Climate change is an important predictor of extinction risk on macroevolutionary timescales

Cooper M. Malanoski[1*], Alex Farnsworth[2], Daniel J. Lunt[2], Paul J. Valdes[2], Erin E. Saupe[1*]

[1]*Department of Earth Sciences, Oxford University, South Parks Road, Oxford, OX1 3AN, U.K.*

[2]*School of Geographical Sciences, University of Bristol, Bristol, UK*

**Corresponding authors*

Climate change matters

Abstract

Anthropogenic climate change is increasing rapidly and already impacting biodiversity. Despite the importance for future projections, understanding of the underlying mechanisms by which climate mediates extinction remains limited. We present an integrated approach examining the role of intrinsic traits vs. extrinsic climate change in mediating extinction risk for marine invertebrates over the past 485 million years. We found that a combination of physiological traits and the magnitude of climate change are necessary to explain marine invertebrate extinction patterns. Our results suggest that taxa previously identified as extinction resistant may still succumb to extinction if the magnitude of climate change is great enough.

Introduction

Climate has changed rapidly over the last several decades (*1*) and is projected to continue into the coming centuries (*2*, *3*). Changes in climate are already impacting modern biodiversity (*4–6*) and are expected to impact biodiversity in the future (*7*). Significant biodiversity loss has also been linked to climate change in the past (*7–10*). However, relatively little is known regarding the impact of climate change on extinction risk for taxa over Phanerozoic timescales, or how the rate and magnitude of climate change affects extinction risk compared to other known predictors. Determining the traits that promote or inhibit extinction provides critical insight on the causal mechanisms generating biodiversity over geological time scales (*11*, *12*) and may help to identify species at risk of extinction today (*13–16*).

Previous work has identified correlates of extinction in modern and ancient taxa, including abundance (*17*, *18*), body size (*15*, *19*, *20*), niche breadth (*21–26*), thermal tolerance (*15*, *22*, *27*, *28*), and geographic range size (*11*, *12*, *17*, *18*, *22*, *28–30*). The latter trait in particular has been identified as a key predictor of extinction over the Phanerozoic (*11*, *14*, *17*, *20*, *29*). Despite this important past work, tests of extinction determinants over macroevolutionary timescales have been conducted for only a few predictor variables, and are often tested in isolation, limiting our understanding of evolutionary drivers and our ability to forecast the effects of anthropogenic climate change on biodiversity.

Thermal tolerance, for example, has been posited to impact the vulnerability of species to anthropogenic climate change today (*31–34*) and in the past (*27*), but has not been assessed over Phanerozoic timescales relative to other extinction risk predictors. Here we hypothesize that the thermal tolerance of a taxon influences its risk of extinction, since this trait mediates responses to climate change, and fossil data suggest extinction patterns tend to vary latitudinally (*35–37*). We additionally hypothesise that the magnitude of climate change experienced by a taxon will affect its

extinction risk. The rate and magnitude of current climate change is considered a significant threat to global biodiversity (*2*), with myriad impacts already observed (*5*). Critically, however, the degree to which taxon-specific climate change estimates are able to predict extinction remains unknown, including whether this extrinsic factor is a stronger predictor of extinction than traits intrinsic to taxa, such as geographic range size or body size.

Paleontological occurrence data and predictors of extinction

Here we use three novel approaches to quantify thermal tolerance and the magnitude of climate change experienced by taxa for 9,264 unique genera in 81 stages across the Phanerozoic (Fig. 1; fig. S2). We assessed thermal tolerance by estimating the realized thermal preference and realized thermal niche breadth for each taxon. Realized thermal preference is estimated as the absolute deviation in occupied temperature from the median occupied temperature for a stage, representing thermal preference, while realized thermal niche breadth is estimated as the occupied range of temperatures for a taxon, representing degree of climate specialization (*38*) (see SI Methods and Materials (*39*)).

We compare proxies for thermal physiology to known correlates of extinction, including geographic range size and body size, to examine their relative weights in governing macroevolutionary dynamics over the last 485 million years, while recognizing the complex interactions between variables (*40*, *41*). Using the Paleobiology Database (*42*), we focus on nine classes of marine invertebrates belonging to Cephalopoda, Gastropoda, Echinoidea, Crinoidea, Trilobita, Ostracoda, Bivalvia, and the brachiopods Rynchonellata and Strophemenata. Marine invertebrates are well suited for our analyses, since they are regarded as the most complete and reliable constituents of the fossil record (*43*), and their geographic ranges are thought to closely approximate their true thermal tolerances based on the correlation between realized and fundamental thermal niches calculated through lab experiments in marine ectotherms (*32*, *44*).

Paleoclimate reconstruction

Spatial estimates of past climate used in the calculation of extinction risk predictors (see SI Methods and Materials (*39*); data S1; fig. S1) were derived from new HadCM3 climate simulations for 81 stages over the Phanerozoic (Fig. 1). These simulations improve on those in Valdes et al. (*45*) in a number of ways, including tuning of climate model variables (*46*) so that the model has increased polar amplification during past climates (*47*) and more realistic prescribed pCO_2 during the Cenozoic (*48*) (fig. S1). The climate model is appropriate for this study in that it can successfully simulate polar warmth of the early Eocene and Cretaceous and produce accurate simulations of the Last Glacial Maximum and pre-industrial climates (*49*). We use several versions of the climate simulations to test the sensitivity of our results to the climate model input (see SI Methods and Materials (*39*); fig. S5).

Statistical model parametrization

We examined patterns of extinction selectivity over the Phanerozoic for a combined total of 292,940 spatially- and temporally-unique fossil occurrences using generalized linear mixed effects models that accounted for variance in selectivity temporally and taxonomically (see SI Methods and Materials (*39*); table S1; fig. S2,S3). Patterns of extinction were modelled as a function of five predictors, both with (table S8) and without interaction terms, including geographic range size, body size, absolute realized thermal preference, realized thermal niche breadth, and the absolute value of taxon-specific climate change estimates (Fig. 1). Trait estimates were based on both jackknife and bootstrap subsampling to mitigate potential spatial and sample size biases (see SI Methods and Materials (*39*); fig. S4). Analyses were performed at both the generic- and species-level (*50*). We

focus here on generic-level patterns because species-level measures and temporal ranges are often poorly constrained on macroevolutionary timescales (*13*, *51*), although modelled results were consistent across both approaches (table S6).

Extinction selectivity patterns

The best model for all combinations of predictors based on AIC (*52*, *53*) was the most saturated, with all predictors significant at $\alpha < 0.05$ and a model weight of 0.98 (*52*) (Fig. 2; table S3). Model results were robust to the choice of subsampling approach (see SI Methods and Materials (*39*); fig. S4; table S4,S5), to the taxonomic rank of analysis (species or genus; table S6), and to different climate model simulations used to characterize our thermal traits (see SI Methods and Materials (*39*); fig. S5), corroborating the importance of these variables in regulating macroevolutionary patterns over 485 million years. Furthermore, power analyses conducted on simulated datasets suggest our model can reliably detect selectivity signals for all five predictors (see SI Methods and Materials (*39*)).

The absolute magnitude of climate change experienced by a taxon emerged as an important predictor of extinction risk in the best-supported model, suggesting that taxa exposed to greater climate change preferentially went extinct. This extrinsic environmental trait showed the same strength of correlation with extinction risk as did realized thermal niche breadth, realized thermal preference, and body size, evidenced by the overlap in confidence intervals for coefficient estimates (Fig. 2, fig. S4; table S3). The only trait with a significantly stronger signal of selectivity was geographic range size. Geographic range size was also the most important predictor when controlling for sample size (fig. S4), and the coefficient of -0.44 was larger than the second largest of -0.29 for body size and realized thermal niche breadth (Fig. 2; table S3). This supports past work that found smaller-ranged species are consistently more vulnerable to extinction on geological timescales (*15*, *18*, *22*, *29*) (Fig. 2).

Using spatio-temporal estimates of climate change experienced by each taxon (see SI methods and materials (*39*); Fig. 1), we find that a localized mean annual temperature change of ~7^{o}C could greatly increase a taxon's probability of extinction, calculated as the temperature when the marginal effects are greater than zero (Fig. 2). This threshold is slightly higher than that found by Song et al. (*54*), who suggested an increase of 5.2^{o}C could result in a mass extinction in the modern. However, Song and colleagues (*54*) used globally-averaged rates of climate change to predict magnitude of extinction over the Phanerozoic, rather than the taxon-specific magnitude of localized climate change used here.

Realized thermal niche breadth and realized thermal tolerance were both strong and consistent predictors of extinction risk over the Phanerozoic (Fig. 1; Fig. 2). These results are robust to the number of occurrences used to estimate thermal tolerance (see SI Methods and Materials (*39*); fig. S4). Taxa that occupied climatic extremes were more extinction prone than taxa occupying more thermally intermediate isotherms (Fig. 2), suggesting a significant evolutionary advantage for taxa living in temperate isotherms over those living near the equator or at the poles. This pattern is congruent with past studies that found taxa adapted to living in climatic extremes were more prone to extinction during hyperthermal events (*27*, *35*, *55*, *56*), which could be explained by the loss of thermal habitat, or ecophysiotypes, at the poles and equator due to climate change (*9*, *56*).

Taxa with narrower thermal breadths were more susceptible to extinction than taxa with broader thermal tolerances in the best-supported model (Fig. 2), and this trait was as important as realized thermal tolerance in explaining patterns of extinction over the Phanerozoic (Fig. 2). The strength of this correlate suggests extinctions are physiologically selective based on thermal limits. However, it

is important to note that the thermal breadth of a taxon is mediated by both temperature and oxygen availability, which has increased over the Phanerozoic, as well as other biotic factors such as competition and nutrient availability (*8*, *56–58*). Based on our models, taxa with narrow realized thermal breadths of less than 15^{o}C may be at greatest risk of extinction (Fig. 2).

Body size was a significant predictor of extinction in the best-supported model, with smaller-bodied marine invertebrates more prone to extinction on geological time scales (Fig. 2; table S3), consistent with previous studies (*15*, *20*). Body size, however, had a relatively small effect size on patterns of extinction over the Phanerozoic compared to geographic range size, especially when controlling for sample size (fig. S4). This pattern is surprising, considering that body size is often regarded as one of the primary traits selected for during extinction events (*11*, *20*, *59*, *60*). The relatively weak predictive performance of body size could result from methodological constraints, which required us to assume body size was constant over time for a taxon. Thus, we were unable to identify any reductions in body size that may have occurred prior to extinction events. We also tested whether the inclusion of small-bodied taxa, such as ostracods, had an impact on body size selectivity, and found no change in overall patterns when excluding ostracods (table S3,S7).

Many of the predictors considered here may interact to influence taxon extinction risk (*34*, *40*, *41*, *61*), despite their low collinearity (table S2). For example, species that have both small geographic ranges and narrow niche breadths may be even more susceptible to extinction than a taxon with either a small geographic range or a narrow niche breadth. We therefore ran an additional stepwise model selection procedure considering all combinations of two-way and three-way interaction terms for the five predictor variables (n=32 predictors). All five main effect predictors remain statistically significant in these models, with significant interactions found between geographic range and realized thermal tolerance, geographic range size and realized thermal niche breadth, and geographic range size and body size (see SI Methods and Materials (*39*)). The overall best model was found to contain the five main effect variables plus the three significant interaction terms (table S8). Therefore, there are likely cumulative impacts of these variables on extinction risk, where the importance of one variable in predicting extinction outcomes can increase or decrease depending on the level of another term.

Discussion

Our analyses suggest that intrinsic, taxon-specific traits were insufficient to explain empirical patterns in extinction dynamics over the Phanerozoic, and that the magnitude of climate change should also be considered. Climate change affected taxa within a stage, regardless of their traits (table S8), although its impact varied spatially. For example, localized temperature changes may have exceeded 14^{0}C (data S1) for some taxa during the Late Permian, which potentially explains why some species with large geographic ranges and realized thermal breadths (i.e., species predicted to be buffered from extinction) go extinct, especially during mass extinction events that are often characterized by severe climate perturbations (fig. S2). However, we do find a decrease in extinction selectivity during mass extinction intervals (fig. S2), similar to past studies (*20*). The effect of climate change on a taxon's risk of extinction is likely mediated by additional factors, including a taxon's dispersal ability and the shape and orientation of shallow marine habitat (*62*).

Although proxies of thermal physiology emerged as important predictors in our model, geographic range size was the strongest correlate of extinction over the Phanerozoic, consistent with previous studies that found realized thermal niche breadth and geographic range size can be decoupled (*22*, *63*). However, the effects of geographic range are complex and may be more or less important based on the other predictors included here (table S8). Geographic range size may be a strong and

consistent predictor of extinction because this trait emerges from the effects of many other factors, including climatic tolerances, dispersal ability, nutrient availability, and biotic interactions (*17*, *18*). Moreover, a larger geographic range size could serve as a buffer against localized perturbations (be they abiotic or biotic in nature), since regional events would be unlikely to affect all populations equally (*22*, *29*). Thus, large geographic range size may buffer taxa from stochastic events, which could lead to extinction, independent of thermal niche breadth.

Model uncertainty

The extinction risk models presented here represent our current best estimate for constraints on marine invertebrate extinction over the Phanerozoic and serve as a framework for future work incorporating additional parameters. For example, other intrinsic traits and factors, such as abundance (*17*), might affect extinction risk over geological timescales. These factors, however, are difficult to characterise for extinct taxa spanning millions of years. Abundance, in particular, is challenging to determine even in the modern (*17*), and thus we focus only on traits that can be estimated more reliably. In addition, there are many other potential extrinsic drivers not considered here, such as anoxia (*8*), euxinia, and ocean acidification (*64*). These extrinsic factors may be especially important during mass extinction events (*64*), but geochemical proxies are not available in our climate models and difficult to estimate over the Phanerozoic.

Our analyses may also be affected by uncertainty in both paleogeographic reconstructions, especially pre-Jurassic (*45*, *65*), and in climate simulations (*49*), since the physiological and extrinsic predictors are based on proxies for spatiotemporal climate that contain uncertainty (see SI Methods and Materials (*39*)). New paleoclimate estimates may impact our understanding of extinction correlates over Phanerozoic timescales. However, the insensitivity of our results to different climate model boundary conditions suggests our patterns may be robust to some degree of climate model uncertainty (fig. S5).

Finally, the temporal resolution of our stage-level analyses may affect our ability to compare climate changes between the past and present (*66*, *67*). However, our analyses find a robust selectivity signal that has persisted over 485 Ma (Fig. 2, fig. S2), despite major changes in boundary conditions (*62*, *68*, *69*), suggesting the determinants of extinction risk studied here may be constant regardless of temporal scale.

Conclusions

The predictable nature of extinction on geological timescales, despite major variations in boundary conditions, environment, biota, and extinction magnitude, is striking. Our results have implications for understanding the fate of biodiversity in response to changing climate today and into the future. Based on our models, taxa with narrow realized thermal breadths of less than 15^{o}C, living predominantly in the poles or tropics, are likely to be at greatest risk of extinction (Fig. 2). These results reinforce the importance of climate in driving extinction on macroevolutionary timescales (*8*–*10*, *70*). We find that extinction dynamics over the past 485 Ma cannot be explained fully without considering the magnitude of climate change in addition to the other physiological and taxonomic trait predictors. Our baseline extinction risk estimates derived from the geological past suggest that biodiversity may face a harrowing future given projected climate change estimates (*2*), which could be made worse when interacting with other anthropogenic extinction drivers (*5*, *13*).

References

1. Intergovernmental Panel on Climate Change (IPCC), Ed., "Changing State of the Climate System" in *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, 2023; https://www.cambridge.org/core/books/climate-change-2021-the-physical-science-basis/changing-state-of-the-climate-system/8B8FB442BA38A2C314ADD4136A9FE2E8), pp. 287–422.

2. C. Lyon, E. E. Saupe, C. J. Smith, D. J. Hill, A. P. Beckerman, L. C. Stringer, R. Marchant, J. McKay, A. Burke, P. O'Higgins, A. M. Dunhill, B. J. Allen, J. Riel-Salvatore, T. Aze, Climate change research and action must look beyond 2100. *Glob. Change Biol.* **28**, 349–361 (2022).

3. Intergovernmental Panel on Climate Change (IPCC), Ed., "Future Global Climate: Scenario-based Projections and Near-term Information" in *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, 2023; https://www.cambridge.org/core/books/climate-change-2021-the-physical-science-basis/future-global-climate-scenariobased-projections-and-nearterm-information/309359EDDCFABB031C078AE20CEE04FD), pp. 553–672.

4. C. D. Thomas, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. Townsend Peterson, O. L. Phillips, S. E. Williams, Extinction risk from climate change. *Nature* **427**, 145–148 (2004).

5. Global Assessment Report on Biodiversity and Ecosystem Services | IPBES secretariat (2019). https://www.ipbes.net/node/35274.

6. M. Yasuhara, C.-L. Wei, M. Kucera, M. J. Costello, D. P. Tittensor, W. Kiessling, T. C. Bonebrake, C. R. Tabor, R. Feng, A. Baselga, K. Kretschmer, B. Kusumoto, Y. Kubota, Past and future decline of tropical pelagic biodiversity. *Proc. Natl. Acad. Sci.* **117**, 12891–12896 (2020).

7. M. Yasuhara, C. A. Deutsch, Paleobiology provides glimpses of future ocean. *Science* **375**, 25–26 (2022).

8. R. G. Stockey, A. Pohl, A. Ridgwell, S. Finnegan, E. A. Sperling, Decreasing Phanerozoic extinction intensity as a consequence of Earth surface oxygenation and metazoan ecophysiology. *Proc. Natl. Acad. Sci.* **118**, e2101900118 (2021).

9. J. L. Penn, C. Deutsch, J. Payne, E. A. Sperling, Aerobic Marine Habitat Loss During the Late Permian Extinction. **2016**, PP31A-2267 (2016).

10. C. Pietsch, E. Petsios, D. J. Bottjer, Sudden and extreme hyperthermals, low-oxygen, and sediment influx drove community phase shifts following the end-Permian mass extinction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **451**, 183–196 (2016).

11. M. L. McKinney, Extinction Vulnerability and Selectivity: Combining Ecological and Paleontological Views. *Annu. Rev. Ecol. Syst.* **28**, 495–516 (1997).

12. D. Jablonski, Background and Mass Extinctions: The Alternation of Macroevolutionary Regimes. *Science* **231**, 129–133 (1986).

13. S. Finnegan, S. C. Anderson, P. G. Harnik, C. Simpson, D. P. Tittensor, J. E. Byrnes, Z. V. Finkel, D. R. Lindberg, L. H. Liow, R. Lockwood, H. K. Lotze, C. R. McClain, J. L. McGuire, A. O'Dea, J. M. Pandolfi, Paleontological baselines for evaluating extinction risk in the modern oceans. *Science* **348**, 567–570 (2015).

14. J. L. Payne, A. M. Bush, N. A. Heim, M. L. Knope, D. J. McCauley, Ecological selectivity of the emerging mass extinction in the oceans. *Science* **353**, 1284–1286 (2016).

15. K. S. Collins, S. M. Edie, G. Hunt, K. Roy, D. Jablonski, Extinction risk in extant marine species integrating palaeontological and biodistributional data. *Proc. R. Soc. B Biol. Sci.* **285**, 20181698 (2018).

16. J. L. Penn, C. Deutsch, Avoiding ocean mass extinction from climate warming. *Science* **376**, 524–526 (2022).

17. M. M. Casey, E. E. Saupe, B. S. Lieberman, The effects of geographic range size and abundance on extinction during a time of "sluggish"' evolution. *Paleobiology* **47**, 54–67 (2021).

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

19. J. L. Payne, N. A. Heim, Body size, sampling completeness, and extinction risk in the marine fossil record. *Paleobiology* **46**, 23–40 (2020).

20. P. M. Monarrez, N. A. Heim, J. L. Payne, Mass extinctions alter extinction and origination dynamics with respect to body size. *Proc. R. Soc. B Biol. Sci.* **288**, 20211681 (2021).

21. E. E. Saupe, N. Barve, H. L. Owens, J. C. Cooper, P. A. Hosner, A. T. Peterson, Reconstructing Ecological Niche Evolution When Niches Are Incompletely Characterized. *Syst. Biol.* **67**, 428–438 (2018).

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

23. T. W. Kammer, T. K. Baumiller, W. I. Ausich, Species longevity as a function of niche breadth: Evidence from fossil crinoids. *Geology* **25**, 219 (1997).

24. K. A. Carscadden, N. C. Emery, C. A. Arnillas, M. W. Cadotte, M. E. Afkhami, D. Gravel, S. W. Livingstone, J. J. Wiens, Niche Breadth: Causes and Consequences for Ecology, Evolution, and Conservation. *Q. Rev. Biol.* **95**, 179–214 (2020).

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

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

300 27. C. J. Reddin, M. Aberhan, N. B. Raja, Á. T. Kocsis, Global warming generates
301 predictable extinctions of warm- and cold-water marine benthic invertebrates via
302 thermal habitat loss. *Glob. Change Biol.* **28**, 5793–5807 (2022).

303 28. C. Chen, T. A. Jefferson, B. Chen, Y. Wang, Geographic range size, water
304 temperature, and extrinsic threats predict the extinction risk in global cetaceans. *Glob.*
305 *Change Biol.* **28**, 6541–6555 (2022).

306 29. J. L. Payne, S. Finnegan, The effect of geographic range on extinction risk during
307 background and mass extinction. *Proc. Natl. Acad. Sci.* **104**, 10506–10511 (2007).

308 30. M. E. Clapham, J. L. Payne, Acidification, anoxia, and extinction: A multiple logistic
309 regression analysis of extinction selectivity during the Middle and Late Permian.
310 *Geology* **39**, 1059–1062 (2011).

311 31. F. T. Dahlke, S. Wohlrab, M. Butzin, H.-O. Pörtner, Thermal bottlenecks in the life
312 cycle define climate vulnerability of fish. *Science* **369**, 65–70 (2020).

313 32. J. M. Sunday, A. E. Bates, N. K. Dulvy, Thermal tolerance and the global
314 redistribution of animals. *Nat. Clim. Change* **2**, 686–690 (2012).

315 33. K. Duffy, T. C. Gouhier, A. R. Ganguly, Climate-mediated shifts in temperature
316 fluctuations promote extinction risk. *Nat. Clim. Change* **12**, 1037–1044 (2022).

317 34. P. B. Day, R. D. Stuart-Smith, G. J. Edgar, A. E. Bates, Species' thermal ranges
318 predict changes in reef fish community structure during 8 years of extreme temperature
319 variation. *Divers. Distrib.* **24**, 1036–1046 (2018).

320 35. C. J. Reddin, Á. T. Kocsis, W. Kiessling, Climate change and the latitudinal selectivity
321 of ancient marine extinctions. *Paleobiology* **45**, 70–84 (2019).

322 36. M. G. Powell, B. R. Moore, T. J. Smith, Origination, extinction, invasion, and
323 extirpation components of the brachiopod latitudinal biodiversity gradient through the
324 Phanerozoic Eon. *Paleobiology* **41**, 330–341 (2015).

325 37. B. J. Allen, M. E. Clapham, E. E. Saupe, P. B. Wignall, D. J. Hill, A. M. Dunhill,
326 Estimating spatial variation in origination and extinction in deep time: a case study
327 using the Permian–Triassic marine invertebrate fossil record. *Paleobiology*, 1–18
328 (2023).

329 38. J. Soberón, Grinnellian and Eltonian niches and geographic distributions of species.
330 *Ecol. Lett.* **10**, 1115–1123 (2007).

331 39. Materials and methods are available as supplementary materials.

40. K. J. Gaston, T. M. Blackburn, Global Scale Macroecology: Interactions between Population Size, Geographic Range Size and Body Size in the Anseriformes. *J. Anim. Ecol.* **65**, 701–714 (1996).

41. J. M. Sunday, A. E. Bates, N. K. Dulvy, Global analysis of thermal tolerance and latitude in ectotherms. *Proc. R. Soc. B Biol. Sci.* **278**, 1823–1830 (2011).

42. Paleobiology Database (2023). https://paleobiodb.org.

43. D. Jablonski, W. G. Chaloner, Extinctions in the Fossil Record [and Discussion]. *Philos. Trans. Biol. Sci.* **344**, 11–17 (1994).

44. T. J. Webb, A. Lines, L. M. Howarth, Occupancy-derived thermal affinities reflect known physiological thermal limits of marine species. *Ecol. Evol.* **10**, 7050–7061 (2020).

45. P. J. Valdes, E. Armstrong, M. P. S. Badger, C. D. Bradshaw, F. Bragg, M. Crucifix, T. Davies-Barnard, J. J. Day, A. Farnsworth, C. Gordon, P. O. Hopcroft, A. T. Kennedy, N. S. Lord, D. J. Lunt, A. Marzocchi, L. M. Parry, V. Pope, W. H. G. Roberts, E. J. Stone, G. J. L. Tourte, J. H. T. Williams, The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geosci. Model Dev.* **10**, 3715–3743 (2017).

46. N. Sagoo, P. Valdes, R. Flecker, L. J. Gregoire, The Early Eocene equable climate problem: can perturbations of climate model parameters identify possible solutions? *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **371**, 20130123 (2013).

47. P. Ross, "Deep ocean circulation during the early Eocene: A model-data comparison," thesis, University College London (2023).

48. J. W. B. Rae, Y. G. Zhang, X. Liu, G. L. Foster, H. M. Stoll, R. D. M. Whiteford, Atmospheric CO2 over the Past 66 Million Years from Marine Archives. *Annu. Rev. Earth Planet. Sci.* **49**, 609–641 (2021).

49. D. C. Wade, N. L. Abraham, A. Farnsworth, P. J. Valdes, F. Bragg, A. T. Archibald, Simulating the climate response to atmospheric oxygen variability in the Phanerozoic: a focus on the Holocene, Cretaceous and Permian. *Clim. Past* **15**, 1463–1483 (2019).

50. J. R. Hendricks, E. E. Saupe, C. E. Myers, E. J. Hermsen, W. D. Allmon, The Generification of the Fossil Record. *Paleobiology* **40**, 511–528 (2014).

51. M. Foote, K. A. Ritterbush, A. I. Miller, Geographic ranges of genera and their constituent species: structure, evolutionary dynamics, and extinction resistance. *Paleobiology* **42**, 269–288 (2016).

52. K. P. Burnham, D. R. Anderson, K. P. Huyvaert, AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* **65**, 23–35 (2011).

53. A. F. Zuur, E. N. Ieno, C. S. Elphick, A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3–14 (2010).

54. H. Song, D. B. Kemp, L. Tian, D. Chu, H. Song, X. Dai, Thresholds of temperature change for mass extinctions. *Nat. Commun.* **12**, 4694 (2021).

55. C. J. Reddin, P. S. Nätscher, Á. T. Kocsis, H.-O. Pörtner, W. Kiessling, Marine clade sensitivities to climate change conform across timescales. *Nat. Clim. Change* **10**, 249–253 (2020).

56. J. L. Penn, C. Deutsch, J. L. Payne, E. A. Sperling, Temperature-dependent hypoxia explains biogeography and severity of end-Permian marine mass extinction. *Science* **362**, eaat1327 (2018).

57. H. O. Pörtner, R. Knust, Climate Change Affects Marine Fishes Through the Oxygen Limitation of Thermal Tolerance. *Science* **315**, 95–97 (2007).

58. E. A. Sperling, T. H. Boag, M. I. Duncan, C. R. Endriga, J. A. Marquez, D. B. Mills, P. M. Monarrez, J. A. Sclafani, R. G. Stockey, J. L. Payne, Breathless through Time: Oxygen and Animals across Earth's History. *Biol. Bull.*, 000–000 (2022).

59. F. A. Smith, J. L. Payne, N. A. Heim, M. A. Balk, S. Finnegan, M. Kowalewski, S. K. Lyons, C. R. McClain, D. W. McShea, P. M. Novack-Gottshall, P. S. Anich, S. C. Wang, Body Size Evolution Across the Geozoic. *Annu. Rev. Earth Planet. Sci.* **44**, 523–553 (2016).

60. J. L. Payne, A. M. Bush, N. A. Heim, M. L. Knope, D. J. McCauley, Ecological selectivity of the emerging mass extinction in the oceans. *Science* **353**, 1284–1286 (2016).

61. S. J. Jacquemin, J. C. Doll, Body Size and Geographic Range Do Not Explain Long Term Variation in Fish Populations: A Bayesian Phylogenetic Approach to Testing Assembly Processes in Stream Fish Assemblages. *PLoS ONE* **9**, e93522 (2014).

62. E. E. Saupe, H. Qiao, Y. Donnadieu, A. Farnsworth, A. T. Kennedy-Asser, J.-B. Ladant, D. J. Lunt, A. Pohl, P. Valdes, S. Finnegan, Extinction intensity during Ordovician and Cenozoic glaciations explained by cooling and palaeogeography. *Nat. Geosci.* **13**, 65–70 (2020).

63. A. Tomašových, D. Jablonski, S. K. Berke, A. Z. Krug, J. W. Valentine, Nonlinear thermal gradients shape broad-scale patterns in geographic range size and can reverse Rapoport's rule. *Glob. Ecol. Biogeogr.* **24**, 157–167 (2015).

64. D. P. G. Bond, S. E. Grasby, On the causes of mass extinctions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **478**, 3–29 (2017).

65. L. Buffan, L. A. Jones, M. Domeier, C. R. Scotese, S. Zahirovic, S. Varela, Mind the uncertainty: Global plate model choice impacts deep-time palaeobiological studies. *Methods Ecol. Evol.* **n/a**.

66. C. Spalding, P. M. Hull, Towards quantifying the mass extinction debt of the Anthropocene. *Proc. R. Soc. B Biol. Sci.* **288**, 20202332 (2021).

67. D. B. Kemp, K. Eichenseer, W. Kiessling, Maximum rates of climate change are systematically underestimated in the geological record. *Nat. Commun.* **6**, 8890 (2015).

68. A. B. Smith, A. J. McGOWAN, The Shape of the Phanerozoic Marine Palaeodiversity Curve: How Much Can Be Predicted from the Sedimentary Rock Record of Western Europe? *Palaeontology* **50**, 765–774 (2007).

69. A. Zaffos, S. Finnegan, S. E. Peters, Plate tectonic regulation of global marine animal diversity. *Proc. Natl. Acad. Sci.* **114**, 5653–5658 (2017).

70. M. W. Schwartz, L. R. Iverson, A. M. Prasad, S. N. Matthews, R. J. O'Connor, Predicting Extinctions as a Result of Climate Change. *Ecology* **87**, 1611–1615 (2006).

71. S. E. Peters, M. McClennen, The Paleobiology Database application programming interface. *Paleobiology* **42**, 1–7 (2016).

72. Á. T. Kocsis, C. J. Reddin, J. Alroy, W. Kiessling, The r package divDyn for quantifying diversity dynamics using fossil sampling data. *Methods Ecol. Evol.* **10**, 735–743 (2019).

73. N. A. Heim, M. L. Knope, E. K. Schaal, S. C. Wang, J. L. Payne, Cope's rule in the evolution of marine animals. *Science* **347**, 867–870 (2015).

74. S. A. F. Darroch, M. M. Casey, G. S. Antell, A. Sweeney, E. E. Saupe, High Preservation Potential of Paleogeographic Range Size Distributions in Deep Time. *Am. Nat.* **196**, 454–471 (2020).

75. M. M. Casey, E. E. Saupe, B. S. Lieberman, The effects of geographic range size and abundance on extinction during a time of "sluggish"' evolution. *Paleobiology* **47**, 54–67 (2021).

76. F. Gradstein, J. Ogg, "The Chronostratigraphic Scale" (2020), pp. 21–32.

77. G. Guinot, F. L. Condamine, Global impact and selectivity of the Cretaceous-Paleogene mass extinction among sharks, skates, and rays. *Science* **379**, 802–806 (2023).

78. J. L. Payne, S. Finnegan, The effect of geographic range on extinction risk during background and mass extinction. *Proc. Natl. Acad. Sci.* **104**, 10506–10511 (2007).

79. S. Finnegan, S. C. Anderson, P. G. Harnik, C. Simpson, D. P. Tittensor, J. E. Byrnes, Z. V. Finkel, D. R. Lindberg, L. H. Liow, R. Lockwood, H. K. Lotze, C. R. McClain, J. L. McGuire, A. O'Dea, J. M. Pandolfi, Paleontological baselines for evaluating extinction risk in the modern oceans. *Science* **348**, 567–570 (2015).

80. M. E. Clapham, J. L. Payne, Acidification, anoxia, and extinction: A multiple logistic regression analysis of extinction selectivity during the Middle and Late Permian. *Geology* **39**, 1059–1062 (2011).

77. D. Jablonski, Body size and macroevolution. *Evol. Paleobiology*, 256–289 (1996).

82. C. Deutsch, J. L. Penn, W. C. E. P. Verberk, K. Inomura, M.-G. Endress, J. L. Payne, Impact of warming on aquatic body sizes explained by metabolic scaling from microbes to macrofauna. *Proc. Natl. Acad. Sci.* **119**, e2201345119 (2022).

83. E. Ramirez Llodra, Fecundity and life-history strategies in marine invertebrates. *Adv. Mar. Biol.* **43**, 87–170 (2002).

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

85. J. Soberon, A. T. Peterson, Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodivers. Inform.* **2** (2005).

86. S. E. Lester, B. I. Ruttenberg, S. D. Gaines, B. P. Kinlan, The relationship between dispersal ability and geographic range size. *Ecol. Lett.* **10**, 745–758 (2007).

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

88. Kowalewski, P. Novack-Gottshall, Resampling Methods in Paleontology. *Paleontol. Soc. Spec. Pap.* **16**, 19–54 (2010).

89. rgeos package - RDocumentation. https://www.rdocumentation.org/packages/rgeos/versions/0.6-4.

90. R. R. Wilcox, How many discoveries have been lost by ignoring modern statistical methods? *Am. Psychol.* **53**, 300–314 (1998).

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

92. B. Braschler, G. A. Duffy, E. Nortje, S. Kritzinger-Klopper, D. du Plessis, N. Karenyi, R. I. Leihy, S. L. Chown, Realised rather than fundamental thermal niches predict site occupancy: Implications for climate change forecasting. *J. Anim. Ecol.* **89**, 2863–2875 (2020).

93. R. J. Hijmans, J. van Etten, M. Sumner, J. Cheng, D. Baston, A. Bevan, R. Bivand, L. Busetto, M. Canty, B. Fasoli, D. Forrest, A. Ghosh, D. Golicher, J. Gray, J. A. Greenberg, P. Hiemstra, K. Hingee, A. Ilich, I. for M. A. Geosciences, C. Karney, M. Mattiuzzi, S. Mosher, B. Naimi, J. Nowosad, E. Pebesma, O. P. Lamigueiro, E. B. Racine, B. Rowlingson, A. Shortridge, B. Venables, R. Wueest, raster: Geographic Data Analysis and Modeling, version 3.6-26 (2023); https://cran.r-project.org/web/packages/raster/index.html.

94. F. Rodriguez-Sanchez, rSDM: Species distribution and niche modelling in R, (2020); https://github.com/Pakillo/rSDM.

95. P. J. Valdes, E. Armstrong, M. P. S. Badger, C. D. Bradshaw, F. Bragg, M. Crucifix, T. Davies-Barnard, J. J. Day, A. Farnsworth, C. Gordon, P. O. Hopcroft, A. T. Kennedy, N. S. Lord, D. J. Lunt, A. Marzocchi, L. M. Parry, V. Pope, W. H. G. Roberts, E. J. Stone, G. J. L. Tourte, J. H. T. Williams, The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geosci. Model Dev.* **10**, 3715–3743 (2017).

96. I. S. Fenton, T. Aze, A. Farnsworth, P. Valdes, E. E. Saupe, Origination of the modern-style diversity gradient 15 million years ago. *Nature* **614**, 708–712 (2023).

97. P. M. Cox, R. A. Betts, C. B. Bunton, R. L. H. Essery, P. R. Rowntree, J. Smith, The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dyn.* **15**, 183–203 (1999).

98. P. M. Cox, R. A. Betts, C. B. Bunton, R. L. H. Essery, P. R. Rowntree, J. Smith, The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dyn.* **15**, 183–203 (1999).

99. M. Collins, S. Tett, C. Cooper, The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **17**, 61–81 (2001).

100. A. Farnsworth, D. J. Lunt, C. L. O'Brien, G. L. Foster, G. N. Inglis, P. Markwick, R. D. Pancost, S. A. Robinson, Climate Sensitivity on Geological Timescales Controlled by Nonlinear Feedbacks and Ocean Circulation. *Geophys. Res. Lett.* **46**, 9880–9889 (2019).

101. M. Kageyama, P. Braconnot, S. P. Harrison, A. M. Haywood, J. H. Jungclaus, B. L. Otto-Bliesner, J.-Y. Peterschmitt, A. Abe-Ouchi, S. Albani, P. J. Bartlein, C. Brierley, M. Crucifix, A. Dolan, L. Fernandez-Donado, H. Fischer, P. O. Hopcroft, R. F. Ivanovic, F. Lambert, D. J. Lunt, N. M. Mahowald, W. R. Peltier, S. J. Phipps, D. M. Roche, G. A. Schmidt, L. Tarasov, P. J. Valdes, Q. Zhang, T. Zhou, The PMIP4 contribution to CMIP6 – Part 1: Overview and over-arching analysis plan. *Geosci. Model Dev.* **11**, 1033–1057 (2018).

102. D. J. Lunt, M. Huber, E. Anagnostou, M. L. J. Baatsen, R. Caballero, R. DeConto, H. A. Dijkstra, Y. Donnadieu, D. Evans, R. Feng, G. L. Foster, E. Gasson, A. S. von der Heydt, C. J. Hollis, G. N. Inglis, S. M. Jones, J. Kiehl, S. Kirtland Turner, R. L. Korty, R. Kozdon, S. Krishnan, J.-B. Ladant, P. Langebroek, C. H. Lear, A. N. LeGrande, K. Littler, P. Markwick, B. Otto-Bliesner, P. Pearson, C. J. Poulsen, U. Salzmann, C. Shields, K. Snell, M. Stärz, J. Super, C. Tabor, J. E. Tierney, G. J. L. Tourte, A. Tripati, G. R. Upchurch, B. S. Wade, S. L. Wing, A. M. E. Winguth, N. M. Wright, J. C. Zachos, R. E. Zeebe, The DeepMIP contribution to PMIP4: experimental design for model simulations of the EECO, PETM, and pre-PETM (version 1.0). *Geosci. Model Dev.* **10**, 889–901 (2017).

103. P. J. Valdes, C. R. Scotese, D. J. Lunt, Deep ocean temperatures through time. *Clim. Past* **17**, 1483–1506 (2021).

104. G. L. Foster, D. L. Royer, D. J. Lunt, Future climate forcing potentially without precedent in the last 420 million years. *Nat. Commun.* **8**, 14845 (2017).

105. D. O. Gough, "Solar Interior Structure and Luminosity Variations" in *Physics of Solar Variations*, V. Domingo, Ed. (Springer Netherlands, Dordrecht, 1981), pp. 21–34.

106. J. T. Kiehl, C. A. Shields, Sensitivity of the Palaeocene–Eocene Thermal Maximum climate to cloud properties. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **371**, 20130093 (2013).

107. P. J. Irvine, L. J. Gregoire, D. J. Lunt, P. J. Valdes, An efficient method to generate a perturbed parameter ensemble of a fully coupled AOGCM without flux-adjustment. *Geosci. Model Dev.* **6**, 1447–1462 (2013).

108. E. J. Judd, J. E. Tierney, B. T. Huber, S. L. Wing, D. J. Lunt, H. L. Ford, G. N. Inglis, E. L. McClymont, C. L. O'Brien, R. Rattanasriampaipong, W. Si, M. L. Staitis, K. Thirumalai, E. Anagnostou, M. J. Cramwinckel, R. R. Dawson, D. Evans, W. R. Gray, E. L. Grossman, M. J. Henehan, B. N. Hupp, K. G. MacLeod, L. K. O'Connor, M. L. Sánchez Montes, H. Song, Y. G. Zhang, The PhanSST global database of Phanerozoic sea surface temperature proxy data. *Sci. Data* **9**, 753 (2022).

109. A. Farnsworth, P. J. Valdes, R. A. Spicer, L. Ding, C. Witkowski, V. Lauretano, T. Su, S. Li, S. Li, Z. Zhou, Paleoclimate model-derived thermal lapse rates: Towards increasing precision in paleoaltimetry studies. *Earth Planet. Sci. Lett.* **564**, 116903 (2021).

110. J. N. Bahcall, M. H. Pinsonneault, S. Basu, Solar Models: Current Epoch and Time Dependences, Neutrinos, and Helioseismological Properties. *Astrophys. J.* **555**, 990 (2001).

111. D. J. Lunt, A. Farnsworth, C. Loptson, G. L. Foster, P. Markwick, C. L. O'Brien, R. D. Pancost, S. A. Robinson, N. Wrobel, Palaeogeographic controls on climate and proxy interpretation. *Clim. Past* **12**, 1181–1198 (2016).

112. N. J. Burls, C. D. Bradshaw, A. M. De Boer, N. Herold, M. Huber, M. Pound, Y. Donnadieu, A. Farnsworth, A. Frigola, E. Gasson, A. S. von der Heydt, D. K. Hutchinson, G. Knorr, K. T. Lawrence, C. H. Lear, X. Li, G. Lohmann, D. J. Lunt, A. Marzocchi, M. Prange, C. A. Riihimaki, A.-C. Sarr, N. Siler, Z. Zhang, Simulating Miocene Warmth: Insights From an Opportunistic Multi-Model Ensemble (MioMIP1). *Paleoceanogr. Paleoclimatology* **36**, e2020PA004054 (2021).

113. P. J. Valdes, E. Armstrong, M. P. S. Badger, C. D. Bradshaw, F. Bragg, M. Crucifix, T. Davies-Barnard, J. J. Day, A. Farnsworth, C. Gordon, P. O. Hopcroft, A. T. Kennedy, N. S. Lord, D. J. Lunt, A. Marzocchi, L. M. Parry, V. Pope, W. H. G. Roberts, E. J. Stone, G. J. L. Tourte, J. H. T. Williams, The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geosci. Model Dev.* **10**, 3715–3743 (2017).

114. E. Hawkins, R. Sutton, The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bull. Am. Meteorol. Soc.* **90**, 1095–1108 (2009).

115. lme4 package - RDocumentation. https://www.rdocumentation.org/packages/lme4/versions/1.1-33.

116. L. H. Liow, Estimating Rates and Probabilities of Origination and Extinction Using Taxonomic Occurrence Data: Capture-Mark-Recapture (CMR) Approaches. (2010).

117. J. J. Sepkoski, A Kinetic Model of Phanerozoic Taxonomic Diversity. III. Post-Paleozoic Families and Mass Extinctions. *Paleobiology* **10**, 246–267 (1984).

118. S. E. Peters, Environmental determinants of extinction selectivity in the fossil record. *Nature* **454**, 626–629 (2008).

119. J. Alroy, M. Aberhan, D. J. Bottjer, M. Foote, F. T. Fürsich, P. J. Harries, A. J. W. Hendy, S. M. Holland, L. C. Ivany, W. Kiessling, M. A. Kosnik, C. R. Marshall, A. J. McGowan, A. I. Miller, T. D. Olszewski, M. E. Patzkowsky, S. E. Peters, L. Villier, P. J. Wagner, N. Bonuso, P. S. Borkow, B. Brenneis, M. E. Clapham, L. M. Fall, C. A. Ferguson, V. L. Hanson, A. Z. Krug, K. M. Layou, E. H. Leckey, S. Nürnberg, C. M. Powers, J. A. Sessa, C. Simpson, A. Tomašových, C. C. Visaggi, Phanerozoic Trends in the Global Diversity of Marine Invertebrates. *Science* **321**, 97–100 (2008).

120. R. A. Close, R. B. J. Benson, E. E. Saupe, M. E. Clapham, R. J. Butler, The spatial structure of Phanerozoic marine animal diversity. *Science* **368**, 420–424 (2020).

121. S. C. Wang, A. M. Bush, Adjusting global extinction rates to account for taxonomic susceptibility. *Paleobiology* **34**, 434–455 (2008).

122. J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, *Multivariate Data Analysis* (Cengage, 2019).

123. L. S. Aiken, S. G. West, R. R. Reno, *Multiple Regression: Testing and Interpreting Interactions* (SAGE, 1991).

124. MuMIn package - RDocumentation. https://www.rdocumentation.org/packages/MuMIn/versions/1.47.5.

116. K. P. Burnham, Model selection and multimodel inference. *Pract. Inf.-Theor. Approach* (1998).

126. S. A. Richards, Testing Ecological Theory Using the Information-Theoretic Approach: Examples and Cautionary Results. *Ecology* **86**, 2805–2814 (2005).

127. D. I. Warton, L. Thibaut, Y. A. Wang, The PIT-trap—A "model-free" bootstrap procedure for inference about regression models with discrete, multivariate responses. *PLOS ONE* **12**, e0181790 (2017).

128. F. Hartig, L. Lohse, DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models, version 0.4.6 (2022); https://cran.r-project.org/web/packages/DHARMa/index.html.

129. P. Green, C. J. MacLeod, SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* **7**, 493–498 (2016).

130. G. T. Antell, R. B. J. Benson, E. E. Saupe, Spatial standardization of taxon occurrence data—a call to action. (2023).

131. B. M. Wolach Kevin R. Murphy, Allen, *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests, Third Edition* (Routledge, New York, ed. 3, 2008).

132. D. Lüdecke, ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *J. Open Source Softw.* **3**, 772 (2018).

133. M. Foote, J. J. Sepkoski, Absolute measures of the completeness of the fossil record. *Nature* **398**, 415–417 (1999).

134. P. C. Fitzgerald, S. J. Carlson, Examining the latitudinal diversity gradient in Paleozoic terebratulide brachiopods: should singleton data be removed? *Paleobiology* **32**, 367–386 (2006).

135. H.-H. M. Huang, M. Yasuhara, D. J. Horne, V. Perrier, A. J. Smith, S. N. Brandão, Ostracods in databases: State of the art, mobilization and future applications. *Mar. Micropaleontol.* **174**, 102094 (2022).

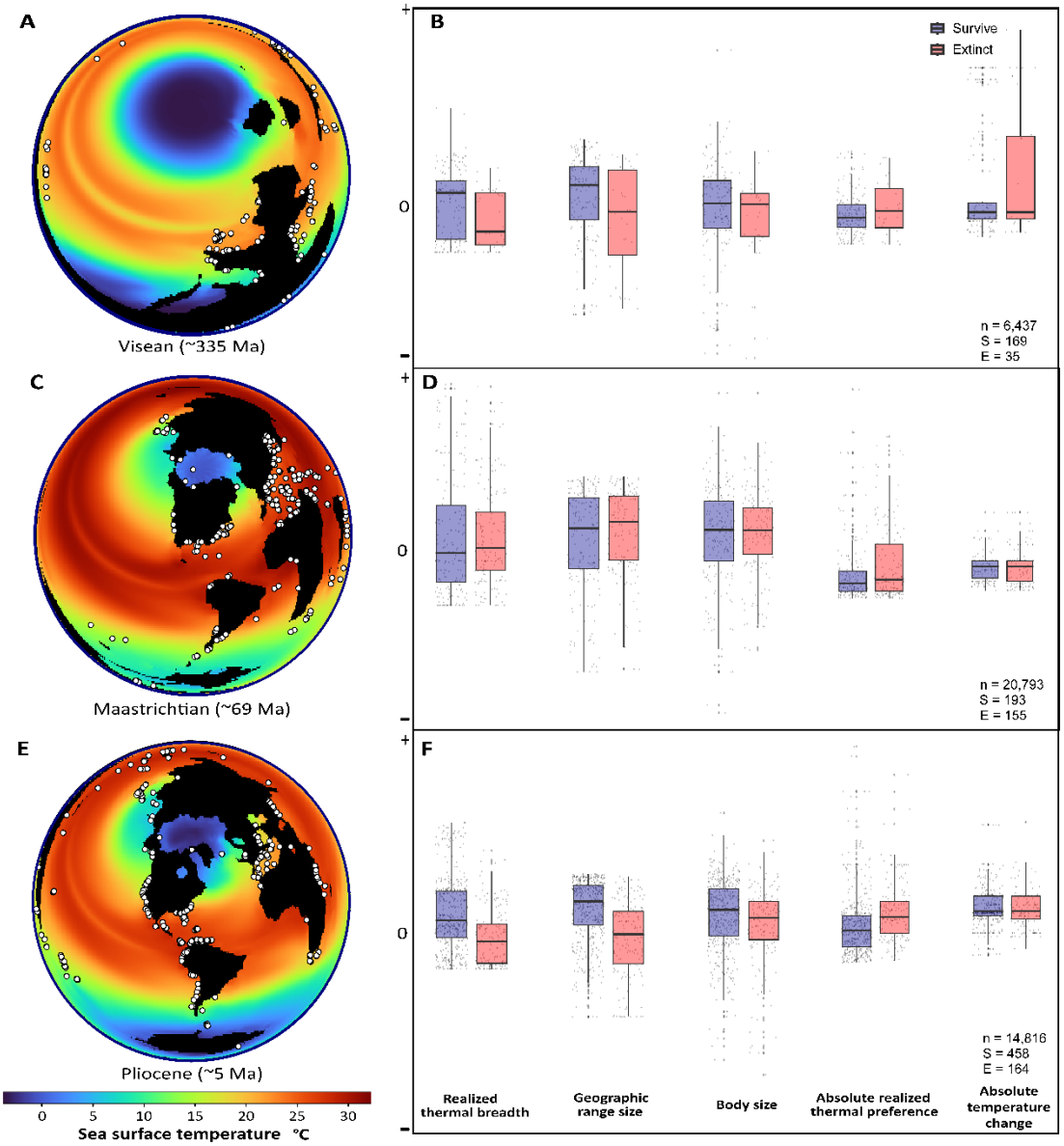
136. D. Lüdecke, M. Ben-Shachar, I. Patil, P. Waggoner, D. Makowski, performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *J. Open Source Softw.* **6**, 3139 (2021).

Acknowledgements

We thank Roger Benson (AMNH), Roger Close (University of Oxford), and Gawain Antell for helpful discussions that informed this contribution, and John Payne, Noel Heim, and Pedro Monnarez for providing the body size data. We are grateful to the contributors of the Paleobiology Database who provided the occurrence data for our analyses. **Funding:** This research was supported by the Leverhulme Prize, Chinese Academy of Sciences Visiting Professorship for Senior International Scientists 2021FSE0001, NERC grant NE/V011405/1, and NERC grant NE/X000222/1. This is Paleobiology Database publication no. xx. **Author contributions:** C.M.M. and E.E.S. conceived the study and wrote the manuscript. C.M.M. conducted the analyses under the supervision of E.E.S.. A.F., P.J.V., and D.J.L. ran all the climate models. All authors provided feedback and valuable discussion on the manuscript text. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The occurrence data used in this study was downloaded from the Paleobiology Database (*42*) and are publicly available. The climate model data and code are available in data S1.

Supplementary Materials

Materials and Methods
Figs. S1 to S5
Tables S1 to S8
References (*71–136*)
Data S1



633

634 **Fig. 1. Sea surface temperature estimates and predictor variable distributions for Phanerozoic**
635 **taxa. (A,C,E)** Sea surface temperature estimates for the Visean (335 Ma), Maastrictian (69 Ma), and
636 Pliocene (5 Ma) using the HadCM3 coupled atmosphere-ocean generalized circulation model,
637 overlain with taxon occurrences (white) used in estimating predictors in (**B,D,F)**. (**B,D,F)** Boxplots of
638 taxon trait estimates for taxa that survived *versus* those that went extinct in the Visean,
639 Maastrichtian, and Pliocene, respectively, where the line represents the median, the black dots the
640 unique estimates for each genus, and the whiskers the 95% confidence interval estimates. The
641 variables were centered and standardized to enable direct comparison between distributions for
642 each predictor, so a value of 0 indicates the mean estimate, positive values represent values greater
643 than the mean, and negative values indicate values less than the mean. N = the number of unique
644 taxon occurrences in the stage, S = the number of survivors, and E = the number of taxa that go
645 extinct in the respective time interval.

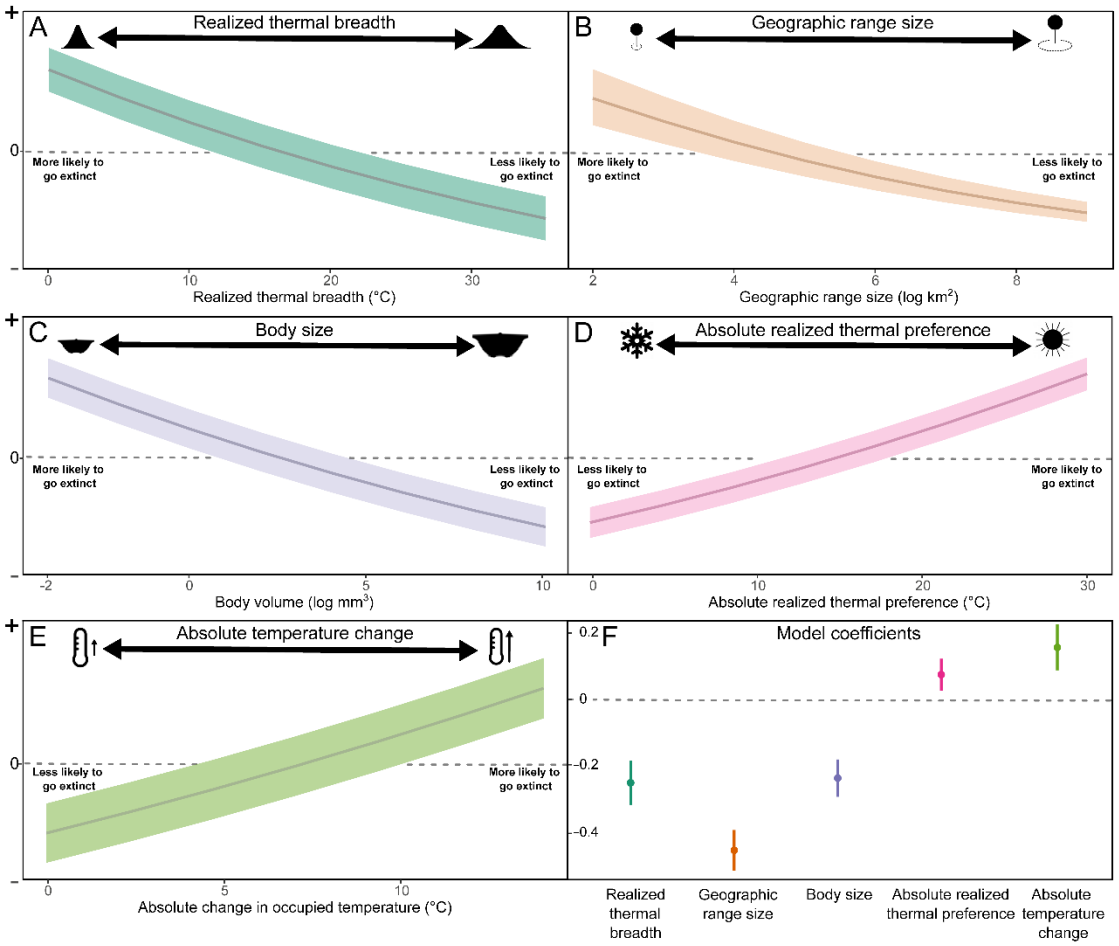


Fig. 2. Extinction risk estimates for each predictor variable included in the best-supported model. **(A** to **E)** Marginal effects plots for each predictor variable in the best model. The y-axis corresponds to the scaled probability of a taxon going extinct with a given trait value, holding the other predictors constant (see SI Methods and Materials (*39*)). The dark lines correspond to the median estimate and the shaded regions the 95% confidence intervals. Positive values indicate greater probability for extinction at the given predictor values, and negative values indicate a taxon is less likely to go extinct at the given predictor values. **(D)** Realized thermal preference was measured as the absolute value of the deviation of a taxon's median occupied temperature from the median occupied temperatures of all taxon occurrences within a stage. **(E)** The absolute value of climate change experienced by each unique taxon was calculated from a given stage (n) to stage n+1. **(F)** Extinction selectivity in log-odds for each predictor variable. The y-axis corresponds to the probability of extinction in log-odds. Positive values indicate a positive relationship with extinction risk, and negative values indicate an inverse relationship with extinction risk (refer to table S3 for the detailed coefficients).

Science
AAAS

Supplementary Materials for

Climate change is an important predictor of extinction risk on macroevolutionary timescales

Cooper M. Malanoski*, Alex Farnsworth, Daniel J. Lunt, Paul J. Valdes, Erin E. Saupe*

Corresponding authors: cooper.malanoski@wolfson.ox.ac.uk, erin.saupe@earth.ox.ac.uk

The PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Methods and Materials

Taxon occurrence data

We focus on marine invertebrates belonging to nine classes, including Cephalopoda, Gastropoda, Echinoidea, Crinoidea, Trilobita, Ostracoda, Bivalvia, and the brachiopods Rynchonellata and Strophemenata. Spatio-temporal occurrence data for these groups were downloaded from the Paleobiology Database on January 22nd, 2023 (*42*, *71*), and processed and cleaned using the specifications in Kocsis et al. (*72*). Downloaded data were filtered to only include genera with associated body size information, belonging to classes that contain more than 200 genera, from Heim et al. (*73*) and Monarrez et. al. (*20*). Bony fish were removed from our dataset, since the focus of our study is on marine invertebrate ectotherms. We included genera with three or more occurrences, to ensure robust calculation of geographic range sizes (*74*, *75*). The resulting dataset contained 292,940 spatiotemporally unique occurrences for 9,264 unique genera across 81 Phanerozoic stages. Taxon occurrence data were binned per stage using the stage names and durations found in Gradstein and Ogg (*76*) in the divDyn v. 0.8.2 R package and the pipeline found in Kocsis et al. (*72*). We excluded the Cambrian from our analyses, following previous studies (*20*, *27*), due to the difficulties in stratigraphic resolution and the scarcity of occurrence data for this time interval.

Model parameters

We examine patterns of extinction selectivity across 485 million years of evolution using five predictor variables at the genus level: body size, geographic range size, absolute realized thermal preference, realized thermal niche breadth, and absolute temperature change at occupied sites. Models were re-run at the species level without body size, since body size was measured at the genus-level. All analyses were conducted in the R v. 4.3.1 statistical environment.

Extinction

We determined when each genus went extinct using the modeltab() function within divDyn to calculate first and last occurrences for each genus. The Cambrian and Holocene were only included in range calculation to remove edge effects (*20*, *27*). Analyses were run both with (table S3) and without (table S4) (*10*–*12*) singletons to test the effects of potentially-spurious occurrences on extinction selectivity patterns (*18*, *77*, *78*). The range-through method was used to calculate extinction and determine temporal ranges for generic-level taxa. Since we estimate spatio-temporally unique traits that vary across time bins, we excluded occurrences for stages where we could not estimate their traits, following previous studies (*78*–*80*). Species-level analyses were run using the temporal ranges of cleaned, accepted species names, determined using the specifications in Kocsis et al. (*72*).

Body size

Body size is thought to be a significant predictor of extinction due to its metabolic and life history implications (*19*, *20*, *59*, *73*, *81*, *82*) and is correlated with fecundity and dispersal ability for marine invertebrates (*20*, *83*). We downloaded body size data for marine invertebrates from Monarrez et al. (*20*) to test the relative importance of body size compared to other predictors. Body size was approximated as the $\log_{10}$ of the biovolume of a genus in mm^3, calculated from an ellipsoid based on axis lengths in 3-dimensions from primarily figured specimens (*73*). The maximum body size achieved by each genus over their temporal range was used as proxy for body size. Body size was

assumed to be constant over the range of a genera (*20*, *73*) and was the only variable that did not vary among stages.

Geographic range size

Geographic range size is often regarded as the most important predictor of extinction risk (*18*, *28*, *77*, *78*, *80*, *84*). Geographic range size may serve as a proxy for the ability to withstand a broad array of environmental and biotic conditions, and may also signal good dispersal ability (*85*, *86*). A large geographic range size could serve as a buffer against localized climatic or biotic perturbations, as these changes would be unlikely to affect every population equally (*78*, *84*). Geographic range size was estimated by fitting a minimum convex hull (*17*, *74*, *87*) to generic (or species) level occurrence data in each stage using the Albers equal area conic map projection in km^2 (*17*). Analyses were performed using spatially-unique occurrences for each unique stage x taxon x paleocoordinate combination. Minimum convex hulls were used for area calculations, as they showed good performance for fossil data, whereas alpha convex hull and great circle distance performed less well at identifying geographic range patterns at smaller sample sizes (*87*).

To account for spatial and sample size biases across taxa and time, we employed two subsampling approaches. In the first approach, we aimed to account for spatial outliers / erroneous occurrences, which may lead to inflated trait estimates using jackknife subsampling (*88*). In this approach, we used jackknife subsampling to iteratively remove one occurrence from each unique stage-by-taxon occurrence dataset and calculate geographic range size (*88*). If there were only three occurrences, a 10 km buffer was implemented using the gbuffer() function within the rgeos v. 0.6 R package (*89*) to obtain a non-zero estimate for geographic range (*17*). Minimum convex hull areas were calculated using gconvexhull() and garea() within the rgeos v. 0.6 R package (*89*). The median estimate from the jackknife iterations was used as the estimate of geographic range size for a given taxon and stage. The median was chosen over the mean, since it is better suited for dealing with outliers and changes in data distributions that can occur in paleontological datasets (*90*).

In the second approach, bootstrap subsampling was used to control for both sample size and spatial heterogeneity (*88*). For each stage-by-taxon dataset, we subsampled down to a set number of occurrences and used this subsample to calculate geographic range size. The upper 95% confidence interval from the 100 bootstrap replicates served as the estimate of geographic range size for a given taxon and stage. The upper confidence interval was used because it may better approximate the maximum geographic extent that can be reconstructed using only a minimal number (three, four, or five) of occurrences over 100 iterations. We also ran analyses using the median of the bootstrap replicates, which produced similar results. The number of subsampled occurrences was held constant across taxa and stages, and we repeated analyses by subsampling from three to five occurrences, in increments of one. The main text analyses present selectivity patterns using three subsampled occurrences, to reduce data loss and avoid exclusion of genuinely rare taxa (*91*). However, use of higher subsampled numbers of occurrences produced congruent results (Figs. S4).

Realized thermal niche breath.

Realized niches are those conditions actually occupied by a taxon, which typically represents a subset of the fundamental niche (*38*). The realized niche is governed by myriad factors including climate, biotic interactions, and dispersal ability (*85*). Realized niche breadth in the modern (*92*) and recent geologic past (*84*) has been found to be an important determinant of extinction risk.

Realized thermal niche breath was estimated as the temperature range occupied by a given taxon in a given stage (species and genus level). For example, if a taxon was found in five grid cells

characterized by temperatures of 11°C, 10°C, 11 °C, 14°C, and 20°C for a given stage, the taxon would be estimated to have a realized thermal niche breadth of 10°C.

Spatio-temporally unique temperature estimates for each occurrence were derived from AOGCM's (see "Paleoclimate Simulations" below) for 81 stages over the Phanerozoic (data S1). AOGCM's were assigned to stages based on the midpoints of each stage and the age of the climate model. Taxon occurrences were projected onto the AOGCM's, and the temperature values were extracted using the extract() function from the raster v. 3.6-20 R package (*93*). Temperatures were interpolated for inland occurrences using the points2nearestcell() function from the rSDM v. 0.3.9 R package (*94*); erroneous occurrences more than 1000 km from a data-complete grid cell were removed.

As with geographic range size, we used both jackknife and bootstrap subsampling to correct for spatial and sample size biases. For jackknifing, we iteratively removed one occurrence and calculated the temperature range (°C) occupied by a given taxon in a given stage; the median from the jackknife iterations was used as our realized niche breadth estimate.

The bootstrapping subsampling procedure followed that for geographic range size, wherein each unique taxon-by-stage combination was subsampled down to a set number of occurrences, and the realized thermal breadth calculated (°C). Bootstrap subsampling was repeated 100 times for a unique taxon-by-stage combination, and the upper confidence interval estimate was used to approximate the thermal range.

Realized thermal preference

We tested the selectivity of realized thermal preference using the absolute value of the deviation of a taxon's median occupied temperature from the median occupied temperatures of all taxon occurrences within a stage. This approach was used as a means of standardizing thermal preferences within each stage, since a temperature preference of 35^0C, for example, may not have the same relationship with extinction risk for each stage over the Phanerozoic, and since our statistical modelling framework is not optimized for polynomials and complex relationships. Therefore, the deviation from the median occupied temperature within a stage is a more consistent way of identifying whether a taxon is occupying extreme temperatures prior to extinction. We also ran our analyses using the median occupied temperature as a proxy for thermal preference and our results and conclusions were unchanged. As with realized thermal niche breadth, temperature estimates were derived from AOGCM's for 81 stages over the Phanerozoic. We used both jackknife and bootstrap subsampling to calculate median estimates at both generic and species level, following the approach outlined for realized thermal niche breadth. These realized thermal preferences were used to test whether there is an evolutionary advantage to living in moderate isotherms over extreme isotherms.

Absolute temperature change

We compared the effects of intrinsic trait estimates vs. extrinsic factors such as climate on extinction risk. The relationship between the magnitude of climate change and extinction rates has been studied on a per-stage basis over the Phanerozoic (*54*). However, previous studies (*10*, *54*) have been limited to time series analyses where temperature and climate change were estimated using stage-averaged climate proxy estimates. We calculated the absolute value of climate change experienced by each unique taxon in a given stage from that stage (n) to stage n+1. We extracted the occupied temperatures for stage n and projected the coordinates for that stage onto the climate model for the next stage to calculate the absolute difference between each point in stage n+1 and stage n. The median value of the climate change experienced by each population was taken to

represent the magnitude of occupied climate change for each genus (or species). Both jackknife and bootstrap subsampling approaches were applied to account for potential sampling biases, following the methodology outlined for geographic range size.

Paleoclimate simulations

To estimate realized thermal preference, realized thermal niche breadth, and absolute temperature change, we utilise a new set of paleoclimate model simulations, developed from the simulations presented in Valdes et al. (*95*).

Here we summarise the climate model and simulations presented in Valdes et al. (*95*) and Fenton, et al. (*96*), and describe the developments that have been made since then.

Valdes et al. (*95*) *climate model and simulations*

The Valdes et al. (*95*) paleoclimate model simulations were carried out using an updated version of the UK Met Office coupled Atmosphere-Ocean General Circulation Model (AOGCM), HadCM3. Specifically, we used HadCM3BL-M2.1D, with the Nomenclature of Valdes, et al. (*95*), which has been heavily utilised in both contemporary and paleoclimate studies (see below).

HadCM3L-M2.1D (HadCM3L hereafter) models the atmosphere and ocean using equations solved on the Arakawa B-grid. The climate model has a resolution of 3.75° longitude × 2.5° latitude in the atmosphere and ocean (~250 km grid squares in the tropics), containing 20 vertical levels in the ocean and 19 hybrid levels in the atmosphere. Parameterization is a process whereby small-scale, often complex, features of the climate system, which cannot be physically represented either spatially or temporally at the model grid size, are simplified. For example, cloud microphysics, convection, and oceanic eddies are parameterized, as they can occur at much finer scales from one meter (or less) to several kilometres. Unlike contemporary climate modelling studies, global-scale soil texture, porosity, and albedo are unknown in the deep past. Thus, a globally uniform medium loam soil from the model land surface scheme of MOSES 2.1 (*97*) is utilised. A major advantage of using HadCM3L is the ability to predict vegetation coverage through deep time (which is not common in many other paleoclimate models where vegetation is prescribed, often with little understanding of global vegetation coverage for deep time periods). Our dynamical vegetation model, TRIFFID (Top-Down Representation of Interactive Foliage and Flora Including Dynamics) produces a generalised vegetation representation that is appropriate for deep time simulation due to its simplified representation of common vegetation structures using five Plant Functional Types (PFTs), in this case, i) broadleaf trees, ii) deciduous trees, iii) shrubs, iv) C3, and v) C4 grasses. TRIFFID predicts the properties of global vegetation, and their distribution, in the form of fractional coverage (and thus PFT co-existence) within a grid cell. The PFT co-existence is based on competition equations based on the climate tolerance of these five plant functional types.

The ocean sub-model is the same as Cox et al. (*98*)—a fully primitive equation, three-dimensional model of the ocean. This model includes a second-order numerical scheme and a centred advection scheme to remove nonlinear instabilities. Critically, flux adjustments (such as artificial heat and salinity adjustments in the ocean component model (*99*) to prevent it from drifting to unrealistic values) are not required in this model. This is crucial for long paleoclimate simulations (*100*), which require in excess of 5000 (often more) model years to reach an equilibrium state in the surface and deep ocean, and would not be scientifically justifiable. Sea ice was calculated assuming a consistent salinity for ice using a zero-layer model. Partial sea ice coverage was possible.

The UKMO HadCM3 suite of models has been used in many international contemporary projects from the Coupled Model Intercomparison Project (CMIP 1-5) experiments to paleo-specific intercomparison projects, such as the latest Paleoclimate Modelling Intercomparison Project (PMIP4) (*101*) and the new Deep-time Model Intercomparison Project (DeepMIP) (*102*). The model has demonstrated expertise at simulating the modern-day climate (*95*) and a range of paleoclimate time periods (*100*, *103*). A paleoclimate simulation is only scientifically robust once it has reached an equilibrium state—i.e., once the simulation has fully adjusted to the paleo-time specific boundary conditions (topography, bathymetry, land ice sheet, greenhouse gas concentrations in the atmosphere, solar luminosity), all of which need to be supplied to the model. Many thousands of model years are often required for the deep ocean to reach an equilibrium state where ocean circulation is fully representative of these boundary conditions (*100*, *103*). Although the atmosphere and land surface respond more quickly (in the order of hundreds of years) due to a lower heat capacity than the ocean, both require the ocean to be in an equilibrium state first due to ocean-land-atmosphere feedbacks that can modify one another. Unfortunately, due to computer and cost limitations, contemporary CMIP6 level models, which are often more complex and at a higher spatial resolution, are incapable of being fully equilibrated for paleo time periods currently. Moreover, multi-sensitivity studies cannot be run with such models to understand the processes and mechanisms responsible for the observed change in climate.

A set of 109 'snapshot' climate simulations, covering 540 Ma of the Phanerozoic, were utilised in this study following Valdes et al. (*95*), with time-specific temporal boundary conditions. The EarthByte group, which is part of the PALEOMAP project (*48*) produced each time-specific paleogeography, whereby each paleogeography informed the next to produce the set of Paleogeographic digital elevation models (DEMs).

Each 1°x1° DEM was interpolated onto the HadCM3L 3.75°x2.5° grid. Ice sheet height (m) on land was estimated assuming a simple parabolic shape on the grid. Time-specific average pCO_2 concentrations produced from the proxy-record (*104*), were used for each simulation from the Fortunian (~540 Ma) to the Campanian (~75 Ma), with a further update to the Valdes et al. (*95*) simulations for the Maastrichtian (~69 Ma) to the Pre-industrial (fig. S1). Time-specific solar luminosity is calculated from the solution of Gough (1981) (*105*). Orbital parameters and aerosol concentrations were held constant at pre-industrial values. For this paper, only the last 95 simulations were used, covering the last 485 million years. An equilibrated Pre-industrial state was chosen as the initialisation point for each simulation in the atmosphere and ocean. Vegetation was set uniformly as shrub globally and allowed to evolve via the interactive dynamic vegetation sub-model, TRIFFID (called every 10 model days).

A three-stage protocol was used for each simulation and was fully equilibrated. i) The global volume-integrated annual mean ocean temperature trend was less than 1°C per 1000 years; ii) trends in surface air temperature were less than 0.3°C per 1000 years; and iii) net energy balance at the top of the atmosphere, averaged over a 100-year period at the end of the simulation, was less than 0.25 W m^2. These simulations, in their entirety, spanned over 9,000 model years to guarantee a complete equilibrium within the Earth system. The last 100 years were then used to calculate the climate mean to avoid any influence from interannual variability.

Developments since Valdes et al. (*95*)

Compared to the simulations presented in Valdes et al. (*95*), the simulations used in this paper have undergone a number of developments, associated with improvements to the model itself, and to the

experimental design. The most significant of these are summarised here. For a full discussion, see Ross (PhD thesis) (*47*).

The version of the climate model used in this paper has an update that includes modification to Cloud Condensation Nuclei (CCN) density and cloud droplet effective radius, following the work of Sagoo et al. (*46*) and Kiehl and Shields (*106*). This update increases higher latitude temperatures without significantly changing tropical temperatures, thereby reducing the pole-to-equator temperature gradient, and better aligning higher latitude temperatures with proxy observations. This update has been verified to be effective under a range of climate conditions, including hot, cool, and icehouse conditions, as well as under pre-industrial boundary conditions. As a result, it is suitable for application throughout the Phanerozoic. In addition, other internal model parameters were tuned, following Irvine et al. (*107*). The ocean barotropic flow solver requires all islands in the paleogeography to be manually defined, to allow non-zero flow through ocean gateways. Although Antarctica was defined as an island in the Valdes et al. (*95*) post-Eocene simulations, no other islands were defined. For the simulations in this paper, all islands were defined through the Phanerozoic.

Uncertainties

Modelling of climate niches in deep time requires global-scale datasets of key environmental variables, such as mean annual sea surface temperature Global proxy data with broad spatiotemporal coverage exist for some past time periods (*108*). However, proxy evidence becomes scarcer and increasingly less well constrained in deep time, and thus niche modelling studies cannot be conducted without the use of paleoclimate models to extract site-specific climate information.

Paleoclimate simulations of deep-time climates are uncertain. These uncertainties need to be considered when interpreting a paleoclimate model simulation which, depending on the time period, can be relatively unconstrained by proxy-observations. Uncertainties can generally be categorised into two main sources, i) boundary condition uncertainty, and ii) climate model uncertainty.

i) Boundary conditions need to be prescribed by the user and are a product of both proxy-evidence, quantitative data (e.g. paleoaltimetry studies, pCO2 reconstructions), qualitative data (e.g. ice sheet height), and numerical techniques (e.g. tectonic plate models). Boundary conditions for the model that vary on timescales of the Phanerozoic include (a) paleogeographic reconstructions, (b) ice sheet height and extent, (c) vegetation, (d) solar luminosity, (e) orbital configuration, and (f) greenhouse gas concentrations.

(a) One source of uncertainty arises from the paleogeographic reconstructions (topography, bathymetry, and land-sea distribution), which are produced as a DEM. These reconstructions are produced using plate models that make assumptions concerning spreading, collisional, deformational, and weathering rates, that become more difficult to constrain the further back in the past. However, paleogeographic topography can often be constrained using geologic data, in particular paleoaltimetry, giving a reasonable approximation of height, with associated uncertainties and assumptions (*109*).

(b) Ice sheet reconstructions are also difficult to estimate, in particular their height. The amount of land ice will have an associated impact on paleogeography through sea-level change, which in certain time intervals could be critical to whether an ocean gateway is open or closed. Whether gateways are open or closed can have a material effect on both ocean and atmospheric circulation globally. Most deep-time paleoclimate simulations will prescribe the ice sheet extent and height based on available geologic evidence. However, this can be

challenging further back in the past, where any direct evidence is not preserved in the rock record or currently inaccessible for investigation (.e.g. covered by land ice in Antarctica).

(c) Vegetation reconstructions are an additional source of uncertainty since the TRIFFID dynamic vegetation model assumes modern PFTs. However, these PFTs likely changed significantly over the Phanerozoic.

(d) Solar luminosity, the amount of incoming solar energy received at the top of Earth's atmosphere (W/m^2) is well constrained (*105*), However, there is decreasing precision through the Phanerozoic (*110*), but this is generally a second-order source of uncertainty compared to paleogeography and pCO_2 (*111*).

(e) Orbital configuration can have a large impact on the climate system over large time spans through varying the amount and location of incident solar radiation. This can, in turn, impact the seasonal and latitudinal climate signal, ice formation (land and sea), sea level change, etc. However, the simulations used in this paper employ a modern-day orbital configuration for two reasons: i) there is large uncertainty associated with the dating of geologic proxies in deep time; and ii) each simulation in any case represents a long-term average of several million years, which consists of multiple orbital cycles; the modern orbit with relatively low eccentricity therefore can be considered representative of a long-term average over multiple orbital cycles.

(f) Greenhouse gas concentrations (GHG) (e.g. pCO_2 and CH_4 concentrations) have shown large swings in concentrations throughout Earth's history (*48*, *104*). There are many assumptions and uncertainties made when converting geologic data into a GHG estimate due to the proxy type, location, age, techniques, and calibration, all of which can lead to widely varying estimates, making constraining past GHGs challenging (*104*).

ii) Climate model uncertainty can be one of the most difficult sources of uncertainty to estimate. Due to the cost and complexity of running paleoclimate models and the expertise needed, it is often prohibitive for one research group to run more than one climate model. However, understanding this source of uncertainty is fundamental, since previous studies have shown contrasting results in different models (*112*), even though different paleoclimate models essentially use the same equations of state to simulate the large-scale behaviour of the Earth system. There can be differences in how these equations are implemented in different climate models, as well as differences in complexity and spatial resolution that will have a material impact on the climate simulated between different models using the same boundary conditions.

It is outside the scope of this manuscript to fully address questions of model uncertainty, due to the computational resources required to run multiple sets of Phanerozoic simulations to full Earth system equilibrium. There is confidence in the robustness of these results because: (1) The HadCM3 family of models still shows good skill at reproducing modern climate when compared to previous IPCC Coupled CMIP fifth assessment generation models (*113*). (2) There has been continual development of HadCM3BL-M2.1D (*113*), such as the addition of the new CCN scheme. Current generation paleoclimate models still suffer from a decades-old problem, whereby simulated high-latitude temperatures are too cool compared to the proxies (known as the 'cold-pole-paradox'), in particular in the winter. Our new scheme now produces much warmer high latitude temperatures, without overheating the tropics. (3) Our climate simulations have been run in excess of 9000 model years, reaching a fully equilibrated state for each time period. (4) While constraining model

uncertainty is important, scenario uncertainty (i.e., climate model boundary conditions), is often a larger source of climate model error than model uncertainty(*114*). (5) Critically, we found that the effect of changes in boundary conditions does not impact our model results (see Section "Sensitivity analyses", below).

Statistical model framework

To model extinction risk over the Phanerozoic as a function of the five predictor variables (body size, geographic range size, absolute realized thermal preference, realized thermal niche breadth, and absolute temperature change experienced by a taxon), we used generalized linear mixed effects models (glmm) using the Glmer() function in the lme4 v.1.1-33 R package (*115*). We chose a glmm framework over other statistical models, such as capture mark recapture Pradel Seniority (*20*, *116*), to account for the hierarchical structure in paleontological data by incorporating temporal and phylogenetic variance in our models. Temporal and phylogenetic hierarchical structure was incorporated using stage and family as random effects. By accounting for the temporal grouping structure in the data, we are partially accounting for differences in boundary conditions and biases among stages, such as changes in paleogeography (*62*), changes in tectonics (*69*), changes in macroevolutionary regime (*17*, *20*, *78*), changes in biota and biotic interactions over time (*117*), changes in outcrop area (*68*), changes in lithology and environments (*118*, *119*), differences in spatial coverage (*120*), and differences in sampling intensity over the Phanerozoic (*119*). The variance explained and conditional modes estimates (fig. S2, table S1) support the inclusion of stage as a random effect.

Families were used for taxonomic grouping in our mixed-effect modelling framework as a proxy for phylogenetic relatedness. Taxa within each family are more likely to share physiological, ecological, and life history traits than more distant relatives (*79*, *121*). Thus, accounting for phylogenetic structure and selectivity differences among groups allows for evaluation of broad-scale patterns in selectivity over geologic time (*121*).

The slope was held constant and the intercept was allowed to vary for each level of the random effects. Variance inflation factors (VIF) were calculated for each of the predictors in the model to check for multicollinearity. We retained all predictors in our model, since VIF's were <2.5 (Table S2) (*122*).

Predictor variables were centered and standardized prior to running models by subtracting the mean and dividing by the standard deviation for each variable. Standardization enables comparison of the strength of model coefficients and will not affect the model selection process (*123*) (fig. 2).

We tested all possible combinations of predictor variables using the dredge() function within the MuMIn v. 1.47 R package (*124*) and compared model estimates using the Akaike Information Criterion (AIC) (*125*). We calculated AIC weights to select the best model following (*52*). AIC was used for the analysis, since the ratio of our observations and model parameters was >40 (*126*). Species-level models were assessed using AICc, since observations within our hierarchical model framework did not always meet this suggested threshold.

For the best model, residual qq plots and model diagnostics were used to assess the model residuals and whether model assumptions were met. We used the plotResiduals() function within the DHARMa v. 0.4.6 R package (*127*, *128*), which is designed specifically for hierarchical modeling frameworks, to confirm that our statistical model met the assumptions required for generalized linear mixed effects models.

To assess the ability of our model to detect an extinction selectivity signal, we performed a power analysis using the powerSim() function within the SIMR v. 1.0.7 R package (*129*). Power analyses are a critical tool for paleontologists, since they can directly test the effect size and probability of incurring a type-II error, while considering sample size and spatial biases, for a statistical model (*130*). We constructed 100 null models with a significant selectivity signal for all five predictor variables by simulating data based on the fixed-effect and random-effect model structure. We subsequently evaluated our ability to refute the null hypothesis, which posits a lack of selectivity, by determining whether a significant selectivity signal was detected (indicated by p-values less than 0.05) when such a signal actually exists. We found that geographic range size, realized thermal niche breadth, body size, and absolute temperature change had a power of 100%, whereas realized thermal preference had a power of 89%, which indicates we would reject the null-hypothesis 100/100 and 89/100 times, respectively. Thus, our model has sufficient power to detect a selectivity signal for all five of our predictors based on the commonly used threshold of 80% (*131*).

The individual effects of each predictor variable on extinction risk were evaluated by examining the marginal effects whilst holding all other predictors constant. We calculated marginal effects specifically for a generalized mixed modeling framework by accounting for the variance in the random effects using the ggeffects() function within the ggeffects v. 1.2.2 R package (*132*). The probability of extinction, whilst holding all other predictors constant, was centered and standardized by subtracting the mean and dividing by the standard deviation, similar to previous studies (*79*) (fig. 2). This procedure allows for comparison of trends in marginal effects despite differences in the scaling of the predictors, while still allowing for the interpretation of risk associated with certain values.

Sensitivity analyses

Sample size is a difficult factor to account for when studying extinction, since a small number of occurrences for a taxon could result from preservational or taxonomic biases, or could reflect true population size or rarity prior to extinction (*18*). Instead of removing taxa with a small number of occurrences, we controlled for sample size using our bootstrap subsampling method. We ran our analyses at three, four, and five occurrences per stage x taxon combination. That is, all calculations of the predictor values for each unique taxon by stage combination were made with only three (or four or five) occurrences per taxon in each stage. We found that the relative importance and magnitude of coefficients did not substantially vary between sampling thresholds, with variation of up to 0.02 log-odds between 3 and 5 samples (fig. S4).

Taxa that occur in one stage (or 'singletons') have been posited to result from potential sampling and taxonomic biases (*133*). However, these singleton taxa could also reflect true rarity (*18*) and the exclusion of these taxa could result in the loss of valuable information on the selectivity of short-lived taxa (*18*, *36*, *134*). Thus, we followed the approaches of previous studies (*18*, *77*, *78*) and ran our models both with and without singletons to test the sensitivity of our results to their inclusion. We found that patterns in selectivity were largely the same when including and excluding singletons in our models (table S3, S4).

Taxonomic rank is also important to consider when quantifying extinction risk over the Phanerozoic. Extinction rates and selectivity are often evaluated at the genus level due to the poor temporal constraint on species ranges (*51*, *79*). However, a genus is an imperfect proxy for species-level dynamics on geologic timescales (*50*). Thus, we ran our analyses at both the species- and genus-level to assess whether results remained congruent. Results of species-level analyses (table S6) were mostly consistent with genus-level results, with all five predictors included in the best model (table

S3). However, the selectivity of geographic range was muted relative to the genus-level analyses. The muted signal between the species- and genus-level results is likely due to differences in sample size, reduced number of spatially-unique temporal occurrences per species, and the taxonomic and temporal uncertainty when studying species over the Phanerozoic (*13*).

Another potential source of uncertainty when studying extinction on macroevolutionary timescales is data quality and reliability. Certain fossil groups may be more well-studied and reliable than other groups, and this reliability could vary over the Phanerozoic. To test the impact of data quality on our results, we re-ran our analyses when excluding ostracods, since ostracods are often regarded as an underrepresented group in the PBDB, with uneven spatial and temporal data contributions (*135*). Ostracods are also significantly smaller in body size than the other classes included in our study, which could potentially bias our extinction selectivity estimates for body size (*19*). However, we found no difference in our results when including (table S3) or excluding (table S7) ostracods, despite a slight increase in body size selectivity from -0.29 to -0.31.

The glmm framework used to study extinction risk did not include interaction terms. However, although the predictors were not highly collinear (table S2), the importance of one variable in predicting extinction outcomes may increase or decrease depending on the level of another term. For example, traits such as body size and geographic range (*40*, *61*), and thermal niche and geographic range (*63*), are not entirely independent. To test the impact of interactions on our extinction risk models over the Phanerozoic, we repeated the model selection procedure including all possible two-way and three-way interactions between variables (table. S7).

Our results are heavily reliant on the climate model data. The climate model used in this analysis are state-of-the-art for long-term paleoclimate simulations, and represent our current best estimates of climate change over the Phanerozoic. The boundary conditions and methodology used to model physical processes have an impact on the reliability of the climate model simulations. We tested the climate model's sensitivity to pCO_2 and the physical processes by repeating our analyses using climate model simulations with fixed concentrations of atmospheric CO_2 (560 ppm and 1,120 ppm) and using the original Valdes et al climate simulation (see Paleoclimate simulations). We also tested our climate model's sensitivity using the new model but without the correct definition of islands (see Paleoclimate simulations). We found that our results were largely consistent regardless of the climate model used (fig. S5), although the strength of the climate change coefficient was more variable than the coefficients of the other predictors, since it relied on two climate models for its estimation (fig. S5). The strength of the climate change predictor also improved with the quality of climate model, suggesting we may be better able to approximate climate change between stages using this 'best' climate model simulation.

All analyses detailed in the main text were performed including singletons with jackknife subsampling and were run at the genus-level, with parameter estimates that relied on the 'best' climate model that included island paleogeography. All subsequent sensitivity tests were performed by varying one of these variables for direct comparison with the main text model. The sample size sensitivity test was performed with the exclusion of singletons (fig. S4), since we aimed to test the impact of sample size on our analyses without the confounding effect of single-stage taxon occurrences.

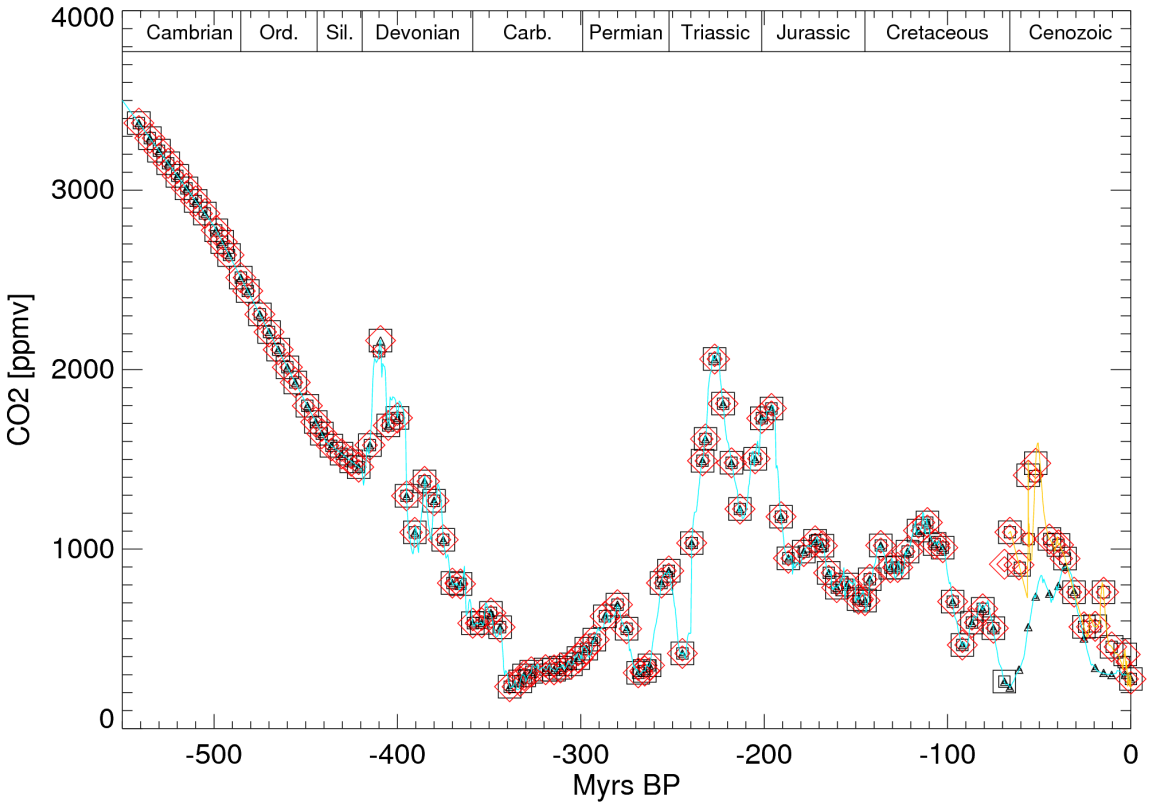


Fig. S1. Phanerozoic pCO_2 estimates used to estimate boundary conditions for our best climate model. The blue pCO_2 line is reproduced from Foster et al. (*104*) and the yellow pCO_2 line is from Rae et al. (*48*). Red box/triangle show the realistic target pCO_2 for each snapshot simulation over a 9000-year simulation period.

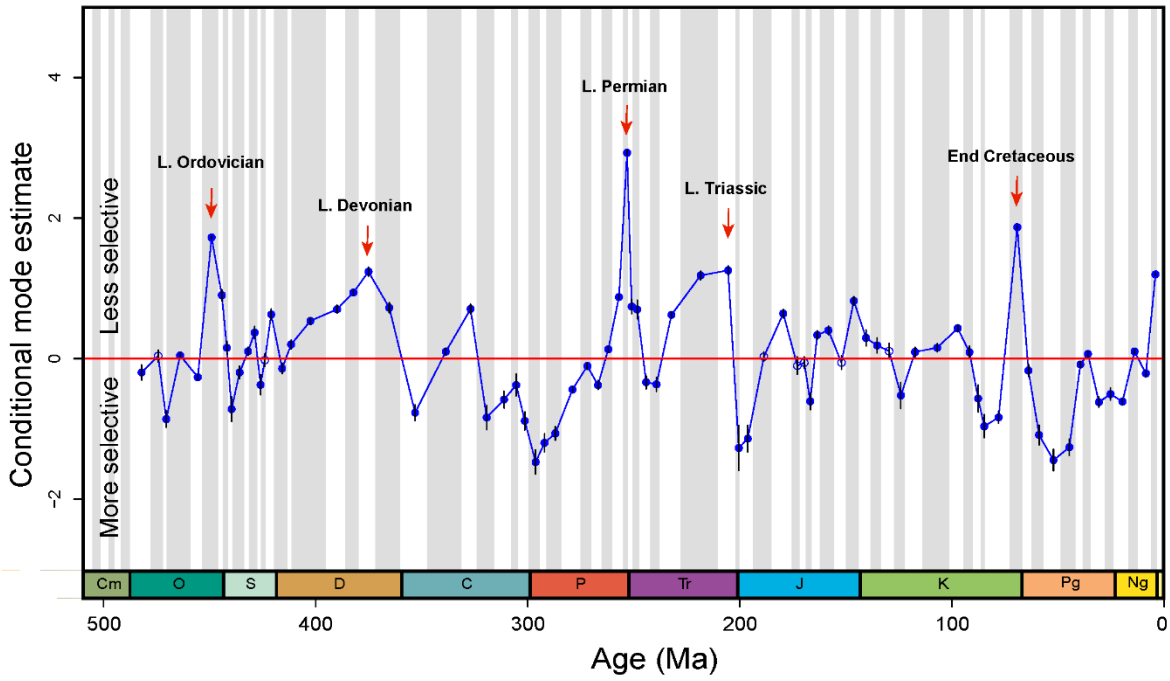


Fig. S2. Conditional mode estimates over the Phanerozoic. Estimates of conditional modes for each level of our temporal random effect, where each point represents the deviation of the intercept measured at each level from the intercept for the best model, which was -1.6. Positive conditional modes represent a weaker relationship between the predictor and response for that level. Model results are presented for genus-level data that included singletons; parameter estimates relied on jackknife subsampling and used the 'best' climate model with islands included in paleogeographic estimates.

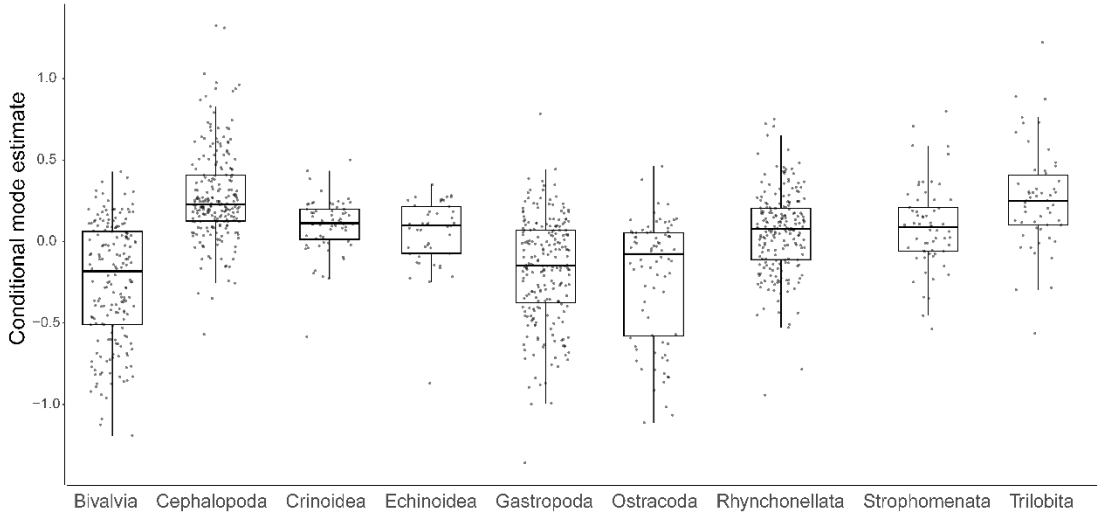


Fig. S3. Conditional mode estimates for each taxonomic group. Estimates of conditional modes for each family within 9 major taxonomic groups showing the distribution of family-level estimates for each group. Each point represents the deviation of the intercept measured at each level from the intercept for the best model, which was -1.6. Positive conditional modes represent a weaker relationship between the predictor and response for that level. The line represents the median estimate for each taxonomic grouping and the whiskers represent the 95th percentile estimates. Model results are presented for genus-level data that included singletons; parameter estimates relied on jackknife subsampling and used the 'best' climate model with islands included in paleogeographic estimates.

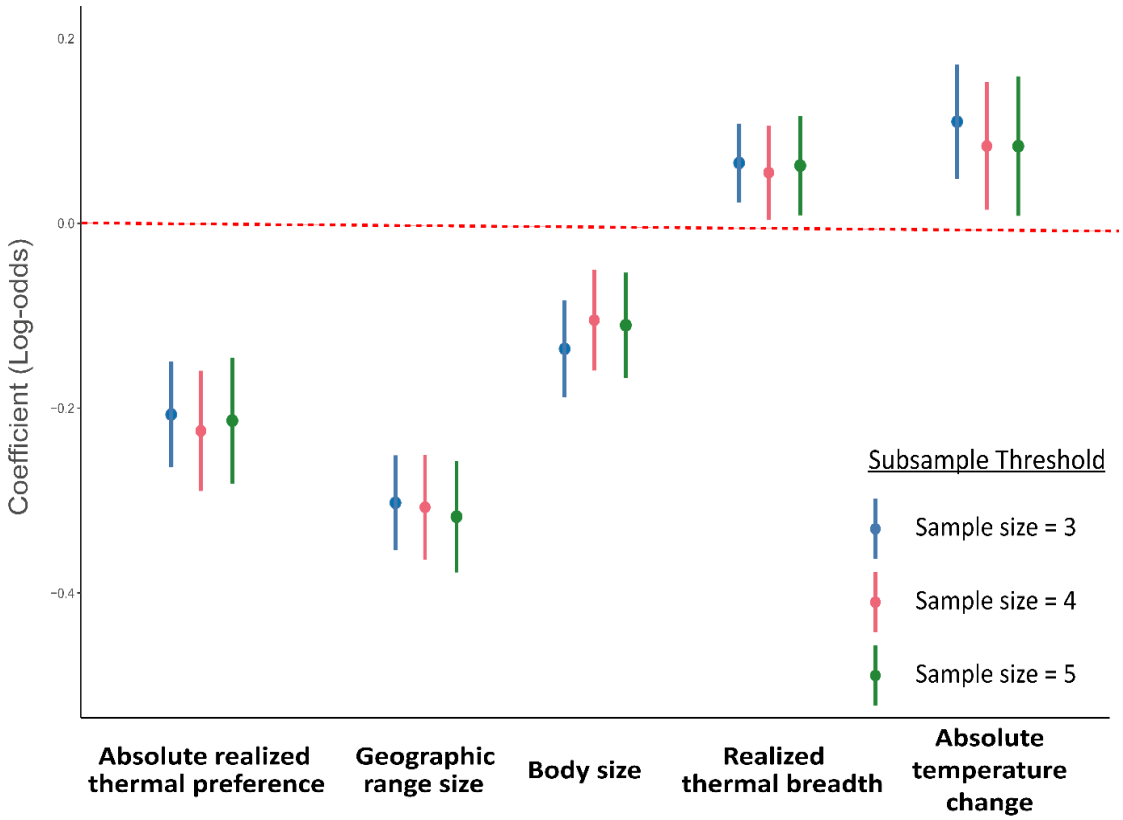


Fig. S4. Model sensitivity test for different sampling thresholds. Model coefficient estimates for the most saturated model while controlling for sample size using bootstrapped replicates at three, four, and five sample size cut-offs. Singletons were excluded from analysis. The points represent the coefficient estimates, and the lines represent the 95% confidence intervals. Parameter estimates relied on the 'best' climate model with islands included in paleogeographic estimates.

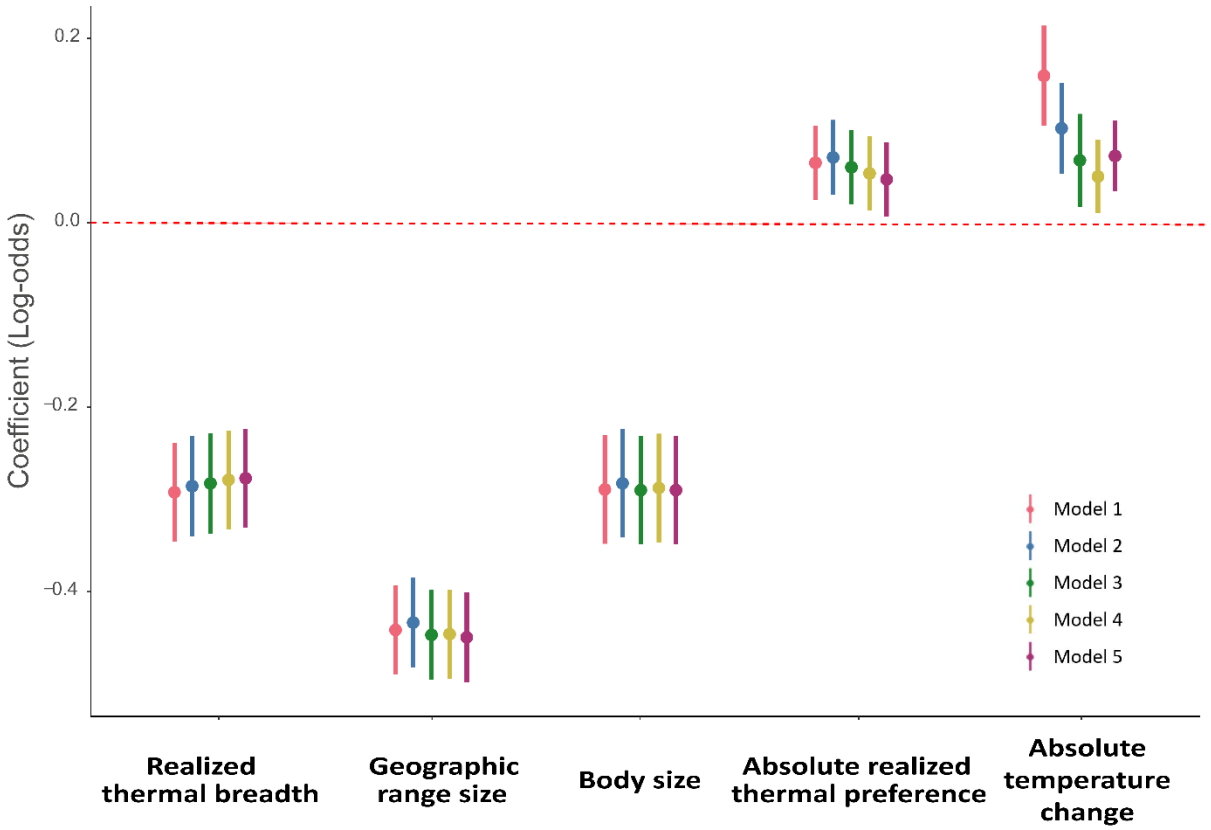


Fig. S5. Model sensitivity test for different climate models. Model coefficient estimates for the most saturated model using different AOGCM simulations. Model 1 is the best climate model used in the main text analyses, which incorporates the new physical parameters, realistic CO_2 estimates, and includes island geography. Model two had the same parameters as model one, except it does not incorporate island geography. Models three, four, and five were run using the old physical parameters, with realistic pCO_2, at 560 ppm pCO_2 and 1,120 ppm pCO_2, respectively. The points represent the model coefficient estimates and the lines indicate the 95% confidence intervals. Model results are presented for genus-level data that included singletons; parameter estimates relied on jackknife subsampling and used the 'best' climate model with islands included in paleogeographic estimates.

Random Effect	Variance	Levels	Standard deviation
(1|Family)	0.3701	1,138	0.6084
(1|Stage)	0.8714	81	0.9335

Table S1 Variance, standard deviation and sample size for each random effect. Variance and standard deviation in the random effects for stage and family for the best model, where the random effects column shows the structure used in the model formula, and levels shows the number of unique levels to the random effect structure. Model results are presented for genus-level data that included singletons; parameter estimates relied on jackknife subsampling and used the 'best' climate model with islands included in paleogeographic estimates.

Term	VIF	VIF_CI_low	VIF_CI_high	Tolerance	Tolerance_CI_low	Tolerance_CI_high
Geographic range size	1.62	1.59	1.66	0.61	0.60	0.62
Body size	1.02	1.01	1.04	0.98	0.96	0.99
Realized thermal preference	1.02	1.01	1.04	0.97	0.95	0.98
Realized thermal breadth	1.60	1.57	1.63	0.62	0.61	0.64
Absolute temperature change	1.00	1.00	1.04	0.99	0.95	0.99

Table S2. Variance inflation factor and tolerance estimates for the best model. VIF and tolerance were calculated for each variable in our mixed-effects model. Predictors showed little collinearity between variables (VIF<2.5; Tolerance > 0.2)**.** Estimates and confidence intervals were estimated using the check_collinearity() function within the Performance v. 0.010 package within R (*136*). Results are presented for genus-level data that included singletons; parameter estimates relied on jackknife subsampling.

Model rank	Geographic range size	Absolute realized thermal preference	Body size	Absolute temperature change	Realized thermal breadth	Df	AIC	Delta	Weight
1	-0.44	0.07	-0.29	0.14	-0.29	8	21773.65	0.00	0.98
2	-0.45	NA	-0.29	0.15	-0.29	7	21781.55	7.90	0.02
3	-0.45	0.07	-0.29	NA	-0.29	7	21797.38	23.73	0.00
4	-0.46	NA	-0.29	NA	-0.29	6	21807.54	33.89	0.00
5	-0.45	0.06	NA	0.15	-0.31	7	21868.75	95.09	0.00
6	-0.46	NA	NA	0.15	-0.30	6	21874.22	100.57	0.00
7	-0.60	0.06	-0.31	0.13	NA	7	21887.69	114.04	0.00
8	-0.61	NA	-0.30	0.14	NA	6	21894.08	120.43	0.00
9	-0.46	0.06	NA	NA	-0.30	6	21894.19	120.53	0.00
10	-0.47	NA	NA	NA	-0.30	5	21901.71	128.06	0.00
11	-0.61	0.07	-0.31	NA	NA	6	21908.88	135.23	0.00
12	-0.61	NA	-0.31	NA	NA	5	21917.32	143.66	0.00
13	-0.62	0.05	NA	0.14	NA	6	21997.15	223.50	0.00
14	-0.63	NA	NA	0.14	NA	5	22000.99	227.34	0.00
15	-0.63	0.06	NA	NA	NA	5	22019.90	246.25	0.00
16	-0.64	NA	NA	NA	NA	4	22025.53	251.88	0.00
17	NA	0.11	-0.32	0.17	-0.59	7	22092.16	318.50	0.00
18	NA	NA	-0.32	0.18	-0.60	6	22117.78	344.13	0.00
19	NA	0.12	-0.32	NA	-0.59	6	22125.90	352.24	0.00
20	NA	NA	-0.32	NA	-0.60	5	22156.32	382.67	0.00
21	NA	0.10	NA	0.17	-0.62	6	22213.71	440.06	0.00
22	NA	NA	NA	0.18	-0.62	5	22235.50	461.84	0.00
23	NA	0.11	NA	NA	-0.62	5	22250.25	476.60	0.00
24	NA	NA	NA	NA	-0.63	4	22276.67	503.02	0.00
25	NA	0.14	-0.43	0.17	NA	6	22851.21	1077.56	0.00
26	NA	0.15	-0.43	NA	NA	5	22887.23	1113.58	0.00
27	NA	NA	-0.42	0.18	NA	5	22901.41	1127.76	0.00
28	NA	NA	-0.42	NA	NA	4	22944.64	1170.99	0.00
29	NA	0.13	NA	0.17	NA	5	23082.42	1308.77	0.00
30	NA	0.14	NA	NA	NA	4	23122.59	1348.94	0.00
31	NA	NA	NA	0.19	NA	4	23126.42	1352.77	0.00
32	NA	NA	NA	NA	NA	3	23173.79	1400.14	0.00

Table S3. Model selection results for the jackknife subsampling approach. All possible model combinations of our five predictor variables are ranked based on AIC weight. Model coefficient estimates in log-odds are listed under each predictor variable, and all listed predictors had standard error estimates that were exclusive of zero.

Model rank	Geographic range size	Absolute realized thermal preference	Body size	Absolute temperature change	Realized thermal breadth	Df	AIC	Delta	Weight
1	-0.33	0.05	-0.12	0.12	-0.26	8.00	17035.36	0.00	0.80
2	-0.34	NA	-0.12	0.12	-0.26	7.00	17038.13	2.77	0.20
3	-0.33	0.05	-0.12	NA	-0.26	7.00	17046.50	11.13	0.00
4	-0.34	NA	-0.12	NA	-0.25	6.00	17050.20	14.84	0.00
5	-0.34	0.04	NA	0.12	-0.27	7.00	17054.67	19.31	0.00
6	-0.34	NA	NA	0.12	-0.27	6.00	17056.40	21.04	0.00
7	-0.34	0.05	NA	NA	-0.27	6.00	17066.15	30.79	0.00
8	-0.35	NA	NA	NA	-0.26	5.00	17068.72	33.36	0.00
9	-0.47	0.04	-0.14	0.11	NA	7.00	17107.94	72.58	0.00
10	-0.48	NA	-0.14	0.11	NA	6.00	17109.33	73.97	0.00
11	-0.48	0.05	-0.14	NA	NA	6.00	17117.18	81.81	0.00
12	-0.48	NA	-0.14	NA	NA	5.00	17119.33	83.97	0.00
13	-0.49	0.03	NA	0.11	NA	6.00	17134.22	98.85	0.00
14	-0.49	NA	NA	0.11	NA	5.00	17134.53	99.17	0.00
15	-0.49	0.04	NA	NA	NA	5.00	17143.74	108.37	0.00
16	-0.50	NA	NA	NA	NA	4.00	17144.71	109.35	0.00
17	NA	0.08	-0.14	0.14	-0.49	7.00	17179.05	143.69	0.00
18	NA	NA	-0.14	0.15	-0.49	6.00	17191.45	156.09	0.00
19	NA	0.09	-0.15	NA	-0.49	6.00	17196.57	161.21	0.00
20	NA	0.08	NA	0.14	-0.51	6.00	17206.85	171.49	0.00
21	NA	NA	-0.14	NA	-0.50	5.00	17211.25	175.88	0.00
22	NA	NA	NA	0.15	-0.51	5.00	17217.39	182.02	0.00
23	NA	0.08	NA	NA	-0.51	5.00	17225.07	189.71	0.00
24	NA	NA	NA	NA	-0.51	4.00	17237.78	202.41	0.00
25	NA	0.11	-0.24	0.14	NA	6.00	17605.08	569.72	0.00
26	NA	0.12	-0.24	NA	NA	5.00	17623.69	588.32	0.00
27	NA	NA	-0.23	0.15	NA	5.00	17629.30	593.94	0.00
28	NA	NA	-0.23	NA	NA	4.00	17651.28	615.91	0.00
29	NA	0.10	NA	0.15	NA	5.00	17684.20	648.84	0.00
30	NA	0.11	NA	NA	NA	4.00	17704.15	668.79	0.00
31	NA	NA	NA	0.16	NA	4.00	17704.97	669.61	0.00
32	NA	NA	NA	NA	NA	3.00	17728.19	692.83	0.00

Table S4. Model selection results for the jackknife subsampling approach without singletons. All possible model combinations of our five predictor variables ranked based on AIC weight. Model coefficient estimates in log-odds are listed under each predictor variable, and all listed predictors had standard error estimates that were exclusive of zero.

Model rank	Geographic range size	Absolute realized thermal preference	Body size	Absolute temperature change	Realized thermal breadth	Df	AIC	Delta	Weight
1	-0.30	0.07	-0.14	0.11	-0.21	8.00	17186.25	0.00	0.96
2	-0.31	NA	-0.13	0.11	-0.20	7.00	17192.96	6.72	0.03
3	-0.31	0.07	-0.14	NA	-0.20	7.00	17195.87	9.63	0.01
4	-0.32	NA	-0.13	NA	-0.19	6.00	17203.56	17.31	0.00
5	-0.31	0.06	NA	0.11	-0.21	7.00	17210.59	24.35	0.00
6	-0.32	NA	NA	0.12	-0.21	6.00	17215.87	29.62	0.00
7	-0.32	0.06	NA	NA	-0.21	6.00	17220.48	34.24	0.00
8	-0.33	NA	NA	NA	-0.20	5.00	17226.67	40.42	0.00
9	-0.42	0.05	-0.15	0.09	NA	7.00	17235.00	48.75	0.00
10	-0.42	NA	-0.14	0.10	NA	6.00	17238.35	52.11	0.00
11	-0.42	0.05	-0.15	NA	NA	6.00	17241.25	55.00	0.00
12	-0.42	NA	-0.14	NA	NA	5.00	17245.34	59.09	0.00
13	-0.44	0.04	NA	0.09	NA	6.00	17263.79	77.55	0.00
14	-0.44	NA	NA	0.10	NA	5.00	17265.80	79.55	0.00
15	-0.44	0.05	NA	NA	NA	5.00	17270.16	83.91	0.00
16	-0.44	NA	NA	NA	NA	4.00	17272.81	86.57	0.00
17	NA	0.09	-0.16	0.13	-0.41	7.00	17315.08	128.83	0.00
18	NA	0.10	-0.16	NA	-0.41	6.00	17330.18	143.93	0.00
19	NA	NA	-0.16	0.14	-0.41	6.00	17331.06	144.81	0.00
20	NA	NA	-0.16	NA	-0.40	5.00	17348.19	161.94	0.00
21	NA	0.09	NA	0.13	-0.43	6.00	17351.93	165.68	0.00
22	NA	NA	NA	0.14	-0.43	5.00	17365.83	179.59	0.00
23	NA	0.09	NA	NA	-0.43	5.00	17367.73	181.48	0.00
24	NA	NA	NA	NA	-0.42	4.00	17383.60	197.36	0.00
25	NA	0.08	-0.24	0.10	NA	6.00	17629.83	443.59	0.00
26	NA	0.09	-0.24	NA	NA	5.00	17637.60	451.35	0.00
27	NA	NA	-0.23	0.11	NA	5.00	17642.14	455.89	0.00
28	NA	NA	-0.23	NA	NA	4.00	17651.28	465.03	0.00
29	NA	0.07	NA	0.10	NA	5.00	17709.60	523.36	0.00
30	NA	0.08	NA	NA	NA	4.00	17717.64	531.40	0.00
31	NA	NA	NA	0.11	NA	4.00	17718.90	532.65	0.00
32	NA	NA	NA	NA	NA	3.00	17728.19	541.95	0.00

Table S5. Model selection results for the bootstrap subsampling approach, which excluded singletons. All possible model combinations of our five predictor variables are ranked based on AIC weight. Model coefficient estimates in log-odds are listed under each predictor variable, and all listed predictors had standard error estimates that were exclusive of zero.

Model rank	Geographic range size	Absolute realized thermal preference	Absolute temperature change	Realized thermal breadth	Df	AICc	Delta	Weight
1	-0.08	0.12	0.16	-0.28	7	18765.37	0.00	0.99
2	NA	0.12	0.16	-0.29	6	18774.46	9.09	0.01
3	-0.08	0.13	NA	-0.27	6	18789.48	24.11	0.00
4	-0.08	NA	0.17	-0.28	6	18791.78	26.41	0.00
5	NA	0.13	NA	-0.28	5	18798.71	33.34	0.00
6	NA	NA	0.17	-0.29	5	18800.18	34.81	0.00
7	-0.08	NA	NA	-0.27	5	18819.12	53.75	0.00
8	NA	NA	NA	-0.28	4	18827.60	62.23	0.00
9	-0.14	0.13	0.14	NA	6	19000.66	235.29	0.00
10	-0.14	0.13	NA	NA	5	19018.22	252.86	0.00
11	-0.13	NA	0.14	NA	5	19032.13	266.76	0.00
12	NA	0.13	0.14	NA	5	19033.26	267.89	0.00
13	NA	0.13	NA	NA	4	19050.66	285.29	0.00
14	-0.13	NA	NA	NA	4	19052.30	286.94	0.00
15	NA	NA	0.14	NA	4	19063.49	298.12	0.00
16	NA	NA	NA	NA	3	19083.40	318.03	0.00

Table S6. Model selection results for the jackknife subsampling approach at the species level. All possible model combinations of our five predictor variables are ranked based on AICc weight. Model coefficient estimates in log-odds are listed under each predictor variable, and all listed predictors had standard error estimates that were exclusive of zero. Singletons were included in parameter estimation and relied on the 'best' climate model that included islands in paleogeographic estimates.

Model rank	Geographic range size	Absolute realized thermal preference	Body size	Absolute temperature change	Realized thermal breadth	Df	AIC	Delta	Weight
1	-0.44	0.08	-0.32	0.14	-0.29	8	21528.77	0.00	1.00
2	-0.45	NA	-0.31	0.15	-0.29	7	21541.15	12.38	0.00
3	-0.45	0.09	-0.32	NA	-0.29	7	21552.71	23.94	0.00
4	-0.46	NA	-0.31	NA	-0.29	6	21567.85	39.08	0.00
5	-0.60	0.07	-0.33	0.14	NA	7	21639.76	110.99	0.00
6	-0.61	NA	-0.33	0.14	NA	6	21650.18	121.41	0.00
7	-0.61	0.08	-0.34	NA	NA	6	21661.30	132.52	0.00
8	-0.46	0.07	NA	0.14	-0.31	7	21668.12	139.34	0.00
9	-0.62	NA	-0.33	NA	NA	5	21674.22	145.44	0.00
10	-0.47	NA	NA	0.15	-0.31	6	21676.53	147.76	0.00
11	-0.46	0.07	NA	NA	-0.30	6	21692.88	164.11	0.00
12	-0.47	NA	NA	NA	-0.30	5	21703.67	174.90	0.00
13	-0.63	0.06	NA	0.14	NA	6	21795.09	266.32	0.00
14	-0.64	NA	NA	0.14	NA	5	21801.45	272.68	0.00
15	-0.63	0.07	NA	NA	NA	5	21817.21	288.44	0.00
16	-0.64	NA	NA	NA	NA	4	21825.65	296.88	0.00
17	NA	0.12	-0.34	0.17	-0.60	7	21842.91	314.14	0.00
18	NA	NA	-0.34	0.18	-0.60	6	21875.52	346.75	0.00
19	NA	0.13	-0.35	NA	-0.60	6	21877.02	348.25	0.00
20	NA	NA	-0.34	NA	-0.61	5	21915.09	386.32	0.00
21	NA	0.11	NA	0.17	-0.63	6	22010.95	482.18	0.00
22	NA	NA	NA	0.18	-0.63	5	22037.66	508.88	0.00
23	NA	0.12	NA	NA	-0.63	5	22046.77	518.00	0.00
24	NA	NA	NA	NA	-0.63	4	22078.53	549.76	0.00
25	NA	0.16	-0.44	0.17	NA	6	22614.05	1085.28	0.00
26	NA	0.16	-0.44	NA	NA	5	22651.68	1122.91	0.00
27	NA	NA	-0.43	0.19	NA	5	22673.28	1144.51	0.00
28	NA	NA	-0.43	NA	NA	4	22718.92	1190.15	0.00
29	NA	0.14	NA	0.18	NA	5	22900.41	1371.63	0.00
30	NA	0.15	NA	NA	NA	4	22940.54	1411.76	0.00
31	NA	NA	NA	0.19	NA	4	22951.12	1422.35	0.00
32	NA	NA	NA	NA	NA	3	22998.90	1470.13	0.00

Table S7. Model selection results with the removal of ostracods. All possible model combinations of our five predictor variables are ranked based on AIC weight. Model coefficient estimates in log-odds are listed under each predictor variable, and all listed predictors had standard error estimates that were exclusive of zero. Singletons were included in parameter estimation and relied on the 'best' climate model that included islands in paleogeographic estimates.

Model predictor	Coefficient	Standard error	P-value
Grs	-0.49	0.03	0.000***
Bsz	-0.28	0.04	0.000***
Artp	0.11	0.03	0.000***
Rtb	-0.25	0.04	0.000***
Atc	0.10	0.03	0.002***
Grs:Bsz	0.10	0.03	0.003***
Grs:Artp	0.08	0.03	0.005**
Grs:Rtb	-0.08	0.03	0.007**
Bsz:Artp	-0.01	0.03	0.785
Bsz:Rtb	-0.04	0.04	0.309
Artp:Rtb	0.01	0.03	0.702
Grs:Atc	-0.01	0.03	0.786
Bsz:Atc	0.03	0.03	0.265
Artp:Atc	0.01	0.02	0.530
Rtb:Atc	-0.05	0.03	0.116
Grs:Bsz:Artp	-0.04	0.04	0.312
Grs:Bsz:Rtb	0.03	0.03	0.366
Grs:Artp:Rtb	-0.01	0.03	0.701
Bsz:Artp:Rtb	-0.03	0.04	0.394
Grs:Bsz:Atc	-0.01	0.03	0.855
Grs:Artp:Atc	0.03	0.02	0.133
Bsz:Artp:Atc	-0.01	0.02	0.804
Grs:Rtb:Atc	0.04	0.03	0.147
Bsz:Rtb:Atc	0.01	0.04	0.872
Artp:Rtb:Atc	0.02	0.03	0.451
Grs:Bsz:Artp:Rtb	0.02	0.03	0.533
Grs:Bsz:Artp:Atc	0.03	0.03	0.334
Grs:Bsz:Rtb:Atc	-0.04	0.03	0.169
Grs:Artp:Rtb:Atc	0.00	0.02	0.875
Bsz:Artp:Rtb:Atc	0.00	0.03	0.940
Grs:Bsz:Artp:Rtb:Atc	0.02	0.03	0.423

Table S8. Model coefficient table for all interaction terms among predictor variables. All possible model combinations of our five predictor variables and 27 interaction terms. Model coefficient estimates in log-odds, standard error, and p-values are listed for each variable. Levels of statistical significance are marked as follows: * $\alpha<0.05$, ** $\alpha<0.01$, and *** $\alpha<0.001$. Interaction terms are indicated by colons: "Grs" stands for Geographic range size, "Bsz" for Body size, "Artp" for Absolute realized thermal preference, "Atc" for Absolute temperature change, and "Rtb" for Realized thermal breadth. The best model includes Grs, Bsz, Artp, Atc, Grs:Bsz, Grs:Artp, Grs:Rtb based on a 0.96 AIC weight, after testing all possible model combinations.

Data S1. Dryad data repository containing all data for the main analyses in this paper.

All data and code to run the main text analyses are included on the Dryad repository at https://doi.org/10.5061/dryad.1ns1rn91g.