PNAS RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES

# The Long chain Diol Index: A marine palaeotemperature proxy based on eustigmatophyte lipids that records the warmest seasons

Sebastiaan W. Rampen[a,1], Thomas Friedl[b], Nataliya Rybalka[b], and Volker Thiel[a]

Edited by Roger Summons, Massachusetts Institute of Technology, Cambridge, MA; received September 13, 2021; accepted March 11, 2022

**Long chain 1,13- and 1,15-diols are lipids which are omnipresent in marine environments, and the Long chain Diol Index (LDI), based on their distributions, has previously been introduced as a proxy for sea surface temperature. The main biological sources for long chain 1,13- and 1,15-diols have remained unknown, but our combined lipid and 23S ribosomal RNA (23S rRNA) analyses on suspended particulate matter from the Mediterranean Sea demonstrate that these lipids are produced by a marine eustigmatophyte group that originated before the currently known eustigmatophytes diversified. The 18S rRNA data confirm the existence of early-branching marine eustigmatophytes, which occur at a global scale. Differences between LDI records and other paleotemperature proxies are generally attributed to differences between the seasons in which the proxy-related organisms occur. Our results, combined with available LDI data from surface sediments, indicate that the LDI primarily registers temperatures from the warmest month when mixed-layer depths, salinity, and nutrient concentrations are low. The LDI may not be applicable in areas where *Proboscia* diatoms contribute 1,13-diols, but this can be recognized by enhanced contributions of $C_{28}$ 1,12 diol. Freshwater input may also affect the correlation between temperature and the LDI, but relative $C_{32}$ 1,15-diol abundances help to identify and correct for these effects. When taking those factors into account, the calibration error of the LDI is 2.4 °C. As a well-defined proxy for temperatures of the warmest seasons, the LDI can unlock important and previously inaccessible paleoclimate information and will thereby substantially improve our understanding of past climate conditions.**

Eustigmatophyceae | long chain diols | LDI | biomarkers | palaeotemperature proxy

## Significance

Organisms adjust their lipid compositions to environmental conditions like temperature. Some adaptations involve changes in ratios between unique lipids, which can still be determined after millions of years. Analyzing such lipid ratios in ancient sediments can be used for climate reconstruction. One example is the Long chain Diol Index (LDI). The lipids, produced by unknown biological sources, occur ubiquitously in marine sediments, and the LDI correlates with temperature. However, LDI records often differ from other temperature reconstructions. We now show that a clade of eustigmatophyte algae produces these lipids, providing information on their evolutionary history. We further demonstrate that the LDI registers temperatures of the warmest months instead of annual mean temperatures, which is valuable information for climate research.

Long chain alkyl diols (LCDs) are lipids consisting of linear alkyl chains with an alcohol group at the primary carbon atom and one at a midchain carbon position. These lipids predominantly consist of even-numbered carbon chains with 28 to 32 carbon atoms and midchain alcohol groups positioned at C-13, C-14, or C-15. Upon being reported in Black Sea sediments (1), it became clear that these lipids occur ubiquitously in both marine and lacustrine environments (2–4). Volkman et al. were the first to report the occurrence of long chain 1,13- and 1,15-diols in both marine and freshwater eustigmatophyte algae (Eustigmatophyceae, Stramenopiles) (5, 6). Later, Sinninghe Damsté et al. identified long chain 1,14-diols in species of the marine diatom genus *Proboscia* (Bacillariophyceae, Stramenopiles) (7), and Rampen et al. reported the occurrence of 1,14-LCDs in the algal species *Apedinella radians* (Dictyochophyceae, Stramenopiles) (8). Species of *Proboscia* diatoms and 1,14-LCDs have both been reported to be abundantly present in seasonal upwelling areas like the Arabian Sea (9, 10), the Peruvian coast (11–13), and Antarctic waters (14, 15). Similarities in fluxes of 1,14-LCDs and the presence of *Proboscia* cell walls (valves) in the Arabian Sea further support the assumption that *Proboscia* spp. are a likely source for 1,14-LCDs (16). So far, there are no indications of *Apedinella* contributing to environmental LCDs (4).

Originally, eustigmatophytes were considered to comprise a small group of algae (17, 18), but with the use of more advanced molecular characterization methods, a more complex phylogeny of the Eustigmatophyceae class has been revealed. An environmental DNA survey in the East African Lake Challa, for example, demonstrated a diversity of eustigmatophytes which far exceeded the classification based on cultivated species at that time (19). Furthermore, the number of eustigmatophyte gene copies detected in that study mirrored LCD concentrations at various depths, confirming the importance of eustigmatophytes as the producers of LCDs in Lake Challa. The current knowledge suggests that these algae primarily occur in freshwater and soil and that it is in these environments where eustigmatophytes emerged (20–22). Only the closely related genera *Nannochloropsis* and *Microchloropsis* are known to contain marine species.

Author contributions: S.W.R., T.F., and V.T. designed research; S.W.R. and N.R. performed research; T.F. and V.T. contributed new reagents/analytical tools; S.W.R. analyzed data; and S.W.R., T.F., and V.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: sebastiaanrampen@yahoo.com.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116812119/-/DCSupplemental.

Published April 11, 2022.

**Table 1. Sample locations, sampling dates, and surface water temperatures (°C) at the time of sampling**

| Station | Location | Surface water temperature (°C) on sampling days in 2019 | | | | | |
|---|---|---|---|---|---|---|---|
| | | 30 Apr | 25 Jun | 2 Aug | 27 Aug | 17 Sep | 8 Oct |
| Med. 1 | 43.048N 5.800E | 16.0 | 21.7 | 19.7 | 26.0 | 19.9 | 19.9 |
| Med. 2 | 43.013N 5.841E | 14.3 | 21.0 | 21.7 | 26.0 | 20.7 | 20.7 |
| Med. 3 | 43.026N 5.751E | 15.0 | 21.4 | 20.4 | 26.1 | 20.0 | 20.0 |
| Med. 4 | 43.102N 5.727E | 15.2 | 21.8 | 20.6 | 25.9 | 19.7 | 19.7 |

Based on their phylogenetic position within the eustigmatophytes, it can be concluded that these marine species recently evolved from freshwater ancestors (21, 23–25). Therefore, they cannot account for the oldest LCDs, which have been identified up to the Cretaceous (26).

So far, there are very few reports on eustigmatophytes in marine environments. In a study on LCDs and their possible sources in the western tropical North Atlantic Ocean, sequencing analyses only identified low amounts of eustigmatophytes in 8 out of 68 suspended particulate matter (SPM) samples, whereas all samples contained LCDs, dominated by $C_{30}$ 1,15-diol (27). Currently known marine eustigmatophytes contain relatively high abundances of monounsaturated and saturated $C_{32}$ 1,15-diol. In marine environments, however, the concentration of monounsaturated $C_{32}$ 1,15-diol is generally very low or below the detection limit, and saturated $C_{32}$ 1,15-diol generally contributes <30% of the total LCDs (3, 6). A few other marine algae have been shown to produce LCDs, but again, their LCD distributions do not match general distributions in marine environments (27). LCDs have also been reported in environmental samples with high contributions of the cyanobacterium *Aphanizomenon flos-aquae*, but analyses of lipids in cultures did not confirm the presence of LCDs in these organisms (28, 29). Hence, so far, no major source could be linked to 1,13- and 1,15-LCDs in marine environments.

Several LCD indices have been proposed as proxies for past environmental conditions. The diol index, a ratio between $C_{30}$ and $C_{32}$ 1,15-diols, was introduced as a proxy for freshwater input or upwelling conditions (30). Ratios of $C_{28}$ + $C_{30}$ 1,14-diols vs. $C_{30}$ 1,15-diol or $C_{28}$ + $C_{30}$ 1,13-diols have also been suggested as upwelling proxies (15, 16). In addition, the nutrient diol index, another index based on 1,14 vs. 1,13-, 1,14-, and 1,15-LCDs, has been introduced as a proxy for sea surface nutrients (31). The fractional abundance of $C_{32}$ 1,15-diol vs. combined $C_{28}$ and $C_{30}$ 1,13-diol + $C_{30}$ and $C_{32}$ 1,15-diol has been suggested to be a marker for riverine input in the marine environment (32, 33), and the Long chain Diol Index (LDI) has been proposed as a proxy for temperature for both freshwater and marine environments (3, 34–38).

$$\text{Long chain Diol Index (LDI)} = [C_{30}\,1,15\text{−diol}]/([C_{28}\,1,13- + C_{30}\,1,13- + C_{30}\,1,15\text{−diols}])$$

$C_{28}$ 1,13- and $C_{30}$ 1,15-diols have been reported in samples up to 74 Myr old, but LDI values in those samples were probably affected by *Proboscia* spp. contributing additional $C_{28}$ 1,13-diol (26). This is one out of many examples demonstrating that the use of LCD indices remains challenging, and knowing the main sources of marine 1,13- and 1,15-LCDs (hereafter indicated as LDI-LCDs) is critical for improving the application of these proxies. The current study was aimed to increase our knowledge on LDI-LCDs and proxies that are based on these lipids. To identify the LDI-LCD sources, we analyzed LCD content and plastid 23S rRNA (p23S rRNA) gene compositions in SPM from surface waters of the Mediterranean Sea near the Gulf of Lion (*SI Appendix*, Fig. S1 and Table 1) collected in spring, summer, and fall of 2019. With the knowledge obtained from these analyses, we critically examined existing DNA data to test our conclusions and to expand on them. Finally, we combined datasets of published LCD distributions from marine surface sediments and environmental data from their sampling locations to examine the correlation between the LDI and temperature in light of our insights. This resulted in a better-defined application of the LDI as a proxy recording the temperatures of the warmest seasons.

## Results and Discussion

**Occurrence of LCDs and Potential Source Organisms in the Mediterranean Sea.** In order to gain insight into the seasonal occurrence of LCDs and their source organisms, Mediterranean Sea surface water samples were taken from mid-spring to early autumn in 2019 (Table 1). The coldest in situ temperatures were measured on 30 April, when sea surface temperatures (SSTs) ranged between 14.3 and 16.0 °C, and the warmest SSTs of 26 °C were measured on 27 August. On the other sampling days, SSTs ranged between 20 and 22 °C. Measured particulate organic carbon (POC) values gradually decreased from 248 μg $L^{-1}$ in April to 167 μg $L^{-1}$ in August, with a minor increase to 198 μg $L^{-1}$ in September (*SI Appendix*, Table S1).

LCDs were detected as free lipids in all samples from the Mediterranean Sea, and concentrations were not affected by acid or base hydrolysis (see *SI Appendix*, Supplementary Note 1). LDI-LCD concentrations increased substantially between 2 and 27 August, when a strong increase in SSTs was observed (Fig. 1 *A* and *G*, Table 1, and *SI Appendix*, Table S1). In accordance with the SSTs, LDI-LCD concentrations on 17 September were decreased again, suggesting that the sources of LDI-LCDs thrive under warm water conditions. In the Mediterranean Sea water, LDI-LCD concentrations, as well as $C_{30}$ 1,15-keto-ol concentrations, followed a distinct pattern over time, reaching maximum values for 27 August (Fig. 1 *A* and *B* and *SI Appendix*, Supplementary Note 2). These concentrations were in the same range as sterols (Fig. 1 *D*–*F*), suggesting a relatively high abundance of LDI-LCD sources at that time. Comparison of LDI-LCD concentration patterns with p23S rRNA read numbers for individual amplicon sequence variants (ASVs) was performed to ascertain the LDI-LCD sources. The ASV numbering is based on the order of total read counts. Read numbers of ASV_003 (i.e., the third most abundant ASV) showed the strongest resemblance to the LDI-LCD concentrations (Fig. 1 and *SI Appendix*, Tables S1 and S2) and was assigned to the cyanobacterium *Synechococcus* sp. ASV_001 and ASV_002 were also assigned to *Synechococcus* cyanobacteria, but their reads showed high numbers on 25 June and 2 August, when LCD concentrations were low. An explanation is that

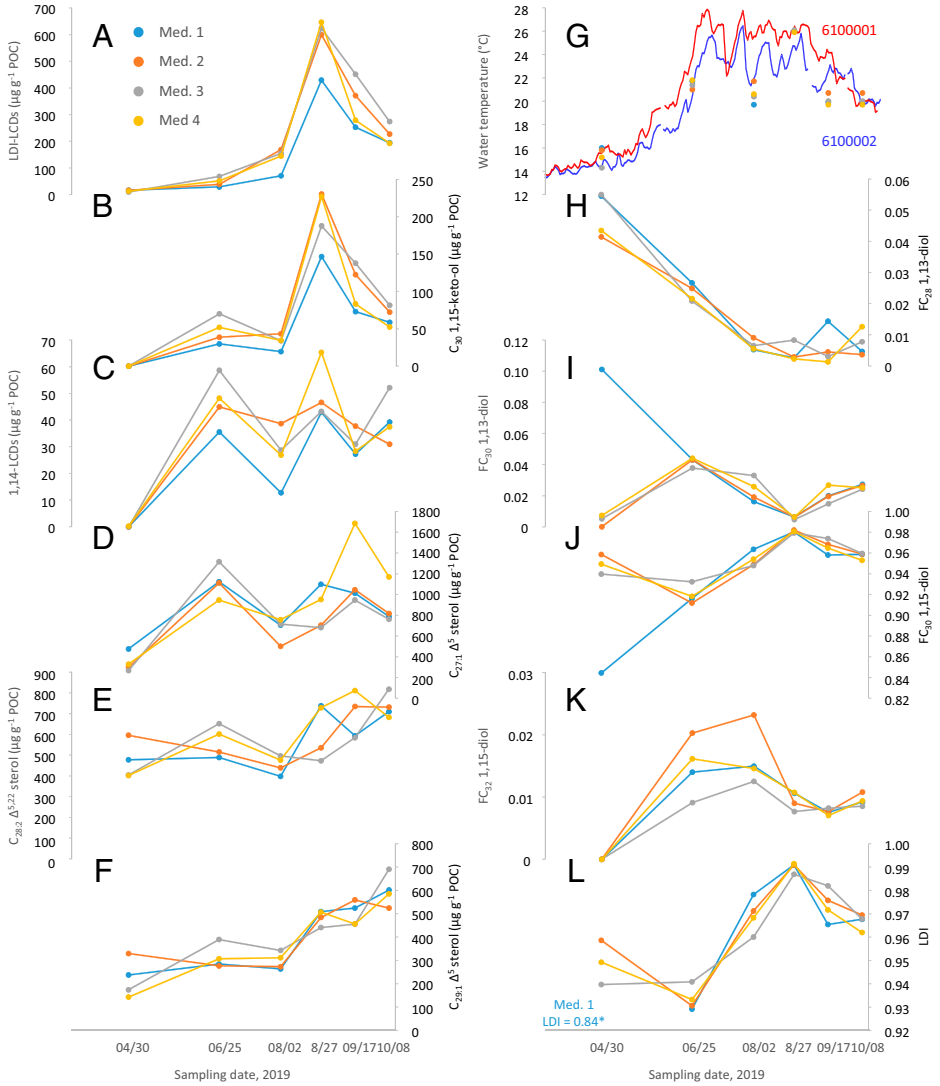


**Fig. 1.** Plots of lipid concentrations and fractional abundances of lipids from Mediterranean Sea samples and surface water temperatures (also see *SI Appendix*, Table S1). (*A–F*) Concentrations of combined LDI-LCDs, $C_{30}$ 1,15-keto-ol, 1,14-LCDs, cholesterol (cholest-5-en-3β-ol), brassicasterol (24-methylcholesta-5,22-dien-3β-ol), and β-sitosterol (24-ethylcholest-5-en-3β-ol). (*G*) Surface water temperatures (°C) measured in situ at the time of sampling (circles) and daily mean temperatures obtained by marine climate observation buoys (https://donneespubliques.meteofrance.fr) located northeast (6100001; BOUEE_COTE D'AZUR) and southwest (61000002; BOUEE_LION) of the sampling locations. (*H–K*) Fractional $C_{28}$ 1,13-diol, $C_{30}$ 1,13-diol, $C_{30}$ 1,15-diol, and $C_{32}$ 1,15-diol concentrations, relative to the combined concentrations of these LCDs. (*L*) LDI values. The LDI value for Med. 1 on 30 April 2019 was 0.84.

only a select group of *Synechococcus* species produces LCDs, something similar as observed for the diatoms, where only *Proboscia* spp. produce 1,14-LCDs (7).

ASV_005, which was not directly assigned to a specific organism by the basic local alignment search tool (BLAST) procedure, also showed a pattern similar to LDI-LCD concentrations (Figs. 1 and 2 and *SI Appendix*, Tables S1 and S2). These genes were absent in samples from 30 April and 25 September, whereas 64 reads were detected on 2 August, and 868 reads were recorded on 27 August. The highest number of ASV_005 reads was identified on 17 September, both for absolute numbers (1,117 reads) and relative to the total reads per sample (13%). Comparing this gene sequence with data from GenBank (https://www.ncbi.nlm.nih.gov/genbank/), the closest available organism, at 91.6% identity was the eustigmatophyte *Nannochloropsis oceanica* strain LAMB2011 (accession number CP038136). To get a more accurate identification of ASV_005, we performed phylogenetic analyses on our p23S rRNA dataset combined with relevant corresponding sequences obtained from GenBank. In the obtained phylogenetic trees (Fig. 2 and *SI Appendix*, Figs. S2–S5), ASV_005 formed a cluster with ASV_026 and ASV_031 which branched off from the known eustigmatophytes before their species diversified. Separating eukaryotes from the prokaryote reads, ASV_005 contributed 76% of all eukaryote reads for 27 August and 59% of the reads for 17 September, a pattern similar to the LDI-LCD concentrations (Fig. 2). ASV_005 was the only ASV tentatively linked to eustigmatophytes for 27 August; ASV_026 contributed an additional 27 and 3% of all eukaryote reads on 2 August and 17 September, respectively; and ASV_031 contributed 1, 9, and 5% of all eukaryote reads for 30 April, 25 June, and 2 August, respectively. These results suggest a shift in LDI-LCD producing species over time, which could be related to SST, nutrient availability, or other environmental conditions. Given the similarity in sequence reads and LDI-LCD concentrations over time and the occurrence of LDI-LCDs in eustigmatophytes, these algae are a more likely source for the LDI-LCDs than cyanobacteria.

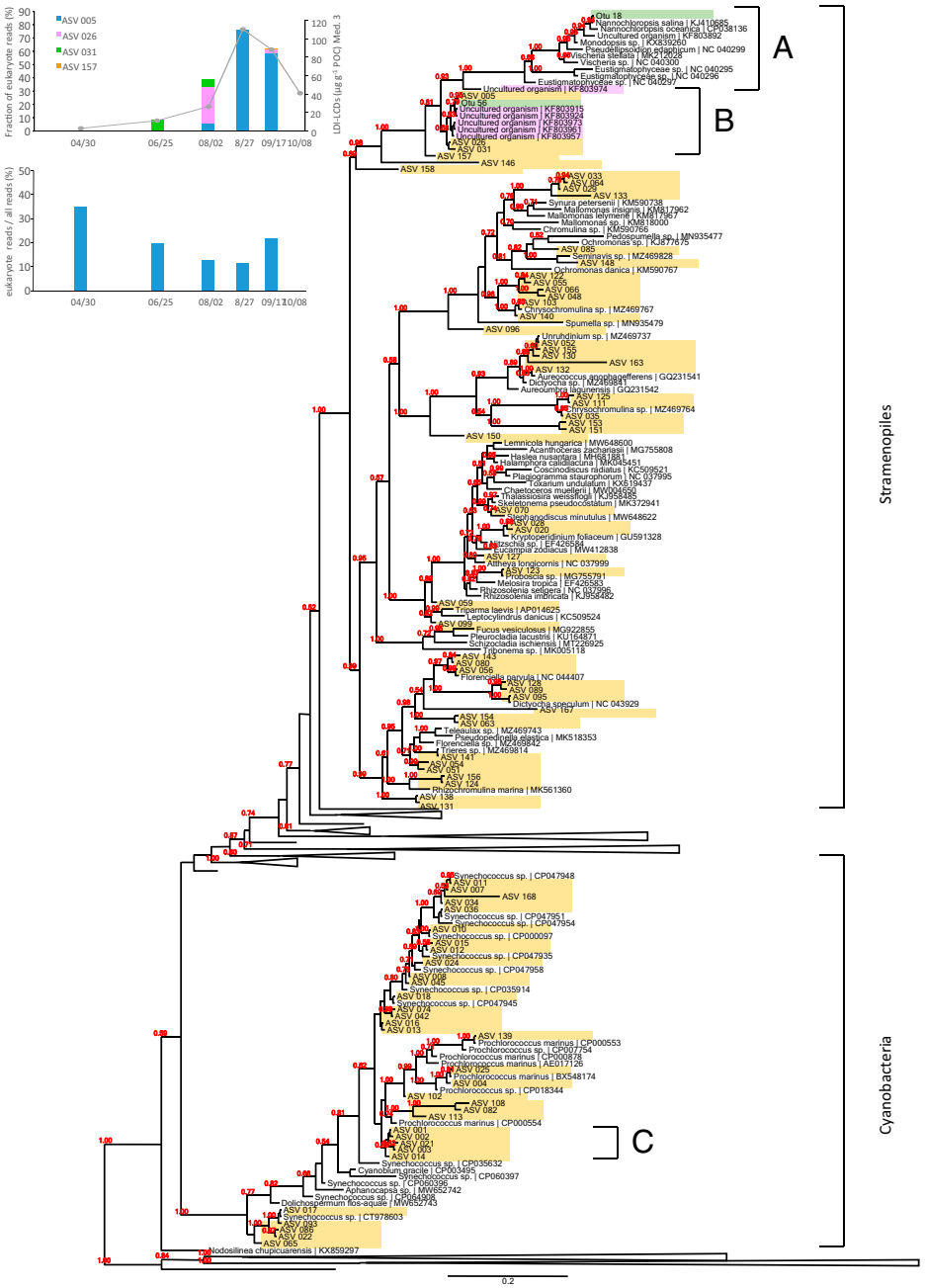


**Fig. 2.** Simplified MrBayes phylogenetic tree based on p23S rRNA sequences recovered from Mediterranean Sea SPM (orange background), combined with sequences obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). The purple background indicates sequences from the salt lake Qinghai (39), and the green background indicates OTUs from the Mariana Trench (40). The red numbers at the nodes indicate PP values. PP values below 0.5 are omitted. Here, "A" indicates the clade consisting of known eustigmatophyte algae, predominantly belonging to freshwater and soil environments; "B" indicates the marine eustigmatophytes; and "C" indicates the clade of cyanobacteria including ASV_001, ASV_002, and ASV_003. (*Inset*) The number of reads for potential eustigmatophyte OTUs, relative to the total number of eukaryote reads; the concentration of LDI-LCDs in SPM from sample location Med. 3; and the number of eukaryote reads relative to the combined eukaryote and prokaryote reads. See *SI Appendix*, Fig. S2 for the complete phylogenetic tree.

In the calculated phylogenetic trees based on p23S sequences, the distinctive ASVs clustered together with six gene fragments extracted from Holocene and Late Pleistocene sediments from the salt lake Qinghai (39) and with one operational taxonomic unit (OTU) from water column samples from the Mariana Trench (40) (Fig. 2 and *SI Appendix*, Figs. S2–S5, cluster b). Indeed, LDI-LCDs have previously been reported in sediments from both environments (41, 42). In the study of the Mariana Trench, eustigmatophyte genes were reported from the bathypelagic and hadal zones (40). We followed the procedures described in that study to cluster their 23S rRNA data into OTUs (40) but used the improved USEARCH version 11.0 instead of version 7.0 (43). With that approach, we obtained only one eustigmatophyte OTU (OTU_18) clustering with *Nannochloropsis* species instead of two OTUs that were reported to have high read numbers at 8,320- and 4,000-m depth. In addition to this *Nannochloropsis*-related sequence, we obtained OTU_56, clustering within the eustigmatophyte sister group identified in our Mediterranean Sea study. More than 85% of this OTU was recorded in the upper epipelagic zone at 4-m depth, and the remaining reads were distributed over all other depths.

Whereas a link between eustigmatophytes and LDI-LCDs seems highly plausible in the current work, the study by

Balzano and coworkers (27) aimed at identifying the sources of LCDs in the western tropical North Atlantic Ocean and detected eustigmatophyte reads in only 2 out of 22 surface water samples, and the read numbers were in no proportion to the measured LDI-concentrations. Acknowledging the limitations of the automated BLAST results, we performed phylogenetic analyses on their DNA data for improved identification (Fig. 3 and *SI Appendix*, Figs. S6–S9). The dataset from Balzano et al. was combined with 18S DNA data from 19 sampling sites across the coastal areas of the Mediterranean Sea (44) and with 18S V4 region sequences from eukaryote species obtained from GenBank. Our results confirm the identification of denovo2075 as a species closely related to *Nannochloropsis*. This species was exclusively found in the surface waters of station 12, one of the shallowest stations located on the continental shelf near the coast (27). Our results do not support the identification of denovo229 as an unspecified eustigmatophyte. On the other hand, our phylogenetic analyses suggest that denovo162, denovo1571, and denovo2431 belonged to an early-branching clade of eustigmatophytes. Denovo162 was identified in almost half of the surface and mixed-layer (15-m depth) water samples from locations farther off the coast (27), but nevertheless, the number of reads is far too low to link the detected LCD concentrations directly to these species. Based on the concentrations of LCDs detected in *N. oceanica* (45), Balzano et al. indicated that even if all of the nanoplankton organisms detected in their study produced LCDs, the total cell number would still have been too low to explain the measured LCD concentrations (27). It was therefore hypothesized that the LCDs originated predominantly from degraded material whose DNA was not preserved.

In the dataset from Penna et al., eustigmatophyte sequences were detected in 18 out of 20 Mediterranean Sea samples (44). An additional BLAST search in GenBank targeting the early-branching eustigmatophyte sequences revealed several other related sequences from various marine settings. These results support the existence of a globally present, yet still unidentified, group of marine eustigmatophytes that originated before the known clade of predominantly freshwater eustigmatophytes diversified. Phylogenetic analyses on an 18S rRNA dataset consisting of sequences with >1,000 base pairs further substantiate the existence of this early-branching group of marine eustigmatophytes (*SI Appendix*, Figs. S10 and S11). In all 18S rRNA phylogenetic trees, the potential LDI-LCD producers cluster as a sister group to the currently known eustigmatophytes (Fig. 3 and *SI Appendix*, Figs. S6–S11), and the same is observed for the Maximum Likelihood phylogenetic tree based on a subset of p23S rRNA sequences (*SI Appendix*, Fig. S5). Bootstrap or posterior probability (PP) values for the p23S rRNA phylogenetic trees in Fig. 2 and *SI Appendix*, Figs. S2–S4 do not reject this possibility. However, there is still the possibility that the clade of potential LDI-LCD producers represents another lineage of stramenopile algae independent of Eustigmatophyceae. Additional sequence, morphological, and physiological data are still required to test if these LDI-LCD producers meet the criteria needed for inclusion in the Eustigmatophyceae.

**Constraining the LDI as a Paleotemperature Proxy.** At our sampling locations in the Mediterranean Sea, LDI-LCDs were predominant at the time with the highest SSTs (Fig. 1 *A* and *G*), which is consistent with earlier observations that the LDI temperature proxy shows the strongest correlation with summer


**Fig. 3.** Eustigmatophyte strains and a sister group extracted from the MrBayes phylogenetic tree based on a subset of 18S V4 region sequences obtained by Balzano et al. from the western North Atlantic Ocean (ref. 27; orange background) and Penna et al. from the Mediterranean Sea (ref. 44; green background), combined with closest relative sequences obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). The red numbers at the nodes indicate PP values. PP values below 0.5 are omitted. Here, "A" indicates the clade consisting of known eustigmatophyte algae, predominantly belonging to freshwater and soil environments, and "B" indicates the marine eustigmatophytes. See *SI Appendix*, Fig. S7 for the complete phylogenetic tree.

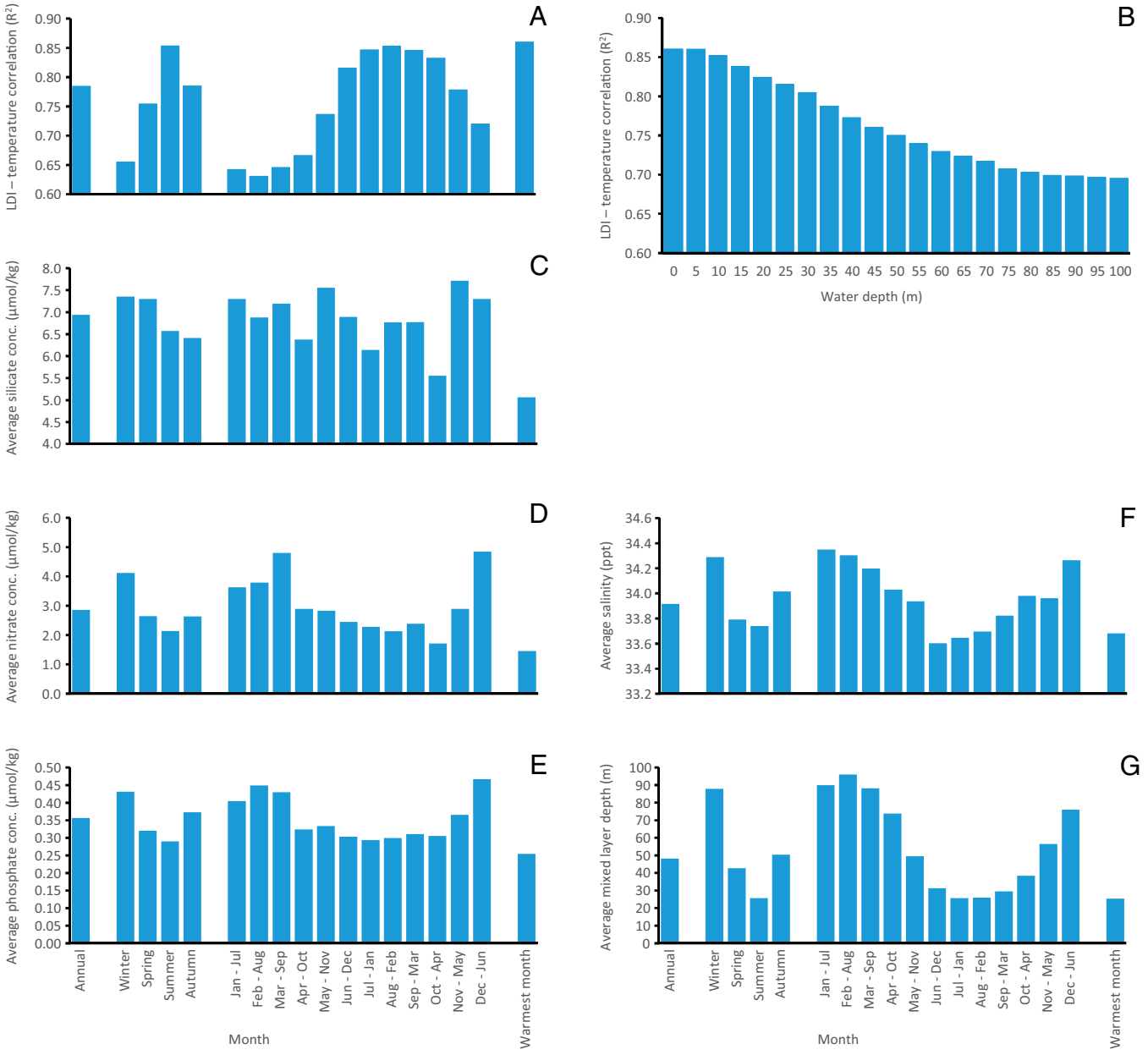


**Fig. 4.** Temperature, nutrient, salinity, and mixed-layer depth data (50) in relation to the combined LDI-LCD dataset obtained from the literature. (*A*) $R^2$ values for LDI values correlated with annual, seasonal, and monthly mean SST data and with the warmest mean monthly temperature data of the overlying water layers at 0-m depth at the individual sampling sites. (*B*) $R^2$ values for LDI values correlated with the warmest mean monthly temperature of the water layers at 0- to 100-m depth for the individual sampling sites. (*C*–*G*) Mean silicate, nitrate, and phosphate concentrations; salinity values; and mixed-layer depths from overlying water layers calculated for annual, seasonal, and monthly time periods and for the data obtained during the months with the maximum surface water temperature at the sampling locations. For seasonal and monthly data, the different seasonality of the northern and southern hemispheres was taken into account: winter is January, February, and March for the northern hemisphere and July, August, and September for the southern hemisphere, whereas the first month in the figure contains data from January for the northern hemisphere and July for the southern hemisphere.

SSTs (34, 35, 46). To explore this in more depth, we combined the surface sediment LCD datasets from a number of studies (34, 47–49) to compare LCD distributions with mean annual, seasonal, and monthly values of temperature, nutrients, salinity, and mixed-layer depths (Fig. 4) at a 1° and 0.25° grid resolution, obtained from the *World Ocean Atlas 2018* (50). We observed the strongest correlations between the LDI and surface water temperatures when using 1° grid datasets with data from all available years combined; if no data directly corresponding to the sample coordinates were available, data from the nearest adjacent grid cell were selected. Neither the use of data from 0.25° grid datasets nor the use of datasets from specific time spans resulted in stronger correlations.

We observed the strongest correlations for the LDI with SSTs from the warmest month at each of the sampling locations (Figs. 4*A* and 5 *A* and *B*). The water conditions during these months were characterized by, on average, low surface water salinities, low nutrient concentrations, and shallow mixing-layer depths (Fig. 4). These findings further corroborate that LDI-LCD producers mainly occur in the upper water layers during periods of enhanced stratification. The results also support the rationale of the 1,14-LCD seasonal upwelling indices; these are based on the assumption that 1,14-LCDs are mainly produced during periods of seasonal upwelling and high nutrient concentrations, whereas the LDI-LCDs are linked to nonupwelling periods (15, 16). The predominance of LDI-

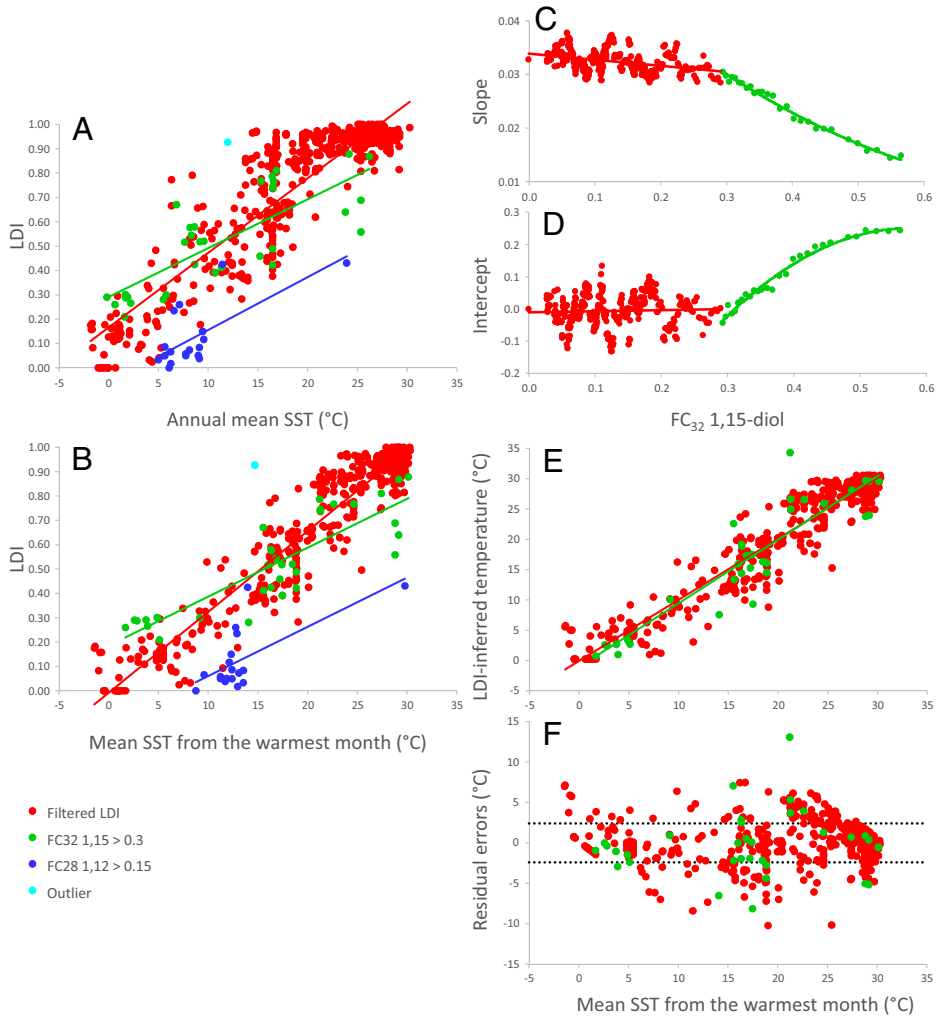


**Fig. 5.** Calibrating LDI values with temperature. Red circles (filtered LDI) indicate samples with $FC_{28}$ 1,12-diol values ≤0.15 and $FC_{32}$ 1,15-diol values ≤0.3. Sample M2001-15 BC from Porcupine Bank, North Atlantic Ocean (outlier), measured by De Bar et al. (35) was excluded from the dataset as its LCD distribution stood out (estimated − measured temperature >3 × SD), even though this sample did not meet the criteria for exclusion as discussed in this study. (*A*) LDI values versus annual mean SST at 0-m depth. (*B*) LDI values versus mean SST values from the warmest month at 0-m depth of each sampling location. (*C* and *D*) Slope and intercept values for subsets of LDI and mean SST values for the warmest month of each sampling location, plotted against $FC_{32}$ 1,15-diol values. Slopes and intercepts were calculated for subsets comprising 21 successive samples, which had been ordered on their $FC_{32}$ 1,15-diol values. For subsets with $FC_{32}$ 1,15-diol values <0.3 ($n$ = 444), slope = $-0.012 \times FC_{32}$ 1,15-diol + 0.034, $R^2$ = 0.131; intercept = $0.029 \times FC_{32}$ 1,15-diol − 0.010, $R^2$ = 0.001. For sample sets with $FC_{32}$ 1,15-diol >0.3 ($n$ = 28), slope = $0.072 \times (FC_{32}$ 1,15-diol$)^2 - 0.123 \times FC_{32}$ 1,15-diol + 0.061, $R^2$ = 0.987; intercept = $-3.561 \times (FC_{32}$ 1,15-diol$)^2 + 4.128 \times FC_{32}$ 1,15-diol − 0.943, $R^2$ = 0.981. (*E*) Temperatures inferred with sample-specific LDI calibrations versus mean SST values from the warmest month at 0-m depth of each sampling location. (*F*) Residual SST (LDI-derived temperatures − measured mean SST values from the warmest month at 0-m depth of each sampling location) versus mean SST values from the warmest month at 0-m depth of each sampling location. Dashed lines reflect the SD of the residuals (±2.4 °C).

LCD producers in surface waters, as proposed in previous LDI calibration studies (34, 35), is also reflected by the predominance of OTU 56 from the Mariana Trench in samples at 4-m depth (40) and the highest LCD concentrations and numbers of eustigmatophyte reads in the upper water layers of the western tropical North Atlantic Ocean (27). Furthermore, the predominance of LDI-LCD producers at times with the shallowest mixed-layer depths matches the observation that in Lake Geneva, LCDs were most abundant during the period of thermal stratification of the water column (51). The low nutrient conditions prevailing during these months are consistent with the finding that eustigmatophytes were predominant in Qinghai Lake during periods with oligotrophic conditions (39).

The LDI–temperature correlation greatly improved when selecting SSTs from the warmest months instead of annual mean SSTs (Figs. 4*A* and 5 *A* and *B*), but the significant scatter that remained suggests there are additional factors besides temperature affecting the LDI. It has been suggested that contributions of 1,13-LCDs from specific *Proboscia* diatom species may affect the LDI (52), and the contribution of these species can be quantified by the fractional abundance of $C_{28}$ 1,12-diol ($FC_{28}$ 1,12-diol) (34):

$$FC_{28}\,1,12\text{−diol} = [C_{28}\,1,12\text{−diol}]/[C_{28}\,1,12- + C_{28}\,1,13- + C_{28}\,1,14- + C_{30}\,1,13- + C_{30}\,1,14- + C_{30}\,1,15\text{−diols}]$$

So far, few studies have reported $FC_{28}$ 1,12-diol values. For the 525 samples reported in the study by de Bar et al. (34), high $FC_{28}$ 1,12-diol values were predominantly identified in areas with mean SSTs from the warmest months ranging between 9 and 14 °C (e.g., Fig. 5 *A* and *B*), and their abundances showed no clear trend with temperature, salinity, or nutrients. The correlation of the LDI with mean SSTs from the warmest months increased from $R^2$ 0.88 to 0.90 by excluding 18 samples with $FC_{28}$ 1,12-diol values >0.15, while excluding additional

samples with lower $FC_{28}$ 1,12-diol values had little effect (*SI Appendix*, Fig. S12*A*).

It has also been indicated that freshwater input or low salinity may affect the LDI (3, 32, 33, 35), and it was shown that the LDI–temperature correlation can be improved by excluding samples from areas with annual mean salinities <32 ppt (34). For the combined LDI dataset, excluding samples with annual mean or warmest month salinities <32 ppt resulted in an $R^2$ increase from 0.91 to 0.92 (*SI Appendix*, Fig. S13). This minor improvement of the correlation may not justify the considerable number of excluded samples (79 and 119, respectively, from a dataset of 688 samples in total). Moreover, exclusion of samples deposited under low salinity conditions is often not feasible for paleoclimate studies where salinity data are typically unavailable. When dividing the dataset into sample sets with annual mean salinities above and below 32 and correlating LDI values of each group individually to annual mean SSTs, it is remarkable that the slopes of the two correlations are almost identical, whereas the intercept for the lower salinity samples is 0.17 higher (*SI Appendix*, Fig. S13). Inferring LDI-derived SSTs using the >32 salinity correlation, thus, results in temperatures that are more than 5 °C higher than when using the <32 salinity correlation. This can be interpreted as an indication that also in low-salinity areas, the LDI is affected by temperature but reflecting colder-than-average periods of the year. When using the SSTs from the warmest months instead of annual mean SSTs, the offset between the two LDI calibrations is no longer constant, and the sample sets are more comparable to each other, which results in the smaller effect on the correlation when excluding low-salinity samples (*SI Appendix*, Fig. S13).

It has been noted that LCD distributions from lacustrine environments and near rivers are often characterized by relatively high $C_{32}$ 1,15-diol abundances (2, 3, 30, 32, 33, 53, 54). Whereas $FC_{32}$ 1,15-diol values were <0.03 in our Mediterranean Sea samples, values >0.1 have been reported in central North and South Atlantic Ocean sediments from areas that are unlikely to be affected by freshwater input (34). This suggests that at least some marine LDI-LCD producers also biosynthesize $C_{32}$ 1,15-diol. In a previous study, it was reported that the removal of samples with high $C_{32}$ 1,15-diol abundances had little effect on the LDI–temperature calibration (34). However, samples with high $FC_{32}$ 1,15-diol values seem to have a different correlation with SST compared to samples with lower $FC_{32}$ 1,15-diol values (Fig. 5 *A* and *B*). We have arranged the data of the combined LCD dataset in the order of their $FC_{32}$ 1,15-diol values and calculated moving slopes and intercepts for sets of 21 samples along the list. Data from sample sets with temperature ranges <15 °C were excluded, and the remaining values were plotted against their average $FC_{32}$ 1,15-diol values (Fig. 5 *C* and *D*). For samples with $FC_{32}$ 1,15-diol values <0.3, the calculated slope values showed a weak negative trend and the corresponding intercept values a weak positive trend with increasing $FC_{32}$ 1,15-diol, but those trends were of no proportion to the nonlinear variation observed for those samples. For samples with $FC_{32}$ 1,15-diol values >0.3, slope values showed a distinct decrease with higher $FC_{32}$ 1,15-diol, whereas the intercept showed a distinct increase.

Based on these observations, we propose to use the average slope of 0.033 and the average intercept of −0.008 when reconstructing SSTs for samples with $FC_{32}$ 1,15-diol values <0.3. For samples with $FC_{32}$ 1,15-diol values >0.3, the slope and intercept for each sample can be calculated individually, using the following equations:

$$\text{Slope} = 0.072 * \text{FC}_{32}\ 1,15\text{−diol}^2 - 0.123 * \text{FC}_{32}\ 1,15\text{−diol} + 0.061$$

$$\text{Intercept} = -3.561 * \text{FC}_{32}\ 1,15\text{−diol}^2 + 4.128 * \text{FC}_{32}\ 1,15\text{−diol} - 0.943$$

Using these values, SSTs from the warmest months can then be estimated by

$$\text{SST} = (\text{LDI–Intercept})/\text{Slope}.$$

For the resulting dataset, estimated versus measured SSTs for the warmest months have an $R^2$ of 0.92 (Fig. 5*E*). For samples with $FC_{32}$ 1,15 diol <0.3, the SD for estimated − measured temperatures is 2.4 °C, and for samples with $FC_{32}$ 1,15-diol >0.3, this SD is 4.6 °C. Sample MT20 from the Gulf of Lion, showing an anomalously high LDI-inferred temperature of 34.3 °C, stood out from the rest of the dataset (estimated − measured temperature >3 × SD). In the original paper, this sample was left out when displaying the spatial distribution of $C_{32}$ 1,15-diol (33), possibly for showing an anomalous $FC_{32}$ 1,15-diol value. By excluding this sample from our dataset, the SD for estimated − measured temperatures is reduced to 3.3 °C. The LDI temperature calibration for samples with $FC_{32}$ 1,15 diol values >0.3 is currently based on only 33 samples, with a relatively large contribution of Baltic Sea samples (8) and Kara Sea samples (7); therefore, one should remain critical when including the results from samples with high $FC_{32}$ 1,15-diol values.

We also used Bayesian statistics following MATLAB code that has been used for previous proxy-calibration studies (55) to calibrate the relationship between the LDI and annual mean SSTs, mean summer SSTs, or mean SSTs from the warmest months for samples with $FC_{32}$ 1,15-diol values <0.3. The obtained slopes, intercepts, and root-mean-squared errors for the obtained Bayesian calibration models were similar to those obtained with the linear regression models. For samples with $FC_{32}$ 1,15-diol values <0.3, LDI temperatures that had previously been inferred with the original calibration (35) can simply be adjusted to warmest month temperatures by adding 2.64 °C to the initial value.

As evident from their identical slope values, LDI calibrations to annual mean SST and mean SST from the warmest month both have a 30.3 °C temperature range, but the warmest month temperatures are more evenly distributed along the calibration line (Fig. 5 *A* and *B*). Perhaps the most obvious improvement is the plateau with LDI values 0.95 and corresponding annual mean SST values ranging from 15.5 to 30.3 °C, whereas for the new calibration, >0.95 values were observed in areas where the maximum month temperatures ranged between 23.3 and 30.3 °C. An alternative calibration has been proposed for South China Sea samples with annual mean SST values >27 °C, where a negative correlation was observed between the LDI and annual mean SST (49). When correlating the LDI data to summer temperatures, a weak positive correlation over a very narrow range of measured temperatures had been observed. For our new calibration, the residual errors of those samples range from −2.5 to 1.8 °C with no obvious trend, thereby disputing the added value of the alternative calibration.

Temperatures calculated from the LDI values of our Mediterranean Sea water samples exceeded the measured SST (*SI Appendix*, Fig. S14*A*). Although the highest LDI values were indeed obtained at the warmest sampling day, the corresponding temperature variations were in no proportion to the measured data. Possibly, organisms living deeper in the water column have different LDI-LCD distributions, and the LDI

preserved in sediments may be the result of mixed signals, but LDI-derived temperatures from surface sediment samples from the Gulf of Lion—near our sampling locations—also generally overestimated measured temperatures (*SI Appendix*, Fig. S14*B*). Some sediment trap studies also reported minimal seasonal LDI variation and LDI-derived temperatures that did not reflect the temperatures of the overlying waters, and similar observations are known for other organic temperature proxies (56). For LDI values obtained in Mozambique Channel sediment trap samples, it was hypothesized that resuspended material affected the sediment trap signal (56), but for our current study, the similar patterns observed for LDI-LCDs and eustigmatophyte sequence reads suggest that LDI-LCDs in our Mediterranean Sea samples were predominantly derived from fresh biomass.

For other well-established paleotemperature proxies, it often remains uncertain how exactly the sediment distributions originate. Sediment trap studies on the $U^{K}_{37}$, an index based on the proportion of di-unsaturated to triunsaturated $C_{37}$ alkenones (57, 58), show that there is no consistent seasonal pattern in alkenone fluxes (59), and yet, the global $U^{K}_{37}$ values strongly correlate with mean annual SSTs (60). Moreover, the global $U^{K}_{37}$ calibration is statistically similar to the first published calibration derived from culturing experiments with the haptophyte alga *Emiliania huxleyi* (61). It has been observed that lipid distributions in different alkenone-producing species are differently affected by temperature, and correlations of the $U^{K'}_{37}$ for freshwater species do not correspond to the global marine $U^{K'}_{37}$ calibration (62–64). Eustigmatophyte culturing experiments have also demonstrated that these algae do adjust their LCD composition to the growth temperature (3). However, different eustigmatophyte families show different responses to temperature, and so far, none of the obtained LDI–temperature correlations corresponded to the marine LDI calibration (3). This demonstrates the relevance of isolating species from the eustigmatophyte clade for culturing experiments to further test the LDI calibration.

## Conclusions

The results of this study change our understanding of the evolution of eustigmatophyte algae. Our data reveal the existence of a marine eustigmatophyte group which has evolved before the currently known eustigmatophytes diversified. These marine eustigmatophytes have remained unnoticed, even though sequences have been detected in a wide variety of marine environments. The omnipresence of 1,13- and 1,15-LCDs in marine sediments can now be attributed to this eustigmatophyte group, and this indicates that these algae comprise a ubiquitous and considerable fraction of marine phytoplankton on a global scale. Our findings provide valuable background information for understanding and applying the LDI, a palaeotemperature proxy based on 1,13- and 1,15-LCDs. In some areas, contribution from additional sources may affect the LDI, but this can be recognized by enhanced $FC_{28}$ 1,12-diol or $FC_{32}$ 1,15-diol values, and $FC_{32}$ 1,15-diol values may even be used to correct for additional input. With the LDI constrained as a warm-season water temperature proxy, previously observed differences between LDI-derived and other paleotemperature records can now be explained and may even be applied to infer the amplitude of seasonal variations. Future research may focus on the spatial and temporal diversity of marine eustigmatophytes. Isolation and culturing experiments may determine which factors affect the LDI distribution in these algae, and further studies may explore the utility of the LDI as a warm-season SST proxy.

## Materials and Methods

**Sampling.** Surface water samples (0-m depth) were taken at four locations near Toulon, at a maximum distance of 5.6 km from the coast (Table 1 and *SI Appendix*, Fig. S1). Med. 2 and 3 were located farthest from the coast into the Mediterranean Sea, and Med. 4 was located at a shallower location at the edge of the Gulf of Lion. Surface water was collected using a bucket and, after measuring the temperature, stored for transport in 20-L Nalgene polycarbonate carboys. SPM matter was filtered within 2 d using precombusted 0.7 μm GF/F filters (Whatman). Filters used for lipid analyses were stored at −20 °C, and samples for DNA analyses were stored at −80 °C.

**Organic Carbon Analyses.** For the determination of POC, aliquots of freeze-dried filters were analyzed with a LECO RC 612 multiphase carbon analyzer. Filters were placed in silver cups and heated from 120 to 1,000 °C in a stream of $O_2$ (750 mL/min). The time- and temperature-resolved $CO_2$ release was detected with an infrared absorption cell.

**Lipid Extraction.** Filters containing SPM from 8.7 to 29.8 L water were freeze-dried and ultrasonically extracted with 3× methanol (MeOH), 3 mL, 3× MeOH: dichloromethane (DCM) 1:1, 3 mL, and 3× DCM, 3 mL. The extracts were combined and dried under $N_2$. Milli-Q water was added to remove the salts, lipids were extracted with DCM, and samples were split into two equal aliquots and dried under $N_2$. Milli-Q water and one drop of 6% KOH in MeOH were added to one aliquot of each sample, and neutral lipids were extracted with *n*-hexane (3×). Thereafter, the water was acidified to pH 1 with 10% HCl, and polar lipids were extracted with *n*-hexane (5×). We added 3 mL 6% KOH in MeOH in a 7-mL vial to the remaining half of the sample extracts and to the extracted filters for base hydrolysis. The vials were closed and heated for 3 h at 80 °C in a heating block. After cooling down, neutral lipid fractions were extracted with *n*-hexane (3×), the remaining samples were acidified to pH 1 with 10% HCl, and polar lipids were extracted with *n*-hexane (5×). A second set of filters containing SPM collected 27 August 2019 and 17 September 2019 were extracted as described above, after which 3 mL 2n HCl in MeOH were added to half of the dried extracts and to the extracted filters for acid hydrolysis. Sample vials were closed and heated for 3 h at 80 °C in a heating block, and after cooling down, lipids were extracted with *n*-hexane (5×). $C_{22}$ 5,17-diol was added as an internal standard (52) to each of the obtained lipid fractions, and the solvent was evaporated under a stream of $N_2$. Fatty acid groups in the polar fractions were methylated according to Poerschmann et al. (65); 1 mL trimethylchlorosilane (TMCS) MeOH (1:9) was added to the dried extracts, and the sample vials were closed and heated for 90 min at 80 °C in a heating block. After cooling down, lipids were extracted with *n*-hexane (3×) and dried under $N_2$. Prior to gas chromatography–mass spectrometry (GC-MS) analyses, alcohol groups of all samples were silylated by adding 15 μl BSTFA [N,O-bis(trimethylsilyl)trifluoroacetamide] containing 5% TMCS and 15 μl pyridine and heating for 30 min at 60 °C. After cooling down, 100 μl *n*-hexane were added to all samples.

**GC-MS Analyses.** GC-MS analyses were performed using a Thermo Fisher Trace 1310 GC coupled to a Thermo Fisher Quantum XLS Ultra MS. The GC was equipped with a capillary column (Phenomenex Zebron ZB-5, 30-m length, 0.25-μm film thickness, inner diameter 0.25 mm), using helium as a carrier gas (flow rate of 1.5 mL/min). Derivatized extracts were injected into a splitless injector and transferred to the GC column at 270 °C. The GC oven temperature was kept at 80 °C for 1 min and then increased to 310 °C at 5 °C/min, at which it was held for 20 min. Electron ionization MS were recorded in full scan mode at 70 eV, a mass range of *m/z* 50 to 600 and a scan time of 0.42 s, or in selected ion monitoring mode of the characteristic fragmentation ions of the environmental LCDs and the LCD standard (i.e., *m/z* 187, 299, 313, 327, and 341; refs. 35 and 52) for the quantification of the LCDs. Correction factors of 0.124, 0.068, and 0.230, obtained from full scan MS, were applied for calculating the concentrations of the saturated and unsaturated environmental LCDs and the LCD standard, respectively. As small amounts of the neutral lipids were identified in the polar fractions–in particular, after base hydrolysis–lipids were quantified in both fractions, and the results of this study consist of the combined data.

With the methods used for the current study, $C_{29}$ 12-hydroxy methyl alkanoate coeluted with $C_{28}$ diols, complicating the identification of $C_{28}$ 1,12-diol (see *SI Appendix, Supplementary Text*).

**DNA Extraction and Sequencing.** On ice, 1/4 sections of the GF/F filters from location Med. 3, containing SPM from ca. 10 L seawater, were cut into small pieces using sterile scalpels and tweezers. Filter pieces were transferred into 2 mL microtubes, and DNA was extracted using a DNEasy Powerlyzer PowerSoil Kit (Qiagen) following the manufacturer's instructions. The universal primers NGS-p23SrV_f1 (5′-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GGA CAG AAA GAC CCT ATG AA-3′) and NGS-p23SrV_r1 (5′-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTC AGC CTG TTA TCC CTA GAG-3′) (66) were used to amplify the plastid 23S rRNA as described by Sherwood and Presting (67). Sequencing was conducted at the G2L: Next-Generation Sequencing at the Göttingen Genomics Laboratory, Department of Microbiology, University of Göttingen (Germany). All the sequence data generated in this study have been deposited in NCBI BioSample database under the accession numbers SAMN24256204, SAMN24256205, SAMN24256206, SAMN24256207, and SAMN24256208. Eustigmatophyte sequences ASV_005, ASV_026, ASV_031, and ASV_157 have also been deposited individually under the accession numbers OL958649, OL958650, OL958651, and OL958652, respectively.

**Bioinformatic Analyses.** We obtained 1,531,336 raw sequences, which were further processed using appropriate packages in the R software environment (68). Sequences containing an "N" and sequences without primer detection were discarded, using the DADA2 (69), Biostrings (70), and ShortReads (71) packages. Thereafter, primers were removed from the reads using the Cutadapt software (72). DADA2 quality filtering was performed using a maximum expected error threshold of 2 for the forward reads and 4 for the reverse reads, and the dereplication of identical reads, sample inference, merging of paired-end reads, and chimera removal were executed using default values (69). The taxonomic assignment of the ASV sequences was performed using the nucleotide–nucleotide basic local alignment search tool (BLASTN) software version 2.10.1+ (73) with the BLASTN algorithm and the June 2021 NCBI GenBank nucleotide database version (74), using the Tidyverse package (75).

In order to verify sequence identifications, phylogenetic analyses were performed by Bayesian Inference algorithms, using the program MrBayes 3.2.7-WIN (76). The generalized time-reversible (GTR) model was used, with gamma-distribution rate variation across sites and a proportion of invariable sites, whereby five hot chains were run in addition to the standard (cold) chain. The Bayesian search was run for 2,500,000 generations, saving every thousandth tree, and the first 700 trees were discarded. In addition, phylogenetic analyses were performed by Maximum Likelihood algorithms, using the program MEGA-X (77) with standard settings and 1,000 bootstrap replications.

**Data Availability.** DNA sequence data have been deposited in the National Center for Biotechnology Information Sequence Read Archive database (accession nos. SAMN24256204–SAMN24256208 and OL958649–OL958652). All other study data are included in the article and/or *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank S. Zeman-Kuhnert and F. P. Kerschhofer for their assistance during the sampling trips; J. Waniek, B. Olberg, and D. Nimptsch for their assistance with the DNA and bioinformatic analyses; C. Conradt for assistance in the organic geochemistry laboratory; and B. Röring for assisting with the organic carbon analyses. The editor and three reviewers are thanked for their constructive comments. This research has been funded by the German Research Foundation (Project RA 2970/1-1).

Author affiliations: [a]Geobiology, Geoscience Centre, University of Göttingen, 37077 Göttingen, Germany; and [b]Experimental Phycology and Culture Collection of Algae (SAG), University of Göttingen, Göttingen, 37073 Germany

1. J. W. de Leeuw, W. I. C. Rijpstra, P. A. Schenck, The occurrence and identification of $C_{30}$, $C_{31}$ and $C_{32}$ alkan-1,15-diols and alkan-15-one-1-ols in Unit I and Unit II Black Sea sediments. *Geochim. Cosmochim. Acta* **45**, 2281–2285 (1981).
2. G. J. M. Versteegh, J. H. F. Jansen, J. W. de Leeuw, R. R. Schneider, Mid-chain diols and keto-ols in SE Atlantic sediments: A new tool for tracing past sea surface water masses? *Geochim. Cosmochim. Acta* **64**, 1879–1892 (2000).
3. S. W. Rampen *et al.*, Sources and proxy potential of long chain alkyl diols in lacustrine environments. *Geochim. Cosmochim. Acta* **144**, 59–71 (2014).
4. S. W. Rampen *et al.*, Evaluation of long chain 1,14-alkyl diols in marine sediments as indicators for upwelling and temperature. *Org. Geochem.* **76**, 39–47 (2014).
5. J. K. Volkman, S. M. Barrett, S. I. Blackburn, Eustigmatophyte microalgae are potential sources of $C_{29}$ sterols, $C_{22}$–$C_{28}$ *n*-alcohols and $C_{28}$–$C_{32}$ *n*-alkyl diols in freshwater environments. *Org. Geochem.* **30**, 307–318 (1999).
6. J. K. Volkman, S. M. Barrett, G. A. Dunstan, S. W. Jeffrey, $C_{30}$–$C_{32}$ alkyl diols and unsaturated alcohols in microalgae of the class Eustigmatophyceae. *Org. Geochem.* **18**, 131–138 (1992).
7. J. S. Sinninghe Damsté *et al.*, A diatomaceous origin for long-chain diols and mid-chain hydroxy methyl alkanoates widely occurring in Quaternary marine sediments: Indicators for high nutrient conditions. *Geochim. Cosmochim. Acta* **67**, 1339–1348 (2003).
8. S. W. Rampen, S. Schouten, J. S. Sinninghe Damsté, Occurrence of long chain 1,14 diols in *Apedinella radians*. *Org. Geochem.* **42**, 572–574 (2011).
9. E. Koning *et al.*, Selective preservation of upwelling-indicating diatoms in sediments off Somalia, NW Indian Ocean. *Deep Sea Res. 1 Oceanogr. Res. Pap.* **48**, 2473–2495 (2001).
10. S. G. Wakeham, M. L. Peterson, J. I. Hedges, C. Lee, Lipid biomarker fluxes in the Arabian Sea, with a comparison to the equatorial Pacific Ocean. *Deep Sea Res. 2 Top. Stud. Oceanogr.* **49**, 2265–2301 (2002).
11. S. Contreras *et al.*, A rainy northern Atacama Desert during the last interglacial. *Geophys. Res. Lett.* **37**, L23612 (2010).
12. L. Herrera, R. Escribano, Factors structuring the phytoplankton community in the upwelling site off El Loa River in northern Chile. *J. Mar. Syst.* **61**, 13–38 (2006).
13. J. Tarazona, D. Gutiérrez, C. Paredes, A. Indacochea, Overview and challenges of marine biodiversity research in Peru. *Gayana (Concepc.)* **67**, 206–231 (2003).
14. M. Brichta, E.-M. Nöthig, *Proboscia inermis*: A key diatom species in Antarctic autumn. https://epic.awi.de/id/eprint/9570/. Accessed 20 November 2020.
15. V. Willmott *et al.*, Holocene changes in *Proboscia* diatom productivity in shelf waters of the northwestern Antarctic Peninsula. *Antarct. Sci.* **22**, 3–10 (2010).
16. S. W. Rampen, S. Schouten, E. Koning, G.-J. A. Brummer, J. S. Sinninghe Damsté, A 90 kyr upwelling record from the northwestern Indian Ocean using a novel long-chain diol index. *Earth Planet. Sci. Lett.* **276**, 207–213 (2008).
17. D. J. Hibberd, G. F. Leedale, Eustigmatophyceae–A new algal class with unique organization of the motile cell. *Nature* **225**, 758–760 (1970).
18. D. J. Hibberd, Notes on the taxonomy and nomenclature of the algal classes Eustigmatophyceae and Tribophyceae (synonym Xanthophyceae). *Bot. J. Linn. Soc.* **82**, 93–119 (1981).
19. L. Villanueva *et al.*, Potential biological sources of long chain alkyl diols in a lacustrine system. *Org. Geochem.* **68**, 27–30 (2014).
20. M. Eliás *et al.*, "Eustigmatophyceae" in *Handbook of the Protists*, J. M. Archibald, A. G. B. Simpson, C. H. Slamovits, Eds. (Springer, 2017), pp. 367–406.
21. R. Amaral *et al.*, Toward modern classification of eustigmatophytes, including the description of Neomonodaceae fam. nov. and three new genera. *J. Phycol.* **56**, 630–648 (2020).
22. M. W. Fawley, K. P. Fawley, A. B. Cahoon, Finding needles in a haystack–extensive diversity in the Eustigmatophyceae revealed by community metabarcode analysis targeting the rbcL gene using lineage-directed primers. *J. Phycol.* **57**, 1636–1647 (2021).
23. T. Ševcíková *et al.*, Plastid genomes and proteins illuminate the evolution of eustigmatophyte algae and their bacterial endosymbionts. *Genome Biol. Evol.* **11**, 362–379 (2019).
24. K. P. Fawley, M. W. Fawley, Observations on the diversity and ecology of freshwater *Nannochloropsis* (Eustigmatophyceae), with descriptions of new taxa. *Protist* **158**, 325–336 (2007).
25. M. W. Fawley, I. Jameson, K. P. Fawley, The phylogeny of the genus *Nannochloropsis* (Monodopsidaceae, Eustigmatophyceae), with descriptions of *N. australis* sp. nov. and *Microchloropsis* gen. nov. *Phycologia* **54**, 545–552 (2015).
26. M. W. De Bar, S. W. Rampen, E. C. Hopmans, J. S. Sinninghe Damsté, S. Schouten, Constraining the applicability of organic paleotemperature proxies for the last 90 Myrs. *Org. Geochem.* **128**, 122–136 (2019).
27. S. Balzano *et al.*, A quest for the biological sources of long chain alkyl diols in the western tropical North Atlantic Ocean. *Biogeosciences* **15**, 5951–5968 (2018).
28. R. J. Morris, S. C. Brassell, Long-chain alkanediols: Biological markers for cyanobacterial contributions to sediments. *Lipids* **23**, 256–258 (1988).
29. J. W. de Leeuw, W. I. C. Rijpstra, L. R. Mur, The absence of long-chain alkyl diols and alkyl keto-1-ols in cultures of the cyanobacterium *Aphanizomenon flos-aquae*. *Org. Geochem.* **18**, 575–578 (1992).
30. G. J. M. Versteegh, H. J. Bosch, J. W. de Leeuw, Potential palaeoenvironmental information of $C_{24}$ to $C_{36}$ mid-chain diols, keto-ols and mid-chain hydroxy fatty acids; a critical review. *Org. Geochem.* **27**, 1–13 (1997).
31. J. K. Gal, J. H. Kim, K. H. Shin, Distribution of long chain alkyl diols along a south-north transect of the northwestern Pacific region: Insights into a paleo sea surface nutrient proxy. *Org. Geochem.* **119**, 80–90 (2018).
32. M. W. De Bar *et al.*, Constraints on the application of long chain diol proxies in the Iberian Atlantic margin. *Org. Geochem.* **101**, 184–195 (2016).
33. J. Lattaud *et al.*, The $C_{32}$ alkane-1,15-diol as a tracer for riverine input in coastal seas. *Geochim. Cosmochim. Acta* **202**, 146–158 (2017).
34. M. W. de Bar *et al.*, Global temperature calibration of the Long chain Diol Index in marine surface sediments. *Org. Geochem.* **142**, 103983 (2020).
35. S. W. Rampen *et al.*, Long chain 1,13- and 1,15-diols as a potential proxy for palaeotemperature reconstruction. *Geochim. Cosmochim. Acta* **84**, 204–216 (2012).
36. A. M. Ferreira *et al.*, Formation of mid-chain alkane keto-ols by post-depositional oxidation of mid-chain diols in Mediterranean sapropels. *Org. Geochem.* **32**, 271–276 (2001).
37. S. Reiche, S. W. Rampen, D. J. C. Dorhout, J. S. Sinninghe Damsté, S. Schouten, The impact of oxygen exposure on long-chain alkyl diols and the long chain diol index (LDI) – A long-term incubation study. *Org. Geochem.* **124**, 238–246 (2018).

38. G. J. M. Versteegh, J. Lipp, Detection of new long-chain mid-chain keto-ol isomers from marine sediments by means of HPLC APCI-MS and comparison with long-chain mid-chain diols from the same samples. *Org. Geochem.* **133**, 92–102 (2019).
39. G. Li *et al.*, Temporal succession of ancient phytoplankton community in Qinghai Lake and implication for paleo-environmental change. *Sci. Rep.* **6**, 19769 (2016).
40. R. Guo *et al.*, Insight Into the pico- and nano-phytoplankton communities in the deepest biosphere, the Mariana Trench. *Front. Microbiol.* **9**, 2289 (2018).
41. C. Harrison, *Millennial Indian Monsoon Variability and Forcing During the Holocene as Revealed by the Qinghai Lake* (Yunnan Province Record, 2019).
42. H. Guan *et al.*, Composition and origin of lipid biomarkers in the surface sediments from the southern Challenger Deep, Mariana Trench. *Geosci. Front.* **10**, 351–360 (2019).
43. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
44. A. Penna, S. Casabianca, A. F. Guerra, C. Vernesi, M. Scardi, Analysis of phytoplankton assemblage structure in the Mediterranean Sea based on high-throughput sequencing of partial 18S rRNA sequences. *Mar. Genomics* **36**, 49–55 (2017).
45. S. Balzano, L. Villanueva, M. de Bar, J. S. Sinninghe Damsté, S. Schouten, Impact of culturing conditions on the abundance and composition of long chain alkyl diols in species of the genus *Nannochloropsis*. *Org. Geochem.* **108**, 9–17 (2017).
46. R. A. Lopes dos Santos *et al.*, Comparison of organic ($U^{K'}_{37}$, $TEX^{H}_{86}$, LDI) and faunal proxies (foraminiferal assemblages) for reconstruction of late Quaternary sea surface temperature variability from offshore southeastern Australia. *Paleoceanography* **28**, 377–387 (2013).
47. L. He, M. Kang, D. Zhang, G. Jia, Evaluation of environmental proxies based on long chain alkyl diols in the East China Sea. *Org. Geochem.* **139**, 103948 (2020).
48. B. Wei *et al.*, Comparison of the $U^{K'}_{37}$, LDI, $TEX^{H}_{86}$, and RI-OH temperature proxies in sediments from the northern shelf of the South China Sea. *Biogeosciences* **17**, 4489–4508 (2020).
49. Y. Yang *et al.*, Assessing the applicability of the long-chain diol (LDI) temperature proxy in the high-temperature South China Sea. *Org. Geochem.* **144**, 104017 (2020).
50. H. E. Garcia *et al.*, World Ocean Atlas 2018: Product documentation. https://www.ncei.noaa.gov/archive/accession/NCEI-WOA18. Accessed 18 December 2020.
51. J. Lattaud *et al.*, Sources and seasonality of long-chain diols in a temperate lake (Lake Geneva). *Org. Geochem.* **156**, 104223 (2021).
52. M. Rodrigo-Gámiz *et al.*, Constraints on the applicability of the organic temperature proxies $U^{K'}_{37}$, $TEX_{86}$ and LDI in the subpolar region around Iceland. *Biogeosciences* **12**, 6573–6590 (2015).
53. J. Lattaud *et al.*, The $C_{32}$ alkane-1,15-diol as a proxy of late Quaternary riverine input in coastal margins. *Clim. Past* **13**, 1049–1061 (2017).
54. M. Shimokawara, M. Nishimura, T. Matsuda, N. Akiyama, T. Kawai, Bound forms, compositional features, major sources and diagenesis of long chain, alkyl mid-chain diols in Lake Baikal sediments over the past 28,000 years. *Org. Geochem.* **41**, 753–766 (2010).
55. E. Dearing Crampton-Flood, J. E. Tierney, F. Peterse, F. M. S. A. Kirkels, J. S. Sinninghe Damsté, BayMBT: A Bayesian calibration model for branched glycerol dialkyl glycerol tetraethers in soils and peats. *Geochim. Cosmochim. Acta* **268**, 142–159 (2020).
56. M. W. De Bar *et al.*, Long-chain diols in settling particles in tropical oceans: Insights into sources, seasonality and proxies. *Biogeosciences* **16**, 1705–1727 (2019).
57. S. C. Brassell, G. Eglinton, I. T. Marlowe, U. Pflaumann, M. Sarnthein, Molecular stratigraphy: A new tool for climatic assessment. *Nature* **320**, 129–133 (1986).
58. F. G. Prahl, S. G. Wakeham, Calibration of unsaturation patterns in long-chain ketone compositions for palaeotemperature assessment. *Nature* **330**, 367–369 (1987).
59. A. Rosell-Melé, F. G. Prahl, Seasonality of $U^{K'}_{37}$ temperature estimates as inferred from sediment trap data. *Quat. Sci. Rev.* **72**, 128–136 (2013).
60. P. Müller, G. Kirst, G. Ruhland, I. von Storch, A. Rosell-Melé, Calibration of the alkenone palaeotemperature index $U^{K'}_{37}$ based on core-tops from the eastern South Atlantic and the global ocean (60°N-60°S). *Geochim. Cosmochim. Acta* **62**, 1757–1772 (1998).
61. F. G. Prahl, L. A. Muehlhausen, D. B. Zahnle, Further evaluation of long-chain alkenones as indicators of paleoceanographic conditions. *Geochim. Cosmochim. Acta* **52**, 2303–2310 (1988).
62. I. T. Marlowe *et al.*, Long-chain ($n$-$C_{37}$-$C_{39}$) alkenones in the Prymnesiophyceae. Distribution of alkenones and other lipids and their taxonomic significance. *Br. Phycol. J.* **19**, 203–216 (1984).
63. M. H. Conte, A. Thompson, D. Lesley, R. P. Harris, Genetic and physiological influences on the alkenone/alkenoate versus growth temperature relationship in *Emiliania huxleyi* and *Gephyrocapsa oceanica*. *Geochim. Cosmochim. Acta* **62**, 51–68 (1998).
64. G. J. M. Versteegh, R. Riegman, J. W. de Leeuw, J. H. F. Jansen, $U^{K'}_{37}$ values for *Isochrysis galbana* as a function of culture temperature, light intensity and nutrient concentrations. *Org. Geochem.* **32**, 785–794 (2001).
65. J. Poerschmann, R. Carlson, New fractionation scheme for lipid classes based on "in-cell fractionation" using sequential pressurized liquid extraction. *J. Chromatogr. A* **1127**, 18–25 (2006).
66. G. G. Presting, Identification of conserved regions in the plastid genome: Implications for DNA barcoding and biological function. *Can. J. Bot.* **84**, 1434–1443 (2006).
67. A. R. Sherwood, G. G. Presting, Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. *J. Phycol.* **43**, 605–608 (2007).
68. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
69. B. J. Callahan *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
70. H. Pagès, P. Aboyoun, R. Gentleman, S. DebRoy, Biostrings: Efficient manipulation of biological strings. https://rdrr.io/bioc/Biostrings/. Accessed 7 December 2020.
71. M. Morgan *et al.*, ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25**, 2607–2608 (2009).
72. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10 (2011).
73. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
74. R. Agarwala *et al.*; NCBI Resource Coordinators, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
75. H. Wickham *et al.*, Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
76. F. Ronquist *et al.*, MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
77. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).