Planktonic Archaeal Ether Lipid Origins in Surface Waters of the North Pacific Subtropical Gyre

Fuyan Li[1], Andy Leu[1], Kirsten Poff[1], Laura T. Carlson[2], Anitra E. Ingalls[2] and Edward F. DeLong[1]*

[1] Daniel K. Inouye Center for Microbial Oceanography: Research and Education, University of Hawai'i at Mânoa, Honolulu, HI, United States, [2] School of Oceanography, University of Washington, Seattle, WA, United States

Thaumarchaeota and Thermoplasmatota are the most abundant planktonic archaea in the sea. Thaumarchaeota contain tetraether lipids as their major membrane lipids, but the lipid composition of uncultured planktonic Thermoplasmatota representatives remains unknown. To address this knowledge gap, we quantified archaeal cells and ether lipids in open ocean depth profiles (0–200 m) of the North Pacific Subtropical Gyre. Planktonic archaeal community structure and ether lipid composition in the water column partitioned into two separate clusters: one above the deep chlorophyll maximum, the other within and below it. In surface waters, Thermoplasmatota densities ranged from 2.11×10^6 to 6.02×10^6 cells/L, while Thaumarchaeota were undetectable. As previously reported for Thaumarchaeota, potential homologs of archaeal tetraether ring synthases were present in planktonic Thermoplasmatota metagenomes. Despite the absence of Thaumarchaeota in surface waters, measurable amounts of intact polar ether lipids were found there. Based on cell abundance estimates, these surface water archaeal ether lipids contributed only 1.21×10^{-9} ng lipid/Thermoplasmatota cell, about three orders of magnitude less than that reported for Thaumarchaeota cells. While these data indicate that even if some tetraether and diether lipids may be derived from Thermoplasmatota, they would only comprise a small fraction of Thermoplasmatota total biomass. Therefore, while both MGI Thaumarchaeota and MGII/III Thermoplasmatota are potential biological sources of archaeal GDGTs, the Thaumarchaeota appear to be the major contributors of archaeal tetraether lipids in planktonic marine habitats. These results extend and confirm previous reports of planktonic archaeal lipid sources, and further emphasize the need for Thermoplasmatota cultivation, to better characterize the membrane lipid constituents of marine planktonic Thermoplasmatota, and more precisely define the sources and patterns of archaeal tetraether lipid distributions in marine plankton.

Keywords: planktonic Thermoplasmatota, archaeal ether lipids, archaeal GDGT ring synthases, euphotic zone, NPSG

INTRODUCTION

Marine planktonic Archaea are abundant in diverse marine planktonic environments (DeLong, 1992, 1998; Fuhrman et al., 1992; DeLong et al., 1994), and play important roles in the biogeochemical cycles of carbon and nitrogen (Ingalls et al., 2006; Falkowski et al., 2008; Martin-Cuadrado et al., 2008; Iverson et al., 2012). Three major groups of planktonic archaea are distributed
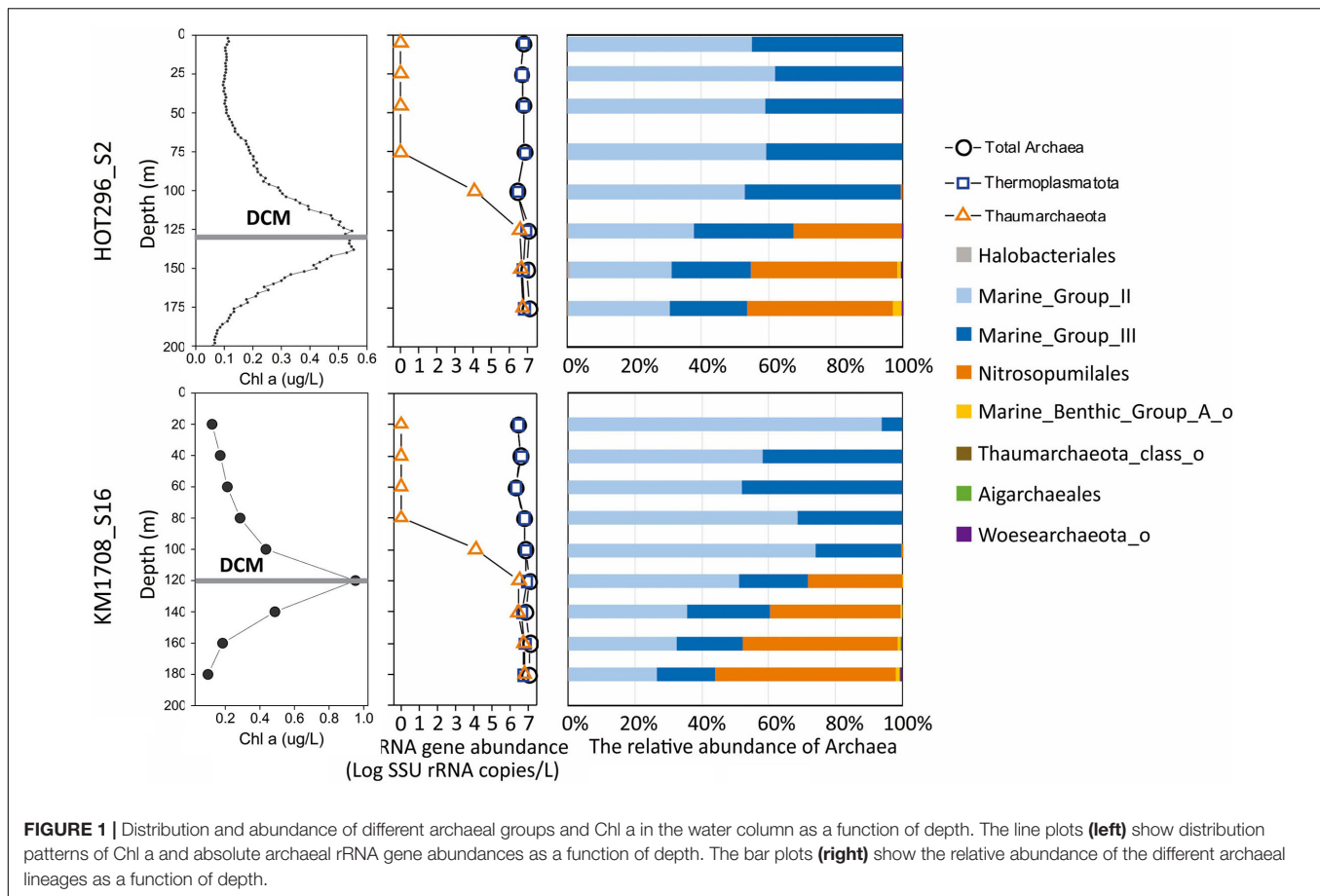
in the marine water column. The most abundant group was originally called Marine Group I (MGI) planktonic Crenarchaeota (DeLong, 1992). This general phylogenetic group was later referred to as Thaumarchaeota as genomic analyses and protein phylogenies became available (Preston et al., 1996; Hallam et al., 2006a,b; Brochier-Armanet et al., 2008). Thaumarchaeota are more abundant below the base of the epipelagic zone and in the meso- and bathypelagic (Karner et al., 2001; Mincer et al., 2007; Brown et al., 2009; Church et al., 2010). The two other most abundant planktonic archaeal groups belong to the Thermoplasmatota (formerly Thermoplasmatales in Euryarchaeota, Rinke et al., 2019), and include MGII Archaea (DeLong, 1992), which are more abundant in the surface and in the deep chlorophyll maximum (DCM) (Massana et al., 2000; DeLong et al., 2006; Mincer et al., 2007; Hugoni et al., 2013); and MGIII Archaea (Fuhrman and Davis, 1997), which can be detected throughout the water column but are found more frequently in deeper waters (Massana et al., 1997; DeLong et al., 2006; Haro-Moreno et al., 2017). To date, only the Thaumarchaeota have been cultivated, and all cultivars are chemolithotrophic ammonia oxidizers (Könneke et al., 2005; Kim et al., 2016; Qin et al., 2017). Some deeply branching lineages affiliated with Thaumarchaeota pSL12-like clade however (Mincer et al., 2007; Church et al., 2010), appear to be capable of heterotrophic growth (Aylward and Santoro, 2020; Reji and Francis, 2020). Symbiotic or cultured MGI Thaumarchaeota synthesize glycerol dialkyl glycerol tetraethers (GDGTs) with zero through four cyclopentane rings (DeLong et al., 1998; Schouten et al., 2008) and crenarchaeol, a distinct GDGT containing four cyclopentane and one cyclohexane ring (Schouten et al., 2008). Among isolated Archaea, crenarchaeol is unique to Thaumarchaeota and is also ubiquitous within this group. As such, crenarchaeol is used extensively as a biomarker for ammonia-oxidizing Archaea (AOA) (Schouten et al., 2008, 2013; Qin et al., 2015; Kim et al., 2016). The other GDGTs with cyclopentane rings are also synthesized by other archaea including members of the Crenarchaeota and Euryarchaeota (Pearson and Ingalls, 2013).

Glycerol dialkyl glycerol tetraethers produced by planktonic Thaumarchaeota within and below the DCM (primarily between 80 and 250 m) are incorporated into sinking particles, and accumulate in sediments (Hurley et al., 2018). Intact polar lipids (IP-GDGTs), core lipids (C-GDGTs) with one to two phosphatidic, glycosidic, or glycophosphatidic head groups attached, are assumed to rapidly degrade with cleavage of polar head groups after cell death (White et al., 1979; Harvey et al., 1986), and are used as quantitative indicators of living AOA in the marine water column (Schubotz et al., 2009; Pitcher et al., 2011; Ingalls et al., 2012; Basse et al., 2014). C-GDGTs are more recalcitrant to degradation and can be preserved in sediments over millions of years, which enables the application of these lipids in paleoenvironmental studies (Kuypers et al., 2001; Pearson et al., 2001). The sediment core-top TEX$_{86}$ proxy, based on the ratio between C-GDGTs with one to three cyclopentane moieties and the crenarchaeol isomer, is observed to correlate with sea surface temperature (SST) (Schouten et al., 2002; Kim et al., 2010). This proxy is used to reconstruct the paleotemperature of the sea surface (Zachos et al., 2006; Tierney et al., 2010; Schefuß et al., 2011). MGI Thaumarchaeota and their lipids occur preferentially in the subsurface and deeper, however, where temperatures are much lower than SST, especially from the tropical to subtropical regions ($\sim$30°N–30°S) (Ho and Laepple, 2016; Hurley et al., 2018; Zhang and Liu, 2018).

The mismatch between the habitat and depth profiles of known GDGT producing Thaumarchaeota and the use of GDGTs as a proxy of SST, has led to speculation that other planktonic archaeal groups, in particular the Thermoplasmatota, may also produce GDGTs in surface waters and contribute to sedimentary GDGTs (Lincoln et al., 2014; Ma et al., 2020). It was proposed that the prevalence of MGII Archaea at the ocean's surface might help explain why surficial sediment TEX$_{86}$ values reflect SST temperatures, even though most modern-day Thaumarchaeota do not reside in the sea surface and their lipids are only present in trace quantities. The TEX$_{86}$ signatures of Thaumarchaeota could also be influenced by oxygen availability (Qin et al., 2015) and ammonia oxidation rate (Hurley et al., 2016). Environmental metagenome assembly data show that genomes of MGII/III Archaea harbor genes coding for catalysis of ether bond formation (i.e., geranylgeranylglyceryl phosphate synthase, GGGP; digeranylgeranylglyceryl phosphate synthase, DGGGP), and that these groups of archaea may have the genetic potential to synthesize GDGTs (Iverson et al., 2012; Deschamps et al., 2014; Villanueva et al., 2017). Station ALOHA offers an opportunity to use a well-stratified water column to directly compare lipids and genes.

Current best estimates for the origin of the TEX$_{86}$ proxy indicate that archaeal lipids produced at or near the base of the euphotic zone are the major contributors to sedimentary GDGTs (Hurley et al., 2018). This also often corresponds with the base of the nitrite maximum. The euphotic layer of the North Pacific Subtropical Gyre (NPSG) is on average at $\sim$173 m ($\pm$7 m) (Letelier et al., 1996; Karl and Church, 2017). The upper primary nitrite maximum is located within the DCM or just below it, i.e., near the base of the euphotic layer (Dore and Karl, 1996). To further investigate the contribution of MGI Thaumarchaeota and MGII/III Thermoplasmatota to the production of ether lipids in the euphotic zone, we examined planktonic archaeal community profiles determined by amplicon sequence analyses, and the distribution of ether lipids (i.e., GDGTs and archeol) in high resolution depth profiles (between 0 and 200 m) in the NPSG (**Figure 1** and **Supplementary Table 1**). Amplicon sequencing, and droplet digital Polymerase Chain Reaction (ddPCR) analyses of small subunit ribosomal RNA (SSU rRNA) genes, were used to determine archaeal community profiles in the upper 200 m. Using the same filter samples, we measured in parallel ether lipids including C-, monohexose- (MH-GDGTs), dihexose- (DH-GDGTs), phosphohexose- (PH-GDGTs), and hexose-phosphohexose-GDGTs (HPH-GDGTs), and C-archeol (see structures in **Supplementary Figure 1**). We explore the relationship between GDGT occurrence and planktonic archaeal communities found at the same depths.

FIGURE 1 | Distribution and abundance of different archaeal groups and Chl a in the water column as a function of depth. The line plots **(left)** show distribution patterns of Chl a and absolute archaeal rRNA gene abundances as a function of depth. The bar plots **(right)** show the relative abundance of the different archaeal lineages as a function of depth.

MATERIALS AND METHODS

Sampling and Water Parameters

Seawater samples from the euphotic zone (between surface and 200 m) were collected at the Hawaii Ocean Time-Series Station ALOHA [Station 2 (S2)] during HOT cruise 296 (October 5–8, 2017), and at the center of a cyclonic eddy [Station 16 (S16)] during the MESO-SCOPE cruise KM1709 (June 26 to July 15, 2017) (**Supplementary Figure 2** and **Supplementary Table 1**). The filtration was subsequently conducted through 0.22 μm Durapore PVDF membrane filters with 25 mm diameter for DNA sample collection and 47 mm diameter for lipid sample collection (EMD Millipore, Billerica, MA, United States) (see detailed information in **Supplementary Table 1**). Samples of 2 L were collected for DNA analysis whereas samples of 20–30 L from the remaining water were collected for lipid analysis. After filtering was completed, DNA filter samples were immediately placed in RNAlater stabilization solution (Ambion, Carlsbad, CA, United States). All filters were frozen at −80°C on board, then transported to lab using dry ice and preserved at −80°C prior to extraction of DNA and lipids.

Environmental parameters such as temperature, salinity, density and Chl a were retrieved from CTD data available on the HOT-DOGS website http://hahana.soest.hawaii.edu/hot/hot-dogs/ for station HOT296_S2 and on the SCOPE website http://scope.soest.hawaii.edu/data/mesoscope/mesoscope.html for KM1709_S16.

DNA Extraction, Sequencing and Quantification

DNA was extracted using the PowerBiofilm DNA Isolation kit (MO BIO, Carlsbad, CA, United States) according to the manufactures protocol after RNAlater was removed. The V4 region of the SSU rRNA gene was amplified with universal primers 515F (5′-GTGYCAGCMGCCGCGGTAA-3′, Parada et al., 2016) and 806RB (5′-GGACTACNVGGGTWTCTAAT-3′, Apprill et al., 2015), modified from Caporaso et al. (2011). PCR amplification procedures were modified from the SSU rRNA illumina amplicon protocol of Earth Microbiome Project (2018). Cycling conditions are as follows: 94°C for 3 min; 28 cycles of 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s; followed by 72°C for 10 min. Amplicons were purified with AMPure XP beads (Beckman Coulter, Brea, CA, United States) and their size was verified using agarose gel electrophoresis. Following this step, all purified PCR products were pooled together and sequenced on the illumina MiSeq platform using the MiSeq Reagent Kits v2 300-cycles chemistry (Illumina, San Diego, CA, United States). All SSU rRNA amplicon sequencing data have deposited in the NCBI Sequence Read Archive under project number PRJNA614540.

The ddPCR assays were performed with the same barcoded primers. Briefly, the reaction mixtures contained 1 × EvaGreen ddPCR Supermix (Bio-Rad, Hercules, CA, United States), 100 nM of each primer and 0.03 ng of template DNA in 20 µL volume. Control reactions replaced the DNA template with nuclease-free water (Ambion, Austin, TX, United States). Each reaction was mixed with 70 µL of droplet generation oil (Bio-Rad) and partitioned into 14657–19526 droplets in a QX200 droplet generator (Bio-Rad). After processing, droplets were transferred into a 96-well ddPCR plate (Bio-Rad) and heated-sealed with pierceable foil (Bio-Rad). The PCRs were initiated at 95°C for 5 min DNA polymerase activation, followed by 40 cycles of a two-step thermal profile of 30 s at 95°C for denaturation and 2 min at 54.6°C for annealing and extension, and one cycle of 4°C for 5 min and 90°C for 5 min for droplet stabilization. After amplification, the fluorescence intensity of individual droplets was detected with QX200 droplet reader (Bio-Rad). Data were analyzed with QuantaSoft droplet reader software (Bio-Rad). Positive and negative signals were detected automatically or manually and the target DNA concentrations were calculated. Nothing was obtained in control reaction.

Processing of Sequence Data

Illumina pair-end sequences were processed with Trimmomatic v0.36 (Bolger et al., 2014), QIIME 1.9.1 (Caporaso et al., 2010), and DADA2 v1.6 (Callahan et al., 2016). Trimmomatic was used to remove adapters and trim low quality bases using the associated threshold parameters: the minimum leading quality = 3, the minimum trailing quality = 3, the number of bases within sliding window = 4, the average quality required for sliding window = 15, the minimum length = 100. The trimmed reads were demultiplexed and split into individual sample files with QIIME. After trimming and demultiplexing, amplicon sequences were further processed using the R package DADA2. We removed any sequences with length shorter than 110 or longer than 220, with ambiguous nucleotide assignment, or with greater than 2 expected errors. All identical sequences were combined into unique sequences through the step of dereplication. An error model was generated to evaluate the probability of each unique sequence being a real sequence versus an artifact sequence produced from PCR or sequencing. The error model was subsequently parameterized and applied for denoising forward and reverse reads independently. The denoised forward and reverse reads were merged together with overlap by at least 12 bases. Amplicon sequence variants (AASVs) were constructed based on exact sequence variants inferred from merged sequences, and then chimeras were removed. The taxonomy was assigned to individual AASVs using the SILVA v132 database. Only sequences restricted to Archaea were included for the following analysis and discussion in this study.

Phylogenetic Analysis

BLASTP searches of archaeal GDGT ring synthases GrsA (Saci_1585) and GrsB (Saci_0240), and Saci_1785 (as an outgroup) homologs from archaeal isolates in **Supplementary Table 1** of Zeng et al. (2019) was performed against the Genome Taxonomy Database using an e-value cutoff of $\leq 1e^{-5}$ and amino acid identity cutoff of $\geq 20\%$. Hits against reference MGII and MGIII genomes were collected. In addition, the ALOHA gene catalog (Mende et al., 2017) was searched for putative GrsA and GrsB homologs from MGII and MGIII with an amino acid identity cutoff of $\geq 90\%$.

The Grs homologs from MGII and MGIII metagenome assembled genomes (MAGs) in GTDB and the ALOHA gene catalog, and from the archaeal isolates were filtered by length (≥ 400 aa) and aligned using mafft v7.407 with the–auto function and the alignment trimmed using trimal v1.4[1] "-automated1" option. A phylogenetic tree was constructed using FastTree V. 2.1.10 with the LG + GAMMA models and rooted by midpoint. Bootstrap values were calculated via non-parametric bootstrapping with 100 replicates. The tree was visualized using the Interactive Tree of Life (iTOL) online tool[2] (Letunic and Bork, 2016).

The metagenomic sequence read coverage value for the putative Grs homologs of MGII/III from the ALOHA gene catalog 2.0 was estimated using coverM 0.3.1 with the -m trimmed_mean –min-read-percent-identity 0.95 –min-read-aligned-percent 0.75 option. The coverage profile was reported here with the cutoff of 1.

Lipid Extraction and Analysis

Lipids were ultrasonically extracted from filters using a modified Bligh-Dyer method with a solvent mixture of methanol, dichloromethane and aqueous buffer (2:1:0.8) (Sturt et al., 2004). Phosphate buffer (50 mM) was used for the first two extractions and trichloroacetic acid buffer (50 mM) was used for the remaining two extractions (Sturt et al., 2004). The total lipid extracts (TLEs) were dried under a stream of N_2, dissolved in methanol with the addition of the internal standard C_{46} and then split into two equal aliquots. One aliquot was subjected to acid hydrolysis with 5% HCl in methanol at 100°C overnight to cleave polar head groups.

Both non-hydrolyzed and hydrolyzed aliquots were analyzed with a Waters I-class Acquity ultra-high performance liquid chromatograph (Waters, Milford, MA, United States) coupled to a Q Exactive HF Orbitrap mass spectrometer (UHPLC-ESI-MS$_{Q\ Exactive}$ Thermo Fisher Scientific, Waltham, MA, United States) equipped with an electrospray ionization source. Samples were run with the mass spectrometer in full scan mode (450–1800 m/z) and again in selected ion monitoring (SIM) mode in order to enhance sensitivity for very low abundance lipids detected in the full scan mode. Analyte separation was achieved on an ACE 3 C18 column (2.1 × 150 mm, 3 um particle size, Advanced Chromatography Technologies Ltd., Aberdeen, Scotland) maintained at 45°C with a flow rate of 0.2 mL/min (Zhu et al., 2013). Solvent A was methanol, formic acid, and ammonia (100/0.04/0.10) and Solvent B was Isopropyl alcohol, formic acid, ammonia (100/0.04/0.10). In brief, GDGTs were eluted with the following gradient: 100% A for 1 min, followed by a gradient to 24% B over 5 min and a gradient to 65% B over

---

[1]https://github.com/scapella/trimal
[2]http://itol.embl.de/

55 min. The column was flushed with 75% B for 5 min and re-equilibrated with 100% A for 8 min.

In order to minimize the noise and collect enough data points per peak based on the Q Exactive resolution, multiple time windows related to each lipid class were applied in the selected ion monitoring (SIM), with each time window having the maximum of 20 compounds targeted using the scan mass point of 0.5 Da. Target lipids (**Supplementary Figure 1**) were determined with SIM [i.e., ammoniated ($[M + NH_4]^+$), sodiated ($[M + Na]^+$), and protonated ($[M + H]^+$) molecules]. The further MS/MS fragmentation was carried out only for $[M + NH_4]^+$ and $[M + H]^+$ molecules on selected samples using collision energy 20 or 25.

Glycerol dialkyl glycerol tetraethers and archeol were identified by retention time, molecular mass and MS/MS fragmentation patterns. Corresponding core lipids before and after acid hydrolysis were quantified relative to the response of the internal standard C_{46}. A standard mix solution was constructed with C_{46}, the purchased archeol standard (1,2-di-O-phytanyl-sn-glycerol from Avanti Lipids, Alabaster, AL, United States), and crenarchaeol which was purified from a marine sediment extract in Ann Pearson's lab in Harvard University. Dilution series of the standard mix solution ranged in concentration from 1 to 5 ng of each standard. Response factors of 1.5 for crenarchaeol and 4.0 for archeol were applied to core lipids, based on the averaged relative response of crenarchaeol and archeol relative to the C_{46} internal standard, respectively.

Only compounds with signal to noise ratio (SN) ≥ 5 (calculated automatically in the Xcalibur with the algorithm ICIS) were reported. Due to the lack of authentic internal standards, absolute quantification of MH-, DH-, PH-, and HPH-GDGTs can't be calculated. Instead, relative distribution of these GDGTs in individual intact polar lipid forms was reported separately here based on the peak area. Abundances of IP-GDGTs and -archeol were obtained by subtraction of core lipids between non-hydrolyzed and hydrolyzed fractions (Ingalls et al., 2012). When using this hydrolysis method, negative calculated values of intact polar lipids were obtained in the surface, likely due to differences in matrix effects related to variability in sample preparation. We treated the negative value as zero for individual IP-ethers when calculating the sum of IP-ethers, and only reported the concentration of total IP-ethers for each sample when comparing with the archaeal biomass (see **Supplementary Table 1**).

RESULTS

Water column profiles in the epipelagic zone in the NPSG were investigated from the HOT Station ALOHA (S2) during HOT cruise 296 and from the center of a cyclonic eddy (S16) during the MESO-SCOPE cruise KM1709 (**Supplementary Figure 2**). The hydrographic features (temperature, salinity, density, and chlorophyll a) were in general similar between the two samplings (**Supplementary Figure 2**). Station HOT296_S2 had a shallow salinity maximum at around 140 m, due to the subduction of high salinity waters into the pycnocline (Lukas, 2001). The DCM

occurred at 120 m in station HOT296_S2, and at 125 m in station KM1709_S16 (**Supplementary Figure 2**).

Vertical Profiles of the Archaeal Community Based on Investigation of the SSU rRNA Gene

A total of 162,748 high-quality archaeal sequences were recovered from the two vertical profiles at stations HOT296_S2 and KM1709_S16, which ranged from 1,695 to 40,915 sequences per sample, with an average of 9,573 per sample (**Supplementary Table 2**). Rarefaction curves of archaeal AASVs detected relative to sequencing effort, displayed clear signs of saturation, suggesting that archaeal diversity was adequately sampled (**Supplementary Figure 3**). In all, 231 AASVs were identified, which ranged from 18 to 23 AASVs per sample (a mean value of 21) above DCM of HOT296_S2 and from 9 to 18 (a mean value of 14) above DCM of KM1709_S16 (**Supplementary Figure 4A**). The richness observed within and below DCM varied between 33 and 78 AASVs per sample (a mean value of 52) in HOT296_S2, and 48–154 (a mean value of 105) AASVs in KM1709_S16 (**Supplementary Figure 4A**). The Shannon index estimated alpha diversity ranged from 2.34 to 2.82 (a mean value of 2.54) above the DCM in HOT296_S2, and from 1.82 to 2.40 (a mean value of 2.04) in KM1709_S16 (**Supplementary Figure 4B**). Within and below DCM, the Shannon index varied between 2.33 and 3.14 (a mean value of 2.66) in HOT296_S2 and between 2.82 and 4.08 (a mean value of 3.63) in KM1709_S16 (**Supplementary Figure 4B**). Diversity estimates of richness and Shannon index both were generally lower above DCM, relative to those within and below DCM.

Archaeal community structure varied with depth. Above the DCM, Thermoplasmatota were by far the predominant, if not exclusive, planktonic archaeal group present by several orders of magnitude. Thaumarchaeota represented less than 0.5% Archaea at 100 m at both stations, Halobacterota [the phylum proposed based on genome phylogenies (Parks et al., 2018)[3]] represented less than 0.4% Archaea at 75 and 100 m at the HOT296_S2 station and Nanoarchaeota represented less than 0.1% of Archaea at 25 and 45 m at the HOT296_S2 station (**Figure 1** and **Supplementary Table 2**). Within and below the DCM, planktonic Thermoplasmatota accounted for 53.3–67.2% of the total Archaea in HOT296_S2 and 43.9–71.8% in KM1709_S16. Thaumarchaeota were significantly enriched in and below the DCM, comprising 32.2–46.1% of archaeal rRNAs in HOT296_S2, and 28.2–55.2% in KM1709_S16 (**Figure 1** and **Supplementary Table 2**). Planktonic Thermoplasmatota were dominated by MGII (51.9–93.9% Archaea above DCM, and 26.4–51.0% Archaea within and below DCM) being relatively more abundant, compared to MGIII (6.1–48.1% Archaea above DCM, and 17.6–30.0% Archaea within and below DCM) (**Figure 1** and **Supplementary Table 2**).

Non-metric multidimensional scaling (NMDS) analysis was used to visualize Bray-Curtis dissimilarity profiles of AASV distributions among the different samples. NMDS plots showed

---

[3] http://gtdb.ecogenomic.org/tree

Depth Distribution of Archaeal Ether Lipids

Hexose- phosphohexose-, PH-, DH-, MH-, and C-GDGTs and C-archeol (**Supplementary Figure 1**) were identified here and their structures were confirmed by MS/MS analysis (see **Supplementary Text** and **Supplementary Figures 5–10**). Although we did not determine a limit of detection or quantification for each GDGT, a peak that was five times the noise (a reasonable value for a limit of quantification using SIM) was estimated to be 0.00060 ng/L (crenarchaeol isomer in the acid-hydrolyzed fraction at 100 m in KM1709_S16) and was the lowest concentration we calculated. The quantification limit was applied to core lipids of GDGTs detected in hydrolyzed and non-hydrolyzed fractions. The peak for GDGT-1 at 20 m in KM1709_S16 had a SN close to 5 and was therefore less than the limit of quantification of 0.0006 ng/L. This compound however, was likely a real peak showing the presence of GDGT-1, and thus was reported here. All sampling depths showed the existence of GDGTs, except the 45 m station of HOT296_S2 where GDGTs were below the quantification limit (0.00060 ng/L) (**Figure 3** and **Supplementary Figure 5**). In this sample, a mass feature with the same exact mass as GDGT-0 was detected after acid hydrolysis of the sample, but it could not be differentiated from noise (signal to noise ratio = 2, **Supplementary Figure 5**). This lipid sample could be treated as the control collected *in situ*, compared to other lipid samples. Subsequent GDGT analyses discussed here therefore refer to all but this one sample. The identity of core and intact polar lipids within shallow depths ($\leq$40 m) was validated by detection of characteristic fragment ions in MS/MS spectra (see **Supplementary Figures 11–14**). The intensity of the mass spectral signal was lower in shallow samples compared to deep ones (e.g., those below DCM), but dominant product ions were still observed in near-surface samples (**Supplementary Figures 11–14**).

Hexose- phosphohexose-, PH-, DH-, MH-, and C-GDGTs and C-archeol were detected in the non-hydrolyzed fractions while total GDGTs and total archeol were detected in acid hydrolyzed fractions (**Figure 3**). Above the DCM, concentrations of total GDGTs were very low, ranging from 0.015 to 0.077 ng/L in HOT296_S2 and from 0.0058 to 0.19 ng/L in KM1709_S16 (**Figure 3**, **Supplementary Table 1**, and **Supplementary Figure 15**). GDGTs under the quantification limit (0.00060 ng/L) would not have a significant impact on the GDGT distributions at the surface (see the measured concentrations of individual total GDGTs compared to the limit in **Supplementary Table 1**). Within and below the DCM, total GDGT concentrations increased greatly by 100–1000 fold, and varied between 6.73 and 11.32 ng/L in HOT296_S2, and 9.22–18.59 ng/L in KM1709_S16 (**Figure 3**, **Supplementary Table 1**, and **Supplementary Figure 15**). Along depth profiles, the total GDGTs reached maximum values at 175 m at the HOT296_S2 station, and at 160 m at the KM1709_S16 station (**Figure 3**,
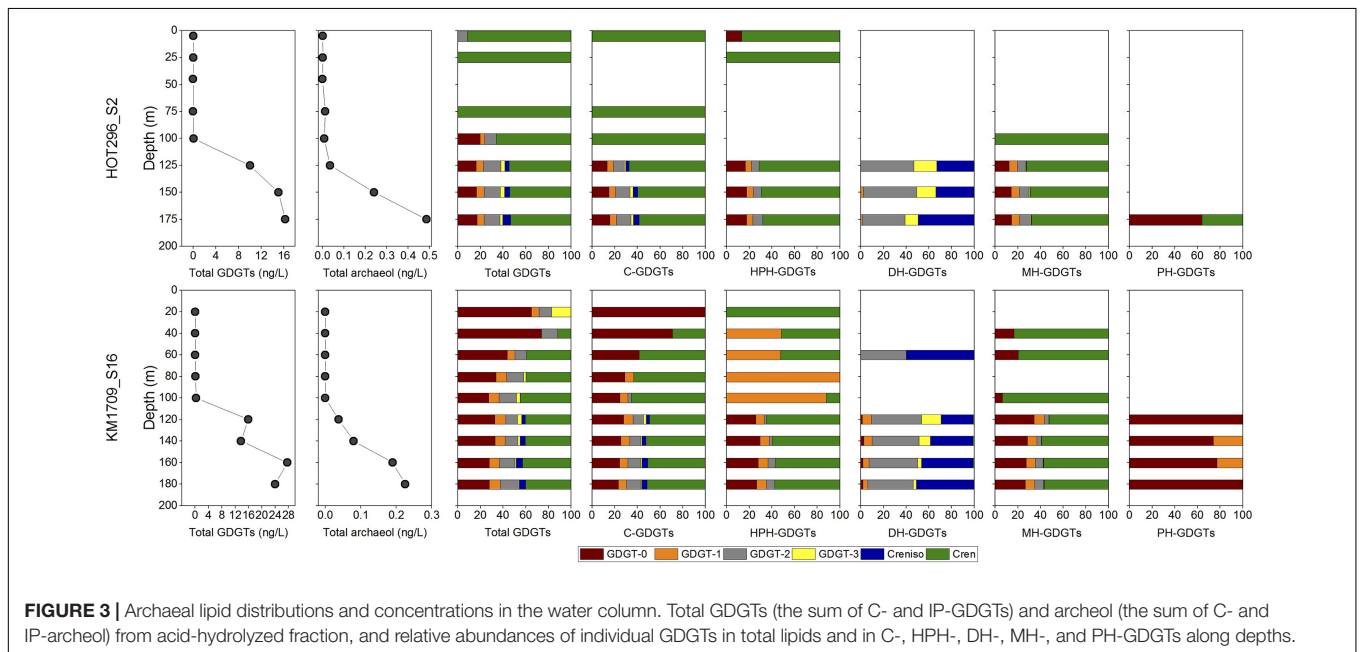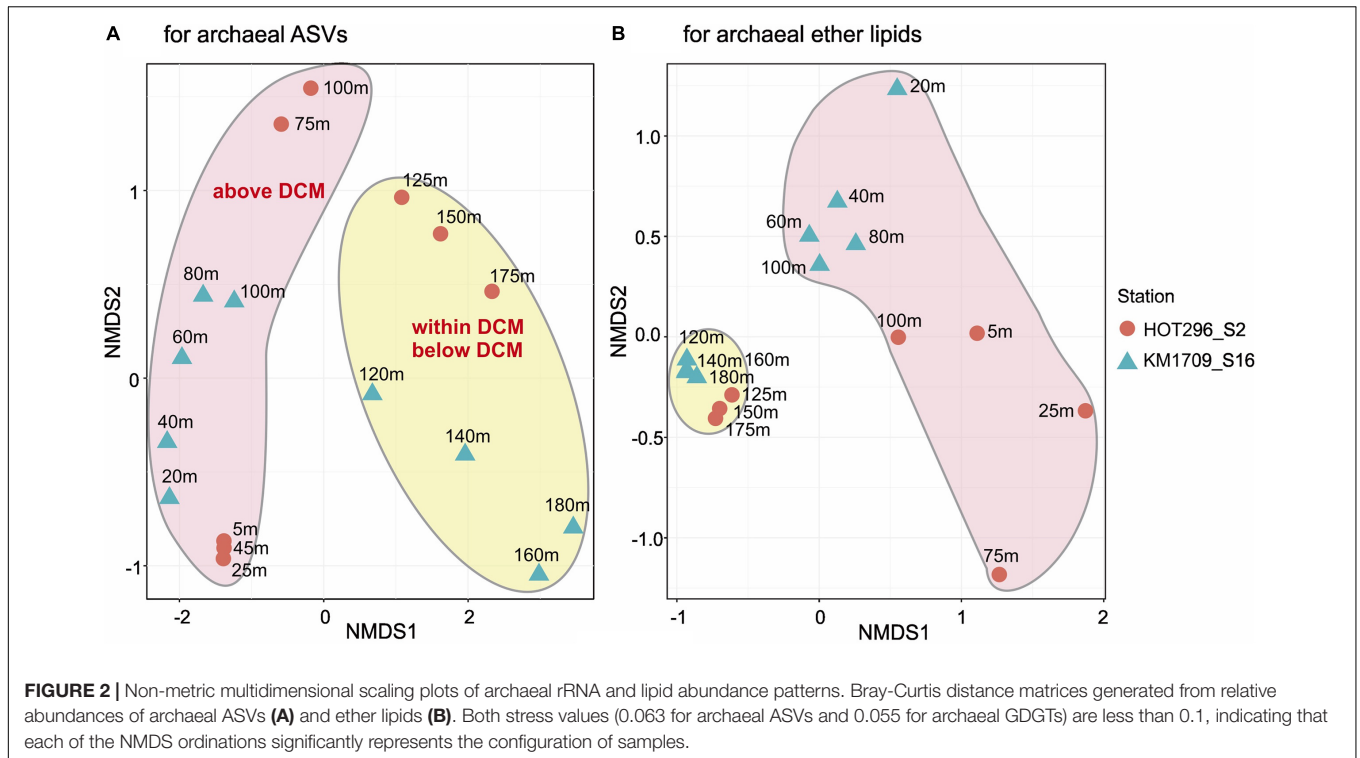
Supplementary Table 1, and **Supplementary Figure 15**). Profiles of total archeol showed the same pattern in depth profiles at both stations, except that the maximum value at the KM1709_S16 station was at 180 m (**Figure 3**, **Supplementary Table 1**, and **Supplementary Figure 15**). Archeol was either undetectable, or accounted for only a minor fraction of archaeal lipids (mostly less than 5%), at all depths except 75 m of HOT296_S2, where the diethers comprised 78% of all archaeal ether lipids (**Figure 3**, **Supplementary Table 1**, and **Supplementary Figure 15**). For the intact polar counterparts of archeol, only monohexose-archeol was detected in our samples.

In the total GDGT pool and in the C-, HPH-, and MH-GDGTs pools, crenarchaeol was either the only compound, or the predominant compound detected above the DCM in HOT296_S2. Other GDGTs were present in lesser abundance, or below detection limits. For example, GDGT-2 from hydrolyzed fractions and GDGT-0 in the HPH-GDGTs were at lower abundance at 5 m depth (**Figure 3**). The distributional patterns of GDGTs above DCM in KM1709_S16 were more complex. In the total GDGTs and in C-GDGTs, GDGT-0 was the most abundant molecule at 20 and 40 m, whereas samples at 60, 80, and 100 m were dominated by GDGT-0 and crenarchaeol (**Figure 3**). In the HPH-GDGT fraction, crenarchaeol was the only lipid detected at 20 m, GDGT-1 was predominant at 80 and 100 m, and samples at 40 and 60 m were composed of almost equal GDGT-1 and crenarchaeol (**Figure 3**). GDGT-2 and crenarchaeol isomer were observed abundantly at 60 m in the DH-GDGT fraction (**Figure 3**). The MH-GDGT patterns above the DCM were more similar to those found at the HOT296_S2 station, and were dominated by crenarchaeol (**Figure 3**).

Within and below the DCM, GDGTs at the HOT296_S2 station closely resembled those of KM1709_S16, and contained a high abundance of GDGT-0 and crenarchaeol in the total GDGTs and in C-, HPH-, and MH-GDGT fractions, with GDGT-2 and the crenarchaeol isomer dominating the DH-GDGTs, and GDGT-0 being the most abundant lipid in PH-GDGTs (**Figure 3**).

Non-metric multidimensional scaling analyses of the depth distributions of GDGTs and archeol in the total lipids, in core lipids, and in HPH-, MH-, DH-, and PH-GDGT fractions were performed (**Figure 3**). Similar to the NMDS results from AASV analyses, samples within and below DCM were clustered together and were well separated from samples above the DCM, but grouped together less tightly (**Figure 2B**).

A non-targeted full scan analysis was also applied here, as well as targeted screening method for the derivatives of GDGTs and archeol including unsaturated, hydroxy and methoxy counterparts as reported in Elling et al. (2014, 2017), with one exception for unsaturated archeol which was not targeted. This lipid class was not identified in our samples from full scan method. And none of these derivatives or lipids related to the production from Archaea were detected in surface water. Only Glycerol Dibiphytanol Diethers (GDDs) were observed at and below 100 m in KM1709_S16 and within and below the DCM in HOT296_S2, which were reported to be components of membrane lipids in Thaumarchaeota (Elling et al., 2014, 2017).

FIGURE 2 | Non-metric multidimensional scaling plots of archaeal rRNA and lipid abundance patterns. Bray-Curtis distance matrices generated from relative abundances of archaeal ASVs **(A)** and ether lipids **(B)**. Both stress values (0.063 for archaeal ASVs and 0.055 for archaeal GDGTs) are less than 0.1, indicating that each of the NMDS ordinations significantly represents the configuration of samples.

FIGURE 3 | Archaeal lipid distributions and concentrations in the water column. Total GDGTs (the sum of C- and IP-GDGTs) and archeol (the sum of C- and IP-archeol) from acid-hydrolyzed fraction, and relative abundances of individual GDGTs in total lipids and in C-, HPH-, DH-, MH-, and PH-GDGTs along depths.

Detection and Analyses of the Putative Homologs of Archaeal GDGT Ring Synthases (GrsA and GrsB) in MGII/III Archaea, and Their Distributions With Depth

The radical S-adenosylmethionine (SAM) genes encoding GrsA and GrsB in *Sulfolobus acidocaldarius* were recently experimentally verified as having authentic archaeal GDGT ring synthase activity in this archaeal thermophile, and homologs were identified in Thaumarchaeota genomes (Zeng et al., 2019). We therefore used these genes to identify candidate homologous proteins from planktonic MGII/III Archaea. A total of 120 homologous sequences from MGII/III Archaea were detected in the GTDB database (Parks et al., 2018) and 52 representatives of putative proteins were in the Station ALOHA

gene catalog, at the identity of 20–35% with Grs sequences from *S. acidocaldarius* (**Figure 4**, **Supplementary Figure 16**, and **Supplementary Table 3**). An *E*-value cutoff of $\leq 1e^{-5}$ was used here for retrieving homologs at low stringency, since mesophilic archaeal proteins may be only distantly related to (hyper)thermophilic GrsA and GrsB proteins identified in Zeng et al. (2019). Consistent with our results in Station ALOHA surface waters (data not shown), Zeng et al. (2019) reported the near absence of Thaumarchaeota GrsA/B homologs in surface waters in Station ALOHA metagenomes, with the vast majority of Thaumarchaeota GrsA/B homologs appearing within and below the DCM.

To further explore the phylogenetic relationships among GrsA and GrsB protein homologs, we constructed a rooted phylogenetic tree that included inferred protein sequences from MGI, II, and III and other archaeal isolates, as described in **Supplementary Table 1** of Zeng et al. (2019) and retrieved from the GTDB database and ALOHA gene catalog. The phylogenetic tree showed that GrsA and GrsB clades consisted exclusively of putative homologs from thermophilic isolates that are known to produce cyclized GDGTs (**Figure 4A**). The putative homologs from MGI grouped together, and were only distantly related to archaeal thermophile GrsA and GrsB (**Figure 4A**). Compared to the outgroup of Saci_1785 radical SAM homologs, all of the putative MGII/III homologs clustered together as an ingroup, although they were more distantly related to clades of GrsA and GrsB from known and cultivated Archaea (**Figure 4A** and **Supplementary Figure 16**).

The metagenomic sequence read coverage of putative homologs from MGII/III was calculated for each sample from different depths within the euphotic zone at Station ALOHA collected during November 2014 to March 2015. The coverage values for each homolog per sample were lower above DCM, ranging between 1 and 5.22X coverage with an average of 1.86X, and were slightly elevated within and below the DCM, ranging between 1.01X and 6.43X coverage, with an average of 2.32X (**Figure 4B**). Since multiple GrsA/B copies have been found in a variety of Archaea (Zeng et al., 2019 and **Figure 4** and **Supplementary Table 3** in this study), we did not attempt to estimate the abundance of MGII/III Archaea based on the Grs homolog abundances.
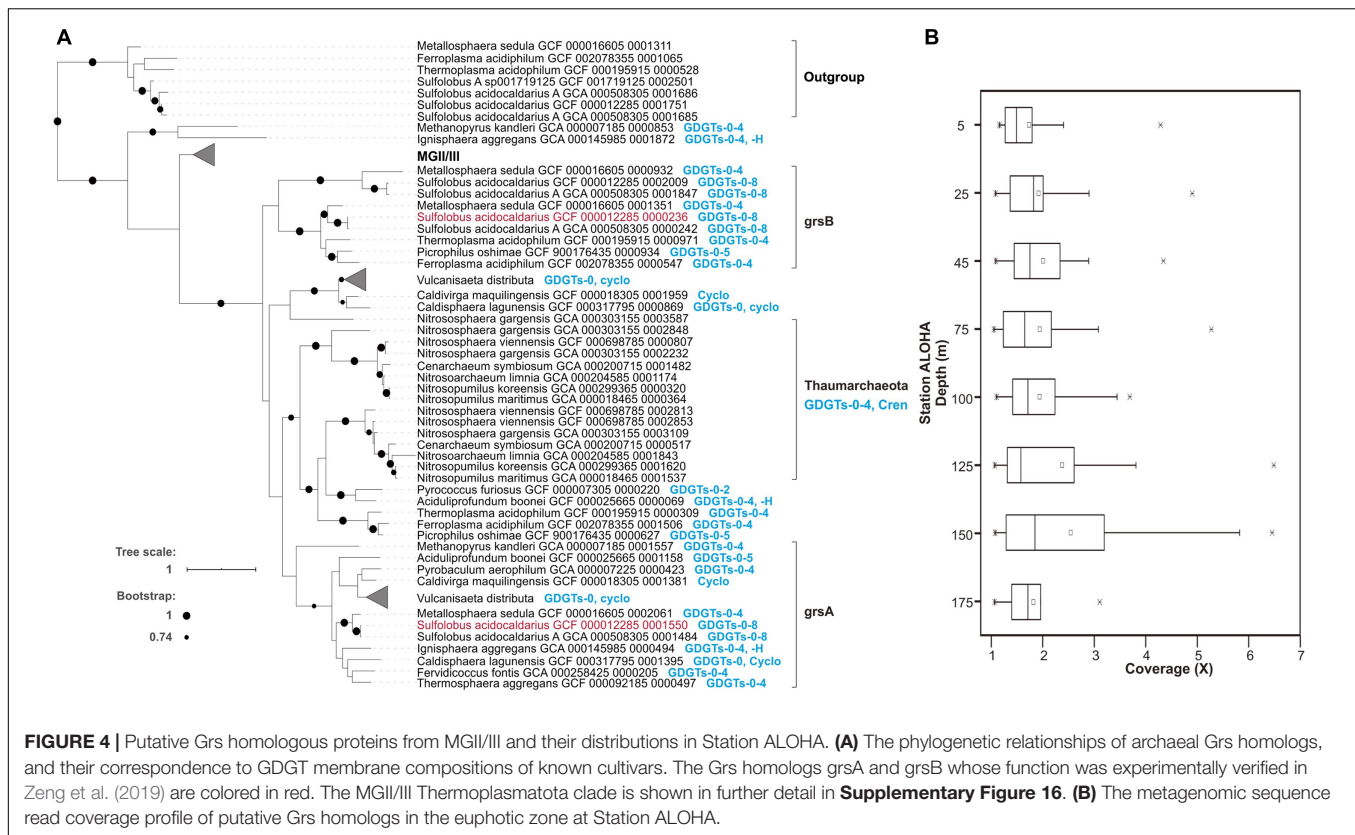
Calculation of Possible Cellular GDGTs and Archeol for Planktonic Thermoplasmatota

Quantitative estimates of Archaea, Thermoplasmatota, and Thaumarchaeota SSU rRNA gene abundances were obtained via ddPCR assays, using universal SSU rRNA primers, in combination with microbial community structure analyses achieved by high-throughput sequencing of the same PCR products. The absolute abundance of Thaumarchaeota changed markedly in depth profiles (ranging between 0 and 6.42×10^6 SSU rRNA copies/L), whereas total archaeal SSU rRNA genes (ranging between 2.11×10^6 and 1.22×10^7 SSU rRNA copies/L) and total Thermoplasmatota SSU rRNA genes (ranging between 2.11×10^6 and 8.55×10^6 SSU

rRNA copies/L) varied much less with depth (**Figure 1** and **Supplementary Table 1**).

At depths shallower than 45 m at HOT296_S2 and 100 m at KM1709_S16, only Thermoplasmatota rRNAs were detectable, and their abundances ranged from 2.11×10^6 to 6.02×10^6 SSU rRNA copies/L (**Figure 1** and **Supplementary Table 1**). Within these depths, concentrations of total IP-ethers ranged from 0.0010 to 0.015 ng/L (**Figure 5A** and **Supplementary Table 1**). We estimated the potential cellular ether lipid content of Thermoplasmatota, assuming they were the origin of the co-occurring ether lipids. The large majority of marine planktonic Bacteria and Archaea are considered to contain one to two copies of SSU rRNA gene per cell (Brown and Fuhrman, 2005; Needham et al., 2018), which may be their ecological strategies to grow in the oligotrophic environments (Klappenbach et al., 2000). Pure cultures of marine Thaumarchaeota, *Nitrosopumilus maritimus*, contain only one copy in the genome (Spang et al., 2010). Here, one copy was used for Thermoplasmatota and Thaumarchaeota. Using the ratio of total IP-ethers to total SSU rRNA gene abundance of Thermoplasmatota, the lipid content was 2.5×10^{-10} to 3.34×10^{-9} ng/cell, with an average of 1.21×10^{-9} ng/cell (**Figure 5B** and **Supplementary Table 1**). Lipids in Archaea with a cell diameter of 500 nm are estimated to be 1.4×10^{-6} ng/cell (Lipp et al., 2008). Based on this assumption, GDGTs appeared to comprise only 0.017–0.22% of the expected intact polar lipid content per Thermoplasmatota cell, while archeol constituted 0–0.019% (**Figure 5C** and **Supplementary Table 1**).

To estimate the abundance of IP-ethers derived from the biomass of MGII/III Thermoplasmatota and MGI Thaumarchaeota within and below DCM in comparison with the measured counter parts, we used the conversion factor for Thermoplasmatota ethers in the upper water column (1.21×10^{-9} ng/cell), and the published value for Thaumarchaeota ethers (1.27×10^{-6} ng/cell; Elling et al., 2014). Halobacterota probably contributing to the intact polar lipid pool within and below the DCM, only constituted less than 1% of archaeal populations, and thus their producing IP-ethers were not considered here. Within and below DCM, the estimated IP-ethers derived from MGII/III Archaeal biomass ranged between 5.18×10^{-3} and 1.03×10^{-2} ng/L, and estimated thaumarchaeotal IP-ethers were between 3.57 and 8.15 ng/L (**Figure 5D** and **Supplementary Table 1**). Hence, less than 1% of estimated total IP-GDGTs/ethers could be attributable to MGII/III Archaea within and below the DCM if Thermoplasmatota are in fact a source of these ether lipids. Within the same depths, the estimated abundances of total IP-ethers from all sources varied from 4.53–6.64 ng/L in HOT296_S2, and from 3.57 to 8.16 ng/L in KM1709_S16; whereas the measured total IP-ethers ranged between 6.18 and 10.21 ng/L in HOT296_S2 and between 7.54 and 16.92 ng/L in KM1709_S16 (**Figure 5D** and **Supplementary Table 1**). The measured concentrations of the total IP-ethers were about two to three-fold greater than expected. Here, we estimated the cellular content of ether lipids for Thaumarchaeota based on the concentration of IP-ethers relative to the SSU rRNA gene abundance of Thaumarchaeota detected within and

FIGURE 4 | Putative Grs homologous proteins from MGII/III and their distributions in Station ALOHA. **(A)** The phylogenetic relationships of archaeal Grs homologs, and their correspondence to GDGT membrane compositions of known cultivars. The Grs homologs grsA and grsB whose function was experimentally verified in Zeng et al. (2019) are colored in red. The MGII/III Thermoplasmatota clade is shown in further detail in **Supplementary Figure 16**. **(B)** The metagenomic sequence read coverage profile of putative Grs homologs in the euphotic zone at Station ALOHA.

below the DCM. This value ranged between 1.74×10^{-6} and 2.93×10^{-6} ng/cell with an average of 2.31×10^{-6} ng/cell.

DISCUSSION

The Near-Exclusive Presence of Thermoplasmatota in Oligotrophic Surface Waters

The NPSG occupies ~40% of the Earth's surface and represents the largest circulation feature on Earth (Karl and Church, 2014), characterized by near-permanent stratification and low seasonality (Bingham and Lukas, 1996; Bryant et al., 2016; Karl and Church, 2017). The general characteristics of the NPSG generate strong depth structuring of microbial community composition in the water column, with most near-surface variability attributable to seasonal changes in day-length and light, and sporadic wind speed variation (DeLong et al., 2006; Bryant et al., 2016).

To better understand the ecology and biogeochemistry of planktonic archaeal populations, we performed synoptic measurements of their abundance patterns and membrane ether lipid distributions in the epipelagic zone of the NPSG. Planktonic Thermoplasmatota and planktonic Thaumarchaeota were confirmed as the two main groups of Archaea in the water column. Thermoplasmatota were abundant at all depths, while Thaumarchaeota increased in abundance within and just

below the DCM. These observations are generally consistent with previous reports in the NPSG (DeLong et al., 2006; Brown et al., 2009; Church et al., 2010) and other oceanic regions (Massana et al., 1997, 2000; Mincer et al., 2007; Hugoni et al., 2013; Parada and Fuhrman, 2017).

The archaeal communities observed within the samples collected tended to separate into two general clusters, one above the DCM and one within and below the DCM. Thermoplasmatota (MGII and MGIII) dominated above the DCM, while within and below the DCM both Thermoplasmatota and MGI Thaumarchaeota were prevalent. Thaumarchaeota were not detected at depths shallower than 45 m at HOT296_S2 and 100 m at KM1709_S16 (**Figure 1**), again consistent with previous reports in this region (Brown et al., 2009). Based on Station ALOHA gene catalog data, the analysis of ribosomal protein S3 (RpS3) showed no presence of MGI Thaumarchaeota from samples shallower than 45 m during November 2014 to April 2016 (**Supplementary Table 4**). One reason may be the inhibitory influences of sunlight on growth of nitrifying Thaumarchaeota in surface waters (Church et al., 2010; Merbt et al., 2012; Horak et al., 2018).

The universal SSU rRNA gene PCR primers used here have been reported to remove biases against Thaumarchaeota (Parada et al., 2016). Based on ddPCR and sequencing using this universal primer set, MGI Thaumarchaeota abundances were approximately 10^6 SSU rRNA copies/L within and below DCM (**Figure 1** and **Supplementary Table 1**), similar to amoA gene densities reported at these depths in the same

FIGURE 5 | The estimates of cellular content of GDGTs and archeol from Thermoplasmatota and their contribution to the intact polar lipid pool. **(A)** The measured abundances of total IP-GDGTs at the surface. **(B)** The cellular content of GDGTs and archeol estimated for Thermoplasmatota. **(C)** The relative abundance of GDGTs and archeol in cellular membrane lipids of Thermoplasmatota. **(D)** The comparison of measured and estimated intact polar ether lipids within and below DCM, showing the latter IP-ethers are dominantly from Thaumarchaeota.

oceanic region (Church et al., 2010). These observations were also supported using an archaeal-specific primer set (ARC344FB/ARC519R; Raskin et al., 1994; Sørensen and Teske, 2006; Teske and Sørensen, 2008). These archaeal-specific primers also showed the complete absence of Thaumarchaeota in NPSG surface waters (**Supplementary Figure 17**). In general, total archaeal SSU rRNA amplicon-based richness and Shannon indices were much lower above the DCM in both stations (**Supplementary Figure 4**), compared to those within and below the DCM. The rarefaction curves also reached an asymptote (**Supplementary Figure 3**).

The MGIII planktonic Archaea detected in this study accounted for an average of approximately 30% of total archaeal sequences recovered, with a maximum abundance of 48.1% at 60 m of KM1709_S16 (**Figure 1**). While few studies have reported the presence of MGIII Archaea in the euphotic zone (Brown et al., 2009; Galand et al., 2010; Haro-Moreno et al., 2017), they represented up to 10% of all archaeal sequences in samples from the Mediterranean DCM (Galand et al., 2010), and 11.8% at 10 m depth at station ALOHA from HOT cruise 169 (Brown

et al., 2009). Haro-Moreno et al. (2017) reconstructed epipelagic MGIII genomes by binning the assembled fragments, which contained numerous photolyase and rhodopsin genes as well as genes coding for degradation and uptake of protein and lipids. These metabolic features seem complementary to oligotrophic conditions found in surface waters, and may allow MGIII with photoheterotrophic lifestyles to thrive, like their MGII relatives, in the ocean's euphotic zone (Brown et al., 2009; Galand et al., 2010; Haro-Moreno et al., 2017). MGIII Archaea therefore appear to have the potential to live in the euphotic zone, and they may at times be comparable in abundance to MGII Archaea.

Planktonic Thermoplasmatota Contributions to Surface Water GDGTs

From larger volumes of seawater (>3000 L/sample), Hurley et al. (2018) reported HPH-, PH-, DH-, MH-, and C-GDGT distribution patterns along depth profiles between 0 and 1000 m. The total lipid concentration reported by Hurley et al. (2018) exhibited a similar pattern to those found here with low

concentrations detected above the DCM (minimum 0.0015 ng/L at 5 m) and high concentrations within and below the DCM (maximum 11.84 ng/L at 158 m) (**Supplementary Figure 18**). Meanwhile, concentrations of the total intact polar lipids also showed the minimum (0.00021 ng/L) and maximum (8.94 ng/L) values at these same depths (Hurley et al., 2018). GDGT compositions within and below the DCM in our study here resembled those of Hurley et al. (2018) whereas above the DCM, similarities were seen mainly in HPH- and MH-GDGTs from HOT296_S2, and in DH- and MH-GDGTs in KM1709_S16 (**Supplementary Figure 18**). Compared to the study in Hurley et al. (2018), the depth profiles of GDGTs in our study showed a larger difference, likely due to the oligotrophic condition and near-permanent stratification in our stations. We used NMDS analysis to decipher the relationship of the GDGT distribution among the different samples reported by Hurley et al. (2018) and found the same clustering patterns as we report here, except that all their samples at Station 2 grouped together (**Supplementary Figure 19**). In depth profiles of the relative distributions of different GDGT classes (see the last plot for each station in **Supplementary Figure 18**), C-GDGTs were much more abundant above, than within and below the DCM in Station 2, which was also observed for their Stations 7, 15, and 23 (Hurley et al., 2018).

A Mantel test was conducted to explore the correlation between the compositions of the archaeal community (based on AASVs) and ether lipid distribution (based on all lipids) by Bray-Curtis distance, with 999 permutations. The significant relationship identified (significance = 0.001) suggests that the biosynthesis of GDGTs and archeol is attributable to specific archaeal populations, as reported in previous studies (Pearson and Ingalls, 2013; Schouten et al., 2013; Lincoln et al., 2014; Sollai et al., 2018). MGII and MGIII planktonic Archaea are therefore still potentially a source of these lipids in shallow open ocean waters, consistent with the findings of Ma et al. (2020) in Northwestern Pacific Ocean surface water. The different depth profiles of GDGTs observed in the surface water between two stations are probably due to the divergent phylogenies in Thermoplasmatota (Rinke et al., 2019), which could also be seen in the distant relationships of archaeal ASVs detected at these two sites (**Figure 2**). Our results contrast with a recent report by Besseling et al. (2020) in surface waters of the coastal North Sea and the North Atlantic Ocean, where MGII/III Thermoplasmatota dominated. However, the threshold of detection for archaeal IP-GDGTs using their analytical methods was reported to be about 7×10^6 Thaumarchaeota SSU rRNA copies/L (Besseling et al., 2020). In our study here, IP-GDGTs were detected in surface waters where MGI Thaumarchaeota were absent and MGII/III Thermoplasmatota were very abundant. The potential for Thermoplasmatota tetraether lipid production was also supported by the presence of genes encoding enzymes associated with ether bond formation and isoprenoid saturation in MAGs of representatives of these two archaeal groups (Iverson et al., 2012; Deschamps et al., 2014; Villanueva et al., 2017). While a small SSU rRNA signal was detected from Woesearchaeota in surface waters on the HOT296_S2 cruise (**Figure 1**), no ether bond formation genes

have yet been reported in MAGs from this group of Archaea (Villanueva et al., 2017).

Detection of GDGT Ring Synthase Homologs in Planktonic Thermoplasmatota

Zeng et al. (2019) recently identified two GDGT ring synthases in *S. acidocaldarius*, GrsA and GrsB (having similarity of 31%), which were found to introduce rings separately at the C-7 and C-3 positions of the core GDGT lipids, with GrsB preferring substrates with existing C-7 rings. Phylogenetic analysis indicates that GrsA and GrsB clades are found predominantly in (hyper)thermophiles, with a few sequences from mesophilic Archaea that mostly belonged to Bathyarchaeota and Thermoplasmatales, and many of them contain the two paralogs (Zeng et al., 2019). The putative homologs from planktonic Thaumarchaeota clustered into a group separate from (hyper)thermophilic archaeal GrsA and GrsB clades (Zeng et al., 2019). Given that Thaumarchaeota are phylogenetically affiliated with the Crenarchaeota, it is perhaps not surprising that their Grs homologs share significant protein sequence identity, even though these genes originate from mesophiles or psychrophiles.

Our phylogenetic observations are consistent with those of Zeng et al. (2019), and we report here additionally that GrsA homologs are also found in planktonic Thermoplasmatota (**Figure 4B**, **Supplementary Figure 16**, and **Supplementary Table 3**). It is possible that since MGII/III Thermoplasmatota are mesophilic organisms in ocean surface waters, they do not produce multiple cyclopentane rings (i.e., five cyclopentane rings or more) that are characteristic of membrane lipids in acidophiles, halophiles, and thermophiles (Pearson and Ingalls, 2013; Schouten et al., 2013). The planktonic Thermoplasmatota ring synthase homologs belonged to the Grs clade, and were more closely related to GrsA and GrsB clades, than some other archaeal homologs, for example those from extremely thermophilic isolates *Methanopyrus kandleri* and *Ignisphaera aggregans* (also known to produce cyclized GDGTs) which although have not been tested yet for the same function as GrsA and GrsB (**Figure 4A**). The catalytic properties of mesophilic MGII/III Thermoplasmatota ring synthase may be distinct from GrsA and GrsB of (hyper)thermophiles identified in Zeng et al. (2019), since their candidate homologs exhibit only distant relationships to GrsA and GrsB families. Future studies need to put more focus on the distribution and function of GDGT synthases from mesophiles. At present however, we cannot exclude the possibility that the putative MGII/III Thermoplasmatota homologs identified here are other radical SAM proteins.

Cellular Content of Ether Lipids in Planktonic Thermoplasmatota

While we assumed that trace MGI Thaumarchaeota would exist and be contributor to the IP-ether pools within depths shallower than 45 m at HOT296_S2 and 100 m at KM1709_S16, the surface MGI thaumarchaeotal abundance estimated would range from 7.87×10^2 to 1.17×10^4 SSU rRNA copies/L with an average of 4.03×10^3 SSU rRNA copies/L, based on concentrations of

the total IP-ethers measured and the thaumarchaeotal cellular content (1.27×10^{-6} ng/cell, Elling et al., 2014) of GDGTs and archeol (**Supplementary Table 1**), Therefore, the estimated abundance of Thaumarchaeota relative to the total archaeal biomass detected would vary between 0.020 and 0.26% (a mean value of 0.10%), which exceeds the theoretical detection limit of our method (approximately 0.021–0.059% with a mean value of 0.038%) using the universal PCR primers. Similarly, the archaeal-specific primer set yielded 90,104–127,794 high-quality reads per sample (a mean value of 107,817) at the surface, with a detection limit of approximately 0.00078–0.0011% (a mean value of 0.00094%). Therefore, if Thaumarchaeota SSU rRNA genes were present in surface waters, they should have been detectable, yet they were not.

We also calculated the cellular content of ether lipids in planktonic Thermoplasmatota inhabiting the surface or MGI Thaumarchaeota within and below the DCM, assuming the ether lipids originate from these groups of Archaea. The average value of 1.21×10^{-9} ng/cell for Thermoplasmatota in surface waters, was three orders of magnitude lower than reported cellular ether content for cultivated MGI Thaumarchaeota (1.27×10^{-6} ng/cell, Elling et al., 2014). Subsurface MGI Thaumarchaeota had a mean lipid content of 2.31×10^{-6} ng/cell, two fold higher than the culture estimate, suggesting the need to derive more accurate lipid/cell conversation factors in future studies. Another explanation would be that cell diameters of Thaumarchaeota within and below the DCM in Station ALOHA may be larger than the cultivated one in Elling et al. (2014). The difference in the analytical procedures used in our study and in Elling et al. (2014) may also be responsible.

Ecological and Geological Implications

There are several potential explanations for the combined absence of Thaumarchaeota rRNA-genes in surface waters, along with low estimated cellular IP-ether lipids in MGII/III Thermoplasmatota found in the same habitat. It is possible that small numbers of Thaumarchaeota cells may be transported upward, either by mixing of thaumarchaeotal cells into surface waters, by vertical migrators or buoyant transport from greater depths (Yayanos and Nevenzel, 1978; Smith et al., 1989; Sommer et al., 2017). In inhospitable surface waters, upwardly transported thaumarchaeotal cells may not survive, and their nucleic acids might decay more rapidly than their lipids. This could explain the absence of thaumarchaeotal SSU rRNA genes in surface waters, while small traces of their lipids remain. A related scenario invokes surface water GDGTs as thaumarchaeotal membranes or micelles, upwardly transported in similar ways. Cultured MGI Thaumarchaeota have been reported to synthesize crenarchaeol as their major membrane lipid with lower abundances of GDGTs-0 to 3 (Sinninghe Damsté et al., 2002; Elling et al., 2014, 2015, 2017), consistent with the GDGT distributions within and below the DCM in our samples. At 5 and 25 m in HOT296_S2, crenarchaeol was determined to be 12-fold greater than GDGT-2 and at least 23-fold greater than the remaining GDGTs (calculated based on the estimated quantification limit of 0.00060 ng/L); whereas GDGT-0 was at least 6- or 9-fold greater than crenarchaeol at 20 and 40 m

in KM1709_S16 (**Figure 3** and **Supplementary Table 1**). The GDGT profiles observed in the surface water samples are quite distinct from those produced by cultured MGI Thaumarchaeota however, so potential explanations invoking upward transport of Thaumarchaeota cells or lipids to the surface waters seem unlikely, or would have to include a mechanism for selectively transporting a structural subpopulation of the lipid pool.

The possibility still remains then, that planktonic Thermoplasmatota were the sole source of surface water GDGTs we report here in the NPSG samples. The facts that planktonic Thermoplasmatota were the only detectable Archaea in surface waters, and that their genomes encoded ether lipid biosynthetic gene homologs (Iverson et al., 2012; Deschamps et al., 2014; Villanueva et al., 2017) as well as GrsA homologs (this study), support this hypothesis.

If planktonic Thermoplasmatota were the source of surface water GDGTs and archeol, why did these lipids comprise less than an estimated 0.3% of their cellular lipids (**Figure 5C**)? This could be explained by the presence or predominance of lipids other than GDGTs or archeol that were not measured in our study, in keeping with a recent study of Besseling et al. (2020). It is possible that other common archaeal lipids such as other phytanyl ether lipids comprise the dominant components of MGII and MGIII planktonic archaeal membrane lipids (Villanueva et al., 2017).

With respect to implications for archaeal lipid export flux, the surface ocean is characterized by higher particle density and enhanced rates of protistan grazing and mesoplankton packaging, which could increase the incorporation of prokaryotic biomass into sinking flux (Richardson and Jackson, 2007; Close et al., 2013). The particle sinking rates reported can be as great as 820 m/d at Station ALOHA (Trull et al., 2008). Phytoplankton in the surface ocean have rapid growth and short generation times. Under these conditions, the contribution of surface MGII/III to subsurface and deeper water GDGTs could not be ignored. This may be part of reason that the estimated cellular content of ether lipids in Thaumarchaeota within and below the DCM is two-fold higher than the corresponding culture estimates. The biological origins of GDGTs exported to sediments have to be interpreted with caution. If MGII/III Thermoplasmatota are the source of surface water GDGTs, this implies that they are also a source of crenarchaeol, challenging the assumption that crenarchaeol is a biomarker of Thaumarchaeota, which until recently were thought to be exclusively ammonia oxidizing chemoautotrophs (Aylward and Santoro, 2020). If this was in fact true, our understanding of nitrogen cycling processes or the activity of Thaumarchaeota reconstructed using crenarchaeol in paleo studies (Dong et al., 2019; Elling et al., 2019), may be more complex than current models would suggest. Future efforts, including the investigation of Thermoplasmatota growth and lipid content, as well as processes such as grazing, packaging, and export from surface waters, may better constrain the biological origins of sedimentary archaeal lipids. The enrichment and isolation of representatives of MGII/III Thermoplasmatota as well as approaches like BONCAT (Dieterich et al., 2006), isotope labeling experiments and natural carbon isotope ratio detection have potential to further improve our understanding of ecological and physiological functions of planktonic archaeal groups.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

ED and AI designed the work. FL and LC conducted the lipid experiments and collected the lipid data. FL conducted the DNA experiment, collected the DNA data, and analyzed the lipid data. FL and KP analyzed the amplicon sequence data. AL analyzed the metagenomic data. FL and ED wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.610675/full#supplementary-material

Supplementary Figure 1 | Structures of core and intact polar GDGTs. Crenarchaeol isomer has a different stereochemistry in one of cyclopentane rings, compared to crenarchaeol (Sinninghe Damsté et al., 2018).

Supplementary Figure 2 | Sampling locations and associated hydrographic data. Hydrographic data is displayed along depth profiles with an interval of 1 m for HOT206_S2, and similarly for KM1709_S16.

Supplementary Figure 3 | Rarefaction curves of rRNA gene recoveries. The collectors curve derived for each sample was based on PCR reactions using universal primers **(A)**; or archaeal specific primers **(B)**. Archaeal specific primers, ARC344FB (5′-ACGGGGCGCAGCAGGCGCGA-3′, Raskin et al., 1994) and

ARC519R (5′-GGTDTTACCGCGGCKGCTG-3′, Sørensen and Teske, 2006), were used to target the V3 region of the SSU rRNA gene.

Supplementary Figure 4 | Archaeal community diversity and richness. Community metrics were calculated as richness **(A)** and Shannon index **(B)** along depths in both stations.

Supplementary Figure 5 | Fragmentation patterns of protonated and ammoniated acid-hydrolyzed C-GDGT archaeal lipid fractions for the 140 m depth sample in KM1709_S16 and the extracted ion chromatograms for the 45 m depth sample in HOT296_S2. LC-MS chromatograms show distributions of and MS/MS spectra show characteristic fragment ions of protonated **(A)** and ammoniated **(B)** C-GDGTs from acid-hydrolyzed fraction of sample at 140 m of KM1709_S16. The retention time for each C-GDGTs is also shown in panel **(A)**. **(C)** Enlarged MS/MS spectra of protonated C-crenarchaeol from panel **(A)**. **(D)** The extracted LC-MS chromatograms of C-GDGTs from both non-hydrolyzed and acid-hydrolyzed fractions of sample at 45 m of HOT296_S2. Three compound peaks are labeled here for the potential assignment to GDGTs. The green colored one with m/z 1302 has the same retention time as protonated GDGT-0. The dark red colored one with m/z 1318 has the retention time very close to sodiated GDGT-3. The MS/MS spectra however, are not able to show the characterized fragmentation pattern of sodiated GDGT-3, which would be dominated by the precursor ion $[M + Na]^+$ (m/z 1318), since this GDGT cannot be fragmented by our MS/MS method. The dark yellow colored one with m/z 1314 is eluted 1 min later than sodiated crenarchaeol.

Supplementary Figure 6 | Fragmentation patterns of protonated and ammoniated acid-hydrolyzed archeol. LC-MS chromatograms show distributions of and MS/MS spectra show characteristic fragment ions of protonated and ammoniated archeol from acid-hydrolyzed fraction of sample at 120 m of KM1709_S16.

Supplementary Figure 7 | Fragmentation patterns of protonated and ammoniated non-hydrolyzed HPH-GDGT archaeal lipid fractions. LC-MS chromatograms show distributions of and MS/MS spectra show characteristic fragment ions of protonated **(A)** and ammoniated **(B)** HPH-GDGTs from non-hydrolyzed fraction of sample at 175 m of HOT296_S2. **(C)** Enlarged MS/MS spectra of ammoniated HPH-GDGT-0 from panel **(B)**.

Supplementary Figure 8 | Fragmentation patterns of protonated non-hydrolyzed PH-GDGT archaeal lipid fractions. **(A)** LC-MS chromatograms showing distributions of and MS/MS spectra showing characteristic fragment ions of protonated PH-GDGTs from non-hydrolyzed fraction of sample collected from a sediment trap at 4000 m in station ALOHA (Boeuf et al., 2019). **(B)** Enlarged MS/MS spectra of protonated PH-GDGT-0 from panel **(A)**. **(C)** LC-MS chromatograms and MS/MS spectra of protonated PH-GDGTs from non-hydrolyzed fraction of sample at 175 m of HOT296_S2.

Supplementary Figure 9 | Fragmentation patterns of ammoniated DH-GDGT archaeal lipids. **(A)** LC-MS chromatograms showing distributions of and MS/MS spectra showing characteristic fragment ions of ammoniated DH-GDGTs from non-hydrolyzed fraction of sample at 175 m of HOT296_S2. **(B)** Enlarged MS/MS spectra of ammoniated DH-crenarchaeol from panel **(A)**.

Supplementary Figure 10 | Fragmentation patterns ammoniated non-hydrolyzed MH-GDGT archaeal lipid fractions. **(A)** LC-MS chromatograms showing distributions of and MS/MS spectra showing characteristic fragment ions of ammoniated MH-GDGTs from non-hydrolyzed fraction of sample at 175 m of HOT296_S2. **(B)** Enlarged MS/MS spectra of ammoniated MH-crenarchaeol from panel **(A)**.

Supplementary Figure 11 | Fragmentation patterns of C-GDGT and HPH-GDGT archaeal lipids from 5 m depth. LC-MS chromatograms showing distributions of and MS/MS spectra showing characteristic fragment ions of C-GDGTs **(A)** and HPH-GDGTs **(B)** on sample at 5 m of HOT296_S2.

Supplementary Figure 12 | Fragmentation patterns of HPH-GDGT archaeal lipids from 25 m depth. LC-MS chromatograms showing distributions of and MS/MS spectra showing characteristic fragment ions of HPH-GDGTs on sample at 25 m of HOT296_S2.

Supplementary Figure 13 | Fragmentation patterns of C-GDGT and HPH-GDGT archaeal lipids from 20 m depth. LC-MS chromatograms showing distributions of

and MS/MS spectra showing characteristic fragment ions of C-GDGTs **(A)** and HPH-GDGTs **(B)** on sample at 20 m of KM1709_S16.

Supplementary Figure 14 | Fragmentation patterns of C-GDGT, HPH-GDGT, and MH-GDGT archaeal lipids from 40 m depth. LC-MS chromatograms showing distributions of and MS/MS spectra showing characteristic fragment ions of C-GDGTs **(A)**, HPH-GDGTs **(B)** and MH-GDGTs **(C)** on sample at 40 m of KM1709_S16.

Supplementary Figure 15 | Concentrations of individual GDGTs and archeol in total lipids from acid-hydrolyzed fraction along depths at both sampling stations.

Supplementary Figure 16 | The cluster of Grs homologs originating from MGII/III as identified in **Figure 4**. Putative homologs from Station ALOHA are labeled in red.

Supplementary Figure 17 | Abundance of specific archaeal groups as a function of depths. The relative distribution of archaeal community was estimated by amplicon sequencing of ddPCR products using archaeal specific primers.

Supplementary Figure 18 | Concentrations of total GDGTs and relative abundances of GDGTs from Hurley et al. (2018). The sum of C-, HPH-, DH-, and MH-GDGTs, and C-, HPH-, DH-, MH-, and IP-GDGTs, as well as relative abundances of different classes of GDGTs at different depths as reported in data in Hurley et al. (2018). The DCM occurred at 50–75 m in Station 2, at 125 m in Station 7, at 60–70 m in Station 15, and at 70 m in Station 23.

Supplementary Figure 19 | Non-metric multidimensional scaling plots of archaeal GDGT relative abundances from Hurley et al. (2018). Based on Bray-Curtis distance matrices generated from the relative abundances of archaeal GDGTs, as reported in data in Hurley et al. (2018). The DCM depths for each station were described in **Supplementary Figure 18**.

Supplementary Table 1 | Sampling, hydrographic data, archaeal abundance estimates, estimated cellular ether lipid content for Thermoplasmatota, estimated and measured ether lipid concentrations.

Supplementary Table 2 | High quality reads of the archaeal community.

Supplementary Table 3 | The Grs homologs in MGII/III from the Genome Taxonomy Database and ALOHA gene catalog.

Supplementary Table 4 | The count profile of Rps3 sequences from the ALOHA gene catalog at the surface of Station ALOHA.

REFERENCES

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecol.* 75, 129–137. doi: 10.3354/ame01753

Aylward, F. O., and Santoro, A. E. (2020). Heterotrophic Thaumarchaeota with small genomes are widespread in the dark ocean. *mSystems* 5:e00415-20. doi: 10.1128/mSystems.00415-420

Basse, A., Zhu, C., Versteegh, G. J. M., Fischer, G., Hinrichs, K.-U., and Mollenhauer, G. (2014). Distribution of intact and core tetraether lipids in water column profiles of suspended particulate matter off Cape Blanc, NW Africa. *Org. Geochem.* 72, 1–13. doi: 10.1016/j.orggeochem.2014.04.007

Besseling, M. A., Hopmans, E. C., Bale, N. J., Schouten, S., Sinninghe Damsté, J. S., and Villanueva, L. (2020). The absence of intact polar lipid-derived GDGTs in marine waters dominated by marine group II: implications for lipid biosynthesis in Archaea. *Sci. Rep.* 10:294.

Bingham, F. M., and Lukas, R. (1996). Seasonal cycles of temperature, salinity and dissolved oxygen observed in the Hawaii Ocean Time-series. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* 43, 199–213. doi: 10.1016/0967-0645(95)00090-9

Boeuf, D., Edwards, B. R., Eppley, J. M., Hu, S. K., Poff, K. E., Romano, A. E., et al. (2019). Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. *Proc. Natl. Acad. Sci. USA.* 116, 11824–11832. doi: 10.1073/pnas.1903080116

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* 6, 245–252. doi: 10.1038/nrmicro1852

Brown, M. V., and Fuhrman, J. A. (2005). Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol.* 41, 15–23. doi: 10.3354/ame041015

Brown, M. V., Philip, G. K., Bunge, J. A., Smith, M. C., Bissett, A., Lauro, F. M., et al. (2009). Microbial community structure in the North Pacific ocean. *ISME J.* 3, 1374–1386.

Bryant, J. A., Aylward, F. O., Eppley, J. M., Karl, D. M., Church, M. J., and DeLong, E. F. (2016). Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. *ISME J.* 10, 1308–1322. doi: 10.1038/ismej.2015.221

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per samples. *Proc. Natl. Acad. Sci. U S A.* 108, 4516–4522. doi: 10.1073/pnas.1000080107

Church, M. J., Wai, B., Karl, D. M., and DeLong, E. F. (2010). Abundances of crenarchaeal amoA genes and transcripts in the Pacific Ocean. *Environ. Microbiol.* 12, 679–688. doi: 10.1111/j.1462-2920.2009.02108.x

Close, H. G., Shah, S. R., Ingalls, A. E., Diefendorf, A. F., Brodie, E. L., Hansman, R. L., et al. (2013). Export of submicron particulate organic matter to mesopelagic depth in an oligotrophic gyre. *Proc. Natl. Acad. Sci. U S A.* 110, 12565–12570. doi: 10.1073/pnas.1217514110

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U S A.* 89, 5685–5689. doi: 10.1073/pnas.89.12.5685

DeLong, E. F. (1998). Archaeal means and extremes. *Science* 280, 542–543. doi: 10.1126/science.280.5363.542

DeLong, E. F., King, L. L., Massana, R., Cittone, H., Murray, A., Schleper, C., et al. (1998). Dibiphytanyl ether lipids in nonthermophilic crenarchaeotes. *Appl. Environ. Microbiol.* 64, 1133–1138. doi: 10.1128/aem.64.3.1133-1138.1998

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311, 496–503. doi: 10.1126/science.1120250

DeLong, E. F., Wu, K. Y., Prezelin, B. B., and Jovine, R. V. M. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature* 371, 695–697. doi: 10.1038/371695a0

Deschamps, P., Zivanovic, Y., Moreira, D., Rodriguez-Valera, F., and López-García, P. (2014). Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic thaumarchaea and euryarchaea. *Genome Biol. Evol.* 6, 1549–1563. doi: 10.1093/gbe/evu127

Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. U S A.* 103, 9482–9487. doi: 10.1073/pnas.0601637103

Dong, L., Li, Z., and Jia, G. (2019). Archaeal ammonia oxidation plays a part in late Quaternary nitrogen cycling in the South China Sea. *Earth Plan. Sci. Lett.* 509, 38–46. doi: 10.1016/j.epsl.2018.12.023

Dore, J. E., and Karl, D. M. (1996). Nitrite distributions and dynamics at station ALOHA. *Deep Sea Res. Part II: Top. Stud. Oceanography* 43, 385–402. doi: 10.1016/0967-0645(95)00105-0

Elling, F. J., Gottschalk, J., Doeana, K. D., Kusch, S., Hurley, S. J., and Pearson, A. (2019). Archaeal lipid biomarker constraints on the Paleocene-Eocene carbon isotope excursion. *Nat. Commun.* 10:4519.

Elling, F. J., Könneke, M., Lipp, J. S., Becker, K. W., Gagen, E. J., and Hinrichs, K.-U. (2014). Effects of growth phase on the membrane lipid composition of the thaumarchaeon *Nitrosopumilus maritimus* and their implications for archaeal lipid distributions in the marine environment. *Geochim Cosmochim Acta* 141, 579–597. doi: 10.1016/j.gca.2014.07.005

Elling, F. J., Könneke, M., Mumann, M., Greve, A., and Hinrichs, K.-U. (2015). Influence of temperature, pH and. (salinity)on membrane lipid composition and TEX86 of marine planktonic thaumarchaeal isolates. *Geochim Cosmochim Acta* 171, 238–255. doi: 10.1016/j.gca.2015.09.004

Elling, F. J., Könneke, M., Nicol, G. W., Stieglmeier, M., Bayer, B., Spieck, E., et al. (2017). Chemotaxonomic characterization of the thaumarchaeal lipidome. *Environ. Microbiol.* 19, 2681–2700. doi: 10.1111/1462-2920.13759

Falkowski, P. G., Fenchel, T., and DeLong, E. F. (2008). The microbial engines that drive earth's biogeochemical cycles. *Science* 320, 1034–1039. doi: 10.1126/science.1153213

Fuhrman, J. A., and Davis, A. A. (1997). Widespread archaea and novel bacteria from the deep sea as shown by 16S rRNA gene sequences. *Mar. Ecol. Prog. Ser.* 150, 275–285. doi: 10.3354/meps150275

Fuhrman, J. A., McCallum, K., and Davis, A. A. (1992). Novel major archaebacterial group from marine plankton. *Nature* 356, 148–149. doi: 10.1038/356148a0

Galand, P. E., Gutiérrez-Provecho, C., Massana, R., Gasol, J. M., and Casamayor, E. O. (2010). Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol. Oceanogr.* 55, 2117–2125. doi: 10.4319/lo.2010.55.5.2117

Hallam, S. J., Konstantinidis, K. T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., et al. (2006a). Genomic analysis of the uncultivated marine crenarchaeote, Cenarchaeum symbiosum. *Proc. Natl. Acad. Sci U.S.A.* 103, 18296–18301. doi: 10.1073/pnas.0608549103

Hallam, S. J., Mincer, T. J., Schleper, C., Preston, C. M., Roberts, K., Richardson, P. M., et al. (2006b). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine *Crenarchaeota*. *PloS Biol.* 4:e95. doi: 10.1371/journal.pbio.0040095

Haro-Moreno, J. M., Rodríguez-Valera, F., López-Garcia, P., Moreira, D., and Martin-Cuadrado, A. B. (2017). New insights into marine group III Euryarchaeota, from dark to light. *ISME J.* 11, 1102–1117. doi: 10.1038/ismej.2016.188

Harvey, H. R., Fallon, R. D., and Patton, J. S. (1986). The effect of organic matter and oxygen on the degradation of bacterial membrane lipids in marine sediments. *Geochim. Cosmochim. Acta* 50, 795–804. doi: 10.1016/0016-7037(86)90355-8

Ho, S. L., and Laepple, T. (2016). Flat meridional temperature gradient in the early Eocene in the subsurface rather than surface ocean. *Nat. Geosci.* 9, 606–610. doi: 10.1038/ngeo2763

Horak, R. E. A., Qin, W., Bertagnolli, A. D., Nelson, A., Heal, K. R., Han, H., et al. (2018). Relative impacts of light, temperature, and reactive oxygen on thaumarchaeal ammonia oxidation in the North Pacific Ocean. *Limnol. Oceanogr.* 63, 741–757. doi: 10.1002/lno.10665

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Dufournel, I. J., Bronner, G., et al. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc. Natl. Acad. Sci. U S A.* 110, 6004–6009. doi: 10.1073/pnas.1216863110

Hurley, S. J., Elling, F. J., Könneke, M., Buchwald, C., Wankel, S. D., Santoro, A. E., et al. (2016). Influence of ammonia oxidation rate on thaumarchaeal lipid composition and the TEX86 temperature proxy. *Proc. Natl. Acad. Sci. U S A.* 113, 7762–7767. doi: 10.1073/pnas.1518534113

Hurley, S. J., Lipp, J. S., Close, H. G., Hinrichs, K.-U., and Pearson, A. (2018). Distribution and export of isoprenoid tetraether lipids in suspended particulate matter from the water column of the Western Atlantic Ocean. *Org. Geochem.* 116, 90–102. doi: 10.1016/j.orggeochem.2017.11.010

Ingalls, A. E., Huguet, C., and Truxal, L. T. (2012). Distribution of intact and core membrane lipids of archaeal glycerol dialkyl glycerol tetraethers among size-fractionated particulate organic matter in Hood Canal. *Puget Sound. Appl. Environ. Microbiol.* 78, 1480–1490. doi: 10.1128/aem.07016-11

Ingalls, A. E., Shah, S. R., Hansman, R. L., Aluwihare, L. I., Santos, G. M., Druffel, E. R. M., et al. (2006). Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proc. Natl. Acad. Sci. U S A.* 103, 6442–6447. doi: 10.1073/pnas.0510157103

Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L., and Armbrust, E. V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335, 587–590. doi: 10.1126/science.1212665

Karl, D. M., and Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean time-series programme. *Nat. Rev. Microbiol.* 12, 699–713. doi: 10.1038/nrmicro3333

Karl, D. M., and Church, M. J. (2017). Ecosystem structure and dynamics in the North Pacific Subtropical Gyre: new views of an old ocean. *Ecosystems* 20, 433–457. doi: 10.1007/s10021-017-0117-0

Karner, M. B., DeLong, E. F., and David, M. K. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409, 507–510. doi: 10.1038/35054051

Kim, J. G., Park, S. J., Sinninghe Damsté, J. S., Schouten, S., Rijpstra, W. I. C., Jung, M.-Y., et al. (2016). Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *PNAS* 113, 7888–7893. doi: 10.1073/pnas.1605501113

Kim, J.-H., Van der Meer, J., Schouten, S., Helmke, P., Willmott, V., Sangiorgi, F., et al. (2010). New indices and calibrations derived from the distribution of crenarchaeol isoprenoid tetraether lipids: implications for past sea surface temperature reconstructions. *Geochim Cosmochim Acta* 74, 4639–4654. doi: 10.1016/j.gca.2010.05.027

Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* 66, 1328–1333. doi: 10.1128/aem.66.4.1328-1333.2000

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437, 543–546. doi: 10.1038/nature03911

Kuypers, M. M., Blokker, P., Erbacher, J., Kinkel, H., Pancost, R. D., Schouten, S., et al. (2001). Massive expansion of marine archaea during a mid-Cretaceous oceanic anoxic event. *Science* 293, 92–95. doi: 10.1126/science.1058424

Letelier, R. M., Dore, J. E., Winn, C. D., and Karl, D. M. (1996). Seasonal and interannual variations in photosynthetic carbon assimilation at Station ALOHA. *Deep-Sea Res II* 43, 467–490. doi: 10.1016/0967-0645(96)00006-9

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, 242–245.

Lincoln, S. A., Wai, B., Eppley, J. M., Church, M. J., Summons, R. E., and DeLong, E. F. (2014). Planktonic *Euryarchaeota* are a significant source of archaeal tetraether lipids in the ocean. *Proc. Natl. Acad. Sci. U S A.* 111, 9858–9863. doi: 10.1073/pnas.1409439111

Lipp, J. S., Morono, Y., Inagaki, F., and Hinrichs, K.-U. (2008). Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Science* 454, 991–994. doi: 10.1038/nature07174

Lukas, R. (2001). Freshening of the upper thermocline in the North Pacific subtropical Gyre associated with decadal changes of rainfall. *Geophys. Res. Lett.* 28, 3485–3488. doi: 10.1029/2001gl013116

Ma, C., Coffinet, S., Lipp, J. S., Hinrichs, K. U., and Zhang, C. (2020). Marine Group II Euryarchaeota contribute to the archaeal lipid pool in Northwestern Pacific Ocean surface waters. *Front. Microbiol.* 11:1034. doi: 10.3389/fmicb.2020.01034

Martin-Cuadrado, A., Rodriguez-Valera, F., Moreira, D., Alba, J., Ivars-Martìnez, E., Henn, M., et al. (2008). Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J.* 2, 865–886. doi: 10.1038/ismej.2008.40

Massana, R., DeLong, E. F., and Pedrós-Alió, C. (2000). A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl. Environ. Microbiol.* 66, 1777–1787. doi: 10.1128/aem.66.5.1777-1787.2000

Massana, R., Murray, A. E., Wu, K. Y., and DeLong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl. Environ. Microbiol.* 63, 50–56. doi: 10.1128/aem.63.1.50-56.1997

Mende, D. R., Bryant, J. A., Aylward, F. O., Eppley, J. M., Nielsen, T., Karl, D. M., et al. (2017). Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* 2, 1367–1373. doi: 10.1038/s41564-017-0008-3

Merbt, S. N., Stahl, D. A., Casamayor, E. O., Martí, E., Nicol, G. W., and Prosser, J. I. (2012). Differential photoinhibition of bacterial and archaeal ammonia oxidation. *FEMS Microbiol. Lett.* 327, 41–46. doi: 10.1111/j.1574-6968.2011.02457.x

Mincer, T. J., Church, M. J., Taylor, L. T., Preston, C., Karl, D. M., and DeLong, E. F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ. Microbiol.* 9, 1162–1175. doi: 10.1111/j.1462-2920.2007.01239.x

Needham, D. M., Fichot, E. B., Wang, E., Berdjeb, L., Cram, J. A., Fichot, C. G., et al. (2018). Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. *ISME J.* 12, 2417–2432. doi: 10.1038/s41396-018-0169-y

Parada, A. E., and Fuhrman, J. A. (2017). Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. *ISME J.* 11, 2510–2525. doi: 10.1038/ismej.2017.104

Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi: 10.1111/1462-2920.13023

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Pearson, A., and Ingalls, A. E. (2013). Assessing the use of archaeal lipids as marine environmental proxies. *Ann. Rev. Earth Planet Sci.* 41, 359–384. doi: 10.1146/annurev-earth-050212-123947

Pearson, A., McNichol, A. P., Benitez-Nelson, B. C., Hayes, J. M., and Eglinton, T. I. (2001). Origins of lipid biomarkers in Santa Monica Basin surface sediment: a case study using compound-specific Δ14C analysis. *Geochim Cosmochim Acta* 65, 3123–3137. doi: 10.1016/s0016-7037(01)00657-3

Pitcher, A., Wuchter, C., Siedenberg, K., Schouten, S., and Sinninghe Damsté, J. S. (2011). Crenarchaeol tracks winter blooms of ammonia-oxidizing Thaumarchaeota in the coastal North Sea. *Limnol. Oceanogr.* 5, 2308–2318. doi: 10.4319/lo.2011.56.6.2308

Preston, C. M., Wu, K. Y., Molinski, T. F., and DeLong, E. F. (1996). A psychrophilic crenarchaeon inhabits a marine sponge; *Cenarchaeum symbiosum*, gen. nov. sp. nov. *Proc Natl Acad Sci U S A.* 93, 6241–6246. doi: 10.1073/pnas.93.13.6241

Qin, W., Carlson, L. T., Armbrust, E. V., Devol, A. H., Moffett, J. W., Stahl, D. A., et al. (2015). Confounding effects of oxygen and temperature on the TEX86 signature of marine Thaumarchaeota. *Proc. Natl. Acad. Sci. U S A.* 112, 10979–10984. doi: 10.1073/pnas.1501568112

Qin, W., Meinhardt, K. A., Moffett, J. W., Devol, A. H., Armbrust, E. V., Ingalls, A. E., et al. (2017). Influence of oxygen availability on the activities of ammonia-oxidizing archaea. *Environ. Microbiol. Rep.* 9, 250–256. doi: 10.1111/1758-2229.12525

Raskin, L., Stromley, J. M., Rittmann, B. E., and Stahl, D. A. (1994). Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl. Environ. Microbiol.* 60, 1232–1240. doi: 10.1128/aem.60.4.1232-1240.1994

Reji, L., and Francis, C. A. (2020). Metagenome-assembled genomes reveal unique metabolic adaptations of a basal marine Thaumarchaeota lineage. *ISME J.* 14, 2105–2115. doi: 10.1038/s41396-020-0675-6

Richardson, T. L., and Jackson, G. A. (2007). Small phytoplankton and carbon export from the surface ocean. *Science* 315, 838–840. doi: 10.1126/science.1133471

Rinke, C., Rubino, F., Messer, L. F., Youssef, N., Parks, D. H., Chuvochina, M., et al. (2019). A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME J.* 13, 663–675. doi: 10.1038/s41396-018-0282-y

Schefuß, E., Kuhlmann, H., Mollenhauer, G., Prange, M., and Pätzold, J. (2011). Forcing of wet phases in southeast Africa over the past 17,000 years. *Nature* 480, 509–512. doi: 10.1038/nature10685

Schouten, S., Hopmans, E. C., Baas, M., Boumann, H., Standfest, S., Könneke, M., et al. (2008). Intact membrane lipids of "Candidatus *Nitrosopumilus maritimus*," a cultivated representative of the cosmopolitan mesophilic group I Crenarchaeota. *Appl. Environ. Microbiol.* 74, 2433–2440. doi: 10.1128/aem.01709-07

Schouten, S., Hopmans, E. C., Schefuß, E., and Sinninghe Damsté, J. S. (2002). Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet Sci. Lett.* 204, 265–274. doi: 10.1016/s0012-821x(02)00979-2

Schouten, S., Hopmans, E. C., and Sinninghe Damsté, J. S. (2013). The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: a review. *Organ. Geochem.* 54, 19–61. doi: 10.1016/j.orggeochem.2012.09.006

Schubotz, F., Wakeham, S. G., Lipp, J. S., Fredricks, H. F., and Hinrichs, K.-U. (2009). Detection of microbial biomass by intact polar membrane lipid analysis in the water column and surface sediments of the Black Sea. *Environ. Microbiol.* 11, 2720–2734. doi: 10.1111/j.1462-2920.2009.01999.x

Sinninghe Damsté, J. S. S., Rijpstra, W. I. C, Hopmans, E. C., den Uijl, M. J., Weijers, J. W. H., and Schouten, S. (2018). The enigmatic structure of the crenarchaeol isomer. *Organ Geochem.* 124, 22–28. doi: 10.1016/j.orggeochem.2018.06.005

Sinninghe Damsté, J. S., Schouten, S., Hopmans, E. C., van Duin, A. C. T., and Geenevasen, J. A. J. (2002). Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. *J. Lipid Res.* 43, 1641–1651.

Smith, K. L. Jr., Williams, P. M., and Druffel, E. R. M. (1989). Upward fluxes of particulate organic matter in the deep North Pacific. *Nature* 337, 724–726. doi: 10.1038/337724a0

Sollai, M., Villanueva, L., Hopmans, E. C., Reichart, G.-J., and Sinninghe Damsté, J. S. (2018). A combined lipidomic and 16S rRNA gene amplicon sequencing approach reveals archaeal sources of intact polar lipids in the stratified Black Sea water column. *Geobiology* 17, 91–109. doi: 10.1111/gbi.12316

Sommer, S. A., Van Woudenberg, L., Lenz, P. H., Cepeda, G., and Goetze, E. (2017). Vertical gradients in species richness and community composition across the twilight zone in the North Pacific Subtropical Gyre. *Mol. Ecol.* 26, 6136–6156. doi: 10.1111/mec.14286

Sørensen, K. B., and Teske, A. (2006). Stratified communities of active archaea in deep marine subsurface sediments. *Appl. Environ. Microbiol.* 72, 4596–4603. doi: 10.1128/aem.00562-06

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., et al. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum *Thaumarchaeota*. *Trends Microbiol.* 18, 331–340. doi: 10.1016/j.tim.2010.06.003

Sturt, H. F., Summons, R. E., Smith, K., Elvert, M., and Hinrichs, K.-U. (2004). Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry–new biomarkers for biogeochemistry and microbial ecology. *Rapid Commun. Mass Spectrom.* 18, 617–628. doi: 10.1002/rcm.1378

Teske, A., and Sørensen, K. B. (2008). Uncultured archaea in deep marine subsurface sediments: have we caught them all? *ISME J.* 2, 3–18. doi: 10.1038/ismej.2007.90

Tierney, J. E., Mayes, M. T., Meyer, N., Johnson, C., Swarzenski, P. W., Cohen, A. S., et al. (2010). Late-twentieth-century warming in Lake Tanganyika unprecedented since AD 500. *Nat. Geosci.* 3, 422–425. doi: 10.1038/ngeo865

Trull, T. W., Bray, S. G., Buesseler, K. O., Lamborg, C. H., Manganini, S., Moy, C., et al. (2008). In situ measurement of mesopelagic particle sinking rates and the control of carbon transfer to the ocean interior during the Vertical Flux in the Global Ocean (VERTIGO) voyages in the North Pacific. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* 55, 1684–1695. doi: 10.1016/j.dsr2.2008.04.021

Villanueva, L., Schouten, S., and Sinninghe Damsté, J. S. (2017). Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide'. *Environ. Microbiol.* 19, 54–69. doi: 10.1111/1462-2920.13361

White, D., Davis, W., Nickels, J., King, J., and Bobbie, R. (1979). Determination of the sedimentary microbial biomass by extractible lipid phosphate. *Oecologia* 40, 51–62. doi: 10.1007/BF00388810

Yayanos, A. A., and Nevenzel, J. C. (1978). Rising-particle hypothesis: rapid ascent of matter from the deep ocean. *Naturwissenschaften* 65, 255–256. doi: 10.1007/bf00368567

Zachos, J. C., Schouten, S., Bohaty, S., Quattlebaum, T., Sluijs, A., Brinkhuis, H., et al. (2006). Extreme warming of mid-latitude coastal ocean during the Paleocene-Eocene thermal maximum: inferences from TEX86 and isotope data. *Geology* 34:737. doi: 10.1130/g22522.1

Zeng, Z., Liu, X. L., Farley, K. R., Wei, J. H., Metcalf, W. W., Summons, R. E., et al. (2019). GDGT cyclization proteins identify the dominant archaeal sources of tetraether lipids in the ocean. *Proc. Natl. Acad. Sci. U S A.* 116, 22505–22511. doi: 10.1073/pnas.1909306116

Zhang, Y., and Liu, X. (2018). Export depth of the TEX86 signal. *Paleoceanography Paleoclimatol.* 33, 666–671. doi: 10.1029/2018pa0 03337

Zhu, C., Lipp, J. S., Wormer, L., Becker, K. W., Schroder, J., and Hinrichs, K. (2013). Comprehensive glycerol ether lipid fingerprints through a novel revered phase liquid chromatography-mass spectrometry protocol. *Organ. Geochem.* 65, 53–62. doi: 10.1016/j.orggeochem.2013. 09.012