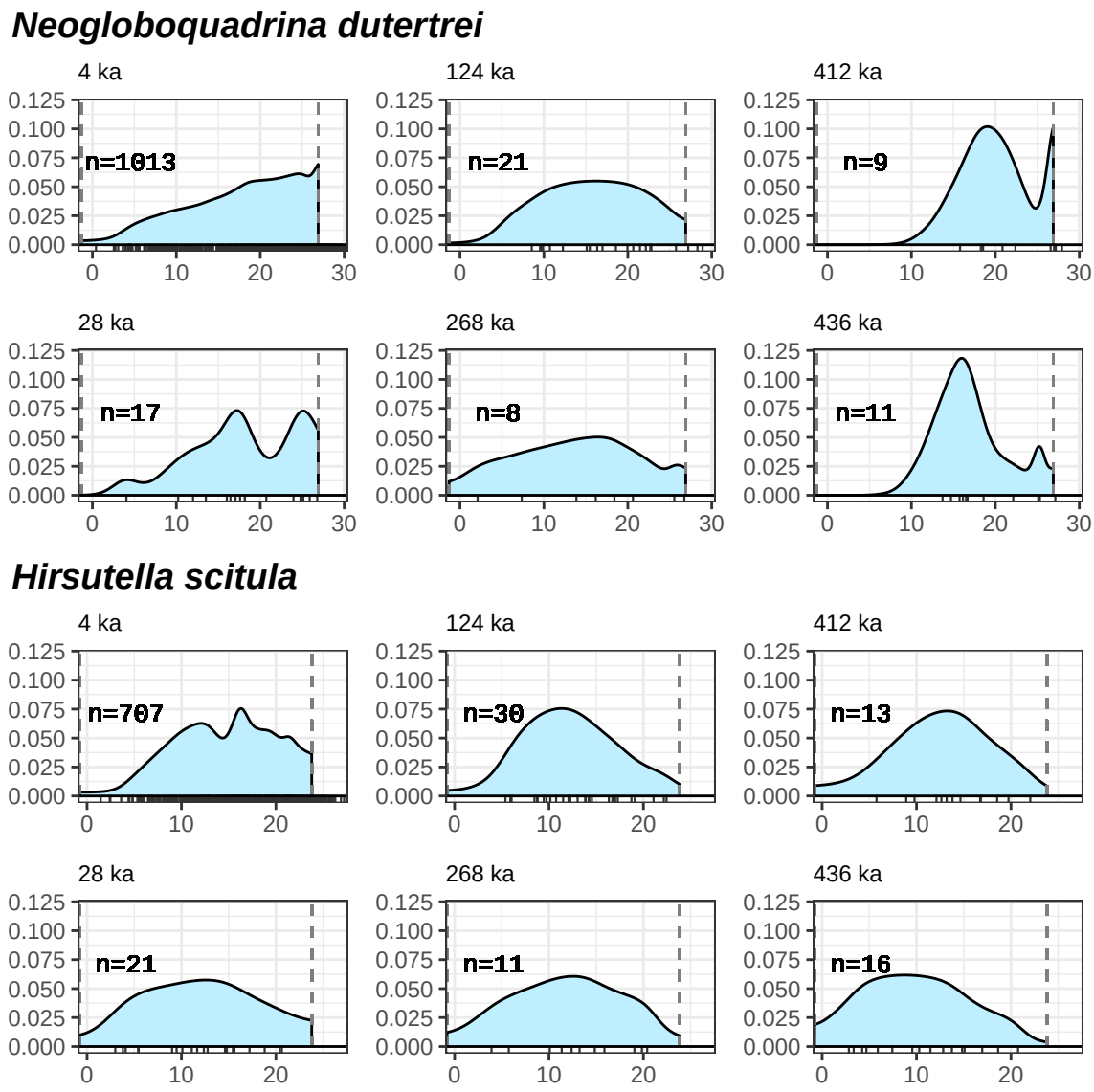


Figure S1. Kernel density estimates for two species in three interglacial peaks (top row for each species) and three glacial maxima (bottom row for each species). Temperature was extracted from the depth habitat of each species: surface for *Neogloboquadrina dutertrei* and subsurface for *Hirsutella scitula*. The original observations used for kernel density estimation are plotted as a rug at bottom of each plot. To compare niches from different time intervals, all densities of a given habitat depth were cut (and rescaled) to the same upper and lower bounds, marked by gray dashed lines. The interval of comparison was -1.3–26.9 °C for surface species and 0.8–23.8 °C for subsurface species; the limits for all habitat depths are listed in the caption to Fig. S5. Note the much larger sample size at 4 ka than any other age. The density estimate is correspondingly rougher for some species (e.g., *H. scitula*), which could artificially inflate the Hellinger distance between niches in the most recent time bin and older interglacial intervals. *Methods* section "Niche lability among extreme climate intervals" discusses this; Figs. 3C and S9B repeat analysis excluding the youngest interglacial (4 ka) and oldest glacial (668 ka).

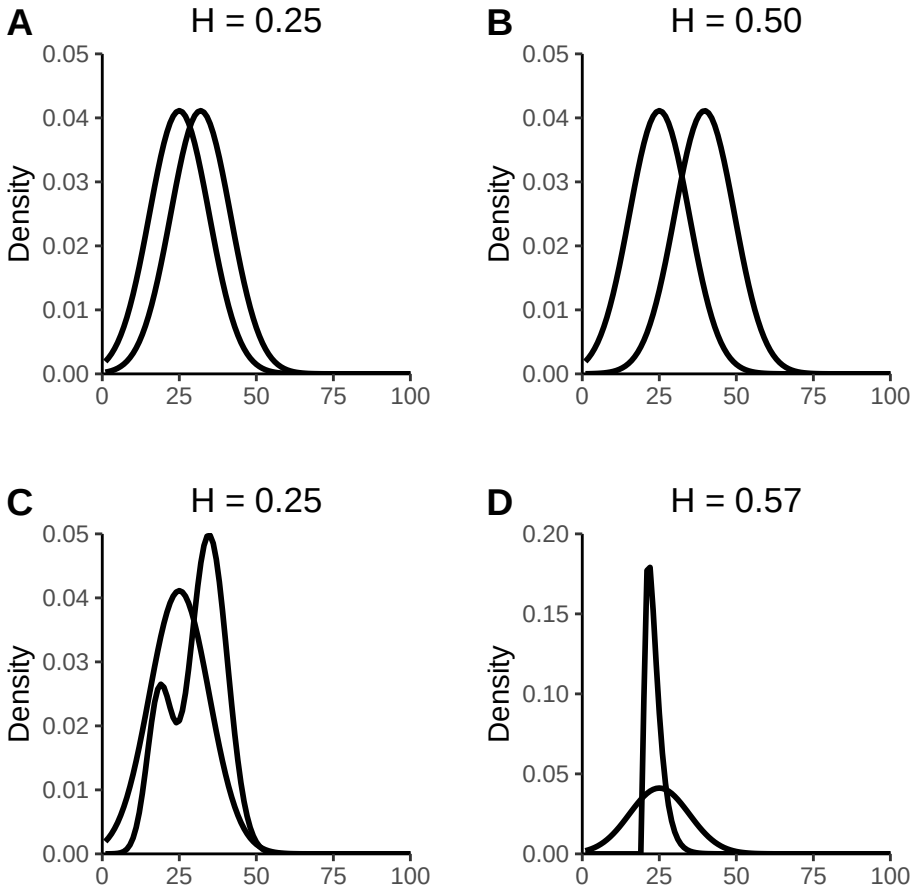


Figure S2. Examples of H distances between standard statistical distributions. H ranges from 0 (identical distributions) to 1 (non-overlapping distributions). In **A** and **B**, one of two identical Normal distributions is translated horizontally to give an H distance of 0.25 and 0.50, respectively. **C,** the original Normal distribution is compared with a mixture of two Normal distributions, with H distance of 0.25. **D,** the original Normal distribution is compared with a chi-squared distribution (4 degrees of freedom), with H distance of 0.57.

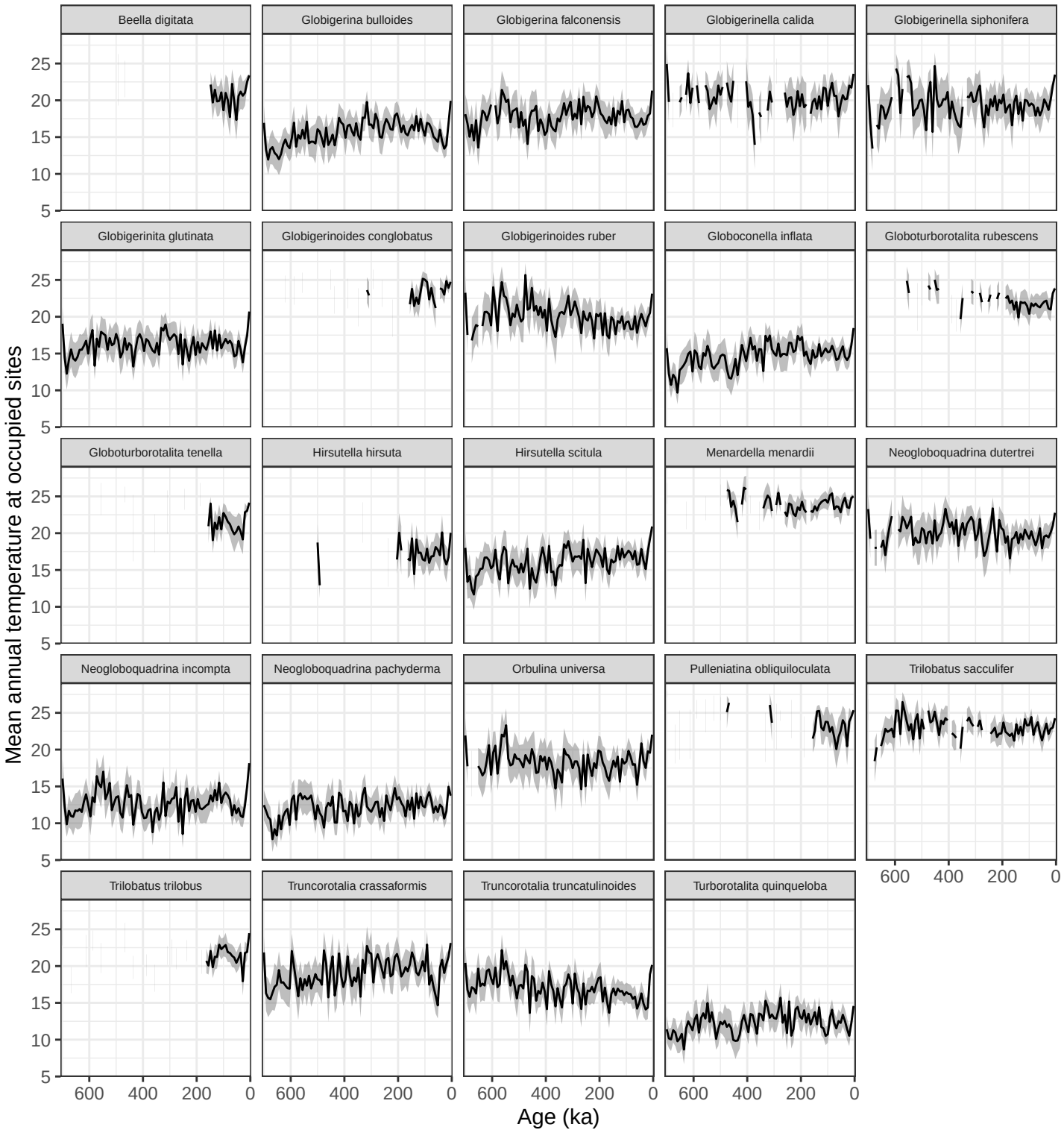


Figure S3. Time series of mean annual temperature at the sea surface of all occupied sites for a focal species. The gray band covers one standard error, which trait evolution models use to estimate parameter values and confidence intervals.

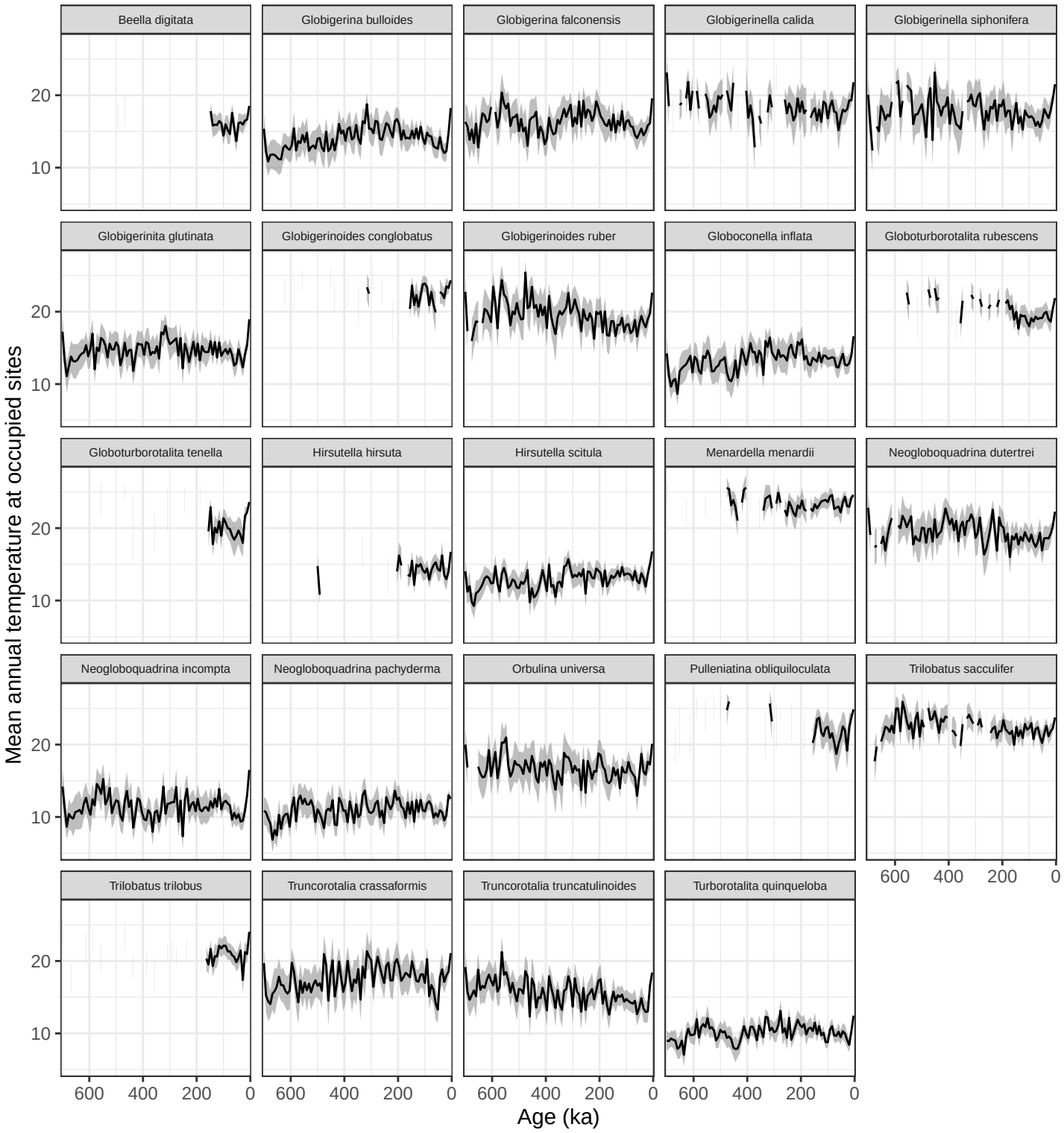


Figure S4. Time series of mean annual temperature at all occupied sites for a focal species, within habitat depths. The gray band covers one standard error, which trait evolution models use to estimate parameter values and confidence intervals.

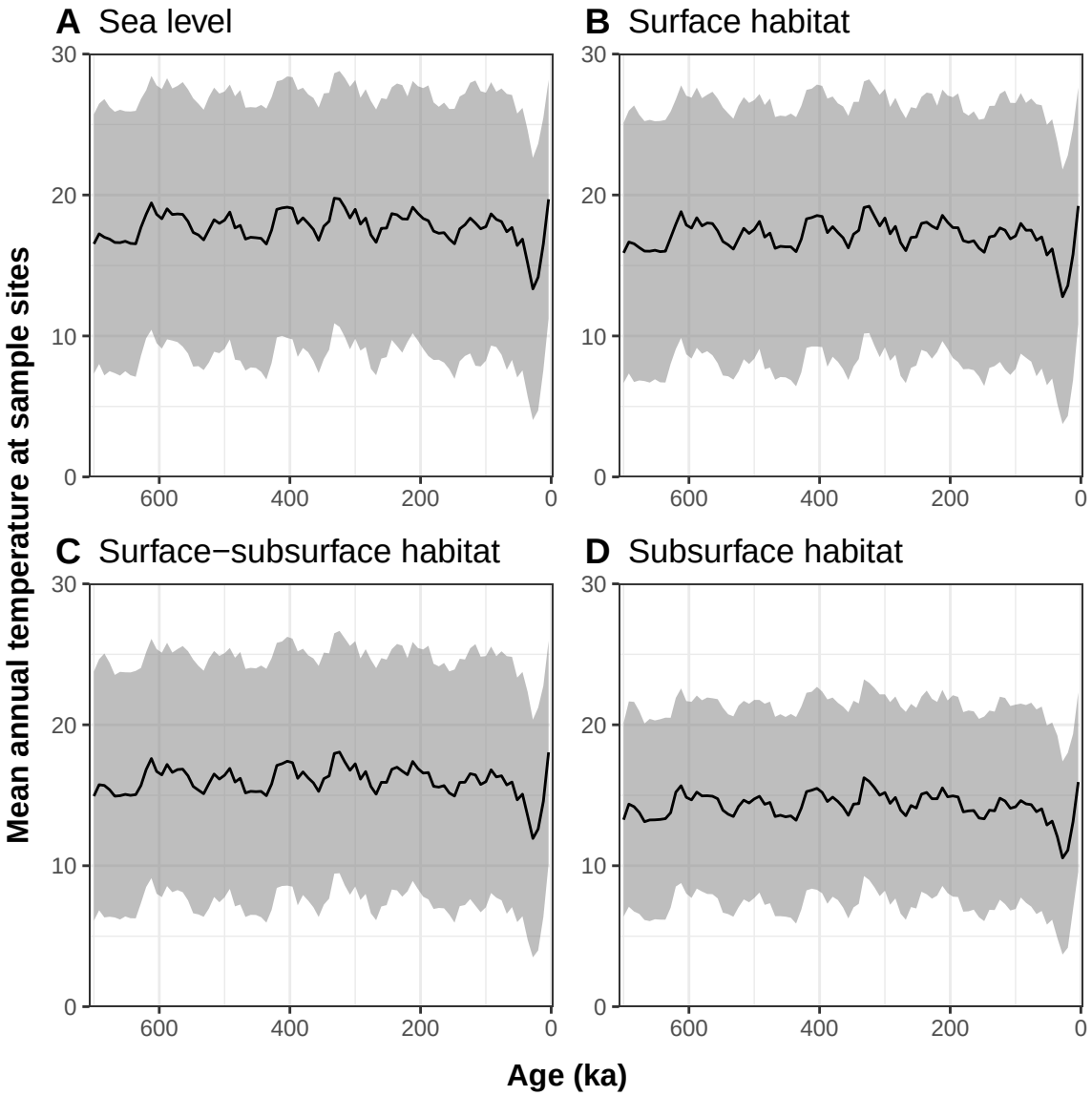
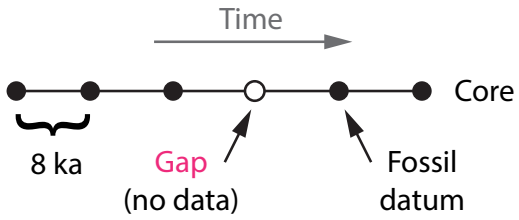
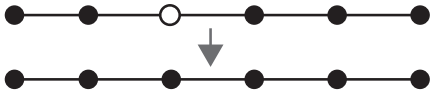


Figure S5. Mean temperature across sampling sites is plotted as a line within an envelope of one standard deviation (not standard error as in Figs. S3 and S4), to indicate the spread of observed global temperatures. The range of temperatures from all sample sites vs. all marine areas is plotted through time in Movie S1. The threshold for maximum temperature of niche comparisons, in order of depth layers, was 27.1, 26.9, 25.9, and 23.8 °C. The threshold for minimum temperature was -1.4, -1.3, -0.6, and 0.8 °C, respectively. See Fig. S6 and *Methods* section "Niche dissimilarity measure" for discussion of threshold calculations.

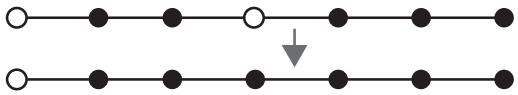
A. Each sediment core records nearly-continuous deposition over its length. Sometimes there are gaps in the fossil record reported from the core. Many gaps are due to failure to report a complete absence of fossils (or the presence of only one species) from sediment examined at that point.



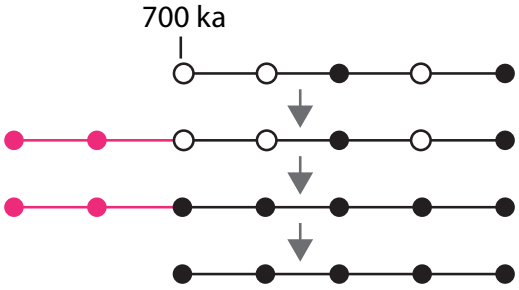
B. We can infer that sediment may have been examined (but not reported) by filling in the points that "range through" an interval bracketed by observations.



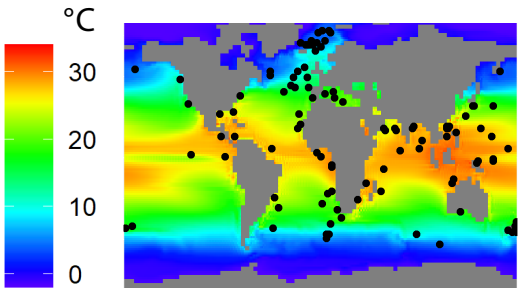
C. A range-through approach can only interpolate in the middle of a sequence, not extrapolate past the end-points. This can mean the ends of the series look less sampled than the middle.



D. The young end of the sediment cores (near 4 ka) is very well sampled, but the old end (near 700 ka) is not. Therefore, for the range-through interpolation step we added fossil and climate data from 800–700 ka, beyond the end of the study interval, so the amount of inference was spread equally across time intervals.



E. We combined the ages and coordinates of each core sample with Atmosphere-Ocean General Circulation Models to convert observations to estimated environmental conditions (mean annual temperature, MAT).



F. We compared core records from across the world to estimate the available and observable temperature range for species in each time bin. The segment of the temperature axis shared by all time bins is the standard interval across which species' niches can be fairly compared.

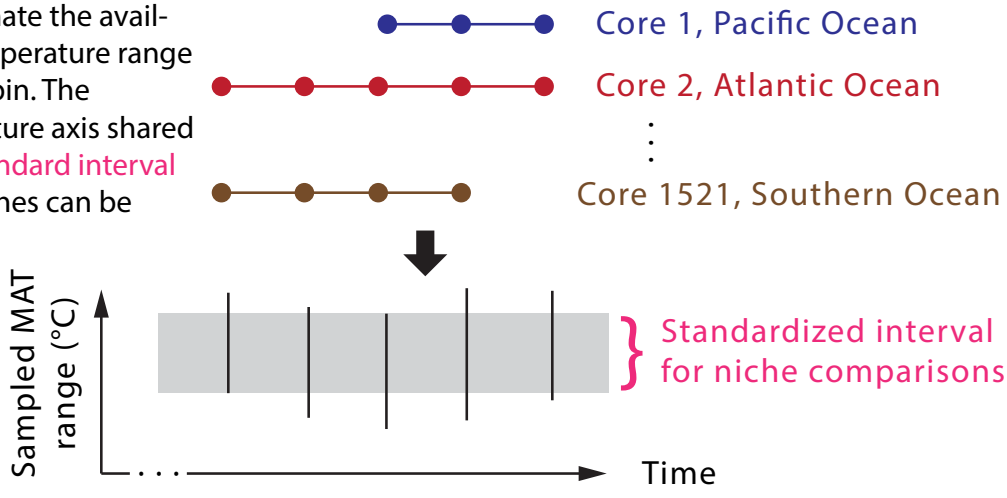


Figure S6. This diagram explains the steps to standardize the temperature interval over which thermal niches were compared. The method uses range-through inference on all available foraminifera occurrence data (presence and absence records) to determine the breadth of environmental conditions (i.e., range in mean annual temperature) that was available and observable for species in each time bin. The comparison interval is the range of environmental conditions shared across every time bin (gray band in F). The bounds of the comparison interval were calculated separately for each depth habitat and are listed in the caption of Fig. S5. Range-through interpolation adds only "pseudo-absence" points to the estimation of sampling distributions, not presence points to species distributions. Most of the pseudo-absence and recorded absence sampling locations were from polar regions during interglacial intervals. Hence, by including all absence points in the definition of sampling distributions, the lower limit of niche estimation was extended to colder temperature values. One reason for gaps in reported data is samples that contain no species (as in present-day Antarctica) or only the polar species *Neogloboquadrina pachyderma*. A single species' abundance is useless for many paleoclimate applications, so records for samples with single species might not have been entered even where sediment was present and observed.

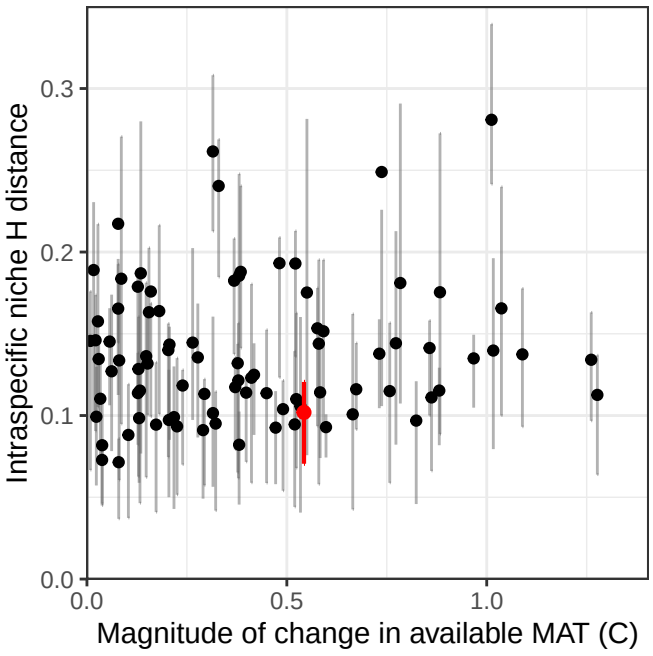


Figure S7. Hellinger's H distance (mean and interquartile range among species) between niches in each pair of consecutive time bins is plotted against absolute differences in available temperature (surface Mean Annual Temperature averaged among global marine sites). Niches were constructed from temperature extracted within depth habitats (see Fig. 2 for sea-surface data). Correcting for temporal autocorrelation in the H time series, Pearson's r was 0.18 (95% CI = $[-0.03, 0.38]$). Without pre-whitening, $r = 0.098$ (95% CI = $[-0.12, 0.30]$).

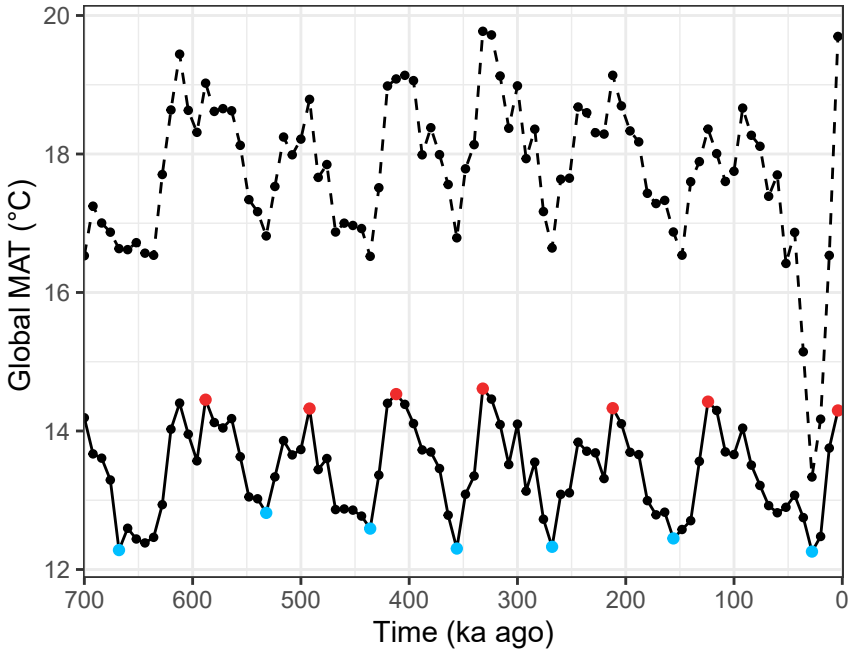


Figure S8. Time series of mean annual temperature (MAT) averaged over either a global marine grid of 10°-resolution (solid line) or all sampled sediment core sites (dashed line). Blue points identify glacial maxima, and red points highlight interglacial warm peaks (ages listed in Table S3).

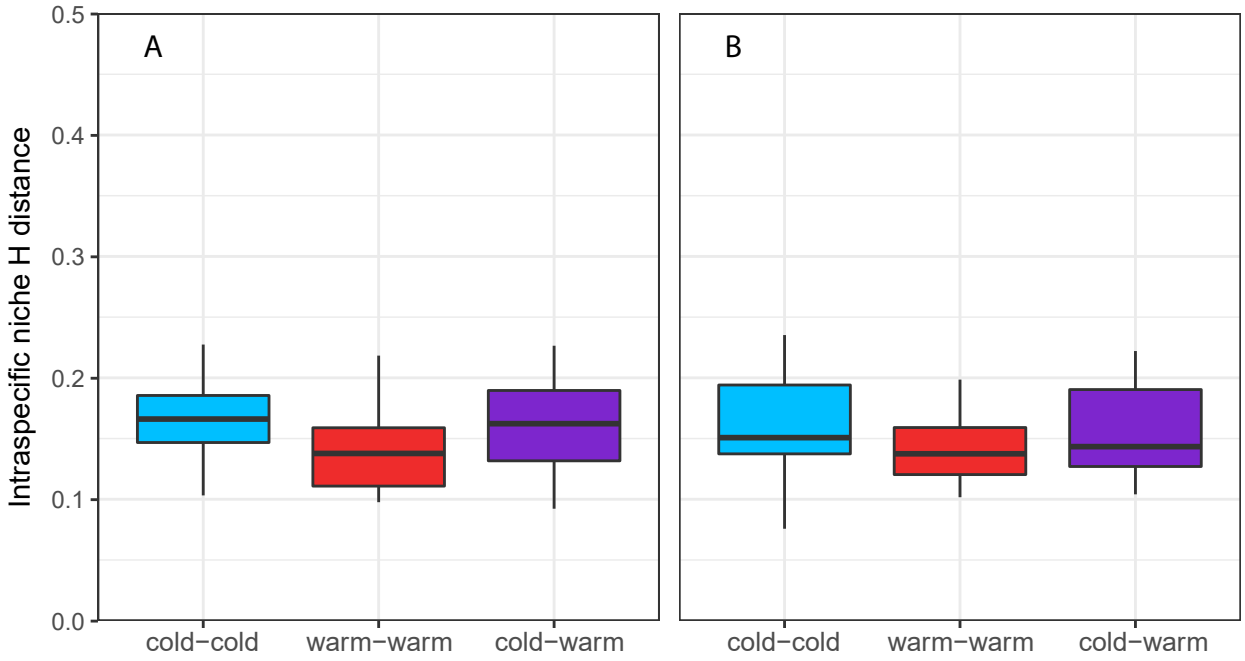


Figure S9. Hellinger's H distances were computed for (1) each different pair of glacial periods ("cold–cold"), (2) each different pair of glacial periods ("warm–warm"), or (3) each different pair of one glacial and one interglacial period ("cold–warm"). Temperature estimates were extracted in species' habitat depths (see Fig. 3 for sea-surface data). Boxplots depict the variation in mean species' H among all pairs of time bins in each category. In **A** calculations were based on all 7 glacial and interglacial intervals (Table S3). In **B** the most recent interglacial (4 ka ago) and oldest glacial interval (668 ka ago) were omitted. The exceptionally dense and sparse sampling in these intervals, respectively, could influence H as described in *Methods* section "Niche lability among extreme climate intervals."

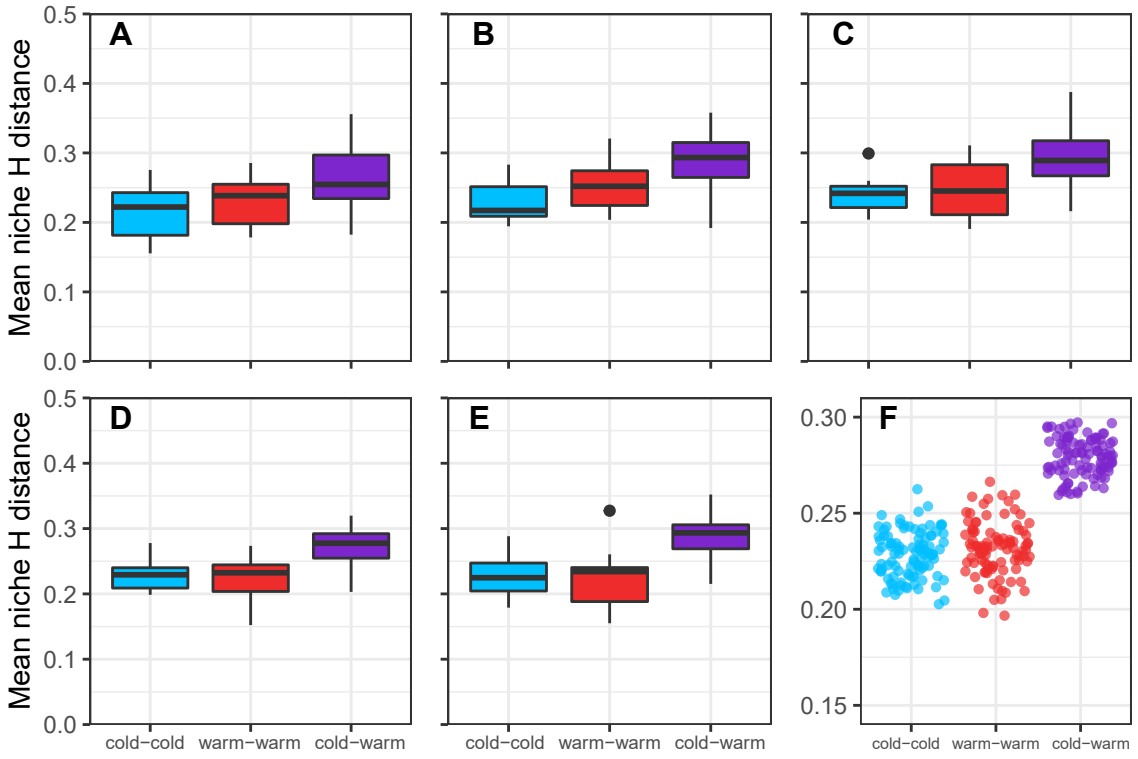


Figure S10. A–E, the first 5 (of 100) replicates of mean Hellinger's H distance comparisons simulated under a hypothesis of strict geographic stasis (i.e., complete niche adaptation to local conditions during environmental change). Boxplots are as in Figs. 3 and S9: H was averaged among all species sampled in a given pair of time bins, and then averaged among all time bin pairs in each of three climate-comparison categories. In **F**, the mean for each climate-comparison category from every simulation replicate is plotted as a point, e.g., the middle line in each boxplot from **A**–**E** appears again as a point in **F**. Note the expanded y-axis in **F**. In each of 100 simulations, the mean H distance was larger on average between a glacial and interglacial interval ("cold–warm") than between two glacial intervals ("cold–cold") or two interglacial intervals ("warm–warm"). The simulation steps are detailed in *Methods* section "Test for type II error." Temperature values were extracted from the sea surface, and the most recent interglacial (4 ka ago) and oldest glacial interval (668 ka ago) were excluded from comparison, as in Fig. 3C.

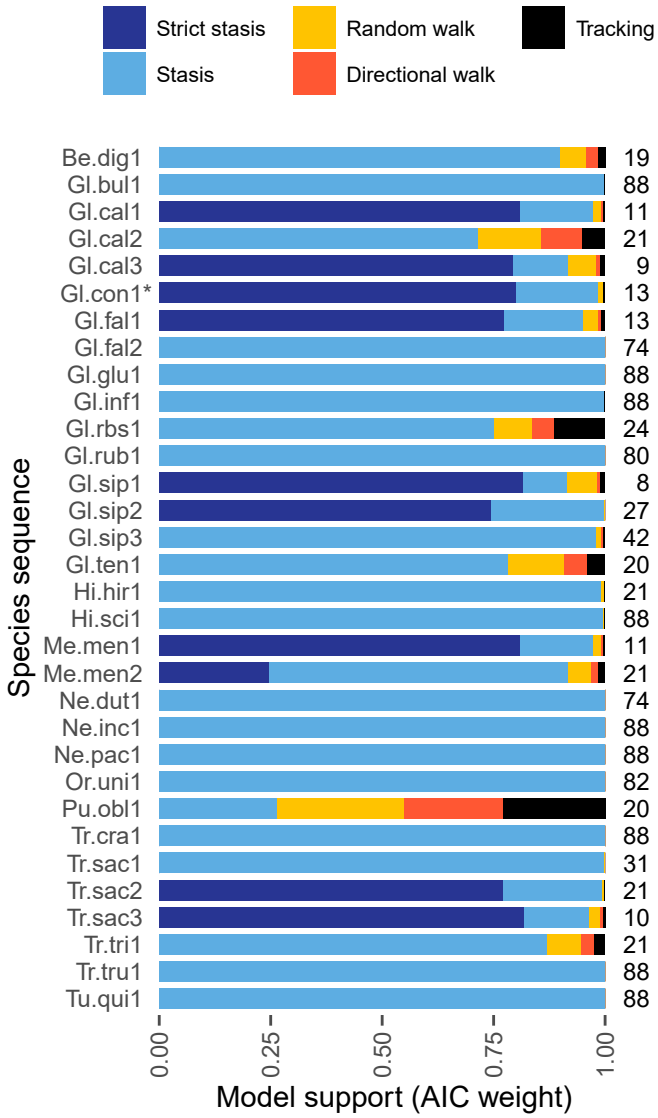


Figure S11. Relative support for five evolutionary models fit to time series of species' occupied temperature values under the depth-habitat approach. Fig. 4 presents results under the sea-surface approach. The length of each sequence (number of 8-ka time bins) is listed at right. Abbreviated species names are listed at left; discontinuous time series for a single species are distinguished with numeric suffices. (Note that *rubescens* is abbreviated rbs to distinguish it from *ruber*.) Table S2 lists full scientific names. An asterisk marks *Globigerinoides conglobatus* because the covariate-tracking model for this species was modified by pooling variances, for successful optimization. AIC, Akaike Information Criterion.

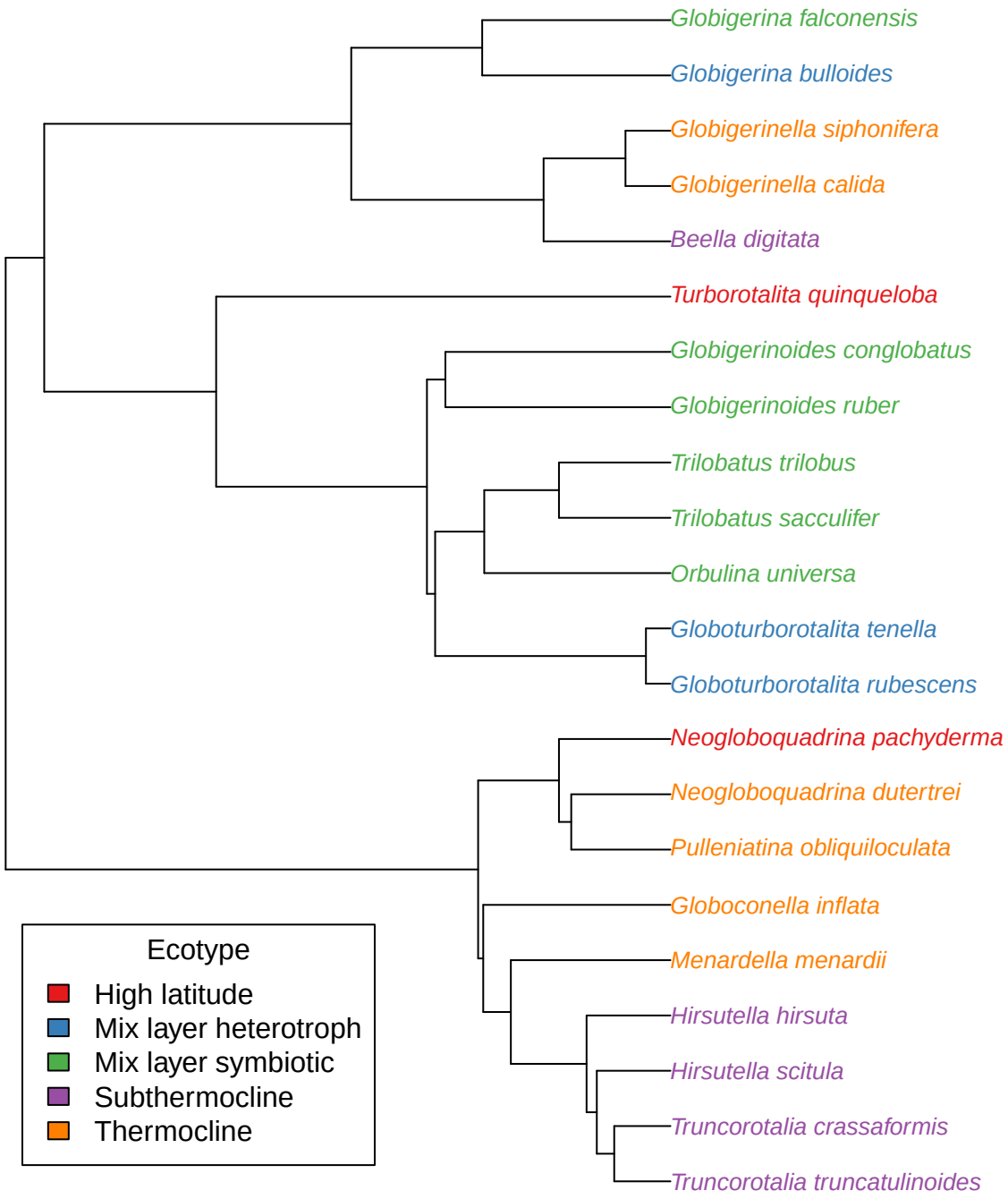


Figure S12. The 22 species used in ANOVA (Table S5) are plotted in the phylogenetic tree of Aze and others (2011). Tips are color-coded according to ecotype.

Analysis	Response variable	Sample size	Result
Time series correlation	Time series of MAT optima for each species	32 species time series	Optima are independent of global MAT
Niche lability *vs* climate change magnitude	H between consecutive bins for each species; mean H among species for each boundary	87 time bin boundaries	Niche lability is independent of the magnitude of global MAT change (Fig. 2)
Niche lability among extreme climate intervals	H between glacial-interglacial intervals for each species; mean among species for each bin pair	21 pairs of glacial-glacial/interglacial-interglacial bins; 49 pairs of glacial-interglacial bins	Niche distance is not greater between opposite climate extremes (Fig. 3)
Trait evolution models	Time series of mean and SE of MAT for each species	32 species time series	Stasis is the most-supported mode of evolution (Fig. 4)
Phylogenetic ANOVA	H between consecutive bins for each species; mean H among all boundaries for each species	22 species in an evolutionary tree	Ecotypes do not explain variance in within-species niche lability

Table S1. Overview of the five analyses to investigate thermal niche lability. Each analysis was repeated for mean annual temperature (MAT) extracted at the sea surface and within species' depth habitats. Hellinger's H is the distance between two probability density functions.

Species	Ecotype	Ecotype ref.	Depth	Depth ref.	H	Optimum
Beella digitata	Subthermocline	Coxall et al. 2007	Subsurface	Schiebel and Hemleben 2017	0.10	22.6
Globigerina bulloides	Mix layer heterotroph	Vergnaud-Grazzini 1976, Durazzi 1981	Surface-subsurf	Rebotim et al. 2017	0.09	13.0
Globigerina falconensis	Mix layer symbiotic	Aze et al. 2011	Surface-subsurf	Rebotim et al. 2017	0.13	16.8
Globigerinella calida	Thermocline	Aze et al. 2011	Surface-subsurf	Rebotim et al. 2017	0.16	20.2
Globigerinella siphonifera	Thermocline	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.14	19.8
Globigerinita glutinata	-		Surface-subsurf	Rebotim et al. 2017	0.09	13.5
Globigerinoides conglobatus	Mix layer symbiotic	Keller 1985	Surface	Watkins et al. 1996	0.19	24.2
Globigerinoides ruber	Mix layer symbiotic	Keller 1985, Pearson et al. 2001, Pearson and Shackleton 1995	Surface	Rebotim et al. 2017	0.15	20.5
Globoconella inflata	Thermocline	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.10	13.2
Globoturborotalita rubescens	Mix layer heterotroph	Aze et al. 2011	Surface-subsurf	Rebotim et al. 2017	0.12	22.3

Globoturborotalita tenella	Mix layer heterotroph	Aze et al. 2011	Surface	Rebotim et al. 2017	0.12	23.7
Hirsutella hirsuta	Subthermocline	Aze et al. 2011	Subsurface	Rebotim et al. 2017	0.16	16.5
Hirsutella scitula	Subthermocline	Shackleton and Vincent 1978, D. R. M. Stewart unpublished data	Subsurface	Rebotim et al. 2017	0.10	14.7
Menardella menardii	Thermocline	Shackleton and Vincent 1978, Pearson and Shackleton 1995, D. R. M. Stewart unpublished data	Surface	Watkins et al. 1996	0.17	23.3
Neogloboquadrina dutertrei	Thermocline	Kahn 1979, Shackleton and Vincent 1978	Surface	Rebotim et al. 2017	0.16	19.5
Neogloboquadrina incompta	High latitude	Darling et al 2006	Surface-subsurf	Rebotim et al. 2017	0.10	9.0
Neogloboquadrina pachyderma	High latitude	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.08	7.7
Orbulina universa	Mix layer symbiotic	Vergnaud-Grazzini 1976	Surface-subsurf	Rebotim et al. 2017	0.13	16.4

Pulleniatina obliquiloculata	Thermocline	Shackleton and Vincent 1978, Pearson and Shackleton 1995	Surface	Rebotim et al. 2017	0.22	25.3
Trilobatus sacculifer	Mix layer symbiotic	Keller 1985, D. R. M. Stewart unpublished data	Surface	Rebotim et al. 2017	0.18	22.4
Trilobatus trilobus	Mix layer symbiotic	Keller 1985, Pearson et al. 2001, Pearson and Shackleton 1995	Surface	Loncaric et al. 2006	0.17	20.1
Truncorotalia crassaformis	Subthermocline	D. R. M. Stewart unpublished data	Surface-subsurf	Ezard et al. 2015, Rebotim et al. 2017, Schiebel 2017, Meilland et al. 2019	0.15	18.5
Truncorotalia truncatulinoides	Subthermocline	Vergnaud-Grazzini 1976, Shackleton and Vincent 1978	Surface-subsurf	Rebotim et al. 2017	0.14	15.7
Turborotalita quinqueloba	High latitude	Pearson and Wade 2009	Subsurface	Rebotim et al. 2017	0.11	10.9

Table S2. Attributes are listed and referenced for each species in the dataset. H is the mean Hellinger's distance between pairs of niches in consecutive time bins; H = 0 if identical. Optimum is the mean annual temperature (°C) associated with the largest probability of occurrence, averaged over the time bins in which the species was sampled.

Table S3. Median age of time bins (ka) associated with climatic extremes. In Figs. 3A and S8, glacial maxima are marked in blue and interglacial peaks are marked in red. Interglacial onset estimates are from Lisiecki and Raymo (2005).

Glacial maximum	Interglacial peak	Interglacial onset
28	4	14
156	124	130
268	212	243
356	332	337
436	412	424
532	492	533
668	588	621

Table S4. Parameter estimates are listed for trait evolution models that received the most support (Fig. 4). The "best-fit" models for species' mean occupied sea surface temperature were all either strict stasis or broad-sense stasis. The parameters for these models are (1) the estimated static mean and (2) the estimated variance about that mean, in ℃. For strict stasis models (listed first), the estimated variance is zero.

	Est. Mean	Est. Variance
Globigerina falconensis1	16.89	0.00
Globigerinella calida1	20.67	0.00
Globigerinella calida2	19.84	0.00
Globigerinella siphonifera1	18.74	0.00
Globigerinella siphonifera2	20.12	0.00
Globigerinoides conglobatus1	23.79	0.00
Menardella menardii11	23.31	0.00
Menardella menardii12	24.77	0.00
Trilobatus sacculifer2	23.60	0.00
Trilobatus sacculifer3	23.92	0.00
Beella digitata1	21.20	1.33
Globigerina bulloides1	15.82	1.31
Globigerina falconensis2	18.15	0.92
Globigerinella calida3	21.10	1.28
Globigerinella siphonifera3	19.90	1.59
Globigerinita glutinata1	16.38	1.19
Globigerinoides ruber1	20.39	1.09
Globoconella inflata1	14.85	1.18
Globoturborotalita rubescens1	22.08	0.65
Globoturborotalita tenella1	22.03	1.28
Hirsutella hirsuta1	17.59	1.27
Hirsutella scitula1	16.43	1.52
Neogloboquadrina dutertrei1	20.39	0.70
Neogloboquadrina incompta1	12.77	1.81
Neogloboquadrina pachyderma1	12.20	0.55
Orbulina universa1	18.61	1.11
Pulleniatina obliquiloculata1	24.23	0.51
Trilobatus sacculifer1	23.02	0.32
Trilobatus trilobus1	21.82	1.20
Truncorotalia crassaformis1	19.34	1.55
Truncorotalia truncatulinoides1	17.10	1.02
Turborotalita quinqueloba1	12.38	0.56

Table S5. Phylogenetic ANOVA compared variation in mean species' H distance among ecotypes. Coefficient values are relative to the reference group of open ocean thermocline species. Negative estimates imply smaller Hellinger's H values, i.e., greater niche conservatism. Mixed layer species are from tropical/subtropical regions and are separated into those with symbionts and those without them. Ecotypes are plotted on the phylogenetic tree in Fig. S12.

Ecotype	Estimate	Std. Err.
A. Sea surface		
Intercept	0.160	0.012
Thermocline	0	-
Mix layer symbiotic	-0.001	0.017
Mix layer heterotroph	-0.051	0.021
Subthermocline	-0.029	0.018
High latitude	-0.066	0.025
B. Depth habitat		
Intercept	0.155	0.014
Thermocline	0	-
Mix layer symbiotic	0.019	0.020
Mix layer heterotroph	-0.048	0.025
Subthermocline	-0.008	0.021
High latitude	-0.047	0.029

Movie S1. Sea surface temperature estimates (°C) from Atmosphere–Ocean Global Circulation Models are mapped at each 8-ka time step. Black dots indicate sediment cores, some of which are discontinuously represented in the occurrence dataset. Land and ice cover are greyed out. At right, a range chart plots the minimum, mean, and maximum temperature from a global marine grid of 10°-resolution vs. from all sampling locations. Sampling sites capture most of the global temperature variation and mean changes through time.