Sources and genetic controls of lipid biomarkers involved in paleotemperature proxies
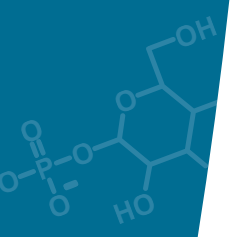
Marc Arnold Besseling

Sources and genetic controls of lipid biomarkers involved in paleotemperature proxies

Marc Arnold Besseling

Sources and genetic controls of lipid biomarkers involved in paleotemperature proxies

Bronnen en genetische controles op lipide biomarkers die toegepast worden in paleotemperatuur proxy's

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de

Universiteit Utrecht

op gezag van de

rector magnificus, prof.dr. H.R.B.M. Kummeling,

ingevolge het besluit van het college voor promoties

in het openbaar te verdedigen op

vrijdag 1 november 2019 des middags te 2.30 uur

door

Marc Arnold Besseling

geboren op 5 oktober 1987

te Utrecht

PROMOTOREN

Prof. dr. J. S. Sinninghe Damsté

Prof. dr. S. Schouten

COPROMOTOR

Dr. L. Villanueva

CONTENTS

CHAPTER | 1

Introduction

Climate reconstruction

The Earth is warming due the anthropogenic (human induced) input of greenhouse gasses, such as carbon dioxide and methane, into the atmosphere (Arrhenius, 1896; Callendar, 1938; Intergovernmental Panel on Climate Change (IPCC), 2007). Our current atmospheric CO_2 levels surpassed the 400 parts per million (ppm; measured at Mauna Loa Observatory on Hawaii) threshold in 2013. These are the highest levels since 800.000 years (Hönisch et al., 2009) and perhaps even for the last 10 million years (Tripati et al., 2009). Even if greenhouse gas emissions would suddenly halt today the Earth will keep warming due to the elevated CO_2 levels (Gillett et al., 2011; Matthews and Zickfeld, 2012). The effects and the severity of this warming on Earth's climate system and our everyday life's is uncertain (IPCC, 2007). To better predict and understand our future in a warmer and rapid changing environment we can look at past climate fluctuations and "sudden" (on a geological timescale) warming events.

Changes in past climate systems and events are recorded in climate archives. The most commonly known archives are tree rings, glacial ice, corals and lacustrine and marine sediments. Some of these archives contain climate data spanning the last thousands of years (tree rings) or up to a million years with the use of ice cores (800.000 years, Augustin et al., 2004). However, one of the longest and continues climate records can be obtained from marine sediments, capturing Earth's climate's history up to tens of million years (e.g. Zachos et al., 2008). The marine sediment climate archive is composed of erosion material from terrestrial origin but also of (in)organic compounds that are transported from the water column towards the seafloor. This flux of "marine snow" forms continuous layers on the seafloor and contains fossil remains of marine organisms such as the calcareous ($CaCO_3$) shells (also called tests) of Foraminifera (unicellular eukaryotes) and coccolithophores (group of algae) and the siliceous (SiO_2) capsules of diatoms (group of algae) and Radiolaria (unicellular eukaryotes).

Besides these mineral-based fossils, organic material is also deposited, mostly originating from biomass produced in the water column. The bulk of this material

is composed out of carbohydrates, proteins and lipids, which substantially degrade when transported to the seafloor (Wakeham et al., 1997) and are subsequently incorporated in the sedimentary archive (see for an overview, Arndt et al., 2013). Lipids have relatively slow degradation rates compared to other organic molecules and these molecular fossils can even be found in sediments from Archean times (older than 2500 million years; Brocks et al., 1999). These fossils can be used to reconstruct past microbial communities as well as climate variables such as Sea Surface Temperature (SST). The latter is important due to the fact that oceans have the largest surface area on Earth and can store large amount of energy, due to its size and high heat capacity (Cess and Goldenberg, 1981).

Lipid-based organic proxies for sea surface temperatures reconstruction

Several proxies exists to reconstruct SST such as those based on assemblages of fossilized (remains of) marine organisms like coccoliths, radiolarians, diatoms, Foraminerifa and dinoflagellates (Moore et al., 1980; Vernal et al., 2005). Changes in SST can also impact the chemical composition of carbonate shells such as Mg/Ca ratios (Chave, 1954; Nürnberg et al., 1996) and the isotopic fractionation of oxygen isotopes (δ^{18}O; Urey, 1947; Emiliani, 1955). The use of lipids to reconstruct past temperatures has substantially increased in the recent decades, largely due to analytical developments improving the sensitivity and specificity of molecular detection. These lipids are discussed in more detail below.

a

$C_{37:2}$ and $C_{37:3}$ alkenone

b

GDGT-0

crenarchaeol

HPH-crenarchaeol

c

C_{30} 1,15, C_{30} 1,13 and C_{28} 1,13 long-chain diols
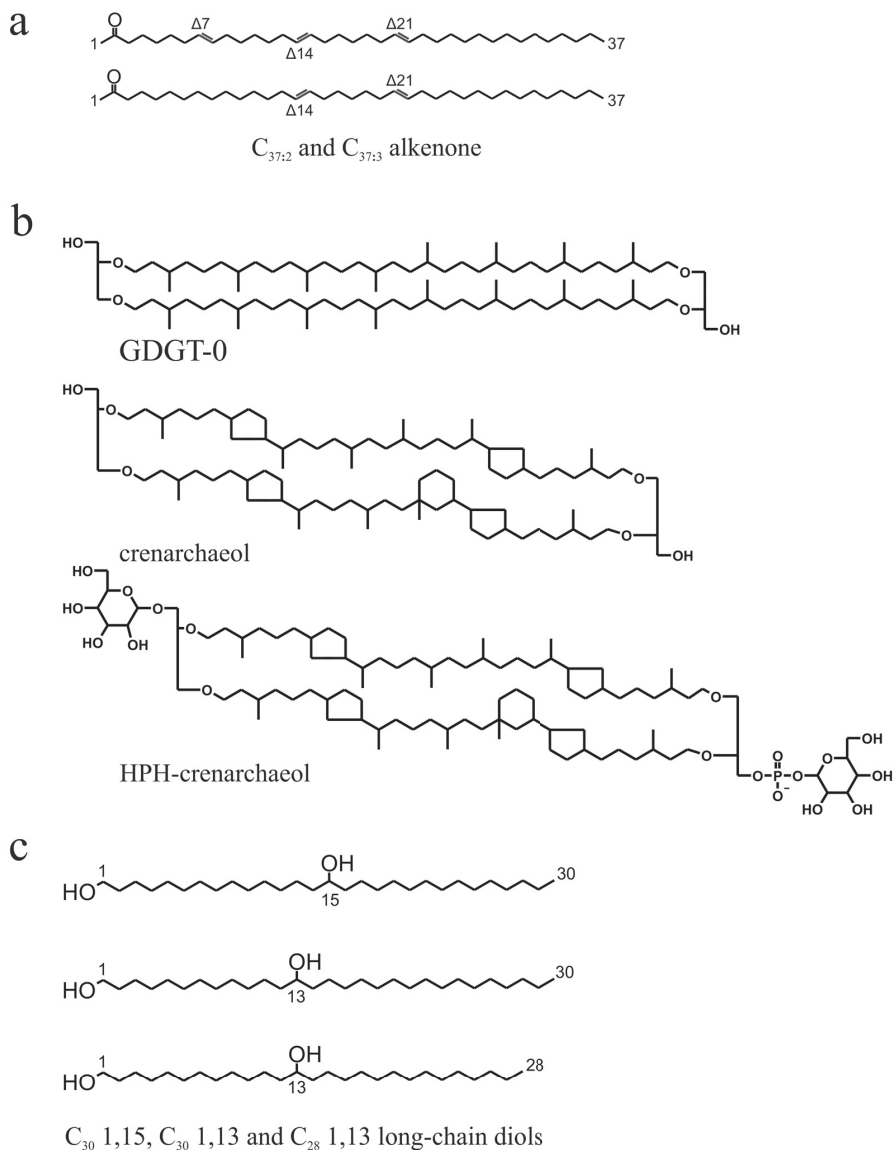
Figure 1. Examples of lipids that are used with organic paleotemperature proxies or as biomarkers for specific organisms. a) Alkenones ($C_{37:2}$ and $C_{37:3}$) used in the unsaturated ketone index ($U_{37}^{K'}$) b) GDGT-0 and crenarchaeol, examples of membrane lipids produced by Archaea, and HPH-crenarchaeol, an example of an intact polar lipid (IPL-) GDGT, used as biomarker for living or intact Thaumarchaeota c) long-chain diols (C_{30} 1,15, C_{30} 1,13 and C_{28} 1,13) used in the long chain diol index (LDI).

U^k_{37}

The first organic paleotemperature proxy developed for reconstructing SSTs was based on the degree of unsaturation of long-chain ketones with a length of 37 carbon atoms (called alkenones; Figure 1). These alkenones are the basis of the Unsaturated Ketone index (U^K_{37}), equation 1, developed by Brassell et al. (1986), later simplified with the $U^{K'}_{37}$ index (equation 2) by Prahl and Wakeham, (1987).

$$U^K_{37} = \frac{[C_{37:2}]-[C_{37:4}]}{[C_{37:2}]+[C_{37:3}]+[C_{37:4}]} \tag{1}$$

$$U^{K'}_{37} = \frac{[C_{37:2}]}{[C_{37:2}]+[C_{37:3}]} \tag{2}$$

These alkenones are produced by a number of haptophyte algae (Marlowe et al., 1984a; Volkman et al., 1980), within the Isochrysidales order (Bendif et al., 2013). *Emiliania huxleyi* is the most common coccolithophore in the open oceans within the Isochrysidales, and are believed to be the most dominant source of marine alkenones, together with *Gephyrocapsa oceanica* (Conte et al., 1995; 2006; Volkman et al., 1995). They are also produced by other marine haptophytes such as *Isochrysis galbana* (Sukenik and Wahnon, 1990) and *Chrysotila lamellosa* (Marlowe et al., 1990).

It was initially believed that alkenones were membrane lipids and that the degree of unsaturation would influence the membrane fluidity (e.g. Prahl et al., 1988). However, alkenones were not found within or bound to membranes within the cell (Sawada and Shiraiwa, 2004). Microscopic analyses and isolation of cellular compartments showed that alkenones accumulate in specific lipid bodies, with such high concentrations that they are called alkenone bodies (Eltgroth et al., 2005). Within these alkenone bodies the amount of alkenones is decreasing when the cells are transferred to the dark, strongly suggesting that alkenones function as a carbon/energy storage to maintain metabolism (Epstein et al., 2001; Pan & Sun, 2011). A remarkable feature of alkenones is that the double bonds occur with the unusual

trans-confirmation (Rechka and Maxwell, 1988), which does not alter the spatial dimension of alkenones in contrary to cis-conformations in membrane fatty acids and are therefore perhaps more suited for storage. Alkenones have double bonds that are spaced with a regular pattern across the carbon chain, i.e. separated by 7 methylene units (e.g. $\Delta^{14,21}$, $\Delta^{7,14,21}$ and $\Delta^{7,14,21,28}$, numbered from the carbonyl position, for alkenones with 2-4 double bonds,; López & Grimalt, 2006). However, there are exceptions, e.g. alkenones from E. huxleyi CCMP2758 contain double bonds at the positions $\Delta^{12,19}$ and $\Delta^{7,12,19}$ (Prahl et al., 2006), and alkenones detected in Black Sea sediments have double bonds at position $\Delta^{14,19}$ (Xu et al., 2001).

The different sources of alkenones in the marine environment could potentially negatively affect the U^K_{37} proxy, because it is known that different algae and their genotypes produce alkenones in a different composition with the same growth temperature (Conte et al., 1998). Other physiological factors than temperature seem to affect the unsaturation ratios to some degree such as salinity (Ono et al., 2012) and growth phase of the algae (Yamamoto et al., 2000). Another disadvantage of the alkenone based proxy is that E. huxleyi first appeared around 250-270 kyrs ago (Thierstein et al., 1977) while long chain alkenones with the length of 37 carbon atoms have been detected in sediments dating back to the Eocene (dating back 56 to 39 million years ago; Marlowe et al., 1984b; van der Smissen and Rullkötter, 1996; Pagani et al., 2011). SSTs calculations from older sediments are therefore based on the assumption that alkenones produced by the biological precursor of E. huxleyi have a similar correlation with temperature.

TEX$_{86}$

Another organic proxy for reconstructing SST is based on archaeal membrane lipids called isoprenoid Glycerol Dialkyl Glycerol Tetraethers (GDGTs; Figure 1). GDGTs are synthesized by the archaeal phylum Thaumarchaeota, as well as other phyla such as Euryarchaeota and Crenarchaeota (see for a review: Schouten et al., 2013; Figure 2). Thaumarchaeota are detected in freshwater environments (Auguet and Casamayor, 2013), in soils (Leininger et al., 2006), on the human skin (Caporaso et al., 2011) and are also one of the most dominant components in the

marine picoplankton community (up to 20 %, Karner et al., 2001; Schattenhofer et al., 2009). Thaumarchaeota are able to synthesize GDGTs with 0-4 cyclopentane moieties, archaeol and the GDGT crenarchaeol, which contains a cyclohexane moiety in addition to 4 cyclopentane moieties (Schouten et al., 2008; Sinninghe Damsté et al., 2002; Figure 1). Crenarchaeol is considered to be solely produced by Thaumarchaeota (e.g. Pitcher et al., 2011) unlike other GDGTs which can be formed by other archaeal phyla (Figure 2). Within living and intact archaeal cells are GDGTs predominantly bound to polar head groups, called intact polar lipid (IPL-) GDGTs (e.g. Langworthy et al., 1974; Sugai et al., 1995; Sinninghe Damsté et al., 2012; Figure 1).
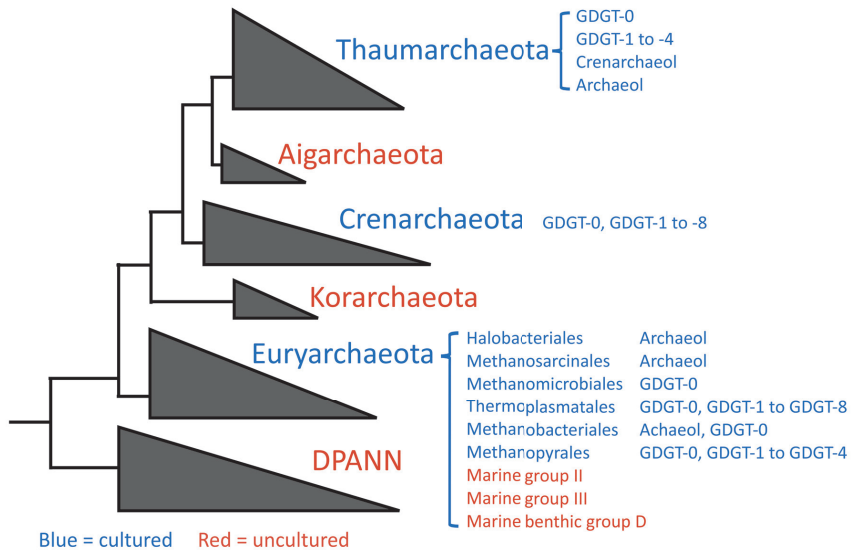
Figure 2. Schematic archaeal phylogenetic tree based on the 16S rRNA gene of Archaea (Modified from Rinke et al. (2013)). The group names colored in blue contain at least one cultured representative from which the lipid membrane composition is known. The group names colored in red have no cultured representative and their lipid membrane composition is unknown.

Common head groups are monohexose (MH), dihexose (DH) and hexose phosphohexose (HPH) (Schleper et al., 1995; Schouten et al., 2008; Shimada et al., 2008; Pitcher et al., 2010; Figure 1). These IPL-GDGTs can be used as biomarkers for living Archaea, e.g. HPH-crenarchaeol has been found to correlate

with Thaumarchaeota within the marine environment (Pitcher et al., 2011b). The composition of GDGTs in the marine environment has been shown to be affected by temperature, i.e. in warm environments there is a high relative abundance of GDGTs with cyclopentane moieties (Figure 3).
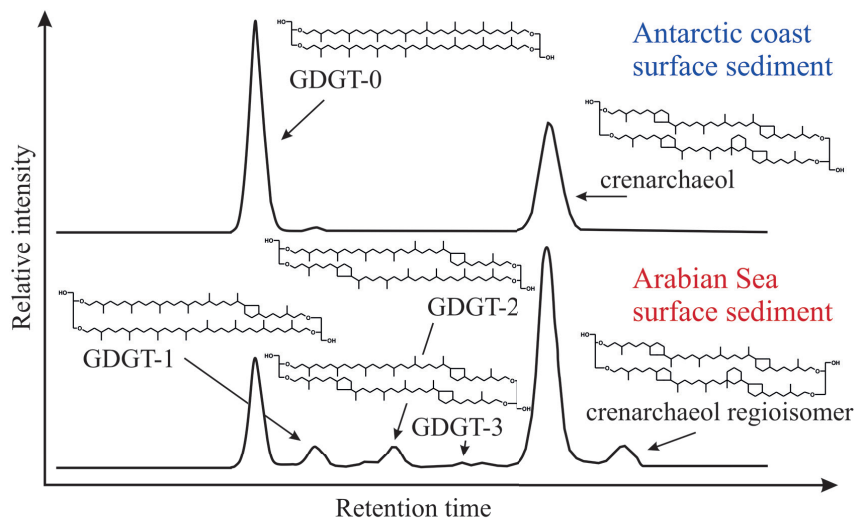
Figure 3. Base peak chromatograms showing the isoprenoid glycerol dialkyl glycerol tetraether (GDGT) distribution in a surface sediment from the coast of Antarctica (top part) and from the Arabian Sea (bottom part; modified from Schouten et al. (2002). The chromatograms illustrate the difference in GDGT composition between locations with substantial different annual mean temperatures.

Based on this phenomenon the TEX_{86} (Tetra Ether indeX of tetraethers consisting of 86 carbon atoms) paleotemperature proxy developed (Schouten et al., 2002; equation 3).

$$\mathrm{TEX}_{86} = \frac{[\mathrm{GDGT-2}]+[\mathrm{GDGT-3}]+[\mathrm{cren'}]}{[\mathrm{GDGT-1}]+[\mathrm{GDGT-2}]+[\mathrm{GDGT-3}]+[\mathrm{cren'}]} \qquad (3)$$

GDGT-0 and crenarchaeol, the two most abundant GDGTs, are left out of the equation because their large abundance would overpower the index. Another

reason why GDGT-0 is exclude is that methanogenic Euryarchaeota, living in anoxic sedimentary environments, are believed to synthesize this GDGT and this would potentially bias the reconstructed temperature (Schouten et al., 2002). The TEX$_{86}$ paleotemperature proxy was further developed and calibrated, with the use of hundreds of marine surface sediment samples (Kim et al., 2008; Kim et al., 2010), showing a strong relationship between TEX$_{86}$ and annual mean SST (between 5°C and 30°C, $r^2 = 0.94$; Kim et al., 2008). Kim et al. (2010) defined the TEX$_{86}$ proxy in the TEX$_{86}^{L}$(equation 4), for low temperature regions (below 15°C) and the TEX$_{86}^{H}$ (equation 5) for high temperature regions (above 15°C).

$$\text{TEX}_{86}^{L} = \log \frac{[\text{GDGT}-2]}{[\text{GDGT}-1]+[\text{GDGT}-2]+[\text{GDGT}-3]} \tag{4}$$

$$\text{TEX}_{86}^{H} = \log \frac{[\text{GDGT}-2]+[\text{GDGT}-3]+[\text{cren}']}{[\text{GDGT}-1]+[\text{GDGT}-2]+[\text{GDGT}-3]+[\text{cren}']} \tag{5}$$

Although TEX$_{86}$ correlates with SST, it has been suggested that the TEX$_{86}$ proxies are recording subsurface ($30-200$ meters below sea level; mbsl) instead of surface sea temperatures (Lopes dos Santos et al., 2010; Kim et al., 2012b). Furthermore, correlations of TEX$_{86}$ with temperature in the polar oceans have shown to be poor (Kim et al., 2008; 2010) while in the Red Sea a calibration between SST and TEX$_{86}$ was found different from that of the global calibration line (Kim et al., 2008; Trommer et al., 2009; Kim et al., 2010). It has been suggested that the TEX$_{86}$ signals from the Red Sea are likely affected by a different Thaumarchaeota community composition compared to the open ocean (Trommer et al., 2009). Also the input of GDGTs by other pelagic archaeal groups may be affecting theTEX$_{86}$, e.g. methanotrophic archaea are not only producing GDGT-0 but also GDGT-1 and -2 (Pancost et al., 2001; Wakeham et al., 2004; Rossel et al., 2008; Liu et al., 2011) which could potentially disturb the TEX$_{86}$ signal. Thaumarchaeota abundance also vary during the year at some locations (Murray et al., 1998; Wuchter, 2006; Galand

et al., 2010; Pitcher et al., 2011c), and thus seasonality can affect the ability of the TEX_{86} proxy to reconstruct annual SSTs.

LDI

The most recent developed organic proxy for reconstructing SSTs is based on Long-Chain alkyl Diols (LCDs), which are lipids that contain an alkyl chain with alcohol groups at carbon positon 1 and at a mid-chain carbon position (Figure 1). Rampen et al. (2012) found, with an extensive core top study, that some LCDs have a strong correlation with annual mean SST, leading to the Long chain Diol Index (LDI; equation 6)

$$LDI = \frac{[C_{30}\ 1,15-diol]}{[C_{28}\ 1,13-diol]+[C_{30}\ 1,13-diol]+[C_{30}\ 1,15-diol]} \qquad (6)$$

The LDI showed the strongest correlations ($r^2 > 0.95$) with temperatures in the upper 30 meters of the water column, which suggests that the LCDs are formed in the top layer of the water column and probably by photoautotrophic algae (Rampen et al., 2012). LCDs have been found in multiple marine algae cultures belonging to the Eustigmatophyceae phylum, such as *Nannochloropsis gaditana* (Méjanelle et al., 2003), *N. oculata* (Volkman et al., 1992; Shimokawara et al., 2010) and *N. salina* (Volkman et al., 1992; Gelin et al., 1996; 1997a). However, there are also LCD producers outside the Eustigmatophyceae phylum such as the marine diatoms *Proboscia indica*, *P. alata* (Sinninghe Damsté et al., 2003) and *P. inermis* (Rampen et al., 2007). Because the detected LCD composition within the marine sediments do not match the LCD distributions in *Nannochloropsis* cultures it remains uncertain what the biological source of LCDs is in the marine environment (Volkman et al., 1992; Versteegh et al., 1997). The biological role of LCDs organism is also unknown: Gelin et al. (1997b) postulated that LCDs are building blocks of algaenans, larger insoluble biopolymers (Tegelaar et al., 1989). Algaenans within *N. granulata* and *N. oculata* are most likely forming a trilaminar outer cell wall around the inner polysaccharidic wall of the algae (Gelin et al., 1999). It is speculated that this outer cell wall could improve

resistance against environmental stress (Allard and Templier, 2000; Versteegh and Blokker, 2004; Kodner et al., 2009).

Producers of the biomarker lipids and biosynthetic pathways

Although the above discussed organic proxies have now been frequently applied to reconstruct past SST, for all of the above discussed lipids there is some uncertainty who the producer(s) is/are in the marine environment. Cultivation of algae and Archaea has partly revealed some of the lipid sources but since most of the marine microorganisms are not yet cultured (cf. Amann et al., 1995; Curtis et al., 2002; Baker and Dick, 2013), the main producers are often uncertain or unknown. To overcome this, the overall community composition of Archaea and/or algae in environmental samples can be studied using genomic techniques and linked to the lipid composition in the same environmental samples. Targeting the Archaea and/or algae community can be done by examining the 16S (Bacteria and Archaea) and 18S (Eukaryotes; including algae) rRNA gene, which are genes that are coding for components of the ribosomal small subunits. These genes are widely used in phylogenetic studies because they are highly conserved and differences in the gene sequence can be used to identify, and group, species (e.g. Woese and Fox, 1977). This has been done for e.g. archaeal community composition to the GDGT composition in the Black Sea (Stadnitskaia et al., 2005; Schubert et al., 2006; Coolen et al., 2007), an Antarctic Lake (Coolen et al., 2004), the Arabian Sea (Schouten et al., 2012) and the North Sea (Herfort et al., 2007). This is also the case with the identification of the haptophyte community composition, with the use of the 18S rRNA gene phylogeny, and comparing this with the LCA composition in numerous lakes (Coolen et al., 2004; D'Andrea and Huang, 2005; Theroux et al., 2010), the Black Sea (Coolen et al., 2006) and the Mediterranean Sea (Boere et al., 2011a; 2011b). However, investigating the community composition with ribosomal genes only highlights potential producers. A more conclusive way is targeting genes that directly links the producing organism with the biomarker lipid. This can be done by using the genes that are involved in the lipid biosynthetic pathway(s). An example is the lipid pathway that forms hopanoids from squalene (Woodward and Bloch, 1953),

with the enzyme squalene-hopene cyclase (*shc*) catalyzing the initial step (Ochs et al., 1990; 1992). The phylogeny of the *shc* enzyme or gene can be used to unravel the diversity of hopanoid producers in the environment, as done for various marine and freshwater systems (Pearson et al., 2007; 2009). However the *shc* enzyme can use multiple substrates and this results in a wide variety of squalene cyclization products (Siedenburg and Jendrossek, 2011; and references therein). More recently, other genes were identified that are involved in the formation of hopanoids, e.g. the *hpnP* gene that is needed methylate hopanoids at the C-2 position (Welander et al., 2010) or the *hpnR* gene to methylate at the C-3 position (Welander and Summons, 2012). These more specific biosynthetic genes can be targeted in the environment to find producers of specific methylated hopanoids.

With respect to the three organic temperature proxies, probably the compounds which have the best constrained biosynthetic pathway are GDGTs. GDGTs are characterized by an ether linkage between the glycerol-1-phosphate(G1P) moiety and the isoprene-based alkyl chain, in contrary to membrane lipids of Bacteria and Eukarya which contain an ester linkage between the glycerol-3-phosphate (G3P) and acetate-based straight or branched alkyl chain (instead of isoprenoid chains). Furthermore genes involved in the formation of the G1P backbone, the acyltransferases mediating the ether bonds and the reductases saturating the isoprenoid side chain are well constrained (see Koga and Morii (2007) for an overview). However, the gene(s) involved in the formation of the tetraether lipid GDGT, supposedly from archaeol, are still unknown.

The biosynthetic pathway leading to LCAs is unknown although a hypothetical pathway for LCAs has been proposed (Rontani et al., 2006). This pathway starts with acetyl-CoA or propionyl-CoA as a primer and involves successive chain elongations with malonyl-CoA, followed by an unknown thioesterase and decarboxylase. Rontani et al. (2006) also proposed that the formation of the atypical *trans* double bonds in the LCAs implicates a series of Δ-desaturases that have active sites at a fixed distance from a carbonyl binding site (i.e. Δ^7 and Δ^{14}). The reason why the degree of double bonds in LCAs changes with growth temperature remains a mystery, especially

1

since the lipids are stored within the cell in specific lipid bodies unlike membrane lipids such as GDGTs which are in contact with the extracellular environment.

Finally, the biosynthetic pathway of LCDs is fully unknown, i.e. the precursors of LCDs remain uncertain and also the enzymes involved in the synthesize of these specific molecules.

Scope and framework of this thesis

Although frequently applied, organic temperature proxies still have some uncertainty, in particular as for none of these organic paleotemperature proxies it is exactly known who the producers are or how the lipids are synthesized. Therefore, in this thesis it was investigated what the sources of lipids were by combining advanced mass spectrometry techniques with high throughput sequencing on samples from different marine and lake environments. By comparison of distributions it was attempted to uncover the sources of lipids used in paleotemperature proxies and implications for the temperature proxies were discussed. Furthermore, the genes responsible for the synthesis of alkenones were investigated.

In **Chapter two** the benthic archaeal community composition was compared to the IPL-GDGT composition of sediments located within, just below, and well below the Oxygen Minimum Zone (OMZ) of the Arabian Sea. This was done in both surface (0–0.5 cm) and subsurface (10–12 cm) sediments. The archaeal community of oxygenated sediments located below the OMZ was dominated by members of Thaumarchaeota (Marine Group I, MGI). The MGI were also living and active, based on the abundance of ammonia monoxygenase (*amo*A) transcripts and the presence of hexose phosphohexose crenarchaeol, an IPL-GDGT used a specific biomarker for MGI. The most dominant archaeal group within the anoxic sediments was the Miscellaneous Crenarchaeota Group (MCG). The MCG was diverse and contained multiple subgroups, subgroup MCG-12 coincided with high relative abundances of GDGT-0 with an unknown polar head group. This suggests that these MCG groups could be a potential source of this archaeal lipid.

In **Chapter three** the pelagic archaeal community composition was investigated in the west, middle and east part of the Mediterranean Sea and compared to the IPL-GDGT and the Core Lipid (CL-) GDGT composition. These different regions are harboring different water masses with different physical conditions. Upper epipelagic waters (0-100 meters deep) were dominated by Marine Euryarchaeota Group II and III (MGII and MGIII), while members of MG-I were dominating the subsurface and the deeper waters. This shift in archaeal community composition coincided with a decrease in IPL GDGT-0 and increase of IPL crenarchaeol. The CL GDGT-2/-3 ratio increased with water depth, however with lower values compared to deep marine waters of some other regions. The increase of the GDGT-2/GDGT-3 ratio coincided with the high relative abundance of deep-water MGI, which may be linked to the high temperature and salinity found in specific water masses of the Mediterranean Sea. These particularities of the Mediterranean Sea could be responsible for the overestimation of SST based on TEX$_{86}$.

In **Chapter four** the pelagic archaeal community composition and the IPL-GDGT composition is described of (sub)surface of the North Atlantic Ocean and of summer coastal North Sea waters. Within the upper surface waters of the (open ocean) North Atlantic Ocean only members of the Euryarchaeota MGII and MGIII were detected. while in surface waters from the coastal North Sea only members of MGII were detected. No IPL-GDGTs were detected in these waters. In contrast, coastal and subsurface waters of the North Atlantic Ocean had a relative high abundance of MGI Archaea, coinciding with the detection of varies IPL-GDGTs, predominantly IPL-crenarchaeol. Analyses with acid hydrolyses, which would remove any polar head groups from the IPL-GDGT, did not reveal any missing IPL-GDGTs in the coastal North Sea samples, populated by abundant MGII Euryarchaeota. This suggests that MGII Euryarchaeota are likely not capable of synthesizing GDGTs or other known archaeal lipids, as previously proposed by others.

In **Chapter five**, the pelagic archaeal community composition of the eastern tropical South Pacific (ETSP) and the Arabian Sea, both containing an oxygen minimum zone (OMZ) were compared to community composition of the fully oxygenated water

column of the North Atlantic Ocean. Thaumarchaeota were the most dominant Archaea, coinciding with the presence of IPL-GDGTs including crenarchaeol. MGIII Archaea were more relative abundant in the OMZs of the ETSP and the Arabian Sea. Archaeol-based IPLs, previously linked to MGIII Archaea residing in the OMZ of the ETSP, were not detected within the OMZ of the Arabian Sea. These results suggest that archaeol-based IPLs are produced by specific MGIII species or synthesized under different physicochemical factors within OMZs.

In **Chapter six**, the Eustigmatophyceae diversity and abundance was compared to the Long Chain Diol (LCD) composition and abundance in the water column of lake Challa, located at the border of Tanzania and Kenya. At 9 meters depth the highest abundance of Eustigmatophyceae and the maximum of LCDs were found. Furthermore, two new (sub)groups of the Eustigmatophyceae were detected which may be a source for LCDs. Analysis of sedimenting particles revealed seasonal changes in LCD abundance. This suggests that there are several blooms of LCD producers over the annual cycle or a change in the LCD composition of the same producer.

In **Chapter seven**, two cold shock experiments with the algae *Emiliania huxleyi* were performed. Within the longest experiment (covering 24 hours), a rapid increase in the $C_{37:3}/C_{37:2}$ LCA ratio occurred within 5 hours after the transfer to colder temperatures. To capture any changes in gene differentiation induced by the cold shock, a second, shorter, cold shock experiment was conducted. A rapid increase in the $C_{37:3}/C_{37:2}$ ratio was again observed and analyses of the transcriptomics data shed light on potential genes involved in the LCA biosynthetic pathway. Most interesting was the swift up-regulation of EMIHUDRAFT_454486, a gene coding for the likely homologue of the putative LCA desaturase Akd1. This suggests that it may be the desaturase in *E. huxleyi* which is responsible for transforming $C_{37:2}$ into $C_{37:3}$ LCAs.

The results described in this thesis show that the archaeal community in the marine water column, and underlying sediments, is highly diverse. As a result, linking this with the ether lipid composition in the environment is challenging as well

as due to complex interacting environmental factors. Nevertheless, some results could be obtained. For example, by studying the (sub)surface of the North Atlantic Ocean and summer coastal North Sea water, MGII Archaea could be excluded as potential producers of GDGTs. However, the question remains what kind of lipids they produce, which will only be answered after successful isolating a MGII member. Another result showing the usefulness of the approach is from Lake Challa where it was shown that uncultivated Eustigmatophyceae may produce LCDs in a lacustrine environment. Finally, culture experiments such as the cold shock experiments on *E. huxleyi* may be useful to identify genes in biomarker lipid pathways which can subsequently be targeted in e.g. metagenomics datasets.

CHAPTER 2

Benthic Archaea as potential sources of tetraether membrane lipids in sediments across an oxygen minimum zone

Marc A. Besseling, Ellen C. Hopmans, R. Christine Boschman, Jaap S. Sinninghe Damsté and Laura Villanueva

ABSTRACT

Benthic Archaea comprise a significant part of the total prokaryotic biomass in marine sediments. Recent genomic surveys suggest they are largely involved in anaerobic processing of organic matter but the distribution and abundance of these archaeal groups is still largely unknown. Archaeal membrane lipids composed of isoprenoid diethers or tetraethers (glycerol dibiphytanyl glycerol tetraether, GDGT) are often used as archaeal biomarkers. Here, we compare the archaeal diversity and intact polar lipid (IPL) composition in both surface (0−0.5 cm) and subsurface (10−12 cm) sediments recovered within, just below, and well below the oxygen minimum zone (OMZ) of the Arabian Sea. Archaeal 16S rRNA gene amplicon sequencing revealed a predominance of Thaumarchaeota (Marine Group I, MG-I) in oxygenated sediments. Quantification of archaeal 16S rRNA and ammonia monoxygenase (*amo*A) of Thaumarchaeota genes and their transcripts indicated the presence of an active *in situ* benthic population, which coincided with a high relative abundance of hexose phosphohexose crenarchaeol, a specific biomarker for living Thaumarchaeota. On the other hand, anoxic surface sediments within the OMZ and all subsurface sediments were dominated by Archaea belonging to the Miscellaneous Crenarchaeota Group (MCG), the Thermoplasmatales and Archaea of the DPANN (superphylum grouping Micrarchaeota, Diapherotrites, Aenigmarchaeota, Nanohaloarchaeota, Parvarchaeota, Nanoarchaeota, Pacearchaeota and Woesearchaeota). Members of the MCG were diverse with a dominance of subgroup MCG-12 in anoxic surface sediments. This coincided with a high relative abundance of IPL GDGT-0 with an unknown polar head group. Subsurface anoxic sediments were characterized by higher relative abundance of GDGT-0, 2 and 3 with dihexose IPL-types, as well as GDGT-0 with a cyclopentanetetraol molecule and a hexose, as well as the presence of specific MCG subgroups, suggesting that these groups could be the biological sources of these archaeal lipids.

INTRODUCTION

Archaea are ubiquitous microorganisms in the marine system (DeLong et al., 1994; Delong and Pace, 2001; Schleper et al., 2005). They occur in diverse environments, e.g. hydrothermal vents (Stetter et al., 1990), the marine water column (Karner et al., 2001; Massana et al., 2004), in the underlying sediments (Teske and Sørensen, 2008; Lloyd et al., 2013), and well below the seafloor (Biddle et al., 2006; Lipp et al., 2008), where they are considered key players in diverse biogeochemical processes (Offre et al., 2013, and references cited therein). Specifically marine sediments have been shown to contain a highly diverse archaeal community (Teske and Sørensen, 2008; Teske, 2013; Lloyd et al., 2013; Spang et al., 2017). The ammonia-oxidizing Thaumarchaeota of the marine group I.1a (further referred to as MG-I) is probably the most widely studied archaeal group in marine sediments. However, in comparison with studies of marine pelagic Thaumarchaeota, the diversity and distribution of benthic Thaumarchaeota is still not well established (e.g. Durbin & Teske, 2010; Jorgenson et al., 2012; Learman et al., 2016). Genomic studies have revealed the existence of uncultured archaeal groups other than Thaumarchaeota in marine, predominantly anoxic, sediments such as the Miscellaneous Crenarchaeota Group (MCG; Meng et al., 2014), Archaea of the DPANN superphylum (composed of Micrarchaeota, Diapherotrites, Aenigmarchaeota, Nanohaloarchaeota, Parvarchaeota, Nanoarchaeota, Pacearchaeota and Woesearchaeota; Castelle et al., 2015; Rinke et al., 2013) and the Marine Benthic Group (MBG) B (Teske & Sørensen, 2008), and D (Lloyd et al., 2013). In the case of the Archaea belonging to the groups of the MCG and MBG-D, metagenomic studies suggest that they are able to degrade extracellular proteins and aromatic compounds (Lloyd et al., 2013; Meng et al., 2014).

Archaeal diversity is currently determined through nucleic acid-based methods but the characterization of other cellular biomarkers such as membrane lipids has proven to be also effective in tracking the presence of archaeal groups in different ecosystems (e.g. Coolen et al., 2004a; Ingalls et al., 2012; Meador et al., 2015; Pitcher et al., 2011b; Sturt et al., 2004). One of the advantages of using lipid-based methods to determine the presence of archaeal groups is that lipids can be preserved in the

sedimentary record. Therefore, they can also be used as biomarkers of the presence and metabolic potential of these microorganisms in past environments. On the contrary, other biomolecules like DNA have a more rapid turnover and they cannot be used for this purpose. In recent years, intact polar lipids (IPLs) have increasingly been applied for tracing 'living' bacteria and Archaea in the environment (Lipp et al., 2008; Lipp and Hinrichs, 2009; Rossel et al., 2008). IPLs with polar head groups are present in living cells but upon cell lysis the polar head groups are lost, releasing core lipids (CLs) that may be preserved in the fossil record. Since IPLs degrade relatively quickly after cell death (Harvey et al., 1986), it is possible to associate the presence of IPLs in the environment with the occurrence of their living producers (Lipp and Hinrichs, 2009; Schubotz et al., 2009).

Archaeal membrane lipids are typically a variation of two main structures, sn-2,3-diphytanylglycerol diether (archaeol) with phytanyl (C_{20}) chains in a bilayer structure, and sn-2,3-dibiphytanyl diglycerol tetraether (glycerol dibiphytanyl glycerol tetraether, GDGT), in which the two glycerol moieties are connected by two C_{40} isoprenoid chains, allowing the formation of a monolayer membrane (Koga and Morii, 2007). GDGTs containing 0–4 cyclopentane moieties (Figure S1) are usually not exclusive to a specific archaeal group (Schouten et al., 2013) with the exception of the GDGT crenarchaeol, containing 4 cyclopentane and one cyclohexane moiety, which is deemed to be exclusive to the Thaumarchaeota phylum (De La Torre et al., 2008; Sinninghe Damsté et al., 2002, 2012). Recently, Lincoln et al. (2014) proposed the Marine Group II as potential producers of crenarchaeol. However, this is still debated (Lincoln et al., 2014b; Schouten et al., 2014). The newly described archaeal groups detected by genetic methods are yet uncultured, therefore, their membrane lipid composition remains unknown.

In this study, we determined the archaeal diversity in a marine benthic system along a strong gradient in bottom water oxygen concentrations and compared it with the diversity of archaeal lipids. We aimed to characterize changes in the archaeal benthic community under different physicochemical conditions, as well as to provide clues on the potential archaeal lipid biomarkers produced by uncultured benthic Archaea.

We analyzed sediments (surface 0-0.5 cm, and subsurface 10-12 cm) of the Murray ridge in the Arabian Sea, which is impinged by one of the strongest present-day oxygen minimum zones (OMZ). Previous studies observed changes in the diversity of archaeal lipids in the same environmental setting in sediments under different oxygen and nutrient concentrations (Lengger et al., 2012; 2014). In our study, we expand the repertoire of archaeal lipid diversity previously detected by Lengger et al. (2012; 2014) by analyzing these sediments with High Resolution Accurate Mass/ Mass spectrometry (UHPLC-HRAM MS). In addition, we determined the archaeal diversity by means of 16S rRNA gene amplicon sequencing, as well as the abundance and potential activity of specific archaeal groups by quantitative PCR (QPCR) of 16S rRNA and the metabolic gene coding for the ammonia monoxygenase (*amo*A gene) of Thaumarchaeota.

MATERIAL and METHODS

Sampling

Sediments were collected in the Northern Arabian Sea during the PASOM cruise in January 2009 with *R/V Pelagia*. Sediment cores obtained with a multicorer were taken on the Murray ridge at four depths, 885 m below sea level (mbsl) (within the OMZ), at 1306 mbsl (just below the OMZ), at 2470 mbsl and 3003 mbsl (both well below the OMZ) as previously described by Lengger et al. (2012). Upon retrieval the cores were sliced in 0.5 cm resolution for the first 2 cm and at 2 cm resolution beyond 10 cm below the surface, and stored at -80°C until further analysis. For an overview of the surface sediments physicochemical conditions see Table 1.

Table 1. Bottom water temperature and bottom water oxygen (BWO) concentration, oxygen penetration depth (OPD) in the sediment, and TOC content and pore water composition of the surface (0-0.5 cm) sediment[a]

Station (mbsl)	T (°C)	BWO (µmol·L⁻¹)	OPD (mm)	TOC (wt %)	NH_4^+ (µM)	NO_2^- (µM)	NO_3^- (µM)	HPO_4^{2-} (µM)
885	10	2.0	0.1	5.6 (± 0.2)	2	1.2	1.3	9.2
1306	6.7	14.3	2.9	2.9 (± 0.1)	2.6*	0.1*	36.2*	5.6
2470	2.1	63.8	9.8	0.8 (± 0.1)	-[b]	-	-	-
3003	1.4	82.9	19	0.7 (± 0.1)	55.6	8.3	46.2	3.8

[a] Data from Kraal et al. (2012) and Lengger et al. (2014). [b] No data available.

Lipid extraction and analysis

Total lipids were extracted from surface (upper 0-0.5 cm) and subsurface (10-12 cm) sediments after freeze-drying using a modified Bligh and Dyer method (Bligh and Dyer, 1959) as previously described by Lengger et al. (2014). C_{16}-PAF (1-O-hexadecyl-2-acetoyl-sn-glycero-3-phosphocholine) was added to the extracts as an internal standard and the extracts were dried under a stream of nitrogen. The extracts with the added standard were then dissolved by adding solvent (hexane:isopropanol:H_2O 718:271:10 [v/v/v/v]) and filtered through a 0.45 µm, 4 mm-diameter True Regenerated Cellulose syringe filter (Grace Davison, Columbia, MD, USA).

IPLs were analyzed according to Sturt et al. (2004) with some modifications. An Ultimate 3000 RS UHPLC, equipped with thermostated auto-injector and column oven, coupled to a Q Exactive Orbitrap MS with Ion Max source with heated electrospray ionization (HESI) probe (Thermo Fisher Scientific, Waltham, MA), was used. Separation was achieved on a YMC-Triart Diol-HILIC column (250 x 2.0 mm, 1.9 µm particles, pore size 12 nm; YMC Co., Ltd, Kyoto, Japan) maintained at 30 °C. The following elution program was used with a flow rate of 0.2 mL min⁻¹: 100% A for 5 min, followed by a linear gradient to 66% A: 34% B in 20 min, maintained for 15 min, followed by a linear gradient to 40% A: 60% B in 15 min, followed by a linear gradient to 30% A: 70% B in 10 min, where A = hexane/2-propanol/formic acid/14.8 M NH_{3aq} (79:20:0.12:0.04 [v/v/v/v]) and B = 2-propanol/water/formic

acid/ 14.8 M NH$_{3aq}$ (88:10:0.12:0.04 [v/v/v/v]). Total run time was 70 min with a re-equilibration period of 20 min in between runs. HESI settings were as follows: sheath gas (N$_2$) pressure 35 (arbitrary units), auxiliary gas (N$_2$) pressure 10 (arbitrary units), auxiliary gas (N$_2$) T 50 °C, sweep gas (N$_2$) pressure 10 (arbitrary units), spray voltage 4.0 kV (positive ion ESI), capillary temperature 275 °C, S-Lens 70 V. IPLs were analyzed with a mass range of m/z 375 to 2000 (resolving power 70,000 at m/z 200), followed by data dependent MS^2 (resolving power 17,500 ppm at m/z 200)), in which the ten most abundant masses in the mass spectrum (with the exclusion of isotope peaks) were fragmented (stepped normalized collision energy 15, 22.5, 30; isolation window 1.0 m/z). A dynamic exclusion window of 6 sec was used as well as an inclusion list with a mass tolerance of 3 ppm to target specific compounds (Table S1). The Q Exactive Orbitrap MS was calibrated within a mass accuracy range of 1 using the Thermo Scientific Pierce LTQ Velos ESI Positive Ion Calibration Solution (containing a mixture of caffeine, MRFA, Ultramark 1621, and N-butylamine in an acetonitrile-methanol-acetic acid solution).

Peak areas for each individual IPL were determined by integrating the combined mass chromatogram (within 3 ppm) of the monoisotopic and first isotope peak of all relevant adducts formed (protonated, ammoniated and/or sodiated adducts may be formed in different proportions depending on the type of IPL). PAF was used as internal standard to continuously monitor MS performance and to assess matrix effects. Reported peak areas have been corrected for these effects. Absolute quantification of IPL GDGTs was not possible due to a lack of standards. Peak areas were not corrected for any possible differences in response factors between the various classes of IPL-crenarchaeol. IPLs with the same headgroup but with the regioisomer of crenarchaeol instead of crenarchaeol as the CL co-elute on the chromatographic system used here and any peak area reported for a crenarchaeol IPL thus represents the sum of both isomers.

To rule out any degradation of the GDGT-IPLs during storage of the sediments at -20°C, the anoxic surface sediment sample at 885 mbsl was also analyzed according to the method previously used by Lengger et al. (2012). The IPL fraction was

separated from the CLs with the use of a silica column and elution with MeOH (Lengger et al., 2012). This IPL fraction was hydrolyzed for 3 h and analyzed by HPLC-APCI/MS (according to Hopmans et al., 2016) and the IPL derived CL-GDGT distribution was compared with previously published data.

Nucleic acids extraction, cDNA synthesis and quantitative PCR (QPCR) analyses

Sediment was centrifuged and the excess of water was removed by pipetting before proceeding with the extraction of nucleic acids from the sediment. DNA/RNA of surface (0-0.5 cm) and subsurface (10-12 cm) sediments was extracted with the RNA PowerSoil® Total Isolation Kit plus the DNA elution accessory (Mo Bio Laboratories, Carlsbad, CA). Concentration of DNA and RNA were quantified by Nanodrop (Thermo Scientific, Waltham, MA) and Fluorometric with Quant-iT™ PicoGreen® dsDNA Assay Kit (Life technologies, Netherlands). RNA extracts were treated with DNAse and reverse-transcribed to cDNA as described by Pitcher et al. (2011). Quantification of archaeal 16S rRNA gene copies and *amo*A gene copies were estimated by QPCR by using the following primers; Parch519F and ARC915R (archaeal 16S rRNA gene), CrenAmoAQ-F and CrenAmoAModR (*amo*A gene), as previously described (Pitcher et al., 2011). For details on the QPCR conditions, efficiency and R^2 of the QPCR assays see Table S2.

16S rRNA gene amplicon sequencing, analysis, and phylogeny

PCR reactions were performed with the universal, Bacteria and Archaea, primers S-D-Arch-0159-a-S-15 and S-D-Bact-785-a-A-21 (Klindworth et al., 2013) as previously described in Moore et al. (2015). The archaeal 16S rRNA gene amplicon sequences were analyzed by QIIME v1.9 (Caporaso et al., 2010). Raw sequences were demultiplexed and then quality-filtered with a minimum quality score of 25, length between 250-350, and allowing maximum two errors in the barcode sequence. Taxonomy was assigned based on blast and the SILVA database version 123 (Altschul et al., 1990; Quast et al., 2013). Representative operational taxonomic units (OTUs, clusters of reads with 97% similarity) of archaeal groups were extracted through filter_taxa_from_otu_table.py and filter_fasta.py with QIIME (Caporaso et al.,

2010). The phylogenetic affiliation of the partial archaeal 16S rRNA gene sequences was compared to release 123 of the Silva NR SSU Ref database (http://www.arb-silva.de/; Quast et al., 2013) using the ARB software package (Ludwig et al., 2004). Sequences were added to the reference tree supplied by the Silva database using the ARB Parsimony tool. MCG intragroup phylogeny for representative sequences of OTUs affiliated to the MCG lineage was carried out in ARB (Ludwig et al., 2004). Sequences were added by parsimony to a previously-built phylogenetic tree composed of reference sequences of the 17 MCG subgroups known so far (Kubo et al., 2012). Affiliation of any 16S rRNA gene sequences to a given subgroup was done assuming a similarity cutoff of ≥85%.

Cloning, sequencing and phylogeny of the archaeal *amo*A gene

Amplification of the archaeal *amo*A gene was performed as described by Yakimov et al., (2011). PCR reaction mixture was the following (final concentration): Q-solution 1× (PCR additive, Qiagen); PCR buffer 1×; BSA (200 µg ml^{-1}); dNTPs (20 µM); primers (0.2 pmol µl^{-1}); MgCl$_2$ (1.5 mM); 1.25 U Taq polymerase (Qiagen, Valencia, CA, USA). PCR conditions for these amplifications were the following: 95°C, 5 min; 35 × [95°C, 1 min; 55°C, 1 min; 72°C, 1 min]; final extension 72°C, 5 min. PCR products were gel purified (QIAquick gel purification kit, Qiagen) and cloned in the TOPO-TA cloning® kit from Invitrogen (Carlsbad, CA, USA) and transformed in *E. coli* TOP10 cells following the manufacturer's recommendations. Recombinant clones plasmid DNAs were purified by Qiagen Miniprep kit and screening by sequencing (n ≥ 30) using M13R primer by Macrogen Europe Inc. (Amsterdam, The Netherlands). Obtained archaeal *amo*A protein sequences were aligned with already annotated *amo*A sequences by using the Muscle application (Edgar, 2004). Phylogenetic trees were constructed with the Neighbor-Joining method (Saitou and Nei, 1987) and evolutionary distances computed using the Poisson correction method with a bootstrap test of 1,000 replicates.

RESULTS

In this study, we analyzed both IPLs and DNA/RNA extracts from sediments previously collected along the Arabian Sea Murray Ridge within the OMZ (885 mbsl), just below the lower interface (1306 mbsl), and well below the OMZ (2470 and 3003 mbsl). The surface sediment (0-0.5 cm) at 885 mbsl was fully anoxic, however, the surface sediments below the OMZ were partly oxygenated (1306 mbsl), and fully oxygenated at 2470 and 3003 mbsl (Table 1). The subsurface sediments (10-12 cm) were fully anoxic at all stations (Table 1). For more details on the physicochemical conditions in these sediments see Table 1.

Archaeal IPL-GDGTs in the surface and subsurface sediments

A range of IPL-GDGTs (GDGT-0 to 4 and crenarchaeol) with the IPL-types monohexose (MH), dihexose (DH) and hexose-phosphohexose (HPH) was detected in surface and subsurface sediments across the Arabian Sea OMZ (Table 2).

For the DH GDGT-0 two structural isomers (type-I with two hexose moieties at both ends of the CL, and type-II with one dihexose moiety; Table 2) were detected and identified based on their mass spectral characteristics (Figure S2). These isomers were previously also reported in thaumarchaeotal cultures (Elling et al., 2014; Elling et al., 2017). In addition, GDGT-0 with both an ether-bound cyclopentanetetraol moiety and a hexose moiety as head groups was identified (Figure S2) in some sediments (Table 2). This IPL was previously reported as a glycerol dibiphytanyl nonitol tetraether (GDNT; de Rosa et al. 1983) but was later shown to contain a 2-hydroxymethyl-1-(2,3-dihydroxypropoxy)-2,3,4,5-cyclopentanetetraol moiety by Sugai et al., (1995) on the basis of NMR spectroscopy characterization.

Table 2. Total IPL abundance and heatmap[a] of the relative abundance (%) of the detected IPLs and sum (not color coded) per IPL-GDGT in the sediments studied.

Sediment	Depth (mbsl)	GDGT-0 MH	GDGT-0 DH I[b]	GDGT-0 HCP[c] II[b]	GDGT-0 HPH	GDGT-0 Sum	GDGT-1 MH	GDGT-1 DH I[b]	GDGT-1 HPH	GDGT-1 Sum	GDGT-2 MH	GDGT-2 DH I[b]	GDGT-2 HPH	GDGT-2 Sum
Surface (0–0.5 cm)	885	0.3	ND[d]	ND	ND	0.3	0.1	1.6	ND	1.7	0.1	29.5	ND	29.6
	1306	1.1	ND	ND	36.6	37.6	0.1	1.5	0.2	1.7	ND	15.4	ND	15.4
	2470	0.2	0.1	ND	71.5	71.9	0.0	0.1	0.4	0.5	ND	0.8	ND	0.8
	3003	0.5	0.1	ND	80.3	80.8	ND	0.2	ND	0.2	ND	0.9	ND	0.9
Subsurface (10–12 cm)	885	0.3	ND	7.8	2.1	11.9	0.1	1.7	0.1	1.9	0.2	27.0	ND	27.1
	1306	2.2	0.9	1.8	2.1	7.4	0.2	6.7	ND	6.9	0.1	29.7	ND	29.7
	2470	4.3	2.7	ND	18.6	25.6	0.1	5.8	ND	5.9	ND	23.2	ND	23.2
	3003	9.1	3.4	ND	13.0	25.5	0.2	4.3	ND	4.6	ND	21.9	ND	21.9

Sediment	Depth (mbsl)	GDGT-3 MH	GDGT-3 DH I[b]	GDGT-3 HPH	GDGT-3 Sum	GDGT-4 MH	GDGT-4 DH I[b]	GDGT-4 HPH	GDGT-4 Sum	Crenarchaeol MH	Crenarchaeol DH I[b]	Crenarchaeol HPH	Crenarchaeol Sum	IPL abundance [au. g sed dw⁻¹][e]
Surface (0–0.5 cm)	885	ND	17.8	ND	17.8	ND	6.1	ND	6.1	1.3	43.1	0.3	44.6	3.3E+08
	1306	0.0	6.9	ND	6.9	ND	2.7	ND	2.7	1.4	15.5	18.7	35.6	2.6E+09
	2470	ND	0.2	ND	0.2	ND	0.0	ND	0.0	0.2	0.6	25.8	26.6	4.9E+09
	3003	ND	0.4	ND	0.4	ND	0.0	ND	0.0	0.4	0.2	17.1	17.6	9.4E+09
Subsurface (10–12 cm)	885	0.1	15.9	ND	15.9	ND	9.4	ND	9.4	1.1	31.1	1.5	33.8	5.2E+08
	1306	0.0	14.5	ND	14.5	ND	6.1	ND	6.1	2.7	32.4	0.4	35.5	2.8E+08
	2470	ND	9.6	ND	9.6	ND	2.9	ND	2.9	3.5	28.3	1.0	32.8	9.4E+07
	3003	ND	9.7	ND	9.7	ND	5.6	ND	5.6	8.2	23.9	0.6	32.7	1.9E+08

[a] Green colors indicate a low relative abundance, red colors indicate a high relative abundance. [b] DH isomers were detected as a GDGT with a glycosidically-bound hexose moiety on both ends of the core (I) and with one glycosidically-bound dihexose moiety on one end (II). [c] HCP is an IPL-type with an ether-bound cyclopentanetetraol moiety on one end and an hexose moiety on the other (previously reported as GDNT; e.g. De Rosa and Gambacorta, 1988; Sturt et al., 2004). [d] ND = not detected. [e] Response area of summed IPLs given in au (arbitrary units) per gram of dry weight (dw) sediment.

2

In the surface sediment at 885 mbsl, crenarchaeol IPLs were dominant (44.7% of all detected IPL-GDGTs), occurring predominantly with DH as IPL-type (with a hexose head group on both ends; 43.1%; Table 2). IPL-GDGT-2 was the second most abundant (29.6%), also mainly consisting of the IPL-type DH (29.5%; Table 2). IPL-GDGT-0, -1, -3 and -4 were occurring with relative abundances of 0.3%, 1.7%, 17.8% and 6.1%, respectively (Table 2). Overall, the majority (98.1%; Table 3) of IPL-GDGTs in surface sediment at 885 mbsl with IPL-type DH (all with a hexose molecule on both ends of the CL).

Table 3. Relative abundance of IPL-GDGTs grouped by polar head group[a].

Sample	Depth (mbsl)	MH	DH	HCP	HPH
	885	1.7%	98.1%	0.0%	0.3%
Surface	1306	2.6%	42.0%	0.0%	55.4%
(0-0.5 cm)	2470	0.5%	1.8%	0.0%	97.7%
	3003	0.8%	1.8%	0.0%	97.4%
	885	1.8%	92.9%	1.6%	3.7%
Subsurface	1306	5.2%	91.9%	0.4%	2.5%
(10-12 cm)	2470	7.9%	72.6%	0.0%	19.6%
	3003	17.6%	68.8%	0.0%	13.6%

[a] Polar head group types detected: MH = monohexose, DH = dihexose, both isomers combined, HCP = monohexose and cyclopentanetetraol, HPH = monohexose and phosphohexose.

The surface sediment at 1306 mbsl contained mostly IPL-GDGT-0 (37.6% of all detected IPL-GDGTs), almost entirely with the IPL-type HPH (36.6% of the total; Table 2). Slightly less abundant was the IPL-crenarchaeol (35.6%), with the IPL-types HPH (18.7%) and DH type-I (15.5%) in equal amounts and with a minor relative abundance with MH (1.4%). Overall, the IPL-GDGTs in surface sediment at 1306 mbsl mainly contained the IPL-types HPH (55.4%; Table 3) and DH (42.0%; Table 3).

Well below the OMZ, surface sediments from 2470 and 3003 mbsl were both dominated by IPL-GDGT-0 (71.9 and 80.8% of all detected IPL-GDGTs, respectively), predominantly with IPL-type HPH (Table 2; Figure 1a).
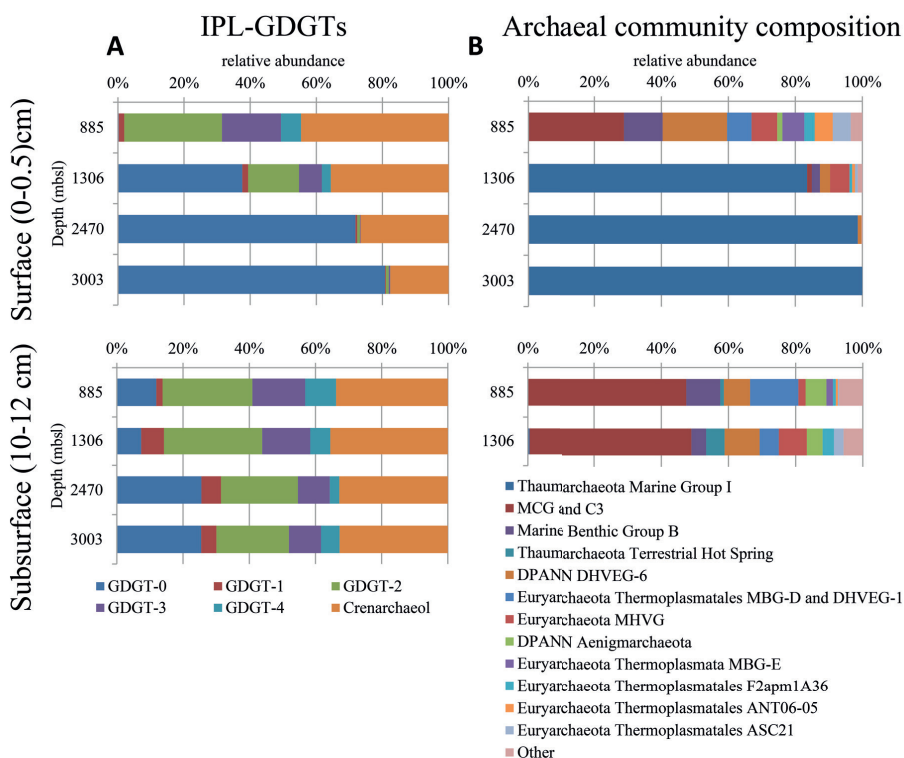
Figure 1. (A) Relative abundances of the IPL-GDGTs (sum of IPL types MH, DH and HPH) for the different core GDGTs in the surface (0–0.5 cm) and subsurface sediments (10–12 cm) and (B) the archaeal community composition as revealed by 16S rRNA gene reads (with average abundance above of > 1 %) in the surface sediments at 885, 1306, 2470, and 3003 mbsl and in the subsurface sediments at 885 and 1306 mbsl.

The IPL-crenarchaeol had a lower relative abundance (26.6 and 17.6%, respectively) and again was dominated by the member with IPL-type HPH (Table 2). The other IPL-GDGTs occurred in minor quantities (<1%). Overall, IPL-type HPH was, thus, by far the most abundant head group detected in surface sediments at 2470 and 3003 mbsl (97.7% and 97.4%, respectively), in contrast to the other two surface sediments studied (Table 3).

In all subsurface (10-12 cm) sediments (i.e. at 885, 1306, 2470 and 3003 mbsl) the most abundant IPL-GDGTs were DH-crenarchaeol (28.9±3.8%; Table 2) and DH-GDGT-2 (25.5±3.5%; Table 2). DH was also the most commonly observed IPL-type attached to GDGT-3 and GDGT-4 (Table 2). Overall the distributions of

the IPL-GDGTs in all subsurface sediments were relatively similar (Figure 1a) in comparison to the substantial changes observed at the surface (cf. Figure 1a). Overall, the IPL-type DH was the predominant one detected in subsurface sediment with a relative abundance ranging from 68.8% at 3003 mbsl to 92.9% at 885 mbsl (Table 3). In contrast to all other sediments, in the subsurface sediments at 885 mbsl and 1306 mbsl, two different isomers (Figure S2) of the DH-GDGT-0 were detected (Table 2). DH type-I (0.9% at 1306 mbsl) is also found in the other surface and subsurface sediments and in combination with other core GDGT structures, whereas the other isomer (DH type-II) only occurs in subsurface sediments at 885 and 1306 mbsl (7.8% at 885 mbsl; 1.8% at 1306 mbsl; Table 2; Figure S2b). In addition, these subsurface sediments also contain small amounts of GDGT-0 with cyclopentanetetraol and MH head groups (IPL-type HCP; 1.6% at 885 mbsl; 0.4% at 1306 mbsl; Table 2; Figure S2c).

We also determined the IPL-derived CL-GDGTs in the 885 mbsl surface sediment following the method of Lengger et al. (2012), in order to exclude IPL degradation within the stored samples. The CL-GDGTs composition derived from freshly obtained IPL showed the same distribution ($r= 0.99$, $\rho < 0.001$) as reported previously (Lengger et al., 2012).

Archaeal diversity in the surface and subsurface sediment

Different archaeal groups were detected in surface and subsurface sediment across the Arabian sea OMZ. The surface sediment at 885 mbsl, contained archaeal 16S rRNA gene sequences that were assigned to several archaeal groups (Figure 1b). The most dominant group was MCG (Total 30.5%, 12.2% attributed to C3; also known as MCG-15, Kubo et al., 2012). Another major group found was the DPANN Woesearchaeota Deep sea Hydrothermal Vent Group 6 (DHVEG-6, 20.3%; Figure 1b; Castelle et al., 2015). Marine Benthic Group (MBG) -B, -D and -E were also present with 12.2%, 7.7% and 6.9% of the archaeal 16S rRNA gene reads, respectively (Figure 1b). Sequences affiliated to the Marine Hydrothermal Vent Group (MHVG,

8.1%) of the phylum Euryarchaeota were also detected (Figure 1b). Other groups, with lower relative abundances, were Thermoplasmatales groups ANT06-05 (5.7%) and F2apm1A36 (3.3%) and the DPANN Aenigmarchaeota (previously named Deep Sea Euryarchaeotic Group, DSEG; 1.6%; Figure 1b).

Below the OMZ, in partly and fully oxygenated surface sediments at 1306, 2470 and 3003 mbsl (Table 1), the most dominant archaeal group was Thaumarchaeota MG-I with relative abundances of 81.5%, 89.7% and 100%, respectively (Figure 1b). At 1306 mbsl other archaeal groups, such as MHVG (5.6%), Thermoplasmatales ASC21 (3.2%), DHVEG-6 (2.9%), MBG-B (2.4%) and MCG (1.3%) made up the rest of the archaeal community (Figure 1b). At 2470 mbsl DHVEG-6 (1.1%) was still detectable besides the MG-I (Figure 1b).

In the subsurface sediments (10-12 cm), only the DNA extracted from the sediments at 885 and 1306 mbsl gave a positive amplification signal. The archaeal composition of the subsurface (10-12 cm) sediments at 885 mbsl and 1306 mbsl was similar (Figure 1b; Pearson correlation coefficient of 0.95), with most of the 16S rRNA gene reads classified within the MCG (47.5% and 48.4%, respectively). Other archaeal groups, such as MBG-D (14.4% and 5.7%, respectively), MBG-B (10.1% and 4.4%), the Woesearchaeota (7.8% and 10.4%), were also detected with comparable relative abundances (Figure 1b). Other archaeal groups such as Thaumarchaeota Terrestrial hot spring, the Euryarchaeota MHVG, MBG-E and the Aenigmarchaeota were detected but at low (< 10%) relative abundance (Figure 1b). Only minor amount of reads were classified as Thaumarchaeota MG-I (0.5% at 1306 mbsl) (Figure 1b).

Considering the high relative abundance of the MCG detected in the surface sediment at 885 mbsl, as well as in the subsurface (10-12 cm) sediments at 885 mbsl and 1306 mbsl (between 30.5-48.4% of total archaeal 16S rRNA gene reads detected in those samples), we performed phylogenetic analyses to determine the diversity of subgroups of the MCG within these sediments. A total of 57 representative 16S rRNA gene reads assigned to MCG were extracted from the dataset and incorporated in a MCG phylogenetic tree of Fillol et al. (2015) (Figure 2).
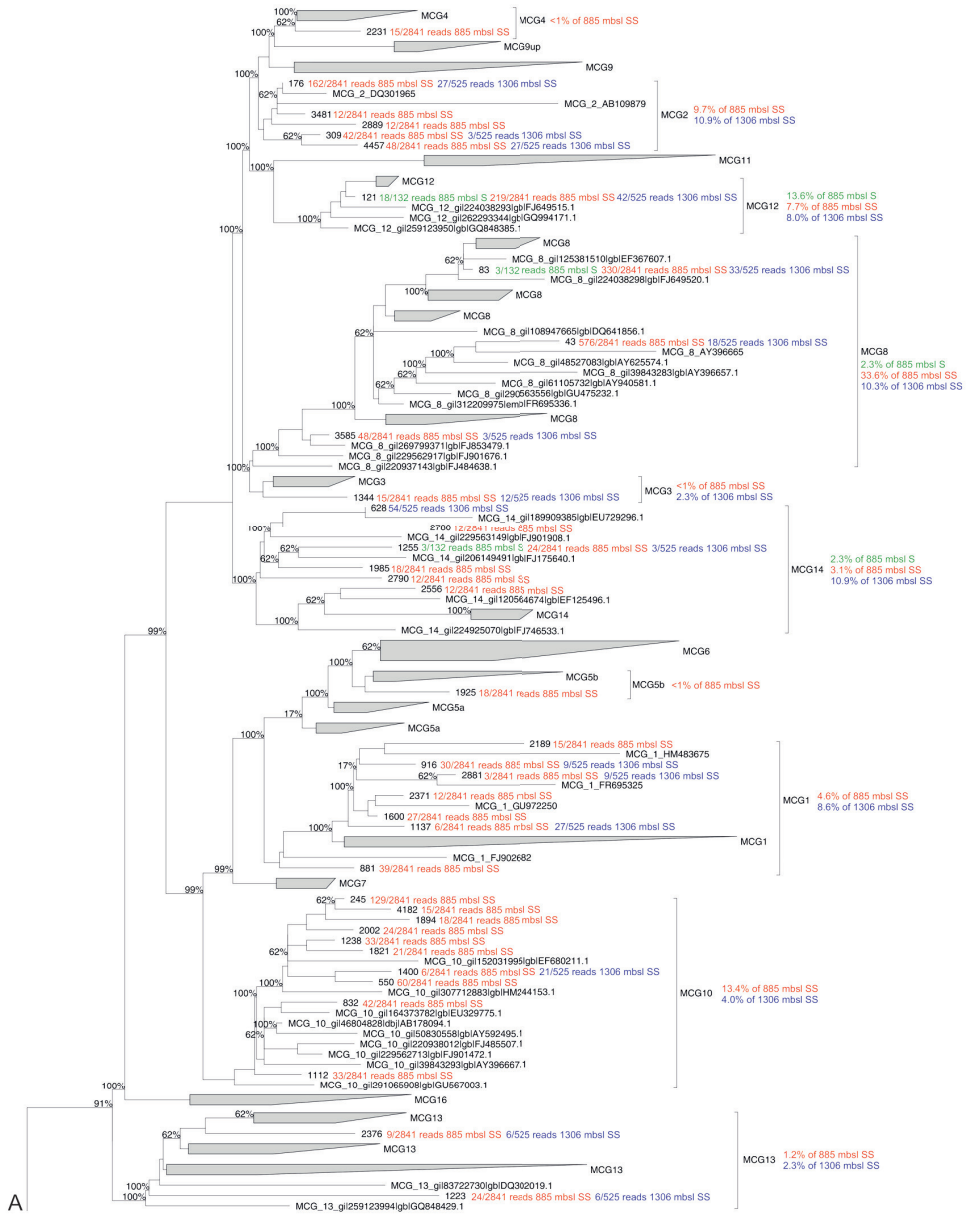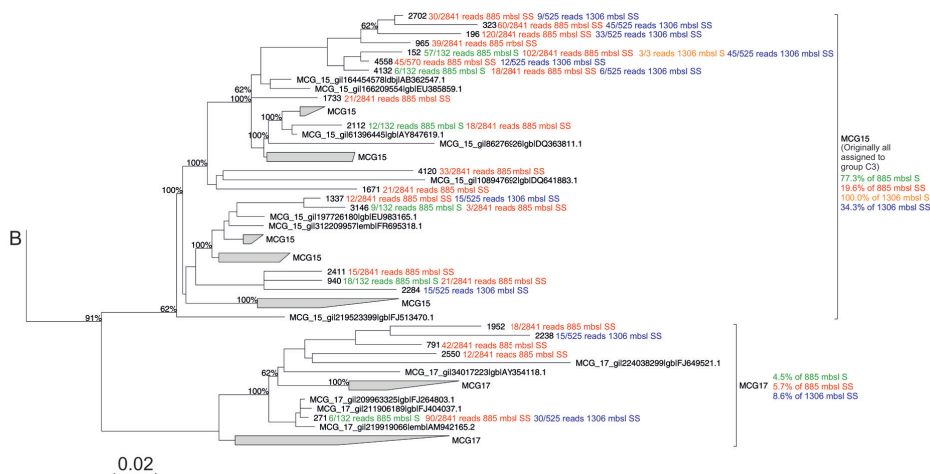
Figure 2. Maximum likelihood phylogenetic tree of the archaeal groups MCG+C3 (modified from Fillol et al., 2015). Extracted OTUs from the Arabian Sea sediments assigned as MCG were inserted in the tree. The number of detected reads per OTU per samples are indicated. Per MCG subgroup the relative abundance is given as detected at the different stations and sediments depths, this is also noted in Table 4. Scale bar represents a 2% sequence dissimilarity.

Table 4. Relative abundance (in %) of MCG- and C3-assigned 16S rRNA gene reads relative to total archaeal reads and distribution (in %) of various subgroups for a station within and a station just below the OMZ

Subgroup	885 mbsl		1306 mbsl	
	Surface (0-0.5 cm)	Subsurface (10-12 cm)	Surface (0-0.5 cm)	Subsurface (10-12 cm)
Total	30.5	47.5	1.3	48.8
1		4.6		8.6
2		9.7		10.9
3		<1		2.3
4		<1		
5b		<1		
8	2.3	33.6		10.3
10		13.4		4.0
12	13.6	7.7		8.0
13		1.2		2.3
14	2.3	3.1		10.9
15	77.3	19.6	100	34.3
17	4.5	5.7		8.6

The majority of MCG 16S rRNA gene reads from the 885 mbsl surface sediment (77.3%; Table 4) clustered in subgroup 15. In the 885 mbsl subsurface sediment, the majority of MCG reads clustered within subgroups 8 and 15 (33.6% and 19.6%, respectively; Table 4). In the 1306 mbsl surface sediment there was only a low relative abundance of MCG (Figure 1b); all MCG Archaea detected clustered in subgroup 15 (Table 4). On the other hand, in the 1306 mbsl subsurface sediment the reads clustered in subgroups 15, 2 and 14 (34.3%, 10.9% and 10.9%, respectively; Figure 2).

As the Thaumarchaeota MG-I was dominant in oxygenated sediments at 1306, 2470 and 3003 mbsl (Figure 1b), we further analyzed the diversity of this group by performing a more detailed phylogeny of the recovered 16S rRNA gene reads attributed to this group. Five OTUs dominated the Thaumarchaeota MG1 (Table 5); we will refer to them as OTU-1 to -5. OTU-1, 2, 3 and 5 were phylogenetically closely related to other known benthic Thaumarchaeota MG-I species, such as '*Ca*. Nitrosoarchaeum koreensis MY1*'* or environmental 16S rRNA gene sequences from marine sediments (Figure 3).

Table 5. Total Thaumarchaeota MG-I 16S rRNA gene reads and distribution per OTU (%) in surface sediments.

	Depth (mbsl)			
	885	1306	2470	3003
Total reads	0	915	1341	1305
OTU ID #1	n.a.[a]	4.3	2.5	3.0
OTU ID #2	n.a.	3.9	8.1	13.6
OTU ID #3	n.a.	43.6	67.6	61.8
OTU ID #4	n.a.	35.1	1.6	0
OTU ID #5	n.a.	3.3	4.7	2.1

[a] n.a. = not applicable
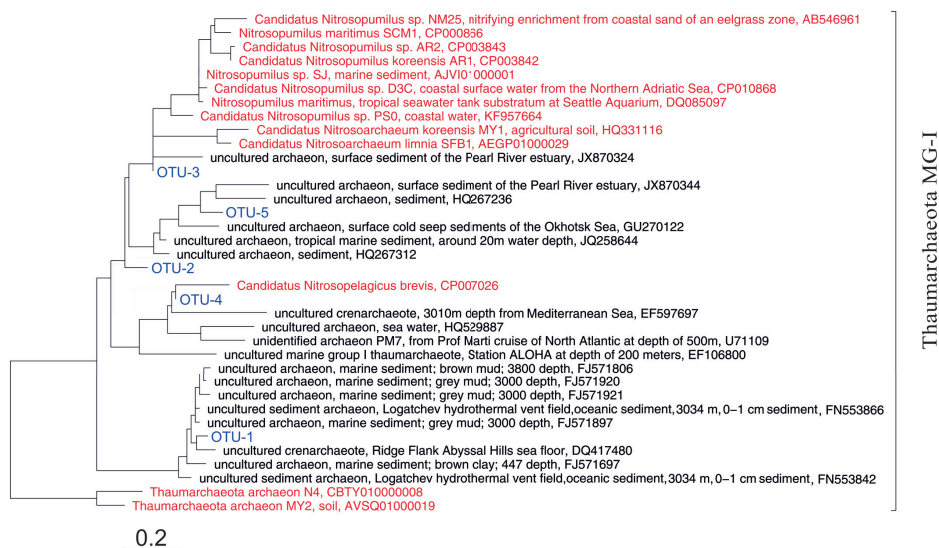
Figure 3. Maximum likelihood phylogenetic tree of MG-I OTUs recovered within the sediment based on the 16S rRNA gene (colored in blue). Sequences from cultured representatives of Thaumarchaeota MG-I are indicated in red. Environmental sequences of MG-I members are indicated in black with their origin specified. The relative abundances of the various OTUs are listed in Table 4. The scale bar represents a 2% sequence dissimilarity.

On the other hand, OTU-4 clustered with 16S rRNA gene sequences from pelagic Thaumarchaeota MG-I species, like *Ca*. Nitrosopelagicus brevis, and also clustered with 16S rRNA sequences recovered from seawater SPM (Figure 3). OTU-3 was the most abundant OTU in the surface sediment at 1306, 2470, and 3003 mbsl with a relative abundance of 44–68% (Table 5). At 1306 mbsl OTU-4 was the second most abundant (35.1%). This OTU had a much lower relative abundance (1.6% and 0.0%) at 2470 and 3003 mbsl, respectively (Table 5). The relative abundance of OTU-2 increased with increasing sampling station depth (Table 5), OTU-1 and 5 had an abundance <5% in the surface sediments (Table 5).

The diversity of Thaumarchaeota MG1 was further assessed by amplification, cloning and sequencing of the archaeal *amo*A gene. Most of the *amo*A gene sequences from surface (27 out of 29 clones) and subsurface sediment at 885 mbsl (9 out of 10 clones) and just one from the surface sediment from 1306 mbsl (1 out of 58 clones) were closely related with *amo*A gene sequences previously recovered from

SPM at 1050 mbsl from this area of the Arabian Sea (Villanueva et al., 2014). Phylogenetically they fall within the 'Water column B, subsurface water' *amo*A clade as defined by Francis et al. (2005) (Figure 4).
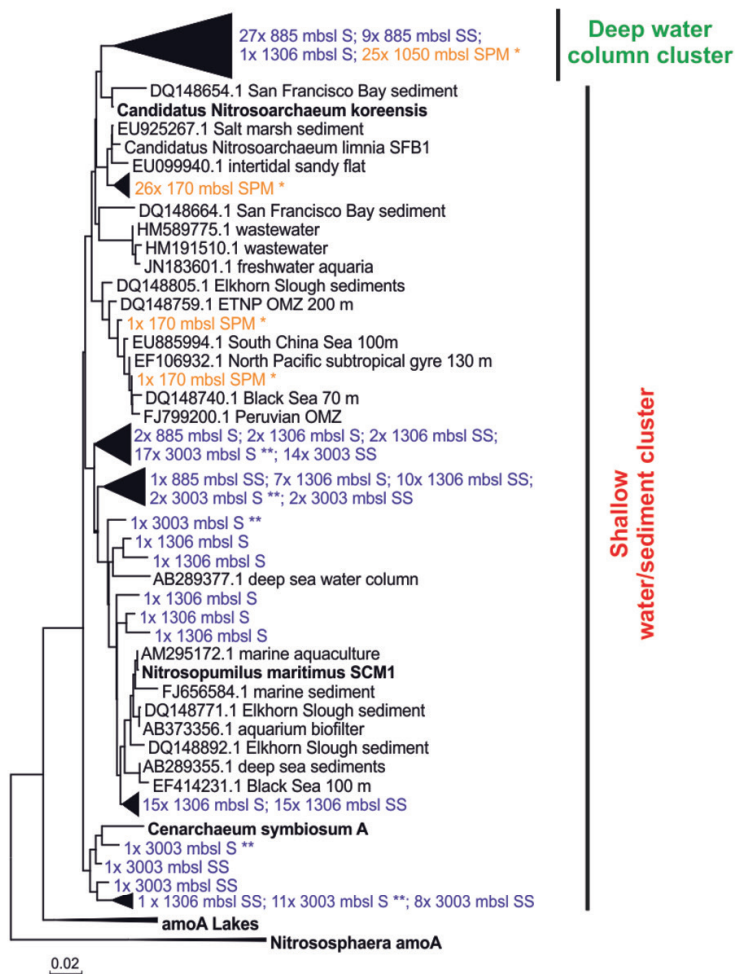
Figure 4. Maximum likelihood phylogenetic tree of *amo*A gene coding sequences recovered from surface (S; 0-0.5 cm) and subsurface (SS; 10-12 cm) sediments (colored in blue) at 885 mbsl, 1306 mbsl and 3003 mbsl (155 clones). *Amo*A gene coding sequences recovered from SPM (colored in orange) at 170 mbsl (28 clones), SPM at 1050 (25 clones) reported by Villanueva et al. (2014). ** indicates *amo*A gene sequences recovered from surface sediments at 3003 mbsl previously reported in Villanueva et al., (2015). Scale bar represents a 2% sequence dissimilarity.

At 1306 and 3003 mbsl (surface and subsurface) the majority of recovered *amo*A gene sequences clustered within the 'shallow water/sediment' clade (100 and 98.3%, respectively) and are closely related with *amo*A gene sequences from water column SPM at 170 mbsl (Villanueva et al., 2014) as well as *amo*A gene coding sequences previously detected in sediments (Villanueva et al., 2014; Figure 4). Of all recovered *amo*A gene sequences from 885 mbsl only a small fraction (8.3%) clustered within the 'shallow water/sediment' clade (Figure 4).

Abundance and potential activity of Archaea in surface and subsurface sediments

The abundance of archaeal 16S rRNA gene copies in the surface sediments of different stations varied slightly: it was lowest at 1306 mbsl (9.8×10^9 copies g^{-1} sediment) and highest at 2470 mbsl (1.5×10^{11}; Figure 5a).
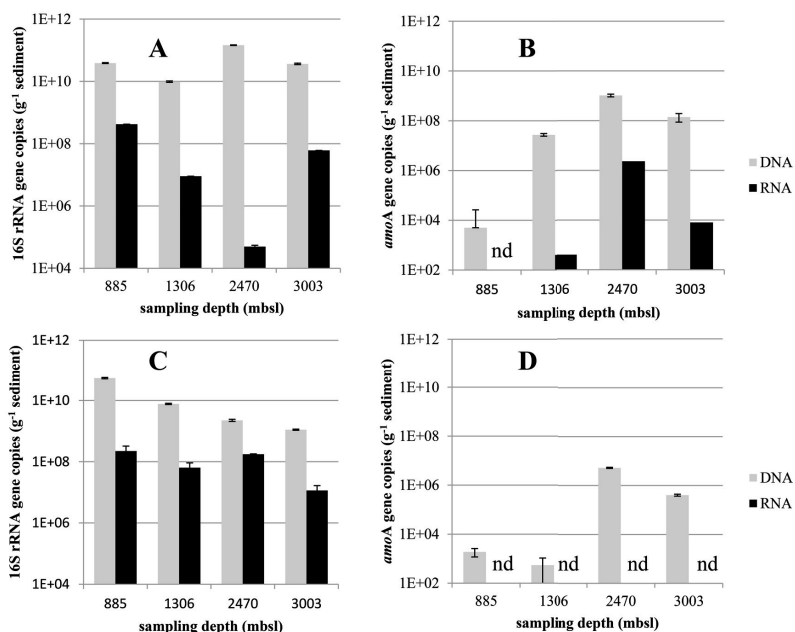
Figure 5. Abundance of Thaumarchaeotal 16S rRNA (A,C) and *amo*A (B,D) gene fragment copies per gram of dry weight in the surface sediment (0-0.5 cm) (A,B) and the subsurface sediment (10-12 cm) (C,D). Black bars indicate the amount of DNA 16S rRNA or *amo*A gene fragment copies and the gray bars indicate the RNA (gene transcripts) of 16S rRNA or *amo*A gene fragment copies. Error bars indicate standard deviation based on *n* = 3 experimental replicates.

The potential activity, based on the 16S rRNA gene transcripts of the archaeal 16S rRNA gene, was the lowest at 2470 mbsl (5×10^4 transcripts g^{-1} sediment), while a higher potential activity was detected at 885, 1306 and 3003 mbsl (0.9-42×10^7; Figure 5a). The abundance of archaeal 16S rRNA gene copies in the subsurface sediment varied also within one and a half order of magnitude (1.1-54×10^9; Figure 5c), with a decrease with increasing water depth. The potential activity showed less variation within the subsurface sediments (1.2-22×10^7 16S rRNA gene transcripts g^{-1} of sediment; Figure 5c) than in the surface sediments.

The abundance of Thaumarchaeota was estimated by quantifying the archaeal *amo*A gene copies. The highest abundance of *amo*A gene copies in surface sediment was detected at 2470 mbsl (1.0×10^9 copies g^{-1} sediment), and the lowest at 885 mbsl (5×10^4; Figure 5b). *Amo*A gene transcripts in surface sediment were under the detection limit at 885 mbsl but were detected below the OMZ with 4×10^2, 2.3×10^6 and 8×10^3 gene transcripts g^{-1} of sediment at 1306, 2470 and 3003 mbsl, respectively (Figure 5b). In subsurface sediments, the abundance of *amo*A gene copies was low at 885 and 1306 mbsl (5.4-19×10^2 gene transcripts g^{-1} sediment) and higher at 2470 and 3003 mbsl (4.1×10^5, 5.4×10^6, respectively; Figure 5d). *Amo*A gene transcripts were not detected in the subsurface sediments (Figure 5d).

DISCUSSION

In this study, we assessed the changes in benthic archaeal diversity and abundance in sediments of the Arabian Sea oxygen minimum zone along a gradient in bottom water oxygen concentrations. The steep Murray Ridge protrudes the OMZ, allowing the study of sediments deposited under varying bottom water oxygen concentrations. All these sediments receive organic matter (OM), the most important fuel for benthic prokaryotic activity in sediments. This OM is produced in a relatively small area of the ocean (i.e. the station within the OMZ, at 885 mbsl, and well below the OMZ, at 3003 mbsl, are only 110 km apart) and, therefore, is likely composed of the same primary photosynthate. However, due to differences in the degree of mineralization resulting from different exposure to oxic conditions in the water column, there

were differences in OM quality. OM in the sediments within the OMZ has a higher biochemical "quality" based on amino acid composition and intact phytopigments compared to OM in the sediments below the OMZ (Koho et al., 2013). Therefore, changes in the quality and flux of OM received by the different sediment niches could also affect the archaeal community composition as several of the archaeal groups (i.e. MCG and MBG-D) reported here have been suggested to use OM as carbon source in anoxic conditions (Lloyd et al., 2013).

Effect of oxygen availability on archaeal diversity and abundance in the surface sediments

We detected large differences in archaeal diversity between the surface sediment deposited within the OMZ and those deposited below the OMZ. In contrast to the diverse anaerobic archaeal community in the surface of 885 mbsl, in surface sediments at 1306, 2470 and 3003 mbsl, Thaumarchaeota MGI were dominant, representing 80-100% of the archaeal population (Figure 1). This clear difference in the benthic archaeal population in the surface sediments can be attributed to the oxygen availability as Thaumarchaeota are known to require oxygen for their metabolism (i.e. nitrification; Könneke et al., 2005). In fact, the oxygen penetration depth (OPD) was observed to be 3, 10, and 19 mm in sediments at 1306, 2470, and 3003 mbsl, respectively, while in sediments at 885 mbsl, the OPD was barely 0.1 mm (Table S1; Kraal et al., 2012). The surface (0-5 mm) sediment at 1306 mbsl was not fully oxygenated (OPD of 3 mm), which probably explains the detection in relatively low abundance (ca. 20%) of the anaerobic Archaea that thrive in the anoxic sediment from 885 mbsl. The low OPD at 1306 mbsl also explains the low *amo*A gene expression in comparison with the deeper surface sediments (Figs. 5b,d). Overall this indicates the presence of Thaumarchaeota with lower activity in the surface sediments at 1306 mbsl (Figure 5). Within the Thaumarchaeota MG1 group, we also detected changes in the relative abundance of specific OTUs in the surface sediments at 1306, 2470 and 3003 mbsl (Table 5). For example, OTU-2 becomes progressively more abundant with increasing water depth, suggesting that this OTU is favored at the higher oxygen concentrations found in the surface sediment

at 3003 mbsl. OTU-4 was closely affiliated with '*Ca.* Nitrosopelagicus brevis', a pelagic MG-I member, which indicates that this DNA is most likely derived from the overlying water column (Table 5), and thus should be considered to represent fossil DNA.

High *amo*A gene abundances were detected in the surface sediment at 2470 and 3003 mbsl, while values in the surface of 885 mbsl were approximately three orders of magnitude less. The lack of oxygen in the surface sediments at 885 mbsl and in the subsurface sediments, as well as undetectable *amo*A gene transcripts at those depths, suggest that in these cases the *amo*A gene DNA signal is fossil. It is well known that under anoxic conditions DNA of marine pelagic microbes may become preserved in sediments even for periods of thousands of years (Marco J.L. Coolen et al., 2004; A.C. Boere et al., 2011). The fossil origin of the Thaumarchaeotal *amo*A gene is supported by the phylogenetic affiliation of the *amo*A gene fragments amplified from the 885 mbsl surface sediment, as those sequences were closely related to *amo*A gene sequences detected in the suspended particulate matter in the subsurface water column (Villanueva et al., 2014), thus suggesting that the detected DNA originated from pelagic Thaumarchaeota present in the subsurface water column, as proposed for the presence of OTU-4 16S rRNA gene sequences in the surface sediment (see earlier).

There is a discrepancy between the 16S rRNA gene copy numbers and the *amo*A gene copy numbers within the sediments (Figure 5). *Amo*A gene copies were consistently lower than the 16S rRNA gene copies, even within sediments that were completely dominated by Thaumarchaeota MG-I. This may be caused by the *amo*A gene primer mismatches and/or the disparity of gene copy numbers within the archaeal genomes (Park et al., 2008).

In the anoxic surface sediment at 885 mbsl (within the OMZ), we detected a highly diverse archaeal population composed of MCG, Thermoplasmatales, MBG-B, -D and -E, Woesearchaeota, and MHVG. Archaeal groups such as MCG and MBG-B and E have been previously described in anoxic marine sediments, where they have been suggested to be involved in anaerobic OM degradation (e.g. Biddle et al., 2006;

Inagaki et al., 2003; Castelle et al., 2015). Members of the DPANN Woesearchaeota were only present in the surface sediment at 885 mbsl but not in the subsurface anoxic sediments at 885 and 1306 mbsl, suggesting that their presence here is not solely dependent on the absence of oxygen but possibly also on the OM composition and availability in surface and subsurface sediments. Alternatively, the DPANN Woesearchaeota 16S rRNA gene signal could also originate from the water column and deposited in the surface sediment at 885 mbsl as fossil DNA as observed for the case of Thaumarchaeota as mentioned above.

Archaeal community composition in the anoxic subsurface sediments

The archaeal diversity in the subsurface sediment (10-12 cm) from both 885 and 1306 mbsl (i.e. dominated by MCG, MBG-B, -D and −E) is similar to that observed in the surface sediment at 885 mbsl. This supports that oxygen availability is an important factor for determining the diversification of archaeal groups (Figure 1b). MCG, one of the dominant archaeal groups in these sediments, showed substantial differences in the distribution of its subgroups (Table 4). All subsurface sediments had a high intra-group diversity of MCG. This is in contrast with the surface sediment at 885 and 1306 mbsl where a high relative abundance of the subgroup MCG-15 is detected (Table 4). A recent study with enrichments of estuarine sediments showed evidence that MCG-8 are capable of growth with lignin as an energy source (Yu et al., 2018). MCG-8 was the most or one of the most abundant MCG groups in our studied subsurface sediments (Table 4). Substantial lignin concentration were previously not detected in the Arabian Sea, which indicated a relatively invariant terrestrial OM contribution (Cowie et al., 1999). Therefor it remains uncertain which physicochemical factor(s) influence the MCG-8 subgroup in the studied subsurface sediments. A recent survey of the ecological niches and substrate preferences of the MCG in estuarine sediments based on genomic data pointed to MCG-6 Archaea as degraders of complex extracellular carbohydrate polymers (plant-derived), while subgroups 1, 7, 15 and 17 have mainly the potential to degrade detrital proteins (Lazar et al., 2016). Lazar et al. (2016) also described the presence of aminopeptidases coded in the genome bin of MCG-15, suggesting

that this subgroup could be specialized in degradation of extracellular peptides in comparison with the other MCG subgroups, which would be restricted to the use of amino acid and oligopeptides. Considering the dominance of the MCG-15 subgroup in the surface sediments analyzed in this study (Table 4), we hypothesize that the proteinaceous OM deposited in the surface sediment, which mainly originates from photosynthate, is still quite undegraded. This would favor the MCG-15 in this niche, fueling its metabolism by the degradation of peptides extracellularly, while in subsurface sediments, other MCG groups such as 2, 8 and 14 would be more favored.

The archaeal 16S rRNA gene abundance in the subsurface sediments progressively declined with increasing water depth, while the potential activity was similar. This can be due to the expected decrease in the flux of OM being delivered to these anoxic sediments layers attributed to higher degradation of OM in oxygenated bottom waters and the progressively larger oxic zone in the sediments (Lengger et al., 2012; Nierop et al., 2017). This results in lower organic carbon concentrations and a decreased biochemical quality of the OM (Koho et al., 2013; Nierop et al., 2017) to sustain the heterotrophic archaeal population inhabiting the anoxic subsurface sediments. Also the presence or lack of macrofauna in the analyzed sediments would have an effect on the OM composition, sediments within the OMZ are less prone to bioturbation which most likely resulted in higher OM preservation (Koho et al., 2013). Differences in the OM biochemical composition can influence the microbial community composition as was shown recently for North Sea sediments (Oni et al., 2015).

Benthic Archaea as potential sources for archaeal IPLs

Archaeal lipids in surface and deeper sediments of the Murray Ridge (Lengger et al., 2012; Lengger et al., 2014), as well as in the overlaying water column (Pitcher et al., 2011; Schouten et al., 2012), have been studied earlier. The study by Lengger et al. (2012) was limited to the determination of MH-, DH- and HPH-crenarchaeol with HPLC/ ESI-MS^2 using a specific selected reaction monitoring method (SRM; Pitcher et al., 2011). A follow-up study of Lengger et al. (2014) reported MH-, DH- and HPH-IPLs with multiple CL-GDGTs. In our study, we expanded the screening

for IPLs carrying different polar head groups in combination with multiple CLs using high resolution accurate mass/mass spectrometry (see Table S1). By applying this method, we were able to target a broader range of IPLs in these sediments. This allows a more direct comparison with the archaeal diversity detected by gene-based methods. Note that the study of Lengger et al. (2014) used a different sampling resolution (surface sediment used was 0-2 compared to our 0-0.5 cm) and our results can, therefore, not be directly compared.

The fully oxygenated surface sediments showed a dominance of GDGT-0 and crenarchaeol mostly with HPH as IPL-type (Table 2). This is the expected IPL-GDGT signature of Thaumarchaeota as previously observed in pure cultures (Elling et al., 2017; Pitcher et al., 2010; Schouten et al., 2008; Sinninghe Damsté et al., 2012). The predominance of the HPH IPL-type in surface (0-2 cm) sediments was previously interpreted as an indication of the presence of an active Thaumarchaeotal population synthesizing membrane lipids *in situ* (Lengger et al., 2012, 2014), taking into account the labile nature of sedimentary phospholipids (Harvey et al., 1986; Schouten et al., 2010; Xie et al., 2013). Elling et al. (2014) showed, in a Thaumarchaeota culture experiment, that a high HPH-crenarchaeal abundance was an indication of metabolically active Thaumarchaeota. The hypothesis by Lengger et al. (2012, 2014) is strongly supported by our data because (i) the archaeal community in the oxygenated surface (0-0.5 cm) sediments is dominated by Thaumarchaeota (Figure 1) and (ii) the high abundance of thaumarchaeotal *amo*A gene copies and gene transcripts detected in the oxygenated surface sediments. On the other hand, in the anoxic surface sediment at 885 mbsl, crenarchaeol was predominantly present with DH as the predominant IPL-type (Table 2). This is considered to be a fossil signal of Thaumarchaeota deposited from the water column due to a higher preservation potential of glycolipid head groups (as present in DH) as previously suggested (Lengger et al., 2012, 2014). Xie et al. (2013) showed a high preservation potential for glyosidic ether lipids in a 300-day incubation study. However the differences in degradation rates between phospholipid versus glycolipid GDGTs still need to be determined, especially on longer time scales that apply to sediments. Nevertheless,

the presence of *amo*A gene sequences in the 885 mbsl surface sediment, which are derived from the deeper water column, as well as the much lower *amo*A gene abundance and lack of *amo*A gene expression (Figure 5b, d) supports the contention that the crenarchaeol IPLs in the surface sediment at 885 mbsl are predominantly fossil since evidence for active Thaumarchaeota is lacking.

The low relative abundance of GDGT-0 IPLs in the surface sediment at 885 mbsl (Table 2) is remarkable. Only MH-GDGT-0 was detected in low relative abundance (0.3 %), whereas any other of the IPL-types with GDGT-0 as CL that were screened for in our study (Table S2; Figure 1b) was absent. In contrast, Lengger et al. (2014) reported a significant amount of IPL-derived CL-GDGT-0 (i.e. 18.5% of total CL GDGTs) when the head groups of the IPLs are released by acid hydrolysis and the remaining CLs were analyzed in a surface (0-0.5 cm) sediment from the same site. We re-analyzed the IPL-derived CL-GDGT composition in the surface (0-0.5 cm) sediment at 88 mbsl and recovered an identical CL-GDGT distribution as reported by Lengger et al. (2014). The discrepancy between CL and IPL distribution may partly be explained by the underestimation of MH IPLs by our method. To assess the MH IPLs underestimation, two North Atlantic suspended particulate matter samples analyzed with the normal phase method (this study) and the reverse phase method (Wörmer et al., 2013). This underestimation of MH IPLs with the used normal phase method could be on average 10 fold compared to the reverse phase method (Figure S4). However, the difference in response factor for the different IPL types is not sufficient to explain the discrepancy between IPL-derived CL-GDGT-0 and IPL GDGT-0 relative abundances. Therefore, we speculate it is due to the presence of an IPL-type with unknown head groups not included in our analytical window. This unknown IPL GDGT-0 may originate from any of the archaeal groups present in the surface sediment at 885 mbsl, such as MCG, Thermoplasmatales, MBG-B, MBG-E and Euryarchaeota MHVG. DPANN Woesearchaeota is also relatively abundant in the surface sediments at 885 mbsl (Figure 1) but recent studies suggest that their small genomes lack the genes coding for the enzymes of the archaeal lipid biosynthetic pathway (Waters et al., 2003; Jahn et al., 2004;

Podar et al., 2013; Villanueva et al., 2017). Therefore, they are not expected to contribute to the IPL-GDGT pool. Ruling out the Woesearchaeota as a possible source of IPL-GDGTs, the IPL GDGT-0 with unknown polar head group(s) in the surface sediment at 885 mbsl may be attributed to the MCG, which make up 30.5% of the archaeal 16S rRNA gene reads in this sediment. Most of these MCG Archaea fall into subgroup MCG-15 (Table 4). Previous studies proposed butanetriol dibiphytanyl glycerol tetraethers (BDGTs) as putative biomarker of the MCG based on the correlation between the presence of these components and MCG in estuarine sediments (Meador et al., 2014). However, we did not detect any IPL BDGTs in the sediments analyzed in our study. Buckles et al. (2013) suggested that members of the MCG and Crenarchaeota group 1.2 could be the biological source of IPL GDGT-0 found in the anoxic hypolimnion of a tropical lake. Considering these evidences, it is possible that the unknown IPL GDGT-0 present in the surface sediment at 885 mbsl could be a biomarker for MCG.

In subsurface sediments, the IPL GDGT distribution was remarkably different from that detected in the surface oxygenated sediment as higher relative abundances of GDGT-1, 2, 3 and 4 were detected in detriment of GDGT-0, similar to the distribution detected in the surface sediment at 885 mbsl. This may represent new archaeal production in the anoxic sediments, selective preservation of archaeal lipids produced in the water column and surface sediments, or both. The HCP GDGT-0 and two isomers of the DH GDGT-0 (Figure S2) were detected in the subsurface sediments at 885 and 1306 mbsl but not in those from deeper waters (Table 2). Since these IPLs were not detected in the surface sediments, it is likely that they are produced *in situ*. Unfortunately, we only obtained information on the archaeal community composition of the subsurface sediments at shallow water depth, so we cannot compare these with the subsurface sediments from deeper waters that lack these DH moieties, which could have provided a clue towards the archaeal source of these IPLs. An IPL composed of GDGT-0 with a cyclopentanetetraol head group has been previously detected in cultures of the hyperthermophilic crenarchaeal *Sulfolobales* (Langworthy et al., 1974; Sturt et al., 2004). However, members of the

Sulfolobales were not detected in our 16S rRNA gene amplicon sequencing data. We also detected a high relative abundance of MCG (up to 48.4% relative abundance) in the subsurface sediment at 885 and 1306 mbsl (Figure 1). The diversity of the MCG population in the subsurface sediments was much higher in comparison with the diversity in surface sediments at 885 mbsl as sequences closely related to the MCG subgroups, 2, 8, 10, 14, 5b, 15, and 17 were detected both in the 885 mbsl and 1306 mbsl subsurface sediments (Figure 2). This presence of these different MCG members, likely caused by niche differentiation (see before), may be the origin of the unusual DH-GDGT-0 isomer and the HCP-GDGT-0 IPL that we detected within the subsurface sediments at 885 and 1306 mbsl.

CONCLUSIONS

By using a combined 16S rRNA gene amplicon sequencing and IPL analysis with high-resolution accurate mass/mass spectrometry we have unraveled the high diversity of benthic Archaea harbored in oxygenated and anoxic sediments of the Arabian Sea, as well as widening our detection window of archaeal intact polar lipids. DNA-based analyses revealed a dominance of active benthic *in situ* Thaumarchaeota in those sediment where oxygen was present, which coincided with high relative abundance of the HPH-crenarchaeol previously suggested to be a marker of living Thaumarchaeota. In the anoxic marine sediments analyzed here, members of the MCG, DPANN and Euryarchaeota Thermoplasmatales dominated. We also observed a high diversity within the MCG with a more diverse population in subsurface sediments. Subsurface anoxic sediments had a high relative abundance of IPL GDGT-1, 2, and 3 with DH headgroups, which could either be attributed to fossil signal due to the more recalcitrant nature of the glycosidic bonds or being IPLs synthesized by the archaeal groups detected in those sediments. In addition, IPL GDGT-0 was also detected with a hexose head group on both ends of the core lipid, two hexoses on one end, and a cyclopentanetetraol molecule bound to the core lipid and a hexose attached to it. Members of the DPANN could possibly be ruled out of making those lipids due to the lack of lipid biosynthetic pathway. Dominant archaeal members in those sediments such as the MCG and Thermoplasmatales,

could be potential biological sources of these IPLs. To conclude, this combined approach has shed light on the possible biological sources of specific archaeal IPLs and also detected a highly diverse benthic archaeal community.

ACKNOWLEDGMENTS

SUPPLEMENTAL FIGURES AND TABLES

Sampling was carried out during a research cruise (64PE424) onboard the R/V Pelagia from July - August 2017. The cruise followed a North to South transect in the North Atlantic. The two samples shown here are from a northern station (61.71°, -20.49°) and a southern station (30.66°, -14.25°). McLane in situ pumps (McLane Laboratories Inc., Falmouth) fitted with pre-ashed 0.7 µm x 142 mm, GF/F filters (Pall Corporation, Port Washington, NY) were used to collect suspended particulate matter (SPM) from the water column. Between 400 and 500 L was filtered at 200 m depth at each station using a cut-off at a pre-programmed pressure threshold. The filters were stored at −80 °C until extraction. Filters were freeze-dried and extracted and further treated as described in Materials and Methods in the main text. The extracts were analyzed in two ways, first using the identical method as described in the main text. In addition, analysis was performed using a reversed phase method based on Wörmer et al. (2013) with some modifications. Identical eluents, gradient and column conditions were used but the analysis was performed on the UHPLC-Q Exactive system described here. HESI settings were as follows: sheath gas (N2) pressure 40 (arbitrary units), auxiliary gas (N2) pressure 10 (arbitrary units), auxiliary gas (N2) T 50 °C, sweep gas (N2) pressure 5 (arbitrary units), spray voltage 4.5 kV (positive ion ESI), capillary temperature 300 °C, S-Lens 70 V. IPLs were analyzed with a mass range of m/z 400 to 2000 and data dependent approach as described in the main text. Peak integration and area corrections based on the PAF internal standard were performed as described in main text.

Table S1. PCR primers, efficiency and R2 of the QPCR assays

QPCR (gene)	Primers	Ref	Efficiency	R^2
16S rRNA	Parch519F/ARC915R	Pitcher et al., 2011	99.8	0.96
16S rRNA	Parch519F/ARC915R	Pitcher et al., 2011	87.5	1.00
16S rRNA	Parch519F/ARC915R	Pitcher et al., 2011	90.0	0.99
*Amo*A	CrenAmoAQ-F/CrenAmoAModR	Pitcher et al., 2011	97.5	1.00
*Amo*A	CrenAmoAQ-F/CrenAmoAModR	Pitcher et al., 2011	78.5	1.00
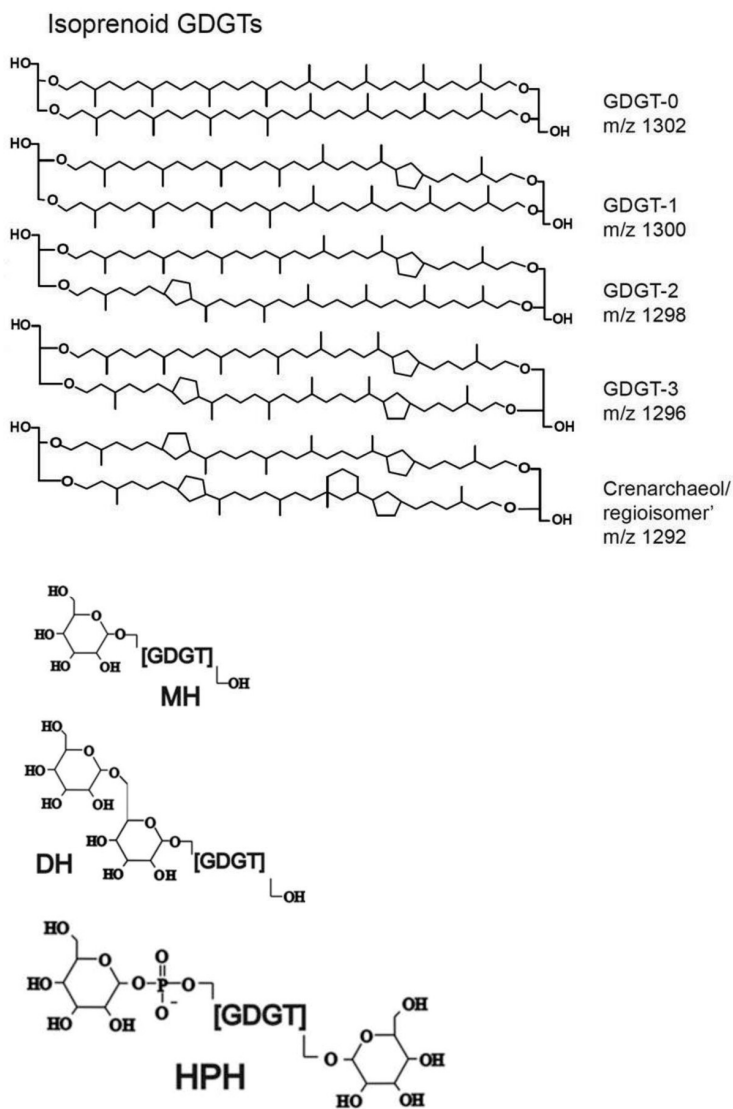
Figure S1. Structures of isoprenoid core-GDGTs discussed in the text (left) and the most commonly detected IPL types (right).
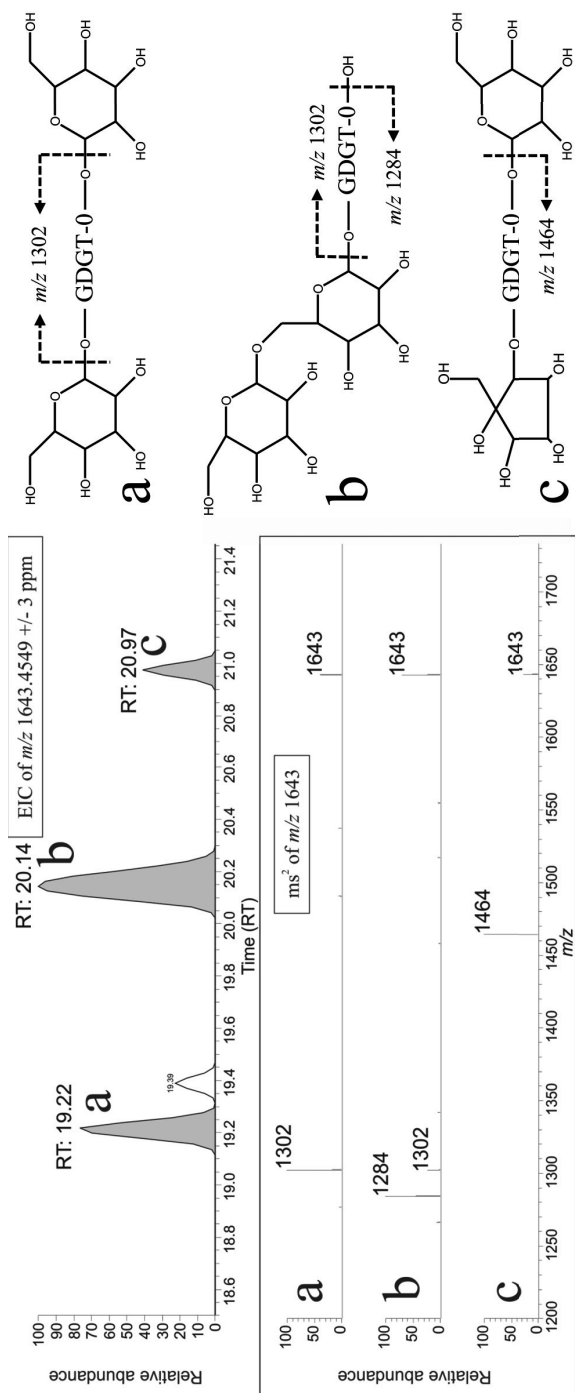
Figure S2. HPLC-ESI/MS Extracted ion current (EIC) of m/z 1643.4549 within 3 ppm mass accuracy representing the ammoniated molecule of DH-GDGT-0 in subsurface sediment (10-12 cm) at 1306 mbsl. The chromatogram shows three separate peaks representing three different isomers (the peak at 19.39 min is the signal of the second isotopologue of GDGT-1). MS2 mass spectra of each of the three isomers are shown in panel a-c in order of retention time. Panel a shows a single fragment ion at m/z 1302.3280. Panel b shows an additional fragment ion at m/z 1284.3152. Panel c shows a main fragmentation product at m/z 1464.3739. The fragmentation product at m/z 1302.3280 shown in panel a represents the core lipid of GDGT-0 resulting from the loss of 2 hexose moieties from DH-GDGT-0, with 1 hexose moiety attached to either side of the molecule. The additional fragment ion at m/z 1284 shown in panel b represents the loss of an additional water from the GDGT-0 core lipid, which can only occur if a free alcohol moiety is present and thus both hexose moieties are postulated to be attached to one side of the GDGT-0 molecule. Panel C shows the loss of a single hexose from the m/z 1643 parent, but no further fragmentation. suggesting that this isomer is in fact the IPL-type HCP GDGT, where the hexose is positioned at the opposite side from the cyclopentanetetraol moiety. If the hexose moiety would have been bound to the cyclopentanetetraol moiety, as suggested by Sturt et al., 2004, an additional loss of water from the MH-GDGT-0 would have been expected, similar to what was observed in the fragmentation of DH-GDGT-0 shown in panel b.
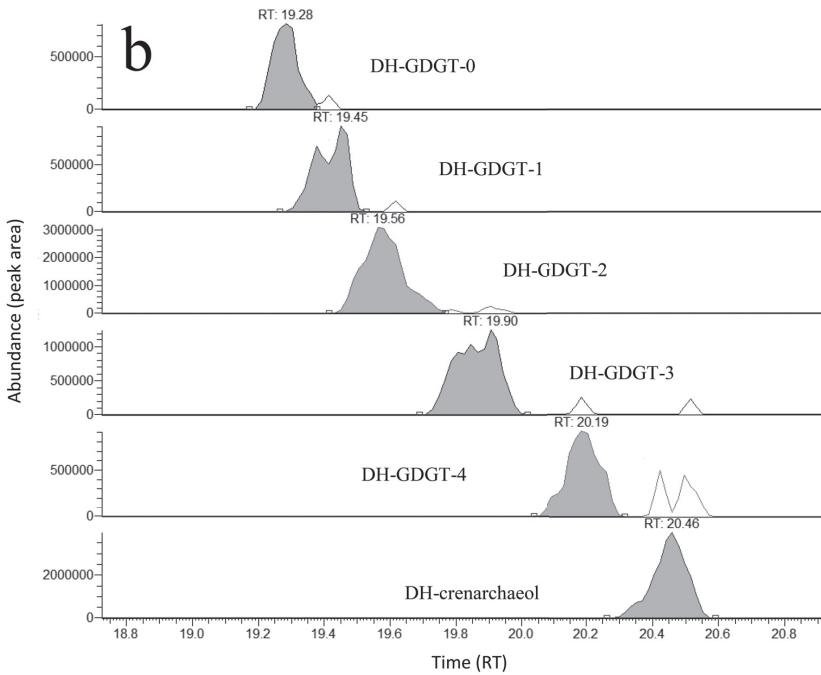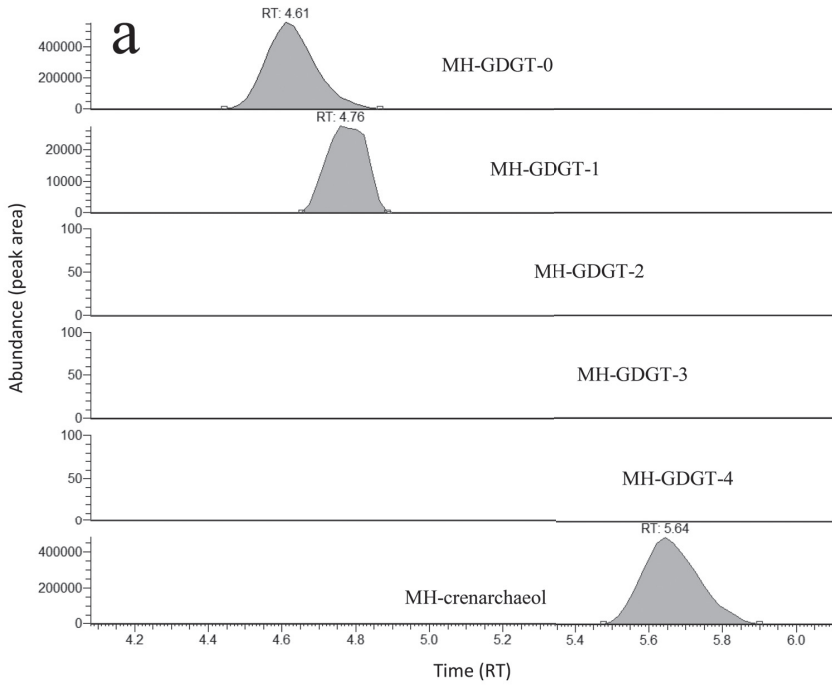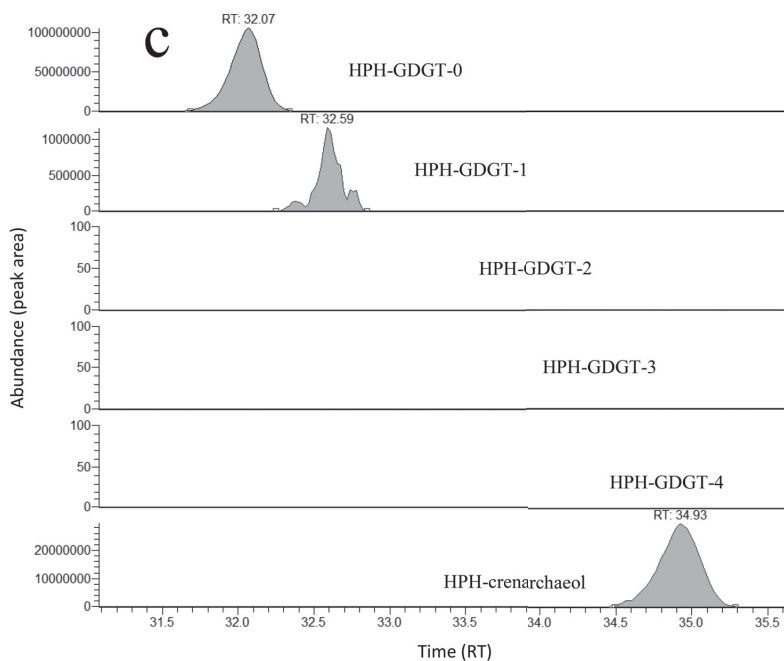
Figure S3. Summed EICs of monoisotope plus first isotopologue in diverse Arabian Sea sediments of a) MH-GDGTs (e.g. MH-GDGT-0; [M+NH4]+, m/z 1481.4020 plus 1482.4054, respectively), b) DH-GDGTs (e.g. DH-GDGT-0; [M+NH4]+, m/z 1643.4549 plus 1644.4582, respectively) and c) HPH-GDGTs (e.g. HPH-GDGT-0; [M+NH4]+, m/z 1723.4212 plus 1724.4246, respectively) within 3 ppm mass accuracy. The used internal standard, 1-O-hexadecyl-2-acetoyl-sn-glycero-3-phosphocholine (PAF; not shown), elutes at 31.4 minutes.
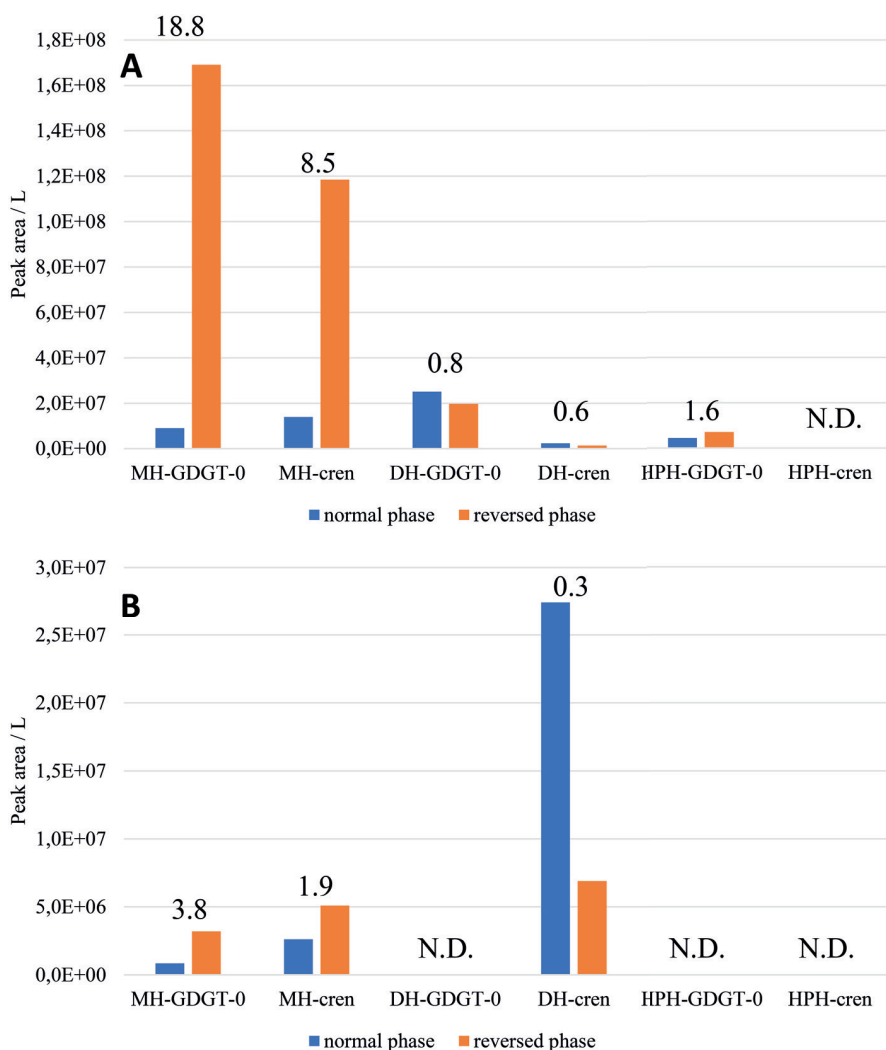
Figure S4. Peak area per liter for six IPL-GDGTs (GDGT-0 and crenarchaeol with MH (mono-hexose), DH (dihexose) and HPH (hexose phosphohexose) headgroups) in two suspended particulate matter (SPM) samples from the North Atlantic (A = Northern station, B = Southern station) analysed using a normal phase (blue bars) and reversed phase HPLC/MS method (orange bars). Numbers in graph indicate the ratio between peak areas measured with normal phase vs. reversed phase method. N.D. = not detected.

CHAPTER 3

Depth-related differences in archaeal populations impact the isoprenoid tetraether lipid composition of the Mediterranean Sea water column

Marc A. Besseling, Ellen C. Hopmans,
Michel Koenen, Marcel T.J. van der Meer,
Sanne Vreugdenhil, Stefan Schouten,
Jaap S. Sinninghe Damsté and Laura Villanueva

ABSTRACT

Thaumarchaeota are one of the most abundant groups of Archaea in the marine water column. Their membrane consists of isoprenoid glycerol dibiphytanyl glycerol tetraethers (GDGTs) which are applied in the widely used TEX_{86} proxy to reconstruct past sea surface temperatures (SSTs). However, in some specific marine systems, such as the Mediterranean Sea, core-top TEX_{86}-derived temperatures do not seem to reflect annual mean SSTs. This has been attributed to contributions of deep-water dwelling Thaumarchaeota. Here, we investigate the potential causes of this bias by studying both the archaeal diversity as well as the intact polar lipid (IPL) GDGT composition in the Mediterranean water column by a combined 16S rRNA gene amplicon sequencing and a lipidomic approach on suspended particulate matter (SPM) at different water depths. The archaeal distribution showed a dominance of archaea other than Thaumarchaeota, i.e. Marine Euryarchaeota group II and III in the upper epipelagic waters (0–100 meters deep), while Thaumarchaeota (Marine group I; MGI) dominated the subsurface and the deeper waters. This shift in the archaeal community composition coincided with a decrease in IPL GDGT-0 and increase of IPL crenarchaeol. The ratio of GDGT-2/GDGT-3 increased with water depth, but values were lower than observed in deep marine waters of some other regions. The increase of the GDGT-2/GDGT-3 ratio coincided with the high relative abundance of deep-water MGI, which may be linked to the high temperature and salinity found in specific water masses of the Mediterranean Sea. We conclude that these particularities of the Mediterranean Sea are responsible for the overestimated SST based on $TEX_{86.}$

INTRODUCTION

Archaea of the phylum Thaumarchaeota (also known as Marine group I, MGI) are among the most widespread and abundant prokaryotes in the marine realm (Karner et al., 2001; Francis et al., 2005; Teira et al., 2006). Their membrane lipids comprise archaeal glycerol dibiphytanyl glycerol tetraethers (GDGTs) with 0−4 cyclopentane moieties as well as the specific GDGT, crenarchaeol, with 4 cyclopentane moieties and a cyclohexane moiety, which is considered a specific biomarker for Thaumarchaeota (Sinninghe Damsté et al., 2002; for a review see Schouten et al., 2013). Members of the MGI are also capable of producing archaeol and methoxy-archaeol (Schouten et al., 2008; Elling et al., 2017). Importantly, the distribution of thaumarchaeotal GDGTs in the marine environment has been shown to be affected by temperature, i.e. with increasing temperature there is an increase in the relative abundance of cyclopentane-containing GDGTs and a specific isomer of crenarchaeol (cren´; Sinninghe Damsté et al., 2018). This has led to the development of the paleotemperature proxy TEX_{86} (TetraEther indeX of tetraethers consisting of 86 carbon atoms) for reconstruction of past sea surface temperatures (SSTs; Schouten et al., 2002). More recent and extended sediment core-top studies resulted in additional indices and novel calibrations (Kim et al., 2008; 2010; Tierney and Tingley, 2015).

Although there are many indications that the TEX_{86} signal of settling particles reaching the seafloor reflects the temperature of the upper epipelagic waters of the oceans, there are two main uncertainties. Firstly, the TEX_{86} signal may also be affected by input of archaea other than Thaumarchaeota, which could potentially synthesize some of the GDGTs involved in the TEX_{86} calculations. An earlier study has suggested that Marine group II Euryarchaeota (MGII) are significant contributors of GDGTs in oceanic surface waters based on the combination of core lipid (CL)-GDGT and 16S rRNA gene amplicon sequencing analyses (Lincoln et al., 2014). However, these findings have been challenged (Schouten et al., 2014) and there are no cultured representatives of the MGII group to confirm this hypothesis. Secondly, it has been suggested that TEX_{86} reflects subsurface temperature instead

3

of SST because most MGI archaea reside in subsurface waters (50–200 m) (e.g., Huguet et al., 2007; Lopes dos Santos et al., 2010; Kim et al., 2012). Considering this, several studies have noted changes in the relative abundances of certain GDGTs (e.g., an increase of the relative abundances of GDGT with 2 vs 3 cyclopentane moieties) in marine suspended particulate matter (SPM) and surface sediments with increasing water depth (Taylor et al., 2013; Hernández-Sánchez et al., 2014; Kim et al., 2015). Although the increase of the GDGT-2/GDGT-3 ratio alone with depth does not affect TEX_{86} temperature, it does suggest the presence of a (thaum)archaeotal population in deeper waters with a different GDGT membrane composition than archaea existing in the subsurface waters and this may ultimately influence the sedimentary TEX_{86} signal.

In the Mediterranean Sea, this effect is particularly strong: the GDGT distribution in surface sediments strongly varies with water depth, even at relatively small spatial scales such as in the Balearic Sea (Kim et al., 2015). When the four GDGTs used in the TEX_{86} proxy were considered, the relative abundances of both GDGT-2 (and hence the GDGT-2/GDGT-3 ratio) and the crenarchaeol isomer increased with increasing water depth. These increases result in a bias towards higher TEX_{86} reconstructed temperatures compared to annual SSTs in deep-water surface sediments. The fact that this trend was also apparent in SPM collected at different water depths suggested that this change might be due to a change in the relative abundance of Thaumarchaeota species having different GDGT distributions in deep waters and that this signal is, at least partly, incorporated into sedimentary GDGTs. Kim et al. (2016) investigated the GDGT distribution in tandem with the diversity of Thaumarchaeota in the SPM of the Mediterranean outflow water sampled at the Portuguese margin at water depths of 100–1000 m. Similar to observations in the Mediterranean Sea itself, the relative abundances of both GDGT-2 and the crenarchaeol isomer increased with increasing water depth both in the core lipids (CLs) as well as in the intact polar lipids (IPLs; based on IPL-derived CL-GDGT analyses). Analysis of the genes encoding for the thaumarchaeotal ammonia monooxygenase (*amo*A) and an enzyme involved in the GDGT biosynthetic pathway

revealed that there was an increasing contribution of 'deep water-dwelling' Thaumarchaeota with increasing water depth that could potentially explain the increase in relative abundance of both GDGT-2 and the crenarchaeol isomer. The combination of IPL-derived CL-GDGT analysis with a DNA-based approach targeting Thaumarchaeota in Kim et al. (2016) proved to be extremely useful for discerning the contribution of different thaumarchaeotal populations, but a potential contribution of other archaeal sources was not addressed.

Here, we study the archaeal community in the Mediterranean Sea water column by using 16S rRNA gene amplicon sequencing of DNA extracted from the SPM obtained from three stations (western, middle and eastern). We combined the analysis of the archaeal community composition with GDGT analyses by high resolution accurate mass/mass spectrometry [UHPLC–HRMS] with the aim to determine the archaeal sources of the isoprenoid GDGT pool present and their potential role explaining the bias of the TEX_{86} in the Mediterranean Sea.

MATERIALS AND METHODS

Sampling and physicochemical analyses

SPM was collected in the Mediterranean during the NESSC cruises in January and February 2016 on board of the R/V *Pelagia*. SPM was sampled at three stations (western station; 36°44'59.9"N 1°29'59.4"W, middle station; 34°58'09.0"N 18°06'47.1"E and the eastern station; 33°18'08"N 33°23'43"E, respectively; further referred to as stations 1, 2 and 3; Figure 1). Samples were obtained from different water depths corresponding to different water masses in the Mediterranean Sea (Figure 1). The upper water mass, the Modified Atlantic Water (MAW) is formed as a result of mixing Atlantic water with surface waters of the Alboran Sea (Gascard and Richez, 1985). Below the MAW the more saline intermediate water mass, the Levantine Intermediate Water (LIW), is present, which is formed in the Levantine Sea (Wüst, 1961). The LIW spreads out across the Mediterranean Sea and is the major contributor of the Mediterranean outflow into the Atlantic. The Mediterranean Sea is also known to contain open-ocean deep convections (e.g.,

Houpert et al., 2016). These convective processes result in deep water mass formation in the Mediterranean Sea (e.g., Castellari et al., 2000). The Western Mediterranean Deep Water (WMDW) formed in the Gulf of Lions (MEDOC group, 1970) and the Eastern Mediterranean Deep Water (EMDW) formed of Adriatic Deep Water (Pollak, 1951; Artegiani et al., 1997) and the Aegean Deep Water (Roether et al., 1996).
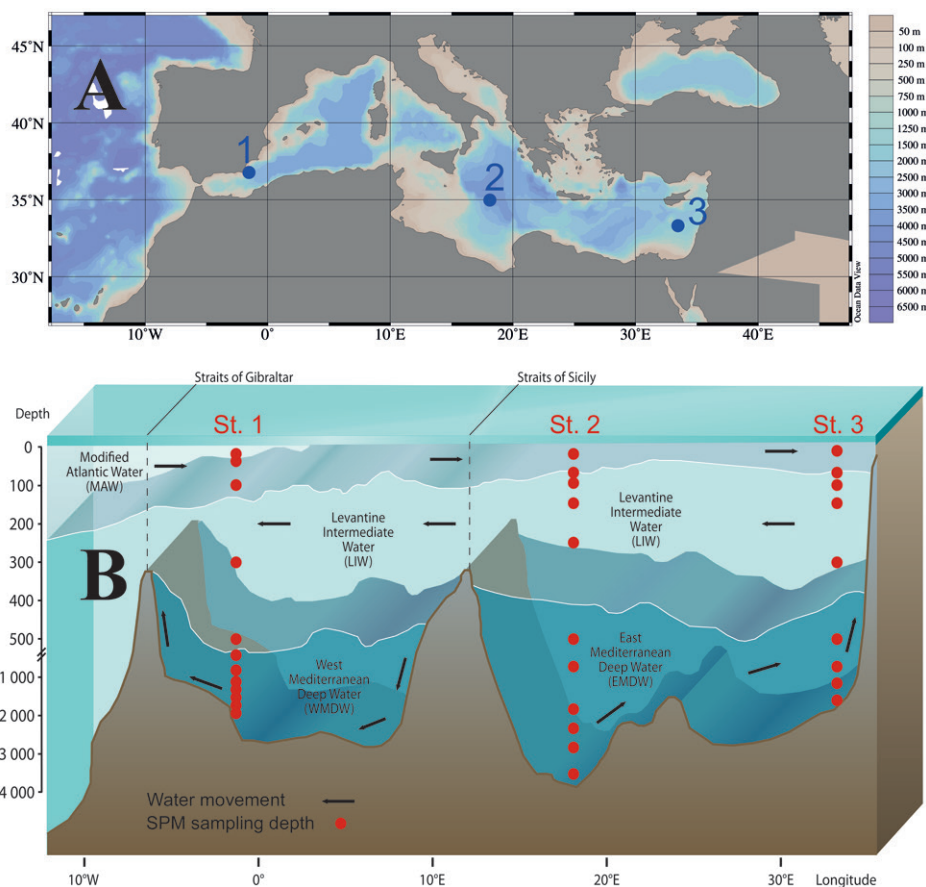
Figure 1. (A) Map of the Mediterranean Sea revealing the locations of the sampling stations. Water depth is color indicated by the bathymetry legend on the right. (B) Vertical distribution of Mediterranean Sea water masses. Modified from GRID-Arendal (http://www.grida.no/resources/5885) based on Hopkins (1985) and Zavaterelli and Mellor (1995). Note: depth axis is not to scale and note the break in the axis scale.

Three McLane WTS-LV in situ pumps (McLane Laboratories Inc., Falmouth, MA, USA) were deployed at various depths (Supplementary Table S1). Water (68–1545 L) was filtered using pre-ashed glass fiber (GF75) filters (Whatman, 0.3 μm pore size and 142 mm in diameter; Supplementary Table S1). Larger pore size (0.7 μm; GF/F) filters were also used at station 1 at 25, 40 and 100 meters below sea level (mbsl; Supplementary Table S1). The filters were immediately stored at -80 °C after retrieval of the pumps.

A Sea-Bird SBE911+ conductivity-temperature-depth (CTD) system equipped with a 24×12 L Niskin bottle rosette sampler was used to collect continuous profiles of temperature, salinity and dissolved oxygen (O_2), the latter using a Sea-Bird oxygen electrode (Sea-Bird, WA, USA). A Chelsea Aquatracka MKIII fluorometer was used to record fluorescence (Supplementary Table S1). Dissolved inorganic nutrients were analyzed in water from Niskin bottles sub-sampled with 60 mL high-density polyethylene syringes with a three-way valve and filtered over Acrodisc PF syringe filters (0.8/0.2 μm Supor Membrane, PALL Corporation) into pre-rinsed 5 mL polyethylene vials. Water samples for the determination of the dissolved phosphate (PO_4^{3-}) and inorganic nitrogen (NO_3^-, NO_2 and NH_4^+) concentrations were stored frozen at -20 °C until analyses upon return at NIOZ. Samples for dissolved reactive silicate (Si) analysis (Strickland and Parsons, 1968) were stored in the dark at 4 °C until analyses at NIOZ. The detection limits were PO_4^{3-} 0.007 μmol L^{-1}, NH_4^+ 0.03 μmol L^{-1}, NO_3^- + NO_2^- 0.01 μmol L^{-1} and NO_2^- 0.002 μmol L^{-1}.

Lipid extraction and analyses

Total lipids were extracted from SPM collected on 0.3 and 0.7 μm pore size glass fiber filters (samples taken at 25, 40 and 100 mbsl at station 1 were collected on 0.7 μm pore size glass fiber filters, the rest of the samples on 0.3 μm pore size glass fiber filters). After freeze-drying the filters were extracted using a modified Bligh and Dyer method as previously described by Schouten et al. (2008). C16-PAF (1-*O*-hexadecyl-2-acetoyl-*sn*-glycero-3-phosphocholine) was added as an internal standard after which the extracts were dried under a stream of nitrogen. The extracts with the added standard were then dissolved by adding solvent

(hexane:isopropanol:H$_2$O 718:271:10, v/v/v) and filtered through a 0.45 μm, 4 mm diameter True Regenerated Cellulose syringe filter (Grace Davison, Columbia, MD, USA).

IPLs were analyzed according to Besseling et al. (2018), a method based on Sturt et al. (2004). Briefly, samples were analyzed using an Ultimate 3000 RS UHPLC, equipped with thermostated auto-injector and column oven, couple to a Q Exactive Orbitrap MS with Ion Max source and heated electrospray ionization (HESI) probe (Thermo Fisher Scientific, Waltham, MA), was used. Separation was achieved on a YMC-Triart Diol-HILIC column (150 × 2.0 mm, 1.9 μm particles, pore size 12 nm; YMC Co., Ltd, Kyoto, Japan) with a guard column of the same material (10 × 2.1 mm). The same inclusion list was used as in Besseling et al. (2018). All samples were run in duplicate and the average IPL-GDGT distribution is reported, assuming identical response factors for all IPLs.

Peak areas for each individual IPL were determined by integrating the combined mass chromatogram (within 3 ppm) of the monoisotopic and first isotope peak of all relevant adducts formed (protonated, ammoniated and/or sodiated adducts may be formed in different proportions depending on the type of IPL). PAF was used to assess LC–MS performance and potential matrix effects. Absolute quantification of IPL GDGTs was not possible due to a lack of standards. IPLs of crenarchaeol and crenarcheol isomer are not separated and thus peak areas reported for any crenarchaeol IPL represents the sum of both isomers.

Bligh and Dyer extracts were dried under a stream of nitrogen and dissolved in a mixture of hexane:propanol (99:1, v/v) and filtered through 0.45 mm polytetrafluorethylene (PFTE) filters. GDGTs present in these extracts were analyzed on an Agilent 1260 UHPLC coupled to a 6130 quadrupole MSD in selected ion monitoring (SIM) mode, according to Hopmans et al. (2016). The injection volume was 10 μl and selected ion monitoring of the [M+H]$^+$ was used to detect the different GDGTs. C$_{46}$ GDGT was used as an internal standard to quantify GDGTs (Huguet et al., 2006b).

For sea surface temperature (SST) reconstructions the $\text{TEX}_{86}^{\text{H}}$ index was used as proposed by Kim et al. (2010). The $\text{TEX}_{86}^{\text{H}}$ index is a logarithmic function of the TEX_{86} index developed by Schouten et al. (2002):

$$\text{TEX}_{86}^{\text{H}} = \log \frac{[\text{GDGT}-2]+[\text{GDGT}-3]+[\text{cren}']}{[\text{GDGT}-1]+[\text{GDGT}-2]+[\text{GDGT}-3]+[\text{cren}']} \quad (1)$$

3

$\text{TEX}_{86}^{\text{H}}$ values were converted to SSTs applying the deep restricted basin core top calibration (composed of Mediterranean Sea (Kim et al., 2015) and the northern Red Sea (Trommer et al., 2009) surface sediments collected at > 1000 mbsl) as suggested by Kim et al. (2015):

$$\text{SST} = 56.3 \times \text{TEX}_{86}^{\text{H}} + 30.2 \quad (2)$$

Nucleic acids extraction and quantitative PCR (QPCR) analyses

DNA was extracted from the glass fiber filters with the RNA PowerSoil® Total Isolation Kit plus the DNA elution accessory (Mo Bio Laboratories, Carlsbad, CA). The concentration of DNA was quantified by Nanodrop (Thermo Scientific, Waltham, MA). Archaeal 16S rRNA gene copies were estimated by quantitative PCR (QPCR) using the following primers: Parch519F and ARC915R (archaeal 16S rRNA gene), as previously described (Pitcher et al., 2011). For details on the QPCR conditions, efficiency and R^2 of the QPCR assays see Supplementary Table S2.

16S rRNA gene amplicon sequencing, analysis, and phylogeny

PCR reactions were performed with the general prokaryotic primers S-D-Arch-0159-a-S-15 and S-D-Bact-785-a-A-21 (Klindworth et al., 2013) using 454 GLX sequencing with a single-ended approach as described in Moore et al. (2015). The archaeal 16S rRNA gene amplicon sequences were analyzed by QIIME v1.9 (Caporaso et al., 2010). Raw sequences were demultiplexed and then quality filtered with a minimum quality score of 25, length in the range 250–350, and allowing

up to two mismatches in the barcode sequence. Taxonomy was assigned based on blast and the SILVA database version 128 (Altschul et al., 1990; Quast et al., 2013). Representative operational taxonomic units (OTUs; based on 97% sequence identity) sequences of archaeal groups were extracted through filter_taxa_from_otu_table. py and filter_fasta.py with QIIME (Caporaso et al., 2010). OTUs comprised of < 50 16S rRNA gene amplicon reads were discarded. The phylogenetic affiliation of the partial archaeal 16S rRNA gene sequences was compared to release 128 of the Silva NR SSU Ref database (http://www.arb-silva.de/; Quast et al., 2013) using the ARB software package (Ludwig et al., 2004). Sequences were added to the reference tree supplied by the Silva database using the ARB Parsimony tool. Affiliation of any 16S rRNA gene sequences to a given subgroup was done assuming a similarity cutoff of ≥ 85%.

The water column was not evenly sampled in our study (with a higher density of sampling points in the (sub)surface in comparison to the deeper part of the water column; Figure 1). In order to give a more realistic estimation of abundances of the archaeal OTUs in the Mediterranean water column, the archaeal copy number (16S rRNA gene copies L^{-1}) at a certain water depth was multiplied by the relative abundance of that specific OTU (in % of all archaeal reads). To minimize the bias of sampling resolution, the meters of the water column covering the distance between sampling points were multiplied by the abundance. The estimated abundances were then divided by the water column depths to get depth-integrated estimated abundances per OTU. These depth-integrated estimated abundances were used to rank the OTUs within each archaeal group, i.e. MGI, MGII and marine euryarchaeota group III (MGIII) and named accordingly (i.e. OTU-1 represents the most abundant within an archaeal group, OTU-2 second most, etc.). The three MG compositions, based on the depth-integrated abundances, at the three different stations were compared by Spearman's rank correlation coefficient analyses.

Clone library, sequencing and phylogenetic analyses of ammonia monooxygenase (amoA) coding gene fragments

The *amo*A gene fragments were amplified as described by Yakimov et al. (2011). The PCR mixture was the following (final concentration): Q-solution 1 × (PCR additive, Qiagen, Valencia, CA, USA); PCR buffer 1× ; BSA (200 μg ml^{-1}); dNTPs (20 μM); primers (0.2 pmol μl^{-1}); MgCl$_2$ (1.5 mM); 1.25 U Taq polymerase (Qiagen). The PCR conditions for these amplifications were the following: 95 °C, 5 min; 35 × (95 °C, 1 min; 55 °C, 1 min; 72 °C, 1 min); final extension 72 °C, 5 min. The PCR products were gel-purified (QIAquick gel purification kit, Qiagen) and cloned in the TOPO-TA cloning® kit from Invitrogen (Carlsbad, CA, USA) and transformed in *Escherichia coli* TOP10 cells following the manufacturer's recommendations. Recombinant clones plasmid DNAs were purified by Qiagen Miniprep kit and screening by sequencing (n = 288) using M13R primer by BaseClear (Leiden, The Netherlands). Obtained archaeal *amo*A protein sequences were quality checked and aligned with already annotated *amo*A sequences by using the ClustalW Multiple alignment (Thompson et al., 1994): this resulted in n = 244 good quality sequences. The phylogenetic trees were constructed with MEGA7 (Kumar et al., 2016) with the neighbor-joining method (Saitou and Nei, 1987) with a bootstrap test of 1000 replicates. Phylogenetic trees were drawn using iTOL (Letunic and Bork, 2016).

Data availability

The CL- and IPL-GDGT data have been submitted to Pangea. Sequence data of the partial 16S rRNA gene sequences have been submitted to the BioSample database under accession no: SAMN12106735–SAMN12106766. Sequence data of partial *amo*A gene sequences have been submitted to the GenBank database under accession no: MK592163–MK592407.

RESULTS

Physicochemistry of the water column

At stations 1, 2 and 3 (western, middle and eastern Mediterranean Sea; respectively; Figure 1) water column SPM samples were taken at > 10 different water depths, ranging between 25 and 3530 mbsl, during the winter of 2016. The water depth at these stations varied greatly (1963, 3609 and 1720 mbsl, respectively) due to the topography of the basin. The water column was fully oxygenated at all stations (oxygen levels of 172-252 µmol kg^{-1}) at the time of sampling. Salinity ranged between 37.9 practical salinity units (psu) and 39.1 psu (Supplementary Table S1). Ammonia concentrations declined with increasing depth at stations 1 and 3 (from 0.13 µmol L^{-1} at the surface to 0.10 µmol L^{-1} at the greatest depth and from 0.20 to 0.15 µmol L^{-1}, respectively), while in station 2 the levels increased from 0.10 to 0.18 µmol L^{-1} (Supplementary Table S1). Temperatures in the water column were in the range 13.1–14.1 °C at station 1, 13.9-17.2 °C at station 2 and 13.8-19.0 °C at station 3 (Supplementary Table S1). Concentration maxima of nitrite (NO_2^-) were detected at 40, 90 and 100 mbsl at stations 1, 2 and 3, respectively (Supplementary Table S1).

Archaeal abundance and diversity

Archaeal 16S rRNA gene abundance was estimated by quantitative PCR using general archaeal primers. At the surface (25 mbsl) of station 1, the copy number was 1.1×10^8 gene copies L^{-1} and increased sharply to 6.3×10^8 gene copies L^{-1} at 100 mbsl (Figure 2). Maximum abundance was reached at 500 mbsl (8.3×10^8 gene copies L^{-1}), with a subsequent decline with increasing depth to 3.5×10^7 gene copies L^{-1} at 1300 mbsl (Figure 2). At 1500 mbsl, another maximum was detected with 1.9×10^8 gene copies L^{-1}, and subsequently a decrease to 4.6×10^7 gene copies L^{-1} at 1900 mbsl (Figure 2). At the surface (25 mbsl) of station 2, archaeal 16S rRNA gene abundance values were 5.4×10^7 gene copies L^{-1} and reached a maximum at 150 mbsl (3.4×10^8 gene copies L^{-1}). With increasing depth, values decreased to 1.7×10^7 gene copies L^{-1} at 1800 mbsl (Figure 2). The abundance declined further to 7.7×10^5 gene copies L^{-1} at 3530 mbsl (Figure 2). At station 3, the abundance at the

surface (25 mbsl) was 4.6×10^7 gene copies L^{-1}, which was slightly higher than at 75 and 100 mbsl (Figure 2). Maximum abundance was detected at 300 mbsl with 4.3×10^8 gene copies L^{-1}. It then declined to 2.3×10^5 gene copies L^{-1} at 1300 mbsl and sharply increased to 6.6×10^7 gene copies L^{-1} at 1700 mbsl (Figure 2).
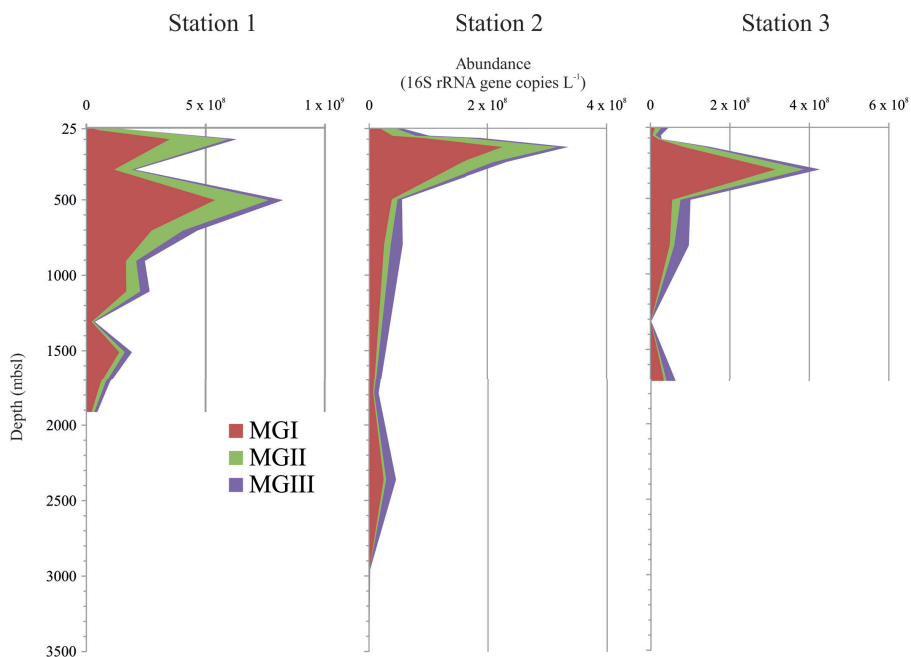
Figure 2. Archaeal abundance profiles showing the absolute abundance (in archaeal 16S rRNA genes copies L^{-1} as determined by Q-PCR) for the three stations. Archaeal community composition for the three major groups of archaea is also given and based on 16S rRNA gene amplicon sequencing.

Based on the results of the 16S rRNA gene amplicon sequencing approach, the archaeal 16S rRNA gene reads were predominantly closely related to three archaeal lineages, the MGI Thaumarchaeota and MGII and III Euryarchaeota (Figure 2). Minor groups, representing < 4% of all archaeal reads, i.e. the Woesearchaeota from the superphylum DPANN (Rinke et al., 2013; Castelle et al., 2015) and marine benthic group A (MBG-A), were also detected. MGI was the major archaeal group for most of the water column (100 mbsl and deeper waters) representing 39–74% of all archaea (Figure 2; Supplementary Table S6). In contrast, in surface waters,

which contain an order of magnitude lower concentrations of archaeal cells, they represented only 15–33% of all archaea of the three stations. The surface waters (25–40 mbsl) at stations 1 and 2 were dominated by MGII (47–84%; Figure 2 and Supplementary Table S6), while in surface waters (25 mbsl) of station 3 MGIII was the most abundant archaeal group (40%; Figure 2). With increasing water depth, the relative abundance of MGII decreased to 24% (1900 mbsl), 15% (3530 mbsl) and 10% (1700 msbl) at stations 1, 2 and 3, respectively (Figure 2). MGIII was less dominant at the subsurface (100–500 mbsl) compared to the surface of stations 2 and 3 with 3–25% relative abundance, MGIII was more dominant at the subsurface of station 1 (up to 7%) compared to the surface (Figure 2). Relative abundances of MGIII increased at all three stations with increasing depth to 27%, 41% and 35% at stations 1, 2 and 3, respectively (Figure 2).

Based on the 16S rRNA gene amplicon sequencing approach, 30 operational taxonomic units (OTUs; based on $\geq$ 97% sequence identity and with a > 50 gene read cut-off) belonging to MGI were detected; these OTUs were present at all three stations (Supplementary Table S3). The five most abundant MGI OTUs in our dataset (estimated by multiplying the relative abundance (%) of reads for an OTU with archaeal 16S RNA abundance; see Materials and methods for details) together accounted for 67% of the estimated depth-integrated abundance of the 30 MGI OTUs detected. These OTUs differed in their niches in the water column as revealed in the heatmap of Figure 3. Heatmaps for all 30 OTUs are shown in Supplementary Figure S2 for each station. OTU-1, the overall most abundant OTU in our dataset (Supplementary Figure S2), dominated in the bottom waters (for stations 1, 2 and 3 at 1900, 1800 and 1700 mbsl, respectively; Figure 3 and Supplementary Figure S2), while OTU-5 was predominant at the subsurface depths 25, 90 and 75 mbsl (Figure 3 and Supplementary Figure S2). Correlation between estimated abundances for the 30 OTUs at each station, based on Spearman's rho rank correlation coefficients (r_s), showed strong correlations between the MGI community composition of station 1 and of station 2 (r_s = 0.80, ρ < 0.001) and station 3 (r_s = 0.74,

ρ < 0.001; Supplementary Table S3). MGI community composition between stations 2 and 3 were also highly correlated (r_s = 0.89, ρ < 0.001; Supplementary Table S3).

3
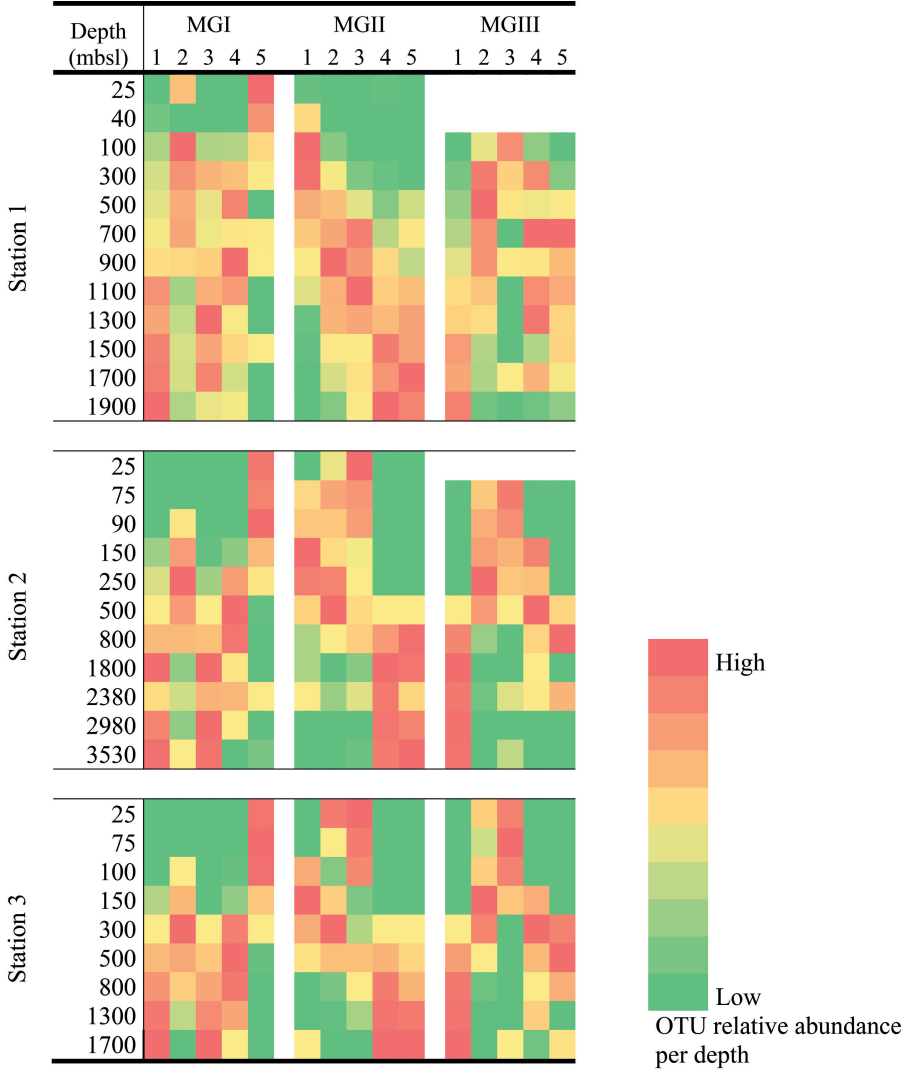
Figure 3. Heatmap of the five most abundant operational taxonomic units (OTUs; based on ≥97% 16S rRNA gene similarity) of the three main archaeal groups (MGI–III) per station. Heatmap colors are indicative of the relative abundance per OTU per marine archaeal group (green=low relative abundance, red=high). MGIII OTUs (1–5) at station 1 (25 and 40 mbsl) and at station 2 (25 mbsl) are not shown because of low read numbers (<20 sequence reads). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The phylogeny of the sequences of the MGI OTUs (Figure 4) revealed that only a minor proportion (3 out of 30) of the OTUs were related to most cultivated MGI Thaumarchaeota (e.g., *Nitrosopumilus maritimus*) falling in cluster I, while the majority of OTUs were related to uncultured MGI species falling in clusters II and III (Figure 4). The uncultured MGI OTUs are quite diverse and in many cases closely related to other marine uncultured MGI 16S rRNA gene sequences, often retrieved from greater water depths. The diversity of MGI OTUs was further investigated by sequencing the thaumarchaeotal *amo*A gene, encoding the alpha-subunit of the enzyme ammonia monooxygenase. We amplified and sequenced a fragment of the *amo*A gene by PCR at three different depths at each station (one in epipelagic, one in mesopelagic and one in bathypelagic waters; for station 1: at 100, 500 and 1500 mbsl, for station 2: at 150, 800 and 2380 mbsl, for station 3: at 25, 300 and 1700 mbsl). Phylogeny and classification was based on a recent *amo*A gene database (La Cono et al., 2018). Throughout the water column, diverse MGI communities were detected and classified as 'shallow' and 'deep' water type based on their *amo*A gene sequence (Figure 5) and amino acid differences (e.g., Supplementary Figure S8) (cf. Francis et al., 2005; Hallam et al., 2006; Mincer et al., 2007; Beman et al., 2008). We detected a clear shift from dominance of the 'shallow water' *amo*A type to predominance of the 'deep water' *amo*A type with increasing depth at all three stations (Figure 5). The 'shallow water' group of *amo*A gene sequences contained 2 subclusters and the 'deep water' group contained 4 subclusters (Supplementary Figure S7A and B; cf. La Cono et al., 2018; Sintes et al., 2016). A minor proportion of *amo*A gene sequences predominantly present in the upper surface water (25 mbsl, obtained at station 3) grouped into subcluster 1 (Supplementary Figure S7A), comprising *amo*A gene sequences of surface dwelling *Nitrosopumilus* (i.e. *Candidatus* N. adriaticus and *Ca.* N. piranensis; Bayer et al., 2016), as well as *amo*A gene sequences of *Ca.* Nitrosopumilus koreensis and *Ca.* N. sediminis isolated from marine sediments (Park et al., 2014). Most of the 'surface-dwelling' MGI Thaumarchaeota detected in the epipelagic zone were grouped into subcluster 2 (Supplementary Figure S7A). This subcluster contains *amo*A sequences from all three stations and from all depths (except from 1700 mbsl at station 3; Figure 5). Most of the obtained *amo*A gene sequences were

clustered in the 'deep water' group, with the majority grouped into subcluster 3, lower relative abundances were clustered in subgroup 4, 5 and 6 (Supplementary Figure S7B). As the 'deep water' group only contains uncultured Thaumarchaeota, their *amo*A sequences cannot be linked to cultured Thaumarchaeota as observed for the MGI 16S rRNA gene sequences.
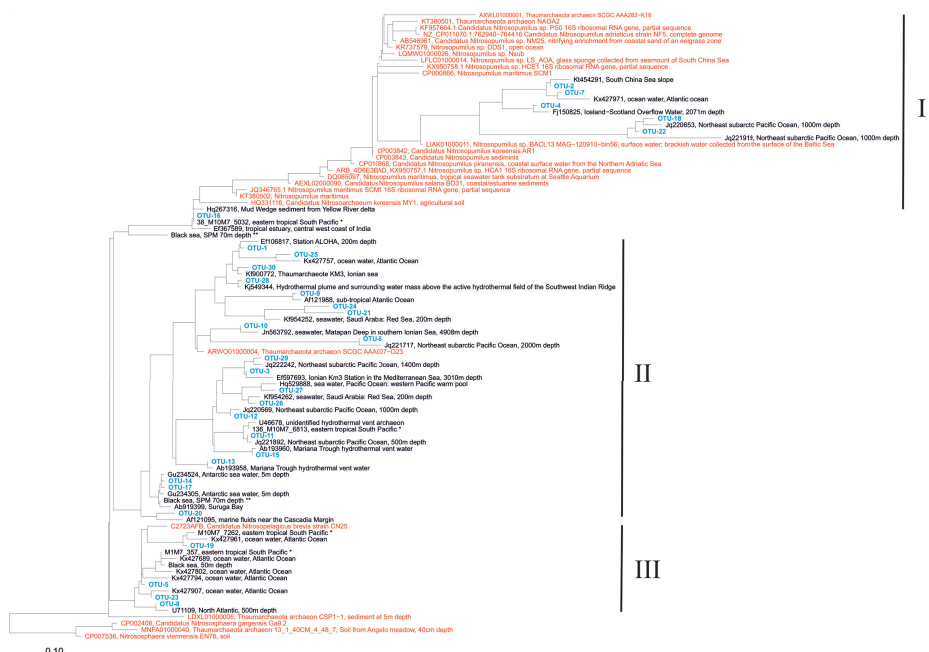
Figure 4. Phylogenetic tree composed of 16S rRNA gene amplicon sequences of marine group I (MGI) operational taxonomic units (OTUs; in blue) detected in suspended particulate matter (SPM) of the Mediterranean Sea. Closely affiliated sequences from uncultured MGI members (in black) and from MGI cultures and from with metagenomics assembled MGI environmental genomes (both in red) are also included. Three distint subgroups are indicated which were used for the comparison with other archaeal subgroups and intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) distributions, shown in Table 1. Scale bar represents a 10% sequence dissimilarity.
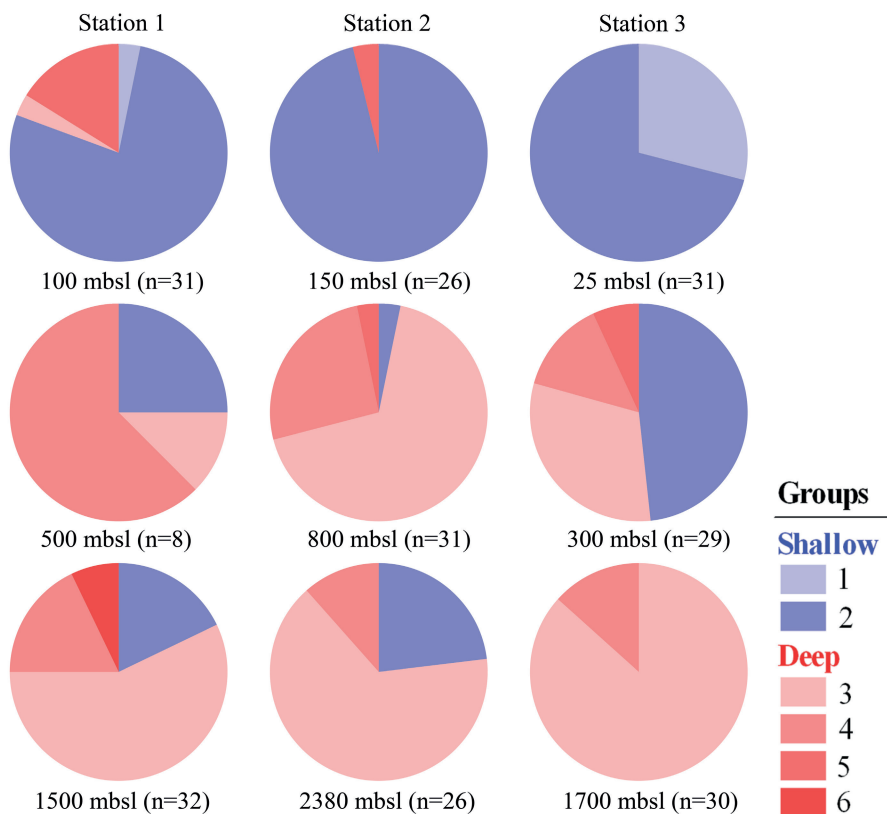
Figure 5. Pie charts showing the fractional abundance of specific groups of sequences of the α-subunit of the ammonia monooxygenase (*amo*A) gene from clone libraries obtained at three different depths at all three stations. Groups are clustered as 'shallow' *amo*A gene (indicated in blue; groups 1 and 2) and 'deep' *amo*A gene (indicated in red; groups 3–6) types. Colors resemble the groups as indicated in the phylogenetic trees (Supplementary Fig. S1A and S1B). Sampling depth is given underneath the pie charts with the number (n=) of sequences that were obtained in the clone libraries.

Thirteen MGII OTUs (again defined on the basis of ≥ 97% sequence identity) were identified at the three stations (Figure 3 and Supplementary Figure S3). OTU-1 of MGII (the most dominant MGII OTU overall) was most abundant within the bathypelagic zone (> 1000 mbsl) for all three stations (Figure 3). OTU-2 was most dominant at the subsurface (100–300 mbsl) for all three stations (Figure 3). OTU-3 was dominant (based on the relative abundance) in surface waters of stations 2 and 3 but was mostly detected at greater depths (500–1300 mbsl) at station 1 (Figure 3). MGII communities between the three stations were compared by correlating

the estimated abundances of the 19 OTUs per station. The MGII community composition was only weakly correlated between station 1 and stations 2 and 3 (r_s = 0.60, ρ < 0.05 and r_s = 0.49, ρ < 0.1, respectively; Supplementary Table S3). However, between station 2 and 3, the MGII community composition showed a stronger correlation (r_s = 0.87, ρ < 0.001; Supplementary Table S3).

Six MGIII OTUs were detected in the water column of the three stations (Figure 3 and Supplementary Figure S4). OTU-1 of MGIII was detected mostly within the bathypelagic zone (> 1000 mbsl) at all three stations (Figure 3). The highest relative abundance of MGIII OTU-2 was detected at the subsurface (250 and 150 mbsl for stations 2 and 3, respectively) or within the mesopelagic zone (500 mbsl at station 1; Figure 3). OTU-3 was most dominant in surface waters (0–100 mbsl) at all three stations (Figure 3). A significant correlation, based on Spearman's rho rank correlation coefficients, of the estimated abundances of the six MGIII OTUs was observed between the three stations (r_s = 0.89, ρ < 0.05; Supplementary Table S8).

Archaeal lipid abundance and distribution in the water column

The archaeal intact polar lipids (IPLs) consisted of GDGT-0 to GDGT-4 and crenarchaeol with monohexose (MH), dihexose (DH) and hexose–phosphohexose (HPH) headgroups (Figure 6). Other archaeal membrane IPLs (including archaeol IPLs) were targeted in the UHPLC–HRMS analysis but not detected (for a list of targeted IPLs, see Besseling et al., 2018). The IPL-GDGTs were detected in all SPM samples with the exception of the surface waters of station 2 (25 and 75 mbsl) and at 75 mbsl of station 3 (Figure 6). Maximum concentrations of IPL-GDGTs (expressed in response units L^{-1}) were detected in subsurface waters (i.e. at stations 1 and 3 at 300 mbsl and at station 2 at 250 mbsl) (Figure 6). For station 1, only HPH-GDGTs with GDGT-0 and crenarchaeol core lipid were detected (53% and 47% relative abundance, respectively; Figure 6) at 25 mbsl. At 40 mbsl, the headgroup DH was also observed in combination with GDGT-1, GDGT-2 and crenarchaeol (Figure 6). At 100 mbsl, the majority of IPL-GDGTs detected had the headgroup DH but also IPL-GDGTs with the headgroup MH, in combination with CLs GDGT-0 and crenarchaeol (Figure 6) were detected. For station 2, IPLs with GDGT-0,

GDGT-2 and crenarchaeol were detected with the headgroups DH (DH–GDGT-2) or with HPH (HPH–GDGT-0 and HPH–crenarchaeol) at 90 mbsl (Figure 6). At 150 mbsl, the majority of IPLs contained a DH headgroup and this remained the dominant headgroup with increasing depth (Figure 6). At station 3, IPL-GDGTs were detected at 25 mbsl but not at 75 mbsl. The IPLs at 25 mbsl consisted mostly of the headgroup HPH in combination with GDGT-0 and crenarchaeol. A minor abundance of DH GDGT-2 was also observed. Generally, with increasing depth, the relative abundance of HPH-IPLs decreased and DH-IPLs increased (Figure 6).
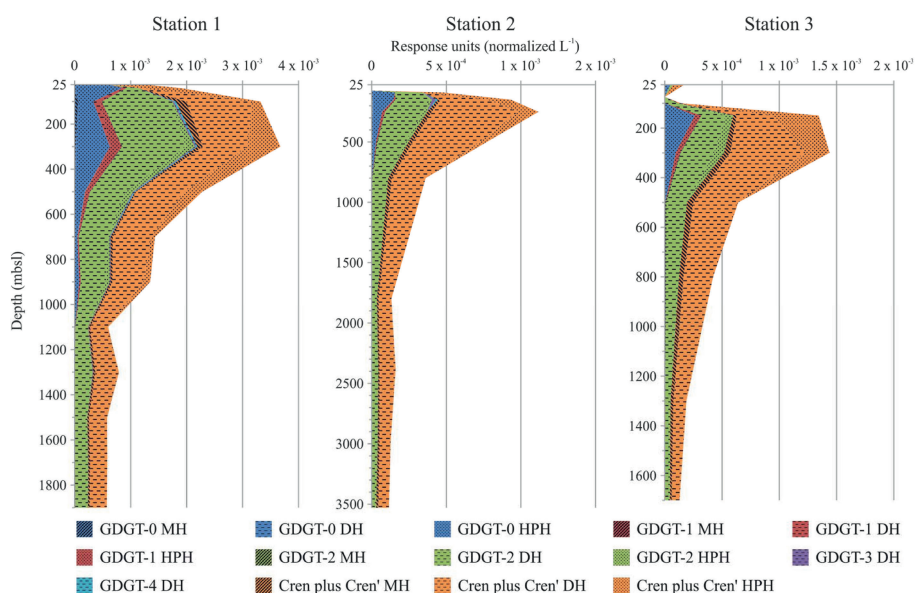
Figure 6. Depth profiles of the concentration (indicated by the response units per L) of the most abundant intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) for the three stations. The overall profile shows the sum of the IPLs.

The summed concentration of core lipid (CL-) GDGTs (i.e. GDGTs present without a head group) was on average highest at station 1 (ranging from 70 to 820 pg L^{-1}) with a maximum at 100 mbsl (Figure 7). Summed concentrations of CL GDGTs in stations 2 and 3 ranged from 20 to 160 pg L^{-1} and 30 to 190 pg L^{-1}, respectively, (Figure 7) with maxima at 250 and 300 mbsl, respectively (Figure 7). After the maximum in the concentration profile, all stations showed a steady decrease in CL GDGT concentrations with increasing depth (Figure 7). CL archaeol was

detected in concentrations < 5% of the summed GDGTs. The distribution of the CL-GDGTs (Figure 8) showed less variation over the water column than the GDGT core distribution of the IPLs (see above). Crenarchaeol was the most abundant GDGT with a relative abundance ranging from 50 to 59%, followed by GDGT-0, ranging from 14 to 32% (Figure 8). Other detected GDGTs were GDGT-1 to GDGT-4 and cren′ (Figure 8). GDGT-3 and GDGT-4 were generally only minor compounds, with the exception of GDGT-3 at 75 mbsl at station 3 (12%). The relative abundance of GDGT-2 increased with greater depth at stations 1 and 3, up to 300 mbsl. However, this increase was much less apparent at station 2 (Figure 8). GDGT-1 relative abundances increased at stations 2 and 3, up to 250 and 300 mbsl, respectively (Figure 8). Cren′ relative abundances increased up to 250 mbsl at station 2 and up to 300 mbsl at stations 1 and 3, but for station 1 this increase was less apparent (Figure 8). The relative abundance of GDGT-0 decreased with greater increasing depth (up to 250 mbsl at station 2 and up to 300 mbsl at stations 1 and 3) and increased with greater depth below 500 mbsl at station 2; this increase was not observed at the other stations (Figure 8).

Surface sediments (0–1 cm) underlying the sampled stations were also analyzed for their CL-GDGT distribution. Sedimentary CL GDGTs were dominated by crenarchaeol followed by GDGT-0 at all three stations. The distribution of CL GDGTs in the surface sediments highly correlates ($r \geq 0.99$) with the CL GDGTs composition of the overlying water column.
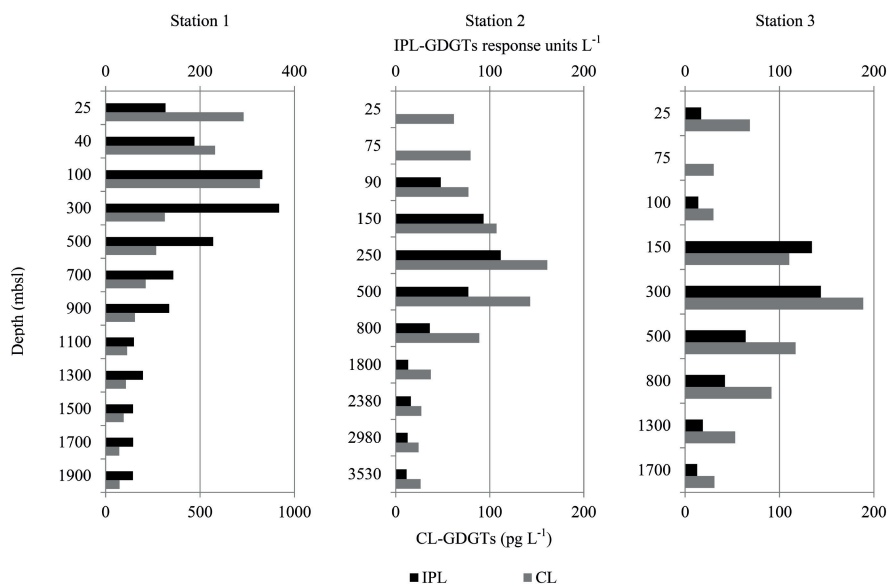
Figure 7. Absolute abundance profiles of intact polar lipid (IPL-) and core lipid (CL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) abundance in the water column of the three stations. The IPL-GDGT abundance (black bars) is indicated in response units per L (upper axis), CL-GDGT abundance (grey bars) is indicated in pg L^{-1} (lower axis).
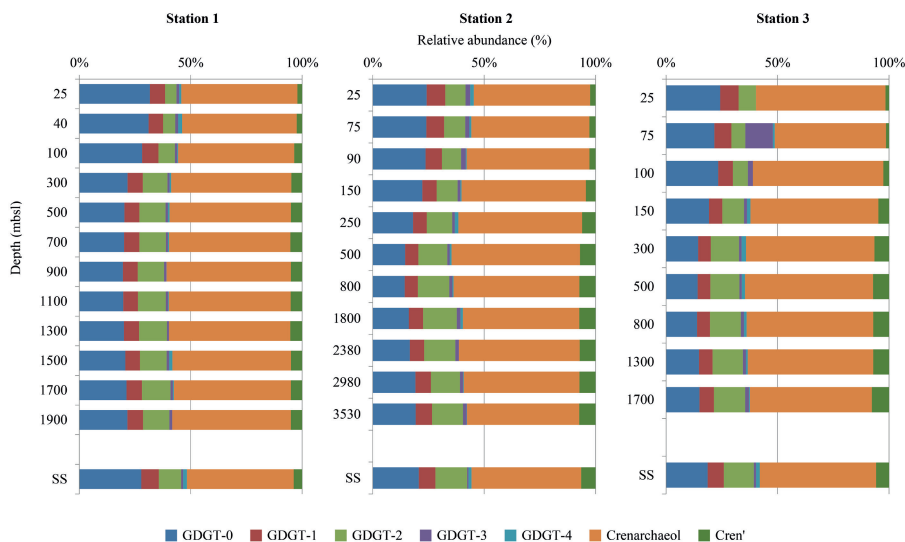
Figure 8. Depth profiles of the relative abundance of seven detected core lipid (CL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) for the three stations and the CL-GDGT composition of the underlying surface sediments (SS).

TEX_{86}^H temperature reconstructions

TEX_{86}^H values were calculated based on the CL GDGTs in the surface sediments collected below the studied water columns. At station 1, the TEX_{86}^H value was -0.187±0.001 (standard deviation, obtained by analyzing three different surface sediment samples per station). This resulted in a TEX_{86}^H temperature of 19.7 °C (calculated using the specific Mediterranean calibration of Kim et al., 2015). At station 2, the TEX_{86}^H value was -0.131±0.001, which corresponds to a TEX_{86}^H temperature of 22.8 °C. At station 3, the TEX_{86}^H value was -0.132±0.002, with a TEX_{86}^H temperature of 22.8 °C.

3

DISCUSSION

A diverse archaeal community composition throughout the water column

In this study, we observed distinctly different distributions of the archaeal groups MGI, II and III throughout the Mediterranean water column. In our study, the same dominant archaeal OTUs were detected with fairly similar distributions among the three stations (Figure 3; Supplementary Figs. S2–S4), despite the vast horizontal distance between them, suggesting that the distribution of these OTUs at our three stations can be taken as representative of the archaeal diversity and distribution throughout the Mediterranean Sea. In agreement with our work, De Corte et al. (2009) showed that the archaeal community composition was highly stratified with depth in the eastern Mediterranean Sea and associated with different water masses. Galand et al. (2010) reported a strong seasonality of the archaeal community composition within shallow coastal NW Mediterranean Sea water, with a dominance of Euryarchaeota in winter. Our samples were also obtained in winter (January and February), and within surface waters (25 mbsl) either MGII (stations 1 and 2) or MGIII (station 3) Euryarchaeota were dominant (Figure 2 and Supplementary Figure S4), in agreement with the study of Galand et al. (2010). The dominance of Euryarchaeota within the surface waters of the Mediterranean Sea has also previously been observed (Tamburini et al., 2009; Zhou et al., 2018). This

is compatible with their potential metabolism, as determined from metagenomics, which suggests that they might be able to use light-derived energy and cope well with UV radiation stress (Frigaard et al., 2006; Iverson et al., 2012; Martin-Cuadrado et al., 2014). From 100 m downwards, MGII and MGIII Euryarchaeota still contribute a substantial part of the total archaeal population with values ranging from 27 to 56% of the total archaea (Figure 2). Certain OTUs of the MGII and MGIII Euryarchaeota show, however, a preference for deep waters (> 1500 mbsl), i.e. OTU-1 (Figure 3). Deep-water samples collected in this study (from 500 m downwards) correspond to the West Mediterranean Deep Water (WMDW, station 1), and the East Mediterranean Deep Water (EMDW, stations 2 and 3) water masses. These stations are all characterized by relatively high bottom water temperatures (e.g., 13–14 °C; Supplementary Table S1), which remain > 12.5 °C year round (Wüst, 1961), which is considerably warmer than deep water temperatures in the global ocean (Sverdrup et al., 1942).

The higher irradiance and UV-induced stress in surface waters, which favors the Euryarchaeota, is believed to have a negative influence on the surface water abundance of MGI Thaumarchaeota, which are known to have a niche preference for subsurface waters (approximately 50–500 m depth, (Massana et al., 2000; Pitcher et al., 2011b; Beman et al., 2012; Lincoln et al., 2014a; Sollai et al., 2018). Indeed, our study shows that MGI archaea have a much higher abundance (both relative and absolute) in subsurface waters (100–300 mbsl) compared to the surface (0–100 mbsl), but are also found up to 500 m depth. This is different from the North Atlantic where the maximum abundance of Thaumarchaeota is typically located at around 100 m depth and then consistently decreased with increasing depth (Herndl et al., 2005). Thus, in the Mediterranean there exists an expanded, deeper niche for MGI Thaumarchaeota, which coincides with the Levantine Intermediate Waters (LIW) and the underlying deep water masses in the Mediterranean Sea water column (Figure 2). Indeed, Techtmann et al. (2015) already reported MGI Thaumarchaeota as dominant members of the prokaryotic community within the LIW and the Eastern Mediterranean Deep Water (EMDW) and concluded

that their high relative abundance was indicative of the LIW. The LIW which is flowing from the eastern to the western basin at approximately 100–600 m depth is characterized by a (relatively) high salinity and temperature (38.5–39.1 psu and 13.3–18.9 °C, respectively) in comparison with similar depths in other ocean basins. These could be the physicochemical factors favoring an increased abundance of MGI Thaumarchaeota in these deeper water masses.

Specific MG1 OTUs, i.e. OTU-5, were present from the surface to approximately 500 m depth, comprising both the surface Modified Atlantic Water (MAW) as well as the LIW (Figure 3). The 16S rRNA gene sequences of these OTUs are closely related to thaumarchaeotal 16S rRNA gene sequences previously detected in the Black Sea surface (Sollai et al., 2018) and in the Atlantic Ocean (Bergauer et al., 2018) (Figure 4). The close phylogenetic association of these thaumarchaeotal OTUs with thaumarchaeotal 16S rRNA gene sequences previously detected in marine waters with high salinity and temperature regimes suggests that these OTUs might be thaumarchaeotal species better adapted to these conditions. Specific thaumarchaeotal MGI OTUs were preferentially found in deep waters (e.g., OTU-1 and OTU-2; Figure 3), indicating a clear niche difference for various thaumarchaeotal OTUs. Changes in the diversity of MGI Thaumarchaeota throughout the water columns are also evident from the observed variations in ammonia monooxygenase (*amo*A) gene sequences, which have been previously used to classify Thaumarchaeota as 'shallow' and 'deep water' type (e.g., Francis et al., 2005; Beman et al., 2008; Yakimov et al., 2011; Sintes et al., 2013). In the Mediterranean Sea, the shallow *amo*A cluster dominated at the surface (25 mbsl, station 3) and subsurface (100 and 150 mbsl, stations 1 and 2, respectively; Figure 5), with sequences closely related to sequences previously detected in, for example, the Mediterranean (accession number MF662834 in Figure 6; La Cono et al., 2018), Antarctic coastal waters (EU239020; Kalanetra et al., 2009) and Monterey Bay, California (DQ148839; Francis et al., 2005). In the deeper waters (> 300 mbsl) *amo*A gene sequences falling in subclusters 3–6 dominated; these subclusters are affiliated to those of deep-water Thaumarchaeota (Sintes et al., 2016; La Cono et al., 2018; Supplementary Figure S7).

Little is known about the physiology of the deep-water MGI Thaumarchaeota since there are currently no cultured representatives. Their niche has been previously linked to much reduced ammonium concentrations in deep waters requiring distinct *amo*A genes (Sintes et al., 2016; Supplementary Figure S8). An adaptation of the ammonia monooxygenase enzyme could increase the affinity towards ammonia or even broaden the spectrum of available substrates as proposed by Smith et al. (2016). However, the Mediterranean Sea is characterized by (ultra)oligotrophic conditions (Krom et al., 1991) and the ammonia concentrations are low throughout the water column (e.g., at station 1: 0.02-0.13 µmol L^{-1}; Supplementary Table S1). It could be that the high affinity for ammonia by Thaumarchaeota is causing these low ammonia concentrations. However, this conversion of ammonia likely would have resulted in an increase of nitrite concentrations in the water column which was not observed (Supplementary Table S1). Therefore it is hard to imagine that the observed niche differentiation of Thaumarchaeota in the Mediterranean Sea is actually linked to differences in ammonia affinity attributed to the deep-water Thaumarchaeota.

Although the deep-water masses West Mediterranean Deep Water (WMDW) and East Mediterranean Deep Water (EMDW) originate from different locations, the deeper water MGI community compositions were similar (Figure 3); i.e., in the three stations the dominant Thaumarchaeotal MGI OTU was OTU-1 (> 50% average relative abundance in the deeper water masses; Supplementary Table S2). The physical properties of the WMDW and the EMDW were similar at the sampling locations and during the time of sampling (Supplementary Table S1), including NH_4^+ concentrations, suggesting the Thaumarchaeota MGI OTUs found in the deep water basins of the Mediterranean Sea are also adapted to this niche, independent of the original provenance of the water mass. Time of sampling in this case, even for the deep-water masses, was probably a determining factor for the archaeal community composition. Winter et al. (2009) also showed that the deep archaeal community was affected by vertical mixing in the Mediterranean Sea and that this seasonal effect could be as dynamic as observed at the surface.

Archaeal lipid composition in the water column of the Mediterranean Sea

In our study, we screened for a wide range of known and potential head group combined with archaeal core lipids (see Besseling et al. (2018) for a list of targeted compounds). However, only the IPLs with head groups MH, DH and HPH in combination with different CL-GDGTs were detected (Figure 6). The concentration of these summed IPLs peaked at all three stations at 250–300 mbsl and were substantially lower in surface (0–100 m; except for Station 1) and deeper (> 500 m) waters (Figure 7), generally in good agreement with the depth distribution of the abundance of archaeal copy numbers (Figure 2). No archaeal IPLs were detected at 25 and 75 mbsl at station 2 and at 75 mbsl at station 3, which is probably due to the analytical detection limits. A lower abundance of most IPL-GDGTs in surface (< 80 mbsl) than in deeper waters has been previously observed in the Black Sea (Schubotz et al., 2009) and in the Western Atlantic Ocean (Hurley et al., 2018).

The IPL-GDGTs in the surface waters at 25–40 mbsl in station 1 and at 25 m at station 3 were mostly dominated by GDGT-0 and crenarchaeol carrying the HPH headgroup. HPH crenarchaeol has been hypothesized as a biomarker for living (Pitcher et al., 2011a; 2011b; Buckles et al., 2013) and active Thaumarchaeota (Elling et al., 2014). MGI DNA was detected at these depths, but in much lower amounts compared to the subsurface (Figure 2). However, the dominance of HPH crenarchaeol suggest that these GDGTs are produced by present, albeit less abundant, MGI cells. With increasing depth (i.e. > 40 mbsl, through the MAW, the LIW and the deepest water masses), the relative abundance of DH IPLs increased, especially for the IPLs with GDGT-2 and crenarchaeol and its isomer as CLs (Figure 6). This may be due to a preferential synthesis of glycosidic IPL-GDGTs due to a physiological adaptation of the producers. Elling et al. (2014) showed differences in the IPL-GDGT composition in different growth phases of the thaumarchaeon *Nitrosopumilus maritimus* with HPH-GDGTs preferentially produced in the early growth phase compared to later growth phases. Furthermore, *N. maritimus* also shows different head group compositions with respect to pH, with relatively more HPH head groups

compared to glycosidic head groups at a higher pH (Elling et al., 2015). However, *N. maritimus* and two other Thaumarchaeota isolates showed clear differences in head group compositions when grown under identical conditions, implying that species composition also affect the IPL compostition (Elling et al., 2015).

Interestingly, the relative abundances of IPLs with a GDGT-0 core decreased with increasing depth in all three stations (Figure 8). This is contrary to other marine waters in which the archaeal IPL composition in SPM has been determined. For example, the relative abundance of IPL-GDGT-0 increased with water depth in the Arabian Sea (Schouten et al., 2012), and in parts of the Western Atlantic Ocean (Hurley et al., 2018). A possible explanation is the headgroup composition of GDGT-0 IPLs, which is predominantly the (presumably) more labile headgroup HPH instead of MH and DH (Figure 6). These differences may be also be caused by differences of the archaeal community composition present at various water depths in the Mediterranean Sea.

Biological sources of archaeal lipids in the water column

To assess potential sources of specific IPLs, we correlated the abundance of archaea falling in specific subgroups (OTUs were grouped in subgroups based on phylogenetic similarity; Figure 4, Supplementary Figs. S2 and S3) with the most abundant archaeal IPLs (Table 1) following the approach of Sollai et al. (2018). The absolute abundance of specific archaeal groups (estimated by multiplying the total archaeal 16S rRNA gene copies L^{-1} by the relative abundances of the archaeal groups and by assuming one 16S rRNA gene copy number per genome) was correlated with the absolute abundances of archaeal IPLs expressed as response units per liter (Table 1). Significant positive correlations (ρ-value < 0.05) for all three stations (Table 1) were found between MGI subgroup I and HPH-GDGTs and DH-GDGTs, driven mostly by DH-GDGT-2 and DH-crenarchaeol. On the other hand, MGI subgroup II is also correlated with DH-crenarchaeol (Table 1). MGI subgroup III was only significantly correlated with HPH-GDGT-0 at all three stations (Table 1; correlations with p-values < 0.05 are highlighted). The observed correlations suggest that these thaumarchaeotal groups are the biological source of most IPLs.

However, culture studies have shown that Thaumarchaeota synthesize mainly IPLs with GDGT-0 and crenarchaeol as core lipids, as well as GDGT-1 to GDGT-4 in lower relative abundance (Sinninghe Damsté et al., 2002a; Schouten et al., 2008; Pitcher et al., 2011a; Sinninghe Damsté et al., 2012; Elling et al., 2014). However, our data reveal that GDGT-2 and crenarchaeol, and perhaps its isomer (cren′), are the main core lipids in IPL-GDGTs, especially from 90 m downwards (Figure 6), while IPL-GDGT-0 would be expected to be also high, as seen before in other marine waters, such as the Portuguese margin (Kim et al., 2016). This particularity seems to be restricted to the MGI populations in the Mediterranean Sea. It is possible that these Thaumarchaeota regulate their lipid membrane composition in a different way, or alternatively, their lipid membrane composition varies between the different phylogenetic groups within the MGI archaea present in the Mediterranean Sea. In this regard, the similarity between major OTUs of MGI found in the Mediterranean Sea water masses with those found in similar physicochemical conditions (i.e. high salinity and temperature as in the Red Sea) could explain the observed similarity in GDGT composition between the Mediterranean and the northern Red Sea (Kim et al., 2015) and the dissimilarity with open ocean systems and Thaumarchaeotal cultures studied so far.

Table 1. Table with correlation values of the abundance of specific archaeal subgroups with those of most common archaeal intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) for the three stations. The correlation data were obtained by applying Spearman ranking correlation coefficient analyses. The IPLs with a dihexose (DH) or a hexose-phosphohexose (HPH) head group were also summed and correlated against the MG subgroups. Correlation coefficient values with a ρ-value<0.05 are highlighted.

	HPH	DH	HPH-GDGT-0	DH-GDGT-2	HPH-GDGT-2	DH-crenarchaeol	HPH-crenarchaeol
Station 1							
MGI Group I	**0.86**	**0.82**	0.42	**0.78**	0.57	**0.83**	0.49
MGI Group II	0.46	**0.79**	-0.05	**0.76**	**0.80**	**0.82**	0.05
MGI Group III	**0.94**	**0.60**	**0.71**	0.56	0.28	**0.60**	**0.75**
MGII Group I	**0.84**	**0.75**	0.50	**0.73**	**0.68**	**0.78**	0.56
MGII Group II	**0.67**	**0.96**	0.15	**0.97**	0.56	**0.95**	0.14
MGII Group III	-0.49	0.10	**-0.74**	0.07	0.10	0.12	**-0.65**

Table 1. Continued

	HPH	DH	HPH-GDGT-0	DH-GDGT-2	HPH-GDGT-2	DH-crenarchaeol	HPH-crenarchaeol
MGII Group IV	**0.58**	-0.01	**0.73**	0.00	-0.06	0.00	**0.70**
MGIII Group I	**0.60**	**0.88**	0.07	**0.87**	**0.74**	**0.90**	0.12
MGIII Group II	-0.35	0.22	**-0.66**	0.16	0.30	0.25	-0.54
Station 2							
MGI Group I	**0.88**	**0.88**	**0.70**	**0.84**	**0.75**	**0.72**	**0.75**
MGI Group II	**0.68**	**0.68**	0.07	0.55	0.43	**0.81**	0.14
MGI Group III	0.36	0.36	**0.77**	0.46	0.49	0.07	**0.78**
MGII Group I	**0.72**	**0.72**	**0.68**	**0.72**	**0.69**	0.55	**0.74**
MGII Group II	**0.75**	**0.75**	**0.79**	**0.75**	**0.77**	0.57	**0.83**
MGII Group III	0.53	0.53	0.14	0.43	-0.01	0.42	0.10
MGII Group IV	-0.45	-0.45	-0.23	-0.45	-0.19	-0.40	-0.23
MGIII Group I	**0.72**	**0.72**	0.56	**0.67**	**0.73**	0.60	**0.61**
MGIII Group II	0.39	0.39	-0.27	0.26	0.05	0.60	-0.21
Station 3							
MGI Group I	**0.98**	**0.98**	0.48	**0.90**	**0.84**	**0.98**	0.58
MGI Group II	**0.95**	**0.97**	0.33	**0.88**	**0.83**	**1.00**	0.50
MGI Group III	0.23	0.31	**0.71**	0.51	0.62	0.18	0.66
MGII Group I	**0.80**	**0.88**	0.61	**0.90**	**0.84**	**0.79**	**0.73**
MGII Group II	**0.90**	**0.90**	0.41	**0.83**	**0.84**	**0.93**	0.51
MGII Group III	0.44	0.37	-0.26	0.21	0.02	0.48	-0.28
MGII Group IV	0.05	-0.11	**0.71**	-0.07	0.12	-0.22	0.65
MGIII Group I	0.65	0.63	0.64	0.62	**0.82**	0.60	**0.82**
MGIII Group II	0.53	0.49	-0.32	0.24	0.05	0.58	-0.22

The abundance of MGII and III archaea were also positively correlated with the concentrations of HPH and DH-GDGTs (Table 1), which may suggest that they also contribute to the GDGT pool. However, the absence of IPL-GDGTs in the surface waters dominated by MGII/MGIII suggest that these archaeal groups are not a major source for these lipids, in contrast to previous suggestions (Lincoln et al., 2014a), and instead might synthesize membrane lipids not included in our analytical window. Indeed, it has been suggested that MGII/MGIII might synthesize mixed

bacterial/archaeal-like membranes based on the presence of bacterial membrane lipid biosynthetic genes compared to archaeal ones (Villanueva et al., 2017). As cultures of these archaea are still lacking, no definite conclusions can be made at this point regarding the possible contribution of the MGII/MGIII to the IPL-GDGT pool present in the marine water column. The positive correlation between MGII and III and the specific GDGTs mentioned above could be due to the fact they share their niche with MGI Thaumarchaeota as indicated in the correlation matrix of the archaeal groups (based on the summed relative abundances of the grouped OTUs; Supplementary Figure S9).

Implications for the TEX$_{86}$ paleothermometer

The TEX$_{86}$ paleothermometer is based on the fractional abundances of the CLs GDGT-1–3, and the crenarchaeol isomer in sediments (Schouten et al., 2002). The assumption for the use of this proxy is that the CL-GDGTs in sediments are primarily derived from GDGTs produced in surface waters, presumably because the GDGTs produced in surface waters are more easily transported to the sediment than those produced in deeper waters through grazing and fecal pellet formation (Huguet et al., 2006a). Kim et al. (2015, 2016) showed that in the Mediterranean Sea and in waters of the Portuguese margin, strongly affected by the outflowing Mediterranean LIW and WMDW, the GDGT distribution in sediments was unusually affected by GDGTs produced in deeper waters. Our detailed study allows us to further evaluate this. The CL-GDGT distribution in SPM changes clearly with depth (Supplementary Figure S12), with an increase in the relative abundance of GDGT-2 and the crenarchaeol isomer and a decrease of GDGT-1, in agreement with scattered SPM data reported by Kim et al. (2015). In the surface sediments the relative abundance of GDGT-2 and the crenarchaeol isomer is less than in SPM from deeper waters but still higher than in SPM from the surface waters. The opposite is true for GDGT-1, with a higher relative abundance in surface waters compared to that in the surface sediments. Although we don't know if this distribution is influenced by seasonal factors, the data suggest that surface CL-GDGTs are transported to the sediment but that CL-GDGTs from deeper waters also contribute to some extent.

This is in good agreement with the IPL data, which show that IPL-GDGTs are primarily produced in subsurface waters (with a maximum at 250–300 mbsl) and that the main CLs are GDGT-2 and crenarchaeol. Transformation to CL-GDGTs from these IPLs by loss of the polar head groups could explain the observed changes with depth in the CL-GDGT profiles in the water column.

An increase in the GDGT-2/GDGT-3 ratio, based on the CL-GDGTs, with increasing water depth has been previously reported for multiple marine settings (e.g., Taylor et al., 2013; Hernández-Sánchez et al., 2014) including the Mediterranean Sea (Kim et al., 2015). Although changes in this ratio do not affect values of TEX_{86}, they have been interpreted as an indicator of GDGT production by Archaea living in the deeper part of the water column. In our data set, the GDGT-2/GDGT-3 ratio also increases with increasing depth (Figure 9). However, GDGT-2/GDGT-3 values within the deeper part (300 mbsl and downwards; located in the LIW, the WMDW and the EMDW) of the Mediterranean Sea were generally lower than observed in SPM from other regions, such as the North Atlantic Ocean (Turich et al., 2007; Basse et al., 2014; Hurley et al., 2018), the South Atlantic Ocean (Hernández-Sánchez et al., 2014; Hurley et al., 2018), the Cariaco Basin (Zhu et al., 2016), the Black Sea (Zhu et al., 2016), the North Pacific (Zhu et al., 2016) and the equatorial Pacific (Zhu et al., 2016) (Figure 9). Furthermore, it remained stable from 300 to 500 mbsl downwards, which is different from observations in the Atlantic Ocean, where the GDGT-2/GDGT-3 ratio increases with depth (Figure 9). The relatively low GDGT-2/GDGT-3 values and the constant values with depth in the Mediterranean Sea could be due to the relatively high and constant in situ temperatures throughout the water column compared to the global open ocean.
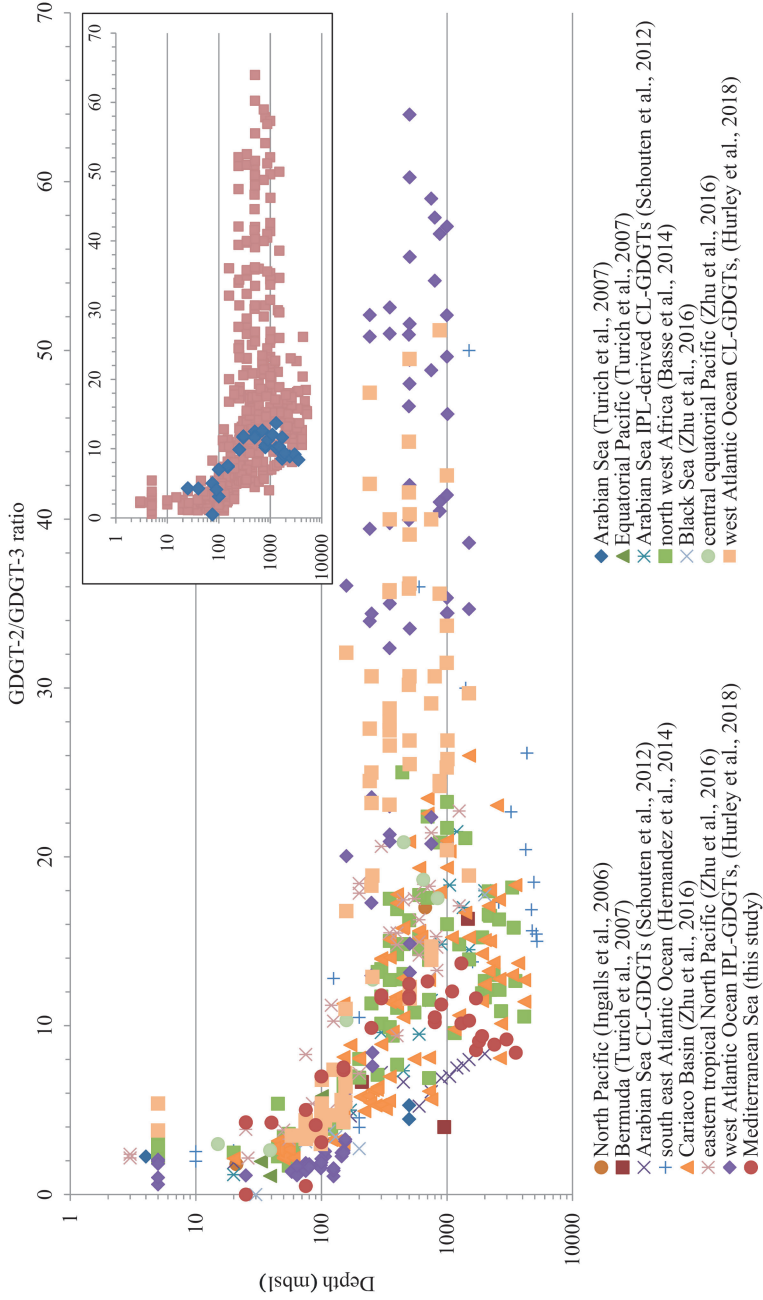
Figure 9. Depth profile of the ratio between glycerol dialkyl glycerol tetraether lipids (GDGTs) with two and three cyclopentane moieties (GDGT-2/GDGT-3) from core lipid (CL-) GDGTs from all three stations (indicated by red circles). For reference, the GDGT-2/GDGT-3 ratios (from CL-GDGTs, intact polar lipid (IPL-) GDGTs and IPL derived CL-GDGTs) for other water columns in multiple marine environments are also plotted. Note the logarithmic Y-axis. The inset distinguishes the data points from this study (blue diamonds) and the data points from the other studies (red squares).

3

As noted before for the Mediterranean Sea, the deviating CL-GDGT distribution in sediments due to a larger contribution of deep water dwelling Thaumarchaeota has consequences for the TEX_{86} paleotemperature proxy (Kim et al., 2015). Accordingly, Kim et al. (2015) developed a specific Mediterranean Sea core top calibration using only deep water sediments. We applied this calibration to the underlying surface sediments, but the reconstructed temperatures were only slightly higher (1.1 °C at both stations) than the recorded annual mean SSTs at station 1 and 3, respectively (Supplementary Figure S5), which is just outside the calibration error of 1 °C (Kim et al., 2015). Only at the station with the deepest water column there is a larger offset (4 °C; Supplementary Figure S5). This station (3609 m deep) was the deepest station and perhaps the expanded deep water column attributed to the larger offset of the reconstructed SST. In fact sediments at > 3000 m water depth were not included in the deep water calibration by Kim et al. (2015). Furthermore, the estimated total abundance of archaea at depths > 1000 mbsl is approximately one order of magnitude lower than the maximum abundance detected in the subsurface waters (100-500 m depth), which suggests that the potential contribution to the archaeal lipid pool of deep water archaea may still be significant. This potential contribution of archaeal lipids, formed in the deeper parts of the Mediterranean Sea and the Portuguese Margin, to the sedimentary record has previously been proposed by Kim et al. (2015, 2016)

Surprisingly, Menzel et al. (2006) reported for Mediterranean Pliocene sapropels lower than expected TEX_{86}-based SST values using the common Schouten et al. (2004) calibration and suggested that this could be caused by the oxycline rising up into the photic zone, which substantially affects the composition of the Thaumarchaeotal community. A recent study by Polik et al. (2018) has confirmed these results with additional analyses of Mediterranean Pliocene and Pleistocene sapropels. The characteristic GDGT distributions within these sapropels and the modern Mediterranean were interpreted to be caused by different Thaumarchaeota community compositions (Polik et al., 2018). Our study highlights the complexity and internal variability within the modern thaumarchaeotal community that is not

limited to the "shallow" and "deep" water types previously described. Furthermore, when a shallow oxycline develops during times of sapropel formation, the Thaumarchaeotal composition and distribution may be much more like that of the present day Black Sea (cf. Sollai et al., 2018) than the present-day oxic Mediterranean with major consequences for the type of calibration that should be used for TEX_{86} paleothermometry. The niche occupancy and thriving of these MGI subgroups in different water masses in the Mediterranean Sea may explain the deviations in SST-derived from TEX_{86} observed both in the past and in the present, and also shows the importance of understanding the thaumarchaeotal community dynamics in situ to get a better understanding on the effects of environmental factors on the GDGT distribution and therefore on the TEX_{86} paleothermometer.

3

CONCLUSIONS

We studied the IPL and CL-GDGT distribution in combination with the archaeal community composition of the Mediterranean Sea water column at three different stations, one in the western basin and two in the eastern basin. We detected a dominance of MGII and MGIII archaea at the surface of the water column, whereas MGI dominated the subsurface and the deeper parts of the water column. We detected specific thaumarchaeotal OTUs, which could be adapted to the relative high temperature and salinity found in the LIW mass of the Mediterranean Sea. These thaumarchaeotal OTUs, which seem to proliferate in the specific conditions in the LIW water mass of the Mediterranean Sea, are most likely causing the different archaeal lipid signals that translate to higher TEX_{86} values than observed in the open ocean with similar SSTs. No evidence was found that archaeal groups such as MGII and MGIII archaea, shown to be present by genetic evidence, contribute substantially to the GDGT signal, and they are therefore not expected to affect TEX_{86} calculations. The CL-GDGT distribution was characterized by an increase of the GDGT-2/GDGT-3 ratio with depth as observed in other marine systems, but with relatively low GDGT-2/GDGT-3 values compared to other oceans may possibly be due to the particularities of the Mediterranean Sea (i.e. high salinity and temperature of specific water masses). Future studies focused on the enrichment/

culture of the specific deep water MGI OTUs found in in this study, followed by lipid analysis, should clarify their role in the deep water archaeal GDGT distribution in the Mediterranean Sea.

ACKNOWLEDGMENTS

SUPPLEMENTAL FIGURES AND TABLES

Table S1. Physicochemical characteristics in the Mediterranean Sea station at the time of sampling, nd = not determined. Station #1.

Depth (mbsl)	Filter (pore size, μm)	Temp (°C)	Sal (PSU)	Oxygen (umol kg⁻¹)	Fluorescence (μg L⁻¹)	Si (μmol L⁻¹)	PO₄³⁻ (μmol L⁻¹)	NH₄⁺ (μmol L⁻¹)	NO₃⁻ + NO₂⁻ (μmol L⁻¹)	NO₂⁻ (μmol L⁻¹)	NO₃⁻ (μmol L⁻¹)
25	0.7	14.08	37.89	251.69	0.42	1.19	0.02	0.13	1.34	0.04	1.30
40	0.7	13.96	37.96	246.01	0.36	1.75	0.05	0.13	2.81	0.04	2.77
100	0.7	13.41	38.31	191.70	0.02	3.61	0.27	0.07	7.39	0.02	7.36
300	0.3	13.32	38.52	172.73	0.02	6.67	0.43	0.08	9.94	0.01	9.92
500	0.3	13.20	38.51	178.04	0.01	8.12	0.44	0.05	9.60	0.01	9.60
700	0.3	13.14	38.50	181.62	0.01	8.65	0.43	0.02	9.37	0.01	9.37
900	0.3	13.11	38.49	184.99	0.02	8.84	0.43	0.07	9.24	0.01	9.23
1100	0.3	13.10	38.48	186.97	0.02	8.92	0.42	0.09	9.10	0.01	9.10
1300	0.3	13.11	38.47	189.72	0.01	8.84	0.41	nd	8.97	0.01	8.96
1500	0.3	13.14	38.48	190.52	0.01	8.82	0.41	0.08	8.92	0.01	8.91
1700	0.3	13.17	38.48	191.22	0.01	8.80	0.41	0.07	8.89	0.01	8.88
1900	0.3	13.20	38.48	191.47	0.01	8.90	0.41	0.10	8.96	0.01	8.95

Station #2

Depth (mbsl)	Filter (pore size, µm)	Temp (°C)	Sal (PSU)	Oxygen (umol kg⁻¹)	Fluorescence (µg L⁻¹)	Si (µmol L⁻¹)	PO₄³⁻ (µmol L⁻¹)	NH₄⁺ (µmol L⁻¹)	NO₃⁻ + NO₂⁻ (µmol L⁻¹)	NO₂⁻ (µmol L⁻¹)	NO₃⁻ (µmol L⁻¹)
25	0.3	17.17	38.86	236.54	0.07	0.83	0.01	0.10	0.02	0.01	0.01
75	0.3	17.09	38.85	236.25	0.07	0.83	0.00	0.12	0.04	0.01	0.03
90	0.3	16.68	38.92	228.41	0.05	0.89	0.00	0.12	0.37	0.02	0.35
150	0.3	15.88	38.98	218.47	0.01	1.23	0.01	0.13	1.55	0.01	1.54
250	0.3	15.17	38.96	207.51	0.01	2.20	0.07	0.15	3.28	0.01	3.27
500	0.3	14.28	38.88	187.69	0.01	5.65	0.19	0.13	5.53	0.01	5.52
800	0.3	13.88	38.78	182.67	0.01	7.67	0.21	0.16	5.75	0.00	5.74
1800	0.3	13.79	38.74	191.13	0.01	8.30	0.20	0.17	5.34	0.00	5.34
2380	0.3	13.83	38.73	192.13	0.01	7.53	0.17	0.13	4.88	0.00	4.88
2980	0.3	13.92	38.73	191.71	0.00	7.46	0.16	0.18	4.82	0.01	4.82
3530	0.3	14.01	38.73	190.66	0.01	7.58	0.16	0.18	4.79	0.01	4.78

Station #3

Depth (mbsl)	Filter (pore size, μm)	Temp (°C)	Sal (PSU)	Oxygen (umol kg⁻¹)	Fluorescence (μg L⁻¹)	Si (μmol L⁻¹)	PO₄³⁻ (μmol L⁻¹)	NH₄⁺ (μmol L⁻¹)	NO₃⁻ + NO₂⁻ (μmol L⁻¹)	NO₂⁻ (μmol L⁻¹)	NO₃⁻ (μmol L⁻¹)
25	0.3	19.01	39.10	230.17	0.06	0.80	0.01	0.20	0.03	0.01	0.02
75	0.3	18.92	39.07	229.82	0.07	0.79	0.00	0.22	0.06	0.02	0.04
100	0.3	17.68	38.95	232.09	0.05	0.84	0.00	0.16	0.25	0.10	0.16
150	0.3	16.83	39.08	214.92	0.02	1.10	0.02	0.18	1.07	0.01	1.05
300	0.3	15.02	39.00	189.35	0.01	4.70	0.15	0.15	4.97	0.01	4.96
500	0.3	14.03	38.83	176.48	0.01	7.96	0.23	0.19	6.22	0.01	6.22
800	0.3	13.78	38.76	176.97	0.01	9.20	0.24	0.17	5.96	0.00	5.96
1300	0.3	13.79	38.76	182.55	0.01	9.18	0.20	0.19	5.35	0.00	5.35
1700	0.3	13.87	38.77	184.10	0.01	9.08	0.19	0.15	5.12	0.01	5.11

Table S2. Polymerase chain reaction (PCR) primers, efficiency and R^2 of the quantitative PCR (QPCR) assays

QPCR (gene)	Station	Primers	Ref	Efficiency	R^2
16S rRNA	1	Parch519F/ARC915R	Pitcher et al., 2011	86.8	0.996
16S rRNA	2	Parch519F/ARC915R	Pitcher et al., 2011	92.6	0.996
16S rRNA	3	Parch519F/ARC915R	Pitcher et al., 2011	88	0.993
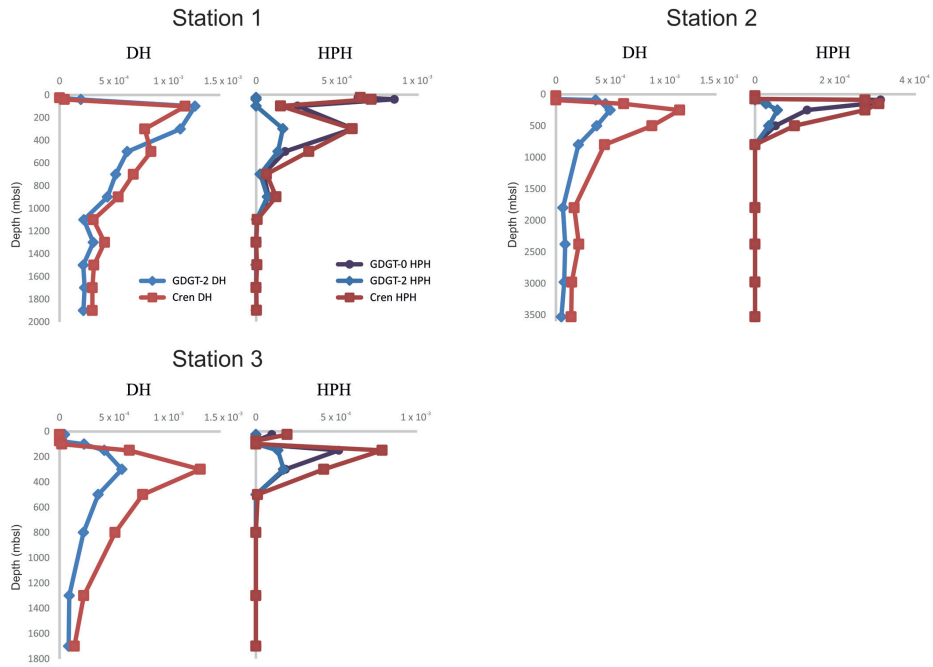
Figure S1. Depth profiles of the 5 most detected intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) at the three stations.

Depth (mbsl)	OTU ID# 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
25	0	8	0	0	12	0	48	12	4	4	0	0	0	0	0	8	0	0	0	0	0	0	4	0	0	0	0	0	0	0
40	9	0	0	0	8	0	36	12	0	0	0	0	1	0	0	19	0	0	1	5	0	0	6	0	3	0	0	0	0	0
100	27	15	5	3	2	3	13	10	2	0	3	1	2	0	0	4	1	1	1	2	0	1	1	1	4	0	0	0	0	0
300	37	12	9	7	0	6	4	3	2	2	2	2	2	1	1	1	1	1	2	2	0	1	1	1	0	0	1	0	0	0
500	41	10	7	8	0	6	3	1	3	2	3	3	1	2	1	0	1	2	1	0	1	1	0	1	0	1	0	1	0	1
700	46	10	8	6	0	5	2	1	3	2	3	3	2	1	2	0	1	1	1	0	1	1	0	1	0	0	1	1	1	1
900	49	6	9	9	0	6	1	1	2	2	2	3	1	1	2	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0
1100	57	3	9	8	0	5	0	1	2	2	1	3	1	1	1	1	1	1	1	0	1	0	0	1	0	0	0	1	1	0
1300	54	4	11	5	0	3	1	1	3	2	3	4	1	2	2	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0
1500	59	4	9	6	0	4	0	1	2	2	2	2	1	2	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	1
1700	59	4	10	4	0	4	0	1	2	2	1	2	0	1	4	0	1	0	0	0	1	0	1	1	0	1	0	0	1	1
1900	61	3	8	5	0	5	0	1	2	2	1	3	0	1	2	0	2	1	1	0	0	0	0	1	0	1	0	0	0	0

Depth (mbsl)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
25	0	0	0	0	73	0	15	4	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	64	0	25	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90	1	1	0	0	80	0	12	3	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150	23	5	0	1	29	3	5	14	3	0	1	0	3	0	0	3	0	1	1	1	0	1	3	1	1	0	0	0	0	0
250	42	8	3	6	5	3	3	7	4	1	2	0	2	2	0	1	1	2	3	1	0	1	1	1	1	0	1	0	0	0
500	51	5	6	8	0	1	0	2	4	3	2	1	3	1	2	0	2	1	2	0	1	1	0	1	0	0	1	1	0	1
800	56	3	8	8	0	2	0	0	3	7	2	1	0	1	2	0	2	1	1	0	1	1	0	1	0	1	0	1	1	1
1800	66	0	12	4	0	0	0	0	6	3	0	4	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
2380	52	1	8	5	2	6	1	0	6	8	2	1	0	1	0	0	1	1	0	0	2	1	0	0	0	1	0	0	1	1
2980	63	0	12	4	0	2	0	0	5	4	0	3	0	0	1	0	2	1	0	0	2	0	0	0	0	0	0	0	0	0
3530	66	1	12	0	0	1	0	0	3	9	0	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0

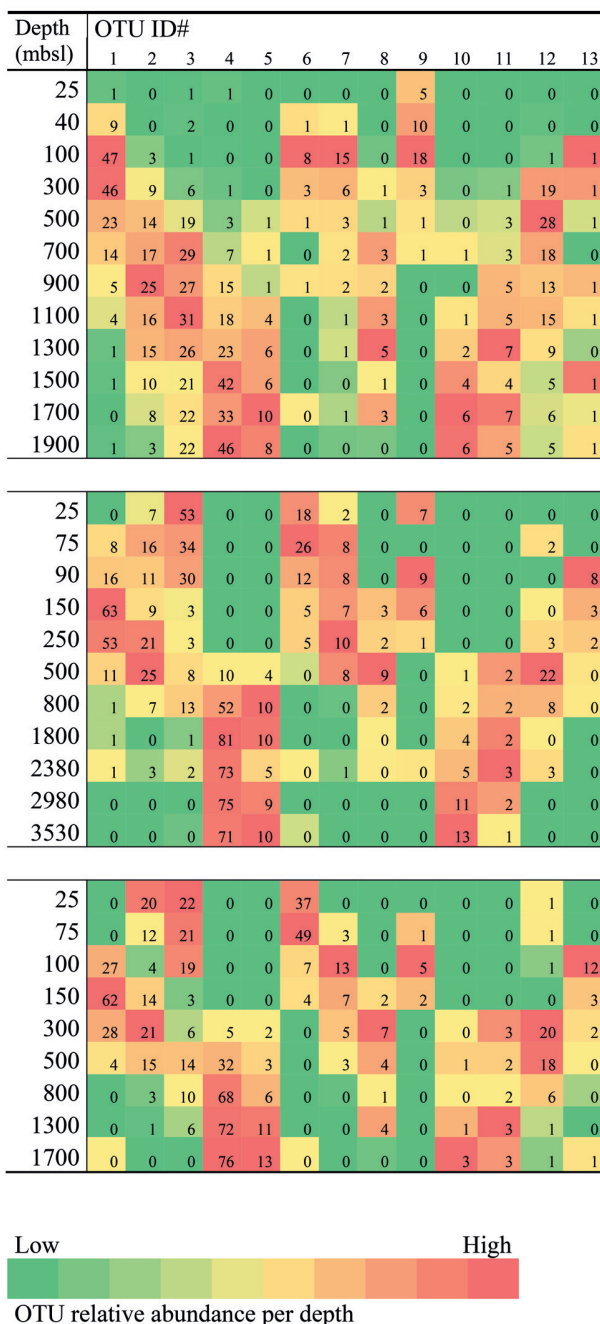| Depth (mbsl)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
25	0	0	0	0	74	0	3	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100	2	0	0	0	79	0	6	2	0	0	0	0	0	0	0	9	0	0	0	0	0	0	2	0	0	0	0	0	0	0
150	29	3	0	2	20	2	4	19	3	1	2	0	4	1	0	0	0	2	1	2	0	2	3	1	1	0	0	0	0	0
300	48	9	4	9	1	3	0	2	2	3	1	1	3	2	2	1	1	2	2	0	1	1	0	0	0	0	1	1	0	0
500	51	4	7	10	0	2	0	1	2	6	2	1	1	3	2	0	1	0	1	0	2	0	0	1	0	0	1	1	0	0
800	53	2	10	9	0	1	0	0	3	5	1	2	0	2	2	0	2	1	1	0	1	0	0	1	0	1	0	1	1	1
1300	55	0	13	7	0	2	0	0	2	7	0	2	0	1	2	0	2	1	0	0	4	0	0	0	0	1	0	0	1	1
1700	56	0	17	5	0	2	0	0	1	6	0	4	0	1	1	0	1	1	0	0	2	0	0	0	0	1	0	0	1	1

Low   High

OTU relative abundance per depth

Figure S2. Heatmap of marine group I (MGI) operational taxonomic units (OTUs) importance factors per depth at all three stations. Heatmap colors are indicative of the importance factor per OTU (green = low relative abundance, red = high).

Depth (mbsl)	OTU ID# 1	2	3	4	5	6	7	8	9	10	11	12	13
25	1	0	1	1	0	0	0	0	5	0	0	0	0
40	9	0	2	0	0	1	1	0	10	0	0	0	0
100	47	3	1	0	0	8	15	0	18	0	0	1	1
300	46	9	6	1	0	3	6	1	3	0	1	19	1
500	23	14	19	3	1	1	3	1	1	0	3	28	1
700	14	17	29	7	1	0	2	3	1	1	3	18	0
900	5	25	27	15	1	1	2	2	0	0	5	13	1
1100	4	16	31	18	4	0	1	3	0	1	5	15	1
1300	1	15	26	23	6	0	1	5	0	2	7	9	0
1500	1	10	21	42	6	0	0	1	0	4	4	5	1
1700	0	8	22	33	10	0	1	3	0	6	7	6	1
1900	1	3	22	46	8	0	0	0	0	6	5	5	1

Depth (mbsl)	1	2	3	4	5	6	7	8	9	10	11	12	13
25	0	7	53	0	0	18	2	0	7	0	0	0	0
75	8	16	34	0	0	26	8	0	0	0	0	2	0
90	16	11	30	0	0	12	8	0	9	0	0	0	8
150	63	9	3	0	0	5	7	3	6	0	0	0	3
250	53	21	3	0	0	5	10	2	1	0	0	3	2
500	11	25	8	10	4	0	8	9	0	1	2	22	0
800	1	7	13	52	10	0	0	2	0	2	2	8	0
1800	1	0	1	81	10	0	0	0	0	4	2	0	0
2380	1	3	2	73	5	0	1	0	0	5	3	3	0
2980	0	0	0	75	9	0	0	0	0	11	2	0	0
3530	0	0	0	71	10	0	0	0	0	13	1	0	0

Depth (mbsl)	1	2	3	4	5	6	7	8	9	10	11	12	13
25	0	20	22	0	0	37	0	0	0	0	0	1	0
75	0	12	21	0	0	49	3	0	1	0	0	1	0
100	27	4	19	0	0	7	13	0	5	0	0	1	12
150	62	14	3	0	0	4	7	2	2	0	0	0	3
300	28	21	6	5	2	0	5	7	0	0	3	20	2
500	4	15	14	32	3	0	3	4	0	1	2	18	0
800	0	3	10	68	6	0	0	1	0	0	2	6	0
1300	0	1	6	72	11	0	0	4	0	1	3	1	0
1700	0	0	0	76	13	0	0	0	0	3	3	1	1

Low   High

OTU relative abundance per depth

Figure S3. Heatmap of marine group II (MGII) operational taxonomic units (OTUs) importance factors per depth at all three stations. Heatmap colors are indicative of the importance factor per OTU (green = low relative abundance, red = high).

Depth (mbsl)	OTU ID# 1	2	3	4	5	6
25	100	0	0	0	0	0
40	0	0	100	0	0	0
100	1	26	71	1	0	0
300	8	63	18	9	1	1
500	16	69	5	4	5	2
700	23	55	0	10	10	1
900	33	55	1	4	7	1
1100	44	39	0	8	7	1
1300	49	32	0	10	6	4
1500	73	17	0	2	6	2
1700	69	16	0	6	5	3
1900	89	6	0	1	2	3

Depth (mbsl)	1	2	3	4	5	6
25	0	27	73	0	0	0
75	0	38	62	0	0	0
90	0	48	52	0	0	0
150	0	54	29	17	0	0
250	0	76	18	6	0	0
500	22	56	1	21	1	0
800	80	11	0	3	4	2
1800	98	0	0	0	0	2
2380	91	4	1	1	1	3
2980	95	0	0	0	0	5
3530	95	0	0	0	0	5

Depth (mbsl)	1	2	3	4	5	6
25	0	35	65	0	0	0
75	0	20	80	0	0	0
100	0	35	65	0	0	0
150	0	71	20	9	0	0
300	14	62	0	20	4	0
500	59	27	0	7	5	2
800	89	4	0	2	2	4
1300	92	1	0	5	0	2
1700	96	0	0	0	0	3

Low   High

OTU relative abundance per depth
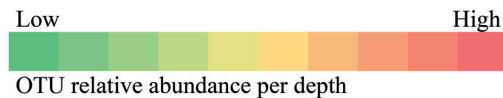
Figure S4. Heatmap of marine group III (MGIII) operational taxonomic units (OTUs) importance factors per depth at all three stations. Heatmap colors are indicative of the importance factor per OTU (green = low relative abundance, red = high).

Table S3. Correlation tables of the marine group I, II and III (MGI, MGII and MGIII) operational taxonomic units (OTUs) within the three stations based on their importance factors. Correlations coefficients were obtained by Spearman ranking correlation coefficient analyses.

Marine group I

Station	Station 1	Station 2	Station 3
Station 1	1		
Station 2	0.80*	1	
Station 3	0.74*	0.89*	1

*significant at $\rho < 0.001$.

Marine group II

	Station 1	Station 2	Station 3
Station 1	1		
Station 2	0.60*	1	
Station 3	0.49	0.87*	1

*significant at $\rho < 0.05$.

Marine group III

	Station 1	Station 2	Station 3
Station 1	1		
Station 2	0.89*	1	
Station 3	0.89*	0.89*	1

*significant at $\rho < 0.05$.
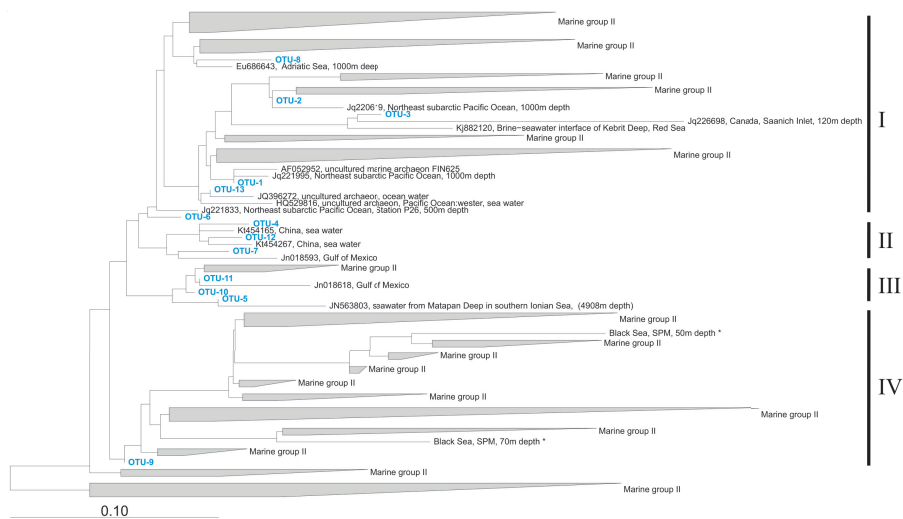
Figure S5. Phylogenetic tree composed of 16S rRNA gene amplicon sequences of marine group II (MGII) operational taxonomic units (OTUs; in blue) detected in the suspended particulate matter (SPM) of the Mediterranean Sea. Closely affiliated sequences from uncultured MGII (in black). Scale bar represents a 10% sequence dissimilarity. * indicates sequences recovered from SPM collected in the water column of the Black Sea (Sollai et al., 2018)
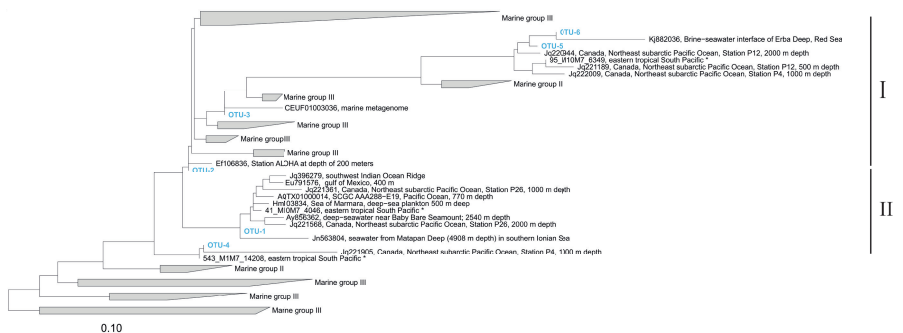
Figure S6. Phylogenetic tree composed of 16S rRNA gene amplicon sequences of marine group III (MGIII) operational taxonomic units (OTUs; in blue) detected in the suspended particulate matter (SPM) of the Mediterranean Sea. Closely affiliated sequences from uncultured MGIII (in black). Scale bar represents a 10% sequence dissimilarity. * indicates sequences recovered from SPM collected in the water column of the eastern tropical south Atlantic (ETSP; Sollai et al., 2015).

Table S4. Archaeal abundance based on 16S rRNA gene copies per liter filtered. *Abundance of marine groups (MGI, II and III) were determined by multiplying the relative abundance obtained by the 16S rRNA gene amplicon sequencing with the archaeal abundance based on the abundance of 16S rRNA gene copies obtained by QPCR.

Depth (mbsl)	Archaeal 16S rRNA gene copies L⁻¹	16S rRNA gene copies L⁻¹ *			Relative abundance (%)		
		MGI	MGII	MGIII	MGI	MGII	MGIII
25	1.1E+08	1.6E+07	9.2E+07	6.1E+05	15	84	1
40	2.0E+08	5.0E+07	1.5E+08	1.3E+06	25	74	1
100	6.3E+08	3.5E+08	2.5E+08	3.3E+07	55	39	5
300	2.0E+08	1.2E+08	7.4E+07	1.4E+07	57	36	7
500	8.3E+08	5.4E+08	2.2E+08	6.0E+07	65	27	7
700	4.7E+08	2.7E+08	1.3E+08	6.1E+07	59	28	13
900	2.5E+08	1.7E+08	4.3E+07	3.6E+07	67	17	15
1100	2.7E+08	1.7E+08	5.8E+07	4.0E+07	63	22	15
1300	3.5E+07	2.3E+07	6.6E+06	5.9E+06	64	19	17
1500	1.9E+08	1.4E+08	2.3E+07	3.0E+07	71	12	16
1700	1.0E+08	6.1E+07	2.0E+07	1.8E+07	59	19	17
1900	4.7E+07	2.2E+07	1.1E+07	1.2E+07	48	24	27

Depth (mbsl)	Archaeal 16S rRNA gene copies L⁻¹	16S rRNA gene copies L⁻¹ *			Relative abundance (%)		
		MGI	MGII	MGIII	MGI	MGII	MGIII
25	5.4E+07	1.8E+07	2.6E+07	1.1E+07	33	47	20
75	1.0E+08	3.9E+07	3.9E+07	2.3E+07	39	39	23
90	1.9E+08	9.1E+07	7.5E+07	2.3E+07	48	40	12
150	3.4E+08	2.3E+08	9.2E+07	1.8E+07	66	27	5
250	2.3E+08	1.6E+08	5.4E+07	2.1E+07	67	23	9
500	5.5E+07	3.8E+07	9.6E+06	7.5E+06	68	17	14
800	5.6E+07	2.5E+07	1.1E+07	2.0E+07	44	20	36
1800	1.7E+07	7.2E+06	1.6E+06	7.3E+06	43	9	44
2380	4.7E+07	2.5E+07	4.1E+06	1.6E+07	53	9	35
2980	2.0E+06	8.3E+05	2.6E+05	8.4E+05	41	13	41
3530	7.7E+05	3.0E+05	1.2E+05	3.2E+05	39	15	41

Depth (mbsl)	Archaeal 16S rRNA gene copies L⁻¹	16S rRNA gene copies L⁻¹ *			Relative abundance (%)		
		MGI	MGII	MGIII	MGI	MGII	MGIII
25	4.6E+07	1.1E+07	1.6E+07	1.9E+07	25	35	40
75	2.6E+07	5.3E+06	7.6E+06	1.3E+07	20	29	51
100	3.0E+07	2.0E+07	7.5E+06	9.6E+05	67	25	3
150	1.5E+08	8.1E+07	5.3E+07	1.3E+07	55	36	9
300	4.3E+08	3.2E+08	7.2E+07	4.0E+07	74	17	9
500	1.0E+08	5.5E+07	2.1E+07	2.5E+07	54	20	25
800	9.8E+07	4.9E+07	1.1E+07	3.7E+07	50	12	37
1300	2.3E+06	1.4E+06	1.6E+05	5.6E+05	63	7	25
1700	6.6E+07	3.5E+07	6.3E+06	2.3E+07	53	10	35
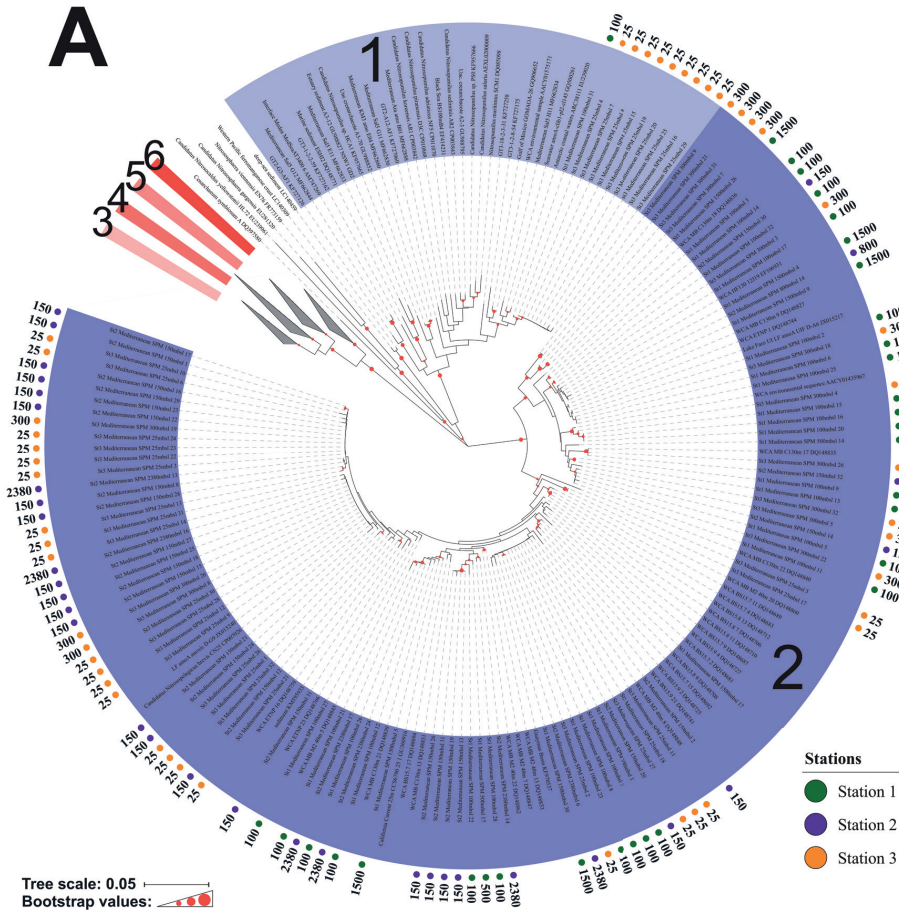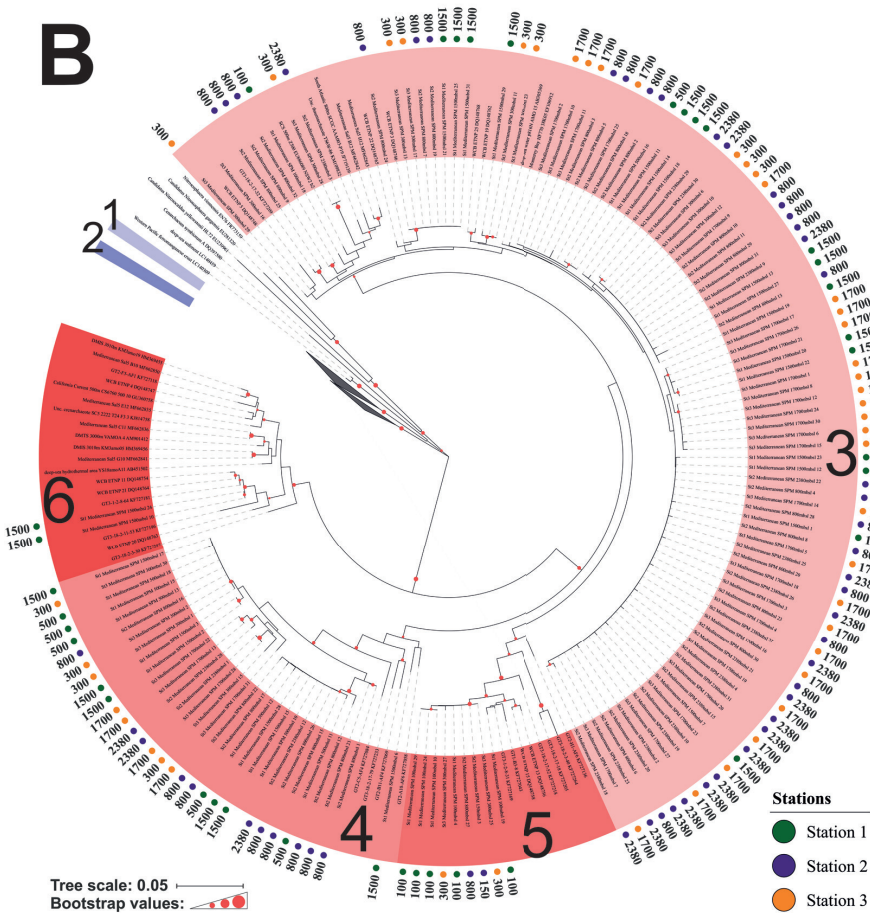
Figure S7. Neighbor-joining phylogenetic nucleotide tree of thaumarchaeotal α-subunit of the ammonia monooxygenase (*amo*A) gene fragments obtained from the sequenced clone libraries. Stations are specified by a colored circle in the outer ring, adjacent is the sampling depth (meters below sea level). (**A**) Phylogenetic tree with the expanded "shallow" water clusters 1 and 2 (indicated in purple), the "deep" water clusters (3-6; indicated in red) are collapsed. (**B**) Phylogenetic tree with the expanded "deep" water clusters 3–6 (indicated in red), the "shallow" clusters (1 and 2; indicated in purple) are collapsed. Tree was composed with a bootstrap test of 1,000 replicates (values higher than 50% are shown on the branches; indicated by red circles). Trees are based on La Cono et al. (2018) and Sintes et al. (2016).

3

Figure S8. Alignment of ammonia monooxygenase protein coding sequences. The amino acids that differed among the compared sequences are shown. Amino acid residues defining the 'shallow' and 'deep water' (*amoA* clusters are indicated by color that resemble the subgroups of Figure 5 and Figure S4). From a couple of subgroups are multiple representative sequences shown. Key: N, asparagine; S, serine; L, leucine; I, isoleucine; A, alanine; G, glycine; V, valine; R, arginine; K, lysine; F, phenylalanine; Q, glutamine; T, threonine; Y, tyrosine; H, histidine; C, cysteine; M, methionine.
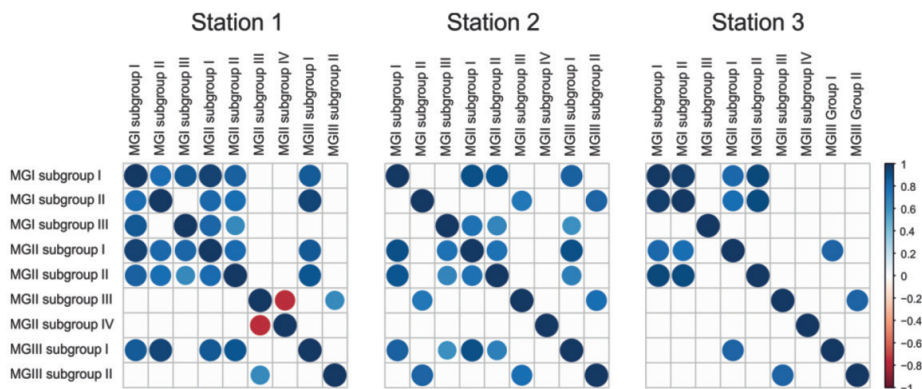
Figure S9. Dot plots of the correlation matrix for the three stations obtained by applying Spearman ranking correlation coefficient analyses to the marine group (MG) subgroups. Correlation coefficient values with a ρ-value >0.05 are not shown. The size of the dot and the intensity of its color (blue represents a positive correlation, red represent a negative correlation) relate to the degree of correlation.
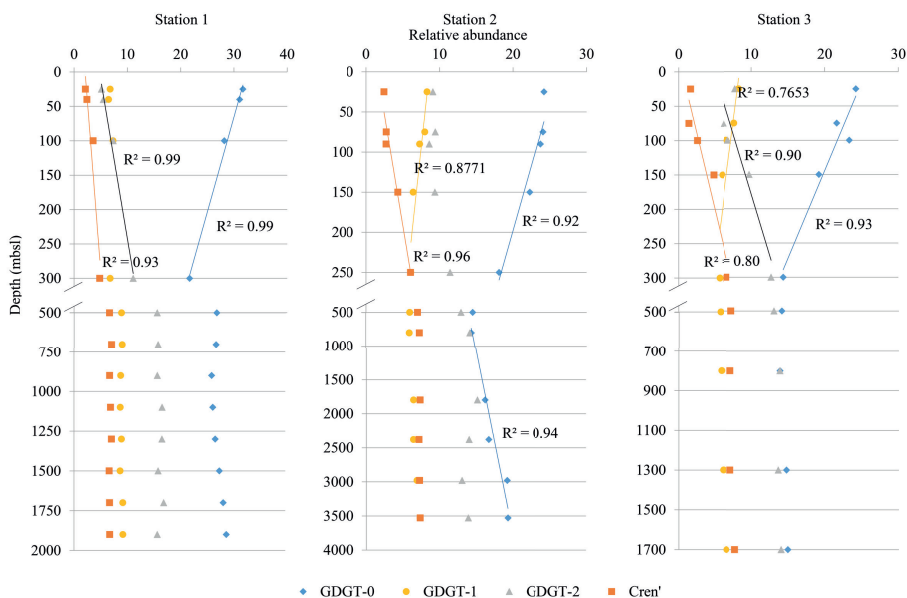
Figure S10. Depth profiles of core lipid (CL-) glycerol dialkyl glycerol tetraether lipids (GDGTs), GDGT-0 (indicated by blue diamants), GDGT-2 (grey triangles) and crenarchaeol isomers (orange squares) at the three sampling stations. Significant regressions (ρ-value <0.05) were given for CL-GDGT relative abundances up to 250 mbsl (station 2) and 300 mbsl (station 1 and 3), any significant regression was also noted from 500 mbsl downwards at all three stations. Note that the y-axes are cut-off, this to represent two different depth resolutions.
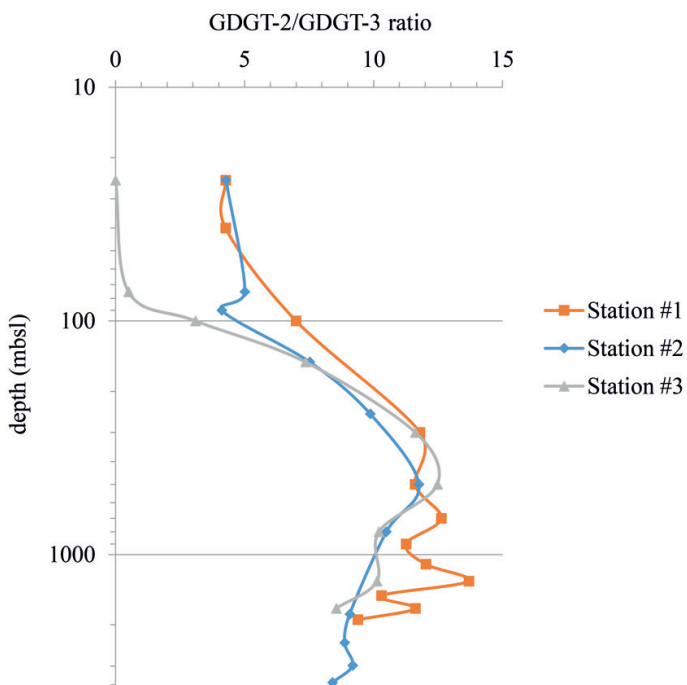
Figure S11. Depth profiles of the core lipid (CL-) glycerol dialkyl glycerol tetraether lipid (GDGT) 2/3 ratio of the three stations.
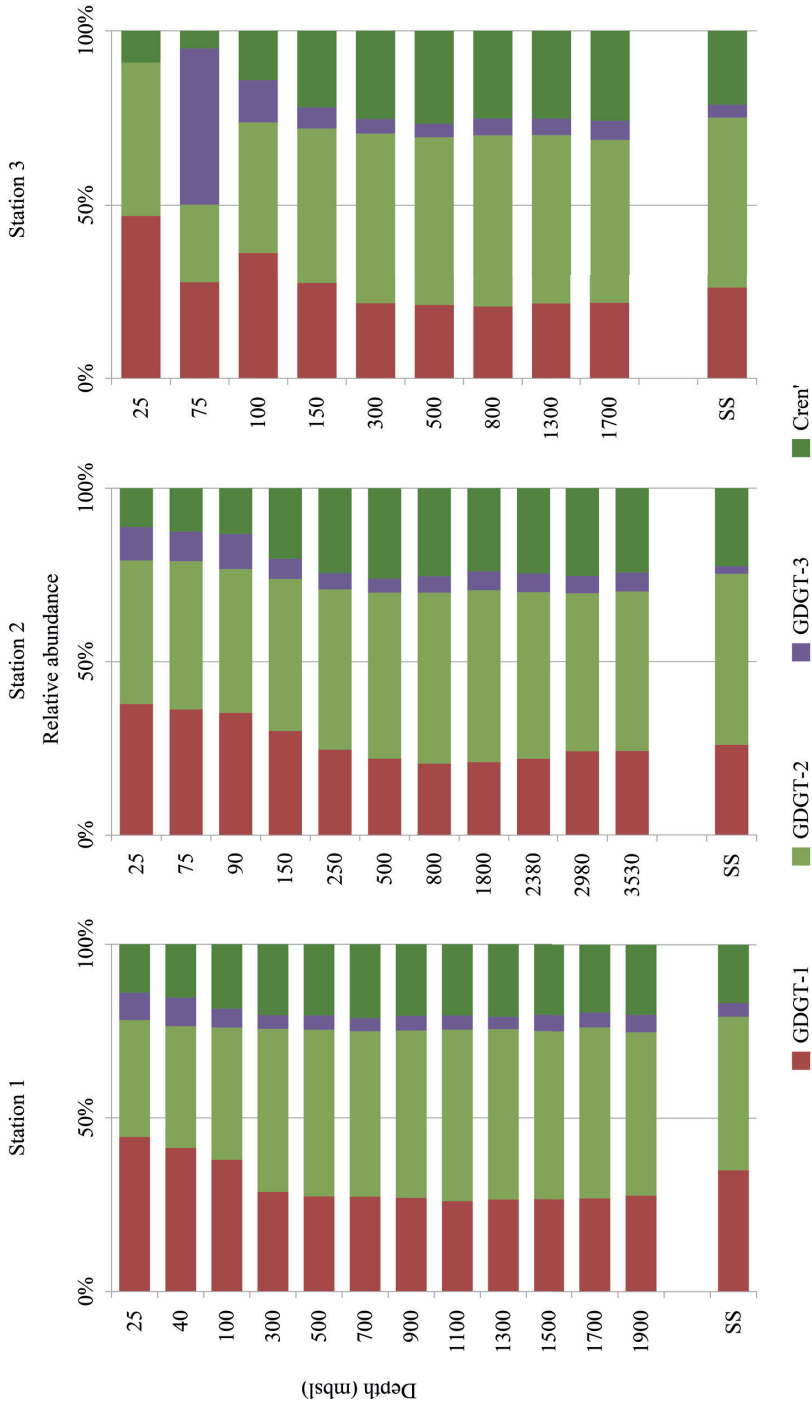
Figure S12. Depth profiles of the relative abundance of the core lipid (CL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) used in the TEX$_{86}$ paleo-temperature proxy for the three stations and the TEX$_{86}$ CL-GDGT composition of the underlying surface sediments (SS).

CHAPTER 4

A search for the membrane lipids of the uncultivated Marine Group II and III Euryarchaeota

Marc A. Besseling, Ellen C. Hopmans, Nicole J. Bale, Stefan Schouten , Jaap S. Sinninghe Damsté and Laura Villanueva

ABSTRACT

The marine pelagic environment is dominated by three major archaeal groups, the Thaumarchaeota marine group I (MGI), and the marine Euryarchaeota groups II and III (MGII and MGIII). Studies of both MGI cultures and the environment have shown that the MGI membrane core lipids are predominantly composed of the glycerol dialkyl glycerol tetraether lipids (GDGTs) and the diether lipid archaeol. However, there are no cultured representatives of MGII and III and, therefore, both their membrane lipid composition and potential contribution to the marine archaeal lipid pool remains unknown. Here, we show that GDGTs present in suspended particulate matter of the (sub)surface waters of the North Atlantic Ocean and the coastal North Sea are derived from MGI and that MGII and III do not significantly contribute to the pool of GDGTs and archaeol, in contrast to previous suggestions. Using genomic and lipidomic data, we suggest that the nature of the membrane lipids of the MGII may be different than those commonly found within the Archaea.

INTRODUCTION

Archaeal communities in the marine pelagic environment are mostly dominated by three major groups, the Thaumarchaeota marine group I (MGI) and the marine Euryarchaeota groups II and III (MGII and MGIII) (DeLong, 1992; Fuhrman et al., 1992; Li et al., 2015). It is known from culture and environmental studies that the Thaumarchaeota MGI are capable of oxidizing ammonia (Könneke et al., 2005; Wuchter et al., 2006) and that some members are able to use urea as an alternative substrate (Bayer et al., 2016). The metabolism of MGII and MGIII is thought to be (photo-)heterotrophic and these archaea are potentially able to degrade proteins, carbohydrates, fatty acids and other lipids (Iverson et al., 2012; Martin-Cuadrado et al., 2014; Li et al., 2015; Orsi et al., 2015). However, these suggestions are based solely on metagenomic data since pure cultures of MGII and MGIII have as yet not been obtained.

In the marine environment archaeal membrane lipids are used as biomarkers for the presence of archaea in microbial ecology studies (Biddle et al., 2006; Pitcher et al., 2011) but also in the paleotemperature proxy TEX_{86} (Schouten et al., 2002), commonly used in paleoclimatological studies. The membrane lipid composition of the MGI has been widely studied (Sinninghe Damsté et al., 2002; Elling et al., 2014; Schouten et al., 2008; Wuchter et al., 2004) and has been found to be composed of the diether lipid archaeol, glycerol dialkyl glycerol tetraether lipids (GDGTs) with zero to 4 cyclopentane moieties and crenarchaeol, a GDGT with four cyclopentane moieties and a cyclohexane moiety, so far exclusively found in Thaumarchaeota (Elling et al., 2017; Schouten et al., 2008; Sinninghe Damsté et al., 2002). In living and intact cells, GDGTs are mostly bound to polar head groups, termed intact polar lipid GDGTs (IPL-GDGTs). These occur mostly with sugar head groups such as monohexose (MH), dihexose (DH) and hexose phosphohexose (HPH) (Elling et al., 2014, 2015; Pitcher et al., 2010, 2011; Schouten et al., 2008).

In contrast, despite their importance in the marine water column as evidenced by metagenomic surveys (Iverson et al., 2012; Li et al., 2015; Martin-Cuadrado et al., 2014; Orsi et al., 2015; among others), it is not possible to directly determine the

4

lipid membrane composition of MGII and MGIII as cultures are lacking. A previous study based on a combination of metagenomics, sequencing of 16S rRNA gene amplicons and GDGT core lipid data in the North Pacific Subtropical Gyre has suggested that MGII are producers of GDGTs, including crenarchaeol (Lincoln et al., 2014a). However, the low amount of extracted DNA and of archaeal gene reads together with the use of core lipid (CL)-GDGTs (which can also be attributed to cell debris) made the conclusions of this study ambiguous (Schouten, et al., 2014; Lincoln et al., 2014b).

Here, we investigated the potential lipid composition of MGII and MGII by characterizing the abundance and diversity of archaea and composition of archaeal lipids in suspended particulate matter (SPM) from coastal North Sea water and the North Atlantic Ocean, which were previously shown to contain high abundances of MGII and MGIII (Herndl et al., 2005; Wuchter, 2006). We analyzed the abundance of the archaeal community by archaeal 16S rRNA gene quantitative polymerase chain reaction (qPCR), while its diversity was analyzed with 16S rRNA gene amplicon sequencing. The abundance and composition of archaeal lipids were determined by High Resolution Accurate Mass/Mass spectrometry (UHPLC-HRAM/MS). We compared the results of these different analyses in order to constrain the importance of the three main archaeal pelagic groups in their contribution to the marine archaeal lipid pool.

MATERIALS AND METHODS

Sampling and physicochemical analyses

Suspended particulate matter (SPM) was collected from the North Atlantic during the HCC cruise in September 2014 with *R/V* Pelagia (Bale et al., 2018). SPM samples were taken with three McLane WTS-LV *in-situ* pumps (McLane Laboratories Inc., Falmouth, MA, USA). The pumps were deployed at various depths and locations (station numbers are according to Bale et al., 2018), i.e. between 5-200 meters below sea level (mbsl) at station 10 (6° 41.04′N; 47° 29.25′W), as well as from 5-83 mbsl at four nearby stations (station 12; 6° 4.04′N; 52° 27.69′W, station 8; 6° 29.24′N;

45° 26.98′W, station 5; 10° 49.74′N; 40°28.20′W and station 4; 12° 24.42′N; 38° 30.12′W) (Figure S1). SPM was filtered onto pre-ashed glass fiber filters (0.7 μm nominal pore sizes, GF/F, Pall Corporation, Port Washington, NY and 0.3 μm nominal pore sizes, GF75, Avantec, Japan. Both 142 mm in diameter) and between 89 and 399 liters of seawater were filtered (Table S1). Physical data and dissolved inorganic nutrients are reported in Bale et al. 2018). Coastal North Sea SPM was collected in July 2015 at high tide from the NIOZ jetty (53° 0.1′N; 4° 47.2′E). Coastal North Sea water (ranging between 20 and 25 liters) was consecutively filtered in triplicate through 0.7 μm and 0.3 μm pore size glass fiber filters (Whatman, 142 mm in diameter). For all sampling locations, the filters were both used for lipid and nucleic acid analyses and were immediately stored at -80°C after sampling.

IPL extraction and analysis

Total lipids were extracted from freeze dried glass fiber filters using a modified Bligh and Dyer method (Bligh & Dyer, 1959) as previously described by Lengger et al. (2012). C_{16}-PAF (1-O-hexadecyl-2-acetoyl-sn-glycero-3-phosphocholine) was added to the extracts as an internal standard and dried under a stream of nitrogen. The extracts with the added standard were re-dissolved in known quantities of injection solvent (hexane:isopropanol:H_2O 718:271:10 v/v/v) and filtered through a 0.45 μm, 4 mm-diameter True Regenerated Cellulose syringe filter (Grace Davison, Columbia, MD, USA).

IPLs were analyzed using ultra high performance liquid chromatography coupled to high-resolution accurate-mass/mass spectrometry (UHPLC-HRAM/MS), designed for the analysis of a wide range of intact polar lipids. An Ultimate 3000 RS UHPLC, equipped with thermostated auto-injector and column oven, coupled to a Q Exactive Orbitrap MS with Ion Max source with heated electrospray ionization (HESI) probe (Thermo Fisher Scientific, Waltham, MA), was used. Separation was achieved on a YMC-Triart Diol-HILIC column (150 x 2.0 mm, 1.9 μm particles, pore size 12 nm; YMC Co., Ltd, Kyoto, Japan) with a guard column of the same material (10 x 2.1 mm) maintained at 30 °C. The following elution program was used with a flow rate of 0.2 mL min^{-1}: 100% A for 5 min, followed by a linear gradient to 66% A:

34% B in 20 min, maintained for 15 min, followed by a linear gradient to 40% A: 60% B in 15 min, followed by a linear gradient to 30% A: 70% B in 10 min, where A = hexane/2-propanol/formic acid/14.8 M NH_3aq (79:20:0.12:0.04 [v/v/v/v]) and B = 2-propanol/water/formic acid/ 14.8 M NH_3aq (88:10:0.12:0.04 [v/v/v/v]). The samples from the North Atlantic (0.7 µm pore-size glass fiber filters) were run with heptane instead of hexane in the mobile phase. Total run time was 70 min with a re-equilibration period of 20 min in between runs. HESI settings were as follows: sheath gas (N_2) pressure 35 (arbitrary units), auxiliary gas (N_2) pressure 10 (arbitrary units), auxiliary gas (N_2) T 50 ˚C, sweep gas (N_2) pressure 10 (arbitrary units), spray voltage 4.0 kV (positive ion ESI), capillary temperature 275 °C, S-Lens 70 V. IPLs were analyzed with a mass range of m/z 375 to 2000 (resolving power 70,000 at m/z 200), followed by data dependent MS^2 (resolving power 17,500 ppm at m/z 200), in which the ten most abundant masses in the mass spectrum (with the exclusion of isotope peaks) were fragmented (stepped normalized collision energy 15, 22.5, 30; isolation window 1.0 m/z). A dynamic exclusion window of 6 sec was used as well as an inclusion list with a mass tolerance of 3 ppm to target specific compounds (see Besseling et al., (2018) for the list of targeted compounds). The Q Exactive Orbitrap MS was calibrated within a mass accuracy range of 1 using the Thermo Scientific Pierce LTQ Velos ESI Positive Ion Calibration Solution (containing a mixture of caffeine, MRFA, Ultramark 1621, and N-butylamine in an acetonitrile-methanol-acetic acid solution).

Peak areas for each individual IPL were determined by integrating the combined mass chromatogram (within 3 ppm) of the monoisotopic and first isotope peak of all relevant adducts formed (protonated, ammoniated and/or sodiated adducts may be formed in different proportions depending on the type of IPL). C_{16}-PAF was used as internal standard to continuously monitor MS performance and to assess matrix effects. Reported peak areas were corrected for these effects. Absolute quantification of IPL GDGTs was not possible due to a lack of authentic standards. Peak areas were not corrected for possible differences in response factors between the various classes of IPL-crenarchaeol. IPLs with crenarchaeol or its isomer(s) as its core lipid,

but with the same headgroup, co-elute with the chromatographic system used here and any peak area reported for a crenarchaeol IPL thus represents the sum of both crenarchaeol and its isomer(s).

Analysis of GDGTs and IPL derived GDGTs

The Bligh and Dyer extracts from three separate 0.7 µm pore-size glass fiber filters obtained from coastal North Sea water, were dried under nitrogen and hydrolyzed separately with 1.5 N HCl in methanol by reflux at 130°C for 2 h to release the CL-GDGTs from the IPL-GDGTs. The pH was adjusted to 7 by adding 2 N KOH/MeOH (1:1 v/v) and, after addition of water to a final 1:1 (v/v) ratio of H_2O-MeOH, extracted three times with dichloromethane (DCM). The DCM fractions were collected and dried over sodium sulfate. The extracts, containing IPL derived CL-GDGTs and CL-GDGTs in case of the acid hydrolyzed extracts or only CL-GDGTs in case of the not hydrolyzed extracts, were dried under N_2 and dissolved in hexane--2-propanol (99:1, vol/vol) and filtered over a 0.45-µm polytetrafluoroethylene filter. Extracts were analyzed by UHPLC-atmospheric pressure chemical ionization (APCI) MS for archaeol and GDGTs, according to Hopmans et al. (2016) with some modifications. Briefly, analysis was performed on an Agilent 1260 UHPLC coupled to a 6130 quadrupole MSD in selected ion monitoring (SIM) mode. Separation was achieved on two UHPLC silica columns (BEH HILIC columns, 2.1 x 150 mm, 1.7 µm; Waters) in series, fitted with a 2.1 x 5 mm pre-column of the same material (Waters) and maintained at 30°C. Archaeol and GDGTs were eluted isocratically for 10 min with 10% B, followed by a linear gradient to 18% B in 20 min, then a linear gradient to 100% B in 20 min, where A is hexane and B is hexane: isopropanol (9:1). Flow rate was 0.2 ml/min. Total run time is 61 min with a 20 min re-equilibration. Source settings were identical to Schouten et al. (2007). Typical injection volume was 10 µl of a 1 mg/ml solution. The *m/z* values of the protonated molecules of archaeol and isoprenoid GDGTs were monitored. Archaeol and GDGTs were quantified by adding a C_{46} GTGT internal standard (Huguet et al., 2006b).

The Bligh and Dyer extract and the acid hydrolyzed Bligh and Dyer extract were also analyzed using ultra high-performance liquid chromatography coupled

to positive ion atmospheric pressure chemical ionization / Time-of-Flight mass spectrometry (UHPLC-APCI/ToFMS) on an Agilent 1290 Infinity II UHPLC, equipped with automatic injector, coupled to a 6230 Agilent TOF MS and Mass Hunter software. Separation of the ether lipids was achieved according to Hopmans et al. (2016) with some modifications using two silica BEH HILIC columns in series (2.1 x 150 mm, 1.7 µm; Waters) at a temperature of 25 °C. The injection volume was 10 µL. Compounds were isocratically eluted with 90% A and 10% B for the first 10 min, followed by a gradient to 18% B in 15 min, a gradient to 30% B in 25 min and a linear gradient to 100% B in 30 min. A = hexane and B = hexane/isopropanol (9:1, v/v) and the flow rate was 0.2 mL/min. The conditions for the APCI source were identical to Schouten et al. (2007) and Hopmans et al. (2016). In addition, the fragmentor was set at 300 V. The ToFMS was operated in extended dynamic range mode (2 GHz) with a scan rate of 2 Hz. We assessed archaeal lipid distributions by monitoring m/z 600 to 1400. Ether lipids were identified by searching within 10 ppm mass accuracy for relevant $[M+H]^+$ signals.

Nucleic acids extraction and quantitative PCR (qPCR) analyses

DNA was extracted from the glass fiber filters with the RNA PowerSoil® Total Isolation Kit plus the DNA elution accessory (Mo Bio Laboratories, Carlsbad, CA). Concentration of DNA was quantified by Qubit fluorometric quantitation (ThermoFisher Scientific). Quantification of archaeal 16S rRNA gene copies were estimated by qPCR by using the following primers; Parch519F and ARC915R (archaeal 16S rRNA gene) as previously described by Pitcher et al. (2011). For details on the qPCR conditions, efficiency and R^2 of the qPCR assays see Table S2.

16S rRNA gene amplicon sequencing, analysis, and phylogeny

PCR reactions were performed with the universal, Bacteria and Archaea, primers S-D-Arch-0159-a-S-15 and S-D-Bact-785-a-A-21 (Klindworth et al., 2013) as described in Moore et al., (2015). The archaeal 16S rRNA gene amplicon sequences were analyzed by QIIME v1.9 (Caporaso et al., 2010). Raw sequences were demultiplexed and then quality-filtered with a minimum quality score of 25, length

between 250-350, and allowing maximum two errors in the barcode sequence. Taxonomy was assigned based on blast and the SILVA database version 128 128 (Altschul et al., 1990; Quast et al., 2013). Representative operational taxonomic units (OTUs) sequences of archaeal groups were extracted through filter_taxa_from_otu_ table.py and filter_fasta.py with QIIME (Caporaso et al., 2010). The phylogenetic affiliation of the partial archaeal 16S rRNA gene sequences was compared to release 123 of the Silva NR SSU Ref database (http://www. arb-silva.de/; Quast et al., 2013) using the ARB software package (Ludwig et al., 2004). Sequences were added to the reference tree supplied by the Silva database using the ARB Parsimony tool. Affiliation of any 16S rRNA gene sequences to a given subgroup was done assuming a similarity cutoff of ≥85%.

Abundances of the three archaeal marine groups were estimated by multiplying archaeal copy number (16S rRNA gene copies L^{-1}) with the relative abundance of the marine group (in % of all archaeal reads), assuming one copy of 16S rRNA gene per genome (Stoddard et al., 2015).

RESULTS

In this study, we analyzed both lipids and DNA from SPM collected at depth intervals between 5-200 mbsl on glass fiber filters with pore size 0.3 μm in the North Atlantic (station 10 of Bale et al., 2018) and between 5-83 mbsl at four nearby stations using glass fiber filters with pore size 0.7 μm. In addition, SPM of North Sea surface waters collected by sequentially filtering through 0.7 μm and 0.3 μm filters in triplicate, was analyzed for both lipids and DNA.

Archaeal abundance and diversity

Total archaeal abundances in the North Atlantic were low for the shallow samples (<50 mbsl), ranging between 8×10^6 and 8×10^7 16S rRNA gene copies L^{-1}, compared to samples taken at greater depths (>= 50 mbsl), ranging between 2×10^7 and 5×10^8 gene copies L^{-1} (Figure 1A). The only exception was at 200 mbsl of station 10 (0.7 μm), which contained a lower amount of 16S rRNA gene copies L^{-1} compared to 50 mbsl from the same station (Figure 1A).

Various archaeal groups were detected in the SPM in the upper 200 m of the North Atlantic on both the 0.3 and 0.7 μm filters of the different stations (Figure 1B). At depths of <50 mbsl, the archaeal community consisted mainly of MGII with relative abundances ranging between 65 to 93% of total archaeal 16S rRNA gene reads (Figure 1B). At these depths, MGIII archaea were present in relative abundances of 6-18%, while the relative abundance of Thaumarchaeota is generally low (<4% except for coastal station 12. At depths >50 mbsl, thaumarchaeotal MGI relative abundance ranged from 10 to 49%, while other archaeal groups present were MGII and III (ranging between 43-84% and between 5 and 8%, respectively).

Abundances of the three main archaeal groups (MGI, II and III) in the North Atlantic were estimated based on the relative abundance, obtained by 16S amplicon sequencing, multiplied by the amount of 16S gene copy numbers per liter seawater filtered. The estimated abundances of MGI Archaea within North Atlantic SPM increased strongly at depths >=50 mbsl, ranging between 1×10^5 and 1×10^8 16S rRNA gene copies L^{-1} (Table S1). Estimated abundances of MGII Archaea were between 7×10^6 and 5×10^8 16S rRNA gene copies L^{-1}, abundances also increased with depth but not as outspoken as the MG1 Archaea (Table S1). MGIII estimated abundances were between 8×10^5 and 3×10^7 16S rRNA gene copies L^{-1} (Table S1).

The coastal North Sea 0.7 μm glass fiber filter contained 1×10^9 archaeal 16S rRNA gene copies L^{-1} while it was substantial lower on the 0.3 μm glass fiber filter (7×10^3 archaeal 16S rRNA gene copies L^{-1}; Table 2). For coastal North Sea SPM collected on the 0.7 μm filters, 16S rRNA gene reads were almost entirely (99.5%) affiliated with MGII and comprised 6 operational taxonomic units (OTUs), the remaining archaeal 16S rRNA gene reads (0.5%) were assigned to MGI.
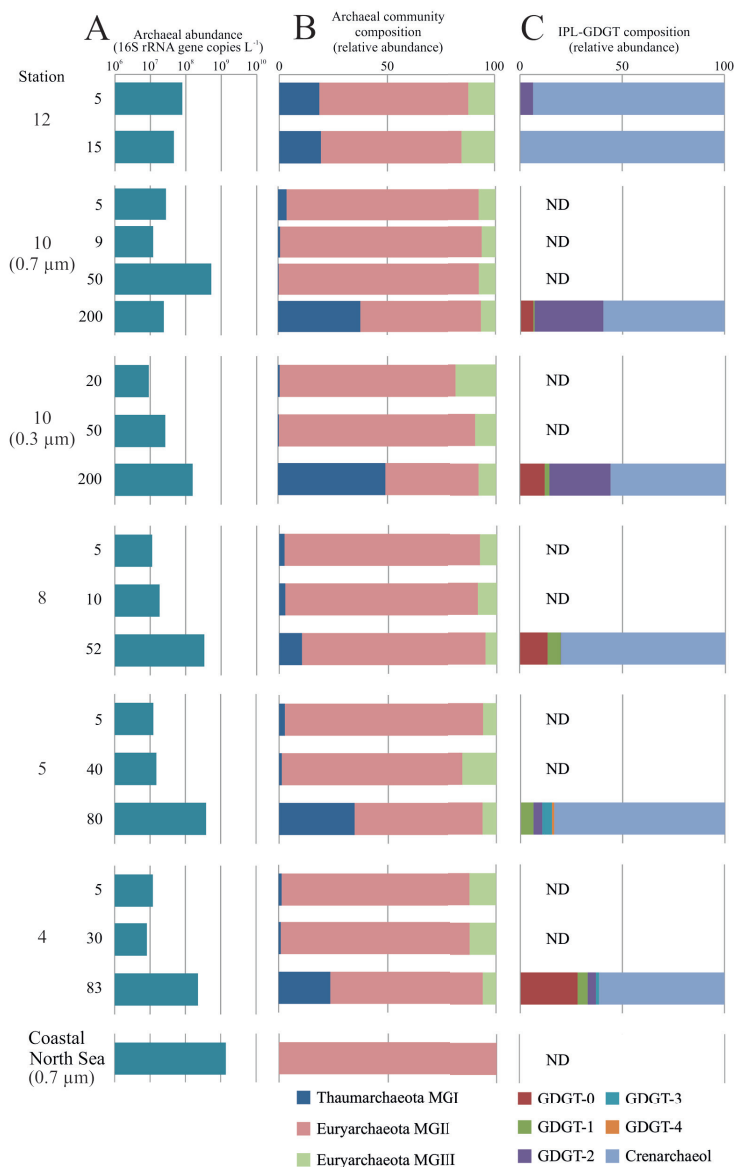
Figure 1. A) Total archaeal abundance determined by quantitative PCR and given as archaeal 16S rRNA gene copies L^{-1} of SPM in the tropical North Atlantic Ocean and North Sea, B) the archaeal community composition based on 16S rRNA gene amplicon sequencing and C) the distribution of the intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) with monohexose (MH), dihexose (DH) and hexose-phosphohexose (HPH) headgroups assuming similar response factors. All North Atlantic stations were sampled with 0.7 µm glass fiber filters, station 10 was also sampled with 0.3 µm glass fiber filters, numbers left of the bars indicate the sampling depth in meters below sea level. ND = not detected.

Archaeal intact polar lipid diversity and distribution

None of the numerous targeted archaeal IPLs (see inclusion list in Besseling et al. 2018) were detected within the upper depth(s) (<50 mbsl) of the North Atlantic (except for station 12) or the summer coastal North Sea SPM (0.7 µm filters; Figure 1C). In the North Atlantic at depths >80 mbsl, archaeal IPLs were detected (Figure 1, 2) with a dominance of the HPH-crenarchaeol (0-84%; relative abundance of the total detected IPL-GDGTs; Table S3) and DH-crenarchaeol (0-51%), with lower relative abundances of DH-GDGT-2 (0-33%) and HPH-GDGT-0 (0-28%). IPLs with crenarchaeol as their CL dominated the IPL-GDGTs in all cases (Figure 1C). This was especially apparent at station 12, the station closest to the shore, which contained a relative abundance of 94 and 100% IPL-crenarchaeol at 5 and 15 mbsl, respectively (Figure 1C).

To investigate if there are IPL-GDGTs that were not detected in our analytical window, the IPL fractions of the North Sea SPM collected on 0.7 µm pore size glass fiber filters, containing a high abundance of 16S rRNA gene reads attributed to the MGII (1.4×10^9 16S rRNA gene copies L^{-1}) and a substantially lower abundance of MGI affiliated reads (6.9×10^6 16S rRNA gene copies L^{-1}), were acid-hydrolyzed (in triplicate) to release IPL-derived GDGTs. Analysis by UHPLC-MS using SIM mode showed that there was a small, but statistically not significant, increase in the concentration after acid hydrolysis for both GDGT-0 (from 4±2 to 11±6 ng L^{-1}) and crenarchaeol (from 4±2 to 7±4 ng L^{-1}) (Figure 2). Small amounts of archaeol were also detected (0.4±0.1 ng L^{-1}), which did not significant increase after acid hydrolysis (1.8±1.8 ng L^{-1}; Figure 2). To check whether archaeal core lipids other than archaeol and GDGTs were present in the summer coastal North Sea SPM, we analyzed both the Bligh and Dyer total lipid extract and an acid-hydrolyzed lipid extract with UHPLC-ToFMS. Additional archaeal lipids detected were hydroxylated isoprenoidal GDGTs (OH-GDGTs), glycerol dialkanol diethers (GDDs) and butanetriol dibiphytanyl glycerol tetraethers (BDGTs), though in very low abundance compared to the GDGTs (Figure S3). Furthermore, no significant increase of these lipids after acid hydrolysis was observed (Figure S3).
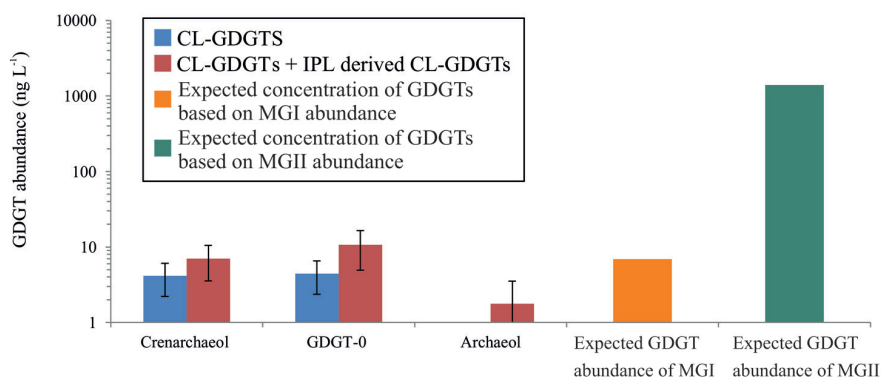
Figure 2. Concentration of intact polar lipid (IPL-) derived core lipid (CL), crenarchaeol and glycerol dialkyl glycerol tetraether lipid (GDGT) without cyclopentane moieties (GDGT-0), the two most abundant GDGTs and CL-GDGT abundance in coastal North Sea suspended particulate matter (SPM), filtered on 0.7 um glass fiber filters, and the estimated abundance of IPL-GDGTs based on the gene copy numbers of MGI and MGII archaea. Error bars indicate standard deviation of triplicates. Estimated amounts of IPL-GDGTs are based on the presence of 1 fg of IPL-GDGTs cell^{-1}, an estimated abundance of 1.4×10^9 marine Euryarchaeota group II (MGII) and 6.9×10^6 MGI cells L^{-1} (see text for details). Note the log-scale on the y-axis.

DISCUSSION

Archaea of the MGII cluster were detected with high relative abundances at the upper-surface (between 0 and 50 mbsl) in the North Atlantic (Figure 1). This is similar to other studies in which members of the MGII were more (relative) abundant in the surface compared to deeper waters (Delong et al., 2006; Herndl et al., 2005; Massana et al., 2000). Furthermore, MGII abundance in the North Atlantic (estimated by multiplying their relative abundance with total archaeal DNA copy abundance and assuming one DNA per cell) ranged between 7×10^6 and 5×10^8 cells L^{-1}, which is similar to previous reported values of abundance of MGII in the surface of the North Atlantic Ocean (i.e. 3×10^7 cells L^{-1}; Herndl et al., 2005). The MGII abundance in summer coastal North Sea was much higher, i.e. 1×10^9 cells L^{-1}, which is similar to previously reported MGII abundances and coincided within the known MGII summer bloom at the same location[5].

Both the relative (Figure 1) and absolute (Table S1) abundance of Thaumarchaeota MGI increased with depth in the North Atlantic Ocean. This is in line with the reported niche preference of MGI archaea for subsurface waters (approximately 50-500 m depth, Lincoln et al., 2014; Liu et al., 2017; Massana et al., 2000; Beman et al., 2012; Pitcher et al., 2011). The archaeal community composition at station 10 at 200 mbsl (sampled both on 0.3 μm and 0.7 μm pore-size filters) was comparable, albeit that the 0.3 μm filter showed a slightly higher relative abundance of MGI 16S rRNA gene reads compared to the 0.7 μm filter (Figure 1). Therefore, the filter pore-size does not seem to affect the archaeal community composition but only the amount of captured archaeal cells. The MGI abundance in summer coastal North Sea was 7 x 10^6 cells L^{-1}, similar to earlier studies Wuchter (2006) and Pitcher et al. (2011b) reporting MGI abundances of <1 x 10^7 cells L^{-1} in summer coastal North Sea SPM. Hence, the archaeal community composition and abundances in the SPM studied here compare well with those of previous studies and indicates an archaeal community composition that almost entirely consists out of MGII archaea, comprised of multiple MGII OTUs.

IPLs with a crenarchaeol core lipid were only detected in SPM of the North Atlantic in which Thaumarchaeota are present in substantial abundance (i.e. >9 x 10^6 cells L^{-1}; Figure 3A). In contrast, in the upper-surface water of the North Atlantic and in the coastal water of the North Sea, the low abundances of MGI cells L^{-1} (between 7 x 10^4 and 7 x 10^6 cells L^{-1}; Figure 1) coincided with the lack of detection of IPL-crenarchaeol, as well as other IPL-GDGTs and IPL-archaeol (Figure 3A). Likely below a certain MGI abundance (<~9 x 10^6 cells L^{-1}), the IPL-GDGTs and IPL-archaeol are below the detection limit of our IPL analysis method (Figure 3A). These observations support previous reports that Thaumarchaeota are the dominant source of IPL-GDGTs (cf. review of Schouten et al., 2013). Besides IPL-GDGTs, a variety of other IPLs were detected in the North Atlantic (cf. Bale et al., 2018) and the coastal North Sea SPM, containing diacyl glycerids and common headgroups such as phosphatidylglycerol (PG), phosphatidylcholine (PC) and diacylglyceryl-

carboxyhydroxymethylcholine (DGCC), among others, which likely have a bacterial or eukaryotic origin (Figure S2).
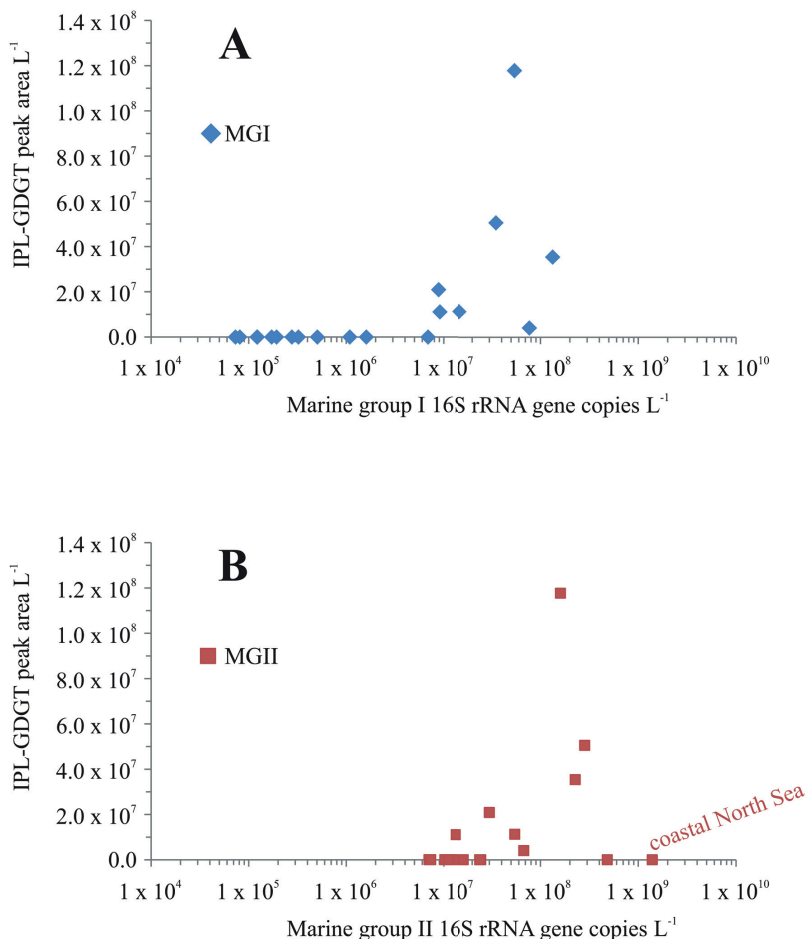
Figure 3. (A) Marine Thaumarchaeota (MGI) and (B) Marine Euryarchaeota group II (MGII) abundances (in cells L^{-1}, estimated from the amount of archaeal 16S rRNA gene copies per liter multiplied by the archaeal group relative abundance) versus intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipid (GDGT) peak area per liter. IPL-GDGT peak areas are the summed abundance of IPL-GDGTs with monohexose (MH), dihexose (DH) and hexose-phosphohexose (HPH) as headgroup. Note the log-scale on the x-axis.

A different picture emerges for MGII archaea. At all stations in the North Atlantic and North Sea where the MGII abundance was >7 x 10^6 cells L^{-1}, i.e. the detection limit for MGI archaea, and MGI archaeal abundances were low (<7 x 10^6), no IPL-GDGTs could be detected, nor any IPLs with an archaeol CL (Figure 3B). Since

the cell size and, therefore, the amount of IPL per cell, are similar for MGI and MGII archaea, this strongly suggests that IPLs of MGII are either not the same as commonly known archaeal IPLs or not detectable in our analytical window. To rule out the latter, acid hydrolysis was performed on three individual filters of SPM collected from the coastal North Sea (0.7 μm pore-size). This only resulted in a minor but statistically insignificant increase of crenarchaeol and GDGT-0 compared to the CL-GDGTs already present (Figure 2), which suggests very low amounts of IPL-GDGTs.

Further support for a predominant Thaumarchaeotal source of GDGTs in the marine environment comes from an estimation of the IPL-GDGT concentration of MGI and MGII, based on the amount of GDGTs expected to be present in MGI and MGII cells. MGI archaeal cells contain approximately 1 fg cell^{-1} of GDGTs based on culture and environmental studies (ranged from 1 fg cellular lipids cell^{-1}, Sinninghe Damsté et al., 2002; 1.4 fg IPL-GDGTs cell^{-1} Lipp et al., 2008; 0.1 − 8.5 fg IPL-GDGTs cell^{-1}, Huguet et al., 2010; 0.8 − 1.8 fg IPL-GDGTs cell^{-1} according to Elling et al., 2014). Thus, we would predict the presence of approximately 7 ng GDGTs per liter in coastal North Sea water derived from MGI, in good agreement with the small, but statistically insignificant, increase in concentration of crenarchaeol and GDGT-0 after acid hydrolysis (Figure 2), which converts IPL-GDGTs into CL-GDGTs. The size of MGII archaeal cells is similar as that of MGI (Orsi et al., 2015) and, therefore, it would be expected to contain approximately the same amount of IPL-GDGTs as estimated for MGI archaeal cells (i.e. 1 fg cell^{-1}). Because of the orders of magnitude higher abundance of MGII archaeal cells in coastal North Sea SPM compared to MGI, the expected concentration of IPL-GDGTs (or other archaeal lipids) attributed to MGII cells would be ca. 1,400 ng L^{-1}. This is three order of magnitudes higher than what is released after acid hydrolysis (Figure 2). This mismatch indicates that if MGII archaea would produce IPL-GDGTs, acid hydrolysis would have indicated much higher concentrations of IPL-GDGTs than observed (Figure 2).

Our data strongly suggests that MGII do not produce known archaeal lipids such as GDGTs (or analogues thereof such as OH-GDGTs, BDGTs and GDD) or diether lipids. Our results strongly contrast those of Lincoln et al. (2014a), who stated that MGII are able to synthesize archaeal GDGTs and, therefore, contribute substantially to the total archaeal GDGT pool, potentially affecting the paleotemperature proxy TEX_{86}. In contrast, our study suggests that MGII do not contribute to the GDGT pool and thus are unlikely to affect the TEX_{86} proxy used in paleoclimatic research.

Consequently, the question arises which lipids MGII Archaea do produce. Phylogenomic analyses by Villanueva et al. (2017) of MGII genomes and metagenomes (Iverson et al., 2012) showed that MGII Archaea lack the gene coding for glycerol-1-phosphate dehydrogenase (G1PDH), the enzyme responsible for the formation of glycerol-1-phosphate (G1P). This implies that they are not able to synthesize the G1P backbone found in archaeal membrane lipids, at least not by the classical biosynthetic pathway. However, MGII Archaea still harbor other archaeal membrane lipid biosynthetic pathway genes such as geranylgeranylglyceryl phosphate synthase (GGGP) and digeranylgeranylglyceryl phosphate synthase (DGGGP), among others (Villanueva et al., 2017). Therefore, they could potentially be capable of synthesizing isoprenoid-based ether lipids, but not with G1P as the glycerol building block, in contrary to other archaeal groups (Villanueva et al., 2017). However, GDGTs with a different glycerol stereochemistry would likely still be in our analytical window. Interestingly, MGII Archaea also possess a putative fatty-acid biosynthetic pathway and putative homologs of the bacterial 1-acylglycerol-3-phosphate O-acyltransferase (PlsC; Villanueva et al., 2017) involved in esterifying fatty acids to glycerol-3-phosphate (G3P) in bacterial membrane lipids. MGII Archaea could potentially produce a bacterial-like or a 'chimeric' lipid membrane, i.e. with a G3P backbone linked to di- or tetraether-linked isoprenoidal lipids, or lipids with one ether-linked isoprenoidal chain at position *sn*-1 and one ester-bound fatty acid at position *sn*-2 of the G3P backbone (Villanueva et al., 2017). In fact, a mixed bacterial/archaeal lipid membrane could potentially favor the MGII in variable environments. Caforio et al. (2018) recently demonstrated that the bacterium *Escherichia coli*, when modified

with archaeal/ether lipid biosynthetic genes, is able to synthesize a mixed lipid membrane, comprised of archaetidylglycerol (with a G1P backbone) alongside the bacterial lipids phosphatidylglycerol and phosphatidylethanolamine (both with a G3P backbone). The engineered *E. coli* with mixed membranes were more resistant to several chemical/physical stresses compared to the wild type *E. coli* (Caforio et al., 2018). However, careful inspection of the IPL composition as revealed by UHPLC–HRAM/MS did not reveal any unusual IPLs in high abundances or those which could be tentatively related to the hypothetical mixed lipids discussed above. This suggest that they either synthesize IPLs commonly found in marine environments or more unusual IPLs which escape our analytical window. Clearly, the nature of the lipid membrane of MGII (and MGIII) will need to ultimately be confirmed by analyzing the membrane lipid composition of MGII isolates.

CONCLUSION

Analysis of North Atlantic Ocean and summer coastal North Sea SPM showed a dominance of MGII and MGIII archaea with lower amounts of MGI. Only in SPM with a sufficient abundance of MGI, we were able to detect known archaeal intact polar lipids, but not in those with a dominance of MGII such as summer coastal North Sea SPM which contained abundant archaeal 16S RNA gene reads attributed to MGII. Acid hydrolysis followed by core lipid analysis ruled out the presence of intact polar archaeal lipids with unknown headgroup(s) in samples with dominant MGII populations. Based on our results, we conclude that it is unlikely that members of MGII are capable of synthesizing GDGTs or other known archaeal lipids, as previously proposed by others. Therefore, the question remains as to what lipids MGII membranes do comprise. Further research focusing on the cultivation of the elusive MGII archaea and analysis of their membrane are required to resolve this enigma.
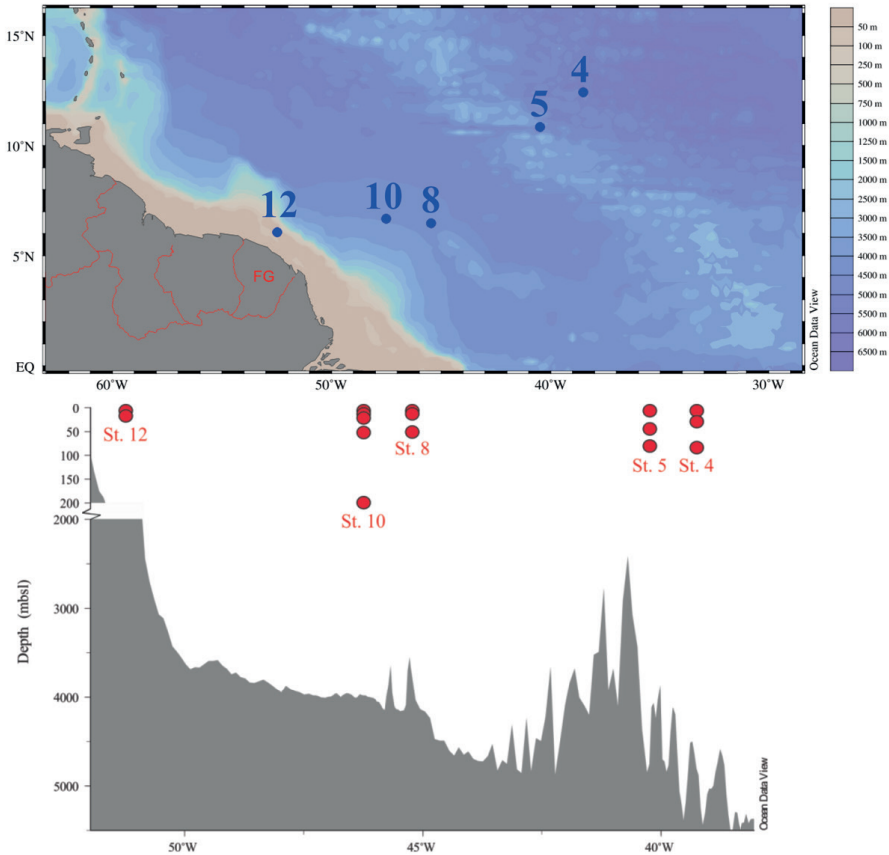
ACKNOWLEDGEMENTS

4

AUTHOR CONTRIBUTIONS

M.A.B., J.S.S.D., and L.V. designed the study; M.A.B. wrote the paper; M.A.B., E.C.H., S.S., J.S.S.D. and L.V. analyzed the data; M.A.B., N.J.B., E.C.H. performed experiments. All authors participated in data interpretation and provided editorial comments on the manuscript.

SUPPLEMENTARY FIGURES

Figure S1. (A) Map of the sampling locations (numbered and marked with blue dots) in the Northern Atlantic Ocean close to the coast of French Guiana (FG) and in the open ocean. Color coded bathymetry scale on the right. (B) Depth profile of the sampled stations with the sampling depths and the seafloor profile. Note that the depth axis is cut off in between 200 and 2000 mbsl.
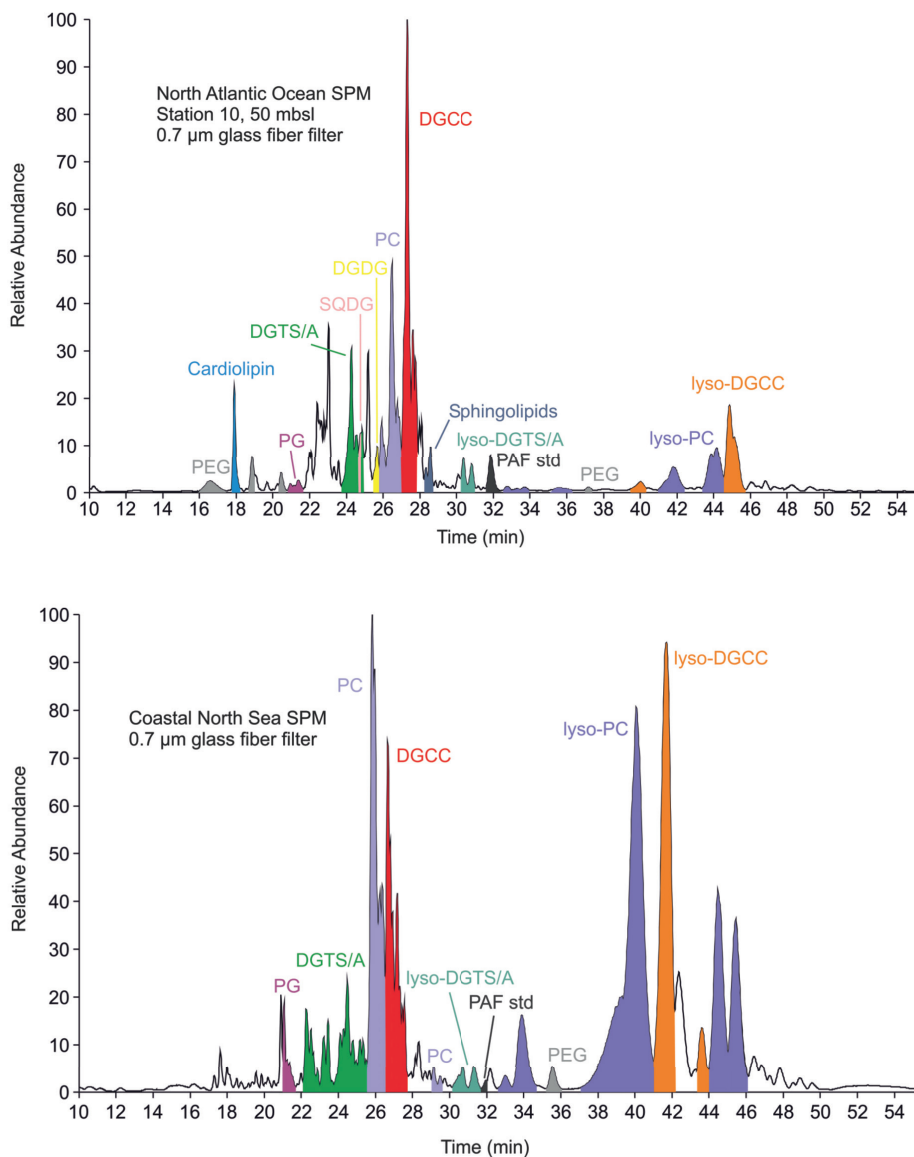
Figure S2. Partial base peak chromatogram of UHPLC-HRAM/MS analysis of intact polar lipids (IPLs) in suspended particulate matter (SPM) from A) North Atlantic Ocean SPM collected at station 10 at 50 meters below sea level (mbsl) on 0.7 μm glass fiber filters and B) coastal North Sea suspended particulate matter. Abbreviations in peak labels: PC = phosphatidylcholine, PG = phosphatidylglycerol, SQDG = sulfoquinovosyldiacylglycerol, DGTS = diacylglyceryl-trimethyl-homoserine, DGTA = diacylglyceryl-trimethyl-alanine, DGCC = diacylglyceryl-carboxyhydroxymethyl-choline, DGDG = digalactosyldiacylglycerol, PEG = polyethylene glycol (contaminant), PAF = platelet activating factor (internal standard; 1-O-hexadecyl-2-acetyl-sng-lycero-3-phosphocholine). Lyso components are IPLs in which one of the fatty acids (FA) chains is not present. We were unable to distinguish between DGTA and DGTS betaine lipids using our analytical methods so the group is referred to as DGTS/A.

Table S1. North Atlantic Ocean sampling stations (cf. Bale et al., 2018) and the Coastal North Sea with sampling depth (mbsl), the pore size of the used glass fiber filters, the amount of seawater (liter) filtered, the relative and estimated 16S rRNA gene abundances per liter of seawater filtered for the three main archaeal groups.

Station	Depth (mbsl)	Filter (pore size, µm)	Amount (l)	Relative abundance (%)			16S rRNA gene copies L^{-1}		
				MGI	MGII	MGIII	MGI	MGII	MGIII
12	5	0.7	90	18.5	69.3	12.2	1E+07	5E+07	1E+07
	15	0.7	107	19.3	65.4	15.3	9E+06	3E+07	7E+06
	5	0.7	271	3.9	88.4	7.7	1E+06	2E+07	2E+06
10	9	0.7	89	1.0	92.7	6.3	1E+05	1E+07	8E+05
	50	0.7	361	0.3	92.0	7.7	2E+06	5E+08	4E+07
	200	0.7	399	37.7	55.6	6.7	9E+06	1E+07	2E+06
	20	0.3	319	0.8	80.7	18.5	7E+04	7E+06	2E+06
10	50	0.3	307	0.5	90.0	9.5	1E+05	2E+07	3E+06
	200	0.3	381	49.1	43.0	8.0	8E+07	7E+07	1E+07
	5	0.7	234	2.4	89.9	7.7	3E+05	1E+07	9E+05
8	10	0.7	247	2.8	88.6	8.7	5E+05	2E+07	2E+06
	52	0.7	381	10.4	84.4	5.2	3E+07	3E+08	2E+07
	5	0.7	312	2.6	91.3	6.1	3E+05	1E+07	8E+05
5	40	0.7	341	1.3	83.2	15.5	2E+05	1E+07	2E+06
	80	0.7	382	34.6	59.1	6.3	1E+08	2E+08	2E+07
	5	0.7	306	1.4	86.4	12.2	2E+05	1E+07	1E+06
4	30	0.7	331	1.0	86.9	12.1	8E+04	7E+06	1E+06
	83	0.7	380	23.7	70.2	6.0	5E+07	2E+08	1E+07
Coastal North Sea	Surface	0.7	7.5 – 10	0.5	99.5	0.0	7E+06	1E+09	0E+00
	Surface	0.3	10	0.1	99.8	0.2	6E+00	7E+03	1E+01

Table S2. PCR primers, efficiency and correlation coefficient of the quantitative QPCR assays

QPCR (gene)	Primers	Ref	Efficiency	R^2
Archaeal 16S rRNA	Parch519F/ARC915R	Pitcher et al., 2011	91.8	0.94
Archaeal 16S rRNA	Parch519F/ARC915R	Pitcher et al., 2011	91.4	1

4

Table S3. Intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipid (GDGT) relative abundances within North Atlantic Ocean suspended particulate matter (SPM). Detected IPL-GDGTs were detected with the following headgroups, headgroups; monohexose (MH), dihexose (DH) and hexose-phosphohexose (HPH).

Station	Depth (mbsl)	GDGT-0		GDGT-1		GDGT-2			GDGT-3	GDGT-4	Crenarchaeaol (and Cren')		
		MH	HPH	MH	HPH	MH	DH	HPH	DH	DH	MH	DH	HPH
12	5	0.0	0.0	0.0	0.0	0.0	6.4	0.0	0.0	0.0	9.9	0.0	83.7
15	15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.7	0.0	81.3
10 (0.7 µm pore size)	200	2.1	4.1	0.7	0.0	0.7	33.0	0.0	0.0	0.0	8.1	51.3	0.0
10 (0.3 µm pore size)	200	0.2	13.1	0.0	2.9	0.0	28.2	0.0	0.0	0.0	0.6	37.2	17.7
8	52	0.0	13.6	0.0	6.3	0.2	0.0	0.0	0.0	0.0	3.7	5.5	70.7
5	80	0.3	0.0	0.0	6.1	0.0	4.3	0.0	4.8	1.1	1.3	4.9	77.3
4	83	0.0	28.1	0.0	4.9	0.0	3.2	0.8	1.5	0.0	1.0	3.1	57.3
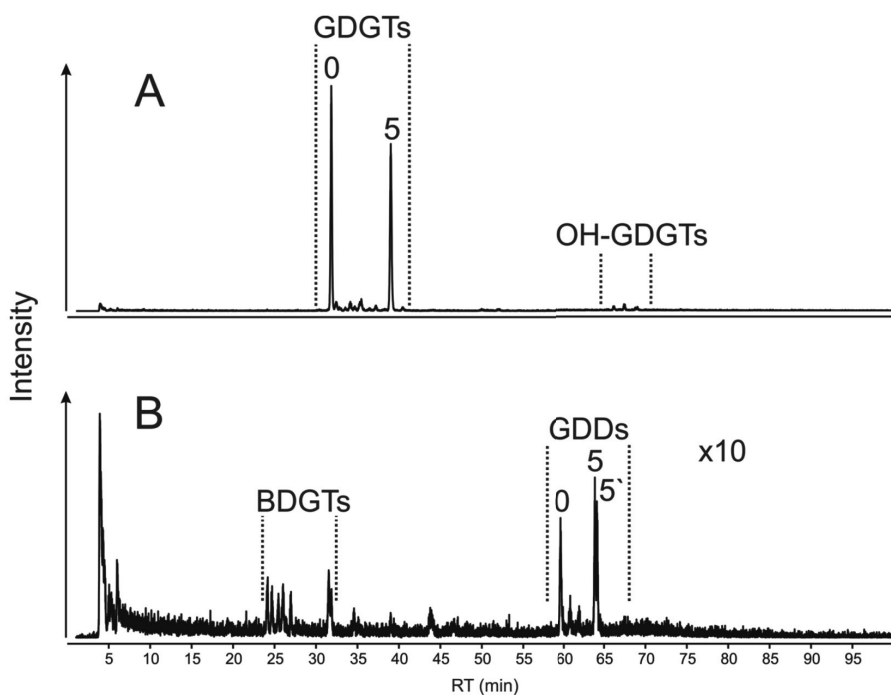
Figure S3. UHPLC-ToFMS chromatograms of the acid hydrolyzed Bligh and Dyer extract of summer coastal North Sea suspended particulate matter. (A) Combined mass chromatograms or the [M + H]+ ions of isoprenoid glycerol dialkyl glycerol tetraether lipids (GDGTs 0-8) and hydroxyl GDGTs (OH-GDGTs 0-8), within 10 ppm mass accuracy, (B) Combined mass chromatograms of [M + H]+ ions of butanetriol dibiphytanyl glycerol tetraethers (BDGTs 0-8) glycerol dialkanol diethers (GDDs 0-8), within 10 ppm mass accuracy, and 10x magnified compared to window A. Number label of peak indicates the number ofnumber of cyclic moieties, where the peaks labeled '5' are the analogues of crenarchaeol with four cyclopentyl and one cyclohexyl moiety.

CHAPTER 5

Impact of archaeal diversity and oxygen availability on ether lipid composition in marine pelagic environments

Marc A. Besseling, Ellen C. Hopmans, Stefan Schouten, Jaap S. Sinninghe Damsté and Laura Villanueva

ABSTRACT

Pelagic Archaea comprise a large part of the prokaryotic community and play an important role in elemental cycling in marine environments. The most abundant archaeal group are the Thaumarchaeota (marine group I; MGI), which membrane lipids consists of isoprenoid glycerol dibiphytanyl glycerol tetraethers (GDGTs) with various head group combinations forming intact polar lipid (IPL-) GDGTs. However, the membrane lipid composition of many other ubiquitous archaeal groups remains unknown. Furthermore, the impact of environmental variables, such as oxygen availability, on archaeal distribution and their lipids is largely unexplored. Here we investigated the archaeal community and lipid composition of the Arabian Sea oxygen minimum zone (OMZ) and the oxygenated water column of the North Atlantic Ocean and compared this to previous studies including the Eastern Tropical South Pacific (ETSP) OMZ. Thaumarchaeota were the most dominant group in the studied water columns coinciding with the presence of GDGTs including crenarchaeol, known to be synthesized by MGI. Oxygen levels affected the relative abundance of MGI with lower abundances at lower oxygen levels. The lack of cultures of MGII and the co-occurrence with MGI compromise the possibility of assigning membrane lipids to MGII Archaea. In contrast to MGI, oxygen availability did not seem to have an impact on the niche differentiation of MGII. Marine benthic group A Archaea were found both in the North Atlantic Ocean water column and within the OMZ of the Arabian Sea suggesting their presence is independent of oxygen availability while DPANN Pacearchaeota and marine Euryarchaeota group III (MGIII) were predominantly present in the oxygen-depleted waters of the ETSP and in Arabian Sea. Archaeol-based IPLs were detected in the ETSP OMZ and attributed to the MGIII as the genomes of DPANN Archaea lack genes within the lipid biosynthetic pathway. However, the lack of archaeol-IPLs in the Arabian Sea OMZ suggests that the synthesis of these type of membrane lipids are either associated to specific species of MGIII or to different environmental conditions between these two oxygen-depleted marine systems.

INTRODUCTION

Archaea are among the most widespread and abundant prokaryotes in the marine realm (DeLong, 1992; Fuhrman et al., 1992; DeLong, 2007). Three major archaeal groups are reported in the marine water column: the Thaumarchaeota Marine Group I (MGI; Francis et al., 2005) and the Euryarchaeota MGII and III. They contribute substantially to global biogeochemical cycles (see Offre et al. 2013 for an overview), e.g. the MGI play an important role in the marine nitrogen cycling with the aerobic oxidation of ammonium (Wuchter et al., 2006) and degradation of urea (Alonso-Saez et al., 2012). From metagenomic studies it is suggested that (deep water) MGII and MGIII might be aerobic heterotrophs potentially being able to degrade proteins, polysaccharides and fatty acids (Iverson et al., 2012; Li et al., 2015; Haro-Moreno et al., 2017). However, so far there are no cultured representatives of MGII and MGIII and therefore their physiology and ecological role remains speculative. MGI are mostly found at the subsurface of the water column (50-200 m; e.g. Pitcher et al., 2011a; Techtmann et al., 2015; Xia et al., 2017; Sollai et al., 2018), while MGII often dominate the archaeal community composition in surface waters (Massana et al., 2000; DeLong et al., 2006; Besseling et al., 2019). Other studies have reported the presence of other phylotypes of MGI as well as MGII and MGIII in deeper parts of the water column (e.g. Galand et al., 2009; Quaiser et al., 2011).

While DNA studies have been used to infer archaeal community diversity and abundance, archaeal lipid composition has also been used to infer their occurrence, in particular in past environments as lipids preserve better than DNA. The best studied are MGI as these make isoprenoid Glycerol Dialkyl Glycerol Tetraethers (GDGTs) with 0-4 cyclopentane moieties as membrane lipids (Schouten et al., 2008; Pitcher et al., 2010; Elling et al., 2017) and the characteristic GDGT crenarchaeol, which contains a cyclohexane moiety in addition to 4 cyclopentane moieties and is considered a specific biomarker for MGI (Sinninghe Damsté et al., 2002). Within the cell GDGTs occur as intact polar lipid (IPL) i.e. with various polar head groups (e.g. Langworthy et al., 1974; Sugai et al., 1995; Sinninghe Damsté et al., 2012). Some of these IPL-GDGTs are used as a biomarker for living Archaea, e.g. hexose

phosphohexose (HPH) crenarchaeol for Thaumarchaeota (Pitcher et al., 2011a). Some studies have observed that Thaumarchaeota MGI gene abundances match GDGT crenarchaeol profiles within different environments (Leininger et al., 2006; Pitcher et al., 2011a & 2011b). However, other studies have reported lack of correspondence between HPH-GDGTs and thaumarchaeotal gene abundances suggesting a potential contribution of fossil lipid material (Schouten et al., 2012; Buckles et al., 2013; Sollai et al., 2018 & 2019). The ether lipid composition of MGII and MGIII is so far unknown due to a lack of culture representatives. MGII has been suggested by Lincoln et al. (2014) to produce GDGTs, including crenarchaeol. However, these findings have been questioned by others (Schouten et al., 2014; Lincoln et al., 2014b; Besseling et al., submitted). Furthermore, environmental factors are also known to impact archaeal lipid composition, e.g. temperature (Gliozzi et al., 1983; Schouten et al., 2002; Elling et al., 2015) and pH (Shimada et al., 2008; Boyd et al., 2011; Elling et al., 2015). Oxygen concentration is known to effect the GDGT composition of MGIs (Qin et al., 2014), although this may be an indirect effect of a change in ammonium oxidation rates (Hurley et al., 2016). Thus, the combined effect of different archaeal GDGT producers and environmental factors can affect the GDGT lipid composition within the marine water column but to which extend is not well constrained.

Here, we studied the archaeal community diversity by using 16S rRNA gene amplicon sequencing and IPL-GDGT composition of the oxygen minimum zone of the Arabian Sea, and compare these results with those previously obtained from the oxygen minimum zone of the eastern tropical South Pacific (ETSP; Sollai et al., 2019). Furthermore, we studied the archaeal community diversity and IPL-GDGT composition of a fully oxygenated water column, i.e. the North Atlantic Ocean, which allows us to infer the impact of oxygen availability on the archaeal community composition and ether lipid composition in marine waters.

MATERIALS AND METHODS

Sampling and physicochemical analyses

Suspended particulate matter (SPM) samples were collected in the northern Arabian Sea during the PASOM cruise in January 2009 with *R/V* Pelagia. SPM was collected as previously described by Pitcher et al (2011). SPM from the North Atlantic was collected during the HCC cruise in September 2014 with *R/V* Pelagia as reported in Bale et al. (2018). All filters were stored at -80°C until further analysis. For an overview of the filters used and physicochemical conditions, see Supplementary Table S1.

Lipid extraction

Total lipids were extracted from freeze dried GF/F filters using a modified Bligh and Dyer method (Bligh & Dyer, 1959) as previously described by (Lengger et al., 2012). A C_{16}-PAF (1-O-hexadecyl-2-acetoyl-sn-glycero-3-phosphocholine) standard was added to the extracts as an internal standard and dried under a stream of nitrogen. The extracts with the added standard were then dissolved by adding IPL injection solvent (hexane:isopropanol:H_2O 718:271:10 vol/vol/vol) and filtered through a 0.45 μm, 4 mm-diameter True Regenerated Cellulose syringe filter (Grace Davison, Columbia, MD, USA).

IPL extraction and analysis

IPLs were analyzed according to Besseling et al. (2018), a method based on Sturt et al. (2004). Briefly, samples were analyzed using an Ultimate 3000 RS UHPLC, equipped with thermostated auto-injector and column oven, couple to a Q Exactive Orbitrap MS with Ion Max source and heated electrospray ionization (HESI) probe (Thermo Fisher Scientific, Waltham, MA), was used. Separation was achieved on a YMC-Triart Diol-HILIC column (150 × 2.0 mm, 1.9 μm particles, pore size 12 nm; YMC Co., Ltd, Kyoto, Japan) with a guard column of the same material (10 × 2.1 mm). The same inclusion list was used as in Besseling et al. (2018). All samples were run in duplicate.

5

Peak areas for each individual IPL were determined by integrating the combined mass chromatogram (within 3 ppm) of the monoisotopic and first isotope peak of all relevant adducts formed (protonated, ammoniated and/or sodiated adducts may be formed in different proportions depending on the type of IPL). The PAF standard was used to assess LC-MS performance and potential matrix effects. IPLs of crenarchaeol and crenarchaeol isomer are not separated and thus peak areas reported for any crenarchaeol IPL thus represents the sum of both isomers.

Absolute quantification of IPL GDGTs was not possible due to the large differences in response factors between IPLs and a lack of standards. Thus, it is not possible to calculate the relative abundance of each IPL in the total pool of IPLs. Therefore we calculated for each IPL the response units L^{-1} of filtered sea water and then normalized this over all samples with depth. This approach allows to see at which depth each IPL maximizes in abundance.

Ring index values were calculated according to Pearson et al. (2004):

$$\text{Ring index} = \frac{[\text{GDGT} - 1] + 2 \times [\text{GDGT} - 2] + 3 \times [\text{GDGT} - 3] + 5 \times [\text{Crenarchaeol}]}{[\text{GDGT} - 0] + [\text{GDGT} - 1] + [\text{GDGT} - 2] + [\text{GDGT} - 3] + [\text{Crenarchaeol}]}$$

DNA extraction, 16S rRNA gene amplicon sequencing and phylogeny

DNA from Arabian Sea and North Atlantic Ocean were extracted from the GF/F filters with the RNA PowerSoil® Total Isolation Kit plus the DNA elution accessory (Mo Bio Laboratories, Carlsbad, CA). Concentration of DNA of the Arabian Sea and the North Atlantic Ocean was quantified by Qubit fluorometric quantitation (ThermoFisher Scientific).

The PCR reactions were performed with the universal, Bacteria and Archaea, primers S-D-Arch-0159-a-S-15 and S-D-Bact-785-a-A-21 (Klindworth et al., 2013) as described in Moore et al., (2015). The archaeal 16S rRNA gene amplicon sequences were analyzed by QIIME v1.9 (Caporaso et al., 2010). Raw sequences were demultiplexed and then quality-filtered with a minimum quality score of

25, length between 250-350, and allowing maximum two errors in the barcode sequence. Taxonomy was assigned based on blast and the SILVA database version 123 (ESTP) and 128 (Arabian Sea and North Atlantic Ocean) (Altschul et al., 1990; Christian Quast et al., 2013). Representative operational taxonomic units (OTUs) sequences of archaeal groups were extracted through filter_taxa_from_otu_table.py and filter_fasta.py with QIIME (Caporaso et al., 2010). The phylogenetic affiliation of the partial archaeal 16S rRNA gene sequences was compared to release 123 of the Silva NR SSU Ref database (http://www.arb-silva.de/; Quast et al., 2013) using the ARB software package (Ludwig et al., 2004). Sequences were added to the reference tree supplied by the Silva database using the ARB Parsimony tool. Affiliation of any 16S rRNA gene sequences to a given subgroup was done assuming a similarity cutoff of ≥85%. Based on the phylogenetic trees OTUs were grouped into subgroups. Relative abundances of the subgroups were calculated by summing the total amount of reads of each subgroup and dividing this by the total amount of archaeal reads at each depth interval. This allows to infer the dominance of each subgroup at each depth interval.

RESULTS

In this study, we analyzed the IPL ether lipid and the archaeal community composition of the oxygen minimum zone of the Arabian Sea and of the fully oxygenated water column in the North Atlantic Ocean (see map; Supplementary Figure S1). A full description of lipid diversity and archaeal community composition of the ETSP can be found in Sollai et al. (2019).

Archaeal diversity

The SPM of the studied regions contained multiple archaeal groups which could be subdivided into different subgroups using phylogenetic trees (see Supplementary Figures S2-4). In the Arabian Sea the main archaeal groups were MGI, MGII and MGIII with reads belonging to Marine Benthic Group A (MBG-A), mostly detected in the center of the OMZ (up to 10 % of total archaeal reads; Figure 1). The *Ca. Nitrosopumilus* group (Supplementary Figure S2) was the most dominant

MGI subgroup within and below the OMZ (ranging between 21 and 78 % of the total archaeal reads; Figure 1). Above the OMZ, MGI subgroup 1 was the most abundant archaeal group (38 − 53 %; Figure 1). Several subgroups of MGII were detected with MGII subgroup 2 predominantly at the top of and above the OMZ (6 − 19 %), while subgroup 4 had the highest relative abundance at 2000 mbsl (13 %) and subgroup 5 dominated at and below the lower boundary of the OMZ (7 − 12 %; Figure 1). Subgroup 1 of the MGIII was detected throughout the water column (3 − 10 %), while MGIII subgroup 3 was relative abundant in the center of the OMZ (11 − 26%; Figure 1).

The archaeal community within the water column of the North Atlantic Ocean consisted of MGI, MGII and MGIII and minor amounts (<3%) of DPANN (superphylum composed of Micrarchaeota, Diapherotrites, Aenigmarchaeota, Nanohaloarchaeota, Parvarchaeota, Nanoarchaeota, Pacearchaeota (also known as Deep Hydrothermal Vent group 6 (DHVE-6)) and Woesearchaeota; Rinke et al., 2013; Castelle et al., 2015)) and MBG-A (Figure 1). The MGI consisted of only MGI subgroup 1 and MGI reads were not detected in the (sub)surface (20 and 50 mbsl), as previously reported by Besseling et al. (2019). MGII consisted of 4 subgroups with subgroups 2 and 3 relative abundant at the (sub)surface (ranging between 17 and 62 %), whereas subgroup 4 had the highest relative abundance at 3000 mbsl (25 %; Figure 1). Reads belonging to subgroup 5 were not detected at 20 and 50 mbsl but were present throughout the rest of the water column. Reads belonging to five MGIII subgroups were detected at depths ranging from 200 − 3000 mbsl, while only subgroup 4b was present in the (sub)surface (Figure 1).

Region	Depth (mbsl)	MGI			MGII					MGIII					DPANN DHVE-6	MBG-A
		Ca. Nitrosopelagicus	*Ca.* Nitrosopumilus	subgroup-1	subgroup-1	subgroup-2	subgroup-3	subgroup-4	subgroup-5	subgroup-1	subgroup-2	subgroup-3	subgroup-4a	subgroup-4b		
Arabian Sea (OMZ)	20			52	ND	19		4		2	ND			2	1	
	170	2	24	38		6		0		4	2			2		
	300	12	60	7		9	2	0	2	3	1	3	0		0	
	450	7	45	3		2	0	0	3	8		0	11		15	4
	600	7	30			0	0	0	1	9		0	20		19	8
	750	3	21			0	0	0	0	9			26		21	9
	900	6	68			1	0	0	7	6			7		2	2
	1050	7	69	0		0	1	1	12	4		2	2		0	
	1200	4	70			0	5	2	12	3		0	1		1	
	1350	4	64			0	4	1	9	5		0	2		0	
	1500	3	78	0		0	2		3	3		0	3		1	
	2000	3	63			2		13	1	7		1	4		1	1
North Atlantic Ocean	20	ND	ND	ND	36	34	7						20			
	50				17	62	1	1					7			
	200			48	23	7	0		9	1	1	4	0		1	
	400			73	3	1	0		15	1	0	1	1	0	0	0
	600			67	1	0	0		17	3	1		1		1	
	800			67	0	0	0		13	3	1		3	0	1	0
	1000			52	0	0	0		21	4	1	0	4		2	0
	1200			57	0	1	0		19	4	1	0	3	0	1	1
	1500			65	0	3	1		13	2	0	0	4	0	0	1
	2000			49	0	9	1		13	1	1	0	6	1	1	1
	2500			69	0	6	2		5	1	2		5	0	2	2
	3000			58	3	7	25		2		1		1		1	

5

	MGI			MGII					MGIII					DPANN DHVE-6	MBG-A
Depth (mbsl)	Ca. Nitrosopelagicus	Ca. Nitrosopumilus	subgroup-1	subgroup-1	subgroup-2	subgroup-3	subgroup-4	subgroup-5	subgroup-1	subgroup-2	subgroup-3	subgroup-4a	subgroup-4b		
ETSP coastal															
22	55	25	4	7	7		ND	ND		2	1	ND	ND	0	ND
75	17	19	3	41	6				0	7	3			2	
83	28	36	5	17	5				1	2	3			2	
OMZ: 100	11	17	16	7	6				3	6	16			12	
135	4	6	12	1	4				8	8	23			24	
150	1	3	7		10				8	10	29			22	
175	2	2	13	4	6				10	10	22			23	
200	1	1	13		5				15	12	25			18	
235	2	2	3		2				17	15	37			22	
250		3	13	5	4				14	6	22			26	
300		0	11	0	3				14	13	31			16	
350	0	2	8		5				21	14	27			15	
500	0	5	53	0	32				1	2	5			0	
1000		6	53	1	29				2	2	7			1	
ETSP open ocean															
100	33	16	16	4	18	5				5	2				
125	23	18	21	4	17	5			0	8	2			0	
OMZ: 155	5	13	21	4	14	3			7	9	12			10	
175	2	2	13	4	6				10	10	22			23	
200	1	1	13		5				15	12	25			18	
235	2	2	3		2				17	15	37			22	
300		0	5		3				31	14	22			17	
500	0		55	14	13	1			2	3	5			1	
1000	0		52	21	6	1			4	1	9			2	

Figure 1. Detected archaeal groups and subgroups per depth (in meters below sea level) in the Arabian Sea, the North Atlantic and Eastern Tropical South Pacific (ETSP). Color indicates the relative abundance archaeal subgroup (percentage of total archaeal reads at a particular depth interval) , with no color indicating a low and red indicating a high percentage of the archaeal subgroup at a particular depth. Oxygen minimum zones (OMZs) within the ETSP and Arabian Sea water columns are indicated with dashed lines. Archaeal subgroups are defined based on 16S rRNA gene phylogeny as shown in supplementary figures S2-4. ND = none detected.

Archaeal intact polar lipid diversity and distribution

In the Arabian Sea, multiple IPL-GDGTs were detected, i.e. GDGT-0 to -4 and crenarchaeol with monohexose (MH), dihexose (DH) and hexose-phosphohexose (HPH) headgroups (Figure 2). The highest relative abundance of IPLs with

GDGT-3 and -4 core lipids were detected in the (sub)surface (at 170 mbsl) above the OMZ and declined within the core of the OMZ (Figure 2). IPL-crenarchaeol with DH and HPH headgroups were reaching maximum relative abundances above the OMZ at 170 mbsl (Figure 2). In contrast, the crenarchaeol with an MH headgroup had maximum relative abundance in the OMZ, at 600 mbsl (Figure 2). Relative abundances of GDGT-0 and -1 with a DH headgroup were increasing with depth with highest abundance at 2000 msbl (Figure 2). In contrast, GDGT-0 and GDGT-1 with an MH or with an HPH headgroup were more equally distributed with a maximum within the OMZ (Figure 2). IPLs with an archaeol core lipid were not detected in the Arabian Sea water column (Figure 2).

The North Atlantic Ocean water column contained a similar set of IPL-GDGTs as those in the Arabian Sea (Figure 2), but no IPLs were detected with GDGT-3 and -4 as core lipids (Figure 2). Furthermore, no IPL-GDGTs were detected at all at 20 and 50 mbsl (cf Besseling et al., 2019a). GDGT-0, -1, -2 and crenarchaeol with the headgroup HPH were all most abundant at 400 mbsl (Figure 2). DH-GDGT-2 and DH-crenarchaeol were detected with a maximum relative abundance at 200 mbsl and this decreased with greater depth (Figure 2). MH-GDGT-0 and MH-crenarchaeol, were also peaking at 200 mbsl and their relative abundance decreased with depth. However MH-GDGT-1 and -2 were observed with a maximum relative abundance at 800 mbsl (Figure 2). No IPLs with an archaeol core lipid were detected (Figure 2).

5

Depth (mbsl)	GDGT-0			GDGT-1			GDGT-2			GDGT-3		GDGT-4		Crenarchaeol			Archaeol			
	MH	DH	HPH	MH	DH	HPH	MH	DH	HPH	DH	HPH	DH	HPH	MH	DH	HPH	MH	DH	PG	PE
Arabian Sea (OMZ)																				
20	2		2	1		3	1	0	1	15	13	14		5	1	5	ND	ND	ND	ND
170	8		11	3	7	11	7	10	13	34	22	45	48	9	24	18				
300	9		7	8		9	10	8	16	9	12	12	3	11	15	10				
450	8		12	9	2	15	11	10	19	14	9	14	22	9	15	15				
600	10		10	13		13	15	8	17	3	10	1	3	20	8	12				
750	10		6	14		8	19	8	12	5	8	2	7	16	7	8				
900	16		6	25	4	8	32	12	12	8	12	4	9	17	9	7				
1050	5	4	14	3	10	13	2	12	5	4	12	4	4	2	8	11				
1200	4	15	11	3	15	8	1	11	3	3	2	2	3	2	6	6				
1350	10	12	8	9	17	6	2	9	2	1		1	1	4	3	3				
1500	9	27	8	6	19	5	1	8	1	2		2		4	3	2				
2000	8	42	5	4	28	3	0	5	0	1				2	1	1				
North Atlantic Ocean																				
20										ND	ND	ND	ND				ND	ND	ND	ND
50																				
200	15		6	7		4	17	22						29	50	9				
400	13	32	62	7	24	62		21	79					16	23	67				
600	13	24	4	9	17	4		15	1					12	12	3				
800	15	31	19	27	18	22	34	13	20					15	8	17				
1000	10		4	15	7	3	19	8						8	4	2				
1200	8		2	11	5	2	15	7						7	2	1				
1500	9	7	2	12	7	2	15	6						6	2	1				
2000	4	3	1		5	0	2	4						1	0	0				
2500	7		0	3	6			1						3						
3000	6	2	0	9	11	0		3						3	0	0				

Figure 2. Detected archaeal intact polar lipid (IPL-) glycerol dialkyl glycerol tetraether lipids (GDGTs) in the water columns of the Arabian Sea, the North Atlantic and Eastern Tropical South Pacific (ETSP). Relative abundances are normalized per IPL-GDGT over depth. Color indicates the relative abundance per IPL-GDGT over the water column, with no color indicating a low and red indicating a high relative abundance of the IPL-GDGT at a particular depth. Oxygen minimum zones (OMZs) within the ETSP and Arabian Sea water columns are indicated with dashed lines. ND = none detected.

DISCUSSION

In this study, we analyzed the IPL-GDGTs and the archaeal community composition of the oxygen minimum zone of the Arabian Sea and of the fully oxygenated water column in the North Atlantic Ocean. We compare our results with those of previous studies of the same regions as well as with our previous study of archaeal genes and lipid composition of the ETSP OMZ (Sollai et al., 2019)

Archaeal diversity in oxygen minimum zones

The water column of the Arabian Sea was dominated by MGI Archaea except for the core of the OMZ (Figure 1). The presence of Thaumarchaeota within the OMZ is likely because Thaumarchaeota are able to cope with low levels of oxygen present at the redoxcline of the OMZ (Lam et al., 2009; Stewart et al., 2012; Bristow et al., 2017). Our results show that the MGI community within the Arabian Sea OMZ is composed predominantly of species related to *Ca. Nitrosopumilus* and, in lower relative abundance, those related to *Ca. Nitrosopelagicus* (Figure 1). MGI subgroup-1, comprising uncultivated species, was a relatively more abundant component of the archaeal community in the (sub)surface in the oxygenated water above the Arabian Sea OMZ.

MGII Archaea occurred at the same depths as MGI in the Arabian Sea and were also in lower relative abundance in the OMZ, similar to a previous metagenomic study of a selection of the same samples as studied here (Lüke et al., 2016). The high relative abundance of MGII at the surface is most likely caused by the photoheterotrophic lifestyle of the Euryarchaeota (Iverson et al., 2012; Martin-Cuadrado et al., 2014; Orsi et al., 2015). Below the OMZ, MGII were again a substantial part of the archaeal community composition, as these deep-sea MGII are most likely utilizing proteins, carbohydrates and lipids for growth (Li et al., 2015).

Within the core of the Arabian Sea OMZ a high relative abundance of MGIII Archaea was detected (Figure 1), in agreement with DGGE results of Schouten et al. (2012). MGIII Archaea have been associated with oxygenated meso- and bathypelagic environments (Fuhrman and Davis, 1997; Massana et al., 2000; López-

García et al., 2001; Bano et al., 2004) but are also detected within the photic zone of (sub)-surface waters (Delong et al., 2006; Haro-Moreno et al., 2017), in OMZs of e.g. the ETSP (Belmar et al., 2011; Stewart et al., 2012; Sollai et al., 2019) and of the Black Sea (Sollai et al., 2018). Metagenomic data of MGIII does not indicate a facultative anaerobic lifestyle (Haro-Moreno et al., 2017) which has been suggested for MGII (Martin-Cuadrado et al., 2008). However, the presence of MGIII OTUs within OMZs, and the observation that MGIII is sometimes almost exclusively detected within the OMZ core (Belmar et al., 2011; Sollai et al., 2019), would suggest that some MGIII members have a(n) (facultative) anaerobic lifestyle.

DPANN Pacearchaeota were detected with a high relative abundance within the Arabian Sea OMZ (Figure 1). These Archaea have been detected in other oxygen-depleted marine systems such as the Black Sea (Sollai et al., 2018) and in a deep anoxic layer water layer within the Mediterranean Sea (Yakimov et al., 2007), although they have also been reported in oxygenated waters such as surface coastal waters from the Mediterranean Sea (Hugoni et al., 2013). Metagenomic studies of DPANN Archaea suggest that they have small genome size and limited metabolic capabilities (Castelle et al., 2015), and therefore it has been proposed that these Archaea have a symbiotic or parasitic lifestyle (Rinke et al., 2013; Castelle et al., 2015). This would mean that their occurrence is connected to certain unknown host organisms.

Reads belonging to MBG-A were detected within the OMZ of the Arabian Sea (Figure 1). MBG-A Archaea are mostly detected in marine sediments (Vetriani et al., 1999; among others) but also they have been detected in marine waters (e.g. Belmar et al., 2011; Kaneko et al., 2016). MBG-A seem to be closely affiliated to MGI, based on 16S rRNA gene phylogeny (Kaneko et al., 2016), but their role in both benthic and pelagic marine environment remains unknown.

The results of the Arabian Sea OMZ can be compared with those of the ETSP, obtained using identical methods, previously reported by Sollai et al. (2019) (Figure 1). The archaeal community in the ETSP is also dominated by Thaumarchaeota MGI (except at 75 mbsl at the coastal station) in the oxygenated part of the water

column (Figure 1), consistent with Belmar et al. (2011), who shows a dominance of MGI at the upper oxycline, and with Quinones et al. (2009), who reports a high abundance of MGI in the upper part of coastal water located near central-southern Chile. Both in the Arabian Sea and ETSP OMZs MGI Archaea are relatively more prominent at the OMZ borders (Figure 1), in agreement with earlier reports of these regions (Molina et al., 2010; Newell et al., 2011; Pitcher et al., 2011a). This seems contradictory with their role as ammonia-oxidizers. However, members of the Thaumarchaeota group are able to thrive within OMZs (Lam et al., 2007; Beman et al., 2008; Lam et al., 2009; Swan et al., 2014; Gillies et al., 2015). Therefore, it has been suggested that Thaumarchaeota have the capacity to grow in oxygen depleted waters (Lam et al., 2009; Stewart et al., 2012). The MGI community compositions between the Arabian Sea and the ETSP differ substantially, e.g. MGI subgroup-1 dominated the surface waters of the Arabian Sea and are not detected at greater depths, whereas it dominated the community at greater depths of the ETSP (Figure 1). These differences in the MGI community composition within the Arabian Sea and ETSP have also been reported by Peng et al. (2013), showing significantly different ammonia oxidizing archaeal communities, based on the ammonia monooxygenase gene subunit A phylogeny, between the two environments. It has been suggested that temperature was a likely parameter to cause this difference, while other physicochemical factors had a lesser effect on the community compositions (Peng et al., 2013). The temperature of the Arabian Sea water column was substantially higher (> 4°C) compared to the ETSP water column, and could thus explain the differences in the MGI community composition (Supplementary Table S1). Also, the salinity is somewhat higher in the Arabian Sea (Supplementary Table S1). There were no major differences observed in nutrients, with only the nitrite levels were elevated at some depths of the ETSP, compared to the Arabian Sea (Supplementary Table S1).

Both the Arabian Sea and the ETSP OMZs contained a high relative abundance of MGIII (Figure 1), although the MGIII community was largely composed of different subgroups between the two OMZs (Figure 1). Higher relative abundances of MGIII

Archaea were also observed within the euxinic zone of the Black sea (Sollai et al., 2018).

The ETSP OMZ contained a high relative abundance of DPANN Pacearchaeota Archaea, which were also detected within the OMZ of the Arabian Sea (Figure 1). In both OMZs the occurrence of Pacearchaeota overlapped with the presence of MGIII subgroup 1 (Figure 1). Pacearchaeota have also been detected within anoxic sediments in high relative abundances, but in absence of MGIII Archaea (Lipsewers et al., 2017; Besseling et al., 2018), making it unlikely that this group is associated with MGIII Archaea. It could be that the DPANN Pacearchaeota within the sediments are substantially different then their pelagic counterparts and perhaps rely on (a) different microbial group(s) for their symbiotic or parasitic lifestyle.

Thus, in summary, the overall archaeal community composition between OMZs was generally similar although there were substantial intergroup differences and MBG-A Archaea were not detected within the OMZ of the ETSP (Figure 1).

Comparison of archaeal diversity between OMZs and North Atlantic

Similar to the water columns of the Arabian Sea and the ETSP, both containing an OMZ, the archaeal community in the fully oxygenated water column of the North Atlantic Ocean is also dominated by MGI (Figure 1). However, the MGI community was less diverse and composed solely of subgroup-1 (Figure 1). This subgroup was also in high relative abundances detected in the (sub)surface of the Arabian Sea and below the OMZ of the ETSP, both oxygen-rich parts of the water column (Figure 1). Oxygen levels within the North Atlantic Ocean were substantially higher in the water column from 200 msbl downwards compared to the Arabian Sea and the ETSP, which would explain the dominance of MGI Archaea from 200 mbsl downwards (Figure 1). Temperatures were similar to that of the ETSP, between 200 and 1000 mbsl (Supplementary Table S1). It thus remains uncertain what causes the differences in the MGI community composition between the different regions.

Similar to the archaeal community composition in the Arabian Sea and ETSP, there was a high relative abundance of MGII Archaea in (sub)surface waters of the North

5

Atlantic Ocean (Figure 1). Furthermore, clear depth niches of the MGII subgroups within the North Atlantic Ocean are observed, with e.g. MGII subgroup-5 having the highest relative abundance at in the mesopelagic zone and subgroup-4 at the greatest depth, similar to what was observed in the Arabian Sea (Figure 1). Therefore, it seems that the MGII community is minimally affected by the presence or absence of oxygen. In the North Atlantic Ocean, subgroup-4b was the most dominant MGIII, with the highest relative abundance at the surface (Figure 1). Within the fully oxygenated water column of the North Atlantic Ocean there were less well-defined depth niches of MGIII subgroups compared to the Arabian Sea and the ETSP which would imply that oxygen has an effect on the MGIII community composition (Figure 1).

DPANN Pacearchaeota were also detected within the North Atlantic Ocean but in low relative abundances compared to the core of the OMZs (Figure 1). These Archaea are also observed in oxygenated fresh water in high relative abundance, e.g. within surface waters of high-altitude Pyrenean lakes (Ortiz-Alvarez and Casamayor, 2016). DPANN Pacearchaeota within the North Atlantic co-occurred with MGIII subgroup-2, instead of MGIII subgroup-1, as was the case with the Arabian Sea and ETSP OMZs (Figure 1). This could mean that they had different hosts or symbiosis with other Archaea, driven by the oxygen gradient.

IPL-GDGT diversity in oxygen minimum zones

The use of high resolution mass spectrometry has broadened the scope of the IPL-GDGT composition as multiple IPL-GDGTs were detected within the water column of the Arabian Sea, in contrast to earlier studies in e.g. the Arabian Sea (Pitcher et al. (2011) and Schouten et al. (2012)) which targeted specific IPLs only, such as HPH-crenarchaeol. Within the water column of the Arabian Sea different distributions of crenarchaeol-based IPLs are observed, e.g. MH-crenarchaeol was peaking below the OMZ and DH- and HPH-crenarchaeol within OMZ, as also observed by Pitcher et al. (2011) and Schouten et al. (2012). This was suggested to be caused by diagenesis, e.g. the loss of a phosphate headgroup in HPH-crenarchaeol resulting in the formation MH-crenarchaeol. However, this diagenesis would then also apply

to IPLs with GDGT-0 as a core lipid, but the maximum relative abundance of MH-GDGT-0 is detected higher in the water column of the Arabian Sea compared to the maximum of HPH-GDGT-0, opposite to that of IPL crenarchaeol (Figure 2). It could be that MH-GDGT-0 resulted from a different pool of IPL-GDGT-0, with unknown and therefore undetected headgroup(s). However, the lack of MH-GDGT-3 and -4 and the occurrence of HPH-GDGT-3 and -4 within the Arabian Sea does not support the hypothesis of diagenesis and MH-GDGTs are most likely formed by Archaea living at these depths

HPH-GDGT-3 and -4 were only detected in the Arabian Sea water column and not within the ETSP and the Black Sea (Sollai et al., 2018). However, they were reported in the water column of the Western Atlantic Ocean (Hurley et al., 2018) and these IPL-GDGTs are known to be produced by marine pelagic Thaumarchaeota (Elling et al., 2014; Elling et al., 2017). Therefore, it is likely that the HPH-GDGT-3 and -4 were produced by Thaumarchaeota in the Arabian Sea water column but perhaps not in the other environments because of different physicochemical factors influencing the IPL composition or the absence of specific Thaumarchaeota species, which could not be differentiated based on 16S rRNA gene phylogeny.

Sollai et al. (2019) observed multiple archaeol-based IPLs mainly within the OMZs of the ETSP (Figure 2). These archaeol-based IPLs were also detected in the OMZ of the Black Sea (Rossel et al., 2008; Elling et al., 2016; Sollai et al., 2018). The archaeol-based IPLs were not detected within the OMZs of the eastern tropical North Pacific (ETNP; Xie et al., 2014) nor in the Arabian Sea (Schouten et al., 2012; this study; Figure 2). Sollai et al. (2019) hypothesized that members of MGIII are producers of the archaeol-based IPLs. MGIII reads that were closely affiliated to those detected in ETSP OMZs were also detected in the OMZ of the Arabian Sea but archaeol-based IPLs were not detected (Figure 2). However, the reads observed in the ETSP OMZ belonged to MGIII subgroup-2 which were not observed in the Arabian Sea water column (MGIII reads there belonged to subgroups 1,2,4a and b), Thus, it could be that this subgroup of MGIII is responsible for the presence of archaeol-based IPLs within the ETSP. However, this subgroup has been detected

5

in low relative abundance in the North Atlantic (Figure 1), where no archaeol-based IPLs were detected, suggesting that this group is also not responsible for these IPLs. Archaeol-based IPLs with MH and phosphohexose (PH) as headgroups have also been observed in multiple Thaumarchaeota, including marine Thaumarchaeota belonging to the *Nitrosopumilus* genus (Elling et al., 2015; Elling et al., 2017). Reads belonging to this group were detected within the water column of the ETSP and the Arabian Sea, and thus the archaeol-based IPLs within the ETSP could be attributed to specific species of *Nitrosopumilus*, or the effect of different physicochemical factors on the lipid composition.

Reads belonging to MBG-A were detected with high relative abundance (up to 10 % of total reads) within the core of the Arabian Sea OMZ (Figure 1), but there were no specific archaeal lipids maximizing at these depths. Finally, DPANN Pacearchaeota members were only in minor (0 − 2 %) relative abundances and are unlikely to be an important source for lipids (Figure 1). Most DPANN members (except the Ca. Micrarchaeota and the genome of *Ca. Iainarchaeum andersonii*) lack the biosynthetic genes needed for synthesizing archaeal lipids (Jahn et al., 2004; Villanueva et al., 2017).

Comparison of IPL-GDGT diversity in OMZs with North Atlantic

Most of the IPL-GDGTs that were detected within the water columns of the Arabian Sea and the ETSP were also detected within the North Atlantic Ocean and thus there seems no major impact of O_2 on archaeal lipid diversity (Figure 2). The only difference was that in the North Atlantic Ocean no IPLs were detected with GDGT-3 or -4 as core lipids, in contrary to the water columns of the Arabian Sea and the ETSP (Figure 2). However, Hurley et al. (2018) reported the presence of IPL-GDGT-3 and -4 in the water column in the North Atlantic Ocean, while IPL-GDGT-3 and -4 were detected in other fully oxygenated water columns (Kim et al., 2016; Besseling et al., 2019b).

Interestingly, for cultures of marine Thaumarchaeota it was observed that the relative abundance of GDGT-1 decreases and that the relative abundance of GDGT-2

and GDGT-3 increases with decreasing O_2 concentrations (Qin et al., 2015). This resulted in increasing ring index values with decreasing O_2 concentrations (Qin et al., 2015). To evaluate this, we calculated the ring index values for GDGTs with the same head group, which was feasible for DH-GDGTs and HPH-GDGTs (Figure 3). For the Arabian Sea an increase is observed of ring index values with decreasing oxygen concentrations for the DH- and HPH-GDGTs (Figure 3). Also within the fully oxygenated water column of the North Atlantic Ocean it seems that DH- and HPH-GDGT based ring index values are increasing with decreasing oxygen concentrations. For the water column of the ETSP this effect was less obvious. These results suggests that oxygen concentrations may have an effect on the (IPL-) GDGT composition of Thaumarchaeota, although other physicochemical factors or archaeal community differences could also be a cause.

5
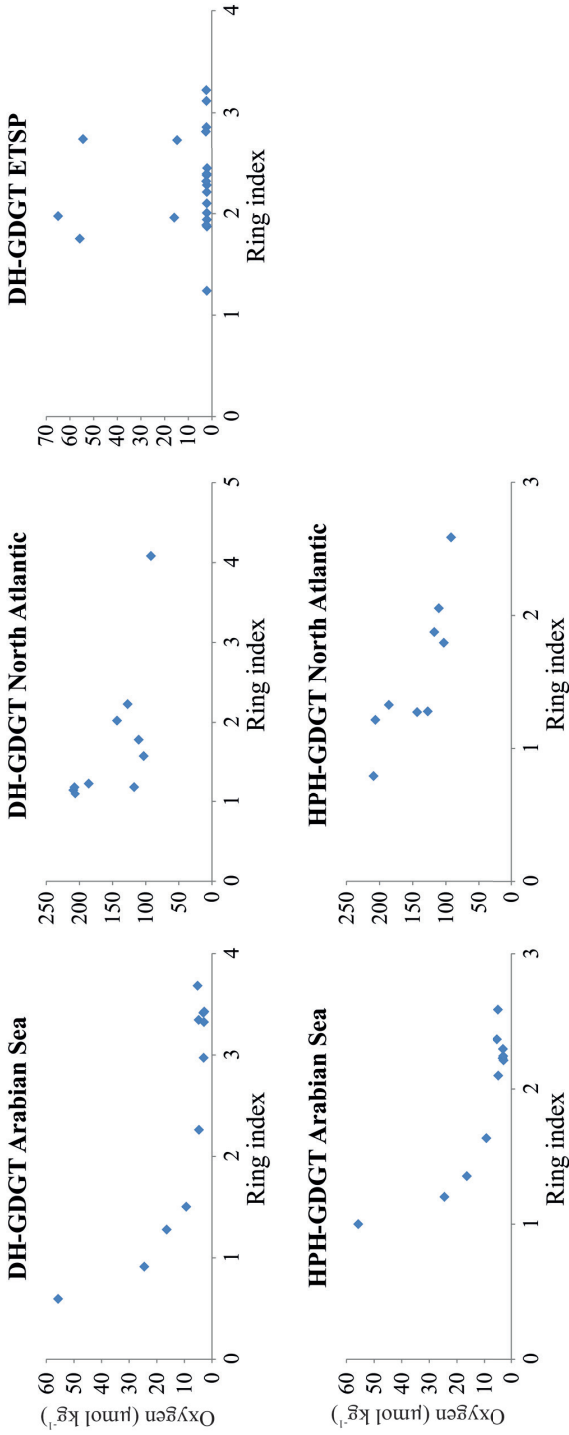
Figure 3. Oxygen concentrations and ring index values of the dihexose (DH-) GDGT in the Arabian Sea, North Atlantic Ocean and the Eastern Tropical South Pacific (ETSP) and the hexose phosphohexose (HPH-) GDGT in the Arabian Sea and North Atlantic Ocean. Data of the coastal and open ocean sites of the ETSP are combined.

Conclusion

By comparing the archaeal community composition with the IPL diversity, using a combined 16S rRNA gene amplicon sequencing and IPL analyses with high-resolution accurate mass/mass spectrometry, we observed a high diversity of pelagic Archaea and of archaeal lipids in the Arabian Sea OMZ and the oxygenated Atlantic Ocean. Phylogeny based on the 16S rRNA gene showed that Archaea belonging to MGI were abundant throughout the water column and co-occurred with MGII members. MGI were also detected within the OMZs strengthening the idea that MGI Archaea are able to grow in oxygen depleted waters. The presence of MGI and MGII coincided with the detection of IPL-GDGTs including crenarchaeol, the biomarker lipid for MGI. However, the lack of cultures of MGII and the co-occurrence with MGI makes difficult to assign specific lipids to the MGII Archaea. Members of the MGB-A Archaea were detected both in the oxygenated water column of the North Atlantic Ocean as well as in the OMZ of the Arabian Sea suggesting that their distribution is not dependent of oxygen availability. High relative abundances of MGIII Archaea together with DPANN Pacearchaeota were detected within the OMZs suggesting an anaerobic lifestyle. DPANN Archaea have been suggested to be unable to synthesize their own membrane lipids as they lack the genes of the lipid biosynthetic pathways, and are not expected to contribute to the lipid pool. The presence of MGIII within the OMZ of the ETSP has previously been linked to the occurrence of archaeol-based IPLs, but these lipids were not detected in the OMZ of the Arabian Sea although they did contain MGIII Archaea, suggesting they may not be the source of these IPLs. Finally, the relative abundance of GDGTs from Thaumarchaeota seem to change in the relative number of cyclopentane moieties with oxygen concentrations. Our results shed some new light on the impact of archaeal diversity and oxygen availability on marine ether lipid compositions.

5

SUPPLEMENTARY FIGURES

Table S1. Table with physicochemical characteristics and filter types used in the Eastern Tropical South Pacific (ETSP), Arabian Sea and the North Atlantic Ocean at the time of sampling.

ETSP

Depth (mbsl)	Filter (pore size, μm)	Temp (°C)	Sal (PSU)	Oxygen (μmol kg^{-1})	Si (μmol L^{-1})	PO$_4^{3-}$ (μmol L^{-1})	NH$_4^+$ (μmol L^{-1})	NO$_2^-$ (μmol L^{-1})	NO$_3^-$ (μmol L^{-1})
22	0.7	17.0	35.1	211.8	2.5	1.1	0.35	0.16	4.3
75	0.7	13.3	34.8	2.1	23.2	2.9	0.08	2.79	13.6
83	0.7	13.0	34.8	2.2	26.7	2.9	0.02	3.51	11.8
100	0.7	12.9	34.9	2.2	28.1	2.9	0.03	5.28	10.2
135	0.7	13.5	34.7	2.2	12.5	2.5	0.05	0.08	23.3
150	0.7	12.7	34.9	2.2	29.1	2.9	0.03	7.09	10.9
175	0.7	12.4	34.9	2.2	28.8	2.9	0.03	7.66	12.5
200	0.7	12.3	34.9	2.2	28.6	2.9	0.01	7.15	14.0
235	0.7	12.1	34.8	2.2	28.6	2.8	0.03	6.71	14.3
250	0.7	11.8	34.8	2.3	28.7	2.9	0.01	6.64	17.0
300	0.7	11.5	34.8	2.4	28.8	2.9	0.06	7.35	17.5
350	0.7	10.7	34.7	2.5	29.7	2.9	0.01	4.85	24.2
500	0.7	8.0	34.6	14.6	34.6	3.0	0.06	0.04	41.2
1000	0.7	4.4	34.5	65.0	75.7	3.1	0.02	0.01	44.0

Table S1. Continued

Depth	Filter	Temp	Sal	Oxygen	Si	PO_4^{3-}	NH_4^+	NO_2^-	NO_3^-
(mbsl)	(pore size, μm)	(°C)	(PSU)	(μmol kg^{-1})	(μmol L^{-1})	(μmol L^{-1})	(μmol L^{-1})	(μmol L^{-1})	(μmol L^{-1})
100	0.7	14.4	34.8	54.4	9.4	2.1	0.01	0.06	21.9
125	0.7	12.7	34.8	2.1	20.5	2.6	0.05	0.01	21.7
155	0.7	12.5	34.8	2.1	25.9	2.8	0.01	0.34	19.8
175	0.7	12.7	34.9	2.1	27.4	2.7	0.02	4.33	14.7
200	0.7	12.2	34.9	2.2	27.9	2.7	0.03	4.91	19.0
235	0.7	11.9	34.8	2.2	28.1	2.7	0.01	5.34	20.4
300	0.7	10.7	34.8	2.4	30.2	2.8	0.01	4.01	28.0
500	0.7	7.6	34.6	15.9	39.4	3.0	0.01	0.02	42.5
1000	0.7	4.4	34.5	55.8	34.5	3.2	0.03	0.00	44.3

5

Arabian Sea

Depth (mbsl)	Filter (pore size, μm)	Temp (°C)	Sal (PSU)	Oxygen (μmol kg⁻¹)	Fluorescence (μg L⁻¹)	Si (μmol L⁻¹)	PO_4^{3-} (μmol L⁻¹)	NH_4^+ (μmol L⁻¹)	NO_2^- (μmol L⁻¹)	NO_3^- (μmol L⁻¹)	DIC (μmol L⁻¹)
20	0.3 and 0.7	24.8	36.3	188	0.16	1.6	0.6	0.09	0.04	3.3	2121
170	0.3 and 0.7	19	35.9	4.8	0.02	17	2.3	0.14	0.62	21.7	2265
300	0.3 and 0.7	15.8	36.1	5.2	0.01	25.5	2.4	0.01	0.03	24.2	2296
450	0.3 and 0.7	14.2	36	3.1	0.04	32.7	2.5	0.03	0.02	24.4	2313
600	0.3 and 0.7	12.2	35.7	2.8	0.05	43.8	2.7	0.06	0.5	26.7	2327
750	0.3 and 0.7	11	35.6	2.9	0.04	55	2.8	0.05	0.02	32	2339
900	0.3 and 0.7	9.6	35.5	3	0.03	65.1	2.9	0.1	0.02	35.4	2344
1050	0.3 and 0.7	8.6	35.4	4.7	0.04	86	3.1	0.1	0.02	37.9	2376
1200	0.3 and 0.7	7.6	35.3	9.3	0.03	96.5	3.1	0.08	0.02	39.6	2384
1350	0.3 and 0.7	6.7	35.2	16.4	0.03	109	3.1	0.07	0.02	39.3	2390
1500	0.3 and 0.7	5.6	35.1	24.5	0.04	120	3.1	BDL	0.01	39.7	2388
2000	0.3 and 0.7	3.2	34.8	55.7	0.01	142	2.8	0.04	0.03	38.1	2400

North Atlantic Ocean

Depth (mbsl)	Filter (pore size, µm)	Temp (°C)	Sal (PSU)	Oxygen (µmol kg⁻¹)	Fluorescence (µg L⁻¹)	PAR (mol m⁻² s⁻¹)	Si (µmol L⁻¹)	PO_4^{3-} (µmol L⁻¹)	NH_4^+ (µmol L⁻¹)	NO_2^- (µmol L⁻¹)	NO_3^- (µmol L⁻¹)	DIC (µmol L⁻¹)
20	0.3	28.8	34.8	165.1	0.09	8.37	1.93	0.04	0.08	BDL	0.02	2022.7
50	0.3	27.4	36.4	166.0	0.12	5.86	1	0.05	0.06	BDL	0.02	2088.7
200	0.3	11.4	35.2	92.0	0.03	0.05	10.65	1.65	0.05	BDL	25.51	2237.9
400	0.3	7.5	34.7	110.6	0.03	0.05	18.66	2.10	0.05	BDL	31.61	2237.3
600	0.3	6.2	34.6	103.0	0.03	0.05	23.82	2.33	0.05	BDL	35.12	2261.1
800	0.3	5.0	34.6	117.4	0.03	0.05	28.56	2.37	0.05	BDL	34.91	2259.0
1000	0.3	4.8	34.7	127.3	0.03	0.05	28.85	2.24	0.06	BDL	32.88	2237.7
1200	0.3	4.8	34.9	143.2	0.02	0.05	24.85	1.98	0.05	BDL	29.09	2234.9
1500	0.3	4.6	35.0	185.9	0.02	0.05	16.53	1.43	0.05	BDL	21.32	2208.4
2000	0.3	3.6	35.0	209.0	0.02	0.05	16.52	1.26	0.05	BDL	18.78	2192.4
2500	0.3	3.0	34.9	207.4	0.02	0.05	23.26	1.31	0.05	BDL	19.47	2198.2
3000	0.3	2.6	34.9	206.2	0.01	0.05	28.97	1.33	0.05	BDL	19.82	2204.6

BDL = below detection limit

PAR = photosynthetically active radiation

5

Figure S1. Map of sampling locations within the Eastern Tropical South Pacific (ETSP), Arabian Sea and the North Atlantic Ocean.
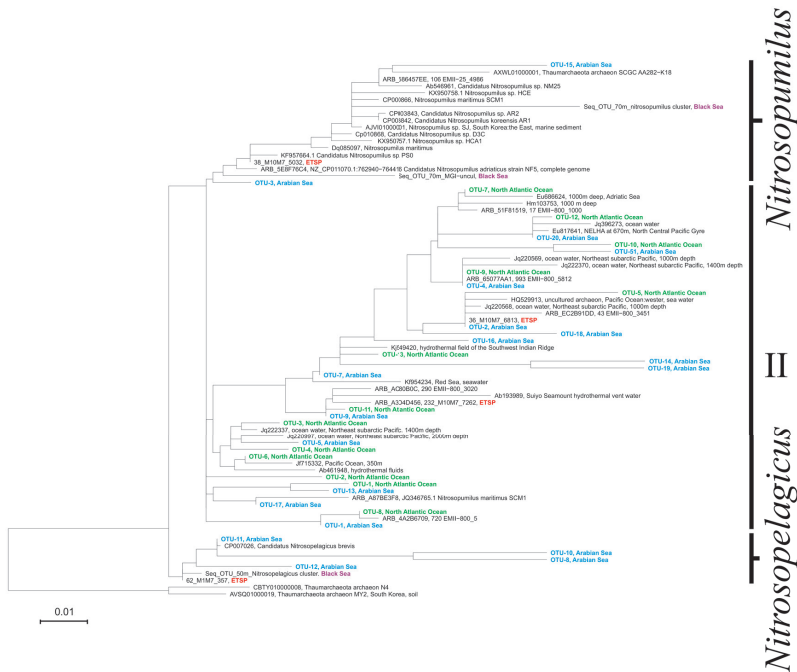
Figure S2. Phylogenetic tree of MGI showing the representative OTU 16S rRNA gene sequences present in the Eastern Tropical South Pacific (ETSP), Arabian Sea and the North Atlantic Ocean. Scale bar indicates estimated sequence divergence of 1%.
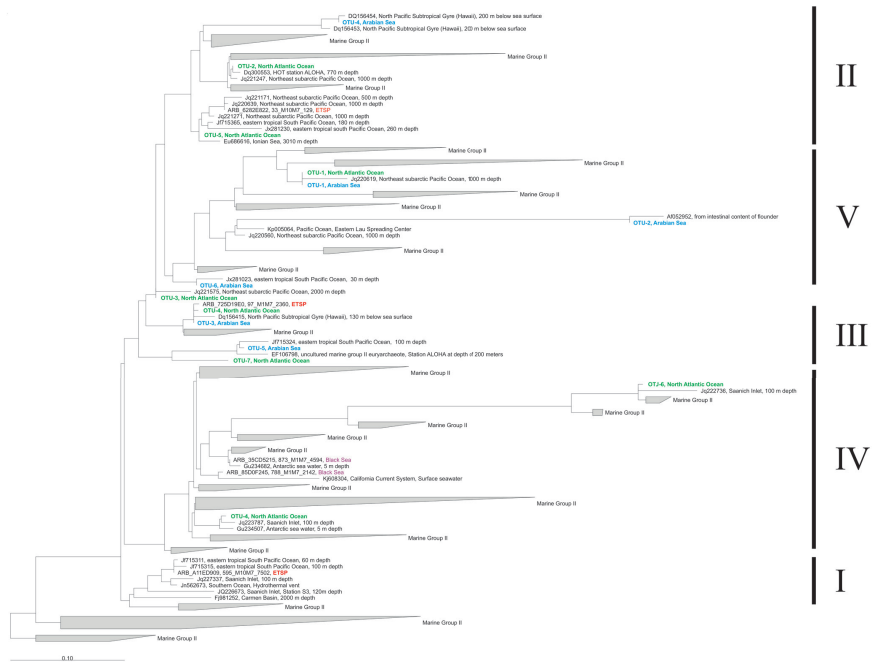
Figure S3. Phylogenetic tree of MGII showing the representative OTU 16S rRNA gene sequences present in the Eastern Tropical South Pacific (ETSP), Arabian Sea and the North Atlantic Ocean. Scale bar indicates estimated sequence divergence of 5%.
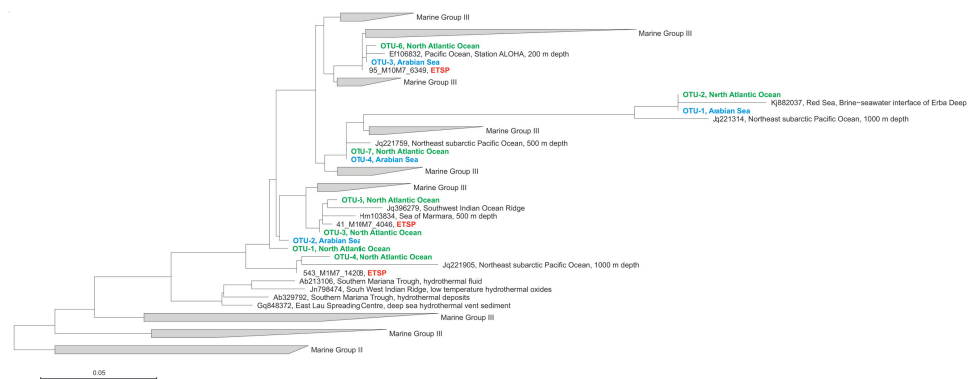
Figure S4. Phylogenetic tree of MGIII showing the representative OTU 16S rRNA gene sequences present in the Eastern Tropical South Pacific (ETSP), Arabian Sea and the North Atlantic Ocean. Scale bar indicates estimated sequence divergence of 5%.

CHAPTER 6

Potential biological sources of long chain alkyl diols in a lacustrine system

Laura Villanueva, Marc A. Besseling, Marta Rodrigo-Gámiz, Sebastiaan W. Rampen, Dirk Verschuren, Jaap S. Sinninghe Damsté

ABSTRACT

Long chain alkyl diols (LCDs) have been detected in a wide range of marine and lacustrine environments, as well as in several algal cultures. However, the identity of the producers, their preferred ecological niche and seasonality are uncertain. We applied a gene-based approach to determine the identity and abundance of Eustigmatophyceae 18S rRNA genes and compared the data with the distribution of LCDs in the water column of Lake Challa (East Africa). Gene-based analysis revealed three known and two novel Eustigmatophyceae groups. Maxima in the number of gene copies and LCD concentration coincided at 9 m water depth, signifying Eustigmatophyceae as important producers of LCDs. In addition, seasonal changes in LCD abundance in sedimenting particles revealed several blooms of LCD producers over the annual cycle.

INTRODUCTION

Long chain alkyl diols (LCDs) consist of an alkyl chain with OH groups at C-1 and at a mid-chain position. LCDs with 28–32 carbons atoms and OH groups at C-1,13 and C-1,15 have been found in Eustigmatophyceae cultures of marine (*Nannochloropsis* sp., *Eustigmatophyceae* sp.; Volkman et al., 1992) and freshwater species (*Vischeria* sp., *Eustigmatos* sp.; Volkman et al., 1999). Other sources are some members of the *Proboscia* diatom genus (Sinninghe Damsté et al., 2003) and the alga *Apedinella radians* of the Dictyochophyceae phylum, both of which produce 1,14-diols (Rampen et al., 2011). LCDs have been found in marine and lacustrine sediments (e.g. Versteegh et al., 1997). Recently, Rampen et al. (2012) proposed the long chain diol index (LDI) as a novel marine paleotemperature proxy based on the C_{30} 1,15-diol abundance relative to the C_{28} 1,13-diol, and C_{30} 1,13-, 1,15-diols. Eustigmatophyceae are generally considered to be major producers of LCDs in lakes but the identity of lacustrine LCD producers, their preferred niche in the water column, and their seasonality is uncertain. Such information could improve the predictive power of the LDI proxy.

We have developed a genetic-based approach to identify and quantify the abundance of potential LCD producers based on the 18S rRNA gene of members of the Eustigmatophyceae and its comparison with the distribution, abundance and seasonality of LCDs in a lake system.

Study site and sampling

Lake Challa is a permanently stratified crater lake on the southeastern flank of Mt. Kilimanjaro (East Africa). Suspended particulate matter (SPM) was collected at 5 and 10 m intervals throughout the water column in early February 2010 (see Buckles et al., 2013 for details and physicochemical conditions at the time of sampling); here we focus on samples comprised between 0.5 and 24 m depth, i.e. within and just below the photic zone. A mid-lake sediment trap at 35 m depth provided monthly samples of settling particles between August 2009 and August 2010.

6

MATERIAL AND METHODS

DNA methods

DNA was extracted from SPM filtered on GF/F 0.7 μm filters as described by Buckles et al. (2013). Primer pair Eust287F (5'- CGA CRA MTC ATT CAA GYT TCT GCC-3'), Eust810R (5'-CCA TGC TAR TGT ATT CAS GGC CT-3') was designed manually, and tested computationally and in PCRs. Gradient PCR was performed with melting temperature (Tm) ranging from 52–63 °C with genomic DNA extracted from different algal cultures (optimal Tm 58 °C). Quantitative PCR (qPCR) using the Eust287F/810R primer pair was performed at Tm of 61 °C and 45 cycles following the conditions described by Buckles et al. (2013). A phylogenetic tree was inferred from the Neighbour-joining method and distances computed with the Jukes-Cantor method. Sequences NCBI accession numbers are KF765160 – KF765375.

Lipid methods

Filters from the SPM and the sediment trap were base hydrolyzed (cf. de Leeuw et al. (1983) by refluxing for 1 h with 1 N KOH in MeOH (96%). After cooling, the solvent was acidified with 2 N HCl/MeOH (1:1; v/v) to pH 2 and transferred to a separatory funnel. The filters were extracted using $MeOH/H_2O$ (1:1 v/v; 1x), MeOH and dichloromethane (DCM; 3x). Solvent was collected in a separatory funnel containing ca. 25 ml bidistilled H_2O. The DCM layer was separated from the H_2O/MeOH layer and the remaining H_2O/MeOH layer was extracted (3x) with DCM. The extracts were combined and rotary evaporated to near dryness. The resulting extract and the residual filters were hydrolyzed with acid (3 h reflux, 2 N HCl/MeOH, 1:1; v/v) and neutralized with 1 N KOH in MeOH (96%). Filters were extracted as above while for the extracts, 3 ml bidistilled H_2O was added and the lipids extracted using DCM (4x). All extracts were combined, dried under N_2, eluted in DCM over a pipette column containing Na_2SO_4, dried under N_2, methylated in DCM using CH_2N_2 in Et_2O and dried under N_2. An internal standard C_{22} 7,16-diol was added to the total lipid extracts and each extracts were fractionated into apolar

and polar fractions using a glass pipette column with activated Al_2O_3 and eluted with hexane/DCM (9/1; v/v) and DCM/MeOH (1/1; v/v). Each polar fraction was silylated prior to gas chromatography-mass spectrometry (GC-MS). LCD analysis was as per Rampen et al. (2012).

RESULTS AND DISCUSSION

Eustigmatophyceae and LCD diversity and abundance

To determine eustigmatophyte diversity in Lake Challa SPM, clone libraries were generated by cloning 18S rRNA gene fragments generated by the primers Eust287F/Eust810R. Sequences from 0.5, 9, and 19 m water depth all clustered into five distinctive phylogenetic groups (Figure 1). No clustering of sequences according to depth was observed as those recovered from the three depths were distributed throughout the tree. Group 1 sequences were closely related to those sequences of the *Goniochloridaceae* family (Pribyl et al., 2012), while groups 4 and 5 sequences clustered with sequences of the *Monodopsidaceae* and *Eustigmataceae* families. Sequences falling in groups 2 and 3 diverged from sequences of cultured representatives, supporting their assignment to one or more unknown Eustigmatophyceae families. Quantification of Eustigmatophyceae gene copies showed a distinctive peak at 9 m depth (Figure 2A). The most abundant LCDs in the February SPM samples were C_{32} 1,15 (138 ng l^{-1}), C_{30} 1,15 (54 ng l^{-1}), and C_{34} 1,17-diols (23 ng l^{-1}). Of these, the C_{34} 1,17-diols may be produced by the novel Eustigmatophyceae with group 2 and 3 sequences, since these diols have previously been found in lake samples (Versteegh et al., 1997; Zhang et al., 2011), but have not been detected in freshwater eustigmatophyte cultures (Volkman et al., 1999). Maximum LCD abundance was at 9 m (62 ng l^{-1}; Table 1, Figure 2B), coinciding with the maximum abundance of Eustigmatophyceae 18S rRNA gene copies (Figure 2). This correlation supports the Eustigmatophyceae as important LCD producers in this lake system. High LCD abundance (38–46 ng l^{-1}) coincides with little or no Eustigmatophyceae 18S rRNA gene copies in the uppermost part of the water column (0–5 m). This pattern may be explained by wind-driven and convective

mixing of preserved LCDs throughout the epilimnion, whereas living algal cells adjust their buoyancy to their preferred habitat at slightly greater depth.
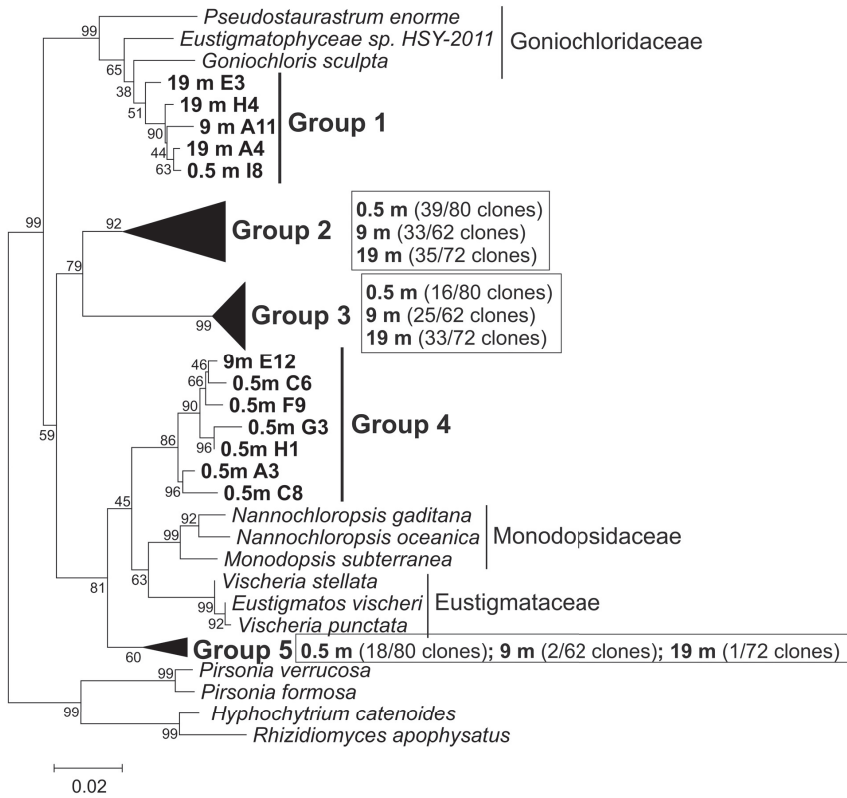
Figure 1. Phylogenetic tree for 18S rRNA gene sequences recovered and closest relatives in the Eustigmatophyceae phylum. Branch support (%) is indicated on the branches. Scale bar indicates 0.02 substitutions per site. Letter and number code, e.g. 19 m E3 is an arbitrary code assignation to the sequences recovered after cloning.
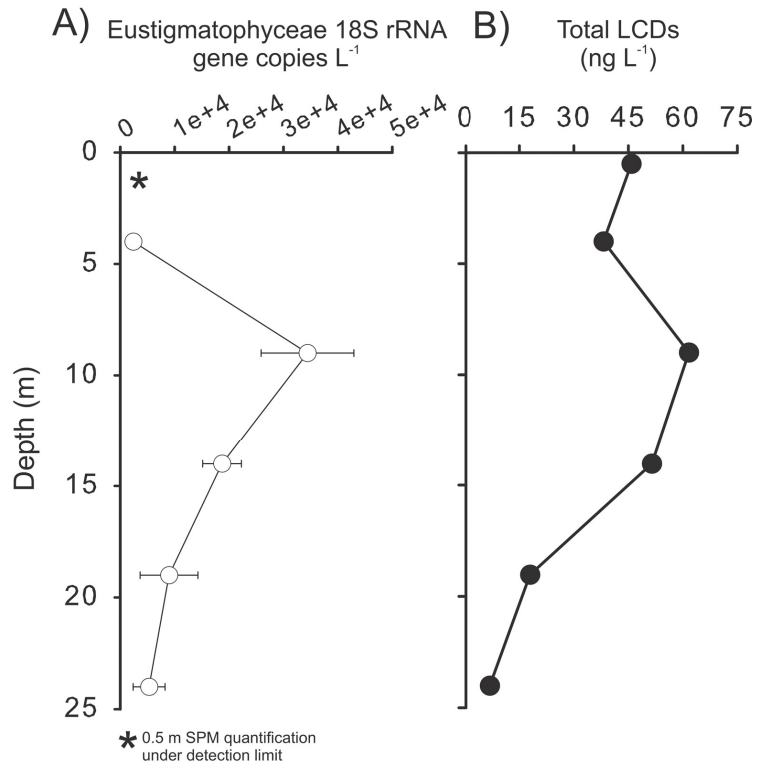
Figure 2. Quantification of Eustigmatophyceae 18S rRNA gene copies and total LCDs in SPM from the upper water column of Lake Challa collected in early February 2010.

Table 1. Quantification of LCDs (ng/l filtered) in Lake Challa SPM samples collected in February 2010.

Depth (m)	C_{30} 1,14	C_{30} 1,15	C_{30} 1,16	C_{32} 1,15	C_{32} 1,16	C_{34} 1,15	C_{34} 1,17	Total
0.5	0.5	9.8	0.9	29.6	0.5	0.0	4.4	45.8
4	0.6	9.5	1.1	22.9	0.4	0.0	3.6	38.1
9	1.0	14.9	0.0	38.1	0.0	0.4	7.3	61.6
14	0.8	12.9	0.0	32.2	0.0	0.4	5.1	51.5
19	0.0	5.2	0.0	10.9	0.0	0.0	1.9	17.9
24	0.0	2.1	0.0	4.1	0.0	0.0	0.6	6.9

Seasonality of LCDs

Peak LCD fluxes in descending particles were detected in February, April and June 2010 (Table 2), with C_{32} 1,15, C_{30} 1,15, and C_{34} 1,17-diols accounting for >85% of total LCD abundance. LCD in settling particles during February (Table 2) were similar to those in the SPM in early February (Table 1). In April, the most abundant LCD was the C_{30} 1,15-diol, while in February and June it was the C_{32} 1,15-diol. These differences may reflect temporary blooms of different LCD producers or a change in the distribution of LCDs within the same producer. Successive seasonal blooming of different Eustigmatophyceae could indicate niche separation controlled by temperature variation in the upper water column (peaking at ca. 27 °C in February), or seasonal nutrient dynamics influenced by the timing of rainfall and water column stratification.

Table 2. Long chain diols (LCDs) fluxes (mg m^{-2} day^{-1}) determined from particles settling in a sediment trap in the center of the Lake Challa.

Date	C_{30} 1,13	C_{30} 1,14	C_{30} 1,15	C_{31} 1,15	$C_{32:1}$ 1,15	C_{32} 1,13	C_{32} 1,15	C_{32} 1,16	C_{34} 1,15	C_{34} 1,17	Total
Aug'09	0.2	1.1	3.8	0.3	0.0	0.0	3.8	0.1	0.0	0.6	9.9
Sep'09	1.8	2.7	5.3	1.4	0.6	0.0	9.5	0.2	0.0	0.8	22.2
Oct'09	0.3	0.5	2.9	0.1	0.0	0.0	1.0	0.0	0.0	0.2	5.0
Nov'09	1.9	1.8	23.0	1.5	0.2	0.0	9.4	0.3	0.0	0.6	38.6
Dec'09	0.6	1.5	18.4	0.5	0.3	0.0	7.6	0.3	0.0	1.8	30.9
Jan'10	0.1	0.3	3.5	0.2	0.0	0.2	10.0	0.3	0.0	3.3	18.1
Feb'10	0.3	2.5	35.0	2.2	0.6	1.6	83.7	2.1	0.5	35.3	164.7
Mar'10	2.4	3.5	38.1	1.6	0.2	0.2	26.1	0.9	0.1	6.1	79.3
Apr'10	10.1	11.2	111.8	5.1	0.7	0.4	48.2	2.3	0.1	16.1	206.2
May'10	1.5	2.3	25.0	0.9	0.2	0.2	23.3	0.6	0.1	7.4	61.6
Jun'10	1.3	2.3	20.0	1.8	0.4	0.4	57.4	0.8	0.2	22.5	107.4
July'10	0.6	2.0	10.9	0.6	0.2	0.0	15.2	0.3	0.0	5.7	35.5
Aug'10	0.0	1.5	5.8	0.4	0.0	0.0	8.7	0.0	0.0	3.2	19.6

CONCLUSIONS

Application of a 18S rRNA gene-based method has revealed the presence of both known and novel groups of Eustigmatophyceae in Lake Challa. Maximum abundance of Eustigmatophyceae gene sequences coincided with maximum LCD abundance at 9 m water depth, suggesting an important role of eustigmatophytes as LCD producers. Seasonal variation in LCD distributions suggests that successive LCD-producing blooms were due to different eustigmatophyte algae or changes in the LCDs produced by a unique algal population in evolving abiotic conditions.

ACKNOWLEDGMENTS

6

CHAPTER 7

Genetic response of the long chain alkenone biosynthetic pathway in the haptophyte *Emiliania huxleyi* upon temperature shock

Marc A. Besseling, Stefan Schouten, Jaap S. Sinninghe Damsté and Laura Villanueva

ABSTRACT

Long chain alkenones (LCAs) are lipids that are synthesized by specific haptophyte algae and contain multiple double bonds with an unusual *trans*-configuration. LCAs were first detected in marine sediments and are widely used in paleoceanographic studies because of their role in the sea surface temperature proxy, the U^K_{37} index, based on the ratio between the di- and tri-unsaturated LCA. Despite their importance in paleoclimatology, the biosynthetic pathway of LCAs remains largely unknown, this includes the enzymes responsible for the possible chain length elongation(s) and for the formation of di- and tri-unsaturated LCAs. To constrain potential candidate genes encoding these enzymes, we conducted two cold shock experiments with cultures of the haptophyte *Emiliania huxleyi*. The first experiment, with sampling during 24 h, revealed a rapid increase in the $C_{37:3}/C_{37:2}$ LCA ratio within 5 h when the haptophyte was transferred from 20 to 15 °C. To increase the LCA alteration effect, we performed a second shorter (3 h) cold shock experiment with a steeper temperature gradient from 20 to 10 °C. A rapid increase in the $C_{37:3}/C_{37:2}$ ratio was again observed and transcriptomic data from this experiment was obtained to shed light on potential genes involved in the LCA biosynthetic pathway. A clear differentiation of gene expression was observed after the cold-shock. Most striking was the swift up-regulation of EMIHUDRAFT_454486, a gene coding for the likely homologue of the putative LCA desaturase Akd1, suggesting that it may be the desaturase responsible for transforming $C_{37:2}$ into $C_{37:3}$ LCAs. Our results can be used to target this gene in gene manipulation experiment as well as environmental studies to constrain potential producers of alkenones.

INTRODUCTION

Long chain alkenones (LCAs) are C_{35}–C_{40} unsaturated methyl- and ethyl- ketones, synthesized by a number of haptophyte algae (Marlowe et al., 1984; Volkman et al., 1980) within the Isochrysidales order (Bendif et al., 2013). Among the Isochrysidales is the alga *Emiliania huxleyi*, the most common coccolithophore in the open oceans, and together with *Gephyrocapsa oceanica* they are believed to be the most dominant sources of marine alkenones (Conte et al., 1995; 2006; Volkman et al., 1995). However, other marine haptophytes such as *Isochrysis galbana* (Sukenik and Wahnon, 1990) and *Chrysotila lamellosa* (Marlowe et al., 1990) are also capable of synthesizing polyunsaturated LCAs. These lipids are widely used in paleoceanographic studies because of their role in the sea surface temperature (SST) proxy U^K_{37} and its derivative the U^K_{37} index ($U^K_{37} = [C_{37:2}]/([C_{37:2}] + [C_{37:3}])$; Brassell et al., 1986; Müller et al., 1998; Prahl & Wakeham, 1987). The principle of the U^K_{37} index is based on the observation that alkenones vary in level of unsaturation depending on culture or seawater temperature in which the organisms reside, i.e. increasing temperatures result in increasing levels of the $C_{37:2}$ and decreasing levels of the $C_{37:3}$ (Prahl & Wakeham, 1987). However, physiological factors other than temperature seem to affect the unsaturation ratios to some degree, such as growth phase (Yamamoto et al., 2000), salinity (Ono et al., 2012), and genetic factors such as species composition and their genotypes (Conte et al., 1998; Zheng et al., 2016). Despite these uncertainties, the strong correlation of U^K_{37} with annual mean SST in different oceanic settings is robust (Müller et al., 1998; Tierney and Tingley, 2018) and the estimated SSTs from paleoclimate studies match those from other paleotemperature proxies (e.g. Bard, 2001; Herbert, 2003 and references cited therein).

Despite their importance in paleoclimatology, it remains speculative what the function of LCAs is within the haptophyte cell. Alkenones were first suspected to be membrane lipids; the degree of unsaturation was thought to influence the membrane fluidity (e.g. Prahl et al., 1988). However, LCAs do not occur bound to membranes within the cell (Sawada and Shiraiwa, 2004), and microscopic analyses and isolation of cellular compartments showed that LCAs accumulate in so-called alkenone bodies

(Eltgroth et al., 2005; Shi et al., 2015). The fact that the LCA content decreases when the cells are grown in the dark, strongly suggests that LCAs function as a carbon/energy storage to maintain metabolism in the dark (Epstein et al., 2001; Pan & Sun, 2011). Furthermore, LCA occur with unusual *trans*-double bonds (Rechka and Maxwell, 1988), something that does not alter their spatial dimensions in contrary to typical unsaturated membrane lipids with *cis* stereochemistry which fulfill the requirements for adaptation of the membrane structure. Studies have shown that these double bonds are spaced with a regular pattern across the LCA linear carbon chain, i.e. typically separated by 7 methylene units (e.g. $\Delta^{14,21}$, $\Delta^{7,14,21}$ and $\Delta^{7,14,21,28}$, numbered from the carbonyl position, for LCAs with 2-4 double bonds,; López & Grimalt, 2006). Nevertheless, there are exceptions to this pattern, for example LCAs in *E. huxleyi* CCMP2758 (e.g. $\Delta^{12,19}$; and $\Delta^{7,12,19}$; Prahl et al., 2006; Zheng et al., 2016) and those in Black Sea and estuarine sediments (e.g. $\Delta^{14,19}$; Xu et al., 2001; van Soelen et al., 2014) possess a different double bond position and spacing.

The biosynthetic pathway leading to LCAs is unknown although several hypothetical pathways for LCAs have been proposed (Rontani et al., 2006; Zheng et al., 2016). These pathways start from a primer of acetyl-CoA or propionyl-CoA involving successive chain elongations with malonyl-CoA, followed by an unknown thioesterase and decarboxylase. Rontani et al. (2006) also proposed a mechanism for the formation of the atypical *trans* double bonds in the LCAs by evaluating several types of alkenones, suggesting that the formation of double bonds in LCAs implicates a series of Δ-desaturases that have active sites at a fixed distance from a carbonyl binding site (i.e. Δ^7; and Δ^{14}). Recently, Kitamura and colleagues (2018) observed, by applying a cold-shock to *E. huxleyi* and labeling with ^{13}C-bicarbonate, that a significant portion of the $C_{37:2}$ ($\Delta^{14,21}$) alkenone pool was converted into $C_{37:3}$ ($\Delta^{7,14,21}$). They concluded that desaturation at the Δ^7 position occurs after the final chain elongation. Building up on that study, a novel alkenone desaturase (Akd1) has been identified in the alkenone-producing haptophyte *Tisochrysis lutea* by showing that overexpression of Akd1 resulted in the conversion of $C_{37:2}$ into $C_{37:3}$ (Endo et al., 2018). This supported the existence of a Δ^7 desaturase reaction transforming the di-

unsaturated into tri-unsaturated LCAs, suggesting that the formation of LCA $C_{37:3}$ is the result of a desaturation of a previously synthesized LCA $C_{37:2}$. However, it is not yet known if this mechanism is common to the rest of LCA-producing haptophytes, if several desaturases could be involved in the same desaturation step, and how the other double bonds of the carbon-chain are formed.

Here we perform cold shock experiments with the haptophyte *E. huxleyi* ubiquitous in the ocean and analyze the impact on the abundance and distribution of $C_{37:2}$ and $C_{37:3}$ alkenones. Transcriptomic data of the second cold shock experiment was used to investigate potential genes involved in the LCA biosynthetic pathway.

MATERIAL AND METHOD

Organism and Culture Conditions

The organism used in this study was the haptophyte *Emiliania huxleyi* (CCMP1516), obtained from the National Center for Marine Algae and Microbiota (NCMA; http://ncma.bigelow.org; formerly CCMP, the Provasoli-Guillard National Center for Culture of Marine Phytoplankton). Cells were grown in 250 ml glass Erlenmeyer flasks in L1-Si medium (NCMA, East Boothbay, USA; Guillard & Hargraves, 1993) based on artificial seawater, at 20°C, under a light:dark cycle of 16:8 h with a light regime of 100 μmol photons m^{-2} s^{-1}.

Cold shock experiments

In a first cold shock experiment, algal cells were initially grown in batch culture in 5 L glass Erlenmeyer flasks at 20°C, under a light:dark cycle of 16:8 h with irradiance. Cells were grown up to a cell density of 8.2 x 10^5 cells ml^{-1}, thoroughly mixed, and 260 ml of culture was transferred to fifty-one 250 ml glass Erlenmeyer flasks with a wide opening. All 250 ml Erlenmeyer flasks were kept at 20°C for 1 day, after which 24 Erlenmeyer flasks were randomly selected and immediately transferred to a room at 15°C with the same light regime and light-dark cycle (T-0, 4.5 h after the dark cycle finished). The change of in-situ growth temperatures of the cultures were monitored using a thermometer inside an additional Erlenmeyer flask containing

E. huxleyi At T-0 (i.e. directly after transfer to 15°C) the cells of three Erlenmeyer flasks, maintained at 20°C, were filtered over different pre-combusted glass fiber (GF/F) filters (pore-size of 0.7 µm, 47 mm in diameter, Whatman) using a vacuum pump system. At the later sampling points (T-1, -2, -3, -4, -5, -6, -12 and -24, where the number indicates the hour(s) after T-0) the cells of three Erlenmeyer flasks maintained at 20°C and those of three Erlenmeyer flasks, transferred to 15°C were filtered using the procedure described for T-0. Filters where flash-frozen in liquid nitrogen immediately after filtration and stored at -80°C until further processing.

In a second cold shock experiment, *E. huxleyi* cells were also initially grown in batch culture within 5 L glass Erlenmeyer flasks at 20°C. Cells were grown up to cell densities of ca. 6.5 x 10^5 cells ml^{-1} (Figure S1), thoroughly mixed, and subsequently transferred into twenty-one 250 ml Erlenmeyer flasks with a wide opening. The Erlenmeyer flasks were kept at 20°C for 1 day, then nine out of twenty-one Erlenmeyer flasks were randomly selected and transferred to 10°C with the same light-dark cycle and level of irradiance (T-0; 2.5 h after the dark cycle finished). The change in *in-situ* growth temperatures of the cultures were monitored using a thermometer within an additional Erlenmeyer flask with *E. huxleyi*. At T-0, cells of three Erlenmeyer flasks maintained at 20°C, were filtered over separate pre-combusted glass fiber (GF/F) filters (pore-size of 0.7 µm, 47 mm in diameter) using a vacuum pumping system. At later sampling points (T-1, -2, and -3, where the number indicates the hour(s) after T-0) the cells of three Erlenmeyer flasks, maintained at 20°C and the cells of three Erlenmeyer flasks transferred to 10°C, were filtered. Filters where flash-frozen in liquid nitrogen immediately after filtration and stored at -80°C until further processing.

Cell abundance and culture conditions

Cell abundance was monitored using flow cytometric determination using a BD Accuri™ C6 cytometer (BD Biosciences, New Jersey, US), with the trigger on chlorophyll red autofluorescence (FL4 filter, 675±25 nm). The phytoplankton cells were distinguished in a scatter plot of chlorophyll versus forward side scatter. All samples were analyzed in triplicate.

Alkenone extractions and analyses

Half of the glass fiber filter collected from each Erlenmeyer flask was used for alkenone analysis. Extractions were done according to Chivall et al. (2014). Briefly, after addition of nonadecan-10-one as internal standard added directly to cut-up freeze-dried filters, the filters were extracted with 2:1 v/v dichloromethane (DCM)/ methanol (MeOH) (4 x 10 ml; 4 x 5 min sonication). The extract was separated using an Al_2O_3 column into three fractions: an apolar (eluted with 9:1 v/v hexane/DCM), an alkenone (1:1 v/v hexane/ DCM) and a polar (1:1 v/v DCM/MeOH) fraction, respectively. The alkenone fraction was analyzed using gas chromatography (GC) with an Agilent 6890 instrument equipped with an Agilent CP-Sil 5 CB column (50 m x 0.32 mm i.d.; 0.12 μm film thickness). The temperature program was: 70-200 °C at 20 °C min^{-1}, then at 3 °C min^{-1} to 320 °C (held for 25 min).

RNA isolation and sequencing

RNA was isolated from the second 3 h cold shock experiment using two quarters of each glass fiber filter obtained at T-0 until T-3 of the experiment (with the number indicating the hour(s) after the start of the experiment, containing triplicates per time point per condition). RNA was isolated for each quarter of the filter and the end solutions were pooled. The Qiagen RNeasy mini kit (Qiagen, Valencia, CA, USA) was used for the RNA isolation, according to the manufacturer's instructions, 10 μL of β-mercaptoethanol and 50 μL of Plant RNA isolation aid (Thermo Scientific, Waltham, MA) were added to optimize the extraction.

RNA concentrations were measured by NanoDrop (Thermo Scientific). The quality of the RNA was determined with a Bioanalyzer (Agilent, Santa Clara CA). Library preparations were done by Utrecht sequencing facility (USEQ, Utrecht, the Netherlands) and samples were sequenced by USEQ using two runs of 1x 75bp on an Illumina Nextseq 500 platform.

7

Reference genome, differential gene expression and visualization

The obtained Illumina single-end data was trimmed (Phred score ≥ 33, minimal length = 36 bp) using Trimmomatic (version 0.36; Bolger et al., 2014) and quality checked with FastQC (Andrews, 2010). RNA-seq data was aligned using the Spliced Transcripts Alignment to a Reference (STAR) tool (Dobin et al., 2013) using the *E. huxleyi* 1516 reference genome (Read et al., 2013). One BAM file, containing the read alignments, was obtained for each sample, the function MergeSamFiles from the Picard tool (http://broadinstitute.github.io/picard) was used to merge the BAM files. From the gene models, provided by ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/372/725/GCA_000372725.1_Emiliana_huxleyi_CCMP1516_main_genome_assembly_v1.0/, we generated a transcript database in R (R Core Team, 2018) using the function makeTxDbFromGFF (GenomicFeatures package, version 1.28.6). Then, reads were assigned to *E. huxleyi* transcripts using the summarizeOverlaps function from the R package GenomicAlignments (version 1.12.2; Lawrence et al., 2013). Normalized read counts were calculated using the DESeq2 package (version 1.16.1, Love et al., 2014) and DESeq2 was used to perform a differential expression analysis between the control (20°C) and the cold shocked (10°C) samples. The list of differentially expressed transcripts with a false discovery rate-adjusted ρ-value <0.01, a shrunken log fold change estimate (LFC) of ≥1 or ≤ -1 (generated with apeglm; Zhu et al., 2018) is shown in supplementary file 1.

The R package ggplot2 (version 3.1.0; Wickham, 2016) was used for the principal component analyses (PCA) and the pheatmap package (version 1.0.10; Kolde, 2018) was used for generating heatmap figures.

Protein identification and subcelullar localizations

The top 100 genes (ranked from lowest to highest adjusted ρ-value) that were up and down regulated (ρ-value <0.01 and a LFC of ≥1 and ≤ -1) in the 3 h cold shock experiments were annotated using nucleotide blast and blastx against the NCBI nucleotide collection and non-redundant protein sequence databases (Altschul et al., 1990; 1997; Johnson et al., 2008(Johnson et al., 2008)(Johnson et al., 2008)).

SignalP 5.0 was used to search for signal peptides (Almagro Armenteros et al., 2019). Subcellular localization of proteins was done with WoLFPSORT (Horton et al., 2007), Loctree3 (Goldberg et al., 2014) and DeepLoc-1.0 (Almagro Armenteros et al., 2017).

RESULTS AND DISCUSSION

Impact of cold shocks on alkenone abundance and distribution

In a first cold shock experiment, *E. huxleyi* cultures were transferred from 20 to 15 °C (Figure 1). The *in-situ* temperature of the transferred cultures decreased exponentially from 19.4°C at T-0 to 14.4°C at T-5 (5 h after T-0). From this point the temperature remained relative stable at ca. 14°C until T-24 (Figure S1). The cell density remained stable at 1.1×10^6 cells ml^{-1}, except after 24 h when cell concentrations had increased to 1.8×10^6 cells ml^{-1} (Figure S2). During this experiment, a change in the $C_{37:3}$ / $C_{37:2}$ LCA ratio was observed (Figure 1). The $C_{37:3}$ / $C_{37:2}$ LCA ratio of the cold-shocked cultures increased almost linearly to 1.40±0.03 after 6 h, increased further to 1.64±0.05 after 12 h, and to 1.91±0.03 after 24 h (Figure 1). The $C_{37:3}$ / $C_{37:2}$ ratio of the control cultures that remained at 20°C increased from 1.08±0.02 at T-0 to 1.25±0.02 after 6 h, but then declined to 1.07±0.06 after 12 h and 0.94±0.02 after 24 h (Figure 1). The fluctuations of the $C_{37:3}$ / $C_{37:2}$ LCA ratio in the control were unexpected as temperature remained constant but are most likely caused by the dark/light rhythm of *E. huxleyi*, as previous studies have shown that LCA concentrations of *E. huxleyi* cells increase during the light phase and decrease when the cells were kept in the dark (Eltgroth et al., 2005; Epstein et al., 2001; Pan & Sun, 2011; Prahl et al., 2003). This is due to the fact that LCAs function as carbon/energy storage and are used during the dark phase to sustain metabolism (Tsuji et al., 2015). Substantial changes in the $C_{37:3}$ / $C_{37:2}$ LCA ratio were also observed during prolonged darkness (5 days or longer) in earlier studies (e.g. Prahl et al., 2003). However, to the best of our knowledge, nothing has been reported on an hourly resolution.

7
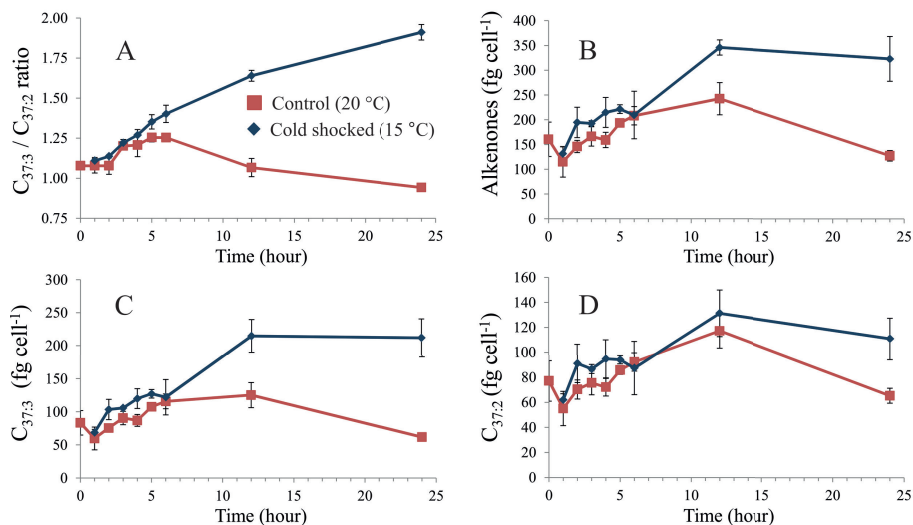
Figure 1. A) $C_{37:3}/C_{37:2}$ ratios of the 24 h cold shock experiment (24 h, at 20°C; colored in red and from 20°C to 15°C; colored in blue). B) alkenone ($C_{37:2} + C_{37:3}$) concentration (fg cell⁻¹) of the 20°C grown cultures (red) and the transferred to 15°C cultures (in blue). C) $C_{37:3}$ and D) $C_{37:2}$ alkenone concentrations (fg cell⁻¹) of the 20°C grown cultures (red) and the transferred to 15°C cultures (in blue). Bars indicate standard deviations of measurements of triplicate cultures.

The consistent increase in the $C_{37:3} / C_{37:2}$ LCA ratio of transferred cultures before 24 h (Figure 1A) indicated a cellular response to the cold shock. This observation is similar to that of Kitamura et al. (2018), who performed a similar experiment with the haptophyte *E. huxleyi*. The total LCA concentration per cell ($C_{37:2}$ plus $C_{37:3}$) increased slightly over time both within the control and cold-shocked cultures. However, after 6 h the concentration of LCAs in the cold-shocked cultures was higher than those remaining at 20°C (Figure 1B). This difference was primarily due to the increase of the concentration of the $C_{37:3}$ LCA in the cold-shocked cultures that started after 6 h (Figure 1C), whereas the $C_{37:2}$ LCA abundance remained comparable between the control and cold-shocked cultures (Figure 1D). Although the concentration of $C_{37:3}$ LCA increased in the transferred cultures, there was no apparent decrease in the concentration of $C_{37:2}$ LCA compared to the control cultures (Figure 1C-D), making it unclear if the $C_{37:2}$ LCA was converted into $C_{37:3}$ LCA, as

previously shown by Kitamura et al. (2018), or that the cold-shocked *E. huxleyi* cells synthesized de novo $C_{37:3}$ LCA (i.e. without the $C_{37:2}$ LCA as intermediate).

Because of the successful rapid induced change in the $C_{37:3}$ / $C_{37:2}$ LCA ratio, we performed a shorter second cold shock experiment (3 h) with a steeper temperature gradient (20 to 10 °C) to capture changes in gene expression. The cell concentrations ranged again between 1.2-1.3 x 10^6 cells ml^{-1} for both the control and transferred cultures and did not change over time (Figure S3). Compared to the 24 h cold shock temperature experiment, the in-situ growth temperature decreased much sharper, from 20.8°C at T-0 to 13.2°C at T-3 (Figure S1). The $C_{37:3}$ / $C_{37:2}$ LCA ratio increased in the control cultures at 20°C from 0.68±0.03 at the start to 0.84±0.02 after 3 h (Figure 2). The cold-shocked cultures showed a strong increase in the $C_{37:3}$ / $C_{37:2}$ LCA ratio from 0.68±0.03 at the start to 0.91±0.05 after 3 h (Figure 2), with the $C_{37:3}$ / $C_{37:2}$ LCA ratio of the cold-shocked cultures at 3 h (0.91±0.05) being significantly higher (one-tailed t-test, ρ-value<0.05) compared to the ratio of the control cultures (0.84±0.02). The $C_{37:3}$ / $C_{37:2}$ LCA ratio increase was less apparent in the concentration of $C_{37:3}$ and $C_{37:2}$ $cell^{-1}$, which did not differ significantly when comparing the control with the transferred cultures (Figure 2C-D). Similar to the LCA concentrations $cell^{-1}$ during the onset of the 24 h cold-shock experiment (Figure 1C-D). In this case, the observed change in the $C_{37:3}$ / $C_{37:2}$ LCA ratio has most likely been caused by an enzymatic conversion of the $C_{37:2}$ into the $C_{37:3}$ LCA (cf. Kitamura et al., 2018). This is expected to be revealed in an alteration of RNA expression level of the respective enzyme(s), which we investigated below.

7
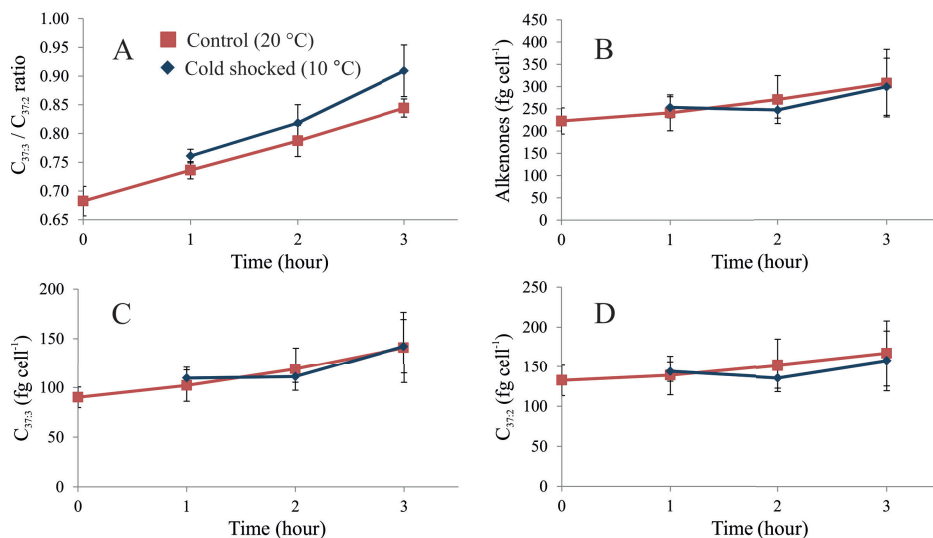
Figure 2. A) $C_{37:3}/C_{37:2}$ ratios of the 3 hour cold shock experiment (3 hours, at 20°C; colored in red and from 20°C to 10°C; colored in blue). B) alkenone ($C_{37:2} + C_{37:3}$) concentration (fg cell^{-1}) of the 20°C grown cultures (red) and the transferred to 10°C cultures (in blue). C) $C_{37:3}$ and D) $C_{37:2}$ alkenone concentrations (fg cell^{-1}) of the 20°C grown cultures (red) and the transferred to 10°C cultures (in blue). Bars indicate standard deviations of measurements of triplicate cultures.

Differential gene expression induced by the cold shock

Because of the anticipated change in RNA expression, RNA from cultures used in the 3 h cold shock experiment were analyzed and the control (20°C) and the cold shocked (from 20°C to 10°C) cultures were compared. To identify differences in gene expression between the control and cold-shocked cultures, we first performed principal component analysis (PCA) on variance stabilizing transformated (vst) normalized gene counts. This showed grouping of the control versus the transferred cultures but also one outlier (one of the transferred cultures obtained after 3 h; Figure S4). This outlier was further confirmed with Cook's distance (Cook, 1977; Figure S5). Therefore, we excluded this outlier and repeated the PCA on the normalized gene counts (Figure S6). This again showed a clear separation on the first principal component (PC1, describing 71% of the variance) between the control and the cold-shocked cultures (Figure S6). The time of sampling had a relatively small effect (PC2, 8% variance) on gene expression. (Figure S6).

Differential gene expression analysis revealed that the cold shock resulted in a total of 132 up-regulated genes and 331 down-regulated genes (≥1 and ≤-1 LFC, ρ-value <0.01; Figure S7 and listed in Table S1). The top 100 (ranked on the adjusted ρ-value from low (most significant) to high (least significant) genes that were affected by the cold shock, either down (n=86) or upregulated (n=14), are plotted in Figure 3. Out of these 100 genes there were 61 that did not match a known protein and 39 with a match (Table S2). The majority of affected genes with a positive match with (a) known protein(s), were associated with the stress response induced by the temperature shock, e.g. enzymes such as 3-dehydroquinate synthase (López-Cristoffanini et al., 2015) and HNH endonuclease (Liu et al., 2014) (Table S2). Interesting is the fact that multiple down-regulated genes matched with aldo-keto reductases (AKRs), a family of enzymes that mostly reduce aldehydes and ketones to alcohols, and are known to be involved in stress reactions in plants (Sengupta et al., 2015; and references therein). However, they are most likely not (directly) involved in the alkenone biosynthetic pathway.

7
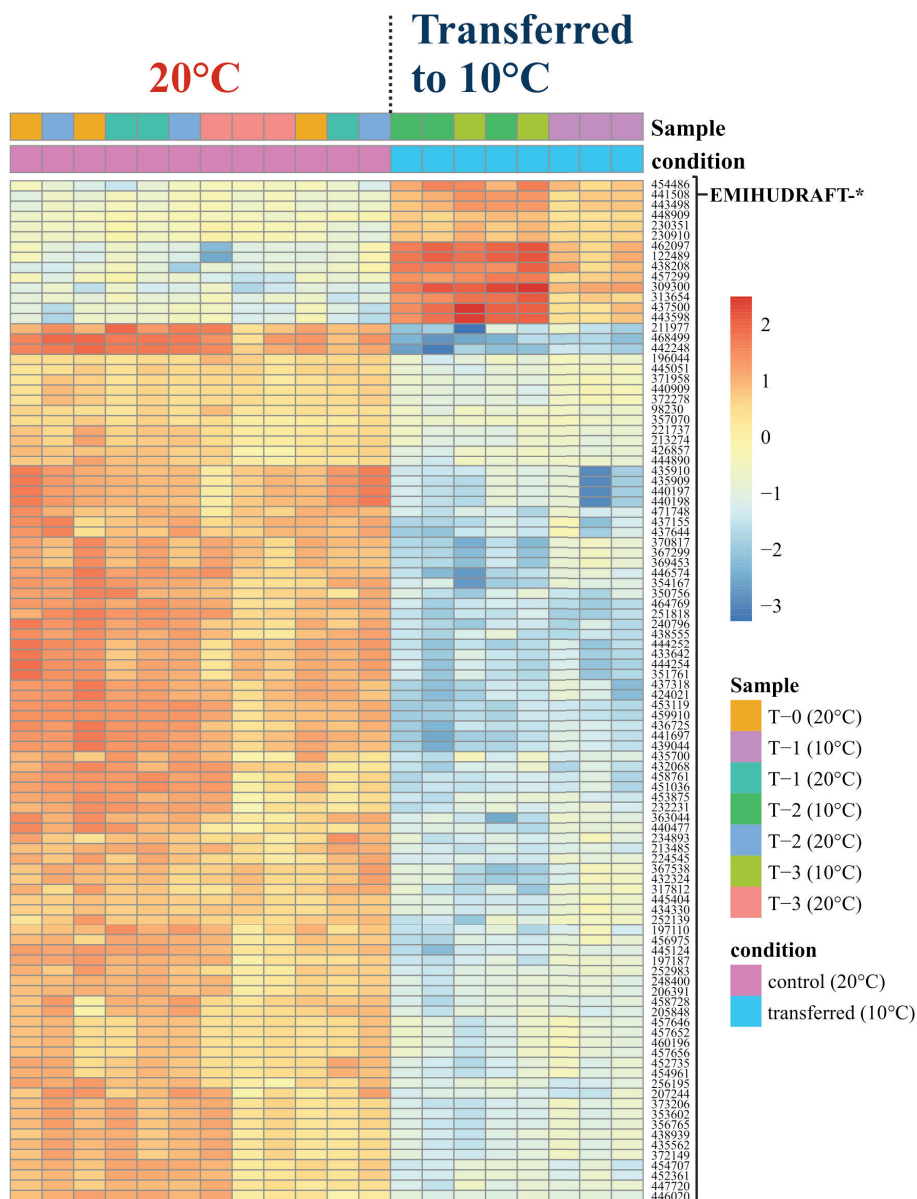
Figure 3. Heatmap of the top 100 differentially expressed genes (based on the adjusted ρ-values) between the control (grown at 20°C) and the transferred cultures (transferred to 10°C). Shown are DESeq2 variance stabilizing transformation (vst) normalized counts for each sample and gene. Genes were scaled to mean zero and standard deviation 1. Genes that significantly deviated in a specific sample from the gene's average across all samples are shown, up-regulated genes are shown in red and down-regulated genes in blue. Samples were hierarchically clustered according to differential gene expression between the control and transferred cultures.

Recently, Endo et al. (2018) identified an alkenone desaturase (Akd1) within *T. lutea* likely responsible of converting $C_{37:2}$ LCA into $C_{37:3}$ LCA by creating the *trans* double bond at the Δ^7 position. The closest gene homologues of this desaturase in *E. huxleyi* are EMIHUDRAFT-196284 (54% protein identity; as also shown by Endo et al. (2018)) and EMIHUDRAFT-454486 (43%). EMIHUDRAFT-196284 was not significantly up or down-regulated (Figure S8). However, EMIHUDRAFT-454486 was indeed one of the major up-regulated genes after the cold shock (Figure 4; Table S2). Its expression increased with time, doubling normalized counts from 1 to 3 h after transferring, while the gene counts within the control cultures remained significantly lower and stable over time (Figure 4). Our results thus support the idea that the EMIHUDRAFT-454486 gene encodes for a desaturase involved in the conversion of $C_{37:2}$ to $C_{37:3}$ alkenones. However, more studies with *E. huxleyi* with regards to this gene should be conducted to confirm this idea, perhaps by overexpression of EMIHUDRAFT-454486 in mutant strains, similar to the experiments performed by Endo et al. (2018) with *T. lutea*. EMIHUDRAFT-454486 may also be a good target in marine metagenomics and transcriptomics in combination with alkenone analysis in order to constrain the sources and functioning of alkenones, which would further improve the use of the LCA-based paleothermometer U_{37}^{K}. Also targeting homologues of EMIHUDRAFT-454486, and other potential parts of the LCA biosynthetic pathway, in other haptophytes could perhaps explain the differences in LCA composition between the different alkenone producing haptophytes (e.g. Conte et al., 1998).

7
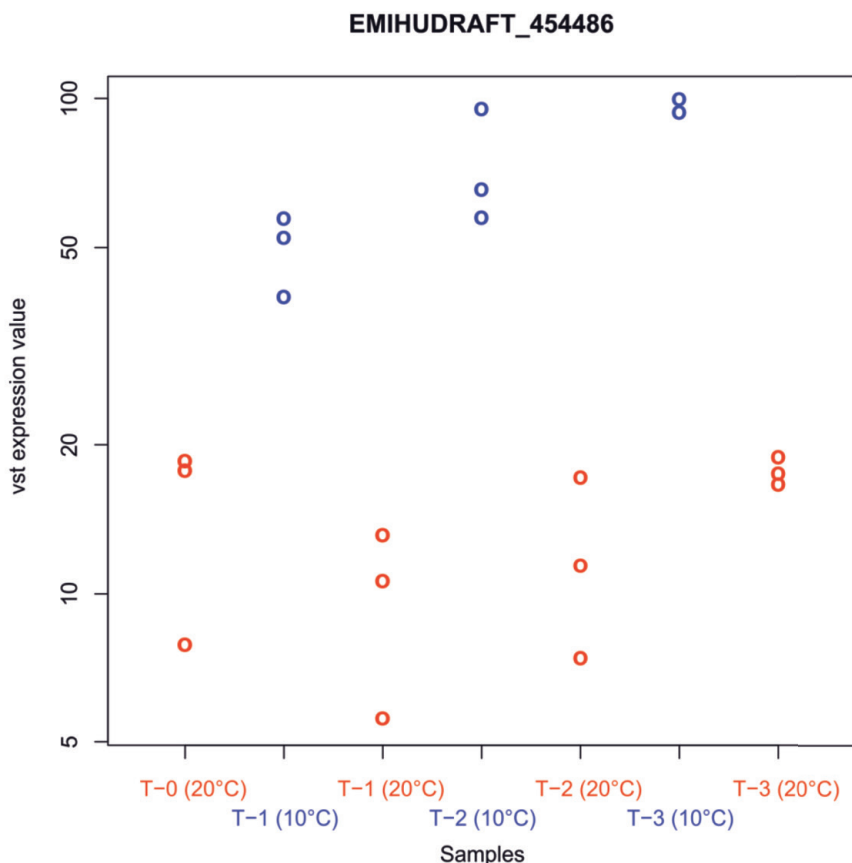
Figure 4. Gene expression of Emihudraft-454486, homologue of Akd1. Gene counts were normalized with variance stabilizing transformation (vst) using DESeq2 to yield expression values.

Our results also shed light on the location of the desaturase gene/protein in *E. huxleyi* and thus the site of desaturation. Shi et al. (2015) did not detect Akd1 in the alkenone bodies of *T. lutea*, suggesting that it operates outside these bodies. Furthermore, Endo et al. (2018) reported that Akd1 has no translocation signal sequences such as chloroplast-localization or endoplasmic reticulum (ER) retrieval signals. Similar to Akd1, no translocation signal sequences were detected within the translated amino acid sequence of EMIHUDRAFT-454486. Subcellular localization predictions based on the amino acid sequence resulted in the endoplasmic reticulum (66.9% likelihood, DeepLoc-1.0) and membrane bound (99.9% likelihood, DeepLoc-1.0) as

the most likely localization (Figure S9-10). This strengthens the idea of Shi et al. (2015) that the alkenone bodies are derived from the endoplasmic reticulum (ER) and potentially locates EMIHUDRAFT_454486 within the alkenone bodies. One of the most abundant proteins in *T. lutea* alkenone bodies was EMIHUDRAFT_465517 and speculated to be involved with alkenone metabolism (Shi et al., 2015). Most recently Shi (2019) reported the overexpression of the EMIHUDRAFT_465517 gene when transferring *T. lutea* cells from N-sufficient to N-deficient medium, which resulted in an increase of alkenone bodies and alkenone content within the cells. However, this gene was not significantly up- or down- regulated in our study but it should be noted that the concentration of LCAs ($C_{37:3} + C_{37:2}$) increased only to a minor extent during the experiment. Therefore, it is perhaps not surprising that proteins involved in the alkenone metabolism are not up- or down- regulated.

CONCLUSION

$C_{37:3}/C_{37:2}$ LCA ratios were altered within a couple of hours by applying a cold shock to *E. huxleyi* cultures. The cold-shock was also reflected in the gene expression of the cold-induced cultures where most of the affected genes that were up- or down-regulated were coupled to cellular stress responses. The up-regulated gene expression of the only known haptophyte alkenone desaturase, Akd1 and its *E. huxleyi* homologue EMIHUDRAFT-454486, coincides with the on-set of the change in the LCA ratio. This strengthens the hypothesis that EMIHUDRAFT-454486 is a potential LCA desaturase within *E. huxelyi*. The likely localization of EMIHUDRAFT-454486 within the membrane of the endoplasmic reticulum could link the protein to the alkenone bodies within *E. huxleyi*. Future proteomic isolations could potentially support or prove this. Our results could be used to specifically target EMIHUDRAFT-454486, e.g. within metagenomic datasets, to further unravel the link between the alkenone biosynthetic pathway and the often diverse LCA family in pelagic environments.

7

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

M.A.B., and L.V. designed the study; M.A.B. wrote the paper; M.A.B., S.S., J.S.S.D. and L.V. analyzed the data; M.A.B., performed experiments. All authors participated in data interpretation and provided editorial comments on the manuscript.

SUPPLEMENTARY FIGURES
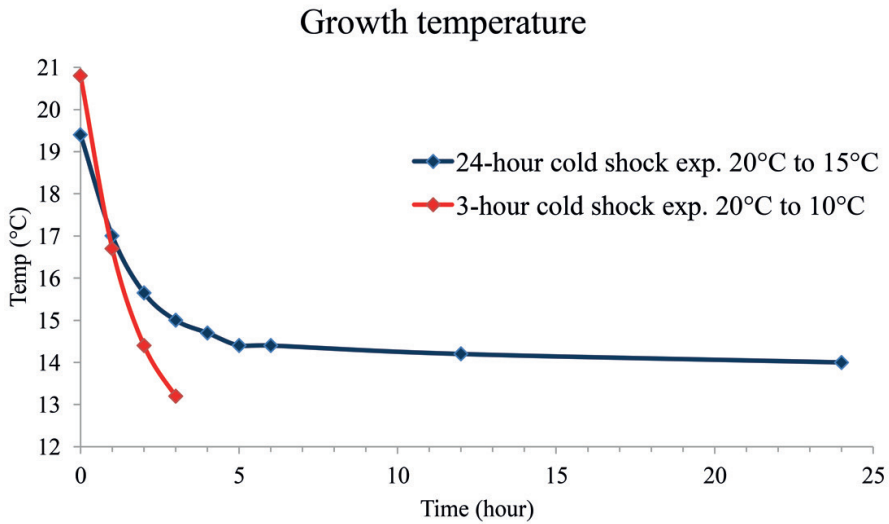
Figure S1. In-situ growth temperatures of the first cold shock experiment (in blue, 24 hours, from 20°C to 15°C) and of the second shorter cold shock experiment (in red, 3 hours, from 20°C to 10°C).
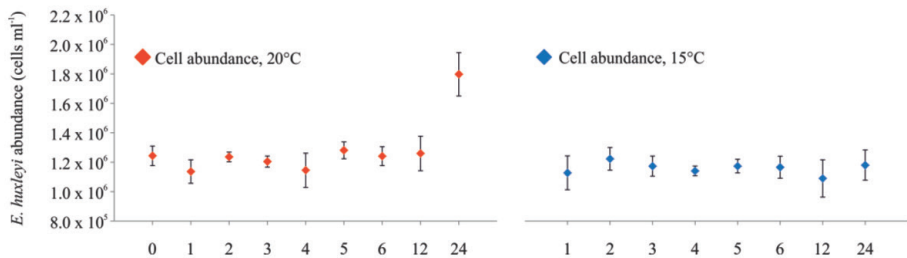
Figure S2. *E. huxleyi* cell abundance in the 24-hour experiment in the control cultures (20 °C) and the cold-shocked cultures (15 °C).
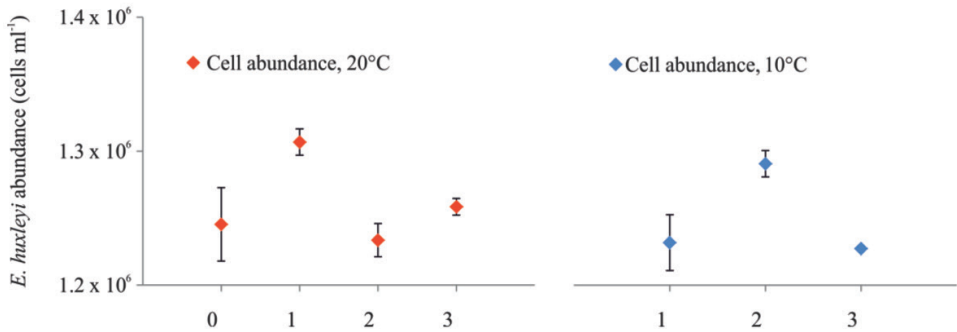
Figure S3. *E. huxleyi* cell abundance in the 3-hour experiment in the control cultures (20 °C) and the cold shocked cultures (10 °C).
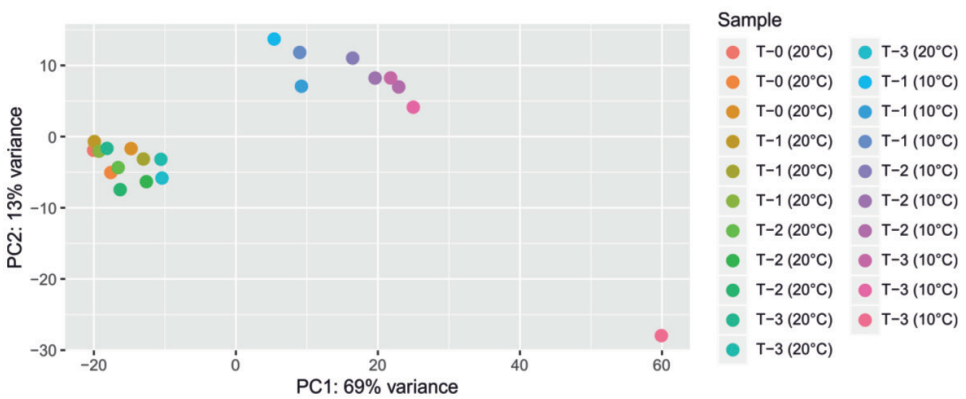
Figure S4. Principal component analysis (PCA) plot using the variance stabilizing transformed (vst) values obtained with DESeq2. Plotted are all analyzed cultures. PCA plot was obtained with ggplot2 and DESeq2.
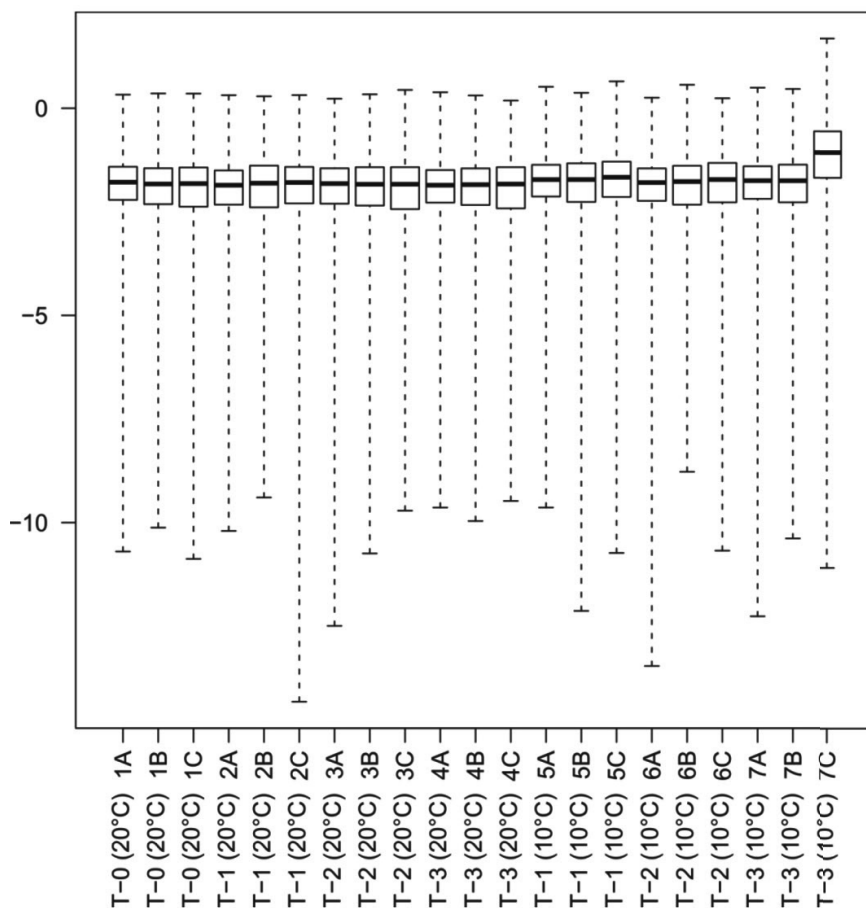
Figure S5. Cook's distance for transcriptomic data from the DeseqDataSet obtained with DESeq2. Rows within a sample that have an extreme count outlier will have the ρ-value and adjusted ρ-value set to NA. These counts are detected by Cook's distance. Sample 7C (T-3, 10 °C) has a higher Cook's distance and is removed from further analyses.

7
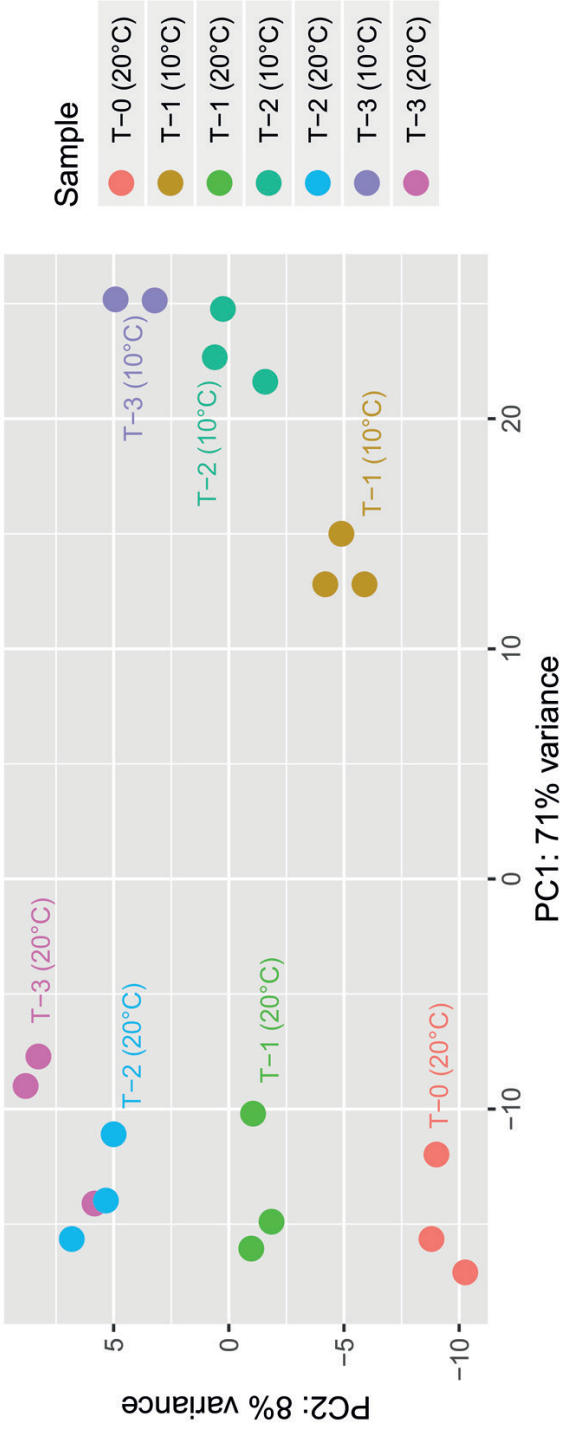
Figure S6. Principal component analysis (PCA) plot using the variance stabilizing transformed (vst) values obtained with DESeq2. Plotted are all analyzed cultures, expect the earlier discussed T-3 outlier (Figure S4-5), grouped per triplicate. Triangles are indicating the control cultures (maintained at 20 °C) and the rounds are indicating the transferred cultures (to 10 °C). PCA plot was obtained with ggplot2 and DESeq2.

C

Distribution of the log2FoldChange, control vs experiment
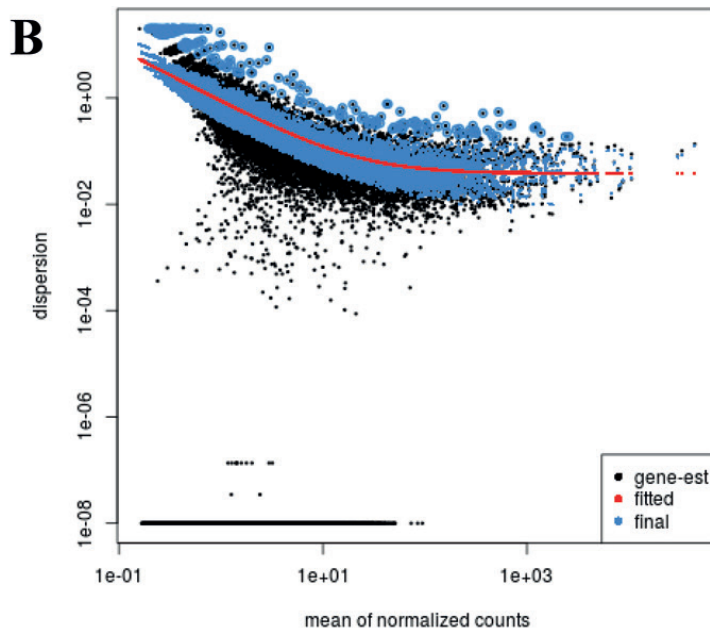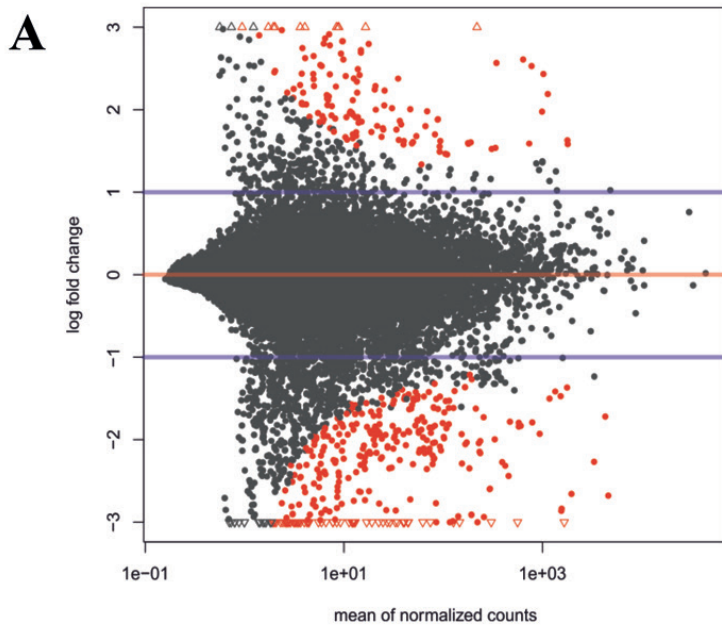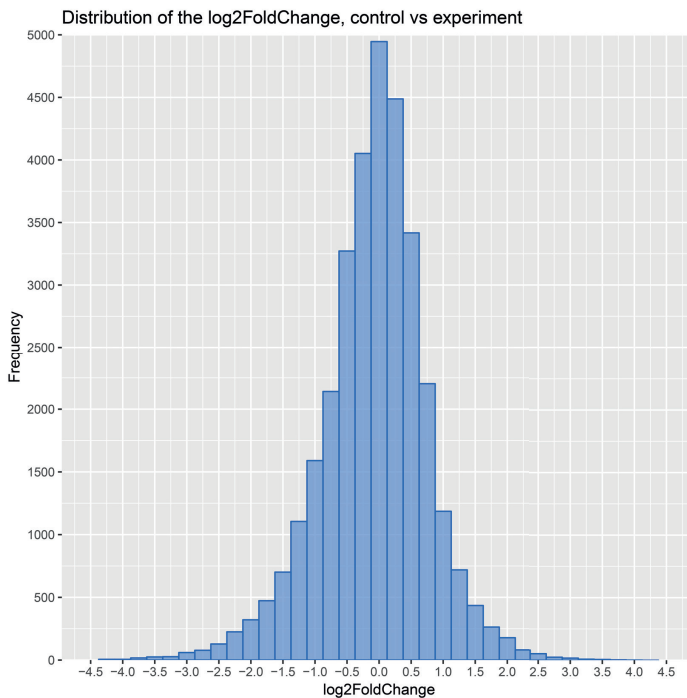
Figure S7. A) MA-plot showing the log2 fold changes (LFC) over the mean of normalized counts between the control and transferred cultures. Genes with an adjusted ρ-value <0.01 and a LFC of <-1 or >1 are indicated in red. Points which fall out of the window are plotted as open triangles pointing either up or down. Made with DESeq2. B) Plot of dispersion estimates, showing a fitted smooth curve (red line), black dots indicate genes with an associated mean expression level and maximum likelihood estimation (MLE) of the dispersion, genes with extremely high dispersion values are not shrunken towards the curve and are indicated by black dots with blue circles around. Made with DESeq2 C) Histogram showing the frequency of genes with respect to the LFC. Made with ggplot2.
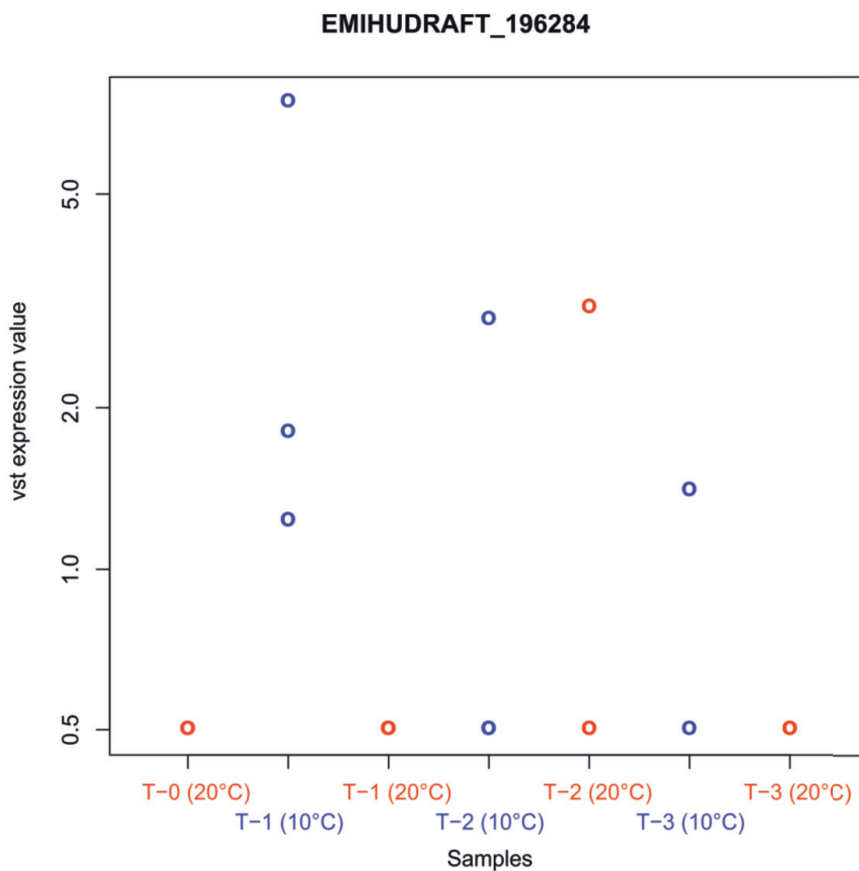
Figure S8. Gene expression of Emihudraft_196284, homologue of Akd1. Gene counts were normalized with variance stabilizing transformation (vst) using DESeq2 to yield expression values.

7

Abbreviations: CHL, chloroplast; CHLM, chloroplast membrane; CYT, cytosol; ER, endoplasmic reticulum; ERM, endoplasmic reticulum membrane; EXT, extra-cellular; GOL, Golgi apparatus; GOLM, Golgi apparatus membrane; MIT, mitochondria; MITM, mitochondria membrane; NUC, nucleus; NUCM, nucleus membrane; PER, peroxisome; PERM, peroxisome membrane; PM, plasma membrane; PLAS, plastid; VAC, vacuole; VACM, vacuole membrane.

Figure S9. Decision tree and decision path of the predicted localization in the cell of EMI-HUDRAFT_454486. Figure generated using LocTree3.

Figure S10. Decision tree and decision path of the predicted localization in the cell of EMI-HUDRAFT_454486. Diagram generated using DeepLoc-1.0. ER= endoplasmic reticulum.

7

SYNTHESIS

The marine sediment archive contains climate information spanning up to 10^7 years (Bradley, 2015). This archive can be used to reconstruct important climate variables such as salinity (e.g. Rohling, 2007; Mezger et al., 2018; Weiss and Bar, 2019), CO_2 concentrations (e.g. Pagani et al., 1999; Pearson and Palmer, 2000; Witkowski et al., 2018) and Sea Surface Temperature (SST; e.g. Zachos et al., 2001; Lopes dos Santos et al., 2013; Smith et al., 2013) among others. Reconstruction of SSTs is important because the oceans represent the largest surface area on Earth and influence global climate. In the recent decades, multiple paleotemperature proxies to reconstruct SSTs have been developed that are based on sedimentary lipids (Brassell et al., 1986; Prahl and Wakeham, 1987; Schouten et al., 2002; Pearson et al., 2004; Rampen et al., 2012; Zhang et al., 2016). E.g. the U^{K}_{37}, based on the degree of unsaturation of long chain alkenones synthesized by haptophytes, the TEX_{86}, based on the relative abundance of glycerol dialkyl glycerol tetraethers (GDGTs) produced by Thaumarchaeota and lastly the LDI, based on specific long chain diols produced by Eustigmatophyte algae. These proxies have been proven to be valuable in reconstructing past climate (e.g. Pelejero et al., 1999; Jenkyns et al., 2012). However, there are still uncertainties in these proxies, e.g. the origin (sources) of these lipids in the marine environment. Also, the lipid biosynthetic pathways leading to the lipids involved in these proxies are largely unknown. The understanding of these pathways could improve our knowledge on other environmental factors, other than temperature which could be affecting the lipid composition. In this thesis, I approach these problems by applying genomics and lipidomics approaches.

Archaea diversity and their lipids in the marine environment

All the above-mentioned SST reconstruction proxies are based on lipids formed in the upper parts of the water column, which then get preserved in the sedimentary record. However, for the TEX_{86} there is also the possibility that non-planktonic sources of GDGTs contribute to the GDGT pool present in the sedimentary record. One of these sources could be the archaeal populations residing within the marine sediments, the benthic population. These benthic archaeal communities

contain often a high diversity of archaeal groups (Teske and Sørensen, 2008; Teske, 2013; Lloyd et al., 2013; Spang et al., 2017). Compared to the pelagic marine environment, the Thaumarchaeota community composition in sediments is still not fully known. Also uncultured archaeal groups such as the Miscellaneous Crenarchaeota Group (MCG) and Archaea belonging to the superphylum DPANN (composed of Micrarchaeota, Diapherotrites, Aenigmarchaeota, Nanohaloarchaeota, Parvarchaeota, Nanoarchaeota, Pacearchaeota and Woesearchaeota) and Marine Benthic Group B and D (MBG-B, MBG-D) are often detected in (anoxic) marine sediments (Teske and Sørensen, 2008; Rinke et al., 2013; Lloyd et al., 2013; Meng et al., 2014; Castelle et al., 2015). To get an insight in the archaeal community composition and the IPL-GDGT composition within marine sediment we studied (sub)surface sediments located within or below the Oxygen Minimum Zone (OMZ) of the Arabian Sea (chapter 2). A high diverse archaeal community composition was detected within the anoxic sediments, with a dominance of MCG Archaea (belonging to the Bathyarchaeota phylum). The lipid composition of the yet-uncultured Bathyarchaeota is still unknown but this study suggest that they might be able to synthesize GDGT-0 with an unknown polar head group. This would imply that a potential contribution of Bathyarchaeota lipids to the anoxic sediment pool would not influence the TEX_{86} estimates. Within the oxygenated surface sediments mostly Thaumarchaeota were detected, belonging to several Operational Taxonomical Units (OTUs). This coincided with a high relative abundance of hexose phosphohexose (HPH-) crenarchaeol, which has been suggested as a biomarker for living Thaumarchaeota therefor it is likely that Thaumarchaeota were intact and living within these sediments.

Ultimately, more studies trying to enrich or isolate benthic Archaea will be needed to determine the exact lipid composition of these uncultured groups. Unfortunately, cultivation approaches are still challenging. Alternatively, other methodological approaches could be investigated such as cell sorting coupled with high sensitivity lipidomic detection to determine lipid composition of particles or enrichments without the need of achieving pure cultures. In addition, the question remains of

what fraction of the sedimentary lipids are produced in situ within the sediment or deposited from the overlying water column. A way to test the amount of in situ production could be achieved by studies based on labeled substrates which could be consumed by Archaea and incorporated in their lipids in a similar way to that described by Lengger et al. (2014). Nevertheless, the low incorporation rates within IPL-GDGTs indicating a low activity of sedimentary Archaea and/or a low turnover rate of IPL-GDGTs would complicate this approach. Most recently Evans et al. (2019) showed, with ^{14}C-bicarbonate and ^{14}C-acetate, relative low production rates of IPL-GDGTs with a monohexose headgroup, suggesting a predominantly pelagic origin. Evans et al. (2019) also detected the production of CL-GDGTs within sediments suggesting an active benthic archaeal community. e.g. enrichments of marine sediments led to a dominance of anaerobic methane oxidizing group-1 (ANME-1) Archaea (Wegener et al., 2016).

The diversity and lipid composition of Archaea in pelagic systems was also investigated in this thesis. The major source of GDGTs in the marine environment are Thaumarchaeota and their lipids are part of the TEX_{86} proxy, assuming that their lipids are primarily produced at the surface of the water column (Schouten et al., 2002). However it is has been observed that TEX_{86} often reflect subsurface temperatures (ca. 30-200 meters below sea level), which was later attributed to the fact that most Thaumarchaeota reside in subsurface waters (Huguet et al., 2007; Jia et al., 2012; Ho and Laepple, 2016; Park et al., 2018). Besides, analyses of GDGTs Δ^{14}C values indicated that Thaumarchaeota residing in deeper waters are also contributing to the GDGT pool (Pearson et al., 2001; Ingalls et al., 2006; Shah et al., 2008). The increase of GDGT-2 over GDGT-3 with increasing depth has first been described by Taylor et al. (2013) and suggested to be caused by deep dwelling Thaumarchaeota. Kim et al. (2015) also observed an increase of the GDGT-2/3 ratio with increasing depth in the Mediterranean Sea and linked this to a possible presence of a Thaumarchaeota community producing GDGTs that are also incorporated within the sedimentary GDGT pool. This input of deep water produced GDGTs could potentially disturb the TEX_{86} signal and bias SST reconstructions,

possible causing the warm bias of TEX_{86} in the Mediterranean Sea (Leider et al., 2010; Castañeda et al., 2010; Huguet et al., 2011). To further elucidate the causes of this bias in the Mediterranean Sea, the archaeal diversity and (IPL) GDGT composition in the Mediterranean Sea water column at three different locations (west, middle and east) was investigated (chapter 3). This showed a dominance of Euryarchaeota, MGII and MGIII, in upper pelagic waters (0 − 100 meters below sea level) and Thaumarchaeota dominating the subsurface and deeper parts of the water column. This shift coincided with an increase of IPL crenarchaeol and a decrease of IPL GDGT-0. Also, the GDGT-2/-3 ratio increased with increasing water depth, which coincided with a high relative abundance of deep-water MGI Archaea. This may be linked to particularities of the Mediterranean Sea with high temperature and salinity within specific water masses. This increase of GDGT-2/-3 ratio does not imply an effect on the TEX_{86} signal however it suggests the presence of deep-water Thaumarchaeota, or other Archaea, with a different GDGT membrane composition. The increase of the GDGT-2/-3 ratio has been observed in multiple marine environments and is likely a global marine phenomenon, however the GDGT-2/-3 ratios observed in the Mediterranean Sea were not as high as e.g. in the Atlantic Ocean (Basse et al., 2014; Hernández-Sánchez et al., 2014; Hurley et al., 2018). This could be caused by the relatively high and constant in situ temperatures throughout the water column of the Mediterranean Sea. However, it is not resolved what it causing the shift in the GDGT(-2/-3) composition with increasing depth and should be investigated in future studies. There is also currently no cultured Thaumarchaeota from the deeper-water column, this would also greatly benefit our knowledge on the membrane composition of deep-water Thaumarchaeota and their potential contribution to the GDGT lipid pool.

Besides Thaumarchaeota it has also been suggested that other marine archaeal groups are capable of producing GDGTs, e.g. Lincoln et al. (2014a) suggested, based on environmental data, that Euryarchaeota Marine Group II (MGII) Archaea are capable of producing archaeal GDGTs, including crenarchaeol. This could potentially contribute significantly to the total marine GDGT pool therefore compromising

TEX$_{86}$. However because of the low amounts of extracted DNA and archaeal reads in combination with the use of core lipid (CL-) GDGTs (can be derived from cell-debris) the conclusion made by Lincoln et al. (2014a) rendered some discussion (Lincoln et al., 2014b; Schouten et al., 2014). In chapter 4 we analyzed suspended particulate matter from the North Atlantic Ocean and from coastal North Sea waters. Here we found that the archaeal community in North Atlantic Ocean surface waters was dominated by Euryarchaeota MGII and MGIII Archaea (< 50 mbsl) and an archaeal community entirely populated out of MGII Archaea in coastal North Sea water. In none of these samples dominated by Euryarchaeota, we were able to detect any of the known intact polar lipid (IPL-) GDGTs. Also, after removing the polar head groups with acid hydrolyses, we could not detect any substantial amount of IPL-derived GDGTs, which would be expected if MGII Euryarchaeota were able to synthesize GDGTs. This implies that MGII/MGIII do not synthesize GDGTs and they do not affect the TEX$_{86}$ estimates. However, the question remains what kind of membrane lipids are synthesized by MGII Archaea. Further research with the use of enrichments and cultivation of MGII Archaea could resolve this enigma.

Apart from the different archaeal groups which can contribute to both the benthic and pelagic archaeal lipid pool, it is also important to determine the effect of environmental variables in the existing archaeal community, for example oxygen availability. In order to tackled this problem, we investigated the archaeal community and lipid composition in the water column of the Arabian Sea and the Eastern Tropical South Pacific (ETSP; both containing an Oxygen Minimum Zone; OMZ) and compared this with the composition within the fully oxygenated water column of the North Atlantic (chapter 5). Oxygen levels affected the relative abundances of Thaumarchaeota, with lower relative abundances within the OMZs, however still occurring within the OMZs which strengthens the idea that Thaumarchaeota can grow in oxygen depleted waters. The presence of MGIII Euryarchaeota has previously been linked to archaeol-based IPLs however these lipids were not detected in the OMZ of the Arabian Sea, suggesting that the production of archaeol by MGIII could be environmentally induced or species-linked. Also, an increase was

observed of the ring-index (based on the amount of ring moieties within GDGTs) with decreasing oxygen concentrations for the dihexose (DH-) and HPH-GDGTs. However, this pattern was more obvious in the Arabian Sea and North Atlantic Ocean water column. This suggests that oxygen concentrations affect the IPL-GDGT composition, although other physicochemical factors or differences in the archaeal communities could also be the cause. The relation between oxygen concentrations and (IPL-) GDGT composition in the marine environment must studied into more detail.

The understanding of the (marine) archaeal diversity has greatly expanded in the last decades with the increasing use and development of DNA sequencing tools (e.g. Woese and Fox, 1977; Barns et al., 1996; Hershberger et al., 1996; Spang et al., 2015; Spang et al., 2017). For our archaeal community composition we sequenced fragments of the 16S rRNA gene, however it is known that this approach can "miss" archaeal groups due to biases caused by primer mismatches (Bahram et al., 2019). We have no evidence for missing certain groups, as for example in the archaeal communities within the Arabian Sea sediments (chapter 2) we detected a high diversity of Archaeal groups belonging to multiple different phyla. In future studies it would be beneficial to switch to or include (to compare) archaeal community compositions derived from metagenomics to exclude any potential biases caused by primer mismatches. Metagenomics can also provide additional clues to the lipid producing capabilities of detected Archaea, e.g. Villanueva et al. (2017) showed by analyzing metagenomic data that MGII Euryarchaeota are lacking the gene for glycerol-1-phosphate dehydrogenase (G1PDH), the enzyme needed to form the glycerol-1-phosphate (G1P) backbone in archaeal membrane lipids. Therefore, it could be that the MGII Euryarchaeota are producing membrane lipids with a different backbone compared to other archaeal groups or even fatty acid-based membranes, which would fit well with the fact that not known IPL archaeal lipids were detected in the samples of coastal North Sea water. Nevertheless, metagenomic data only hints towards the possible membrane lipid composition, and ultimately

pure cultures or enrichments should be obtained to prove the predictions based on environmental studies.

Sources and lipid biosynthesis of algal lipids related to paleothermometry

As seen in the other proxies, the producers of long chain diols involved in the LDI paleotemperature proxy are not known in the marine environment. LCD have been seen to be synthesized by eustigmatophyte algae but still little is known about the diversity of this group of LCD-producers nor of their niche preference in the environment. In order to understand more of the potential LCD producers, the water column of a freshwater lake, lake Challa (chapter 6) was investigated for the LCD composition as well as the diversity of eustigmatophytes both in the water column in depth as well as in different seasons. A maximum abundance of LCDs was observed at 9 meters depth, coinciding with the maximum abundance of Eustigmatophyceae 18S rRNA gene copies suggesting that this depth can be the preferential niche of this algal group or an accumulation of LCD-producers while sinking. Besides, the application of a 18S rRNA gene based approach also highlighted the fact that the diversity of eustigmatophytes was higher than expected and covered yet-uncultured groups. Changes in seasonal LCD abundances and distributions also suggested different blooms of LCD-producers or changing physiochemical factors influencing the LCD composition. The study in Lake Challa revealed an unknown diversity of LCD-producers, however, LCD-producers from freshwater systems are likely different than those found in marine environments, according to a recent study focused on LCD compositions in river and in marine systems (Lattaud et al., 2018). Currently it is still unknown who the actual producers of LCDs in the marine environment are. A recent study by Balzano et al. (2018) detected high amounts of LCDs in Suspended Particulate Matter (SPM) in the North Atlantic Ocean. Nevertheless, known LCD producers were only detected, often in very low relative abundances, in half of the studied samples, therefore it remains uncertain who the source(s) of LCDs in marine environment is. More culture studies could resolve this problem such has been done for marine and freshwater producers (e.g. Volkman et al.,

1999; Rampen et al., 2009; 2011), as well as environmental studies such as reported in chapter 6 and by Lattaud et al. (2016) and Balzano et al. (2018), which have been proven to be helpful in suggesting potential producers. The next step in the study of LCD-producers could be linked to the detection of genes of the LCD biosynthetic pathway. In this regard, the LCD pathway is still pending to be fully resolved but a recent study by Balzano et al. 2019) has proposed several pathways leading to the LCD biosynthesis. Once a key LCD biosynthetic gene has been identified in the future, this could be then used to screen the presence of the potential ability to synthesize LCD both in sequenced genomes and in environmental samples.

For the case of the Long Chain Alkenones (LCAs), there are also uncertainties regarding their main producers as well as how they are synthesized. The biosynthetic pathway of LCAs is largely unknown but some proposals have been put forward by e.g. Rontani et al. (2006). Besides, Kitamura et al. (2018) recently described the conversion of di-unsaturated ($C_{37:2}$) into tri-unsaturated ($C_{37:3}$) alkenones by cold-shocking *Emiliania huxleyi* cultures. Building on to this study, Endo et al. (2018), found that overexpression of a novel desaturase (Akd1) within *Tisochrysis lutea* resulted in a conversion of $C_{37:2}$ into $C_{37:3}$ alkenones. In order to better understand the LCA biosynthetic pathway as well as determining if these proposed enzymes are also found in other haptophytes, in chapter 7 we performed a cold-shock *E. huxleyi* cultures which coincided with a rapid increase in the $C_{37:3}$/$C_{37:2}$ ratio. With the use of transcriptomics there was a clear differentiation of gene expression detected after the cold-shock, among the differentiated genes was also gene EMIHUDRAFT-454486 swiftly up-regulated. This gene is a homologue of the putative LCA desaturase Akd1, likely responsible for the transformation of $C_{37:2}$ into $C_{37:3}$ LCAs in *Tisochrysis lutea*. This method could be used to target other genes involved in the biosynthetic pathway of LCAs, e.g. on the formation of other double bonds in LCAs. The homologue of Akd1, EMIHUDRAFT-454486, could also be targeted in the future in metagenomic datasets to further unravel the link between the biosynthetic pathway of LCAs and the often diverse LCA diversity in pelagic environments.

Outlook

Organic paleotemperature proxies will, most certainly, remain widely used in the upcoming future for the reconstruction of past temperature. It is essential to keep applying the latest technology both in the analysis of lipid biomarkers and also of the identity and ecophysiology of the producers. By taking into account the identity of the lipid biomarker producer, its biology and niche preference, it will be possible to better understand the predictions made by the paleotemperature proxies as well as to correct certain interpretations. Thus, advancement of the field of organic geochemistry should come hand by hand with the application of genomic and biological approaches. The use genomic approaches and the combination of genomics and lipidomics in environmental samples as seen in this thesis, has provided very useful information on the possible biological sources of specific lipids. Still, these studies keep having a degree of uncertainty as we are comparing molecules (i.e. DNA and lipids), which in some conditions could have different residence times in the environmental samples and therefore could not be connected. In this regard, the ideal situation would be to obtain enrichments or cultures of yet-uncultured microbial species and check their lipid synthesis. Several studies and initiatives nowadays are seeing the importance of cultivation and it is foresee that the next years will bring developments in the isolation of fastidious strains. For example, recently it has been reported the isolation of the first Asqard archaeum, "*Candidatus* Prometheoarchaeum syntrophicum strain MK-D1" by Imachi et al. (2019) which will most likely shed light on the membrane lipid composition of this ubiquitous archaeal group. It is expected that other archaeal groups, such as the MGII studied in this thesis, will become available in culture in the near future with the aid of new cultivation approaches and by using the information contained in their genomes.

Apart from new developments in culturing approaches, it is expected that other methodological advancements will help in this field of research. In this thesis, high-mass-accuracy HPLC-MS methods have been applied to determine the lipid composition in environmental samples. This method generated large datasets from which we only analyzed a limited amount of features. However, this data could be

further analyzed with the use of lipidomics software pipelines such as LOBSTAHS (Collins et al., 2016). It would be recommended to combine this information with genomic data obtained from the same samples containing not only microbial identity (rRNA gene taxonomy) but also metabolic potential and lipid biosynthetic pathways, which can also be related to the final lipid product. Other analytical developments, which would advance this field, would be instruments that could analyze single cell lipids or determine the lipids in sorted cells or aggregates. This could be combined with the identification of the cells by fluorescence in situ hybridization approaches and metabolic measurements with Nanoscale Secondary Ion Mass Spectrometry (NanoSIMS) and partially alleviate the needs of cultivation.

REFERENCES

IPCC, (2007). Climate change 2007: synthesis report. Contribution of working groups I, II and III to the fourth assessment report of the intergovernmental panel on climate change [core writing team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp.

Allard, B., & Templier, J. (2000). Comparison of neutral lipid profile of various trilaminar outer cell wall (TLS)-containing microalgae with emphasis on algaenan occurrence. Phytochemistry, 54, 369–380.

Almagro Armenteros, Jose Juan, Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). Sequence analysis DeepLoc: Prediction of protein subcellular localization using deep learning. Bioinformatics, 33(21), 3387-3395. https://doi.org/10.1093/bioinformatics/btx431

Almagro Armenteros, José Juan, Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S.,   Nielsen, H. (2019). SignalP5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology. https://doi.org/10.1038/s41587-019-0036-z

Alonso-Saez, L., Waller, A. S., Mende, D. R., Bakker, K., Farnelid, H., Yager, P. L., … Bertilsson, S. (2012). Role for urea in nitrification by polar marine Archaea. Proceedings of the National Academy of Sciences, 109(44), 17989–17994. https://doi.org/10.1073/pnas.1201914109

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389-3402.

Amann, R. I., Ludwig, W., & Schleifer, K. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological Reviews, 59(1), 143–169.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from http://bioinformatics.babraham.ac.uk/projects/fastqc

Arndt, S., Jørgensen, B. B., Larowe, D. E., Middelburg, J. J., Pancost, R. D., & Regnier, P. (2013). Quantifying the degradation of organic matter in marine sediments: a review and synthesis. Earth Science Reviews, 123, 53-86. https://doi.org/10.1016/j.earscirev.2013.02.008

Arrhenius, S. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. Philosophical Magazine and Journal of Science, 41(5), 237–276.

Artegiani, A., Paschini, E., Russo, A., Bregant, D., Raicich, F., Pinardi, N., … Pinardi, N. (1997). The Adriatic Sea general circulation. Part II: Baroclinic circulation structure. Journal of Physical Oceanography, 27(8), 1515–1532. https://doi.org/10.1175/1520-0485(1997)027<1515:TASGCP>2.0.CO;2

Auguet, J.-C., & Casamayor, E. O. (2013). Partitioning of Thaumarchaeota populations along environmental gradients in high mountain lakes. FEMS Microbiology Ecology, 84(1), 154–164. https://doi.org/10.1111/1574-6941.12047

Augustin, L. C., Barbante, P. R. F., Barnes, J. M., Barnola, M., Bigler, E., Castellano, O., … 2004. (2004). Eight glacial cycles from an Antarctic ice core. Nature, 429(6992), 623–628.

Bahram, M., Anslan, S., Hildebrand, F., Bork, P., & Tedersoo, L. (2019). Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. Environmental Microbiology Reports, 11(4), 487–494. https://doi.org/10.1111/1758-2229.12684

Baker, B. J., & Dick, G. J. (2013). Omic approaches in microbial ecology: Charting the unknown. Microbe, 8(9), 353–360.

Bale, N. J., Villareal, T. A., Hopmans, E. C., Brussaard, C. P. D., Besseling, M., Dorhout, D., … Schouten, S. (2018). C5 glycolipids of heterocystous cyanobacteria track symbiont abundance in the diatom Hemiaulus hauckii across the tropical North Atlantic. Biogeosciences, 15(4), 1229–1241. https://doi.org/10.5194/bg-15-1229-2018

Balzano, S., Lattaud, J., Villanueva, L., Rampen, S., Brussaard, C. P. D., van Bleijswijk, J., … Schouten, S. (2018). A quest for the biological sources of the ubiquitous long chain alkyl diols in the marine realm. Biogeosciences, 15, 5951–5968. https://doi.org/10.5194/bg-2018-97

Balzano, S., Villanueva, L., Bar, M. De, Sahonero Canavesi, D. X., Yildiz, C., Engelmann, J. C., … Schouten, S. (2019). Biosynthesis of long chain alkyl diols and long chain alkenols in Nannochloropsis spp. (Eustigmatophyceae). Plant and Cell Physiology, 60(8), 1666–1682. https://doi.org/10.1093/pcp/pcz078

Bano, N., Ruffin, S., Ransom, B., & Hollibaugh, J. T. (2004). Phylogenetic composition of Arctic Ocean archaeal assemblages and comparison with Antarctic assemblages. Applied and Environmental Microbiology, 70(2), 781–789. https://doi.org/10.1128/AEM.70.2.781

Bard, E. (2001). Comparison of alkenone estimates with other paleotemperature proxies. Geochemistry Geophysics Geosystems, 2(1). https://doi.org/10.1029/2000GC000050

Barns, S. M., Delwiche, C. F., Palmer, J. D., & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proceedings of the National Academy of Sciences, 93, 9188–9193.

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

Bayer, B., Vojvoda, J., Offre, P., Alves, R. J. E., Elisabeth, N. H., Garcia, J. A. L., … Herndl, G. J. (2016). Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. The ISME Journal, 10(5), 1051–1063. https://doi.org/10.1038/ismej.2015.200

Becker, K. W., Lipp, J. S., Versteegh, G. J. M., Wörmer, L., & Hinrichs, K.-U. (2015). Rapid and simultaneous analysis of three molecular sea surface temperature proxies and application to sediments from the Sea of Marmara. Organic Geochemistry, 85, 42-53. https://doi.org/10.1016/j.orggeochem.2015.04.008

References

Belmar, L., Molina, V., & Ulloa, O. (2011). Abundance and phylogenetic identity of archaeoplankton in the permanent oxygen minimum zone of the eastern tropical South Pacific. FEMS Microbiology Ecology, 78, 314–326. https://doi.org/10.1111/j.1574-6941.2011.01159.x

Beman, J. M., Popp, B. N., & Alford, S. E. (2012). Quantification of ammonia oxidation rates and ammonia-oxidizing archaea and bacteria at high resolution in the Gulf of California and eastern tropical North Pacific Ocean. Limnology and Oceanography, 57(3), 711–726. https://doi.org/10.4319/lo.2012.57.3.0711

Beman, J. M., Popp, B. N., & Francis, C. A. (2008). Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. The ISME Journal, 2(4), 429–441. https://doi.org/10.1038/ismej.2007.118

Bendif, E. M., Probert, I., & Schroeder, D. C. (2013). On the description of Tisochrysis lutea gen. nov. sp. nov. and Isochrysis nuda sp. nov. in the Isochrysidales, and the transfer of Dicrateria to the Prymnesiales (Haptophyta). Journal of Applied Phycology, 25, 1763–1776. https://doi.org/10.1007/s10811-013-0037-0

Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., … Herndl, G. J. (2018). Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. Proceedings of the National Academy of Sciences of the United States of America, 115(3), E400–E408. https://doi.org/10.1073/pnas.1708779115

Besseling, M. A., Hopmans, E. C., Boschman, R. C., Sinninghe Damsté, J. S., & Villanueva, L. (2018). Benthic archaea as potential sources of tetraether membrane lipids in sediments across an oxygen minimum zone. Biogeosciences, 15(13), 4047-4064. https://doi.org/10.5194/bg-15-4047-2018

Biddle, J. F., Lipp, J. S., Lever, M. a, Lloyd, K. G., Sørensen, K. B., Anderson, R., Hinrichs, K.-U. (2006). Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. Proceedings of the National Academy of Sciences of the United States of America, 103(10), 3846-3851. https://doi.org/10.1073/pnas.0600035103

Bligh, E. G., Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. Canadian Journal of Biochemistry and Physiology, 37(8), 911–917. https://doi.org/10.1139/o59-099.

Boere, A. C., Rijpstra, W. I. C., de Lange, G. J., Sinninghe Damsté, J. S., & Coolen, M. J. L. (2011). Preservation potential of ancient plankton DNA in Pleistocene marine sediments. Geobiology, 9, 377-393. https://doi.org/10.1111/j.1472-4669.2011.00290.x

Boere, A. C., Rijpstra, W. I. C., Lange, G. J. De, Malinverno, E., Sinninghe Damsté, J. S., & Coolen, M. J. L. (2011). Exploring preserved fossil dinoflagellate and haptophyte DNA signatures to infer ecological and environmental changes during deposition of sapropel S1 in the eastern Mediterranean. Paleoceanography, 26, 1-16. https://doi.org/10.1029/2010PA001948

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boyd, E. S., Pearson, A., Pi, Y., Li, W.-J., Zhang, Y. G., He, L., … Geesey, G. G. (2011). Temperature and pH controls on glycerol dibiphytanyl glycerol tetraether lipid composition in the hyperthermophilic crenarchaeon Acidilobus sulfurireducens. Extremophiles, 15, 59–65. https://doi.org/10.1007/s00792-010-0339-y

Bradley, R. S. (2015). Paleoclimatic reconstruction. In Paleoclimatology (3rd ed., pp. 1–11). Elsevier Inc. https://doi.org/10.1016/B978-0-12-386913-5.00001-6

Brassell, S. C., Eglinton, G., Marlowe, I. T., Pflaumann, U., & Sarnthein, M. (1986). Molecular stratigraphy: a new tool for climate assessment. Nature, 320, 129–133.

Bristow, L. A., Callbeck, C. M., Larsen, M., Altabet, M. A., Dekaezemacker, J., Forth, M., … Canfield, D. E. (2017). N2 production rates limited by nitrite availability in the Bay of Bengal oxygen minimum zone. Nature Geoscience, 10, 24–31. https://doi.org/10.1038/NGEO2847

Brocks, J. J., Logan, G. A., Buick, R., & Summons, R. E. (1999). Archean molecular fossils and the early rise of eukaryotes. Science, 285, 1033–1036.

Buckles, L. K., Villanueva, L., Weijers, J. W. H., Verschuren, D., & Sinninghe Damsté, J. S. (2013). Linking isoprenoidal GDGT membrane lipid distributions with gene abundances of ammonia-oxidizing Thaumarchaeota and uncultured crenarchaeotal groups in the water column of a tropical lake (Lake Challa, East Africa). Environmental Microbiology, 2, 1-18. https://doi.org/10.1111/1462-2920.12118

Caforio, A., Siliakus, M. F., Exterkate, M., Jain, S., Jumde, V. R., Andringa, R. L. H., … van der Oost, J. (2018). Converting Escherichia coli into an archaebacterium with a hybrid heterochiral membrane. Proceedings of the National Academy of Sciences, 115(14), 3704–3709. https://doi.org/10.1073/pnas.1721604115

Callendar, G. S. (1938). The artificial production of carbon dioxide and its influence on temperature. Quarterly Journal of the Royal Meteorological Society, 64(275), 223–240.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7(5), 335–336. https://doi.org/10.1038/NMETH.F.303

Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-lyons, D., Gonzalez, A., Stombaugh, J., … Knight, R. (2011). Moving pictures of the human microbiome. Genome Biology, 12(5), 1–8. https://doi.org/10.1186/gb-2011-12-5-r50

Castañeda, I. S., Schefuß, E., Pätzold, J., Sinninghe Damsté, J. S., Weldeab, S., & Schouten, S. (2010). Millennial-scale sea surface temperature changes in the eastern Mediterranean (Nile River Delta region) over the last 27,000 years. Paleoceanography, 25(1), PA1208. https://doi.org/10.1029/2009PA001740

Castellari, S., Pinardi, N., & Leanman, K. (2000). Simulation of water mass formation processes in the Mediterranean Sea: Influence of the time frequancy of the atmospheric forcing. Journal of Geophysical Research, 105(C10), 24157–24181.

References

Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., … Banfield, J. F. (2015). Genomic expansion of domain Archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Current Biology, 25, 1–12. https://doi.org/10.1016/j.cub.2015.01.014

Cess, R. D., & Goldenberg, S. D. (1981). The effect of ocean heat capacity upon global warming due to increasing atmospheric carbon dioxide. Journal of Geophysical Research, 86, 498–502.

Chave, K. E. (1954). Aspects of the biogeochemistry of magnesium 1. calcareous marine organisms. The Journal of Geology, 62(3), 266–283.

Chivall, D., M'Boule, D., Sinke-Schoen, D., Sinninghe Damsté, J. S., Schouten, S., & van der Meer, M. T. J. (2014). Impact of salinity and growth phase on alkenone distributions in coastal haptophytes. Organic Geochemistry, 67, 31-34. https://doi.org/10.1016/j.orggeochem.2013.12.002

Collins, J. R., Edwards, B. R., Fredricks, H. F., & Van Mooy, B. A. S. (2016). LOBSTAHS: An adduct-based lipidomics strategy for discovery and identification of oxidative stress biomarkers. Analytical Chemistry, 88(14), 7154–7162. https://doi.org/10.1021/acs.analchem.6b01260

Conte, M. H., Thompson, A., Eglinton, G., & Green, J. C. (1995). Lipid biomarker diversity in the coccolithophorid Emiliania huxleyi (Prymnesiophyceae) and the related species Gephyrocapsa oceanica. Journal of Phycology, 31, 272–282.

Conte, M. H., Thompson, A., Lesley, D., & Harris, R. P. (1998). Genetic and physiological influences on the alkenone/alkenoate versus growth temperature relationship in Emiliania huxleyi and Gephyrocapsa oceanica. Geochimica et Cosmochimica Acta, 62(1), 51–68.

Conte, M. H., Weber, J. C., Schulte, S., & Blanz, T. (2006). Global temperature calibration of the alkenone unsaturation index (Uk37) in surface waters and comparison with surface sediments. Geochemistry Geophysics Geosystems, 7(2), 1–22. https://doi.org/10.1029/2005GC001054

Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics, 19(1), 15–18.

Coolen, M. J. L., Abbas, B., van Bleijswijk, J., Hopmans, E. C., Kuypers, M. M. M., Wakeham, S. G., & Sinninghe Damsté, J. S. (2007). Putative ammonia-oxidizing Crenarchaeota in suboxic waters of the Black Sea: a basin-wide ecological study using 16S ribosomal and functional genes and membrane lipids. Environmental Microbiology, 9(4), 1001-1016. https://doi.org/10.1111/j.1462-2920.2006.01227.x

Coolen, M. J. L., Boere, A., Abbas, B., Baas, M., Wakeham, S. G., & Damste, J. S. S. (2006). Ancient DNA derived from alkenone-biosynthesizing haptophytes and other algae in Holocene sediments from the Black Sea. Paleoceanography, 21, 1–17. https://doi.org/10.1029/2005PA001188

Coolen, M. J. L., Hopmans, E. C., Rijpstra, W. I. C., Muyzer, G., Schouten, S., Volkman, J. K., & Sinninghe Damsté, J. S. (2004). Evolution of the methane cycle in Ace Lake (Antarctica) during the Holocene: response of methanogens and methanotrophs to environmental change. Organic Geochemistry, 35(10), 1151-1167. https://doi.org/10.1016/j.orggeochem.2004.06.009

Coolen, M. J. L., Muyzer, G., Rijpstra, W. I. C., Schouten, S., Volkman, J. K., & Sinninghe Damsté, J. S. (2004). Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic lake. Earth and Planetary Science Letters, 223(1-2), 225-239. https://doi.org/10.1016/j.epsl.2004.04.014

Curtis, T. P., Sloan, W. T., & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of the United States of America, 99(16), 10494–10499.

D'Andrea, W. J., & Huang, Y. (2005). Long chain alkenones in Greenland lake sediments: Low delta13C and exceptional abundance. Organic Geochemistry, 36, 1234–1241. https://doi.org/10.1016/j.orggeochem.2005.05.001

De Corte, D., Yokokawa, T., Varela, M. M., Agogué, H., & Herndl, G. J. (2009). Spatial distribution of Bacteria and Archaea and amoA gene copy numbers throughout the water column of the Eastern Mediterranean Sea. The ISME Journal, 3, 147-158. https://doi.org/10.1038/ismej.2008.94

De La Torre, J. R., Walker, C. B., Ingalls, A. E., Könneke, M., & Stahl, D. A. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. Environmental Microbiology, 10(3), 810-818. https://doi.org/10.1111/j.1462-2920.2007.01506.x

de Leeuw, J. W., Rijpstra, W. I. C., Schenck, P. A., & Volkman, J. K. (1983). Free, esterified bound sterols in Black Sea Unit I sediments. Geochimica et Cosmochimica Acta, 47, 455–465.

De Rosa, M., & Gambacorta, A. (1988). The lipids of archaebacteria. Progress in Lipid Research, 27(3), 153–175. https://doi.org/https://doi.org/10.1016/0163-7827(88)90011-2

DeLong, E. F. (1992). Archaea in coastal marine environments. Proceedings of the National Academy of Sciences, 89, 5685–5689.

Delong, E. F., & Pace, N. R. (2001). Environmental diversity of Bacteria and Archaea. Society of Systematic Biologists, 50(4), 470–478. https://doi.org/10.1080/10635150118513

Delong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N., … Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. Science, 311, 496–503.

DeLong, E. F., Ying Wu, K., Prézelin, B. B., & Jovine, R. V. M. (1994). High abundance of Archaea in Antarctic marine picoplankton. Nature, 371, 695-697. https://doi.org/10.1038/371695a0

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Durbin, A. M., & Teske, A. (2010). Sediment-associated microdiversity within the Marine Group I Crenarchaeota. Environmental Microbiology Reports, 2(5), 693–703. https://doi.org/10.1111/j.1758-2229.2010.00163.x

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

References

Elling, F. J., Becker, K. W., Könneke, M., Schröder, J. M., Kellermann, M. Y., Thomm, M., & Hinrichs, K. U. (2016). Respiratory quinones in Archaea: Phylogenetic distribution and application as biomarkers in the marine environment. Environmental Microbiology, 18(2), 692-707. https://doi.org/10.1111/1462-2920.13086

Elling, F. J., Konneke, M., Lipp, J. S., Becker, K. W., Gagen, E. J., & 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. Geochimica et Cosmochimica Acta, 141, 579–597. https://doi.org/10.1016/j.gca.2014.07.005

Elling, F. J., Könneke, M., Mußmann, M., Greve, A., & Hinrichs, K.-U. (2015). Influence of temperature, pH, and salinity on membrane lipid composition and TEX86 of marine planktonic thaumarchaeal isolates. Geochimica et Cosmochimica Acta, 171, 238-255. https://doi.org/10.1016/j.gca.2015.09.004

Elling, F. J., Könneke, M., Nicol, G. W., Stieglmeier, M., Bayer, B., Spieck, E., Hinrichs, K. U. (2017). Chemotaxonomic characterisation of the thaumarchaeal lipidome. Environmental Microbiology, 19(7), 2681-2700. https://doi.org/10.1111/1462-2920.13759

Eltgroth, M. L., Watwood, R. L., & Wolfe, G. V. (2005). Production and cellular localization of neutral long-chain lipids in the haptophyte algae Isochrysis galbana and Emiliania huxleyi. J. Phycol., 41, 1000–1009.

Emiliani, C. (1955). Pleistocene temperatures. Journal of Geology, 63, 538–578.

Endo, H., Hanawa, Y., Araie, H., Suzuki, I., & Shiraiwa, Y. (2018). Overexpression of Tisochrysis lutea Akd1 identifies a key cold-induced alkenone desaturase enzyme. Scientific Reports, 8(1), 1–10. https://doi.org/10.1038/s41598-018-29482-8

Epstein, B. L., Hondt, S. D., & Hargraves, P. E. (2001). The possible metabolic role of C37 alkenones in Emiliania huxleyi. Organic Geochemistry, 32, 867–875.

Evans, T. W., Coffinet, S., Konneke, M., Lipp, J. S., Becker, K. W., Elvert, M., … Hinrichs, K.-U. (2019). Assessing the carbon assimilation and production of benthic archaeal lipid biomarkers using lipid-RIP. Geochimica et Cosmochimica Acta.

Fillol, M., Auguet, J.-C., Casamayor, E. O., & Borrego, C. M. (2015). Insights in the ecology and evolutionary history of the Miscellaneous Crenarchaeotic Group lineage. The ISME Journal, 1–13. https://doi.org/10.1038/ismej.2015.143

Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E., & Oakley, B. B. (2005). Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. Proceedings of the National Academy of Sciences of the United States of America, 102(41), 14683–14688. https://doi.org/10.1073/pnas.0506625102

Frigaard, N. U., Martinez, A., Mincer, T. J., & DeLong, E. F. (2006). Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. Nature, 439(7078), 847–850. https://doi.org/10.1038/nature04435

Fuhrman, J A, McCallum, K., & Davis, A. A. (1992). Novel major archaebacterial group from marine plankton. Nature, 356(6365), 148–149. https://doi.org/10.1038/356148a0

Fuhrman, Jed A, & Davis, A. A. (1997). Widespread Archaea and novel Bacteria from the deep sea as shown by 16s rRNA gene sequences. Marine Ecology Progress Series, 150, 275–285.

Galand, P. E., Casamayor, E. O., Kirchman, D. L., Potvin, M., & Lovejoy, C. (2009). Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. The ISME Journal, 3(7), 860–869. https://doi.org/10.1038/ismej.2009.23

Galand, P. E., Gutiérrez-Provecho, C., Massana, R., Gasol, J. M., & Casamayor, E. O. (2010). Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). Limnology and Oceanography, 55(5), 2117-2125. https://doi.org/10.4319/lo.2010.55.5.2117

Gascard, J. C., & Richez, C. (1985). Water masses and circulation in the western Alboran Sea and in the Straits of Gibraltar. Progress in Oceanography, 15(3), 157–216. https://doi.org/10.1016/0079-6611(85)90031-X

Gelin, F., Boogers, I., Noordeloos, A. A. M., Sinninghe Damsté, J. S., Riegman, R., & de Leeuw, J. W. (1997). Resistant biomaeromolecules in maline microalgae of the classes Eustigmatophyceae and Chlorophyceae: Geochemical implications. Organic Geochemistry, 26, 659-675.

Gelin, F., Noordeloos, A. A. M., Sinninghe Damsté, J. S., Hatcher, P. G., & de Leeuw, J. W. (1996). Novel, resistant microalgal polyethers: An important sink of organic carbon in the marine environment? Geochimica et Cosmochimica Acta, 60(7), 1275-1280.

Gelin, F., Volkman, J. K., de Leeuw, J. W., & Sinninghe Damsté, J. S. (1997). Mid-chain hydroxy long-chain fatty acids in microalgae from the genus nannochloropsis. Phytochemistry, 45(4), 641-646.

Gelin, F., Volkman, J. K., Largeau, C., Derenne, S., Sinninghe Damsté, J. S., & de Leeuw, J. W. (1999). Distribution of aliphatic, nonhydrolyzable biopolymers in marine microalgae. Organic Geochemistry, 30(2-3), 147-159. https://doi.org/10.1016/S0146-6380(98)00206-X

Gillett, N. P., Arora, V. K., Zickfeld, K., Marshall, S. J., & Merryfield, W. J. (2011). Ongoing climate change following a complete cessation of carbon dioxide emissions. Nature Geoscience, 4(2), 83–87. https://doi.org/10.1038/ngeo1047

Gillies, L. E., Thrash, J. C., deRada, S., Rabalais, N. N., & Mason, O. U. (2015). Archaeal enrichment in the hypoxic zone in the northern Gulf of Mexico. Environmental Microbiology, 17(10), 3847–3856. https://doi.org/10.1016/j.jcsr.2018.02.024

Gliozzi, A., Paoli, G., De Rosa, M., & Gambacorta, A. (1983). Effect of isoprenoid cyclization on the transition temperature of lipids in thermophilic archaeabacteria. Biochimica et Biophysica Acta, 735, 234–242.

Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., … Rost, B. (2014). LocTree3 prediction of localization. Nucleic Acids Research, 42(May), 350–355. https://doi.org/10.1093/nar/gku396

Guillard, R. R. L., & Hargraves, P. E. (1993). Stichochrysis immobilis is a diatom, not a chrysophyte. Phycologia, 32(3), 234–236.

References

Hallam, S. J., Mincer, T. J., Schleper, C., Preston, C. M., Roberts, K., Richardson, P. M., & DeLong, E. F. (2006). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. PLoS Biology, 4(4), 520–536. https://doi.org/10.1371/journal.pbio.0040095

Haro-Moreno, J. M., Rodriguez-Valera, F., López-García, P., Moreira, D., & Martin-Cuadrado, A.-B. (2017). New insights into marine group III Euryarchaeota, from dark to light. The ISME Journal, 1-16. https://doi.org/10.1038/ismej.2016.188

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

Herbert, T. D. (2003). Alkenone Paleotemperature Determinations. In H. D. Holland & K. K. Turekian (Eds.), Treatise on Geochemistry (pp. 391–432). Elsevier Science. https://doi.org/https://doi.org/10.1016/B0-08-043751-6/06115-6

Herfort, L., Schouten, S., Abbas, B., Veldhuis, M. J. W., Coolen, M. J. L., Wuchter, C., Sinninghe Damsté, J. S. (2007). Variations in spatial and temporal distribution of Archaea in the North Sea in relation to environmental variables. FEMS Microbiology Ecology, 62(3), 242-257. https://doi.org/10.1111/j.1574-6941.2007.00397.x

Hernández-Sánchez, M. T., Woodward, E. M. S., Taylor, K. W. R., Henderson, G. M., & Pancost, R. D. (2014). Variations in GDGT distributions through the water column in the South East Atlantic Ocean. Geochimica et Cosmochimica Acta, 132, 337-348. https://doi.org/10.1016/j.gca.2014.02.009

Herndl, G. J., Reinthaler, T., Teira, E., Aken, H. Van, Veth, C., Pernthaler, A., & Pernthaler, J. (2005). Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. Applied and Environmental Microbiology, 71(5), 2303–2309. https://doi.org/10.1128/AEM.71.5.2303

Hershberger, K. L., Barns, S. M., Reysenbach, A. L., Dawson, S. C., & Pace, N. R. (1996). Wide diversity of Crenarchaeota. Nature. https://doi.org/10.1038/384420a0

Ho, S. L., & Laepple, T. (2016). Flat meridional temperature gradient in the early Eocene in the subsurface rather than surface ocean. Nature Geoscience, 9(august), 606–610. https://doi.org/10.1038/NGEO2763

Hönisch, B., Hemming, N. G., Archer, D., Siddall, M., & McManus, J. F. (2009). Atmospheric carbon dioxide concentration across the Mid-Pleistocene transition. Science, 324, 1551-1554. https://doi.org/10.1126/science.1171477

Hopkins, T. S. (1985). Physics of the sea. In R. Margalef (Ed.), Key Environments: West Mediterranean (pp. 100–125). New York: Pergamon Press.

Hopmans, E. C., Schouten, S., & Sinninghe Damsté, J. S. (2016). The effect of improved chromatography on GDGT-based palaeoproxies. Organic Geochemistry, 93, 1-6. https://doi.org/10.1016/j.orggeochem.2015.12.006

Horton, P., Park, K., Obayashi, T., Fujita, N., Harada, H., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Research, 35(17), 585–587. https://doi.org/10.1093/nar/gkm259

Houpert, L., Durrieu de Madron, X., Testor, P., Bosse, A., D'Ortenzio, F., Bouin, M. N., … Raimbault, P. (2016). Observations of open-ocean deep convection in the northwestern Mediterranean Sea: Seasonal and interannual variability of mixing and deep water masses for the 2007-2013 period. Journal of Geophysical Research: Oceans, 121, 8139–8171.

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Jouan Dufournel, I., Bronner, G., … Galand, P. E. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proceedings of the National Academy of Sciences of the United States of America, 110(15), 6004–6009. https://doi.org/10.1073/pnas.1216863110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1216863110

Huguet, C., Cartes, J. E., Sinninghe Damsté, J. S., & Schouten, S. (2006). Marine crenarchaeotal membrane lipids in decapods: Implications for the TEX86 paleothermometer. Geochemistry, Geophysics, Geosystems, 7(11). https://doi.org/10.1029/2006GC001305

Huguet, C., Hopmans, E. C., Febo-Ayala, W., Thompson, D. H., Sinninghe Damsté, J. S., & Schouten, S. (2006). An improved method to determine the absolute abundance of glycerol dibiphytanyl glycerol tetraether lipids. Organic Geochemistry, 37(9), 1036-1041. https://doi.org/10.1016/j.orggeochem.2006.05.008

Huguet, C., Martrat, B., Grimalt, J. O., Sinninghe Damsté, J. S., & Schouten, S. (2011). Coherent millennial-scale patterns in UK37' and TEX86H temperature records during the penultimate interglacial-to-glacial cycle in the western Mediterranean. Paleoceanography, 26(2), 1-10. https://doi.org/10.1029/2010PA002048

Huguet, C., Schimmelmann, A., Thunell, R., Lourens, L. J., Sinninghe Damsté, J. S., & Schouten, S. (2007). A study of the TEX86 paleothermometer in the water column and sediments of the Santa Barbara Basin, California. Paleoceanography, 22, 1-9. https://doi.org/10.1029/2006PA001310

Huguet, C., Urakawa, H., Martens-Habbena, W., Truxal, L., Stahl, D. A., & Ingalls, A. E. (2010). Changes in intact membrane lipid content of archaeal cells as an indication of metabolic status. Organic Geochemistry, 41(9), 930–934. https://doi.org/10.1016/j.orggeochem.2010.04.012

Hurley, S. J., Elling, F. J., Martin, K., Buchwald, C., Wankel, S. D., Santoro, A. E., … Pearson, A. (2016). Influence of ammonia oxidation rate on thaumarchaeal lipid composition and the TEX86 temperature proxy. Proceedings of the National Academy of Sciences of the United States of America, 113(28), 7762–7767. https://doi.org/10.1073/pnas.1518534113

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

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., … Takai, K. (2019). Isolation of an archaeon at the prokaryote-eukaryote interface. BioRxiv, 1–36.

References

Inagaki, F., Suzuki, M., Takai, K., Oida, H., Sakamoto, T., Aoki, K., … Horikoshi, K. (2003). Microbial communities associated with geological horizons in coastal subseafloor sediments from the Sea of Okhotsk. Applied and Environmental Microbiology, 69(12), 7224–7235. https://doi.org/10.1128/AEM.69.12.7224–7235.2003

Ingalls, A. E., Huguet, C., & 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. Applied and Environmental Microbiology, 78(5), 1480–1490. https://doi.org/10.1128/AEM.07016-11

Ingalls, A. E., Shah, S. R., Hansman, R. L., Aluwihare, L. I., Santos, G. M., Druffel, E. R. M., & Pearson, A. (2006). Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. Proceedings of the National Academy of Sciences of the United States of America, 103(17), 6442–6447.

Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L., & Armbrust, E. V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science, 335(6068), 587–590. https://doi.org/10.1126/science.1212665

Jahn, U., Summons, R., Sturt, H., Grosjean, E., & Huber, H. (2004). Composition of the lipids of Nanoarchaeum equitans and their origin from its host Ignicoccus sp. strain KIN4/I. Archives of Microbiology, 182(5), 404–413. https://doi.org/10.1007/s00203-004-0725-x

Jenkyns H. C., Schouten-Huibers L., Schouten S. and Sinninghe Damsté J. S. (2012) Warm middle Jurassic–early Cretaceous high-latitude sea-surface temperatures from the Southern Ocean. Climate of the Past, 8, 215–226. https://doi.org/10.5194/cp-8-215-2012

Jia, G., Zhang, J., Chen, J., Peng, P., & Zhang, C. L. (2012). Archaeal tetraether lipids record subsurface water temperature in the South China Sea. Organic Geochemistry, 50, 68–77. https://doi.org/10.1016/j.orggeochem.2012.07.002

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., Mcginnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Research, 36, 5–9. https://doi.org/10.1093/nar/gkn201

Jorgenson, S. L., Hannisdal, B., Lanzén, A., Baumberger, T., Flesland, K., Fonseca, R., Schleper, C. (2012). Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. Proceedings of the National Academy of Sciences, 109(42), E2846-E2855. https://doi.org/10.1594/PANGAEA.786687

Kalanetra, K. M., Bano, N., & Hollibaugh, J. T. (2009). Ammonia-oxidizing Archaea in the Arctic Ocean and Antarctic coastal waters. Environmental Microbiology, 11(9), 2434–2445. https://doi.org/10.1111/j.1462-2920.2009.01974.x

Kaneko, R., Nagata, T., Suzuki, S., & Hamasaki, K. (2016). Depth-dependent and seasonal variability in archaeal community structure in the subarctic and subtropical western North Pacific. Journal of Oceanography, 72(3), 427–438. https://doi.org/10.1007/s10872-016-0372-2

Karner, M. B., DeLong, E. F., & Karl, D. M. (2001). Archaeal dominance in the mesopelagiczone of the Pacific Ocean. Nature, 409(6819), 507–510. https://doi.org/10.1038/35054051

Kellermann, M. Y., Yoshinaga, M. Y., Wegener, G., Krukenberg, V., & Hinrichs, K. (2016). Tracing the production and fate of individual archaeal intact polar lipids using stable isotope probing. Organic Geochemistry, 95, 13–20. https://doi.org/10.1016/j.orggeochem.2016.02.004

Kim, J.-H., Crosta, X., Willmott, V., Renssen, H., Bonnin, J., Helmke, P., Sinninghe Damsté, J. S. (2012). Holocene subsurface temperature variability in the eastern Antarctic continental margin. Geophysical Research Letters, 39, 1-6. https://doi.org/10.1029/2012GL051157

Kim, J.-H., Romero, O. E., Lohmann, G., Donner, B., Laepple, T., Haam, E., & Sinninghe Damsté, J. S. (2012). Pronounced subsurface cooling of North Atlantic waters off Northwest Africa during Dansgaard-Oeschger interstadials. Earth and Planetary Science Letters, 339-340, 95-102. https://doi.org/10.1016/j.epsl.2012.05.018

Kim, J.-H., Schouten, S., Hopmans, E. C., Donner, B., & Sinninghe Damsté, J. S. (2008). Global sediment core-top calibration of the TEX86 paleothermometer in the ocean. Geochimica et Cosmochimica Acta, 72, 1154-1173. https://doi.org/10.1016/j.gca.2007.12.010

Kim, J.-H., Schouten, S., Rodrigo-Gámiz, M., Rampen, S., Marino, G., Huguet, C., Sinninghe Damsté, J. S. (2015). Influence of deep-water derived isoprenoid tetraether lipids on the TEX86H paleothermometer in the Mediterranean Sea. Geochimica et Cosmochimica Acta, 150, 125-141. https://doi.org/10.1016/j.gca.2014.11.017

Kim, J.-H., van der Meer, J., Schouten, S., Helmke, P., Willmott, V., Sangiorgi, F., Sinninghe Damsté, J. S. (2010). New indices and calibrations derived from the distribution of crenarchaeal isoprenoid tetraether lipids, Implications for past sea surface temperature reconstructions. Geochimica et Cosmochimica Acta, 74, 4639-4654.

Kim, J.-H., Villanueva, L., Zell, C., & Sinninghe Damsté, J. S. (2016). Biological source and provenance of deep-water derived isoprenoid tetraether lipids along the Portuguese continental margin. Geochimica et Cosmochimica Acta, 172, 177-204. https://doi.org/10.1016/j.gca.2015.09.010

Kitamura, E., Kotajima, T., Sawada, K., Suzuki, I., & Shiraiwa, Y. (2018). Cold-induced metabolic conversion of haptophyte di- to tri- unsaturated C37 alkenones used as palaeothermometer molecules. Scientific Reports, 8, 1–11. https://doi.org/10.1038/s41598-018-20741-2

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glockner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research, 41(1), 1–11. https://doi.org/10.1093/nar/gks808

Kodner, R. B., Summons, R. E., & Knoll, A. H. (2009). Phylogenetic investigation of the aliphatic, non-hydrolyzable biopolymer algaenan, with a focus on green algae. Organic Geochemistry, 40, 854–862. https://doi.org/10.1016/j.orggeochem.2009.05.003

Koga, Y., & Morii, H. (2007). Biosynthesis of ether-type polar lipids in Archaea and evolutionary considerations. Microbiology and Molecular Biology Reviews, 71(1), 97–120. https://doi.org/10.1128/MMBR.00033-06

Koho, K. A., Nierop, K. G. J., Moodley, L., Middelburg, J. J., Pozzato, L., Soetaert, K., … Reichart, G. J. (2013). Microbial bioavailability regulates organic matter preservation in marine sediments. Biogeosciences, 10(2), 1131–1141. https://doi.org/10.5194/bg-10-1131-2013

References

Kolde, R. (2018). pheatmap: Pretty Heatmaps. Retrieved from https://cran.r-project.org/package=pheatmap

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., & Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature, 437(22), 543-546. https://doi.org/10.1038/nature03911

Kraal, P., Slomp, C. P., Reed, D. C., Reichart, G.-J., & Poulton, S. W. (2012). Sedimentary phosphorus and iron cycling in and below the oxygen minimum zone of the northern Arabian Sea. Biogeosciences, 9(7), 2603–2624. https://doi.org/10.5194/bg-9-2603-2012

Krom, M. D., Kress, N., Brenner, S., & Gordon, L. I. (1991). Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. Limnology and Oceanography, 36(3), 424–432. https://doi.org/10.4319/lo.1991.36.3.0424

Kubo, K., Lloyd, K. G., F Biddle, J., Amann, R., Teske, A., & Knittel, K. (2012). Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. The ISME Journal, 6(10), 1949–1965. https://doi.org/10.1038/ismej.2012.37

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution, 33(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

La Cono, V., Ruggeri, G., Azzaro, M., Crisafi, F., Decembrini, F., Denaro, R., … Yakimov, M. M. (2018). Contribution of bicarbonate assimilation to carbon pool dynamics in the deep Mediterranean Sea and cultivation of actively nitrifying and CO2-fixing bathypelagic prokaryotic consortia. Frontiers in Microbiology, 9(January), 1–18. https://doi.org/10.3389/fmicb.2018.00003

Lam, P., Jensen, M. M., Lavik, G., Mcginnis, D. F., Muller, B., Schubert, C. J., … Kuypers, M. M. M. (2007). Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea. Proceedings of the National Academy of Sciences of the United States of America, 104(17), 7104–7109.

Lam, P., Lavik, G., Jensen, M. M., Vossenberg van de, J., Schmid, M., Woebken, D., … Kuypers, M. M. M. (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. Proceedings of the National Academy of Sciences, 106(12), 4752–4757. https://doi.org/10.3969/j.issn.1001-1749.2015.03.13

Langworthy, T. A., Mayberry, W. R., & Smith, P. F. (1974). Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of Sulfolobus acidocaldarius. Journal of Bacteriology, 119(1), 106–116.

Lattaud, J., Kim, J.-H., De Jonge, C., Zell, C., Sinninghe Damsté, J. S., & Schouten, S. (2016). The C32 alkane-1,15-diol as a tracer for riverine input in coastal seas. Geochimica et Cosmochimica Acta, 202, 146-158. https://doi.org/10.1016/j.orggeochem.2019.103904

Lattaud, J., Kirkels, F., Peterse, F., Freymond, C. V, Eglinton, T. I., Hefter, J., … Schouten, S. (2018). Long-chain diols in rivers: distribution and potential biological sources. Biogeosciences, 15, 4147–4161. https://doi.org/10.5194/bg-15-4147-2018

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., … Carey, V. J. (2013). Software for computing and annotating genomic ranges. PLOS Computational Biology, 9(8), 1–10. https://doi.org/10.1371/journal.pcbi.1003118

Lazar, C. S., Baker, B. J., Seitz, K., Hyde, A. S., Dick, G. J., Hinrichs, K. U., & Teske, A. P. (2016). Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. Environmental Microbiology, 18(4), 1200–1211. https://doi.org/10.1111/1462-2920.13142

Learman, D. R., Henson, M. W., Thrash, J. C., Temperton, B., Brannock, P. M., Santos, S. R., … Halanych, K. M. (2016). Biogeochemical and microbial variation across 5500 km of Antarctic surface sediment implicates organic matter as a driver of benthic community structure. Frontiers in Microbiology, 7, 1–11. https://doi.org/10.3389/fmicb.2016.00284

Leider, A., Hinrichs, K. U., Mollenhauer, G., & Versteegh, G. J. M. (2010). Core-top calibration of the lipid-based UK37' and TEX86 temperature proxies on the southern Italian shelf (SW Adriatic Sea, Gulf of Taranto). Earth and Planetary Science Letters, 300(1–2), 112–124. https://doi.org/10.1016/j.epsl.2010.09.042

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., … Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature, 442(7104), 806–809. https://doi.org/10.1038/nature04983

Lengger, S. K., Hopmans, E. C., Reichart, G.-J., Nierop, K. G. J., Sinninghe Damsté, J. S., & Schouten, S. (2012). Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids in the Arabian Sea oxygen minimum zone. Part II: Selective preservation and degradation in sediments and consequences for the TEX86. Geochimica et Cosmochimica Acta, 98, 244-258. https://doi.org/10.1016/j.gca.2012.05.003

Lengger, S. K., Hopmans, E. C., Sinninghe Damsté, J. S., & Schouten, S. (2014). Impact of sedimentary degradation and deep water column production on GDGT abundance and distribution in surface sediments in the Arabian Sea: Implications for the TEX86 paleothermometer. Geochimica et Cosmochimica Acta, 142, 386-399. https://doi.org/10.1016/j.gca.2014.07.013

Lengger, S. K., Lipsewers, Y. A., Haas, H. de, Sinninghe Damsté, J. S., & Schouten, S. (2014). Lack of 13C-label incorporation suggests low turnover rates of thaumarchaeal intact polar tetraether lipids in sediments from the Iceland shelf. Biogeosciences, 11, 201-216. https://doi.org/10.5194/bg-11-201-2014

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research, 44(W1), W242–W245. https://doi.org/10.1093/nar/gkw290

Li, M., Baker, B. J., Anantharaman, K., Jain, S., Breier, J. A., & Dick, G. J. (2015). Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nature Communications, 1–6. https://doi.org/10.1038/ncomms9933

References

Lincoln, S. A., Wai, B., Eppley, J. M., Church, M. J., Summons, R. E., & DeLong, E. F. (2014). Planktonic Euryarchaeota are a significant source of archaeal tetraether lipids in the ocean. Proceedings of the National Academy of Sciences of the United States of America, 111(27), 9858–9863. https://doi.org/10.1073/pnas.1409439111

Lincoln, S. A., Wai, B., Eppley, J. M., Curch, M. J., Summons, R. E., & Delong, E. F. (2014). Reply to Schouten et al .: Marine Group II planktonic Euryarchaeota are significant contributors to tetraether lipids in the ocean. Proceedings of the National Academy of Sciences of the United States of America, 111(41), 4286. https://doi.org/10.1073/pnas.1416736111

Lipp, Julius S, Morono, Y., Inagaki, F., & Hinrichs, K.-U. (2008). Significant contribution of Archaea to extant biomass in marine subsurface sediments. Nature, 454, 991–994. https://doi.org/10.1038/nature07174

Lipp, Julius Sebastian, & Hinrichs, K.-U. (2009). Structural diversity and fate of intact polar lipids in marine sediments. Geochimica et Cosmochimica Acta, 73(22), 6816–6833. https://doi.org/10.1016/j.gca.2009.08.003

Lipsewers, Y. A., Hopmans, E. C., Sinninghe Damsté, J. S., & Villanueva, L. (2017). Potential recycling of thaumarchaeotal lipids by DPANN Archaea in seasonally hypoxic surface marine sediments. Organic Geochemistry. https://doi.org/10.1016/j.orggeochem.2017.12.007

Liu, F., Wang, W., Sun, X., Liang, Z., & Wang, F. (2014). RNA-Seq revealed complex response to heat stress on transcriptomic level in Saccharina japonica (Laminariales, Phaeophyta). Journal of Applied Phycology, 26(3), 1585–1596. https://doi.org/10.1007/s10811-013-0188-z

Liu, H., Zhang, C. L., Yang, C., Chen, S., Cao, Z., Zhang, Z., & Tian, J. (2017). Marine group II dominates planktonic Archaea in water column of the northeastern South China Sea. Frontiers in Microbiology, 8, 1–11. https://doi.org/10.3389/fmicb.2017.01098

Liu, X., Lipp, J. S., & Hinrichs, K.-U. (2011). Distribution of intact and core GDGTs in marine sediments. Organic Geochemistry, 42(4), 368–375. https://doi.org/10.1016/j.orggeochem.2011.02.003

Lloyd, K. G., Schreiber, L., Petersen, D. G., Kjeldsen, K. U., Lever, M. A., Steen, A. D., Jørgensen, B. B. (2013). Predominant Archaea in marine sediments degrade detrital proteins. Nature, 496(7444), 215-218. https://doi.org/10.1038/nature12033

Logemann, J., Graue, J., Köster, J., Engelen, B., Rullkötter, J., & Cypionka, H. (2011). A laboratory experiment of intact polar lipid degradation in sandy sediments. Biogeosciences, 8(9), 2547-2560. https://doi.org/10.5194/bg-8-2547-2011

Lopes dos Santos, R. A., Prange, M., Castañeda, I. S., Schefuß, E., Mulitza, S., Schulz, M., Schouten, S. (2010). Glacial-interglacial variability in Atlantic meridional overturning circulation and thermocline adjustments in the tropical North Atlantic. Earth and Planetary Science Letters, 300, 407-414. https://doi.org/10.1016/j.epsl.2010.10.030

Lopes dos Santos, R. A., Spooner, M. I., Barrows, T. T., De Deckker, P., Sinninghe Damsté, J. S., & Schouten, S. (2013). Comparison of organic (UK37, TEX86, LDI) and faunal proxies (foraminiferal assemblages) for reconstruction of late Quaternary sea-surface temperature variability from offshore southeastern Australia. Paleoceanography, 28, 1-11. https://doi.org/10.1002/palo.20035

López-Cristoffanini, C., Zapata, J., Gaillard, F., Potin, P., Correa, J. A., & Contreras-Porcia, L. (2015). Identification of proteins involved in desiccation tolerance in the red seaweed Pyropia orbicularis (Rhodophyta, Bangiales). Proteomics, 15(23-24), 3954-3968. https://doi.org/10.1002/pmic.201400625

López-García, P., Moreira, D., López-López, A., & Rodríguez-Valera, F. (2001). A novel haloarchaeal-related lineage is widely distributed in deep oceanic regions. Environmental Microbiology, 3(1), 72-78. https://doi.org/10.1046/j.1462-2920.2001.00162.x

López, J. F., & Grimalt, J. O. (2006). Reassessment of the structural composition of the alkenone distributions in natural environments using an improved method for double bond location based on GC-MS analysis of cyclopropylimines. Journal of the American Society for Mass Spectrometry, 17, 710-720. https://doi.org/10.1016/j.jasms.2006.01.015

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(550), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, … Schleifer, K.-H. (2004). ARB: a software environment for sequence data. Nucleic Acids Research, 32(4), 1363–1371. https://doi.org/10.1093/nar/gkh293

Lüke, C., Speth, D. R., Kox, M. A. R., Villanueva, L., & Jetten, M. S. M. (2016). Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. PeerJ, 4, e1924. https://doi.org/10.7717/peerj.1924

Marlowe, I. T., Brassell, S. C., Eglinton, G., & Green, J. C. (1984). Long chain unsaturated ketones and esters in living algae and marine sediments. Organic Geochemistry, 6, 135–141.

Marlowe, I. T., Brassell, S. C., Eglinton, G., & Green, J. C. (1990). Long-chain alkenones and alkyl alkenoates and the fossil coccolith record of marine sediments. Chemical Geology, 88(3–4), 349–375. https://doi.org/10.1016/0009-2541(90)90098-R

Marlowe, I. T., Green, J. C., Neal, A. C., Brassell, S. C., Eglinton, G., & Course, P. A. (1984). Long chain (n-C37–C39) alkenones in the Prymnesiophyceae. Distribution of alkenones and other lipids and their taxonomic significance. British Phycological Journal, 19, 203–216. https://doi.org/10.1080/00071618400650221

Martin-Cuadrado, A.-B., Garcia-Heredia, I., Moltó, A. G., López-Úbeda, R., Kimes, N., López-García, P., Rodriguez-Valera, F. (2014). A new class of marine Euryarchaeota group II from the mediterranean deep chlorophyll maximum. The ISME Journal, 9, 1-16. https://doi.org/10.1038/ismej.2014.249

Martin-Cuadrado, A.-B., Rodriguez-Valera, F., Moreira, D., Alba, J. C., Ivars-Martínez, E., Henn, M. R., López-García, P. (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. The ISME Journal, 2(8), 865-886. https://doi.org/10.1038/ismej.2008.40

References

Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A., … Pedros-Alio, C. (2004). Phylogenetic and ecological analysis of novel marine stramenopiles. Applied and Environmental Microbiology, 70(6), 3528–3534. https://doi.org/10.1128/AEM.70.6.3528–3534.2004

Massana, R., Delong, E. F., & Pedrós-Alió, C. (2000). A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. Applied and Environmental Microbiology, 66(5), 1777-1787. https://doi.org/10.1128/AEM.66.5.1777-1787.2000

Matthews, H. D., & Zickfeld, K. (2012). Climate response to zeroed emissions of greenhouse gases and aerosols. Nature Climate Change, 2(5), 338–341. https://doi.org/10.1038/nclimate1424

Meador, T. B., Bowles, M., Lazar, C. S., Zhu, C., Teske, A., & Hinrichs, K.-U. (2015). The archaeal lipidome in estuarine sediment dominated by members of the Miscellaneous Crenarchaeotal Group. Environmental Microbiology, 17(7), 2441–2458. https://doi.org/10.1111/1462-2920.12716

MEDOC group. (1970). Observation of formation of deep water in the Mediterranean Sea, 1969. Nature, 227(5262), 1037–1040. https://doi.org/10.1038/2271037a0

Méjanelle, L., Sanchez-Gargallo, A., Bentaleb, I., & Grimalt, J. O. (2003). Long chain n-alkyl diols, hydroxy ketones and sterols in a marine eustigmatophyte, Nannochloropsis gaditana, and in Brachionus plicatilis feeding on the algae. Organic Geochemistry, 34(4), 527-538. https://doi.org/10.1016/S0146-6380(02)00246-2

Meng, J., Xu, J., Qin, D., He, Y., Xiao, X., & Wang, F. (2014). Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. The ISME Journal, 8(3), 650–659. https://doi.org/10.1038/ismej.2013.174

Menzel, D., Hopmans, E. C., Schouten, S., & Sinninghe Damsté, J. S. (2006). Membrane tetraether lipids of planktonic Crenarchaeota in Pliocene sapropels of the eastern Mediterranean Sea. Palaeogeography, Palaeoclimatology, Palaeoecology, 239(1-2), 1-15. https://doi.org/10.1016/j.palaeo.2006.01.002

Mezger, E. M., de Nooijer, L. J., Siccha, M., Brummer, G.-J. A., Kucera, M., & Reichart, G.-J. (2018). Taphonomic and ontogenetic effects on Na/Ca and Mg/Ca in spinose planktonic Foraminifera from the Red Sea. Geochemistry Geophysics Geosystems, 19, 4174–4194. https://doi.org/10.1029/2018GC007852

Mincer, T. J., Church, M. J., Taylor, L. T., Preston, C., Karl, D. M., & DeLong, E. F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. Environmental Microbiology, 9(5), 1162–1175. https://doi.org/10.1111/j.1462-2920.2007.01239.x

Molina, V., Belmar, L., & Ulloa, O. (2010). High diversity of ammonia-oxidizing archaea in permanent and seasonal oxygen-deficient waters of the eastern South Pacific. Environmental Microbiology, 12(9), 2450–2465. https://doi.org/10.1111/j.1462-2920.2010.02218.x

Moore, E. K., Villanueva, L., Hopmans, E. C., Rijpstra, W. I. C., Mets, A., Dedysh, S. N., & Sinninghe Damsté, J. S. (2015). Abundant trimethylornithine lipids and specific gene sequences are indicative of planctomycete importance at the oxic/anoxic interface in Sphagnum-dominated northern wetlands. Applied and Environmental Microbiology, 81(18), 6333-6344. https://doi.org/10.1128/AEM.00324-15

Moore, T. C., Burckle, L. H., Geitzenauer, K., Luz, B., Molina-Cruz, A., Robertston, J. H., … Wenkam, C. (1980). The reconstruction of sea surface temperatures in the Pacific Ocean of 18,000 bp. Marine Micropaleontology, 5, 215–247.

Müller, P. J., Kirst, G., Ruhland, G., von Storch, I., & Rossel-Melé, A. (1998). Calibration of the alkenone paleotemperature index U37K' based on core-tops from the eastern South Atlantic and the global ocean (60°N-60°S). Geochimica et Cosmochimica Acta, 62(10), 1757–1772.

Murray, A. E., Preston, C. M., Massana, R., Taylor, L. T., Blakis, A., Wu, K., & DeLong, E. F. (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. Applied and Environmental Microbiology, 64(7), 2585–2595.

Newell, S. E., Babbin, A. R., Jayakumar, A., & Ward, B. B. (2011). Ammonia oxidation rates and nitrification in the Arabian Sea. Global Biogeochemical Cycles, 25, 1–10. https://doi.org/10.1029/2010GB003940

Nierop, K. G. J., Reichart, G., Veld, H., & Sinninghe Damsté, J. S. (2017). The influence of oxygen exposure time on the composition of macromolecular organic matter as revealed by surface sediments on the Murray Ridge (Arabian Sea). Geochimica et Cosmochimica Acta, 206, 40-56.

Nürnberg, D., Bijma, J., & Hemleben, C. (1996). Assessing the reliability of magnesium in foraminiferal calcite as a proxy for water mass temperatures. Geochimica et Cosmochimica Acta, 60(5), 803-814.

Ochs, D., Kaletta, C., Entian, K.-D., Beck-Sickinger, A., & Poralla, K. (1992). Cloning, expression, and sequencing of squalene-hopene cyclase, a key enzyme in triterpenoid metabolism. Journal of Bacteriology, 174(1), 298–302.

Ochs, D., Tappe, C. H., Gartner, P., Kellner, R., & Poralla, K. (1990). Properties of purified squalene-hopene cyclase from Bacillus acidocaldarius. Europian Journal of Biochemistry, 194, 75–80.

Offre, P., Spang, A., & Schleper, C. (2013). Archaea in biogeochemical cycles. Annual Review of Microbiology, 437–457. https://doi.org/10.1146/annurev-micro-092412-155614

Oni, O. E., Schmidt, F., Miyatake, T., Kasten, S., Witt, M., Hinrichs, K.-U., & Friedrich, M. W. (2015). Microbial communities and organic matter composition in surface and subsurface sediments of the Helgoland Mud Area, North Sea. Frontiers in Microbiology, 6(November), 1–16. https://doi.org/10.3389/fmicb.2015.01290

Ono, M., Swada, K., Shiraiwa, Y., & Kubota, M. (2012). Changes in alkenone and alkenoate distributions during acclimatization to salinity change in Isochrysis galbana: Implication for alkenone-based paleosalinity and paleothermometry. Geochemical Journal, 46, 235–247.

References

Orsi, W. D., Smith, J. M., Wilcox, H. M., Swalwell, J. E., Carini, P., Worden, A. Z., & Santoro, A. E. (2015). Ecophysiology of uncultivated marine euryarchaea is linked to particulate organic matter. The ISME Journal, 9(8), 1747–1763. https://doi.org/10.1038/ismej.2014.260

Ortiz-Alvarez, R., & Casamayor, E. O. (2016). High occurrence of Pacearchaeota and Woesearchaeota (Archaea superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. Environmental Microbiology Reports, 8, 210–217. https://doi.org/10.1111/1758-2229.12370

Pagani, M., Arthur, M. A., & Freeman, K. H. (1999). Miocene evolution of atmospheric carbon dioxide. Paleoceanography, 14(3), 273–292. https://doi.org/10.1029/1999PA900006

Pagani, M., Huber, M., Liu, Z., Bohaty, S. M., Henderiks, J., Sijp, W., … Deconto, R. M. (2011). The role of carbon dioxide during the onset of Antarctic glaciation. Science, 334, 1261–1264. https://doi.org/10.1126/science.1203909

Pan, H., & Sun, M. (2011). Variations of alkenone based paleotemperature index (UK37) during Emiliania huxleyi cell growth, respiration (auto-metabolism) and microbial degradation. Organic Geochemistry, 42(6), 678–687. https://doi.org/10.1016/j.orggeochem.2011.03.024

Pancost, R. D., Hopmans, E. C., Sinninghe Damsté, J. S., & Party, T. M. S. S. (2001). Archaeal lipids in Mediterranean Cold Seeps: Molecular proxies for anaerobic methane oxidation. Geochimica et Cosmochimica Acta, 65(10), 1611-1627.

Park, E., Hefter, J., Fischer, G., & Mollenhauer, G. (2018). TEX86 in sinking particles in three eastern Atlantic upwelling regimes. Organic Geochemistry, 124, 151–163. https://doi.org/10.1016/j.orggeochem.2018.07.015

Park, S.-J., Ghai, R., Martín-Cuadrado, A.-B., Rodríguez-Valera, F., Chung, W.-H., Kwon, K., Rhee, S.-K. (2014). Genomes of two new ammonia-oxidizing archaea enriched from deep marine sediments. PLoS ONE, 9(5), 1-10. https://doi.org/10.1371/journal.pone.0096449

Park, S.-J., Park, B.-J., & Rhee, S.-K. (2008). Comparative analysis of archaeal 16S rRNA and amoA genes to estimate the abundance and diversity of ammonia-oxidizing archaea in marine sediments. Extremophiles, 12(4), 605–615. https://doi.org/10.1007/s00792-008-0165-7

Pearson, A., Huang, Z., Ingalls, A. E., Romanek, C. S., Wiegel, J., Freeman, K. H., … Zhang, C. L. (2004). Nonmarine crenarchaeol in Nevada hot springs. Applied and Environmental Microbiology, 70(9), 5229–5237. https://doi.org/10.1128/AEM.70.9.5229

Pearson, A, McNichol, A. P., Benenitez-Nelson, B. C., Hayes, J. M., & Eglinton, T. I. (2001). Origins of lipid biomarkers in Santa Monica Basin surface sediment: A case study using compound-specific 14C analysis. Geochimica et Cosmochimica Acta, 65(18), 3123–3137.

Pearson, Ann, Leavitt, W. D., Sáenz, J. P., Summons, R. E., Tam, M. C., & Close, H. G. (2009). Diversity of hopanoids and squalene-hopene cyclases across a tropical land-sea gradient. Environmental Microbiology, 11(5), 1208-1223. https://doi.org/10.1111/j.1462-2920.2008.01817.x

Pearson, Ann, Page, S. R. F., Jorgenson, T. L., Fischer, W. W., & Higgins, M. B. (2007). Novel hopanoid cyclases from the environment. Environmental Microbiology, 9(9), 2175–2188. https://doi.org/10.1111/j.1462-2920.2007.01331.x

Pearson, P. N., & Palmer, M. R. (2000). Atmospheric carbon dioxide concentrations over the past 60 million years. Nature, 406, 695–699. https://doi.org/10.1038/35021000

Pelejero C., Grimalt J. O., Heilig S., Kienast M. and Wang L. (1999). High-resolution UK37 temperature reconstructions in the South China Sea over the past 220 kyr. Paleoceanography, 14, 224–231. https://doi.org/10.1029/1998PA900015

Peng, X., Jayakumar, A., Ward, B. B., Xiao, X., & Tong, S. J. (2013). Community composition of ammonia-oxidizing archaea from surface and anoxic depths of oceanic oxygen minimum zones. Frontiers in Microbiology, 4, 1–12. https://doi.org/10.3389/fmicb.2013.00177

Pitcher, A., Hopmans, E. C., Mosier, A. C., Park, S.-J., Rhee, S.-K., Francis, C. A., Sinninghe Damsté, J. S. (2011). Core and intact polar glycerol dibiphytanyl glycerol tetraether lipids of ammonia-oxidizing archaea enriched from marine and estuarine sediments. Applied and Environmental Microbiology, 77(10), 3468-3477. https://doi.org/10.1128/AEM.02758-10

Pitcher, A., Hopmans, E. C., Schouten, S., & Sinninghe Damsté, J. S. (2009). Separation of core and intact polar archaeal tetraether lipids using silica columns: Insights into living and fossil biomass contributions. Organic Geochemistry, 40(1), 12-19. https://doi.org/10.1016/j.orggeochem.2008.09.008

Pitcher, A., Rychlik, N., Hopmans, E. C., Spieck, E., Rijpstra, W. I. C., Ossebaar, J., Sinninghe Damsté, J. S. (2010). Crenarchaeol dominates the membrane lipids of Candidatus Nitrososphaera gargensis, a thermophilic group I.1b Archaeon. The ISME Journal, 4(4), 542-552. https://doi.org/10.1038/ismej.2009.138

Pitcher, A., Villanueva, L., Hopmans, E. C., Schouten, S., Reichart, G.-J., & Sinninghe Damsté, J. S. (2011). Niche segregation of ammonia-oxidizing archaea and anammox bacteria in the Arabian Sea oxygen minimum zone. The ISME Journal, 5(12), 1896-1904. https://doi.org/10.1038/ismej.2011.60

Pitcher, A., Wuchter, C., Siedenberg, K., Schouten, S., & Sinninghe Damsté, J. S. (2011). Crenarchaeol tracks winter blooms of ammonia-oxidizing Thaumarchaeota in the coastal North Sea. Limnology and Oceanography, 56(6), 2308-2318. https://doi.org/10.4319/lo.2011.56.6.2308

Podar, M., Makarova, K. S., Graham, D. E., Wolf, Y. I., Koonin, E. V, & Reysenbach, A.-L. (2013). Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. Biology Direct, 8(1), 1–20. https://doi.org/10.1186/1745-6150-8-9

Polik, C. A., Elling, F. J., & Pearson, A. (2018). Impacts of paleoecology on the TEX86 sea surface temperature proxy in the pliocene-pleistocene Mediterranean Sea. Paleoceanography and Paleoclimatology, 33, 1–18. https://doi.org/10.1029/2018PA003494

Pollak, M. I. (1951). The sources of the deep water of the eastern Mediterranean Sea. Journal of Marine Research, 10(1), 128–152.

Prahl, F. G., Sparrow, M. A., & Wolfe, G. V. (2003). Physiological impacts on alkenone paleothermometry. Paleoceanography, 18(2), n/a-n/a. https://doi.org/10.1029/2002PA000803

References

Prahl, Fredrick G., Rontani, J.-F., Volkman, J. K., Sparrow, M. a., & Royer, I. M. (2006). Unusual C35 and C36 alkenones in a paleoceanographic benchmark strain of Emiliania huxleyi. Geochimica et Cosmochimica Acta, 70(11), 2856–2867. https://doi.org/10.1016/j.gca.2006.03.009

Prahl, Fredrick G., & Wakeham, S. G. (1987). Calibration of unsaturation patterns in long-chain ketone compositions for paleotemperature assessment. Nature, 330, 367–369.

Prahl, Fredrick G, Muehlhausen, L. A., & Zahnle, D. L. (1988). Further evaluation of long-chain alkenones as indicators of paleoceanographic conditions. Geochimica et Cosmochimica Acta, 52, 2303–2310.

Přibyl, P., Eliáš, M., Cepák, V., Lukavský, J., & Kaštánek, P. (2012). Zoosporogenesis, morphology, ultrastructure, pigment composition, and phylogenetic position of Trachydiscus minutus (Eustigmatophyceae, Heterokontophyta). Journal of Phycology, 48(1), 231–242. https://doi.org/10.1111/j.1529-8817.2011.01109.x

Qin, W, Amin, S. A., Martens-Habbena, W., Walker, C. B., Urakawa, H., Devol, A. H., … Stahl, D. A. (2014). Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. Proceedings of the National Academy of Sciences of the United States of America, 111(34), 12504–12509. https://doi.org/10.1073/pnas.1324115111

Qin, Wei, Carlson, L. T., Armbrust, E. V., Devol, A. H., Moffett, J. W., Stahl, D. A., & Ingalls, A. E. (2015). Confounding effects of oxygen and temperature on the TEX86 signature of marine Thaumarchaeota. Proceedings of the National Academy of Sciences of the United States of America, 112(35), 10979–10984. https://doi.org/10.1073/pnas.1501568112

Quaiser, A., Zivanovic, Y., Moreira, D., & López-García, P. (2011). Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. The ISME Journal, 5(2), 285-304. https://doi.org/10.1038/ismej.2010.113

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., … Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41(November 2012), D590-596. https://doi.org/10.1093/nar/gks1219

Quast, Christian, Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41(Database issue), D590-6. https://doi.org/10.1093/nar/gks1219

Quinones, R. A., Levipan, H. A., & Urrutia, H. (2009). Deep-Sea Research II Spatial and temporal variability of planktonic archaeal abundance in the Humboldt Current System off Chile. Deep-Sea Research II, 56, 1073–1082. https://doi.org/10.1016/j.dsr2.2008.09.012

Rampen, S. W., Schouten, S., Schefuß, E., & Sinninghe Damsté, J. S. (2009). Impact of temperature on long chain diol and mid-chain hydroxy methyl alkanoate composition in Proboscia diatoms: Results from culture and field studies. Organic Geochemistry, 40(11), 1124-1131. https://doi.org/10.1016/j.orggeochem.2009.08.005

Rampen, S. W., Schouten, S., & Sinninghe Damsté, J. S. (2011). Occurrence of long chain 1,14-diols in Apedinella radians. Organic Geochemistry, 42(5), 572-574. https://doi.org/10.1016/j.orggeochem.2011.03.009

Rampen, S. W., Schouten, S., Wakeham, S. G., & Sinninghe Damsté, J. S. (2007). Seasonal and spatial variation in the sources and fluxes of long chain diols and mid-chain hydroxy methyl alkanoates in the Arabian Sea. Organic Geochemistry, 38(2), 165-179. https://doi.org/10.1016/j.orggeochem.2006.10.008

Rampen, S. W., Willmott, V., Kim, J.-H., Uliana, E., Mollenhauer, G., Schefuß, E., Schouten, S. (2012). Long chain 1,13- and 1,15-diols as a potential proxy for palaeotemperature reconstruction. Geochimica et Cosmochimica Acta, 84, 204-216. https://doi.org/10.1016/j.gca.2012.01.024

Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., … Grigoriev, I. V. (2013). Pan genome of the phytoplankton Emiliania underpins its global distribution. Nature, 499(7457), 209–213. https://doi.org/10.1038/nature12221

Rechka, J. A., & Maxwell, J. R. (1988). Characterisation of alkenone temperature indicators in sediments and organisms. Advances in Organic Geochemistry, 13, 727–734.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., … Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature, 499(7459), 431–437. https://doi.org/10.1038/nature12352

Roether, W., Manca, B. B., Klein, B., Bregant, D., Georgopoulos, D., Beitzel, V., … Luchetta, A. (1996). Recent changes in Eastern Mediterranean deep waters. Science, 271(January), 30–32.

Rohling, E. J. (2007). Progress in paleosalinity: Overview and presentation of a new approach. Paleoceanography, 22, 1–9. https://doi.org/10.1029/2007PA001437

Rontani, J. F., Prahl, F. G., & Volkman, J. K. (2006). Re-examination of the double bond positions in alkenones and derivatives: biosynthetic implications. J. Phycol., 42, 800–813. https://doi.org/10.1111/j.1529-8817.2006.00251.x

Rossel, P. E., Lipp, J. S., Fredricks, H. F., Arnds, J., Boetius, A., Elvert, M., & Hinrichs, K. U. (2008). Intact polar lipids of anaerobic methanotrophic archaea and associated bacteria. Organic Geochemistry, 39(8), 992–999. https://doi.org/10.1016/j.orggeochem.2008.02.021

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4), 406–425. https://doi.org/citeulike-article-id:93683

Sawada, K., & Shiraiwa, Y. (2004). Alkenone and alkenoic acid compositions of the membrane fractions of Emiliania huxleyi. Phytochemistry, 65, 1299–1307. https://doi.org/10.1016/j.phytochem.2004.03.015

Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., & Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environmental Microbiology, 11(8), 2078–2093. https://doi.org/10.1111/j.1462-2920.2009.01929.x

Schleper, C., Jurgens, G., & Jonuscheit, M. (2005). Genomic studies of uncultivated archaea. Nature Reviews Microbiology, 3(6), 479–488. https://doi.org/10.1038/nrmicro1159

References

Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., … Zillig, W. (1995). Picrophilus gen. nov., fam. nov.: A novel aerobic, heterotrophic, thermoacidophilic genus and family comprising Archaea capable of growth around pH 0. Journal of Bacteriology, 177(24), 7050–7059. https://doi.org/10.1128/jb.177.24.7050-7059.1995

Schouten, S., Hopmans, E. C., Baas, M., Boumann, H., Standfest, S., Könneke, M., Sinninghe Damsté, J. S. (2008). Intact membrane lipids of "Candidatus Nitrosopumilus maritimus," a cultivated representative of the cosmopolitan mesophilic group I crenarchaeota. Applied and Environmental Microbiology, 74(8), 2433-2440. https://doi.org/10.1128/AEM.01709-07

Schouten, S., Hopmans, E. C., Schefuβ, E., & Sinninghe Damsté, J. S. (2002). Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? Earth and Planetary Science Letters, 204, 265–274.

Schouten, S., Hopmans, E. C., & Sinninghe Damsté, J. S. (2013). The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. Organic Geochemistry, 54, 19-61. https://doi.org/10.1016/j.orggeochem.2012.09.006

Schouten, S., Huguet, C., Hopmans, E. C., Kienhuis, M. V. M., & Sinninghe Damsté, J. S. (2007). Analytical methodology for TEX86 paleothermometry by high-performance liquid chromatography/atmospheric pressure chemical ionization-mass spectrometry. Analytical Chemistry, 79(7), 2940-2944. https://doi.org/10.1029/2004PA001110.This

Schouten, S., Middelburg, J. J., Hopmans, E. C., & Sinninghe Damsté, J. S. (2010). Fossilization and degradation of intact polar lipids in deep subsurface sediments: A theoretical approach. Geochimica et Cosmochimica Acta, 74(13), 3806-3814. https://doi.org/10.1016/j.gca.2010.03.029

Schouten, S., Pitcher, A., Hopmans, E. C., Villanueva, L., van Bleijswijk, J., & Sinninghe Damsté, J. S. (2012). Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids in the Arabian Sea oxygen minimum zone: I. Selective preservation and degradation in the water column and consequences for the TEX86. Geochimica et Cosmochimica Acta, 98, 228-243. https://doi.org/10.1016/j.gca.2012.05.002

Schouten, S., Villanueva, L., Hopmans, E. C., van der Meer, M. T. J., & Sinninghe Damsté, J. S. (2014). Are Marine Group II Euryarchaeota significant contributors to tetraether lipids in the ocean? Proceedings of the National Academy of Sciences of the United States of America, 111(41), E4285. https://doi.org/10.1073/pnas.1416176111

Schubert, C. J., Coolen, M. J. L., Neretin, L. N., Schippers, A., Abbas, B., Durisch-kaiser, E., … Kuypers, M. M. M. (2006). Aerobic and anaerobic methanotrophs in the Black Sea water column. Environmental Microbiology, 8(10), 1844–1856. https://doi.org/10.1111/j.1462-2920.2006.01079.x

Schubotz, F., Wakeham, S. G., Lipp, J. S., Fredricks, H. F., & 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. Environmental Microbiology, 11(10), 2720–2734. https://doi.org/10.1111/j.1462-2920.2009.01999.x

Sengupta, D., Naik, D., & Reddy, A. R. (2015). Plant aldo-keto reductases (AKRs) as multi-tasking soldiers involved in diverse plant metabolic processes and stress defense: a structure-function undate. Journal of Plant Physiology, 179, 40–55.

Shah, S. R., Mollenhauer, G., Ohkouchi, N., Eglinton, T. I., & Pearson, A. (2008). Origins of archaeal tetraether lipids in sediments: Insights from radiocarbon analysis. Geochimica et Cosmochimica Acta, 72(18), 4577–4594. https://doi.org/10.1016/j.gca.2008.06.021

Shi, Q. (2019). Expression profiling of genes coding for abundant proteins in the alkenone body of marine haptophyte alga Tisochrysis lutea. BMC Microbiology, 19(56), 1–8.

Shi, Q., Araie, H., Bakku, R. K., Fukao, Y., Rakwal, R., Suzuki, I., & Shiraiwa, Y. (2015). Proteomic analysis of lipid body from the alkenone-producing marine haptophyte alga Tisochrysis lutea. Proteomics, 0, 1–14. https://doi.org/10.1002/pmic.201500010

Shimada, H., Nemoto, N., Shida, Y., Oshima, T., & Yamagishi, A. (2008). Effects of pH and temperature on the composition of polar lipids in Thermoplasma acidophilum HO-62. Journal of Bacteriology, 190(15), 5404–5411. https://doi.org/10.1128/JB.00415-08

Shimokawara, M., Nishimura, M., Matsuda, T., Akiyama, N., & Kawai, T. (2010). Bound forms, compositional features, major sources and diagenesis of long chain, alkyl mid-chain diols in Lake Baikal sediments over the past 28,000 years. Organic Geochemistry, 41(8), 753–766. https://doi.org/10.1016/j.orggeochem.2010.05.013

Siedenburg, G., & Jendrossek, D. (2011). Squalene-hopene cyclases. Applied and Environmental Microbiology, 77(12), 3905–3915. https://doi.org/10.1128/AEM.00300-11

Sinninghe Damsté, J. S., Rampen, S., Irene, W., Rijpstra, C., Abbas, B., Muyzer, G., & Schouten, S. (2003). A diatomaceous origin for long-chain diols and mid-chain hydroxy methyl alkanoates widely occurring in quaternary marine sediments: indicators for high-nutrient conditions. Geochimica et Cosmochimica Acta, 67(7), 1339-1348. https://doi.org/10.1016/S0016-7037(02)01225-5

Sinninghe Damsté, J. S., Rijpstra, W. I. C., Hopmans, E. C., den Uijl, M. J., Weijers, J. W. H., & Schouten, S. (2018). The enigmatic structure of the crenarchaeol isomer. Organic Geochemistry, 124, 22-28. https://doi.org/10.1016/j.orggeochem.2018.06.005

Sinninghe Damsté, J. S., Rijpstra, W. I. C., Hopmans, E. C., Jung, M.-Y., Kim, J.-G., Rhee, S.-K., Schleper, C. (2012). Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids of group I.1a and I.1b Thaumarchaeota in soil. Applied and Environmental Microbiology, 78(19), 6866-6874. https://doi.org/10.1128/AEM.01681-12

Sinninghe Damsté, J. S., Rijpstra, W. I. C., Hopmans, E. C., Prahl, F. G., Wakeham, S. G., & Schouten, S. (2002). Distribution of membrane lipids of planktonic crenarchaeota in the Arabian Sea. Applied and Environmental Microbiology, 68(6), 2997-3002. https://doi.org/10.1128/AEM.68.6.2997

Sinninghe Damsté, J. S., Schouten, S., Hopmans, E. C., van Duin, A. C. T., & Geenevasen, J. A. J. (2002). Crenarchaeol: the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. Journal of Lipid Research, 43, 1641-1651. https://doi.org/10.1194/jlr.M200148-JLR200

References

Sintes, E., Bergauer, K., De Corte, D., Yokokawa, T., & Herndl, G. J. (2013). Archaeal amoA gene diversity points to distinct biogeography of ammonia-oxidizing Crenarchaeota in the ocean. Environmental Microbiology, 15(5), 1647–1658. https://doi.org/10.1111/j.1462-2920.2012.02801.x

Sintes, E., De Corte, D., Haberleitner, E., & Herndl, G. J. (2016). Geographic distribution of archaeal ammonia oxidizing ecotypes in the Atlantic Ocean. Frontiers in Microbiology, 7, 1–14. https://doi.org/10.3389/fmicb.2016.00077

Smith, J. M., Damashek, J., Chavez, F. P., & Francis, C. A. (2016). Factors influencing nitrification rates and the abundance and transcriptional activity of ammonia-oxidizing microorganisms in the dark northeast Pacific Ocean. Limnology and Oceanography, 61(2), 596–609. https://doi.org/10.1002/lno.10235

Smith, M., De Deckker, P., Rogers, J., Brocks, J., Hope, J., Schmidt, S., … Schouten, S. (2013). Comparison of Uk37, TEX86 and LDI temperature proxies for reconstruction of south-east Australian ocean temperatures. Organic Geochemistry, 64, 94–104. https://doi.org/10.1016/j.orggeochem.2013.08.015

Sollai, M., Villanueva, L., Hopmans, E. C., Keil, R. G., & Sinninghe Damsté, J. S. (2019). Archaeal sources of intact membrane lipid biomarkers in the oxygen deficient zone of the eastern tropical South Pacific. Frontiers in Marine Science, 10, 1-20. https://doi.org/10.3389/fmicb.2019.00765

Sollai, M., Villanueva, L., Hopmans, E. C., Reichart, G.-J., & 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, 1-19. https://doi.org/10.1111/gbi.12316

Spang, A., Caceres, E. F., & Ettema, T. J. G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science, 357(6351), 1–10. https://doi.org/10.1126/science.aaf3883

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature, 521, 173-185. https://doi.org/10.1038/nature14447

Stadnitskaia, A., Muyzer, G., Abbas, B., Coolen, M. J. L., Hopmans, E. C., Baas, M., Sinninghe Damsté, J. S. (2005). Biomarker and 16S rDNA evidence for anaerobic oxidation of methane and related carbonate precipitation in deep-sea mud volcanoes of the Sorokin Trough, Black Sea. Marine Geology, 217, 67-96. https://doi.org/10.1016/j.margeo.2005.02.023

Stetter, K. O., Fiala, G., Huber, G., Huber, R., & Segerer, a. (1990). Hyperthermophilic microorganisms. FEMS Microbiology Reviews, 75, 117–124. https://doi.org/10.1111/j.1574-6968.1990.tb04089.x

Stewart, F. J., Ulloa, O., & Delong, E. F. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. Environmental Microbiology, 14(1), 23–40. https://doi.org/10.1111/j.1462-2920.2010.02400.x

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., & Schmidt, T. M. (2015). rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Research, 43(D1), D593–D598. https://doi.org/10.1093/nar/gku1201

Strickland, J. D. H., & Parsons, T. R. (1968). A practical handbook of seawater analysis. Ottawa: Queen's Printer.

Sturt, H. F., Summons, R. E., Smith, K., Elvert, M., & 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 Communications in Mass Spectrometry, 18(6), 617–628. https://doi.org/10.1002/rcm.1378

Sugai, A., Sakuma, R., Fukuda, I., Kurosawa, N., & Itoh, Y. H. (1995). The structure of the core polyol of the ether lipids from Sulfolobus acidocaldarius. Lipids, 30(4), 339–344. https://doi.org/10.1007/BF02536042

Sukenik, A., & Wahnon, R. (1990). Biochemical quality of marine unicellular algae with special emphasis on lipid composition. I. Isochrysis galbana. Aquaculture, 97, 61–72.

Sverdrup, H. U., Johnson, M. W., & Flemming, R. H. (1942). The oceans - their physics, chemistry, and general biology. New York: Prentice-Hall.

Swan, B. K., Chaffin, M. D., Martinez-Garcia, M., Morrison, H. G., Field, E. K., Poulton, N. J., … Stepanauskas, R. (2014). Genomic and metabolic diversity of Marine Group I Thaumarchaeota in the mesopelagic of two subtropical gyres. PloS One, 9(4), e95380. https://doi.org/10.1371/journal.pone.0095380

Tamburini, C., Garel, M., Al Ali, B., Mérigot, B., Kriwy, P., Charrière, B., & Budillon, G. (2009). Distribution and activity of Bacteria and Archaea in the different water masses of the Tyrrhenian Sea. Deep-Sea Research Part II: Topical Studies in Oceanography, 56(11-12), 700-712. https://doi.org/10.1016/j.dsr2.2008.07.021

Taylor, K. W. R., Huber, M., Hollis, C. J., Hernandez-Sanchez, M. T., & Pancost, R. D. (2013). Re-evaluating modern and Paleogene GDGT distributions: Implications for SST reconstructions. Global and Planetary Change, 108, 158–174. https://doi.org/10.1016/j.gloplacha.2013.06.011

Team, R. C. (2018). R: A language and environment for statistical computing. Retrieved from https://www.r-project.org/.

Techtmann, S. M., Fortney, J. L., Ayers, K. A., Joyner, D. C., Linley, T. D., Pfiffner, S. M., & Hazen, T. C. (2015). The unique chemistry of Eastern Mediterranean water masses selects for distinct microbial communities by depth. PLoS ONE, 10(3), 1–22. https://doi.org/10.1371/journal.pone.0120605

Tegelaar, E. W., de Leeuw, J. W., Derenne, S., & Largeau, C. (1989). A reappraisal of kerogen formation. Geochimica et Cosmochimica Acta, 53, 3103–3106.

Teira, E., Lebaron, P., Van Aken, H., & Herndl, G. J. (2006). Distribution and activity of Bacteria and Archaea in the deep water masses of the North Atlantic. Limnology and Oceanography, 51(5), 2131–2144. https://doi.org/10.4319/lo.2006.51.5.2131

References

Teske, A. (2013). Marine deep sediment microbial communities. The Prokaryotes: Prokaryotic Communities and Ecophysiology. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30123-0

Teske, A., & Sørensen, K. B. (2008). Uncultured archaea in deep marine subsurface sediments: have we caught them all? The ISME Journal, 2(1), 3-18. https://doi.org/10.1038/ismej.2007.90

Theroux, S., Andrea, W. J. D., Toney, J., Amaral-zettler, L., & Huang, Y. (2010). Phylogenetic diversity and evolutionary relatedness of alkenone-producing haptophyte algae in lakes: Implications for continental paleotemperature reconstructions. Earth and Planetary Science Letters, 300(3–4), 311–320. https://doi.org/10.1016/j.epsl.2010.10.009

Thierstein, H. R., Geitzenauer, K. R., Molfino, B., & Shackleton, N. J. (1977). Global synchroneity of late Quaternary coccolith datum levels: Validation by oxygen isotopes. Geology, 5, 400–404. https://doi.org/10.1130/0091-7613(1977)5<400

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22(22), 4673–4680.

Tierney, J. E., & Tingley, M. P. (2015). A TEX86 surface sediment database and extended Bayesian calibration. Scientific Data, 2, 1–10. https://doi.org/10.1038/sdata.2015.29

Tierney, J. E., & Tingley, M. P. (2018). BAYSPLINE: a new calibration for the alkenone paleothermometer. Paleoceanography, 1–21. https://doi.org/10.1002/2017PA003201

Tripati, A. K., Roberts, C. D., & Eagle, R. A. (2009). Coupling of CO2 and ice sheet stability over major climate transitions of the last 20 million years. Science, 326, 1394–1397. https://doi.org/10.1126/science.1178296

Trommer, G., Siccha, M., van der Meer, M. T. J., Schouten, S., Sinninghe Damsté, J. S., Schulz, H., Kucera, M. (2009). Distribution of Crenarchaeota tetraether membrane lipids in surface sediments from the Red Sea. Organic Geochemistry, 40(6), 724-731. https://doi.org/10.1016/j.orggeochem.2009.03.001

Tsuji, Y., Yamazaki, M., Suzuki, I., & Shiraiwa, Y. (2015). Quantitative analysis of carbon flow into photosynthetic products functioning as carbon storage in the marine coccolithophore, Emiliania huxleyi. Marine Biotechnology, 17, 428–440. https://doi.org/10.1007/s10126-015-9632-1

Turich, C., Freeman, K. H., Bruns, M. A., Conte, M., Jones, A. D., & Wakeham, S. G. (2007). Lipids of marine Archaea: Patterns and provenance in the water-column and sediments. Geochimica et Cosmochimica Acta, 71(13), 3272–3291. https://doi.org/10.1016/j.gca.2007.04.013

Urey, H. C. (1947). The thermodynumic properties of isotopic substances. Journal of the Chemical Society, 562–581.

van der Smissen, J. H., & Rullkötter, J. (1996). 18 . Organofacies variations in sediments from the continental slope and rise of the New Jersey continental margin (Sites 903 and 905). Proceedings of the Ocean Drilling Program, Scientific Results, 150, 329-344.

van Soelen, E. E., Lammers, J. M., Eglinton, T. . I., Sinninghe Damsté, J. S., & Reichart, G.-J. (2014). Organic Geochemistry Unusual C35 to C38 alkenones in mid-Holocene sediments from a restricted estuary (Charlotte Harbor, Florida). Organic Geochemistry, 70, 20-28. https://doi.org/10.1016/j.orggeochem.2014.01.021

Vernal, A. De, Eynaud, F., Henry, M., Hillaire-Marcel, C., Londeix, L., Mangin, S., … Turon, J.-L. (2005). Reconstruction of sea-surface conditions at middle to high latitudes of the Northern Hemisphere during the Last Glacial Maximum (LGM) based on dinoflagellate cyst assemblages. Quaternary Science Reviews, 24, 897–924. https://doi.org/10.1016/j.quascirev.2004.06.014

Versteegh, G. J. M., Bosch, H.-J., & de Leeuw, J. W. (1997). Potential palaeoenvironmental information of C24 to C36 mid-chain diols, keto-ols and mid-chain hydroxy fatty acids; a critical review. Organic Geochemistry, 27(97), 1–13.

Versteegh, Gerard J M, & Blokker, P. (2004). Resistant macromolecules of extant and fossil microalgae. Phycological Research, 52, 325–339.

Vetriani, C., Jannasch, H. W., MacGregor, B. J., Stahl, D. A., & Reysenbach, A.-L. (1999). Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. Applied and Environmental Microbiology, 65(10), 4375–4384.

Villanueva, L., Schouten, S., & Sinninghe Damsté, J. S. (2014). Depth-related distribution of a key gene of the tetraether lipid biosynthetic pathway in marine Thaumarchaeota. Environmental Microbiology, 17(10), 3527-3539. https://doi.org/10.1111/1462-2920.12508

Villanueva, L., Schouten, S., & Sinninghe Damsté, J. S. (2017). Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the "lipid divide." Environmental Microbiology, 19(1), 54-69. https://doi.org/10.1111/1462-2920.13361

Volkman, J. K., Barrett, S. M., & Blackburn, S. I. (1999). Eustigmatophyte microalgae are potential sources of C29 sterols, C22-C28 n-alcohols and C28-C32 n-alkyl diols in freshwater environments. Organic Geochemistry, 30, 307–318.

Volkman, J. K., Barrett, S. M., Blackburn, S. I., & Sikes, E. L. (1995). AIkenones in Gephyrocapsa oceanica: Implications for studies of paleoclimate. Geochimica et Cosmochimica Acta, 59(3), 513–520.

Volkman, J. K., Barrett, S. M., Dunstan, G. A., & Jeffrey, S. W. (1992). C30-C32 alkyl diols and unsaturated alcohols in microalgae of the class Eustigmatophyceae. Organic Geochemistry, 18(1), 131–138.

Volkman, J. K., Eglinton, G., Corner, E. D. S., & Forsberg, T. E. V. (1980). Long-chain alkenes and alkenones in the marine coccolithophorid Emiliania Huxleyi. Phytochemistry, 19, 2619–2622.

Wakeham, S. G., Hopmans, E. C., Schouten, S., & Sinninghe Damsté, J. S. (2004). Archaeal lipids and anaerobic oxidation of methane in euxinic water columns: a comparative study of the Black Sea and Cariaco Basin. Chemical Geology, 205, 427-442. https://doi.org/10.1016/j.chemgeo.2003.12.024

References

Wakeham, S. G., Lee, C., Hedges, J. I., Hernes, P. J., & Peterson, M. L. (1997). Molecular indicators of diagenetic status in marine organic matter. Geochimica et Cosmochimica Acta, 61(24), 5363–5369.

Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., … Noordewier, M. (2003). The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. Proceedings of the National Academy of Sciences of the United States of America, 100(22), 12984–12988. https://doi.org/10.1073/pnas.1735403100

Wegener, G., Krukenberg, V., Ruff, S. E., & Kellermann, M. Y. (2016). Metabolic capabilities of microorganisms involved in and associated with the anaerobic oxidation of methane. Frontiers in Microbiology, 7, 1–16. https://doi.org/10.3389/fmicb.2016.00046

Weiss, G. M., & Bar, M. W. De. (2019). Paleosensitivity of hydrogen isotope ratios of long-chain alkenones to salinity changes at the Chile margin. Paleoceanography and Paleoclimatology, 978–989. https://doi.org/10.1029/2019PA003591

Welander, P. V., Coleman, M. L., Sessions, A. L., Summons, R. E., & Newman, D. K. (2010). Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. Proceedings of the National Academy of Sciences of the United States of America, 107(19), 8537–8542. https://doi.org/10.1073/pnas.0912949107

Welander, P. V., & Summons, R. E. (2012). Discovery, taxonomic distribution, and phenotypic characterization of a gene required for 3-methylhopanoid production. Proceedings of the National Academy of Sciences of the United States of America, 109(32), 12905–12910. https://doi.org/10.1073/pnas.1208255109

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer-Verlag New York. Retrieved from http://ggplot2.org

Winter, C., Kerros, M.-E., & Weinbauer, M. G. (2009). Seasonal changes of bacterial and archaeal communities in the dark ocean: Evidence from the Mediterranean Sea. Limnology and Oceanography, 54(1), 160–170. https://doi.org/10.4319/lo.2009.54.1.0160

Witkowski, C. R., Weijers, J. W. H., Blais, B., Schouten, S., & Sinninghe Damsté, J. S. (2018). Molecular fossils from phytoplankton reveal secular PCO2 trend over the Phanerozoic. Science Advances, 4, 1-7.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proceedings of the National Academy of Sciences of the United States of America, 74(11), 5088–5090. https://doi.org/10.1073/pnas.74.11.5088

Woodward, R. B., & Bloch, K. (1953). The cyclization of squalene in cholesterol synthesis. Journal American Chemical Society, 75, 2023–2024.

Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., van Bleijswijk, J., Timmers, P., Sinninghe Damsté, J. S. (2006). Archaeal nitrification in the ocean. Proceedings of the National Academy of Sciences of the United States of America, 103(33), 12317-12322. https://doi.org/10.1073/pnas.0600756103

Wuchter, Cornelia. (2006). Ecology and membrane lipid distribution of marine Crenarchaeota: Implications for TEX86 paleothermometry.

Wuchter, Cornelia, Schouten, S., Coolen, M. J. L., & Sinninghe Damsté, J. S. (2004). Temperature-dependent variation in the distribution of tetraether membrane lipids of marine Crenarchaeota: Implications for TEX 86 paleothermometry. Paleoceanography, 19(4), 1-10. https://doi.org/10.1029/2004PA001041

Wüst, G. (1961). On the vertical circulation of the Mediterranean Sea. Journal of Geophysical Research, 66(10), 3261-3271. https://doi.org/10.1029/JZ066i010p03261

Xia, X., Guo, W., & Liu, H. (2017). Basin scale variation on the composition and diversity of Archaea in the Pacific Ocean. Frontiers in Microbiology, 8, 1–15. https://doi.org/10.3389/fmicb.2017.02057

Xie, S., Lipp, J. S., Wegener, G., Ferdelman, T. G., & Hinrichs, K.-U. (2013). Turnover of microbial lipids in the deep biosphere and growth of benthic archaeal populations. Proceedings of the National Academy of Sciences of the United States of America, 110(15), 6010–6014. https://doi.org/10.1073/pnas.1218569110

Xie, Sitan, Liu, X., Schubotz, F., Wakeham, S. G., & Hinrichs, K. (2014). Organic Geochemistry Distribution of glycerol ether lipids in the oxygen minimum zone of the Eastern Tropical North Pacific Ocean. Organic Geochemistry, 71, 60–71. https://doi.org/10.1016/j.orggeochem.2014.04.006

Xu, L., Reddy, C. M., Farrington, J. W., Frysinger, G. S., Gaines, R. B., Johnson, C. G., … Eglinton, T. I. (2001). Identification of a novel alkenone in Black Sea sediments. Organic Geochemistry, 32, 633–645.

Yakimov, M. M., Cono, V. La, Smedile, F., DeLuca, T. H., Juárez, S., Ciordia, S., Giuliano, L. (2011). Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production in Tyrrhenian deep waters (Central Mediterranean Sea). The ISME Journal, 5(6), 945-961. https://doi.org/10.1038/ismej.2010.197

Yakimov, M. M., La Cono, V., Denaro, R., D'Auria, G., Decembrini, F., Timmis, K. N., … Giuliano, L. (2007). Primary producing prokaryotic communities of brine, interface and seawater above the halocline of deep anoxic lake L'Atalante, Eastern Mediterranean Sea. The ISME Journal, 1(8), 743–755. https://doi.org/10.1038/ismej.2007.83

Yamamoto, M., Shiraiwa, Y., & Inouye, I. (2000). Physiological responses of lipids in Emiliania huxleyi and Gephyrocapsa oceanica (Haptophyceae) to growth status and their implications for alkenone paleothermometry. Organic Geochemistry, 31, 799–811.

Zachos, J. C., Dickens, G. R., & Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature, 451(7176), 279–283. https://doi.org/10.1038/nature06588

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. Science, 292, 686–693.

References

Zavaterelli, M., & Mellor, G. L. (1995). A numerical study of the Mediterranean Sea circulation. Journal of Physical Oceanography, 25(June), 1384–1414. https://doi.org/10.1175/1520-0485(1995)025<1384

Zhang, Y. G., Pagani, M., & Wang, Z. (2016). Ring Index: A new strategy to evaluate the integrity of TEX86 paleothermometry. Paleoceanography, 31(2), 220–232. https://doi.org/10.1002/2015PA002848

Zhang, Z., Metzger, P., & Sachs, J. P. (2011). Co-occurrence of long chain diols, keto-ols, hydroxy acids and keto acids in recent sediments of Lake El Junco, Galápagos Islands. Organic Geochemistry, 42(7), 823-837. https://doi.org/10.1016/j.orggeochem.2011.04.012

Zheng, Y., Dillon, J. T., Zhang, Y., & Huang, Y. (2016). Discovery of alkenones with variable methylene-interrupted double bonds: implications for the biosynthetic pathway. Journal of Phycology, 1050, 1037–1050. https://doi.org/10.1111/jpy.12461

Zhou, J., Song, X., Zhang, C. Y., Chen, G. F., Lao, Y. M., Jin, H., & Cai, Z. H. (2018). Distribution patterns of microbial community structure along a 7000-mile latitudinal transect from the Mediterranean Sea across the Atlantic Ocean to the Brazilian coastal sea. Microbial Ecology, 76, 1–18. https://doi.org/10.1007/s00248-018-1150-z

Zhu, A., Ibrahim, J. G., & Love, M. I. (2018). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics, 1–26. https://doi.org/10.1093/bioinformatics/bty895

Zhu, C., Wakeham, S. G., Elling, F. J., Basse, A., Mollenhauer, G., Versteegh, G. J. M., … Hinrichs, K.-U. (2016). Stratification of archaeal membrane lipids in the ocean and implications for adaptation and chemotaxonomy of planktonic archaea. Environmental Microbiology, 18(12), 4324–4336. https://doi.org/10.1111/1462-2920.13289

SUMMARY

The Earth is warming due to the anthropogenic (human induced) input of greenhouse gasses into the atmosphere. To better understand and predict the effects in a rapid changing environment we can learn from past climate fluctuations. These changes in past climate systems are recorded in multiple climate archives. Marine sediments are one of the most used and they can provide a long and continuous climate record, up to tens of million years. The marine sediment archive contains, besides terrestrial erosion material, (in)organic compounds that originate from the water column and are deposited within the sediment. Among these organic compounds, lipids derived from marine organisms. These lipids have been used to reconstruct sea surface temperature (SST), which is one of the most important climate variables to reconstruct as the oceans represent the largest surface area on Earth and influence global climate.

The first developed lipid-based SST paleotemperature reconstruction proxy, called the U^{K}_{37} index, is based on the unsaturation degree of alkenones, long-chain unsaturated methyl and ethyl ketone lipids. These alkenones are mainly produced, in the open ocean, by haptophyte algae such as *Emiliania huxleyi* and *Gephyrocapsa oceanica*, however other algal alkenone producers are also known such as *Isochrysis galbana* and *Chrysotila lamellosa*. These alkenones are stored within specific cell organelles called alkenone bodies and it is hypothesized that alkenones have a function in carbon/energy storage to maintain metabolism in dark periods. Because of the wide diversity of alkenones and the often mismatch between sedimentary alkenone compositions with those from algal cultures it is still unclear who the primary producers are in the marine environment. Also physiological factors other than temperature are known to affect the level of alkenone unsaturation, which could potentially bias alkenone-based temperature reconstructions.

Another widely used organic SST reconstruction proxy is the TEX_{86} (Tetra Ether indeX of tetraethers consisting of 86 carbon atoms), based on archaeal membrane lipids, called isoprenoid Glycerol Dialkyl Glycerol Tetraethers (GDGTs). The

composition of GDGTs within the marine environment has been shown to be affected by temperature. These membrane lipids are produced by Archaea, belonging to multiple phyla, such as the Thaumarchaeota, Euryarchaeota and Crenarchaeota. Thaumarchaeota are the most abundant pelagic marine archaea and synthesize GDGTs with 0 to 4 cyclopentane moieties, archaeol and the GDGT crenarchaeol. Crenarchaeol is so far only detected in archaeal cultures belonging to the Thaumarchaeota and contains a cyclohexane moiety in addition to 4 cyclopentane moieties. The GDGTs are predominantly bound to polar head groups in living and intact archaeal cells, thus known as intact polar lipid (IPL-) GDGTs. Although TEX_{86} values correlate with SSTs, several uncertainties have been claimed over the years. For example, it has been suggested that the TEX_{86} proxies are recording subsurface (30-200 meters below sea level) temperatures. Besides, in specific regions, such as the polar oceans and the Red Sea, TEX_{86} temperature reconstructions deviate from the global calibration. Lastly, Archaea other than Thaumarchaeota could potentially synthesize GDGTs therefore affecting the TEX_{86} reconstructions.

The third and most recently developed organic proxy for SST reconstructions is the Long chain Diol Index (LDI) based on Long-Chain alkyl Diols (LCDs). The LCDs are lipids that contain an alkyl chain with alcohol groups at the first carbon position and at a mid-chain carbon position. From culture studies, it is known that phototrophic algae, belonging to the Eustigmatophyceae phylum, are capable of producing these LCDs. However there are also LCD producers outside the Eustigmatophyceae phylum such as certain marine diatoms. Because of the discrepancy between the Eustigmatophyte culture LCD compositions and the composition within marine sediments it is uncertain who the primarily producer(s) of LCDs is/are. It is speculated that LCDs are building blocks of insoluble polymers called algeanans found in the outer cell wall, which could potentially improve the resistance against environmental stress.

Although the above discussed organic proxies have been frequently applied to reconstruct past SST, there is still some uncertainty regarding the biological producers of the lipids that are involved in those proxies. Also their biosynthetic

pathways are largely unknown. The characterization and detection of key genes or enzymes of these synthetic pathways are expected to clarify their biological sources and the effects of environmental factors on the lipid production and composition.

This thesis focus on investigating the biological sources and biosynthetic pathways of lipids used in the above described organic paleotemperature proxies. This was done by applying high resolution methods to determine the diversity of producers and their lipids on samples from marine, lake environments and lab cultures.

In order to determine differences in the distribution of archaeal groups and their potential membrane lipids, (sub)surface sediments located within, below and below the Arabian Sea oxygen minimum zone (OMZ) were investigated. Thaumarchaeota (Marine Group I, MGI) dominated in oxygenated sediments outside of the OMZ coinciding with a high relative abundance of hexose phosphohexose crenarchaeol, a specific biomarker for living Thaumarchaeota. Detection of transcripts of the ammonia monoxygenase (*amo*A) gene suggested that MGI were active in these sediments. On the other hand, in anoxic sediments members of the Miscellaneous Crenarchaeota Group (MCG; Bathyarchaeota) were dominant as expected by their predicted anaerobic metabolism. A specific group of MCG, subgroup MCG-12, coincided with high relative abundances of IPL GDGT-0 with (an) unknown headgroup(s), which could be used as a biomarker of the MCG-12 group in future studies.

Reconstructions of SST based on TEX_{86} in the Mediterranean Sea have previously lead to overestimated values which have been attributed to different populations of MGI inhabiting the water column. To better understand the causes of this bias in the SST-TEX_{86} values, the total archaeal pelagic population together with the core lipid (CL-) and IPL-GDGT composition in three station in the west, middle and east of the Mediterranean Sea were determined. Marine Euryarchaeota Group II and III (MGII and MGIII) dominated the upper pelagic part (0-100 meters deep) of the water column while MGI dominated the subsurface and deeper waters. This shift in the archaeal community coincided with a decrease in the IPL GDGT-0 and an increase of IPL crenarchaeol relative abundances. Also the CL GDGT-2/-3

ratio increased with water depth, coinciding with a higher relative abundance of deep-water MGI thriving at higher temperature and salinity values encountered in specific water masses in the Mediterranean Sea.

There is a general lack of knowledge of the archaeal community and archaeal lipid distribution in marine oxygenated water columns. In this regard, the pelagic archaeal community and lipid content in the North Atlantic Ocean and in coastal North Sea waters were characterized. Members of the MGII and MGIII dominated the upper surface of the water column in both systems, coinciding with lack of detection of archaeal IPLs This suggests that MGII Euryarchaeota are most likely not capable of synthesizing any known archaeal membrane lipids, therefore they are not expected to contribute to the GDGTs used in the TEX_{86} calculation.

To be able to determine patterns of archaea distribution and their lipids in different marine water masses with different oxygen regimes, we compared different dataset including the OMZs of the eastern tropical South Pacific (ETSP) and the Arabian Sea, to the fully oxygenated water column of the North Atlantic Ocean. MGIII Euryarchaeota were predominant in the anoxic waters within the OMZs together with DPANN Archaea. In the Arabian Sea OMZ, archaeol-based IPLs were detected while absent in the ETSP OMZ suggesting that these lipids are either produced by the specific MGIII species present in the Arabian Sea OMZ or alternatively or synthesized only in this OMZ setting.

As pointed out above, the diversity and distribution of LCD-producers in environmental systems is quite unknown. In order to determine the potential LCD-producers in a freshwater system, the diversity and abundance of Eustigmatophyceae was determined in the water column of the Lake Challa, located on the border of Tanzania and Kenya. These results were compared to the LCD composition in the same samples. The maximum abundance of Eustigmatophyceae and LCDs were detected at 9 meters depth, suggesting that this is the preferred niche of LCD-producers in the water column of this lake. Phylogenetic analyses indicated that uncultured groups Eustigmatophyceae were the most likely source for LCDs. An analysis of sedimenting particles revealed seasonal changes in LCD abundance and

composition, suggesting the occurrence of several blooms of LCD producers over the annual cycle or alternatively a change in the LCD composition of the same producer.

Finally, in order to clarify the lipid biosynthetic pathway of long chain alkenones produced by haptophytes and used to determine SST by the U^K_{37} proxy, the haptophyte algae *Emiliania huxleyi* was subjected to a cold shock to induce a rapid increase in the $C_{37:3}/C_{37:2}$ alkenone ratio. Transcriptomic analysis were performed with samples collected during the cold shock experiment to characterize potential biosynthetic genes that would be linked to the change of the $C_{37:3}/C_{37:2}$ alkenone ratio. The upregulation of a gene coding for a homologue of a putative alkenone desaturase in the haptophyte *Tisochrysis lutea* indicates that that specific gene could code for the desaturase responsible for transforming $C_{37:2}$ into $C_{37:3}$ in *E. huxleyi*.

The results described within this thesis show the usefulness of combining high resolution methods for the characterization of archaeal populations and lipids in environmental samples in order to determine not only the distribution and potential physiological strategies of Archaea but also their potential membrane lipids. These lipids can be further used as biomarkers of the presence of these Archaea in present and past environments. Besides, the application of a 18S rRNA gene-based analysis to determine the diversity of Eustigmatophyte algae and determination of LCD, has shed some light on the niche, seasonality and diversity of LCD-producers in a freshwater system. Finally, lipidomics and transcriptomic analyses of culture experiments, where specific conditions known to change the type or abundance or a given lipid biomarker, have been proven to be an excellent experimental set up to identify the lipid biosynthetic pathway of a lipid biomarker of interest.

SAMENVATTING

De aarde warmt op door antropogene (door de mens) input van broeikasgassen in de atmosfeer. Om de effecten van een snel veranderende omgeving beter te begrijpen en kunnen voorspelen kunnen we leren van eerdere klimaat fluctuaties. Deze veranderingen in vroegere klimaatsystemen zitten opgeslagen in meerdere klimaat archieven. Marine sedimenten worden veelvuldig gebruikt en bevatten een lange en ononderbroken geschiedenis van het klimaat, tot tientallen miljoenen jaren. De marine sedimenten archieven bevatten, naast continentaal erosiemateriaal, (in) organische stoffen die afkomstig zijn vanuit de waterkolom en opgeslagen zitten in de marine sedimenten. Onder deze organische stoffen zitten lipiden afkomstig van marine organismen. Deze lipiden worden gebruikt om zee oppervlaktetemperaturen (SST) te reconstrueren, dit is een van de belangrijkste klimaat variabelen om te reconstrueren omdat de oceanen de grootste oppervlakte hebben op aarde en het globale klimaat beinvloeden.

De eerste op lipiden gebaseerde SST reconstructie proxy, genaamd de U_{37}^{K} index, is gebaseerd op de hoeveelheid onverzadigheden in alkenonen, lange-keten onverzadigde methyl en ethyl ketonen lipiden. Deze alkenonen worden voornamelijk geproduceerd, in de open oceaan, door haptophyten algen zoals *Emiliania huxleyi* en *Gephyrocapsa oceanica*, echter andere alkenone producenten zijn bekend zoals *Isochrysis galbana* en *Chrystila lamellosa*. Deze alkenonen zitten opgeslagen in specifieke celorganellen genaamd alkenone lichamen en er is gehypothiseerd dat alkenonen functioneren in de koolstof/energie opslag om metabolisme gaande te houden tijdens donkere periodes. Vanwege de uitgebreide diversiteit van alkenonen en de meerdere mismatches tussen alkonenen composities in de sedimenten en in die in algen culturen is het nog onduidelijke wie de belangrijkste producenten zijn in het marine milieu. Buiten temperatuur zijn er ook andere fysiologische factoren bekend die een effect hebben op de hoeveelheid verzadigingen in alkenonen, dit kan mogelijk een systematische fout veroorzaken in temperatuur reconstructies gebaseerd op alkenonen.

Een andere veel gebruikte organische SST reconstructie proxy is de TEX_{86} (Tetra Ether indeX van tetraethers bestaande uit 86 koolstof atomen), gebaseerd op archaeale membraan lipiden, genaamd isoprenoide Glycerol Dialkyl Glycerol Tetraethers (GDGTs). Het is bewezen dat de compositie van GDGTs binnen het marine milieu beïnvloed wordt door temperatuur. Deze membraanlipiden worden geproduceerd door Archaea, behorend tot meerdere phyla, zoals de Thaumarchaeota, Euryarchaeota en Crenarchaeota. Thaumarchaeota zijn de meest voorkomende pelagische marine Archaea en produceren GDGTs met 0 tot 4 cyclopentaan groepen, archaeol en de GDGT crenarchaeol. Crenarchaeol is tot zo ver alleen gedetecteerd in Thaumarchaeota culturen en bevat een cyclohexane groep, in additie tot 4 cyclopentaan groepen. In levende en intacte Archaea zijn de GDGTs voornamelijk gebonden aan polaire hoofdgroepen, genaamd de intacte polaire lipiden (IPL-) GDGTs. Ondanks dat TEX_{86} waarden correleren met SST, zijn er over de jaren meerdere onduidelijkheden verschenen. Bijvoorbeeld is er gesuggereerd dat dat de TEX_{86} proxies temperaturen onder de oppervlakte (30-200 meter diep) reflecteren. Hiernaast wijken temperaturen gereconstrueerd met TEX_{86} af van de globale kalibratie in specifieke regionen zoals de pool oceanen en de Rode Zee. Tevens, Archaea buiten de Thaumarchaeota groep kunnen in potentie ook GDGTs produceren die op hun beurt de TEX_{86} reconstructies kunnen beïnvloeden.

De derde en meest recent ontwikkelde organische proxy voor SST reconstructies is de lange keten diol index (Long chain Diol Index, LDI) gebaseerd op lange keten alkyl diolen (LCDs). De LCDs zijn lipiden die een alkyl keten bevatten met alcoholgroepen op de eerste koolstof positie en op een koolstof te midden van de keten. Uit cultuurstudies is bekend dat photosynthetiserende algen, binnen de Eustigmatophyceae phylum, deze LCDs kunnen produceren. Hoewel er zijn ook LCD producenten buiten de Eustigmatophyceae phylum bekend zoals bepaalde marine diatomeeen. Vanwege de tegenstrijdigheid tussen de LCD compositie in Eustigmatophyceae culturen en die in marine sedimenten is het nog onduidelijk wie de voornaamste producent(en) is/zijn van LCDs. Het is gesuggereerd dat LCDs bouwstenen zijn van onoplosbare polymeren genaamd algeananen, die gevonden

worden in de buitenste celwand, dit kan mogelijk de weerbaarheid verhogen tegen externe factoren.

Ondanks dat de bovenstaande beschreven organische proxies veelvuldig worden toegepast om SST te reconstrueren is er nog veel onduidelijk met betrekking tot de producenten van de lipiden die gebruikt worden in deze proxies. Ook zijn de biosynthetische processen die de lipiden vormen grotendeels onbekend. De detectie en kenmerking van belangrijke genen of enzymen die onderdeel zijn van deze processen zullen hoogstwaarschijnlijk een hoop ophelderen wat betreft de producenten en welke effecten externe factoren hebben op de lipiden productie en compositie.

Deze thesis focust op het onderzoek naar de biologische bronnen en biosynthetische processen van lipiden die gebruikt worden in bovenstaand beschreven organische paleotemperatuur proxies. Dit is gedaan door het toepassen van hoge resolutie methoden om de diversiteit van de producenten en hun lipiden vast te stellen, in marine, zoetwater omgevingen en culturen in het lab.

Om verschillen vast te stellen in de distributie van archaeale groepen en hun potentiele membraanlipiden, werden oppervlakte sedimenten en sedimenten daaronder in de Arabische Zee bestudeerd. Deze sedimenten bevonden zich in, onder en nog verder beneden de zuurstof minimum zone (OMZ) die zich in de Arabische Zee bevindt. Thaumarchaeota (Marine groep I, MGI) domineerde in zuurstofrijke sedimenten buiten de OMZ, dit viel samen met een hoge relatieve hoeveelheid aan hexose phosphohexose crenarchaeol, een specifieke biomarker voor levende Thaumarchaeota. Detectie van transcripten van het ammonia monoxygenase (*amo*A) gen suggereerde dat MGI actief waren in deze sedimenten. Aan de andere kant, in zuurstofloze sedimenten waren soorten uit de Miscellaneous Crenarchaeota Group (MCG; Bathyarchaeota) dominant zoals verwacht vanwege hun anaerobische metabolisme. Een specifieke groep van de MCG, subgroep MCG-12, werd waargenomen samen met hoge relatieve hoeveelheden van IPL GDGT-0 met (een) onbekende hoofdgroep(en), die wellicht gebruikt kan worden als biomarker voor de MCG-12 groep in toekomstige studies.

Op TEX$_{86}$ gebaseerde SST reconstructies in de Middellandse Zee hebben in het verleden geleid tot overschattende waardes, dit werd toegedicht aan verschillende populaties van MGI in de waterkolom. Om de oorzaken van deze afwijkende waardes beter te begrijpen is de totale pelagische Archaea populatie samen met die van de kern lipiden en IPL-GDGT compositie in drie stations, in het westen, midden en oosten van de Middellandse Zee bepaald. Marine Euryarchaeota groep II en III (MGII en MGIII) domineerde de opperste pelagische zone (0-100 meter diep) van de waterkolom waar de MGI merendeel onder het oppervlakte en in diepere waters voorkwamen. Deze verandering in de Archaea populatie viel samen met een afname van de IPL GDGT-0 en een toename van IPL crenarchaeol relatieve hoeveelheden. Ook de GDGT-2/-3 ratio nam toe met water diepte, dit viel samen met de hogere relatieve hoeveelheden van diep-water MGI die wellicht gedijde bij een hogere temperatuur en zoutgehalte, wat karakteristiek is voor de specifieke water massa's in de Middellandse Zee.

Er is een algemeen gemis aan kennis over de Archaea populatie en archaeale lipiden distributie in marine, zuurstofhoudende waterkolommen. Daarom werd de pelagische Archaea populatie en de aanwezige lipiden in de Noord Atlantische Oceaan en in kust water van de Noordzee bepaald. MGII en MGIII Archaea domineerde de opperste laag van de waterkolom in beide systemen, dit viel samen met het ontbreken van archaeale IPLs. Dit suggereert dat MGII Euryarchaeota zeer waarschijnlijk niet in staat zijn om archaeale membraanlipiden te synthetiseren, daarom wordt niet aangenomen dat ze bijdragen aan de GDGTs die gebruikt worden in de TEX$_{86}$ berekening.

Om patronen in Archaea distributie en hun lipiden te kunnen bepalen in verschillende marine water massa's met verschillende zuurstof gehaltes, hebben we datasets vergeleken van gebieden met een OMZ, zoals de oostelijke tropische Zuid Pacific (ETSP) en de Arabische Zee met die van de Noord Atlantische Oceaan, een waterkolom die volledig zuurstof bevat. MGIII Euryarchaeota waren voornamelijk in de anoxische wateren binnen de OMZs samen met DPANN Archaea. In de Arabische Zee OMZ, waren IPLs met archaeol als basis waargenomen echter waren

ze absent in de ETSP OMZ, dit suggereert dat deze lipiden of geproduceerd worden door specifieke MGIII soorten die aanwezig zijn in de Arabische Zee of dat ze alleen geproduceerd worden in deze specifieke OMZ omstandigheden.

Zoals hier boven beschreven, de diversiteit en distributie van LCD producenten in milieu systemen is vrij onbekend. Om de potentiele LCD producenten te bepalen in een zoetwater systeem, werden de diversiteit en hoeveelheid van Eustigmatophyceae algen bepaald in de waterkolom van het Challa meer, dit meer bevindt zich op de grens tussen Tanzania en Kenya. De resultaten werden vergeleken met de LCD compositie in dezelfde monsters. De grootste hoeveelheid aan Eustigmatophyceae algen en LCDs werden gedetecteerd op 9 meter diepte, dit suggereert dat dit het geprefereerde niche is van LCD producenten in de water kolom van dit meer. Fylogenetische analyses wezen erop dat ongecultiveerde groepen van Eustigmatophyceae algen de meest waarschijnlijk bron waren van deze LCDs. Een analyse van deeltjes die naar de bodem van het meer zakte liet zien dat er seizoen veranderingen waren in de LCD hoeveelheid en compositie, dit suggereert de aanwezigheid van meerdere oplevingen van LCD producenten gedurende het jaar of een verandering in de LCD compositie binnen dezelfde producent.

Tenslotte, om de biosynthetische productie op te helderen van alkenonen, geproduceerd door haptophyten en gebruikt om SST te reconstrueren met de U^K_{37} proxy, werd de haptophyte *Emiliania huxleyi* onderworpen aan een koude shock. Dit om een snelle verhoging te induceren in de $C_{37:3}/C_{37:2}$ alkenone ratio. Het transcriptoom van de monsters die werden genomen tijdens het koude shock experiment werden geanalyseerd, dit om mogelijke genen te vinden die een rol hebben in de verandering van de $C_{37:3}/C_{37:2}$ alkenone ratio. De opregulatie van een gen dat codeert voor een homoloog van een vermeend alkenone desaturase in de haptophyte *Tisochrysis lutea* geeft aan dat dat specifieke gen kan coderen voor de desaturase verantwoordelijk voor de transformatie van $C_{37:2}$ naar $C_{37:3}$ in *E. huxelyi*.

De resultaten beschreven in deze thesis laten de bruikbaarheid zien van het combineren van hoge resolutie methoden voor het karakteriseren van archaeale populaties en lipiden in milieu monsters om niet alleen de distributie en mogelijke

fysiologische strategieën van Archaea te bepalen maar ook om hun mogelijke membraam lipiden in kaart te brengen. Deze lipiden kunnen verder gebruikt worden als biomarkers voor de aanwezigheid van deze Archaea in huidige en vroegere milieus. Hiernaast hebben analyses van het 18S rRNA gen om de diversiteit te bepalen van Eustigmatophyceae algen en de bepaling van LCD'en geleid tot enig inzicht op het niche, seizoensgebondenheid en diversiteit van LCD producenten in een zoet water systeem. Tot slot, lipidome en transcriptoom analyses van cultuur experimenten, waarin specifieke condities het type en de hoeveelheid van een lipide biomarker veranderen, zijn bewezen nuttig om meer te weten over de biosynthese van een bepaalde lipide biomarker.

DANKWOORD

Allereerst wil ik Laura bedanken. Bedankt dat je mij de kans gaf om mijn master stage te lopen bij het NIOZ. De vaartocht met de Pelagia naar en rondom IJsland en het verwerken van de samples op het NIOZ waren prachtige ervaringen. Ik heb onze samenwerking altijd als zeer prettig ervaren en hoefde er dan ook niet lang over na te denken toen er de mogelijkheid was om je PhD student te worden. Hier heb ik geen moment spijt van gehad. Ik heb ontzettend veel van je geleerd, over Archaea en de productie van lipiden, het uitvoeren van experimenten, het maken en geven van presentaties en zoveel meer. Ik heb bewondering hoe je altijd tijd had voor vragen naast al je andere studenten, vele projecten en gezinsuitbreidingen.

Jaap, bedankt voor de begeleiding tijdens mijn master en tijdens mijn PhD, voor je oog voor details en voor je creatieve en briljante inbreng in besprekingen, posters, presentaties en manuscripten. Ik bewonder dat je deur altijd open staat voor vragen over werk, of opmerkingen over Ajax. Ik had me geen betere promotor kunnen wensen en ik hoop ook dat je dit in de toekomst weer kan doen voor anderen.

Stefan, je was oorspronkelijk niet mijn promotor maar ik had me geen betere "invaller" kunnen wensen voor de laatste eindsprint van mijn thesis. Ik had het geluk dat je over alle onderwerpen kennis hebt, over Archaea en GDGTs maar ook over algen experimenten en transcriptomics, je input was van onschatbare waarde. Zonder jou had deze thesis er nu nog niet gelegen. Ik heb vele bewondering hoe je het afgelopen jaar de afdeling hebt geleid en nog steeds zo snel feedback gaf op mijn manuscripten.

Ellen, al mijn kennis over massa spectrometrie, chromatografie en de verschillende IPL mogelijkheden heb ik aan jou te danken. Bedankt voor je al je tijd en het beantwoorden voor al mijn vragen over lipiden.

I also want to thank the members of the dissertation committee, for reading and evaluation of this thesis.

Ook wil ik iedereen bedanken met wie ik heb samengewerkt maar in het bijzonder. Elda, voor het extraheren van DNA en RNA van menig sediment en filter. Sanne,

je was mijn beste student ;) maar ook de beste student die ik kon wensen. Het NIOZ mag blij zijn dat je hier bleef werken. Michel bedankt voor het extraheren en analyseren van GDGTs en alkenonen van vele filters. Marianne bedankt voor al je hulp in het lipiden lab.

Van het moleculaire lab wil ik graag Anneke, Harry, Judith en Maartje bedanken. Voor alle hulp maar bovenal de gezellige tijd. Ik weet zeker dat er in de toekomst nog veel uitmuntend onderzoek gedaan wordt in het nieuwe lab.

Ook wil ik graag mijn kantoorgenoten bedanken. Cindy, voor de gezellige tijd toen ik begon aan mijn PhD. Sergio, we were office mates for years and I could not think of a better office mate, thanks for the nice time at the Gordon conference in the US, all the discussions on living in the Netherlands and football. But also for everything I learned from you about bioinformatics.

I also want to thank the whole BGC/MMB department for the great work environment, all the nice lunches and Friday afternoon drinks. Also the fellow PhD's during all these years, Caitlyn, Christine, Clara, Claudia, Dina, Ella, Fons, Gabriella, Julie, Laura, Lisa, Maaike, Marijke, Martina, Milou, Nadine, Nomikos, Ruth, Saara, Sandra, Sebastian, Sigrid, Sophie, Tim, Yvonne and Zoe. But also all the postdocs, permanent staff and students, Alle, Angelique, Annika, Caglar, Darci, David, Denise, Douwe, Eli, Jessica, Jort, Julia, Kevin, Kim, Marcel, Marta, Michele, Monique, Philip, Sabine, Sebastiaan and Subhash, among many others, thank you.

Thanks to the Potvis people, for all the drinks in the bar, bbq's, games and diners. Zeynep, thanks for all the nice dinners, fun time on the island and in our vegetable garden. Diana and Alejandro, thank you for joining our Potvis community with your own zoo, thanks for hosting amazing breakfasts, lunches and dinners. Alejandro thank you for being my paranymph, a good friend and also for all the help with bioinformatics, you are a true wizard when it comes to coding. Many thanks to all the other nice people in the Potvis, Linda, Lisette, Marten, Sigrid and Siham, Ulrike, among many others. Rineke bedankt dat ik jarenlang een deel van je moestuin mocht gebruiken.

bedankt. Paul, ik ben blij dat we zoveel dezelfde interesses delen, veel succes met het afronden van jouw PhD.

Bedankt Eveline voor alles. Voor de prachtige tijd op Texel, het wonen in Den Burg en op de Potvis. Ik ben blij dat we elkaar nog zo vaak spreken en zien. Ook bedankt aan je moeder.

CURRICULUM VITAE

Marc Arnold Besseling was born on the 5th of October 1987 in Utrecht, the Netherlands. In 2005 he graduated from secondary school (havo) at Leidsche Rijn College in Utrecht. After this he enrolled in the Bachelor Molecular Biology at the Institute for Life Science and Chemistry in Utrecht. During this Bachelor he did his first internship (2008) at the Royal Netherlands Institute of Sea Research (NIOZ) on Texel, working on the identification of bivalve larvae with the use of molecular techniques within the Wadden Sea and North Sea. After this he did his second internship at the Observatoire Océanologique in Banyuls-sur-Mer, France. Working on the influence of Saharan dust on the bacterial community composition within the Mediterranean Sea. Because of his growing interest in teaching and years of teaching scientific courses as a tutor, Marc decided to do a premaster in 2010, followed by the Master in Science Education and Community at Utrecht University. During his Masters he did a scientific internship at the biogeochemistry department of the NIOZ under the supervision of dr. Laura Villanueva, which included a cruise around Iceland (2011). His MSc thesis was titled "Developing molecular tools to estimate the biological sources of long chain diols". The second year of the Master was filled by teaching biology at two secondary schools in Utrecht (Amadeus Lyceum) and in Culemborg (Koningin Wilhelmina College) to become a first degree licensed biology teacher. However, after the two-year Master he started a PhD position on potential sources and genetic controls of lipid biomarkers involved in paleotemperature proxies under the supervision of dr. Laura Villanueva and prof. dr. Jaap S. Sinninghe Damsté.