ETH Zurich
Department of Biology
Master in Biology
Major in Microbiology and Immunology

Taxonomic profiling of microbes co-cultured with protists in the Roscoff Culture Collection

Semester Project Report

Submitted by:
Zuzanna Błocka

Supervisors:
Prof. Dr. Shinichi Sunagawa
Dr. Lucas Paoli
Dr. Samuel Miravet-Verde

Study advisor:
Prof. Dr. Wolf-Dietrich Hardt

Zurich, February 2023

ETH zürich

Contents

1 Abstract **2**

2 Introduction **3**

2.1 The untapped potential of marine microbiota 3

2.2 The Roscoff Culture Collection 3

3 Project goals **4**

4 Materials and Methods **5**

4.1 Sample collection 5

4.2 Genomic DNA extraction 5

4.3 16S rDNA amplification 5

4.4 Illumina MiSeq sequencing 6

4.5 Data processing 6

4.6 Bioinformatic analyses 7

5 Results **7**

5.1 Sample composition 7

5.2 General microbial composition of the samples 8

5.3 Intra-sample diversity is similar between host genera 10

5.4 Inter-sample diversity can be best explained by host phylogeny 13

5.5 Taxonomical composition of samples is specific to clusters 15

6 Discussion **19**

6.1 Negative controls are necessary in every extraction batch. 19

6.2 General diversity 19

6.3 Hierarchical clustering reveals differences in taxonomic composition. 20

7 Conclusions **21**

8 Acknowledgements **21**

9 References **22**

10 Supplemental Information **26**

1. Abstract

The ocean hosts an enormous diversity of microorganisms. However, most of these are not at present available for culture, which makes them difficult to study. We know that many marine organisms rely on complex nutrient cycles which are difficult to unravel. Microalgae produces soluble carbon-, sulfur-, and nitrogen-containing compounds which are taken up by heterotrophic bacteria. The photosynthesizers, on the other hand, often rely on co-factors of bacterial origin, such as vitamin B_{12}. Since many ocean bacteria may rely on algal organic products to obtain essential nutrients, searching for microbes of interest in algal strain cultures can help pinpoint the method of cultivation. In this study, 16S meta-barcoding was performed on bacteria associated to sixty strains from the Roscoff Culture Collection. The bacterial taxonomic composition differed between the algal divisions. Hierarchical clustering revealed inter- and intra-division differences. The microbiota of the samples mostly belonged to Bacteroidota, Alpha- and Gammaproteobacteria. This is consistent with previous studies on the cycling of nutrients in the ocean, as these groups are metabolically flexible and ubiquitous heterotrophs. Among the sequenced reads, there appear members of poorly described or undescribed prokaryotic lineages at high taxonomic levels, which could be subject to further study and cultivation efforts. Obtaining these organisms in pure culture could allow the description of new lineages and their unique metabolisms, and potentially, the production of their bio-products of interest.

2. Introduction

2.1. The untapped potential of marine microbiota

Globally, oceans host an immense diversity of microorganisms from all domains of life. The study of these organisms interests researchers in both pure and applied biology. Some of them show the evolutionary stages of life on Earth that are not known on land. An example of this is the discovery of the Asgardarcheota, a phylum of marine archaea which helped explain the divergence of eukaryotes from prokaryotes [1]. Others are of interest to pharmaceutical and other industries looking for new natural products such as antibiotics, antivirals and cancer drugs [2].

Unfortunately for researchers hoping to learn about ocean ecosystems, studies reveal that only about a small fraction (estimates vary, 1-45%) of the marine microbiome is presently available in culture [3]. At the same time, metagenomic research has recently revealed the immense biosynthetic potential of many of the unculturable marine organisms [4]. One of the main hypotheses for why they do not show on plating experiments is that these organisms require a symbiont for nutrient production or growth signaling [5]. It is likely that the majority of these symbionts are phytoplankton. The exploration of algal-bacterial symbioses [1] can help elucidate the biological pathways and environmental roles of both of these groups, as well as discover compounds useful for humans. Unfortunately, most of the current experimental setups for studying algae-bacterial interactions are either relatively simple (e.g., one species of bacteria and algae each [6]) or too complicated (e.g., microcosms or direct sampling from the environment [7, 8]) to untangle the specificities of these ecological niches [9]. The existing research suggests that algae often rely on specific strains of bacteria for the production of growth factors, such as vitamin B_{12}, needed for methionine synthesis [10]. The growth of the bacteria is also dependent, to an extent, on the algal symbiont. In particular, organosulfur compounds, such as taurine [6], dimethylsulfoniopropionate (DMSP) and 2,3-dihydroxypropane-1-sulfonate (DHPS) [11].

2.2. The Roscoff Culture Collection

The Roscoff Culture Collection (RCC) in Roscoff (Brittany, France) maintains strains of marine algae, protists, bacteria and viruses [12]. These strains are generally available for researchers and cultured in constant and controlled conditions, including medium components and concentration, incubation temperature and light conditions. This provides a wide array of test subject provenance and at the same time ensures a high potential for reproducibility of the results, as compared to culturing strains independently for each experiment. Using the microalgae cultures as mini-ecosystems which are, first, stable in terms of environmental conditions, and second, include a primary producer and their symbionts, could be one step towards disentangling the

[1] The term *symbiosis* is understood in this text as a close, long-term interaction between two species, with no indication as to whether it is positive or negative.

interactions between the present organisms. Algal cultures from the RCC explored in this study represent groups described below. All of them play an important part in the global cycling of carbon, sulfur, and nitrogen, providing organic forms of these nutrients to other ocean microorganisms through various methods of release [11].

Chlorophyta There are three classes represented in this study from the division Chlorophyta: Chloropicophyceae, Mamiellophyceae and Pyramimonadophyceae. Chlorophyta are green algae which are known to produce a variety bioactive substances (including steroids, alkaloids, fatty acids, glycerides and terpenes [13]). These could either be nutrient sources for other microorganisms or chemical signaling factors used for e.g. quorum sensing or chemotaxis [14].

Haptophyta The Haptophyta division is represented in this study by the genus *Phaeocystis*. It is an important player in the carbon and sulfur cycles in the ocean, periodically producing large blooms [15]. This genus is referred to as "sea foam", relating to the colonies it forms, glued together with a polysaccharide matrix which traps nutrients and helps the algal cells to absorb them [16]. This abundance of diverse organic matter should attract heterotrophs who degrade it.

Other algae Other genera used in this study include *Minutocellus* and *Prasinoderma*. The former is an ubiquitous diatom known to aggregate with the help of bacteria to form particles of marine snow [17]. The latter was initially described to belong in Prasinophyta, which is a synonym for Chlorophyta [18], but more recent genomics research has named the new phylum Prasinodermophyta as a sister clade to Chlorophyta [19].

3. Project goals

The main goals of this project were to establish a methodology for a larger screening of the microbiomes of the RCC and to explore tendencies of their micro-ecosystems. The algal strains likely sustain stable bacterial communities as a result of being cultured for many years in the same conditions. Microbes found in co-cultivation with some of these algae, could likely take advantage of the algal metabolites produced, such as carbon-, sulfur- and nitrogen-containing compounds. Understanding these complex cycles is needed to better describe the symbioses in marine ecosystems and the importance of different players on the ecosystem as a whole: who produces which compounds, who do they need to synthesize them, and who degrades them. Additionally, many microorganisms which are rare or otherwise unculturable with today's methods could be growing in the RCC cultures thanks to an interaction with the host species. This way, the culture collection could be used as a starting point for obtaining highly desirable microbial strains for the production of pharmaceuticals and other chemicals: a strain found by genomic sequencing in one experiment is much more likely to be found in the same algal strain again than when obtaining samples from the environment.

In order to contribute to this effort, the following aspects were analyzed in this work:

- How diverse are the microbiomes of RCC cultures?
- How much of that diversity is shared between groups?
- What groups can be defined?
- How specific are these interactions?
- Are there microorganism lineages that are shared between all the studied algal strains?

4. Materials and Methods

4.1. Sample collection

Strains used in this study came from the Roscoff Culture Collection (RCC), France. For negative control, Milli-Q water available at the station was first filtered with a single-use 0.22 μm filter and then processed in the same way as other samples. Nitrocellulose membrane filters (MF-Millipore^{TM}) were inserted into re-usable filter holders (Sartorius). The holders had previously been washed in 70% ethanol, then Milli-Q water, DNA AWAY^{TM} Surface Decontaminant (Thermo Scientific^{TM}), and water again. Next, the assembled syringe filters were autoclaved. Cultures were transferred from flasks to sterile syringes and a 3 μm filter attached. 800 μl of culture was filtered into 200 μl 50% glycerol to obtain a 10% glycerol cryostock. Next, a 0.22 μm filter was attached to the end of the 3 μm filter and the rest of the culture was filtered (or, if the filters clogged, however much was possible). The supernatant was collected in 50 ml Falcon tubes, and the 0.22 μm filter was folded into a cryotube. The tubes containing the filter and cryostock were then flash-frozen in liquid nitrogen and then transferred first to -20 °C, and then to -80 °C at the end of sampling. The cryotubes were shipped to Zurich on dry ice, while the supernatants were shipped in a cooled Styrofoam box.

4.2. Genomic DNA extraction

DNA was extracted from the 0.22 μm filters with the DNeasy Power Water Kit (Qiagen). Extraction success was measured with Qubit^{TM} dsDNA High Sensitivity. A test extraction was performed first on filters from LB cultures of *Escherichia coli* K12 and cultures of *Symbiodinium* algae, collected in the same way as the experimental samples. After the extraction method proved effective, DNA from the Roscoff samples was also extracted.

4.3. 16S rDNA amplification

In order to obtain the fourth variable region of 16S for sequencing, two PCR reactions were employed. First, the region was amplified with 515F [20] and 806R [21] primers. After a clean-up with Cytiva Sera-Mag Select beads, a second PCR was performed using Illumina UD barcoded

adapters and the product was cleaned with the same magnetic bead kit. The PCR conditions were first tested on a variety of sample types and DNA concentrations, including some of the *E. coli* K12 and *Symbiodinium* test samples as well as some from the failed extraction. Input DNA elute volume was adjusted between samples so that its final concentration would be similar in each reaction mixture. The total concentration of the product was measured with the NanoDrop Eight Spectrophotometer (ThermoFisher Scientific) and the size of the 16S peak was measured with a Fragment Analyzer^{TM} Automated CE System, using the DNF-473 Standard Sensitivity NGS Fragment Analysis Kit (1 bp - 6,000 bp).

A combination of 13 cycles in PCR1 and 22 cycles in PCR2 proved to work well and produce a clean, high peak at around 450 bp, but only in samples where the DNA concentration had previously been detected with Qubit^{TM}. Because of a tight timeline and inability to get more samples shipped from Roscoff quickly, it was decided to try to amplify the low DNA samples using 35 cycles of PCR1 instead of 13 cycles. After the whole library was amplified and again measured on the Fragment Analyzer^{TM}, these low-DNA samples had no peaks still. Therefore, only the negative extraction control was used for downstream application from this batch.

Ten additional controls were used in the PCR reactions: six molecular-grade water samples (NC) and four positive controls (PC) in the form of a mock community (ZymoBIOMICS Microbial Community DNA Standard II (Log Distribution)). Four NC and two PC were amplified with the 13-cycle PCR1 and two NC and two PC in the 35-cycle PCR1.

4.4. Illumina MiSeq sequencing

The library was pooled equimolarly based on the concentration of the 400-500 bp peak as calculated by the Fragment Analyzer. The volume of the controls was the average volume of samples. This pool was then cleaned up again with the Cytiva magnetic beads and the final concentration measured using a 4200 TapeStation System (Agilent), with a D1000 ScreenTape Assay. The pool solution was then sequentially diluted and measured again until it reached the 4 nmol/l required. The sequencing was performed at the Genome Engineering and Measurement Lab of the ETHZ, on an Illumina MiSeq machine. The V2 2x250 bp (500 cycles) kit was used.

4.5. Data processing

The raw data obtained from the Genome Engineering and Measurement Lab was in the form of demultiplexed read files. The processing pipeline is part of a standard operating procedure written by Dr. Hans-Joachim Ruscheweyh. The first step of this pipeline includes matching and removing primers using Cutadapt [22]. Next, the Dada2 pipeline [23] is used to perform quality control, assess errors, infer amplicon sequence variants (ASVs), merge reads and remove bimeras. An ASV table is generated with IDtaxa2 [24], which classifies the sequences to their appropriate taxonomies. An additional OTU table is made by clustering the ASVs by 97% similarity using usearch [25]. These tables were the basis of further analyses.

Table 1. Algal phylogeny (according to the RCC) and number of samples.

Domain	Division	Class	Order	Family	Genus	n
Eukaryota	Chlorophyta	Chloropicophyceae	Chloropicales	Chloropicales	*Chloroparvula*	5
					Chloropicon	10
		Mamiellophyceae	Mamiellales	Bathycoccaceae	*Bathycoccus*	3
				Mamiellaceae	*Micromonas*	5
		Pyramimonadophyceae	Pseudoscourfieldiales	Pycnococcaceae	*Pycnococcus*	1
	Heterokontophyta	Mediophyceae	Cymatosirales	Cymatosiraceae	*Minutocellus*	1
	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	*Phaeocystis*	33
	Prasinodermophyta	Prasinodermophyceae	Prasinodermales	Prasinodermaceae	*Prasinoderma*	2

4.6. Bioinformatic analyses

Diversity measures and statistics were calculated using QIIME 2 [26] Python API. Data frames were prepared to appropriate formats using Pandas [27]. Alpha diversity was measured using the following metrics: Pielou evenness [28], Simpson index, Shannon entropy [29], and Faith phylogenetic diversity [30]. The last metric utilized a phylogenetic tree calculated by aligning the ASVs with Mafft [31] and inferring phylogeny with Fasttree 2 [32]. To obtain a beta diversity distance matrix, the unweighted UniFrac metric was applied [33], using the phylogeny obtained in the previous step. Significance of groupings of distances was calculated with the PERMANOVA test [34]. Multi-dimensional scaling was also performed to visualize the grouping of data by various meta-data categories [35]. Data visualizations were produced in Seaborn [36] and Plotly [37], with additions based on Matplotlib [38]. Additional identification of 16S sequences was performed in NCBI BLAST [39] on the nr/nt database.

5. Results

5.1. Sample composition

Sixty-five cultures were prepared by the RCC for the experiment. They were filtered through 3 μ and 0.22 μ filters in Roscoff. After transport to Zurich, DNA was extracted from the smaller-pored filters. While most DNA extraction batches resulted in detectable levels of DNA in a range from $\approx$0.3 ng/μl to 66 ng/μl, one six-sample batch resulted in undetectable DNA. It was the batch which contained the negative control. Despite many efforts, including increasing PCR amplification cycle number, obtaining amplicons of these samples was not possible. Time constraints prevented re-extraction. Only the control sample from this batch was used for sequencing. The other samples yielded DNA amounts sufficient for amplification and sequencing. The metadata of all samples, including RCC strain numbers, is held in Supplementary Information (Table 3). Overall, the sequenced samples include sixty strains from at least 28 species (many are only assigned to the genus level) in 8 genera (Table 1). The most abundant genera are *Phaeocystis* (n = 33) and *Chloropicon* (n = 10), while *Pycnococcus* and *Minutocellus* only contain one replicate each.

5.2. General microbial composition of the samples

Sequence variant counts were normalized by rescaling their distributions by the sum of counts in the sample. To visualize the most abundant operational taxonomic units (OTUs), these relative counts were then added over the OTUs and shown in Figure 1. The most abundant OTUs belong to *Marinobacter* (Gammaproteobacteria), *Balneola* (Rhodothermia) and *Marinoscillum* (Bacteroidia). The most common phyla are Proteobacteria and Bacteroidota, while some abundant OTUs also represent Cyanobacteria, Bdellovibrionota, Actinobacteriota and Planctomycetota. The Cyanobacteria classification refers to the identified 16S of the chloroplast. *Marinobacter* (OTU1) is the most abundant OTU in the samples overall and it has been identified in every one of 70 sequenced samples (60 algal strains + 10 controls), while in 58 of them it is at above 1% abundance. This OTU appears in only one sequence variant. In the water controls, its relative abundance ranges between 0.7 and 7.5%.

Interestingly, there is an OTU (OTU48) within the one hundred most common ones that has not been classified even at phylum level. Submitting the sequence to NCBI BLAST yielded >91% identity to some uncultured bacteria (GenBank accession IDs MF949365.1, MF550395.1, MF014318.1, for example) and only 86% identity to *Leptospira* sp. isolates. All of these alignments had E values < 0.0001. This OTU was identified in RCC1737 and RCC3538 *Phaeocystis globosa* strains as well as in RCC6429 *Phaeocystis jahnii*, with 199, 22462, and 3 total counts and 0.0169%, 7.3571% 0.0013% relative counts, respectively. Other OTUs unclassified at such high level are mostly spurious, with total counts between 1 and 100. An exception is OTU176, which accounts for 1059 total counts in RCC2686 *Prasinoderma coloniale*, which is 0.1297% of relative abundance in this sample. However, when searched by BLAST, this aligns with 100% identity, an E value of $6{\cdot}10^{-128}$ and 100% query cover to the 16S rDNA of the chloroplast of *Prasinoderma coloniale*.

In general, there are unclassified lineages at lower than phylum level in Actinobacteriota, Bacteroidota. Proteobacteria, and SAR324 clade (Marine group B). All of them are present in very low abundances (all of them at 0.75% sum of relative counts or lower).

Sequences which represent unknown lineages suggest the potential for discovery in these samples. A next step could be shotgun metagenomic sequencing to assemble their metagenomes, which could help specify their identification.

The ASV table was filtered by relative abundances $\geq$0.1% and total count $\geq$5. Reads coming from PCR controls were filtered out, then subset by samples where an ASV was found in at least 3 replicates from the same host genus, in order to minimize spurious identification. The sizes of intersections are shown in Figure 2. The effect of replicate number is visible, as *Phaeocystis* displays the most unique ASVs. To check the distribution of relative abundances in these ASVs, a histogram was also prepared (Figure 3). Many of the unique ASVs show a range of abundances within the *Phaeocystis* samples, including ones of >1%.

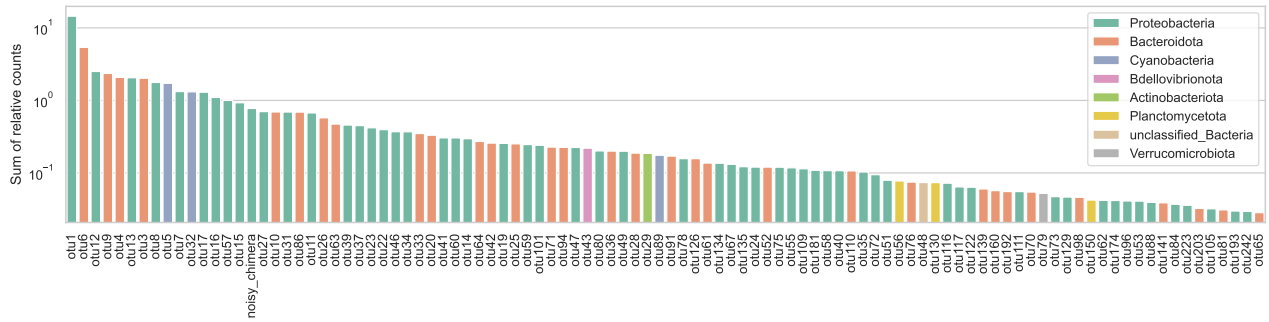

Figure 1. Bar plot representing the sums of normalized counts in identified OTUs. Notice chimeric amplicons were found. It could be the result of high hybridization affinity of two different 16S fragments. Shown are the first 100 most abundant OTUs after filtering out data from PCR controls.

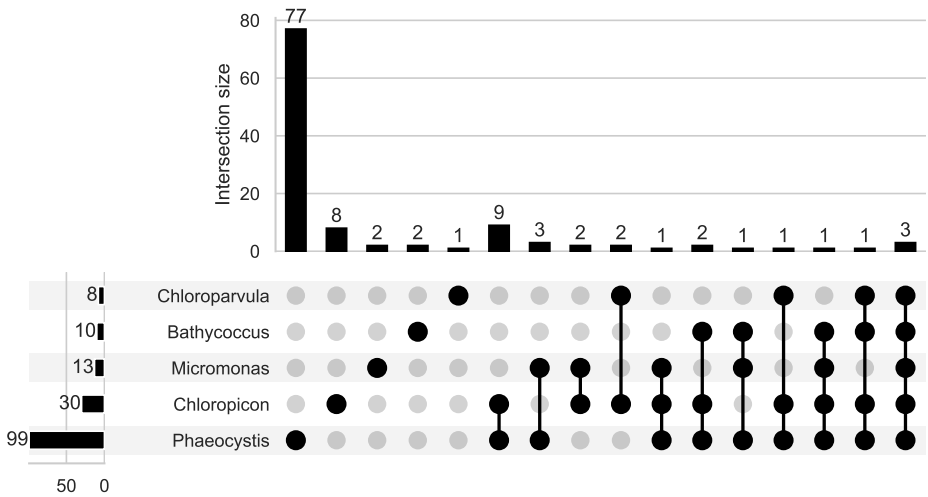

Figure 2. Intersection sizes of identified ASVs by algal host. Black dots in the bottom chart represent presence in the respective host genus; bar height in the top plot indicate intersection size.

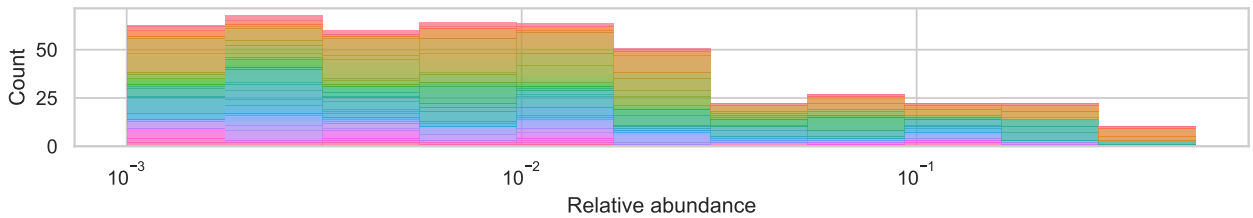

Figure 3. Histogram of normalized counts in ASVs unique to *Phaeocystis* samples. Bins are defined in logarithmic space. Colors represent ASVs. Notice there are ASVs which appear continually from small to large abundances - they are ubiquitous, but their amounts vary widely from one sample to another.

5.3. Intra-sample diversity is similar between host genera

Alpha diversity measures the differences within a community. The ASV profile table was used as a basis for calculating the following diversity metrics: Pielou evenness, Shannon entropy, Simpson index, and Faith phylogenetic diversity (Figure 4). The pairwise Kruskal-Wallis test [40] for genera revealed several statistically significant differences ($p \geq 5\%$) in all the calculated metrics 2. The significance is different in every metric, but most hosts differ from the positive and/or negative controls (Zymo and Water, respectively). *Pycnococcus* and *Minutocellus* cannot show significant differences due to group size (n = 1). *Bathycoccus*, *Chloroparvula*, and *Chloropicon* differ consistently across all metrics from the controls. *Micromonas* and *Phaeocystis* differ from both controls in evenness, entropy and Simpson index but only from the positive control in phylogenetic diversity. *Chloroparvula* differs from *Phaeocystis* with a low p-value in the first three metrics, but not the last one, and from *Micromonas* in evenness, Simpson index, and phylogenetic diversity. The most abundant differences between genera are observed in Shannon entropy, where *Phaeocystis* is distinguished from *Chloroparvula*, *Chloropicon*, *Micromonas*, and *Prasinoderma*. When looking at differences in alpha diversity between the sampling ocean, RCC incubation temperature and growing medium, the groups mostly differed from the controls and not from each other. The significant differences measured were between 15°C and 20°C, as well as between 15°C and 22°C, in Shannon entropy and Simpson index.

In general, the smallest number of differences can be found in the phylogenetic diversity of samples, which is consistent with the fact that all the hosts are marine algae and most likely interact with related prokaryotes. In some pairs, the q-value, which is an adjustment on the p-value, indicates a lower significance level.

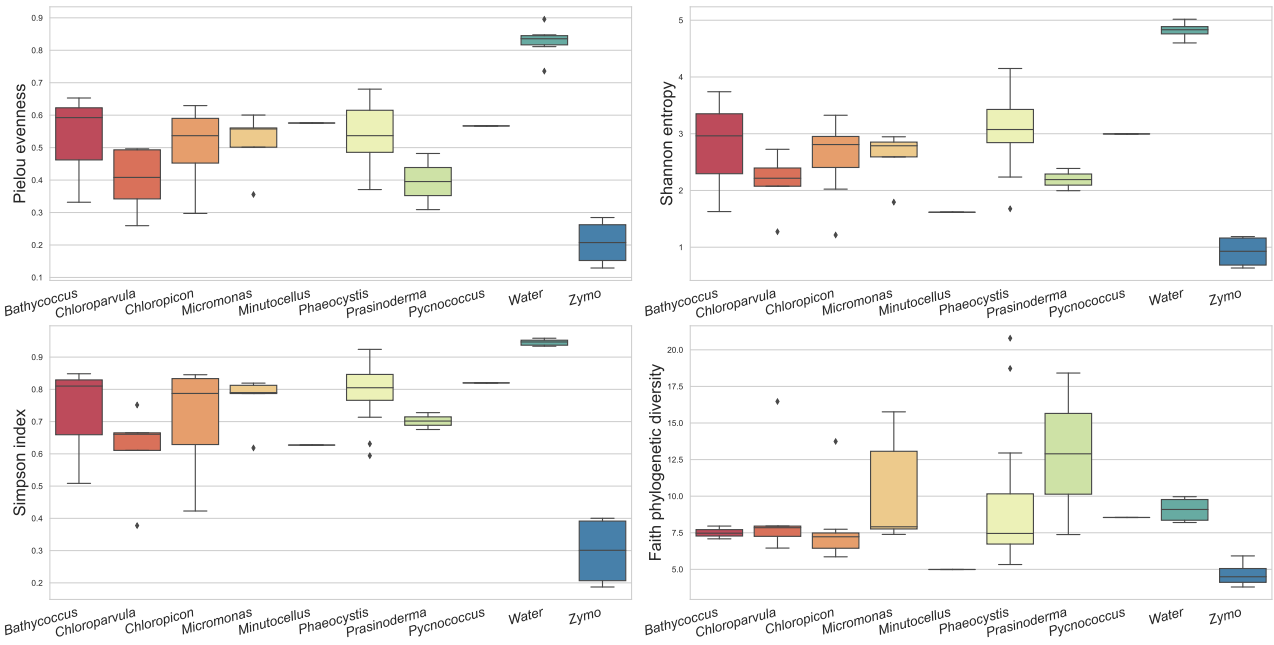

Figure 4. Alpha diversity metrics grouped by host genus.

Table 2. Significant ($p \leq 0.05$) pairwise differences found by the Kruskal-Wallis test, based on different alpha diversity metrics. Grouping based on algal host genus.

(a) Pielou evenness

Group 1	Group 2	H	p-value	q-value
Bathycoccus (n=3)	Water (n=6)	5.4000	0.0201	0.0906
Bathycoccus (n=3)	Zymo (n=4)	4.5000	0.0339	0.1271
Chloroparvula (n=5)	*Micromonas* (n=5)	3.9382	0.0472	0.1501
Chloroparvula (n=5)	*Phaeocystis* (n=33)	7.2844	0.0070	0.0447
Chloroparvula (n=5)	Water (n=6)	7.5000	0.0062	0.0447
Chloroparvula (n=5)	Zymo (n=4)	4.8600	0.0275	0.1124
Chloropicon (n=10)	Water (n=6)	10.5882	0.0011	0.0187
Chloropicon (n=10)	Zymo (n=4)	8.0000	0.0047	0.0447
Micromonas (n=5)	Water (n=6)	7.5000	0.0062	0.0447
Micromonas (n=5)	Zymo (n=4)	6.0000	0.0143	0.0715
Phaeocystis (n=33)	Water (n=6)	14.8500	0.0001	0.0052
Phaeocystis (n=33)	Zymo (n=4)	10.4211	0.0012	0.0187
Prasinoderma (n=2)	Water (n=6)	4.0000	0.0455	0.1501
Water (n=6)	Zymo (n=4)	6.5455	0.0105	0.0591

(b) Shannon entropy

Group 1	Group 2	H	p-value	q-value
Bathycoccus (n=3)	Water (n=6)	5.4000	0.0201	0.0824
Bathycoccus (n=3)	Zymo (n=4)	4.5000	0.0339	0.1164
Chloroparvula (n=5)	*Phaeocystis* (n=33)	9.5333	0.0020	0.0227
Chloroparvula (n=5)	Water (n=6)	7.5000	0.0062	0.0397
Chloroparvula (n=5)	Zymo (n=4)	6.0000	0.0143	0.0644
Chloropicon (n=10)	*Phaeocystis* (n=33)	4.4041	0.0359	0.1164
Chloropicon (n=10)	Water (n=6)	10.5882	0.0011	0.0187
Chloropicon (n=10)	Zymo (n=4)	8.0000	0.0047	0.0397
Micromonas (n=5)	*Phaeocystis* (n=33)	4.3865	0.0362	0.1164
Micromonas (n=5)	Water (n=6)	7.5000	0.0062	0.0397
Micromonas (n=5)	Zymo (n=4)	6.0000	0.0143	0.0644
Phaeocystis (n=33)	*Prasinoderma* (n=2)	4.2475	0.0393	0.1179
Phaeocystis (n=33)	Water (n=6)	14.8500	0.0001	0.0052
Phaeocystis (n=33)	Zymo (n=4)	10.4211	0.0012	0.0187
Prasinoderma (n=2)	Water (n=6)	4.0000	0.0455	0.1280
Water (n=6)	Zymo (n=4)	6.5455	0.0105	0.0591

(c) Simpson index

Group 1	Group 2	H	p-value	q-value
Bathycoccus (n=3)	Water (n=6)	5.4000	0.0201	0.0906
Bathycoccus (n=3)	Zymo (n=4)	4.5000	0.0339	0.1387
Chloroparvula (n=5)	*Micromonas* (n=5)	3.9382	0.0472	0.1609
Chloroparvula (n=5)	*Phaeocystis* (n=33)	9.5333	0.0020	0.0227
Chloroparvula (n=5)	Water (n=6)	7.5000	0.0062	0.0397
Chloropicon (n=10)	Water (n=6)	10.5882	0.0011	0.0187
Chloropicon (n=10)	Zymo (n=4)	8.0000	0.0047	0.0397
Micromonas (n=5)	Water (n=6)	7.5000	0.0062	0.0397
Micromonas (n=5)	Zymo (n=4)	6.0000	0.0143	0.0715
Phaeocystis (n=33)	Water (n=6)	14.8500	0.0001	0.0052
Phaeocystis (n=33)	Zymo (n=4)	10.4211	0.0012	0.0187
Prasinoderma (n=2)	Water (n=6)	4.0000	0.0455	0.1609
Water (n=6)	Zymo (n=4)	6.5455	0.0105	0.0591

(d) Faith phylogenetic diversity

Group 1	Group 2	H	p-value	q-value
Bathycoccus (n=3)	Water (n=6)	5.4000	0.0201	0.1295
Bathycoccus (n=3)	Zymo (n=4)	4.5000	0.0339	0.1695
Chloroparvula (n=5)	Zymo (n=4)	6.0000	0.0143	0.1073
Chloropicon (n=10)	*Micromonas* (n=5)	4.8600	0.0275	0.1546
Chloropicon (n=10)	Water (n=6)	6.7765	0.0092	0.1073
Chloropicon (n=10)	Zymo (n=4)	7.2200	0.0072	0.1073
Micromonas (n=5)	Zymo (n=4)	6.0000	0.0143	0.1073
Phaeocystis (n=33)	Zymo (n=4)	9.4952	0.0021	0.0927
Water (n=6)	Zymo (n=4)	6.5455	0.0105	0.1073

5.4. Inter-sample diversity can be best explained by host phylogeny

Beta diversity is used to measure differences between communities. In order to control for the influence of microbial phylogeny on sample diversity, the unweighted UniFrac method was used to calculate distances between groups of samples. The variables with the biggest measured influence on sample diversity were determined to be the phylogeny and sampling ocean of host algae 5. The highest distances observed are between *Micromonas* and other genera as well as *Prasinoderma* and others. The empty intersections in plots indicate where the sample sizes were too small to calculate beta diversity. As for the sampling oceans, the communities are mostly different to controls and where the culture provenance is unknown.

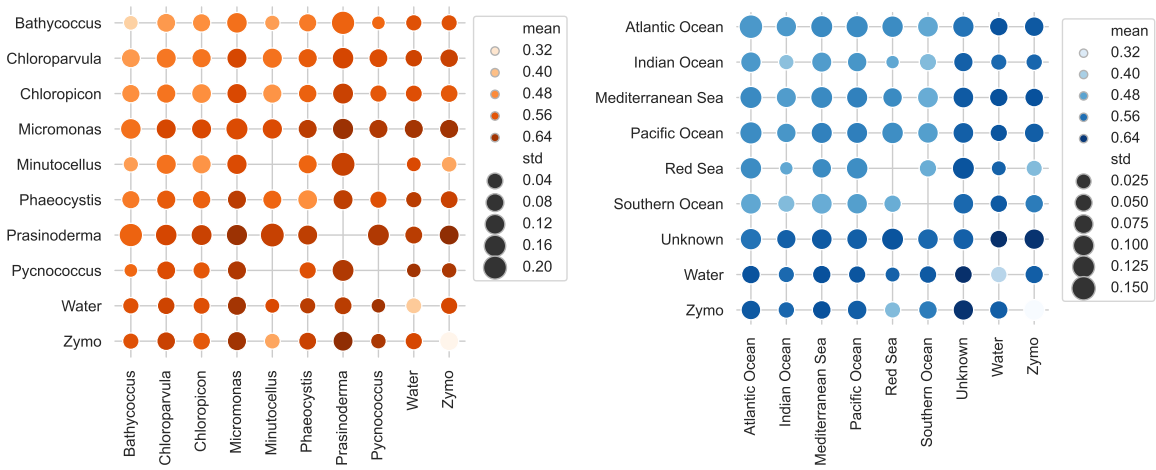

Figure 5. Unweighted UniFrac distances grouped by genus and sampling ocean.

PCoA and Adonis PERMANOVA A principal coordinates analysis (PCoA) was performed to assess the influence of different variables on the variance. Overall, 28 coordinates are needed to explain 90% of the variance, while the three first explain only 33%. Since, when coloring a plot of the first three axes by genus, slight grouping could be observed, using increasingly higher taxonomic levels of the hosts was tried until a clear distinction became visible on division level. This was mostly because 24 out of the 60 samples belong to Chlorophyta, but the numbers for each separate genus are relatively low. Coloring by sampling ocean did not reveal any visible grouping (Figure 6). An Adonis PERMANOVA test [41] revealed that the host division accounts for 29% of the variance (p=0.001), while the medium accounts for 9.3% ($p = 0.001$) and the sampling ocean for 8% ($p = 0.002$). The variance explained by the intersection of division and ocean or medium and ocean are not statistically significant ($p = 0.225$ and 0.116 respectively).

Characteristics of Haptophyta and Chlorophyta samples Subsetting the distance data by representatives of Haptophyta (*Phaeocystis* sp.) and visualizing main PCoA axes did not yield visible grouping when coloring by species, RCC medium, RCC temperature or sampling ocean. Doing the same to *Chlorophyta* samples revealed grouping on the second axis when

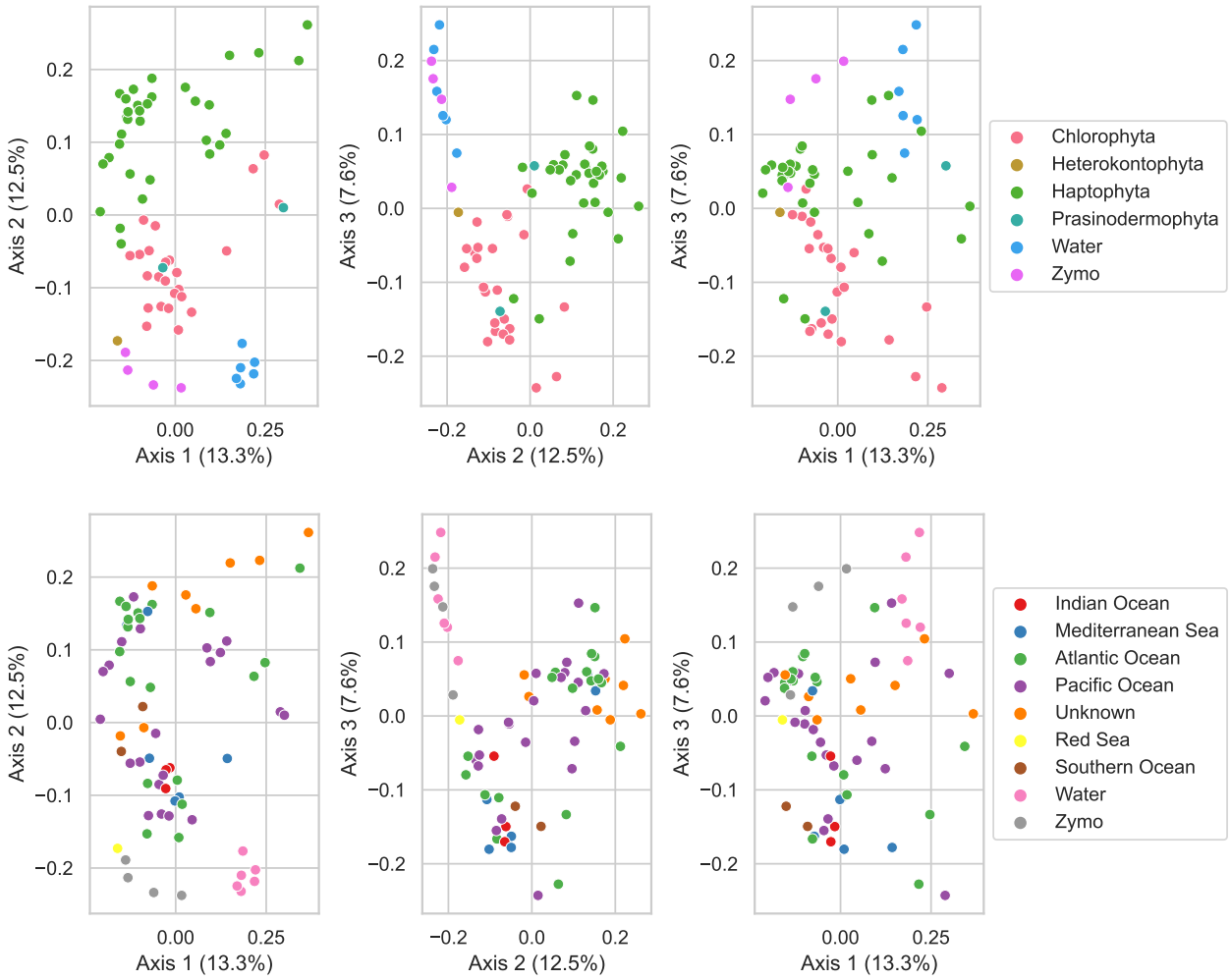

Figure 6. PCoA based on unweighted UniFrac distances.

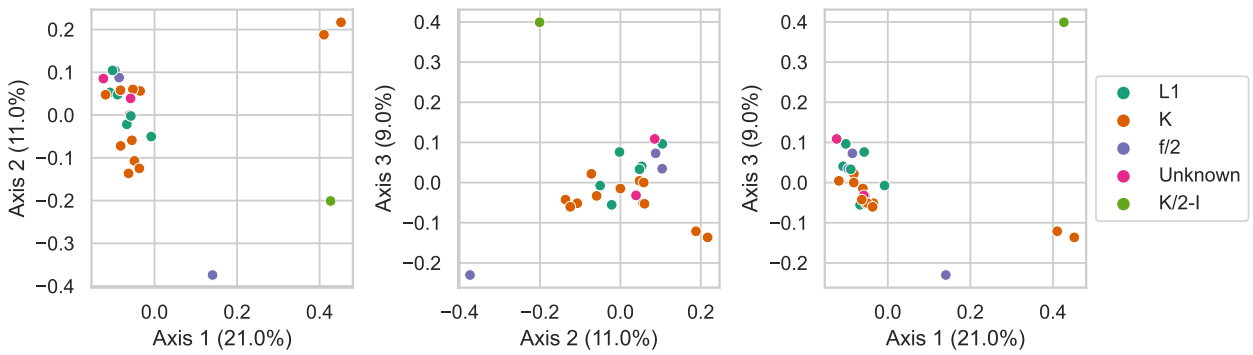

Figure 7. PCoA of Chlorophyta samples.

coloring by algal class - Chloropicophyceae (*Chloroparvula, Chloropicon*) grouped separately from Mamiellophyceae (*Micromonas, Bathycoccus*) (Figure 7). This difference is not, however, significant as tested by Adonis PERMANOVA (R2 = 0.142, p = 0.087).

Distance clustering The unweighted UniFrac distance matrix was hierarchically clustered (Figure 8). There are five main clades of the dendrogram. Both the positive and negative controls group within themselves and separate from the other samples. A sister group to them contains two large clades of *Phaeocystis* samples and one clade composed of mostly Chlorophyta, but which also includes two *Phaeocystis* samples and one each of *Prasinoderma* and *Minutocellus*. One clade forms a sister group to all others. It is composed of a mix of genera. These samples are not only distant from all others, but also from each other. While the control clades and one of the *Phaeocystis* clades are internally not distant, the Chlorophyta clade is not as internally cohesive. The last *Phaeocystis* clade does not display high similarity between members either.

5.5. Taxonomical composition of samples is specific to clusters

The clusters obtained from Figure 8 were used to subset the OTU table. The samples were then grouped by family with aggregation over the sum of relative counts (Figure 9). Their composition is further discussed below (Section 6.3).

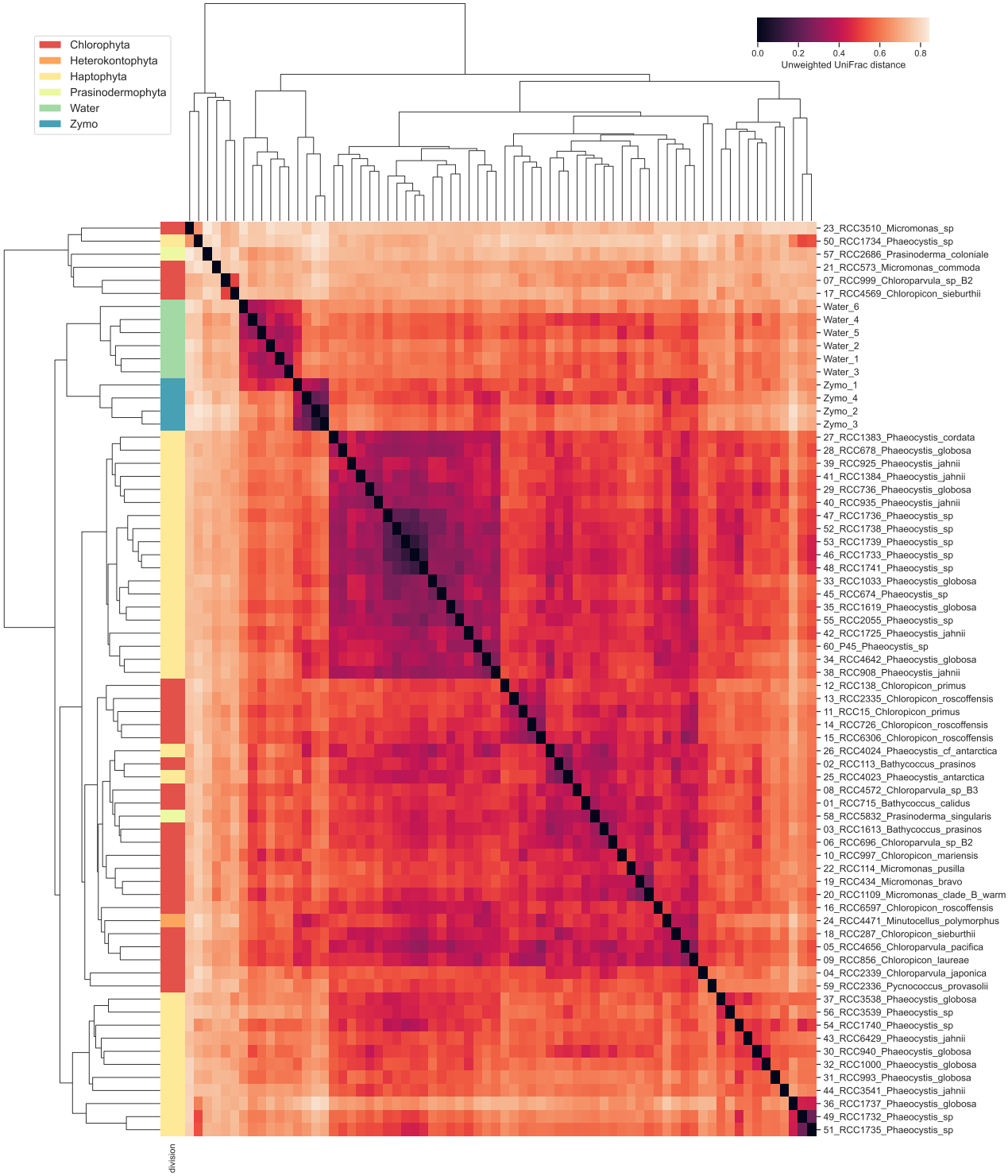

Figure 8. Unweighted UniFrac distance hierarchical clustering.

Figure 9. Family-level taxonomical assignment of OTUs. Plots were grouped by clusters from Figure 8. Traces indicated "Other" have less than 10% relative abundance.

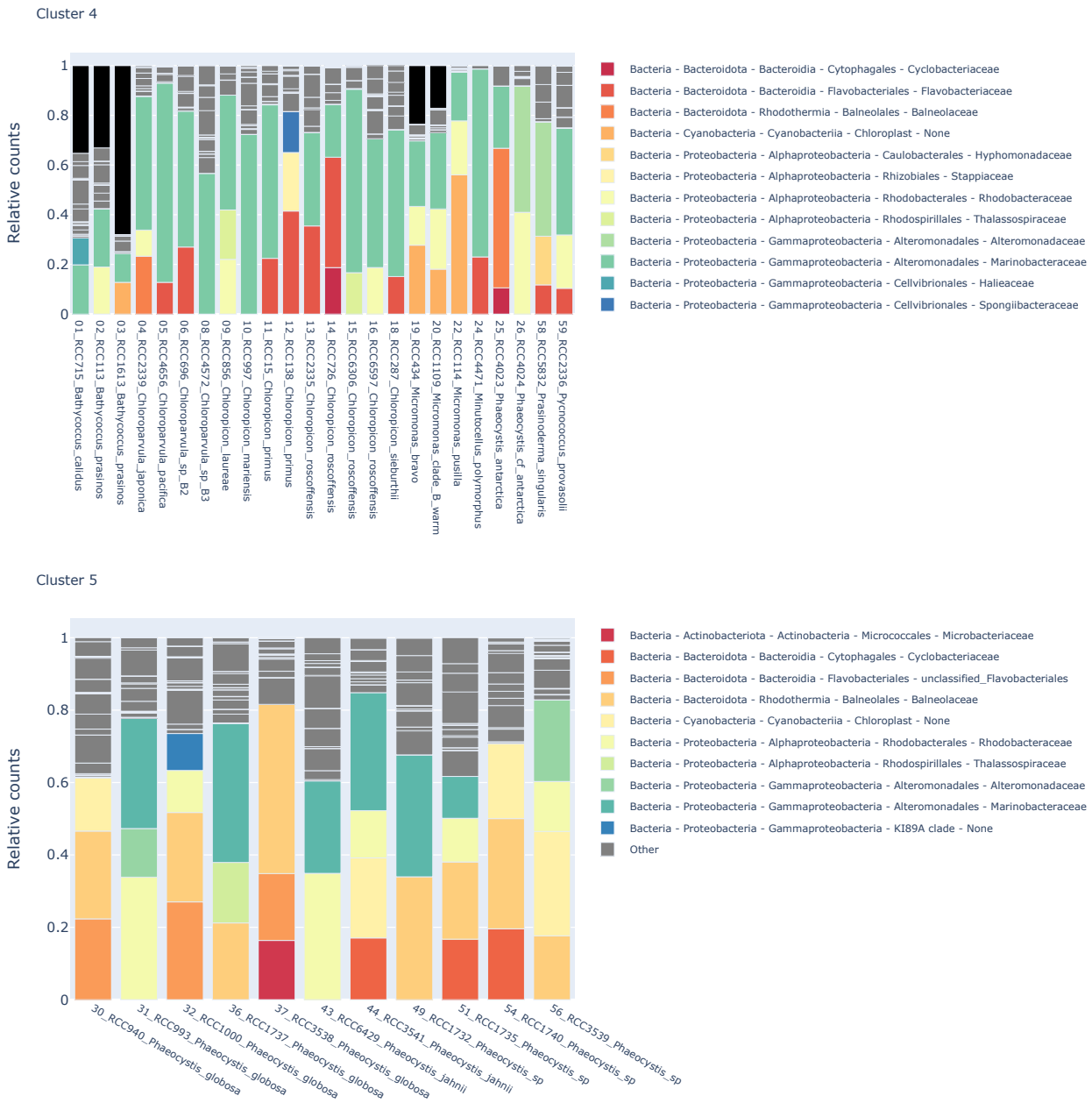

Figure 9. (Cont.) Family-level taxonomical assignment of OTUs. Plots were grouped by clusters from Figure 8. Traces indicated "Other" have less than 10% relative abundance.

6. Discussion

6.1. Negative controls are necessary in every extraction batch.

All the presented results have to be taken with caution since the negative extraction control was in a failed batch. Therefore, we do not know the extent of contamination which occurred during extraction. In a repeat experiment in the future, a water sample should be included in every batch. Some additional medium samples would also be useful to gauge if there are not any bacteria introduced by the periodic passaging of cultures in the RCC. Further, also extracting DNA from the 3 μm pore filter will help find host-associated microbes which, due to attachment, could have been left in the large particle fraction.

In the case of this experiment, we do see some amplification of the PCR negative control, which is nucleic acid-free water, also used in the reaction mixture of all samples. However, there are few counts in the water samples detected by sequencing. Most of the contamination is classified as *Prauserella* (otu115) and *Rubrobacter* (otu131), both in Actinobacteria. There is also some signal from OTU77, OTU97 and OTU132, each in a different replicate.

6.2. General diversity

Overall, the samples are relatively similar in their evenness, entropy, richness and phylogenetic diversity. While pairwise tests reveal differences between a number of pairs, the significance of this is not excessively high. There are also differences in distances between genera, but the amount of variance they explain is low (Figure 6).

The genera with larger sample sizes show larger numbers of unique ASVs (Figure 2), which is consistent with the fact that low abundance species will have a certain chance of detection which increases with attempts. The relative abundance of ASVs shows an exponential distribution, meaning they appear often in small amounts and very rarely in high abundance. Additionally, while it can be postulated that the RCC provides very stable growing conditions for their cultures, their microbiomes could still be subject to temporal variations caused by, for example, the time of day, the season, or a cycle we are not currently aware of. Therefore, there might be bacteria we are missing in some host samples because currently their number is low while at another time it would be much higher. The health or moment in the life cycle of the host could also play a role in the availability of nutrients for the microbes [42] - while all the cultures used had been grown by the RCC to sufficient levels that they were deemed healthy, some might have started dying before they were sampled over the course of a week.

Novel microbes Most of the collected ASVs were successfully identified at high taxonomic level. The only exception is OTU48, which was found in three *Phaeocystis* samples. This likely *Leptospira* relative might be an undescribed species, since its most closely related strains are not identified either. The sequences coming from unidentified classes are so low in abundance that it could be extremely difficult to obtain their cultures if they were desired.

6.3. Hierarchical clustering reveals differences in taxonomic composition.

- **Cluster 1**, which does not display inter-group similarity (Figure 8), also does not have any characteristic microbial families in common. *Chloropicon sieburthii* RCC4569 includes a large fraction of a Bacteroidia OTU which is not classified lower than class level, so this might be a novel lineage of bacteria.

- **Cluster 2** shows that the composition of positive controls (Zymo) was largely correctly identified - according to the supplier, it should contain 89.1% DNA from *Listeria monocytogenes*, 8.9% from *Pseudomonas aeruginosa*, 0.89% from *Bacillus subtilis* and low abundances of other microorganisms. Therefore, the identification of taxa can be reasonably trusted. The negative PCR controls had very low overall read abundances as well as concentrations of the 16S peak after amplification as measured on an automatic electrophoresis system. Additionally, most of the identified families have abundances lower than 10% in the water samples and the more abundant families in multiple of these samples were Pseudonocardiaceae and Rubrobacteriaceae, which were not found in large amounts elsewhere. Thus, it can be concluded that contamination in the PCR step was minimal.

- **Cluster 3** displays the highest level of similarity between samples (Figure 8). It consists only of *Phaeocystis* samples, which suggests selection for a set group of microorganisms by that genus. The most common are Bacteroidota representatives, as well as Alpha- and Gammaproteobacteria. Studies show that these groups are associated with algal blooms and fill in a similar ecological niche, namely the degradation of polysaccharides and other metabolites [43]. This is consistent with the previously mentioned ability of *Phaeocystis* to periodically produce large blooms with a polysaccharide extracellular matrix connecting their cells, which binds inorganic carbon and other nutrients into more bioavailable forms. One sample, *Phaeocystis globosa* RCC736, contains a relatively large percentage of *Methylophagaceae*, which are C1 degraders [43].

- **Cluster 4** is mostly made up of Chlorophyta samples, with the addition of both *Phaeocystis antarctica* replicates, as well as *Minutocellus polymorphus* and *Prasinoderma singularis*. This group has on average a larger proportion of Marinobacteraceae and closely related Alteromonadaceae than the previous one. The first family is in most part made up of *Marinobacter*, which, as mentioned above, was found in all samples. Indeed, this is a psychrophilic genus ubiquitous in the oceans and other salt-rich habitats. It is mostly heterotrophic and able to metabolize a wide variety of substrates [44].

- **Cluster 5** includes a similar proportion of Bacteroidota, but much less *Marinobacteraceae* proportionally to clades 3 and 4. Although, like cluster 3, it consists only of *Phaeocystis* samples, the inter-sample distance is much higher - the group is less homogeneous in its

microbiome. *Phaeocystic globosa* RCC1737 contains a further unclassified member of the KI89A clade of Gammaproteobacteria.

7. Conclusions

This project helped establish an effective method of obtaining microbial samples from pure but non-axenic algal strains. It revealed the main bacterial groups associated with these cultures. It also elucidated potential pitfalls in a larger-scale experiment, like not including enough extraction batches. The hosts included in the study show statistically significant differences in phylogeny-aware diversity metrics. Grouping by inter-sample distances reveals separate clades. These groups show visible differences in taxonomic composition. Overall, samples are associated with heterotrophic bacteria which have previously been reported to degrade organic compounds produced by algal blooms. Reproducing this study on a larger scale will help more closely specify interactions between marine algae and prokaryotes, as well as help identify new taxa, currently unknown to science. Knowing the host biochemistry could help culture difficult to obtain, but biosynthetically precious, bacterial strains.

8. Acknowledgements

I would like to thank Prof. Shinichi Sunagawa for allowing me to complete this project in his group, as well as my supervisors, Drs. Laucas Paoli and Samuel Miravet-Verde, for guiding me in this study and sharing their immense knowledge and experience. I am also very grateful to our lab manager Charlotte Carlström and technicians Jeremy Nathan and Nataliia Polishchuk for helping me in the experimental part. Finally, this research would not be possible without the collaboration with the Roscoff Culture Collection, who provided the algal strains and infrastructure to sample from them on-site.

9. References

[1] Katarzyna Zaremba-Niedzwiedzka, Eva F Caceres, Jimmy H Saw, Disa Bäckström, Lina Juzokaite, Emmelien Vancaester, Kiley W Seitz, Karthik Anantharaman, Piotr Starnawski, Kasper U Kjeldsen, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, 541(7637):353–358, 2017.

[2] José Diogo Santos, Inês Vitorino, Fernando Reyes, Francisca Vicente, and Olga Maria Lage. From ocean to medicine: Pharmaceutical applications of metabolites from marine bacteria. *Antibiotics*, 9(8):455, 2020.

[3] Carlos JC Rodrigues and Carla CCR de Carvalho. Cultivating marine bacteria under laboratory conditions: Overcoming the "unculturable" dogma. *Frontiers in Bioengineering and Biotechnology*, page 1476, 2022.

[4] Lucas Paoli, Hans-Joachim Ruscheweyh, Clarissa C Forneris, Florian Hubrich, Satria Kautsar, Agneya Bhushan, Alessandro Lotti, Quentin Clayssen, Guillem Salazar, Alessio Milanese, et al. Biosynthetic potential of the global ocean microbiome. *Nature*, 607(7917):111–118, 2022.

[5] Samay Pande and Christian Kost. Bacterial unculturability and the formation of intercellular metabolic networks. *Trends in microbiology*, 25(5):349–361, 2017.

[6] SA Amin, LR Hmelo, HM Van Tol, BP Durham, LT Carlson, KR Heal, RL Morales, CT Berthiaume, MS Parker, B Djunaedi, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, 522(7554):98–101, 2015.

[7] Jianming Zhu, Guofu Chen, Jin Zhou, Yanhua Zeng, Keke Cheng, and Zhonghua Cai. Dynamic patterns of quorum sensing signals in phycospheric microbes during a marine algal bloom. *Environmental Research*, 212:113443, 2022.

[8] Mariana Câmara dos Reis, Sarah Romac, Florence Le Gall, Dominique Marie, Miguel J Frada, Gil Koplovitz, Thierry Cariou, Nicolas Henry, Colomban de Vargas, and Christian Jeanthon. Exploring the phycosphere of emiliania huxleyi: from bloom dynamics to microbiome assembly experiments. *Molecular Ecology*, 2022.

[9] Jonathan J Cole. Interactions between bacteria and algae in aquatic ecosystems. *Annual review of Ecology and systematics*, 13(1):291–314, 1982.

[10] Martin T Croft, Andrew D Lawrence, Evelyne Raux-Deery, Martin J Warren, and Alison G Smith. Algae acquire vitamin b12 through a symbiotic relationship with bacteria. *Nature*, 438(7064):90–93, 2005.

[11] Mary Ann Moran and Bryndan P Durham. Sulfur metabolites in the pelagic ocean. *Nature Reviews Microbiology*, 17(11):665–678, 2019.

[12] The Roscoff Culture Collection. URL: https://roscoff-culture-collection.org, accessed: 2023-01-30.

[13] Sayed Asmat Ali Shah, Syed Shams Ul Hassan, Simona Bungau, Yongsheng Si, Haiwei Xu, Md Habibur Rahman, Tapan Behl, Daniela Gitea, Flavia-Maria Pavel, Raluca Anca Corb Aron, et al. Chemically diverse and biologically active secondary metabolites from marine phylum chlorophyta. *Marine drugs*, 18(10):493, 2020.

[14] Jin Zhou, Yihua Lyu, Mindy L Richlen, Donald M Anderson, and Zhonghua Cai. Quorum sensing is a language of chemical signals and plays an ecological role in algal-bacterial interactions. *Critical reviews in plant sciences*, 35(2):81–105, 2016.

[15] Walker O Smith Jr, Louis A Codispoti, David M Nelson, Thomas Manley, Edward J Buskey, H Joseph Niebauer, and Glenn F Cota. Importance of phaeocystis blooms in the high-latitude ocean carbon cycle. *Nature*, 352(6335):514–516, 1991.

[16] Kebi Wu, Si Tang, Xiaotian Wu, Jianming Zhu, Junting Song, Yanlin Zhong, Jin Zhou, and Zhonghua Cai. Colony formation of phaeocystis globosa: A case study of evolutionary strategy for competitive adaptation. *Marine Pollution Bulletin*, 186:114453, 2023.

[17] Bianca N Cruz and Susanne Neuer. Particle-associated bacteria differentially influence the aggregation of the marine diatom minutocellus polymorphus. *ISME Communications*, 2(1):73, 2022.

[18] T Hasegawa, H Miyashita, M Kawachi, H Ikemoto, N Kurano, S Miyachi, and M Chihara. Prasinoderma coloniale gen. et sp. nov., a new pelagic coccoid prasinophyte from the western pacific ocean. *Phycologia*, 35(2):170–176, 1996.

[19] Linzhou Li, Sibo Wang, Hongli Wang, Sunil Kumar Sahu, Birger Marin, Haoyuan Li, Yan Xu, Hongping Liang, Zhen Li, Shifeng Cheng, et al. The genome of prasinoderma coloniale unveils the existence of a third phylum within green plants. *Nature ecology & evolution*, 4(9):1220–1231, 2020.

[20] Alma E Parada, David M Needham, and Jed A Fuhrman. Every base matters: assessing small subunit rrna primers for marine microbiomes with mock communities, time series and global field samples. *Environmental microbiology*, 18(5):1403–1414, 2016.

[21] Amy Apprill, Sean McNally, Rachel Parsons, and Laura Weber. Minor revision to v4 region ssu rrna 806r gene primer greatly increases detection of sar11 bacterioplankton. *Aquatic Microbial Ecology*, 75(2):129–137, 2015.

[22] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1):10–12, 2011.

[23] Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson, and Susan P Holmes. Dada2: High-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7):581–583, 2016.

[24] Adithya Murali, Aniruddha Bhargava, and Erik S Wright. Idtaxa: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, 6(1):1–14, 2018.

[25] Robert C Edgar. Search and clustering orders of magnitude faster than blast. *Bioinformatics*, 26(19):2460–2461, 2010.

[26] Evan Bolyen, Jai Ram Rideout, Matthew R Dillon, Nicholas A Bokulich, Christian C Abnet, Gabriel A Al-Ghalith, Harriet Alexander, Eric J Alm, Manimozhiyan Arumugam, Francesco Asnicar, et al. Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature biotechnology*, 37(8):852–857, 2019.

[27] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[28] Evelyn C Pielou. The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, 13:131–144, 1966.

[29] Claude E Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[30] Daniel P Faith. Conservation evaluation and phylogenetic diversity. *Biological conservation*, 61(1):1–10, 1992.

[31] Kazutaka Katoh and Daron M Standley. Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–780, 2013.

[32] Morgan N Price, Paramvir S Dehal, and Adam P Arkin. Fasttree 2–approximately maximum-likelihood trees for large alignments. *PloS one*, 5(3):e9490, 2010.

[33] Catherine Lozupone, Manuel E Lladser, Dan Knights, Jesse Stombaugh, and Rob Knight. Unifrac: an effective distance metric for microbial community comparison. *The ISME Journal*, 5:169–172, 2011.

[34] Marti J Anderson. A new method for non-parametric multivariate analysis of variance. *Austral ecology*, 26(1):32–46, 2001.

[35] Nathan Halko, Per-Gunnar Martinsson, Yoel Shkolnisky, and Mark Tygert. An algorithm for the principal component analysis of large data sets. *SIAM Journal on Scientific Computing*, oct 2011. Publisher: Society for Industrial and Applied Mathematics.

[36] Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[37] Plotly Technologies Inc. Collaborative data science, 2015.

[38] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[39] Mark Johnson, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L Madden. Ncbi blast: a better web interface. *Nucleic acids research*, 36(suppl_2):W5–W9, 2008.

[40] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[41] Marti J Anderson. A new method for non-parametric multivariate analysis of variance. *Austral ecology*, 26(1):32–46, 2001.

[42] Justin R Seymour, Shady A Amin, Jean-Baptiste Raina, and Roman Stocker. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nature microbiology*, 2(7):1–12, 2017.

[43] Ben Francis, Tim Urich, Annett Mikolasch, Hanno Teeling, and Rudolf Amann. North sea spring bloom-associated gammaproteobacteria fill diverse heterotrophic niches. *Environmental Microbiome*, 16:1–16, 2021.

[44] Kim M Handley and Jonathan R Lloyd. Biogeochemical implications of the ubiquitous colonization of marine habitats and redox gradients by marinobacter species. *Frontiers in microbiology*, 4:136, 2013.

10. Supplemental Information

Table 3. Sample metadata

RCC ID	Species	Sampling ocean	RCC medium	RCC temp.
715	*Bathycoccus calidus*	Indian Ocean	L1	22 °C
113	*Bathycoccus prasinos*	Mediterranean Sea	K	20 °C
1613	*Bathycoccus prasinos*	Atlantic Ocean	K	15 °C
2339	*Chloroparvula japonica*	Pacific Ocean	L1	22 °C
4656	*Chloroparvula pacifica*	Pacific Ocean	L1	22 °C
696	*Chloroparvula* sp. B2	Indian Ocean	K	20 °C
999	*Chloroparvula* sp. B2	Pacific Ocean	K	20 °C
4572	*Chloroparvula* sp. B3	Atlantic Ocean	K	22 °C
856	*Chloropicon laureae*	Pacific Ocean	K	20 °C
997	*Chloropicon mariensis*	Pacific Ocean	K	20 °C
15	*Chloropicon primus*	Atlantic Ocean	f/2	22 °C
138	*Chloropicon primus*	Pacific Ocean	f/2	20 °C
2335	*Chloropicon roscoffensis*	Pacific Ocean	L1	22 °C
726	*Chloropicon roscoffensis*	Indian Ocean	L1	20 °C
6306	*Chloropicon roscoffensis*	Atlantic Ocean	Unknown	22 °C
6597	*Chloropicon roscoffensis*	Unknown	Unknown	Unknown
4569	*Chloropicon sieburthii*	Atlantic Ocean	K	22 °C
287	*Chloropicon sieburthii*	Pacific Ocean	K	20 °C
434	*Micromonas bravo*	Mediterranean Sea	K	20 °C
1109	*Micromonas* clade B warm	Mediterranean Sea	K	20 °C
573	*Micromonas commoda*	Mediterranean Sea	f/2	20 °C
114	*Micromonas pusilla*	Atlantic Ocean	K	20 °C
3510	*Micromonas* sp.	Atlantic Ocean	K/2-I	15 °C
4471	*Minutocellus polymorphus*	Red Sea	K	22 °C
4023	*Phaeocystis antarctica*	Southern Ocean	K/2ET	4 °C
4024	*Phaeocystis cf antarctica*	Southern Ocean	K/2ET	4 °C
1383	*Phaeocystis cordata*	Mediterranean Sea	K/2ET	15 °C
678	*Phaeocystis globosa*	Atlantic Ocean	K/2ET	15 °C
736	*Phaeocystis globosa*	Atlantic Ocean	K/2ET	15 °C
940	*Phaeocystis globosa*	Pacific Ocean	K/2ET	15 °C
993	*Phaeocystis globosa*	Pacific Ocean	K/2ET	15 °C
1000	*Phaeocystis globosa*	Pacific Ocean	K/2ET	15 °C
1033	*Phaeocystis globosa*	Pacific Ocean	K/2ET	15 °C
4642	*Phaeocystis globosa*	Pacific Ocean	K/2ET	15 °C
1619	*Phaeocystis globosa*	Atlantic Ocean	K/2ET	15 °C

Continued on next page

Table 3. Sample metadata

RCC ID	Species	Sampling ocean	RCC medium	RCC temp.
1737	*Phaeocystis globosa*	Atlantic Ocean	K/2ET	15 °C
3538	*Phaeocystis globosa*	Unknown	K/2ET	15 °C
908	*Phaeocystis jahnii*	Pacific Ocean	K/2ET	15 °C
925	*Phaeocystis jahnii*	Pacific Ocean	K/2ET	15 °C
935	*Phaeocystis jahnii*	Pacific Ocean	K/2ET	15 °C
1384	*Phaeocystis jahnii*	Mediterranean Sea	K/2ET	15 °C
1725	*Phaeocystis jahnii*	Atlantic Ocean	K/2ET	15 °C
6429	*Phaeocystis jahnii*	Pacific Ocean	Unknown	Unknown
3541	*Phaeocystis jahnii*	Unknown	K/2ET	15 °C
674	*Phaeocystis* sp.	Atlantic Ocean	K/2ET	15 °C
1733	*Phaeocystis* sp.	Atlantic Ocean	K/2ET	15 °C
1736	*Phaeocystis* sp.	Atlantic Ocean	K/2ET	15 °C
1741	*Phaeocystis* sp.	Pacific Ocean	K/2ET	15 °C
1732	*Phaeocystis* sp.	Unknown	K/2ET	15 °C
1734	*Phaeocystis* sp.	Unknown	K/2ET	15 °C
1735	*Phaeocystis* sp.	Unknown	K/2ET	15 °C
1738	*Phaeocystis* sp.	Atlantic Ocean	K/2ET	15 °C
1739	*Phaeocystis* sp.	Atlantic Ocean	K/2ET	15 °C
1740	*Phaeocystis* sp.	Atlantic Ocean	K/2ET	15 °C
2055	*Phaeocystis* sp.	Atlantic Ocean	K/2ET	15 °C
3539	*Phaeocystis* sp.	Unknown	K/2ET	15 °C
2686	*Prasinoderma coloniale*	Pacific Ocean	L1	22 °C
5832	*Prasinoderma singularis*	Pacific Ocean	L1	15 °C
2336	*Pycnococcus provasolii*	Pacific Ocean	L1	22 °C
P-45	*Phaeocystis* sp.	Unknown	Unknown	Unknown