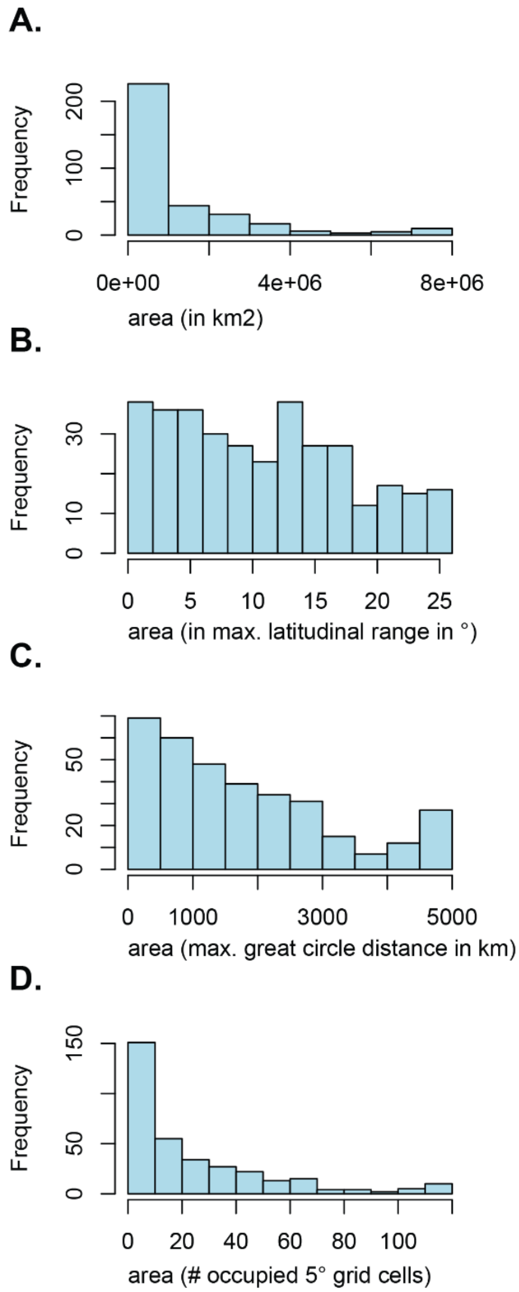
The American Naturalist

Supplementary figures and data

S1: Isolating sedimentary rock units in ArcMap: data were derived from USGS state-by-state geological maps (https://mrdata.usgs.gov/geology/state/). Sedimentary rock polygons were isolated for a given geological stage in ArcMap and assigned to stage-level time bins based on the formal lithostratigraphic name of their encompassing geological group, formation, or member using data from the National Geological Map Database (https://ngmdb.usgs.gov/Geolex/search) and extensive review of the literature (see 'Stratigraphic References' section at end of Supplementary Material document). We kept the polygon as a 'sedimentary unit' if the rock type matched any of the following: 'black shale', 'calcarenite', 'claystone', 'conglomerate', 'dolostone (dolomite)', 'graywacke', 'limestone', 'sand', 'sandstone', 'shale', 'siltstone', 'arkose', 'breccia', 'calc-silicate rock', 'mudstone', 'arenite', or 'arkose'. We excluded any metamorphic or igneous rock types and any unconsolidated sediment, as they are not likely to contain or preserve fossils in the rock record. We evaluated the list of formations returned and assigned each formation to a stage based on a stratigraphic database for the Carboniferous compiled by EES (references included at the end of this document), and the information available in the National Geological Map Database.

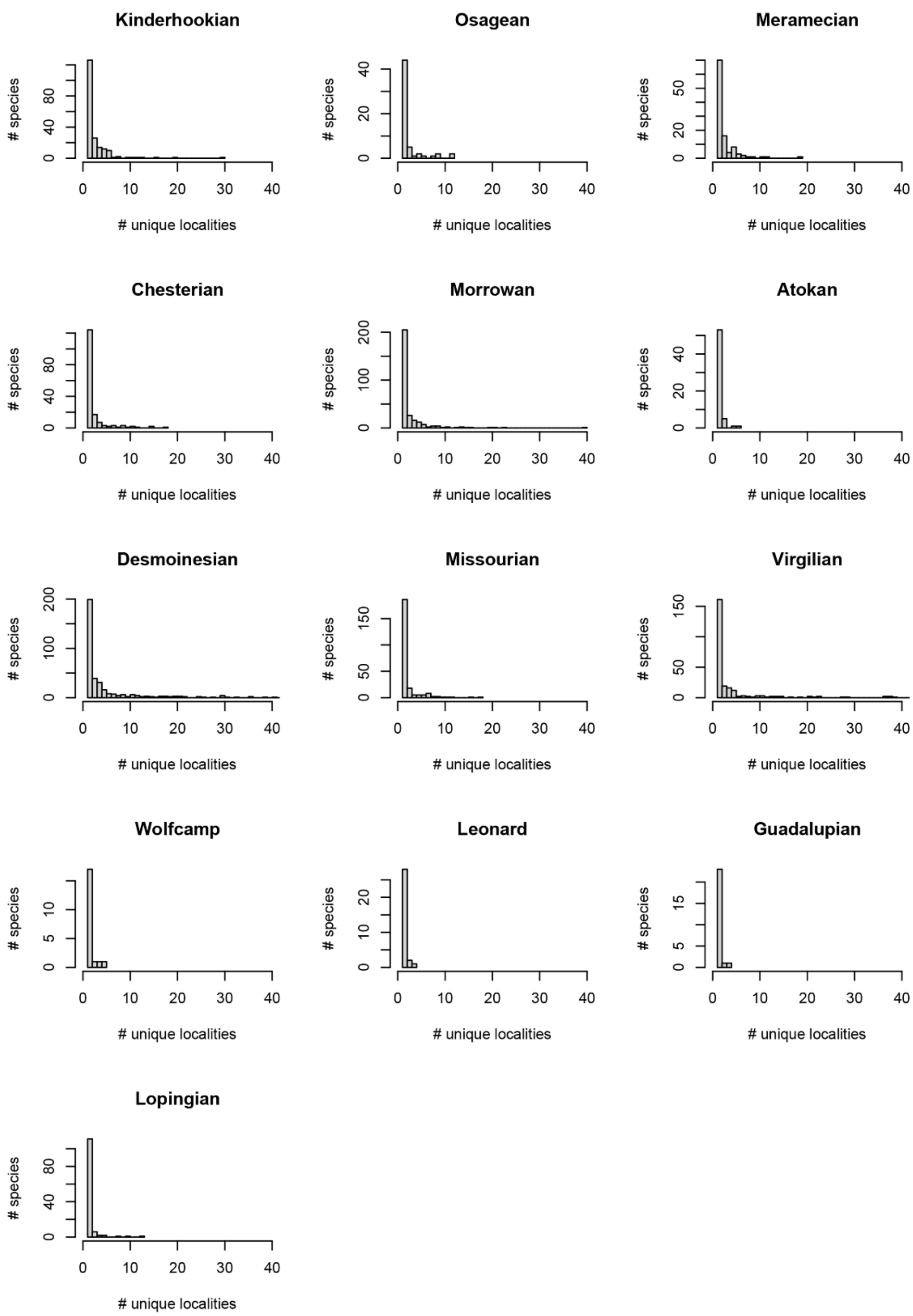


S2: Frequency histograms illustrating the distribution of 341 range sizes used in our analyses, measured in four ways: A) raw area (km^2); B) maximum latitudinal range (in degrees); C) maximum great circle distance (km); and, D) number of occupied 5° grid cells.


A.
Frequency
0
100
200
0e+00
4e+06
8e+06
area (in km2)
B.
Frequency
0
10
30
0
5
10
15
20
25
area (in max. latitudinal range in °)
C.
Frequency
0
20
50
0
1000
3000
5000
area (max. great circle distance in km)
D.
Frequency
0
50
150
0
20
40
60
80
100
area (# occupied 5° grid cells)

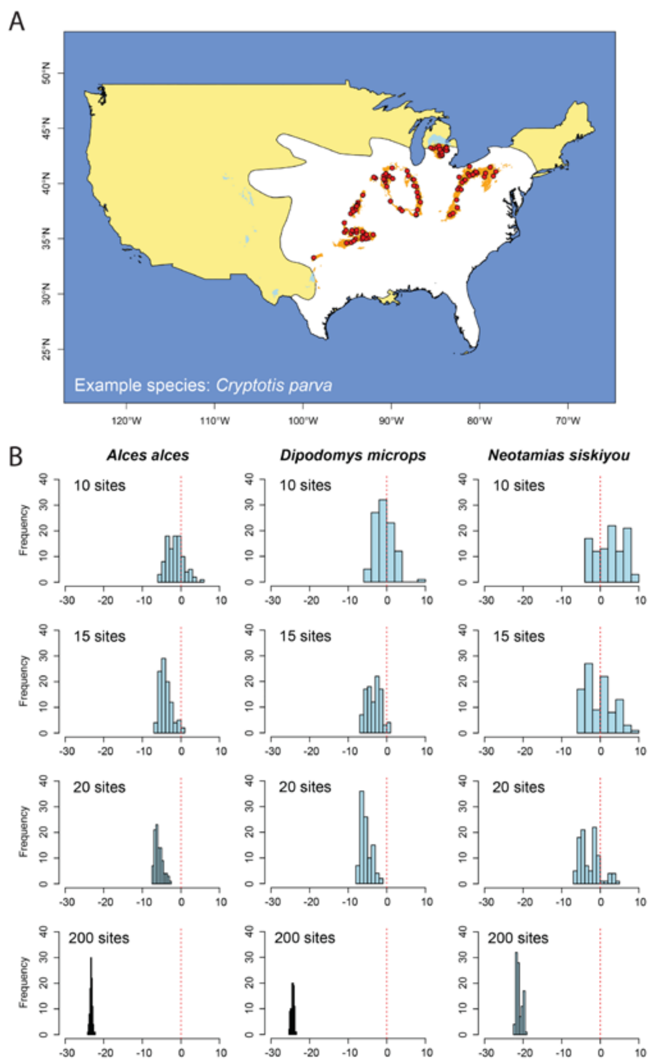
Supplemental Material (not copyedited or formatted) for: Simon A. F. Darroch, Michelle M. Casey, Gawain T. Antell, Amy Sweeney, Erin E. Saupe. 2020. "High Preservation Potential of Paleogeographic Range Size Distributions in Deep Time." The American Naturalist 196(4). DOI: https://doi.org/10.1086/710176.

S3: Frequency histograms illustrating the number of sites (unique latitude and longitude coordinates) recorded for benthic marine invertebrate (brachiopods, mollusks, and echinoderms) species from the Carboniferous and Permian. Data from the Paleobiology Database (paleobiodb.org). Histograms illustrate that the vast majority of species are found in only one site, with decreasing numbers found at larger numbers of sites. With rare exceptions, the vast majority of species are represented by < 20 sites.

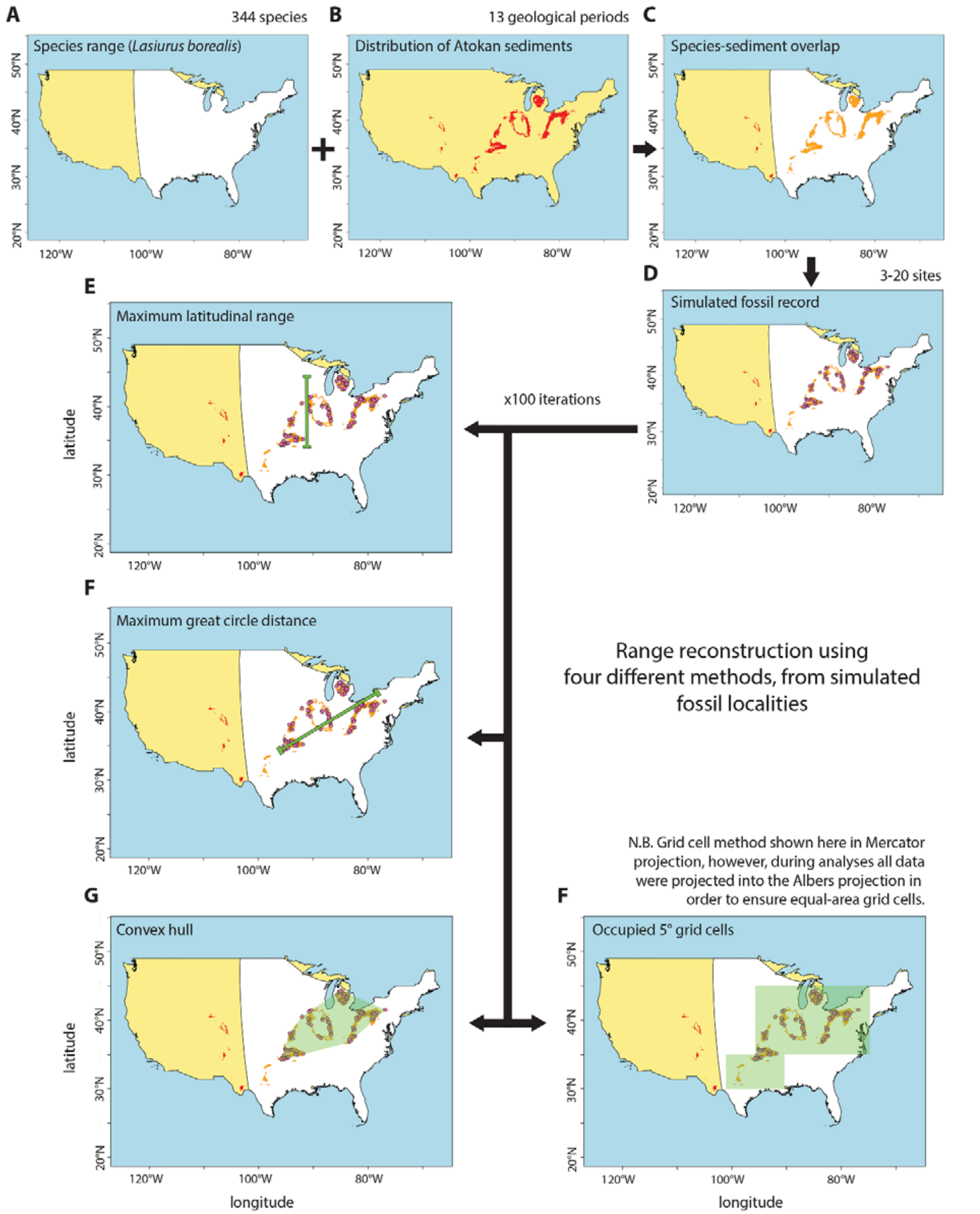


Supplemental Material (not copyedited or formatted) for: Simon A. F. Darroch, Michelle M. Casey, Gawain T. Antell, Amy Sweeney, Erin E. Saupe. 2020. "High Preservation Potential of Paleogeographic Range Size Distributions in Deep Time." The American Naturalist 196(4). DOI: https://doi.org/10.1086/710176.

S4: Fossil sites were placed randomly in our simulations. In reality, however, fossil sites are aggregated and 'patchy' at all spatial scales (Plotnick, 2017). We therefore quantified site aggregation to ensure our simulations replicated empirical patterns. We iteratively (100 times) generated fossil records for three test species and quantified relative aggregation of sites using z-values, which indicate whether occurrences are over- or under-dispersed compared to a purely-random distribution (with values < 0 indicating clustering). Panel A (below) illustrates a typical clustered occurrence pattern for the Least Shrew *Cryptotis parva* generated using our method. Panel B illustrates the distribution of z-values for 100 sets of fossil localities, for three other example species (*Alces alces*, *Dipodomys microps*, and *Neotamias siskiyou*). Results indicate that the majority of fossil placements were statistically clustered, particularly when considering >10 sites. This 'clustered' pattern of fossil localities observed in nature (and replicated in our simulations) likely reflects the intersection of independent phenomena operating at multiple spatial scales identified by Plotnik (2017), including the formation of sedimentary basins, the distribution of environments suitable for fossil preservation, tectonic and erosional processes, taphonomic variables, and human factors such as the presence of roadcuts or quarries.



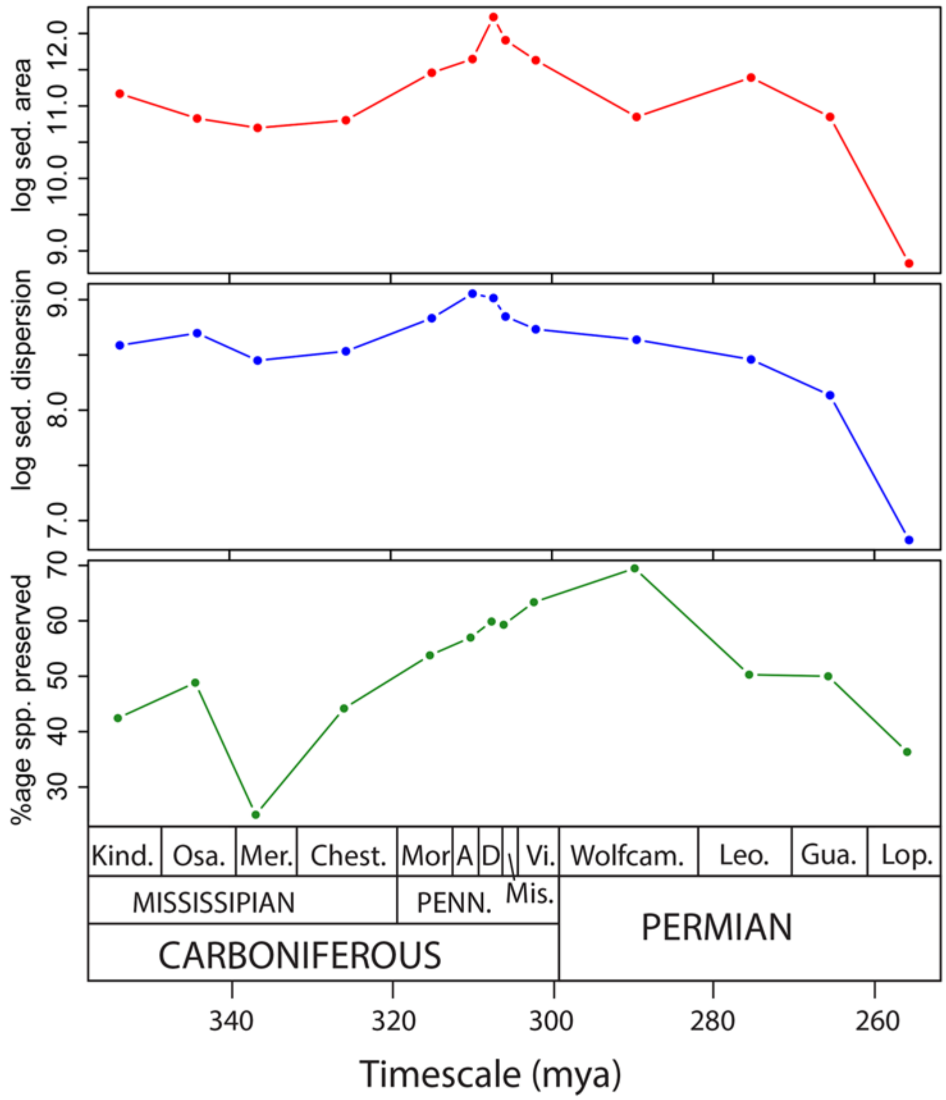
S5: Methodology for iteratively obtaining range-size estimates for 341 species using different thresholds for numbers of simulated fossil localities. Range-size estimates were repeated for each of 13 geological stages spanning the Carboniferous–Permian (Atokan used here as an example). Species range polygons are represented in white, sediments in red, and the area of overlap between the sediments and species ranges in orange.



S6: Potential correlates for the accuracy of extinction experiments, including preserved sedimentary area in each stage, sedimentary 'dispersion' in each stage, and number of species recovered in the simulations in each stage (out of 341 species simulated).

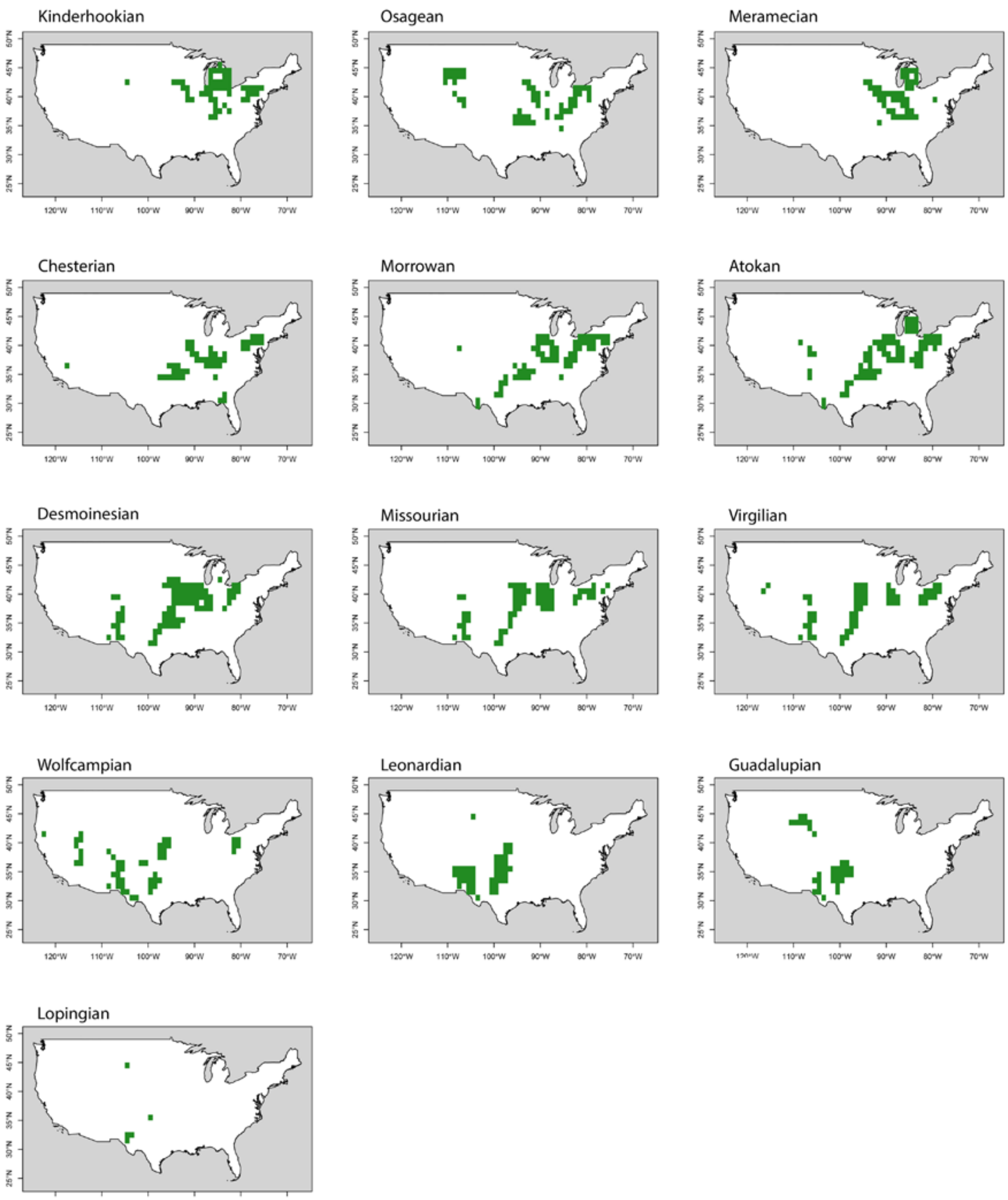
Stage	**Duration (Ma)**	**Midpoint age (Ma)**	**Sedimentary area (km^2)**	**Dispersion (km)**	**Number of fossil species**
Kinderhookian	11.2	354	70906	5362	144
Osagean	8.0	344	50410	5986	166
Meramecian	7.0	337	44242	4673	85
Chesterian	14.9	326	49170	5082	151
Morrowan	6.4	315	94673	6854	183
Atokan	3.7	310	114347	8561	194
Desmoinesian	1.5	307	204217	8224	204
Missourian	1.5	306	148350	6958	202
Virgilian	6.0	302	112515	6202	216
Wolfcampian	19.0	290	51464	5640	237
Leonardian	9.4	275	88600	4717	171
Guadalupian	10.2	266	51483	3411	170
Lopingian	9.4	256	6818	920	123

S7: Time series illustrating log sedimentary area (top), log sedimentary dispersion (middle), and log number of species preserved (bottom) in each of 13 Stages.

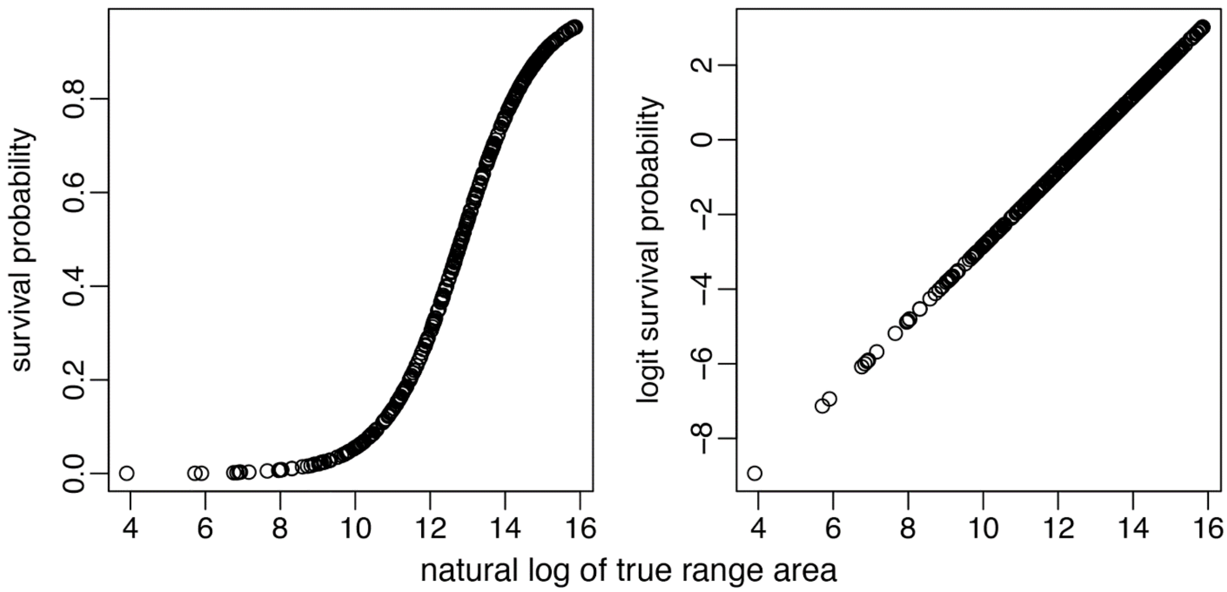


Supplemental Material (not copyedited or formatted) for: Simon A. F. Darroch, Michelle M. Casey, Gawain T. Antell, Amy Sweeney, Erin E. Saupe. 2020. "High Preservation Potential of Paleogeographic Range Size Distributions in Deep Time." The American Naturalist 196(4). DOI: https://doi.org/10.1086/710176.

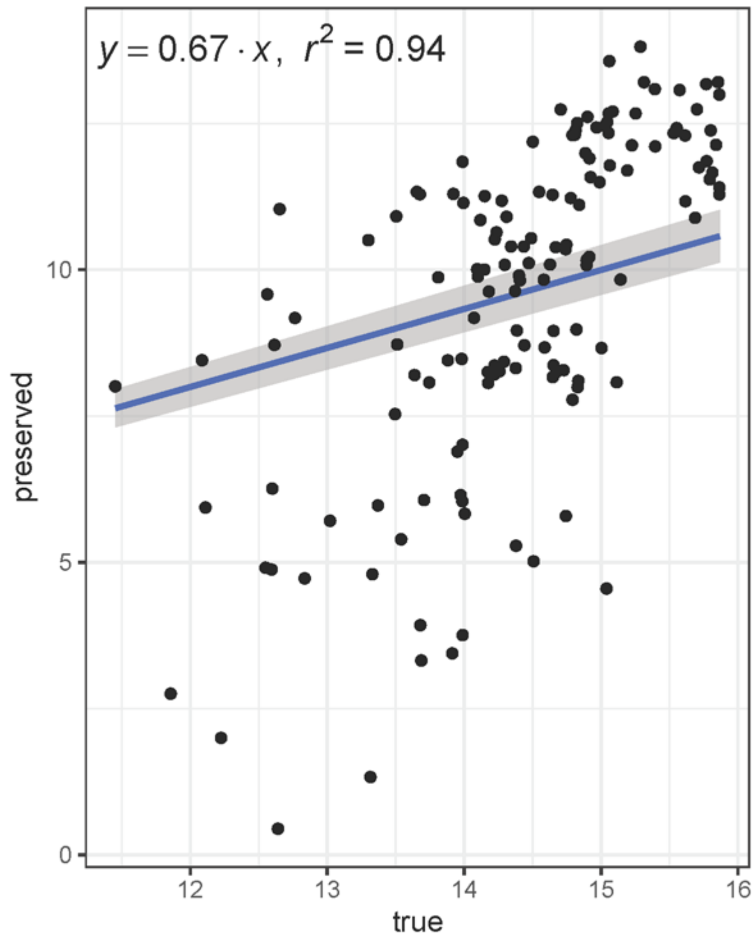
S8: Method for measuring sedimentary 'dispersion' using minimum spanning trees. We rasterized sedimentary areas in each stage into 1° latitude/longitude grid cells; a cell was included if any part of it intersected with sediments (illustrated below in green). Latitude/longitude centroids were estimated for all occupied cells, and a minimum spanning tree calculated from these centroids (i.e., the shortest distance between all points). Sedimentary 'dispersion' was recorded as the summed distance of minimum spanning trees.



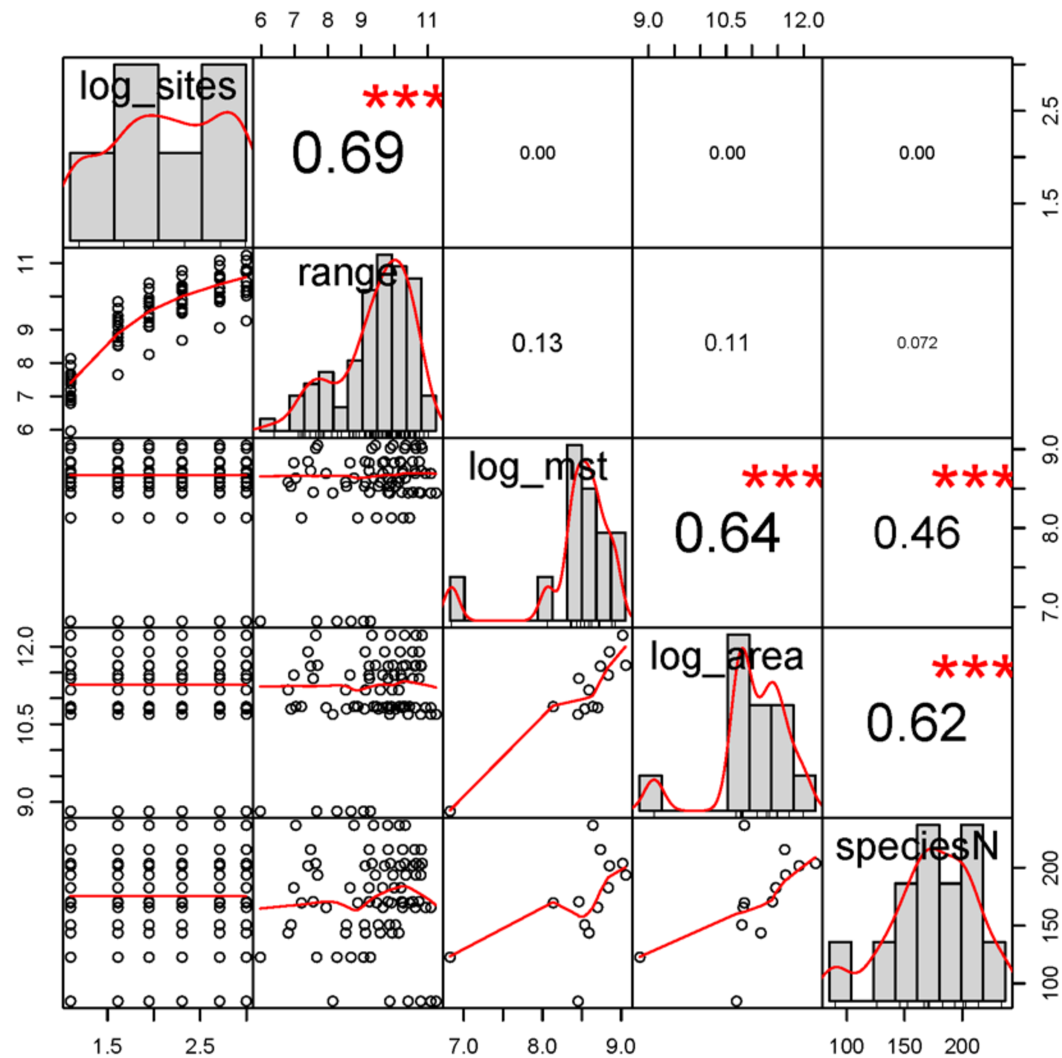
S9: Probability of species' survival was simulated as the inverse logit of geographic range size, such that widespread species were more likely to survive. Probability of survival is plotted on the y-axis (original scale on left; logit scale on right) for each species' IUCN range area (convex hull method only). One iteration of site placement with 50% extinction rate is plotted, although we also considered 25% and 75% extinction rate.



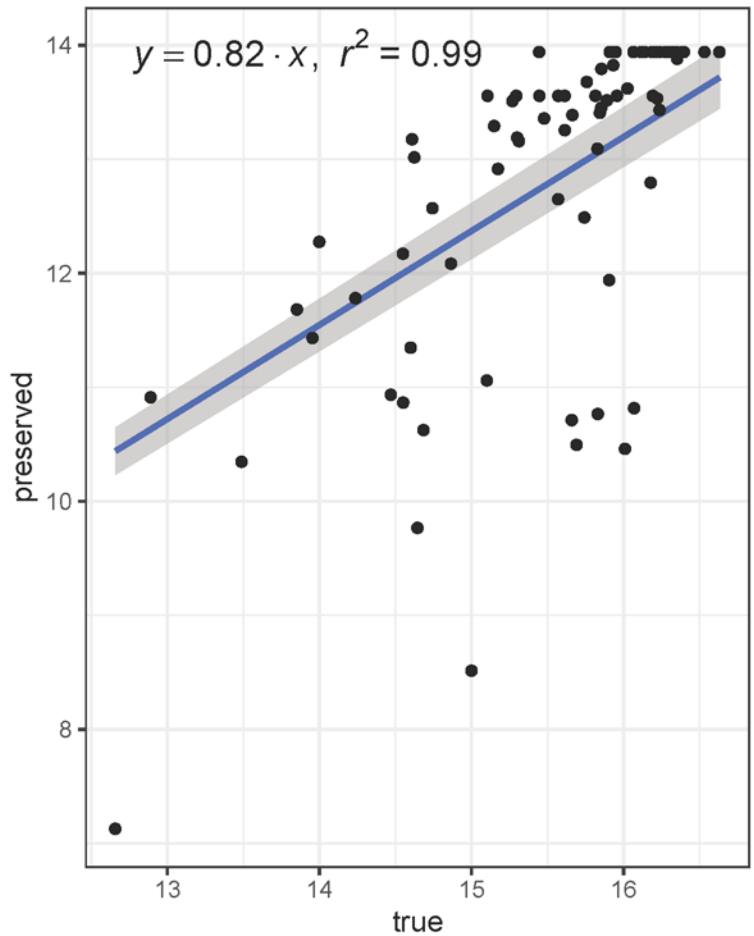
S10: Species' range sizes as measured from fossil data are plotted relative to true IUCN range size. Geographic range size is natural-log-transformed convex hull area (km2). In this figure, 10 sites were placed in Kinderhookian sedimentary areas according to the by-species method. The simulated preservation process resulted in average range truncation of 0.67 on a log scale.



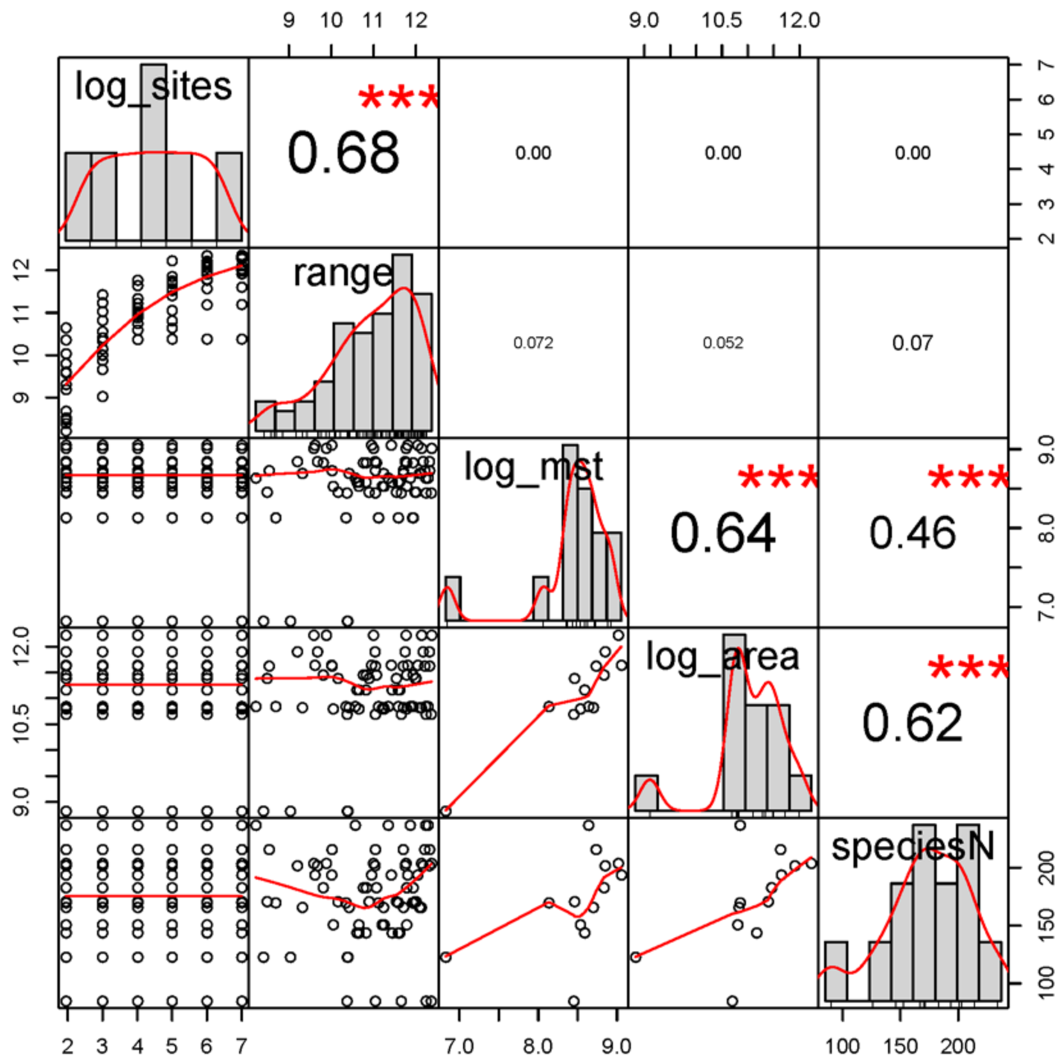
S11: Kendall's tau correlations (upper right triangle) are calculated for each pair of candidate predictors (Table 1) to be included in Model 2 (Table 2). Since mean fossil species' range area correlated strongly with the number of sites, the former was excluded from the model. Similarly, sedimentary area was excluded due to strong correlation with sedimentary dispersion (mst) and species count (species N). Univariate histograms are plotted along the diagonal, and pairwise scatterplots are plotted in the lower left triangle. Data in this plot were generated with by-species site placement.



S12: Species' range sizes as measured from fossil data are plotted relative to true IUCN range size. Geographic range size is natural-log-transformed convex hull area (km2). In this figure, 55 sites were placed in Kinderhookian sedimentary areas according to the by-species method. The simulated preservation process resulted in average range truncation of 0.82 on a log scale.



S13: Kendall's tau correlations (upper right triangle) are calculated for each pair of candidate predictors (Table 1) to be included in Model 2 (Table 2). Since mean fossil species' range area correlated strongly with the number of sites, the former was excluded from the model. Similarly, sedimentary area was excluded due to strong correlation with sedimentary dispersion (mst) and species count (species N). Univariate histograms are plotted along the diagonal, and pairwise scatterplots are plotted in the lower left triangle. Data in this plot were generated with by-sediment site placement.



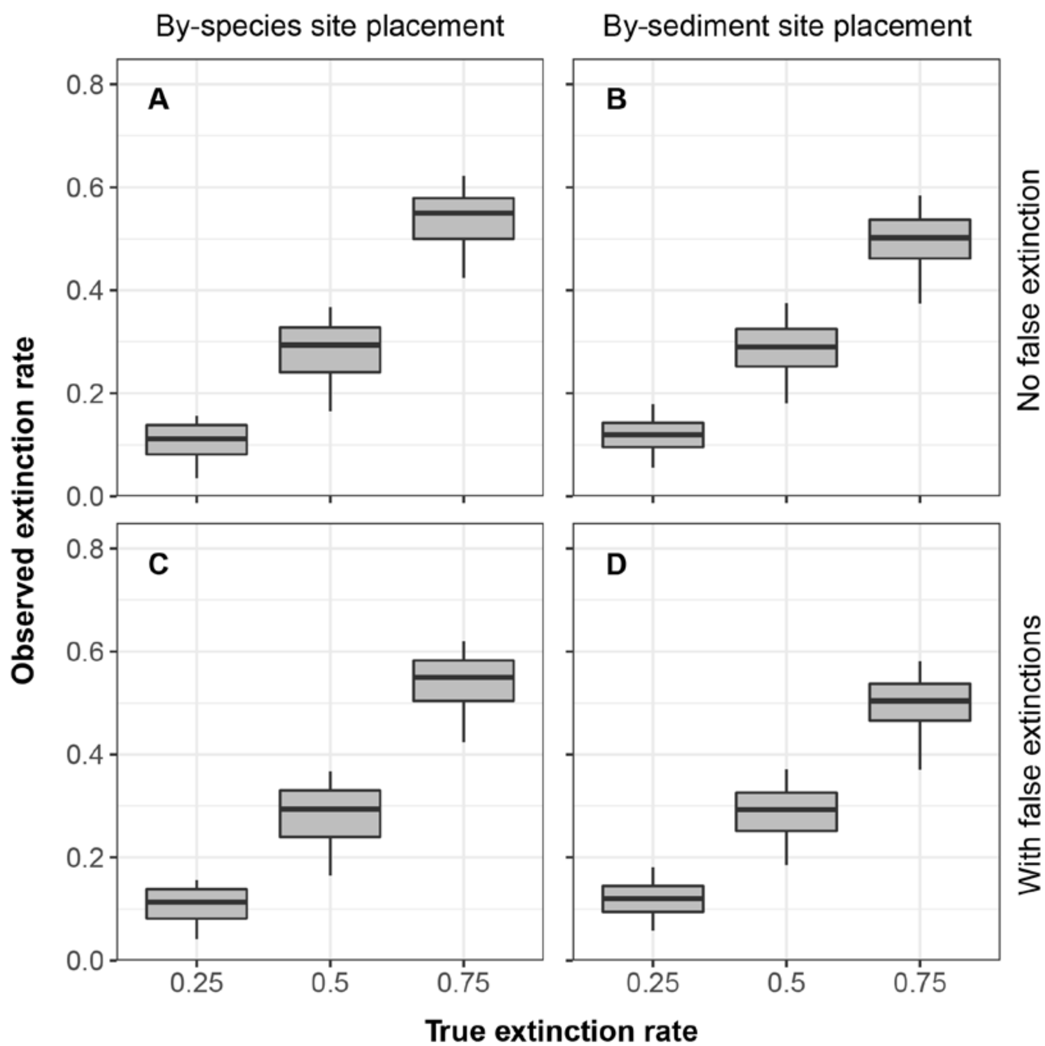
S14: Accuracy of selectivity reconstruction (i.e., mean of beta, the per-species increase in extinction risk per unit range size) was modeled as a function of 6 predictor variables: dispersion of sedimentary rock (natural log km), number of species preserved, number of sites, extinction rate, and number of false extinctions. The four models (Models 3, 5, 7, and 9, Table 2) are based on different sets of simulations: random (null) extinction vs. selective extinction, and including vs. excluding false extinctions. Tables 1 and 2 give further details on the models and predictors. The estimate and its standard error are given for each variable, above the diving line; below the line are reported summary data for the model (model number, R^2, and mean and 90% quantile of the response variable). A, by-species site placement data; B, by-sediment site placement data.

Predictor variable	**Random, no false E**	**Random, with false E**	**Selective, no false E**	**Selective, with false E**
		A. By-species site placement		
Intercept	-0.008 ± 0.008	0.085 ± 0.047	-0.153 ± 0.040	-0.288 ± 0.068
Dispersion	-0.000 ± 0.001	-0.014 ± 0.006	0.025 ± 0.005	0.023 ± 0.009
N sampled	0.000 ± 0.000	0.000 ± 0.000	0.001 ± 0.000	0.001 ± 0.000
Sites	0.000 ± 0.001	0.023 ± 0.004	0.043 ± 0.004	0.087 ± 0.006
Extinction %	0.013 ± 0.002	0.009 ± 0.013	0.005 ± 0.012	0.018 ± 0.020
N false e	NA	0.009 ± 0.000	NA	0.006 ± 0.000
Model # (Table 2)	3	5	7	9
R^2	0.048	0.671	0.348	0.571
Mean beta	-0.000	0.108	0.270	0.394
[90% CI]	[-0.020, 0.020]	[0.001, 0.370]	[0.170, 0.420]	[0.198, 0.705]
		B. By-sediment site placement		
Intercept	-0.005 ± 0.006	-0.684 ± 0.070	-0.235 ± 0.038	-1.373 ± 0.112
Dispersion	-0.000 ± 0.001	0.071 ± 0.008	0.015 ± 0.005	0.131 ± 0.013
N sampled	-0.000 ± 0.000	-0.000 ± 0.000	0.001 ± 0.000	0.001 ± 0.000
Sites	0.000 ± 0.000	0.022 ± 0.003	0.042 ± 0.002	0.067 ± 0.005
Extinction %	0.015 ± 0.002	0.060 ± 0.020	0.013 ± 0.012	0.116 ± 0.031
N false e	NA	0.014 ± 0.001	NA	0.017 ± 0.001
Model # (Table 2)	3	5	7	9
R^2	0.091	0.614	0.681	0.665
Mean beta	-0.000	0.102	0.169	0.303
[90% CI]	[-0.016, 0.016]	[-0.003, 0.488]	[0.061, 0.402]	[0.070, 0.887]

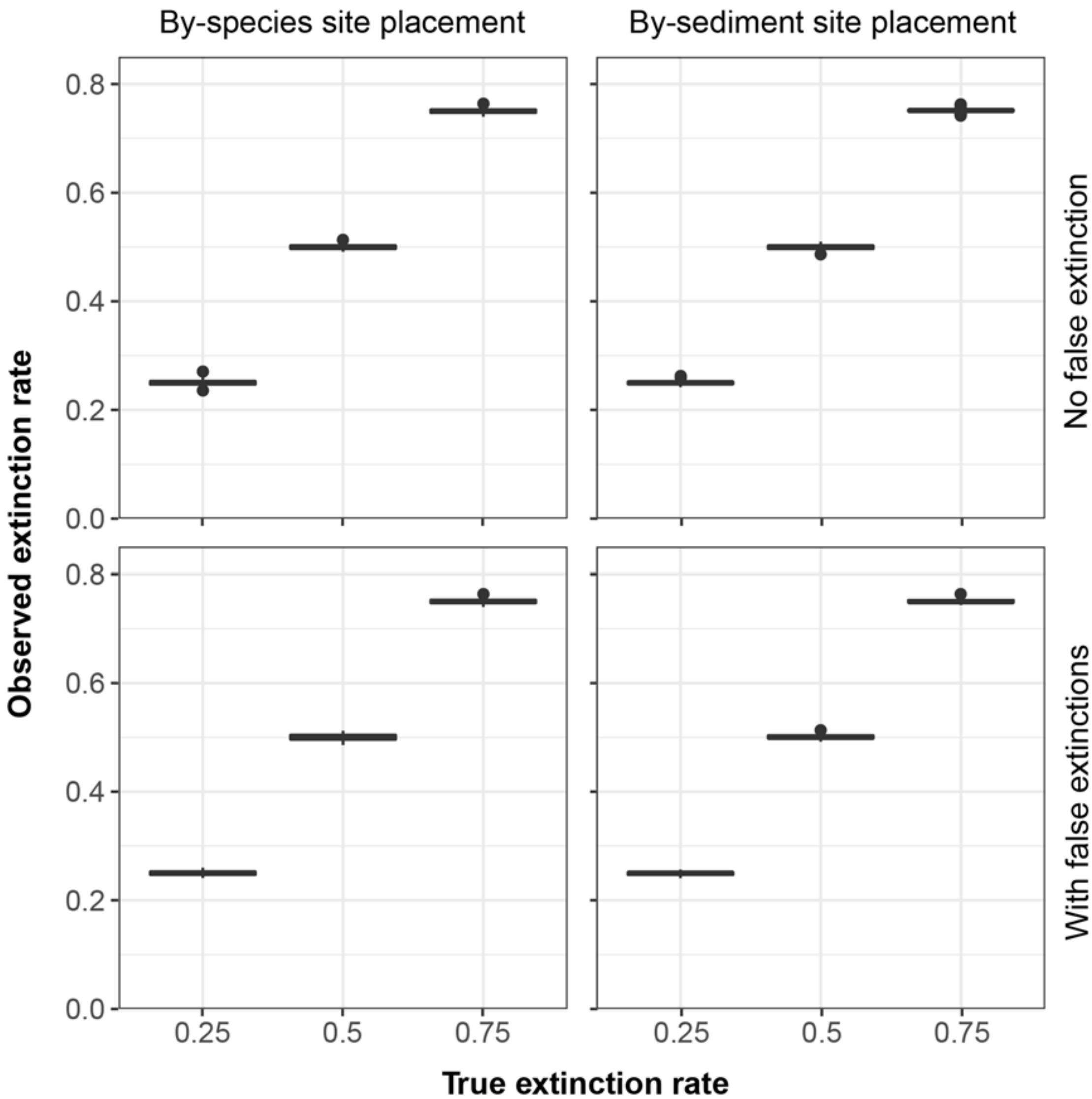
S15: Precision of selectivity reconstruction (i.e., standard deviation of beta) was modeled as a function of 6 predictor variables. The reported values are as in S14. A, by-species site placement data; B, by-sediment site placement data.

Predictor variable	Random, no false E	Random, with false E	Selective, no false E	Selective, with false E
A. By-species site placement				
Intercept	0.131 ± 0.029	0.121 ± 0.043	-51.910 ± 74.264	-33.624 ± 151.23
Dispersion	0.010 ± 0.004	0.007 ± 0.005	16.503 ± 9.517	21.883 ± 19.196
N sampled	-0.001 ± 0.000	-0.001 ± 0.000	-0.481 ± 0.133	-0.681 ± 0.265
Sites	0.010 ± 0.003	0.016 ± 0.004	9.792 ± 6.887	3.589 ± 13.509
Extinction %	-0.001 ± 0.009	0.051 ± 0.012	-39.092 ± 21.690	-60.400 ± 44.031
N false e	NA	0.000 ± 0.000	NA	-0.381 ± 0.713
Model #	4	6	8	10
R^2	0.306	0.208	0.026	0.013
SD beta [95% CI]	0.098 [0.050, 0.170]	0.105 [0.050, 0.214]	6.728 [0.055, 0.372]	9.512 [0.058, 0.304]
B. By-sediment site placement				
Intercept	-2.479 ± 6.151	-13.387 ± 22.669	3.449 ± 8.948	10.203 ± 13.639
Dispersion	1.476 ± 0.726	1.594 ± 2.652	3.420 ± 1.057	2.256 ± 1.599
N sampled	-0.075 ± 0.012	-0.075 ± 0.042	-0.163 ± 0.018	-0.152 ± 0.025
Sites	0.461 ± 0.292	1.165 ± 0.967	0.846 ± 0.425	0.864 ± 0.592
Extinction %	-0.698 ± 1.842	16.109 ± 6.356	-20.333 ± 2.680	-15.914 ± 3.817
N false e	NA	0.119 ± 0.163	NA	-0.119 ± 0.084
Model #	4	6	8	10
R^2	0.065	0.013	0.198	0.096
SD beta [95% CI]	1.650 [0.048, 4.925]	3.827 [0.051, 9.846]	3.996 [0.054, 24.910]	3.931 [0.056, 18.800]

Supplemental Material (not copyedited or formatted) for: Simon A. F. Darroch, Michelle M. Casey, Gawain T. Antell, Amy Sweeney, Erin E. Saupe. 2020. "High Preservation Potential of Paleogeographic Range Size Distributions in Deep Time." The American Naturalist 196(4). DOI: https://doi.org/10.1086/710176.



S16: The extinction rates observed from fossil data are plotted for each threshold of true extinction rate (25%, 50%, or 75%). Observed extinction rate always underestimates the true rate because geographically restricted species are more likely to go extinct without being detected. In A and B, false extinction was simulated; in C and D, no false extinctions were included. Data were from by-species site placement in A and C, and from by-sediment site placement in B and D. Extinction was selective on all species' range sizes (see S16 for null selectivity).



S17: The extinction rates observed from fossil data are plotted for each threshold of true extinction rate (25%, 50%, or 75%). Observed extinction rate always underestimates the true rate because geographically restricted species are more likely to go extinct without being detected. In A and B, false extinction was simulated; in C and D, no false extinctions were included. Data were from by-species site placement in A and C, and from by-sediment site placement in B and D. Extinction was random with respect to species' range sizes (see Figure 3 for selective extinction).

R scripts to replicate all results are archived along with copies of the data files at (https://github.com/GawainAntell/ExtSelectivity2020) and are archived at Zenodo (https://doi.org/10.5281/zenodo.3762196).