Flat meridional temperature gradient in the early Eocene in the subsurface rather than surface ocean

Sze Ling Ho[*][†] and Thomas Laepple[*]

The early Eocene (49–55 million years ago) is a time interval characterized by elevated surface temperatures and atmospheric CO_2 (refs 1,2), and a flatter-than-present latitudinal surface temperature gradient[3,4]. The multi-proxy-derived flat temperature gradient has been a challenge to reproduce in model simulations[5-7], especially the subtropical warmth at the high-latitude surface oceans[4,8], inferred from the archaeal lipid-based palaeothermometry, TEX_{86}^H. Here we revisit the TEX_{86}^H interpretation by analysing a global collection of multi-proxy temperature estimates from sediment cores spanning millennia to millions of years. Comparing the variability between proxy types, we demonstrate that the present TEX_{86}^H interpretation[9] overestimates the magnitude of past climate changes on all timescales. We attribute this to an inappropriate calibration, which reflects subsurface ocean but is calibrated to the sea surface, where the latitudinal temperature gradient is steeper. Recalibrating the proxy to the temperatures of subsurface ocean, where the signal is probably formed, yields colder TEX_{86}^H-temperatures and latitudinal gradient consistent with standard climate model simulations of the Eocene climate[10], invalidating the apparent, extremely warm polar sea surface temperatures. We conclude that there is a need to reinterpret TEX_{86}^H-inferred marine temperature records in the literature, especially for reconstructions of past warm climates that rely heavily on this proxy as reflecting subsurface ocean.

The combination of global warming and high atmospheric CO_2 during the early Eocene renders this time interval a potential analogue for anthropogenic climate change. A long-standing, unresolved issue in simulating warmer-than-present climates is the failure of models in reproducing the flat Equator-to-pole surface temperature gradient often observed in proxy data[5,6]. The model–proxy match is improved in the terrestrial realm in high-CO_2 scenarios[6] or via tuning model parameters[7], but the extreme surface oceanic warmth in polar regions inferred from TEX_{86}^H has been irreconcilable. Although this discrepancy was attributed to missing processes in climate models[8], inconsistencies in proxy data also suggest deficiencies in proxy understanding[11,12]. For example, TEX_{86}^H-inferred sea surface temperatures (SST) in deepwater formation areas are $>10\,^{\circ}C$ higher than independently inferred temperatures for deep waters[13] and the Antarctic coast[14].

TEX_{86}^H palaeothermometry[15] is based on the relationship between the relative distribution of archaeal lipids and temperature, which holds true even at temperatures higher than the modern tropical SSTs[16]. Confidence in TEX_{86}^H for past warm climates reconstruction culminated in TEX_{86}^H being used to correct Mg/Ca-derived tropical Pliocene temperature variations[17].

However, it was recently shown that the glacial–interglacial amplitude of TEX_{86}^H-derived SST change in the tropics is overestimated relative to other proxy evidence[18], a result also independently found by a multi-proxy study in the subpolar region[19]. If this discrepancy stems from a systematic bias in TEX_{86}^H, it would have implications for TEX_{86}^H-based reconstructions across timescales, including the flat thermal gradient during the early Eocene.

To investigate this hypothesis, we examine 22 pairs of $U_{37}^{K'}$- and TEX_{86}^H-inferred temperature records (in total 5,528 samples) measured in tandem on sediment cores from sites spanning diverse oceanographic settings and timescales from thousands to millions of years (Supplementary Figs 1 and 2). Temporal leads and lags between proxy record pairs are possible, as these proxies might reflect different water depths, where the timing of temperature changes differ[20], and proxy-specific sedimentary processes[21] can create temporal offsets. This inhibits a direct comparison of the time series, as timing differences generally destroy coherency.

Instead, power spectra of each of the paired $U_{37}^{K'}$ and TEX_{86}^H records are estimated and the mean power spectral estimate (PSD) for each proxy type is compared. Spectra averaged over multiple regions attenuate local effects, thereby facilitating inter-comparison between proxy types. This technique is insensitive to temporal offsets between the records[22] and allows fingerprinting the reasons for discrepancies between the proxies[23].

The power spectra of $U_{37}^{K'}$ and TEX_{86}^H show a very similar shape, with increased energy towards lower frequencies (Fig. 1). In addition to the known power-law scaling between 1/500 yr and 1/100 kyr (ref. 23) and a flatter spectrum between 1/100 kyr and 1/5 Myr (Fig. 1a), the imprint of the orbital variability is visible as local maxima of the PSD at 1/41 kyr, and more prominently at 1/100 kyr (Fig. 1b).

Despite the similarity in the shape, there is an evident offset in the power spectra of both proxies, with the variability in TEX_{86}^H records being approximately three times as strong as that in $U_{37}^{K'}$ records across millennial to million-year timescales (Fig. 1). This remarkable finding is insensitive to the choice of any of the commonly used TEX_{86}^H calibrations (Supplementary Fig. 4), and also applies to the single records: 20 of 22 pairs show more variability in TEX_{86}^H than $U_{37}^{K'}$ (Supplementary Fig. 3).

One possible explanation is that non-climate effects on the proxies, post-sedimentary processes, or a temperature limitation of the $U_{37}^{K'}$ calibration account for the observed discrepancy. However, independent additive noise processes such as analytical uncertainty would flatten only the fast-frequency tail of the spectrum, as there the noise would represent a considerable fraction of the signal (Supplementary Fig. 3). In contrast, mixing processes such as

---

Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Telegrafenberg A43, D-14473 Potsdam, Germany. [†]Present address: University of Bergen and Bjerknes Centre for Climate Research, Allegaten 41, 5007 Bergen, Norway. [*]e-mail: ling.ho@uib.no; tlaepple@awi.de
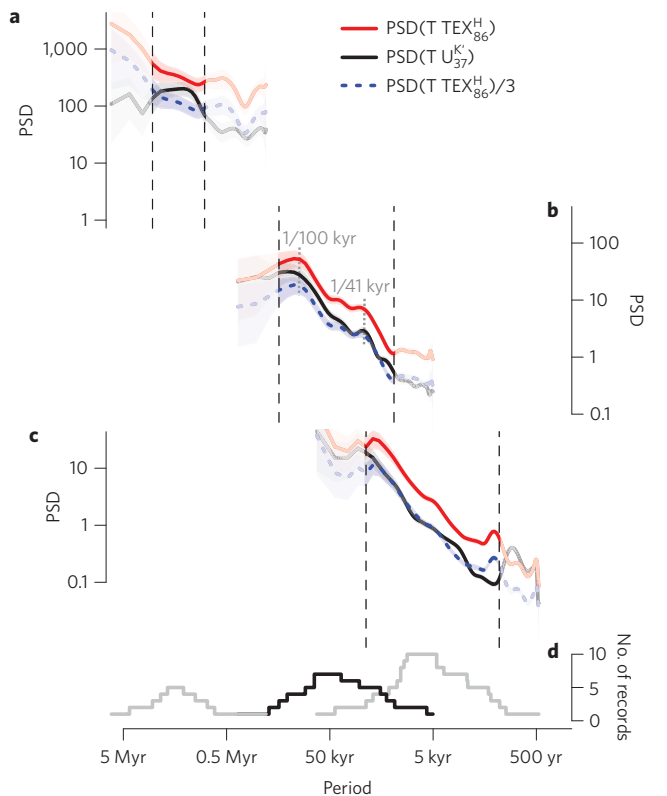
Figure 1 | Spectral estimates of $U^{K'}_{37}$- and TEX^{H}_{86}-inferred temperature variability. a–c, Mean power spectral density (PSD) estimate of the globally distributed single records for three time intervals: multi-million years (**a**), past one million years (**b**), last glacial cycle (**c**). **d**, Number of records contributing to each frequency. Dividing the TEX^{H}_{86}-inferred variability by three (dashed blue line) results in a spectral estimate similar to that derived from $U^{K'}_{37}$ SST. As averaging across records is needed to damp the effect of regionally differing surface and subsurface variability (Supplementary Fig. 3), estimates based on fewer than three records are shaded white. Proxy records constituting the composite spectra are listed in Supplementary Fig. 2. Shading indicates 95% confidence intervals representing the uncertainty of the spectral estimate.

bioturbation would dampen the fast frequencies[24]. Non-thermal effects, such as temporal oxygen changes[25] (Supplementary Information) could lead to additional TEX^{H}_{86} variability, but this should be strongly timescale- and setting-dependent, unlike the results we obtain. Reduced sensitivity of $U^{K'}_{37}$ at the warm temperature end could dampen the $U^{K'}_{37}$-derived variability, but we obtain similar results when excluding high temperatures (see Methods).

Another explanation for the observed TEX^{H}_{86}–$U^{K'}_{37}$ discrepancy is that the actual temperature variability seen by TEX^{H}_{86} is larger than that seen by $U^{K'}_{37}$, as both proxies might reflect water temperatures of a different season or depth range. General circulation model (GCM) experiments, however, indicate that long-term variability does not show a clear seasonal distinction persisting across timescales, and similar variability is on average expected at intermediate depths (~600 m) compared to the sea surface (Methods, Supplementary Fig. 5). The latter is independently confirmed by analysing a global compilation (16 sites, in total 7,691 samples) of paired multi-species Mg/Ca records, which shows similar variability (Supplementary Fig. 6) for mixed-layer-dwelling and thermocline-dwelling foraminifera across timescales.

Given the systematic nature of the discrepancy; scaling the TEX^{H}_{86} variability by a constant factor leads to a very similar variability pattern between proxies (Fig. 1), and also the variability mismatch

is consistent over diverse climate regimes, oceanic settings and a large temperature range, we argue that the most likely reason for the discrepancy is a systematic bias of the calibration slope, for instance due to an inappropriate choice of target temperature for calibration.

The target parameter for the commonly used calibrations for both proxies studied here is SST. This is probably appropriate for the $U^{K'}_{37}$ calibration because the producers of alkenones are restricted to the photic zone and a similar calibration was obtained independent of habitat assumptions[26]. The depth habitat is less constrained for the source organisms of TEX^{H}_{86}, given the occurrence of marine archaea throughout the water column[27] and their generally low abundance in the upper photic zone, potentially due to light inhibition[28]. The exact depth origin of archaeal lipids exported to marine sediments remains unclear[11,12], but they probably do not originate exclusively from the mixed layer[29].

As noted previously[15,20], any global marine sediment core-top calibration is sensitive to the depth range of the calibration target temperature, as the zonal marine temperature gradient decreases with depth (Fig. 2b), affecting the calibration slope (Fig. 2c).

To quantify the sensitivity of the calibration to the habitat depth, we derive all possible calibrations of the global core-top TEX^{H}_{86} data against depth-integrated seawater temperatures, with start and end depth between 0 and 1,000 m. This results in numerous calibrations with an explained variance indistinguishable (Methods) from the widely applied global SST calibration (Fig. 2a), demonstrating that the global TEX^{H}_{86} calibration is largely unconstrained without additional knowledge. Thus, we determine the slope independent of the core-top calibration by matching the $U^{K'}_{37}$ and TEX^{H}_{86} variability within their uncertainties. This is possible, as the $U^{K'}_{37}$ calibration is better constrained, and as both proxies are expected to show similar variability when averaged over a large number of sites (Methods). Given the distribution of TEX^{H}_{86}-to-temperature slopes (Supplementary Fig. 7), we use the set of core-top calibrations (hatched area in Fig. 2a) to constrain the depth range and calibration intercept (Fig. 2b) by choosing all calibrations with the corresponding slopes. The resultant subsurface calibration ensemble (mean calibration: $T = 40.8 \times TEX^{H}_{86} + 22.3$, full ensemble in Supplementary Information) is by construction consistent with both the modern spatial proxy–temperature relationship and the multi-proxy evidence of past variability. The ensemble includes calibrations with integrated water depths ranging from a 50 m depth interval (for example, 150–200 m) to a 950 m depth interval (for example, 0–950 m), with 0–550 m having the highest occurrence. Notably, all water depths are far greater than typically assumed for the interpretation of TEX^{H}_{86} (refs 9,20).

It has been proposed that deeper living archaea might produce lipids with a different temperature relationship and affect the sedimentary signal[29]. Repeating our variability comparison for shallow (<1,000 m water depth) and deeper (≥1,000 m) sites does not affect the resulting calibration (<5% slope change), suggesting similar sedimentary TEX^{H}_{86}–temperature relationships on a global scale regardless of the water depth at the sites.

Our findings have a strong impact on the interpretation of TEX^{H}_{86}-based reconstructions. First, the absolute value and amplitude of change in all TEX^{H}_{86}-inferred temperatures are approximately halved, as both slope and intercept of the calibration are reduced by 40–47% and 42–48%, respectively (67% quantiles). This affects TEX^{H}_{86}-based studies on all timescales, including the reconciliation of TEX^{H}_{86} and $U^{K'}_{37}$ for the glacial cooling in the tropics[18] and the Middle Eocene Climatic Optimum warming in the Southern Ocean, as well as an improved variability match in most (20 of 22) of the paired records in our collection (Supplementary Fig. 8). As we now compare reconstructed surface (for example, $U^{K'}_{37}$-based) and subsurface (TEX^{H}_{86}) temperatures, we do not expect a complete reconciliation of $U^{K'}_{37}$- and TEX^{H}_{86}-derived records at every site (Supplementary Fig. 8). Further, although our
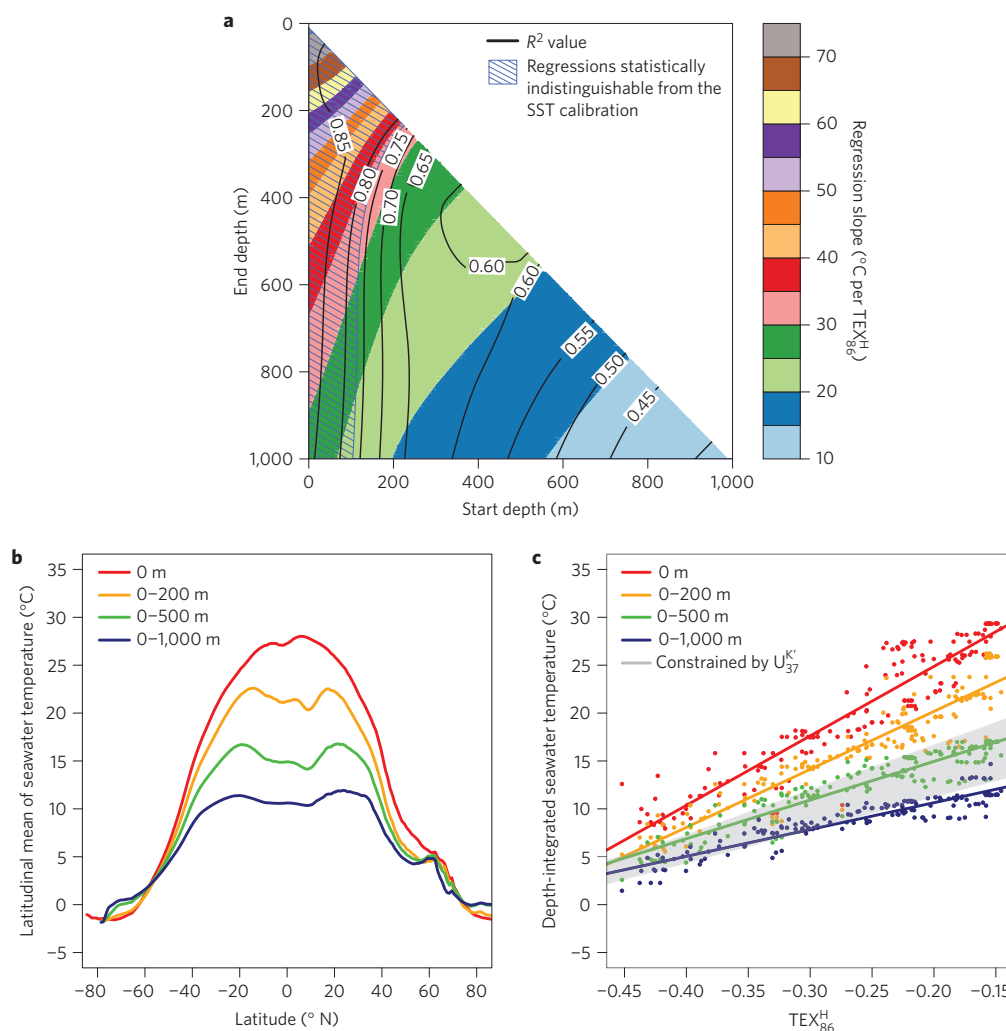
Figure 2 | Depth dependence of the global TEX$_{86}^H$ calibration. a, Slope (colour shading) and R^2 (contour) of all possible TEX$_{86}^H$–temperature regressions using integrated seawater temperatures from World Ocean Atlas 2009 between a start depth (*x*-axis) and end depth (*y*-axis). The hatched area denotes regressions that are statistically indistinguishable from the widely applied global sea surface temperature TEX$_{86}^H$ calibration[9]. **b**, Zonal mean seawater temperatures from selected depth ranges show that the Equator-to-pole gradient decreases with increasing depths. **c**, Global core-top TEX$_{86}^H$ values shown against temperature of selected integrated water depths, respectively. Every colour shows the same full core-top data set. Grey shaded area denotes regressions that are both statistically indistinguishable from the SST calibration (hatched area in panel **a**) and constrained by U$_{37}^{K'}$ variability (95% confidence interval).

global approach provides a first-order correction to the calibration, individual records may still be affected by site-specific non-thermal effects (Supplementary Information).

Second, the attribution to subsurface temperatures instead of SSTs leads to different physical interpretations, especially when interpreting the timing of climate changes derived from multiple proxies and when analysing spatial patterns such as the latitudinal temperature gradient.

Both facts change the interpretation of early Eocene temperature reconstructions. The subsurface calibrations (Fig. 2a) yield an overall, significant decrease in TEX$_{86}^H$-inferred temperature estimates, invalidating the apparent tropical warmth at both poles, and reconciling the temperature mismatch between the deepwater and its source—that is, subducted surface waters from the high southern latitudes[30] (Fig. 3). They further bring TEX$_{86}^H$-inferred temperature into agreement with independent temperature estimates for bottom waters at shallow sites, where a relatively well-mixed water column below the mixed layer is expected (Fig. 3b). Consequently, our interpretation leads to a reduced latitudinal gradient in the Eocene warming, which is

expected, as even the modern temperature gradient at depths is considerably flatter than that at the sea surface (Fig. 2b). This is in stark contrast with the previous interpretation of TEX$_{86}^H$ as SST, where the weak latitudinal TEX$_{86}^H$ temperature gradient would suggest a fundamentally different climate from the modern day.

Our finding thus substantially alters the model–data comparison of TEX$_{86}^H$-derived early Eocene temperatures. Analysing a state-of-the-art model simulation of the early Eocene (ECHAM5/MPI-OM[10]; 2 × CO$_2$) shows, as previously noted[5], that the SST interpretation of TEX$_{86}^H$ is irreconcilable even with the warmest month SST (Fig. 3a). In contrast, our subsurface estimates are much closer (65% RMSE reduction) to the subsurface temperatures simulated for the corresponding water depths (Fig. 3b). Although model and data agree within uncertainties, the tendency of the recalibrated temperatures to be colder (warmer) in the tropics (mid/high southern latitudes) suggests that site-specific non-thermal effects on TEX$_{86}^H$ (Supplementary Information) or deficiencies in the model or forcing formulation are still in place, and call for a continued effort from both proxy and model communities.
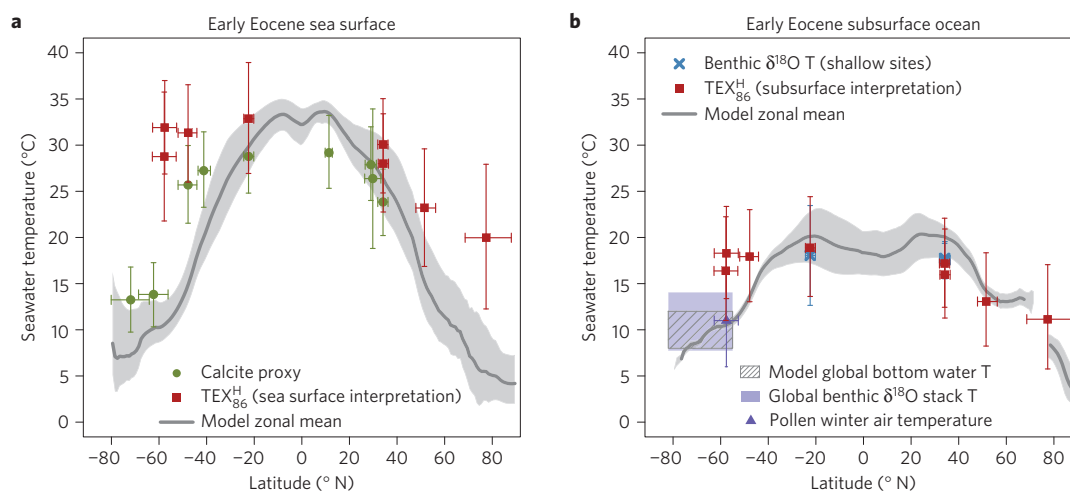
Figure 3 | Proxy–model comparison of early Eocene seawater temperatures. a, At sea surface using the standard TEX$_{86}^H$ calibration[9]. **b**, At subsurface ocean using our proposed calibrations. Shading of the simulated zonal mean temperature[10] indicates the seasonal range in **a** and the uncertainty (95% confidence interval) in the depth range recorded by TEX$_{86}^H$ in **b**. Error bars are 95% confidence intervals and show the combined uncertainty of the proxy temperature data (vertical), and the uncertainty in palaeolatitude estimates for 49–55 Myr (horizontal). The size of the blue box is the uncertainty of the temperature estimate (vertical) and the assumed latitudinal range of the deepwater formation in the southern high latitudes (horizontal)— that is, one of the major source regions in the global ocean[30]. The hatched box shows the range of simulated global bottom water temperatures, demonstrating the agreement between bottom water temperatures and southern high-latitude upper ocean temperatures in the model. Details on study sites are in Supplementary Table 1. Our model results are in agreement with the EoMIP ensemble, both in terms of zonal mean and temperature at the proxy record location (Supplementary Fig. 9).

Nevertheless, our results largely resolve the long-standing proxy–model discrepancy without having to force the models with extremely high CO$_2$ or to invoke missing physics in the models, as previously suggested[6,8]. The results thus not only lead to a more consistent multi-proxy based reconstruction of past climates, but also give credibility to the climate model-based projections of a future warmer world.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of this paper.

References

1. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
2. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–603 (2005).
3. Greenwood, D. R. & Wing, S. L. Eocene continental climates and latitudinal temperature gradients. *Geology* **23**, 1044–1048 (1995).
4. Bijl, P. K. *et al*. Early Palaeogene temperature evolution of the southwest Pacific Ocean. *Nature* **461**, 776–779 (2009).
5. Lunt, D. J. *et al*. A model–data comparison for a multi-model ensemble of early Eocene atmosphere–ocean simulations: EoMIP. *Clim. Past* **8**, 1717–1736 (2012).
6. Huber, M. & Caballero, R. The early Eocene equable climate problem revisited. *Clim. Past* **7**, 603–633 (2011).
7. Sagoo, N., Valdes, P., Flecker, R. & Gregoire, L. J. The early Eocene equable climate problem: can perturbations of climate model parameters identify possible solutions? *Phil. Trans. R. Soc. A* **371**, 20130123 (2013).
8. Sluijs, A. *et al*. Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene Thermal Maximum. *Nature* **441**, 610–613 (2006).
9. Kim, J.-H. *et al*. New indices and calibrations derived from the distribution of crenarchaeal isoprenoid tetraether lipids: implications for past sea surface temperature reconstructions. *Geochim. Cosmochim. Acta* **74**, 4639–4654 (2010).
10. Heinemann, M., Jungclaus, J. H. & Marotzke, J. Warm Paleocene/Eocene climate as simulated in ECHAM5/MPI-OM. *Clim. Past* **5**, 785–802 (2009).
11. Pearson, A. & Ingalls, A. E. Assessing the use of archaeal lipids as marine environmental proxies. *Annu. Rev. Earth Planet. Sci.* **41**, 359–384 (2013).
12. Schouten, S., Hopmans, E. C. & Sinninghe Damsté, J. S. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: a review. *Org. Geochem.* **54**, 19–61 (2013).
13. Sluijs, A. *et al*. Southern ocean warming, sea level and hydrological change during the Paleocene–Eocene thermal maximum. *Clim. Past* **7**, 47–61 (2011).
14. Pross, J. *et al*. Persistent near-tropical warmth on the Antarctic continent during the early Eocene epoch. *Nature* **488**, 73–77 (2012).
15. Schouten, S., Hopmans, E. C., Schefuß, E. & Sinninghe Damsté, J. S. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* **204**, 265–274 (2002).
16. Schouten, S., Forster, A., Panoto, F. E. & Sinninghe Damsté, J. S. Towards calibration of the TEX86 palaeothermometer for tropical sea surface temperatures in ancient greenhouse worlds. *Org. Geochem.* **38**, 1537–1546 (2007).
17. O' Brien, C. L. *et al*. High sea surface temperatures in tropical warm pools during the Pliocene. *Nature Geosci.* **7**, 606–611 (2014).
18. Ho, S. L. & Laepple, T. Glacial cooling as inferred from marine temperature proxies TEX$_{86}^H$ and U$_{37}^{K'}$. *Earth Planet. Sci. Lett.* **409**, 15–22 (2015).
19. Seki, O. *et al*. Assessment and calibration of TEX$_{86}$ paleothermometry in the Sea of Okhotsk and sub-polar North Pacific region: implications for paleoceanography. *Prog. Oceanogr.* **126**, 254–266 (2014).
20. Kim, J.-H. *et al*. Pronounced subsurface cooling of North Atlantic waters off northwest Africa during Dansgaard–Oeschger interstadials. *Earth Planet. Sci. Lett.* **339–340**, 95–102 (2012).
21. Ohkouchi, N., Eglinton, T. I., Keigwin, L. D. & Hayes, J. M. Spatial and temporal offsets between proxy records in a sediment drift. *Science* **298**, 1224–1227 (2002).
22. Rhines, A. & Huybers, P. Estimation of spectral power laws in time uncertain series of data with application to the Greenland Ice Sheet Project 2 δ^{18}O record. *J. Geophys. Res. Atmos.* **116**, D01103 (2011).
23. Laepple, T. & Huybers, P. Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability. *Earth Planet. Sci. Lett.* **375**, 418–429 (2013).
24. Laepple, T. & Huybers, P. Ocean surface temperature variability: large model–data differences at decadal and longer periods. *Proc. Natl Acad. Sci. USA* **111**, 16682–16687 (2014).
25. Qin, W. *et al*. Confounding effects of oxygen and temperature on the TEX$_{86}$ signature of marine Thaumarchaeota. *Proc. Natl Acad. Sci. USA* **112**, 10979–10984 (2015).

26. Conte, M. H. *et al*. Global temperature calibration of the alkenone unsaturation index ($U^{K'}_{37}$) in surface waters and comparison with surface sediments. *Geochem. Geophys. Geosyst.* **7,** Q02005 (2006).
27. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409,** 507–510 (2001).
28. Merbt, S. N. *et al*. Differential photoinhibition of bacterial and archaeal ammonia oxidation. *FEMS Microbiol. Lett.* **327,** 41–46 (2012).
29. Taylor, K. W. R., Huber, M., Hollis, C. J., Hernandez-Sanchez, M. T. & Pancost, R. D. Re-evaluating modern and Palaeogene GDGT distributions: implications for SST reconstructions. *Glob. Planet. Change* **108,** 158–174 (2013).
30. Thomas, D. J., Bralower, T. J. & Jones, C. E. Neodymium isotopic reconstruction of late Paleocene–early Eocene thermohaline circulation. *Earth Planet. Sci. Lett.* **209,** 309–322 (2003).

Author contributions

S.L.H. and T.L. contributed equally to the study.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to S.L.H. or T.L.

Competing financial interests

The authors declare no competing financial interests.

Methods

Proxy compilation and composite power spectra. We collected all available published paired TEX_{86}^H and $U_{37}^{K'}$ records spanning multi-millennial to multi-million-year timescales (Supplementary Fig. 2). As the signal-to-noise ratios of proxy-derived Holocene temperatures are low[23], we selected records that contain at least the last deglaciation (oldest sample >18 kyr BP).

We harmonized proxy temperature records by applying a common global calibration for each proxy type (ref. 31 for $U_{37}^{K'}$ and ref. 9 for TEX_{86}^H). For simplicity, we use in the following the same notation for the calibration of both proxies: $T(TEX86H) = \alpha_{TEX86H} + \beta_{TEX86H} TEX86H$ and $T(UK'37) = \alpha_{UK'37} + \beta_{UK'37} UK'37$, with $\beta_{UK'37} = 1/b$ and $\alpha_{UK'37} = -a/b$, and with intercept a and slope b as given by ref. 31.

For the spectral estimation, we followed the procedure described in ref. 24. Before the spectral analysis, all time series were evenly interpolated on a resolution determined separately for each proxy to minimize variance biases near the Nyquist frequencies (simulations with beta = 1.5, ≥70% of the variance preserved). Spectra were estimated using Thomson's multitaper method[32], with three windows on the detrended time series. The lowest frequency is omitted to minimize the bias caused by the multitaper method. As we are interested in the overall shape of the spectra, power spectra were smoothed using a Gaussian kernel with a constant width in logarithmic frequency space of 0.05 kyr^{-1}.

To avoid discontinuities across frequencies where the number of available estimates change, we scaled the proxy spectra to an average value in the largest common frequency interval. This does not affect the amplitude relationship between proxies, because the spectra of both proxies are rescaled with the same weights. As not all records share a common frequency interval, we separated the records into three independent spectra, covering ranges of 1/0.5 kyr to 1/50 kyr, 1/5 kyr to 1/200 kyr and 1/200 kyr to 1/2,000 kyr. Similar results concerning the $U_{37}^{K'}$ to TEX_{86}^H ratio, albeit with expected artificial jumps at frequencies where the number of records changes, are obtained when using the average spectral energy over all records covering a specific frequency (Supplementary Fig. 6b), or when looking at each single spectra pair (Supplementary Fig. 3).

To test if different noise levels could cause the difference between the $U_{37}^{K'}$ and TEX_{86}^H spectra, we further estimate the spectra of the $U_{37}^{K'}$ records after adding white noise with a standard deviation of 2.2 °C corresponding to the difference of the calibration standard errors of the TEX_{86}^H (ref. 9) and $U_{37}^{K'}$ calibration[31]. Although this brings the overall variability of TEX_{86}^H and $U_{37}^{K'}$ into better agreement, it increases the discrepancy of the spectral shape by resulting in too much high-frequency variability and equal/less low-frequency variability (Supplementary Fig. 3).

Ratio of subsurface over sea surface temperature variability. The production/export depth of TEX_{86}^H is not well constrained, thus complicating the comparison of TEX_{86}^H and SST (for example, $U_{37}^{K'}$) based records. Fortunately, as we will demonstrate in the following, similar temperature variability in the subsurface and the surface ocean are to be expected in the average over many sites and on timescales long enough to allow the propagation of temperature signal to depths.

One piece of evidence comes from proxy data; we collected published records ($N = 16$, oldest sample >18 kyr BP) containing Mg/Ca-derived temperatures measured in parallel on surface and subsurface dwelling foraminifera and applied the same spectral estimation as for the $U_{37}^{K'}$–TEX_{86}^H collection. The paired Mg/Ca records show a mean ratio of subsurface/surface temperature standard deviation of 1.11 (0.91–1.32; 95% confidence interval) and the mean spectrum also confirms that, on most timescales, the subsurface variability in these records has an amplitude similar to the surface variability (Supplementary Fig. 6b). Interestingly, the mean temperature variability recorded by Mg/Ca corresponds better to the mean variability recorded by $U_{37}^{K'}$, and is again lower than the variability estimated from the classical TEX_{86}^H SST interpretation.

Independent information can be gained from climate model simulations. As no continuous simulation covering millions of years exists, and glacial–interglacial changes are an important contribution of the variability in our data set, we study the depth and seasonal structure of the oceanic temperature response on forced changes in the Last Glacial Maximum (LGM; mean of last 50 years of an LGM simulation versus mean of a pre-industrial control run) and high CO_2 (mean of year 140–150 of the 4 × CO_2 scenario simulation versus mean of year 1–10 of the same simulation). The simulations were obtained from the CMIP5/PMIP3 database (only the MPI-ESM results are shown here, but similar results were obtained with any model of the PMIP3 ensemble). We extracted the annual mean, local summer (JJA/DJF) and local winter (DJF/JJA) temperatures of the uppermost layer, as well as the integrated annual seawater temperature between 0 and 600 m.

The simulated mean temperature change at depth in both scenarios is very similar to the mean temperature change at the sea surface (for the 4 × CO_2 simulation: warming[0–600 m]/warming[surface] = 1.08; for LGM simulation: cooling[0–600 m]/cooling[surface] = 0.98), suggesting that comparable climate signal independent of the recording depth of the proxy can be expected when averaging over a large number of sites. This also holds for the recording season,

where we obtain a ratio of summer (winter) to annual mean temperature change for the LGM case of 1.02 (0.97) and for the 4 × CO_2 case of 1.03 (0.99).

We further subsampled the model ocean at random positions and generated distributions of the ratio of subsurface over sea surface mean temperature change depending on the number of sites analysed (Supplementary Fig. 5). For a single site, the range of the ratio of subsurface over sea surface temperature change varies considerably (95% CI; 0.55–1.91 for the LGM case and 0.46–2.01 for the 4 × CO_2 case). However, the more sites are averaged, the closer the mean value converges to one; for 22 sites, the mean ratio is constrained between 0.91 and 1.18 (0.93 and 1.27 for 4 × CO_2). We note that sampling the actual 22 proxy sites used in this study results in a ratio well within the expectations (Supplementary Fig. 5).

Interestingly, the remaining spread between subsurface and surface temperature change caused by sampling a limited number of sites obtained in our model experiment (relative standard deviation, rsd = 7% for the LGM case, 8% for the 4 × CO_2) is comparable to the spread obtained by bootstrapping the T(TEX86H) to T(UK'37) ratio (rsd = 10%), demonstrating that our subsurface hypothesis is consistent with the model results.

Amplitude ratios of TEX_{86}^H over $U_{37}^{K'}$ and test of $U_{37}^{K'}$ temperature limitation. Amplitude ratios between TEX_{86}^H- and $U_{37}^{K'}$-derived temperature variability $\sigma_{T(TEX86H)}/\sigma_{T(UK'37)}$ are defined here as the ratio of the standard deviations of time series, analysing only the time period recorded by both proxies. For 7% of the samples in the paired records, wherein proxies were not measured on the same sediment depth, the records were first brought to a common time basis by linear interpolation of the higher resolved record to the sample ages of the lower resolved record. Different approaches to determine the amplitude ratios, such as integrating the composite frequency spectra, led to similar results. Reduced sensitivity of $U_{37}^{K'}$ at the warm temperature end could dampen the $U_{37}^{K'}$-derived variability and affect the amplitude ratio. However, we obtain similar results when excluding high temperatures (for example, setting a temperature limit of 25 °C removes 40% of the data points, but the amplitude ratio distribution does not change significantly: mean ratio of 1.7 using all data versus 1.9 excluding warm samples, t-test $p = 0.5$).

Estimation of the TEX_{86}^H-to-T slope. The TEX_{86}^H-to-T slope, β_{TEX86H_v}, required to reconcile the variability recorded by both proxies, can be obtained by correcting the standard TEX_{86}^H regression slope β_{TEX86H_v} by the inferred ratio of TEX_{86}^H- and $U_{37}^{K'}$-derived temperature variability:

$$\beta_{TEX86H_v} = \beta_{TEX86H} / \overline{\left(\frac{\sigma_{T(TEX86H)}}{\sigma_{T(UK'37)}} \right)}$$

For proxy calibrations based on linear regression, the standard deviations in proxy temperatures can also be derived by multiplying the standard deviation in index values by the regression slope:

$$\sigma_{T(TEX86H)} = \sigma_{TEX86H} \times \beta_{TEX86H}$$

$$\sigma_{T(UK'37)} = \sigma_{UK'37} \times \beta_{UK'37}$$

Therefore

$$\begin{aligned} \beta_{TEX86H_v} &= \beta_{TEX86H} / \overline{\left(\frac{\sigma_{T(TEX86H)}}{\sigma_{T(UK'37)}} \right)} \\ &= \beta_{TEX86H} / \overline{\left(\frac{\sigma_{TEX86H} \times \beta_{TEX86H}}{\sigma_{UK'37} \times \beta_{UK'37}} \right)} \\ &= \beta_{UK'37} / \overline{\left(\frac{\sigma_{TEX86H}}{\sigma_{UK'37}} \right)} \end{aligned}$$

This is equivalent to directly estimating the calibration slope from the ratio between the variability of the TEX_{86}^H and $U_{37}^{K'}$ indices, as both calibrations are linear.

Using β_{TEX86H_v} as an estimate for the underlying TEX_{86}^H-to-temperature relationship assumes that the main sources of variability are shared by both proxies. This is a reasonable assumption as the total variability (mean variance of TEX_{86}^H inferred $T = 7.01$ °C^2) is much larger than the typical variability caused by the analytical uncertainty or the workup procedure (for example, 0.16–0.36 °C^2; refs 33,34), and TEX_{86}^H records, recalibrated with the new β_{TEX86H_v}, show a timescale dependence similar to $U_{37}^{K'}$ over a wider range of amplitudes (Fig. 1).

Uncertainty of the slope estimate. The uncertainty of the derived slope estimates stems from two independent sources, namely the uncertainty of ($\overline{\sigma_{TEX86H}/\sigma_{UK'37}}$) and the uncertainty of $\beta_{UK'37}$. As we hypothesize that the TEX_{86}^H is recording subsurface temperature, it is expected that, locally, the subsurface and surface variability may differ, leading to variations in the local $\sigma_{TEX86H}/\sigma_{UK'37}$ ratio.

As discussed before, proxy and model evidence both suggest that the subsurface to surface marine temperature variability ratio can be approximated as a distribution with a mean of one (Supplementary Figs 5 and 6). To include the uncertainty caused by the limited number of sites, we bootstrap our 22 ($\sigma_{TEX86H}/\sigma_{UK'37}$) ratios ($1 \times 10^5$ repetitions, drawing 22 ratios with replacement). The uncertainty of $\beta_{UK'37}$ is normally distributed with $\sigma = 0.49$ ($^\circ$C/UK'37) (ref. 31) and is combined with the bootstrapped ratios to form a distribution of β_{TEX86H_v} (Supplementary Fig. 7).

Calibration of the TEX$_{86}^H$ index and depth constraint. Core-top calibrations estimate the relationship between a spatial set of proxy values (here TEX$_{86}^H$) measured in core-tops and associated modern observed temperatures. Both variables, proxy values and temperature data, contain errors. A minimum error estimate is provided by the analytical uncertainty of the proxy index and the instrumental uncertainty of the temperature data set. However, the true errors are considerably larger and also contain, on the proxy side, the heterogeneity within the sediment and potential secondary influences such as nutrients. On the temperature side, they contain the attribution uncertainty to the recording depth and season, as well as the different time period sampled by the sedimentary proxy index and the instrumental data set. As any regression result (with an $R^2 < 1$) depends on the relative contribution of errors on both variables, an accurate estimate of the (relative) errors on TEX$_{86}^H$ and temperature would be needed, but these are largely unknown.

Whereas the widely used calibrations for U$_{37}^{K'}$ (refs 26,31,35) attribute all error to the proxy side by using temperature as the independent variable, the commonly used calibrations for TEX$_{86}^H$ ignore the error on the proxy side by employing a least square regression with temperature as the dependent variable (y-axis) and the TEX$_{86}^H$ index as independent variable (x-axis). This causes an underestimation of the TEX$_{86}^H$ calibration slope, and thus an artificially improved match between the inferred U$_{37}^{K'}$ and TEX$_{86}^H$ temperature amplitude. A recent method includes errors in both variables in a Bayesian approach[36], but this does not circumvent the problem that the relative error on each of the variables is largely unknown.

Here, we take a pragmatic approach and bracket the possible range of results by performing the regression in both directions (with TEX$_{86}^H$ as dependent variable, analogous to the current U$_{37}^{K'}$ approach; and with TEX$_{86}^H$ as the independent variable, analogous to the most commonly used TEX$_{86}^H$ calibrations).

We perform a least square regression between the global data set of TEX$_{86}^H$ values in surface sediments[9] and climatological depth-integrated temperatures extracted from the World Ocean Atlas 2009[37], following the procedure of Kim and colleagues[9]. The regression analysis is performed for any start and end depth level between the sea surface and 1,000 m, at 50 m steps (for instance 0–50 m, 0–100 m, ..., 50–100 m). TEX$_{86}^H$ core-top sites shallower than 1,000 m are excluded to ensure that the comparison of TEX$_{86}^H$–temperature correlation at various depths is based on the same data set. This approach implicitly removes coastal sites at which the water temperature data are more uncertain due to the coarse grid resolution at the ocean–land boundary not resolving features of the dynamic hydrography such as upwelling, which may bias the TEX$_{86}^H$-temperature relationship. In the first step, all regression models that result in an R^2 value not distinguishable from the explained variance of the SST calibration are retained (Steiger's Z test[38] for the difference between two correlations sharing a common variable, $p = 0.05$, assuming 50 effective spatial degrees of freedom to account for the spatial autocorrelation). In the second step, from this remaining subset we choose all calibrations that are consistent with the prescribed slopes (± 1) (Supplementary Fig. 7). To propagate the slope uncertainty, we perform this step using 1,000 slopes drawn from the bootstrapped distribution described in the previous section on 'Uncertainty of the slope estimate', and perform the regressions with TEX$_{86}^H$ as dependent and independent variables, respectively. This results in an ensemble of 7461 TEX$_{86}^H$ calibrations, termed SUBCAL for the sake of simplicity. As the target temperatures of these calibrations are from subsurface water depths, temperature estimates calculated from these calibrations are collectively referred to as subsurface seawater temperatures. We note that assuming a single depth range to describe the global habitat distribution is a simplification, as biophysical parameters such as light availability, nutrients or other water mass properties will ultimately determine where the signal is recorded. However, a constant depth range is a conservative assumption; Preliminary experiments show that following isopycnals or a certain nutrient (nitrate) level results in an even smaller temperature sensitivity of TEX$_{86}^H$ to temperature, as both parameters are partly following the isotherms, decreasing the temperature range spanned by the core-top locations.

Early Eocene model–proxy comparison. The proxy data set is largely based on the EoMIP compilation[5], with the addition of two new TEX$_{86}^H$ data sets[39,40]. We included all sites spanning 55–49 Myr ago, including data covering the Early Eocene Climatic Optimum, but not the Palaeocene–Eocene Thermal Maximum. Proxy temperature data are presented as the mean of the time slice. We revised the palaeolatitude estimates for each site using the online calculator (www.paleolatitude.org)[41] with the palaeomagnetic reference frame of ref. 42, and

show in Fig. 3 both the mean and the largest possible range within the interval of 49–55 Myr ago.

As the seawater δ^{18}O and Mg/Ca values during the early Eocene are not well constrained, we follow the EoMIP approach of considering a range of estimates for temperature estimation, that is, 3 to 5 mol mol^{-1} for seawater Mg/Ca, and -0.81 to -1‰ for seawater δ^{18}O. Calibrations used to estimate temperatures from calcite proxies are as in ref. 5; that is, the calibrations of ref. 43 for planktic foraminifera δ^{18}O, ref. 44 for planktic foraminifera Mg/Ca, and ref. 45 for the clumped isotopes of mollusc.

Temperature calculation from benthic foraminiferal δ^{18}O at shallow and/or neritic sites (that is, Tanzania[46], Bass River[47] and Wilson Lake[48]) and the global stack[49] for 49–55 Myr ago, was synchronized assuming -1‰ for ice-free world seawater δ^{18}O, -0.27‰ for VSMOW to VPDB conversion, and $+0.28$‰ for calcification disequilibrium of *Cibicidoides spp*[50].

We removed all TEX$_{86}^H$ data associated with Branched-Isoprenoid Tetraether (BIT) index values >0.3, as BIT values >0.3 suggest a potential warm bias caused by terrigenous glycerol dialkyl glycerol tetraethers[51]. Consequently, all TEX$_{86}^H$ data from Hatchetigbee Bluff and three data points from Tanzania were excluded. We are aware of alternative TEX$_{86}$ indices such as 1/TEX$_{86}$ (ref. 52) and TEX$_{86}^L$ (ref. 53), but the preference of TEX$_{86}^H$ for the Eocene sites in our compilation is also supported by a systematic site-specific assessment[40], which suggests that TEX$_{86}^H$-inferred early Eocene temperatures are more realistic for these study sites. TEX$_{86}^H$ values were converted to SST by using the global SST calibration of ref. 9 and to subsurface temperatures by using our subsurface calibration ensemble SUBCAL.

Following EoMIP, proxy-inferred early Eocene temperatures at single sites are compared to the latitudinal mean of simulated seawater temperatures as the palaeolongitude, as well as the Eocene model bathymetry, is uncertain. The ECHAM5/MPI-OM simulation of ref. 10 was carried out under $2 \times CO_2$ levels and using the palaeobathymetry of ref. 54. Despite having a larger polar warming than previous studies[10], both the annual mean and the range of monthly mean of latitudinal sea surface Eocene warming of this simulation are within two standard deviations of the mean EoMIP ensemble (Supplementary Fig. 9), suggesting that this simulation is largely comparable to others. Early Eocene SST values are taken from the uppermost level in the simulation, whereas the subsurface seawater temperatures are extracted from the same depth range as the target temperature for the calibrations applied. Simulated bottom water temperatures are extracted as the area-weighted mean of the deepest level for every grid box that is deeper than 1,000 m.

Uncertainty in inter-proxy and proxy–model comparison. As in the EoMIP study, proxy-derived temperature time-slice uncertainties are estimated using the variability within the defined time slice and the calibration model.

Uncertainties can be divided into:

(1) Variability within the defined time slice. This partly includes the uncertainty in boundary conditions of the models. We account for this uncertainty by drawing the proxy values from a normal distribution, with the mean and standard deviation estimated from the proxy estimates within the time slice.
(2) Uncertainty of the proxy to temperature relationship. As no good estimate of the temporal calibration uncertainty is available, we take the common approach of using the residual standard error from the spatial core-top calibration as estimate for the downcore uncertainty. For Mg/Ca and δ^{18}O we additionally account for the secular changes in Mg/Ca and the δ^{18}O of the seawater over time by drawing the Mg/Ca and δ^{18}O values from a uniform distribution covering the range of literature values provided in EoMIP (ref 5).
(3) Attribution uncertainty in season and/or water depth. The seasonal uncertainty is relevant for the SST calibration, and thus included as the range of the model temperatures in Fig. 3a. It has a minor effect on the subsurface calibration (Supplementary Fig. 10), and is therefore omitted for the subsurface temperature uncertainty. The depth uncertainty is considered by applying the depth ensemble calibration SUBCAL. It thus results in an ensemble of TEX$_{86}^H$ temperature estimates as well as an ensemble of model temperatures corresponding to the water depths in SUBCAL (Fig. 3b).

The three uncertainty contributions can be assumed to be independent, and are combined using 10,000 Monte Carlo realizations to derive a probability distribution for each temperature estimate.

Code availability. The R scripts used for the data analysis are available upon request from the authors.

Data availability. The non-Eocene data compilation that support the findings of this study are available in PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.149998). The Eocene data compilation is in the supplement of refs 5,39,40. The subsurface calibration ensemble is in the Supplementary Information.

References

31. Müller, P. J., Kirst, G., Ruhland, G., von Storch, I. & Rosell-Melé, A. Calibration of the alkenone paleotemperature index $U_{37}^{K'}$ based on core-tops from the eastern South Atlantic and the global ocean (60° N–60° S). *Geochim. Cosmochim. Acta* **62,** 1757–1772 (1998).

32. Percival, D. B. & Walden, A. T. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge Univ. Press, 1993).

33. McClymont, E. L. *et al*. Sea-surface temperature records of Termination 1 in the Gulf of California: challenges for seasonal and interannual analogues of tropical Pacific climate change. *Paleoceanography* **27,** PA2202 (2012).

34. Richey, J. N., Hollander, D. J., Flower, B. P. & Eglinton, T. I. Merging late Holocene molecular organic and foraminiferal-based geochemical records of sea surface temperature in the Gulf of Mexico. *Paleoceanography* **26,** PA1209 (2011).

35. Prahl, F. G., Muehlhausen, L. A. & Zahnle, D. L. Further evaluation of long-chain alkenones as indicators of paleoceanographic conditions. *Geochim. Cosmochim. Acta* **52,** 2303–2310 (1988).

36. Tierney, J. E. & Tingley, M. P. A Bayesian, spatially-varying calibration model for the TEX$_{86}$ proxy. *Geochim. Cosmochim. Acta* **127,** 83–106 (2014).

37. Locarnini, R. A. *et al*. *NOAA Atlas NESDIS 68* 184 (US Government Printing Office, 2010).

38. Steiger, J. H. Test for comparing elements of a correlation matrix. *Psychol. Bull.* **87,** 245–251 (1980).

39. Bijl, P. K. *et al*. Eocene cooling linked to early flow across the Tasmanian Gateway. *Proc. Natl Acad. Sci. USA* **110,** 9645–9650 (2013).

40. Frieling, J. *et al*. Paleocene–Eocene warming and biotic response in the epicontinental West Siberian Sea. *Geology* **42,** 767–770 (2014).

41. van Hinsbergen, D. J. J. *et al*. A paleolatitude calculator for paleoclimate studies. *PLoS ONE* **10,** e0126946 (2015).

42. Torsvik, T. H. *et al*. Phanerozoic polar wander, palaeogeography and dynamics. *Earth-Sci. Rev.* **114,** 325–368 (2012).

43. Bemis, B. E., Spero, H. J., Bijma, J. & Lea, D. W. Reevaluation of the oxygen isotopic composition of planktonic foraminifera: experimental results and revised paleotemperature equations. *Paleoceanography* **13,** 150–160 (1998).

44. Anand, P., Elderfield, H. & Conte, M. H. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18,** 1050 (2003).

45. Ghosh, P. *et al*. 13C–18O bonds in carbonate minerals: a new kind of paleothermometer. *Geochim. Cosmochim. Acta* **70,** 1439–1456 (2006).

46. Pearson, P. N. *et al*. Stable warm tropical climate through the Eocene epoch. *Geology* **35,** 211–214 (2007).

47. John, C. M. *et al*. North American continental margin records of the Paleocene–Eocene thermal maximum: implications for global carbon and hydrological cycling. *Paleoceanography* **23,** PA2217 (2008).

48. Zachos, J. C. *et al*. Extreme warming of mid-latitude coastal ocean during the Paleocene–Eocene Thermal Maximum: inferences from TEX$_{86}$ and isotope data. *Geology* **34,** 737–740 (2006).

49. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451,** 279–283 (2008).

50. Dunkley Jones, T. *et al*. Climate model and proxy data constraints on ocean warming across the Paleocene–Eocene Thermal Maximum. *Earth-Sci. Rev.* **125,** 123–145 (2013).

51. Weijers, J. W. H., Schouten, S., Spaargaren, O. C. & Sinninghe Damsté, J. S. Occurrence and distribution of tetraether membrane lipids in soils: implications for the use of the TEX$_{86}$ proxy and the BIT index. *Org. Geochem.* **37,** 1680–1693 (2006).

52. Liu, Z. *et al*. Global cooling during the Eocene–Oligocene climate transition. *Science* **323,** 1187–1190 (2009).

53. Kim, J.-H. *et al*. Holocene subsurface temperature variability in the eastern Antarctic continental margin. *Geophys. Res. Lett.* **39,** L06705 (2012).

54. Bice, K. L. & Marotzke, J. Numerical evidence against reversed thermohaline circulation in the warm Paleocene/Eocene ocean. *J. Geophys. Res. Oceans* **106,** 11529–11542 (2001).