

Developing molecular tools to estimate the biological sources of long chain diols

“Towards the improvement of a paleotemperature
reconstruction proxy based on long chain diols”

By

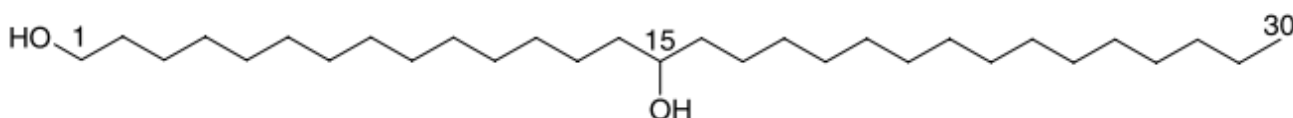
Marc A. Besseling

(BSc of Applied Sciences)

A THESIS

Submitted in partial fulfillment of the requirements for the
degree of

Master of Science



Royal Netherlands Institute for Sea
Research
Dr. Laura Villanueva
Prof. Dr. Jaap S. Sinninghe Damsté

Utrecht University,
Faculty of Science
Prof. Dr. Henk Brinkhuis



Contents

SUMMARY.....	4
SAMENVATTING.....	5
LIST OF ABBREVIATIONS.....	6
1. INTRODUCTION.....	7
1.1. INORGANIC AND ORGANIC PROXIES FOR SEA SURFACE TEMPERATURE RECONSTRUCTION.	7
1.1.1. <i>Long-chain alkenones and the Uk37</i>	8
1.1.2. <i>Thaumarchaeota membrane lipids and TEX₈₆</i>	10
1.2. LONG-CHAIN DIOLS AND THE LCD INDEX	12
1.2.1. <i>The biological source and role of long chain diols</i>	13
1.3. LONG CHAIN DIOLS AS PALEOINDICATORS.....	18
2. AIMS OF THIS STUDY.....	22
3. MATERIALS AND METHODS.....	23
3.1. STUDY SITES AND SAMPLE COLLECTION.....	23
3.1.1. <i>Long chain diol cruise</i>	23
3.1.2. <i>Lake Challa</i>	24
3.2. NUCLEIC ACID-BASED METHODS.....	26
3.2.1. <i>Nucleic acid extraction</i>	26
3.2.2. <i>Amplifying DNA fragments</i>	26
3.2.3. <i>Gradient PCR</i>	27
3.2.4. <i>Denaturing gradient gel electrophoresis</i>	27
3.2.5. <i>Clone libraries</i>	28
3.2.6. <i>Sequencing and phylogenetic analyses</i>	28
3.2.7. <i>Quantitative real-time polymerase chain reaction (qPCR)</i>	30
3.3. LIPID ANALYSES	31
3.3.1. <i>Lipid extractions</i>	31
4. RESULTS.....	33
4.1. DESIGN AND TESTING OF SPECIFIC EUSTIGMATOPHYTE 18S rRNA GENE PRIMERS.....	33
4.1.1. <i>Testing of the developed primers on algal pure cultures</i>	35
4.2. DETECTION OF EUSTIGMATOPHYTES AND LONG CHAIN DIOLS IN ENVIRONMENTAL SAMPLES	37
4.2.1. <i>Long Chain Diol cruise</i>	37
4.2.2. <i>Lake Challa</i>	44
5. DISCUSSION	53
5.2. FUTURE WORK	56
REFERENCES	56
SUPPLEMENTARY FIGURES	64

Summary

In recent years, multiple climate reconstruction proxies based on organic molecules have been developed. These proxies are proven to be of great importance to reconstruct past environmental conditions such as temperature. The Long chain Diol Index (LDI) is a novel paleotemperature proxy, which is based on Long-Chain Diols (LCD). These lipids are detected in marine and lacustrine environments and have been identified in several algal cultures, predominantly belonging to the Eustigmatophyceae, a Stramenopiles phylum. However, the biological source of the LCDs in the environment remains uncertain. In this study, we focused on eustigmatophytes as a possible source of LCDs and surveyed two different environments, with multiple stations around Iceland and Lake Challa, a stratified lake in Eastern Africa. With Quantitative Real Time-Polymerase Chain Reaction (qRT-PCR) in combination with designed primers we were able to quantify the 18S rRNA gene sequences of eustigmatophytes in suspended particulate matter (SPM). These results were compared to the abundance and composition of LCDs in the SPM and also with the LCDs that were obtained from sedimenting particles and surface sediment. We found a similar trend of the LCDs and the eustigmatophyte 18S rRNA gene sequence abundances in the water column. We also detected seasonal distributional changes of LCDs that might indicate different blooms of LCD producers and/or environmental conditions. Also, unique 18S rRNA gene sequences were amplified, which can indicate novel genera, belonging to the Eustigmatophyceae phylum.

Samenvatting

In de afgelopen jaren zijn er meerdere klimaat reconstructie proxies ontwikkeld die gebaseerd zijn op organische moleculen. Deze proxies hebben bewezen van groot belang te zijn om milieu condities zoals temperatuur te kunnen reconstrueren. De lange keten diol index (LDI) is zo'n recente paleotemperatuur reconstructie proxy, gebaseerd op lange keten diolen (LCD). Deze lipiden zijn waargenomen in marine en zoetwater milieus en ook in meerdere algen culturen, waarvan overwegend algen die behoren tot de Eustigmatophyceae, een Stramenopile phylum. Ondanks deze bevindingen blijft de biologische bron van de LCD in de milieus nog onzeker. In dit onderzoek zijn we gericht op eustigmatofyten als de mogelijke bron voor LCD en zijn twee verschillende milieus onderzocht, met meerdere stations rond IJsland en het Challa meer, gelegen in Oost Afrika. Met kwantitatieve polymerase ketting reactie (qRT-PCR) in combinatie met ontwikkelde primers waren we in staat om de 18S rRNA gen sequenties van eustigmatofyten te kwantificeren uit materiaal in suspensie (SPM). Deze resultaten werden vergeleken met de aanwezigheid en compositie van LCD in het SPM maar ook met de LCD in zinkend materiaal en in oppervlakte sediment. We hebben een gelijke trend waargenomen wat betreft de aanwezigheid van LCD en de eustigmatofyte 18S rRNA gen sequenties in de water column. Ook is er een seizoen afhankelijke diversiteit van LCD dat kan wijzen op verschillende opevingen van LCD producenten en/of milieu condities. Tevens zijn er unieke 18S rRNA gen sequenties gedetecteerd die kunnen wijzen op nieuwe genera in het Eustigmatophyceae phylum.

List of abbreviations

$\delta^{18}\text{O}$	delta oxygen 18 isotope (measurement of ratio stable isotopes $^{18}\text{O}/^{16}\text{O}$)
DGGE	Denaturing Gradient Gel Electrophoresis
CTD	water sampler with instruments for Conductivity, Temperature, and Depth
GC	Gas Chromatography
GC/MS	Gas Chromatography/Mass Spectrometry
GDGT	Glycerol Dialkyl Glycerol Tetraether lipids
GFF	Glass Fiber Filter
ITCZ	InterTropical Convergence Zone
kyr	Thousand years
LCD	Long-Chain alkyl Diols
LDI	Long chain Diol Index
Mg/Ca	Magnesium/Calcium ratio
MS	Mass Spectrometry
myr	Million years
PC	PolyCarbonate (filter)
<i>psaA</i> gene	gene coding for photosystem I P700 chlorophyll a apoprotein A1
<i>psbA</i> gene	gene coding for photosystem II P680 reaction center D1 protein
<i>psbC</i> gene	gene coding for photosystem II CP43 chlorophyll apoprotein
qRT-PCR	Quantitative Real Time-Polymerase Chain Reaction
<i>rbcL</i> gene	gene coding for ribulose 1-5 bifosfaat carboxylase Large subunit
rRNA	Ribosomal RiboNucleic Acid
SPM	Suspended Particulate Matter
SST	Sea Surface Temperature
SSU gene	Small SubUnit rRNA gene, can be a 18S rRNA gene (eukaryote) or a 16 rRNA gene (prokaryote).
SWM	South Western Monsoon
TEX ₈₆	Tetra Ether indeX consisting of 86 carbon atoms, an organic geochemical proxy for reconstruction sea surface temperature
TEX ₈₆ ^H	modified Tetra Ether indeX consisting of 86 carbon atoms for sea surface temperature reconstructions between 15 and 30°C
TEX ₈₆ ^L	modified Tetra Ether indeX consisting of 86 carbon atoms for sea surface temperature reconstructions up to 15°C
TLS	TriLaminar Sheath
U ₃₇ ^K	alkenone unsaturation index, a sea surface temperature reconstruction proxy based on alkenones
U ₃₇ ^{K'}	modified alkenone unsaturation index, simplified version of the U ₃₇ ^K , one without the alkenone containing 4 double bonds

1. Introduction

Earth's climate is a continuously changing system and it is currently progressing into a progressively warmer state. It is because of this rapid change that it is of great importance to be able to predict accurate forecasts and be able to anticipate the future climate. To gain knowledge about the behavior of this complex and dynamic system we can learn from our past. In Earth's history many major climatic events occurred, studying these events allows us to understand diverse geochemical processes and feedback mechanisms that shaped Earth's climate. These events are resulting in changes in environmental conditions which are recorded in climate archives. The most commonly used archives are glacial ice, tree rings, corals and sediments. Some of these archives contain climate data from millions of years.

Past environmental conditions have had an impact on all living species in the history of the Earth. These organisms adapt to different conditions and by doing so they change their biochemical pathways and corresponding products. When organisms in the water column die, they become a nutrient source for other life forms, but also a small part of their dead organic material escapes recycling and is transported to the ocean floor and buried into the sediment. During this transportation through the water column and into the sediment, the organic compounds can be altered through a process called *diagenesis*. However, part of the structure of the organic compound can be preserved, so that the original biomolecule from which is derived can still be recognized. Thus, the preserved molecules can be used as biomarker for the presence of the producing organism. As the sediment containing these organic molecules deposits over time, it forms a sedimentary record in the form of organic molecules and other fossils that can be used as biomarkers to reconstruct past environmental conditions. Different proxies are already developed to reconstruct environmental conditions like temperature, salinity, precipitation, etc. based on the biomarkers found in deep sea sediment.

1.1. Inorganic and organic proxies for sea surface temperature reconstruction.

Currently, there is a strong debate on global warming and its impact on human society. Reconstruction of climate conditions in the past enables us to understand climatic processes and their feedback mechanisms. This information can be used to predict the future climate and how this is affected by human activities. This explains the need for proxies able to reconstruct paleotemperature.

One of the most important parameters to reconstruct Earth's past temperature is Sea Surface Temperature (SST). There are multiple possibilities to gain information about past SST, e.g. by application of inorganic temperature proxies. Among inorganic proxies for the reconstruction of SST there are those based on the

isotopic fractionation of oxygen isotopes ($\delta^{18}\text{O}$), isotope composition (Urey 1947) of calcite from planktonic foraminifera (Emiliani 1955), or magnesium to calcium (Mg/Ca) ratio in foraminiferal calcite (Chave 1954; Nürnberg et al., 1996). These inorganic proxies have been widely applied in palaeoclimatic studies (e.g. Garidel-thoron et al., 2005; Zachos et al., 2008). However, there are some uncertainties when applying these proxies for temperature reconstructions. For example, it is known that other factors other than temperature impact the Mg/Ca ratio in calcite shells. In addition, the Mg/Ca ratio of the seawater need to be known in advance for accurate reconstructions (Brown & Elderfield 1996; Ries 2004). Similar considerations have to be made regarding to the $\delta^{18}\text{O}$ isotope composition which is affected by seawater carbonate concentrations (Spero et al., 1997). Another factor is that the main component of the foraminiferal shells is calcium carbonate (CaCO_3), which is affected by dissolution in the deep sea (Peterson 2013). This causes the planktonic foraminiferal calcite shells to dissolve during sedimentation processes (Berger et al., 1982), making reconstructions based on these microfossils impossible.

In general, every proxy is affected by certain uncertainties for example by lack of information on the ecophysiology, preferred niche and seasonality of the producing organism, events of diagenesis of the organic molecule, transportation etc. Studies aimed to clarify these questions are needed to bring a paleoclimatic proxy to maturity.

1.1.1. Long-chain alkenones and the U_{37}^K

The first organic proxy to determine SST was introduced by Brassel et al., (1986) who proposed an index based on the degree of unsaturation of long-chain (C_{37}) ketones (alkenones; Figure 1). These alkenones are predominantly produced by *Emiliana huxleyi* (de Leeuw et al., 1980), an ubiquitous unicellular marine coccolithophorid (Figure 2). Cultures studies demonstrated that the relative abundance of the 37 carbon atom long-chain alkenone with two double bonds increases with growth temperature, whereas the relative abundance of the C_{37} alkenone with three double bonds decreases with an increase in temperature (Prahl & Wakeham 1987) (Figure 3). This observation lead to the formulation of the paleotemperature reconstruction proxy Unsaturated Ketone index (U_{37}^K ; equation [1]):

$$U_{37}^K = \frac{[\text{C}_{37:2}] - [\text{C}_{37:4}]}{[\text{C}_{37:2}] + [\text{C}_{37:3}] + [\text{C}_{37:4}]} \quad [1]$$

Prahl & Wakeham (1987) designed an alternative index called the $U_{37}^{K'}$ (equation [2]; also based on 37 carbon length alkenone molecules):

$$U_{37}^{K'} = \frac{[\text{C}_{37:2}]}{[\text{C}_{37:2}] + [\text{C}_{37:3}]} \quad [2]$$

The major advantage of the $U_{37}^{K'}$ proxy respect to the inorganic-based proxies for SST reconstruction mentioned above is that no additional information is needed about the composition of the seawater. A disadvantage of the $U_{37}^{K'}$ proxy is that *E. huxleyi* appeared around 250-270 kyrs ago (Marlowe et al., 1990; Thierstein et al., 1977), so this limits the use of the proxy from that specific point of time on. However, there are long chain C_{37} alkenones found in older sediments dating back to the Eocene but the SSTs calculated are based on assumptions because there is no information about the biological precursor of *E. huxleyi* (van der Smissen & Rullkötter 1996).

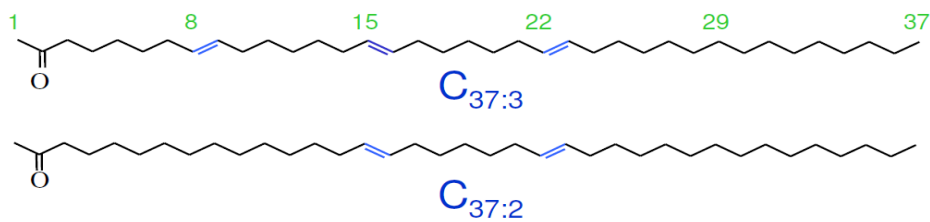


Figure 1. Example of two long-chain unsaturated alkenones. Both molecules consist of 37 carbon atoms and contain two or three unsaturations. $C_{37:2}$ and $C_{37:3}$ are long-chain alkenones identified in sediments and their abundance ratio can be used in SST reconstructions.

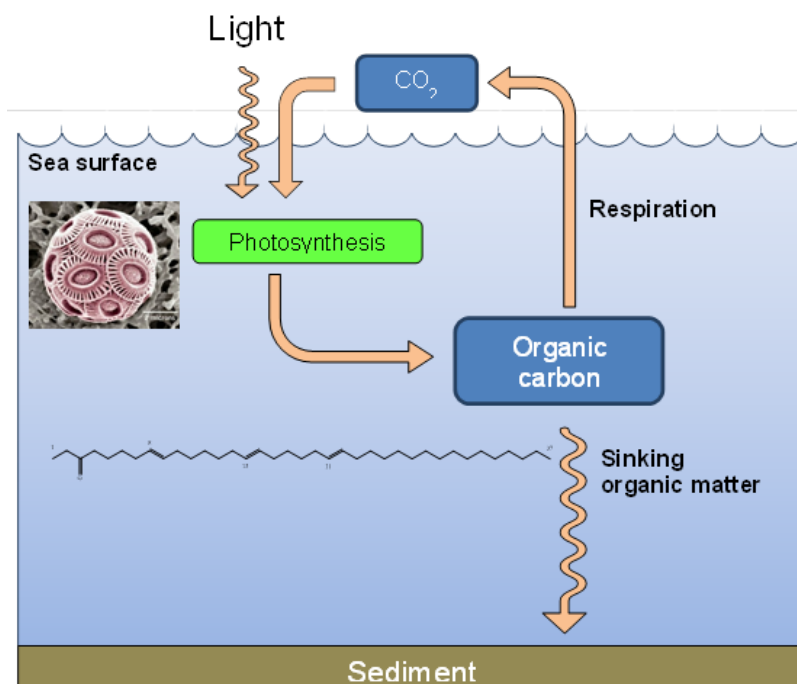


Figure 2. Process of deposition of an organic biomarker. As an example, when the autotrophic marine coccolithophorid algae *E. huxleyi*, synthesizing long-chain alkenones, dies a part of its organic material including the long-chain alkenones sinks to the bottom of the ocean and preserves in the sedimentary record.

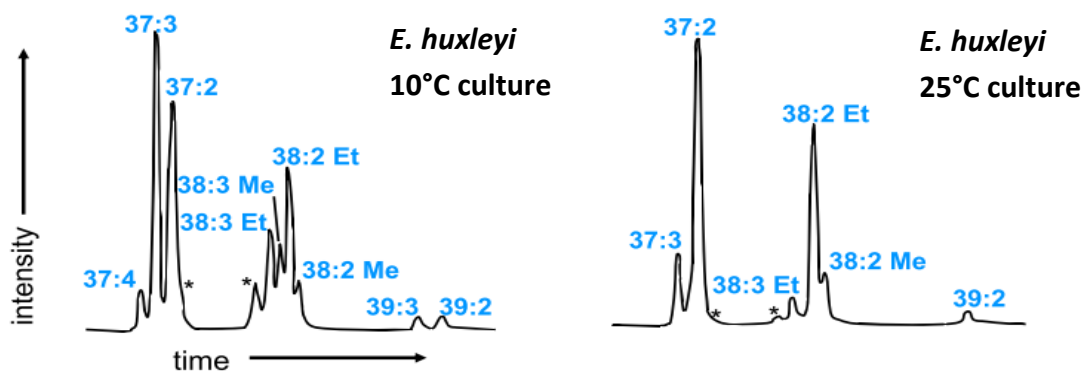


Figure 3. Gas chromatograms of the long-chain, unsaturated alkenone compositions measured in cultures of *E. huxleyi* grown at 10°C and 25°C. The relative abundance of the 37 carbon atom long-chain alkenone with 2 double bonds increases with culture temperature, whereas the relative abundance of the C₃₇ alkenone with 3 double bonds decreases with an increase in temperature. Asterisks mark the di-unsaturated fatty acids (n-C_{36:2}) also produced by this algae as a methyl and ethyl ester. (modified after Prahl & Wakeham 1987)

1.1.2. Thaumarchaeota, membrane lipids and TEX₈₆

Another developed proxy for paleotemperature reconstruction is based on the archaeal membrane ether lipids isoprenoid Glycerol Dialkyl Glycerol Tetraethers (GDGT; Figure 4). GDGT lipids are produced by Archaea of the phylum Thaumarchaeota (recently classified as a new phylum, formerly known as Group I Crenarchaeota, Brochier-Armanet et al., 2008; Spang et al., 2010), as well as other archaeal groups of the euryarchaeota and crenarchaeota phyla (Schouten et al., 2013). Thaumarchaeota are ubiquitous in freshwater ecosystems (e.g. Auguet & Casamayor, 2013), in soil (Leininger et al., 2006), and they are also one of the most important component of the ocean's picoplankton (up to 20%; Karner et al., 2001). Thaumarchaeota synthesize GDGT with 0–4 cyclopentane moieties (Figure 4), and the GDGT crenarchaeol, which contains a cyclohexane moiety in addition to four cyclopentane moieties (Damsté et al., 2002; Schouten et al., 2000). Previous studies have suggested that the GDGT crenarchaeol is exclusively synthesized by Thaumarchaeota (Pitcher et al. 2011). The distribution of thaumarchaeotal GDGTs in the marine environment has also been shown to be affected by temperature, i.e. with increasing temperature there is an increase in the relative abundance GDGT containing cyclopentane moieties (Figure 5) (Schouten et al., 2002; Wuchter et al., 2004; Wuchter et al., 2005). Based on this, the TEX₈₆ paleotemperature proxy was developed and calibrated (e.g. Kim et al., 2010a; Schouten et al., 2002).

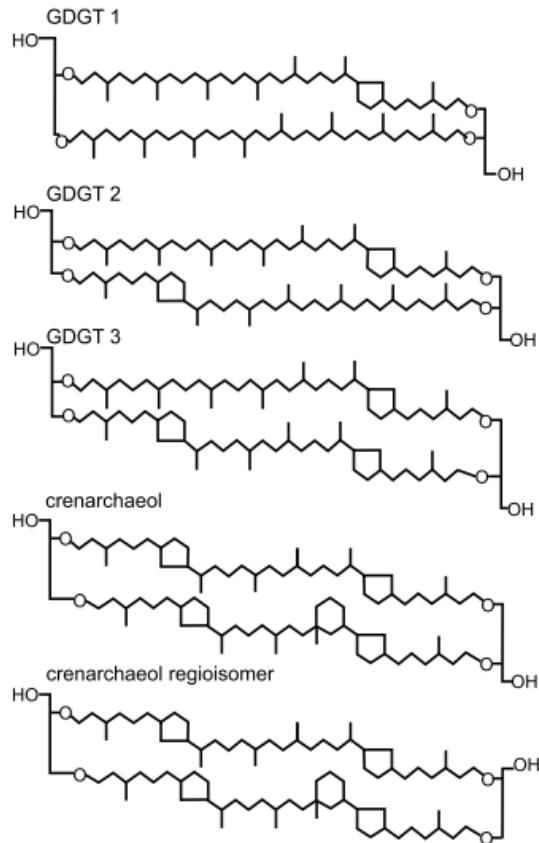


Figure 4. Structures of isoprenoid GDGTs.

TEX_{86} (Tetra Ether index of tetraethers consisting of 86 carbon atoms; equation 3), is defined as the following index:

$$TEX_{86} = \frac{[\text{GDGT} - 2] + [\text{GDGT} - 3] + [\text{Crenarchaeol regio isomer}]}{[\text{GDGT} - 1] + [\text{GDGT} - 2] + [\text{GDGT} - 3] + [\text{Crenarchaeol regio isomer}]} \quad [3]$$

The TEX_{86} has a strong linear relationship with annual mean SST between 5°C and 30°C ($r^2 = 0.935$; Kim et al., 2008). Later on, the proxy was further evaluated by Kim et al. (2010) that proposed a TEX_{86}^L proxy for low temperature regions and a TEX_{86}^H proxy for high temperature regions. A couple of major advantages of the TEX_{86} , compared to the U_{37}^K , is that the TEX_{86} can be applied to older sediments (oldest record is ~160Ma ago, Jenkyns et al., 2012), and can be used for temperature reconstructions above 28°C (e.g. Sluijs et al., 2007).

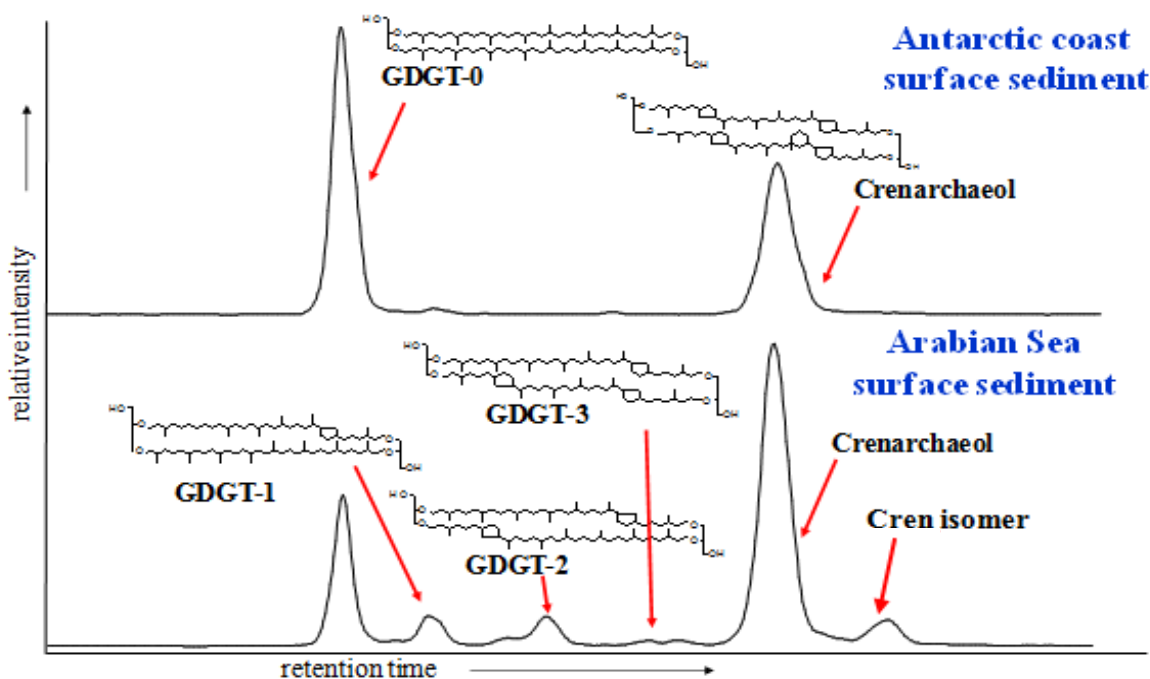


Figure 5. Differences in isoprenoid GDGT distribution in two locations with different temperature regime.

1.2. Long-chain diols and the LCD index

Long-chain alkyl diols (referred from here on as long-chain diols, LCDs or diols) contain an alkyl chain with alcohol groups at C₁ and at a mid-chain carbon position (Figure 6). De Leeuw et al. were the first in identifying LCDs in sediment of the Black Sea (de Leeuw et al., 1981), these encountered LCDs were 1,15-diols with carbon chain lengths of C₃₀, C₃₁ and C₃₂. In addition, the C₃₀ LCD was the most abundant diol detected. Since then, various LCDs have been reported with carbon chain lengths ranging from C₂₄ to C₃₆, and mid-chain alcohol positions ranging from C₁₁ to C₁₉ (Versteegh et al., 1997). Diols have been found in marine sediments (Table 1; Versteegh et al., 1997; Rampen et al., 2007; Willmott et al., 2010; Lopes dos Santos et al., 2012), as well as in lacustrine sediments (Table 1; Romero-Viana et al., 2012; Xu & Jaffé 2009).

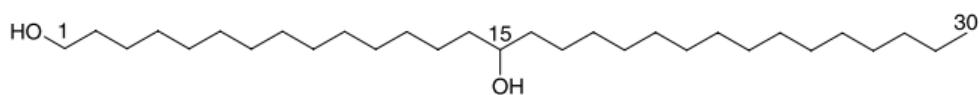


Figure 6. Example of a LCD containing an alkyl chain with an alcohol group at C₁ and at a mid-chain carbon position (in this example at carbon atom number 15, then being a 1,15-diol).

1.2.1. The biological source and role of long chain diols.

Long-chain diols were suggested by de Leeuw et al., (1981) as potential biological marker molecules (de Leeuw et al. 1981). Later, LCDs were detected during a bloom of the cyanobacterium *Phanizomenonflos-aquae* in the Baltic Sea and thus they were believed to be produced by this cyanobacterial group (Morris & Brassell, 1988). However, in cultures of *P. flos-aquae*, grown under different conditions, no LCDs were detected (de Leeuw et al., 1992).

The biological source of LCDs remained unclear until LCDs were identified in cultures of the Eustigmatophyceae algae species *Nannochloropsis oculata*, *N. salina* and *Eustigmatophyceae sp.* (Volkman et al., 1992). These marine eustigmatophytes produce a different composition of LCDs with respect to the LCDs found in marine sediments and in the cyanobacterial bloom. The biodegradation of LCDs after deposition in the sediments has been suggested as a possible reason for that difference (Volkman et al., 1992). However, a major presence of C₃₀ diols in the analyzed sediments while C₃₂ diols were predominant in the eustigmatophyte cultures and the cyanobacterial bloom samples (Versteegh et al. 1997) suggested that other algae were a more likely source of the LCD found in sediments.

In the past decade, several studies has been performed to identify the biological source of LCDs. Gelin et al., (1996, 1997) conducted some studies on the eustigmatophyte *N. salina* and identified more LCDs produced by this algae (Table 1; Gelin et al., 1996; Gelin et al., 1997). Because those LCDs were also found in a lacustrine sediment (e.g. Robinson et al., 1986; Cranwell et al., 1987) it was suggested that the source could also be freshwater algae. So far, three freshwater algae species have been identified as producers of LCDs: *Vischeria punctata*, *Vischeria helvetica* and *Eustigmatos vischeri*, all belonging the Eustigmatophyceae phylum (Volkman et al., 1999).

The Eustigmatophyceae phylum species were previously grouped into the Xantophyceae. However, based on morphological features and chloroplast pigments a new phylum was established (Hibberd & Leedale, 1970, 1971, 1972). The Eustigmatophyceae species are also unique among the chromophyte algae because they lack chlorophyll c (Guillard & Lorenzen, 1972). Later more distinguished features were identified (Andersen 2004; and references therein) and more recently molecular phylogenetic analyses confirmed the unique placing of the Eustigmatophyceae phylum within the photosynthetic stramenopiles group (also known as heterokont algae) (Andersen et al., 1998; Yang et al., 2012). Currently within the Eustigmatophyceae phylum there is one order (Eustigmatales), four families and therein seven genera known (Figure 7).

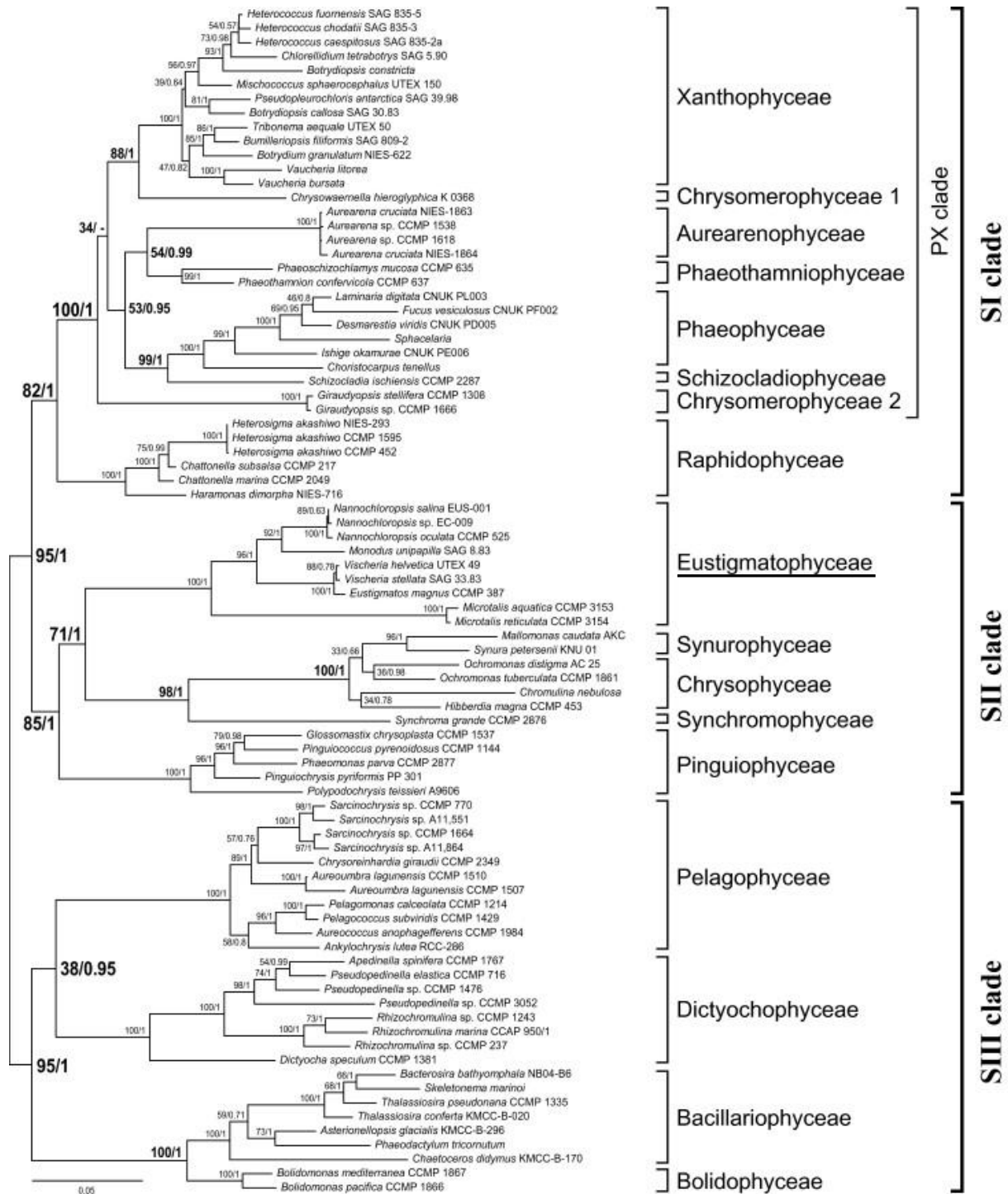


Figure 7. Maximum likelihood tree of the stramenopiles based on five-gene dataset of nuclear SSU rRNA, photosystem I protein (*psaA*), photosystem II proteins (*psbA* and *psbC*), and ribulose 1-5 biphosphate carboxylase Large subunit protein (*rbcl*) (Yang et al. 2012).

As the biological source of LCDs remains unclear there are only speculations of the biochemical role of these compounds. Gelin et al., (1997) postulated that LCDs are building blocks of algaenans (Figure 8), non-hydrolyzable and insoluble biopolymers (Tegelaar et al.,1989). Algaenans have been isolated from numerous Chlorophytes (green algae; Kodner et al.,2009), Eustigmatophytes (Gelin et al.,

1999) and a dinoflagellate (Gelin et al., 1999). In the eustigmatophytes *Nannochloropsis granulata* and *N. oculata*, algaenans are probably forming a trilaminar outer cell wall (TLS, for trilaminar sheath) around the inner polysaccharidic wall of the algae (Figure 9; Gelin et al., 1999). The function of TLS is not fully understood but some hypothesis have suggested that TLS can serve as a protective layer that reinforces cell walls and zygospores (Blokker et al., 1999) and to improve resistance against environmental stress (Allard & Templier, 2000; Kodner et al., 2009; Versteegh & Blokker, 2004).

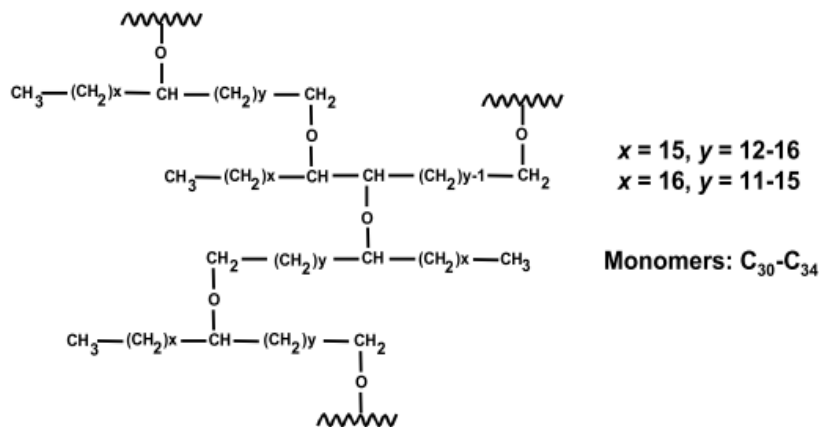


Figure 8. Simplified model that illustrates the algaenan present in the walls of the Eustigmatophyte *Nannochloropsis salina* (modified by Versteegh & Blokker 2004 after Gelin et al., 1997).

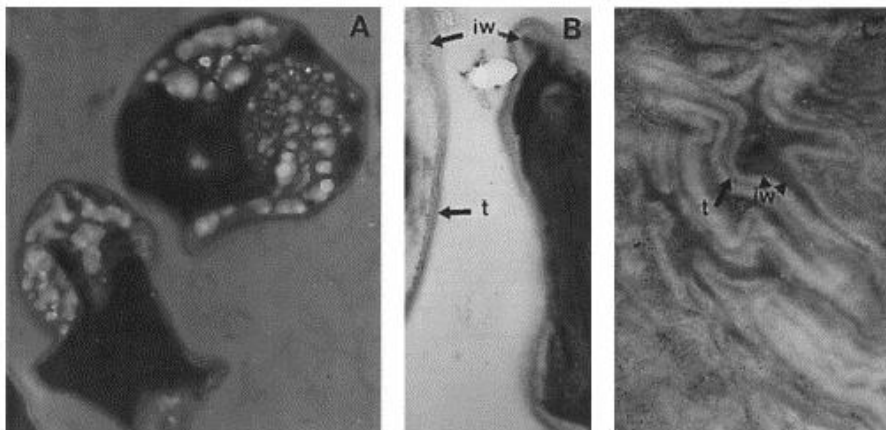


Figure 9. Transmission electron microphotographs of the eustigmatophyte *Nannochloropsis salina*. A ($\times 29,000$): lyophilized cells; B ($\times 72,750$): detail showing the organization of the cell walls (t represents the trilaminar outer cell wall (TLS) and iw represents the inner polysaccharidic wall); C ($\times 72,750$): residue after saponification of the lipid-free biomass, majority of the cell contents is removed however the TLS and inner polysaccharidic walls are still recognizable (modified after Gelin et al., 1999).

Table 1. Occurrence of long-chain diols in sediments reported in the literature.

Location	Environment	Diols											Time scale	Ref		
		C26	C28	C29	C30:1	C30	C31	C32:1	C32	C33	C34	C35			C36	
Geulhemmerberg, The Netherlands	Marine	1,13	1,12 1,13			1,13 1,15			1,15						early Cenozoic (66,5 Myr)	1
Clarkia, Northern Idaho, USA	Lacustrine					1,15			1,15						Miocene (17-20 Myr)	2,3
Guatemala Basin, Pacific Ocean (ODP Site 1241a)	Marine		1,14			1,15									late Miocene (7 Myr)	4
Off shore Northwest Africa	Marine					1,15			1,15						mid Pliocene (3,5 Myr)	5
Off shore Peru (ODP Site 679)	Marine					1,15			1,15						early Pleistocene	5
Off shore Oman (ODP Site 723)	Marine		1,12 1,14			1,14 1,15			1,15						early Pleistocene	5,6
Western flank of the Mid-Atlantic Ridge (IODP site U1313)	Marine		1,13			1,13 1,14 1,15			1,15						early Pleistocene (2,49 Myr)	7
Angola Basin	Marine		1,13 1,14			1,13 1,14 1,15			1,15						mid Pleistocene (147 Kyr)	8
Off shore southern Australia	Marine		1,14			1,14 1,15									mid Pleistocene (130 Kyr)	9
Arabian Sea	Marine		1,14			1,14 1,15									late Pleistocene (90 Kyr)	10
South Atlantic Ocean	Marine		1,13			1,13 1,15									late Pleistocene (43 Kyr)	11
Lake Baikal, Russia	Lacustrine	1,13 1,14 1,15	1,13 1,14 1,15	1,12 1,13		1,13 1,14 1,15	1,14 1,15	1,15	1,15	1,15 1,16 1,17	1,15 1,16 1,17	1,17 1,19			late Pleistocene (28Kyr)	12
Arabian Sea	Marine					1,14 1,15			1,14 1,15						Holocene (15 Kyr)	13
Lake Valencia, Venezuela	Lacustrine								1,15	1,16	1,17	1,18	1,19		Holocene (10 kyr)	14
North-western Antarctic Peninsula	Marine		1,12 1,13 1,14			1,13 1,14									Holocene (8,5 Kyr)	15
Ionian Sea, Eastern Mediterranean	Marine		1,13 1,14 1,15	1,13 1,14	1,14	1,15	1,15		1,15						Holocene (8,2 Kyr)	16
Black Sea	Marine					1,15	1,15		1,15						Holocene (7 Kyr)	17
Lake Malawi, East Africa	Lacustrine					1,15			1,15						Holocene (700 yr)	18

Location	Environment	Diols											Time scale	Ref	
		C26	C28	C29	C30:1	C30	C31	C32:1	C32	C33	C34	C35			C36
Okinawa Trough, China	Marine					1,15			1,15					Recent	19
Lake Priest Pot, UK	Lacustrine					1,13 1,14 1,15			1,15					Recent	20
Lake Kinneret, Israël	Lacustrine					1,15			1,15					Recent	21
Santa Monica Basin, California, USA	Marine					1,15								Recent	22

1. Yamamoto et al., 1996; 2. Huang et al., 1995; 3. Logan & Eglinton 1994; 4. Seki et al., 2012; 5. Ten Haven et al., 1992; 6. Ten Haven & Rullkötter 1991; 7. Naafs et al., 2012b; 8. Versteegh et al., 2000; 9. Lopes dos Santos et al., 2012; 10. Rampen et al., 2008; 11. Rampen et al., 2012; 12. Shimokawara et al., 2010; 13. Schouten et al., 2000; 14. Xu et al., 2007; 15. Willmott et al., 2010; 16. Smith et al., 1983; 17. de Leeuw et al., 1981; 18. Castañeda et al., 2011; 19. Shanchun et al., 1994; 20. Cranwell et al., 1987; 21. Robinson et al., 1986; 22. Pearson et al., 2001

1.3. Long chain diols as paleoindicators

An elaborated study was performed on the algae *Nannochloropsis gaditana* by Méjanelle et al. (2003), which suggested that this algae is a source of a broad range of LCDs. Other producers of LCDs outside the Eustigmatophyceae phylum have been also identified, such as the diatoms *Proboscia indica* and *P. alata*, producers of 1,14-diols (Sinninghe Damsté et al., 2003) (Figure 10). These ubiquitous marine 1,14-diols were suggested to be palaeoindicators for high primary productivity, found in upwelling regions (Sinninghe Damsté et al., 2003). Rampen et al. (2007) confirmed this hypothesis by comparing *Proboscia* cultures with data from sediment traps deployed in the Arabian Sea (Rampen et al., 2007). This resulted in a novel proxy for upwelling intensity during the South Western Monsoon (SWM) (Rampen et al., 2008), defined as the following index:

$$\text{Diol index for SWM upwelling} = \frac{[\text{C}_{28} + \text{C}_{30} \text{ 1,14 - diol}]}{[\text{C}_{28} + \text{C}_{30} \text{ 1,14 - diol}] + [\text{C}_{30} \text{ 1,15 - diol}]} \quad [4]$$

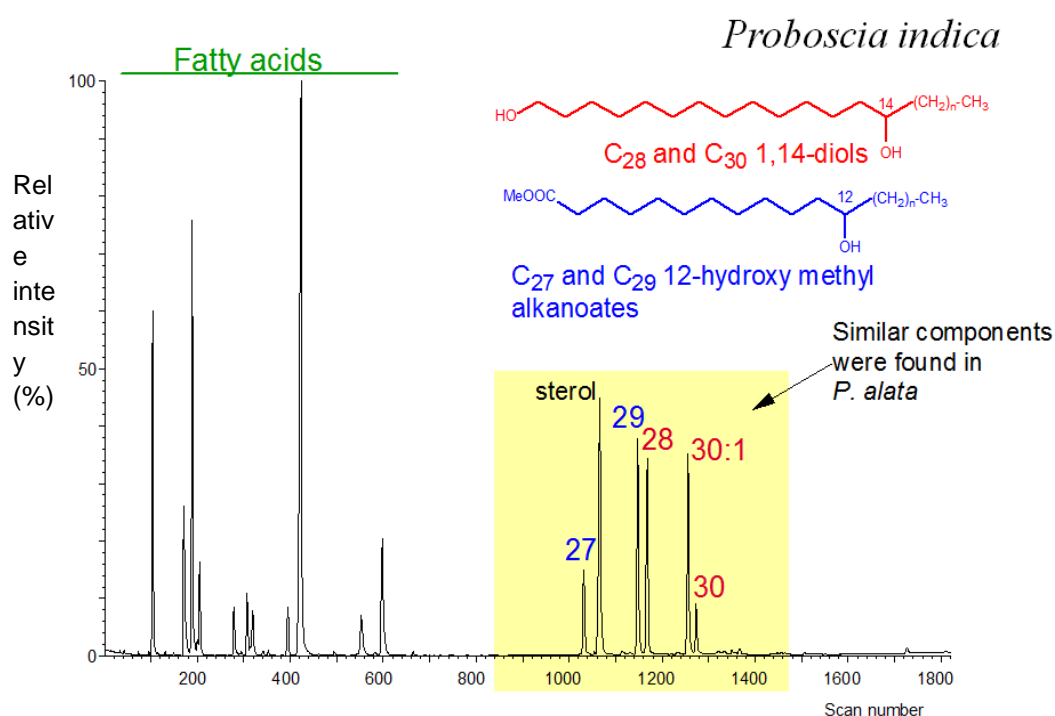


Figure 10. Gas Chromatography (GC)/ Mass Spectrometry (MS) of the total lipid fraction of the diatom *Proboscia indica*. Saturated C_{28} and C_{30} 1,14-diols were detected along with unsaturated C_{30} 1,14-diols and C_{27} and C_{29} 12-hydroxy methyl alkanates. (modified after Sinninghe Damsté et al., 2003).

A strong correlation was shown between 1,14-diols and upwelling in a specific region (Rampen et al., 2008). However, the question remained if 1,14-diols could also be used for palaeotemperature reconstructions. Culture experiments on three different *Proboscia* species (cold and warm water species) was conducted with temperatures ranging from 8°C to 27°C (Rampen et al., 2009). Results showed a positive correlation between temperature and elongation of the chain length, and between temperature and the degree of saturation of 1,14-diols. All species showed similar trends of 1,14-diol modification with an increasing temperature, which suggested that temperature changes induce the differences in the molecule and that it is not species related (Rampen et al., 2009). Rampen et al. (2009) also compared these findings with the LCD composition found in surface sediments from the eastern South Atlantic. However, no correlation between annual mean SST and the degree of unsaturation of the LCDs was observed. Comparing the annual mean SST with the chain length of the 1,14-diols gave a significant but lower correlation than the 1,14-diols derived from the culture experiments in the same temperature range. Rampen et al. (2009) concluded that temperature was probably not the controlling factor of the distribution and biosynthesis of 1,14-diols and that other factors may play a role.

Confirmation of this hypothesis came from the identification of an additional 1,14-diol producer, the algae *Apedinella radians* of the Dictyochophyceae phylum (Rampen et al., 2011). Grown under different growth temperatures *A. radians* produces a different composition of 1,14-diols than the *Proboscia* diatoms. For instance, the *A. radians* grown at low temperature (8°C) predominantly contained the C₃₀ 1,14-diol. However, the 1,14-diols produced by *Proboscia* diatoms grown at similar low temperatures (2°C and 8°C) were dominated by the C₂₈ 1,14-diol (Rampen et al., 2009). This difference in chain length distribution showed that temperature has a different effect on both species when it comes to the production of 1,14-diols. This result has a direct impact on the applicability of a paleotemperature proxy based on LCDs.

Table 2. Producers of long-chain diols as reported in the literature.

Species	Diols																	Ref
	C26	C28:1	C28	C29:1	C29	C30	C31:1	C31	C32:1	C32	C33	C34:1	C34	C35:1	C35	C36:1	C36	
<i>Nannochloropsis oculata</i> ^a						1,13 1,15		1,15	1,15	1,15			1,15					1
<i>Eustigmatophyte CS-246</i> ^a						1,13 1,15		1,15	1,15	1,15			1,15					1
<i>Nannochloropsis salina</i> ^a						1,13 1,15		1,15	1,15	1,15			1,15					1
<i>Nannochloropsis salina</i> ^a					1,13	1,13 1,15		1,15	1,15	1,15	1,17		1,17		1,19	1,19	1,19	2
<i>Nannochloropsis salina</i> ^a			1,13			1,13 1,14 1,15		1,14			1,16		1,17				1,19	3
<i>Vischeria punctata</i> ^b			1,11 1,12 1,13 1,15		1,11 1,12 1,13 1,14	1,13 1,14 1,15		1,13 1,15 1,16		1,13 1,14 1,15 1,17								4
<i>Vischeria helvetica</i> ^b			1,12 1,13 1,15		1,11 1,12 1,13 1,14 1,15	1,13 1,14 1,15		1,12 1,13 1,15 1,16		1,13 1,15 1,17								4
<i>Eustigmatos vischeri</i> ^b			1,13 1,14 1,15		1,13 1,14 1,15	1,13 1,14 1,15		1,12 1,13 1,15 1,16		1,15 1,16 1,17								4
<i>Nannochloropsis gaditana</i> ^a			1,13	1,12	1,12 1,13 1,14 1,16	1,13 1,14 1,15 1,16		1,13 1,14 1,15 1,16	1,14 1,15	1,15 1,16 1,17	1,17 1,18	1,17 1,18	1,15 1,16 1,17 1,18 1,19	1,18 1,19	1,17 1,18 1,19	1,18 1,19 1,20 1,21	1,18 1,19	5
<i>Proboscia indica</i> ^a & <i>Proboscia alata</i> ^a		1,14	1,14			1,14	1,14											6
<i>Proboscia inermis</i>	1,12 1,13 1,14	1,14	1,14			1,13 1,14	1,14											7
<i>Nannochloropsis oculata</i> ^a			1,12			1,13 1,15		1,15	1,15	1,15	1,15 1,16 1,17		1,15		1,19	1,19	1,19	8
<i>Apedinella radians</i> ^a			1,14			1,14				1,14								9

^a = Marine species ^b = Freshwater species

1. Volkman et al. 1992; 2. Gelin et al. 1996; 3. Gelin et al. 1997; 4. Volkman et al. 1999; 5. Méjanelle et al. 2003; 6. Sinninghe Damsté et al. 2003; 7. Rampen et al. 2007; 8. Shimokawara et al. 2010; 9. Rampen et al. 2011

Recently, Rampen et al. (2012) described the occurrence and distribution of 1,13- and 1,15-diols in 164 surface sediments derived from different oceans. Samples were obtained from different water depths (ranging from 20 to 6000 meters) and annual mean SST (ranging from -1.8°C to 28.8°C). When taken the fractional abundances (a specific 1,13- or 1,15-diol relative to the sum of the 1,13- and 1,15-diols), the C₂₈ and C₃₀ 1,13-diols gave a strong negative correlation with the annual mean SST (C₂₈ 1,13-diol r = -0.918, C₃₀ 1,13-diol r = -0.859). The C₃₀ 1,15-diol showed a strong positive correlation with annual mean SST (r = 0.972), while the C₃₂ 1,15-diol fractional abundances did not seem to be controlled by SST (r = 0.189). Based on these results Rampen et al. (2012) proposed the Long chain Diol Index (LDI, equation 5), a novel index for SST reconstructions:

$$\text{LDI} = \frac{[\text{C}_{30} \text{ 1,15-diol}]}{[\text{C}_{28} \text{ 1,13-diol}] + [\text{C}_{30} \text{ 1,13-diol}] + [\text{C}_{30} \text{ 1,15-diol}]} \quad [5]$$

The LDI for the marine surface sediments showed the strongest correlations ($R^2 > 0.95$) with temperatures in the upper 30 meters of the water column. This suggests that the diols are formed in the top layer of the water column and probably by photoautotrophic algae (Rampen et al., 2012).

Although there is a clear correlation between 1,13- and 1,15-diols with SST, no clear reason exists why this is the case. In addition, the biological source of the LCD is still unknown and the recovered distributions of LCDs in marine sediments does not match with the LCD distributions of *Nannochloropsis* algae cultures (Volkman et al., 1992; Versteegh et al., 1997). In these cultures low amounts of the C₂₈ 1,13-diol and high amounts of the C₃₂ 1,15-diol (> 70% of the total amount of LCDs, Volkman et al., 1992; Méjanelle et al., 2003) were detected, whereas lower amounts of C₃₂ 1,15-diol (average of 12%; Rampen et al., 2012) are found in marine sediments. This supports the idea that the main source of 1,13- and 1,15-diols found in marine sediments is not marine eustigmatophytes of the genus *Nannochloropsis*.

The effects of temperature on the composition of long chain diols was proven (Rampen et al., 2012) and therefore LCDs seem to be promising biomarkers to develop temperature proxies. The LCDs are also widely spatial and temporal distributed in the sedimentary record. Application of the LDI in a multi proxy study (Lopes dos Santos et al., 2013) showed that the LDI can be used to provide complementary information to other paleotemperature proxies. However, the biological source(s), the seasonality and niche of the LCD producer(s), as well as the preservation and diagenesis of LCDs, are factors that are still unknown. Further studies aiming to address these questions are key to bring the LDI paleotemperature proxy to maturity.

2. Aims of this study

As indicated above, there is a lack of information of the biological sources of LCDs, as well as the seasonality and niche occupancy of their producers. This reduces the predictive power and complicates the interpretation of the paleotemperature proxy based on LCDs. Considering previous reports mentioned above on the possible biological sources of these organic molecules, members of the Eustigmatophyceae algal family are good candidates to be their major sources in certain environments.

Here, we have developed a genetic-based approach for the estimation of the diversity and abundance of potential LCD-producers by targeting the 18S rRNA gene of the Eustigmatophyceae family. Primers have been developed and tested both for sequencing and quantitative polymerase chain reaction (qPCR) purposes, and compared to previously described genetic tests for the detection of algal groups. This genetic approach has been compared to the LCD diversity and abundance detected in two sample locations: (i) marine suspended particulate matter (SPM) collected around Iceland in the Atlantic Ocean; and (ii) SPM from a stratified tropical lake, Lake Challa (East Africa). By developing and testing this genetic-based method we aim to establish a tool to detect the presence, diversity and abundance of potential-LCD producers in present and past environmental samples. This tool will also be applied to unravel the preferred niche and seasonality of potential LCD-producers, which will in turn be relevant information to improve the predictive nature of the LDI proxy.

3. Materials and methods

3.1. Study sites and sample collection

3.1.1. Long chain diol cruise

The Long-Chain Diols (LCD) cruise with the Royal NIOZ research vessel Pelagia was carried out from 7th to 22nd July 2011. The aim of this cruise was to collect water and sediment samples from the seabed underlying the various water masses present around Iceland. In this manner it is possible to collect, during a relatively short cruise, water and sediments containing organic compounds (in this case LCDs) that are expected to be influenced by different environmental conditions (water temperature, salinity, nutrients, etc).

Samples were obtained from 8 stations (Figure 11 and Table 3) around Iceland. Water samples for DNA analysis were collected from 5 meters depth with the use of Niskin bottles attached on a rosette with Conductivity-Temperature-Depth (CTD) sensors. 20 liters of seawater were filtered in duplicate to collect the SPM through 142 mm diameter, 0.22 μm pore-size polycarbonate (PC) filters (filter code: GTTP; Millipore, Billerica, MA, USA) with a 142 mm diameter, 0.4 μm pore-size PC filter (filter code: HTTP; Millipore) as support (Figure 12). The filtration was done at 4°C in a cooled container. Filters were directly stored at -80°C until further processing.

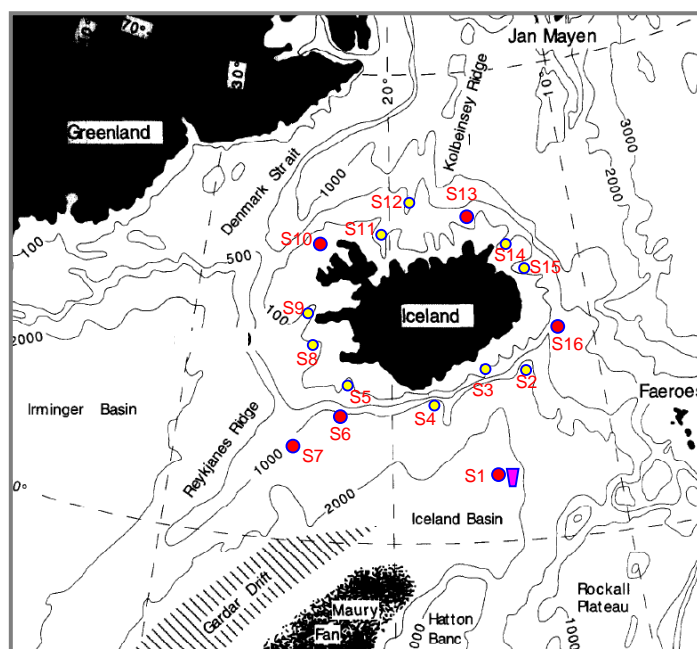


Figure 11. Bathymetry and topographic data around Iceland with sampling stations. The pink circles indicate the main stations and the pink square at S1 is the location of a sediment trap deployment. Map adapted and modified from Lacasse et al.(1996).

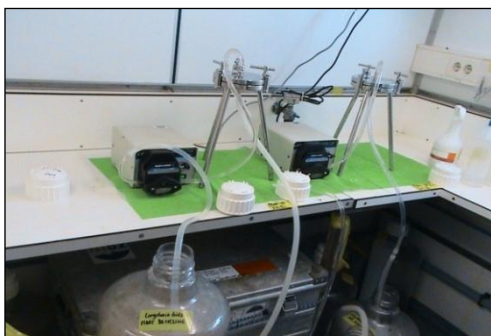


Figure 12. Double filtration system as used on R/V Pelagia during the Long Chain Diol cruise. System was placed in a 4°C cooled container on deck. Photo courtesy of Sebastiaan Rampen.

Samples for lipid analysis were collected with a McLane in situ pump (McLane Laboratories Inc., Falmouth). The pump filtered between 30 and 200 L with a cut-off at a pre-programmed pressure threshold and the SPM was collected on pre-ashed 0.7 µm GF/F filters (Pall Corporation, Washington) and immediately frozen at –80°C.

Table 3. Long-chain diols cruise sampling stations with the date and a description of the work performed on these locations.

Station	Date of sampling	Location	Work performed
1	11-7-2011	62°00'01"N 15°59'98"W	sea surface water filtered
16	14-7-2011	63°59'14"N 12°12'47"W	sea surface water filtered
13	16-7-2011	67°30'08"N 15°04'14"W	sea surface water filtered
11	17-7-2011	66°37'99"N 20°50'05"W	sea surface water filtered
10	18-7-2011	66°40'66"N 24°10'83"W	sea surface water filtered
8	19-7-2011	64°17'58"N 24°08'81"W	sea surface water filtered
5	20-7-2011	63°35'00"N 22°08'65"W	sea surface water filtered
6	21-7-2011	63°14'27"N 22°33'67"W	sea surface water filtered

3.1.2. Lake Challa

Lake Challa (3°19'S, 37°42'E) is a crater lake on the flanks of the Kilimanjaro Mountain (Figure 13). The bottom waters are anoxic and organic material in the sediments is well preserved. A field campaign was undertaken from 29th January to 5th February 2010 to collect depth profiles of suspended particulate matter, nutrients and other physical parameters from the water column of Lake Challa, East Africa (Figure 14). A vertical profile of suspended particulate matter was obtained from lake water samples taken with a Van Dorn water sampler on February 3rd, 2010. Approximately 6 l of water were collected at intervals of 5 m from 0 to 30 m and subsequently every 10 m from 40 to 80 m depth. From 80 m depth, water samples were collected from 82.5, 85 and 89.4 m depth. Samples were filtered through pre-ashed (300°C, 10 h) glass fiber filters (Whatmann GFF; pore size 0.7 µm) within 9 h of their collection and frozen immediately. SPM filters were transported frozen and stored at -20°C (Buckles et al., 2013).

A sediment trap was placed at 35 m depth in the center of Lake Challa. Sediment trap material was collected from January to August 2010. Following its collection the material was left to stand for two days or until all particulates had settled. Excess water was decanted off and samples were subsequently stored frozen. Back in the laboratory the samples were thawed and filtered onto ashed glass fibre filters (Whatmann GF/F; pore size 0.7 μm) prior to extraction.

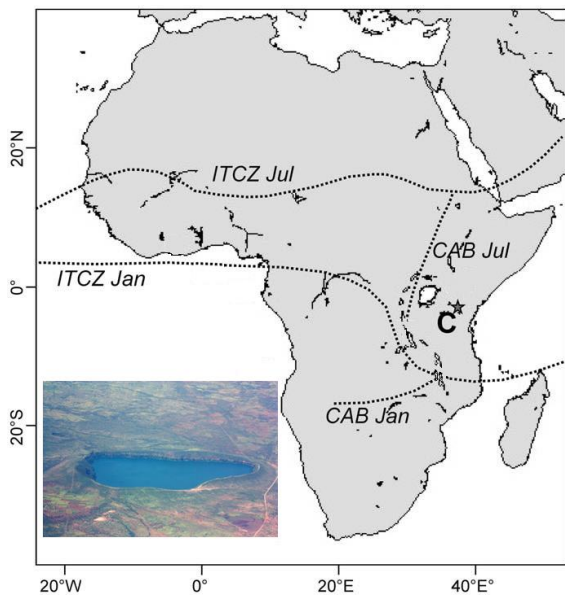


Figure 13. Map of Africa with lake Challa indicated by the star. The dotted lines give the January and July positions of the Intertropical Convergence Zone (ITCZ). Left of lake Challa is Congo air Boundary (CAB) where the Atlantic and Indian Ocean moisture converges. Adapted and modified from Moernaut et al. (2010). The inset shows a photograph of lake Challa from the air.

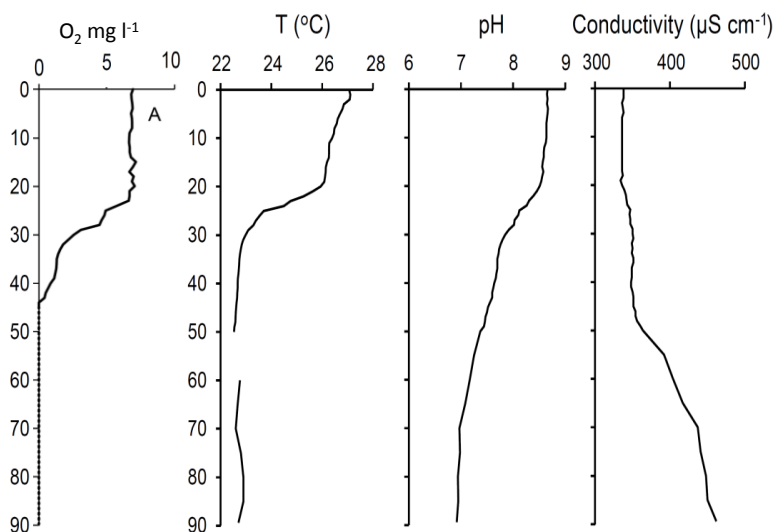


Figure 14. Physicochemical conditions in the Lake Challa water column in the time of sampling. Modified from Buckles et al. (2013).

3.2. Nucleic acid-based methods

3.2.1. Nucleic acid extraction

The glass fiber filters were cut in half while the polycarbonate filters were extracted as a whole for DNA extraction. With a sterile razor blade the filters were cut in smaller pieces and transferred to 50 ml tubes with 1 g glass beads (1:1 w/w 0.1 mm and 0.5 mm glass beads (BioSpec Products Inc., Bartlesville, OK, USA) and 10 ml 6.7% sucrose/TE buffer (sucrose [pH 8.0], TE (50 mM Tris-HCl [pH 8.0], 10 mM EDTA). Tubes were placed on a vortex at maximum speed for 10 minutes. The tubes were centrifuged for 5 min at 3,500 rpm and the supernatant was recovered into a 15 ml tube. Tubes were centrifuged again for 3 min at 3,500 rpm in order to remove excess of the glass beads and the supernatant was transferred to a clean 15 ml tube. Lysozyme (final concentration of 833 µg/ml) was added to the lysate and the tubes were incubated for 30 min at 37°C. Proteinase K (final concentration of 200 µg/ml) and Sodium Dodecyl Sulfate (SDS) (final concentration of 0.5%) were added to the tubes and incubated for 15 min at 65°C. Tubes were cooled down for 5 min to 37°C in a water bath. RNase (final concentration of 85 ng/µl,) was added in order to digest RNA molecules, mixed and incubated for 30 min at 37°C. The lysate was placed on ice for rapid cooling and added an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1 vol/vol/vol), vortexed and centrifuged for 3 min at 35,000rpm. The aqueous phase was recovered into a clean 15 ml tube and an equal volume chloroform/isoamyl alcohol (24:1 vol/vol) was added. Tubes were vortexed and centrifuged for 3 min at 3,500 rpm. The aqueous phase was recovered and mixed with an equal volume of chloroform. The aqueous phase was placed in a clean tube and used for further DNA extraction with the DNeasy® blood and tissue DNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturers' instructions. The extraction was done in duplicate and the extracts were pooled at the end to exclude any possible biases from the extraction procedure. DNA concentration were estimated by using a NanoDrop spectrophotometer (ND-1000, Isogen Life Sciences, De Meern, The Netherlands) and stored in -80°C until further processing.

3.2.2. Amplifying DNA fragments

DNA fragments were amplified by the Polymerase Chain Reaction (PCR) by using different primers. Table 4 gives more details of the primer sets used and the PCR conditions. PCR amplification was performed with a thermal cycler and each sample had a total volume of 50 µl. The PCR mixture containing 5 µM of each deoxyribonucleotide (dNTP's), 6.0 mM MgCl₂, 0.2 µM of each primer, 400 ng/ml Bovine Serum Albumin (BSA), 1.25 U of Taq DNA polymerase (Qiagen) and 1 µl of DNA solution as template. PCR products were purified with the Qiaquick® PCR Purification kit (Qiagen). In cases when PCR byproducts and/or large amounts of primer dimers were observed, the PCR products were loaded onto a 1% (w/vol) agarose gel. The electrophoresis was run for 1 h at 90V with 1X TAE buffer 40 mM Tris, 20 mM acetate, 1 mM EDTA [pH 7.4]) and stained with nucleic acid stain SYBR

Green (4X concentrated in 1X TAE buffer, Invitrogen) for 1 hour. The PCR products were then visualized with a transilluminator (Clare Chemical Research, Dolores, CO, USA) and bands were cut out of the gel with a razor blade. PCR products were cleaned from the agarose and purified with the Qiaquick® gel extraction kit (Qiagen) according to the manufacturers' instructions. Purified products were stored at -20°C until further processing.

Table 4. Primers used for PCR and q-PCR in this study. Degeneracies in primer sequences are colored in red (**R** represent base A and G, **M** base A and C, **Y** base C and T, **S** base C and G, **D** base A, G and T).

Primer	Specific target DNA	Sequence (5' to 3')	PCR conditions	Reference
EUK1A	Eukaryotic 18S rRNA gene	CTG GTT GAT CCT GCC AG	45 cycles	(Díez et al., 2001)
EUK516R	Eukaryotic 18S rRNA gene	ACC AGA CTT GCC CTC C	95°C 1min, 50°C 1min, 72°C 1min	(Díez et al., 2001)
Eust287F	18S rRNA gene	CGA CRA MTC ATT CAA GYT TCT GCC	45 cycles	This study
Eust810R	18S rRNA gene	CCA TGC TAR TGT ATT CAS GGC CT	95°C 1min, 58°C 1min, 72°C 1min	This study
PLA491F	Universal plastid 16S rRNA gene	GAG GAA TAA GCA TCG GCT AA	40 cycles	(Fuller et al., 2006)
OXY1313R	Universal plastid 16S rRNA gene	CTT CAC GTA GGC GAG TTG CAG C	95°C 1min, 58°C 1min, 72°C 1min	(West & Scanlan 1999)

3.2.3. Gradient PCR

Gradient PCR was used to check the optimum annealing temperature for the designed primers. The optimum annealing temperature is the temperature that results in good amplification of the gene from the targeted species. The PCR mixture (25 µl) containing 2.5 µM of each deoxyribonucleotide (dNTP's), 6.0 mM MgCl₂, 0.2 µM of each primer, 400 ng/ml Bovine Serum Albumin (BSA), 1.25 U of Taq DNA polymerase (Qiagen) and 1 µl of DNA solution as template. The cycling conditions for the gradient PCR reaction were the following: 95°C, 5 min; 45 × [95°C, 1 min; annealing temperature, 1 min; 72°C, 1 min]; 72°C, 5 min and 4°C, 15 min. The annealing temperatures for the gradient PCR ranged between 52.3°C and 63°C.

3.2.4. Denaturing gradient gel electrophoresis

Purified PCR samples were measured with the NanoDrop and 25 µl per sample (20 µl of DNA template varying from 50 ng for single band samples to 500 ng for environmental samples and 5 µl of loading buffer) were placed in a 6% gel (w/vol polyacrylamide (37.5:1 acrylamide/bisacrylamide). The gel was casted with denaturing gradients ranging between 20% and 80% (where 100% denaturant contains 7 M urea and 40% formamide). The electrophoresis was run in 1X TAE buffer (40 mM Tris, 20 mM acetate, 1 mM EDTA [pH 7.4]) at 60°C, first for 30 min at

10 V and subsequently for 16 hours at 100 V. After the electrophoresis the gel was stained in the dark for 1.5 hour with the nucleic acid stain SYBR Green (4X concentrated in 1X TAE buffer, Invitrogen). A photo of the gel was made with a UVP AutoChemi Imaging System (UVP, Upland, CA, USA) and a Hamamatsa digital camera (Hamamatsa Corporation, Bridgewater, NJ, USA). DGGE bands were punctured with sterile cut off pipette tips and the tips were placed in Eppendorf tubes with 15 µl of 1X PCR buffer (Qiagen). Samples were stored at 4°C for 2 days and then transferred to -20°C until further processing.

3.2.5. Clone libraries

After purification, 4 µl of each PCR product were mixed with 1 µl of pCR® 2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA) and 1 µl of salt solution from the TOPO TA Cloning® kit (Invitrogen). This is done directly after the purification in order to leave the nucleotide A-overhang on the PCR fragments. The mixture was kept at room temperature for 15 minutes. 3 µl of the ligated mixture was added to One Shot®TOP10 Competent *E. coli* cells (Invitrogen). The *E. coli* cells with the vectors were kept on ice for 30 minutes, heat shocked for 30 seconds at 42°C and then kept on ice for 2 minutes. Under sterile conditions was 250 µl of Super Optimal broth with Catabolite repression (SOC) medium (Invitrogen) added to the *E. coli* cells and shaken at 250 rpm for 1 h at 37°C. The cells were placed in volumes of 10 µl, 25 µl, 50 µl and 100 µl on Luria Broth (LB) agar plates supplemented with ampicillin (50 µg/ml) and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; 40 µl/ml). Plates were incubated overnight at 37°C, followed by 1 h at 4°C for a better discrimination of the colony colors. Isolated colonies were inoculated in 2-3 ml LB medium supplemented with ampicillin in 96-well grow plates or BD Falcon™ tubes (BD, Franklin Lakes, NJ, USA). Plates and tubes were incubated at 37°C overnight at 250 rpm. Cultures were centrifuged for 5 minutes at 2,500 rpm and the supernatant was removed. Plasmid extraction was done according to the manufacturers' instructions (for single tube samples with the Quickclean 5M Miniprep (Genscript, Piscataway, NJ, USA), for the 96 well plates the Quickclean 96-well plasmid Miniprep kit (Genscript) was used.

3.2.6. Sequencing and phylogenetic analyses

DNA fragments were sequenced by Macrogen (Amsterdam, the Netherlands). Raw sequence data was edited with the bioinformatic program BioEdit Sequence Alignment Editor, version 7.1.9. (Hall, 1999). With this program the raw sequences were checked on quality by looking at the electropherogram. Good parts of a sequence has clear single peaks with no background signal (Figure 15). The end of a sequence tend to decrease in quality and with an electropherogram can be selected what part of the sequences has to be discarded. It is recommended to crop and discard before further analysis the end of the sequence reads which usually show background signal (Figure 16).

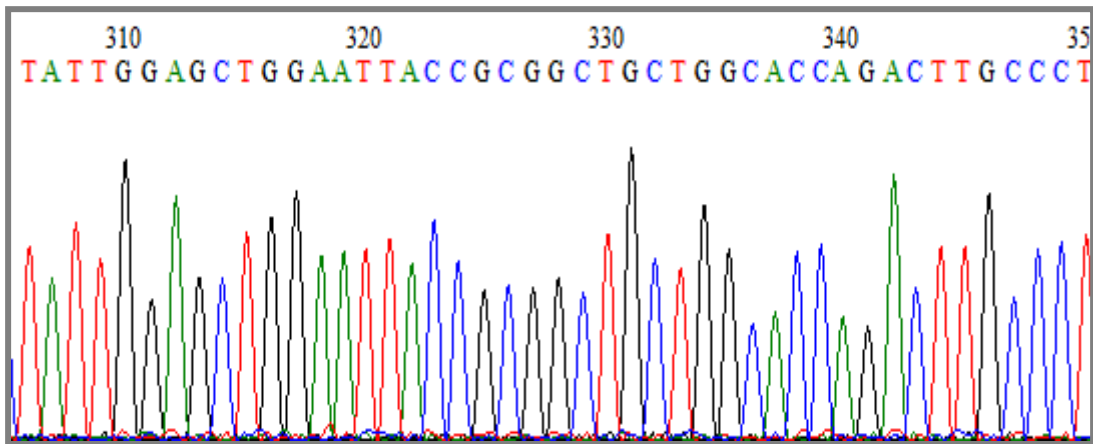


Figure 15. An electropherogram of good quality sequence derived with a clone library from SPM of lake Challa.

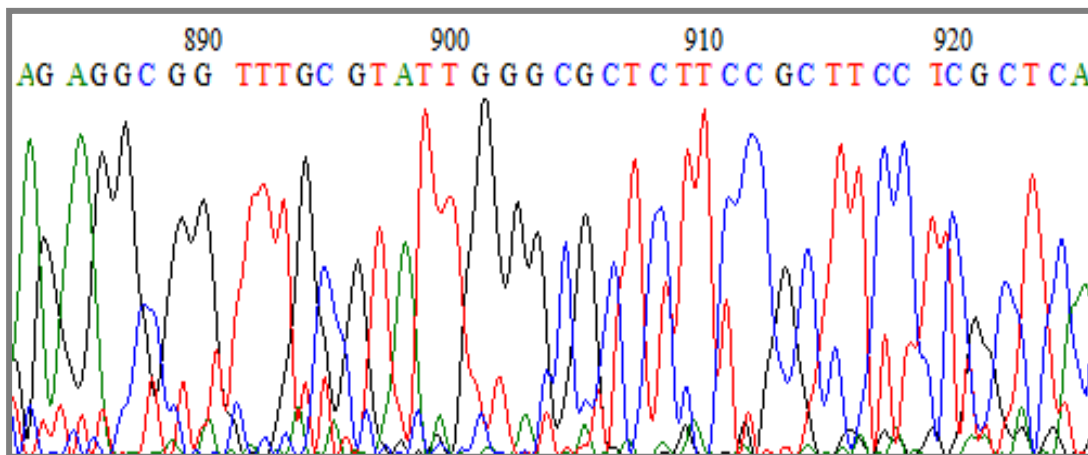


Figure 16. An electropherogram of poor quality sequence derived with a clone library from SPM of lake Challa.

Edited gene sequences were also aligned by ClustalW (Thompson et al., 1994) in BioEdit, to ensure that all the sequences are in the right direction and to delete sequence parts derived from the plasmids and primers used in e.g. the clone library method. Alignments were exported as fasta files to the bioinformatic program MEGA (Molecular Evolutionary Genetics Analysis version 5, Tamura et al., 2011). Phylogenetic trees were constructed with MEGA following the maximum likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993) with a bootstrap test of 1,000 replicates.

Sequences were analysed for the presence of chimeras using the Bellerophon tool at the GreenGenes website (<http://greengenes.lbl.gov/>). The phylogenetic affiliation of the partial eustigmatophyte 18S rDNA gene sequences were also compared to release 111 of the Silva SSU Ref database (<http://www.arb-silva.de/>; Quast et al., 2013) using the ARB software package (Ludwig et al., 2004). The partial sequences generated in this study were added to the reference tree supplied by the Silva database using the ARB Parsimony tool. In addition, the 18S rRNA gene sequence alignment including the Eustigmatales group and the sequences obtained

in our analysis was exported from ARB and used to construct a maximum likelihood tree in PhyML version 3.0 (Guindon et al., 2010) using the GTR model plus gamma distribution. Branch support was calculated with the approximate likelihood ratio test (aLRT).

3.2.7. Quantitative real-time polymerase chain reaction (qPCR)

Quantitative PCR reactions were performed with a CFX96™ real-time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA). The reaction mixture (25 µl) contained 1 U of Pico Maxx high fidelity DNA polymerase (Stratagene, Agilent Technologies, Santa Clara, CA, USA) 2.5 µl of 10x Pico Maxx PCR buffer, 2.5 µl 2.5 mM of each dNTP, 0.5 µl BSA (20 mg/ml), 0.02 pmol/µl of primers, 10,000 times diluted SYBR Green® (Invitrogen) (optimized concentration), 0.5 µl 50 mM of MgCl₂ and ultrapure sterile water. All reactions were performed in iCycler iQ™ 96-well plates (Bio-Rad, Hercules CA, US) with optical tape (Bio-Rad). One µL of diluted environmental DNA was added to 24 µl of mix in each well. Specificity of the reaction was tested with a gradient melting temperature assay. The cycling conditions for the qPCR reaction were the following: 95°C, 4 min; 45 × [95°C, 30 s; 61°C, 40 s; 72°C, 30 s]; final extension 80°C, 25 s. Specificity for qPCR reaction was tested on agarose gel electrophoresis and with a melting curve analysis (50°C–95°C; with a read every 0.5°C held for 1 s between each read) in order to identify unspecific PCR products such as primer dimers or fragments with unexpected fragment lengths. Samples were positive with a Ct (threshold cycle at which a significant increase in fluorescence occurs) lower than the Negative Template Control (NTC) and a primer melting curve similar to those of the standard.

A positive PCR product of the eustigmatophyte 18S rRNA gene reaction (sample 1355SAB000 E3, recovered from SPM collected at 0.5 meter depth) was picked for amplification as a qPCR standard. The fragment 1355SAB000 E3 was amplified in five separate reactions in order to increase DNA abundance of the product and also to avoid biases of a single PCR reaction. The PCR reactions (5 × 50 µl) each contained 2.5 µM of each deoxyribonucleotide (dNTP's), 6.0 mM MgCl₂, 0.2 µM of each primer, 400 ng/ml Bovine Serum Albumin (BSA), 1.25 U of Taq DNA polymerase (Qiagen) and 1 µl of DNA solution as template. The primers used for the PCR amplification were M13F (5'-GTTTCCCAGTCACGAC-3') and M13R (5'-TCACACAGGAAACAGCTATGAC-3'). PCR conditions were the following; 95°C, 5 min; 30 × [95°C, 1 min; 50°C, 1 min; 72°C, 1 min]; 72°C, 7 min and 4°C, 15 min. PCR product was run on a 1% (w/vol) agarose gel 1 h at 90 V. The PCR product band was visualized and cut out of the gel and purified twice with the Qiaquick® PCR Purification kit (Qiagen). DNA concentration was estimated by using a NanoDrop spectrophotometer. The amplified fragment had a length of 737 bp (535 bp 1355SAB000 E3 fragment plus 202 bp of plasmid and primer fragment). The final concentration of the mixture contained 3.22×10^{10} copies per µl. The standard curves were used in logarithmic dilutions ranging from 3.22×10^7 to 3.22×10^0

molecules and were analyzed in triplicate. Gene copy numbers were estimated from the standard curves.

3.3. Lipid analyses

3.3.1. Lipid extractions

Base hydrolysis (saponification) of residue

A quarter of a 142 mm diameter 0.7 µm pore-size glass fiber filter (LCD cruise samples) or half of a 47 mm 0.7 µm GF/F filter (Lake Challa SPM profile) was cut into small pieces. The filter pieces were placed in a 100 ml glass boiling flask with a round bottom with 20 ml of 6% KOH in methanol (MeOH). Glass flask was placed under a glass condenser for 1 hour refluxing at 115°C. The mixture was cooled down to room temperature and 10 ml of HCl/MeOH was added. A glass funnel was filled with approximately 10 ml of double distilled H₂O and a drop of Dichloromethane (DCM). The mixture from the glass flask was added to the funnel. 20 ml of MeOH/H₂O were mixed with the remaining filter parts in the flask and also added to the funnel. Same procedure was followed using 20 ml of MeOH and 3 times with 20 ml of DCM. The funnel was mixed carefully and pH was checked (pH around 2 a 3). The DCM layer was recovered from the funnel in a 100 ml glass flask with round bottom and placed in a Rotavapor (type R-114, Büchi, Flawil, Swiss) under low pressure until most of the liquid was vaporized. Approximately 15 ml of DCM were added to the funnel and the DCM was recovered in the same glass flask used in the previous step and vaporized. The last step was performed three times. The recovered liquid after the last evaporation was transferred to a glass vial and dried under nitrogen flow.

Acid hydrolysis of extracts

The small filter parts left from the base hydrolysis remaining in the glass flask were further treated by acid hydrolysis. 15 ml of HCl/MeOH were added to the filter parts and 2 ml to the vial from the base hydrolysis. Both were placed under a glass condenser for 3 hours of refluxing at 150°C. After cooling down to room temperature 30 ml of KOH/MeOH were added to the filter parts and 2 ml to the vial. The supernatant from the filters was placed in a 100 ml glass funnel with approximately 10 ml of double distilled water and 3 ml of DCM. The flask was rinsed with 20 ml of MeOH and the supernatant was placed in the funnel. Filter pieces were further washed with 20 ml of DCM and the mixture was transferred again into the funnel and vaporized with the Rotavapor (type R-114, Büchi) under low atmospheric pressure. This last step was performed three times. The base and the acid DCM layers were combined in a glass vial and dried under nitrogen gas.

Separation of fractions

The Total Lipid Extract (TLE) was separated into a nonpolar and a ketones fraction with a column made of a glass pipette with aluminum oxide. First the column was

rinsed with Hexane/DCM (9:1 vol/vol). Hexane/DCM (9:1 vol/vol) was added to the TLE extract and the solution was added to the column. The column was rinsed with DCM/MeOH (1:1 vol/vol) and the eluate was recovered as the nonpolar fraction. Then DCM/MeOH was placed into the now seemingly empty TLE vial, carefully shaken and the solution was also added to the column. The flow thru of the column was captured in a glass vial as the ketones fraction. Vials with the two fractions were dried under nitrogen gas and stored in the fridge until further processing.

Silylation

Samples were silylated by adding 15 μl of BSTFA (N,O-Bis(trimethylsilyl)-trifluoroacetamide) and 15 μl of pyridine and heating for 20 min at 60°C. 50 μl of ethyl acetate was added.

Gas Chromatography–Mass Spectrometry (GC-MS)

Compound identification of the long chain diols was conducted using a Thermo Finnigan Trace Ultra GC connected to Thermofinnigan DSQ MS operated at 70 eV, with a mass range m/z 50–800 and 3 scans s^{-1} . Used is a silica column (25 x 0.32 mm) and coated with CP Sil-5 (film thickness of 0.12 μm). The temperature program initiated at 70 °C, increased first at a rate of 20 °C min^{-1} to 130 °C, and next at a rate of 4 °C min^{-1} to the final temperature of 320 °C, which was held for 10 min. The relative abundances of diols were measured using the same GC/MS operated in single ion mode, monitoring ions of m/z 299, 313 and 327, with a dwell time of 100 ms and ionization energy of 70 eV (according to Rampen et al., 2008).

4. Results

4.1. Design and testing of specific eustigmatophyte 18S rRNA gene primers

To design specific primers for the Eustigmatophyceae phylum, all 18S rRNA gene sequences (124 sequences) of the phylum were downloaded from the NCBI database. The selected 18S rRNA gene sequences comprised representative members of the different genera of the Eustigmatophyceae phylum (Table 5). Other 18S rRNA gene sequences from selected algae belonging to the Stramenopiles (Bacillariophyceae, Dictyochophyceae, Chrysophyceae, Pinguiphyceae, Palegophyceae, Rapidophyceae, Phaeophyceae, and Xanthophyceae; total of 80) were also included in the analysis. Sequences aligned with the ClustalW method.

Table 5. Taxonomy of the Phylum Eustigmatophyceae. Uncultured members are not given. * *Trachydiscus minutus* is not yet assigned to a family within the phylum (Přibyl et al., 2012)

Phylum	Eustigmatophyceae
Order	Eustigmatales
Family	Eustigmataceae
Genus	Eustigmatos
Species	<i>Eustigmatos magnus</i>
Species	<i>Eustigmatos polyphem</i>
Species	<i>Eustigmatos vischeri</i>
Genus	Pseudostaurastrum
Species	<i>Pseudostaurastrum enorme</i>
Species	<i>Pseudostaurastrum limneticum</i>
Genus	Vischeria
Species	<i>Vischeria helvetica</i>
Species	<i>Vischeria punctata</i>
Species	<i>Vischeria stellata</i>
Family	Loboceae
Genus	Pseudotetraedriella
Species	<i>Pseudotetraedriella kamillae</i>
Family	Monodopsidaceae
Genus	Nannochloropsis
Species	<i>Nannochloropsis gaditana</i>
Species	<i>Nannochloropsis granulata</i>
Species	<i>Nannochloropsis limnetica</i>
Species	<i>Nannochloropsis maritime</i>
Species	<i>Nannochloropsis oceanic</i>
Species	<i>Nannochloropsis oculata</i>
Species	<i>Nannochloropsis salina</i>
Family	Pseudocharaciopsidaceae
Genus	Ellipsoidion
Species	<i>Ellipsoidion sp. UTEX B SNO113</i>
Genus	Pseudocharaciopsis
Species	<i>Pseudocharaciopsis minuta</i>
Genus	Trachydiscus
Species*	<i>Trachydiscus minutus</i>

18S rRNA gene fragment areas unique in eustigmatophytes were detected and the primer pair Eust287F/810R (numbers correspond to the nucleotide position of the 18S rRNA gene sequence of the eustigmatophyte *Eustigmatos vischeri*, accession number JX274590.1) was designed manually and checked for secondary structures and % G+C in the Sigma DNA calculator (<http://www.sigma-genosys.com/calc/DNACalc.asp>). The amplicon size was approximately 523 bp. The specificity of the primer pair was investigated by nucleotide blast in NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the nucleotide collection and using the highly similar sequence option (megablast). Eust287F/810R primers covered members of the Eustigmatophyceae phylum as indicated in Tables 6 and 7.

Table 6. Sequence of the Eust287F primer designed in this study.

Eust287F (5'-3')	CGAC R AMTCATTCAAG Y TTCTGCC
<i>Nannochloropsis oceanica</i> (JQ315726.1)	CGACGAATCATTCAAGTTTCTGCC
<i>Eustigmatos vischeri</i> (JX274590.1)	CGACGAATCATTCAAGTTTCTGCC
<i>Vischeria helvetica</i> (AB731568.1)	CGACGAATCATTCAAGTTTCTGCC
<i>Stephanopyxis turris</i> (AB430590.1)	CGACAAATCATTCAAGTTTCTGCC
<i>Monodus subterranea</i> (FJ896225.1)	CGACGAATCATTCAAGTTTCTGCC
<i>Incisomonas marina</i> (JN848814.1)	CGACAAATCATTCAAGTTTCTGCC
<i>Gonostomum</i> sp. (JX946277.1)	CGACAAATCATTCAAGTTTCTGCC
<i>Synura echinulata</i> (GU325513.1)	CGAT G CATCATTCAAGTTTCTGCC

*Accession numbers in parentheses. Degeneracies in primer sequences are colored in red (**R** represent base A and G, **M** base A and C, **Y** base C and T). Mismatches from the primer with the gene sequences are shown in blue.

The Eust287F primer matched 100% with the 18S rRNA gene sequences of *Stephanopyxis* sp. (diatoms, Bacillariophyta; Coscinodiscophyceae), and *Monodus subterranea* (Xanthophyceae phylum; Mischooccales; Pleurochloridaceae; Monodus), as well as with some ciliate species (*Incisomonas* and *Gonostomum*), while e.g. 18S rRNA gene sequence of *Synura* sp. (Chrysophyceae phylum; Synurophyceae; Synurales; Mallomonadaceae) had two mismatches with the designed primer (Table 7).

Table 7. Sequence of the Eust810R primer designed in this study.

Eust810R (3'-5')	AGGCC S TGAATACA Y TAGCATGG
<i>Nannochloropsis oceanica</i> (JQ315726.1)	AGGCCCTGAATACATTAGCATGG
<i>Eustigmatos vischeri</i> (JX274590.1)	AGGCCCTGAATACATTAGCATGG
<i>Vischeria helvetica</i> (AB731568.1)	AGGCCCTGAATACATTAGCATGG
<i>Monodus subterranea</i> (FJ896225.1)	AGGCCCTGAATACATTAGCATGG
<i>Synura echinulata</i> (GU325513.1)	AGGCCCTGAATACATTAGCATGG

*Eust810R sequence is reverse complemented (3'-5') for comparison purposes. Accession numbers are in parentheses. Degeneracies in primer sequences are colored in red (**S** represent base C and G, **Y** base C and T).

Primer Eust810R presented possible amplification with members of the Eustigmatophyte phylum but also matched to 18S rRNA gene sequences of the genus *Monodus subterranea* (Xanthophyceae phylum; Mischococcales) and *Synura* sp. (Chrysophyceae phylum; Synurophyceae).

4.1.1. Testing of the developed primers on algal pure cultures.

The novel primer pair Eust287F/810R was tested with a gradient PCR (PCR procedure with different annealing temperatures) using genomic DNA derived from single algal cultures as a template (Table 8). Genomic DNA of four different single algal cultures (covering four different phyla of stramenopiles) were used as a template in a gradient PCR with annealing temperatures ranging from 52.2 °C to 63 °C. This resulted in a positive amplification for the species *Nannochloropsis oceanica* (Eustigmatophyceae) and also for *Hibberdia magna* (Chrysophyceae) (Figure 17).

Table 8. Outcome of PCR amplification with Eust287F/810R primers on genomic DNA of different algal cultures.

Species	Annealing temperature (°C)									
	54.2	55.4	56.6	57.8	59.0	60.1	60.8	61.0	62.0	63.0
<i>Nannochloropsis oceanica</i> (Eustigmatophyceae)	+	+	+	+	+	+	+	+	+	+
<i>Hibberdia magna</i> (Chrysophyceae)	+	+	+	+	+	+/-	+/-	+/-	-	-
<i>Apedinella Radians</i> (Dictyochophyceae)	-	-	-	-	-	-	-	-	n/a	n/a
<i>Pelagomonas calceolate</i> (Pelagophyceae)	-	-	-	-	-	-	-	-	n/a	n/a

+ indicates a positive amplification as shown with the gel electrophoresis. +/- indicates that amplification occurred but shows a fainted amplification with the gel electrophoresis. – indicates that no amplification occurred. n/a = not available.

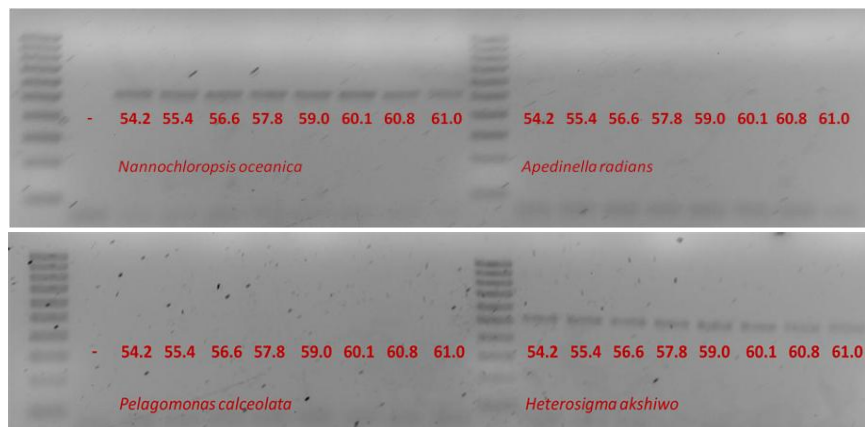


Figure 17. Gel electrophoresis of the DNA fragments obtained with the gradient PCR.

Other genomic DNA obtained from single algal cultures were tested with the novel primer pair, which resulted in positive amplification in the following algae; *Laminarionema elsbetiae* (Phaeophyceae), *Desmarestia aculea* (Phaeophyceae), *Phaeomonas parva* (Pinguiphyceae), *Glossomastix chrysoplasta* (Pinguiphyceae), *Heterosigma akshiwo* (Raphidophyceae) and *Fibrocapsa janponica* (Raphidophyceae) (Table 9, 10, 11). In addition, Two algal species *Phaeothamnion confervicola* (Phaeothamniaceae) and *Tribonema viride* (Tribonemataceae) from two different stramenopile phyla also gave a positive amplification with the novel primers at lower annealing temperatures (ranging from 52.4°C to 57.7°C) (Table 11).

Table 9. Outcome of PCR amplification with Eust287F/810R primers on genomic DNA of different algal cultures.

Species (phylum)	Annealing temperature (°C)	
	62.0	63.0
<i>Chromulina ochromonoides</i> (Chrysophyceae)	-	-
<i>Pseudopedinella elástica</i> (Dictyophyceae)	-	-
<i>Sarcinochrysis marina</i> (Pelagophyceae)	-	-
<i>Aureoumbra lagunensis</i> (Pelagophyceae)	-	-
<i>Pylaiella littoralis</i> (Phaeophyceae)	-	-
<i>Laminarionema elsbetiae</i> (Phaeophyceae)	+/-	+/-
<i>Arthrocladia villosa</i> (Phaeophyceae)	-	-
<i>Desmarestia aculea</i> (Phaeophyceae)	+/-	+/-
<i>Phaeomonas parva</i> (Pinguiphyceae)	+/-	+/-
<i>Pinguicoccus pyrenoidosus</i> (Pinguiphyceae)	-	-
<i>Glossomastix chrysoplasta</i> (Pinguiphyceae)	+/-	+/-
<i>Heterosigma akshiwo</i> (Raphidophyceae)	+/-	+/-
<i>Fibrocapsa janponica</i> (Raphidophyceae)	+/-	+/-

+ indicates a positive amplification as shown with the gel electrophoresis. +/- indicates that amplification occurred but shows a faint amplification with the gel electrophoresis. – indicates that no amplification occurred.

Table 10. Outcome of PCR amplification with Eust287F/810R primers on genomic DNA of different algal cultures.

Species (phylum)	Annealing temperature (°C)			
	59.4	61.1	62.1	62.5
<i>Phaeothamnion confervicola</i> (Phaeothamniaceae)	+	+	+	+
<i>Tribonema viride</i> (Tribonemataceae)	+	+	+	+
<i>Vaucheria terrestris</i> (Vaucheriaceae)	-	-	-	-
<i>Florenciella parvula</i> (Dictyochophyceae)	-	-	-	-
<i>Hibberdia magna</i> (Chrysophyceae)	+	+	+	+
<i>Pinguicoccus pyrenoidosus</i> (Pinguiophyceae)	+	+	+	+
<i>Pylaiella littoralis</i> (Phaeophyceae)	+	+	+	+
<i>Pelagomonas calceolate</i> (Pelagophyceae)	+/-	+/-	-	-
<i>Chromulina ochromonoides</i> (Chrysophyceae)	-	-	-	-
<i>Pseudopedinella elástica</i> (Dictyochophyceae)	-	-	-	-
<i>Apedinella Radians</i> (Dictyochophyceae)	-	-	-	-

+ indicates a positive amplification as shown with the gel electrophoresis. +/- indicates that amplification occurred but shows a faint amplification with the gel electrophoresis. – indicates that no amplification occurred.

Table 11. Outcome of PCR amplification with Eust287F/810R primers on genomic DNA of different algal cultures.

Species (phylum)	Annealing temperature (°C)			
	52.3	54.1	55.9	57.7
<i>Phaeothamnion confervicola</i> (Phaeothamniaceae)	+	+	+	+
<i>Tribonema viride</i> (Tribonemataceae)	+	+	+	+

+ indicates a positive amplification as shown with the gel electrophoresis. +/- indicates that amplification occurred but shows a faint amplification with the gel electrophoresis. – indicates no amplification occurred.

4.2. Detection of Eustigmatophytes and Long chain diols in environmental samples

4.2.1. Long Chain Diol cruise

General eukaryotic diversity

The eukaryotic diversity in SPM of 4 Icelandic stations (Figure 18; all samples were taken at 5 meters depth except for S1.2) was investigated by DGGE by using the picoeukaryotes 18S rRNA gene primers Euk1A/Euk516R. The number of detected bands ranged between a single band (Station 13) and 10 bands (Station 1.1d). The Euk1A/Euk516R PCR amplifications on genomic DNA extracted from the different stations were performed in duplicate and revealed an identical profile on the DGGE gel (Figure 18; duplicates are indicated by a **d** behind the station number).

Selected bands (indicated by numbers in Figure 18) were reamplified and sequenced. The affiliation of the sequences is listed in Table 12.

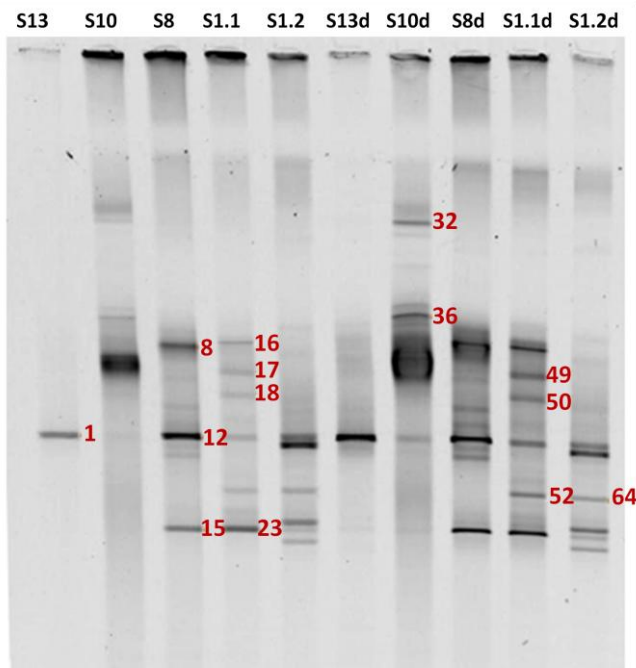


Figure 18. DGGE fingerprint of 18S rRNA gene fragments from the SPM at different stations around Iceland. Numbers indicate the bands that were selected for sequencing. (S1.1 is a seawater surface sample, S1.2 is an intermediate seawater sample; 40 meters depth)

Table 12. Affiliation of DGGE bands recovered from SPM collected at station 1 during the LCD cruise. Sequences were obtained by using the Euk1A/Euk516R primer pair.

Band number	Size (bp)	%	Closest relative and affiliation	Accession number
1	517	100	Uncultured eukaryote clone SHAU581	JQ222973.1
8	501	93	Uncultured Chlorophyta clone RS.12f.10m.00184	KC583052.1
12	517	100	Uncultured eukaryote clone SHAU581	JQ222973.1
15	516	100	Uncultured marine eukaryote clone CNCIII51_10	HM581760.1
16	511	90	Uncultured marine eukaryote clone CNCIII51_5	HM581780.1
17	514	99	Uncultured marine eukaryote isolate JPeuk-71	AY145110.1
18	514	93	Uncultured marine eukaryote clone SGUH1088.FRAG.MO.5m	JX841385.1
23	516	100	Uncultured marine eukaryote clone CNCIII51_10	HM581760.1
32	508	94	Uncultured eukaryote	AB695501.1
36	513	100	<i>Alexandrium fundyense</i> strain IMR_S_182008 (Alveolata)	JF521629.1
49	514	99	Uncultured marine eukaryote isolate JPeuk-71	AY145110.1
50	513	87	Uncultured marine eukaryote clone SGUH433.FRAG.MO.5m	JX841649.1
52	515	99	Uncultured metazoan clone RS.12f.10m.00175	KC582964.1
64	515	100	Uncultured metazoan clone RS.12f.10m.00175	KC582964.1

%. Percentage of homology with the closest relative.

Almost all of the sequences were closely related to uncultured eukaryotes (13 out of the 14 fragments for which sequence data were obtained). One sequence displayed a high homology with 18S rRNA gene sequences of the dinoflagellate *Alexandrium fundyense* (band 36).

In order to further estimate the algal diversity in station 1, we applied plastid 16S rRNA gene primers PLA491F and OXY1313R (Table 4) to the DNA extracted from this station. Forty-three sequences had good quality and were analyzed by BLAST. Twenty-three sequences were closely related to the coccolithophore *Emiliana huxleyi* (Figure 19). Other sequences were homologues to uncultured bacteria (14), cyanobacteria (4) and Haptophytes (2). No sequences were closely related to *Florenciella*, *Dictyocha* or other Stramenopiles.

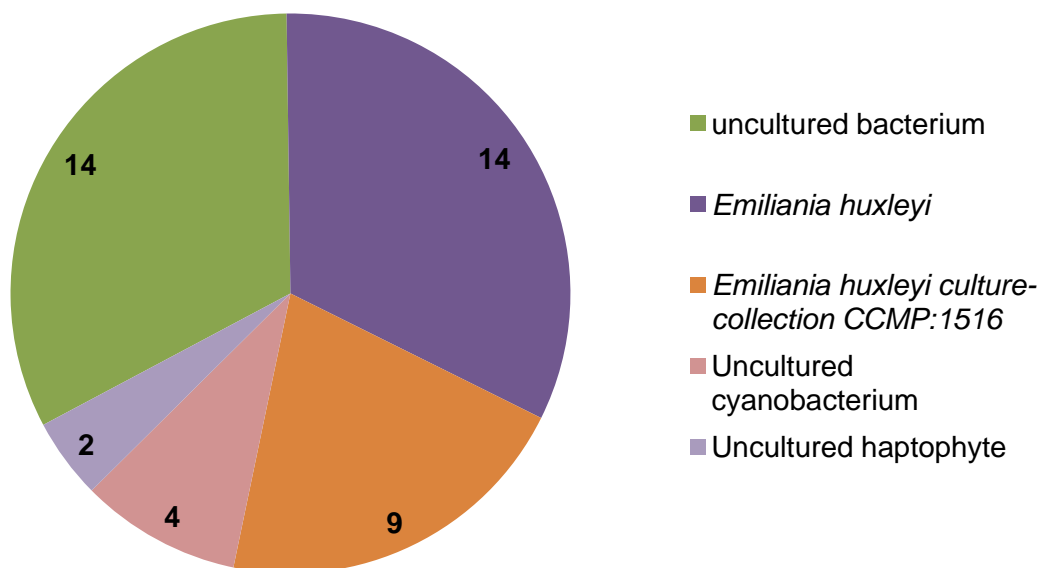


Figure 19. Pie chart showing the numbers of fragment hits from a plastid 16S rRNA gene clone library from station 1 during the Long Chain Diol cruise. Fragments were obtained with the primer combination PLA491F and OXY1313R.

Eustigmatophyceae diversity at the long chain diol cruise station 1

Primers Eust287F/810R, designed in this study for the specific detection of members of the Eustigmatophyte group, were used to amplify an amplicon (500 bp approximately) using genomic DNA extracted from station 1 of the LCD cruise. The PCR product was purified, cloned and individual 18S rRNA gene fragments were sequenced. Out of the 106 fragments sequenced only 99 sequences had good quality for analyzing with BLAST. Of these 99 sequences (Figure 20), 72 gave a positive match with Alveolate species *Parallelostrombidium* sp.(8), *Strombidium* cf. *basimorphum* (4), *Pseudotontonia simplicidens* (3), *Strombidium* sp (1) and uncultured Alveolates (56). Twenty-four sequences had a best match with *Dictyocha globosa* (5) or *speculum* (7) and uncultured *Florenciella* (12) both belonging to the Dictyophyceae phylum (a phylum close related to the Eustigmatophyceae; Figure 7).

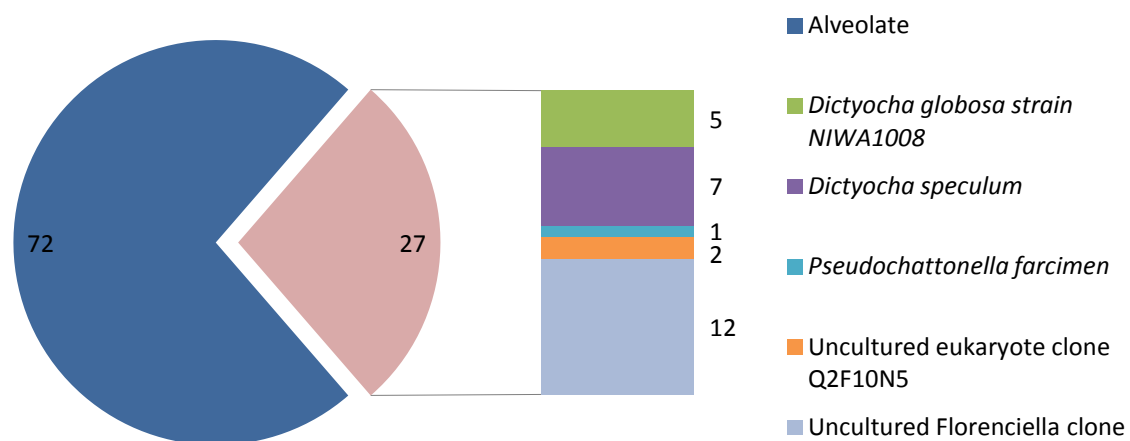


Figure 20. Pie chart showing the numbers of fragment hits from the 18S rRNA gene clone library at station 1 of the long chain diol cruise. Fragments were obtained with the primer combination Eust287F/Eust810R.

A maximum likelihood tree (Figure 21) was constructed with twenty-three 18S rRNA gene fragments from station 1 of the long chain diol cruise. These gene sequences were closely related to the species *Dictyocha sp.* and *Florenciella sp.*. Eleven sequences were closely related to the 18S rRNA gene sequence of *Florenciella sp.*, however the sequences were placed outside the Dictyochophyceae clade and closer to the Eustigmatophyceae clade. In addition, twelve sequences related to *Dictyocha speculum* and *Dictyocha globosa* were placed outside the Dictyochophyceae clade.

The gene fragments that were used in the maximum likelihood tree of Figure 21 were again aligned in ARB and a new phylogenetic tree was made with all available eukaryotic 18S rRNA gene sequences. Figure 22 shows the Stramenopile cluster of this tree. In this phylogenetic tree all 18S gene fragments fall in the Dictyochophyceae phylum, within this phylum there are two groups. The first group comprises the 18S rRNA gene sequences colored in red (Figure 22) and it is grouped closely with *Dictyocha speculum* and *Dictyocha fibula*. The second group comprises the 18S rRNA gene sequences colored in blue (Figure 22), which are closely grouped with the cultured *Florenciella parvula*.

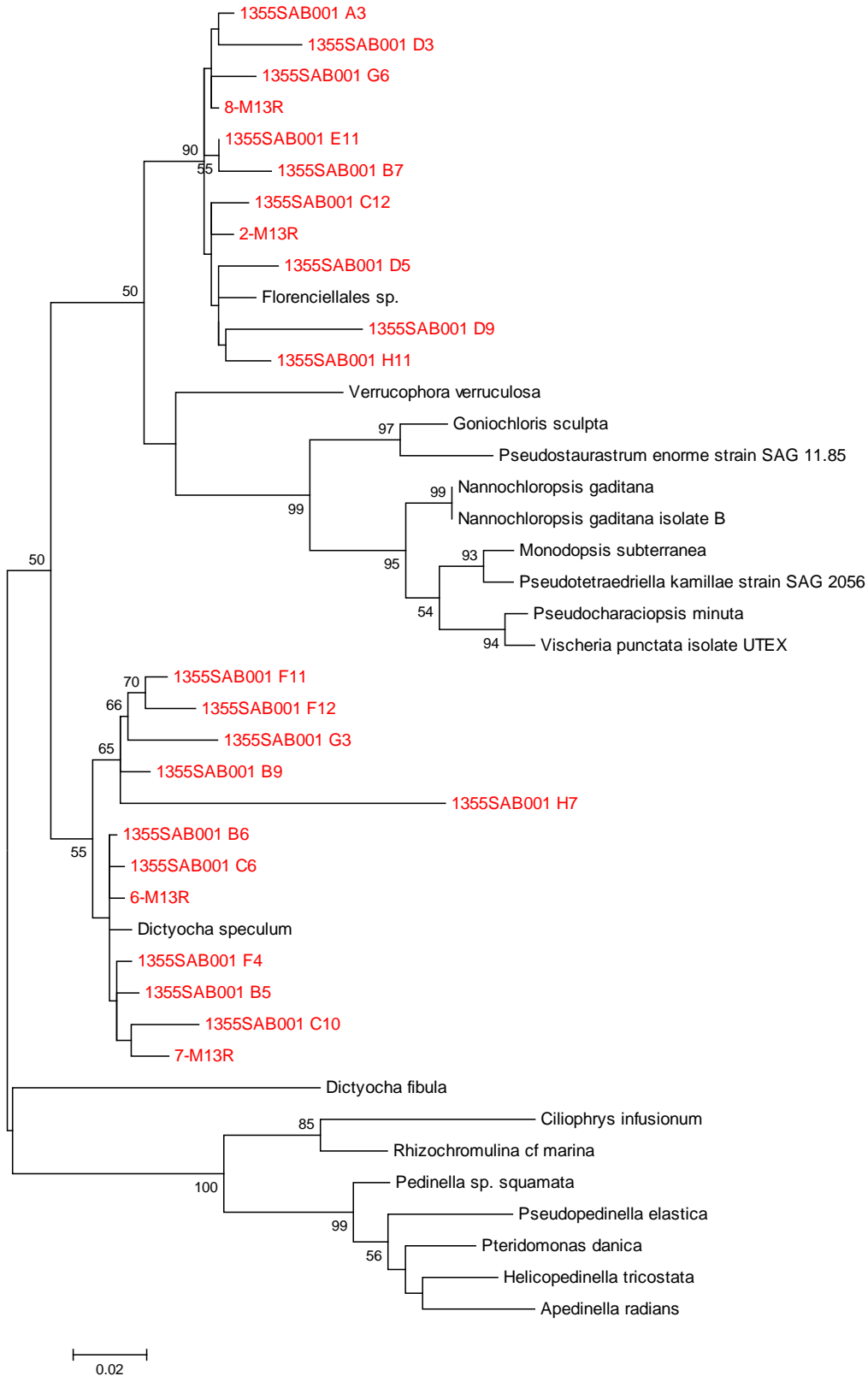


Figure 21. Maximum likelihood tree of the *Florenciella* sp. and *Dictyocha* sp. 18S rRNA gene fragments from station 1 during the Long Chain Diol cruise. 18S rRNA gene fragments from genomic databases (Genbank) were inserted. Bootstrap values >50% are shown on the tree branches. Sequences in red were obtained with the primer combination Eust287F/Eust810R as indicated above.

Stramenopiles

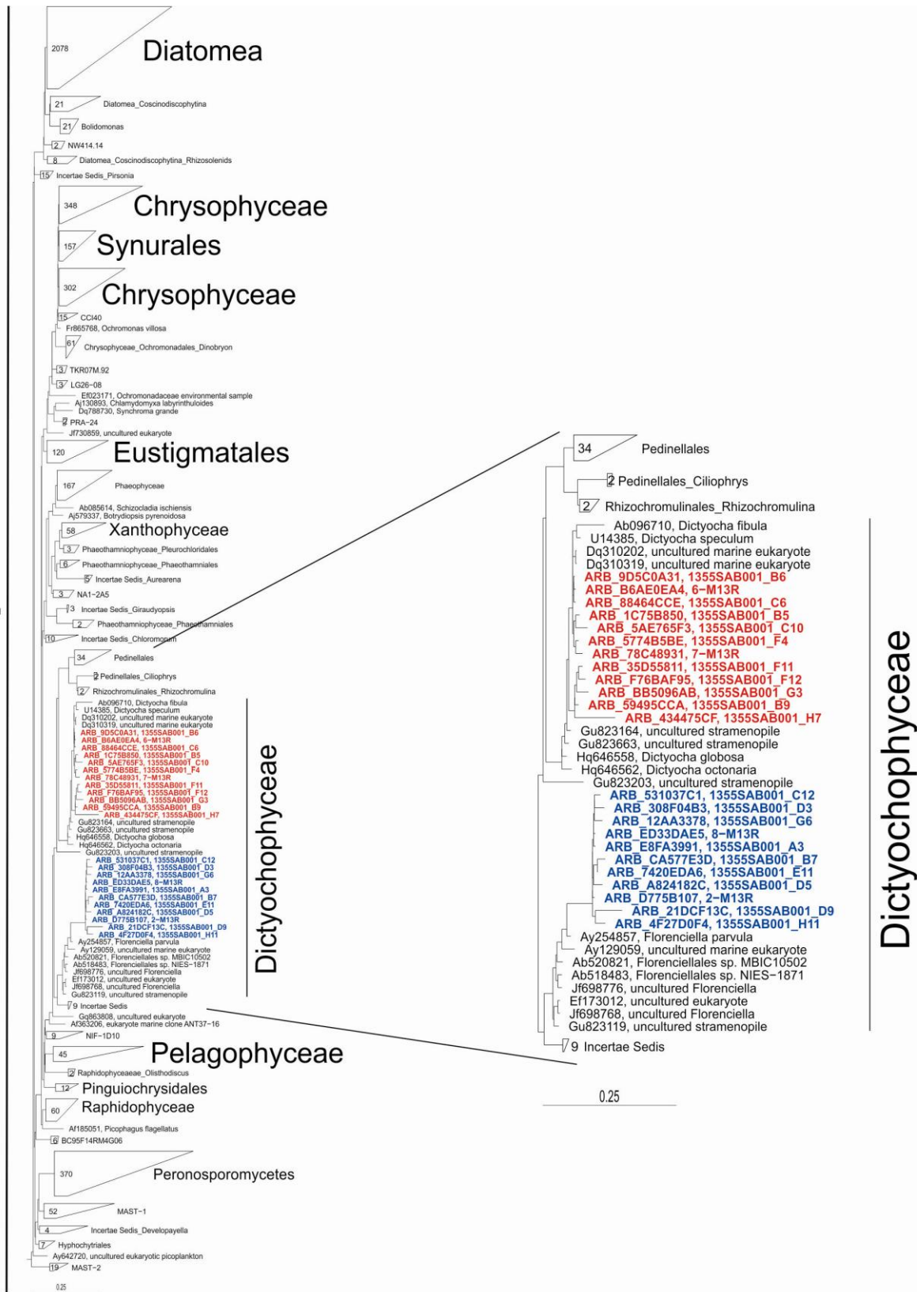


Figure 22. Phylogenetic tree in ARB of the *Florenciella* sp. and *Dictyocha* sp. 18S rRNA gene fragments from station 1 during the long-chain diol cruise. 18S rRNA gene fragments that matched with *Dictyocha* sp. are colored in red, the fragments that matched with *Florenciella* sp. are colored in blue.

LCD diversity and abundance at station 1

This data was kindly provided by Dr. Marta Rodrigo-Gamiz and included in the result section for comparison purposes.

No LCDs were detected in the SPM of any of the stations sampled around Iceland. However, LCDs were detected in sedimenting particles as shown in Figure 23 (collected with a sediment trap, deployed at 2,000 meters depth at Station 1; Figure 11). As observed in Figure 23, C₂₈ 1,13 and the C₃₀ 1,13-diol were abundant in the sedimenting particles collected in July 2011. However, smaller amounts were detected in the following month (August). An increase in the amount of collected C₂₈ 1,13 and C₃₀ 1,13-diol occurred in September and high amounts were detected until mid-October. Until May 2012, the collected amount of LCDs remained relative low with a single increase of the C₂₈ 1,13-diol collected in March 2012. During April 2012, the collected C₂₈ 1,13, C₃₀ 1,13 and 1,15-diol were at a minimum. C₂₈ 1,13 and C₃₀ 1,13-diol were found in the sediment trap material collected in May 2012. In the first collection in May the C₃₀ 1,13-diol was more abundant than the C₂₈ 1,13-diol and in the second collection in May 2012 and the first in June the C₂₈ 1,13-diol is the most abundant one. C₃₀ 1,15-diol was found in relative low amounts in all the sediment trap samples.

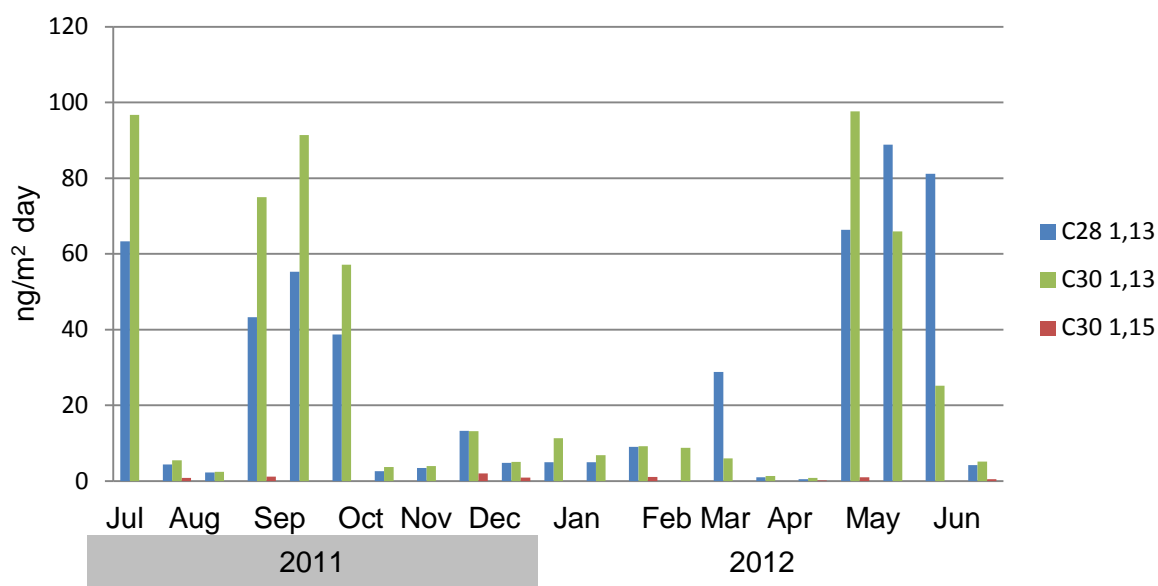


Figure 23. LCDs found in sediment collected from July 2011 until June 2012 with a sediment trap deployed at 2,000 meters depth (data by Marta Rodrigo-Gamiz).

4.2.2. Lake Challa

Eukaryotic diversity in the Lake Challa SPM

The primer pair Euk1A/Euk516R for picoeukaryotes 18S rRNA gene detection was applied to genomic DNA extracts obtained from SPM samples of Lake Challa and the PCR products were run in a DGGE gel (Figure 24). Representative bands were selected, amplified and sequenced. The affiliation of the sequences and percentage of homology with closest relative is listed in Table 13.

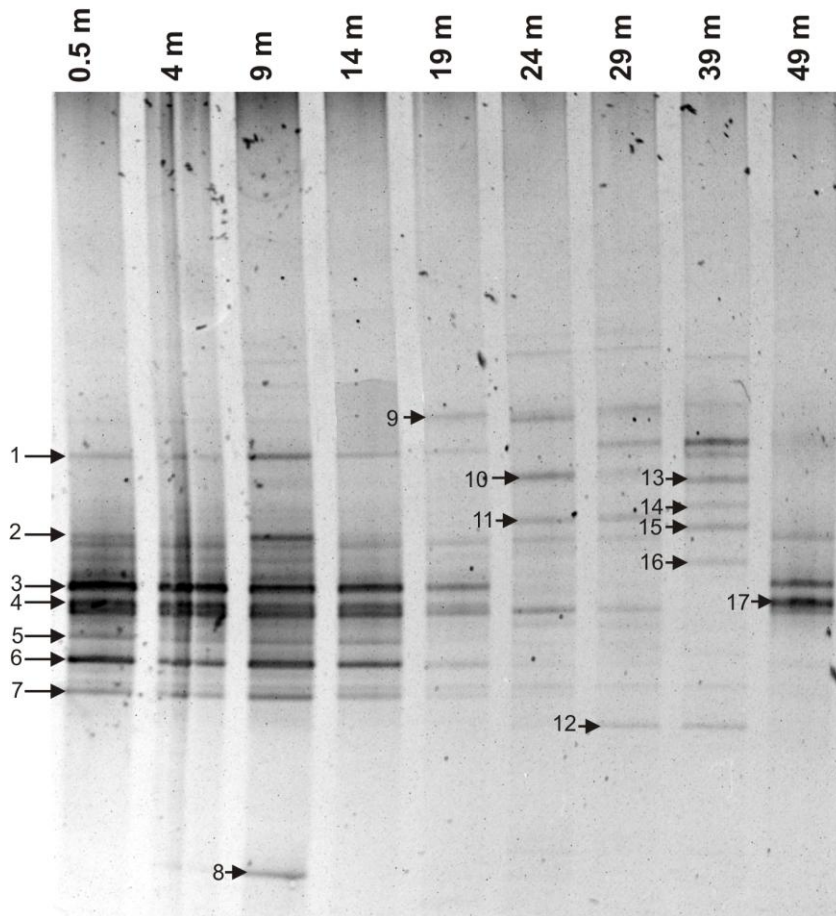


Figure 24. DGGE fingerprints of 18S rRNA gene fragments from the suspended particulate matter taken at different depths in Lake Challa. Bands picked for sequencing are numbered.

Table 13. Affiliation of DGGE bands recovered in Lake Challa SPM by using the Euk1A/Euk516R primer pair.

Picked band	Size (bp)	%	Closest relative and affiliation	Accession number
1	489	97	<i>Meseres corlissi</i> (Ciliate; Ciliophora, Halterriidae)	EU399529.1
2	331	92	Uncultured stramenopile	FN690664.1
3	491	91	<i>Mallomonas papillosa</i> (Stramenopiles; Synurophyceae)	JQ281519.1
4	415	92	<i>Mortierella alpine</i> (Fungi; Mortierales)	JQ040258.1
5	474	85	Uncultured alveolate	GQ844679.1
6	462	85	Uncultured alveolate	GQ844679.1
7	439	96	<i>Raineriophrys erinaceoides</i> (Centrohelids; Acanthocystidae)	AY749633.1
8	368	99	<i>Bythotrephes longimanus</i> (Crustacea; Cercopagididae)	AY075082.1
9	395	99	Didinium sp. KFS-B (Ciliate; Ciliophora, Didiniidae)	AB558118.1
10	265	99	<i>Coleps hirtus</i> (Ciliate; Ciliophora, Colepidae)	HM140399.1
11	368	92	<i>Ochromonas danica</i> (Stramenopiles, Chrysophyceae)	JQ281514.1
12	428	99	Uncultured eukaryote clone TW-A1-1-8g (Protozoa)	EU860733.1
13	469	99	Uncultured ciliate clone MLBA1.54	FJ410555.1
14	424	98	Uncultured chrysophyte clone	GQ330571.1
15	482	100	<i>Synedra ulna</i> (Diatom; Bacillariophyta, Fragilariaceae)	AM497727.1
16	444	98	Uncultured fungus clone WD4-121	GQ844550.1
17	478	94	Uncultured stramenopile clone WD4-73	GQ844505.1

%. Percentage of homolog with the closest relative.

Most of the sequences were closely related to 18S rRNA gene sequences of ciliates and fungi. Only 4 sequences were closely related to algal groups, specifically to the Synurophyceae, Chrysophyceae, and Bacillariophyceae phyla. In the case of the closest relatives to the Synurophyceae and Chrysophyceae phyla, the percentage of homology of the amplified sequences with known representatives was low (91-92%, band 3 and 11). Band 3, affiliated to *Mallomonas* sp. (Synurophyceae phylum; 91%) was detected from 0.5 to 19 m depth being more intense in the surface SPM samples. Band 11 affiliated with *Ochromonas* sp. (Chrysophyceae phylum; 92%) was only clearly detected between 24 to 29 m depth. Band 14, affiliated to an uncultured chrysophyte clone, was only found at 39 m depth. Also, band 15, closely affiliated to the diatom *Synedra* sp. of the Fragilariaceae family, was only detected at 39 m depth.

Eustigmatophyceae diversity in the SPM of Lake Challa

Clone libraries were generated by cloning the PCR fragments generated by primers Eust287F/Eust810R in several Lake Challa SPM samples. From the Lake Challa SPM sample at 0.5 m, 114 sequences were recovered and further analyzed. With a BLAST maximum identity threshold of 90%, 89 sequences were used for further analysis. 55 out of the 89 sequences gave a positive match with species belonging to the Eustigmatophyceae phylum such as *Nannochloropsis* sp., *Pseudostaurastrum enorme*, *Pseudotetraedriella kammillae* and *Vischeria stellata*. (Figure 25A).

The clone library of the PCR fragment generated by the Eust287F/810R primer pair in the Lake Challa SPM sample at 9 m generated 84 good quality sequences of . 31 out of the 84 sequences gave a positive match with species belonging to the Eustigmatophyceae phylum like *Nannochloropsis salina* (Figure 25B). The clone library of the PCR fragment generated by the Eust287F/810R primer pair in the Lake Challa SPM sample at 19 m depth generated 68 sequences of good quality for further analysis. 67 sequences with a BLAST maximum identity threshold of 90% were obtained. 36 out of the 67 sequences gave a positive match with species belonging to the Eustigmatophyceae phylum (Figure 25C).

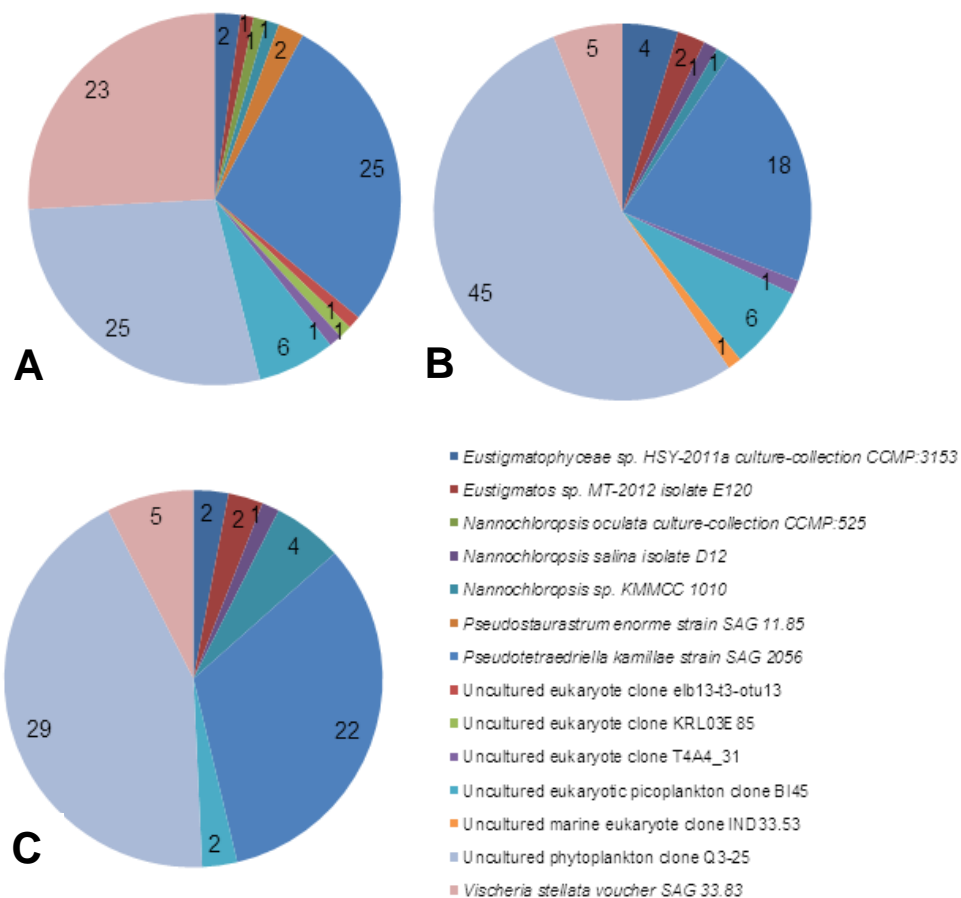


Figure 25. Pie charts showing the numbers of fragment hits from a 18S rRNA gene clone libraries from Lake Challa at 0.5 meter depth (A), 9 meters depth (B) and 19 meters depth (C). Fragments were obtained with the primer combination Eust287F/Eust810R.

To have a better overview of changes in abundance and diversity of the major Eustigmatophyceae species, data was plotted in clusters (Figure 26). The sequences closely related to *Vischeria stellata* showed a decrease in abundance below 0.5 m depth.

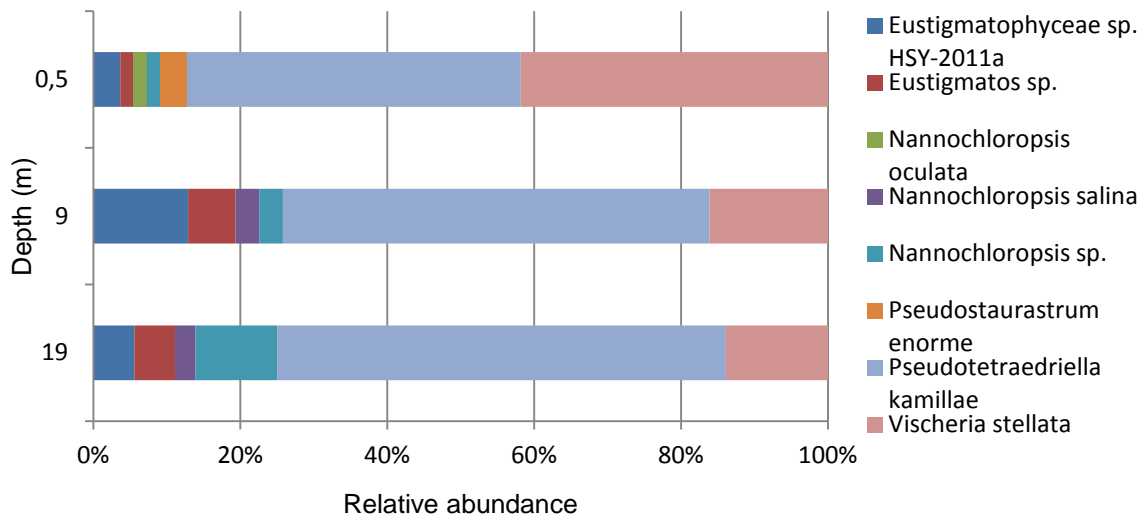


Figure 26. Bar graph showing the relative abundance of the cultured Eustigmatophyceae species hits with BLAST over the three different water column depths. Data is derived from 18S rRNA gene clone libraries with the primer combination Eust287F/Eust810R.

A maximum likelihood tree was constructed with eighty-five 18S rRNA gene fragments recovered from the 0.5 meters depth clone library (Figure 27). These gene fragment were closely related to Eustigmatophyceae species. Twenty-three gene sequences were closely related to sequences belonging to the *Vischeria sp.*, the *Pseudochloropsis minuta* and the *Eustigmatos* species. In this group there was a distinctive subgroup comprising 16 sequences (Group 1a; Figure 27). The second subgroup (Group 1b) was closely related to the *Vischeria* dominating group and contained 4 sequences. All the sequences in these two subgroups and the three in the middle were closely related to *Vischeria stellata voucher SAG 33.83*. In the Eustigmatophyceae clade (Figure 27; colored in green) contained the 18S rRNA gene sequence of the uncultured phytoplankton clone Q3-25, which was closely related to most of the gene sequences obtained from the SPM in Lake Challa (Figure 25).

Group 2 (Figure 27; colored in blue) contained 40 gene sequences, closely related to the Eustigmatophyceae clade, however, not with cultured individual members of this clade. This group is composed by gene sequences which matched the cultured Eustigmatophyceae members *Pseudotetraedriella kamillae* (25 sequences), *Pseudostaurastrum enorme* (2), *Eustigmatos sp.* (1), *Eustigmatophyceae sp. HSY-2011a* (1), and also sequences of some uncultured species like phytoplankton clone Q3-25 (9), eukaryote clone elb13-t3-otu13 (1) and eukaryote clone KRL03E85 (1). Group 3 (colored in purple) comprised by 16 gene sequences, is more distant from the Eustigmatophyceae clade than group 2.

Sequences in this group are closely related to cultured Eustigmatophyceae species *Nannochloropsis oculata* (1) and *Nannochloropsis* sp. *KMMCC 400*. The Chrysophyceae (Figure 27; colored in light blue) and Dictyochophyceae clades (Figure 27; colored in light green) are separated from the Eustigmatophyceae but closely related. None of the obtained gene sequences fell in these two clades. Some of the recovered sequences form a cluster separated to the previous ones (Figure 27, pink and grey cluster). Sequences included in the pink cluster (Figure 27) were closely related to the 18S rRNA gene of the uncultured eukaryotic picoplankton clone BI45 (all with an identity score of 98%, accession number; EF196772.1). Sample 1355SAB000_H7 (colored in grey) is matched with the 18S rRNA gene of the uncultured eukaryote clone T4A4_31 (with an identity score of 99%, accession number; HQ394059.1).

We also compared this tree (Figure 27) with a phylogenetic tree generated in ARB which includes a more complete database of all the 18S rRNA gene sequences available up to date. The 18S rRNA gene sequences recovered from the Eust287F/Eust810R clone libraries at 0.5, 9 and 19 m were added to the ARB tree by using the ARB parsimony tool (Figure 28). Group 1 in Figure 28 contains five sequences derived from SPM collected at three different depths (0.5 m I8; 9 m A11; 19 m A4, E3 and H4). Group 2 (Figure 28; colored in red) contained seventy-five 18S rRNA gene sequences obtained from three different depths. Group 3 is divided into four subgroups (Figure 28; Group 3a, b, c, & d; colored in green), each subgroup contains only 18S rRNA gene sequences retrieved from the SPM at Lake Challa and previously described sequences. Subgroup 3a contains six sequences from 0.5 m and 9 m depth. Subgroup 3b contains twenty-four sequences from three different depths (0.5 m, 9 m and 19 m). Subgroup 3c contains thirteen sequences from three different depths. Subgroup 3d contains sixty-two sequences from three different depths. Group 4 has two subgroups (Figure 28; Group 4a & b; colored in blue). Both subgroups only contain 18S rRNA sequences from the SPM of Lake Challa, which are not related to previously described 18S rRNA gene sequences. Subgroup 4a contains twenty-nine sequences from three different depths. Subgroup 4b contains 7 sequences from two different depths (0.5 m and 9 m). For an overview of the 18S rRNA gene sequences within the different groups see Supplementary Figure 1.

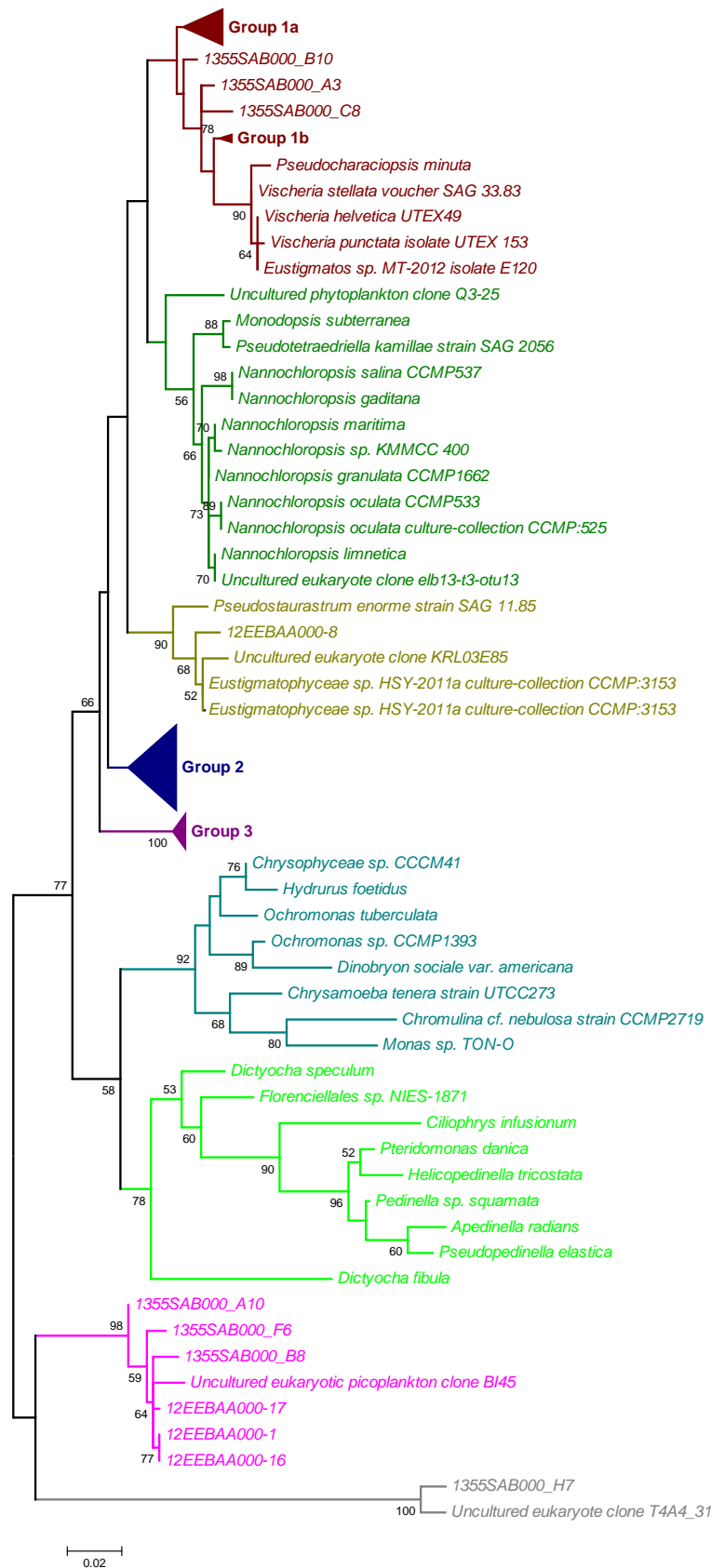


Figure 27. Maximum likelihood tree of the 18S rRNA gene sequences obtained from suspended particulate matter collected at 0.5 m depth at lake Challa. 18S rRNA gene fragments obtained from Genbank were inserted. Bootstrap values >50% are shown at the nodes. Fragments were obtained with the primer combination Eust287F/Eust810R.

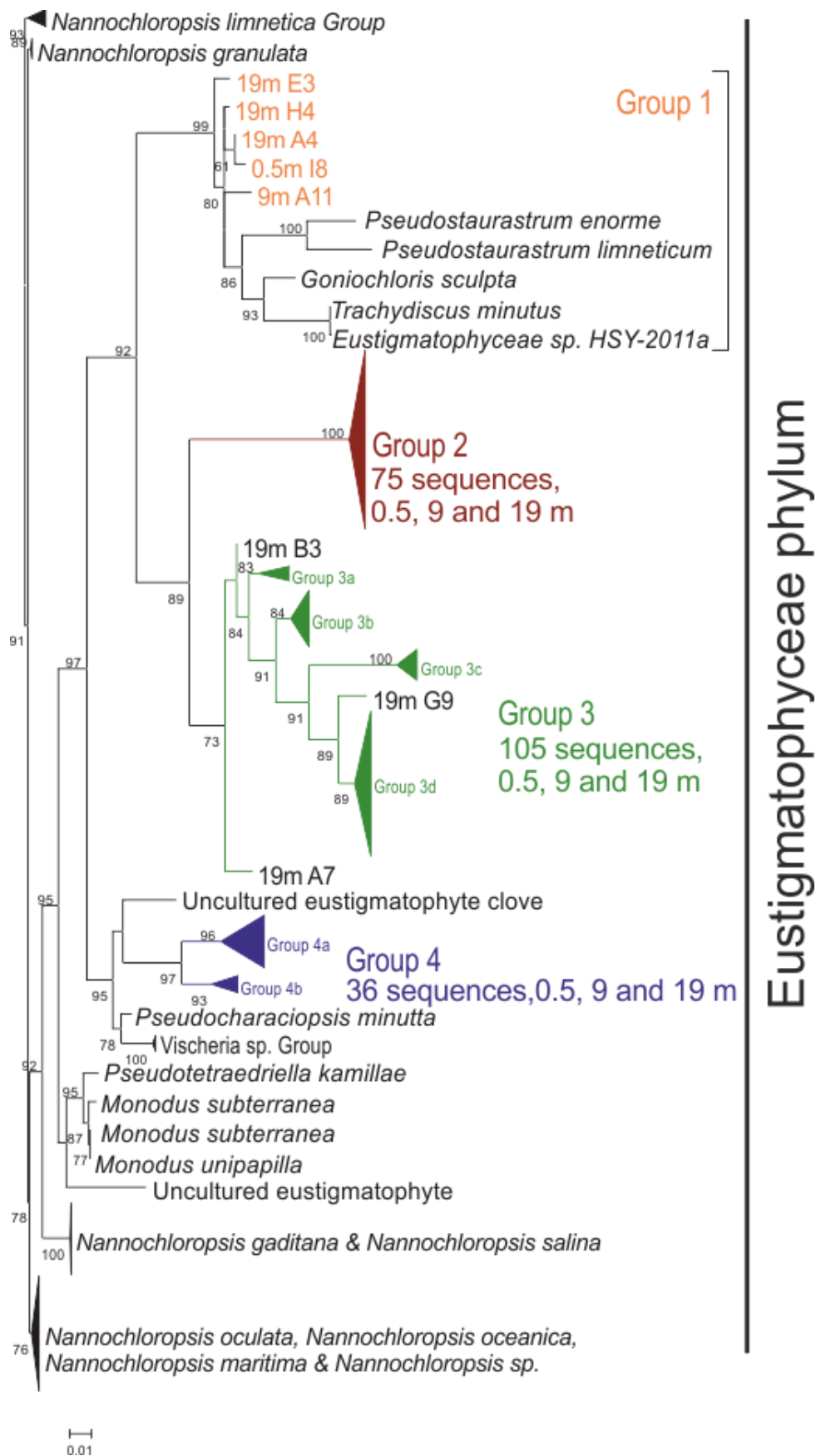


Figure 28. Maximum likelihood tree of the Eust287F/Eust810R 18S rRNA gene sequences obtained from suspended particulate matter collected at 0.5 m, 9 m and 19 m depth at lake Challa.

Eustigmatophyceae 18S rRNA gene and LCD abundances in the Lake Challa water column

The Eustigmatophyceae primers Eust287F/Eust810R were also used for quantification by qPCR. The SPM samples of lake Challa showed a distinctive peak of Eustigmatophyceae 18S rRNA gene copies per liter at 9 m depth (approximately 3.5×10^4 copies/L) (Figure 29, A). Below 9 m the total gene copies per liter quickly declined to 5.3×10^3 copies at a depth of 24 m. Gene copy numbers at 0.5 m samples were under the detection limit of the qPCR method used. The abundance of LCDs in the water column of Lake Challa peaks at 9 m depth (62 ng L^{-1} of total LCDs) (Figure 29, B). After that it declines to 6.9 ng L^{-1} , at 24 m depth. The total amount of recovered LCDs is relative high at water column depths of 0.5 and 4 m with respectively 46 and 38 ng L^{-1} .

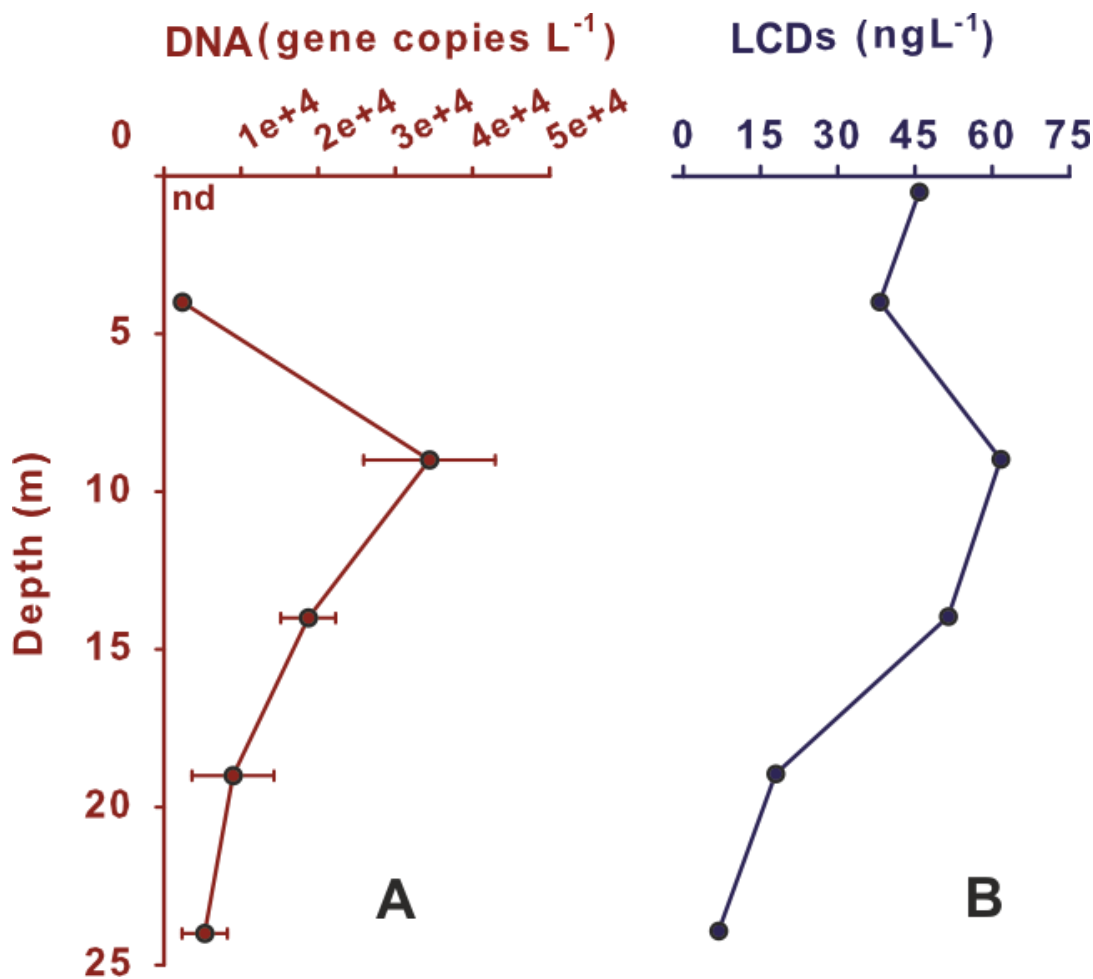


Figure 29. Water column profiles from SPM obtained at Lake Challa. **A** Quantification of Eustigmatophyceae 18S rRNA gene copies. No error bar given for the sample taken at 4 meters depth, only one of the triplicate came positive above the detection limit in the qPCR quantification (nd = non detected). **B** Quantification of the total amount of LCDs.

A broad range of LCDs were identified in the SPM collected in the water column of Lake Challa. The two most abundant LCDs found in the top layer (till 24 m depth) of the lake were the C_{30} 1,15 and the C_{32} 1,15-diol, which accounted respectively for 25% and 62% of the total amount of LCDs detected. The third abundant LCD was the C_{34} 1,17-diol, which accounted for 11% of the total amount of LCDs. Other minor (< 1.5%) LCDs were the C_{30} 1,14, C_{30} 1,16, C_{32} 1,16 and the C_{34} 1,15). However, these LCDs were not detected in all the analyzed water column depths.

LCD quantification in the sediment trap material

In addition, we determined the LCD composition and abundance in a sediment trap material collected at 35 m water depth in the center of the lake from winter to summer 2010. The maximum accumulation of LCDs was detected in March 2010 ($0.6 \mu\text{g}$ of LCDs $\text{m}^{-2} \text{day}^{-1}$), with the C_{30} 1,15 and the C_{32} 1,15-diol as the most abundant LCDs (Figure 30). A decline in the abundance of the C_{30} 1,15-diol was detected from March onwards with a small peak in July. Other LCDs were also detected during these months but in relatively low amounts (Table 14).

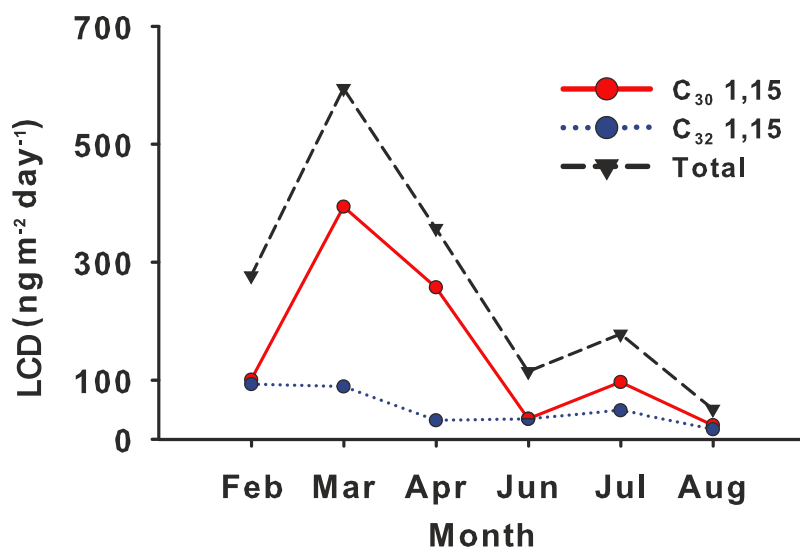


Figure 30. Long-chain diols quantification in a sediment trap deployed in the center of Lake Challa.

Table 14. Abundance of LCDs in sediment trap material (ng of LCD $\text{m}^{-2} \text{day}^{-1}$).

Month of recovery	C_{30} 1,16	C_{30} 1,15	C_{30} 1,14	C_{30} 1,13	C_{32} 1,16	C_{32} 1,15	C_{34} 1,17
Feb	24	101	18	41		93	
Mar	28	394	38	28	4	89	12
Apr	18	257	25	20		32	4
Jun	3	35	35			35	8
Jul	5	97	11			49	17
Aug		24	3			17	7
Total	78	908	130	89	4	315	48

5. Discussion

In this study we designed and tested new 18S rRNA gene primers for the detection of members of the Eustigmatophyceae phylum, which are believed to be important producers of LCDs. By applying these primers in environmental samples (marine and freshwater suspended particulate matter), we have revealed the high diversity of eustigmatophyceae groups and novel eustigmatophyte sequences. Quantification of the eustigmatophyte 18S rRNA gene abundance and the total LCD content in a lake water column were correlated, suggesting a relation between the two parameters. The spatial diversity of LCDs in the lake water column indicates a singular biological source. However the seasonal diversity of LCDs varies over time which might indicate different blooms of LCD producers and or environmental condition affecting the LCD composition.

Design and testing of Eustigmatophyceae phylum 18S rRNA gene primers

The current universal picoeukaryote primers, designed to target the 16S rRNA (Fuller et al., 2006) or the 18S rRNA gene (Díez et al., 2001), are designed to match all the picoeukaryotes in environmental samples. However, in our study we have observed that these 'universal' primers are not covering the whole diversity and are often biased toward specific groups (Potvin & Lovejoy, 2009; Stoeck et al., 2006; among others). The biases introduced by 'universal' 18S rRNA gene primers in environmental studies suggests that the design of novel primers to target specific algal groups might be more informative. Thus, the designing and testing of Eustigmatophyceae phylum 18S rRNA gene primers in this study is timely.

Our designed primer pair Eust287F/Eust810R matched sequences of members of the Eustigmatophyceae phylum and also a limited number of sequences outside this group. In our case, by performing clone libraries of the amplified 18S rRNA gene fragments, we have been able to report a high diversity within the Eustigmatophyceae but also observed a cross hybridization of our primers with other stramenopiles groups. This should be taken into account in environmental studies in which these primers are applied. Unfortunately, the diversity of the 18S rRNA gene is restricted as well as the rather limited 18S rRNA gene database of algal species, which makes difficult to design specific primers. Future studies should be then focused on combining these Eustigmatophyceae 18S rRNA gene primers with the detection of other genes that could be distinctive in algal groups such as genes coding for the lipid biosynthesis pathway, the rubisco or photosystem proteins (psaA, psaB, psbA, etc.). Studying environmental samples with multiple primer pairs would give a more accurate view on the picoeukaryotic and even the Eustigmatophyceae diversity.

LCD diversity and potential producers in Iceland suspended particulate matter

In the SPM samples collected around Iceland, low LCD quantification was reported (M. Rodrigo-Gamiz, personal communication), and no amplification was

detected by applying the designed Eustigmatophyceae 18S rRNA gene primers. The lack of amplification can be due to the low presence of LCD-producers (indicated by the low abundance of LCDs in the samples) or alternatively it could be due to mismatches between the designed primers and yet-unknown marine eustigmatophyceae sequences present in Iceland waters SPM.

The LCD composition was also examined in a sediment trap placed in the Iceland Basin, Southeast of Iceland. The LCDs with higher abundances in the sediment trap were C₂₈ and C₃₀ 1,13-diol (up to 98 ng m⁻² day⁻¹). However their relative abundances changed throughout the year. For example, in the beginning of May 2011 the C₃₀ 1,13-diol was more abundant than the C₂₈ 1,13-diol, while at the end of May and in June 2011 the C₂₈ 1,13-diol was more dominant. These seasonal differences in fractional abundances can indicate a change in the LCD producer or changes in the LCD relative abundances being produced by the same algal group as a response to seasonal changes in the environmental conditions or physicochemical factors in the water column.

The detection of LCDs in sediment trap material but not in the surface water SPM samples could be due to the fact that the LCDs collected in the sediment trap were transported from different areas by influence of the different currents surrounding Iceland. Another reason would be the lack of algal blooms around the time we collected the SPM samples. The marked seasonal differences in LCDs suggest that LCD-producer blooms might be occurring in Spring months rather than in the Summer. To estimate the seasonal diversity of potential LCD-producers, future studies will be focused on the application of molecular methods to sediment trap material and combined to determinations of diversity and abundance of LCDs.

LCD diversity and potential producers in Lake Challa SPM

Eustigmatophyceae 18S rRNA gene sequences were recovered from the Lake Challa SPM samples. Phylogenetic analyses of the Eustigmatophyceae 18S rRNA gene sequences pointed to the diversification of this phylum into different clusters. One of these clusters (Figure 28; Group 4) was closely related to the *Vischeria* genus. In addition, a couple of sequences were also closely related to the eustigmatophytes *Pseudostaurastrum* sp. and *Goniocloris* sp. (Figure 28; Group 1). On the other hand, two distinctive gene sequence clusters (cluster 2, 3, Figure 28) were not associated to any previously described Eustigmatophyceae species. This cluster was supported by high bootstrap values in the maximum likelihood phylogeny. Thus, we conclude that the application of our Eustigmatophyceae 18S rRNA gene primers has unraveled the presence of a distinctive, non-yet cultured eustigmatophyte group.

The qPCR data derived from the SPM in the water column of Lake Challa reveals a peak in Eustigmatophyte 18S rRNA gene copy numbers at 9 meters water depth (Figure 29). Since Eustigmatophyceae are photosynthetic algae, it seems unlikely to find highest concentrations of this algal group deeper than the light penetration depth (5-7 m reported in the SPM sampling by Buckles et al., 2013). An

explanation for this could be that in fact the maximum abundance of Eustigmatophyceae 18S rRNA gene abundance is between 4 and 9 m depth. It is then possible that the preferred niche of Eustigmatophyceae members is located in this depth range, perhaps eustigmatophytes members are driven down from the top of the water column by heavy competition with other photosynthetic organisms. Another possibility is that the recovery of SPM followed an Eustigmatophyceae bloom and that the dead sinking organic material (including LCDs and DNA) was mainly located in that depth range.

In the SPM C_{30} and C_{32} 1,15-diols were the most abundant LCDs (Figure 30). The predominance of these LCD has been previously reported in both marine and lacustrine environmental settings (Schouten et al., 2000; Smith et al., 1983; Ten Haven et al., 1992; Castañeda et al., 2011; Shimokawara et al., 2010; Cranwell et al., 1987; Huang et al., 1995). Moreover in culture studies of marine and freshwater eustigmatophytes were also the C_{30} 1,15 and the C_{32} 1,15-diol detected. In the marine eustigmatophyte cultures the C_{30} 1,15-diol was the major LCD (Méjanelle et al., 2003; Shimokawara, et al., 2010; Volkman et al., 1992) while the C_{30} 1,15-diol was most abundant in the freshwater Eustigmatophyte cultures (Volkman et al., 1999). In Lake Challa SPM samples collected in January 2010, a dominance of the C_{32} 1,15-diol was detected. However, in the sediment trap material from February 2010 onward the most abundant LCD was C_{30} 1,15-diol. This shift in the dominance between C_{32} 1,15-diol and C_{30} 1,15-diol may be caused by changing environmental conditions like lake surface temperature and precipitation. Another possibility is the occurrence of two different algal blooms in the lake during the year contributing with either with a pool of the C_{30} 1,15-diol or the C_{32} 1,15-diol.

One identified peak of the total LCD abundance in Lake Challa (Figure 31) occurs simultaneously with the long rain season (from March to May) induced by twice-yearly passing of the Inter Tropical Convergence Zone (ICTZ) in Eastern Africa. Romero-Viana et al. (2012) concluded that the blooms of algae in Isabel lake (a hypersaline lake on Isabel Island, Eastern Pacific) were triggered by freshwater input caused by the wet season. Thus, it is possible the peak of the total LCD abundance identified in the long rain season is derived from an algal bloom favored upon entrance of rain water, what can induce an increased influx of land-derived nutrients.

Finally, the C_{34} 1,17-diol is the LCD with the longest carbon chain detected in the SPM and in the sediment trap samples. In the SPM (collected in January) it was the third most abundant LCD (23 ng L^{-1} ; 10% of the total amount of LCD detected). So far, to our knowledge, this LCD has only been detected in two other freshwater lakes, lake Baikal in Russia (Shimokawara et al. 2010) and in lake Valencia in Venezuela (Xu et al., 2007). Thus, C_{34} 1,17-diol could be a specific LCD of freshwater algal groups. Future studies should address its provenance and seasonal/spatial variation.

5.2. Future work

In our study we analyzed sediment trap material collected in just half of the year. In these months, we observed important changes in the LCD abundance and composition. In order to detect LCD seasonal changes it is obviously necessary to capture the full annual cycle of the lake, preferable sampling twice a month for a higher sensitivity. For a more detailed study about the correlation between LCDs and their biological source it would be of great importance to simultaneously collect and correctly preserve the samples for both LCDs and nucleic acid analyses. Previous studies have suggested preservation solutions to fixed sediment trap material that could potentially not interfere with nucleic acid extractions (e.g. Karl et al., 2011).

With the outcome of the proposed sediment trap set-up we could determine the timing of highest recovery of LCDs and potentially of LCD-producer bloom. These kind of samples could be further investigated by metagenomic approaches. Metagenomics sequencing would targeted to 18S rRNA gene diversity or also to total DNA sequence without biases introduced by primers. With these kind of studies we would be able to unravel the planktonic diversity in the sampling site and design specific primers based on the detected sequences. Nucleic acid-based analyses should be also complemented by approaches involving the identification of Eustigmatophyceae algae by microscopy and isolation in pure culture. Growing cultures of Eustigmatophyceae algae under different growing conditions can give more insight in the LCD composition under changing conditions and compare to the LCD composition detected in the environment in which the culture was isolated.

References

- 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, pp.369–380.
- Auguet, J.-C. & Casamayor, E.O., 2013. Partitioning of Thaumarchaeota populations along environmental gradients in high mountain lakes. *FEMS microbiology ecology*, 84(1), pp.154–164.
- Berger, W.H., Bonneau, M.-C. & Parker, F.L., 1982. on the deep-sea floor: lysocline and dissolution rate. *Oceanologica acta*, 5(2), pp.249–258.
- Blokker, P. et al., 1999. Molecular structure of the resistant biopolymer in zygosporangium cell walls of *Chlamydomonas monoica*. *Planta*, 207, pp.539–543.
- Brassel, S.C. et al., 1986. Molecular stratigraphy: a new tool for climate assessment. *Nature*, 320, pp.129–133.

- Brochier-Armanet, C. et al., 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature reviews. Microbiology*, 6(3), pp.245–52.
- Brown, S.J. & Elderfield, H., 1996. Variations in Mg/Ca and Sr/Ca ratios of planktonic foraminifera caused by postdepositional dissolution: Evidence of shallow Mg dependent dissolution. *Paleoceanography*, 11(5), pp.543–551.
- Buckles, L.K. et al., 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. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23560451> [Accessed April 29, 2013].
- Castañeda, I.S. et al., 2011. Organic geochemical records from Lake Malawi (East Africa) of the last 700years, part II: Biomarker evidence for recent changes in primary productivity. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 303(1-4), pp.140–154.
- Cranwell, P.A., Eglinton, G. & Robinson, N., 1987. Lipids of aquatic organisms as potential contributors to lacustrine sediments--II *. *Organic Geochemistry*, 11(6), pp.513–527.
- Damsté, J.S.S. et al., 2002. Distribution of Membrane Lipids of Planktonic Crenarchaeota in the Arabian Sea. *Applied and environmental microbiology*, 68(6), pp.2997–3002.
- Díez, B. et al., 2001. Application of Denaturing Gradient Gel Electrophoresis (DGGE) To Study the Diversity of Marine Picoeukaryotic Assemblages and Comparison of DGGE with Other Molecular Techniques. *Applied and environmental microbiology*, 67(7), pp.2942–2951.
- Emiliani, C., 1955. Pleistocene Temperatures. *Journal of Geology*, 63, pp.538–578.
- Fuller, N. et al., 2006. Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using a PCR biased towards marine algal plastids. *Aquatic Microbial Ecology*, 43, pp.79–93.
- Garidel-thoron, T. De et al., 2005. Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature*, 433, pp.294–298.
- Gelin, F. et al., 1999. Distribution of aliphatic, nonhydrolyzable biopolymers in marine microalgae. *Organic Geochemistry*, 30(2-3), pp.147–159.
- Gelin, F., Volkman, J.K., et al., 1997. Mid-chain hydroxy long-chain fatty acids in microalgae from the genus nannochloropsis. *phytochemistry*, 45(4), pp.641–646.
- Gelin, F. et al., 1996. Novel, resistant microalgal polyethers: An important sink of organic carbon in the marine environment? *Geochimica et Cosmochimica Acta*, 60(7), pp.1275–1280.
- Gelin, F., Boogers, I., et al., 1997. Resistant biomacromolecules in marine microalgae of the classes Eustigmatophyceae and Chlorophyceae: Geochemical implications. *Organic Geochemistry*, 26, pp.659–675.
- Guindon, S. et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), pp.307–21.

- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. , pp.95–98.
- Ten Haven, H.L. et al., 1992. Variations in the content and composition of organic matter in sediments underlying active upwelling regimes: a study from ODP Legs 108, 112, and 117. *Geological Society, London, Special Publications*, 64(1), pp.229–246.
- Ten Haven, H.L. & Rullkötter, J., 1991. Preliminary lipid analyses of sediments recovered during Leg 117. *Proceedings of the Ocean Drilling Program, Scientific Results*, 117, pp.561–569.
- Huang, Y. et al., 1995. Molecular and isotopic biogeochemistry of the Miocene Clarkia Formation: hydrocarbons and alcohols. *Organic Geochemistry*, 23(9), pp.785–801.
- Jenkyns, H.C. et al., 2012. Warm Middle Jurassic–Early Cretaceous high-latitude sea-surface temperatures from the Southern Ocean. *Climate of the Past*, 8(1), pp.215–226.
- Karl, D.M. et al., 2011. Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *PNAS*, 109(6), pp.1842–1849.
- Karner, M.B., DeLong, E.F. & Karl, D.M., 2001. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature*, 409(6819), pp.507–510.
- Kim, J.-H. et al., 2008. Global sediment core-top calibration of the TEX86 paleothermometer in the ocean. *Geochimica et Cosmochimica Acta*, 72, pp.1154–1173.
- Kim, J.-H., van der Meer, J., Schouten, S., Helmke, P., Willmott, V., et al., 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, pp.4639–4654.
- Kim, J.-H., van der Meer, J., Schouten, S., Helmke, P., Willmott, V., et al., 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, pp.4639–4654.
- 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, pp.854–862.
- Lacasse, C. et al., 1996. North Atlantic deep-sea sedimentation of Late Quaternary tephra from the Iceland hotspot. *Marine Geology*, 129(3-4), pp.207–235.
- De Leeuw, J.W. et al., 1980. On the occurrence and structural identification of long chain unsaturated ketones and hydrocarbons in sediments. *Physics and Chemistry of the Earth*, 12, pp.211–217.
- De Leeuw, J.W., Rijpstra, W.I.C. & Mur, L.R., 1992. The absence of long-chain alkyl diols and alkyl keto-l-ols in cultures of the cyanobacterium *Aphanizomenon flos-aquae* *. *Organic Geochemistry*, 18(4), pp.575–578.

- De Leeuw, J.W., Rijpstra, W.I.C. & Schenck, P.A., 1981. The occurrence and identification of C30, C31, and C32 alkan-1,15-diols and alkan-15-one-1-ols in Unit I and Unit II Black Sea sediments. *Geochimica et Cosmochimica Acta*, 45, pp.2281–2285.
- Leininger, S. et al., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, 442(7104), pp.806–809.
- Logan, G.A. & Eglinton, G., 1994. Biogeochemistry of the Miocene lacustrine deposit, at Clarkia, northern Idaho, U.S.A. *Organic Geochemistry*, 21(8), pp.857–870.
- Lopes dos Santos, R. a. et al., 2012. Late Quaternary productivity changes from offshore Southeastern Australia: A biomarker approach. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 363-364, pp.48–56.
- Lopes dos Santos, R.A. et al., 2013. Comparison of organic (UK'37, TEX86, LDI) and faunal proxies (foraminiferal assemblages) for reconstruction of late Quaternary sea-surface temperature variability from offshore southeastern Australia. *Paleoceanography*, 28, pp.1–11.
- Ludwig, W. et al., 2004. ARB: a software environment for sequence data. *Nucleic acids research*, 32(4), pp.1363–71.
- Marlowe, I.T. et al., 1990. Long-chain alkenones and alkyl alkenoates and the fossil coccolith record of marine sediments. *Chemical Geology*, 88(3-4), pp.349–375.
- Méjanelle, L. et al., 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), pp.527–538.
- Moernaut, J. et al., 2010. The seismic-stratigraphic record of lake-level fluctuations in Lake Challa: Hydrological stability and change in equatorial East Africa over the last 140kyr. *Earth and Planetary Science Letters*, 290(1-2), pp.214–223.
- Morris, R.J. & Brassell, S.C., 1988. Long-chain alkanediols: biological markers for cyanobacterial contributions to sediments. *Lipids*, 23(3), pp.256–258.
- Naafs, B.D. a., Hefter, J. & Stein, R., 2012. Application of the long chain diol index (LDI) paleothermometer to the early Pleistocene (MIS 96). *Organic Geochemistry*, 49, pp.83–85.
- 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), pp.803–814.
- Pearson, A. et al., 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), pp.3123–3137.
- Peterson, M.N.A., 2013. Calcite: Rates of Dissolution in a Vertical Profile in the Central Pacific. *Science*, 154(3756), pp.1542–1544.

- Pitcher, A. et al., 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), pp.3468–77.
- Potvin, M. & Lovejoy, C., 2009. PCR-Based Diversity Estimates of Artificial and Environmental 18S rRNA Gene Libraries. *Journal of Eukaryotic Microbiology*, 56(2), pp.174–181.
- Prahl, F.G. & Wakeham, S.G., 1987. Calibration of unsaturation patterns in long-chain ketone compositions for paleotemperature assessment. *Nature*, 330, pp.367–369.
- Přibyl, P. et al., 2012. Zoosporogenesis, Morphology, Ultrastructure, Pigment Composition, and Phylogenetic Position of *Trachydiscus Minutus* (Eustigmatophyceae, Heterokontophyta)1. *Journal of Phycology*, 48(1), pp.231–242.
- Quast, C. et al., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(Database issue), pp.D590–6.
- Rampen, S.W. et al., 2008. A 90 kyr upwelling record from the northwestern Indian Ocean using a novel long-chain diol index. *Earth and Planetary Science Letters*, 276(1-2), pp.207–213.
- Rampen, S.W. et al., 2009. Impact of temperature on long chain diol and mid-chain hydroxy methyl alkanolate composition in *Proboscia* diatoms: Results from culture and field studies. *Organic Geochemistry*, 40(11), pp.1124–1131.
- Rampen, S.W. et al., 2012. Long chain 1,13- and 1,15-diols as a potential proxy for palaeotemperature reconstruction. *Geochimica et Cosmochimica Acta*, 84, pp.204–216.
- Rampen, S.W. et al., 2007. Seasonal and spatial variation in the sources and fluxes of long chain diols and mid-chain hydroxy methyl alkanolates in the Arabian Sea. *Organic Geochemistry*, 38(2), pp.165–179.
- Rampen, S.W., Schouten, S. & Sinninghe Damsté, J.S., 2011. Occurrence of long chain 1,14-diols in *Apedinella radians*. *Organic Geochemistry*, 42(5), pp.572–574.
- Ries, J.B., 2004. Effect of ambient Mg/Ca ratio on Mg fractionation in calcareous marine invertebrates: A record of the oceanic Mg/Ca ratio over the Phanerozoic. *Geology*, 32(11), pp.981–984.
- Robinson, N. et al., 1986. Lipid geochemistry of Lake Kinneret. *Organic Geochemistry*, 10, pp.733–742.
- Romero-Viana, L., Kienel, U. & Sachse, D., 2012. Lipid biomarker signatures in a hypersaline lake on Isabel Island (Eastern Pacific) as a proxy for past rainfall anomaly (1942–2006AD). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 350-352, pp.49–61.
- Schouten, S. et al., 2002. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth and Planetary Science Letters*, 204, pp.265–274.

- Schouten, S., Hoefs, M.J.L. & Sinninghe Damsté, J., 2000. A molecular and stable carbon isotopic study of lipids in late Quaternary sediments from the Arabian Sea. *Organic Geochemistry*, 31, pp.509–521.
- Schouten, S., Hopmans, E.C. & Sinninghe Damsté, J.S., 2013. The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. *Organic Geochemistry*, 54, pp.19–61.
- Seki, O. et al., 2012. Paleoceanographic changes in the Eastern Equatorial Pacific over the last 10 Myr. *Paleoceanography*, 27(3), pp.1–14.
- Shanchun, J. et al., 1994. Origins and simulated thermal alteration of sterols and keto-alcohols in deep-sea marine sediments of the Okinawa Trough. *Organic Geochemistry*, 21(3), pp.415–422.
- Shimokawara, M. et al., 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), pp.753–766.
- Sinninghe Damsté, J.S. et al., 2003. A diatomaceous origin for long-chain diols and mid-chain hydroxy methyl alkanooates widely occurring in quaternary marine sediments: indicators for high-nutrient conditions. *Geochimica et Cosmochimica Acta*, 67(7), pp.1339–1348.
- Sluijs, A. et al., 2007. Environmental precursors to rapid light carbon injection at the Palaeocene/Eocene boundary. *Nature*, 450(7173), pp.1218–21.
- 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, pp.329–344.
- Smith, D.J., Eglinton, G. & Morris, R.J., 1983. Occurrence of Long-Chain Alkan-diols and Alkan-15-one-1-ols in a Quaternary Sapropel from the Eastern Mediterranean. *Lipids*, 18(12), pp.902–905.
- Spang, A. et al., 2010. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends in microbiology*, 18(8), pp.331–40.
- Spero, H.J. et al., 1997. Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes. *Nature*, 390, pp.497–501.
- Stoeck, T. et al., 2006. A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist*, 157(1), pp.31–43.
- Tamura, K. et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10), pp.2731–9.
- Tamura, K. & Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, 10(3), pp.512–26.
- Tegelaar, E.W. et al., 1989. A reappraisal of kerogen formation. *Geochimica et Cosmochimica Acta*, 53, pp.3103–3106.

- Thierstein, H.R. et al., 1977. Global synchronicity of late Quaternary coccolith datum levels: Validation by oxygen isotopes. *Geology*, 5, pp.400–404.
- 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), pp.4673–80.
- Urey, H.C., 1947. The Thermodynamic Properties of Isotopic Substances. *Journal of the Chemical Society*, pp.562–581.
- Versteegh, G.J.M. et al., 2000. Mid-chain diols and keto-ols in SE Atlantic sediments: A new tool for tracing past sea surface water masses? *Geochimica et Cosmochimica Acta*, 64(11), pp.1879–1892.
- Versteegh, G.J.M. & Blokker, P., 2004. Resistant macromolecules of extant and fossil microalgae. *Phycological Research*, 52, pp.325–339.
- 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), pp.1–13.
- Volkman, J.K. et al., 1992. C30-C32 alkyl diols and unsaturated alcohols in microalgae of the class Eustigmatophyceae. *Organic Geochemistry*, 18(1), pp.131–138.
- 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, pp.307–318.
- West, N.J. & Scanlan, D.J., 1999. Niche-Partitioning of Prochlorococcus Populations in a Stratified Water Column in the Eastern North Atlantic Ocean Niche-Partitioning of Prochlorococcus Populations in a Stratified Water Column in the Eastern North Atlantic Ocean †. *Applied and environmental microbiology*, 65(6), pp.2585–2591.
- Willmott, V. et al., 2010. Holocene changes in Proboscia diatom productivity in shelf waters of the north-western Antarctic Peninsula. *Antarctic Science*, 22(01), pp.3–10.
- Wuchter, C. et al., 2004. Temperature-dependent variation in the distribution of tetraether membrane lipids of marine Crenarchaeota: Implications for TEX 86 paleothermometry. *Paleoceanography*, 19(4), pp.1–10.
- Wuchter, C. et al., 2005. Temporal and spatial variation in tetraether membrane lipids of marine Crenarchaeota in particulate organic matter: Implications for TEX 86 paleothermometry. *Paleoceanography*, 20(3), pp.1–11.
- Xu, Y. & Jaffé, R., 2009. Geochemical record of anthropogenic impacts on Lake Valencia, Venezuela. *Applied Geochemistry*, 24(3), pp.411–418.
- Xu, Y., Simoneit, B.R.T. & Jaffé, R., 2007. Occurrence of long-chain n-alkenols, diols, keto-ols and sec-alkanols in a sediment core from a hypereutrophic, freshwater lake. *Organic Geochemistry*, 38(6), pp.870–883.

- Yamamoto, M. et al., 1996. Molecular paleontology of the earliest Danian at Geulhemmerberg (the Netherlands). *Geologie en Mijnbouw*, 75, pp.255–267.
- Yang, E.C. et al., 2012. Supermatrix data highlight the phylogenetic relationships of photosynthetic stramenopiles. *Protist*, 163, pp.217–231.
- Zachos, J.C., Dickens, G.R. & Zeebe, R.E., 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, 451(7176), pp.279–83.

Supplementary figures

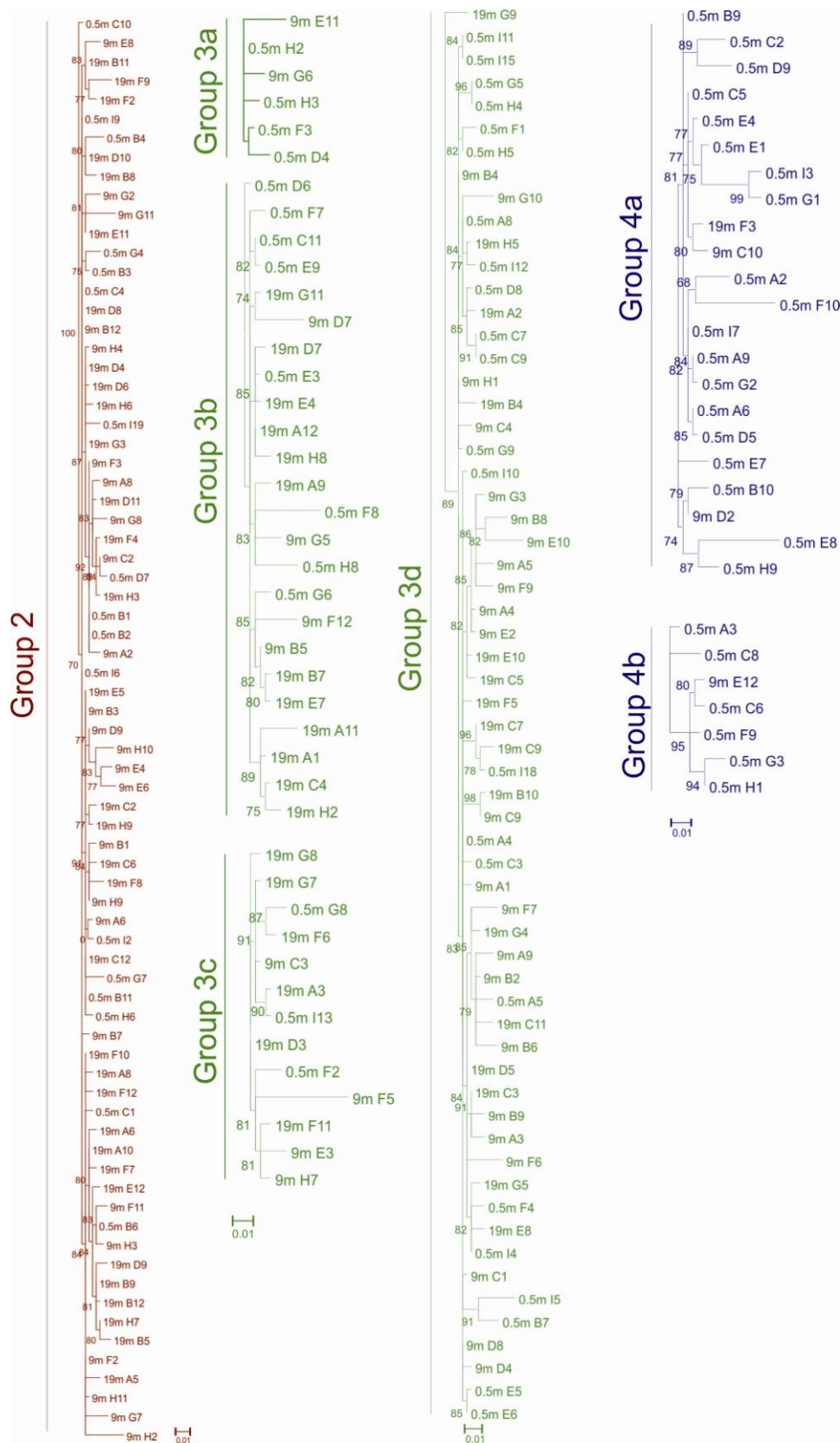


Figure S1. The 18S rRNA gene sequence names that are grouped in the phylogenetic tree of Lake Challa (Figure 28). Sequences are obtained from the SPM at three different depths in the water column (0.5, 9 and 19 meters).