

# Thermal niches of planktonic foraminifera are static throughout glacial–interglacial climate change

Gawain T. Antell<sup>a,1</sup>, Isabel S. Fenton<sup>a</sup>, Paul J. Valdes<sup>b</sup>, and Erin E. Saupe<sup>a,1</sup>

<sup>a</sup> Department of Earth Sciences, University of Oxford, OX1 3AN Oxford, United Kingdom

<sup>b</sup> School of Geographical Sciences, University of Bristol, BS8 1SS Bristol, United Kingdom

<sup>1</sup> To whom correspondence may be addressed. Email: gawainantell@gmail.com or erin.saupe@earth.ox.ac.uk.

**Contributions:** E.E.S. formulated the idea. I.F. compiled species occurrence data. P.J.V. ran climate models. G.T.A. and E.E.S. designed the analysis. G.T.A. programmed the code, visualized the data, and drafted the manuscript. All authors edited and approved the text. The authors declare no competing interests.

**Short title:** Thermal niche stasis during climate change

## Classification

1. Biological Sciences - Ecology
2. Physical Sciences - Earth, Atmospheric, and Planetary Sciences

## Significance Statement

We examined the degree to which temperature tolerances changed on 8,000-year timescales across 700,000 years of glacial–interglacial climate change. We coupled a new fossil occurrence database of planktonic foraminifera, an abundant type of zooplankton, with Atmosphere–Ocean Global Circulation Model reconstructions of past climates. Our suite of analyses demonstrated that foraminiferal species have not shifted their temperature tolerances in response to glacial cycles: species occupied the same temperature conditions regardless of the magnitude of global temperature change. The limited tendency of planktonic foraminifera to change their tolerances suggests that ongoing global change could hasten local or global extinctions of plankton and other widely-dispersing marine species.

## ABSTRACT

Abiotic niche lability reduces extinction risk by allowing species to adapt to changing environmental conditions *in situ*. In contrast, species with static niches must keep pace with the velocity of climate change as they track suitable habitat. The rate and frequency of niche lability have been studied on human timescales (months to decades) and geological timescales (millions of years), but lability on intermediate timescales (millennia) remains largely uninvestigated. Here, we quantified abiotic niche lability at 8-ka resolution across the last 700 ka of glacial–interglacial climate fluctuations, using the exceptionally well-known fossil record of planktonic foraminifera coupled with Atmosphere–Ocean Global Climate Model reconstructions of paleoclimate. We tracked foraminiferal niches through time along the univariate axis of mean annual temperature, measured both at the sea surface and at species’ depth habitats. Species’ temperature preferences were uncoupled from the global temperature regime, undermining a hypothesis of local adaptation to changing environmental conditions. Furthermore, intraspecific niches were equally similar through time, regardless of climate change magnitude on short timescales (8 ka) and across contrasts of glacial and interglacial extremes. Evolutionary trait models fit to time series of occupied temperature values supported widespread niche stasis above randomly wandering or directional change. Ecotype explained little variation in species-level differences in niche lability after accounting for evolutionary relatedness. Together, these results suggest that warming and ocean acidification over the next hundreds to thousands of years could redistribute and reduce populations of foraminifera and other calcifying plankton, which are primary components of marine food-webs and biogeochemical cycles.

**Keywords:** ecological niche conservatism, climate change, planktonic foraminifera, macroecology, macroevolution

## BACKGROUND

Abiotic niche dynamics determine patterns of community composition over space and regulate trajectories of diversity over time (1). Both niche lability (2, 3) and conservatism (1, 4) have been proposed to spur speciation, and abiotic niche lability has been associated with ecological invasions (5-7) and with reduced risk of extinction during times of climate change (8). Thus, a deeper understanding of species' propensity for niche stasis vs. lability could improve predictions of biodiversity restructuring in response to anthropogenic climate change (9).

Stasis in species' abiotic niches through time has been documented in empirical research, but most such studies have been limited to ecological niche modeling on decadal scales (reviewed in 10) or paleoecological examination on  $10^6$ - to  $10^7$ -year scales (5, 11, 12). Since empirical rates of niche change are scarce and difficult to acquire, many studies merely assume that niche evolution occurs at a constant rate along branches of a phylogeny (2, 3, 6, 7). Niche dynamics at intermediate timescales of centuries to millennia are particularly poorly documented (10), and studies at this meso-scale have been restricted to terrestrial systems (e.g., 13, 14, 15) or to comparisons between the present-day and the single historical time step of the Last Glacial Maximum, ~21 ka ago (16-20). Quantifying the rate and relative frequency of niche change in marine species over timescales of  $10^2$ – $10^5$  years is important, however, because species will adapt or go extinct in response to anthropogenic ocean changes over this timescale (21).

Here, we investigated climatic niche lability from the rich sedimentary archive of global planktonic foraminifera across the last 700 ka of glacial–interglacial cycles, at 8-ka resolution. Planktonic foraminifera (Protista) construct “shells” (tests) of calcite, thereby sequestering carbon and recording an isotopic signature of past ocean conditions. Tests readily accumulate over large expanses of the seafloor. Consequently, the fossil record of foraminifera—arguably

“the best fossil record on Earth” (22)—affords an exceptionally high-resolution view into past species distributions. This detailed record fuels studies of biostratigraphy, paleoclimatology, and paleoecology (20, 22-25). Moreover, the complete species diversity of planktonic foraminifera has been described for the Plio-Pleistocene, with good agreement between morphological and molecular phylogenies (22, 25-27). Although some have speculated that foraminifera competitively exclude each other (24), recent work found that planktonic foraminifera species seldom restrict each other’s distributions (28). Presumably, therefore, species occupy the full envelope of existing environmental conditions within their tolerance limits, and geographic distributions are determined almost entirely by physical ocean conditions.

We developed five analyses to investigate the degree of abiotic niche lability in foraminifera. All methods examined the univariate niche axis of temperature, which is the single most important explanatory variable with regard to geographic distributions of foraminifera (20, 29-32) and is a climate-related stressor and extinction driver for diverse marine fauna across timescales (33, 34). The adaptive potential of thermal niches has been taken as a key determinant of global community structure and genetic connectance in plankton (35). Primary productivity and other environmental variables, however, may also structure abiotic niches of plankton (36). Our suite of analyses quantified whether and by how much planktonic foraminiferal niches shifted along a temperature axis. First, we correlated time series of species’ thermal optima with global temperature to determine whether species tracked suitable habitat or experienced environmental fluctuations *in situ*. We then quantified species’ niche dissimilarity between pairs of time bins—either tracking niches across bin boundaries or contrasting niches at climatic extremes of glacial maxima and interglacial thermal peaks. As a new way to characterize abiotic niche change, we applied trait evolution models to continuous time series of temperatures at

occupied sites. Lastly, we explored variation in intraspecific niche lability among ecotypes while accounting for phylogenetic relatedness. SI Appendix Table S1 lists the response variable and sample size for each analysis.

## RESULTS

Occurrence data were derived from a global dataset of Cenozoic planktonic foraminifera records (37). We converted fossil coordinates into presence–absence values within 1.25° raster grid cells at 8-ka intervals over the past 700 ka (i.e.,  $n = 88$  time bins), as detailed in *Methods* sections “Fossil occurrence compilation” and “Spatiotemporal resolution.” Mean annual temperature (MAT) was extracted from a coupled Atmosphere–Ocean Generalized Circulation Model at every cell containing an occurrence for each time interval (Fig. 1, *Methods* section “Paleoclimate reconstructions”). Occurrences from ocean drilling cores covered all oceans and latitudinal bands (SI Appendix Movie S1). We repeated all analyses first using temperature at the sea surface and then using temperatures derived at each species’ preferred depth layer (a “depth-habitat approach,” defined in *Methods* section “Species attributes” and SI Appendix Table S2). The final dataset contained 24 species and 42,233 unique age-cell-species combinations.

Thermal niches were estimated from occurrences using kernel density estimation (KDE; Fig. 1, SI Appendix Fig. S1), except in “Trait evolution models” analysis where the mean of occurrences was used instead. KDE generates a continuous occurrence probability distribution for an organism along an environmental or resource gradient, based on a finite sample of presence points (e.g., 38, 39). We programmed KDE in a way that standardized for differential sampling of environmental conditions through time (*Methods* section “Kernel density estimation”). Thermal optima (preferred temperatures) were defined as the MAT associated with the probability function maximum for a species in a time bin (Fig. 1). Niche dissimilarity was

defined as the Hellinger's H distance between two probability distributions; identical distributions give  $H = 0$  and non-overlapping distributions give  $H = 1$  (SI Appendix Fig. S2). SI Appendix Table S2 lists the ecotype, depth habitat, thermal optimum, and mean H for each species.

### Time series correlation

For the first test of niche lability, we correlated concurrent time series of preferred and available temperatures to quantify the degree of climate tracking on short (8-ka) timescales. A species' preferred temperature was defined as the MAT associated with the largest estimated probability density from KDE in a time bin (Fig. 1). Available environment was summarized as average MAT across global marine points (*Methods* section "Definition of 'available environment'"). We considered separately any discontinuous time series for a focal species, resulting in 32 time series from the 24 study taxa (Figs. S3 and S4; *Methods* section "Time series correlation").

If species adapted to changing environmental conditions *in situ*, one would expect a positive correlation between preferred and available global temperatures. If instead niches remained static and species migrated to track preferred thermal conditions, one would expect a negligible correlation, and results were consistent with this scenario. The mean correlation (Pearson's r coefficient) between species' thermal optima and available temperatures was 0.062 with standard deviation of 0.26. Results were nearly identical under the depth-habitat approach ( $r = 0.031 \pm 0.23$ ).

### Niche lability vs. climate change magnitude

In addition to correlating standing time series of thermal preferences and global temperature (above), we also compared bin-to-bin changes in these two variables. This second

examination of niche lability quantified changes in the entire niche distribution from one time step to the next, rather than tracking a single point estimate of a niche. For each species, we compared thermal niche estimates at consecutive time bins using Hellinger's H distance (Fig. 1; *Methods* section "Niche dissimilarity measure") and then averaged H distances among species that crossed each time bin boundary. The process of building, standardizing, and comparing niche distributions is presented in *Methods* section "Kernel density estimation" and SI Appendix Figs. S5 and S6. As a correlate to the average species' niche shift across each boundary, we calculated the magnitude of change in mean global temperature from one time step to the next.

If species shifted, broadened, or otherwise changed their niches in response to climatic perturbation, one would expect larger H distances with larger climate change magnitude. Instead, mean H appeared to vary independently of global change, under both the sea-surface and depth-habitat approaches (Fig. 2, SI Appendix Fig. S7). Niches were similar (H was small) between every pair of intervals (Fig. 2, SI Appendix Fig. S7).

### Niche lability among extreme climate intervals

In addition to tracking thermal niche change across consecutive time steps, described in the preceding analysis, we compared niches at glacial minima and maxima (Fig. 3A, SI Appendix Fig. S8 and Table S3). For every possible pair of different (1) glacial, (2) interglacial, or (3) glacial–interglacial time bins, we calculated mean H distance among species sampled in both bins (Fig. 1, SI Appendix Fig. S1). If niches adapted to available environmental conditions, one would expect more niche divergence (larger H distances) in the glacial–interglacial category than in either of the two control categories. By comparing niche lability at extreme endpoints of climate-change fluctuations, this test could reveal lability differences that might be too subtle to

detect in the preceding analysis, which evaluated the narrow gradient of change in available environment from one time bin to the next.

Niches were no more different between intervals of climatic extremes than they were in intervals of similar climate (Fig. 3B, SI Appendix Fig. S9A), supporting the preceding results that suggested thermal niches of planktonic foraminifera have been static. Sampling in the youngest interglacial and oldest glacial interval was anomalous, and these bins were leverage points in the analysis (*Methods* section “Niche lability among extreme climate intervals” for details). When only the middle six glacial and interglacial intervals were considered, H distances were smaller (niche estimates were more similar) and less variable, but the pattern across categories remained the same (Fig. 3C, SI Appendix Fig. S9B). Furthermore, we tested the analytical power to detect signals of niche change by simulating fossil occurrences under a hypothesis of local niche adaptation within unchanging geographic distributions (*Methods* section “Test for type II error”). A signal of adaptation was detected in all 100 simulations (SI Appendix Fig. S10), which adds confidence that niche lability would have been detected if present and strong.

### Trait evolution models

We fit five trait evolution models to time series of occupied temperature for each species to quantify relative support for interpreting each series as (1) strict stasis, (2) stasis *sensu lato* (white noise), (3) an unbiased random walk, (4) a biased (directional) random walk, or (5) tracking of mean global temperature (*Methods* section “Trait evolution models” for details). Support for either of the stasis models would suggest niche stability (or white-noise fluctuation about a static mean), support for either of the random walk models would indicate niche lability independent of climate, and support for covariate-tracking would indicate niche shifts that

matched global trends in environmental conditions. Historically, trait evolution models have been applied to phenotypic data but were adapted here to environmental occupancy data. As in “Time series correlation” analysis, we split the data into 32 time series by species. However, in that earlier analysis we tracked the optimum temperature from KDE for each species, whereas present implementation of trait evolution models required us to track the mean of MAT from original occurrence points instead (SI Appendix Figs. S3 and S4).

Stasis models fit all species time series better than random walk or covariate-tracking models: strict stasis was best supported for 10 series and general stasis for 22 series, using temperatures extracted at the sea surface (Fig. 4). Support for static trait models confirms qualitative inspection of time series (SI Appendix Figs. S3 and S4). Results were similar under the depth-habitat approach to MAT estimation (SI Appendix Fig. S11); strict stasis models received the most support for nine series, general stasis for 22 series, and unbiased random walk for one series (*Pulleniatina obliquiloculata*).

Evolutionary models parameterize the variance in a trait mean through time as a parameter called omega. In strict stasis models, the trait mean is invariant through time (i.e., omega is zero). In general stasis models, omega represents the amplitude of “white-noise” biological fluctuation; in practice omega might also include some variation from heterogeneous fossil sampling. Omega estimates from general stasis best-fit models ranged from 0.3–1.8 °C (1.1 °C on average; SI Appendix Table S4). These values are small relative to the 10–20 °C temperature tolerance ranges reported from laboratory studies of planktonic foraminifera (29) or the up-to-30 °C range in ocean MAT both horizontally and vertically.

## Phylogenetic ecotype comparisons

The degree of niche lability can vary among species, and ecological and life history traits could explain some of this variation. For instance, foraminifera that actively prey upon other plankton might incur higher metabolic expenses than autotrophic species. Consequently, these heterotrophs might be more sensitive to ambient temperature changes and less flexible in adapting their thermal niche. Indirect evidence for this hypothesis comes from the observation that symbiont-bearing planktonic foraminifera experience faster speciation rates than those without symbionts (24). Many studies have proposed that faster climatic niche lability causes increased speciation (e.g., 2, 3), although niche conservatism could also promote speciation by stimulating allopatry (1, 4). Alternatively, macroecological differences among ecotypes could be due in part to an advantage for symbiont-bearing foraminifera in warm intervals, when nutrient supply is lower (40).

We used ANOVA to investigate differences in niche lability among ecotypes, while incorporating phylogenetic data to mitigate pseudoreplication due to shared evolutionary history (SI Appendix Fig. S12). Ecotype explained little variation in niche lability among species either in the sea-surface approach ( $F(4, 22) = 3.31$ , phy  $p = 0.34$ , SI Appendix Table S5A) or depth-habitat approach ( $F(4, 22) = 2.55$ , phy  $p = 0.44$ , SI Appendix Table S5B).

## DISCUSSION

Our analyses found prevalent stability of thermal niches in planktonic foraminifera across 700 ka. The independence of preferred and available temperature time series indicated that niche lability was not tied to climate cycles at the scale of  $10^3$  years, and trait evolution models indicated stasis at a temporal scale of  $10^6$  years. Similarly, analyses of intraspecific niche

dissimilarity through time indicated a decoupling of niche lability and climate change magnitude, both on short intervals (8 ka) and across glacial–interglacial extremes. Previous studies have argued that foraminifera do (41) and do not (20) adapt to environmental change on millennial scales, but our results from a much larger number of time intervals suggest planktonic foraminifera track suitable habitats rather than changing their thermal preferences.

Niche stasis could result from failure to adapt at range margins, where individuals would be most likely to encounter novel combinations of abiotic and biotic conditions (42). One hypothesis for limited adaptation is that dispersal of individuals from central populations swamps locally-adapted individuals in the gene pool at range margins (43, 44). Alternatively, lack of emigrants to marginal populations could stifle genetic variation and thereby limit adaptation (43, 44). Outcomes depend on dispersal ability relative to environmental gradient steepness, population density relative to carrying capacity, and selection strength relative to genetic drift (44, 45). The balance of each set of factors differs for marine ecosystems in comparison to their better-studied terrestrial counterparts (42). Foraminifera, like other plankton and animals with pelagic larvae, disperse great distances on directional currents and attain enormous population sizes. Terrestrial organisms with large populations and long dispersal distances, such as some plants, may share the pattern of limited niche lability found here.

Abiotic niche stasis puts species at risk of population declines and extinction from climate-related stressors (8). Survival hinges both on suitable conditions continuing to exist somewhere on Earth, and on populations establishing in those areas at least as quickly as populations elsewhere go extinct. Due to shallower environmental gradients over ocean expanses compared to on land, marine species must migrate quickly to keep pace with preferred conditions (46)— poleward shifts in living marine species have been clocked at nearly six times faster than

those of terrestrial counterparts (47). For calcifying plankton, ocean acidification and stratification add to the stress of warming water temperature (48, 49). Declines and redistributions in this functional group will cascade to affect marine food-webs (48, 50) and biogeochemical cycles (21).

The specific result of thermal niche stability in foraminifera carries important consequences even if conclusions are not extrapolated to other clades. In particular, “transfer functions” in paleoclimatology assume that thermal niches of foraminifera remain constant. These functions convert microfossil census data to proxies of sea surface temperature, which serve as the basis for forcing and validating ocean circulation models (23, 51). Hence, evidence of prolonged niche stability in foraminifera strengthens the theoretical basis for the tools used to understand past Earth conditions and to predict future climate states.

A major aim of macroevolutionary research is pinpointing at what rate(s) niche change accrues along phylogenetic trees. Although we found thermal niches of foraminifera to be stable on scales of hundreds of thousands of years, these niches must evolve at some point since species have distinct environmental preferences. Punctuated equilibrium offers one reconciliation of abiotic niche stability within lineages and niche differences across a tree, by attributing niche change to speciation events (52).

### Acknowledgements

The authors thank Tim Coulson and members of the Ecological and Evolutionary Dynamics Lab (University of Oxford) for discussions on the work. A Leverhulme Grant DGR01020 (I.F., E.E.S.) and the Clarendon Foundation (G.T.A.) supported the authors.

## METHODS

### DATA

#### Paleoclimate reconstructions

The temperature at all 3-dimensional locations in prehistoric oceans is not directly observable but can be reconstructed using a coupled Atmosphere–Ocean General Circulation Model (AOGCM). We obtained mean annual temperature (MAT) estimates from the UK Met Office Unified Model Hadley Centre Coupled Model (HadCM3), with the MOSES 2.1 land-surface model and TRIFFID dynamic global vegetation model (53). The precise configuration was HadCM3b-M2.1aD (54). This AOGCM linked atmospheric circulation (at a resolution of  $3.75^\circ \times 2.5^\circ$  and 19 levels) to ocean circulation (at  $1.25^\circ \times 1.25^\circ$  horizontal resolution and 20 vertical ocean levels). The AOGCM included many environmental variables besides MAT, such as salinity, water age, and N-S and E-W current velocity. Primary productivity and terrestrial nutrient input were not modeled in the AOGCM because these variables are less accurate and more difficult to reconstruct globally, although they are important resource axes for foraminifera in some contexts (36).

A climate simulation was run at 4-ka intervals for the last 700 ka, initialized with climate estimates from similar HadCM3 simulations based on the last glacial–interglacial cycle (55). Each simulation was run for 1,000 years: 900 years of spinup to allow equilibration, and a further 100 years that were averaged to produce the output. The ocean layers of interest (within the upper few hundred meters) required only ~500 years to equilibrate. Deeper waters reached equilibrium more slowly (on the scale of millennia) but impacted surface waters only slightly.

Each simulation at each time step was accorded time-specific boundary conditions for orbital position, greenhouse gas concentrations (carbon dioxide, methane, and nitrous oxide), and

ice sheet extent and volume (and corresponding ocean salinity). The greenhouse gas forcing for each time step was linearly interpolated from ice core records. Ice sheet reconstructions were those of De Boer and others (56, 57), which are not so well constrained as more recent reconstructions (e.g., ICE-6G (58) and GLAC-1D (59-62) for the past 27 ka). However, comparisons of climates at the Last Glacial Maximum (21 ka ago) using these different ice sheets revealed relatively small changes except over the ice sheets themselves, and the DeBoer ice sheet reconstructions are the only available estimates of global ice sheets over the last 1 Ma.

### Definition of “available environment”

We defined “available environment” as the set of ocean temperatures from a standard latitude–longitude grid at a given time step, as reconstructed by AOGCM. The first four analyses compared species’ thermal niches to the mean or first difference of mean available environment. Hence, calculation of available conditions affected many sets of results. One issue is that averaging MAT of all AOGCM grid cells indiscriminately in a time step could falsely dampen climatic signals. During glacial intervals some cells were covered in ice and therefore would not contribute to MAT estimates, bringing up global mean temperature; during warm intervals some of these polar cells would be open water, bringing down global mean temperature. To obviate this issue of ice-volume fluctuation, we laid a 10° grid of points on the Earth and identified the points that consistently recorded open-water conditions throughout the study interval. We then took the mean of MAT at these ocean points to generate a time series of global temperature, plotted in Fig. 3A. For the depth-habitat approach to analyses, we took separate global means at each of the three species’ depth layers.

Available environment could be defined as the set of temperatures at all sampled occurrence points instead of at all marine points. Depending on the direction of the null

hypothesis, one could argue that using sampled sites would lessen the chance of type 1 error. In practice, however, we found that replacing the global mean with the sample-site mean gave equivalent results. This is because the entire study interval was well sampled (SI Appendix Movie S1), and the mean MAT among all sampled points closely tracked the global mean (Pearson's  $r = 0.80$ ), although with slight discrepancies in the timing of some of the interglacial peaks (SI Appendix Fig. S8). Therefore, results herein are derived from the global marine environment.

### Fossil occurrence compilation

We assembled the most comprehensive dataset of Cenozoic foraminiferal occurrences to date (37). We subset the dataset to records from the last 700 ka to match the temporal extent of AOGCM reconstructions. Most occurrence records ( $10^5$ ) were acquired from datasets in the PANGAEA Open Access library ([www.pangaea.de](http://www.pangaea.de)). The compilation also includes records ( $10^3$ ) from the Deep Sea Drilling Project (1968–1983), the Ocean Drilling Program (1985–2004), the Integrated Ocean Drilling Program (2004–2013), and the International Ocean Discovery Program (2013–present); these were accessed from the Neptune database (<http://www.nsb-mfn-berlin.de>) or the program's online data repository. Additionally, our dataset includes occurrences from ForCenS, which itself draws on four compilations and six datasets, including the compilations of CLIMAP, the Brown University Foraminiferal Database, the ATL947 database, and MARGO (63).

We harmonized the taxonomy according to currently valid nomenclature (37). Macro- and micro-perforate species were included, but benthic foraminifera were excluded. The database contains a text column for “original name” to track synonymization. We excluded two

species that originated within the last 700 ka and eight species that went extinct over that span, to ensure all study species were extant throughout the entire study interval.

All specimens were collected from sediment cores and core tops, with no live-collection samples such as from plankton tows. Even the youngest sediment samples are estimated to be centuries or millennia in age (64). Thus, all specimens reflect pre-industrial species distributions, which were unaffected by anthropogenic global warming. Paleo-coordinates were reconstructed by rotating modern coordinates according to the plate model of Matthews and others (65), in accordance with Neptune methods.

We omitted records from sites at shallow depths, where planktonic foraminifera cannot establish viable populations (66). The exact minimum depth for reproduction is unknown and may vary with water conditions. Here, we excluded samples collected at < 100 m water depth under the assumption that any foraminifera at those sites drifted in from elsewhere. Deep-water sites could also be problematic because carbonate dissolution could preferentially destroy fossils of certain species. However, such preservational artifacts had negligible impact in previous studies of abiotic niche estimates in foraminifera (20), and so we retained deep-water sites for this work.

We extracted MAT values from AOGCM raster cells containing foraminiferal occurrences. If the MAT value of a cell was NA (because the cell partially fell on ice or land), we averaged the values of the nine surrounding cells. Finally, we converted presence, absence, and count data to presence-only occurrences for individual species. Absence data were not used subsequently except to estimate distributions of sampling effort (i.e., sampled core locations as in SI Appendix Fig. S6 and Movie S1). Occurrences for a species in a time bin were included only if they totaled six or more, a minimum cutoff for robust probability density estimation. Similarly,

any species not observed to cross at least six consecutive time bin boundaries was excluded; this cutoff was necessary to ensure sufficient per-species sample size to fit trait evolution models, to correlate time series, and to obtain mean H distances for phylogenetic ANOVA. These sample size constraints reduced the scope of the dataset from 58 to 24 planktonic species. Therefore, results might be more representative of species that are widespread and abundant than those that are restricted and rare.

### Spatiotemporal resolution

Occurrence records were binned into 88 intervals of 8 ka, from the recent sub-fossil record to 700 ka ago. The chosen bin length reflects the resolution of the AOGCM outputs (4 ka), the time-averaging of fossil assemblages and imprecision in fossil ages (up to a few thousand years (67)), and the rate of climatic change (fluctuation from glacial minimum to interglacial maximum approximately every 50 ka). To minimize error in bin assignments due to age uncertainty, records were omitted if age estimates were derived from foraminiferal zones or had confidence intervals longer than 2 ka; remaining records could be assigned confidently to a single time bin based on the mean age estimate. Occurrences were paired with AOGCM data from the midpoint of time bins (e.g., climate was modeled at 12 ka ago for occurrences 16- to 8-ka old).

We rasterized point coordinates to a 1.25° latitude-longitude grid and shortened the dataset to unique combinations of species, grid cells, and time bins. This spatial resolution matched the scale of the AOGCM reconstructions and is appropriate for the magnitude of error in occurrence coordinates. Spatial imprecision could occur from ocean currents displacing foraminiferal tests as the dead organisms sank to the seafloor. For sites at < 1 km depth or for large foraminifera, the distance between where the organism died and where it settled may be on

the scale of a single grid cell (68, 69). For depths of 2–3.5 km, a maximum displacement of 100–400 km is reasonable, corresponding to an error of up to 3 grid cells (70, 71). Because MAT values are generally quite similar in adjacent cells (i.e., spatially autocorrelated; Fig. 1, SI Appendix Movie S1), even the largest transport of dead foraminifera to the ocean floor would affect niche estimates undetectably.

### Species attributes

Most living species of foraminifera are benthic, but the ~50 extant planktonic species are an abundant constituent of the zooplankton at the base of the marine food web. Until recently, all planktonic species were thought to reproduce exclusively sexually, with synchronous release of gametes at the mixed-layer depth of the water column (27). Two research teams have now reported asexual reproduction in planktonic species, but the frequency of asexual reproduction in natural conditions is unknown (72, 73). The diversity of ecological strategies in foraminifera includes autotrophy (via association with algal symbionts), herbivory, omnivory, and carnivory (27). In this study, each species was classed into one of five ecotypes (SI Appendix Table S2) following Aze and others (26): (1) tropical/subtropical, mixed layer, bearing symbionts, (2) tropical/subtropical, mixed layer, without symbionts, (3) open ocean, thermocline, (4) open ocean, sub-thermocline, or (5) high latitude. The full Cenozoic foraminifera database also contained a category for upwelling/high productivity (37), but no species in this group occurred in the subset of data used here. Ecotype data were used during phylogenetic analysis of variance in niche lability among species.

Foraminifera species inhabit specific and consistent vertical ranges and do not participate in vertical diel migration as do some other plankton, although these protists do migrate to the mixed layer depth to breed (74). Therefore, we performed each analysis two ways: using MAT at

the sea surface to estimate niche differences, or using MAT extracted from each focal species' depth range (a "depth-habitat approach"). In analyses based on the depth-habitat approach, each species' thermal occupancy was then compared against global ocean temperature from the corresponding depth layer. We searched the literature for quantitative information on each species' modern depth range to assign taxa to one of three depth habitats: surface (40 m in the AOGCM), surface-subsurface (78 m), or subsurface (164 m). SI Appendix Table S2 lists the depth habitat and reference(s) for each species.

Habitat depth estimates are complicated by the fact that vertical structure of the thermocline, halocline, and photic gradient can shift with latitude, longitude, and time. For instance, a surface-dwelling species might live at 30 m in temperate zones but 10 m in equatorial zones. Our analyses simplified this complexity by using a constant depth point to define each habitat zone. Another assumption of the depth-habitat approach is that species' depth ranges have been conserved throughout the study interval. Since the goal of the study was to quantify niche lability, there is danger of theoretical tautology: assuming stability of depth niches to test the stability of thermal niches. Nonetheless, living depth relates intimately to fundamental aspects of foraminiferal autecology, such as whether a species is constrained to the photic zone by its symbionts, and so depth conservatism on the study timescale is probable. On one hand, the depth-habitat approach incorporates biological complexity into consideration, and so results might be more accurate than those based on sea surface data. On the other hand, due to the assumptions of the depth-habitat approach, results from sea surface data might be more precise than those from the depth-habitat approach. Ultimately, the differences between the two methods were small, and temperature values at the sea surface correlated tightly with those from habitat

zones (Pearson's  $r > 0.95$ ). Previous studies of foraminiferal ecology also recovered equivalent results whether analysis was based on data from the sea surface or at depth (32).

### Kernel density estimation

We used kernel density estimation (KDE) to construct niche distributions. KDE generates a continuous probability density function (PDF) from a finite set of observations along one or more independent axes. Unlike a probability mass function for discrete data, the PDF at a given point does not represent the probability of an event (e.g., species occurrence) at that point; rather, the integral of the PDF over a defined interval is the probability of an event occurring within those bounds. KDE is a cornerstone of nonparametric statistics with a robust mathematical foundation and applications in many fields, including ecology (38, 75-77). KDE can be conducted in any number of dimensions, but every additional dimension requires exponential increases in sample size to maintain low mean square error (75). We chose to prioritize high accuracy of niche estimates (low error) and fine temporal resolution (dividing available data into many time bins) at the expense of a multivariate response (e.g., both MAT and salinity).

One of the difficulties of using KDE with ecological data is the prevalence of non-uniform sampling along environmental gradients. In these instances, one must account for the sampling distribution to make fair and accurate inferences about niches (38, 39). This problem is often called “length-bias” in the statistical literature, for the common case where the probability of observing an event scales linearly with the independent variable (e.g., the chance of identifying an animal increases directly with group size). When using KDE to construct abiotic niches, non-linear bias could arise from the unequal land cover of different biomes. For instance, because the curvature of the Earth and ocean-atmosphere circulation patterns lead to larger surface areas of temperate than polar habitats, the probability of observing a species in a

temperate region of niche space may be higher than that of observing the species in a polar region.

We used the **kerneval** package in R to construct niches for each species' occupied MAT in each time bin (version 0.1.2, available at <https://github.com/GawainAntell/kerneval>). This tool can correct for uneven sampling using the method of Barmi and Simonoff (78). First, we estimated the PDF of sampling probability along the MAT axis, based on the MAT values from all sediment core sites in each time bin. Each core is expected to contain a nearly continuous sedimentary record of ocean conditions and fauna, and so we interpolated sampling locations for cores in the middle of their temporal duration for which no records were otherwise present in the database (SI Appendix Fig. S6A–D). These “pseudo-absences” were used only to estimate the sampling density when standardizing niche estimates; no species presence points were added. Next, the MAT axis for each species was transformed by the inverse of the cumulative distribution function of the sampling probability. We performed KDE for each species in the transformed space and back-transformed the resultant PDF to interpret the niche on the original axis scale.

Choice of bandwidth is a critical step in KDE (75), and so we explored densities with rule-of-thumb (“nrd0”), cross-validation (“CV”), and Sheather and Jones solve-the-equation (“SJ-ste”) (79) methods of bandwidth selection. The rule of thumb over-smoothed densities to the point of obscuring differences between species and intervals, while CV generated unrealistically rough PDFs. We choose the SJ-ste method for its middle-ground behavior, although this too could be criticized for under-smoothing densities somewhat (SI Appendix Fig. S1).

## Niche dissimilarity measure

We used Hellinger's H distance to quantify the similarity of niche distributions between time bin pairs (Fig. 1, SI Appendix Fig. S2). This measure has a long history in the statistics literature, treats the two entities of comparison symmetrically, can be adapted for discrete or multidimensional data, and behaves well even when distributions are highly peaked or contain stretches of zero probability. H ranges from 0 to 1, with 0 indicating identical distributions; this is the opposite interpretation of Schoener's D, which is a similarity measure rather than a distance measure.

To compare PDFs in a fair way, the sampling universe (observation probability) of each distribution must be the same. In the context of niche overlap calculations, two biological entities must be able to access the same region of niche space for the niche comparisons to be fair (10). If the sampling universe is not shared, one might infer an entity was absent from a region of niche space where another entity was present when, in reality, the region was only available to the latter species. For instance, a species might occupy extremely cold regions during a glacial period but not during an interglacial period merely because there were no equally-cold locations on Earth during the interglacial. Without accounting for this difference in availability of environment, one would calculate a larger niche discrepancy than would be justified.

One could standardize niche space either prior to or during niche construction. In either case, one must determine the segment of the niche axis accessible to species in every time bin. The upper and lower limits of the comparison interval are listed in SI Appendix Fig. S5 for every depth habitat and plotted as dashed lines on niche distributions in Fig. 1 and SI Appendix Fig. S1. If we were to standardize niche space *a priori*, we would have had to discard any fossil occurrences associated with temperatures outside the interval of comparison (i.e., data points

outside the dashed lines in Fig. 1 and SI Appendix Fig. S1). This truncation would remove 19% of presences, almost all of which would exceed the upper limit. Instead, we used all presence points to estimate niche densities but then trimmed the density estimates to a standard MAT interval. We rescaled density functions over this interval so the probability integrated to unity. Since the standardization took place as the last step of KDE rather than on raw data, time series used in trait evolution models (i.e., the mean and standard error of occupied MAT values for each species) were unaffected.

SI Appendix Fig. S6 outlines the process of calculating the standard temperature interval for niche comparisons. As when estimating the sampling density for niche estimation (“Kernel density estimation” section above), we interpolated gaps in sediment core records with a “range-through” approach (SI Appendix Fig. S6A–D). We extracted MAT values from AOGCM cells corresponding to core sites in each time bin (SI Appendix Fig. S6E). We determined the sampled temperature range in each time bin (SI Appendix Movie S1) and then defined the standard interval for niche comparisons as the temperature range accessible to species in all time bins (SI Appendix Fig. S6F).

For some niche estimation problems, it may be desirable to reflect data across estimation bounds so that the density does not drop towards zero as it approaches those bounds. For example, if the range of sampled values is much narrower than the suspected true breadth of the distribution, then boundary reflection could improve KDE accuracy. We did not reflect boundaries during KDE, however, because the temperature bounds of niche comparison were close to the physiological tolerances of foraminifera. The lower bound for sea-surface data ( $-1.4^{\circ}\text{C}$ ) approaches the freezing point of sea ice ( $-2^{\circ}\text{C}$ ). The upper bound ( $27^{\circ}\text{C}$ ) is slightly lower than the maximum thermal tolerance of foraminifera,  $30\text{--}32^{\circ}\text{C}$  (29). However, under default

KDE settings a distribution generally does not fall to zero exactly at each bound, but rather tapers towards zero at a point three times the bandwidth farther away. Thus, for our data KDE distributions without boundary reflection match closely the laboratory evidence for foraminiferal thermal limits. In addition, 27 °C is the point at which the tight linear relationship between sea surface temperature and thermocline structure begins to break down. Comparing foraminiferal presences to global temperature in a linear way is justified only up to this point (30).

## ANALYSES

### Time series correlation

The first analysis correlated time series of species' thermal optima and global temperature. We constructed separate time series for each species, imposing a minimum length requirement of seven time bins. If a species was sampled in multiple, discontinuous sets of at least seven consecutive bins, then each segment was analyzed separately to match the methods used in the trait evolution models (described below). For each time step in a species' series, we used KDE to define preferred or optimal temperature as the MAT value associated with the PDF maximum. We then correlated each time series of thermal optima with concurrent estimates of mean global MAT. We report the mean and standard deviation of correlation coefficients (Pearson's  $r$ ) across all time series ( $n = 32$ ). A strongly positive correlation coefficient could indicate niche shifts in adjustment to available environment ("Definition of 'available environment'" section above). In the depth-habitat version of time series analysis, MAT across global marine points was extracted from the same depth layer as was preferred by each species.

### Niche lability vs. climate change magnitude

We calculated niche dissimilarity (Hellinger's H distance) for each species across each boundary it crossed. We then averaged H distances among all species that crossed a given boundary, for each of the 87 time bin boundaries. Ultimately, the goal was to relate these niche lability estimates to climate change magnitude, which we defined as the absolute value of the difference in mean global MAT in each pair of consecutive bins. If climate change caused species to alter their niches, one would expect larger niche differences with larger magnitude of environmental change.

We correlated niche dissimilarity and climate change magnitude both with and without removing temporal autocorrelation ("pre-whitening"). First we checked for autocorrelation: an augmented Dickey-Fuller test indicated the time series of climate change magnitudes was stationary but the series of mean H distances was not. Therefore, a first-order autoregressive time series model was fit to the H series; the estimated coefficient of autocorrelation was 0.49. In the case of pre-whitening the data to remove autocorrelation, we used the series of residuals from the autoregressive model in place of the original H values as a correlate to climate change magnitude. Temporal autocorrelation was negligible in the time series of H model residuals and the time series of climate change magnitude, and in the residuals of a linear model of the former on the latter. This last result confirmed that no further correction was needed to account for pseudoreplication in the correlation of interest.

### Niche lability among extreme climate intervals

We calculated Hellinger's H distances not only across consecutive time bins (described in the previous section), but also in pairs of discontinuous time bins representing glacial maxima and minima. We determined the timing of these climatic extremes based on the average global

marine temperature curve (Fig. 3A, SI Appendix Fig. S8 and Table S3). The timings derived from the AOGCM data agreed closely with interglacial onsets reported in the literature (80). For each pair of time bins (e.g. the glacial maximum 28 ka ago and the interglacial peak 212 ka ago, shown in Fig. 1), we took the mean H distance among all species observed in both bins. We used boxplots to summarize variation among bin pairs in each of three climate categories: glacial–glacial, interglacial–interglacial, and glacial–interglacial intervals (Fig. 3, SI Appendix Fig. S9). The mean global temperature difference for time bin pairs in each category was 0.23 °C, 0.14 °C, and 1.99 °C, respectively.

We conducted a variation on the extreme climate comparisons analysis to test sensitivity to sample size deviations through time. One of the benefits of KDE is its invariance to small deviations in sample size; however, the number of observations influences bandwidth calculation to a small degree. The estimate of a complicated distribution will be somewhat rougher with denser sampling than with sparse sampling. This responsiveness to sample size is considered a desirable statistical quality in theory, because it means that estimates will be perfect representations of distributions as sample size goes to infinity.

Unfortunately, in the empirical case of a comparison of foraminifera niches in the most recent interglacial (4 ka ago) against any older interval, the difference in sample size was up to two orders of magnitude: of the 42,233 occurrences, 19,711 (47%) were from the 4-ka-ago time step, 2,079 were from 12 ka ago, and the rest are spread nearly evenly across older time steps. The density estimates for species at 4 ka ago are noticeably rougher than in much older intervals (SI Appendix Fig. S1), which could inflate the H distance between densities. In the 86 time bins older than 12 ka, the correlation between sample size and bandwidth estimate was negligible across density estimates (Pearson's  $r = 0.040$ , 95% CI = [-0.0087, 0.089]), and the sample size

difference between any consecutive pair of time bins was small enough to be safely ignored in time series and bin-to-bin niche lability analyses. Among glacial intervals, the oldest (668 ka) was sampled least and therefore could influence H distances with other bins. Therefore, we repeated the analysis of niche lability among climate extremes after excluding these two temporal endpoints of glacial–interglacial cycles from comparison (Fig. 3C, SI Appendix Fig. S9B).

### Test for type II error

We simulated fossil occurrences through time to test for the possibility that the empirical analyses were unable to detect niche changes that occurred (i.e., the possibility of falsely detecting negative signals of niche change, a type II error). In each simulation, species' occurrence sample sizes and the magnitude of climate change were kept the same as those in the empirical data. We used the geographic occurrences of species at 12 ka ago as a reference for species' "true" distributions and assumed geographic ranges were static (and hence niches adapted to local environmental changes) through time. We chose the time step at 12 ka since global temperatures were close to the long-term mean (Fig. 3A) and sampling was dense throughout species' geographic distributions.

We generated occurrence data for species in glacial and interglacial intervals (listed in Table S3), excluding the youngest and oldest extremes (4 ka and 668 ka) as in the niche lability analysis presented in Fig. 3C and SI Appendix Fig. S9B. For each interval and focal species, we randomly drew the observed number of occurrences for that interval and species, sampling without replacement from the species' occurrences at 12 ka. Then, we extracted the sea surface MAT at the simulated presences based on AOGCM data for the relevant time step. We used the simulated data to run the "Niche lability among extreme climate intervals" analysis as described

in *Methods* to determine whether mean H distance was larger between pairs of glacial and interglacial intervals than either (1) pairs of glacial intervals or (2) pairs of interglacial intervals. The entire simulation and comparison procedure was repeated 100 times. Boxplot comparisons of results from the first five replicates are illustrated in Fig. S10A–E.

The simulation framework considers scenarios of complete niche adaptation, but species could respond in many ways beyond a binary of either niche stasis (and geographic habitat tracking) or niche adaptation (and geographic stasis). For instance, species could expand their niche breadth to both occupy the existing geographic range with new environmental conditions *and* track preferred habitat into new geographic regions. Species could also shrink in realized niche space by occupying only the subset of the original range that remained within tolerance limits. Either of these alternatives would manifest as a difference in the response of the leading and lagging geographic range edge. The simulations conducted here represent a scenario of “lock-step” changes in range edges, which is the most common empirical pattern (47) and allowed us to test whether strong and consistent lability would be detectable with our data and methods.

### Trait evolution models

We fit five competing models of trait evolution (described below) to time series of each species’ MAT occurrences to characterize niche change as static, white noise, wandering, directional, or related to global temperature. We used Akaike weights to compare the five models for each series and defined the “best fit” as the mode of trait change receiving the most support (Fig. 4, SI Appendix Fig. S11). Parameters were estimated with a maximum likelihood approach (81, 82) implemented in the **palaeoTS** package, version 0.5.2 (83). The functions in this package take as arguments the mean, variance, and sample size of a trait at each time step. These

summary statistics must be calculated from the original, finite set of MAT values associated with species' occurrences; inputting continuous PDFs from KDE into trait models would require substantial rewrites of the paleoTS code. Although optimum temperature (from KDE) arguably has more direct biological meaning than mean temperature among a focal species' occupied sites, we chose to follow precedent and use mean values with existing code. In practice, the difference between these two metrics is minor; the Pearson correlation coefficient between mean and optimum MAT across all species-bin combinations was 0.85 (95% CI = [0.83, 0.86]) at the sea-surface.

The **paleoTS** models allow several user-specified arguments. We chose not to pool the variance from time steps in a series, since some species' sample variances were unequal. The singular exception was the covariate-tracking model for *Globigerinoides conglobatus* under the depth habitat approach (marked with an asterisk in SI Appendix Fig. S11), for which we pooled variances—the specific trait and covariate values for this series by chance resulted in an otherwise unsolvable problem for the default L-BFGS-B optimization method. For all series we specified a restricted maximum likelihood approach (“ancestor–descendant”) instead of a full likelihood approach (“joint”), which treats a series as a joint sample from a multivariate normal distribution. Although the joint approach is more sensitive to directional trends in noisy data (84), package documentation warns that ancestor–descendant parameterization is “generally safer” for covariate-tracking models because it avoids issues of induced autocorrelation (83).

The modeled evolutionary modes were: (1) strict stasis, (2) stasis with white noise, (3) unbiased random walk, (4) generalized (i.e., directional) random walk, and (5) tracking of global temperature as a covariate. A strict stasis model assumes that the mean of a trait is constant through time, whereas a general stasis model assumes that the observed mean in each step is

drawn from a distribution with a constant mean but positive variance. In other words, in a model of stasis *sensu latu*, observed trait means tend to fluctuate about a constant underlying mean. An unbiased random walk allows the mean in each time bin to move away from the mean in the previous bin by a step length drawn from a distribution with mean of zero and a positive variance; observed trait means migrate from the initial value in a random direction. A general random walk modifies the distribution for the step length such that the mean is constant and non-zero; observed trait means tend to increase or decrease depending on the sign of the step length mean.

For the covariate-tracking model, we identified the series of global mean temperatures during time bins corresponding to the focal species' series. In the depth-habitat approach, we extracted global temperature values at the preferred depth of the focal species in the respective time bins. The model then estimated a parameter for the correlation between the trait series (species' occupied temperature) and the covariate series (global temperature). Hence, the covariate-tracking model was akin to the "Time series correlation" analysis, which found species' preferred temperature to be independent of mean global temperature. Based on this earlier result, the covariate-tracking model might be expected to receive poor support. However, by including covariate-tracking in the trait evolution models analysis irrespective of earlier results, we could quantify the relative amount by which other models of niche change were more favored.

It is possible to fit more complicated models than the five considered here. We refrained from fitting punctuations and other complex model types on the grounds of both statistical rigor and relevance to our research questions. Our primary focus was niche change magnitude and covariation with climate, which the five simple trait models already quantify. Incidentally, there

is a known bias towards selecting a complicated model when time series are long (85), and the length of time series varied among species in our empirical dataset. If we had included complicated models in comparisons, sampling inconsistencies might have led simple models to be favored more often for rare species and complicated models to be favored more often for common ones.

### Phylogenetic ecotype comparisons

To explore interspecific variation in niche lability, we calculated the mean H across all consecutive bin pairs in which a focal species was observed. We then merged the species-level dataset with the bifurcating morphospecies phylogeny of macroperforate planktonic foraminifera (SI Appendix Fig. S12) from Aze and others (26), available from the **timetree** R package, version 3.3.25 (86). Two species were missing from the phylogeny and so were excluded from ecotype comparisons: *Globigerinita glutinata* and *Neogloboquadrina incompta*. We performed phylogenetic ANOVA on H values by ecotype (detailed in the “Species attributes” section above), following the method of Garland and others (87) as implemented in the R package **geiger**, version 2.0.7 (88). In this approach, the test statistic is calculated in the standard way for ANOVA, while the null distribution is generated by simulating data along the tree. We set the number of simulations at 10,000 to calculate Wilks test values. The evolutionary rate was estimated from the mean squared independent contrasts under Brownian motion.

### Data and code availability

All data and code necessary to run analyses and produce final results have been deposited on the online data archive Zenodo (DOI: [10.5281/zenodo.4658884](https://doi.org/10.5281/zenodo.4658884)). The **kerneval** software package can be installed from <https://github.com/GawainAntell/kerneval>. Analyses were conducted in an R statistical computing environment, version 4.0.3 (89).

## REFERENCES

1. Wiens JJ & Graham CH (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* 36:519-539.
2. Gómez-Rodríguez C, Baselga A, & Wiens JJ (2015) Is diversification rate related to climatic niche width? *Global Ecol. Biogeogr.* 24(4):383-395.
3. Castro-Insua A, Gómez-Rodríguez C, Wiens JJ, & Baselga A (2018) Climatic niche divergence drives patterns of diversification and richness among mammal families. *Scientific Reports* 8(1):1-12.
4. Pyron RA, Costa GC, Patten MA, & Burbrink FT (2015) Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biol. Rev. Camb. Philos. Soc.* 90(4):1248-1262.
5. Stigall AL (2014) When and how do species achieve niche stability over long time scales? *Ecography* 37(11):1123-1132.
6. Velasco JA, Martinez-Meyer E, & Flores-Villela O (2018) Climatic niche dynamics and its role in the insular endemism of *Anolis* lizards. *Evolutionary Biology* 45(3):345-357.
7. Folk RA, *et al.* (2019) Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proc. Natl. Acad. Sci. U.S.A.* 116(22):10874-10882.
8. Lavergne S, Evans ME, Burfield IJ, Jiguet F, & Thuiller W (2013) Are species' responses to global change predicted by past niche evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1610):20120091.
9. Munday PL, Warner RR, Monro K, Pandolfi JM, & Marshall DJ (2013) Predicting evolutionary responses to climate change in the sea. *Ecol. Lett.* 16(12):1488-1500.

10. Peterson AT (2011) Ecological niche conservatism: a time-structured review of evidence. *J. Biogeogr.* 38(5):817-827.
11. Holland SM & Zaffos A (2011) Niche conservatism along an onshore-offshore gradient. *Paleobiology* 37(2):270-286.
12. Saupe E, *et al.* (2014) Macroevolutionary consequences of profound climate change on niche evolution in marine molluscs over the past three million years. *Proc. R. Soc. B* 281(1795):20141995.
13. Banks WE, *et al.* (2008) Neanderthal extinction by competitive exclusion. *PLoS One* 3(12):e3972.
14. Pearman PB, *et al.* (2008) Prediction of plant species distributions across six millennia. *Ecol. Lett.* 11(4):357-369.
15. Record S, Fitzpatrick MC, Finley AO, Veloz S, & Ellison AM (2013) Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change. *Global Ecol. Biogeogr.* 22(6):760-771.
16. Martínez-Meyer E, Townsend Peterson A, & Hargrove WW (2004) Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecol. Biogeogr.* 13(4):305-314.
17. Martínez-Meyer E & Peterson AT (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *J. Biogeogr.* 33(10):1779-1789.
18. Banks WE, d'Errico F, Peterson AT, Kageyama M, & Colombeau G (2008) Reconstructing ecological niches and geographic distributions of caribou (*Rangifer*

- tarandus) and red deer (*Cervus elaphus*) during the Last Glacial Maximum. *Quaternary Science Reviews* 27(27-28):2568-2575.
19. Waltari E & Guralnick RP (2009) Ecological niche modelling of montane mammals in the Great Basin, North America: examining past and present connectivity of species across basins and ranges. *J. Biogeogr.* 36(1):148-161.
  20. Waterson A, Edgar K, Schmidt DN, & Valdes PJ (2017) Quantifying the stability of planktic foraminiferal physical niches between the Holocene and Last Glacial Maximum. *Paleoceanography* 32(1):74-89.
  21. Hutchins DA & Fu F (2017) Microorganisms and ocean global change. *Nature Microbiology* 2(6):17058.
  22. Kucera M (2007) Planktonic foraminifera as tracers of past oceanic environments, Chapter 6. *Developments in marine geology. Elsevier, Oxford*:213-262.
  23. Kucera M, *et al.* (2005) Reconstruction of sea-surface temperatures from assemblages of planktonic foraminifera: multi-technique approach based on geographically constrained calibration data sets and its application to glacial Atlantic and Pacific Oceans. *Quaternary Science Reviews* 24(7-9):951-998.
  24. Ezard TH, Aze T, Pearson PN, & Purvis A (2011) Interplay between changing climate and species' ecology drives macroevolutionary dynamics. *Science* 332(6027):349-351.
  25. Morard R, *et al.* (2015) PFR2: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. *Molecular Ecology Resources* 15(6):1472-1485.
  26. Aze T, *et al.* (2011) A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biol. Rev. Camb. Philos. Soc.* 86(4):900-927.

27. Schiebel R & Hemleben C (2017) *Planktic foraminifers in the modern ocean* (Springer).
28. Rillo MC, *et al.* (2019) On the mismatch in the strength of competition among fossil and modern species of planktonic Foraminifera. *Global Ecol. Biogeogr.*
29. Bijma J, Faber WW, & Hemleben C (1990) Temperature and salinity limits for growth and survival of some planktonic foraminifers in laboratory cultures. *The Journal of Foraminiferal Research* 20(2):95-116.
30. Rutherford S, D'Hondt S, & Prell W (1999) Environmental controls on the geographic distribution of zooplankton diversity. *Nature* 400(6746):749-753.
31. Fenton IS, Pearson PN, Dunkley Jones T, & Purvis A (2016) Environmental predictors of diversity in recent planktonic foraminifera as recorded in marine sediments. *PloS one* 11(11):e0165522.
32. Richter DJ, *et al.* (2019) Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *bioRxiv*:867739.
33. Reddin CJ, Nätscher PS, Kocsis ÁT, Pörtner H-O, & Kiessling W (2020) Marine clade sensitivities to climate change conform across timescales. *Nature Climate Change* 10(3):249-253.
34. Mathes GH, van Dijk J, Kiessling W, & Steinbauer MJ (2021) Extinction risk controlled by interaction of long-term and short-term climate change. *Nature Ecology & Evolution.*
35. Ward BA, Cael B, Collins S, & Young CR (2021) Selective constraints on global plankton dispersal. *Proc. Natl. Acad. Sci. U.S.A.* 118(10).
36. Brombacher A, Wilson PA, Bailey I, & Ezard TH (2018) Temperature is a poor proxy for synergistic climate forcing of plankton evolution. *Proceedings of the Royal Society B: Biological Sciences* 285(1883):20180665.

37. Fenton IS, *et al.* (in review) Triton: a new database of Cenozoic planktonic foraminifera. *Scientific Data*.
38. Broennimann O, *et al.* (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecol. Biogeogr.* 21(4):481-497.
39. Godsoe W & Case BS (2015) Accounting for shifts in the frequency of suitable environments when testing for niche overlap. *Methods Ecol. Evol.* 6(1):59-66.
40. Hannisdal B, Haaga KA, Reitan T, Diego D, & Liow LH (2017) Common species link global ecosystems to climate change: dynamical evidence in the planktonic fossil record. *Proceedings of the Royal Society B: Biological Sciences* 284(1858):20170722.
41. Irwin AJ, Finkel ZV, Müller-Karger FE, & Ghinaglia LT (2015) Phytoplankton adapt to changing ocean environments. *Proc. Natl. Acad. Sci. U.S.A.* 112(18):5762-5766.
42. Donelson JM, *et al.* (2019) Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change. *Philosophical Transactions of the Royal Society B* 374(1768):20180186.
43. Bridle JR & Vines TH (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends Ecol. Evol.* 22(3):140-147.
44. Polechová J & Barton NH (2015) Limits to adaptation along environmental gradients. *Proc. Natl. Acad. Sci. U.S.A.* 112(20):6401-6406.
45. Bridle JR, Polechová J, Kawata M, & Butlin RK (2010) Why is adaptation prevented at ecological margins? New insights from individual-based simulations. *Ecol. Lett.* 13(4):485-494.
46. Burrows MT, *et al.* (2011) The pace of shifting climate in marine and terrestrial ecosystems. *Science* 334(6056):652-655.

47. Lenoir J, *et al.* (2020) Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution*:1-16.
48. Jackson JB (2008) Ecological extinction and evolution in the brave new ocean. *Proc. Natl. Acad. Sci. U.S.A.* 105(Supplement 1):11458-11465.
49. Dutkiewicz S, *et al.* (2015) Impact of ocean acidification on the structure of future phytoplankton communities. *Nature Climate Change* 5(11):1002-1006.
50. Cahill AE, *et al.* (2013) How does climate change cause extinction? *Proceedings of the Royal Society B: Biological Sciences* 280(1750):20121890.
51. CLIMAP Project (1981) *Seasonal reconstructions of the Earth's surface at the last glacial maximum* (Geological Society of America).
52. Gould SJ & Eldredge N (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology*:115-151.
53. Cox PM (2001) Description of the "TRIFFID" dynamic global vegetation model. (Hadley Centre).
54. Valdes PJ, *et al.* (2017) The BRIDGE HadCM3 family of climate models: HadCM@Bristol v1. 0. *Geoscientific Model Development* 10:3715-3743.
55. Davies-Barnard T, Ridgwell A, Singarayer J, & Valdes P (2017) Quantifying the influence of the terrestrial biosphere on glacial–interglacial climate dynamics. *Climate of the Past* 13(10):1381-1401.
56. De Boer B, Van de Wal R, Lourens L, Bintanja R, & Reerink T (2013) A continuous simulation of global ice volume over the past 1 million years with 3-D ice-sheet models. *Climate Dynamics* 41(5-6):1365-1384.

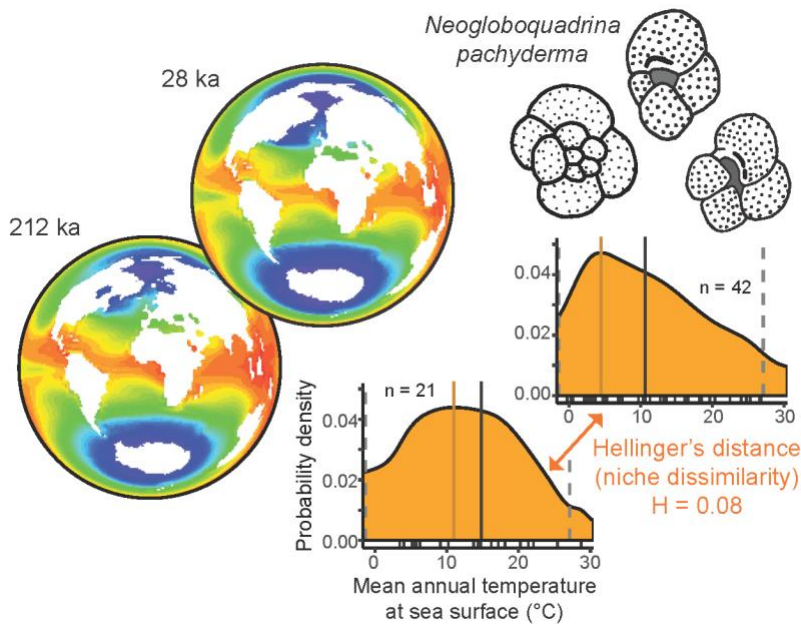
57. De Boer B, Lourens LJ, & Van De Wal RS (2014) Persistent 400,000-year variability of Antarctic ice volume and the carbon cycle is revealed throughout the Plio-Pleistocene. *Nat. Commun.* 5:2999.
58. Peltier WR, Argus D, & Drummond R (2015) Space geodesy constrains ice age terminal deglaciation: The global ICE-6G\_C (VM5a) model. *Journal of Geophysical Research: Solid Earth* 120(1):450-487.
59. Tarasov L & Richard Peltier W (2002) Greenland glacial history and local geodynamic consequences. *Geophysical Journal International* 150(1):198-229.
60. Tarasov L, Dyke AS, Neal RM, & Peltier WR (2012) A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling. *Earth and Planetary Science Letters* 315:30-40.
61. Briggs RD, Pollard D, & Tarasov L (2014) A data-constrained large ensemble analysis of Antarctic evolution since the Eemian. *Quaternary Science Reviews* 103:91-115.
62. Tarasov L, *et al.* (2014) The global GLAC-1c deglaciation chronology, meltwater pulse 1-a, and a question of missing ice. *IGS Symposium: Contribution of Glaciers and Ice Sheets to Sea-Level Change*, pp 26-30.
63. Siccha M & Kucera M (2017) ForCenS, a curated database of planktonic foraminifera census counts in marine surface sediment samples. *Scientific Data* 4:170109.
64. Jonkers L, Hillebrand H, & Kucera M (2019) Global change drives modern plankton communities away from the pre-industrial state. *Nature* 570(7761):372-375.
65. Matthews KJ, *et al.* (2016) Global plate boundary evolution and kinematics since the late Paleozoic. *Global Planet. Change* 146:226-250.

66. Darling KF, Kucera M, & Wade CM (2007) Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proc. Natl. Acad. Sci. U.S.A.* 104(12):5002-5007.
67. Martin RE (1999) Taphonomy and temporal resolution of foraminiferal assemblages. *Modern foraminifera*, (Springer), pp 281-298.
68. Gyldenfeldt A-Bv, Carstens J, & Meincke J (2000) Estimation of the catchment area of a sediment trap by means of current meters and foraminiferal tests. *Deep Sea Research Part II: Topical Studies in Oceanography* 47(9-11):1701-1717.
69. Qiu Z, Doglioli A, & Carlotti F (2014) Using a Lagrangian model to estimate source regions of particles in sediment traps. *Science China Earth Sciences* 57(10):2447-2456.
70. Siegel D & Deuser W (1997) 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(9-10):1519-1541.
71. Waniek J, Koeve W, & Prien RD (2000) Trajectories of sinking particles and the catchment areas above sediment traps in the northeast Atlantic. *J. Mar. Res.* 58(6):983-1006.
72. Takagi H, Kurasawa A, & Kimoto K (2020) Observation of asexual reproduction with symbiont transmission in planktonic foraminifera. *J. Plankton Res.*
73. Davis CV, *et al.* (2020) Extensive morphological variability in asexually produced planktic foraminifera. *Science Advances* 6(28):eabb8930.
74. Meilland J, *et al.* (2019) Highly replicated sampling reveals no diurnal vertical migration but stable species-specific vertical habitats in planktonic foraminifera. *J. Plankton Res.* 41(2):127-141.

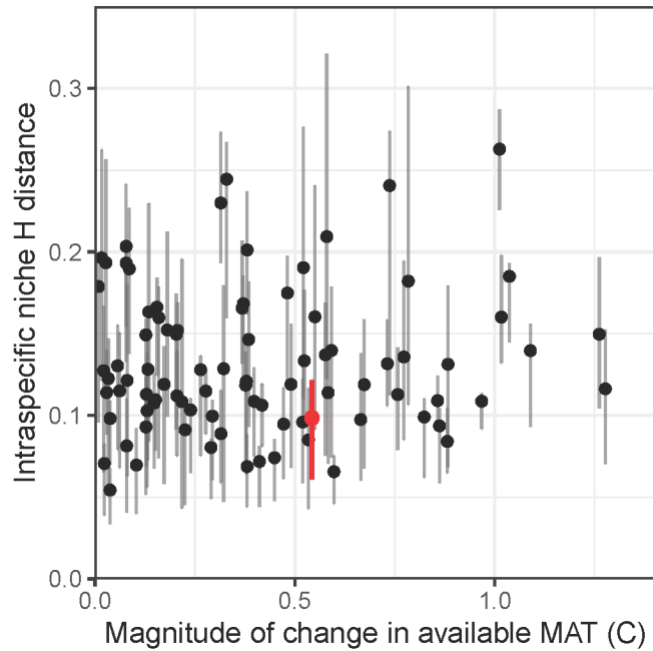
75. Silverman BW (1986) *Density Estimation for Statistics and Data Analysis* (Chapman and Hall Ltd, Bristol) p 175.
76. Worton BJ (1989) Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70(1):164-168.
77. Horne JS, Garton EO, & Sager-Fradkin KA (2007) Correcting home-range models for observation bias. *The Journal of Wildlife Management* 71(3):996-1001.
78. Barmi HE & Simonoff JS (2000) Transformation-based density estimation for weighted distributions. *Journal of nonparametric statistics* 12(6):861-878.
79. Sheather SJ & Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)* 53(3):683-690.
80. Lisiecki LE & Raymo ME (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* 20(1).
81. Hunt G (2006) Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* 32(4):578-601.
82. Hunt G (2012) Measuring rates of phenotypic evolution and the inseparability of tempo and modemeasuring rates of phenotypic evolution. *Paleobiology* 38(3):351-373.
83. Hunt G (2019) paleoTS: Analyze paleontological time-series), R package version 0.5.2.
84. Hunt G (2008) Evolutionary patterns within fossil lineages: Model-based assessment of modes, rates, punctuations and process. *From Evolution to Geobiology: Research Questions Driving Paleontology at the Start of a New Century*, The Paleontological Society Papers, eds Bambach RK & Kelley P (The Paleontological Society, New Haven, CT), pp 117–131.

85. Hunt G, Hopkins MJ, & Lidgard S (2015) Simple versus complex models of trait evolution and stasis as a response to environmental change. *Proc. Natl. Acad. Sci. U.S.A.* 112(16):4885-4890.
86. Bapst DW (2012) paleotree: an R package for paleontological and phylogenetic analyses of evolution. *Methods Ecol. Evol.* 3(5):803-807.
87. Garland Jr T, Dickerman AW, Janis CM, & Jones JA (1993) Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* 42(3):265-292.
88. Harmon LJ, Weir JT, Brock CD, Glor RE, & Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics* 24(1):129-131.
89. R Core Team (2020) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria), Version 4.0.3.

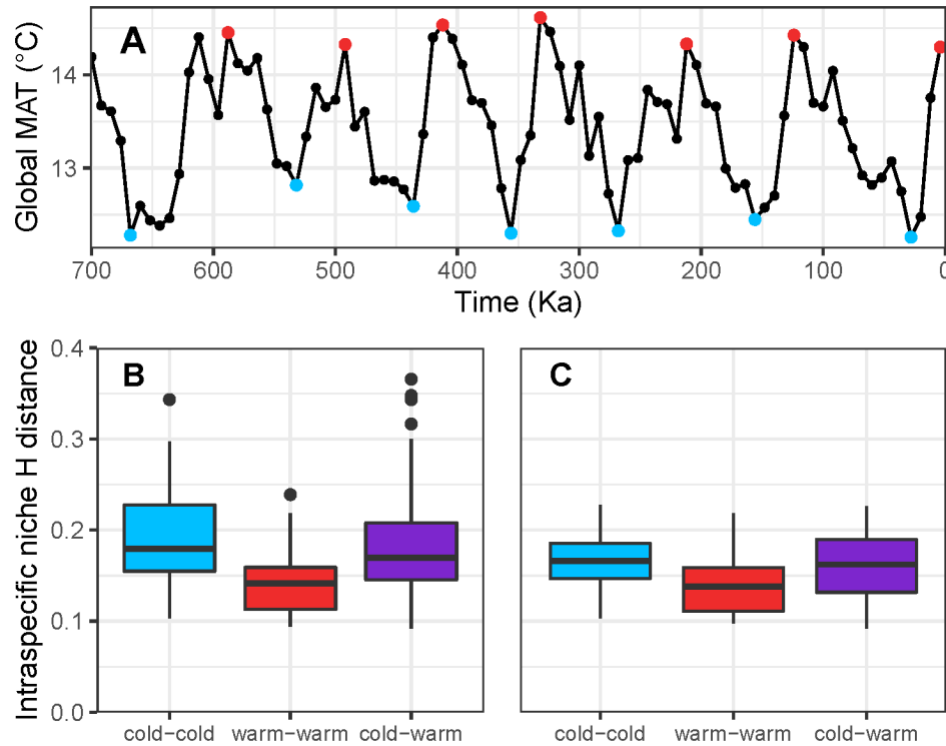
## Figures



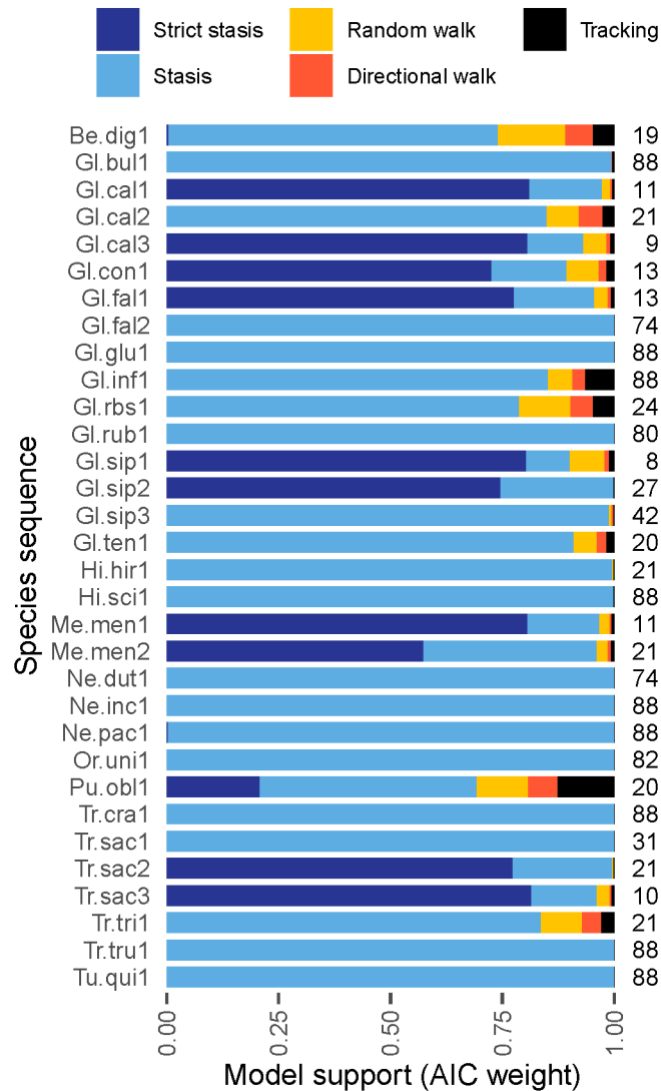
**Figure 1.** This study combined Atmosphere–Ocean Global Circulation Model temperature reconstructions (sea surface estimates at 28 ka and 212 ka shown) with fossil occurrences of planktonic foraminifera (polar species *Neogloboquadrina pachyderma* illustrated at top). Temperatures at occurrence sites are marked with ticks below the x-axis of the plots; these values are the basis for kernel density estimates of thermal niches (orange). Vertical black lines mark the mean of occupied temperatures for a sample; means were used in trait evolution models. Dark orange lines mark optimal or “preferred” temperatures, inferred as the temperature associated with the maximum of the probability density function. Hellinger’s H distance quantifies dissimilarity between a pair of probability density functions, from 0 (identical distributions) to 1 (no overlap). In the figured example, the H distance between *N. pachyderma* niches at the 28 ka glacial and 212 ka interglacial time steps is 0.08. H distances were calculated over a standard temperature interval (bounded by dashed gray lines at -1.4 and 27.1 °C) for all species and time bins.



**Figure 2.** Niche dissimilarity across each time bin boundary (mean Hellinger’s H among boundary-crossing species) was independent of absolute differences in available temperature (surface Mean Annual Temperature averaged among global marine sites) across each boundary. Corrected for temporal autocorrelation, Pearson’s  $r$  correlation coefficient was 0.18, with 95% confidence interval [-0.03, 0.37] (uncorrected  $r = 0.081$ , 95% CI = [-0.13 to 0.29]). The most recent interval (12–4 ka ago) is plotted in red. Grey bars denote interquartile ranges.  $H = 0$  for identical distributions.



**Figure 3.** (A) Time series of mean annual temperature (MAT) averaged over a standardized global grid. Blue points identify glacial maxima, and red points highlight interglacial warm peaks (ages listed in SI Appendix Table S3). (B, C) Hellinger’s H distances were equivalent for niches between pairs of glacial maxima (“cold-cold”), pairs of interglacial peaks (“warm-warm”), and pairs of glacial and interglacial extremes (“cold-warm”). Boxplots depict the variation in mean species’ H among all pairs of time bins in each category. In C the most recent interglacial (4 ka ago) and oldest glacial interval (668 ka ago) were omitted from comparisons due to anomalous sampling in these intervals, as described in *Methods* section “Niche lability among extreme climate intervals”; note the smaller variance in C than B.



**Figure 4.** Relative support for five evolutionary models fit to each species’ mean temperature time series (SI Appendix Fig. S3). Blue indicates support for stasis (*sensu stricto* or *sensu lato*), whereas yellow and orange imply changes in central tendency of occupied temperatures away from the starting value. Black represents a model where species’ occupied temperature correlates with global temperature. The preponderance of blue supports stasis as the dominant evolutionary mode of those tested. Parameter estimates from the most-supported model for each species are listed in SI Appendix Table S4. The length of each series (number of 8-ka time bins) is listed at right. Binomial names are abbreviated at left as the first two and three letters of the genus and species epithets, respectively. Discontinuous time series for a single species are distinguished with numeric suffixes. (Note that *rubescens* is abbreviated *rbs* to distinguish it from *ruber*.) SI Appendix Table S2 lists full scientific names. AIC, Akaike Information Criterion.