Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

Biogeosciences Discussions
Open Access
EGU

A quest for the biological sources of the ubiquitous long chain alkyl diols in the marine realm.

Sergio Balzano[a*], Julie Lattaud[a], Laura Villanueva[a], Sebastiaan Rampen[a], Corina P.D. Brussaard[a], Judith van Bleijswijk[a], Nicole Bale[a], Jaap S. Sinninghe Damsté[a,b], Stefan Schouten[a,b]

[a] Department of Marine Microbiology and Biogeochemistry (MMB) and Utrecht University, NIOZ Royal Netherlands Institute for Sea Research, PO Box 59, 1790 AB Den Burg, The Netherlands

[b] Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands

* Correspondence to Sergio Balzano. E-mail address: Sergio.balzano@nioz.nl (S. Balzano).

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

Abstract

Long chain alkyl diols (LCDs) are widespread in the marine water column and sediments but their biological sources are mostly unknown. Here we combine lipid analyses with 18S rRNA gene amplicon sequencing on suspended particulate matter (SPM) collected in the photic zone of the tropical North Atlantic at 24 stations to infer relationships between LCDs and potential LCD-producers. The C_{30} 1,15-diol was detected in all SPM samples and accounted for >95 % of the total LCDs, while minor proportions of C_{28} and C_{30} 1,13-diols, C_{28} and C_{30} 1,14-diols as well as C_{32} 1,15-diol were found. The concentration of the C_{30} and C_{32} diols was higher in the mixed layer of the water column compared to the deep chlorophyll maximum (DCM), whereas concentrations of C_{28} diols were comparable. Sequencing analyses revealed extremely low contributions (≈ 0.1 % of the 18S rRNA gene reads) of known LCD-producers but the contributions from two taxonomic classes to which known producers are affiliated, i.e. Dictyochophyceae and Chrysophyceae, followed a trend similar to that of the concentrations of C_{30} and C_{32} diols. Statistical analyses indicated that the abundance of 4 operational taxonomic units (OTUs) of the Chrysophyceae and Dictyochophyceae, along with 23 OTUs falling in other phylogenetic groups, were significantly correlated with C_{30} diol concentrations. However, it is not clear whether some of these OTUs might indeed correspond to LCD-producers or whether these correlations are just indirect. Furthermore, based on the average LCD-content measured in cultivated LCD-producing algae, the detected concentrations of LCDs in SPM are too high to be explained by the abundances of the suspected LCD-producing OTUs. This is likely explained by the slower degradation of LCDs compared to DNA in the oxic water column and suggests that some of the LCDs found here were likely to be associated to suspended debris, while the DNA from the related LCD-producers had been already fully degraded. This suggests that care should be taken in constraining biological sources of relatively stable biomarker lipids by quantitative comparisons of DNA and lipid abundances.

Keywords: long chain alkyl diols, HCC, tropical North Atlantic, 18S rRNA gene amplicon sequencing, Eustigmatophyceae, Chrysophyceae, Dictyochophyceae

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

1. INTRODUCTION

Long chain alkyl diols (LCDs) are lipids that consist of a linear alkyl chain with 22–38 carbons which is hydroxylated at the terminal carbon atom and at an intermediate position, and are usually saturated or monounsaturated. LCDs were identified for the first time in Black Sea sediments (de Leeuw et al., 1981) and have subsequently been found to occur widespread in both suspended particulate matter (SPM) and sediments from both coastal and off-shore sites throughout the World Ocean (Jiang et al., 1994; Versteegh et al., 1997; Rampen et al., 2014b). LCDs can be preserved in marine sediments for long periods of time and their distribution can reflect the environmental conditions at the time they were produced. Several indices, based on ratios between the different diols, have been proposed for the reconstruction of past environmental conditions. The Diol Index, reflecting the proportion of C_{28} and C_{30} 1,14-diols over the sum of C_{28} and C_{30} 1,14-diols and C_{30} 1,15-diol, has been proposed to track ancient upwelling conditions since the 1,14-diols are believed to be mostly produced by upwelling diatoms of the genus *Proboscia* (Rampen et al., 2008). Another index, the long chain diol index (LDI), which is based on the proportion of the C_{30} 1,15-diol over the C_{28} and C_{30} 1,13-diols, shows a strong correlation with sea surface temperature (SST) and is used to determine past SST (Plancq et al., 2014; Rampen et al., 2012; Rodrigo-Gámiz et al., 2015). In addition, since the C_{32} 1,15-diol is the major component of the LCDs of freshwater Eustigmatophyceae (Rampen et al., 2014a; Volkman et al., 1992), the fractional abundance of C_{32} 1,15-diol has been suggested to be a marker of riverine input in seawater (Lattaud et al., 2017b; de Bar et al., 2016; Lattaud et al., 2017a). Other markers for riverine inputs in seawater are the $C_{30\text{-}36}$ 1,ω20-diols which are produced by the freshwater fern *Azolla* (Speelman et al., 2009; Mao et al., 2017). However, application of these proxies in the marine realm remains uncertain as long as the sources of the major marine diols are unknown.

The most abundant LCDs in seawater are the saturated C_{28} and C_{30} 1,13-diols, C_{28} and C_{30} 1,14-diols, and C_{30} and C_{32} 1,15-diols (Rampen et al., 2014b), which are all likely produced by phytoplankton. However, the marine biological sources of LCDs are still not fully clear because, in contrast with the widespread occurrence of LCDs in the sediment, few marine taxa have been shown to contain these lipids. Eustigmatophyceae contain C_{30} 1,13-, C_{30} 1,15-, and C_{32} 1,15-diols (Rampen et al., 2014a; Volkman et al., 1992) but they comprise mostly freshwater species and only a few rare marine representatives from the genus *Nannochloropsis* are known (Fawley and Fawley, 2007; Andersen et al., 1998). Furthermore, the distribution of LCDs in the marine environment does not match that of LCDs of marine Eustigmatophytes (Rampen et al., 2012). Species of the diatom genera *Proboscia* and the

dictyocophycean *Apedinella radians* contain C_{28-32} 1,14-diols (Sinninghe Damsté et al., 2003; Rampen et al., 2009; Rampen et al., 2011), with the former accounting for significant proportions of marine biomass mostly in upwelling regions (Moita et al., 2003; Lassiter et al., 2006), whereas the latter has been occasionally observed in estuarine environments (Bergesch et al., 2008; Seoane et al., 2005). Few other marine species from classes genetically related to diatoms and Eustigmatophyceae have been recently shown to produce LCDs (Table S1). All the known LCD-producing phytoplankters belong to the eukaryotic supergroup Heterokontophyta, a division which includes, among others, diatoms and brown seaweeds. The widespread occurrence of LCDs in the marine environment, in spite of the restricted abundance and distribution of marine LCD-producers, suggests that these compounds may be produced by unknown phytoplankton species.

A way of assessing the sources of biomarker lipids is to compare the abundance of lipids in environmental samples with the composition of the microbial community, as determined by genetic methods. For example, Villanueva et al. (2014) analysed both LCDs and eustigmatophycean 18S rRNA gene sequences in a tropical freshwater lake and found five clades of uncultured Eustigmatophyceae in the top 25 m of the water column of the lake, where LCDs were also abundant. Abundance determination by quantitative polymerase chain reaction (qPCR) highlighted that the number of eustigmatophycean 18S rRNA gene copies peaked at the same depth as the LCDs, suggesting that Eustigmatophyceae are a primary source for LCDs in freshwater (Villanueva et al., 2014). However, one of the limitations of this approach is that it relies on specific eustigmatophycean primers designed based on the sequences available in the genetic databases, which could be biased and not target all the existing LCD biological sources. To compensate for this limitation high throughput amplicon sequencing of the 18S rRNA gene allows the exploration of the total marine microbial communities in deep details (de Vargas et al., 2015; Balzano et al., 2015; Christaki et al., 2014; Logares et al., 2012; Massana et al., 2015; Stoeck et al., 2009) and the combination of this analyses with lipid composition may potentially assist in identifying the main LCD producers in marine settings.

In the present study, we quantitatively analysed the composition and abundance of LCDs in suspended particulate matter (SPM) collected along the tropical North Atlantic (Fig. 1A) at different depths in the photic zone (surface, deep chlorophyll maximum and bottom of the wind mixed layer; see also Bale et al., 2018). The 18S rRNA gene abundance and composition of the SPM was also analysed by quantitative PCR (qPCR) and high throughput amplicon sequencing to infer the taxonomic composition and to compare the abundance of the

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

different taxa with that of the LCDs in order to identify the potential marine biological sources of LCDs.

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

2. MATERIAL AND METHODS

2.1 Cruise transect, ancillary data, and SPM collection

Samples were taken during the HCC cruise (64PE393), which took place from 24th August to 21st September 2014 along a transect on the tropical North Atlantic Ocean (see Bale et al. (2018) for details). The transect was from Mindelo (Cape Verde) to a location about 500 km from the Amazon River mouth and then westwards along the coast towards Barbados (Fig. 1A). Temperature, salinity and nutrient data have previously been reported in Bale et *al.* (2018).

Seawater was collected from two or three depths at each station to measure the concentration of chlorophyll *a* (Chl-*a*) and the abundances of photosynthetic pico and nanoeukaryotes. Seawater was collected during the up cast using Niskin bottles mounted on a CTD frame. The sampling depths were determined based on the evaluation of the vertical profiles of temperature, salinity, and Chl-*a* fluorescence after the down cast of the CTD deployment. The depth of the bottom wind mixed layer (BWML) and the deep chlorophyll maximum (DCM) were determined based on the lowest position of the mixed layer and the depth at which the highest values of chlorophyll fluorescence were observed. For Chl-*a* determination seawater was collected from the Niskin bottles and filtered through 0.7 µm pore-size glass-fiber (Whatman GF/F) filters, followed by frozen storage. Chl-*a* was extracted with methanol buffered with 0.5 M ammonium acetate, homogenized for 15 s and analysed by high performance liquid chromatography.

Photosynthetic pico- and nanoeukaryotes were enumerated by flow cytometry according to the protocol by Marie et al. (2005). In short, 1 mL samples were counted fresh using a Becton-Dickinson FACSCalibur (Erembodegem, Belgium) flow cytometer equipped with an air-cooled Argon laser (488 nm, 15 mW). Phytoplankton were discriminated based on their chlorophyll *a* autofluorescence and scatter signature. Cyanobacteria, i.e. *Synechococcus* and *Prochlorococcus*, were not taken into account for the current study. Size fractionation was performed by gravity filtration with >3 µm average cell diameter phytoplankton groups classified as nanoeukaryotic and those <3 µm average cell diameter as picoeukaryotic phytoplankton.

Three McLane *in situ* pumps (McLane Laboratories Inc., Falmouth) were used to collect SPM from the water column for the analysis of both lipids and microbial communities. Similarly to the seawater collected with Niskin bottles for Chl-*a* and flow cytometry analyses,

the *in situ* pumps were deployed at the surface (3 - 5 m depth), the BWML and the DCM (Table S2). Between 100 and 400 L seawater were pumped and the SPM was collected on pre-combusted 0.7 µm GF/F filters (Pall Corporation, Washington) and immediately frozen at –80°C. For the determination of the organic carbon concentrations, SPM was freeze dried and analysis was carried out using a Flash 2000 series Elemental Analyzer (Thermo Scientific) equipped with a thermal conductivity detector.

2.2 Lipid extraction and LCD analyses

Lipids were extracted from the GF/F filters as described previously (Lattaud et al., 2017b). Briefly, ¼ of the filters were dried using a LyoQuest (Telstart, Life Sciences) freeze-dryer and lipids were extracted using base and acid hydrolysis. The base hydrolysis was achieved with 12 mL of a 1 N KOH in methanol (MeOH) solution by refluxing for 1 h. Subsequently, the pH was adjusted to 4 with 2 M HCl: MeOH (1:1, v/v) and the extract was transferred into a separatory funnel. The residues were further extracted once with MeOH: H_2O (1 : 1, v/v), twice with MeOH, and three times with dichloromethane (DCM). The extracts were combined in the separatory funnel and bidistilled water (6 mL) was added. The combined solutions were mixed, shaken and separated into a MeOH : H_2O and a DCM phase, the DCM phase was removed and collected into a centrifuge tube. The aqueous layer was re-extracted twice with 3 mL DCM. The pooled DCM layers were dried over a Na_2SO_4 column and the DCM was evaporated under a stream of nitrogen. The extract was then acid hydrolyzed with 2 mL of 1.5 N HCl in MeOH solution under reflux for 2 h. The pH was adjusted to 4 by adding 2 N KOH : MeOH. 2 mL of DCM and 2 mL of bidistilled water were added to the hydrolyzed extract, mixed and shaken and, after phase separation, the DCM layer was transferred into another centrifuge tube. The remaining aqueous layer was washed twice with 2 mL DCM. The combined DCM layers were dried over a Na_2SO_4 column, the DCM was evaporated under a stream of nitrogen and a C_{22} 5,17-diol was added to the extract as internal standard. The extract was separated on an activated aluminium oxide column into three fractions using the following solvents: hexane: DCM (9: 1, v/v), hexane: DCM (1: 1, v/v) and DCM: MeOH (1: 1, v/v). The latter (polar) fraction containing the diols was dried under a gentle nitrogen stream. Diols were derivatized by silylating an aliquot of the polar fraction with 10 µL N,O-Bis(trimethylsilyl) trifluoroacetamide (BSTFA) and 10 µL pyridine, heating for 30 min at 60 ºC and adding 30 µL of ethyl acetate. The analysis of diols was performed by gas chromatography-mass spectrometry (GC-MS) using an Agilent 7990B GC gas chromatograph, equipped with a fused silica capillary column (2 5m x 320 µm) coated

with CP Sil-5 (film thickness 0.12 µm), coupled to an Agilent 5977A MSD mass spectrometer. The temperature regime for the oven was the same as that used by Lattaud et al. (2017b). The diols were identified and quantified via SIM (Single Ion Monitoring) of the m/z= 299.3 (C_{28} 1,14-diol), 313.3 (C_{28} 1,13-diol, C_{30} 1,15-diol), 327.3 (C_{30} 1,14-diol) and 341.3 (C_{30} 1,13-diol, C_{32} 1,15-diol) ions (Versteegh et al., 1997; Rampen et al., 2012). Absolute concentrations were calculated using the peak area of the internal standard as a reference.

2.3 DNA extraction, PCR, qPCR, and 18S rRNA gene sequencing

On ice a small portion of the GF/F filters, corresponding to 1/16 of their initial size, hence containing SPM from ca. 25 L of seawater, was cut in many small pieces using sterile scissors and tweezers. Filter pieces were then transferred into 2 mL microtubes and the DNA was extracted using a MOBIO powersoil DNA isolation kit (Qiagen) following manufacturer instructions. We amplified the hypervariable V4 region of the 18S rRNA which is considered the best genetic marker for the identification of microbial eukaryotes (Logares et al., 2012; Massana et al., 2015). The V4 is located in a central region (565-584 bp to 964-981 bp for *Saccharomyces cerevisiae*) of the 18S rRNA and it was amplified from the genomic DNA by PCR using the universal eukaryotic primers TAReuk454FWD1 (5'-CCAGCASCYGCGGTAATTCC-3') and TAReuk454REV3 (5'-ACTTTCGTTCTTGATYRA-3') (Stoeck et al., 2010). Primers were modified for multiplex sequencing on a Roche 454 GS FLX system: a 454-adapter A (CCATCTCATCCCTGCGTGTCTCCGACTCAG), a key (TCAG), and a 10 bp sample-specific Multiple Identifier (MID, Table S3) were bound to the 5' end of the forward primer, whereas a 454-adapter 2 (CCTATCCCCTGTGTGCCTTGGCAGTCTCAG) and a unique MID (CGTGTCA) were bound to the 5' end of the reverse primer for all the samples. The PCR mixture included 25 µL Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher Scientific) 19.1 µL deionised water, 1.5 µL Dimethyl sulfoxide, 1.7 µL from each primer and 25 ng genomic DNA and the V4 region was amplified using the same thermal cycling as described by Logares et al. (2012). Amplicons were visualised on a 1% agarose gel and V4 bands were excised and subsequently purified using a QIAquick Gel Extraction Kit (Qiagen) and DNA concentration was measured by Qubit Fluorometric Quantitation (ThermoFisher Scientific). For each sequencing run, 20 samples were pooled in equimolar amount and sequenced using a 454 GS-FLX Plus (Macrogen Korea). Some samples yielded a low

number of reads and were re-sequenced; overall 77 samples were sequenced in 5 sequencing runs.

To determine the concentration of total 18S rRNA genes within the seawater sampled we carried out qPCR using the same primers and the same cycling conditions as described above. qPCR analysis was performed on a Biorad CFX96TM Real-Time System/C1000 Thermal cycler equipped with CFX Manager™ Software. Abundance of 18S rRNA gene sequences was determined with the same primer pair (TAReuk454FWD1/ TAReuk454REV3) used for the 18S rRNA gene diversity analysis. Each reaction contained 12.5 µL MasterMix phusion, 8.25 µL deionised nuclease-free water, 0.75 µL DMSO, 1 µL from each primer and 0.5 µL Sybr green and 1 µL of DNA template. Reactions were performed in iCycler iQTM 96-well plates (Bio-Rad). A mixture of V4 18S rRNA gene amplicons obtained as described above was used to prepare standard solutions. All qPCR reactions were performed in triplicate with standard curves from 6.4 x 10^3 to 6.4 x 10^9 V4 molecules per microliter. Specificity of the qPCR was verified with melting curve analyses (50 °C to 95 °C).

2.4 Bioinformatic analyses

Bioinformatic analyses were carried out using the python based bioinformatic pipeline quantitative insight in microbial ecology (QIIME) (Caporaso et al., 2010). Overall we obtained 372 107 raw sequences; reads with a length comprised between 250 and 500 bp, less than 8 homopolymers, and a phred quality ≥ 25 over 50 bp sliding windows were kept for downstream analyses. Chimeric sequences were then identified by comparison with the Protist Ribosomal Database 2 (PR2) (Guillou et al., 2013)using the Uchime algorithm (Edgar et al., 2011) and removed from the dataset along with singletons (i.e. reads not sharing 100 % identity with at least another read).

A total of 238 564 reads remaining after quality filtering were clustered into 2457 Operational Taxonomic Units (OTUs) based on 95 % sequence identity using Uclust (Edgar, 2010). Samples containing less than 1000 sequencing reads were removed from the dataset. The taxonomic affiliation of the OTUs was then inferred by comparison with the PR2 (Guillou et al., 2013) using BLAST (Altschul et al., 1990) within the QIIME pipeline. Reads from metazoa and multicellular fungi were removed from the dataset which finally contained 1871 OTUs and 184 279 reads. The abundance of the different taxa in each sample were estimated by multiplying the percentage of reads with the concentration of V4 copies measured by qPCR. Taxa containing LCD-producers were extracted from the dataset and plotted using Ocean Data View (Schlitzer, 2017).

248

249 **2.5 Statistical analyses**

250 Linear regression analyses between the concentrations of the different LCDs were
251 performed to assess whether some of the LCDs were likely to derive from a common source.
252 To investigate relationships between LCDs and environmental conditions we calculated the
253 Spearman rank correlation coefficient (ρ) using the R package vegan (Dixon, 2003). The
254 environmental data used were temperature, salinity, TOC, nutrients (nitrate, nitrite,
255 ammonium, phosphate, and silica), as well the concentration of Chl-*a* and the abundance of
256 photosynthetic pico and nanoeukaryotes. Samples containing missing data and outliers were
257 removed from the dataset before the calculations. Both correlation coefficients and p-values
258 were calculated and the latter were corrected for False Discovery Rates (Benjamini and
259 Hochberg, 1995). Correlations were considered significant for p-values <0.01.
260 To investigate the relationships between lipids and microbial taxa we also calculated the
261 Spearman's rank correlation coefficient between the LCD concentrations and the abundance
262 of the different taxa at both OTU and class levels. To this end, taxonomic data were
263 normalized based on the number of V4 copies in the different samples measured by qPCR.
264 Comparisons at class level provide the advantage of pooling distribution data from several
265 closely-related OTUs reducing the number of zeros (samples where a given OTU is absent),
266 as the latter complicate statistical analyses of biological distributions (Legendre and
267 Gallagher, 2001). However pooling OTUs at higher taxonomic levels likely leads to
268 combining the LCD-producers with species falling in the same taxonomic level but which are
269 unable to produce LCDs. We thus removed OTUs observed in fewer than 19 samples (25 %)
270 and compared the resulting OTU table with the LCD concentrations. These analyses were
271 performed using the qiime script *observation_metadata_correlation.py* (Caporaso et al.,
272 2010) and the p-values were corrected for false discovery rates (Benjamini and Hochberg,
273 1995).

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

3. RESULTS

3.1 Ancillary data

The HCC cruise sailed across tropical Atlantic waters (Fig. 1A) in late summer and was targeted at SPM from the photic zone collected at the surface, the bottom water mixed layer (BWML) and the deep chlorophyll maximum (DCM). The extent of the photic zone as well as the depths of both BWML and DCM at each station were assessed based on the vertical profiles of temperature, salinity and chlorophyll fluorescence. Temperature of photic zone waters ranged between 15-29 °C (Fig. 1B), the BWML depth ranged between 9-40 m, whereas the depth of the DCM ranged between 45-105 m. Temperatures varied at the DCM increasing westwards, whereas they were relatively constant at surface and BWML. Salinity varied between 29 and 36.5 PSU (Fig. 1C) at the surface, whereas it was fairly constant in the DCM (36 to 37). The concentration of Chl-*a* varied from 34 to 470 ng L^{-1} (Fig. 1D), with the lowest values measured at the surface of the easternmost (1 to 6) and westernmost (21 to 23) stations and the relatively higher concentrations in surface waters of the shallowest stations (11 to 13) located above the continental shelf and about 500 km off the Amazon River mouth (Fig. 1A). The POC concentration ranged between 0.6 and 13 mg L^{-1} and also peaked at surface for the shallowest stations (Fig. 1E).

Photosynthetic picoeukaryotes, quantified by flow cytometry, were more abundant at the DCM compared to surface and BWML (Fig. 1F). Their abundance peaked at the DCM of Stations 1 and 2 (>1.5 x 10^7 cell L^{-1}), whereas for surface waters the highest values were measured at Stations 11 to 13. In contrast, photosynthetic nanoeukaryotes did not vary substantially through the water column and their abundance peaked at the surface of Stn. 17 reaching a density of 1.4 x 10^5 cell L^{-1} (Fig. 1G).

3.2 Long chain alkyl diols

Six long chain diols were detected, the C_{28} and C_{30} 1,13-diols, C_{28} and C_{30} 1,14-diols and the C_{30} and C_{32} 1,15-diols (Fig. 2). The C_{30} 1,15-diol dominated all samples, accounting for >95 % of the total LCDs, and its concentration ranged from 100 to 1600 pg L^{-1}. The concentration of the C_{28} 1,13-diol ranged from 0 (i.e. undetectable) to 55 pg L^{-1}, whereas the highest concentration measured for the C_{28} 1,14-diol was 64 pg L^{-1}. The other minor diols were usually more abundant than the C_{28} diols, reaching concentrations up to 190 pg L^{-1} for the C_{30} 1,13-diol, 240 pg L^{-1} for the C_{30} 1,14-diol, and 480 pg L^{-1} for the C_{32} 1,15-diol (Fig. 2).

The concentration of the C_{28} 1,13-diol peaked in the surface waters of Station 10 but it was below the detection limit in 19 samples from different depths and stations (Fig. 2A). The C_{28} 1,14-diol reached its highest concentrations at the DCM of Station 12 (64 pg L^{-1}) and at the surface of Station 13 (45 pg L^{-1}) and tended to be more abundant in the waters of the eastern stations (Fig. 2B). The concentrations of both C_{28} 1,13- and C_{28} 1,14-diols did not vary significantly with depth (t-test, p-value >0.1), while those of the C_{30} 1,13-, C_{30} 1,14-, and C_{30} 1,15-diols were higher in the mixed layer (surface and BWML) compared to the DCM (p-value <0.01).

The concentration of the C_{30} 1,13-diol peaked at the surface of Stations 10 and 14 (Fig. 2C), while that of the C_{30} 1,14-diol reached its maximum at the BWML of Stations 7 and 8 (Fig. 2D). The highest concentration of the C_{30} 1,15-diol was measured at the surface of Station 17 (16 ng L^{-1}, Fig. 2E). The concentration of the C_{32} 1,15-diol peaked in the surface waters of Station 10 and 14 and at the DCM of Station 7 (Fig. 2F) and its concentration did not vary significantly with depth. The concentrations of both the C_{30} and C_{32} diols peaked in the mixed layer of Stations 7-10 and 14-17, which are located in close proximity to the Amazon Shelf (Figs. 2C-F).

3.3 Eukaryotic 18S rRNA gene diversity analysis

Sequencing of the hypervariable V4 region of the 18S rRNA gene of 68 SPM samples resulted in 238 564 reads with an average of 4 987 reads per sample (Table S2). Reads were clustered based on 95% sequence identity and, after removal of reads of metazoa and multicellular fungi, we obtained 1871 operational taxonomic units (OTUs). Rarefaction analyses indicate that >90 % of the genetic diversity was captured (Figure S1), suggesting that no sample was undersequenced. Most (>90 %) reads sequenced here were assigned to Dinophyceae, Syndiniales, Metazoa, Haptophyta, and Radiolaria (Fig. 3). Samples were grouped according to the depth layer (surface, BWML, and DCM) and analysis of similarity (anosim) revealed that the average variance between samples from different groups was higher than the average variance between samples from the same group (p-value ≈ 0.001), indicating that the eukaryotic community was mostly influenced by the water depth rather than the geographic location. The proportion of reads from Dinophyceae, Syndiniales, and Haptophyta was slightly higher in the mixed layer compared to the DCM, whereas Radiolaria and Pelagophyceae tended to be slightly more abundant in deeper waters (Fig. 3). All samples except surface waters from Station 12, the BWML from Station 11 and the DCM from Station 22 exhibited high contributions (>50 %) from Dinophyceae and Syndiniales (Supplementary

Figure S2). Radiolaria dominated the DCM at Station 22, diatoms were relatively abundant (≈ 10-20 %) at the surface of Stations 12-14 and the BWML of Station 12 while the contribution of diatom reads was <5 % for all the other samples.

18S rRNA gene reads of only four taxa containing known LCD-producers were detected within our dataset: *Proboscia* spp., Florenciellales, *Heterosigma* spp., and Eustigmatophyceae (Table 1). In 33 out of 68 SPM samples we did not detect any 18S rRNA gene read from known LCD-producers, whereas reads from these taxa accounted for <0.1 % of the total 18S rRNA reads in 24 samples, 0.1 to 0.5 % in 8 samples, 0.5 to 1 % in 2 samples and 1.5 % in one sample (Stn. 20, BWML). The 18S rRNA gene reads from putative LCD-producers were mostly recovered from the mixed layer (Table 1). Florenciellales was the most abundant taxon among the known LCD-producers since it exhibited the highest number of reads (99) and was present in 28 out of 68 samples. The other taxa of putative LCD-producers were detected only in 8 (Eustigmatophyceae) or 2 (*Proboscia* sp. and *Heterosigma akashiwo*) samples (Table 1) accounting from 3 (*Proboscia*) to 45 (Eustigmatophyceae) reads. The Eustigmatophyceae (mostly affiliated to *Nannochloropsis oculata*) were found at surface for the Stations 11, 12, and 13, as well as at the DCM of Station 20 (Fig. 4A).

Since species genetically related to cultivated microalgae known to produce LCD may also contain LCDs, we expanded our community composition analyses to groups at a higher taxonomic level, and focused on those classes or divisions that contain LCD-producers (Table S1). Specifically we investigated the distribution of Eustigmatophyceae, since they are the most well-known class of LCD-producers, Pelagophyceae and Chrysophyceae, which include the LCD-producers *Sarcinochrysis marina* and *Chrysosphaera parvula*, respectively (Table S1), Dictyochophyceae, which includes *Apedinella radians* (Rampen et al., 2011), and Raphidophyceae, which include two LCD-producers, *H. akashiwo* and *Haramonas dimorpha* (Table S1). We did not detect any representative of Pinguiophyceae, a class which include the LCD producer *Phaeomonas parva* (Table S1). Reads associated to Pelagophyceae, and mostly (97 %) affiliated to *Pelagomonas calceolata*, were recovered more frequently as they were present in 55 samples with an average abundance of 85 reads (2 % of total reads) per sample and a maximum value of 935 reads (12 % of total) in the DCM of Stn. 23 (Fig. 4B). Pelagophyceae reads were mostly detected in the DCM and were particularly abundant at the 3 westernmost stations investigated, where they comprised 8 % of total reads (Fig. 4B).

Chrysophyceae and Dictyochophyceae were also detected in most samples (54 and 57 samples, respectively) and their reads were recovered more frequently at the surface and BWML of the westernmost part of the transect (Stns. 20-23) and at the surface of Stations 3-4

(Fig. 4C and D). Their 18S rRNA gene reads reached abundances of up to 55 and 41 reads (0.4 and 0.6 % of total, respectively), for Chrysophyceae and Dictyochophyceae respectively, in the BWML of Station 20 (Table S4). Raphidophyceae were present only in three samples from Stations 11, 12, and 13 (Fig. 4F).

4. DISCUSSION

4.1. Comparison of diol distributions

In general, it is thought that 1,13- and 1,15-diols derive from a different source than 1,14-diols in the marine realm (Sinninghe Damsté et al., 2003; Rampen et al., 2007; Rampen et al., 2011). Indeed, linear regressions showed that the concentration of C_{30} 1,15-diol is significantly correlated with those of the C_{30} 1,13- and C_{32} 1,15-diols (Figs. 5A-B). We did not observe any significant correlation between the concentrations of the C_{28} 1,13- and the C_{30} 1,13- or C_{30} 1,15-diol (Figs. 5C-D), which might be due to the fact that C_{28} 1,13-diol was only present in low abundance and below detection limit in 19 out of 71 samples. This low abundance of C_{28} 1,13-diol is consistent with the relatively high temperatures observed for the tropical Atlantic ocean (Fig. 1B), since the LCD core top calibration study has revealed that the fractional abundance of the C_{30} 1,15-diol is high and that of the C_{28} 1,13-diol is low when SST is relatively high (Rampen et al., 2012).

The concentration of the C_{28} 1,14-diol was not correlated to that of the C_{30} 1,14-diol (Fig. 5E), potentially suggesting a different origin for the C_{28} and the C_{30} 1,14-diols. However, the concentration of C_{30} 1,14-diol was significantly correlated to the C_{30} 1,15-diol (Fig. 5F). This is quite surprising as the 1,14-diols in seawater have been suggested to derive from *Proboscia* spp. (Sinninghe Damsté et al., 2003; Rampen et al., 2009), and to a lesser extent from *A. radians* (Rampen et al., 2011) whereas the 1,13- and 1,15-diols are thought to be associated with Eustigmatophyceae (Rampen et al., 2014 and references cited therein). Previous studies highlighted indeed good correlations in the fluxes of C_{28} and C_{30} 1,14-diols in the water column of the Arabian Sea (Rampen et al., 2007) and the northwestern Indian Ocean (Rampen et al., 2008). *Proboscia* spp. contain also unsaturated 1,14-diols which were not found here; specifically the warm water species *Proboscia indica* is dominated by $C_{28:1}$ and $C_{30:1}$ 1,14-diols (Rampen et al., 2007) suggesting that the 1,14-diols found here do not derive from *Proboscia* spp.. This is confirmed by the absence or very low proportions of 18S rRNA gene reads from the major producers of $C_{28\text{-}30}$ 1,14 diols, that are *Proboscia* spp. and *A. radians* (Table 1). This suggests two different sources for the C_{28} and the C_{30} 1,14-diols.

Since the C_{30} 1,15-diol accounted for >95 % of the LCDs, it is possible that the C_{30} 1,14-diol was biosynthesised in low amounts, along with C_{30} 1,13-diol, by the producers of C_{30} 1,15-diol. This is supported by the fact that Eustigmatophyceae can contain small amounts of 1,14-diols along with large quantities of 1,15 diols (Rampen et al., 2014a): specifically the C_{28} 1,14 diol, accounts for up to the 15% of the total LCDs in *Pseudostarastrum enorme*, and lower proportions (1-5%) of C_{30} 1,14 diols were previously found in *Vischeria punctata* and *Eustigmatos vischeri* (Rampen et al., 2014a).

It has been found that the distributions of LCDs can be affected by riverine input, which is reflected by elevated amounts of the C_{32} 1,15-diol (>10%, de Bar et al., 2016; Lattaud et al. 2017b). However, the fractional abundance of the C_{32} 1,15-diol in the SPM is low (0 to 4 %, data not shown), far lower than the values typically measured in river-influenced ecosystems such as the Iberian Atlantic Margin (de Bar et al., 2016), the Kara Sea (Lattaud et al., 2017b) or the Congo River plume (Versteegh et al., 2000). The HCC cruise took place in a period of the year (August/September) when the water discharge from the Amazon River is typically low (Molleri et al., 2010), thus leading to low inputs of riverine organic matter into the sea. The LCDs in the sampled SPM are thus likely not impacted by terrestrial input of LCDs.

Beyond Heterokontophyta, LCDs may also be produced by lower (Speelman et al., 2009) and higher (Wen and Jetter, 2007; Racovita and Jetter, 2016) plants. However, only 4 reads from our dataset were associated with a plant species, i.e. *Panax ginseng* (Table S4), which is not known to contain LCDs. The near absence of 18S rRNA gene reads from higher plants confirms the low riverine input of organic matter in the SPM of the tropical North Atlantic waters analysed here.

We explored the variations in the concentrations of LCDs with respect to environmental data. The C_{28} 1,13- and 1,14-diols, both occurring in low abundance, did not exhibit significant correlations with any of the environmental data measured here (Table 2). In contrast the concentrations of C_{30} 1,13-, 1,14- and 1,15-diols exhibited significant but weak positive correlations with temperature and dissolved silica and weak negative correlations with salinity and nitrite. The concentration of the C_{32} 1,15-diol revealed a correlation with the same environmental variables as the C_{30} diols except for dissolved silica and nitrite and exhibited a weak negative correlation with the concentration of nitrate. The correlations found here are likely simply due to different water masses: the mixed layer, where the highest proportions of LCDs were measured, exhibited indeed higher temperatures and lower salinities compared to the DCM. We repeated the analyses after excluding DCM samples and did not find strong positive or negative correlations between LCDs and environmental

variables (data not shown). Thus, there does not seem to be a major control of environmental conditions on the concentrations of LCDs.

4.2 Comparison with eukaryotic abundance and diversity

Although LCDs are likely produced by phytoplankton, the variability in LCD abundance is not correlated to that of Chl-*a* concentration, photosynthetic pico- and nanoeukaryotes (Table 2). This lack of correlation suggests that the LCD-producers found here accounted for only a small proportion of phytoplankton. The high proportion of Dinophyceae, Syndiniales, and Radiolaria found within our genetic libraries agree with previous studies on marine microbial communities based on 18S rRNA gene sequencing in different environments (Christaki et al., 2014; Comeau et al., 2011; de Vargas et al., 2015). However, these taxa do not necessarily dominate marine microbial communities and so our results are likely due to a relatively high number of rRNA gene copies per cell (Zhu et al., 2005). Larger-sized dinoflagellates such as *Prorocentrum minimum* and *Amphidinium carterae* can contain up to 1000 gene copies per cell compared to <10 of rRNA gene copies for smaller sized (<3 µm) species of Chlorophyta, Pelagophyceae, and Haptophyta (Zhu et al., 2005).

4.2.1 LCD-producers

Although the primers used in this study have a perfect match with the 18S sequences of most eukaryotes (including all the classes containing LCD-producers), and the rarefaction curves indicate that we sampled an appropriate (i.e. >90%) proportion of the eukaryotic community, the large number (100-1000) of rRNA gene copies per cell present within dinoflagellates and Radiolaria might have somehow affected the detection of LCD producers. In particular *Nannochloropsis salina* has been shown to possess only 1-2 copies of 18S rRNA gene (Zhu et al., 2005), and similarly, the other marine *Nannochloropsis* species which do not differ greatly in size from *N. salina* (Fawley and Fawley, 2007) are also likely to have a low number of 18S rRNA gene copies. Known species of LCD-producers were present in only 51 % of our samples as revealed by sequencing data (Table 1), whereas the major LCD, the C_{30} 1,15-diol, was present in all the samples. This suggests that the LCDs found here were (1) either produced by other species which were not detected using the current methodology or (2) that the LCD-producers were undersampled because of their low number of rRNA gene copies per cell, or (3) that the DNA of the LCD-producers was no longer present in the SPM at the moment of sampling. Specifically, marine Eustigmatophyceae were represented by only two OTUs (denovo2075, *Nannochloropsis oculata*, and denovo229, uncultured

476 Eustigmatophycea, Table S4) detected in only 8 samples, confirming the hypothesis of
477 Volkman et al. (1999) and Rampen et al (2012) that they are not the major producers of LCDs
478 in the marine environment. Even if we expand our analyses of LCD-related species to a
479 higher taxonomic level, we do not find large proportions of 18S rRNA reads (generally <0.9
480 % of total reads) except for the class Pelagophyceae, which accounts for up to 12 % of total
481 reads (Fig. 2A-E). However, Pelagophyceae are unlikely to be the source of any of the LCDs
482 found here because their vertical distribution does not correspond well to that of LCDs, which
483 were either more abundant in the upper layers (C_{30} 1,13-, 1,14-, and 1,15-diols and C_{32} 1,15-
484 diol) or did not vary greatly with depth (C_{28} diols, Fig. 2). Chrysophyceae and
485 Dictyochophyceae were instead more abundant in the upper layers (Fig. 4B-C) and although
486 none of the three known LCD-producers from these taxa (Table S1) seems to contain the C_{30}
487 1,15-diol, which was dominant here, species within the Chrysophyceae and
488 Dictyochophyceae may possibly be a source for the C_{30} diols.
489 The C_{28} diols exhibited higher concentrations at the BWML of Station 12 and at surface
490 in Stn. 13 (Fig. 2A and B), and higher proportions of 18S rRNA gene reads were recovered
491 from Pelagophyceae (2.4 %), and Eustigmatophyceae (0.5 %), at the surface of Stations 11-12
492 (Fig. 4D-F). The scattered occurrences of these groups and the mismatches in distributions
493 when compared to the LCDs suggest that the LCDs in the tropical North Atlantic Ocean are
494 unlikely to derive from Pelagophyceae, radial centric diatoms, Raphidophyceae, and/or
495 Eustigmatophyceae.
496 Overall the abundance of known LCD producers is low and scattered and does not match
497 the observed abundance patterns observed for the LCDs, suggesting that most of the LCDs
498 measured here were not produced by any of these species.
499
500 *4.2.2 Correlations between OTUs and LCDs*
501 Since LCDs have been shown to be present within two genetically distant eukaryotic
502 supergroups, the Heterokontophyta and the Archaeplastida, the latter including plants as well
503 as green and red algae, the genetic and enzymatic machinery required for the biosynthesis of
504 LCDs might be present in other genera and classes, including uncultured species. We,
505 therefore, also compared the concentration of LCDs with the composition of the entire
506 eukaryotic microbial community, normalised with respect to the 18S rRNA gene abundance,
507 at both class and OTU levels to identify co-occurrence patterns. No significant correlation
508 was found at class level (data not shown), whereas the correlations at OTU level were weak
509 (≤ 0.5) but significant (p-value <0.01) for 27 OTUs affiliated to 11 different classes (Table 2).

A reason behind the lack of correlation between taxonomic classes and LCDs can be that pooling OTUs at higher taxonomic levels likely leads to combining the LCD-producers with species which are unable to produce LCDs but are falling in the same taxonomic level. The ability from microorganisms to biosynthesise LCDs can indeed vary even between genetically related species, since it has been shown that some genera include both LCD-producers and species which do not contain LCDs (Table S1).

The C_{30} 1,15 diol exhibited significant correlations ($p < 0.01$) with 23 OTUs and overall, 27 OTUs were significantly correlated with C_{30} or to a lesser extent, C_{32} diols (Table 3). Of the 27 OTUs, 4 OTUs were affiliated to classes containing known LCD-producers (Chrysophyceae and Dictyochophyceae, Table 3). The abundance of the two chrysophycean OTUs (denovo465 and denovo1680, Table 3) exhibited significant correlations with the concentrations of both C_{30} 1,13 and C_{30} 1,15 and accounted for 52 % of the total reads from this class and the only known LCD-producer from this class (*Chrysosphaera parvula*) was found to contain C_{32} 1,15 diol (Rampen, unpublished results). The 2 OTUs affiliated to Dictyochophyceae (denovo873 and denovo958) and exhibiting positive correlation with LCDs, cluster within Pedinellales and Florenciellales families, respectively, and are thus closely related to two known LCD-producers, *Florenciella parvula* and *Apedinella radians*. However, *F. parvula* contains C_{24} 1,13-, C_{24} 1,14-, and C_{24} 1,15-diols (Rampen, unpublished results) and *A. radians* produces C_{28}, C_{30}, and C_{32} 1,14-diols (Rampen et al., 2011), whereas the two dictyochophycean OTUs denovo873 and denovo958 exhibited positive correlation with the C_{30} 1,15-diol (Table 3).

The correlation values found here are nearly all low ($\rho \approx 0.4$-0.5), raising the question on whether these relationships reflect the ability of these species to produce LCDs or that these co-occurrences are simply driven by other environmental conditions leading to similar spatial distributions of OTUs and LCDs. Furthermore, other taxa which showed significant correlations with C_{30} diols occur rarely in the marine environment. For example, Centroheliozoa are mostly known as freshwater predators (Slapeta et al., 2005) and, in seawater, have been sporadically detected in anoxic environments only (Stock et al., 2009; Stoeck et al., 2009), suggesting that the centroheliozoan reads found here are unlikely to derive from active microorganisms. In contrast, other taxa include marine representatives commonly found in the photic zone of seawater and thus the reads found here might derive from living organisms: Syndiniales are intracellular parasites of other marine protists, and the genetic clades found here (Group I Clade 4, Group II Clades 2, 7,8,17, and 23) are commonly detected in the upper 100 m of the water column (Guillou et al., 2008). Spirotrichea instead

include several heterotrophic and mixotrophic marine planktonic ciliates (Santoferrara et al., 2017; Agatha et al., 2004), whereas *Phaeocystis* is a widespread primary producer. The uncultured classes exhibiting significant positive correlations with LCDs (Prasino Clade IX and the HAP-3 clade) are also commonly observed in the photic zone (Lopes dos Santos et al., 2016; Shi et al., 2009; Egge et al., 2015). However, cultivated representatives will be required in order to confirm whether species within these taxa are capable of LCD synthesis.

4.3 Can 18S rRNA gene-based community composition analysis be used to determine LCD biological sources?

The lack of correlations of C_{28} and C_{32} diols with any OTUs as well as the low degree of correlation between OTUs and C_{30} diols and the trace abundance or near absence of known LCD producers suggest that the 18S rRNA genes from the microorganisms which produced these LCDs were either absent, or present below detection level in the seawater sampled. The fact that we sampled >90 % of the OTUs potentially present (Figure S1) and the use of universal eukaryotic primers suggests that LCD-producers have been unlikely to escape detection. However, the relatively low number of rRNA gene copies found for *N. oculata* (Zhu et al., 2005) and likewise also in other smaller-sized marine Eustigmatophyceae, suggest that LCD-producers might have been undersampled with respect to larger-sized species which can contain up to 1000 rRNA copies per cell (Zhu et al., 2005).

It should be considered that both the LCDs and DNA in the SPM might derive not only from active or senescent cells, but also from debris (Not et al., 2009). In addition, LCDs can persist in seawater for likely much longer periods than the DNA of the related LCD-producers. Although the biological function of LCDs is unclear for most species, they have been shown to be the building blocks of cell wall polymers in Eustigmatophyceae. In *Nannochloropsis* cell wall, LCDs and long chain alkenols are likely to be bound together through ester and ether bonds to form highly refractory polymers known as algaenans (Gelin et al., 1997; Scholz et al., 2014). These biopolymers are thought to be quite persistent and accumulate in ancient sediments for millions of years (Derenne and Largeau, 2001; de Leeuw et al., 2006; Tegelaar et al., 1989). Indeed, LCDs are ubiquitous in recent surface sediments (Rampen et al., 2012) and ancient sediments of up to 65 million years old (Yamamoto et al., 1996) showing their recalcitrant nature. Recent laboratory experiments highlighted that LCDs from dead biomass of *Nannochloropsis oculata* can persist in seawater for longer than 250 days under anoxic (Grossi et al., 2001) and oxic conditions (Reiche et al., unpublished results). In contrast, much shorter turnover times (6 h to 2 months) are typically reported for

extracellular DNA in the oxic water column (Nielsen et al., 2007). This suggests that the DNA from LCD-producers likely reflects the living eukaryotic community (recently) present when seawater was sampled, while the LCDs probably represent an accumulation that occurred over longer periods of time (weeks to months or even years).

Because of this large difference in turnover rates between LCDs and the DNA from the LCD-producers, 18S rRNA gene analysis of environmental samples may be unsuccessful for identifying LCD-producers. This is seemingly in contrast to a previous study that showed that the LCD concentration in the upper 25 m of the freshwater lake Challa (Tanzania) was related to the number of eustigmatophycean 18S rRNA gene copies (Villanueva et al., 2014). However, Villanueva et al. (2014) used Eustigmatophyceae-biased primers and since this was a lake system, Eustigmatophyceae are likely to be the major source of LCDs in freshwater ecosystems. Importantly, they found a mismatch for the uppermost part of the water column (0–5 m), where high LCD abundance (38–46 ng L^{-1}) coincided with little or no Eustigmatophyceae 18S rRNA gene copies. This pattern was explained by them to be caused by wind-driven and convective mixing of preserved LCDs, while phytoplankton adjusted its buoyancy at greater depth (Villanueva et al., 2014).

Laboratory experiments carried out under different conditions of temperature, light irradiance, salinity and nitrate concentrations revealed average cellular LCD content of about 23 fg $cell^{-1}$ (Balzano et al., 2017) for *Nannochloropsis oceanica*. The average LCD concentration in the SPM investigated was ca. 2.6 ng L^{-1}, which would correspond to ca. 1.1 x 10^6 pico/nano algal cells L^{-1}. We detected average phytoplankton abundances of 3.3 x 10^6 cell L^{-1} for picoeukaryotes and 3.6 x 10^4 cell L^{-1} for nanoeukaryotes. Although nanoplanktonic Eustigmatophyceae might produce larger amounts of LCDs than those measured in our previous study (Balzano et al., 2017), because of their larger cell size, the nanoplankton abundances measured here are two orders of magnitude lower than the levels required to source the LCDs (1.1 x 10^6 cell L^{-1}). Therefore, if the LCDs measured here were biosynthesised by intact microorganisms in the water column, nanoplankton alone would not be able to source all the LCDs measured, and therefore in addition at least one-third of the picophytoplankton should be able to produce LCDs, which is unrealistic. This supports the idea that most of the LCDs detected here are of fossil nature and not contained in living cells. The higher concentrations of LCDs found in the SPM from the mixed layer compared to the DCM suggest that LCDs were originally produced at a higher frequency in the mixed layer. Moreover their possible fossil nature indicates that LCDs were likely to persist in the mixed layer for long periods, eventually associated with suspended particulate matter.

612

613 The combination of lipid and DNA analyses is often complicated by different turnover
614 rates, especially for refractory compounds such as LCDs. Studies focused on more labile
615 biomarker lipids such as fatty acids or intact polar lipids can be more successful, e.g. with
616 short branched fatty acids (Balzano et al., 2011) or archaeal phospholipids (Buckles et al.,
617 2013; Pitcher et al., 2011). Therefore, care has to be taken in inferring sources of biomarker
618 lipids by the quantitative comparison of DNA abundance with biomarker lipid concentrations.

619

5. Conclusions

The combination of lipid analyses and 18S rRNA gene amplicon sequencing revealed some weak correlations between 23 OTUs and C_{30} diols. Four of these OTUs are affiliated to classes that include few LCD-producing species (i.e. Chrysophyceae and Dictyochophyceae), whereas the remaining 19 OTUs belong to taxa in which the presence of LCDs has not been shown. In both cases it remains unclear whether the correlation between these 27 OTUs and the C_{30} diols reflects novel LCD-producers or is driven by other environmental conditions.

The abundances of photosynthetic pico and nanoeukaryotes measured here suggest that these microbial populations are highly unlikely to source all the LCDs found. Some of the LCDs found here might be associated with suspended debris rather than intact cells, with the DNA from their producers being already degraded at the time of sampling. DNA degradation rates in the oxygenated water column are indeed faster than those of most lipids, including LCDs. The freshness of the organic matter and the turnover rates of both lipids and DNA in a given environment should thus be considered when identifying the biological sources of a specific class of lipids through DNA sequencing. In particular the 18S rRNA gene amplicon sequencing can be suitable to track LCD sources (1) for simple ecosystems or laboratory/*in situ* mesocosms with high proportions of fresh organic matter and (2) for low oxygen/anoxic environments where extracellular DNA can persist for longer periods.

Acknowledgments

We thank the captains and the crew of the R/V Pelagia for their support during the cruise. We thank Sharyn Ossebaar for help in sample collection, H. Witte, E. Panoto, and S. Vreugdenhil for support in molecular biology, H. Malschaert for the bioinformatics and M. Besseling for helpful discussions. This research was funded by the European Research Council (ERC) under the European Union's Seventh Framework Program (FP7/2007-2013) ERC grant agreement [339206]. S.S. and J.S.S.D. receive financial support from the Netherlands Earth System Science Centre (NESSC).

References

Agatha, S., Struder-Kypke, M. C., and Beran, A.: Morphologic and genetic variability in the marine planktonic ciliate *Laboea strobila* Lohmann, 1908 (Ciliophora, Oligotrichia), with notes on its ontogenesis, J. Eukaryot. Microbiol., 51, 267-281, 10.1111/j.1550-7408.2004.tb00567.x, 2004.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J.: Basic local alignment search tool, J. Mol. Bio., 215, 403-410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2, 1990.

Andersen, R. A., Brett, R. W., Potter, D., and Sexton, J. P.: Phylogeny of the Eustigmatophyceae based upon 18S rDNA, with emphasis on *Nannochloropsis*, Protist, 149, 61-74, 1998.

Bale, N. J., Villareal, T. A., Hopmans, E. C., Brussaard, C. P. D., Besseling, M., Dorhout, D., Sinninghe Damsté, J. S., and Schouten, S.: C5 glycolipids of heterocystous cyanobacteria track symbiont abundance in the diatom *Hemiaulus hauckii* across the tropical north Atlantic, Biogeosciences, 15, In press, 10.5194/bg-2017-300, 2018.

Balzano, S., Pancost, R. D., Lloyd, J. R., and Statham, P. J.: Changes in fatty acid composition in degrading algal aggregates, Mar. Chem., 124, 2-13, 10.1016/j.marchem.2010.11.001, 2011.

Balzano, S., Abs, E., and Leterme, S. C.: Protist diversity along a salinity gradient in a coastal lagoon, Aquat. Microb. Ecol., 74, 263-277, 10.3354/ame01740, 2015.

Balzano, S., Villanueva, L., de Bar, M., Sinninghe Damsté, J. S., and Schouten, S.: Impact of culturing conditions on the abundance and composition of long chain alkyl diols in species of the genus *Nannochloropsis*, Org. Geochem., 108, 9-17, https://doi.org/10.1016/j.orggeochem.2017.02.006, 2017.

Benjamini, Y., and Hochberg, Y.: Controlling the false discovery rate. A pratical and powerful approach to multiple testing., J Roy. Stat. Soc. B Met., 57, 289-300, 1995.

Bergesch, M., Odebrecht, C., and Moestrup, O.: Nanoflagellates from coastal waters of southern Brazil (32 degrees S), Bot. Mar., 51, 35-50, 10.1515/bot.2008.003, 2008.

Buckles, L. K., Villanueva, L., Weijers, J. W. H., Verschuren, D., and Sinninghe Damste, J. S.: 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), Environ. Microbiol., 15, 2445-2462, 10.1111/1462-2920.12118, 2013.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Tumbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R.: QIIME allows analysis of high-throughput community sequencing data, Nat. Methods, 7, 335-336, 10.1038/nmeth.f.303, 2010.

688 Christaki, U., Kormas, K. A., Genitsaris, S., Georges, C., Sime-Ngando, T., Viscogliosi, E., and Monchy,
689 S.: Winter-summer succession of unicellular eukaryotes in a meso-eutrophic coastal system,
690 Microb. Ecol., 67, 13-23, 10.1007/s00248-013-0290-4, 2014.

691 Comeau, A. M., Li, W. K. W., Tremblay, J.-E., Carmack, E. C., and Lovejoy, C.: Arctic Ocean microbial
692 community structure before and after the 2007 record sea ice minimum, Plos One, 6,
693 10.1371/journal.pone.0027492, 10.1371/journal.pone.0027492, 2011.

694 de Bar, M. W., Dorhout, D. J. C., Hopmans, E. C., Rampen, S. W., Sinninghe Damsté, J. S., and
695 Schouten, S.: Constraints on the application of long chain diol proxies in the Iberian Atlantic
696 margin, Org. Geochem., 101, 184-195, 10.1016/j.orggeochem.2016.09.005, 2016.

697 de Leeuw, J. W., Irene, W., Rijpstra, C., and Schenck, P. A.: The occurrence and identification of C_{30},
698 C_{31} and C_{32} alkan-1, 15-diols and alkan-15-one-1-ols in Unit I and Unit II Black Sea sediments,
699 Geochim. Cosmochim. Ac., 45, 2281-2285, http://dx.doi.org/10.1016/0016-7037(81)90077-6,
700 1981.

701 de Leeuw, J. W., Versteegh, G. J. M., and van Bergen, P. F.: Biomacromolecules of algae and plants
702 and their fossil analogues, Plant Ecol., 182, 209-233, 10.1007/s11258-005-9027-x, 2006.

703 de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N.,
704 Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron,
705 S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G.,
706 Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A.,
707 Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Acinas, S. G., Bork, P., Bowler, C.,
708 Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J.,
709 Sieracki, M. E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., and
710 Karsenti, E.: Eukaryotic plankton diversity in the sunlit ocean, Science, 348,
711 10.1126/science.1261605, 10.1126/science.1261605, 2015.

712 Derenne, S., and Largeau, C.: A review of some important families of refractory macromolecules:
713 Composition, origin, and fate in soils and sediments, Soil Sci., 166, 833-847,
714 10.1097/00010694-200111000-00008, 2001.

715 Dixon, P.: VEGAN, a package of R functions for community ecology, J. Veg. Sci., 14, 927-930,
716 10.1111/j.1654-1103.2003.tb02228.x, 2003.

717 Edgar, R. C.: Search and clustering orders of magnitude faster than BLAST, Bioinformatics, 26, 2460-
718 2461, 10.1093/bioinformatics/btq461, 2010.

719 Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R.: UCHIME improves sensitivity and
720 speed of chimera detection, Bioinformatics, 27, 2194-2200, 10.1093/bioinformatics/btr381,
721 2011.

722 Egge, E. S., Johannessen, T. V., Andersen, T., Eikrem, W., Bittner, L., Larsen, A., Sandaa, R.-A., and
723 Edvardsen, B.: Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway,
724 explored by high-throughput sequencing, Mol. Ecol., 24, 3026-3042, 10.1111/mec.13160,
725 2015.

726 Fawley, K. P., and Fawley, M. W.: Observations on the diversity and ecology of freshwater
727 *Nannochloropsis* (Eustigmatophyceae), with descriptions of new taxa, Protist, 158, 325-336,
728 10.1016/j.protis.2007.03.003, 2007.

729 Gelin, F., Boogers, I., Noordeloos, A. A. M., Sinninghe Damsté, J. S., Riegman, R., and De Leeuw, J. W.:
730 Resistant biomacromolecules in marine microalgae of the classes Eustigmatophyceae and
731 Chlorophyceae: geochemical implications, Org. Geochem., 26, 659-675, 10.1016/s0146-
732 6380(97)00035-1, 1997.

733 Grossi, V., Blokker, P., and Sinninghe Damste, J. S.: Anaerobic biodegradation of lipids of the marine
734 microalga *Nannochloropsis salina*, Org. Geochem., 32, 795-808, 10.1016/s0146-
735 6380(01)00040-7, 2001.

736 Guillou, L., Viprey, M., Chambouvet, A., Welsh, R. M., Kirkham, A. R., Massana, R., Scanlan, D. J., and
737 Worden, A. Z.: Widespread occurrence and genetic diversity of marine parasitoids belonging
738 to Syndiniales (Alveolata), Environ. Microbiol., 10, 3349-3365, 10.1111/j.1462-
739 2920.2008.01731.x, 2008.

740 Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C.,
741 Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra,
742 W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard,
743 R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D.,
744 Zimmermann, P., and Christen, R.: The Protist Ribosomal Reference database (PR2): a catalog
745 of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy, Nucleic Acid
746 Res., 41, D597-D604, 10.1093/nar/gks1160, 2013.

747 Jiang, S. C., Oleary, T., Volkman, J. K., Zhang, H. Z., Jia, R. F., Yu, S. H., Wang, Y., Luan, Z. F., Sun, Z. Q.,
748 and Jiang, R. H.: Origins and simulated thermal alteration of sterols and keto-alcohols in
749 deep-sea marine sediment of the Okinawa Trough., Org. Geochem., 21, 415-422, 1994.

750 Lassiter, A. M., Wilkerson, F. P., Dugdale, R. C., and Hogue, V. E.: Phytoplankton assemblages in the
751 CoOP-WEST coastal upwelling area, Deep-Sea Research Part II-Topical Studies in
752 Oceanography, 53, 3063-3077, 10.1016/j.dsr2.2006.07.013, 2006.

753 Lattaud, J., Dorhout, D., Schulz, H., Castañeda, I. S., Schefuß, E., Sinninghe Damsté, J. S., and
754 Schouten, S.: The C_{32} alkane-1,15-diol as a proxy of late Quaternary riverine input in coastal
755 margins, Clim. Past, 13, 1049-1061, 10.5194/cp-13-1049-2017, 2017a.

756 Lattaud, J., Kim, J. H., De Jonge, C., Zell, C., Sinninghe Damste, J. S., and Schouten, S.: The C_{32} alkane-
757 1,15-diol as a tracer for riverine input in coastal seas, Geochim. Cosmochim. Ac., 202, 146-
758 158, 10.1016/j.gca.2016.12.030, 2017b.

759 Legendre, P., and Gallagher, E. D.: Ecologically meaningful transformations for ordination of species
760 data, Oecologia, 129, 271-280, 10.1007/s004420100716, 2001.

761 Logares, R., Audic, S., Santini, S., Pernice, M. C., de Vargas, C., and Massana, R.: Diversity patterns and
762 activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing, ISME J,
763 6, 1823-1833, 10.1038/ismej.2012.36, 2012.

764 Lopes dos Santos, A., Gourvil, P., Tragin, M., Noel, M.-H., Decelle, J., Romac, S., and Vaulot, D.:
765 Diversity and oceanic distribution of prasinophytes clade VII, the dominant group of green
766 algae in oceanic waters, ISME J, 11, 512-528, 10.1038/ismej.2016.120, 2016.

767 Mao, S. Y., Zhu, X. W., Wu, N. Y., Jia, G. D., Sun, Y. G., Guan, H. X., and Wu, D. D.: Alcohol compounds
768 in Azolla imbricata and potential source implication for marine sediments, Sci. China Earth
769 Sci., 60, 348-359, 10.1007/s11430-016-5177-6, 2017.

770 Marie, D., Simon, N., and Vaulot, D.: Phytoplankton cell counting by flow cytometry., in: Algal
771 Culturing Techniques, edited by: RA, A., Elsevier Academic Press: Burlington, MA, 253-268,
772 2005.

773 Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R.,
774 Claverie, J. M., Decelle, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Forn, I., Forster, D.,
775 Guillou, L., Jaillon, O., Kooistra, W., Logares, R., Mahe, F., Not, F., Ogata, H., Pawlowski, J.,
776 Pernice, M. C., Probert, I., Romac, S., Richards, T., Santini, S., Shalchian-Tabrizi, K., Siano, R.,
777 Simon, N., Stoeck, T., Vaulot, D., Zingone, A., and de Vargas, C.: Marine protist diversity in
778 European coastal waters and sediments as revealed by high-throughput sequencing, Environ.
779 Microbiol., 17, 4035-4049, 10.1111/1462-2920.12955, 2015.

780 Moita, M. T., Oliveira, P. B., Mendes, J. C., and Palma, A. S.: Distribution of chlorophyll a and
781 Gymnodinium catenatum associated with coastal upwelling plumes off central Portugal, Acta
782 Oecol., 24, S125-S132, 10.1016/s1146-609x(03)00011-0, 2003.

783 Molleri, G. S. F., Novo, E., and Kampel, M.: Space-time variability of the Amazon River plume based
784 on satellite ocean color, Cont. Shelf Res., 30, 342-352, 10.1016/j.csr.2009.11.015, 2010.

785 Nielsen, K. M., Johnsen, P. J., Bensasson, D., and Daffonchio, D.: Release and persistence of
786 extracellular DNA in the environment, Environmental Biosafety Research, 6, 37-53,
787 10.1051/ebr:2007031, 2007.

788 Not, F., del Campo, J., Balague, V., de Vargas, C., and Massana, R.: New Insights into the Diversity of
789 Marine Picoeukaryotes, Plos One, 4, 10.1371/journal.pone.0007143,
790 10.1371/journal.pone.0007143, 2009.

791 Pitcher, A., Villanueva, L., Hopmans, E. C., Schouten, S., Reichart, G. J., and Sinninghe Damste, J. S.:
792 Niche segregation of ammonia-oxidizing archaea and anammox bacteria in the Arabian Sea
793 oxygen minimum zone, ISME J, 5, 1896-1904, 10.1038/ismej.2011.60, 2011.

794 Plancq, J., Mattioli, E., Pittet, B., Simon, L., and Grossi, V.: Productivity and sea-surface temperature
795 changes recorded during the late Eocene-early Oligocene at DSDP Site 511 (South Atlantic),
796 Palaeogeogr. Palaeocl., 407, 34-44, 10.1016/j.palaeo.2014.04.016, 2014.

797 Racovita, R. C., and Jetter, R.: Identification of In-Chain-Functionalized Compounds and Methyl-
798 Branched Alkanes in Cuticular Waxes of Triticum aestivum cv. Bethlehem, Plos One, 11, 25,
799 10.1371/journal.pone.0165827, 2016.

Rampen, S. W., Schouten, S., Wakeham, S. G., and Sinninghe Damste, J. S.: Seasonal and spatial variation in the sources and fluxes of long chain diols and mid-chain hydroxy methyl alkanoates in the Arabian Sea, Org. Geochem., 38, 165-179, 10.1016/j.orggepchem.2006.10.008, 2007.

Rampen, S. W., Schouten, S., Koning, E., Brummer, G.-J. A., and Sinninghe Damsté, J. S.: A 90 kyr upwelling record from the northwestern Indian Ocean using a novel long-chain diol index, Earth Planet. Sc. Lett., 276, 207-213, 10.1016/j.epsl.2008.09.022, 2008.

Rampen, S. W., Schouten, S., Schefuss, E., and Sinninghe Damsté, J. S.: Impact of temperature on long chain diol and mid-chain hydroxy methyl alkanoate composition in *Proboscia diatoms*: results from culture and field studies, Org. Geochem., 40, 1124-1131, 10.1016/j.orggeochem.2009.08.005, 2009.

Rampen, S. W., Schouten, S., and Sinninghe Damsté, J. S.: Occurrence of long chain 1,14-diols in *Apedinella radians*, Org. Geochem., 42, 572-574, 10.1016/j.orggeochem.2011.03.009, 2011.

Rampen, S. W., Willmott, V., Kim, J.-H., Uliana, E., Mollenhauer, G., Schefuss, E., Sinninghe Damsté, J. S., and Schouten, S.: Long chain 1,13-and 1,15-diols as a potential proxy for palaeotemperature reconstruction, Geochim. Cosmochim. Ac., 84, 204-216, 10.1016/j.gca.2012.01.024, 2012.

Rampen, S. W., Datema, M., Rodrigo-Gamiz, M., Schouten, S., Reichart, G.-J., and Sinninghe Damsté, J. S.: Sources and proxy potential of long chain alkyl diols in lacustrine environments, Geochim. Cosmochim. Ac., 144, 59-71, 10.1016/j.gca.2014.08.033, 2014a.

Rampen, S. W., Willmott, V., Kim, J.-H., Rodrigo-Gamiz, M., Uliana, E., Mollenhauer, G., Schefuss, E., Sinninghe Damsté, J. S., and Schouten, S.: Evaluation of long chain 1,14-alkyl diols in marine sediments as indicators for upwelling and temperature, Org. Geochem., 76, 39-47, 10.1016/j.orggeochem.2014.07.012, 2014b.

Rodrigo-Gámiz, M., Rampen, S. W., de Haas, H., Baas, M., Schouten, S., and Sinninghe Damsté, J. S.: Constraints on the applicability of the organic temperature proxies UK037, TEX86 and LDI in the subpolar region around Iceland, Biogeosciences, 12, 6573-6590, 2015.

Santoferrara, L. F., Alder, V. V., and McManus, G. B.: Phylogeny, classification and diversity of Choreotrichia and Oligotrichia (Ciliophora, Spirotrichea), Mol. Phylogenet. Evol., 112, 12-22, 10.1016/j.ympev.2017.03.010, 2017.

Ocean Data View, http://www.awi-bremerhaven.de/GEO/ODV, 2017.

Scholz, M. J., Weiss, T. L., Jinkerson, R. E., Jing, J., Roth, R., Goodenough, U., Posewitz, M. C., and Gerken, H. G.: Ultrastructure and composition of the *Nannochloropsis gaditana* cell wall, Eukaryot. Cell, 13, 1450-1464, 10.1128/ec.00183-14, 2014.

Seoane, S., Laza, A., Urrutxurtu, M., and Orive, E.: Phytoplankton assemblages and their dominant pigments in the Nervion River estuary, Hydrobiologia, 549, 1-13, 10.1007/s10750-005-1736-6, 2005.

837 Shi, X. L., Marie, D., Jardillier, L., Scanlan, D. J., and Vaulot, D.: Groups without Cultured
838 Representatives Dominate Eukaryotic Picophytoplankton in the Oligotrophic South East
839 Pacific Ocean, Plos One, 4, 10.1371/journal.pone.0007657, 2009.

840 Sinninghe Damsté, J. S., Rampen, S., Irene, W., Rupstra, C., Abbas, B., Muyzer, G., and Schouten, S.: A
841 diatomaceous origin for long-chain diols and mid-chain hydroxy methyl alkanoates widely
842 occurring in Quaternary marine sediments: Indicators for high-nutrient conditions, Geochim.
843 Cosmochim. Ac., 67, 1339-1348, 10.1016/s0016-7037(02)01225-5, 2003.

844 Slapeta, J., Moreira, D., and Lopez-Garcia, P.: The extent of protist diversity: insights from molecular
845 ecology of freshwater eukaryotes, P. Roy. Soc. B-Biol. Sci., 272, 2073-2081,
846 10.1098/rspb.2005.3195, 2005.

847 Speelman, E. N., Reichart, G. J., de Leeuw, J. W., Rijpstra, W. I. C., and Sinninghe Damste, J. S.:
848 Biomarker lipids of the freshwater fern Azolla and its fossil counterpart from the Eocene
849 Arctic Ocean, Org. Geochem., 40, 628-637, 10.1016/j.orggeochem.2009.02.001, 2009.

850 Stock, A., Juergens, K., Bunge, J., and Stoeck, T.: Protistan diversity in suboxic and anoxic waters of
851 the Gotland Deep (Baltic Sea) as revealed by 18S rRNA clone libraries, Aquat. Microb. Ecol.,
852 55, 267-284, 10.3354/ame01301, 2009.

853 Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M. J., Chistoserdov, A., Orsi,
854 W., and Edgcomb, V. P.: Massively parallel tag sequencing reveals the complexity of
855 anaerobic marine protistan communities, BMC Biol., 7, 10.1186/1741-7007-7-72, 2009.

856 Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W., and Richards, T. A.:
857 Multiple marker parallel tag environmental DNA sequencing reveals a highly complex
858 eukaryotic community in marine anoxic water, Mol. Ecol., 19, 21-31, 10.1111/j.1365-
859 294X.2009.04480.x, 2010.

860 Tegelaar, E. W., Deleeuw, J. W., Derenne, S., and Largeau, C.: A reappraisal of kerogen formation,
861 Geochim. Cosmochim. Ac., 53, 3103-3106, 10.1016/0016-7037(89)90191-9, 1989.

862 Versteegh, G. J. M., Bosch, H. J., and De Leeuw, J. W.: Potential palaeoenvironmental information of
863 C-24 to C-36 mid-chain diols, keto-ols and mid-chain hydroxy fatty acids; a critical review,
864 Org. Geochem., 27, 1-13, 10.1016/s0146-6380(97)00063-6, 1997.

865 Versteegh, G. J. M., Jansen, J. H. F., De Leeuw, J. W., and Schneider, R. R.: Mid-chain diols and keto-
866 ols in SE Atlantic sediments: A new tool for tracing past sea surface water masses?, Geochim.
867 Cosmochim. Ac., 64, 1879-1892, 10.1016/s0016-7037(99)00398-1, 2000.

868 Villanueva, L., Besseling, M., Rodrigo-Gamiz, M., Rampen, S. W., Verschuren, D., and Sinninghe
869 Damsté, J. S.: Potential biological sources of long chain alkyl diols in a lacustrine system, Org.
870 Geochem., 68, 27-30, 10.1016/j.orggeochem.2014.01.001, 2014.

871 Volkman, J. K., Barrett, S. M., Dunstan, G. A., and Jeffrey, S. W.: C-30-C-32 Alkyl diols and unsaturated
872 alcohols in microalgae of the class Eustigmatophyceae., Org. Geochem., 18, 131-138,
873 10.1016/0146-6380(92)90150-v, 1992.

874 Volkman, J. K., Barrett, S. M., and Blackburn, S. I.: Eustigmatophyte microalgae are potential sources
875 of C-29 sterols, C-22-C-28 *n*-alcohols and C-28-C-32 *n*-alkyl diols in freshwater environments.,
876 Org. Geochem., 30, 307-318, 10.1016/s0146-6380(99)00009-1, 1999.

877 Wen, M., and Jetter, R.: Very-long-chain hydroxyaldehydes from the cuticular wax of Taxus baccata
878 needles, Phytochemistry, 68, 2563-2569, 10.1016/j.phytochem.2007.05.029, 2007.

879 Yamamoto, M., Ficken, K., Baas, M., JH, B., and JW, d. L.: Molecular paleontology of the earliest
880 danian at Geulhemmerberg (the Netherlands), Geol. Mijnbouw, 75, 255-267, 1996.

881 Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D.: Mapping of picoeucaryotes in marine
882 ecosystems with quantitative PCR of the 18S rRNA gene, FEMS Microbiol. Ecol., 52, 79-92,
883 10.1016/j.femsec.2004.10.006, 2005.

884

885

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

886 **Table 1.** Distribution of the18S rRNA gene reads associated with known LCD-producers
887

Taxon	Florenciellales	*Heterosigma*	Eustigmatophyceae	*Proboscia*	Total[a]
No of samples	28	2	8	2	35
Surface	12	1	2	0	12
BWML[b]	11	0	2	2	13
DCM[c]	5	1	4	0	10
No of reads	99	10	45	3	157
% total	0.04	0.004	0.02	0.001	0.06
Surface	48	4	25	0	77
BWML	41	0	9	3	53
DCM	10	6	11	0	27

888

889 [a] Total number of samples where 18S rRNA gene reads from LCD-producers were found or
890 total number of reads associated with LCD-producers. Overall 68 samples were screened for the
891 presence of 18S rRNA genes affiliated to LCD-producers.

892 [b] Bottom water mixed layer

893 [c] Deep chlorophyll maximum

894

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

Table 2. Spearman rank correlation coefficients between LCDs and environmental variables[a].

	C_{28} 1,13	C_{28} 1,14	C_{30} 1,13	C_{30} 1,14	C_{30} 1,15	C_{32} 1,15
POC[b]	0.3	0.2	0.2	0.3	0.3	0.3
Salinity	-0.2	0.0	**-0.5**	**-0.7**	**-0.6**	**-0.6**
Temperature	0.2	-0.1	**0.5**	**0.5**	**0.5**	**0.5**
Phosphate	0.0	0.2	-0.3	-0.2	-0.3	-0.2
Ammonium	0.0	0.1	-0.3	-0.4	-0.4	-0.2
Nitrite	-0.2	0.0	**-0.6**	**-0.5**	**-0.6**	-0.4
Nitrate	0.0	0.2	-0.4	-0.3	-0.3	**-0.5**
Silica	0.1	0.0	0.4	**0.5**	**0.5**	0.4
Chl-*a*	-0.1	0.0	-0.2	-0.2	-0.3	-0.1
Picoeukaryotes	-0.1	-0.1	-0.4	-0.3	-0.4	-0.2
Nanoeukaryotes	0.0	-0.1	0.1	0.2	0.2	0.2

[a] Significant (p-value < 0.01) correlation values are in bold.
[c] Particulate Organic Carbon

Biogeosciences Discussions
Open Access

917 **Table 3.** List of Operational Taxonomic Units (OTUs), representing 95 % of sequence identity, which were correlated[a] with LCDs in the HCC
918 cruise.
919
920

OTU ID	Taxon	Class	C_{30} 1,13	C_{30} 1,14	C_{30} 1,15	C_{32} 1,15	Total 1,13	Total 1,14	Total 1,15
denovo2033	Choreotrichia	Spirotrichea							0.40
denovo2137	*Climacocylis scalaria*		0.45		0.49		0.45		0.49
denovo940	*Laboea strobila*		0.53	0.46	0.60		0.56	0.45	0.59
denovo685	Oligotrichia				0.41				0.40
denovo1804	*Pseudotontonia*		0.56	0.47	0.56	0.47	0.53	0.41	0.57
denovo492	*Blastodinium spinulosum*	Dinophyceae	0.43	0.44	0.46				0.45
denovo720	*Ceratocorys horrida*				0.46		0.44		0.45
denovo1682	*Neoceratium fusus*							0.47	
denovo526	*Protodinium simplex*			0.43	0.44			0.48	0.43
denovo267	*Pyrophacus_steinii*				0.43				0.42
denovo732	Dino Group I Clade 4	Syndiniales			0.40		0.46		0.41
denovo555	Dino Group II Clade 2				0.49		0.41		0.48
denovo1077	Dino Group II Clade 7				0.44		0.42		0.43
denovo1834	Dino Group II Clade 8				0.44		0.45		0.45
denovo1145	Dino Group II Clade 17		0.50		0.49		0.53	0.42	0.48
denovo2080	Dino Group II Clade 23				0.40		0.43		0.40
denovo725	Prasino Clade 9B	Prasino Clade IX			0.42		0.41		0.41
denovo1066	Pterocystida	Centroheliozoa			0.46				0.46
denovo400	HAP3	Haptophyta		0.47	0.49			0.47	0.48
denovo2132	*Phaeocystis*						0.46		
denovo972	*Haptolina*						0.44		
denovo465	**Chrysophyceae Clade G**	Chrysophyceae	0.44		0.43		0.45		0.42
denovo1680	**Chrysophyceae Clade H**		0.44		0.42		0.48		0.42
denovo1988	Raphid pennate	diatoms			0.41				0.41
denovo873	**Pedinellales**	Dictyochophyceae	0.56	0.45	0.55	0.52	0.55		0.56

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

denovo958	**Florenciellales**	Dictyochophyceae			0.43		0.45		0.44
denovo2433	Unidentified picozoan	Picozoa	0.49	0.50	0.55		0.47	0.46	0.55

[a] Only significant (p-value < 0.01 after FDR correction) correlations are shown. OTUs closely related to known LCD-producers are in bold.

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

926 **Figure Legend**

927

928 **Figure 1.** HCC cruise track in the western tropical Atlantic Ocean, physical seawater
929 properties, and biological parameters. (**A**) Map of the sampling stations. Spatial distribution
930 of (**B**) temperature, (**C**) salinity, the concentration of (**D**) chlorophyll-*a*, (**E**) organic carbon
931 concentrations, and the abundance of photosynthetic (**F**) picoeukaryotes and (**G**)
932 nanoeukaryotes. Data were plotted using ocean data view (ODV) software using kriging for
933 interpolation between datapoints (Schlitzer, 2017).

934

935 **Figure 2.** Spatial distribution of the concentration of LCDs: (**A**) C_{28} 1,13, (**B**) C_{28} 1,14,
936 (**C**) C_{30} 1,13, (**D**) C_{30} 1,14, (**E**) C_{30} 1,15, and (**F**) C_{32} 1,15. Data were plotted using ocean data
937 view (ODV) software using kriging for interpolation between datapoints (Schlitzer, 2017).

938

939 **Figure 3.** Average fractional abundance of the reads obtained by 18S rRNA gene
940 sequencing of SPM from the western tropical Atlantic Ocean over the various classes of
941 eukaryotes. The V4 fragment of the 18S rRNA gene was sequenced using universal
942 eukaryotic primers. Samples were pooled according to depth and the average contribution
943 from each group at the different depth is shown. Error bars represent the standard deviation in
944 the data from the various stations.

945

946 **Figure 4.** Spatial distribution of the 18S rRNA gene fragments related to taxa containing
947 LCD-producers at different stations and depth. (**A**) Pelagophyceae, (**B**) Chrysophyceae, (**C**)
948 Dictyochophyceae, (**D**) radial centric diatoms, (**E**) Eustigmatophyceae, and (**F**)
949 Raphidophyceae. Data were plotted using ocean data view (ODV) software using kriging for
950 interpolation between datapoints (Schlitzer, 2017).

951

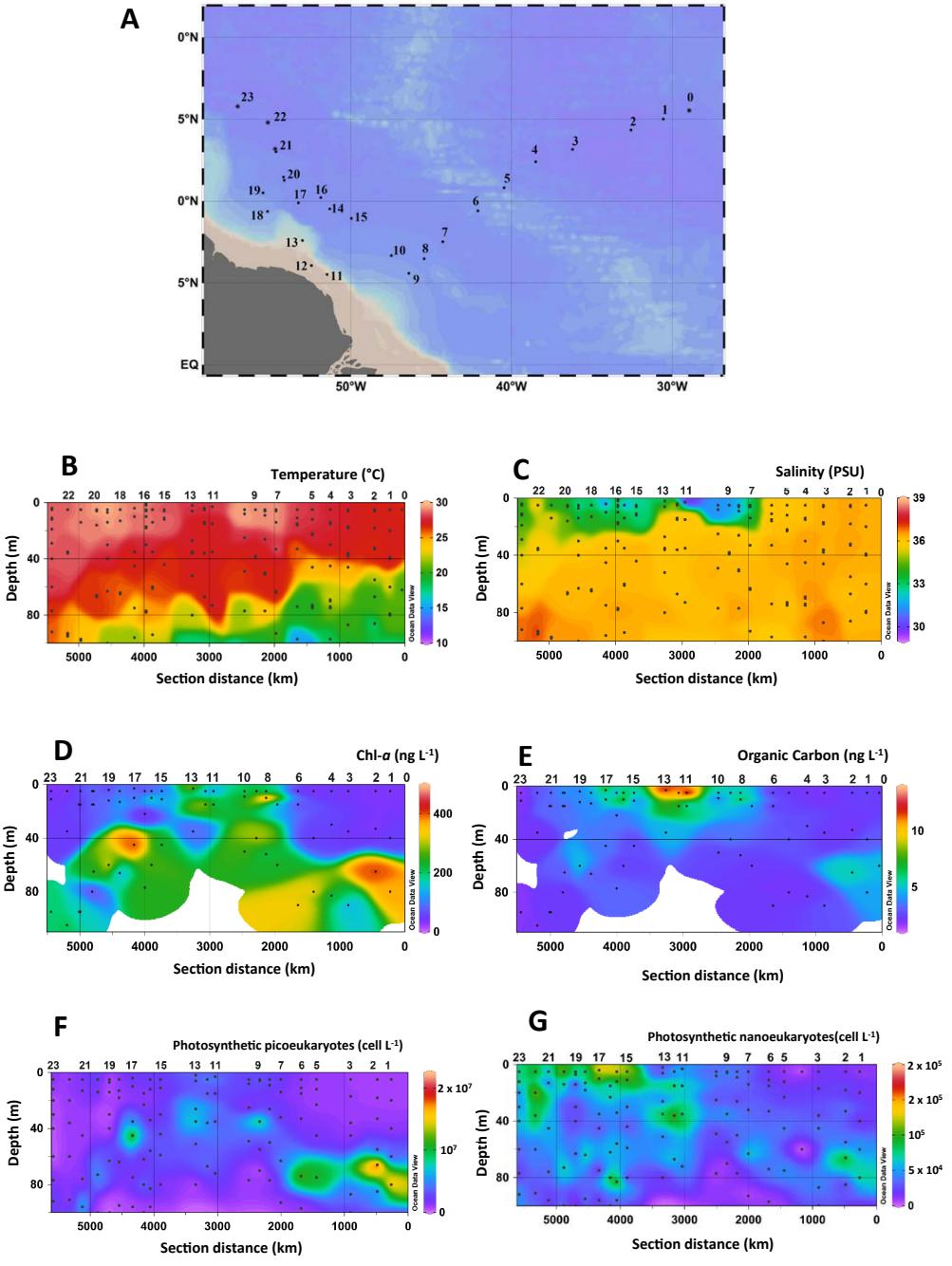
Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

952 **Figure 5.** Scatter plots of the concentrations of the different LCDs in the western tropical
953 Atlantic Ocean. (**A**) C_{30} 1,13 diol vs C_{30} 1,15 diol, (**B**) C_{32} 1,15 diol vs C_{30} 1,15 diol, (**C**) C_{28}
954 1,13 diol vs C_{30} 1,13 diol, (**D**) C_{28} 1,13 diol vs C_{30} 1,15 diol, (**E**) C_{30} 1,14 diol vs C_{28} 1,14
955 diol, (**F**) C_{30} 1,14 diol vs C_{30} 1,15 diol.

956

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

Figure 1



Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

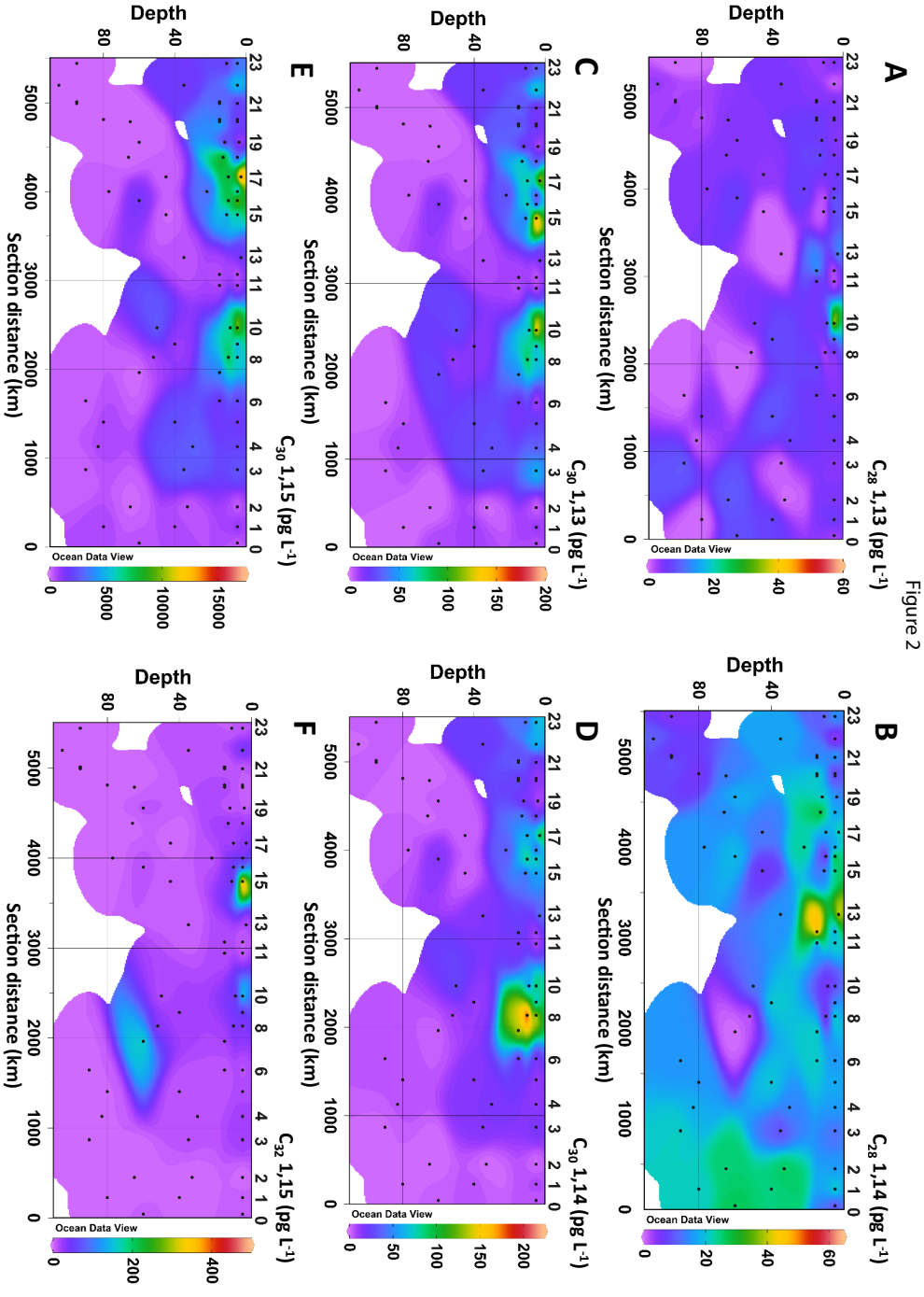
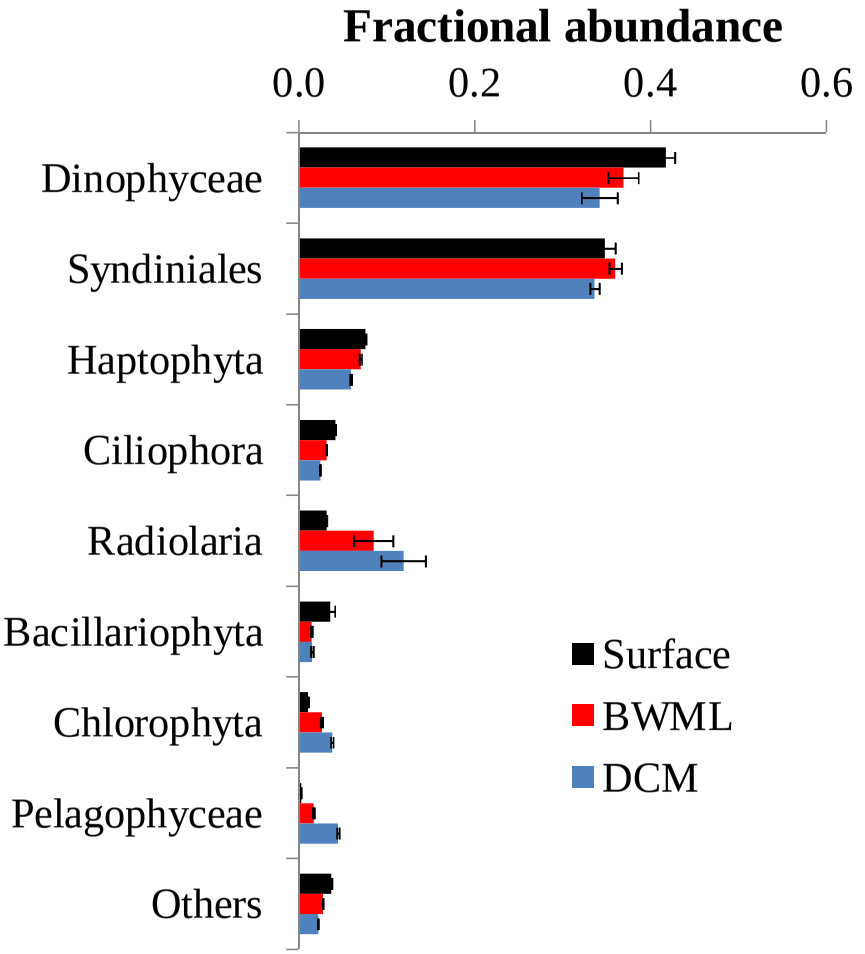


Figure 2

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

Figure 3



Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

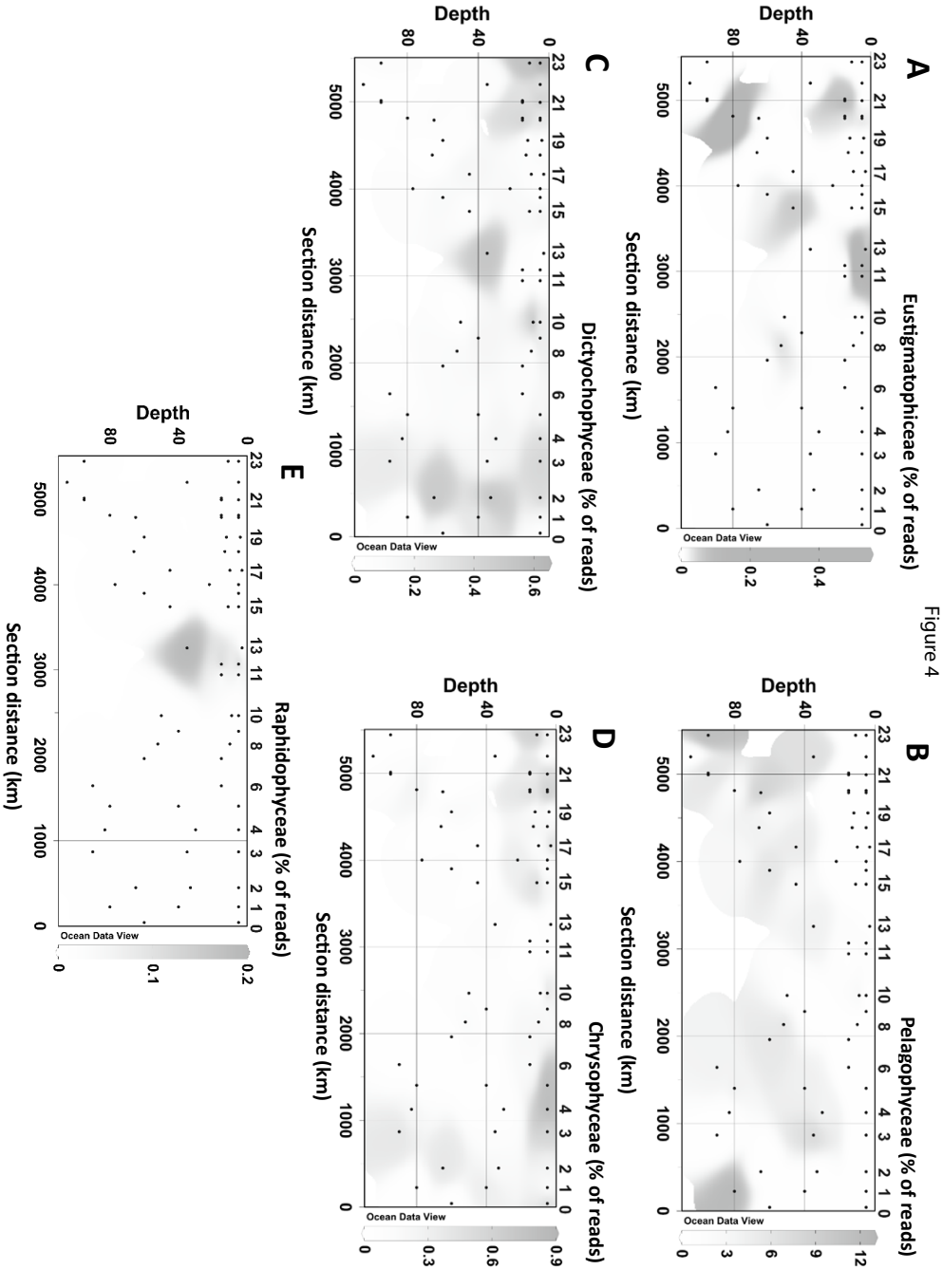


Figure 4

Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-97
Manuscript under review for journal Biogeosciences
Discussion started: 28 March 2018

Figure 5