

Differential association of key bacterial groups with diatoms and *Phaeocystis* spp. during spring blooms in the Southern Ocean

Nyree J. West¹  | Marine Landa² | Ingrid Obernosterer² 

¹CNRS FR3724, Observatoire Océanologique de Banyuls (OOB), Sorbonne Université, Banyuls sur mer, France

²Laboratoire d'Océanographie Microbienne, LOMIC, CNRS, Sorbonne Université, Banyuls sur mer, France

Correspondence

Nyree J. West, Observatoire Océanologique de Banyuls sur mer, Banyuls sur mer F-66650, France.

Email: nyree.west@obs-banyuls.fr

Funding information

Centre National de la Recherche Scientifique, Grant/Award Number: LEFE-CYBER; Institut Polaire Français Paul Emile Victor; Agence Nationale de la Recherche, Grant/Award Number: 10-BLAN-0614

Abstract

Interactions between phytoplankton and heterotrophic bacteria significantly influence the cycling of organic carbon in the ocean, with many of these interactions occurring at the micrometer scale. We explored potential associations between specific phytoplankton and bacteria in two size fractions, 0.8–3 µm and larger than 3 µm, at three naturally iron-fertilized stations and one high nutrient low chlorophyll station in the Southern Ocean. The composition of phytoplankton and bacterial communities was determined by sequencing the *rbcl* gene and 16S rRNA gene from DNA and RNA extracts, which represent presence and potential activity, respectively. Diatoms, particularly *Thalassiosira*, contributed significantly to the DNA sequences in the larger size fractions, while haptophytes were dominant in the smaller size fraction. Correlation analysis between the most abundant phytoplankton and bacterial operational taxonomic units revealed strong correlations between *Phaeocystis* and picoeukaryotes with SAR11, SAR116, *Magnetospira*, and *Planctomarina*. In contrast, most *Thalassiosira* operational taxonomic units showed the highest correlations with *Polaribacter*, *Sulfitobacteria*, *Erythrobacter*, and *Sphingobium*, while *Fragilariopsis*, *Haslea*, and *Thalassionema* were correlated with OM60, *Fluviicola*, and *Ulvibacter*. Our in-situ observations suggest distinct associations between phytoplankton and bacterial taxa, which could play crucial roles in nutrient cycling in the Southern Ocean.

KEYWORDS

algae, cell-microbe interactions, molecular microbial ecology, Ocean microbiology

1 | INTRODUCTION

The Southern Ocean (SO) plays a major role in global climate regulation through both physical and biological mechanisms, notably by the contribution of phytoplankton to the carbon cycle via the biological carbon pump and to the sulphur cycle through dimethylsulfoniopropionate

(DMSP) production. Phytoplankton groups contribute differently to these processes, with diatoms being key players in the SO primary production and CO₂ drawdown (Sarhou et al., 2005) while prymnesiophytes are globally important DMSP producers (Keller et al., 1989; McParland & Levine, 2019). Besides their specific roles in elemental cycles, these groups of phytoplankton could differently affect food web

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *MicrobiologyOpen* published by John Wiley & Sons Ltd.

structures in the SO (Hunt et al., 2021; Krumhardt et al., 2022). Phytoplankton cell size in particular influences grazing, therefore playing a key role in the transfer of cell material toward higher trophic levels and the efficiency of the biological carbon pump (Christaki et al., 2021; Eddy et al., 2021; Moline et al., 2004). For these reasons, the composition and activity of phytoplankton communities are critical determinants of SO ecosystem functioning.

The SO is the largest high nutrient low chlorophyll (HNLC) region of the world's oceans with low phytoplankton biomass during most of the year due to limitation principally by the micronutrient iron (Fe). Spring phytoplankton blooms can, however, develop in regions with natural Fe input, such as over continental shelves or close to islands such as the Kerguelen, Crozet, and Sandwich Islands (Blain et al., 2007; Pollard et al., 2007) and are usually dominated by diatoms (Quéguiner, 2013). The prevalence of diatoms in these naturally Fe-fertilized regions is largely driven by the availability of silicic acid and diatoms' ability to outcompete other phytoplankton taxa for available iron (Marchetti et al., 2012). The onset of silica limitation and subsequent decline of the diatom blooms are usually accompanied by a phytoplankton shift from small, fast-growing to large, slow-growing diatoms (Blain et al., 2021) to an increasing contribution of haptophytes dominated by *Phaeocystis* spp. (Irion et al., 2021; Salter et al., 2007). *Phaeocystis* is generally more abundant in coastal regions (Davidson et al., 2010) and can form extensive prolonged blooms as was reported north-west of the Ross Sea above the Australian Antarctic Ridge (AAR) (Schine et al., 2021) and north of the Crozet plateau (Poulton et al., 2007). Shifts in environmental conditions and ensuing changes in phytoplankton assemblages are expected to impact other components of the microbial community with repercussions on biogeochemical cycles and food web structure.

There is overwhelming evidence of ecological coupling between phytoplankton and heterotrophic bacteria through studies carried out both in the field and the lab. Successional changes through the course of phytoplankton blooms are mirrored in the heterotrophic bacterial community composition (see reviews by Buchan et al., 2014; Bunse & Pinhassi, 2017) that can reveal repeated patterns on multi-annual scales (Fuhrman et al., 2006; Lambert et al., 2019). These interactions may range from simple resource provision to highly specific associations encompassing a multitude of interactions from the mutualistic exchange of info-chemicals or specific nutrients to competition and antagonism (Cirri & Pohnert, 2019; Costas-Selas et al., 2024; Coyne et al., 2022). For some of these exchanges to occur, spatial proximity of the partners has long been hypothesized (Bell & Mitchell, 1972) which led to the coining of the term "phycosphere," analogous to the rhizosphere in plant-soil bacteria interactions (see Seymour et al., 2017) for a review). In the phycosphere, high phytoplankton-derived metabolite concentrations provide a dissolved organic matter (DOM) hotspot that facilitates bacterial growth, particularly for bacteria that are chemotactic or require attachment to their eukaryotic partners. The phycosphere depends on phytoplankton cell size and exudation rates, with large (>50 µm), leaky species expected to provide a much larger phycosphere than picocyanobacteria that are estimated to fall below the theoretical

lower limit of detection by chemotactic bacteria (Seymour et al., 2017). Hence, distinct phytoplankton communities with specific phycosphere properties can exert different selective pressure on coexisting bacterial taxa.

The naturally fertilized region east of Kerguelen Island has been extensively used as a natural laboratory to investigate the effect of Fe on biogeochemistry and ecosystem functioning. Pronounced and specific responses of the bacterial community during the different phases of the Fe-induced phytoplankton blooms were observed (Hernandez-Magana et al., 2021; Landa et al., 2016; Obernosterer et al., 2011; West et al., 2008). However, these previous studies focused mostly on free-living bacteria (0.2–0.8 µm). The major aim of the present study was to identify associations between phytoplankton and particle-attached bacteria by focusing on two size fractions (0.8–3 and >3 µm) that should separate the larger and smaller size classes of phytoplankton (Biggs et al., 2019) and the results presented in this study complement observations in the free-living fraction from the same period and study sites (Dinasquet et al., 2022). To achieve this objective, heterotrophic bacteria and phytoplankton diversity and activity were assessed in spring phytoplankton blooms by metabarcoding of the 16S rRNA gene for bacteria and the RubisCO *rbcL* gene for phytoplankton. As a proxy for activity (protein synthesizing potential; (Blazewicz et al., 2013), cDNA fragments generated from RNA from each marker gene were also analyzed by the same approach.

2 | EXPERIMENTAL PROCEDURES

2.1 | Study area

The samples for the present study were collected in the naturally Fe-fertilized and high nutrient low chlorophyll (HNLC) waters off Kerguelen Island during the KEOPS2 (Kerguelen Ocean and Plateau Study 2) cruise aboard the French R/V Marion Dufresne II ship in the austral spring (10 October to 20 November 2011) (Figure 1).

2.2 | Sample collection

Seawater samples were collected from the four stations at 3–4 discrete depths using a CTD equipped with 12L Niskin bottles (General Oceanics) (See Table 1). One set of samples from the upper 3 depths (20–150/160 m) were size fractionated by prefiltering 9L of seawater through a 65 µm mesh and then filtering onto 10 µm pore-size polycarbonate filters (PC, Nuclepore, 47 mm diameter). The second set of samples was collected from 4 depths (20–300 m), and passed through a 25 µm mesh, before sequentially filtering onto 3 µm and 0.8 µm PC filters followed by a 0.2 µm Sterivex cartridge filter. The volumes filtered ranged between 5 and 7 L except for the 20 and 300 m depths at station R where 2.5 and 3 L were filtered respectively. All filters were stored frozen (–80°C) until further analysis in the laboratory. In this study, the samples from the 10, 3,

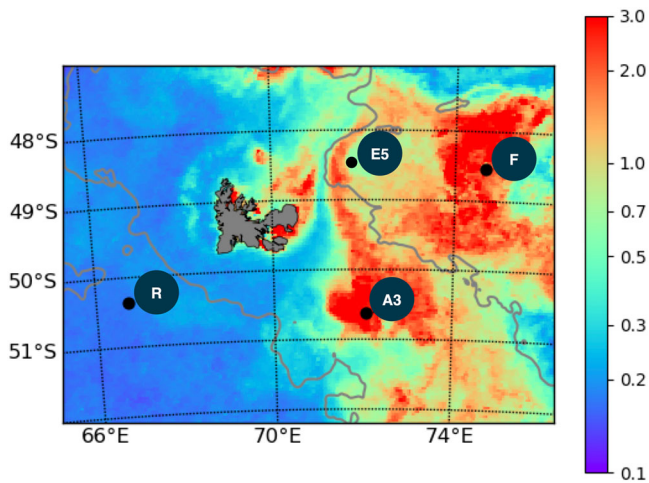


FIGURE 1 Location of the four sampling stations with MODIS-Aqua satellite (CLS-CNES) images of surface chlorophyll a (chl a) concentrations ($\mu\text{g/L}$) indicating the presence of the phytoplankton blooms. A3 is the reference bloom station above the Kerguelen Plateau and R is the reference HNLC station. Station F was situated in a bloom above the Polar Front and E5 was located in the Polar Front meander.

and $0.8\ \mu\text{m}$ filters were analyzed whereas the $0.2\ \mu\text{m}$ Sterivex cartridge filters were analyzed separately (Landa et al., 2016).

2.3 | DNA and RNA extraction

DNA and RNA were extracted simultaneously from the filters as described previously (West et al., 2016). Briefly, bacterial lysis was achieved by adding to the filters 425 μL lysis buffer (40 mM EDTA, 50 mM Tris, 0.75 M sucrose) and carrying out three freeze-thaw cycles (liquid nitrogen -65°C) followed by lysozyme treatment (final concentration 1 mg/mL) at 37°C for 45 min. Proteinase K and SDS were added (final concentration 0.2 mg/mL and 1% respectively) to the filters and incubated at 55°C for 1 h. Purification of DNA and RNA was carried out with the Qiagen AllPrep DNA/RNA extraction kit using 1.55 mL RLT + buffer containing β -mercaptoethanol according to the manufacturer's instructions. The quality of the DNA and RNA was verified by agarose gel electrophoresis and quantified by Picogreen and Ribogreen (Invitrogen) respectively.

2.4 | cDNA synthesis

cDNA was synthesized from RNA immediately after extraction using the Superscript VILO kit (Invitrogen) according to the manufacturer's instructions using 100 ng RNA.

2.5 | Illumina sequencing

To determine bacterial diversity and activity, bacterial 16S rRNA gene fragments of about 450 bp were amplified from 36 DNA samples and

35 cDNA samples (see Table A1) using the primer pair 341F (CCTACGGGNGGCWGCAG) and 805R (GACTACHVGGGTATCTAATCC) described previously (Klindworth et al., 2013). Phytoplankton diversity and activity were assessed by amplifying *rbcl* fragments of 554 bp from 16 surface DNA samples and 4 surface cDNA samples from the 2 bloom stations A3 and F-L (Table A1) using the Form ID Rbcl gene primer set (F: GATGATGARAAYATTAAGT, R: ATTTGDCACAGTGDATACCA) described previously (Wawrik et al., 2003). The 341F and the *rbcl* forward primers were tagged at the 5' end with different 7 bp tags for each sample, and that were chosen from a set of tags designed to be robust to substitution, deletion and insertion errors incurred in massively parallel sequencing (Faircloth & Glenn, 2012). We also included in parallel a control consisting of DNA from a synthetic mock community (Mock) of 20 bacterial species containing equimolar (even) rRNA operon counts (HM-782D; Genomic DNA from Microbial Mock Community B, Even, Low Concentration, BEI Resources, Manassas, VA that was amplified as above with the bacterial primer pair. This standard is now obtainable from LGC Standards S.a.r.l.; reference ATCC[®] MSA-1002[™]).

Before PCR, DNA samples were diluted in molecular grade water (Merck) to a concentration of $10\ \text{ng}/\mu\text{L}$. We decided to pool the $65\text{--}10\ \mu\text{m}$ and $25\text{--}3\ \mu\text{m}$ DNA and cDNA size fraction samples for depths $20\text{--}150/160\ \text{m}$ (see Table A1) to give a combined fraction size of $>3\ \mu\text{m}$ $<65\ \mu\text{m}$ (pooled at equal concentrations of DNA/cDNA). We cannot rule out that phytoplankton falling between 10 and $25\ \mu\text{m}$ may be overestimated due to the overlap between these 2 size fractions. The samples from the $0.8\ \mu\text{m}$ filter corresponded to a fraction size $>0.8\ \mu\text{m}$ $<3\ \mu\text{m}$.

DNA/cDNA samples or Mock DNA ($1\ \mu\text{L}$) were amplified in duplicate $10\ \mu\text{L}$ reactions containing 1X KAPA 2G Fast Ready Mix (Merck) and $0.5\ \mu\text{M}$ of each primer. The PCR cycling conditions were 95°C for 3 min followed by 25 cycles of 95°C for 15 s, 55°C for 15 s and 72°C for 2 s, and a final extension of 72°C for 30 s. Duplicate reactions were pooled and PCR amplification was verified by gel electrophoresis. To normalize the samples before pooling and sequencing, the Sequalprep Normalization Plate (96) kit (Invitrogen) was used according to the manufacturer's instructions. After binding, washing and elution, the environmental and the mock community PCR products were pooled into one tube. The clean-up of the PCR amplicon pool was achieved with the Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions with elution in $30\ \mu\text{L}$ of molecular biology grade water. DNA was quantified with the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen) according to the manufacturer's instructions. Approximately 700 ng of DNA was pooled with 350 ng of barcoded PCR products from a different project (sequencing the 16S rRNA genes of lichen-associated bacteria) and sent out to the sequencing company Fasterris for library preparation and sequencing (described below). Library preparation involved ligation on PCR using the TruSeq DNA Sample Preparation Kit (Illumina) according to the manufacturer's instructions except that 5 PCR cycles were used instead of 10 cycles. The library was sequenced on one Illumina MiSeq run using the $2 \times 300\ \text{bp}$ protocol with MiSeq version 3.0 chemistry and the base-calling

TABLE 1 Brief description of the study sites.

Station	Date (2011)	Z _{ML} (m)	Euphotic layer (m)	Depth (m)	Temp (°C)	Chl a (µg L ⁻¹) ^a	NO ₃ ⁻ +NO ₂ ⁻ (µM) ^b	PO ₄ ³⁻ (µM) ^b	Si(OH) ₄ (µM) ^b	DFe (nM) ^c
R	26/10	105 ± 15	100	20	2.18	0.32	25.71	1.81	12.07	0.088 ^d
				60	2.10	0.27	26.22	1.82	12.25	0.076
				150	1.79	0.07	26.64	1.90	14.34	0.181
				300	1.84	b.d	33.45	2.30	38.67	0.33
F	07/11	38 ± 7	35	20	4.31	2.88	19.15	0.91	7.23	0.26
				70	3.19	0.34	27.10	1.93	14.25	0.30
				150	2.52	0.04	29.94	2.01	21.04	n.a
				300	2.88	b.d	33.79	2.26	37.02	0.40
A3	16/11	153 ± 15	38	20	2.25	1.64	25.51	1.75	18.41	0.18 ^d
				80	2.16	2.12	26.31	1.79	19.24	0.14
				160	2.15	2.29	26.49	1.81	19.46	n.a
				300	1.91	0.05	33.13	2.25	42.29	0.66
E5	19/11	46 ± 13	55	20	3.26	1.10	25.25	1.71	11.53	0.06
				80	2.98	0.92	26.07	1.83	12.59	0.10
				150	1.94	0.20	28.42	2.00	20.10	0.11
				300	1.94	0.12 ^d	34.06	2.36	48.52	0.23

Note: Sampling depths and corresponding environmental parameters are given. The mixed layer depth (Z_{ML}) is the mean ± sd of all CTD casts performed during the occupation of each station. The Z_{ML} is based on a difference in sigma of 0.02 to the surface value. The euphotic layer is defined as the depth with 1% of surface PAR.

Abbreviations: b.d, below detection; n.a, not available.

^aData are from (Lasbleiz et al. 2014).

^bData are from Blain et al. (2015).

^cData are from Quéroué et al. (2015).

^dValues are provided for the depth closest to the sampling depth used for sequence analysis within the Z_{ML} (40 m at R and 37 m at A3-2).

pipeline MiSeq Control Software 2.4.1.3, RTA 1.18.54.0, and CASAVA-1.8.2. The error rate was measured by spiking the library with about 0.5% of a PhiX library and mapping the reads onto the PhiX reference genome.

3 | BIOINFORMATIC ANALYSIS

3.1 | Sequence preprocessing

Paired-end Illumina sequence reads were initially processed using version 8 or version 9 of the USEARCH-64 package (Drive5) for read merging, quality filtering, parsing of the reads according to the project and primer removal, with some commands from QIIME version 1.9.1 (Caporaso et al., 2010) and *mothur* version 1.33.3 (Schloss et al., 2009).

Several criteria were chosen in the preprocessing steps to minimize errors and reduce over-inflation of diversity (Bokulich et al., 2013); (a) only three mismatches were allowed in the overlapping region when merging the paired-end reads, (b) quality filtering was carried out after merging using a stringent expected error of 1.0

(Edgar & Flyvbjerg, 2015), (c) zero mismatches were allowed in the barcode when demultiplexing, and (d) exact matches to both primers required. More details are found in Appendix 3.

3.2 | Definition of zOTUs

The UNOISE3 algorithm from the USEARCH-32 package (version 11) was used to denoise the 16S rRNA or *rbcl* reads (removal of sequencing errors and chimaeras), allowing the recovery of correct biological sequences, also termed zero-radius OTUs (zOTUs) (see Edgar, 2016) for the description and validation of the previous algorithm UNOISE2).

3.3 | Classification of 16S rRNA sequences

zOTUs were classified with the USEARCH-32 v11 classifier SINTAX and the SILVA v123 sequence database using a cutoff of 0.8 (see Appendix 3). zOTUs were aligned using the *mothur* SILVA SEED

alignment (v123) and further filtering was carried out to remove zOTUs unassigned at Kingdom level, or those that were not well aligned. The filtered zOTU table, together with the corresponding taxonomy was converted to biom format for import into the R package Phyloseq (McMurdie & Holmes, 2013) for further sequence manipulation and data exploration.

For bacterial diversity, the PhyloSeq object was filtered to remove Archaea, Eukaryotic, mitochondrial, and plastid sequences.

Of the 232 zOTUs identified as plastids, around 93% were only classified as far as Order level as Chloroplasts using the SILVA 123 database. To obtain a classification at a lower taxonomic level, all zOTUs classified as Chloroplasts were extracted and classified further with the SINTAX classifier as carried out for the bacterial sequences, using the Phytoref Database (Decelle et al., 2015; see Appendix 3) before reimporting into Phyloseq.

3.4 | Taxonomic assignment of rbcL sequences

In the absence of an rbcL sequence database for marine phytoplankton, zOTU sequences were analyzed by NCBI Blastn (Altschul, 1997) and the hit with the highest percentage similarity including a named species was used to assign the taxonomy with the taxonomic levels retrieved from the WoRMS database (WoRMS Editorial Board, 2016). Sequences with >99% sequence identity to a named species were assigned to species level unless two different species showed the same % identity. In this case, the zOTU was assigned as far as the genus level. zOTUs showing a % ID of 97% or greater were assigned to the genus level. Blastn results with hits <97% were assigned to the lowest taxonomic level possible coherent with the different hits proposed in the results table (e.g., for Zotu74, there were only two hits at 96% with the named species *Teleaaulax* sp. and *Phagioselmis* sp., both belonging to the family Geminigeraceae and therefore the zOTU was assigned at this level). In the case of blastn results that retrieved only uncultured sequence hits, even after allowing 250 hits, uncultured clones that had been situated in phylogenetic trees in published data (e.g., Bhadury & Ward, 2009) were used to attempt to assign the sequences at least at the class level.

zOTUS assigned to non-marine groups including the Phylum Streptophyta (land plants) and the Class Trebouxiophyceae as defined by the World Register of Marine Species (WoRMS) were filtered out of the OTU table as most likely representing cross-talk OTUs from the lichen project analyzed in the same run.

The rbcL sequences were imported into ARB (arb-6.0), aligned and a neighbor-joining tree was constructed. The tree was exported in Newick format to be combined with the zOTU table, tax_table, and meta-data using the Phyloseq package.

3.5 | Data analysis

Data manipulation and analysis was performed using various packages of the R (version 3.5.1) platform (ape, phyloseq, vegan, mixOmics,

ggplot2, dplyr, tidyr, factoextra, dendextend, matrixStats, stringr, reshape2, tidyverse). Alpha diversity metrics were calculated with the phyloseq package and the Shannon Entropy values (x) were transformed to effective species diversities according to (Jost, 2006) by taking the exponential of the values (exp(x)). Rarefaction of zOTU tables was chosen before clustering methods since this normalization method may better reflect the clustering of samples according to their biological origin (Weiss et al., 2017). Nonmetric dimensional scaling (NMDS) plots were generated by Phyloseq using Bray–Curtis dissimilarity (upper three depths) and dendrograms were constructed using hclust() and Ward.D2 clustering. The significance of the bacterial community structure differences represented in the nMDS plots was tested by PERMANOVA using the adonis2 routine in vegan after first verifying the homogeneity of group dispersions with betadisper.

For the figures focusing on specific taxonomic groups (Figures 4–6, A6 and A7), a reduced data set was used to eliminate potentially spurious bacterial zOTUs from Illumina cross-talk artefacts. In brief, the variance was calculated for the zOTUs as carried out previously (West et al., 2018), retaining only those zOTUs with a variance $>1 \times 10^{-7}$. This reduced the zOTU table to 679 taxa but still retained $91.3\% \pm 2.3\%$ of the total read count when considering the upper three depths of the four stations or $89.7\% \pm 4.3\%$ when including the 300 m depth samples.

Mantel and partial Mantel tests were performed in vegan using mantel() and mantel.partial() based on the Pearson correlation method. Before correlation analysis, environmental variables (depth, temperature, salinity, oxygen, phosphate, silicic acid, nitrite, nitrate, ammonium, and dissolved organic carbon (DOC)) were z-score transformed and a Euclidean distance matrix was calculated. The bacteria and phytoplankton zOTU matrices underwent Hellinger transformation before calculating Bray–Curtis distance matrices. The correlation heatmaps were generated using spls() and cim() in the mixOmics package with clr-transformed data (Chun & Keleş, 2010; Rohart et al., 2017).

4 | RESULTS AND DISCUSSION

4.1 | Environmental context

Phytoplankton and their associated bacterial communities in the particulate fractions (0.8–3 and $>3 \mu\text{m}$) were characterized at three stations influenced by Fe fertilization to the north-east and south-east the Kerguelen Islands Plateau (stations F, E5, and A3) and at a reference station in the HNLC region west of the plateau (station R) (Figure 1). All stations were located south of the polar front except for station F, and their hydrographic conditions are described in detail in (Park et al., 2014) and (d'Ovidio et al., 2015). Station R was sampled as a reference HNLC site west of the Kerguelen plateau with surface waters characterized by low concentrations of dissolved iron (dFe; $<0.088 \text{ nM}$) (Qu  rou   et al., 2015) and Chl *a* ($<0.32 \mu\text{g L}^{-1}$) (Lasbleiz et al., 2016). The three stations subject to natural Fe fertilization had

overall higher concentrations of dFe (0.18–0.66 nM) in the wind mixed layer (Z_{ML}) and Chl *a* concentration varied considerably, with values ranging from 0.9 (E5) to 4.0 $\mu\text{g L}^{-1}$ (F) among the Fe-fertilized stations. This patchiness in Chl *a* and other biological parameters was due to spatial and temporal variability in the bloom development (Lasbleiz et al., 2016). By contrast, the major inorganic nutrients N and P were similar across sites and characteristic of this region (Blain et al., 2015). Stations R and A3 had deeper Z_{ML} (105 and 153 m, respectively) than stations F and E5 (38 and 46 m, respectively) (Table 1) with the euphotic layer depth (1% surface PAR) roughly equaling the Z_{ML} at all sites except for station A3, where PAR penetrated to only a small fraction of the Z_{ML} .

4.2 | Patterns of bacterial and phytoplankton diversity in size-fractionated samples

Bacterial and phytoplankton diversity were assessed at the molecular level by generating sequence reads from the 16S rRNA gene (region V3-V4) and the RubisCO large subunit gene (*rbcL*) respectively from DNA and RNA samples (reverse-transcribed to cDNA). The *rbcL* gene primers used to target the RubisCO form 1D, present in chromophytic (red) algal groups representing the dominant phytoplankton species in the Southern Ocean (e.g., Wright et al., 2010), but would not target the green algal groups belonging to Chlorophyta and cyanobacteria or dinoflagellates (e.g., Tabita et al., 2008). We chose to use *rbcL* since it is plastid-encoded and therefore excludes heterotrophic protists that are often overrepresented in 18S rRNA sequence libraries (Georges et al., 2014; Heywood et al., 2011). The *rbcL* gene also performs better than the 18S rRNA gene to distinguish closely related taxa (Evans et al., 2007). In addition, since *rbcL* is a key gene involved in photosynthetic carbon fixation, the analysis of *rbcL* RNA transcripts could be used as a proxy for phytoplankton activity (Wawrik et al., 2003).

After paired-end read joining and quality filtering, 0.9 million 16S rRNA gene reads were denoised followed by clustering into 1719 bacterial zOTUs (100% identity; zero-radius OTUs) and 232 plastid zOTUs (see Experimental Procedures). Similarly, just over 0.5 million *rbcL* reads were clustered into 222 *rbcL* zOTUs. Although 16S rRNA zOTUs assigned to plastids are often filtered out in bacterial diversity studies, here they were retained as a useful comparison to the picture of phytoplankton diversity revealed by the *rbcL* sequences. The proportional abundance of the 16S rRNA bacterial sequences was highest in the 0.8–3 μm fraction whereas plastid sequences were proportionally more abundant in the >3 μm size fraction except for the 300 m

samples where bacterial sequences dominated (Figure A1). To examine the diversity and composition of the active fraction of bacterial and phytoplankton communities, 16S rRNA and *rbcL* gene sequences were also generated from reverse-transcribed RNA (cDNA) for the majority of the samples for 16S rRNA and from four surface samples at the A3 and F bloom samples for *rbcL* (see Table A1 for details).

The alpha diversity index Shannon Entropy was calculated from rarefied zOTU tables and then transformed to true diversities or “effective species number” for both phytoplankton and bacteria (see Experimental Procedures). This is defined as the number of evenly represented species to give the value of the calculated index which allows an intuitive comparison between the samples (Jost, 2006). For phytoplankton, the 16S rRNA and *rbcL* data gave diversity estimates of a similar magnitude for a given sample (Table A2). The R and E5 >3 μm size fraction phytoplankton diversity was over threefold higher than that at the bloom stations (A3 and F) whereas the 0.8–3 μm fractions of all stations gave similar true diversities of around 14–20. Similarly, microscopic observations revealed a higher diatom diversity, corresponding mainly to the >3 μm size fraction, at the E5 and R sites as compared to A3 and F (Lasbleiz et al., 2016). Bacterial diversity was generally higher in the >3 μm fraction than in the 0.8–3 μm fraction in the two surface depths. When ranks of bacterial diversity and depth were compared (1–4, lowest to highest diversity or surface to deep; Table 2), diversity was strongly positively correlated with depth ($R^2 = 1$) for the 0.8–3 μm fraction, while for the >3 μm fraction there was no significant correlation between diversity and depth.

4.3 | Distinct phytoplankton community structures at the different stations

Phytoplankton communities clustered by a fraction and by station for the upper two surface depths (20–80 m) for both phytoplankton marker genes which showed good agreement apart from the samples from station F (see tanglegram, Figure A2). For both fractions, the communities at R and E5 were most closely related to each other and fell into a cluster with the communities from A3. Whereas the *rbcL* sequences grouped the 0.8–3 and >3 μm fractions in a separate cluster, the 16S rRNA sequences grouped the A3 and F >3 μm samples with one of the 0.8–3 μm fraction samples. The difference in clustering may be explained by the higher diversity of the *rbcL* sequences compared to the 16S rRNA sequences which allowed the differentiation of two distinct *Thalassiosira* sp. zOTUs at A3 and F whereas a single common zOTU was detected in the 16S rRNA data set (Figure 2, discussed below).

TABLE 2 Impact of different factors on bacterial community structure tested by PERMANOVA analysis.

Factor	Station	Fraction	Depth (m)	Sample type	Residuals
Values	A3,F,E5,R	0.8 μm , >3 μm	20,60–80,150	DNA, RNA	
DNA(<i>n</i> = 24), RNA (<i>n</i> = 24)	0.24 (0.001)	0.14 (0.001)	0.11 (0.001)	0.09 (0.001)	0.42

Note: R values are indicated with the *p* Value in brackets.

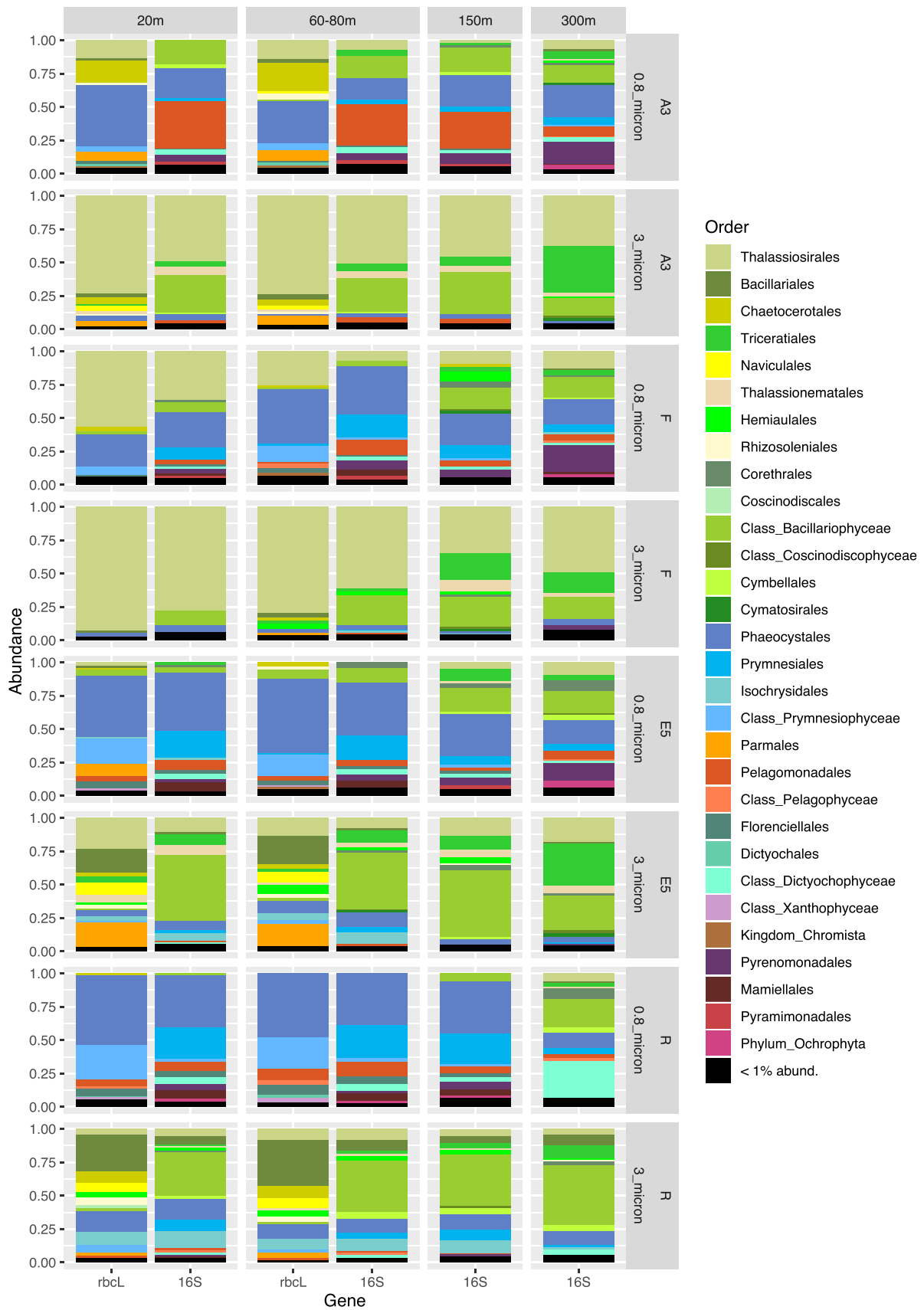


FIGURE 2 Comparison of phytoplankton taxonomic community structure (order level) at different depths at the four sampling stations in the 0.8–3 μm or $>3 \mu\text{m}$ size fractions as revealed by plastid *rbcl* and 16S rRNA marker gene sequencing. Diatom orders are shown in shades of green, yellow, and beige from Thalassiosirales to Cymbellales.

In the $>3\ \mu\text{m}$ fraction of the upper two surface depths, Bacillariophyceae accounted for $>90\%$ of the sequences at stations A3 and F but exhibited lower relative abundances ($>60\%$) in the same size fraction at the E5 and R stations (Figure A3). The lower relative abundance of diatoms at E5 and R compared to A3 and F was also observed by (Lasbleiz et al., 2016) who assessed the abundance of different phytoplankton groups by microscopy counts and flow cytometry. Diatoms accounted for 72%–74% of the phytoplankton groups in the euphotic zone at A3 and F but only 38% and 7% respectively at E5 and R (Lasbleiz et al., 2016). Conversely, in the 0.8–3 μm fraction Prymnesiophyceae largely dominated the R and E5 sequences in the present study. Although Prymnesiophyceae were the most abundant class in the majority of the 0.8–3 μm fraction samples of the bloom stations, attaining $>50\%$ of sequences at 80 m at F and 20 m at A3, Bacillariophyceae sequences remained relatively abundant even in this smaller size fraction. At the two surface depths for which *rbcl* and 16S rRNA plastid sequences were available for comparison (20–80 m, Table A1), the two analyses yielded consistent relative abundances of the two major classes Bacillariophyceae (diatoms) and Prymnesiophyceae (haptophytes). Notable differences between the two marker genes included the higher relative abundance of Pelagophyceae and Cryptophyceae zOTUs in the 16S rRNA data set compared to the *rbcl* data set, whereas the silica-containing Bolidophyceae group that is closely related to diatoms was observed in the *rbcl* but not the 16S rRNA data set. It is not uncommon to see differences in the relative abundance of phytoplankton groups when using different primer pairs due to the effects of primer bias as observed by (Shi et al., 2011). In their study, Pelagophyceae and Dictyophyceae sequences were favored with one pair whereas Prymnesiophyceae dominated the sequences obtained with the second pair, underlining the utility of multiple marker gene comparisons to obtain a fuller picture of phytoplankton diversity.

At the Order level, in the $>3\ \mu\text{m}$ fraction, both marker genes indicated a dominance of Thalassiosirales diatoms in the bloom stations A3 (73% of total sequences) and particularly at F, where these sequences accounted for $>90\%$ of the total *rbcl* sequences (Figure 2). Surprisingly, at A3, higher relative abundances of the diatom order Chaetocerotales were observed in the *rbcl* sequences of the smaller 0.8–3 μm fraction compared to the $>3\ \mu\text{m}$ fraction whereas the converse was true at R. For the 16S rRNA sequences, the A3 and F stations shared the same main zOTU that was assigned at Order level (Thalassiosirales; data not shown) whereas the *rbcl* sequences distinguished the dominant zOTU at these two stations (Figure 3). A single *rbcl* zOTU assigned as *Thalassiosira ritscheri* accounted for 60%–75% of the sequences at F and less than 0.4% at A3, whereas the second most dominant zOTU was assigned as *Thalassiosira* sp. (97% similarity to *Thalassiosira antarctica*) and accounted for 49%–54% of the sequences at A3 and less than 2.1% at F. The $>3\ \mu\text{m}$ phytoplankton communities at the HNLC station R and E5 showed a more diverse and even distribution with higher relative abundances of *Fragilariopsis* spp., *Pseudo-nitzschia* spp., *Chaetoceros* spp. and *Eucampia* spp. The haptophyte *rbcl* sequences recovered from the 0.8–3 μm fraction were dominated by two roughly equally abundant

Phaeocystis zOTUs at the bloom stations A3 and F (zOTU3 and zOTU6) and a single *Phaeocystis antarctica* (zOTU3) at E5 and R. The major *Thalassiosira* zOTU1 also accounted for a significant fraction of the *rbcl* sequences in the 0.8–3 μm fraction at F (Figure 3).

We compared the relative abundance of diatoms as assessed by *rbcl* gene sequencing from the $>3\ \mu\text{m}$ samples to microscopy counts carried out on CTD rosette samples (Lasbleiz et al., 2016 and Table A3). To allow comparisons, the species identified by microscopy and the *rbcl* zOTUs were grouped, when possible, at the order level, apart from a few low-abundance species that did not fall into the orders listed and were grouped instead under “Class_Bacillariophyceae.” The main differences were observed for the Thalassiosirales and Chaetocerotales at the bloom stations where our *rbcl* sequence data set appeared to overestimate Thalassiosirales abundance: in particular the dominance of *Thalassiosira* spp. at A3 diverges from the microscopy counts where *Chaetoceros* (Hyalochaete) spp. accounted for 60% and 80% of carbon biomass and live cell abundance respectively (Lasbleiz et al., 2016, Table A3). This could be explained by our 65 μm prefiltering step that could have reduced the number of the *Chaetoceros* chain-forming diatoms in the sequence datasets or less efficient lysis of certain species with highly silicified frustules at the nucleic extraction step (Luddington et al., 2016). For station R, the diatom order abundances were relatively similar between microscopy counts and sequencing data except for the absence of Thalassionematales in the *rbcl* data.

The analysis of the comparatively small RNA (cDNA) *rbcl* data set obtained from the surface depth of 20 m at the bloom stations A3 and F provided further evidence that the dominant diatoms affiliated to *Thalassiosira* spp. were also active, given the similar relative abundances in the $>3\ \mu\text{m}$ fraction for RNA or DNA. Interestingly the 0.8–3 μm size fraction showed more pronounced differences between the RNA and DNA data sets, notably at station A3. Whereas *Thalassiosira* sequences were also the major contributor to the station F RNA sequences, at A3, *Chaetoceros* spp. accounted for $>65\%$ of the RNA sequences (Figure A4).

Therefore, despite the apparent overestimation of *Thalassiosira* spp. in the *rbcl* DNA sequence data, we were nonetheless able to detect the main diatom species and highlight a dominant Chaetocerotales activity in the RNA data set at A3 in agreement with (Lasbleiz et al., 2016). Furthermore, we were also able to assess the diversity and activity of the smaller haptophyte phytoplankton that are difficult to identify by light microscopy. The evaluation of protist diversity at the KEOPS2 stations by 18S rRNA gene sequencing (Georges et al., 2014) revealed the presence of small diatom species at the bloom stations such as *Thalassiosira* sp., *Minidiscus* sp. and *Coscinodiscus* sp. but failed to detect certain species such as *Fragilariopsis* sp. 18S rRNA Haptophyta sequences were also dominated by *Phaeocystis antarctica* in agreement with our *rbcl* and 16S rRNA data. However, Bacillariophyceae were underestimated seemingly due to the overrepresentation of alveolates that possess a high rRNA operon copy number (Zhu et al., 2005) and that are known to dominate 18S rRNA gene metabarcoding databases (Vaulot et al., 2022).

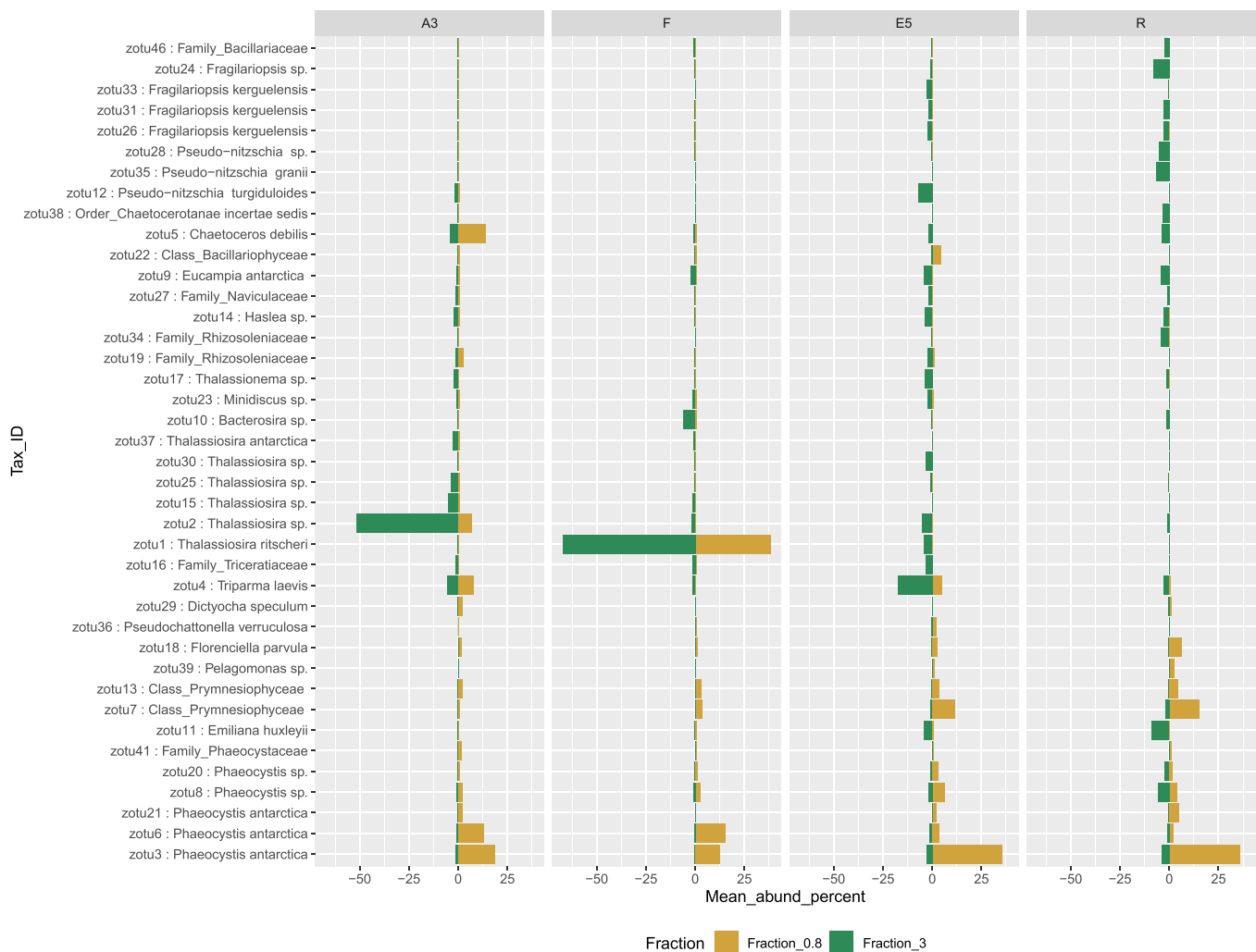


FIGURE 3 Mirror column chart to compare the relative mean abundance (upper two surface depths) of the 40 most abundant rbcL zOTUs in the 0.8–3 μm and >3 μm fractions at the four stations.

4.4 | Bacterial community composition

Beta diversity analysis on the rarefied bacterial 16S rRNA gene (DNA and cDNA) sequences revealed that for the bloom stations A3 and F, the community composition was similar between the DNA and RNA (cDNA) samples except for the A3 0.8–3 μm fraction samples at the three depths from 20 to 150 m where the DNA samples clustered separately from the RNA samples. The converse was observed at station R where all DNA and RNA samples clustered separately apart from the 20 m >3 μm sample (Figure A5). Community structure comparisons of marker gene sequences originating from RNA and DNA to reveal activity (protein synthesizing potential; Blazewicz et al., 2013) and presence, respectively, have been applied to marine and freshwater bacterial communities (Denef et al., 2016; Hunt et al., 2013). 16S rRNA gene sequences from RNA and DNA were also compared with meta-transcriptomic analysis on the samples to investigate the impact of a phytoplankton bloom on bacterial diversity (Wemheuer et al., 2014). The high coupling between the composition of the total (DNA) and the active (RNA) bacterial communities observed only in bloom stations

suggests a stronger role of bottom-up control in the shaping of community composition in response to an increased supply of photosynthate. In contrast, higher fractions of dormant cells and slow-growing bacteria could further explain the distinct clustering between the total and active bacterial communities at the HNLC site.

Bacterial community composition was significantly different between sample stations, size fractions, sample type and depth, as illustrated in the nMDS plots in Figure 4 and supported by the PERMANOVA analysis in Table 2. These differences can also be observed by the relative abundance of the different taxonomic groups in Figure A6 and the hierarchical clustering shown in Figure A5. Communities in both size fractions from below the Z_{ML} clustered separately from those within the Z_{ML} and this was particularly noticeable at station F where the Z_{ML} was relatively shallow (See Z_{ML} in Table 1 and Figure A5). At a glance, even at the order level, there were clear differences in the taxonomic composition of the different size fractions. Our results are in line with studies from other marine environments showing that particles of different sizes harbor distinct bacterial communities, and that specific bacterial taxa are found preferentially associated with certain particle sizes

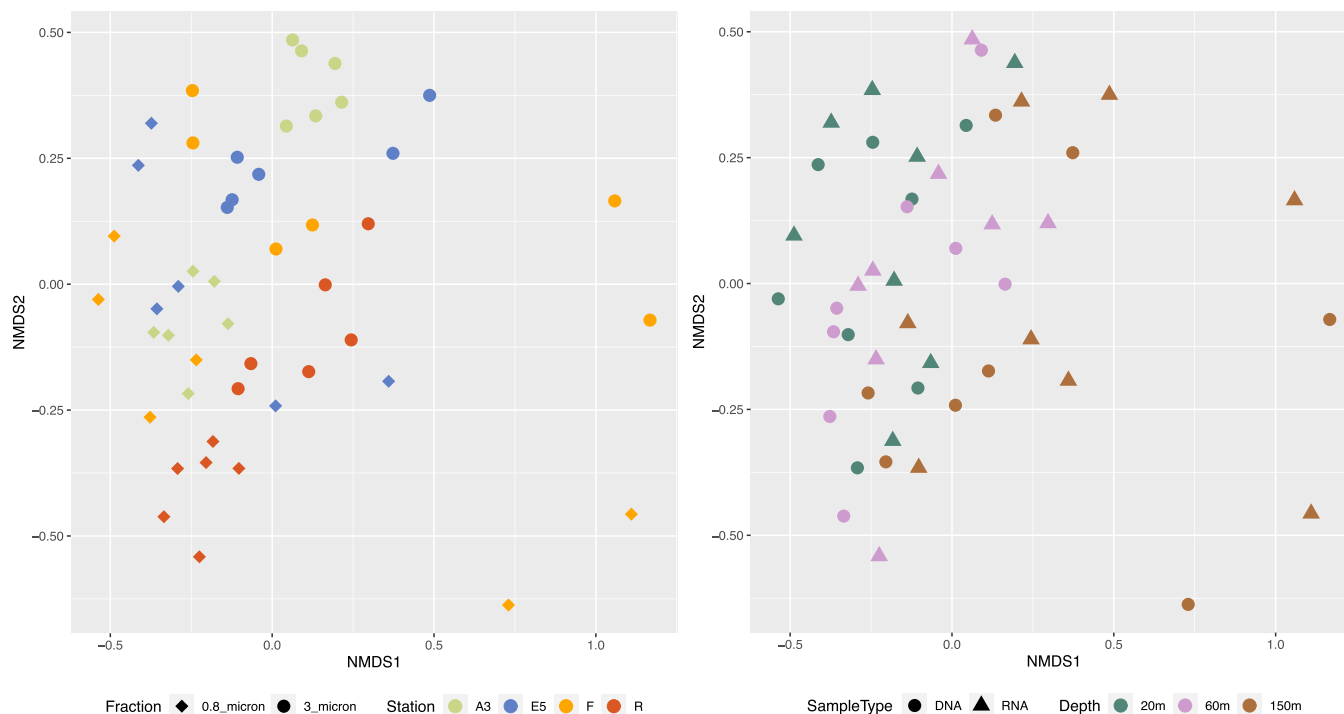


FIGURE 4 The similarity of the bacterial communities in different size fractions collected from the four stations in three depth layers (20, 60–80, and 150 m) assessed by 16S rRNA marker gene analysis of DNA or RNA (cDNA) using Bray–Curtis distances and nMDS. The same nMDS is shown in two separate panels to show the separation of the samples by fraction and station (left) and by depth (right). Size fraction 0.8-micron (0.8–3 μm), 3-micron (>3 μm).

(Mestre et al., 2017, 2020; Yung et al., 2016). To reveal differences in the preferences of different bacterial taxa for the particulate niches represented by the two size fractions, a Particle-Association Niche Index (PAN-Index) was calculated for each zOTU as described previously (Mestre et al., 2018; Salazar et al., 2015). The PAN-Index of taxa found predominantly in the 0.8–3 μm fraction showed a value towards 0 whereas taxa found more often in the >3 μm fraction had a value closer to 1 (Figure 5).

In the mixed layer depths (Z_{ML}), the >3 μm particle niche was characterized by several Proteobacteria families including Alteromonadaceae, Colwelliaceae, Pseudomonadaceae, Halieaceae, Bdellovibrionaceae, Erythrobacteraceae, and Sphingomonadaceae. All the families belonging to the Bacteroidetes phylum were preferentially associated with this larger particle fraction size, as were the Planctomycete families Phycisphaeraceae and Planctomycetaceae and those belonging to the Verrucomicrobia. Low PAN index clades comprised several Oceanospirillales families including SAR86, different SAR11 clades, Moraxellaceae, Marine group B (SAR324), Porticocaceae (SAR92), and the families Rhodospirillaceae and Rhodobacteraceae. *Synechococcus* zOTUs were also most abundant in the 0.8–3 μm fraction as observed previously (Mestre et al., 2017), most likely due to their larger cell size than heterotrophic bacteria. However, relatively high abundances (5.5%–6%) were also observed in the >3 μm fractions at stations R and F, respectively, which could be explained by the formation of aggregates, attachment to particles or due to symbiotic relationships. *Synechococcus*-like symbionts have

been detected in dinoflagellates, tintinnids and radiolarians (Foster et al., 2006) and in Foraminifera (Bird et al., 2017). Furthermore, high abundances of *Synechococcus psbO* marker gene sequences were also recovered in metagenomic libraries from large-size fractions collected by Tara Oceans (Pierella Karlusich et al., 2023).

4.5 | Abiotic versus biotic variables influencing bacterial community structure

To explore the influence of abiotic versus biotic variables on the bacterial community structure, partial Mantel tests were performed between dissimilarity matrices of bacterial 16S rRNA gene zOTUs versus phytoplankton zOTUs assessed either by plastid 16S rRNA gene or by *rbcL* gene diversity (biotic) and versus an environmental variable matrix (abiotic). In the upper two surface depths (20 m and 60–80 m) bacterial communities were significantly correlated with phytoplankton (based on 16S rRNA) for both size fractions, but not to environmental variables (Table A4). When all depth layers were considered, bacterial communities were correlated with phytoplankton but more strongly with environmental variables, highlighting the impact of the environmental conditions set by depth on bacterial community composition. For the *rbcL* data, only the 20 m and 60–80 m surface samples were available, and for both size fractions, bacterial communities showed significant correlations with phytoplankton but not with environmental variables. These observations

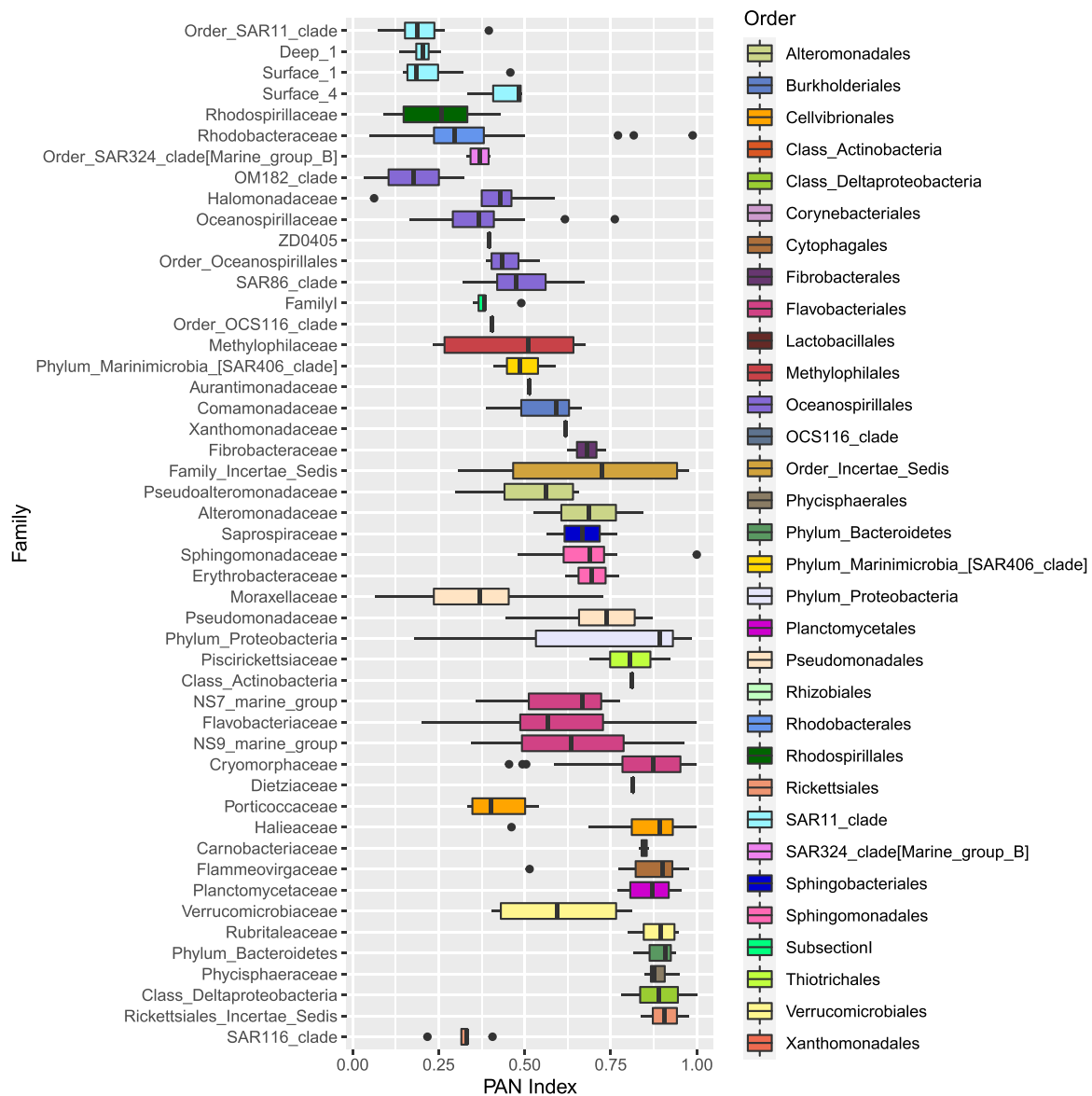


FIGURE 5 The size fraction preference of zOTUs in the mixed layer depths (Z_M) at the four stations inferred by calculating a Particle associated Niche Index (PAN index) and illustrated in a boxplot with zOTUs grouped at the family level. zOTUs with a PAN index close to zero are prevalent in the 0.8–3 μm fraction whereas those with a PAN index toward one are more abundant in the >3 μm fraction. Subsection I, Family I corresponds to *Synechococcus* sp., Family Incertae Sedis to *Marinicella* sp. and the family Dietziaceae is affiliated to the Corynebacteriales order. Families are color-coded according to order as illustrated in the legend.

are in line with the conclusions from several recent studies showing that biotic interactions could be more important for structuring plankton community composition than abiotic factors (Arandia-Gorostidi et al., 2022; Lima-Mendez et al., 2015; Liu et al., 2019; Needham et al., 2018).

4.6 | Phytoplankton-bacteria co-occurrences

Given the strong correlation between the phytoplankton and bacterial communities indicated by the partial Mantel tests, we identified specific correlations between the most relatively abundant phytoplankton

and bacterial zOTUs (representing >1% of the total sequences in at least one sample) for the upper two surface depths (20 m and 60–80 m) by sparse partial least squares (PLS) regression provided by the MixOmics R package. Correlations were done between the *rbcl* zOTUs and 16S rRNA bacterial zOTUs from DNA samples (Figure 6a) and also between the 16S rRNA plastid and bacterial zOTUs from DNA and from RNA (cDNA) potentially representing the active zOTU interactions (Figure A7). Further information on the relative abundance of the strongly correlating bacterial zOTUs and their potential activity is presented below the heatmap (6b).

This correlation analysis presented in Figure 6a identified 56 and 43 strongly positively correlated ($r > 0.6$) phytoplankton and bacterial

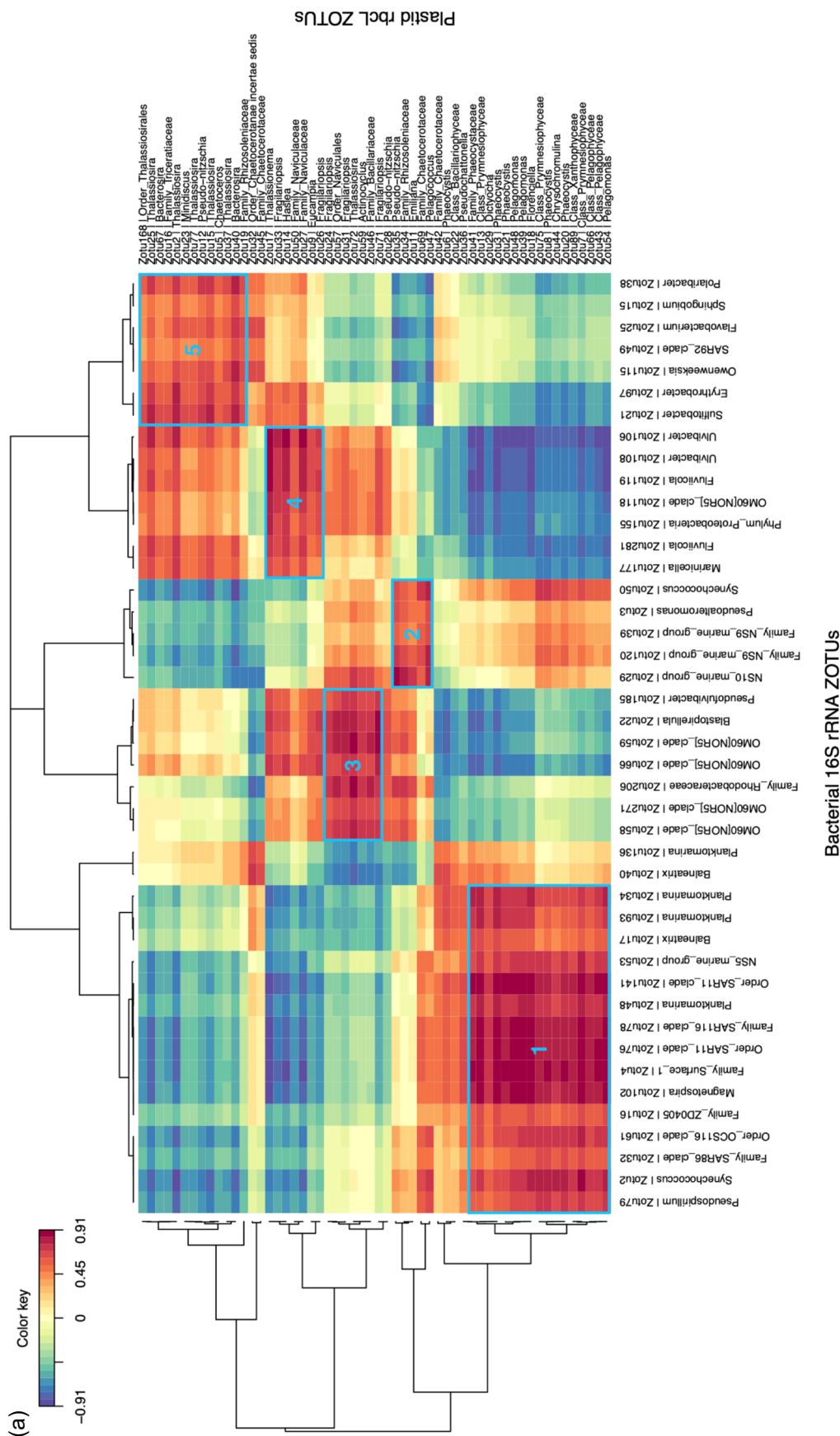


FIGURE 6 (a) Clustered image heatmap showing the most significant correlations between the surface bacterial 16S rRNA zOTUs and phytoplankton rbcL zOTUs (upper 2 surface depths). The main bacterial zOTU clusters are outlined in blue and illustrated similarly in (b). (b) For clarity, the summed relative abundances (DNA) of bacterial surface zOTUs in the 2 size fractions (0.8–3 and >3 μm fraction) and their relative activity (ratio of RNA:DNA relative abundance) is shown only for the two main bloom stations A3 and F (omitting the meander station E5) for comparison with the HNLC station R.

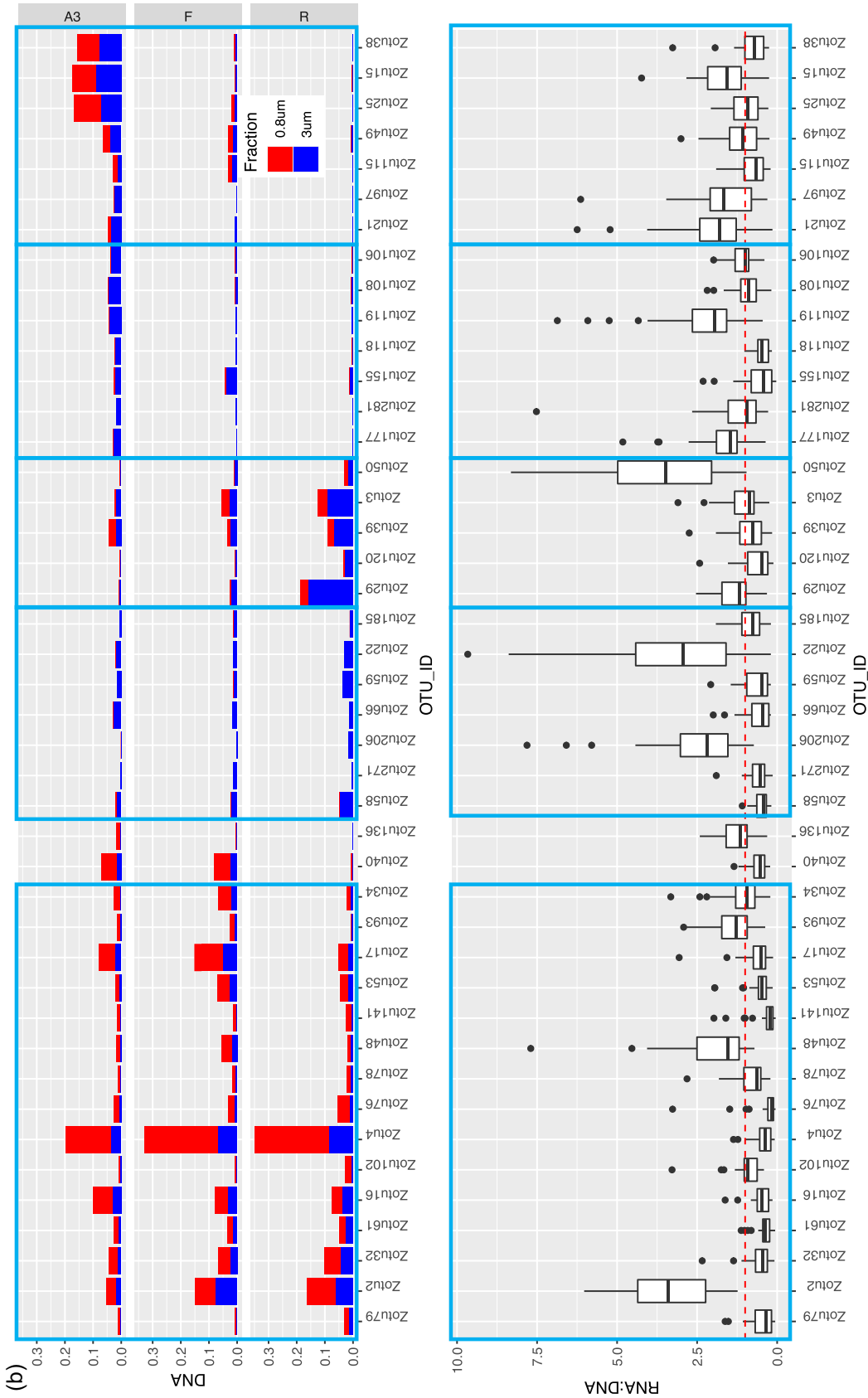


FIGURE 6 (Continued).

zOTUs, respectively. Together, the phytoplankton zOTUs correlating to bacterial zOTUs contributed from 59.8% to 89.9% ($n = 12$) of total phytoplankton communities at the two surface depths when excluding station F, for which the relative abundances of correlating zOTUs was significantly lower (8.9% to 63.7%, $n = 4$), in particular in the $>3 \mu\text{m}$ fraction (Table A5). In contrast, the contribution of bacterial zOTUs correlating to phytoplankton was lower but more stable (31.0%–49.9% $n = 16$) with no particular patterns between stations (Table A5). The difference between phytoplankton and bacteria in the contribution of correlated zOTUs to total communities could reflect a general tendency of phytoplankton to associate with specific bacteria while from the bacterial point of view, a significant proportion of the community could behave like generalists and thrive on particles, irrespective of the identity of phytoplankton present within the particles. This could suggest that the bacterial zOTUs identified as correlated to phytoplankton zOTUs could be involved in true partnerships with specific phytoplankton taxa.

Bacterial and phytoplankton zOTUs with strong positive correlations ($r > 0.6$) were distributed in several distinct clusters driven in part by the compositional differences observed between the two size fractions (Figure 6a). Cluster 1 contained taxa associated with the 0.8–3 μm fraction with the strongest correlations between the Prymnesiophyceae and Pelagophyceae phytoplankton zOTUS with SAR11, SAR116, and OCS116 clades, *Magnetospira* and 3 *Planktomarina* (Roseobacter RCA clade) zOTUS. These latter zOTUs (zOTU48, 93, 34) were the only members of the cluster showing a higher relative activity as observed in Figure 6b (excluding *Synechococcus*). In this cluster, the presence of typical pelagic oligotrophs such as SAR11 and the Roseobacter RCA clade (*Planktomarina*) and their high correlations with Pelagophyte (*Pelagomonas* sp.) and Haptophyte (*Phaeocystis* sp.) zOTUs suggest that for some of these bacterial taxa, interactions with phytoplankton might be more prevalent than previously recognized. Such interactions could rely on the exchange of organic sulfur compounds such as DMSP or sulfonates, which prymnesiophytes are known to produce, that could fulfill the organic sulfur requirements of taxa that have lost the sulphate assimilation pathway such as the SAR11 (Durham et al., 2019; Tripp et al., 2008). Our data suggest that physical association, either transient or more stable, could play a part in the nutrient exchange between these organisms, as was also suggested by a few other studies that identified SAR11 ecotypes in $>0.8 \mu\text{m}$ size fractions (Allen et al., 2012; Mestre et al., 2017). Our findings differ from those of a *Phaeocystis*-dominated bloom in an Antarctic polyna, where the *Phaeocystis* cells were in a colony form, and consequently fell into the larger $>3 \mu\text{m}$ fraction, correlating instead with the Gammaproteobacteria SAR92 (Delmont et al., 2014). In our study, SAR92 was most abundant at A3 but did not exceed more than 2.8% of the total bacterial community and did not show a particular size fraction preference (Figure 5b, Porticocaceae), although the correlation with one SAR92 zOTU was strongest with diatom zOTUs (Figure 6, cluster 5). In a previous study, samples were taken during the peak and decline of the diatom bloom at A3, and SAR92 was both relatively abundant and active in the surface waters (Obernosterer et al., 2011; West et al., 2008), supporting the correlation observed in this study.

Cluster 5 highlighted correlations between several *Thalassiosira*, *Pseudo-nitzschia*, *Bacterosira*, and *Chaetoceros* zOTUs, with two Roseobacter zOTUs (*Sulfitobacteria* and *Erythrobracter*), Bacteroidetes zOTUs (*Polaribacter*, *Flavobacterium* and *Owenweeksia*), SAR92, and *Sphingobium*. These zOTUs were generally more abundant at A3 than F or R and also showed higher relative activity, particularly for the Roseobacters and *Sphingobium* (Figure 6b). *Sphingobium* is often described as a polycyclic aromatic hydrocarbon degrader but reports of the interactions of this taxon with phytoplankton are scarce and would warrant further investigation. Roseobacter OTUs were also amongst the most active members of the bacterial communities dominating the peak of the bloom at the same station A3 sampled during the KEOPS1 study (Obernosterer et al., 2011; West et al., 2008). Associations between diatoms and members of the highly diverse Roseobacter clade are among the best-characterised marine microbial partnerships that have important repercussions for marine biogeochemical cycles (e.g., Buchan et al., 2014). Many phytoplankton are vitamin B₁₂ auxotrophs (Croft et al., 2005) and therefore these associations typically revolve around the trading of essential metabolic currencies, such as the provision of the sulfonates in exchange for vitamin B₁₂, known to be synthesized by many Roseobacter clade members (Durham et al., 2015). There are several reports of the specific association of *Sulfitobacter* strains with *Pseudo-nitzschia* species that promote the growth of the diatoms through their provision of B vitamins (Andrew et al., 2022) or by the secretion of the growth-promoting hormone indole-3-acetic acid (Amin et al., 2015). An extensive study of Roseobacter biogeography from an Australian time series revealed the co-occurrence of *Planktomarina* and *Sulfitobacter* with high DMSP-producing phytoplankton, such as *Phaeocystis* (O'Brien et al., 2022). Although also correlated to *Thalassiosira*, one *Sulfitobacter* amplicon sequence variant (ASV) showed the highest correlations to *Phaeocystis* that exceeded those of SAR11 and SAR86 (O'Brien et al., 2022). This is in contrast to our size fractionated data where the *Planktomarina* and *Sulfitobacteria* were significantly correlated to *Phaeocystis* and diatoms respectively.

Cluster 4 implied potential interactions between *Thalassionema*, *Fragilariopsis*, *Eucampia*, and Naviculaceae with several Flavobacteria zOTUs (*Ulvibacter* and *Fluviicola*) and the Gammaproteobacteria *Marinicella* and OM60[NOR5]. Clusters 2 and 3 revealed correlations between the NS9 and NS10 bacteroidetes groups and *Pseudoalteromonas* with both haptophytes and diatoms (Cluster 2) and between the Gammaproteobacteria OM60[NOR5], the Bacteroidetes *Pseudofulvibacter* zOTU and the Planctomycete *Blasitopirellula* with diatom zOTUs (Cluster 3).

For bacterial zOTUs preferentially found in the $>3 \mu\text{m}$ size fraction and correlating to diatom zOTUs, the majority belonged to the phylogenetic groups Flavobacteria, Gammaproteobacteria, and Roseobacter that are typically found in association with phytoplankton (Abell & Bowman, 2005; Liu et al., 2019; Tran et al., 2023). The distinct associations observed for specific diatoms could result from substrate preferences that vary among bacterial taxa (Bunse et al., 2021; Krüger et al., 2019; Xing et al., 2015) responding to differences in the composition of organic matter released by diatom species

(Mühlenbruch et al., 2018). Genomic analyses of Flavobacteria demonstrate that they possess many genes involved in adhesion and gliding motility as well as a unique secretion system (Fernández-Gómez et al., 2013; Gavriilidou et al., 2020) that allows them to employ their diverse repertoire of enzymatic capabilities for the degradation of complex organic matter (Kappellmann et al., 2019). These traits are thought to facilitate a particle-attached lifestyle, including physical association with phytoplankton. Our observation of several Flavobacteria members with relative abundances correlating with those of several diatoms in the >3 µm size fraction supports the idea that members of this group thrive in the phycosphere and further suggests specific interactions with phytoplankton.

In conclusion, our combined bacterial 16S rRNA and plastid gene metabarcoding approach on size-fractionated samples from DNA and RNA extracts revealed new insights into potential bacteria-phytoplankton associations that could pave the way for further laboratory or in situ studies. In contrast to the 18S rRNA marker gene, the 16S rRNA and *rbcl* plastid marker genes targeted directly the phytoplankton and gave a complementary picture of the diversity. Furthermore, the more variable *rbcl* gene was a more informative marker, allowing the discrimination of the dominant phytoplankton species between the two main bloom stations. In analyzing size-fractionated samples, we were able to reveal distinct groups of bacterial taxa preferentially co-occurring with the different phytoplankton size classes dominated by *Phaeocystis* spp. and diatoms, two phytoplankton groups with key ecological and biogeochemical roles. By including metabarcoding sequencing from bacterial 16S rRNA and phytoplankton *rbcl* transcripts, we were able to add a new dimension to our data by revealing a potential elevated relative activity (RNA:DNA > 1) of all highly correlated Roseobacter zOTUs, *Sphingobium*, and the *Thalassiosira* and *Chaetoceros* zOTUs. The function and importance of the potential interactions we revealed could be investigated in more depth by combining metagenomic and metatranscriptomic analyses of key phytoplankton-bacteria partners in co-culture experiments and natural samples during bloom transitions from diatoms to *Phaeocystis*.

AUTHOR CONTRIBUTIONS

Nyree J. West: Conceptualization (supporting); data curation (lead); formal analysis (lead); investigation (equal); methodology (equal); writing—original draft (lead); writing—review & editing (equal). **Marine Landa:** Investigation (equal); methodology (equal), writing—review & editing (equal). **Ingrid Obernosterer:** Conceptualization (lead); funding acquisition (equal); investigation (equal); methodology (equal); project administration (lead), resources (equal); writing—review & editing (equal).

ACKNOWLEDGMENTS

The financial support for this research was provided by CNRS-INSU-LEFE-CYBER, ANR-10-BLAN-0614, and the French Polar Institute Paul-Émile Victor (IPEV). Our gratitude extends to Stéphane Blain, the KEOPS2 project coordinator, Bernard Quéguiner, the chief scientist, Captain Bernard Lassiette, and the crew of the R/V Marion Dufresne II for their invaluable assistance and support during the project. We are grateful to the Bio2Mar core facility for granting us

access to their equipment. Our thanks go to Karine Leblanc for her contribution to the diatom counts, Yan Liu for her help with data analysis, and Urania Christaki for her insightful discussions. The graphical abstract was created using Biorender.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Sequence data are openly available in the European Nucleotide Archive under accession number PRJEB23734: <https://www.ebi.ac.uk/ena/browser/view/PRJEB23734>.

ETHICS STATEMENT

None required.

ORCID

Nyree J. West  <http://orcid.org/0000-0003-3763-6996>

Ingrid Obernosterer  <http://orcid.org/0000-0002-2530-8111>

REFERENCES

- Abell, G. C. J., & Bowman, J. P. (2005). Colonization and community dynamics of class flavobacteria on diatom detritus in experimental mesocosms based on Southern Ocean seawater. *FEMS Microbiology Ecology*, 53, 379–391.
- Allen, L. Z., Allen, E. E., Badger, J. H., McCrow, J. P., Paulsen, I. T., Elbourne, L. D. H., Thiagarajan, M., Rusch, D. B., Neelson, K. H., Williamson, S. J., Venter, J. C., & Allen, A. E. (2012). Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *The ISME Journal*, 6, 1403–1414.
- Altschul, S. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389–3402.
- Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsek, M. R., Moran, M. A., & Armbrust, E. V. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, 522, 98–101.
- Andrew, S., Wilson, T., Smith, S., Marchetti, A., & Septer, A. N. (2022). A tripartite model system for Southern Ocean diatom-bacterial interactions reveals the coexistence of competing symbiotic strategies. *ISME Communications*, 2, 1–12.
- Arandia-Gorostidi, N., Krabberød, A. K., Logares, R., Deutschmann, I. M., Scharek, R., Morán, X. A. G., González, F., & Alonso-Sáez, L. (2022). Novel interactions between phytoplankton and bacteria shape microbial seasonal dynamics in coastal ocean waters. *Frontiers in Marine Science*, 9, 901201.
- Bell, W., & Mitchell, R. (1972). Chemotactic and growth responses of marine bacteria to algal extracellular products. *The Biological Bulletin*, 143, 265–277.
- Bhadury, P., & Ward, B. B. (2009). Molecular diversity of marine phytoplankton communities based on key functional genes1. *Journal of Phycology*, 45, 1335–1347.
- Biggs, T. E. G., Alvarez-Fernandez, S., Evans, C., Mojica, K. D. A., Rozema, P. D., Venables, H. J., Pond, D. W., & Brussaard, C. P. D. (2019). Antarctic phytoplankton community composition and size structure: Importance of ice type and temperature as regulatory factors. *Polar Biology*, 42, 1997–2015.
- Bird, C., Darling, K. F., Russell, A. D., Davis, C. V., Fehrenbacher, J., Free, A., Wyman, M., & Ngwenya, B. T. (2017). Cyanobacterial

- endobionts within a major marine planktonic calcifier (*Globigerina bulloides*, Foraminifera) revealed by 16S rRNA metabarcoding. *Biogeosciences*, 14, 901–920.
- Blain, S., Capparas, J., Guéneuguès, A., Obernosterer, I., & Oriol, L. (2015). Distributions and stoichiometry of dissolved nitrogen and phosphorus in the iron-fertilized region near kerguelen (Southern Ocean). *Biogeosciences*, 12, 623–635.
- Blain, S., Quéguiner, B., Armand, L., Belviso, S., Bombled, B., Bopp, L., Bowie, A., Brunet, C., Brussaard, C., Carlotti, F., Christaki, U., Corbière, A., Durand, I., Ebersbach, F., Fuda, J. L., Garcia, N., Gerringa, L., Griffiths, B., Guigue, C., ... Wagener, T. (2007). Effect of natural iron fertilization on carbon sequestration in the Southern Ocean. *Nature*, 446, 1070–1074.
- Blain, S., Rembauville, M., Crispì, O., & Obernosterer, I. (2021). Synchronized autonomous sampling reveals coupled pulses of biomass and export of morphologically different Diatoms in the Southern Ocean. *Limnology and Oceanography*, 66, 753–764.
- Blazewicz, S. J., Barnard, R. L., Daly, R. A., & Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: Limitations and uses. *The ISME Journal*, 7, 2061–2068.
- Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A., & Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from illumina amplicon sequencing. *Nature Methods*, 10, 57–59.
- Buchan, A., LeClerc, G. R., Gulvik, C. A., & González, J. M. (2014). Master recyclers: Features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology*, 12, 686–698.
- Bunse, C., Koch, H., Breider, S., Simon, M., & Wietz, M. (2021). Sweet spheres: Succession and CAZyme expression of marine bacterial communities colonizing a mix of alginate and pectin particles. *Environmental Microbiology*, 23, 3130–3148.
- Bunse, C., & Pinhassi, J. (2017). Marine bacterioplankton seasonal succession dynamics. *Trends in Microbiology*, 25, 494–505.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, 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., ... Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335–336.
- Christaki, U., Guéneuguès, A., Liu, Y., Blain, S., Catala, P., Colombet, J., Debeljak, P., Jardillier, L., Irion, S., Planchon, F., Sassenhagen, I., Sime-Ngando, T., & Obernosterer, I. (2021). Seasonal microbial food web dynamics in contrasting Southern Ocean productivity regimes. *Limnology and Oceanography*, 66, 108–122.
- Chun, H., & Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72, 3–25.
- Cirri, E., & Pohnert, G. (2019). Algae–bacteria interactions that balance the planktonic microbiome. *New Phytologist*, 223, 100–106.
- Costas-Selas, C., Martínez-García, S., Delgadillo-Nuño, E., Justel-Díez, M., Fuentes-Lema, A., Fernández, E., & Teira, E. (2024). Linking the impact of bacteria on phytoplankton growth with microbial community composition and co-occurrence patterns. *Marine Environmental Research*, 193, 106262.
- Coyne, K. J., Wang, Y., & Johnson, G. (2022). Algicidal bacteria: A review of current knowledge and applications to control harmful algal blooms. *Frontiers in Microbiology*, 13, 871177.
- Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, 438, 90–93.
- Davidson, A. T., Scott, F. J., Nash, G. V., Wright, S. W., & Raymond, B. (2010). Physical and biological control of protistan community composition, distribution and abundance in the seasonal ice zone of the Southern Ocean between 30 and 80°E. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57, 828–848.
- Decelle, J., Romac, S., Stern, R. F., Bendif, E. M., Zingone, A., Audic, S., Guiry, M. D., Guillou, L., Tessier, D., Le Gall, F., Gourvil, P., Dos Santos, A. L., Probert, I., Vaulot, D., de Vargas, C., & Christen, R. (2015). PhytoREF: A reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Molecular Ecology Resources*, 15, 1435–1445.
- Delmont, T. O., Hammar, K. M., Ducklow, H. W., Yager, P. L., & Post, A. F. (2014). *Phaeocystis antarctica* blooms strongly influence bacterial community structures in the amundsen sea polynya. *Frontiers in Microbiology*, 5, 646.
- Denef, V. J., Fujimoto, M., Berry, M. A., & Schmidt, M. L. (2016). Seasonal succession leads to Habitat-Dependent differentiation in ribosomal RNA:DNA ratios among freshwater lake bacteria. *Frontiers in Microbiology*, 7, 606.
- Dinasquet, J., Landa, M., & Obernosterer, I. (2022). SAR11 clade microdiversity and activity during the early spring blooms off kerguelen island, Southern Ocean. *Environmental Microbiology Reports*, 14, 907–916.
- d'Ovidio, F., Della Penna, A., Trull, T. W., Nencioli, F., Pujol, M.-I., Rio, M.-H., Park, Y. H., Cotté, C., Zhou, M., & Blain, S. (2015). The biogeochemical structuring role of horizontal stirring: Lagrangian perspectives on iron delivery downstream of the kerguelen plateau. *Biogeosciences*, 12, 5567–5581.
- Durham, B. P., Boysen, A. K., Carlson, L. T., Groussman, R. D., Heal, K. R., Cain, K. R., Morales, R. L., Coesel, S. N., Morris, R. M., Ingalls, A. E., & Armbrust, E. V. (2019). Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nature Microbiology*, 4, 1706–1715.
- Durham, B. P., Sharma, S., Luo, H., Smith, C. B., Amin, S. A., Bender, S. J., Dearth, S. P., Van Mooy, B. A. S., Campagna, S. R., Kujawinski, E. B., Armbrust, E. V., & Moran, M. A. (2015). Cryptic carbon and sulfur cycling between surface ocean plankton. *Proceedings of the National Academy of Sciences*, 112, 453–457.
- Eddy, T. D., Bernhardt, J. R., Blanchard, J. L., Cheung, W. W. L., Colléter, M., du Pontavice, H., Fulton, E. A., Gascuel, D., Kearney, K. A., Petrik, C. M., Roy, T., Rykaczewski, R. R., Selden, R., Stock, C. A., Wabnitz, C. C. C., & Watson, R. A. (2021). Energy flow through marine ecosystems: Confronting transfer efficiency. *Trends in Ecology & Evolution*, 36, 76–86.
- Edgar, R. C. (2016). UNOISE2: Improved error-correction for illumina 16S and ITS amplicon sequencing. bioRxiv 081257.
- Edgar, R. C., & Flyvbjerg, H. (2015). Error filtering, pair assembly, and error correction for next-generation sequencing reads. *Bioinformatics* 31, 401.
- Evans, K. M., Wortley, A. H., & Mann, D. G. (2007). An assessment of potential diatom “barcode” genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in sellaphora (Bacillariophyta). *Protist*, 158, 349–364.
- Faircloth, B. C., & Glenn, T. C. (2012). Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. *PLoS One*, 7, e42543.
- Fernández-Gómez, B., Richter, M., Schüller, M., Pinhassi, J., Acinas, S. G., González, J. M., & Pedrós-Alió, C. (2013). Ecology of marine bacteroidetes: A comparative genomics approach. *The ISME Journal*, 7, 1026–1037.
- Foster, R. A., Carpenter, E. J., & Bergman, B. (2006). Unicellular cyanobionts in open ocean dinoflagellates, radiolarians, and tintinnids: Ultrastructural characterization and immuno-localization of phycoerythrin and nitrogenase1. *Journal of Phycology*, 42, 453–463.
- Fuhrman, J. A., Hewson, I., Schwalbach, M. S., Steele, J. A., Brown, M. V., & Naeem, S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proceedings of the National Academy of Sciences*, 103, 13104–13109.
- Gavriliidou, A., Gutleben, J., Versluis, D., Forgiarini, F., van Passel, M. W. J., Ingham, C. J., Smidt, H., & Sijkema, D. (2020). Comparative genomic

- analysis of flavobacteriaceae: Insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. *BMC Genomics*, 21, 569.
- Georges, C., Monchy, S., Genitsaris, S., & Christaki, U. (2014). Protist community composition during early phytoplankton blooms in the naturally iron-fertilized kerguelen area (Southern Ocean). *Biogeosciences*, 11, 5847–5863.
- Hernandez-Magana, A. E., Liu, Y., Debeljak, P., Crispi, O., Marie, B., Koedooder, C., & Obernosterer, I. (2021). Prokaryotic diversity and activity in contrasting productivity regimes in late summer in the kerguelen region (Southern Ocean). *Journal of Marine Systems*, 221, 103561.
- Heywood, J. L., Sieracki, M. E., Bellows, W., Poulton, N. J., & Stepanauskas, R. (2011). Capturing diversity of marine heterotrophic protists: One cell at a time. *The ISME journal*, 5, 674–684.
- Hunt, B. P. V., Espinasse, B., Pakhomov, E. A., Cherel, Y., Cotté, C., Delegrange, A., & Henschke, N. (2021). Pelagic food web structure in high nutrient low chlorophyll (HNLC) and naturally iron fertilized waters in the kerguelen islands region, Southern Ocean. *Journal of Marine Systems*, 224, 103625.
- Hunt, D. E., Lin, Y., Church, M. J., Karl, D. M., Tringe, S. G., Izzo, L. K., & Johnson, Z. I. (2013). Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. *Applied and Environmental Microbiology*, 79, 177–184.
- Irion, S., Christaki, U., Berthelot, H., L'Helguen, S., & Jardillier, L. (2021). Small phytoplankton contribute greatly to CO₂ fixation after the diatom bloom in the Southern Ocean. *The ISME Journal: Multidisciplinary Journal of Microbial Ecology*, 15, 2509–2522.
- Jost, L. (2006). Entropy and diversity. *Oikos*, 113, 363–375.
- Kappelmann, L., Krüger, K., Hehemann, J.-H., Harder, J., Markert, S., Unfried, F., Becher, D., Shapiro, N., Schweder, T., Amann, R. I., & Teeling, H. (2019). Polysaccharide utilization loci of north sea flavobacteria as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *The ISME Journal*, 13, 76–91.
- Keller, M. D., Bellows, W. K., & Guillard, R. R. L. (1989). Dimethyl sulfide production in marine phytoplankton. In E. S. Saltzman & W. J. Cooper (Eds.), *Biogenic sulfur in the environment*. ACS symposium series (pp. 167–182). American Chemical Society.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1. <https://doi.org/10.1093/nar/gks808>
- Krüger, K., Chafee, M., Ben Francis, T., Glavina del Rio, T., Becher, D., Schweder, T., Amann, R. I., & Teeling, H. (2019). In marine bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. *The ISME Journal*, 13, 2800–2816.
- Krumhardt, K. M., Long, M. C., Sylvester, Z. T., & Petrik, C. M. (2022). Climate drivers of Southern Ocean phytoplankton community composition and potential impacts on higher trophic levels. *Frontiers in Marine Science*, 9, 1–18. <https://doi.org/10.3389/fmars.2022.916140>
- Lambert, S., Tragin, M., Lozano, J.-C., Ghigliione, J.-F., Vaulot, D., Bouget, F.-Y., & Galand, P. E. (2019). Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. *The ISME Journal*, 13, 388–401.
- Landa, M., Blain, S., Christaki, U., Monchy, S., & Obernosterer, I. (2016). Shifts in bacterial community composition associated with increased carbon cycling in a mosaic of phytoplankton blooms. *The ISME Journal*, 10, 39–50.
- Lasbleiz, M., Leblanc, K., Armand, L. K., Christaki, U., Georges, C., Obernosterer, I., & Quéguiner, B. (2016). Composition of diatom communities and their contribution to plankton biomass in the naturally iron-fertilized region of kerguelen in the Southern Ocean. *FEMS Microbiology Ecology*, 92, fiw171.
- Lasbleiz, M., Leblanc, K., Blain, S., Ras, J., Cornet-Barthaux, V., Hélias Nunige, S., & Quéguiner, B. (2014). Pigments, elemental composition (C, N, P, and Si), and stoichiometry of particulate matter in the naturally iron fertilized region of Kerguelen in the Southern Ocean. *Biogeosciences*, 11, 5931–5955.
- Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., ... Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, 348, 1262073.
- Liu, Y., Debeljak, P., Rembauville, M., Blain, S., & Obernosterer, I. (2019). Diatoms shape the biogeography of heterotrophic prokaryotes in early spring in the Southern Ocean. *Environmental Microbiology*, 21, 1452–1465.
- Luddington, I. A., Lovejoy, C., & Kaczmarek, I. (2016). Species-rich meta-communities of the diatom order thalassiosirales in the Arctic and Northern Atlantic Ocean. *Journal of Plankton Research*, 38, 781–797.
- Marchetti, A., Schruth, D. M., Durkin, C. A., Parker, M. S., Kodner, R. B., Berthiaume, C. T., Morales, R., Allen, A. E., & Armbrust, E. V. (2012). Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proceedings of the National Academy of Sciences*, 109, E317–E325.
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8, e61217.
- McParland, E. L., & Levine, N. M. (2019). The role of differential DMSP production and community composition in predicting variability of global surface DMSP concentrations. *Limnology and Oceanography*, 64, 757–773.
- Mestre, M., Borrull, E., Sala, M. M., & Gasol, J. M. (2017). Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *The ISME Journal*, 11, 999–1010.
- Mestre, M., Höfer, J., Sala, M. M., & Gasol, J. M. (2020). Seasonal variation of bacterial diversity along the marine particulate matter continuum. *Frontiers in Microbiology*, 11, 1590.
- Mestre, M., Ruiz-González, C., Logares, R., Duarte, C. M., Gasol, J. M., & Sala, M. M. (2018). Sinking particles promote vertical connectivity in the ocean microbiome. *Proceedings of the National Academy of Sciences*, 115, E6799–E6807.
- Moline, M. A., Claustre, H., Frazer, T. K., Schofield, O., & Vernet, M. (2004). Alteration of the food web along the Antarctic peninsula in response to a regional warming trend. *Global Change Biology*, 10, 1973–1980.
- Mühlenbruch, M., Grossart, H.-P., Eigemann, F., & Voss, M. (2018). Mini-review: Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. *Environmental Microbiology*, 20, 2671–2685.
- Needham, D. M., Fichot, E. B., Wang, E., Berdjeb, L., Cram, J. A., Fichot, C. G., & Fuhrman, J. A. (2018). Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. *The ISME Journal*, 12, 2417–2432.
- Obernosterer, I., Catala, P., Lebaron, P., & West, N. J. (2011). Distinct bacterial groups contribute to carbon cycling during a naturally iron fertilized phytoplankton bloom in the Southern Ocean. *Limnology and Oceanography*, 56, 2391–2401.
- O'Brien, J., McParland, E. L., Bramucci, A. R., Siboni, N., Ostrowski, M., Kahlke, T., Levine, N. M., Brown, M. V., van de Kamp, J., Bodrossy, L., Messer, L. F., Petrou, K., & Seymour, J. R. (2022). Biogeographical and seasonal dynamics of the marine roseobacter community and ecological links to DMSP-producing phytoplankton. *ISME Communications*, 2, 16.

- Park, Y.-H., Durand, I., Kestenare, E., Rougier, G., Zhou, M., d'Ovidio, F., Cotté, C., & Lee, J. H. (2014). Polar front around the kerguelen islands: An up-to-date determination and associated circulation of surface/subsurface waters. *Journal of Geophysical Research: Oceans*, 119, 6575–6592.
- Pierella Karlusich, J. J., Pelletier, E., Zinger, L., Lombard, F., Zingone, A., Colin, S., Gasol, J. M., Dorrell, R. G., Henry, N., Scalco, E., Acinas, S. G., Wincker, P., de Vargas, C., & Bowler, C. (2023). A robust approach to estimate relative phytoplankton cell abundances from metagenomes. *Molecular Ecology Resources*, 23, 16–40.
- Pollard, R., Sanders, R., Lucas, M., & Statham, P. (2007). The crozet natural iron bloom and export experiment (CROZEX). *Deep Sea Research Part II: Topical Studies in Oceanography*, 54, 1905–1914.
- Poulton, A. J., Mark Moore, C., Seeyave, S., Lucas, M. I., Fielding, S., & Ward, P. (2007). Phytoplankton community composition around the crozet plateau, with emphasis on Diatoms and phaeocystis. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54, 2085–2105.
- Quéguiner, B. (2013). Iron fertilization and the structure of planktonic communities in high nutrient regions of the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 90, 43–54.
- Quéroué, F., Sarthou, G., Planquette, H. F., Bucciarelli, E., Chever, F., van der Merwe, P., Lannuzel, D., Townsend, A. T., Cheize, M., Blain, S., d'Ovidio, F., & Bowie, A. R. (2015). High variability in dissolved iron concentrations in the vicinity of the Kerguelen Islands (Southern Ocean). *Biogeosciences*, 12, 3869–3883.
- Rohart, F., Gautier, B., Singh, A., & Lê Cao, K. A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology*, 13, e1005752.
- Salazar, G., Cornejo-Castillo, F. M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D., Arrieta, J. M., Duarte, C. M., Gasol, J. M., & Acinas, S. G. (2015). Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Molecular Ecology*, 24, 5692–5706.
- Salter, I., Lampitt, R. S., Sanders, R., Poulton, A., Kemp, A. E. S., Boorman, B., Saw, K., & Pearce, R. (2007). Estimating carbon, silica and diatom export from a naturally fertilised phytoplankton bloom in the Southern Ocean using PELAGRA: A novel drifting sediment trap. *Deep Sea Research Part II: Topical Studies in Oceanography*, 54, 2233–2259.
- Sarthou, G., Timmermans, K. R., Blain, S., & Tréguer, P. (2005). Growth physiology and fate of Diatoms in the ocean: A review. *Journal of Sea Research*, 53, 25–42.
- Schine, C. M. S., Alderkamp, A.-C., van Dijken, G., Gerringa, L. J. A., Sergi, S., Laan, P., van Haren, H., van de Poll, W. H., & Arrigo, K. R. (2021). Massive Southern Ocean phytoplankton bloom fed by iron of possible hydrothermal origin. *Nature Communications*, 12, 1211.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75, 7537–7541.
- Seymour, J. R., Amin, S. A., Raina, J.-B., & Stocker, R. (2017). Zooming in on the phycosphere: The ecological interface for phytoplankton-bacteria relationships. *Nature Microbiology*, 2, 17065.
- Shi, X. L., Lepère, C., Scanlan, D. J., & Vaulot, D. (2011). Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the south Pacific Ocean. *PLoS One*, 6, e18979.
- Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E., & Scott, S. S. (2008). Distinct form I, II, III, and IV rubisco proteins from the three kingdoms of life provide clues about rubisco evolution and structure/function relationships. *Journal of Experimental Botany*, 59, 1515–1524.
- Tran, D. Q., Milke, F., Niggemann, J., & Simon, M. (2023). The diatom *thalassiosira rotula* induces distinct growth responses and colonization patterns of roseobacteraceae, flavobacteria and gammaproteobacteria. *Environmental Microbiology*, 25, 3536–3555.
- Tripp, H. J., Kitner, J. B., Schwabach, M. S., Dacey, J. W. H., Wilhelm, L. J., & Giovannoni, S. J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature*, 452, 741–744.
- Vaulot, D., Sim, C. W. H., Ong, D., Teo, B., Biwer, C., Jamy, M., & Lopes dos Santos, A. (2022). metaPR2: A database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists. *Molecular Ecology Resources*, 22, 3188–3201.
- Wawrik, B., Paul, J., Campbell, L., Griffin, D., Houchin, L., Fuentes-Ortega, A., & Muller-Karger, F. (2003). Vertical structure of the phytoplankton community associated with a coastal plume in the Gulf of Mexico. *Marine Ecology Progress Series*, 251, 87–101.
- Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5, 27.
- Wemheuer, B., Güllert, S., Billerbeck, S., Giebel, H.-A., Voget, S., Simon, M., & Daniel, R. (2014). Impact of a phytoplankton bloom on the diversity of the active bacterial community in the Southern north sea as revealed by metatranscriptomic approaches. *FEMS Microbiology Ecology*, 87, 378–389.
- West, N. J., Lepère, C., Manes, C. L. O., Catala, P., Scanlan, D. J., & Lebaron, P. (2016). Distinct spatial patterns of SAR11, SAR86, and actinobacteria diversity along a transect in the ultra-oligotrophic south Pacific Ocean. *Frontiers in Microbiology*, 7, 234.
- West, N. J., Obernosterer, I., Zemb, O., & Lebaron, P. (2008). Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. *Environmental Microbiology*, 10, 738–756.
- West, N. J., Parrot, D., Fayet, C., Grube, M., Tomasi, S., & Suzuki, M. T. (2018). Marine cyanolichens from different littoral zones are associated with distinct bacterial communities. *PeerJ*, 6, e5208.
- WoRMS Editorial Board. (2016). World register of marine species. Available from <http://www.marinespecies.org> at VLIZ. Accessed July 1, 2023. <https://doi.org/10.14284/170>
- Wright, S. W., van den Eenden, R. L., Pearce, I., Davidson, A. T., Scott, F. J., & Westwood, K. J. (2010). Phytoplankton community structure and stocks in the Southern Ocean (30–80°E) determined by CHEMTAX analysis of HPLC pigment signatures. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57, 758–778.
- Xing, P., Hahnke, R. L., Unfried, F., Markert, S., Huang, S., Barbeyron, T., Harder, J., Becher, D., Schweder, T., Glöckner, F. O., Amann, R. I., & Teeling, H. (2015). Niches of two polysaccharide-degrading polaribacter isolates from the north sea during a spring diatom bloom. *The ISME Journal*, 9, 1410–1422.
- Yung, C.-M., Ward, C. S., Davis, K. M., Johnson, Z. I., & Hunt, D. E. (2016). Insensitivity of diverse and temporally variable particle-associated microbial communities to bulk seawater environmental parameters. *Applied and Environmental Microbiology*, 82, 3431–3437.
- Zhu, F., Massana, R., Not, F., Marie, D., & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, 52, 79–92.

How to cite this article: West, N. J., Landa, M., & Obernosterer, I. (2024). Differential association of key bacterial groups with diatoms and *Phaeocystis* spp. during spring blooms in the Southern Ocean. *MicrobiologyOpen*, 13, e1428. <https://doi.org/10.1002/mbo3.1428>

APPENDIX 1

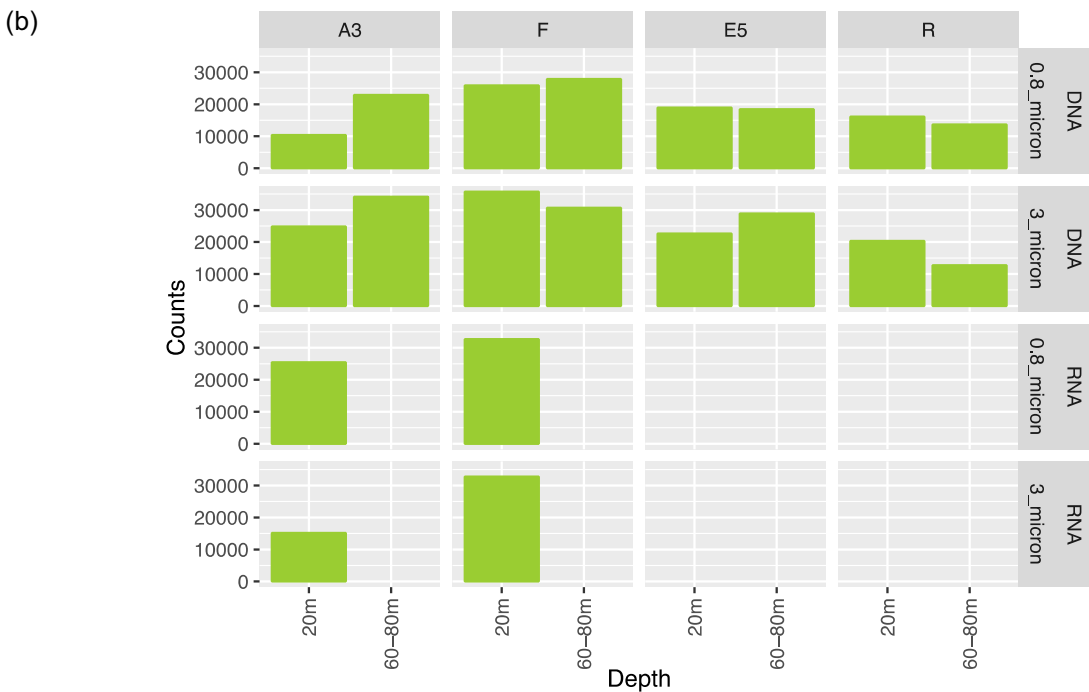
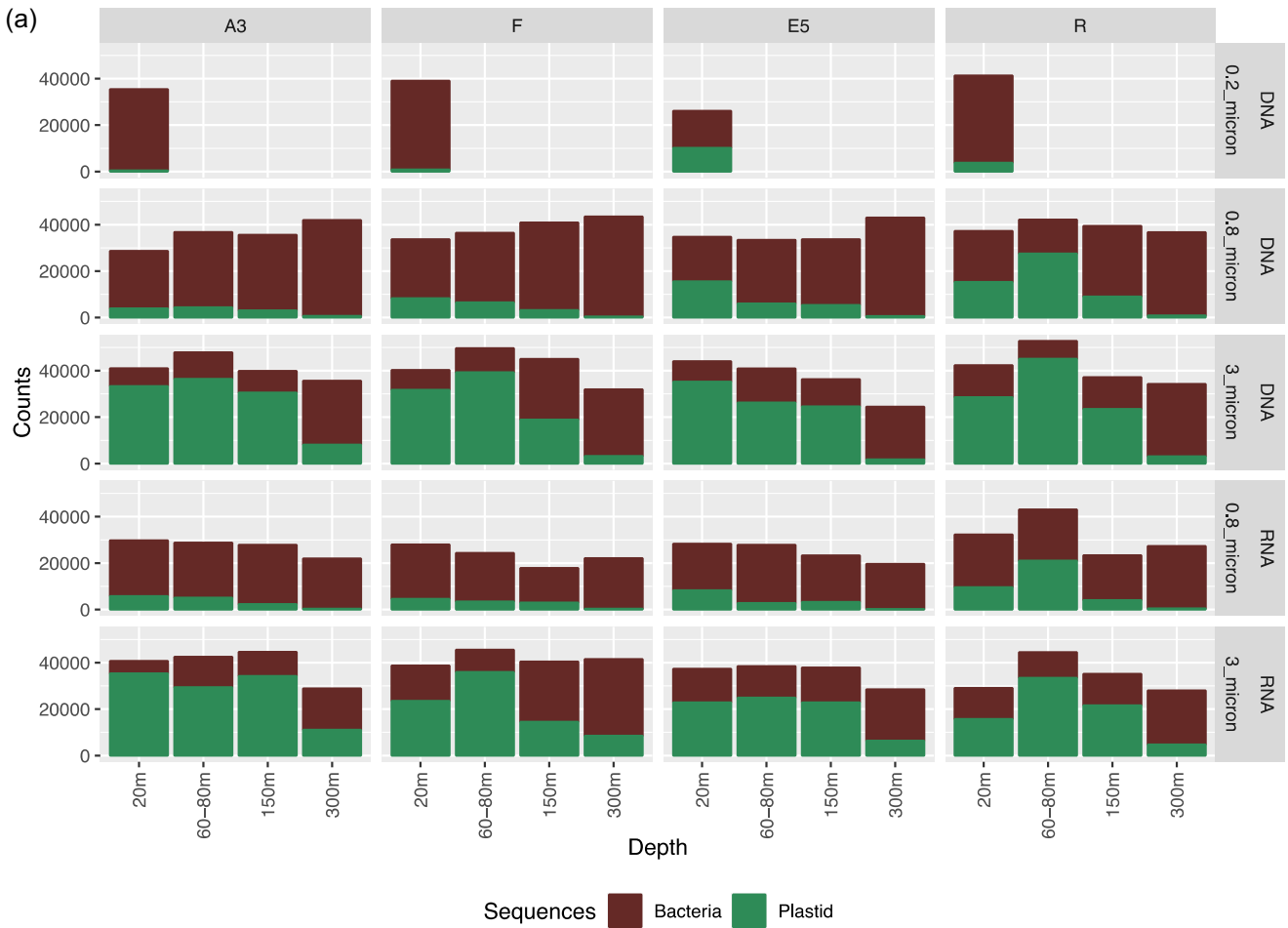


FIGURE A1 16S rRNA gene (a) and rbcL gene (b) sequencing library sizes obtained from the different sampling sites, depths, and size fractions from DNA and RNA (cDNA). The plastid and bacterial 16S rRNA gene sequences were separated in Phyloseq by subsetting out the Class “Chloroplast,” assigned by the Sintax Silva database.

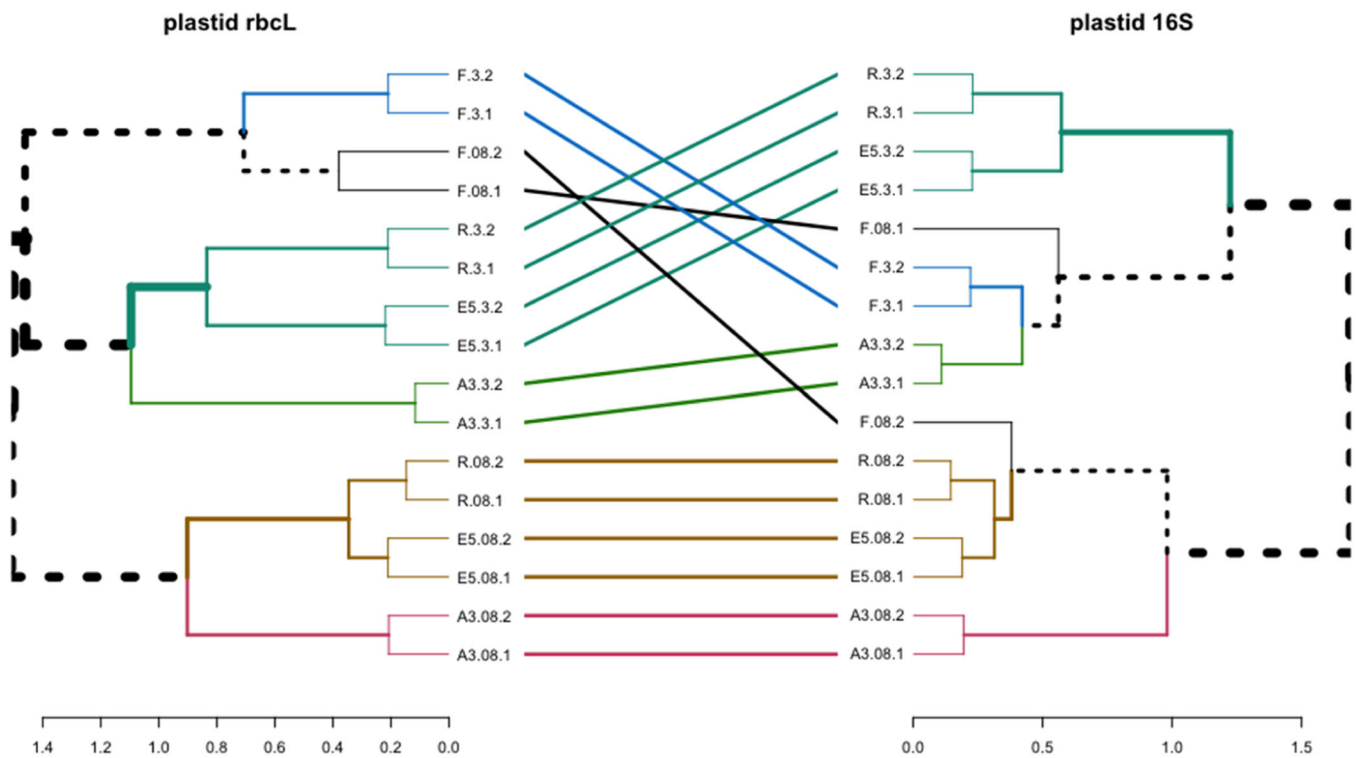


FIGURE A2 Phytoplankton community structure comparisons in the top two surface depths 20 m (1) and 60–80 m (2) at stations A3, F, E5, and R, recovered from two size fractions 0.8–3 μm (0.8) and >3 μm (3) assessed by plastid 16S rRNA and rbcL marker gene sequencing. Dissimilarity dendrograms using Bray–Curtis distances were constructed for each marker gene and their level of congruence was examined by Tanglegram. Sample notation is the Station Name.SizeFraction.Depth.

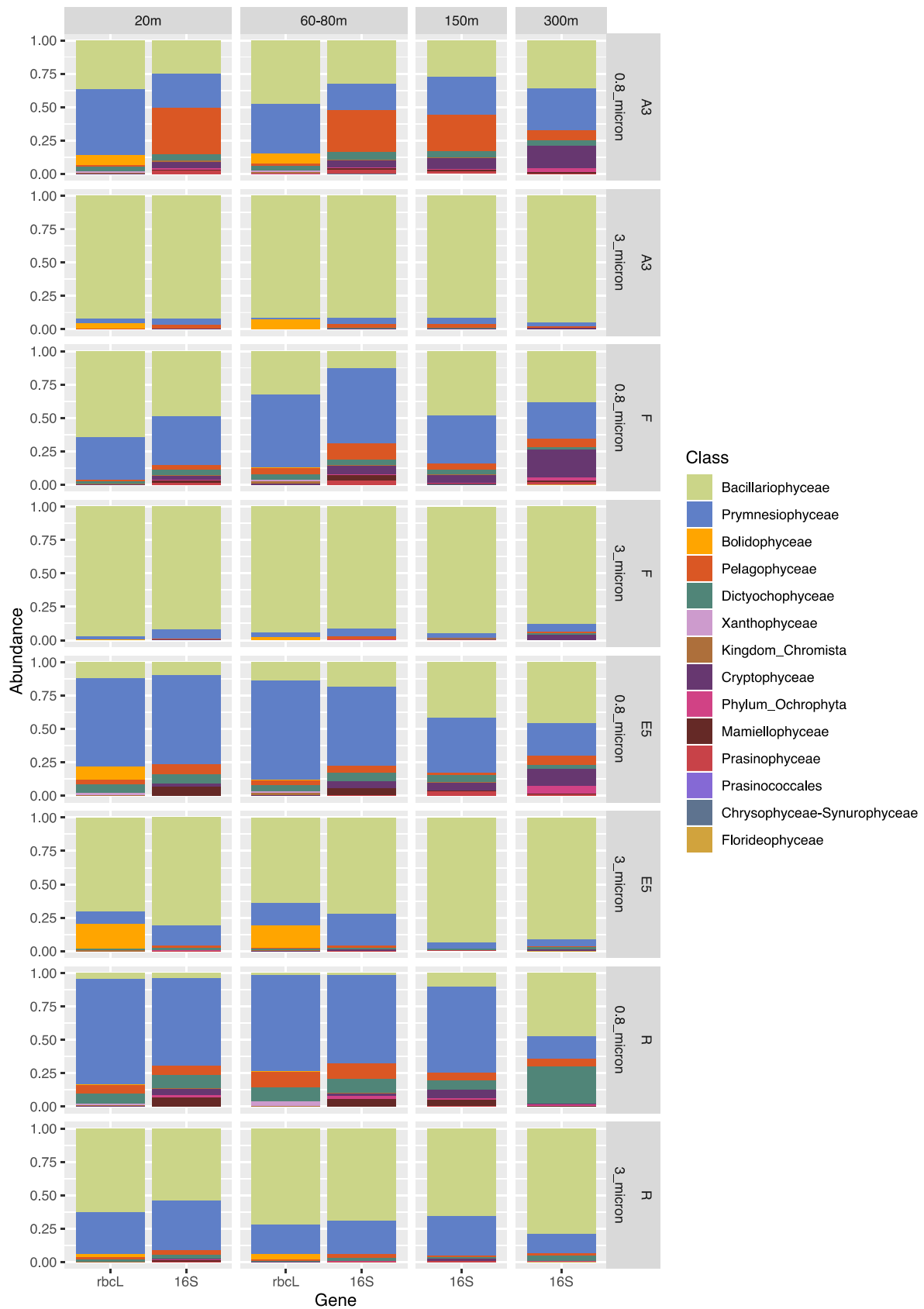


FIGURE A3 Comparison of phytoplankton taxonomic community structure (class level) at different depths at the four sampling stations in the 0.8–3 μm or >3 μm size fractions as revealed by plastid rbcL and 16S rRNA marker gene sequencing.

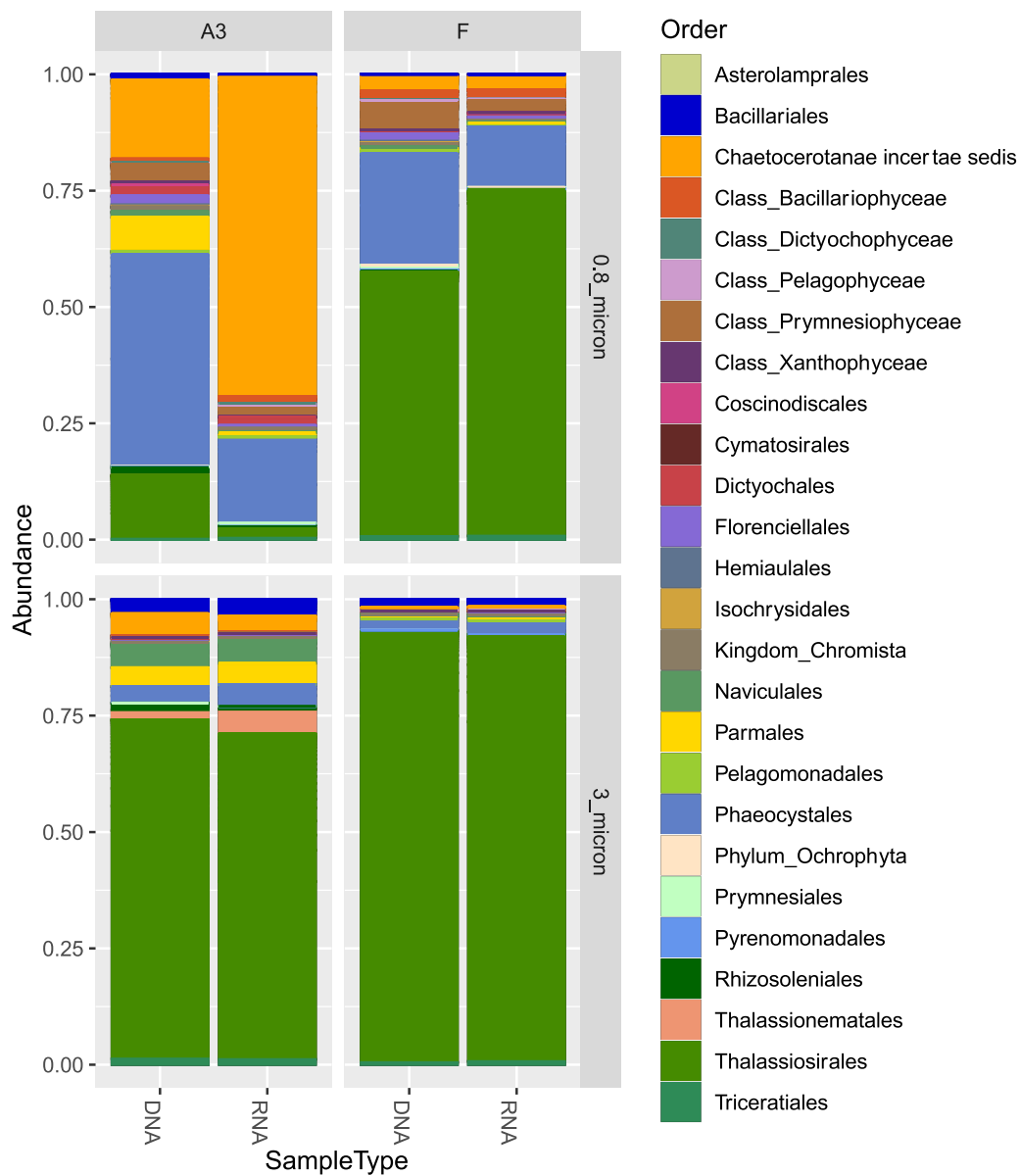


FIGURE A4 Comparison of presence and activity of phytoplankton (order level) by *rbcl* gene sequencing from DNA or RNA (cDNA) extracted from the 0.8–3 μm or >3 μm fractions collected at 20 m at the bloom stations A3 and F.

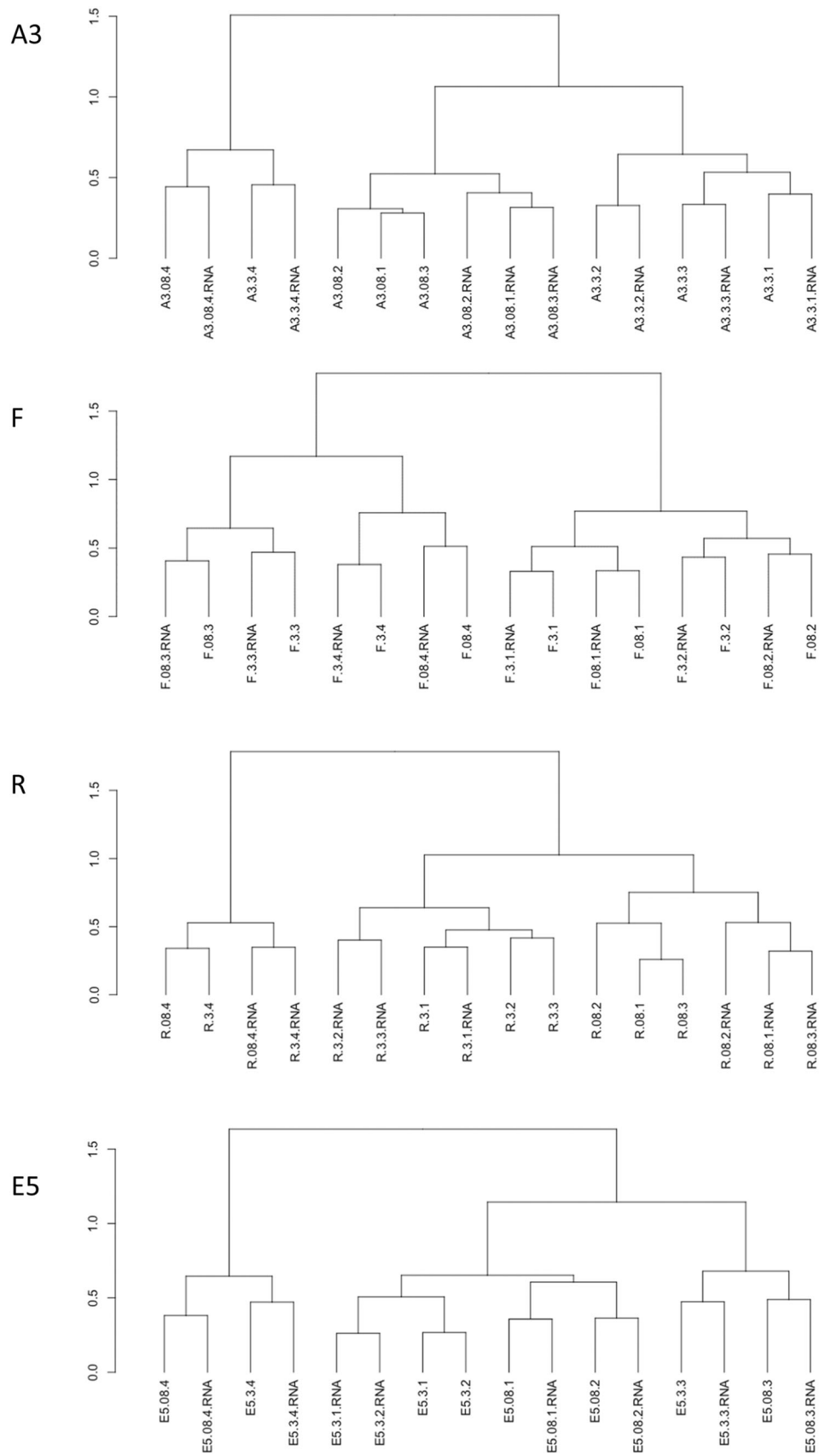


FIGURE A5 Bacterial community structure comparisons at depths 20 (1), 60–80 (2), 150 (3), and 300 m (4) at stations A3, F, E5, and R, recovered from two size fractions 0.8–3 μm (0.8) and >3 μm (3) assessed by 16S rRNA gene sequencing from DNA and RNA (RNA). Dissimilarity dendrograms were constructed using Bray–Curtis distances. The sample notation is StationName.SizeFraction.Depth for DNA samples and Name.SizeFraction.Depth.RNA for RNA samples.

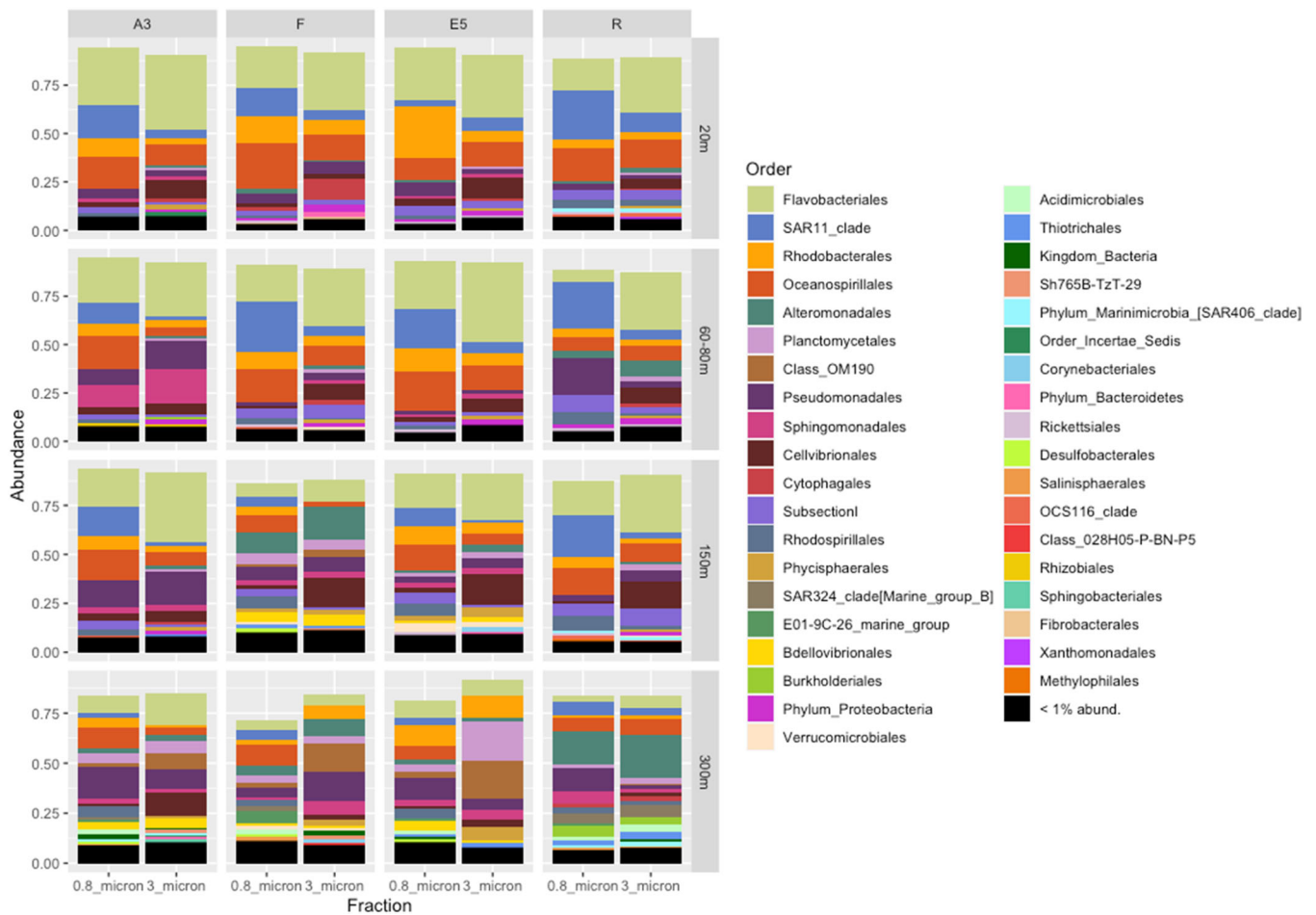


FIGURE A6 Bacterial taxonomic diversity at the order level (DNA) at different depths for the four sampling stations in the 0.8–3 μm (0.8_micron) or >3 μm (3_micron) size fractions.

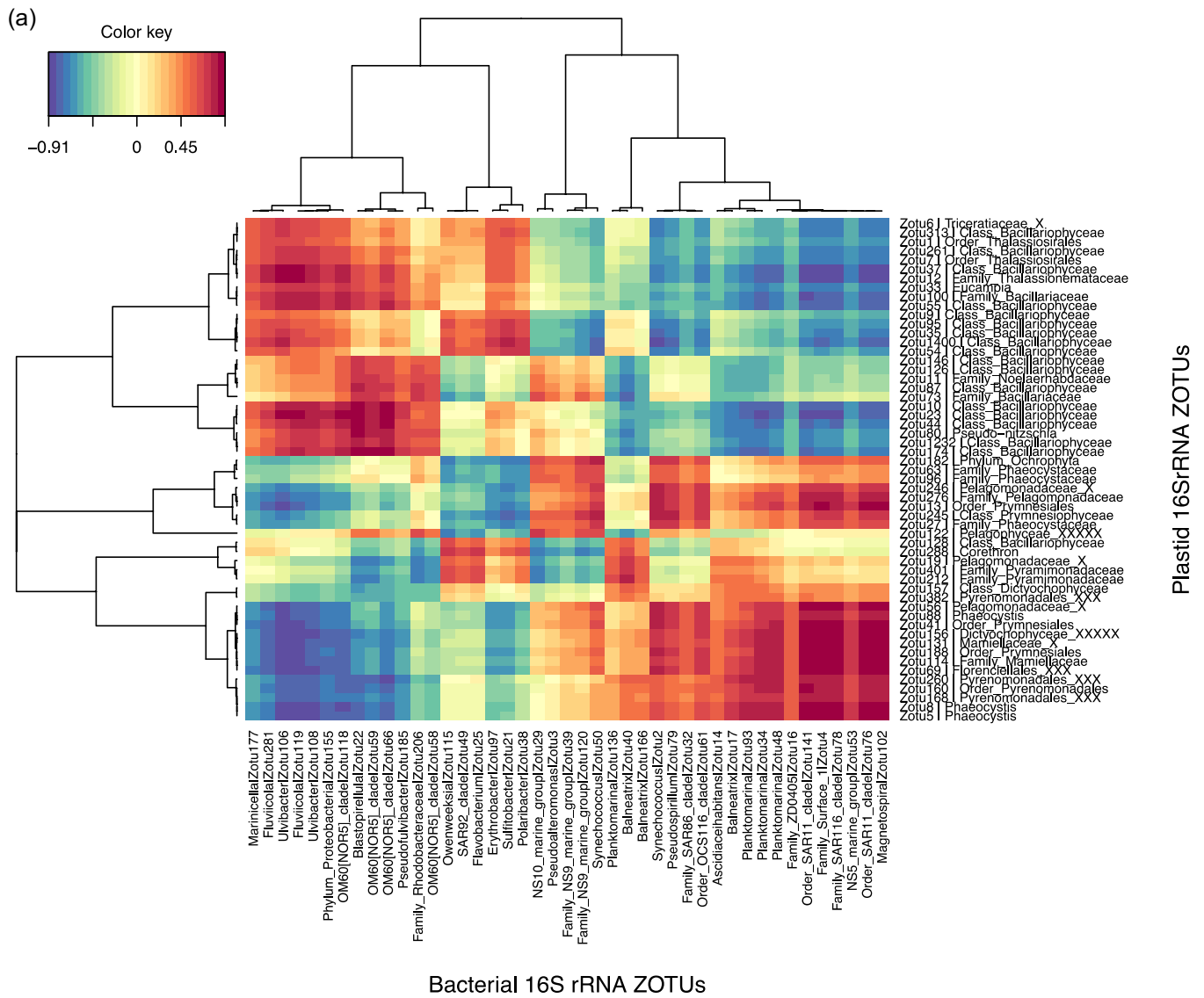


FIGURE A7 Clustered image heatmaps showing the most significant correlations between the bacterial 16S rRNA zOTUs and phytoplankton plastid 16S rRNA ZOTUs, recovered from DNA (a) or RNA (cDNA) (b).

APPENDIX 2

TABLE A1 Description of the samples analyzed at the four stations in terms of sampling depths and size fractions for 16S rRNA marker gene sequencing from DNA and RNA (cDNA).

Station	Depths (m)	Fractions (μm) for DNA sequencing			Fractions (μm) for cDNA sequencing		
A3-2	20	0.2-0.8	0.8-3	3-65	0.2-0.8	0.8-3	10-65 ^a
	80	na	0.8-3	3-65	na	0.8-3	3-65
	150	na	0.8-3	3-65	na	0.8-3	3-65
	300	na	0.8-3	3-25	na	0.8-3	3-25
F	20	0.2-0.8	0.8-3	3-65	0.2-0.8	0.8-3	3-65
	70	na	0.8-3	3-65	na	0.8-3	3-65
	150	na	0.8-3	3-65	na	0.8-3	3-65
	300	na	0.8-3	3-25	na	0.8-3	3-25
E5	20	0.2-0.8	0.8-3	3-65	0.2-0.8	0.8-3	3-65
	80	na	0.8-3	3-65	na	0.8-3	3-65
	150	na	0.8-3	3-65	na	0.8-3	3-65
	300	na	0.8-3	3-25	na	0.8-3	3-25
R	20	0.2-0.8	0.8-3	3-65	na	0.8-3	3-65
	60	na	0.8-3	3-65	na	0.8-3	3-65
	150	na	0.8-3	3-65	na	0.8-3	3-65
	300	na	0.8-3	3-25	na	0.8-3	3-25

Note: *rbcl* sequences were also analyzed for those samples highlighted in green.

Abbreviation: na, not analyzed.

^aThe RNA extraction from the 3 μm filter failed.

TABLE A2 Alpha diversity of phytoplankton and bacterial communities in surface depths at the different stations converted into true diversities or "effective species number" (Jost, 2006) from the Shannon Index calculations (DNA).

Station	Depth	Fraction	<i>rbcl</i> (phytoplankton plastids)	16S rRNA (phytoplankton plastids)	16S rRNA (bacteria)
A3	20	0.8-3 μm	20.69	12.80	115.18
	60-80		26.94	21.60	127.19
	150			18.06	136.14
	300			37.62	394.36
	20	>3 μm	12.21	13.05	159.09
	60-80		9.27	14.83	142.21
	150			14.09	130.03
	300			7.88	286.17
F	20	0.8-3 μm	7.74	16.46	90.35
	60-80		20.70	25.34	113.78
	150			34.44	317.58
	300			39.12	375.17

(Continues)

TABLE A2 (Continued)

Station	Depth	Fraction	<i>rbcL</i> (phytoplankton plastids)	16S rRNA (phytoplankton plastids)	16S rRNA (bacteria)
	20	>3 μm	3.95	4.94	113.99
	60–80		8.00	11.35	177.57
	150			14.95	180.07
	300			14.91	148.19
E5	20	0.8–3 μm	17.19	19.30	80.09
	60–80		16.33	22.36	101.26
	150			30.86	199.08
	300			44.65	354.15
	20	>3 μm	33.90	31.65	148.15
	60–80		36.28	37.19	138.95
	150			30.15	166.93
	300			20.02	115.55
R	20	0.8–3 μm	14.68	21.16	131.80
	60–80		14.13	15.90	117.70
	150			23.03	154.18
	300			28.33	164.07
	20	>3 μm	39.61	44.19	168.96
	60–80		38.48	32.30	151.75
	150			41.51	142.69
	300			38.19	171.77

TABLE A3 Relative abundances of diatoms (primarily order level) at the four stations obtained by microscopy identification (Lasbleiz et al., 2016) of CTD rosette surface samples, compared to those obtained in our >3 μm size fraction samples from 20 m using *rbcL* marker gene sequencing.

Order	A3-20m <i>rbcL</i>	A3-12m Counts	F-20m <i>rbcL</i>	F-25m Counts	E5-20m <i>rbcL</i>	E5-17m Counts	R-20m <i>rbcL</i>	R-28m Counts
Bacillariales	3.4	3.4	1.9	8.1	25.1	22.3	45.1	43.2
Chaetocerotales	5.2	87.5	0.0	26.1	3.8	37.8	13.6	7.8
Class_Bacillariophyceae	0.0	0.0	0.0	0.0	1.7	0.3	3.2	0.2
Coscinodiscales	0.0	0.0	0.0	0.0	0.0	0.0	3.8	0.0
Hemiaulales	0.0	0.0	0.0	2.6	2.2	0.7	6.3	1.9
Naviculales	4.6	0.5	0.0	0.9	13.3	2.0	11.1	4.8
Rhizosoleniales	1.6	0.0	0.0	0.0	3.9	0.9	9.2	4.6
Thalassionematales	2.3	3.5	0.0	2.6	8.4	22.9	0.0	18.7
Thalassiosirales	79.3	5.2	95.5	59.3	33.3	8.8	6.5	16.3
Triceratales	1.4	0.0	0.0	0.3	6.4	0.5	0.0	0.0
Corethrales	0.0	0.0	0.0	0.0	0.0	3.3	0.0	0.8
Leptocylindrales	0.0	0.0	0.0	0.1	0.0	0.5	0.0	1.2
Other	2.3	0.0	2.6	0.0	2.0	0.0	1.3	0.4

Note: Low abundance species or zOTUs (<1% abundance) were grouped into "Other."

TABLE A4 Summary of partial mantel tests showing Pearson's *R* correlations and *p* Values to explore the abiotic (environmental variables; Env) and biotic (Phytoplankton, Phyto) influence on bacterial community structure in the two size fractions for the top two surface depths (bacterial 16S rRNA and phytoplankton rbcL data set) or all depths (bacterial and plastid 16S rRNA).

Surface depths 20 m and 60–80 m		
Comparison: 16S rRNA	0.8–3 μm	>3 μm
Bacteria × Phyto Env	0.4585*	0.4373*
Bacteria × Env Phyto	−0.3868 ns	0.3581 ns
Phyto × Env Bacteria	0.7273**	0.2257 ns
Comparison: rbcL	0.8–3 μm	>3 μm
Bacteria × Phyto Env	0.6021**	0.4784*
Bacteria × Env Phyto	−0.5317 ns	0.06923 ns
Phyto × Env Bacteria	0.816**	0.6121**
Depths 20 m, 60–80, 150, 300 m		
Comparison: 16S rRNA	0.8–3 μm	>3 μm
Bacteria × Phyto Env	0.4365**	0.487**
Bacteria × Env Phyto	0.6116**	0.787**
Phyto × Env Bacteria	0.2116*	−0.1679 ns

Note: Tests are presented as the 2 matrices being compared (e.g., Bacteria × Phyto) while controlling for the effects of the third matrix (e.g., |Env).

Abbreviation: ns, nonsignificant.

***p* < 0.01; **p* < 0.05.

TABLE A5 Proportional abundance of bacterial zOTUs (out of total bacterial zOTUs) correlating to phytoplankton zOTUs and proportional abundance of phytoplankton zOTUs correlating to bacterial zOTUs (out of total phytoplankton zOTUs) for the significantly correlated zOTUs in Figure 6.

Proportional abundance bacterial zOTUs correlating to phytoplankton zOTUs			Proportional abundance phytoplankton zOTUs correlating to bacterial zOTUs				
Sample	bact zOTUs (0.8–3 μm)	bact zOTUs (>3 μm)	Sum bact zOTUs	Sample	phyto zOTUs (0.8–3 μm)	phyto zOTUs (>3 μm)	Sum phyto zOTUs
A3.08.1	38.75	10.73	49.48	A3.08.1	70.87	12.73	83.60
A3.08.2	37.50	10.30	47.80	A3.08.2	63.26	18.60	81.86
E5.08.1	22.39	9.29	31.68	E5.08.1	35.59	2.74	38.33
E5.08.2	44.72	5.22	49.94	E5.08.2	58.94	4.70	63.65
F.08.1	39.11	6.97	46.08	F.08.1	78.34	3.18	81.52
F.08.2	43.28	3.84	47.13	F.08.2	88.01	1.35	89.36
R.08.1	38.42	5.66	44.08	R.08.1	87.93	1.99	89.92
R.08.2	32.62	6.57	39.19	R.08.2	89.45	0.26	89.72
A3.3.1	13.89	30.62	44.51	A3.3.1	9.65	77.49	87.14
A3.3.2	18.64	21.37	40.01	A3.3.2	7.67	78.88	86.54
E5.3.1	22.40	20.23	42.63	E5.3.1	2.97	5.87	8.84
E5.3.2	22.80	16.85	39.65	E5.3.2	5.10	17.81	22.91
F.3.1	19.34	11.63	30.97	F.3.1	14.52	45.31	59.83
F.3.2	20.21	16.28	36.49	F.3.2	23.88	42.27	66.15
R.3.1	23.62	19.51	43.12	R.3.1	37.29	46.38	83.68
R.3.2	11.35	35.30	46.65	R.3.2	28.27	52.64	80.91

APPENDIX 3

Illumina data analysis

16S rRNA reads preprocessing and analysis using USEARCH v8/v9/11 unless QIIME v 1.9 or MOTHUR indicated.

Sequence preprocessing

1. Merge pairs (0 mismatches in overlap region)


```
usearch8 -fastq_mergepairs 141003_SN1126_A_L001_HCI-2_R1.fastq -reverse 141003_SN1126_A_L001_HCI-2_R2.fastq -fastq_truncqual 3 -fastq_maxdiffs 0 -fastqout merged0.fastq
```
2. Quality filtering with usearch8 (maxee 1.0)


```
usearch8 -fastq_filter merged0.fastq -fastq_maxee 1.0 -fastaout reads0_filt.fasta -fasta_cols 0
```
3. Separate eubacterial 16S reads in the forward sense using the 341F primer (ligation on PCR was not directional so reads appear in both orientations)


```
grep -B1 "^\{7\}CCTACGGG.GGC.GCAG" reads0_filt.fasta > eubfor0.fasta
```
4. Separate eubacterial 16S reads in the reverse sense using the 806R primer


```
grep -B1 "GACTAC.GGGTATCTAATCC" reads0_filt.fasta > eubrev0.fasta
```
5. Reverse complement reverse reads with QIIME command


```
adjust_seq_orientation.py -i eubrev0.fasta -r
```
6. Combine forward and reverse complemented reads in one file


```
cat eubrev0_rc.fasta eubfor0.fasta > euball0.fasta
```
7. Remove the double hyphen that is introduced at the beginning of the sequence during the reverse complementing (otherwise the demultiplexing only assigns the forward reads)


```
sed s'/\--//g' euball0.fasta > eub0.fasta
```
8. Demultiplex reads using QIIME command to parse out reads to sample using the mapping file with no mismatches in barcode allowed


```
demultiplex_fasta.py -f eub0.fasta -m mapping_file_eub.txt -b 7 -e 0.5 --retain_unassigned_reads
```
9. Parse out reads to the different projects: KEOPS2 and MOCK data set using the QIIME command


```
extract_seqs_by_sample_id.py -i demultiplexed_seqs.fna -o keops2.fasta -m mapping_file_eub.txt -s "Project:KEOPS2"
extract_seqs_by_sample_id.py -i demultiplexed_seqs.fna -o mock.fasta -m mapping_file_eub.txt -s "Project:MOCK"
```
10. Trim off primer sequences with MOTHUR allowing only exact matches to primers

File: 16Soligos.txt

```
forwardCCTACGGGNGGCWGCAG
reverseGACTACHVGGGTATCTAATCC
mothur > trim.seqs(fasta=keops2.fasta, oligos=16Soligos.txt)
mothur > trim.seqs(fasta=mock.fasta, oligos=16Soligos.txt)
```

Identification of zOTUs (USEARCH-32 v11)

11. Dereplicate sequences


```
usearch -fastx_uniques keops2.trim.fasta -fastaout keops2.uniques2.fasta -sizeout -relabel Uniq
```
12. Denoise using the UNOISE3 algorithm


```
usearch -unoise3 keops2.uniques2.fasta -zotus keops2_zotus.fasta
```
13. Map reads to zOTUs (note that for correct mapping, the read sample names should not contain a period as a delimiter since this is the delimiter expected after the sample name and before the read count, eg A3_08_1.23 for A3, 0.8 µm fraction, depth 1, read count 23)


```
usearch -otutab keops2.trim.fasta -zotus keops2_zotus.fasta -otutabout keops2zotutab2.txt -biomout keops2zotutab2.json -mapout keops2zotusmap2.txt -notmatched keops2unmappedzotus2.fasta -dbmatched keops2zotus_with_sizes2.fasta -sizeout
```

Classification of 16S rRNA zOTUs (USEARCH-64 v9)

14. Classify zOTU sequences using the 16S rRNA SILVA database v.123 and the SINTAX classifier


```
usearch9 -sintax keops2_zotus.fasta -db /users/west/silva_16s_v123 -tabbedout keops2.zotus.silva.sintax -strand plus -sintax_cutoff 0.8
```
15. Convert Sintang taxonomic classification into GreenGenes format (using Excel) for later export into R package Phyloseq by keeping columns 1 (zOTU name) and 4 (taxonomy) (eg k:Bacteria;p:Cyanobacteria;c:Chloroplast replaced with k__Bacteria;p__Cyanobacteria;c__Chloroplast)

16. Remove zOTUs without classification at Kingdom level from syntax classification, from zOTU sequences and zOTU Table (2 zOTUs removed)

Save as `keops2_zotus.mod.fasta`, `keops2zotutab2mod.txt` and `keops2.zotus.silva.syntax.mod.txt`

Alignment of zOTUs (Mothur)

17. Align retained sequences to the mothur 16S rRNA SILVA SEED alignment (version 132)
- ```
mothur> align.seqs(candidate=keops2_zotus.mod.fasta, template=core_set_aligned.fasta.imputed, processors=8)
```
18. Inspect the alignment manually with AliView v 1.18 to identify spurious length zOTU sequences (<300 bp) and remove these and the zOTUs in the flipaccnos (not aligned) file from the alignment
- Save as `keops2_zotus.mod.align.mod.txt`

#### *Filtering of zOTUs from the zOTU table*

19. Make a BIOM table from the corrected zOTU table (step 16)
- ```
biom convert --table-type="OTU table" -i keops2zotutab2mod.txt -o keops2.zotus.biom --to-json
```
20. Add taxonomy to the BIOM table
- ```
biom add-metadata --sc-separated taxonomy --observation-header OTUID,taxonomy --observation-metadata-fp keops2.zotus.silva.syntax.mod.txt -i keops2.zotus.biom -o keops2.zotus.tax.biom
```
21. Remove the 17 spurious zOTU sequences (in the flip.accnos file or short length) identified in steps 17 and 18 (zOTUs names saved as list in file `bad_zotus.txt`). (QIIME)
- ```
filter_otus_from_otu_table.py -i keops2.zotus.tax.biom -o keops2.zotus.tax.filt.biom -e bad_zotus.txt
```
22. Remove low count samples
- ```
filter_samples_from_otu_table.py -i keops2.zotus.tax.filt.biom -o keops2.zotus.tax.filt2.biom -n 100
```

#### *Prepare mapping file and tree for phyloseq object creation*

23. Sort sample order according to the mapping file
- ```
sort_otu_table.py -i keops2.zotus.tax.filt2.biom -o keops2.zotus.tax.filt2.sort.biom -m mapping_file_keopsII_mod2.txt -s SampleOrder
```
24. Make a tree of zOTU sequences
- ```
make_phylogeny.py -i keops2_zotus.mod.align.mod.txt -o keops2.zotus.repset.tre
```
25. Convert biom table to json format for input into R package phyloseq
- ```
biom convert -i keops2.zotus.tax.filt2.sort.biom -o keops2zotus.json.biom --table-type="OTU table" --to-json
```

Create phyloseq object

26. Import into phyloseq in R Studio (phyloseq and ape packages)
- ```
library(ape)
library(phyloseq)
keops2zotu_otutab <- import_biom("keops2zotus.json.biom", parseFunction=parse_taxonomy_greengenes, parallel=TRUE)
keops2map <- import_qiime_sample_data("mapping_file_keopsII_mod2.txt")
keops2zotutree <- read.tree("keops2.zotus.repset.tre")
keops2zotu <- merge_phyloseq(keops2zotu_otutab, keops2map, keops2zotutree)
```

#### *Classification of Plastid sequences with PhytoRef*

To obtain a more informative classification of the 16S rRNA chloroplast sequences than offered by the SILVA database v.123 database, we used the PhytoRef database (Decelle et al., 2015).

1. To be able to use this database with the syntax classifier, the headings of the PhytoRef taxonomy can be modified with the following unix script:
- ```
cat PhytoRef_with_taxonomy.fasta | awk -F "|" '{if($1~/>/) {printf("%s;", $1); printf("tax=k:%s,", $2); printf("p:%s,", $4); printf("c:%s,", $5); printf("o:%s,", $7); printf("f:%s,", $9); printf("g:%s,", $10); printf("s:%s,", $11); printf("%s", "\n")}} else {print $0}}' > phytoref.syntax.txt
```

2. The 16S rRNA zOTUs identified as chloroplast sequences (step 14) are subsetted out to a new phyloseq object and the zOTU ID names are saved as text (phyloseq)


```
keops2zotuchloro <- subset_taxa(keops2zotu, Class=="Chloroplast")
listchlorozotus <- taxa_names(keops2zotuchloro)
```
3. The chloroplast zOTU sequences are extracted from the 16S rRNA fasta file using the zOTU IDs (QIIME)


```
filter_fasta.py -f keops2_zotus.mod.fasta -o keops2zotuchloro.fasta -s listchlorozotus.txt
```
4. Classification with syntax


```
usearch9 -sintax keops2zotuchloro.fasta -db phytoref.sintax.txt -tabbedout keops2zotu-chloro.phytoref.sintax -strand plus -sintax_cutoff 0.8
```
5. Convert syntax taxonomic headers as in step 15 for later importation into phyloseq.
6. Export zOTU table from phyloseq


```
write.csv(otu_table(keops2zotuchloro), file="keops2zotuchlorootutab.csv")
check the table and save it as txt
```
7. Make biom file from zOTU table, add taxonomy and convert to json format for phyloseq reimportation


```
biom convert --table-type="OTU table" -i keops2zotuchlorootutab.txt -o keops2zotuchloro.biom --to-json
biom add-metadata --sc-separated taxonomy --observation-header OTUID,taxonomy --observation-metadata-fp keops2zotuchloro.phytoref.sintax.mod.txt -i keops2zotuchloro.biom -o keops2zotuchlorotax.biom
biom convert -i keops2zotuchlorotax.biom -o keops2zotuchlorotaxjson.biom --table-type="OTU table" --to-json
```
8. Reimport into phyloseq


```
keops2zotuchlorootutab <- import_biom("keops2zotuchlorotaxjson.biom", parseFunction=parse_taxonomy_greenegens, parallel=TRUE)
keops2zotuchloro <- merge_phyloseq(keops2zotuchlorootutab, keops2map, keops2zotutree)
```

Sequence preprocessing of *rbcl* sequence reads (using USEARCH v8/v9 unless QIIME v 1.9 or MOTHUR indicated)

1. *rbcl* sequences were essentially processed as the 16S rRNA reads from step 3 using the *rbcl* primers until step 8 with primer removal achieved with the custom python script `strip_primers_exclude.py` from Tony Walterst (<https://gist.github.com/walterst>).
2. zOTU identification was performed the same as for the 16S rRNA reads from steps 11-13.
3. zOTU sequences were aligned and inspected with AliView and 5 short sequences were removed.
4. zOTUs were classified using `blastn` with further manual taxonomic refinement using the WoRMS database (WoRMS Editorial Board, 2016).