Article:
Fenton, IS, Aze, T orcid.org/0000-0002-4201-7105, Farnsworth, A et al. (2 more authors) (2023) Origination of the modern-style diversity gradient 15 million years ago. Nature, 614 (7949). pp. 708-712. ISSN 0028-0836

https://doi.org/10.1038/s41586-023-05712-6

Origination of the modern-style diversity gradient 15 million years ago

Isabel Fenton[1], Tracy Aze[2], Alexander Farnsworth[3], Paul Valdes[3], Erin E. Saupe*[1]

[1]Department of Earth Sciences, University of Oxford; 3 S Parks Road, Oxford, OX1 3AN, UK

[2]School of Earth and Environment, University of Leeds; Leeds, LS2 9JT, UK

[3]School of Geographical Sciences, University of Bristol; University Road Bristol BS8 1SS, UK

*Corresponding author. Email: erin.saupe@earth.ox.ac.uk

Abstract

The latitudinal diversity gradient (LDG) is a prevalent feature of modern ecosystems across diverse clades[1-4]. Recognized for well over a century, the causal mechanisms for LDGs remain disputed, in part because numerous putative drivers simultaneously covary with latitude[1,3,5]. The past provides the opportunity to disentangle LDG mechanisms, because the relationships among biodiversity, latitude, and possible causal factors have varied over time[6-9]. We quantify the emergence of the LDG in planktonic foraminifera at high spatio-temporal resolution over the last 40 million years, finding a modern-style gradient arose only 15 million years ago. Spatial and temporal models suggest LDGs for planktonic foraminifera may be controlled by the physical structure of the water column. Steepening of the latitudinal temperature gradient over the last 15 Ma, associated with increased vertical temperature structure at low latitudes, may have enhanced niche partitioning and provided more opportunities for speciation equatorially. Supporting this hypothesis, we find higher rates of low latitude speciation steepened the diversity gradient, consistent with spatio-temporal patterns of depth partitioning by planktonic foraminifera. Extirpation of species from high latitudes also strengthened the LDG, but this effect tended to be weaker compared to speciation. Our results provide a step change in understanding the evolution of marine LDGs on long time scales.

The spatial structure of Earth's biodiversity has the potential to provide important insight on evolutionary drivers. Today, species richness peaks at low latitudes in both marine[10-12] and terrestrial[1,4] systems across diverse taxonomic groups, referred to as the 'latitudinal diversity gradient' (LDG). Although first documented over 200 years ago, the causal mechanisms responsible for elevated low-latitude richness are still disputed[1,3,4,13].

LDGs derive from differential rates of speciation, extinction, local extirpation, and dispersal[14,15], which themselves may be controlled by the dynamics of climate[14,16-19], biotic interactions[20-23], energy/primary production[24,25], or available surface area[26]. Identifying which of these factors are key controls on rates of speciation, extinction, extirpation, and dispersal is difficult, because most are collinear with each other today[5]. However, the relationship between latitude and these hypothesized drivers has not been constant over Earth history, and thus intervals of the geological past can provide insight on how biodiversity is generated and maintained[6,7,9].

In marine systems, previous work has suggested that diversity gradients were present for tens of millions of years but varied in strength and shape[4,6-9]. However, limited fossil data has prevented detailed examination of the emergence of modern-day LDGs for more than short temporal intervals [27-29] or coarse spatio-temporal resolutions[7,9,30], leaving key knowledge gaps[31].

Here we utilize our recent compilation of planktonic foraminifera[32], a group of biomineralizing marine plankton, to study the establishment and maintenance of the modern-style LDG at a previously unachieved spatio-temporal resolution. By examining spatial diversity patterns across 40 million years, we provide fundamental insight on the co-evolution of the biosphere and geosphere and test key hypotheses on LDG drivers[12,33,34].

We quantified temporal patterns in LDGs using 464,963 unique species-by-locality-by-time records (Fig. 1A; Fig. S1-4). Records older than 40 Ma were excluded due to poor-quality low-latitude data for earlier time intervals. LDGs were constructed using five subsampling approaches that accounted for sampling biases (SI Methods, Fig. S5), all of which reveal significant changes in the spatial distribution of species over the last 40 Ma (Fig. 1B,C; Figs. S6-10). A modern-style LDG began to emerge gradually beginning ~34 million years ago, coincident with the transition from warmhouse to coolhouse conditions, but remained shallow until around ~15–10 million years ago, contemporaneous with an increase in global cooling[35]. Gradient (i.e., slope) estimates for richness steepened from virtually no gradient at 40 Ma (Fig. 1C) and were insensitive to methodological choice (Figs. S11-24, Table S1). This pattern supports previous suggestions that shallower diversity gradients occur during warmer, greenhouse intervals[7,8,27].

Modern-day diversity for planktonic foraminifera is richest at mid-latitudes, with a slight depression at the equator[16,28]. We fit linear and 2nd order polynomial models to each of our 16 LDGs to test whether they are better characterized as unimodal or bimodal; data for the Northern and Southern Hemispheres were modelled together and individually (see SI 'Estimating LDG Gradients'). For all time periods other than the most recent (0–2.5 Ma), a linear model produced a better fit (Table S2). Peak richness for planktonic foraminifera occurred at higher latitudes from 40–20 Ma, but with generally flatter gradients, after which peak richness shifted to ~10° to 20° latitude, consistent with the diversity pattern observed today[16,28] (Fig. 1B).

To identify potential drivers of LDG changes over the last 40 Ma, we investigated the relationship between richness and climate both spatially within time bins and temporally across time bins. In the first approach, we modelled richness as a function of mean annual sea

surface salinity, mean annual mixed layer depth, mean annual thermocline extent, and mean annual sea surface temperature (SST) using spatial autoregressive models within each of the 2.5 Ma time bins (Fig. S25-27). After correcting for multiple comparisons, only SST exhibited a consistent and strong positive relationship with richness over time (Fig. 2). This relationship persisted for at least the last 15 million years, and perhaps longer, but with confidence intervals that overlap zero. Results are insensitive to permutations of the data (Figs. S28-32). The temperature range of the thermocline is highly correlated with SST (Fig. S26) and therefore could not be included in the multivariate model. When thermocline temperature range was modelled separately, however, the strength of relationship was similar to that of SST (Fig. S33). No other variables exhibited such a relationship (Fig. S33). Richness was not linked to surface ocean area when species richness was modelled as a function of area within 15° latitudinal bins (Fig. S34).

In the second model approach, we investigated the relationship between change in richness and change in climate variables at given locations on Earth. Change was examined over 2.5, 5, 7.5., 10, and 12.5 million-year intervals of time (Fig. S35 & S36). SST change was the only significant predictor of change in richness across these five models, and results were robust to permutations of the data (Table S3). No relationship was found when thermocline temperature range was modelled separately (Table S4), but this could reflect the greater uncertainty in estimating temperatures at depth with paleoclimate models, especially in deep time.

These results and previous work[16,36,37] suggest species richness for planktonic foraminifera could be explained, at least in part, by steepening of the latitudinal temperature gradient and associated increase in vertical temperature structure at low latitudes[35,38] over the last 15 Ma, the latter of which may have enhanced niche-partitioning-mediated speciation in the tropics[39,40]. To further test this hypothesis, we examined the degree to which species of planktonic foraminifera partition by depth within the water column over time and space. We found that low-latitude assemblages of species today are more evenly distributed vertically within the water column across the mixed layer, thermocline, and sub-thermocline than are assemblages at high latitudes (Fig. 3). However, assemblages exhibited greater evenness of depth habitats (mixed layer, thermocline, and sub-thermocline) across latitudes when the gradient was shallower millions of years ago. This pattern implies that warmer waters at high latitudes supported a broader range of vertical temperature habitats within the water column from 40–15 Ma, and that these assemblages collapsed as the high latitudes cooled.

If changes to the vertical structure of the water column facilitated the formation of a modern-style LDG, we would expect higher rates of speciation at low latitudes coincident with the steepening of the diversity gradient, and higher rates of either extirpation and/or extinction at high latitudes. We quantified differential rates of speciation, extinction, extirpation, and dispersal in low *versus* high latitudes, defined as within or exclusive of 30° latitude.

Low-latitude speciation began to exceed high latitude speciation after 30 Ma, suggesting the modern-style LDG is possibly driven by higher rates of low-latitude origination (Fig. 4; Fig. S37-43). These results are consistent with previous findings of higher speciation rates at low latitudes for planktonic foraminifera[15,41,42]. In addition to low-latitude speciation, local extirpation at high latitudes also contributed to a modern LDG, but tended to have a smaller effect size (Fig. 4; Figs. S37-S38, S40-43). In previous work, Yasuhara et al.[29] found that the redistribution of species' ranges, and not speciation, was important to LDG formation, but this was likely due to the short timescale of their study (i.e., last 3 Myrs).

Steepening of the diversity gradient was not driven by extinction, because extinction was higher at low latitudes beginning ~20 Ma (Fig. 4; Figs. S37-S38, S40-43). This heightened

extinction dampened the effect of higher speciation in the tropics, but speciation still tended to outpace extinction, adding to richness. Similarly, dispersal from high to low latitudes did not contribute to the emergence of a modern-day LDG, as dispersal dynamics were reversed (i.e., occurred predominantly from low to high latitudes) from at least the last 10 Ma (Fig. 4; Figs. S37-S38, S40-43). Dispersal was usually rapid, with a mean wait time after speciation of 1.7 Ma (+/- 2 Ma).

The relationship of temperature with richness was dampened for time periods older than 15 Ma. This weakened relationship may reflect climate model inaccuracies that inflate with time, or the fact that richness and temperature do not vary significantly with latitude. In the latter scenario of limited temperature variation, sensitivity analyses suggest no relationship would be found between richness and temperature (Fig. S29). Scarcer data in deep time may also make a relationship with temperature more difficult to obtain (Fig. S30), but LDGs constructed with minimal data still return modern-style LDGs towards the present (Fig. S31). Excluding sites potentially subject to dissolution only served to strengthen patterns (Fig. S32).

We were not able to test the relationship of richness with zooplankton biomass[43] or nutrient availability[44], because high-resolution spatial and temporal data for these variables at global scale do not exist for the last 40 Ma. However, the regions with highest foraminiferal richness, the nutrient-poor subtropical gyres[45], are characterised by lowest population densities of planktonic foraminifera[46]. The subtropical gyres have been incredibly stable over 10's of millions of years[47], and may have accumulated their high diversity due to both niche partitioning and reduced extirpation and extinction. It is possible that competition with diatoms or other groups, especially when interacting with seasonal food availability at high latitudes, also structure the LDG for planktonic foraminifera, but seasonality was found to correlate with SST and thermocline temperature range in many time bins and was therefore not included as a predictor in our models (Fig. S27).

Taken together, our results suggest the modern-day LDG for planktonic foraminifera is controlled, at least in part, by high-latitude cooling that brought colder bottom waters to the tropics, increasing both latitudinal temperature gradients and vertical temperature gradients at low latitudes. The increased vertical temperature structure within the water column at low latitudes may have enhanced niche partitioning, providing more opportunity for speciation over the last 30–15 Ma[16,36,37]. The cooler high-latitude water also served to extirpate regional populations of species. Consistent with our hypothesis, the tropics today are richer than the tropics of the Eocene and Miocene, potentially due to a stronger vertical temperature structure that was weak-to-absent during these warmer time periods. Extreme warmth during the Eocene may have also exceeded species' temperature tolerances equatorially, further depressing diversity, an effect that may occur more in the future as the tropics continue to warm[33,48]. Global diversity for planktonic foraminifera in the Eocene, however, was similar to planktonic foraminiferal diversity today, since species were distributed more evenly across latitudes 40 million years ago (Fig. 3).

Our analyses suggest a role for water column structure in facilitating niche partitioning and therefore the emergence of the LDG in planktonic foraminifera. Other potential mechanisms that may have elevated low-latitude diversity include the total amount of suitable area vertically within the water column[26], metabolic scaling[42,49], or changes to the biological carbon pump that redistributed nutrients at depth and opened new niches as climate cooled over the last 15 Ma[50,51]. This latter mechanism of nutrient redistribution complements the vertical temperature-driven niche separation proposed here.

To conclude, the establishment of the modern LDG is consistent with a cooling climate that allowed for elevated speciation via niche partition at low latitudes, whilst restructuring distributions and removing niches at high latitudes. By resolving how spatial patterns of biodiversity have varied through deep time, we provide valuable information on hypothesised causes crucial for understanding how biodiversity is generated and maintained over geological timescales, beyond the scope of modern-day ecological studies.

References

1 Fine, P. V. Ecological and evolutionary drivers of geographic variation in species diversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 369-392 (2015).

2 Hillebrand, H. On the generality of the latitudinal diversity gradient. *The American Naturalist* **163**, 192-211 (2004).

3 Mittelbach, G. G. *et al.* Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* **10**, 315-331 (2007).

4 Willig, M. R., Kaufman, D. M. & Stevens, R. D. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34**, 273-309 (2003).

5 Pontarp, M. *et al.* The Latitudinal Diversity Gradient: Novel Understanding through Mechanistic Eco-evolutionary Models. *Trends in Ecology & Evolution* **34**, 211-223, doi:https://doi.org/10.1016/j.tree.2018.11.009 (2019).

6 Crame, J. A. Taxonomic diversity gradients through geological time. *Divers Distrib* **7**, 175-189 (2011).

7 Mannion, P. D., Upchurch, P., Benson, R. B. J. & Goswami, A. The latitudinal biodiversity gradient through deep time. *Trends in Ecology & Evolution* **29**, 42-50 (2014).

8 Powell, M. G. Latitudinal diversity gradients for brachiopod genera during late Palaeozoic time: links between climate, biogeography and evolutionary rates. *Glob Ecol Biogeogr* **16**, 519-528 (2007).

9 Powell, M. G., Beresford, V. P. & Colaianne, B. A. The latitudinal position of peak marine diversity in living and fossil biotas. *J Biogeogr* **39**, 1687-1694 (2012).

10 Hillebrand, H. Strength, slope and variability of marine latitudinal gradients. *Marine Ecology Progress Series* **273**, 251-267 (2004).

11 Beaugrand, G., Rombouts, I. & Kirby, R. R. Towards an understanding of the pattern of biodiversity in the oceans. *Glob Ecol Biogeogr* **22**, 440-449 (2013).

12 Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098 (2010).

13 Pianka, E. R. Latitudinal gradients in species diversity: a review of concepts. *Am Nat* **100**, 33-46 (1966).

14 Saupe, E. E. *et al.* Spatio-temporal climate change contributes to latitudinal diversity gradients. *Nature Ecology & Evolution* **3**, 1419-1429 (2019).

15 Stehli, F. G., Douglas, R. G. & Newell, N. D. Generation and maintenance of gradients in taxonomic diversity. *Science* **164**, 947-949 (1969).

16 Rutherford, S., D'Hondt, S. & Prell, W. Environmental controls on the geographic distribution of zooplankton diversity. *Nature* **4000**, 749-752 (1999).

17 Klopfer, P. H. Environmental determinants of faunal diversity. *Am Nat* **93**, 337-342 (1959).

18 Haffer, J. & Prance, G. T. Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation. *Amazoniana* **16**, 579-607 (2001).

19 Dynesius, M. & Jansson, R. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences* **97**, 9115-9120 (2000).
20 Dobzhansky, T. Evolution in the tropics. *American Scientist* **38**, 209-221 (1950).
21 Williams, C. B. *Patterns in the Balance of Nature*. (Academic Press, 1964).
22 Paine, R. T. Food web complexity and species diversity. *Am Nat* **100**, 65-75 (1966).
23 Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M. & Roy, K. Is there a latitudinal gradient in the importance of biotic interactions? *Annu. Rev. Ecol. Evol. Syst.* **40**, 245-269 (2009).
24 Currie, D. J. Energy and large-scale patterns of animal and plant species richness. *Am Nat* **137**, 27-49 (1991).
25 Connell, J. H. & Orias, E. The ecological regulation of species diversity. *Am Nat* **98**, 399-414 (1964).
26 Rosenzweig, M. L. *Species Diversity in Space and Time*. (Cambridge University Press, 1995).
27 Fenton, I. S. *et al.* The impact of Cenozoic cooling on assemblage diversity in planktonic foraminifera. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150224, doi:doi:10.1098/rstb.2015.0224 (2016).
28 Yasuhara, M. *et al.* Past and future decline of tropical pelagic biodiversity. *Proceedings of the National Academy of Sciences* **117**, 12891-12896, doi:doi:10.1073/pnas.1916923117 (2020).
29 Yasuhara, M., Hunt, G., Dowsett, H. J., Robinson, M. M. & Stoll, D. K. Latitudinal species diversity gradient of marine zooplankton for the last three million years. *Ecology Letters* **15**, 1174-1179 (2012).
30 Jablonski, D., Roy, K. & Valentine, J. W. Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science* **314**, 102-106 (2006).
31 Yasuhara, M., Tittensor, D. P., Hillebrand, H. & Worm, B. Combining marine macroecology and palaeoecology in understanding biodiversity: microfossils as a model. *Biological Reviews* **92**, 199-215, doi:https://doi.org/10.1111/brv.12223 (2017).
32 Fenton, I. S. *et al.* Triton, a new species-level database of Cenozoic planktonic foraminiferal occurrences. *Scientific Data* **8**, 160, doi:10.1038/s41597-021-00942-7 (2021).
33 Yasuhara, M. & Deutsch, C. A. Paleobiology provides glimpses of future ocean. *Science* **375**, 25-26, doi:doi:10.1126/science.abn2384 (2022).
34 Yasuhara, M. *et al.* TIME MACHINE BIOLOGY CROSS-TIMESCALE INTEGRATION OF ECOLOGY, EVOLUTION, AND OCEANOGRAPHY. *Oceanography* **33**, 16-28 (2020).
35 Westerhold, T. *et al.* An astronomically dated record of Earth’s climate and its predictability over the last 66 million years. *Science* **369**, 1383-1387, doi:doi:10.1126/science.aba6853 (2020).
36 Al-Sabouni, N., Kucera, M. & Schmidt, D. N. Vertical niche separation control of diversity and size disparity in planktonic foraminifera. *Marine Micropaleontology* **63**, 75-90, doi:https://doi.org/10.1016/j.marmicro.2006.11.002 (2007).
37 Lowery, C. M., Bown, P. R., Fraass, A. J. & Hull, P. M. Ecological Response of Plankton to Environmental Change: Thresholds for Extinction. *Annual Review of Earth and Planetary Sciences* **48**, 403-429, doi:10.1146/annurev-earth-081619-052818 (2020).
38 Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic Deep-Sea Temperatures and Global Ice Volumes from Mg/Ca in Benthic Foraminiferal Calcite. *Science* **287**, 269-272, doi:doi:10.1126/science.287.5451.269 (2000).

39 Weiner, A., Aurahs, R., Kurasawa, A., Kitazato, H. & Kučera, M. Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist. *Molecular Ecology* **21**, 4063-4073 (2012).
40 Schneider, E. & Kennett, J. P. Segregation and speciation in the Neogene planktonic foraminiferal clade Globoconella. *Paleobiology* **25**, 383-395 (1999).
41 Raja, N. B. & Kiessling, W. Out of the extratropics: the evolution of the latitudinal diversity gradient of Cenozoic marine plankton. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20210545, doi:doi:10.1098/rspb.2021.0545 (2021).
42 Allen, A. P. & Gillooly, J. F. Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Ecology Letters* **9**, 947-954 (2006).
43 Irigoien, X., Huisman, J. & Harris, R. P. Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature* **429**, 863-886 (2004).
44 Schiebel, R. & Hemleben, C. *Planktic Foraminifers in the Modern Ocean*. 358 (Springer-Verlag, 2017).
45 Ruddimann, W. F. Recent planktonic foraminifera: dominance and diversity in North Atlantic surface sediments. *Science* **164**, 1164-1167 (1969).
46 Bé, A. W. H. & Tolderlund, D. S. in *Micropaleontology of Marine Bottom Sediments* (eds B M Funnell & W K Riedel) (Cambridge University Press, 1971).
47 Sibert, E., Norris, R., Cuevas, J. & Graves, L. Eighty-five million years of Pacific Ocean gyre ecosystem structure: long-term stability marked by punctuated change. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20160189, doi:doi:10.1098/rspb.2016.0189 (2016).
48 Chaudhary, C., Richardson, A. J., Schoeman, D. S. & Costello, M. J. Global warming is causing a more pronounced dip in marine species richness around the equator. *Proceedings of the National Academy of Sciences* **118**, e2015094118, doi:doi:10.1073/pnas.2015094118 (2021).
49 Worm, B. & Tittensor, D. P. *A Theory of Global Biodiversity*. (Princeton University Press, 2018).
50 Boscolo-Galazzo, F. *et al.* Temperature controls carbon cycling and biological evolution in the ocean twilight zone. *Science* **371**, 1148-1152, doi:doi:10.1126/science.abb6643 (2021).
51 Boscolo-Galazzo, F. *et al.* Late Neogene evolution of modern deep-dwelling plankton. *Biogeosciences* **19**, 743-762, doi:10.5194/bg-19-743-2022 (2022).

Acknowledgments

We thank Adam Woodhouse, Gwen Antell, Pincelli Hull, Helen Johnson, Heather Bouman, and Wolfgang Kiessling for informative discussions. We thank Moriaki Yasuhara and two anonymous referees for constructive feedback that improved this contribution. We also thank contributors to the Triton and Neptune Databases, from which Triton heavily draws. EES was supported by Leverhulme Trust grant RPG-201170, the Leverhulme Prize, and the National Science Research Council grant NE/V011405/1.

Data availability

All data to reproduce our analyses are provided in the following DOI: https://doi.org/10.6084/m9.figshare.21355467. The spatio-temporal planktonic foraminiferal occurrence data were derived from Triton, an open-source database (Fenton et al. 2021. *Scientific Data*): https://www.nature.com/articles/s41597-021-00942-7.

Code availability

All code to reproduce the analyses herein is provided in the following DOI: https://doi.org/10.6084/m9.figshare.21355467. Our custom code relied on the following R packages: HH R package v. 3.1-47, spatialreg R package v.1.2-3, geosphere R package, vegan R package v.2.5-7, and mapast R package v.0.1.

Author contributions

Conceptualization: EES

Methodology: EES, IF

Investigation: IF, EES, AF, PV

Visualization: TA, IF, EES

Funding acquisition: EES

Project administration: EES

Supervision: EES

Writing – original draft: EES

Writing – review & editing: EES, IF, TA, AF, PV

Competing interests

The authors declare that they have no competing interests.

Figure Captions

Fig. 1. The emergence of a modern-day latitudinal diversity gradient in planktonic foraminifera over the last 40 Ma. (**A**) Data used to quantify LDGs. Points are colored by the age of the sample. Distributional maps rotated to paleo-positions for 2.5 Ma intervals are available in Fig. S1. (**B**) LDGs constructed using unique site-by-age richness estimates in 2.5 Ma time bins (the midpoint age is shown). To generate curves, point-level data were binned by 15° latitude and richness estimated using the 75th percentile of the samples in each latitude-by-age bin. Results were robust to LDG construction method (Figs. S6-10) and show the modern-day LDG emerged only ~15 Ma. (**C**) Estimates of the gradient (i.e., slope) for LDGs over the last 40 Ma using the curves from (**B**). Gradients were insensitive to methodological approach (Figs. S11-24) and show a steepening of the gradient towards the present.

Fig. 2. Spatial autoregressive model coefficients from analyses examining the relationship between richness and four environmental predictors within 2.5 Ma time bins. The midpoint age for each bin is shown. Environmental variables include mean annual sea surface temperature, the log of mean annual mixed layer depth, the width of the thermocline, and the log of mean annual salinity. Coefficients are black if significant at an α of 0.05 using the Bejamini & Hochberg (BH) correction for multiple comparisons. Results are shown for data that was latitudinally restricted (within 55° latitude) but are consistent

with models that include all latitudinal data (Fig. S28). Error bars represent 95% confidence intervals around the mean.

Fig. 3. The dynamics of depth partitioning for planktonic foraminiferal assemblages across space and time. (**A**) We quantified evenness of depth habitat within each unique site-by-age bin using Simpson's index. This metric determines how evenly spread species are among the mixed layer, thermocline, and sub-thermocline for a given time and place within the water column. Mean evenness estimates were taken for each 2.5 Ma and 15° latitude bin. Analyses were performed only within 55° latitude, since data were limited at high latitudes earlier in the Cenozoic. (**B**) Mean evenness across latitudes for a given 2.5 Ma time bin, for low latitudes (defined as within 30°), high latitudes, and globally. Low-latitude assemblages of species today are more evenly distributed across the mixed layer, thermocline, and sub-thermocline than are assemblages at high latitudes. However, assemblages exhibited greater evenness across latitudes when the gradient was shallower millions of years ago. Current knowledge of foraminiferal depth preferences only allowed for measurement of evenness across three depths, but our assumption is the thermocline and subthermocline are subdivided to contain multiple niches, with more niches present at low latitudes during the last 15 Ma, and more niches present at higher latitudes during warmer intervals. For the raw evenness data in each sample, see Fig. S44.

Fig. 4. Macroevolutionary processes structuring LDGs over the last 40 Ma. Panels show the number of speciation, extinction, extirpation, and dispersal events in low-latitude (tropical) and high-latitude (temperate) regions. Dispersal quantified the number of species moving from the tropics to temperate regions and *vice versa*. The steepening of the gradient towards the recent coincides with enhanced rates of tropical speciation and with higher extirpation from temperate regions. Temperate and tropical regions were defined by 30° latitude for all time periods. Data show results for the approach including all site-by-age records. For proportional patterns, bootstrap subsampling analyses, and results excluding species present in both temperate and tropical regions within a given 2.5 Ma time bin, see Figs. S36-43.

1. Methods

1.1. Choice of taxon

Planktonic foraminifera are a group of unicellular, bio-mineralising marine plankton that originated ~170 Ma in the Early to Middle Jurassic[44]. They are an important component of the plankton from the high latitudes to the equator in all ocean basins. Species of planktonic foraminifera occupy a range of ecological niches in the upper two kilometers of the open ocean: some live in the surface mixed-layer and host algal photo-symbionts, while others live within or below the thermocline and feed primarily on sinking phytodetritus[44].

Due to their global abundance and preservation potential, planktonic foraminifera have been demonstrated to have the best species-level fossil record of the last 66 million years[52]. The quality of this fossil record permits an exceptionally high-resolution view into past species distributions, ecologies, and life histories. Recent work compiling this information into the single, harmonized Triton database[32] allows us to investigate the drivers and patterns of LDGs at greater fidelity than possible before.

1.2. Description and preparation of the data

We quantified temporal patterns in LDGs using 464,963 unique species-by-locality-by-age records from the Triton database[32], which represents the single largest dataset for any fossil group. Data were curated for taxonomic consistency, and all ages were converted to the GTS 2020 timescale. Following Fenton *et al*[32], occurrences significantly outside a species' known age range were excluded, using a threshold of 2 Ma for the Neogene and 5 Ma for the Paleogene; such records are likely to be taxonomic misidentifications or the result of reworking. Spatial coordinates were rotated to their past position (paleo-coordinates) based on Matthews *et al*[53]. Spatial imprecision in coordinates may occur due to error in the paleo-coordinate rotation plate model, and/or from ocean currents displacing foraminiferal tests as the dead organisms sank to the seafloor. However, neither source of error is likely to have large effects on the LDG patterns quantified herein, given the spatial resolution of our analyses (usually 15° latitude bins, or roughly 1500 km, see section 3.1). For example, for sites <1 km depth or for large foraminifera, the distance between where the organism died and where it settled is unlikely to be larger than ~100 km[54,55]. For depths of 2 km to 3.5 km, a maximum displacement of 100 to 400 km is reasonable[56,57]. We limited analyses to the last 40 Ma because poor-quality data at low latitudes early in the Cenozoic made it challenging to accurately calculate richness for these early intervals.

1.3. Quantifying temporal patterns for LDGs over 40 Ma

We estimated LDGs using percentiles from point-occurrence data and by calculating central tendencies using three different subsampling methods, described below. All analyses used the R v. 4.1.3 computing environment[58].

1.3.1. Spatio-temporal bin selection

LDGs were constructed by calculating richness within spatio-temporal bins. Numerous binning schemes were tested to evaluate the sensitivity of patterns to spatio-temporal resolution. Eight temporal bin durations were tested, from 2.5 Ma to 20 Ma, in 2.5 Ma increments. Temporal binning may inflate species counts[59], and thus bin duration should be narrower than the average duration of species[60]. For the planktonic foraminiferal species studied here, median species' duration was 5.96 Ma, suggesting bin durations longer than this will suffer from greater time averaging. We therefore focused analyses on the shortest temporal duration (2.5 Ma) but present results for other schemes (Figs. S7-9, S14). Longer temporal bins give steeper gradient estimates, but overall patterns remain consistent, as can be seen when rescaling richness in each time bin (Fig. S14). The spatial distribution of data at 2.5 Ma resolution are shown in Figs. S1.

We tested eight latitudinal bin resolutions, from 2.5° to 20°, in steps of 2.5°. We aimed to use the highest spatial bin resolution, whilst ensuring sufficient data to calculate LDGs within 2.5 Ma time bins. Therefore, to choose an 'optimum', we eliminated binning schemes that resulted in an empty set (no data) for latitudinal bins in any 2.5 Ma time bin. We further eliminated any binning scheme with insufficient data (defined here as 15 samples, 5 sites, and 150 records) in at least five latitudinal bins for any time bin. This approach identified 15° latitude bins as the highest 'ideal' resolution for analyses.

Analyses focused on a spatio-temporal resolution of 2.5 Ma and 15° latitude (Figs. S2-4), which was sufficiently sensitive to return the known, modern-day LDG with an equatorial dip (Fig. 1B)[16,28]. Estimates of the LDG using other spatial binning schemes produced similar results (Fig. S7-S9), and latitudinal resolution did not have a strong influence on gradient estimates (Fig. S15).

1.3.2. Point-occurrence method for LDGs

LDGs were quantified using unique site-by-age richness estimates within each 2.5 Ma time bin. Analyses excluded samples where the aim of the study was to identify selected species only, or where preservation was marked as poor (<10% of the data).

To generate LDG curves, point-level data were binned by 15° latitude. Richness was estimated using the 75th percentile of the samples in each latitude-by-age bin. Changes in the LDG through time were not affected by choice of percentile (Fig. S6 & S12). Percentiles were used to estimate LDGs because planktonic foraminiferal assemblages are more likely to be characterized by low richness due to dissolution or incomplete community counts, rather than high richness from time averaging, particularly given the short temporal bins used here. Furthermore, modern diversity patterns are characterized by the most diverse sites in a given latitude; for example, in many clades, the tropics today house both very low and very high species richness dependent on the location and environment[61].

1.3.3. Subsampling method for LDGs

To test whether point-level LDG patterns remained consistent when accounting for variation in data quantity, we estimated LDGs using three subsampling approaches. For each latitude-by-age combination, 1000 subsamples were drawn at random with replacement and the number of unique species counted. The mean of the 1000 subsamples served as the richness estimate for that latitude-by-age bin, with associated 95% confidence intervals.

Three different subsampling methods were employed: (i) by site, (ii) by sample, and (iii) by record (Fig. S5). The 'by site' method divided the dataset based on locality (or site) and selected a set number of sites for a given latitude-by-age bin. Richness was calculated from all species at the selected sites within that bin, irrespective of their exact age. We varied the number of sites selected from 3 to 20 (n = 3, 5, 10, 15, and 20). We focused on five sites because it allowed for the maximum number of spatio-temporal bins and produced narrower confidence intervals than when subsampling using three sites, but results were broadly insensitive to site number (Fig. S9 & S18).

The 'by sample' method of subsampling divided the dataset based on both locality and age and selected a set number of unique site-by-age samples from a given latitude-by-age bin. Each sample therefore represents the foraminiferal assemblage alive at a given time. The richness was calculated from unique species in the selected assemblages. We varied the number of samples selected from 10 to 50 in increments of 5. We selected 15 samples based on the trade-off between maximum number of samples and narrowest confidence intervals (Fig. S7 & S19).

The 'by record' method of subsampling divided the dataset based on locality, age, and species and selected a set number of unique species records from any site or age within a given latitude-by-age bin. Each record represented one row in the Triton dataset. The richness was calculated as the number of unique species within the selected records. We selected from 100 to 500 records in intervals of 50. We focused on 150 records, but results were broadly insensitive to record number (Fig. S8 & S20).

In contrast to point-level LDG construction, we did not exclude data where the aim of the study was to identify selected species, and where preservation was marked as poor, because subsampling does not assume any one sample is representative. Comparisons of point-level and subsampled LDG estimates show similar patterns (Fig. S10 & S23-S24), although subsampling tended to suggest higher average species richness. Higher richness in subsampled LDGs may reflect the influence of spatio-temporal averaging, which can inflate richness estimates. Alternatively or additionally, incomplete sampling or dissolution might lower richness in the point level estimates.

1.3.4. Estimating LDG gradients

The strength (or gradient) of LDGs was estimated using a linear model of richness as a function of latitude. For point-level data, richness estimates from the same site and time bin were characterized by the 50th, 60th, 75th, and 90th percentiles to avoid pseudo-replication. We focused on patterns using the 75th percentile, but results were insensitive to percentile choice (Figs. S6, S12, S13). To ensure gradient estimates were comparable over time, we calculated gradients only within 55° latitude, since older time periods were characterized by sparse high-latitude data. However, results were insensitive to latitudinal extent (Fig. S11).

Gradients were estimated individually for the Northern and Southern Hemispheres, and for both hemispheres combined using absolute latitude. We tested whether a model with separate slopes for each hemisphere provided a significantly better fit to data than a combined model using AICc. For most time periods (n=11 of 16), a model that used data from both hemispheres was preferred over a model with separate gradients (Table S1). Combining the two hemispheres increases the data on which a gradient is calculated.

We calculated gradients on rescaled richness to investigate the influence of variation in richness through time on gradient estimates (Fig. S11-S13). Richness was rescaled by setting maximum richness within a time bin to one. Rescaling affected the steepness of the latitudinal gradient but allowed for more direct comparisons of LDG shape across time. LDGs with peak richness in equatorial regions will have a steeper rescaled gradient than gradients where peak richness is in the mid-latitudes.

We compared gradients estimated from point-level LDGs to gradients estimated from subsampled LDGs. Gradients were calculated for each subsampling iteration and every latitude-by-age bin combination (Fig. S16), from which we derived mean and 95% confidence intervals (Figs. S14-S15, S18-S21). Using the median instead of the mean produced almost identical results (Fig. S17). Latitude-by-time bin resolution influenced gradient estimates but did not mask overall patterns (Figs. S14 & S15). Choice of age bin exaggerated latitudinal differences because broader age bins tended to result in more time averaging, but overall patterns were similar (Fig. S14). Similarly, choice of subsampling method had little effect, although the use of sites tended to produce broader confidence intervals (Figs. S21, S23 & S24). The amount of subsampling affected the steepness of the gradient, but not the overall shape, and rescaling removed this effect (Figs. S18-20). Visual inspection of gradients suggested the slopes of the Northern and Southern Hemisphere, for all time periods, were mirrored (Fig. S16).

Finally, we calculated and compared gradients from raw richness tallied directly from species counts in each latitude-by-age bin (Fig. S22), finding similar overall patterns to the subsampled and point-level gradients.

1.3.5. Estimating bimodality

Shallow gradients may indicate minimal differences in richness across latitudes or may indicate that bimodality was more pronounced during these times. To distinguish between these two possibilities, we compared a linear model with a second order polynomial model using AICc (Table S2).

1.4. Identifying LDG drivers over the last 40 Ma

1.4.1. Diversity dataset

We investigated possible LDG drivers over the last 40 Ma by coupling atmosphere-ocean generalized circulation models (AOGCMs) to point-level richness estimates within 2.5 Ma time bins. To avoid pseudo-replication, a single, mean richness estimate was calculated for each climatic grid cell.

1.4.2. Paleoclimate simulations

To explore whether paleoclimate influenced the biogeographic distribution of planktonic foraminifera, we utilized a newly updated version of a state-of-the-art paleo-general circulation model.

1.4.2.1. Paleoclimate model

Paleoclimate model simulations were carried out using a recent version of the UK Met Office coupled Atmosphere-Ocean General Circulation Model (AOGCM), HadCM3, or HadCM3L-M2.1D following the nomenclature of [62]. HadCM3L-M2.1D has a model resolution of 3.75° longitude × 2.5° latitude in the atmosphere and ocean (~250 km grid squares in the tropics), with 19 hybrid levels in the atmosphere and 20 vertical levels in the ocean. Equations were solved on the Arakawa B-grid. As is common in all climate models, sub-grid scale processes, such as cloud, convection, and oceanic eddies, were parameterized because they cannot be resolved at the scales required (usually meters to several kilometers) of the model resolution.

Due to scarce spatiotemporal data recording land-surface vegetation and soil characteristics in deep time, we used modern-day vegetation expressions for broadleaf trees, deciduous trees, shrubs, C3 and C4 grasses (five in total), and a globally-uniform distribution of medium loam soil characteristics in the model land surface scheme (MOSES 2.1)[63]. The land surface scheme also included evaporation from sub-grid scale lakes, which were prescribed as a lake fraction in each grid box at the start of the simulation. We used a version of the model that included the dynamical vegetation model, TRIFFID (Top-Down Representation of Interactive Foliage and Flora Including Dynamics). TRIFFID predicts the distribution and properties of global vegetation based on plant functional types (PFTs) in the form of fractional coverage (and thus PFT co-existence) within a grid-cell, based on competition equations using the climate tolerance of five plant functional types.

The model included a further update that modified cloud condensation nuclei density and cloud droplet effective radius, following recent work[64,65]. This modification raised temperatures at high latitudes, without significantly changing tropical temperatures, which reduced the pole-to-equator temperature gradient in line with proxy observations. This update was found to work under hot, cool, and icehouse conditions, as well as under pre-industrial boundary conditions, making it appropriate for use across modern and deep time.

The ocean model was based on the model of Cox et al.[66] and is a full primitive equation with a three-dimensional model of the ocean. A second-order numerical scheme was used along with centred advection to remove nonlinear instabilities. Flux adjustments—such as artificial heat and salinity adjustments in the ocean component model[67] to prevent them from drifting to unrealistic values—were not required in this model, which is a crucial feature for long paleoclimate simulations[68]. Sea-ice was calculated on a zero-layer model; partial sea ice coverage was possible, with a consistent salinity assumed for ice.

Each simulation was initialised from an equilibrated pre-industrial state in the atmosphere and ocean. Surface vegetation was uniformly set as shrub everywhere and allowed to evolve via TRIFFID based on the evolution of the local climate.

The HadCM3 family of models has contributed extensively to the Coupled Model Intercomparison Project (CMIP 1-5) experiments and the Paleoclimate Modelling Intercomparison Project (PMIP 1-4), and has demonstrated skill at reproducing the modern-day climate[62] and paleoclimate in an array of experiments[69-71]. Paleoclimate experiments require hundreds of years to reach a near-surface equilibrium state but substantially longer (many thousands of years) for the deep ocean[68,69]; even longer is required for true climate equilibrium, due to the long period of adjustment of ocean circulation to applied forcings. Lower resolution models are less computational expensive, allowing fully equilibrated

simulations of deep time climate, which would not be possible with higher resolution, more complex models.

1.4.2.2. Snapshot simulations, specific boundary conditions and spin-up

We ran nine 'snapshot' simulations over the last 40 Ma of the Cenozoic, with each simulation having time-specific boundary conditions. Paleogeographic digital elevation models (DEMs) were produced by the EarthByte group as part of the PALEOMAP project[72]. Each stage and time-specific DEM were interpolated from a 1°x1° grid onto the HadCM3L 3.75°x2.5° grid. Similarly, land ice was transformed onto the model grid assuming a simple parabolic shape to estimate ice sheet height (m). 'Realistic' pCO_2 concentrations for each simulation were based on Foster et al.[73]. Time-specific solar luminosity for each simulation was based on Gough[74]. All other boundary conditions, such as orbital parameters, volcanic aerosol concentrations, etc., were held constant at pre-industrial values.

To ensure each simulation had fully adjusted to the boundary conditions, we followed a 3-stage spin-up protocol so that each simulation was fully equilibrated: i) The globally and 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 m2. In general, these simulations were run for over 9,000 model years to ensure full Earth system equilibrium. Climate means were produced from the last 100 years of each simulation.

1.4.2.3. Spatiotemporal interpolation techniques

Using the climatic snapshot simulations, we interpolated model data and boundary conditions over the last 40 Ma. First, the DEM (bathymetry & topography) was interpolated linearly between each pair of snapshot simulations at 0.5 million-year-increments onto a 1°x1° longitude by latitude grid. Each increment was time weighted between the two snapshot simulations using the DEMs. The land-sea mask was generated for each newly generated DEM, where any grid box above 0 m was taken as land and any grid box below 0 m as ocean. Isolated ocean points were removed if six of the surrounding grid boxes were land, with the corresponding topography set as the mean of the surrounding grid boxes.

The snapshot model data (e.g., sea surface temperature, sea surface salinity) was interpolated using bicubic remapping for each variable onto the interpolated land-sea mask. Environmental parameter fields (e.g., sea surface temperature, salinity) were extrapolated to fill in grid boxes that were newly created using Poisson's equation (elliptic partial differential equation) over the input domain.

1.4.2.4. Paleo-climate model uncertainties

Global-scale biodiversity data requires global-scale environmental datasets, such as sea surface temperature. Global proxy databases with large spatiotemporal coverage are available for past time periods, but proxy evidence is limited deeper in time and increasingly less well-constrained, necessitating the use of paleoclimate models to provide these global datasets. However, simulating paleoclimate is challenging. This challenge stems, in part, from the many parameters that may be unconstrained for deeper time observations. Uncertainties can be partitioned into two main sources: i) boundary condition uncertainty, and ii) model uncertainty.

i) Boundary conditions are spatiotemporally varying parameters that are required by climate models, but which cannot be calculated internally by the model and instead need to be provided by the user. The most important boundary conditions for the model used here

654 are: (a) paleogeographic reconstructions, (b) ice sheet height and extent, (c) solar luminosity,
655 (e) orbital configuration, and (e) greenhouse gas concentrations.
656
657 (a) A large source of uncertainty arises from paleogeographic reconstruction. DEMs,
658 constrained by paleo-databases, provide topography, bathymetry, and land-sea
659 concentration. These DEMs are crucial for determining local, regional, and global
660 atmospheric and ocean circulation, and therefore the climate in the model. The deeper
661 in time, the less proxy data is available, which results in greater uncertainty in these
662 reconstructions. However, our understanding of plate tectonics, spreading ridges,
663 weathering rates, and basinal deposition allows for accurate first-order approximation
664 of deep-time paleogeography, especially over the last 40 Ma. The largest uncertainties
665 usually result from the height and depth of topographic and bathymetric surfaces and
666 their spatial coverage because of proxy uncertainty, e.g.[75].
667 (b) Ice sheets (and associated sea-level height) can have a large impact on regional
668 climate and global climate, primarily due to changes in land surface area, planetary
669 albedo, and modification to ocean and atmospheric circulation. Currently, most
670 paleoclimate models prescribe the height and extent of ice sheets based on proxy
671 evidence.
672 (c) Solar luminosity, which is the amount of energy received by a planetary body from its
673 parent star, is fairly well known. Gough (1981)[74] approximated the amount of energy
674 based on a simple model using the age of the parent star. Apart from the first 0.2 Gyr
675 of Earth history, this approximation is shown to agree well with observations[76].
676 (d) Orbital configuration (the eccentricity, obliquity, and precession of the planet's orbit
677 around its parent star and rotation on its own axis) can have a substantial impact on
678 the seasonal and latitudinal climate signal, which, in turn, can lead to significant
679 changes in climate state (glacial to interglacial cycles). These effects are due to ice
680 sheet formation and associated changes in global atmospheric and oceanic response.
681 For deep-time simulations, a modern orbital configuration is often imposed. There are
682 several reasons for this. First, chronological uncertainty in proxies that are used in
683 model comparison will cover many orbital cycles, which may result in the proxy
684 being more representative of a mean orbital state (akin to the modern day). Second,
685 using a modern configuration makes it easier to understand how different a deep-time
686 climate is compared to the modern climate. Orbital variation will have its largest
687 impact on climate where ice sheets can grow, which is partly reflective of the amount
688 of pCO_2 in the atmosphere and associated global temperature.
689 (e) Greenhouse gas concentrations, and more specifically pCO_2 concentrations, are
690 variable through the geologic past. Proxy type, age, techniques, and calibration
691 uncertainty when converting to a pCO_2 estimate, as well as the number of records
692 through the geologic past, can make constraining a deep time pCO_2 concentration
693 problematic[73]. Furthermore, although a combination of multiple proxy sources can
694 improve the robustness of the estimated pCO_2 estimates, it may prove problematic in
695 situations where the errors are combined to produce a mean pCO_2 estimate. Time
696 averaging is also an issue. Although paleogeography changes on geological
697 timescales, pCO_2 concentrations can vary on hundreds to million-year timescales.
698 Here we were only interested in the long-term background signal (millions of years),
699 with scatter around the mean, typically ~400ppm[73].
700
701 ii) Although all globally available paleoclimate models use the same well-known
702 equations to simulate the large-scale behavior of the atmosphere and ocean, results from
703 different models can vary, in particular at the local and regional scale. This variation is due to

the complexity of each model, resolution dependencies, spin-up and the applied initial state, and parameterizations used to approximate processes such as cloud formation that cannot be explicitly resolved at the grid-scale of all current paleoclimate models. In an ideal world, such paleoclimate simulations would be run by multiple paleoclimate modelling groups, as is done in the Climate Model Intercomparison Project (CMIP) for near-future climate change studies. Unfortunately, such paleo-climate comparisons are not currently possible; for instance, these simulations took over two years to complete on a high-performance supercomputer, and few paleoclimate modelling groups have the capability to set up such deep time simulations. However, confidence in the robustness of our results can be obtained by the fact that: (a) the HadCM3 family of models, although 20 years older than many contemporaries, still compares reasonably well with models form the previous IPCC Coupled CMIP fifth assessment models[62]. (b) HadCM3L-M2.1D has seen continued development[62]. Here we use an updated version of the model that solves a persistent problem in the majority of paleoclimate models known as the 'cold pole paradox', where simulated higher latitude temperatures were previously much cooler than suggested by proxy-observations. (c) These simulations have been run for over 9000 model years. Paleoclimate simulations usually are run for only a couple of thousands of years due to computational costs. However, it can take upwards of 5,000 years to allow a model simulation to equilibrate to all the applied model forcings, especially for the deep ocean, and, as such, for global ocean circulation to be fully representative of the deep-time period. (d) Although model uncertainty is important to constrain, it has been shown that scenario uncertainty (i.e., the applied boundary conditions) is a larger source of error, at least for future climate simulations[77].

1.4.3. Climate drivers

1.4.3.1. Environmental constraints on planktonic foraminifera

We investigated nine environmental variables previously posited to constrain the distributions of planktonic foraminifera[44], including mean annual sea surface temperature, seasonal variation in sea surface temperature, mean annual mixed layer depth, seasonality of the mixed layer depth, mean annual sea surface salinity, seasonality of sea surface salinity, the width of the thermocline, the temperature range within the mixed layer, and the temperature range within the thermocline (Fig. S25-S27).

Sea surface temperature. Surface temperature estimates were calculated at 5 m water depth. Temperatures within the water column were highly correlated with surface temperatures and therefore not included in model comparisons. Temperature seasonality was estimated as the standard deviation of monthly variations. HadCM3 estimates potential temperatures (i.e., removing the effect of pressure). Planktonic foraminifera are more likely to respond to actual temperature, but the difference between potential and actual temperature is small (< than 1°C at 1000 m) and thus we did not correct for this effect.

Surface salinity. Surface salinity was calculated at 5 m water depth. Salinity deeper in the water column was highly correlated with surface salinity and therefore not included in model comparisons. Seasonality in salinity was estimated as the standard deviation of the monthly variations. Extremes of salinity, whether high or low, may exceed planktonic foraminiferal species' tolerances, whose salinity optimum sits around 35 PSU[44]. Therefore, the expected relationship with salinity is thought to be polynomial on the original scale. However, we have insufficient data to model polynomials for many of the time intervals, and thus to model the effect of salinity, we calculated the absolute difference from this optimal salinity value. By using an absolute difference from the optimum, we converted the relationship to approximately linear, with high richness close to the optimum and low richness at the extremes.

Water column structure. We estimated annual mean width of the mixed layer, seasonal variation in mixed layer depth, and annual mean width of the thermocline. The mixed layer refers to the top of the water column, where environmental conditions are relatively homogeneous and stable due to mixing. Mixed layer depth was derived from HadCM3L output, which used a Kraus-Turner[78] bulk mixed layer approach that calculated a balance between the energy available for mixing the water column and the introduction of buoyancy at the ocean surface. For more detail, see Foreman[79].

The thermocline is the section of the water column below the mixed layer where temperature changes rapidly with depth. Below the thermocline, temperature is relatively stable. On a rescaled temperature profile (where the temperature and depth axes are rescaled from 0 to 1), the base of the thermocline was calculated as the point where the gradient steepens past a -1:1 line using a loess function in the base R stats package[58]. We chose a span $\alpha = 0.4$ based on sensitivity analyses: if the span is too large, the regression will be over-smoothed, whereas if the span is too small, large variance will result.

Water column temperature structure. We calculated the mean temperature range within the mixed layer and thermocline based on their identified boundaries, described above. The temperature range within the thermocline and mixed layer provided a proxy for the diversity of temperature niches at depth within the water column.

Maps for the environmental variables can be found in Fig. S25.

1.4.3.2. Variable selection

The interpolated AOGCM data provided mean global estimates for each of the nine variables at 2.5 Ma intervals. Shallow-water grid cells (water depth < 200m) were excluded for each time bin, based on a Scotese bathymetry model[72]. Removing shallow-water cells excluded coastal environments where planktonic foraminifera communities may be limited by environmental conditions [44]. Variables were centred and rescaled within each 2.5 Ma time bin to better interpret the intercept terms and to ensure similar units for regression coefficients.

The environmental variables were highly collinear with each other in most time periods (Fig. S26). We tested for multicollinearity within each of the 2.5 Ma time bins using variance inflation factors (VIFs). Analyses relied on the 'vif' function in the HH R package v. 3.1-47[80]. To include the same set of environmental predictors across the 16 time bin models, we limited analyses to those variables with VIFs < 2 in each time interval, following[81]. The final variable set included mean sea surface temperature, mean annual mixed layer depth, mean annual sea surface salinity, and width of the thermocline. Mean thermocline temperature range was also considered of interest, but was too strongly correlated with mean sea surface temperature to be included in the same models (i.e., the mean R^2 for a model between these two variables is 0.886 across time bins).

We assessed the relationship of these four variables with richness using both multivariate and univariate models (Figs. S28-S33). For univariate analyses, we included thermocline temperature range alongside the other four predictors (Fig. S33). Before modelling, mixed layer depth (mean and seasonality) and salinity (mean and seasonality) were log transformed to correct for non-linearity.

1.4.4. Model building and calibration

1.4.4.1. Identifying LDG drivers within time intervals

We modelled the relationship between richness and the four selected environmental predictors within each 2.5 Ma time bin using spatial autoregressive models. Models were fit using the errorsarlm function in the spatialreg R package v.1.2-3[82]. We adjusted significance levels (α of 0.05) to account for multiple comparisons by applying the Bejamini & Hochberg

(BH) correction[83]. The optimum neighborhood distance was calculated for each time period, between 500 km and the distance where autocorrelation becomes non-significant, following [84] using AIC. Models were checked for heteroscedasticity. We restricted analyses to within 55° latitude, because older time periods lacked data at high latitudes. Using all latitudinal data, rather than limiting analysis to within 55° latitude, produced similar results (Fig. S28).

1.4.4.2. Model exploration and sensitivity analyses

We explored the effects of sampling and preservation on within-age-bin model results. Sampling in the fossil record varies over time, with older time bins typically characterized by fewer data points. Data loss can potentially mask relationships between the predictors and species richness (Fig. 2). We therefore explored whether the results held when fewer data points were used for within-age-bin models. Specifically, in each time bin, data were subsampled to include the same number of data points as the most data-poor time bin (n=49, for 37.5–40 Ma), iterated 100 times. To summarize these results, we used the mean coefficient and standard error from the 100 subsampling iterations for each time bin. Model results removed any statistically significant relationship between temperature and richness, even for more recent time periods, but coefficients were similar (Fig. S30). Importantly, modern-day LDGs constructed from the more limited data show similar patterns to those constructed on the full dataset (Fig. S31).

Dissolution increases with water depth as the carbonate ion concentration and calcite saturation state decrease. Dissolution can artificially reduce richness at a particular site and time as a function of ocean basin depth and the species present, because certain species are more susceptible to dissolution[44]. The differential impact of dissolution may obscure the relationship between environmental variables and species richness. We assessed the potential effect of dissolution on model results by excluding samples with species richness in the lowest 25th percentile, based on 15° latitude and 2.5 Ma age bins. Model results suggested similar relationships and patterns (Fig. S32).

The warmer, greenhouse conditions on Earth ~40 Ma ago limited temperature variability across latitudes. Reduced temperature variation may obscure any relationship of temperature with species richness. We therefore tested whether a relationship between temperature and richness can be detected in more recent time bins, when analyses were limited to the range of temperatures characteristic of warmer times. We identified the time bin with the smallest temperature range (35–37.5 Ma time bin, with a temperature range from 15.25–33.38°C) and restricted analyses for the other time bins to this temperature range. Using this approach resulted in some time bins having an even narrower temperature range, since time bins towards the present do not have temperatures reaching 33°C.

When a narrow temperature range was used, the relationship between temperature and richness was lost for most time bins, aside from the most recent (Fig. S29). Limiting analyses to a narrower temperature range suggested that no relationship exists between temperature and species richness, even for more recent time bins. Thus, the absence of a relationship between richness and temperature for deeper time bins may reflect the narrower temperature range, which itself elevated species richness across latitudes.

1.4.4.3. Identifying LDG drivers through time

In addition to investigating the relationship between potential environmental drivers and richness within time bins spatially, we examined the degree to which change in environmental variables drives change in richness at given sites. To compare shorter- and longer-term trends, differences were calculated using a range of temporal gaps (from 2.5 Ma to 12.5 Ma, in 2.5 Ma steps). For each temporal gap, we modelled the change in species richness as a linear function of the change in the selected environmental variables. Only a

subset of the data used for spatial, within-time analyses could be used for the across-time analyses: to be included, sites were required to have pairs of samples with the requisite time gap (Fig. S35-36). Analyses with longer time gaps had fewer data points (Table S3).

We ran both multivariate and univariate models; for the latter, we included analysis of thermocline temperature range (Table S4). Richness was estimated as the mean richness for each site within each 2.5 Ma time bin. Analyses used values for the environmental variables on their original scale, rather than rescaled values. Because separate models were run for different time gaps, we adjusted significance values (α of 0.05) following the Bejamini & Hochberg (BH) correction[83]. Results were insensitive to latitudinal extent (Table S3 & S4).

1.4.5. The effect of ocean area on species richness

Open ocean area has been hypothesized to drive the LDG, because area is found to correlate positively with species richness today[26]. To test whether this relationship holds through time for planktonic foraminifera, we modelled richness as a function of open ocean area within 15° latitude and 2.5 Ma time bins (Fig. S34). Open ocean area was defined as the area of grid cells with depths > 200 m based on the Scotese bathymetry models[72]. We quantified area using the areaPolygon function in the geosphere R package[85].

1.5. Depth partitioning across space and time

Species of planktonic foraminifera live at different depths in the water column, down to about two kilometers[44]. Depth preferences for fossil species can be estimated using the isotopic signature of their shells and are usually divided into mixed layer, thermocline, and sub-thermocline dwellers[52]. We investigated how the depth structure of planktonic foraminiferal assemblages changed across space and time. For each unique site and time bin, we partitioned species by three depth categories (mixed layer, thermocline, and sub-thermocline dwellers). We classified the single species inferred to occur across multiple depths (*Chiloguembelina ototara*) as a mixed layer dweller, since it possessed photo-symbionts.

We quantified evenness of depth habitat within each unique site-by-age with Simpson's index using the diversity function in the vegan R package v.2.5-7[86]. This metric determined how evenly spread species are among the three depths for a given time and place. That is, highly even assemblages would have roughly equal numbers of species in each depth class, whereas uneven assemblages would vary more in species count by depth class. We excluded sites most likely to be subject to dissolution, defined as those that fell in the lowest 25th percentile of species richness for a given site and time. The mean of these evenness estimates was taken for each 2.5 Ma and 15° latitude bin (Fig. S44). Analyses were performed only within 55° latitude, given the sparse data at high latitudes earlier in the Cenozoic (Fig. 3).

Current knowledge of foraminiferal depth preferences only allowed for measurement of evenness across three depths over the last 40 million years. Our assumption, however, is that the thermocline and subthermocline are subdivided to contain multiple thermal niches, with more niches available during warmer conditions earlier in the Cenozoic across latitudes, and more niches available at low latitudes during the last 15–10 Ma. The mixed layer is unlikely to be as subdivided, given conditions within are more homogenous.

1.6. Speciation, extinction, and dispersal dynamics

To determine the macroevolutionary processes structuring LDGs, we quantified spatial patterns of speciation, extinction, extirpation, and dispersal in planktonic foraminifera over the last 40 Ma. Specifically, we investigated whether species originated and went extinct in tropical or temperate regions, whether they dispersed primarily into or out of these regions, and whether they were extirpated (went regionally extinct) at higher rates in one region over

the other (Fig. 4, Fig. S37-40). Since our focus is on latitudinal patterns, not climatic patterns, we defined the separation between tropical and temperate regions as 30° for all time periods. We selected 30° because modern-day richness for planktonic foraminifera peaks in the subtropics; present-day richness quantified using the typical cut-off of 23° would result in similar diversity in these regions.

We used two approaches to quantify macroevolutionary dynamics. In the first approach (Fig. S37), we considered all site-by-age records, including those outside of 55° latitude. Speciation was considered to have occurred in the region of first occurrence for a species (Fig. S39). Similarly, extinction was considered to have occurred in the region of last occurrence for a species. That is, even if a species was present in both temperate and tropical regions within a 2.5 Ma time bin, the species was counted as going extinct in the tropics if that region held the last occurrence. Extirpation was documented in the time bin when the species last disappeared from a region (either temperate or tropics) but persisted in the other region. In this way, each species had, at most, one extirpation. Dispersal dynamics were identified by documenting how long it took a species after origination to appear in a region different from where it originated; for example, if a species originated in the tropics, we assessed whether the species ever left the tropics, and, if so, how long it took the species to move to the temperate region. Species that originated and moved within a 2.5 Ma time bin were counted towards both metrics in that time bin. Raw counts (Fig. S37), proportions (Fig. S37), and per capita rates (Fig. S40) were calculated.

In the second approach (Fig. S38), we excluded species present in both temperate and tropical regions within a given 2.5 Ma time bin. These species do not contribute to the development of the LDG. Thus, by focusing only on species present in one region, we can more clearly distinguish the processes contributing to LDG formation. For these analyses, we counted a speciation as the region of first occurrence for a species, but only if the species did not expand its range to another region in the same time bin. Extinction was recorded in the region where the species was last observed, but only if the species was found uniquely in that region. Extirpation occurred when a species was present in both regions but was lost from one region. Thus, extirpation could occur numerous times for each species over time, unlike in the previous approach. Dispersal was counted when a species moved from one region to another. For example, when a species was found only in the tropical region in one time bin, to when it was found only in the temperate region in another time bin. The two approaches produced congruent conclusions, which is that low-latitude speciation and high-latitude regional extirpation, contributed to the steepening of the diversity gradient over the last 40 Ma.

Incomplete sampling may affect the location of first and last appearance for taxa in the fossil record. Although it is unlikely that geographic error in the first and last appearances would be sufficient to mask true patterns given the broad geographic bins used here, we employed a bootstrap resampling protocol to test this assumption. Each species was subsampled down to 75% of available records across its lifetime, and the evolutionary metrics (speciation, extinction, extirpation, dispersal) re-calculated for each time bin. This subsampling process was repeated 100 times, such that 100 temperate and 100 tropical estimates were returned for each metric (speciation, extinction, extirpation, and dispersal) in each time bin. Our goal was to assess how sampling may affect understanding of the location of first and last appearance and dispersal patterns for species. The estimates, however, remained broadly consistent across bootstrap replicates (Figs. S41-43), most likely because planktonic foraminifera are densely sampled with high spatio-temporal resolution—arguably the best of any fossil group[52]. Planktonic foraminifera are employed extensively for biostratigraphy and have been subject to considerable taxonomic revision, resulting in well-

established stratigraphic ranges[87]; accordingly, the geographic locations of first and last occurrences in Triton that correlate with these datums are robust.

Methods References

52 Aze, T. *et al.* A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews* **86**, 900-927 (2011).

53 Matthews, K. J. *et al.* Global plate boundary evolution and kinematics since the late Paleozoic. *Global and Planetary Change* **146**, 226-250, doi:https://doi.org/10.1016/j.gloplacha.2016.10.002 (2016).

54 v. Gyldenfeldt, A.-B., Carstens, J. & Meincke, J. Estimation of the catchment area of a sediment trap by means of current meters and foraminiferal tests. *DSR* **47**, 1701-1717, doi:https://doi.org/10.1016/S0967-0645(00)00004-7 (2000).

55 Qiu, Z., Doglioli, A. M. & Carlotti, F. Using a Lagrangian model to estimate source regions of particles in sediment traps. *Science China Earth Sciences* **57**, 2447-2456 (2014).

56 Siegel, D. A. & Deuser, W. G. Trajectories of sinking particles in the Sargasso Sea: modeling of statistical funnels above deep-ocean sediment traps. *Deep Sea Research Part I: Oceanographic Research Papers* **44**, 1519-1541, doi:https://doi.org/10.1016/S0967-0637(97)00028-9 (1997).

57 Waniek, J., Koeve, W. & Prien, R. D. Trajectories of sinking particles and the catchment areas above sediment traps in the northeast Atlantic. *Journal of Marine Research* **58**, 983-1006 (2000).

58 R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2019).

59 Alroy, J. The Fossil Record of North American Mammals: Evidence for a Paleocene Evolutionary Radiation. *Systematic Biology* **48**, 107-118, doi:10.1080/106351599260472 (1999).

60 Marcot, J. D. The fossil record and macroevolutionary history of North American ungulate mammals: standardizing variation in intensity and geography of sampling. *Paleobiology* **40**, 238-255 (2014).

61 Gaston, K. J., Williams, P. H., Eggleton, P. & Humphries, C. J. Large scale patterns of biodiversity: spatial variation in family richness. *Proceedings of the Royal Society Biological Sciences* **260**, 149-154 (1995).

62 Valdes, P. J. *et al.* The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geoscience Model Development* **10**, 3715-3743 (2017).

63 Cox, P. M. *et al.* The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *ClDy* **15**, 183-203, doi:10.1007/s003820050276 (1999).

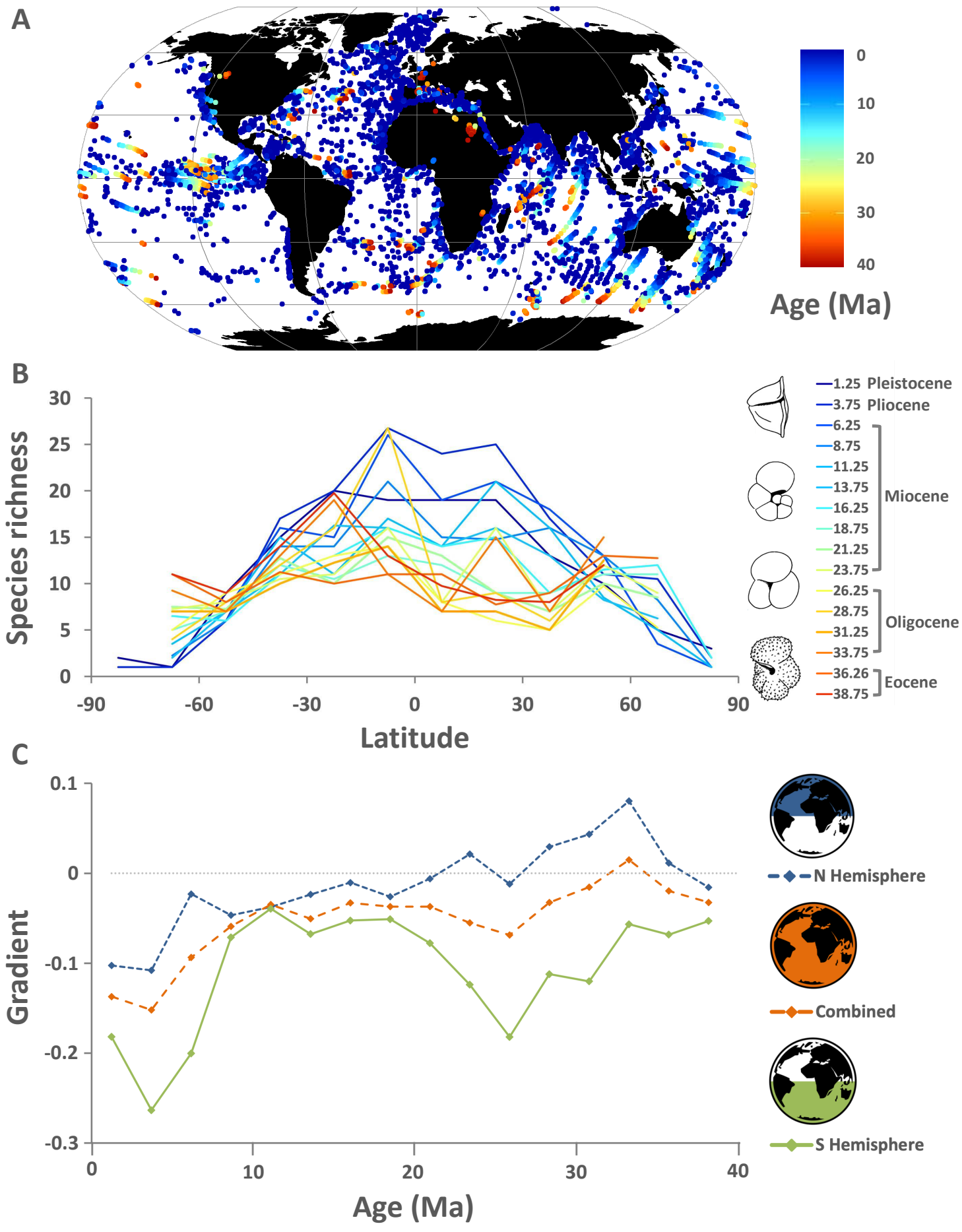
64 Sagoo, N., Valdes, P., Flecker, R. & Gregoire, L. J. The Early Eocene equable climate problem: can perturbations of climate model parameters identify possible solutions? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20130123 (2013).

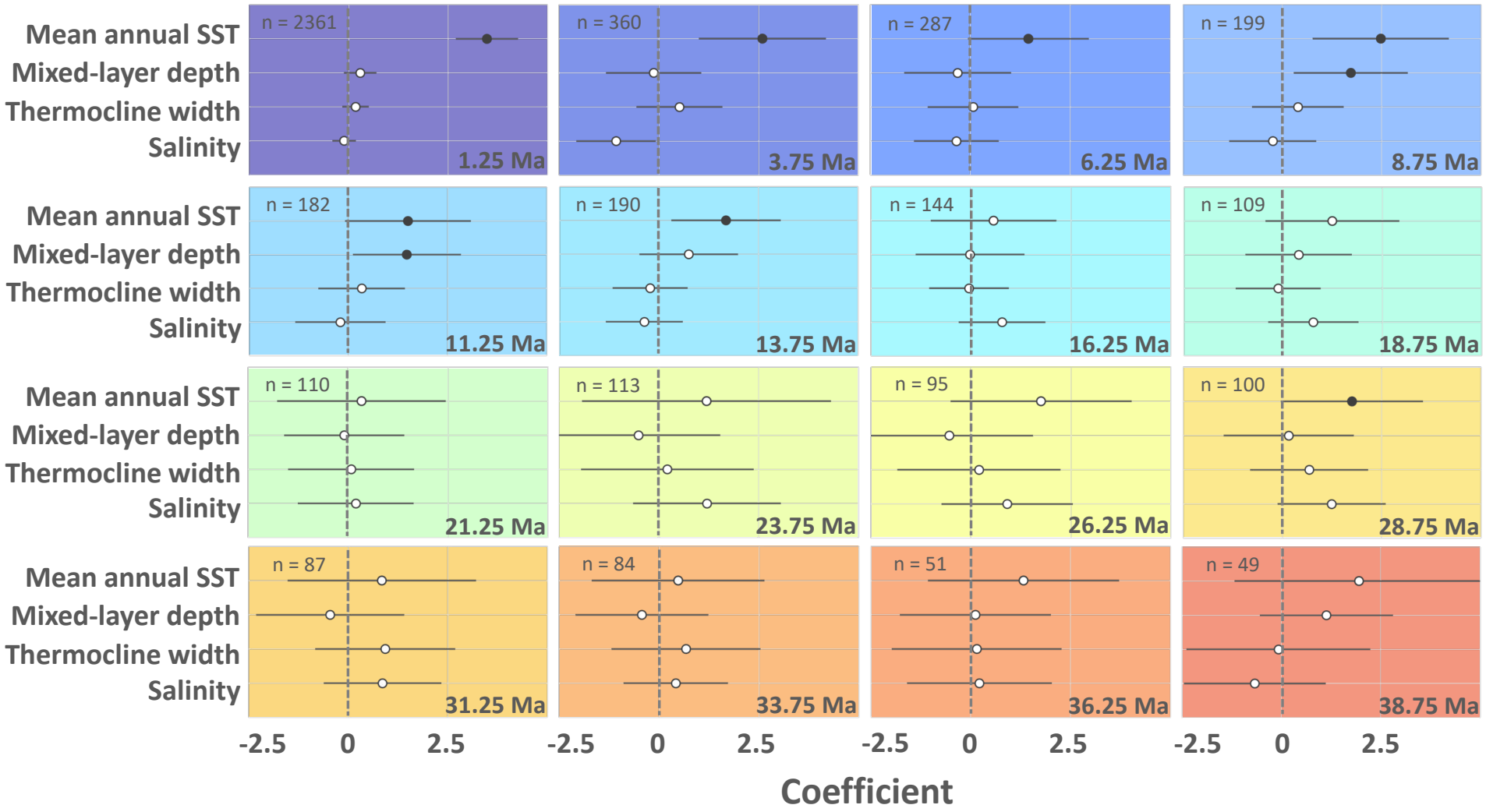
65 Kiehl, J. T. & Shields, C. A. Sensitivity of the Palaeocene–Eocene Thermal Maximum climate to cloud properties. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**, 20130093 (2013).

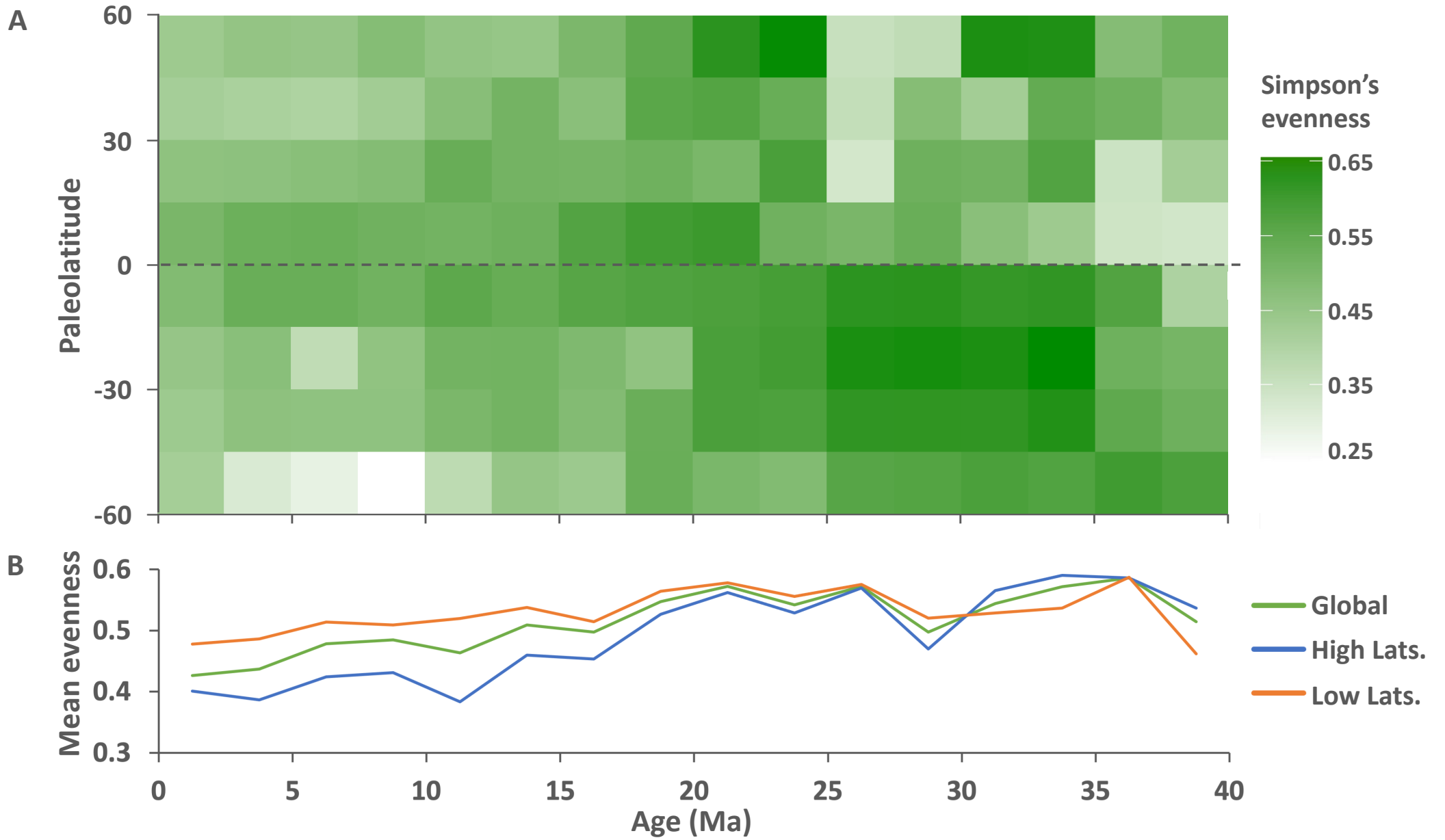
66 Cox, M. D. A primitive equation, 3-dimensional model of the ocean. *GFDL Ocean Group Technical Report No 1, GFDL, Princeton University* (1984).

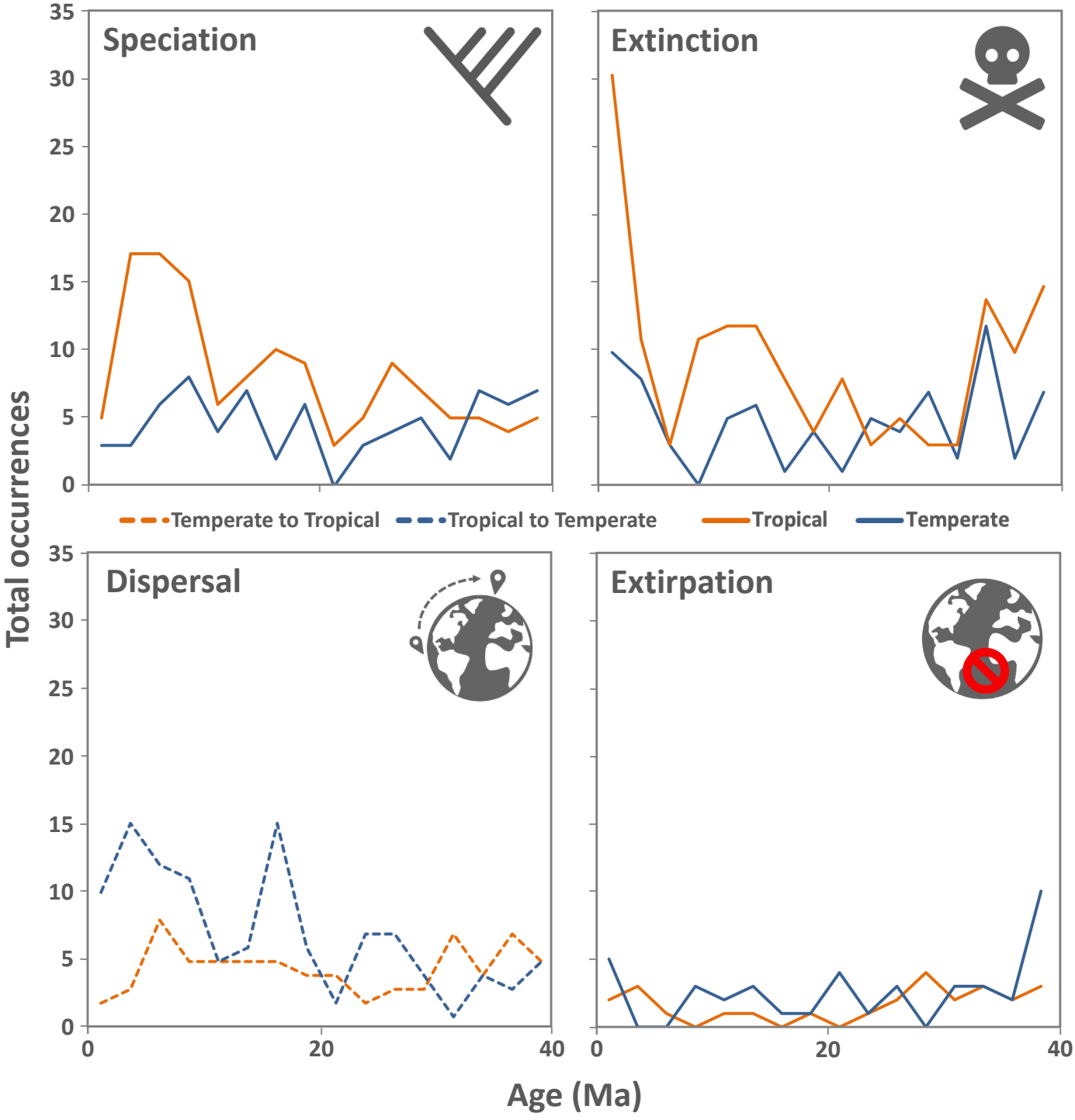
67 Collins, M., Tett, S. F. B. & Cooper, C. The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *ClDy* **17**, 61-81, doi:10.1007/s003820000094 (2001).
68 Farnsworth, A. *et al.* Climate Sensitivity on Geological Timescales Controlled by Nonlinear Feedbacks and Ocean Circulation. *Geophysical Research Letters* **46**, 9880-9889, doi:https://doi.org/10.1029/2019GL083574 (2019).
69 Valdes, P. J., Scotese, C. R. & Lunt, D. J. Deep ocean temperatures through time. *Clim. Past* **17**, 1483-1506, doi:10.5194/cp-17-1483-2021 (2021).
70 Farnsworth, A. *et al.* Past East Asian monsoon evolution controlled by paleogeography, not CO₂. *Science Advances* **5**, eaax1697, doi:doi:10.1126/sciadv.aax1697 (2019).
71 Jones, L. A., Mannion, P. D., Farnsworth, A., Bragg, F. & Lunt, D. J. Climatic and tectonic drivers shaped the tropical distribution of coral reefs. *Nature Communications* **13**, 3120, doi:10.1038/s41467-022-30793-8 (2022).
72 Scotese, C. R. & Wright, N. *PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the Phanerozoic*, 2018).
73 Foster, G. L., Royer, D. L. & Lunt, D. J. Future climate forcing potentially without precedent in the last 420 million years. *Nature Communications* **8**, 14845, doi:10.1038/ncomms14845 (2017).
74 Gough, D. O. Solar interior structure and luminosity variations. *SoPh* **74**, 21-34 (1981).
75 Farnsworth, A. *et al.* Paleoclimate model-derived thermal lapse rates: Towards increasing precision in paleoaltimetry studies. *Earth and Planetary Science Letters* **564**, 116903, doi:https://doi.org/10.1016/j.epsl.2021.116903 (2021).
76 Bahcall, J. N., Pinsonneault, M. H. & Basu, S. Solar Models: Current Epoch and Time Dependences, Neutrinos, and Helioseismological Properties. *The Astrophysical Journal* **555**, 990-1012, doi:10.1086/321493 (2001).
77 Hawkins, E. & Sutton, R. The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society* **90**, 1095-1108, doi:10.1175/2009BAMS2607.1 (2009).
78 Kraus, E. B. & Turner, J. S. A one-dimensional model of the seasonal thermocline, Part II. *Tell* **19**, 98-105 (1967).
79 Foreman, S. J. Unified Model Documentaiton Paper Number 40, The Ocean Model Report. (The Met Office, 2005).
80 HH: Statistical Analysis and Data Display: Heiberger and Holland. R package version 3.1-47 (2022).
81 Zuur, A. F., Ieno, E. N. & Elphick, C. S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3-14 (2010).
82 Bivand, R., Millo, G. & Piras, G. A Review of Software for Spatial Econometrics in R. *Mathematics* **9**, 1276 (2021).
83 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: Series B (Methodological)* **57**, 289-300 (1995).
84 Cooper, N. & Purvis, A. Body size evolution in mammals: complexity in tempo and mode. *Am Nat* **175**, 727-738 (2010).
85 geosphere: Spherical Trigonometry. R package version 1.5-14 (2021).
86 Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.5-7. (2020).
87 Wade, B. S., Pearson, P. N., Berggren, W. A. & Pälike, H. Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to the

geomagnetic polarity and astronomical time scale. *Earth-Sci Rev* **104**, 111-142, doi:https://doi.org/10.1016/j.earscirev.2010.09.003 (2011).


A
0
10
20
30
40
Age (Ma)
B
Species richness
30
25
20
15
10
5
0
-90
-60
-30
0
30
60
90
Latitude
1.25 Pleistocene
3.75 Pliocene
6.25
8.75
11.25
13.75
16.25
18.75
21.25
23.75
Miocene
26.25
28.75
31.25
33.75
Oligocene
36.26
38.75
Eocene
C
Gradient
0.1
0
-0.1
-0.2
-0.3
0
10
20
30
40
Age (Ma)
N Hemisphere
Combined
S Hemisphere


Mean annual SST
Mixed-layer depth
Thermocline width
Salinity
n = 2361
1.25 Ma
n = 360
3.75 Ma
n = 287
6.25 Ma
n = 199
8.75 Ma
n = 182
11.25 Ma
n = 190
13.75 Ma
n = 144
16.25 Ma
n = 109
18.75 Ma
n = 110
21.25 Ma
n = 113
23.75 Ma
n = 95
26.25 Ma
n = 100
28.75 Ma
n = 87
31.25 Ma
n = 84
33.75 Ma
n = 51
36.25 Ma
n = 49
38.75 Ma
-2.5
0
2.5
Coefficient


A
60
30
0
-30
-60
Paleolatitude
Simpson's evenness
0.65
0.55
0.45
0.35
0.25
B
Mean evenness
0.6
0.5
0.4
0.3
0
5
10
15
20
25
30
35
40
Age (Ma)
Global
High Lats.
Low Lats.


Speciation
Extinction
Dispersal
Extirpation
Total occurrences
Age (Ma)
0
5
10
15
20
25
30
35
0
20
40
Temperate to Tropical
Tropical to Temperate
Tropical
Temperate