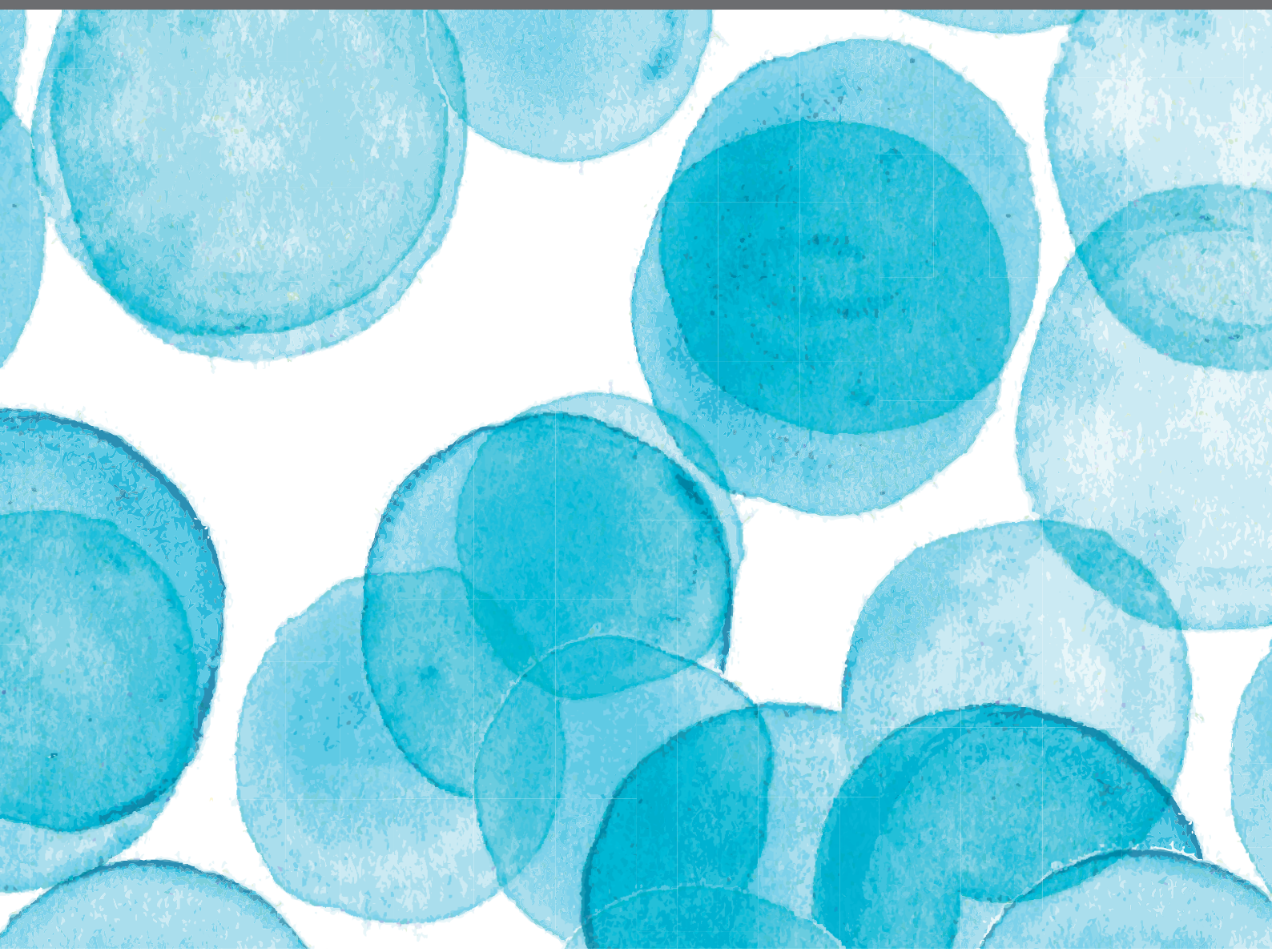
TOWARDS A MECHANISTIC UNDERSTANDING OF MARINE MICROBIAL ECOSYSTEMS

EDITED BY: Julia C. Engelmann, Susanne Wilken, Karoline Faust and Gipsi Lima-Mendez

PUBLISHED IN: Frontiers in Microbiology, Frontiers in Marine Science and Frontiers in Genetics



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-201-1
DOI 10.3389/978-2-88974-201-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

TOWARDS A MECHANISTIC UNDERSTANDING OF MARINE MICROBIAL ECOSYSTEMS

Topic Editors:
Julia C. Engelmann, Royal Netherlands Institute for Sea Research (NIOZ), Netherlands
Susanne Wilken, University of Amsterdam, Netherlands
Karoline Faust, KU Leuven, Belgium
Gipsi Lima-Mendez, Catholic University of Louvain, Belgium

Citation: Engelmann, J. C., Wilken, S., Faust, K., Lima-Mendez, G., eds. (2022). Towards a Mechanistic Understanding of Marine Microbial Ecosystems. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-201-1

Table of Contents

frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 May 2020
doi: 10.3389/fmicb.2020.01011

Sediment Microbial Communities and Their Potential Role as Environmental Pollution Indicators in Xuande Atoll, South China Sea

*Biao Zhang[1], Yan Li[1], Shi-Zheng Xiang[1], Yu Yan[1], Rui Yang[1], Meng-Ping Lin[1], Xue-Mu Wang[2], Yu-Long Xue[2] and Xiang-Yu Guan[1,3]**

[1] School of Ocean Sciences, China University of Geosciences, Beijing, China, [2] Marine Geological Survey Institute of Hainan Province, Haikou, China, [3] Hebei Marine Resource Survey Center, Qinhuangdao, China

In this study, 39 sediment samples were collected from Qilian Island, Iltis Bank, and Yongxing Island in Xuande Atoll in the South China Sea (SCS), and the microbial community structures and distribution were analyzed. The microbial community was influenced by both natural environmental factors and human activities. The abundance of genera *Vibrio* and *Pseudoalteromonas*, which are associated with pathogenicity and pollutant degradation, were significantly higher in Qilian Island than in Yongxing Island and Iltis Bank, suggesting possible contamination of Qilian Island area through human activities. Pathogenic or typical pollutants-degrading bacteria were found to be negatively correlated with most of the commonly occurring bacterial populations in marine sediment, and these bacteria were more likely to appear in the sediment of deep water layer. This co-occurrence pattern may be due to bacterial adaptation to environmental changes such as depth and contaminations from human activities, including garbage disposal, farming, and oil spills from ships. The findings of this study could help in understanding the potential influences of human activities on the ecosystem at the microbial level.

Keywords: environmental pollution indicator, microbial community, network, sediment, South China Sea

OPEN ACCESS

Edited by:
Karoline Faust,
KU Leuven, Belgium

Reviewed by:
Miguel Lurgi,
Swansea University, United Kingdom
Julia C. Engelmann,
Royal Netherlands Institute for Sea Research (NIOZ), Netherlands

****Correspondence:***
Xiang-Yu Guan
guanxy@cugb.edu.cn

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *13 September 2019*
Accepted: *24 April 2020*
Published: *25 May 2020*

Citation:
Zhang B, Li Y, Xiang S-Z, Yan Y, Yang R, Lin M-P, Wang X-M, Xue Y-L and Guan X-Y (2020) Sediment Microbial Communities and Their Potential Role as Environmental Pollution Indicators in Xuande Atoll, South China Sea. Front. Microbiol. 11:1011. doi: 10.3389/fmicb.2020.01011

INTRODUCTION

Marine sediments are an important component of the marine environment and represent the largest organic carbon reservoir on Earth (Parkes et al., 2000). Marine sediments, which are mainly derived from continental transport and sedimentation of biological products, are rich in nutrients and provide important habitats for microorganisms (Kallmeyer et al., 2012; Parkes et al., 2014). Microorganisms in sediments play key roles in the regulation of major geochemical and eco-environmental processes of marine ecosystems, particularly nutrient dynamics and biogeochemical cycles (Fuhrman, 2009; Nogales et al., 2011; Graham et al., 2012). Environmental factors, such as pH, water temperature, silicate, and ocean currents, cause differences in the structure and diversity of marine microbial communities (Gilbert et al., 2009; Hollister et al., 2010; Kirchman et al., 2010). Besides, the dispersion and movement of microorganisms are also driven by hydrography (Hamdan et al., 2013). The grain-size distribution of sediments can reflect the hydrodynamic conditions,

such as tide and ocean current. Bimodal distribution and large difference of sediment particle size indicate complex hydrodynamic forces (Anthony and Héquette, 2007; Hu et al., 2012).

The microbial community composition of sediment samples has been noted to show geographic differences. Examination of the microbial community from several marginal waters and sediments of the Sea of Japan, South China Sea (SCS) Trough, Sea of Okhotsk, Peruvian Marginal Sea, revealed that the structure and diversity of microbial communities vary among different regions (Rochelle et al., 1994; Inagaki et al., 2003; Newberry et al., 2004; Webster et al., 2006). For instance, the proportions of some genera belonging to the phyla *Proteobacteria*, *Firmicutes*, *Actinobacteria,* and *Bacteroides,* which can use petroleum as carbon sources, are increased in coastal sediments that are heavily polluted by human activities such as Hangzhou Bay and Bohai Bay; in addition, classes such as γ-*Proteobacteria* and δ-*Proteobacteria* have been found to dominate such areas (Kochling et al., 2011; Wang et al., 2013; Guan et al., 2014; Lu et al., 2016; Su et al., 2018). These microorganisms are considered to be potential indicator groups (Xiong et al., 2014). In addition, genera such as *Firmicutes* and *Bacilli* have also been noted to be dominant in many areas that have been heavily contaminated and disturbed by human activities (Zhang et al., 2008; Lu et al., 2016). Microbial communities in marine sediments are significantly influenced by changes in spatial distribution and environment factors, and specific microorganisms can be used as environmental indicators (Gillan et al., 2005; Glasl et al., 2019).

The SCS is located in the south of mainland China, and is exposed to complex tidal and ocean currents (Lien et al., 2005; Liu et al., 2010; Xiaodong Deng and Cai, 2013), which might affect migration and the distribution of microbial communities. Microbial communities in the sediments of the SCS are affected by both natural and anthropogenic factors. In the SCS, the sediment microbial diversity is high, and γ-*Proteobacteria*, *Planctomycetes* and *Bacteroidetes* are the main taxa in this area. Moreover, new unnamed bacteria are constantly being identified and different sediments have been shown to have different microbial community structures (Zhu et al., 2013; Yu et al., 2018). There are a large number of protease-producing bacteria in γ-*Proteobacteria*, which play an important role in the degradation of organic nitrogen in sediments (Zhou et al., 2009). The composition and diversity of microbial communities have been noted to show spatial and seasonal changes in the SCS (Du et al., 2011), and total organic carbon and nitrogen also impact bacterial community structure (Li and Wang, 2014). The hypoxic environment of the seabed forms a unique microbial community in which the anaerobic group of the community is more active (Walsh et al., 2016). In recent years, human activities and offshore oil spills have led to serious threats to the marine ecological environment, as well as causing significant changes to the structure and diversity of microbial communities in this region (Yuan, 2015; Limei Guo et al., 2017; Lufei Yang and Li, 2019).

Xuande Atoll is the largest residential area in the SCS, and is possibly influenced by human activities such as garbage disposal, farming, and oil spills from ships. In the present study, different sediment samples from three island reefs in Xuande Atoll area were investigated to elucidate the structure of the sediment microbial community, microbial response to human activities and environmental factors, and the role of the microbial community as indicators of pollution. The goals of this study were: (1) to identify the spatial distribution of microbial community structure in sediments among different islands and reefs, and the main impacts of various environmental factors on microbial communities; (2) to understand the role of sediment microorganisms in the SCS as indicators of environmental pollution associated with human activities; and (3) to reveal the possible response mechanism of sediment microbial communities to natural and anthropogenic factors. The results of this study will provide a reference for future environmental protection in the SCS, as well as impart valuable information on microbial indicators in the SCS.

MATERIALS AND METHODS

Sample Collection and Grouping

A total of 39 surface sediment samples were collected from the Xuande Atoll area of the SCS using a grab sampler from September to October of 2015. The water depth of the sampling points ranged from 16.5 to 97.6 m. To study the microbial community structure among sediments at different water depths, the samples were divided into three groups, namely, L (15–50 m), M (50–70 m), and H (70–100 m) according to the water depth. The sediment samples with three depth layers were located at Qilian Island (QL1-8, QM9-21, QH22-25), Iltis Bank (IL26-31, IM32, IH33-34), and Yongxing Island (YL35, YM36-39) (**Figure 1** and **Table 1**). To investigate the distribution of the regional microbial communities and the differences in microbial community structure among the three islands, all 39 samples were divided into three groups according to the distribution of sampling points on the three islands and three depth groups (**Tables 2**, **3**). The collected sediment samples were quickly packed in 50 mL sterile centrifuge tubes and frozen at −20°C until they were transported to the laboratory, where they were stored at −80°C, and the genomic DNA was extracted as soon as possible.

Measurement of Environmental Factors

The sediment samples were centrifuged at 5000 rpm for 30 min at room temperature, and the supernatant sediment pore water was filtered through a 0.45 μm Millipore filter. The pH and salinity of the sediment pore water were measured using a YSI Professional Plus handheld multi-parameter water quality meter (Yellow Springs Instrument Company, United States). The concentrations of ammonium (NH_4^+), nitrate (NO_3^-), and nitrite (NO_2^-) were measured by using UV-visible spectrophotometer HP 8453 (Hewlett-Packard, United States) at wavelengths of 220–275, 540, and 420 nm, respectively. Ion chromatography was employed for the detection of chloridion (Cl^-), sulfate (SO_4^{2-}), calcium (Ca^{2+}), and potassium (K^+) (Thermo Fisher, United States). The total organic carbon (TOC) of each sample was determined by a TOC analyzer (Shimadzu, Japan) with a detection limit

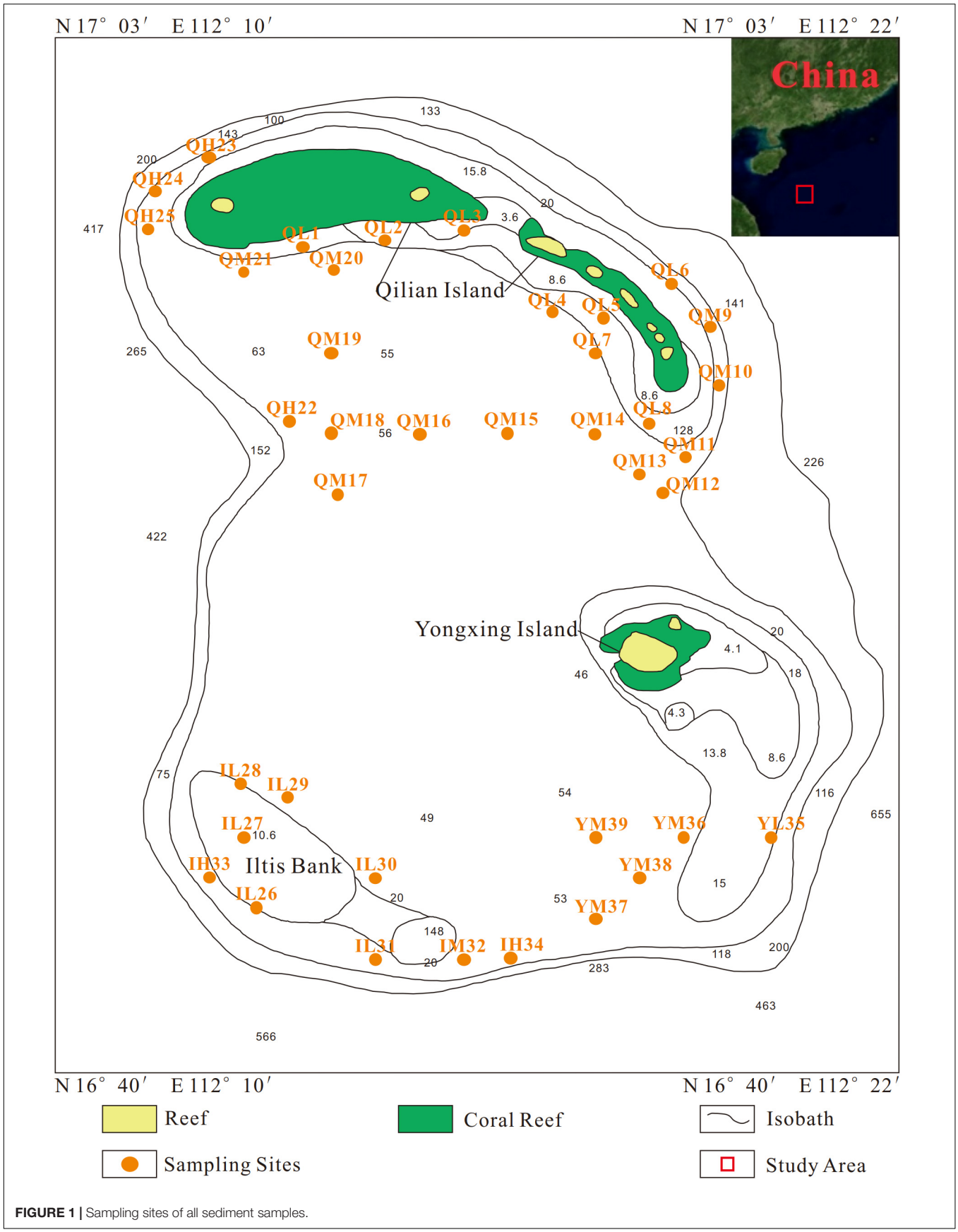


FIGURE 1 | Sampling sites of all sediment samples.

TABLE 1 | Sampling locations and depths.

Station	Longitude (E)	Latitude (N)	Depth (m)	Station	Longitude (E)	Latitude (N)	Depth (m)
QL1	112°14′09.208″	16°57′49.318″	43.0	QM21	112°13′04.165″	16°57′21.176″	60.0
QL2	112°15′37.499″	16°57′56.207″	34.7	QH22	112°13′48.754″	16°54′29.501″	73.6
QL3	112°17′03.659″	16°57′19.796″	18.1	QH23	112°12′26.245″	16°59′31.442″	77.2
QL4	112°18′39.888″	16°56′31.472″	45.4	QH24	112°11′27.665″	16°58′54.056″	97.6
QL5	112°19′31.573″	16°56′23.678″	16.5	QH25	112°11′18.943″	16°58′09.164″	90.3
QL6	112°20′48.502″	16°57′04.015″	39.9	IL26	112°13′11.684″	16°45′18.651″	33.2
QL7	112°19′25.512″	16°55′45.652″	45.5	IL27	112°12′59.184″	16°46′37.231″	16.7
QL8	112°20′23.188″	16°54′25.495″	20.1	IL28	112°12′53.783″	16°47′39.007″	31.1
QM9	112°21′30.552″	16°56′14.587″	59.2	IL29	112°13′47.543″	16°47′22.948″	49.1
QM10	112°21′37.774″	16°55′10.191″	59.8	IL30	112°15′21.089″	16°45′50.219″	43.3
QM11	112°21′02.837″	16°53′47.788″	63.4	IL31	112°15′22.167″	16°44′17.336″	48.1
QM12	112°20′37.648″	16°53′06.029″	60.1	IM32	112°16′58.167″	16°44′19.249″	59.7
QM13	112°20′13.097″	16°53′29.095″	55.6	IH33	112°12′20.320″	16°45′52.874″	88.2
QM14	112°19′25.625″	16°54′14.509″	51.1	IH34	112°17′52.638″	16°44′29.063″	80.8
QM15	112°17′49.744″	16°54′15.635″	57.1	YL35	112°22′31.561″	16°46′33.159″	34.1
QM16	112°16′13.693″	16°54′16.471″	58.3	YM36	112°20′56.833″	16°46′33.766″	52.1
QM17	112°14′43.911″	16°53′05.624″	63.6	YM37	112°19′20.472″	16°45′2.555″	56.5
QM18	112°14′37.603″	16°54′17.186″	68.8	YM38	112°20′9.039″	16°45′48.223″	59.1
QM19	112°14′39.385″	16°55′48.354″	63.1	YM39	112°19′22.209″	16°46′33.138″	59.6
QM20	112°14′39.003″	16°57′21.742″	53.4				

of 0.1 mg/L. Sediment samples were processed as previously described (Shuqing et al., 2010), and sediment particle size was measured using a Mastersizer 2000 laser particle size analyzer (Malvern Instruments, United Kingdom).

16S Amplicon Sequencing and Data Processing

Total DNA from each sample (0.5 g) was extracted using a PowerSoil DNA Extraction Kit (QIAGEN Instruments, United States). The extracted DNA was stored at −20°C for further use and at −80°C for permanent preservation. The quantity and quality of isolated DNA were evaluated using a Nano Drop spectrophotometer (Thermo Fisher Scientific, United States) and agarose gel electrophoresis (Bio-Rad, United States), respectively.

The V4-V5 regions of the bacterial 16S ribosomal RNA gene was amplified by PCR (95°C for 2 min, followed by 25 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s and then final extension at 72°C for 5 min) using the 515F 5′-barcode-GTGCCAGCMGCCGCGG)-3′ and 907R 5′-barcode-CCGTCAATTCMTTTR AGTTT-3′ primers, where the barcode was an eight-base sequence unique to each sample. PCR reactions were performed in triplicate 20 μL mixtures containing 4 μL of 5× FastPfu Buffer, 2 μL of 2.5 mM dNTPs, 0.8 μL of each primer (5 μM), 0.4 μL of FastPfu Polymerase, and 10 ng of template DNA. Amplicons were extracted from 2% agarose gels and purified using an AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, United States), according to the manufacturer's instructions, then quantified using QuantiFluor^{TM}-ST (Promega, United States). Purified amplicons were pooled in equimolar concentrations and paired-end sequenced (2 × 300) on an Illumina MiSeq platform according to the standard protocols.

Raw fastq files were demultiplexed, then quality-filtered using QIIME (version 1.17) with the following criteria: (i) the 300 bp reads were truncated at any site receiving an average quality score <20 over a 50 bp sliding window and truncated reads that were shorter than 50 bp were discarded; (ii) exact barcode matching, two nucleotide mismatches during primer matching, reads containing ambiguous characters were removed, and (iii) only sequences that overlapped more than 10 bp were assembled according to their overlap sequence. Reads that could not be assembled were discarded.

SILVA 132 rRNA database[1] was employed as the reference database, and RDP Classifier 2.11 was used for taxonomy

[1]https://www.arb-silva.de/

TABLE 2 | Samples from different islands.

Island	Samples
Qilian Island	QL1, QL2, QL3, QL4, QL5, QL6, QL7, QL8, QM9, QM10, QM11, QM12, QM13, QM14, QM15, QM16, QM17, QM18, QM19, QM20, QM21, QH22, QH23, QH24, QH25
Iltis Bank	IL26, IL27, IL28, IL29, IL30, IL31, IM32, IH33, IH34
Yongxing Island	YL35, YM36, YM37, YM38, YM39

TABLE 3 | Samples from different depths.

Depth	Samples
L (15–50 m)	QL1, QL2, QL3, QL4, QL5, QL6, QL7, QL8, IL26, IL27, IL28, IL29, IL30, IL31, YL35
M (50–70 m)	QM9, QM10, QM11, QM12, QM13, QM14, QM15, QM16, QM17, QM18, QM19, QM20, QM21, IM32, YM36, YM37, YM38, YM39
H (70–100 m)	QH22, QH23, QH24, QH25, IH33, IH34

annotation[2]. Operational Taxonomic Units (OTUs) were clustered with a 97% similarity cutoff using UPARSE (version 7.1)[3] and chimeric sequences were identified and removed using UCHIME.

Statistical Analysis

Among the 39 samples, the particles of 34 sediment samples were in compliance with the Mastersizer 2000 test. The Udden-Wentworth φ particle size standard was used for the grain size, and the average particle size was calculated by the moment method (McManus, 1988) as follows: Average particle size (Mz) = 1/φ. The remaining five samples, QL6, QL8, QM13, QM14, and SL35, could not be measured because of the large particle size and were therefore assigned values of 0.331, 0.332, 0.333, 0.334, and 0.335, respectively, based on three times the standard deviation.

To determine the alpha diversity, we rarified the OTUs and calculated the Chao index (species abundance), Shannon's diversity (community diversity), and Shannon's evenness (community uniformity). To test whether the sequencing results represented the actual situation of the microorganisms in the sample, the Coverage value reflecting the coverage of the sample library was calculated using Mothur. Bray-Curtis dissimilarity between samples was calculated using the R package ecodist. In addition, Spearman's correlation between environmental factors and microbial diversity indices were calculated using the R package psych. Wilcoxon rank-sum tests of environmental factors among different islands or depths were conducted using the R package stats. Heatmaps to show the composition of the bacterial community and Spearman's correlation between bacterial abundance and environmental factors were generated using the R package psych. Differential testing of bacterial abundance at the genus level among islands or depth groups was conducted using the Kruskal-Wallis H test in the R package statistics, and the *p*-value was checked using false discovery rate (FDR).

To conduct co-occurrence network analysis, the Spearman correlations of bacteria were computed based on the relative abundance of genus, and the networks were visualized using the R igraph package,[4] after which the modularity of the network was calculated and modules were detected using the greedy modularity optimization method (Deng et al., 2012). Non-metric multidimensional scaling (NMDS) analysis based on Bray-Curtis distance was performed using the R package "ecodist," which demonstrated the similarity of the microbial community structure among the 39 samples. To analyze the similarities among samples from different depths and spatial distribution, ANOSIM was employed to examine whether the differences among the groups were significantly higher than those within groups, which was calculated using the R package vegan, function "anosim." Constrained ordination methods (CCA) were calculated using the R package vegan, function "cca" to investigate the key factors affecting the variation in the microbial communities among different sites.

[2]http://sourceforge.net/projects/rdp-classifier/

[3]http://drive5.com/uparse/

[4]http://igraph.org

RESULTS AND DISCUSSION

Physicochemical Factors of Sediments

The particle size distribution curve of the sediments showed two clear components in the grain-size frequency curves (**Supplementary Figure S1**). The finer peaks concentrate at ca. 30 μm, which can be found for all the samples except QM12 (**Supplementary Figure S1A**) and IL27 (**Supplementary Figure S1C**), containing one peak at ∼60 μm. On the contrary, the coarser peaks distribute at ca. 120 μm, existing only in several samples either depositing in proximal or at the steep slope, and all the samples are at Qilian Island except IH34. The finer peak likely represents the mean hydro-energy of the current westward in our study area, while the appearance of the coarse fraction is determined by the submarine topography, location or the source distance. The mean grain-size of the samples collected from the open water area is relatively small and characterized by the solely finer component, might be indicative of the influence of the current transportation. However, as a consequence of proximal deposition, sediments collected from the areas close to the islands and banks are significantly coarser, such as sample QM10 and QM11. Moreover, as representative of the maximum hydro-energy, cumulative 1% of the grain-size for these coarse samples around 1900 μm correlated to the rolling transportation (Passega, 1957, 1964). Therefore, we inferred that the commonly existing fine fraction with peak at 30 μm was indicative of the current hydro-energy, and the coarse fraction is related to the proximal deposition especially in Qilian Island.

The environmental factors are shown in **Supplementary Table S1**. The concentrations of TOC of pore water of sediments at Qilian Island were significantly higher than those at Iltis Bank (Wilcoxon's test, $p < 0.01$), suggesting that the content of organic matter in the sediment of Qilian Island was significantly higher than that at Iltis Bank. The concentrations of NH_4^+ and PO_4^{3-} at Qilian Island were significantly higher than those at Yongxing Island (Wilcoxon's test, $p < 0.01$). The concentrations of TOC in sediments of water depth H (70–100 m) were significantly higher than those of water depth M (50–70 m) (Wilcoxon's test, $p < 0.05$). The concentrations of NO_2^- in sediments of water depth M (50–70 m) were significantly higher than those of water depth L (15–50 m) (Wilcoxon's test, $p < 0.05$). The concentrations of NO_3^- in sediments of water depth H (70–100 m) were significantly higher than those of water depth L (15–50 m) (Wilcoxon's test, $p < 0.05$), suggesting that the concentrations of TOC and NO_3^- in sediments of water depth H (70–100 m) were higher than those of water depth 15–70 m.

Microbial Community Structure and Its Correlation With Environmental Factors

A total of 4,404,200 high-quality sequences were retained for post-run analysis. A total of 9698 OTUs were assigned at a 3% dissimilarity threshold. The microbial diversity indices are shown in **Supplementary Table S2**. The correlation between

the microbial diversity index and environmental factors showed that the Shannon and Shannon even indices were significantly negatively correlated with TOC, NO_3^-, and depth, while Chao index was significantly negatively correlated with TOC and NO_3^- (**Table 4**).

The dominant phyla in the sediments were *Proteobacteria* (55.50%, mainly γ-, δ-, α-), *Actinobacteria* (8.54%), *Bacteroidetes* (6.56%), and *Firmicutes* (6.27%) (**Supplementary Figure S2**). *Proteobacteria* are widely found in sediments and distributed in marine sediments worldwide (Bienhold et al., 2016). The top 20 dominant genera were *Escherichia-Shigella*, *Vibrio, Pseudoalteromonas, Psychrobacter, Bacillus*, and *Lactococcus,* and some norank bacteria (**Figure 2**). We found that *Bacillus* and *Lactococcus* had a high abundance in samples QH25 and YH33. Most *Bacillus* are saprophytic, with a wide range of habitats, and some strains are pathogenic (Ruger et al., 2000). *Lactococcus* is found in some cultured marine fish and can cause fish diseases (Tanekhy, 2013). *Psychrobacter* was in high abundance in sample QL1, and was detected only in oil-contaminated waters (Prabagaran et al., 2007). Additionally, *Escherichia-Shigella* was found in high abundance at eight sampling points from QH23 to YM32 distributed at different water depths and different islands (**Figure 2**). *Escherichia-Shigella* is a pollution indicator and part of the main bacterial group responsible for diarrhea worldwide (Brunette et al., 2015). We speculate that the frequent and complex horizontal and vertical currents in the SCS (Liu and Gan, 2017) and the frequent flow of people in this region are likely responsible for the presence and spread of *Escherichia-Shigella*. *Vibrio* was found to be highly abundant at four sampling points (QM9, QM11, QM12, and QM13). Many strains of *Vibrio* are well-known pathogenic bacteria that can cause disease in humans or marine animals (Xiaohua et al., 2018). The increased abundance of *Vibrio* is associated with algal outbreaks (such as diatoms and brown algae) (Miller et al., 2005; Gilbert et al., 2012). In the Deepwater Horizon oil spill in the Gulf of Mexico, *Vibrio* was dominant in the petroleum-associated microbial community in surface seawater samples (Hamdan and Fulmer, 2011) and salt marsh plant samples contaminated by petroleum foam (Liu and Liu, 2013). In addition, QM9, QM11, QM12, and QM13 at southeast of Qilian Island are exposed to frequent human activities. Therefore, it is very likely that the domestic garbage on the island could lead to eutrophication of the sea, causing high growth of *Vibrio*. *Pseudoalteromonas* was highly abundant in QM14, reaching 93.32%. This genus is mainly distributed in open sea and coastal waters, affecting sedimentation, germination, and metamorphosis of various invertebrates and algae (Holmstrom and Kjelleberg, 1999; Bowman, 2007). In the present study, high abundances of *Escherichia-Shigella, Vibrio, Pseudoalteromonas, Bacillus, Psychrobacter,* and *Lactococcus* were distributed almost around Qilian Island, suggesting that Qilian Island has rapidly developed with frequent human activities, including garbage disposal, farming, and oil spills from ships, which had caused an increase in pathogenic bacteria and pollution indicator bacteria.

TABLE 4 | Correlations between environmental factors and microbial diversity indices.

	Shannon	Chao	Shannoneven
Depth	−0.338*	−0.191	−0.371*
TOC	−0.671***	−0.587***	−0.650***
NH_4^+	−0.268	−0.261	−0.307
PO_4^{3-}	−0.276	−0.176	−0.320*
NO_2^-	−0.302	−0.182	−0.302
NO_3^-	−0.686***	−0.520***	−0.698***
Cl^-	−0.171	−0.050	−0.169
SO_4^{2-}	−0.033	0.120	−0.060
K^+	0.114	0.156	0.116
Ca^{2+}	−0.027	0.113	−0.030
Mz	−0.110	0.013	−0.114
Salinity	0.228	0.121	0.261
pH	0.271	0.270	0.295

*Spearman correlation: *, $p < 0.05$; ***, $p < 0.001$.*

As shown in **Figure 3A**, the top 50 genera were divided into three clusters, G1, G2, and G3, based on the clustering tree. G1 was positively correlated with NH_4^+, PO_4^{3-}, TOC, and Mz, but was negatively correlated with salinity and pH. However, G2 was positively correlated with TOC. G3 contained the largest number of genera, almost all of which were unnamed bacteria in the ocean and negatively correlated with depth, NO_2^-, NH_4^+, PO_4^{3-}, TOC, and NO_3^-. The CCA further revealed the correlation between environmental factors and unique genera (**Supplementary Figure S3**). The genera *Vibrio* and *Pseudoalteromonas* clustered in G1 had positive correlation with Mz, PO_4^{3-}, NO_3^-, and TOC; *Psychrobacter* was mainly affected by NH_4^+ and TOC; and *Escherichia-Shigella*, *Bacillus*, *Pseudomonas,* and *Lactococcus* belonging to G2 presented positive relationship with TOC. These findings indicated that different taxon clusters in the marine sediment showed varied responses to environmental factors. However, the effects of different bacteria and functional groups on environmental factors need to be further investigated, along with the influence of environmental factors on different bacteria and functional groups.

In complex environments, network analysis provides a promising beginning for exploration of the organization and dynamics of microbial interactions and niches (Duran-Pinedo et al., 2011; Zhou et al., 2011; Faust and Raes, 2012). To identify functional groups that may interact with each other or share niches in sediments, we constructed a co-occurrence network of sediment bacteria in the study area. Upon modular analysis of the created network, 50 nodes and 468 edges were divided into seven modules (top 50 genera, $r > 0.6$, $p < 0.05$, **Supplementary Table S3**). There were many unclassified and norank bacteria in modules 1 and 5, and they were widely distributed and dominant in the marine environment (**Figure 3B**). These findings suggest that modules 1 and 5 have high co-occurrence, which may represent two main co-evolutions of functional communities. To investigate the factors affecting the classification and abundance of bacterial clusters, we analyzed the network diagram of the correlation between modules and environmental factors (**Figure 3C**). Modules 1 and 5, which had high diversity, shared positive relationship with each other, but had negative relationship with modules 3, 6, and 7. Unlike modules 1 and 5,

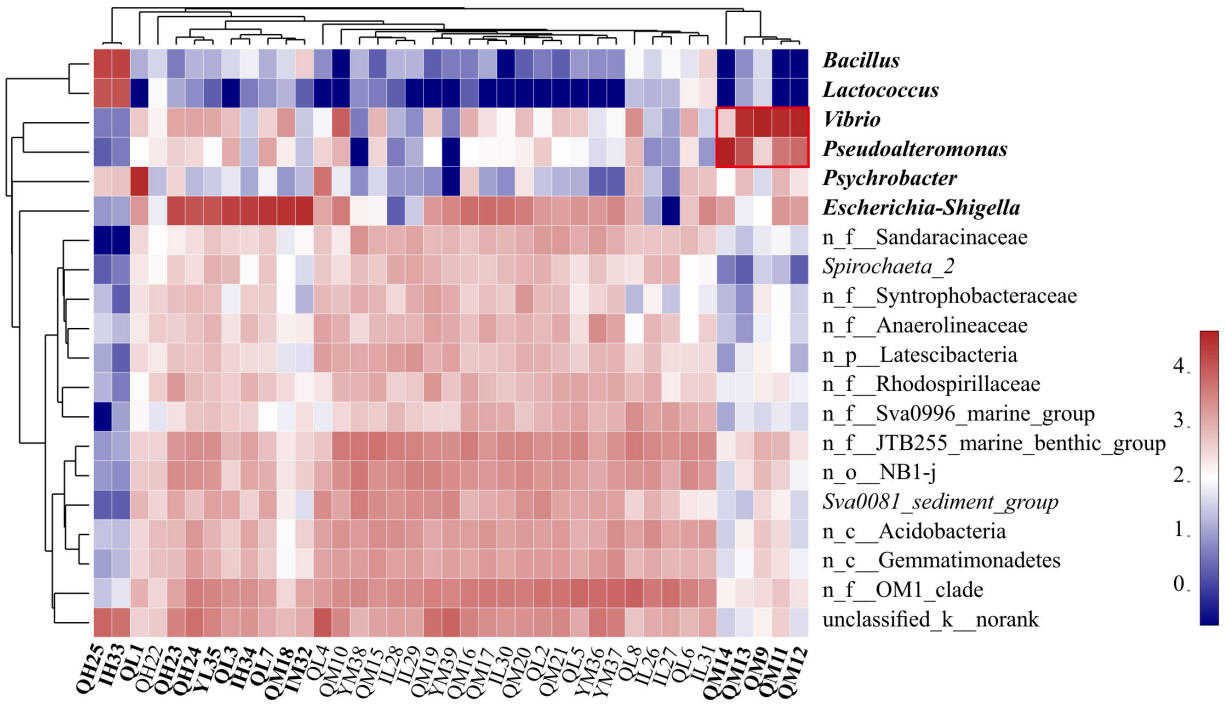


FIGURE 2 | Heatmap of the relative abundance of the top 20 genera of all sediment samples. Clusters based on microbial community composition are above the heatmap, clusters based on the genera are on the left of the heatmap, and special samples and bacteria are shown in bold. n_ indicates no rank. The data were scaled per column. The values represent the relative abundance of genera in the sample.

modules 2, 6, and 7 were positively correlated with TOC. Modules 6 and 7 were positively correlated with Mz.

Polluted areas usually show higher types of pollutant resistance and different community compositions, when compared with pristine areas (Sun et al., 2013; Quero et al., 2015; Gubelit et al., 2016). The presence of potential pathogenicity and pollution indicator bacteria, such as *Vibrio*, *Pseudoalteromonas*, *Escherichia-Shigella, Psychrobacter*, and *Pseudomonas,* along with strong hydrodynamic conditions of the current migrating westward, makes the ecological risks more serious. Different modules reflect the habitat heterogeneity, which is affected by different environmental selection mechanisms (Thompson, 2005; Olesen et al., 2007). In addition, modules 3 and 7 (*Vibrio*, *Pseudoalteromonas*, *Lactococcus, Bacillus,* and *Pseudomonas*) showed negative correlation with modules 1 and 5 (dominant microbial functional groups), indicating that the increase in pathogenic and pollution indicator bacteria may decrease the diversity of pristine bacteria in the SCS. Thus, sediment microbial communities of the SCS disturbed by human activities could present altered interactions of bacteria and correlation between functional groups, which might affect the bacterial element cycle and ecological function in sediment.

Spatial Distribution of Microbial Community in Xuande Atoll Sediments

There were no significant differences in the richness, diversity, and evenness of microbial communities among the three islands (Wilcoxon test, $p > 0.05$, **Supplementary Table S2**). Previous studies have shown that geographical location has a strong impact on the composition of microbial communities, and the β diversity distance attenuation model of bacterial community composition depends on the spatial scale (Martiny et al., 2011; Zinger et al., 2011). A study of the SCS showed that control of the environment and diffusion limiting microorganisms will result in a distribution pattern of biogeography (Zhang et al., 2014). Through ANOSIM analysis of bacterial community composition between the three islands at the OTU level (**Figure 4A**), the within-group difference of Qilian Island was greater than the difference between the islands, while that of Yongxing Island was the lowest. For Iltis Bank, the within-group difference was relatively large and slightly lower than the difference between the groups. The results of NMDS analysis also showed that the sediment samples around Qilian Island were most dispersed, indicating that the microbial community structure near Qilian Island was most diverse (**Figure 4B**). The sample points at Yongxing Island were closest together. The sample points at Iltis Bank were also scattered.

Subsequently, the microbial community structures of three islands were compared at the genus level (**Figure 4C**). The abundance of *Vibrio*, *Pseudoalteromonas,* and *Psychrobacter* at the three islands was significantly different, and the abundance of these three genera in Qilian Island was higher than that in Yongxing Island and Iltis Bank. In Qilian Island, *Vibrio* was abundant in sample QM9 (84.6%), QM11 (73.3%), QM12 (80.3%), and QM13 (66.8%); *Pseudoalteromonas* was abundant in

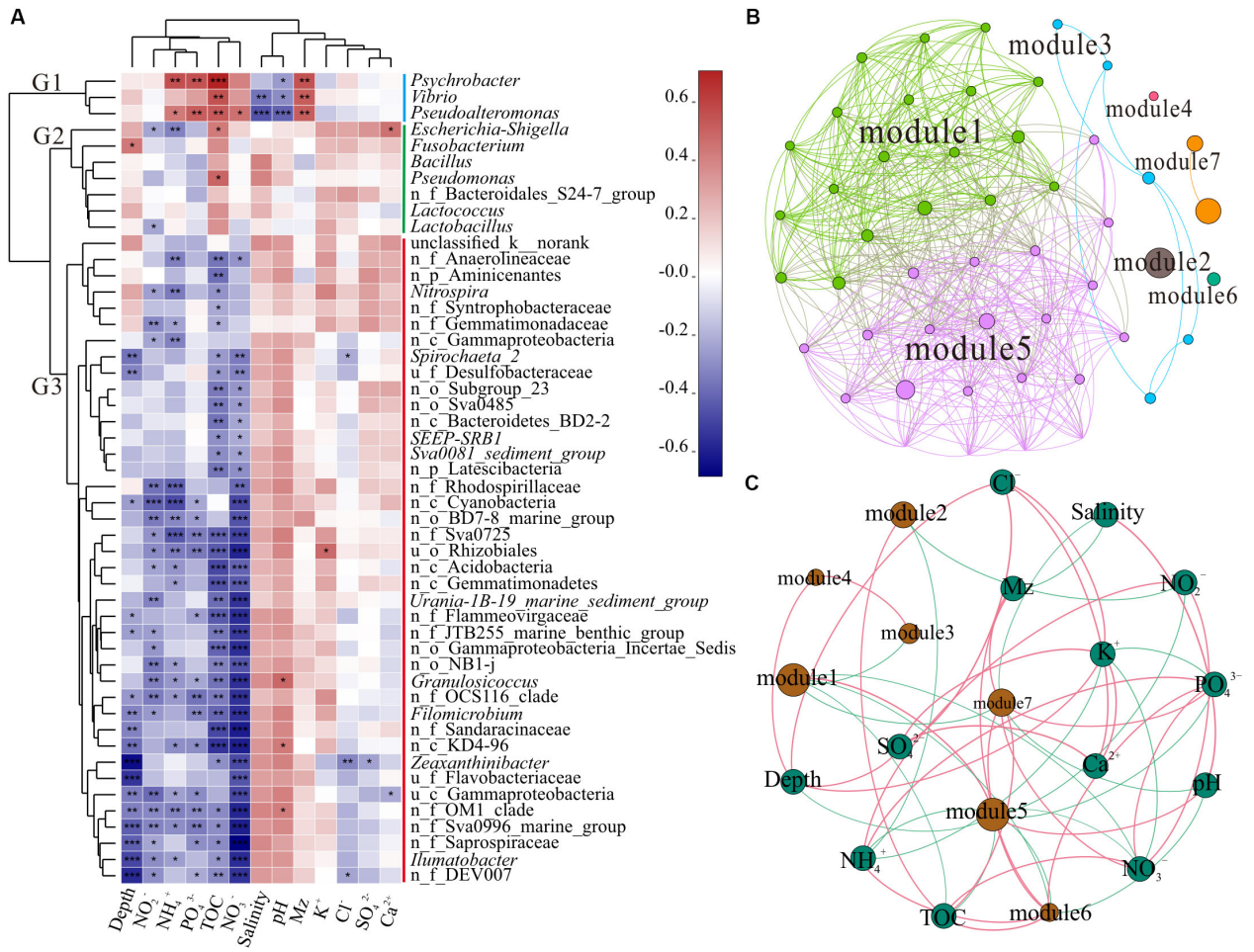


FIGURE 3 | Correlation analysis. **(A)** Correlation heatmap between the top 50 genera and environmental factors. Clusters based on environmental factors are above the heatmap and clusters of genera are on the left of the heatmap; n_ and u_ indicate no rank and unclassified, respectively. The asterisks in the cell represent significant correlation, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. The values denote correlation coefficient. **(B)** Network of the top 50 genera. Top 50 abundant genera in all 39 samples were selected to calculate the correlation among bacteria. Each node represents a genus, the nodes are colored by module, and the connection between the nodes represents a significant correlation (Spearman, $p < 0.05$, $r > 0.6$). **(C)** Network of the correlation between modules and environmental factors (Spearman, $p < 0.05$, $r > 0.3$) across all 39 samples. The sum of all bacterial abundances in the module represents the richness of the module. Brown nodes represent the modules of the microbiome, and the sizes of the brown nodes indicate the relative abundance of the bacteria. Green nodes represent environmental factors. Red edge represents positive correlation and green edge denotes negative correlation.

sample QM11 (9.1%), QM12 (12.5%), QM13 (25.8%), and QM14 (93.3%); and *Psychrobacter* was abundant in sample QL1 (74.3%). These findings suggested that the differences in community structure were mainly between unique sites (QM1, QM9, QM11, QM12, QM13, QM14) and the genera *Psychrobacter*, *Vibrio*, and *Pseudoalteromonas*.

Effects of Different Depths on Microbial Community Structure in Sediments

The depth of water significantly affects the community structure of microorganisms in coastal areas (Hewson et al., 2007; Boer et al., 2009). For instance, sediment depth related patterns in bacterial community structure were detected in a variety of benthic habitats, such as cold seep sediments (Inagaki et al., 2002), warm deep Mediterranean sea (Luna et al., 2004), coral reef sediments (Hewson and Fuhrman, 2006), as well as continental shelf sediments of the southern North Sea (Franco et al., 2007). Environmental parameters, such as wave impacts (Hewson and Fuhrman, 2006), organic carbon and chlorophyll *a* contents in the sediment (Polymenakou et al., 2005), eutrophication associated with fish farms (Vezzulli et al., 2002), and inorganic nutrients enrichment (Hewson et al., 2003), have been found to be related to shifts in the benthic bacterial community structure. In this study, the differences in Wilcoxon rank-sum test of microbial diversity indices in the sediments from different water depths showed that the Shannon indices of L (15–50 m) and H (70–100 m) groups were significantly different (Wilcoxon test, $p < 0.05$, **Supplementary Table S2**). In addition, Chao index of M (50–70 m) and

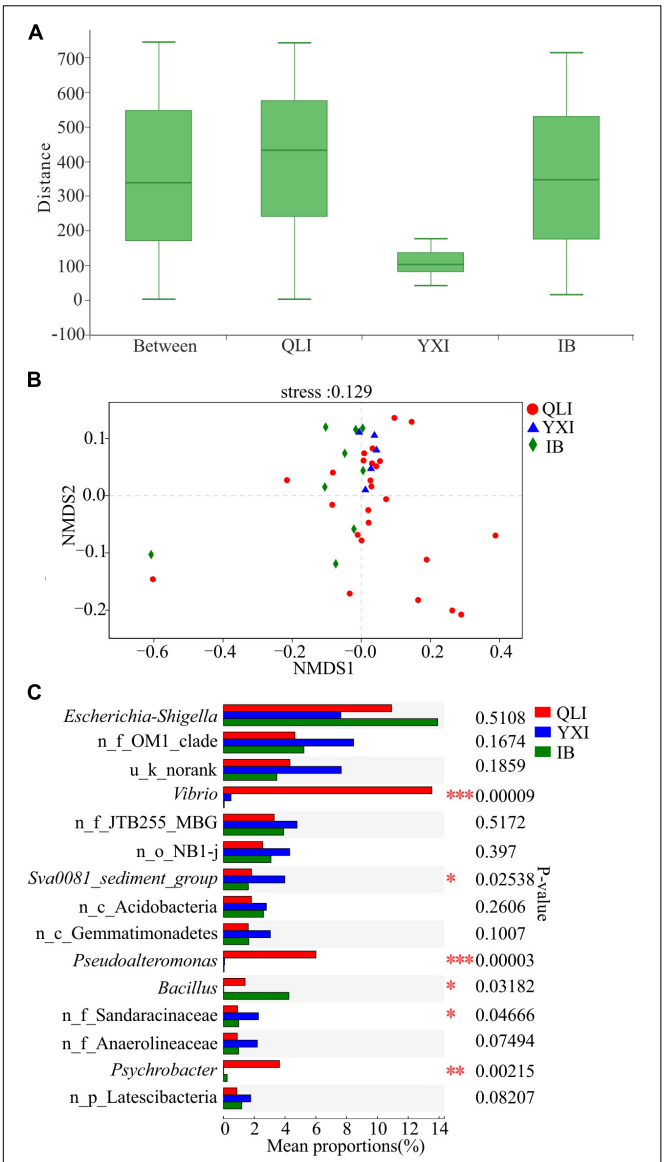


FIGURE 4 | Difference in microbial communities among three islands. QLI, YXI, and IB represent Qilian Island, Yongxing Island, and Iltis Bank, respectively. **(A)** Similarity within and between islands. ANOSIM analysis based on Bray-Curtis distance of bacterial community structure at the OTU level ($p = 0.775$, $r = -0.0785$). **(B)** NMDS analysis based on Bray-Curtis distance of bacterial community structure at the OTU level among all 39 sediment samples. **(C)** Differences in bacterial abundance among three islands based on the Kruskal-Wallis test, and the p-value was checked using FDR. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. n_ and u_ indicate no rank and unclassified, respectively, MBG represents marine_benthic_group.

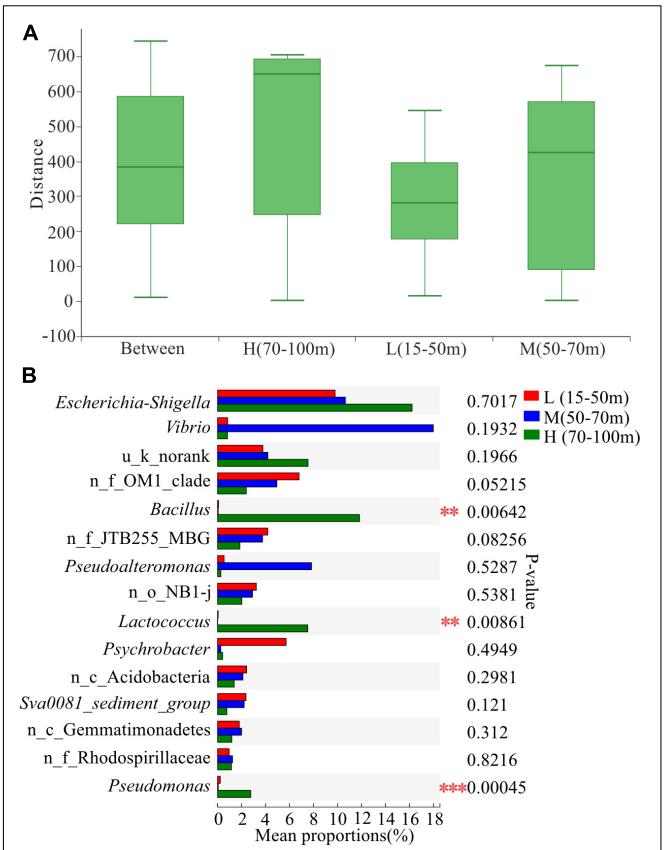


FIGURE 5 | Variation in microbial communities among three different depths. L, M, and H represent samples from depths of 15–50, 50–70, and 70–100 m, respectively. **(A)** Similarity analysis within and between depths. ANOSIM analysis based on Bray-Curtis distance of bacterial community structure at the OTU level ($p = 0.004$, $r = 0.1911$). **(B)** Differences in bacterial abundance among three depths based on the Kruskal-Wallis test and the p-value was checked using FDR. $^{**}p < 0.01$; $^{***}p < 0.001$. n_ and u_ indicate no rank and unclassified, respectively; MBG represents marine_benthic_group.

H (70–100 m) groups were significantly different (Wilcoxon test, $p < 0.05$, **Supplementary Table S2**). These findings indicated a decreased microbial diversity along the vertical depth in the ocean.

The results of ANOSIM at the OTU level (**Figure 5A**) showed that the within-group difference of water depth L (15–50 m) was less than the difference between the groups, indicating that the microbial community structure in the water depth L (15–50 m) was similar. In addition, the sample points of water depth M (50–70 m) were scattered, and the within-group difference was slightly larger than the difference between the groups. Moreover, the within-group difference of water depth H (70–100 m) was significantly larger than the difference between groups. The microbial community structure at water depth H (70–100 m) presented the most significant difference. The proportion of *Vibrio* at water depth M (50–70 m) was the highest, reaching 18.54% (**Figure 5B**). In addition, the proportion of *Bacillus* differed significantly among the three water depths, with the highest level of 12.20% being observed at water depth H (70–100 m). Furthermore, the proportion of *Pseudoalteromonas* at water depth M (50–70 m) was 8.07%. Overall, the proportion of pathogenic and pollutant metabolism associated bacteria at the depth of 50–100 m was significantly higher than that in the shallow water layer of 15–50 m across all the 39 samples, indicating that the study area had been

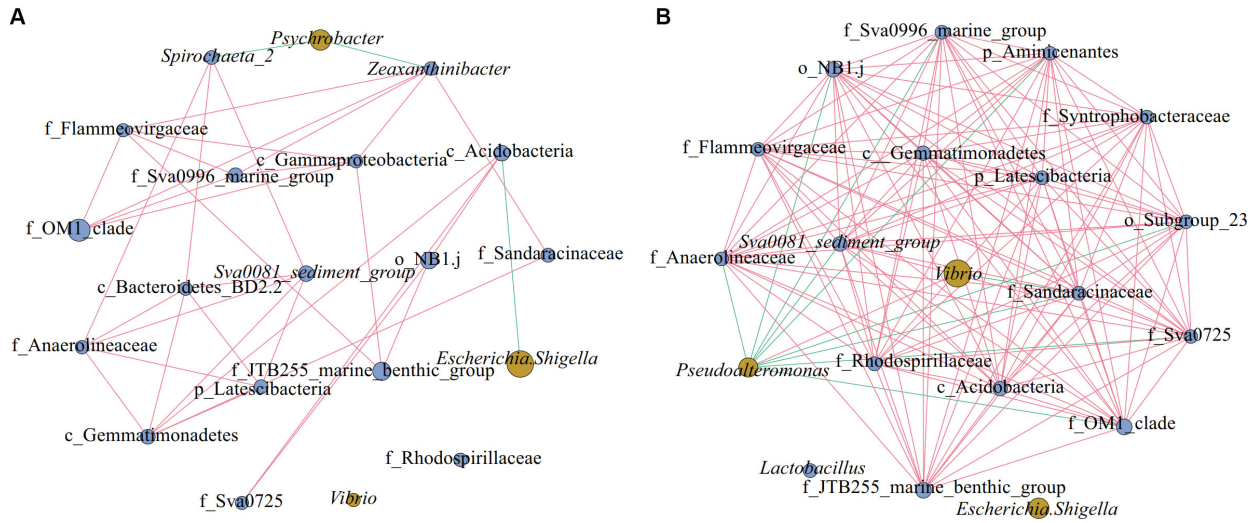


FIGURE 6 | Networks of microorganisms in sediment of different water depths. Co-occurrence of the top 20 genera in sediment of different water depths: **(A)** Depth L (15–50 m), **(B)** Depth M (50–70 m). Network diagrams were constructed using 15 (L) and 15 (M) samples; in the group M (50–70 m), we discarded the samples QM17, QM20, and SM36. Blue nodes represent indigenous bacteria and yellow nodes denote pathogenic and pollutants-degrading bacteria. The size of the nodes represents the relative abundance of the bacteria. The edge between the points represents a significant correlation, with the red edge signifying positive correlation and the green edge denoting a negative correlation (Spearman, $p < 0.01$, $r > 0.6$).

contaminated and the deep water layer is more severely polluted than the shallow water layer.

As shown in **Figure 6**, there were 20 nodes in each of the two network diagrams (A and B; top 20 genera, $p < 0.01$, $r > 0.6$), and there were 35 and 131 edges in the sediment of water depths L (15–50 m) and M (50–70 m), respectively. In addition, there were 32 and 121 edges with positive correlations and 3 and 10 edges with negative correlations, respectively. Moreover, the number of connections in the network increased from the sediment of water depth L (15–50 m) to M (50–70 m), and the proportion of positive correlation edges at the sediment of water depth L (15–50 m) and M (15–50 m) were 91.4 and 92.4%, respectively. Many positive interactions have been found to occur in natural bacterial populations (Morris et al., 2012; Hallam and McCutcheon, 2015). The network represents coordinated variability, where a common change in abundance of members reflects interactions between members or responds to environmental factors (Shi et al., 2016). Studies on macrobiology and microbiology have shown that resources and food supply are important drivers of ecological network structure (Henzi et al., 2009; Foster et al., 2012). In the sediment of water depth L (15–50 m), *Psychrobacter* was negatively correlated with *Spirochaeta_2* and *Zeaxanthinibacter*, and *Escherichia-Shigella* was negatively correlated with c_Acidobacteria (**Figure 6A**). In the sediment of water depth M (50–70 m), *Pseudoalteromonas* was negatively correlated with other bacteria except *Vibrio*, and *Vibrio* was negatively correlated with f_Sandaracinaceae (**Figure 6B**). Besides, another network diagram was constructed based on 39 samples (**Supplementary Figure S4**) using Spearman correlation with $p < 0.01$ and $r > 0.5$, and it was found that *Psychrobacter* was negatively correlated with f_Sandaracinaceae and f_Anaerolineaceae, and *Lactococcus* was negatively correlated with *Sva0081_sediment_group*, but positively correlated with *Bacillus*. There was a significant positive correlation among all indigenous bacteria. *Psychrobacter* and *Lactococcus* are pollutants-degrading bacteria, suggesting that the indigenous bacterial group may have been affected. Therefore, we speculate that, in our study area, the relationship between microbes shows negative correlation between pathogenic and pollutants-degrading bacteria and indigenous bacteria, and that this phenomenon may be affected by pollution and competition among different ecological groups. When the seawater is polluted, the ecological group may change more severely, and the competition between the pathogen and indigenous group would make recovery of the ecological environment difficult.

CONCLUSION

In the Qilian Island area, high abundance of potential pathogenic bacteria, *Escherichia-Shigella*, *Lactococcus*, and *Pseudoalteromonas,* and pollutants-degrading bacteria, *Bacillus* and *Vibrio,* revealed that this study area is likely to have been contaminated, and that the pollution is mainly affected by geographical location and hydrodynamic conditions, with a possible tendency to spread. The existence of unique genera in the study area is the main reason for the difference in community structure between islands. The pathogenic bacterial group was negatively correlated with most of the indigenous bacteria in the ocean, suggesting that the indigenous bacterial group might have been influenced, and that the complexity of this effect is reflected in how different groups of microbes respond to different environmental factors, the pattern of interaction between

ecological groups, and mechanisms of microbial community formation at different depths. Because of the increased human activity in the area, the impact of anthropogenic activities on microbial community structure and the potential changes in the role of microorganisms in the cycle of elements require further research.

DATA AVAILABILITY STATEMENT

The accession number for sequencing data in NCBI is SRP222491.

AUTHOR CONTRIBUTIONS

BZ did the writing and data analysis of this research. YY and S-ZX participated in the experiment of samples used in this study. YL, RY, and M-PL participated in the data analysis and discussion of this research. X-MW and Y-LX did the collection of the samples for this study. X-YG participated in the topic and discussion of the whole study.

FUNDING

This research was supported by the National Natural Science Foundation of China (41731282; granted to X-YG) and China Geological Survey (DD20190323).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01011/full#supplementary-material

FIGURE S1 | Particle size distribution curve of sediment samples.

FIGURE S2 | Proportion of bacteria at the phylum level. u_ indicates unclassified.

FIGURE S3 | CCA analysis of microbial and environmental factors.

FIGURE S4 | Network of microorganisms based on 39 samples.

TABLE S1 | Environmental variables of all sediment samples.

TABLE S2 | Biodiversity indices of bacterial communities.

TABLE S3 | Genera in different modules from network analysis.

REFERENCES

Anthony, E. J., and Héquette, A. (2007). The grain-size characterisation of coastal sand from the somme estuary to belgium: sediment sorting processes and mixing in a tide- and storm-dominated setting. *Sediment. Geol.* 202, 369–382.

Bienhold, C., Zinger, L., Boetius, A., and Ramette, A. (2016). Diversity and biogeography of bathyal and abyssal seafloor bacteria. *PLoS ONE* 11:e0148016. doi: 10.1371/journal.pone.0148016

Boer, S. I., Hedtkamp, S. I., van Beusekom, J. E., Fuhrman, J. A., Boetius, A., and Ramette, A. (2009). Time- and sediment depth-related variations in bacterial diversity and community structure in subtidal sands. *ISME J.* 3, 780–791. doi: 10.1038/ismej.2009.29

Bowman, J. P. (2007). Bioactive compound synthetic capacity and ecological significance of marine bacterial genus pseudoalteromonas. *Mar Drugs* 5, 220–241. doi: 10.3390/md504220

Brunette, G. W., Kozarsky, P. E., and Cohen, N. J. (2015). *CDC Health Information for International Travel 2016: The Yellow Book*. Oxford: Oxford University Press.

Deng, Y., Jiang, Y. H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinform.* 13:113. doi: 10.1186/1471-2105-13-113

Du, J., Xiao, K., Huang, Y., Li, H., Tan, H., Cao, L., et al. (2011). Seasonal and spatial diversity of microbial communities in marine sediments of the South China Sea. *Antonie Van Leeuwenhoek* 100, 317–331. doi: 10.1007/s10482-011-9587-9

Duran-Pinedo, A. E., Paster, B., Teles, R., and Frias-Lopez, J. (2011). Correlation network analysis applied to complex biofilm communities. *PLoS ONE* 6:e28438. doi: 10.1371/journal.pone.0028438

Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi: 10.1038/nrmicro2832

Foster, E. A., Franks, D. W., Morrell, L. J., Balcomb, K. C., Parsons, K. M., van Ginneken, A., et al. (2012). Social network correlates of food availability in an endangered population of killer whales, *Orcinus orca*. *Anim. Behav.* 83, 731–736.

Franco, M. A., De Mesel, I., Demba Diallo, M., Van Der Gucht, K., Van Gansbeke, K., Van Rijswijk, P., et al. (2007). Effect of phytoplankton bloom deposition on benthic bacterial communities in two contrasting sediments in the southern North Sea. *Aquat. Microb. Ecol.* 48, 241–254. doi: 10.1016/j.scitotenv.2011.12.034

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature* 459, 193–199. doi: 10.1186/1475-2859-13-S1-S2

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., et al. (2009). The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* 11, 3132–3139. doi: 10.1111/j.1462-2920.2009.02017.x

Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbruck, L., Reeder, J., Temperton, B., et al. (2012). Defining seasonal marine microbial community dynamics. *ISME J.* 6, 298–308. doi: 10.1038/ismej.2011.107

Gillan, D. C., Danis, B., Pernet, P., Joly, G., and Dubois, P. (2005). Structure of sediment-associated microbial communities along a heavy-metal contamination gradient in the marine environment. *Appl. Environ. Microbiol.* 71, 679–690. doi: 10.1128/AEM.71.2.679-690.2005

Glasl, B., Bourne, D. G., Frade, P. R., Thomas, T., Schaffelke, B., and Webster, N. S. (2019). Microbial indicators of environmental perturbations in coral reef ecosystems. *Microbiome* 7:94. doi: 10.1186/s40168-019-0705-7

Graham, D. E., Wallenstein, M. D., Vishnivetskaya, T. A., Waldrop, M. P., Phelps, T. J., Pfiffner, S. M., et al. (2012). Microbes in thawing permafrost: the unknown variable in the climate change equation. *ISME J.* 6, 709–712.

Guan, X., Zhu, L., Li, Y., Xie, Y., Zhao, M., and Luo, X. (2014). Composition and variation of sediment bacterial and nirS-harboring bacterial communities at representative sites of the Bohai Gulf coastal zone, China. *World J. Microbiol. Biotechnol.* 30, 1291–1300. doi: 10.1007/s11274-013-1553-4

Gubelit, Y., Polyak, Y., Dembska, G., Pazikowska-Sapota, G., Zegarowski, L., Kochura, D., et al. (2016). Nutrient and metal pollution of the eastern gulf of finland coastline: sediments, macroalgae, microbiota. *Sci. Total Environ.* 550, 806–819. doi: 10.1016/j.scitotenv.2016.01.122

Hallam, S. J., and McCutcheon, J. P. (2015). Microbes don't play solitaire: how cooperation trumps isolation in the microbial world. *Environ. Microbiol. Rep.* 7, 26–28. doi: 10.1111/1758-2229.12248

Hamdan, L. J., Coffin, R. B., Sikaroodi, M., Greinert, J., Treude, T., and Gillevet, P. M. (2013). Ocean currents shape the microbiome of Arctic marine sediments. *ISME J.* 7, 685–696. doi: 10.1038/ismej.2012.143

Hamdan, L. J., and Fulmer, P. A. (2011). Effects of COREXIT EC9500A® on bacteria from a beach oiled by the deepwater horizon spill. *Aquat. Microb. Ecol.* 63, 101–109.

Henzi, S., Lusseau, D., Weingrill, T., Van Schaik, C., and Barrett, L. (2009). Cyclicity in the structure of female baboon social networks. *Behav. Ecol. Sociobiol.* 63, 1015–1021.

Hewson, I., and Fuhrman, J. A. (2006). Spatial and vertical biogeography of coral reef sediment bacterial and diazotroph communities. *Mar. Ecol. Prog. Ser.* 306, 79–86.

Hewson, I., Jacobson Meyers, M. E., and Fuhrman, J. A. (2007). Diversity and biogeography of bacterial assemblages in surface sediments across the San Pedro Basin, Southern California Borderlands. *Environ. Microbiol.* 9, 923–933. doi: 10.1111/j.1462-2920.2006.01214.x

Hewson, I., Vargo, G. A., and Fuhrman, J. A. (2003). Bacterial diversity in shallow oligotrophic marine benthos and overlying waters: effects of virus infection, containment, and nutrient enrichment. *Microb. Ecol.* 46, 322–336. doi: 10.1007/s00248-002-1067-3

Hollister, E. B., Engledow, A. S., Hammett, A. J., Provin, T. L., Wilkinson, H. H., and Gentry, T. J. (2010). Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *ISME J.* 4, 829–838. doi: 10.1038/ismej.2010.3

Holmstrom, C., and Kjelleberg, S. (1999). Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. *FEMS Microbiol. Ecol.* 30, 285–293. doi: 10.1111/j.1574-6941.1999.tb00656.x

Hu, L., Shi, X., Yu, Z., Lin, T., Wang, H., Ma, D., et al. (2012). Distribution of sedimentary organic matter in estuarine–inner shelf regions of the east china sea: implications for hydrodynamic forces and anthropogenic impact. *Mar. Chem.* 14, 29–40. doi: 10.1016/j.scitotenv.2018.06.383

Inagaki, F., Sakihama, Y., Inoue, A., Kato, C., and Horikoshi, K. (2002). Molecular phylogenetic analyses of reverse-transcribed bacterial rRNA obtained from deep-sea cold seep sediments. *Environ. Microbiol.* 4, 277–286. doi: 10.1046/j.1462-2920.2002.00294.x

Inagaki, F., Suzuki, M., Takai, K., Oida, H., Sakamoto, T., Aoki, K., et al. (2003). Microbial communities associated with geological horizons in coastal subseafloor sediments from the sea of okhotsk. *Appl. Environ. Microbiol.* 69, 7224–7235. doi: 10.1128/aem.69.12.7224-7235.2003

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., and D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16213–16216. doi: 10.1073/pnas.1203849109

Kirchman, D. L., Cottrell, M. T., and Lovejoy, C. (2010). The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ. Microbiol.* 12, 1132–1143. doi: 10.1111/j.1462-2920.2010.02154.x

Kochling, T., Lara-Martin, P., Gonzalez-Mazo, E., Amils, R., and Sanz, J. L. (2011). Microbial community composition of anoxic marine sediments in the Bay of Cadiz (Spain). *Int. Microbiol.* 14, 143–154. doi: 10.2436/20.1501.01.143

Li, T., and Wang, P. (2014). Bacterial diversity in sediments of core MD05-2902 from the Xisha Trough, the South China Sea. *Acta Oceanol. Sin.* 33, 85–93.

Lien, R.-C., Tang, T. Y., Chang, M. H., and D'Asaro, E. A. (2005). Energy of nonlinear internal waves in the South China Sea. *Geophys. Res. Lett.* 32.

Limei Guo, J. S., Ming, H., and Ji, F. (2017). Study on hydrocarbon degradation and change of bacterial abundance 5 years after dalian "7·16" oil spill. *Ocean Bull.* 36, 311–319.

Liu, J., Cai, S., and Wang, S. A. (2010). Currents and mixing in the northern South China Sea. *Chin. J. Oceanol. Limnol.* 28, 974–980.

Liu, Z., and Gan, J. (2017). Three-dimensional pathways of water masses in the s outh c hina s ea: a modeling study. *J. Geophys. Res. Oceans* 122, 6039–6054.

Liu, Z., and Liu, J. (2013). Evaluating bacterial community structures in oil collected from the sea surface and sediment in the northern Gulf of Mexico after the Deepwater Horizon oil spill. *Microbiologyopen* 2, 492–504. doi: 10.1002/mbo3.89

Lu, X. M., Chen, C., Zheng, T. L., and Chen, J. J. (2016). Temporal-spatial variation of bacterial diversity in estuary sediments in the south of Zhejiang Province, China. *Appl. Microbiol. Biotechnol.* 100, 2817–2828. doi: 10.1007/s00253-015-7103-2

Lufei Yang, L. C., and Li, X. (2019). Ecological effects of seasonal low oxygen on macrobenthos community in muping Marine pasture, yantai. *Biodiversity* 27, 200–210.

Luna, G. M., Dell'Anno, A., Giuliano, L., and Danovaro, R. (2004). Bacterial diversity in deep Mediterranean sediments: relationship with the active bacterial fraction and substrate availability. *Environ. Microbiol.* 6, 745–753. doi: 10.1111/j.1462-2920.2004.00611.x

Martiny, J. B., Eisen, J. A., Penn, K., Allison, S. D., and Horner-Devine, M. C. (2011). Drivers of bacterial beta-diversity depend on spatial scale. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7850–7854. doi: 10.1073/pnas.1016308108

McManus, J. (1988). "Grain size determination and interpretation," in *Techniques in Sedimentology*, 63–85.

Miller, S. D., Haddock, S. H., Elvidge, C. D., and Lee, T. F. (2005). Detection of a bioluminescent milky sea from space. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14181–14184. doi: 10.1073/pnas.0507253102

Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio* 3, 203–216. doi: 10.1128/mBio.00036-12

Newberry, C. J., Webster, G., Cragg, B. A., Parkes, R. J., Weightman, A. J., and Fry, J. C. (2004). Diversity of prokaryotes and methanogenesis in deep subsurface sediments from the Nankai Trough, Ocean Drilling Program Leg 190. *Environ. Microbiol.* 6, 274–287. doi: 10.1111/j.1462-2920.2004.00568.x

Nogales, B., Lanfranconi, M. P., Pina-Villalonga, J. M., and Bosch, R. (2011). Anthropogenic perturbations in marine microbial communities. *FEMS Microbiol. Rev.* 35, 275–298. doi: 10.1111/j.1574-6976.2010.00248.x

Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. (2007). The modularity of pollination networks. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19891–19896.

Parkes, R. J., Cragg, B., Roussel, E., Webster, G., Weightman, A., and Sass, H. (2014). A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: geosphere interactions. *Mar. Geol.* 352, 409–425.

Parkes, R. J., Cragg, B. A., and Wellsbury, P. (2000). Recent studies on bacterial populations and processes in subseafloor sediments: a review. *Hydrogeol. J.* 8, 11–28.

Passega, R. (1957). Texture as characteristic of clastic deposition. *AAPG Bulletin* 41, 1952–1984.

Passega, R. (1964). Grain size representation by CM patterns as a geologic tool. *J. Sedimen. Res.* 34, 830–847.

Polymenakou, P. N., Bertilsson, S., Tselepides, A., and Stephanou, E. G. (2005). Links between geographic location, environmental factors, and microbial community composition in sediments of the Eastern Mediterranean Sea. *Microb. Ecol.* 49, 367–378. doi: 10.1007/s00248-004-0274-5

Prabagaran, S. R., Manorama, R., Delille, D., and Shivaji, S. (2007). Predominance of roseobacter, sulfitobacter, glaciecola and psychrobacter in seawater collected off Ushuaia, Argentina, Sub-Antarctica. *FEMS Microbiol. Ecol.* 59, 342–355. doi: 10.1111/j.1574-6941.2006.00213.x

Quero, G. M., Cassin, D., Botter, M., Perini, L., and Luna, G. M. (2015). Patterns of benthic bacterial diversity in coastal areas contaminated by heavy metals, polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs). *Front. Microbiol.* 6:1053. doi: 10.3389/fmicb.2015.01053

Rochelle, P. A., Cragg, B. A., Fry, J. C., Parkes, R. J., and Weightman, A. J. (1994). Effect of sample handling on estimation of bacterial diversity in marine sediments by 16S rRNA gene sequence analysis. *FEMS Microbiol. Ecol.* 15, 215–225.

Ruger, H. J., Fritze, D., and Sproer, C. (2000). New psychrophilic and psychrotolerant *Bacillus marinus* strains from tropical and polar deep-sea sediments and emended description of the species. *Int. J. Syst. Evol. Microbiol.* 50(Pt 3), 1305–1313. doi: 10.1099/00207713-50-3-1305

Shi, S., Nuccio, E. E., Shi, Z. J., He, Z., Zhou, J., and Firestone, M. K. (2016). The interconnected rhizosphere: high network complexity dominates rhizosphere assemblages. *Ecol. Lett.* 19, 926–936. doi: 10.1111/ele.12630

Shuqing, Q., Xuefa, S., Guoqing, W., Gang, Y., Ningjing, H., Shengfa, L., et al. (2010). Discussion on grain size characteristics and transport trend of sediment deposits in the bohai sea. *Acta Oceanol. Sin.* 32, 139–147.

Su, Z., Dai, T., Tang, Y., Tao, Y., Huang, B., Mu, Q., et al. (2018). Sediment bacterial community structures and their predicted functions implied the impacts from natural processes and anthropogenic activities in coastal area. *Mar. Pollut. Bull.* 131, 481–495. doi: 10.1016/j.marpolbul.2018.04.052

Sun, M. Y., Dafforn, K. A., Johnston, E. L., and Brown, M. V. (2013). Core sediment bacteria drive community response to anthropogenic contamination over multiple environmental gradients. *Environ. Microbiol.* 15, 2517–2531. doi: 10.1111/1462-2920.12133

Tanekhy, M. (2013). Some study on bacterial infection in some cultured marine fish. *J. Arab. Aquacult. Soc.* 8, 163–178.

Thompson, J. N. (2005). Coevolution: the geographic mosaic of coevolutionary arms races. *Curr. Biol.* 15, R992–R994. doi: 10.1016/j.cub.2005.11.046

Vezzulli, L., Chelossi, E., Riccardi, G., and Fabiano, M. (2002). Bacterial community structure and activity in fish farm sediments of the Ligurian sea (Western Mediterranean). *Aquacult. Int.* 10, 123–141.

Walsh, E. A., Kirkpatrick, J. B., Rutherford, S. D., Smith, D. C., Sogin, M., and D'Hondt, S. (2016). Bacterial diversity and community composition from seasurface to subseafloor. *ISME J.* 10, 979–989. doi: 10.1038/ismej.2015.175

Wang, L., Liu, L., Zheng, B., Zhu, Y., and Wang, X. (2013). Analysis of the bacterial community in the two typical intertidal sediments of Bohai Bay, China by pyrosequencing. *Mar. Pollut. Bull.* 72, 181–187. doi: 10.1016/j.marpolbul.2013.04.005

Webster, G., Parkes, R. J., Cragg, B. A., Newberry, C. J., Weightman, A. J., and Fry, J. C. (2006). Prokaryotic community composition and biogeochemical processes in deep subseafloor sediments from the Peru Margin. *FEMS Microbiol. Ecol.* 58, 65–85. doi: 10.1111/j.1574-6941.2006.00144.x

Xiaodong Deng, J. L., and Cai, S. (2013). Analysis on the characteristics of tidal current in the continental shelf of xisha islands, South China sea. *J. Trop. Oceanogr.* 32, 8–12.

Xiaohua, Z., Heyu, L., Xiaolei, W., and Austin, B. (2018). The important role of *Vibrio* in marine organic carbon cycle. *Sci. Sin. Terrae* 48, 5–17.

Xiong, J., Ye, X., Wang, K., Chen, H., Hu, C., Zhu, J., et al. (2014). Biogeography of the sediment bacterial community responds to a nitrogen pollution gradient in the East China Sea. *Appl. Environ. Microbiol.* 80, 1919–1925. doi: 10.1128/AEM.03731-13

Yu, T., Li, M., Niu, M., Fan, X., Liang, W., et al. (2018). Difference of nitrogen-cycling microbes between shallow bay and deep-sea sediments in the South China Sea. Applied microbiology and biotechnology. *Microbiol. Biotechnol.* 102, 447–459. doi: 10.1007/s00253-017-8594-9

Yuan, X. (2015). *Isolation and Diversity Analysis of Petroleum Degrading Bacteria in Coastal Zone*. Qingdao: Qingdao University of Technology.

Zhang, W., Ki, J.-S., and Qian, P.-Y. (2008). Microbial diversity in polluted harbor sediments I: bacterial community assessment based on four clone libraries of 16S rDNA. *Estuar. Coastal Shelf Sci.* 76, 668–681.

Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G. J. (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Mol. Ecol.* 23, 2260–2274. doi: 10.1111/mec.12739

Zhou, J., Deng, Y., Luo, F., He, Z., and Yang, Y. (2011). Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. *MBio* 2:e122-11. doi: 10.1128/mBio.00122-11

Zhou, M.-Y., Chen, X.-L., Zhao, H.-L., Dang, H.-Y., Luan, X.-W., Zhang, X.-Y., et al. (2009). Diversity of both the cultivable protease-producing bacteria and their extracellular proteases in the sediments of the South China Sea. *Microb. Ecol.* 58, 582–590. doi: 10.1007/s00248-009-9506-z

Zhu, D., Tanabe, S. H., Yang, C., Zhang, W., and Sun, J. (2013). Bacterial community composition of South China Sea sediments through pyrosequencing-based analysis of 16S rRNA genes. *PLoS ONE* 8:e78501. doi: 10.1371/journal.pone.0078501

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B., et al. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS ONE* 6:e24570. doi: 10.1371/journal.pone.0024570

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 30 June 2020
doi: 10.3389/fmars.2020.00472

Structure and Functional Diversity of Surface Bacterioplankton Communities in an Overwintering Habitat for Large Yellow Croaker, *Pseudosciaena crocea*, of the Southern East China Sea

*Wen Yang[1], Shi-Zhan Zheng[1], Shou-Heng Zhou[1], Li Zhao[1], Jin-Yong Zhu[1], Betina Lukwambe[1,2], Regan Nicholaus[1,3], Cheng-Hua Li[1] and Zhong-Ming Zheng[1]**

[1] School of Marine Sciences, Ningbo University, Ningbo, China, [2] Department of Food Science and Technology, University of Dar es Salaam, Dar es Salaam, Tanzania, [3] Department of Natural Sciences, Mbeya University of Science and Technology, Mbeya, Tanzania

OPEN ACCESS

Edited by:
Karoline Faust,
KU Leuven, Belgium

Reviewed by:
Bernd Wemheuer,
University of New South Wales, Australia
Stilianos Fodelianakis,
École Polytechnique Fédérale de Lausanne, Switzerland

***Correspondence:**
Zhong-Ming Zheng
zhengzhongming@nbu.edu.cn

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Marine Science

Received: 27 January 2020
Accepted: 27 May 2020
Published: 30 June 2020

Citation:
Yang W, Zheng S-Z, Zhou S-H, Zhao L, Zhu J-Y, Lukwambe B, Nicholaus R, Li C-H and Zheng Z-M (2020) Structure and Functional Diversity of Surface Bacterioplankton Communities in an Overwintering Habitat for Large Yellow Croaker, Pseudosciaena crocea, of the Southern East China Sea. Front. Mar. Sci. 7:472. doi: 10.3389/fmars.2020.00472

Given the biogeochemical functions of marine microbes, the structure and functional diversity of bacterioplankton communities could be regarded as bioindicators that reflect environmental quality. The habitat of the large yellow croaker (*Pseudosciaena crocea*) is suffering from widespread degradation, but the study of the croaker habitat is still insufficient. Here, we used 16S rRNA gene sequencing and functional prediction to investigate the surface bacterioplankton communities in an overwintering ground of *P. crocea* during the overwintering period (November 2018 to April 2019). Our results showed that the taxa and functions of the environmental bacterioplankton community exhibited obvious seasonality. Specifically, the bacterioplankton communities in autumn were characterized by animal parasites or symbionts, such as *Acinetobacter baumannii* and *Moraxella*, while organic substance–degrading bacteria, such as *Rhodococcus*, *Stenotrophomonas*, and *Alcanivorax*, dominated in spring. Moreover, phylogenetic-based mean nearest taxon distance (MNTD) analyses indicated that dispersal limitation was the most important process governing the spatiotemporal assembly of the bacterioplankton communities. The distance decay of similarity also suggested the impact of dispersal limitation on the generation of biogeographical patterns. Furthermore, variation partitioning analysis (VPA) and a partial Mantel test revealed that environmental filters, such as nutrients, temperature, and salinity, had relatively weak deterministic effects on the bacterioplankton community assembly. Overall, these findings provide a basis for understanding the theory of surface microbial community assembly in the overwintering habitat of the large yellow croaker.

Keywords: large yellow croaker, surface water, planktonic bacteria, community assembly, dispersal limitation, FAPROTAX

INTRODUCTION

Abundant fisheries resources are indispensable cornerstones for the sustainable development of the coastal economy (Tursi et al., 2015; Keen et al., 2018). However, as China's most productive fishing area, the East China Sea has been suffering from resource degradation in recent decades due to increasing human impacts, such as overexploitation, pollutant discharge, and climate change (Wang et al., 2012). The Chinese government has implemented a series of strict fishing moratoriums and stock management systems since 1995, and the current status of most natural resources has generally improved. Despite all these programs, the wild stocks of large yellow croaker (*Pseudosciaena crocea*), which is known as China's national fish, have not recovered (Liu and de Mitcheson, 2008). Evidence strongly suggests that coastal pollution, habitat degradation, and possible ecosystem changes may have compromised the potential for stock recovery (Liu and de Mitcheson, 2008). However, the habitat (including the spawning, nursery, and overwintering grounds and migration channels) of the large yellow croaker has not been sufficiently researched. According to inferred migratory routes, croaker aggregate in overwintering grounds from October to April each year (Liu and de Mitcheson, 2008; Xu and Chen, 2011; Chen and Xu, 2012). Because the low temperature may suppress croaker immune functions by influencing multiple physiological processes and thereby make the croaker susceptible to the surrounding environment (Yang et al., 2015; Liu et al., 2019), the habitat conditions of their overwintering grounds are vital for recovering their stock and supporting this marine resource.

Water microorganisms, especially planktonic bacteria, are one of the largest reserves of biodiversity and promote various ecological processes in marine ecosystems (Sunagawa et al., 2015). On the one hand, bacteria have long been used as sensitive and aggregative indicators of changes in environmental conditions due to their wide distribution, short generation time, and diverse functions (Dai et al., 2017; Yang et al., 2018). On the other hand, the bacterioplankton community includes a large number of probiotic organisms and pathogens that can directly affect fish health (Xiong et al., 2015; Hu et al., 2017). Effects of bacterioplankton on fish could also be indirect through the food chain (Kalcheva et al., 2008). For example, a microbial loop mediated by bacteria could affect the upper elements of the food chain of the croaker. Therefore, an accurate representation of the overwintering ground's current conditions can be systematically derived by observing the marine bacterioplankton community. Given the important roles of the bacterioplankton community in the assessment of the habitat environment, it is necessary to reveal the fundamental processes that underlie biogeographic and ecological patterns of bacterioplankton diversity. However, this field of study remains a major challenge in bacterial ecology (Jiao et al., 2019). Historically, bacterial community assembly has been studied from a deterministic perspective, for example, environmental filtering and biotic interactions (Dini-Andreote et al., 2015). However, an increasing number of recent studies support the idea that stochasticity plays a predominant role in some microbial systems (Zhou et al., 2013; Chen W. D. et al., 2019), especially in the biogeographical patterns of marine bacterioplankton systems (Giovannoni and Nemergut, 2014). In fact, community assembly is affected by deterministic and stochastic processes simultaneously, but the relative effects of these two processes vary among different ecosystems. Various methods have been developed to separate the relative effects of deterministic and stochastic processes. Among them, a phylogenetic framework that uses information about phylogenetic relatedness and phylogenetic dissimilarity derived from phylogenetic trees has been widely used (Stegen et al., 2013, 2015). Using this framework will help us to reveal the potential factors underlying the spatiotemporal patterns of bacterial communities in the overwintering grounds for croaker. In addition, the focus of bacterial ecology is turning from its initial focus on pattern description toward functional profiling. The Functional Annotation of Prokaryotic Taxa (FAPROTAX), which focuses on the biogeochemical cycle functions of microorganisms, is commonly used (Louca et al., 2016). This method depends on prokaryotic functional data from culturable bacteria and species classification (OTU table) to obtain community function information and provides opportunities to understand the differences in the structure and function of bacterioplankton communities (Yan et al., 2019).

In the present study, we investigated the surface bacterioplankton communities in the overwintering grounds of *P. crocea* during the overwintering period using high-throughput sequencing of 16S rRNA genes, aiming at (1) revealing the spatiotemporal patterns of taxonomic and functional profiles of bacterioplankton communities in the overwintering grounds and (2) understanding the processes that underlie the observed patterns in species abundances across space and time. These results are essential for understanding microbial community assembly theory in this marine ecosystem and adding to the data on the environmental microbiome in the surface water of croaker habitat.

MATERIALS AND METHODS

Study Site, Sample Collection, and Environmental Variables Analyses

Surface water samples (∼2 m) were collected in sterile containers from the overwintering ground for *P. crocea* (**Figure 1**), which is located in the East China Sea during 17 November–5 December 2018 (November) and 7–30 April 2019 (April). In total, we obtained 50 water samples (25 sites × 2 time points). The samples were stored in an icebox (4°C) and immediately transported to the refrigerator at −20°C. Environmental variables of water samples, including water temperature (WT), salinity (SAL), orthophosphate (PO_4), nitrite (NO_2), nitrate (NO_3), ammonium (NH_4), silicate (SiO_3), total nitrogen (TN), and total phosphorus (TP), were analyzed following the standard methods (AQSIQ, 2007) (**Supplementary Table S1**). Approximately 500 mL of a water sample from each site was filtered-sterilized onto one 0.2-μm pore-sized polycarbonate membrane (47 mm diameter, Millipore, Boston, MA, United States) on the sampling day. The membranes were immediately frozen at −80°C until they were needed.

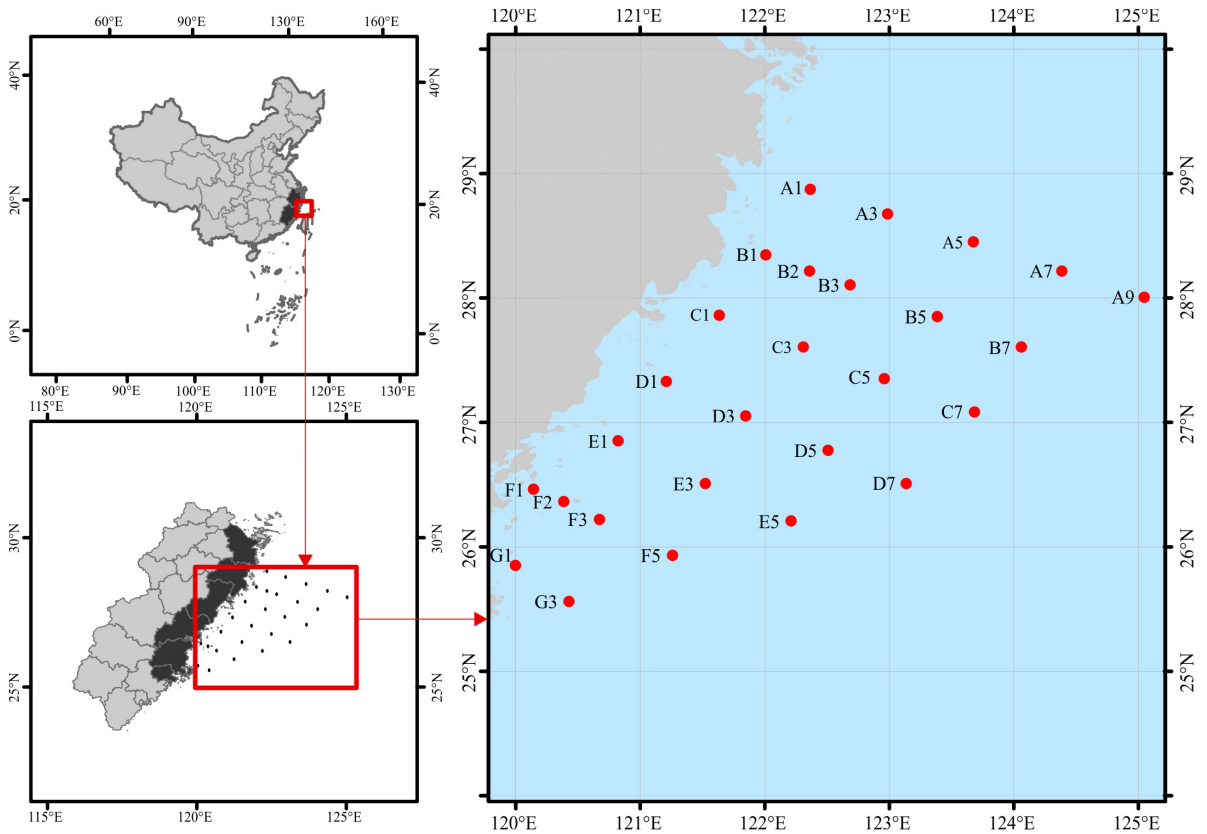


FIGURE 1 | Study area and location of sampling stations in the overwintering ground. To keep the data comparable with future studies, we keep the discontinuous numbers for site names according to the routine-monitoring project. The sampling sites remain the same for both seasons, except for November at B2 and F2 and April at B1 and F1.

DNA Extraction, Bacterial 16S Amplification, and Illumina HiSeq Sequencing

Bacterioplankton DNA was extracted using a DNA extraction kit (MinkaGene Water DNA Kit). A NanoDrop One spectrophotometer was used to measure the DNA concentration and purity (Thermo Fisher Scientific, MA, United States). The V4 regions of the bacterial 16S rRNA genes were amplified (50 μL reaction volume) by thermocycling: 5 min at 94°C for initialization; 30 cycles of 30 s denaturation at 94°C, 30 s annealing at 52°C, and 30 s extension at 72°C, followed by 10 min final elongation at 72°C using the primer sets 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNNGGGTATCTAAT-3') with barcodes (Parada et al., 2015). One pooled sample was combined through an equimolar amount of each PCR product from each sample and sequenced using an Illumina HiSeq 2500 platform to generate 250 bp paired-end reads (Guangdong Magigene Biotechnology Co., Ltd., Guangzhou, China).

Sequencing Data Processing

Sequenced paired-end reads were processed using USEARCH (Version 11) (Edgar, 2010). The paired reads were merged with the fastq_mergepairs function and filtered with a "maxee" value of 1.0 (Edgar, 2010). The unique sequence reads were obtained with the fastx_uniques function and then denoised using the UNOISE3 (unoise_alpha = 2 and minsize = 4 as per default settings) algorithm to correct errors and remove chimeras (Edgar, 2016). This generated a list of zero-radius operational taxonomic units (ZOTUs). A ZOTU abundance table was produced by mapping the total reads to the list of ZOTUs with the -otutab function with an identity threshold of 0.97. The representative sequences for each ZOTU were assigned to taxonomic groups using the RDP classifier within the SILVA database (release 138) clustered at 99% similarity. Raw sequences have been deposited in the NCBI Sequence Read Archive (SRA) database under the BioProject number PRJNA604077 and the accession number SRP247264. Functional profiles of the bacterial community were annotated using FAPROTAX 1.2.2 using python collapse_table.py with the rarefied ZOTU table (Louca et al., 2016).

Bacterioplankton Diversity and Composition Analyses

All statistical analyses were performed in the R software environment (version 3.5.2) unless otherwise noted. To improve

the normality and homoscedasticity of the data, all of the biotic data were transformed by the Hellinger transformation, and the abiotic data were normalized by a chord transformation using the decostand function (Legendre and Gallagher, 2001). The Bray–Curtis dissimilarity matrix for bacterioplankton communities was generated to identify the relationships among samples with the vegdist function. Principal coordinate analysis (PCoA) and cluster analysis were used to visualize the overall structure of the bacterioplankton communities with the hclust (method = "ward.D") function. Permutational multivariate analysis of variance (PERMANOVA) was performed to quantitatively evaluate the contributions of seasons, latitude, and offshore distances to the variation in community structure with the adonis function. Analysis of similarity (ANOSIM) was performed to test whether there was a statistically significant difference in the community composition between the two seasons with the anosim function from the vegan package. To acquire the biomarker functions for each season, we used the randomForest (importance = TRUE, proximity = TRUE) function to generate a classification model for November and April. The biomarker functions were identified by the most important season-discriminating features using tenfold cross-validation implemented with the rfcv function (Zhang et al., 2019).

Ecological Process Estimation and Environmental Driver Analyses

The quantification of the effects of ecological processes on bacterioplankton community assembly was performed according to Stegen et al. (2013) based on the turnover in the phylogenetic community composition and species composition. First, the observed nearest-phylogenetic-neighbor distance ($\beta MNTD_{obs}$) as well as the null expectation of the mean nearest-phylogenetic-neighbor distance ($\beta MNTD_{exp}$) were used to calculate β-nearest taxon index (βNTI) metrics, which quantify the phylogenetic turnover between communities (Wang et al., 2019). The relative contributions of heterogeneous and homogeneous selection were estimated as the fractions of pairwise βNTI values that were $>+2$ and <-2, respectively (Stegen et al., 2015). Subsequently, the outcome of the βNTI analyses in combination with a second null model referred to as a Bray–Curtis-based Raup–Crick (RC_{bray}) (Chase et al., 2011) model was used to estimate the relative contributions of stochastic processes (Stegen et al., 2013). Specifically, a fraction of pairwise $|\beta NTI| < 2$ but $RC_{bray} < -0.95$ or $> +0.95$ indicated the relative contributions of homogenizing dispersal or dispersal limitation, respectively; a fraction of pairwise $|\beta NTI| < 2$ and $|RC_{bray}| < 0.95$ indicated that the community assembly was not dominated by any single process (Stegen et al., 2015).

Linear regressions were performed on Bray–Curtis dissimilarity (dependent variable) versus geographical distance (independent variable) through distance-decay analysis to quantify the distance decay in the bacterioplankton community (Wang et al., 2015). The relative contributions of environmental and spatial variables were evaluated by variation partitioning analysis (VPA). First, the spatial variables were derived from geographic coordinates using both linear trend and principal coordinates of neighbor matrices (PCNM) procedures (Griffith and Peres-Neto, 2006) to capture all spatial scales with the PCNM function. Second, the environmental, trend, and PCNM variables that explained a significant ($P < 0.05$) amount of variation in bacterioplankton community composition were selected by a forward selection model with the forward.sel function (Blanchet et al., 2008). The significance of the selected variables was assessed with 1000 permutations by using the rda and anova.cca functions from the vegan package. Finally, variance partitioning was performed using the varpart function from the vegan package on the Hellinger-transformed bacterial data, the forward-selected environmental variables, and the forward-selected spatial variables, and the data were tested for significance with 999 permutations (Meot et al., 1998). Pairwise comparisons of environmental variables were performed using the corrplot function from the corrplot package. Partial Mantel correlations between environmental variables and bacterioplankton community composition controlled for geographic distance were computed using the mantel.partial function from the corrplot package.

RESULTS

Taxonomic and Functional Characteristics of Bacterioplankton Communities

The bacterioplankton communities were mainly composed by Gammaproteobacteria (mean relative abundance, 24.5%), Alphaproteobacteria (23.7%), Actinobacteria (18.1%), Bacteroidetes (11.9%), and Cyanobacteria (9.9%) during the overwintering period, but the composition of phyla/classes between the two seasons was significantly different [ANOSIM, $R = 0.148$, $p = 0.003$] (**Figure 2**). Similarly, the functional composition of bacterioplankton communities between two seasons was significantly different [ANOSIM, $R = 0.321$, $p = 0.001$] although they were all dominant by chemoheterotrophy (mean relative abundance, 24.7%), aerobic chemoheterotrophy (22.1%), phototrophy (6.13%), photoautotrophy (6.04%), cyanobacteria (6.00%), and oxygenic photoautotrophy (6.00%) (**Figure 2**). Additionally, we observed that the relative *SD* of functional variability was significantly lower than those of variability at different taxonomic levels (**Figure 3**), indicating high taxonomic variability (even at phylum level) accompanied by relatively stable functional distributions in the overwintering ground.

Spatiotemporal Variation and Patterns of Bacterioplankton Taxa and Function

To visualize the overall taxonomic and functional structure of bacterioplankton communities and their underlying driving forces, we used PCoA, Cluster in combination with PERMANOVA on Bray–Curtis dissimilarities among the samples. We found that samples in the overwintering ground display a temporal pattern of divergence along

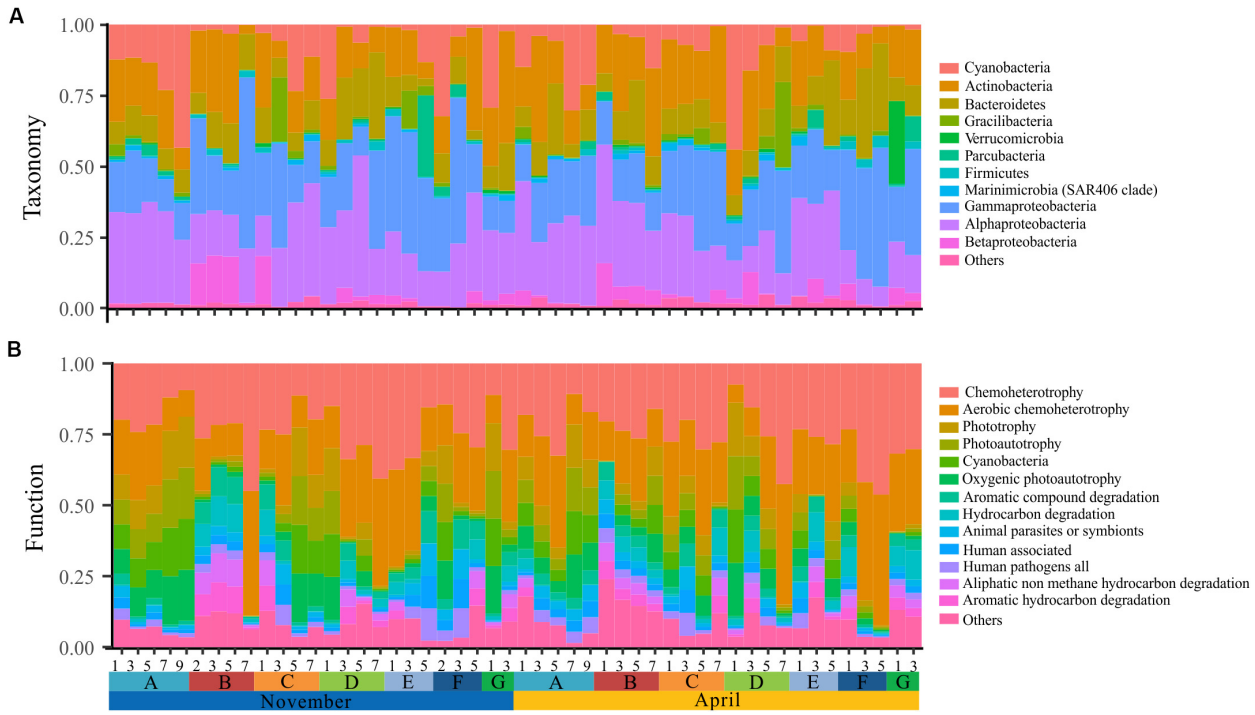


FIGURE 2 | Spatial and temporal variation of taxonomy and function of surface bacterioplankton communities in the overwintering ground. **(A)** Phylum-level (phylum Proteobacteria is grouped into the class level) taxonomic compositions of all samples. **(B)** Functional characterization of all samples.

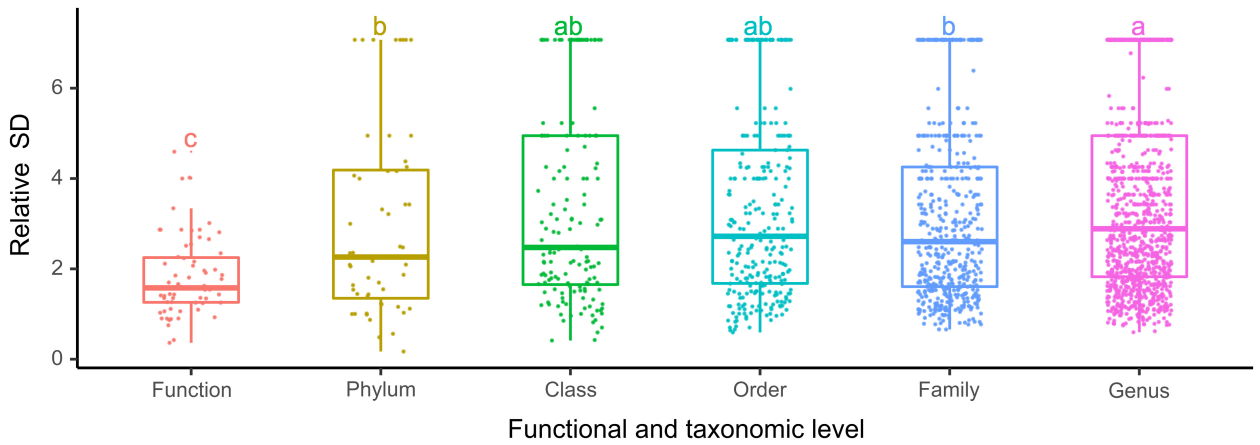


FIGURE 3 | The relative *SD* of bacterioplankton taxonomic and functional variability in the overwintering ground. Different lowercase letters denote significant differences ($p < 0.05$) among levels based on one-way ANOVA.

the first principal coordinate (**Figures 4A,B**), and this divergence was further corroborated by Cluster, where the communities from November and April formed separate clusters (**Figures 4C,D**). PERMANOVA of pairwise dissimilarities between bacterioplankton communities indicated that seasonal compartmentalization comprises the largest source of variation (**Table 1**). Although the latitude and offshore distances also showed significant effects (**Table 1**), the community structure did not exhibit a regular or obvious spatial pattern (**Figures 4A,B**).

Specific Functional Groups and Genera of Bacterioplankton Communities in Two Seasons

In view of the functional differences of the bacterioplankton community between seasons, we further explored the specific functional groups and functional genera for November and April. A model based on the random forest machine-learning algorithm was established to evaluate the importance of indicator

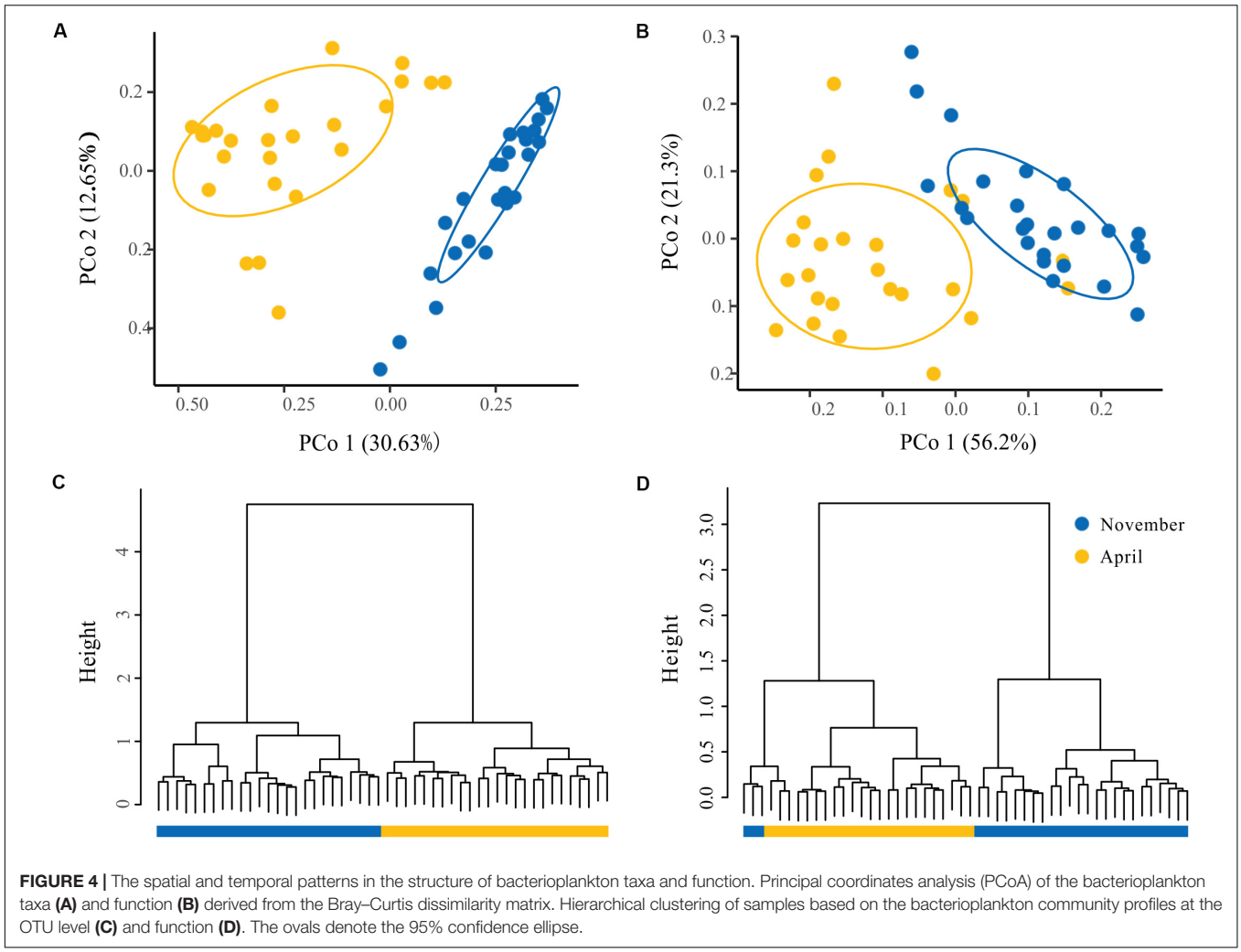


FIGURE 4 | The spatial and temporal patterns in the structure of bacterioplankton taxa and function. Principal coordinates analysis (PCoA) of the bacterioplankton taxa **(A)** and function **(B)** derived from the Bray–Curtis dissimilarity matrix. Hierarchical clustering of samples based on the bacterioplankton community profiles at the OTU level **(C)** and function **(D)**. The ovals denote the 95% confidence ellipse.

TABLE 1 | Quantitative effects of seasons, latitude, and offshore distances on variation in community structure (taxonomic and functional) using non-parametric permutational multivariate analysis of variance (PERMANOVA).

	Taxonomic structure		Functional structure	
	R^2	*p*	R^2	*p*
Seasons	0.287	**0.001**	0.350	**0.001**
Latitude	0.118	**0.012**	0.103	0.176
Offshore distances	0.033	**0.013**	0.040	**0.038**
Seasons: latitude	0.149	**0.002**	0.148	**0.032**
Seasons: offshore distances	0.018	0.147	0.013	0.478
Latitude: offshore distances	0.059	0.837	0.032	0.999
Seasons: latitude: offshore distances	0.070	0.535	0.059	0.893

Data in bold indicate significant correlations ($p < 0.05$).

bacterial functions. Ultimately, five bacterial functions were defined as biomarkers when considering the minimum cross-validation error of the model (**Figure 5A**). Of these, four bacterial functions showed higher relative abundance in April than November, and only one bacterial function showed higher relative abundance in November than April (one-way ANOVA, $p < 0.001$) (**Figure 5A**). Furthermore, we observed that the dynamics of bacterial genera that drive the biomarker functions showed significant temporal divergence (**Figure 5B**). For example, the genus *Acinetobacter* and *Moraxella* were abundant in the bacterioplankton communities collected in November, and the genus *Rhodococcus* and *Stenotrophomonas* were abundant in the bacterioplankton communities collected in April (**Figure 5B**).

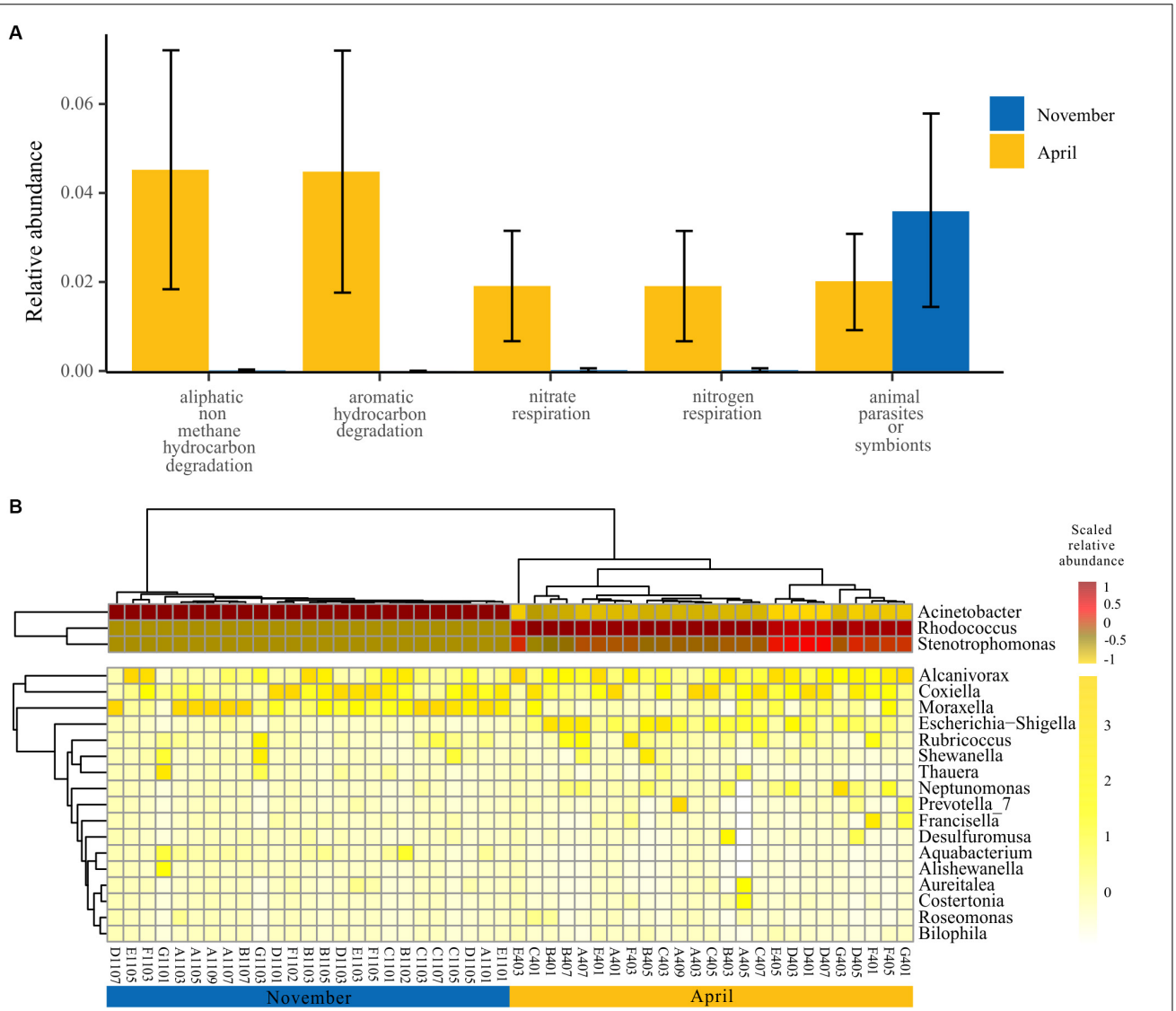


FIGURE 5 | Specific bacterial functions of two seasons. **(A)** Biomarker functions with higher relative abundance in two seasons. The eight biomarker functions were identified by applying the random-forest classification of the relative abundance of the bacterioplankton communities against November and April. Data bars represent means. Error bars represent SEM. **(B)** The spatial and temporal dynamics of bacterioplankton genera that drive the biomarker functions.

Ecological Processes Governing the Spatiotemporal Assembly of Bacterioplankton Communities

To explore the underlying driving forces of bacterioplankton community variation, we used the mean nearest taxon distance (MNTD) to quantify the relative contributions of major ecological processes that regulate the community assembly. The ecological process dominating the community assembly was dispersal limitation (accounting for 67.0–79.7%), which was relatively stable over the seasons (**Figure 6**). Second, 20.3–31.0% of the variation in community composition was primarily due to heterogeneous selection (**Figure 6**). In addition, there was still 0–3.7% of compositional variation unexplained by any ecological processes (**Figure 6**). These results indicate that stochastic rather than deterministic processes structured the bacterioplankton community assembly in the overwintering ground.

Distance-Decay of Similarities in Taxonomy and Function of Bacterioplankton Communities

The spatial dissimilarities in bacterioplankton taxonomic composition were significantly higher than that in functional composition (**Figure 7B**). Furthermore, the taxonomic composition of bacterioplankton communities collected in both November (adjusted R^2 = 0.139, $p < 0.001$) and April (adjusted R^2 = 0.040, $p < 0.001$) exhibit a clear distance-decay pattern in similarities (**Figure 7A**). This pattern was also

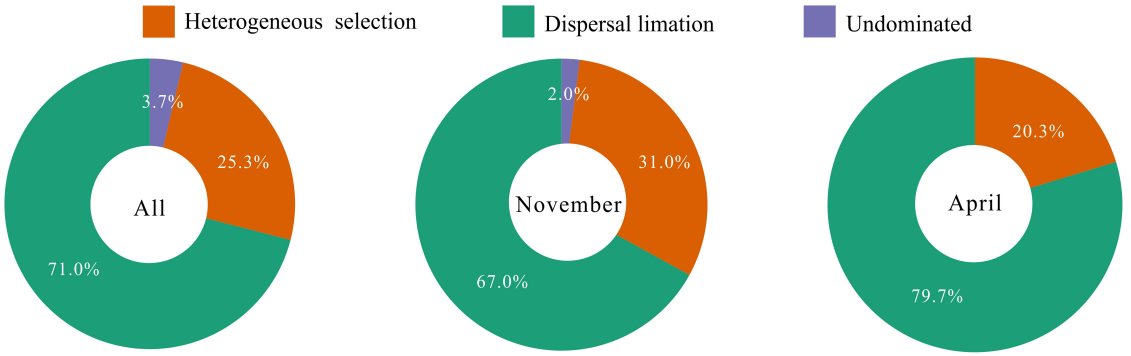


FIGURE 6 | Summary of ecological processes governing the compositional variation of total bacterioplankton communities and the bacterioplankton communities collected in November and April. The percentages are given as the relative contribution of each process to the community variation as indicated by different colors.

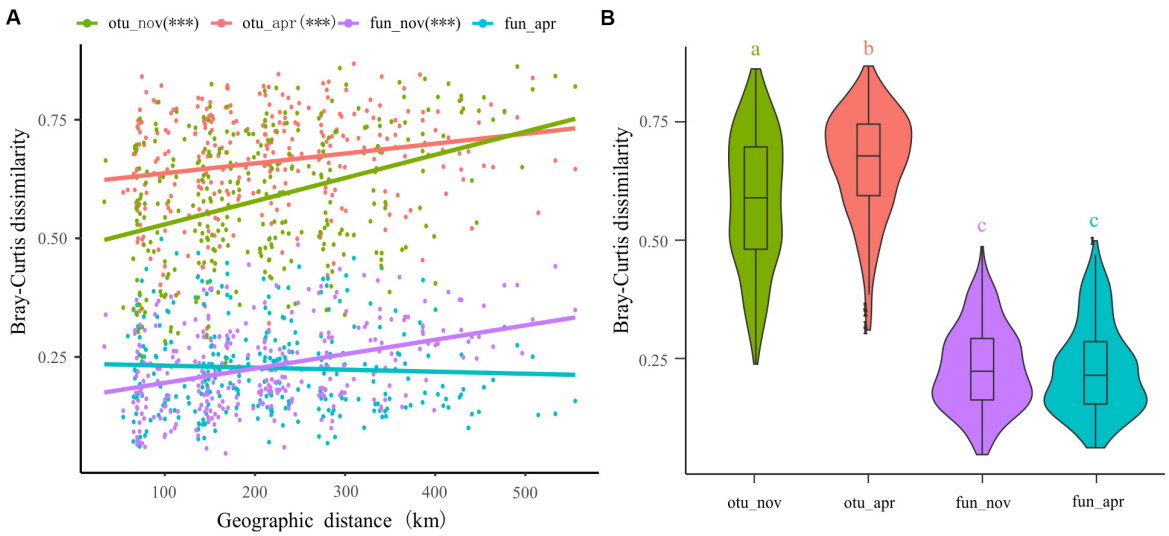


FIGURE 7 | Distance-decay of similarities in taxonomy and function of bacterioplankton communities in different seasons. **(A)** Corrections between Bray–Curtis dissimilarities with geographic distance between sites. **(B)** The spatial dissimilarities in bacterioplankton taxonomic composition and functional composition.

observed in the bacterial functional composition in November (adjusted R^2 = 0.142, p < 0.001) but not in April (adjusted R^2 = -0.001, p = 0.395) (**Figure 7A**).

Drivers of Bacterioplankton Taxonomic and Functional Composition

To explore the deterministic mechanisms underlying community assembly, we used VPA for disentangling the relative role of environmental and spatial variables. Results of the variation partitioning showed that the drivers of bacterioplankton community composition were different between seasons and between taxonomic and functional levels (**Figure 8**). Specifically, the environment-related variables (pure environmental variables, environmental-trend covariation, environmental-PCNM covariation, covariation of all variables) dominated the variation of bacterioplankton taxonomic composition collected in all seasons (both November and April) and November (only November) and the variation of bacterioplankton functional composition collected in April (only April), while space-related variables (pure linear trend, pure PCNM, environmental-trend covariation, environmental-PCNM covariation, trend-PCNM covariation, covariation of all variables) dominated the variation of bacterioplankton taxonomic composition collected in April and the variation of bacterioplankton functional composition collected in all seasons and November (**Figure 8**). Remarkably, there was a large proportion (53.4–76.7%) of variation unexplained by the above variables. To further identify the geographic-independent environmental drivers, we correlated distance-corrected dissimilarities of taxonomic and functional community composition with those of environmental variables (**Figure 9**). Overall, nutrients were the strongest correlates of both taxonomic and functional composition in both seasons. Moreover, significant correlations were found between the temporal differences in community composition (between the seasons) and various environmental variables (**Figure 9A**),

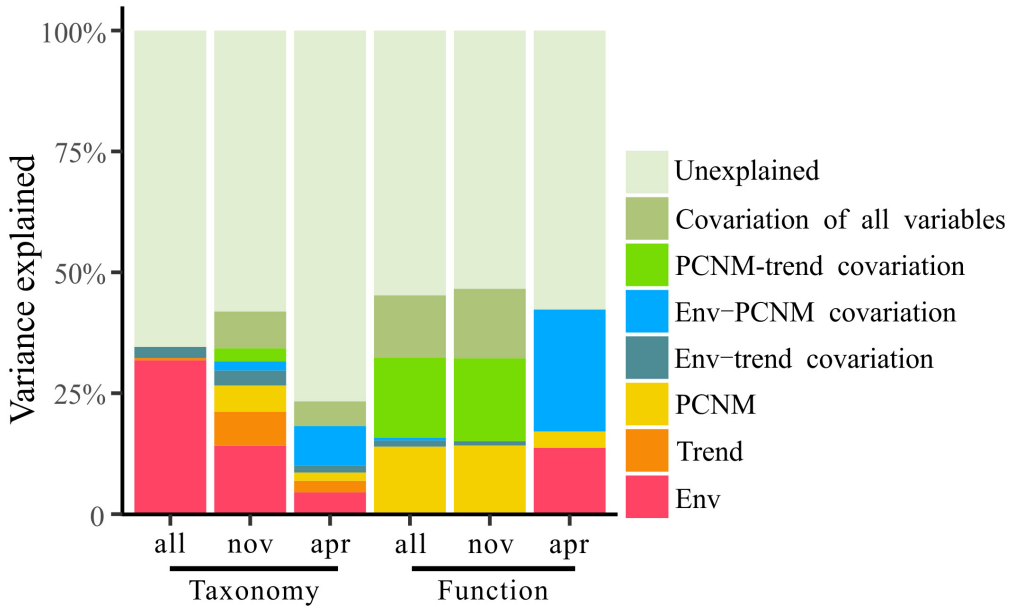


FIGURE 8 | Variation partitioning of the taxonomic and functional structure of total bacterioplankton communities (all) and the bacterioplankton communities collected in November (nov) and April (apr) by the seawater environmental variables (Env) and the spatial variables including linear trend (Trend) and PCNM variables (PCNM).

and the spatial differences in community composition (within the season) were weakly driven by environmental variables (**Figures 9B,C**).

DISCUSSION

The Spatiotemporal Patterns and Characteristics of Surface Bacterioplankton Communities in an Overwintering Ground

The habitats of the large yellow croaker are being degraded due to the increased use of coastal areas for aquaculture operations and development (Kang et al., 2018). The failed recovery of the common snook (*Centropomus undecimalis*) indicates that habitat degradation is considered more significant than exploitation in explaining population declines (Brennan et al., 2008), which indicates the need for the study and restoration of croaker habitat. Planktonic bacteria are generally used as a bioindicator to systematically reflect or record the conditions of the marine environment or fish habitats due to their biogeochemical significance and direct relationship with fish health (Dowle et al., 2015; Dai et al., 2017). Over the past decade, dozens of studies have characterized patterns, associations, and drivers in the bacterioplankton community in various fish habitats and elucidated similarities and differences in the bacterioplankton communities between fish farms and natural areas (Tamminen et al., 2011; Xiong et al., 2015; Jing et al., 2019). However, comprehensive information on the spatial and temporal patterns of bacterioplankton communities in the croaker habitat is scarce.

Our study systematically investigated the surface bacterioplankton communities in an overwintering ground for large yellow croakers. We found that, during the overwintering period, the dominant bacterioplankton phyla and functions remained unchanged, but their relative abundance exhibited seasonality (**Figures 2**, **4** and **Table 1**). Specifically, the bacterioplankton community in autumn was characterized by animal parasites or symbionts (**Figure 5A**), and *Acinetobacter* was the representative genus (**Figure 5B**). *Acinetobacter baumannii*, which was detected in our samples, is a virulent pathogen for many fish species that can trigger serious septicemia outbreaks (Brahmi et al., 2016; Li et al., 2017). More seriously, *A. baumannii* has been proven to be a serious human clinical pathogen involved in a large number of infections, including bacteremia, secondary meningitis, urinary tract infection, and ventilator-associated pneumonia (Mari-Almiralla et al., 2019). Therefore, the presence of *Acinetobacter* in overwintering grounds increases the latent risk of fish disease and the possibility of coinfection between fish and humans through handling or consumption (Xia et al., 2008). In addition, other functional genera found in autumn, such as *Moraxella* (**Figure 5B**), are opportunistic pathogens of fish (Walke et al., 2015). Once the environment deteriorates or suddenly changes, those bacteria might threaten the health of the croaker or other fish. In spring, many nutrient-related functional bacterioplankton were dominant (**Figure 5**). *Rhodococcus*, one of the dominant functional genera in April (**Figure 5B**), is generally regarded as the most versatile and efficient group of organisms that are suitable for the biodegradation of organic compounds (Larkin et al., 2005; Martínková et al., 2009). Given this characteristic, *Rhodococcus* has been found worldwide in

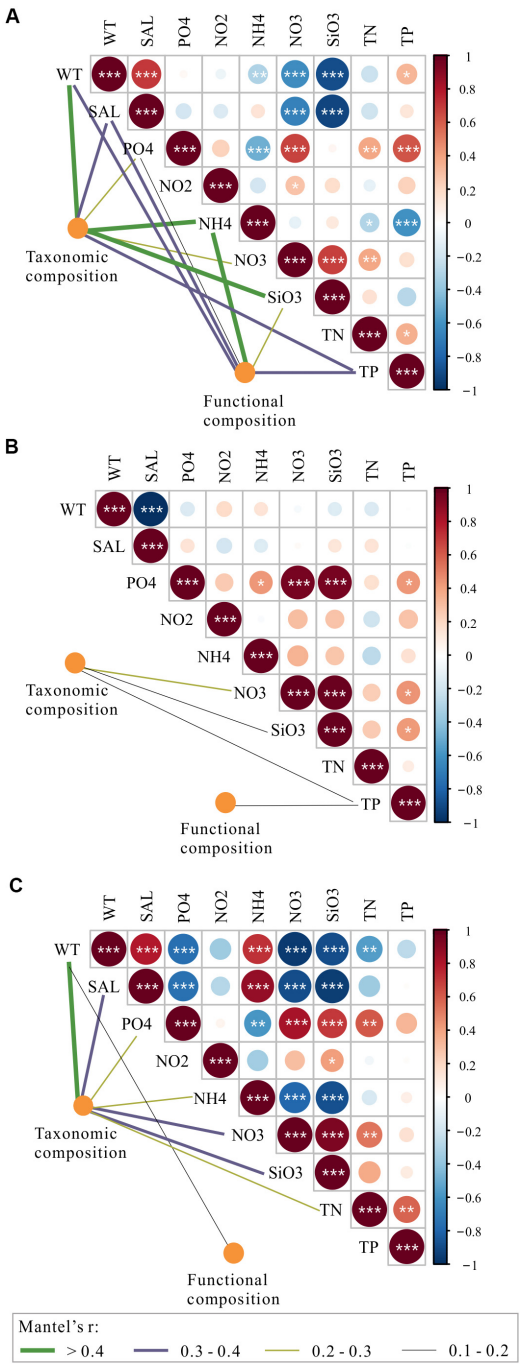


FIGURE 9 | Environmental drivers of taxonomic and functional composition of total bacterioplankton communities **(A)** and the bacterioplankton communities collected in November **(B)** and April **(C)**. Pairwise comparisons of environmental variables are shown with a color gradient denoting Spearman's correlation coefficients. The taxonomic and functional composition of bacterioplankton communities was related to each environment variable by partial (geographic distance-corrected) Mantel tests. Edge width corresponds to Mantel's *r* statistic for the corresponding distance correlations, and edge color denotes the statistical significance based on 9999 permutations. WT, water temperature; SAL, salinity; PO_4, orthophosphate; NO_2, nitrite; NO_3, nitrate; NH_4, ammonium; SiO_3, silicate; TN, total nitrogen; TP, total phosphorus.

eutrophic aquatic ecosystems (Wu et al., 2009; Lee et al., 2010), suggesting that it might be used as an organismal indicator of nutrient enrichment (Rowbotham and Cross, 1977). Moreover, *Alcanivorax* (**Figure 5B**), a hydrocarbon-degrading bacterium, could also be enriched in eutrophic environments and could be related to the presence of pollutants (Patel et al., 2014). Other functional genera in spring, such as *Stenotrophomonas* (Jia et al., 2019) and *Neptunomonas* (Lanfranconi et al., 2010), were also organic-degrading or nitrogen-related organisms that prefer organic-rich or eutrophic conditions. Therefore, the functional composition of the bacterioplankton community in April suggests that the nutrient level in the surface water of the habitat in spring was higher than that in the winter, which was verified by environmental data (**Supplementary Table S1**) and implies that this sea area has a high risk of developing eutrophication.

The Mechanism of Surface Bacterioplankton Community Assembly in the Overwintering Ground for Large Yellow Croaker

One key question in ocean microbial ecology is the extent to which environmental filtering and various biological interactions, on the one hand, and limited dispersal and historical contingency, on the other, are responsible for spatiotemporal patterns (Sunagawa et al., 2015). Nevertheless, the answer to this question in marine ecosystems is still controversial. For example, some studies have proposed that marine bacterioplankton communities are structured by deterministic factors (Pontarp et al., 2012; Sunagawa et al., 2015; Wang et al., 2019), and other studies have demonstrated that stochastic factors can potentially overwhelm the influences of selection on community structure (Beaton et al., 2016; Evans et al., 2017; Zhou and Ning, 2017). The latter theory is supported by our findings as dispersal limitation dominated the processes governing both the spatial and temporal scales of bacterioplankton community assembly (**Figure 6**). It has historically been considered that free-living bacteria are small, numerous, and easily transported and, hence, do not experience passive dispersal limitations (Cottenie, 2005; Shurin et al., 2009). However, numerous studies have recently documented that bacterial communities in water show spatial differences and reveal strong biogeographic patterns (Verreydt et al., 2012; Müller et al., 2014; Wu et al., 2018). The distance-similarity decay relationships for the bacterioplankton communities observed in our study (**Figure 7**) also suggest the impact of dispersal limitation on the generation of biogeographical patterns in marine bacteria (Bell, 2010; Adams et al., 2013). It is interesting to note that, even though the study area is located in the East China Sea and the geographic scale of the overwintering ground we investigated is comparatively broad, the distance-decay patterns in our study were weaker than those in other studies as the distance decay rates (the regression slopes) of our bacterioplankton communities were lower (Wang et al., 2015; Dai et al., 2017). The strong distance-decay patterns in Wang et al. (2015) and Dai et al. (2017) were mainly attributed to spatially structured environmental gradients, which essentially suggests that the spatial heterogeneity of bacterioplankton communities in

their studies was structured by the deterministic heterogeneous selection of abiotic and biotic environmental conditions (Dini-Andreote et al., 2015). However, stochastic dispersal exerted a dominant role in shaping our bacterioplankton communities (**Figure 6**), contributed to the relatively random distribution of the bacterial community compared with that under harsh environmental filters (Evans et al., 2017), and attenuated the heterogeneity of bacterial taxa in our study. Additionally, lower or no distance-similarity decay relationships were found in bacterial functions compared with the relationships found in bacterial taxa (**Figure 7**). This is a universal phenomenon because functional redundancy, which means that different populations share similar or the same functions, appears to be quite high in the microbial community, especially compared with the level of functional redundancy in plant or animal communities (Delgado-Baquerizo et al., 2016; Louca et al., 2018).

Comparatively, environmental stressors wielded less power than stochastic processes in creating temporal and spatial differences in the bacterioplankton communities' taxonomical and functional composition in the overwintering habitat for croaker. Our variation partitioning results showed that over 50% of the variation in the bacterioplankton community structure was unexplained by the environment-related variables (**Figure 8**). The lower power of the environmental variables could be explained by the restraining force of stochastic effects (Zhou and Ning, 2017); dispersal limitation can obscure the relationships between environmental variables and microbial community composition, resulting in apparently weak deterministic effects (Evans et al., 2017).

Among the purely environmental filters that were independent of spatial effects, nutrient-related variables, such as silicate, nitrate, and ammonium, could have exhibited some degree of heterogeneous selection on temporal and spatial variations in the bacterioplankton community composition in the overwintering grounds (**Figure 9**). The mechanism by which silicate influences the composition of the bacterioplankton community could be related to its influence on diatom growth; it may shape bacterioplankton community composition due to the tight coupling between diatoms and bacteria (Topper et al., 2010; Wang et al., 2015). The correlations of bacterioplankton community composition with nitrogenous nutrients could be attributed to the abundant biodegrading bacteria in the habitat (**Figure 5**), which are capable of decomposing organic substances into various inorganic nutrients (Manser et al., 2006; Li and Boyd, 2016). An equally important driver of variation in the bacterioplankton community was temperature (**Figure 9**); this finding is in accordance with those reported in freshwater ecosystems, oceans, and forests at the global scale (Penuelas et al., 2012; Sunagawa et al., 2015; Chen C. Y. et al., 2019). The importance of salinity in shaping the composition of the bacterioplankton community was also observed (**Figure 9**). Salinity may contribute to density gradients that physically separate water masses and their resident bacterial communities from near-shore to offshore areas (Lozupone and Knight, 2007; Fortunato et al., 2012) (**Supplementary Table S1**). Notably, more correlations were found between the temporal differences in community composition and the various environmental variables than between spatial differences and environmental variables (**Figure 9**); this difference may have resulted from the greater differences in environmental variables between seasons. Interestingly, the interactions among community composition and purely environmental filters were more complex and closer in spring than in autumn (**Figures 9B,C**). One possible explanation for this difference is that the high eutrophication level in April (**Supplementary Table S1**), as a long-term and continuous disturbance, could increase the importance of deterministic environmental filtering to bacterioplankton community assembly (**Figure 6**) (Dai et al., 2017).

Overall, we obtained a unique taxonomic and functional profile of the bacterioplankton community in the surface water of the overwintering ground and revealed the assembly mechanisms that structure it on the spatial–temporal scale. The systematic map of the surface bacterioplankton community has implied that current conditions are not conducive to the high primary productivity and the health of upper elements of the croakers' food chain. There are, however, potential limitations to our approach that merit further discussion. First, although the above results and discussion display the patterns and processes of bacterioplankton community assembly in the overwintering grounds for large yellow croaker, samples collected from the surface water can only somewhat reflect habitat conditions for the croaker who live at moderate and lower water depths (Lü et al., 2008). Therefore, we will further study the habitat microbiome of the water masses in which croaker is located. Additionally, we did not sample replicates at the study sites, which might have limited the analysis of the within-site variability and thereby might affect the accuracy of the results of the random forest classification and the distance decay analysis to a certain extent.

DATA AVAILABILITY STATEMENT

All the raw sequences have been uploaded to the NCBI Sequence Read Archive (SRA) database under the BioProject number PRJNA604077 and the accession number SRP247264.

AUTHOR CONTRIBUTIONS

WY and Z-MZ led manuscript writing and data analysis. S-ZZ, S-HZ, and BL contributed with sequence analysis, statistical analyses, and microbiological interpretations. LZ and RN collected and processed samples. J-YZ and C-HL were responsible for organizing field sampling and inter-institution communication. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research was supported by the National Key Research and Development Program of China (Reference No. 2018YFC1406300), the Natural Science Foundation of Zhejiang Province (Reference No. LQ20C190003), the Department of Education Scientific Research Project of Zhejiang

Province (Reference No. Y201839309), Natural Science Foundation of Ningbo (Reference Nos. 2019A610421 and 2019A610443), and the K.C. Wong Magna Fund in Ningbo University.

ACKNOWLEDGMENTS

The authors would like to thank the colleagues from Xiamen University, Second Institute of Oceanography, and other institutions for helping on board. In addition, the authors thank Prof. Ling-feng Huang and Prof. Jian-Yu Hu from Xiamen University for their selfless sharing of environmental data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2020.00472/full#supplementary-material

REFERENCES

Adams, R. I., Miletto, M., Taylor, J. W., and Bruns, T. D. (2013). Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *ISME J.* 7, 1262–1273. doi: 10.1038/ismej.2013.28

AQSIQ (2007). *The Specification for Marine Monitoring of China-Part 4: Seawater Analysis (GB 17378.4e2007)*. Beijing: General administration of quality supervision, inspection and quarantine.

Beaton, E. D., Stevenson, B. S., King-Sharp, K. J., Stamps, B. W., Nunn, H. S., and Stuart, M. (2016). Local and regional diversity reveals dispersal limitation and drift as drivers for groundwater bacterial communities from a fractured granite formation. *Front. Microbiol.* 7:1933. doi: 10.3389/fmicb.2016.01933

Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *ISME J.* 4, 1357–1365. doi: 10.1038/ismej.2010.77

Blanchet, F. G., Legendre, P., and Borcard, D. (2008). Forward selection of explanatory variables. *Ecology* 89, 2623–2632. doi: 10.1890/07-0986.1

Brahmi, S., Touati, A., Cadiere, A., Djahmi, N., Pantel, A., Sotto, A., et al. (2016). First description of two sequence type 2 *Acinetobacter baumannii* isolates carrying OXA-23 carbapenemase in *Pagellus acarne* fished from the Mediterranean Sea near Bejaia, Algeria. *Antimicrob. Agents Chemother.* 60, 2513–2515. doi: 10.1128/Aac.02384-15

Brennan, N. P., Walters, C. J., and Leber, K. M. (2008). Manipulations of stocking magnitude: addressing density-dependence in a juvenile cohort of common snook (*Centropomus undecimalis*). *Rev. Fish. Sci.* 16, 215–227. doi: 10.1080/10641260701689022

Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M., and Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in alpha-diversity. *Ecosphere* 2:24. doi: 10.1890/Es10-00117.1

Chen, C. Y., He, R. Q., Cheng, Z. Y., Han, M. Z., Zha, Y. G., Yang, P. S., et al. (2019). The seasonal dynamics and the influence of human activities on campus outdoor microbial communities. *Front. Microbiol.* 10:1579. doi: 10.3389/fmicb.2019.01579

Chen, J. J., and Xu, Z. L. (2012). Analysis of population division and geographical isolation of *Larimichthys crocea* in the East China Sea and Yellow Sea. *J. Fish. China* 19, 310–320. doi: 10.3724/sp.j.1118.2012.00310

Chen, W. D., Ren, K. X., Isabwe, A., Chen, H. H., Liu, M., and Yang, J. (2019). Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. *Microbiome* 7:138. doi: 10.1186/s40168-019-0749-8

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecol. Lett.* 8, 1175–1182. doi: 10.1111/j.1461-0248.2005.00820.x

Dai, W. F., Zhang, J. J., Tu, Q. C., Deng, Y., Qiu, Q. F., and Xiong, J. B. (2017). Bacterioplankton assembly and interspecies interaction indicating increasing coastal eutrophication. *Chemosphere* 177, 317–325. doi: 10.1016/j.chemosphere.2017.03.034

Delgado-Baquerizo, M., Giaramida, L., Reich, P. B., Khachane, A. N., Hamonts, K., Edwards, C., et al. (2016). Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. *J. Ecol.* 104, 936–946. doi: 10.1111/1365-2745.12585

Dini-Andreote, F., Stegen, J. C., van Elsas, J. D., and Salles, J. F. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1326–E1332. doi: 10.1073/pnas.1414261112

Dowle, E., Pochon, X., Keeley, N., and Wood, S. A. (2015). Assessing the effects of salmon farming seabed enrichment using bacterial community diversity and high-throughput sequencing. *FEMS Microbiol. Ecol.* 91:fiv089. doi: 10.1093/femsec/fiv089

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* [Preprint]. doi: 10.1101/081257

Evans, S., Martiny, J. B. H., and Allison, S. D. (2017). Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J.* 11, 176–185. doi: 10.1038/ismej.2016.96

Fortunato, C. S., Herfort, L., Zuber, P., Baptista, A. M., and Crump, B. C. (2012). Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME J.* 6, 554–563. doi: 10.1038/ismej.2011.135

Giovannoni, S., and Nemergut, D. (2014). Microbes ride the current. *Science* 345, 1246–1247. doi: 10.1126/science.1259467

Griffith, D. A., and Peres-Neto, P. R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology* 87, 2603–2613. doi: 10.1890/0012-9658(2006)87[2603:smietf]2.0.co;2

Hu, X. J., Wen, G. L., Cao, Y. C., Gong, Y. X., Li, Z. J., He, Z. L., et al. (2017). Metabolic and phylogenetic profiles of microbial communities from a mariculture base on the Chinese Guangdong coast. *Fish. Sci.* 83, 465–477. doi: 10.1007/s12562-017-1073-5

Jia, Y., Zhou, M., Chen, Y., Lou, J., and Hu, Y. (2019). Carbon selection for nitrogen degradation pathway by *Stenotrophomonas* maltophilia: based on the balances of nitrogen, carbon and electron. *Bioresour. Technol.* 294:122114. doi: 10.1016/j.biortech.2019.122114

Jiao, S., Xu, Y., Zhang, J., and Lu, Y. (2019). Environmental filtering drives distinct continental atlases of soil archaea between dryland and wetland agricultural ecosystems. *Microbiome* 7:15. doi: 10.1186/s40168-019-0630-9

Jing, X. Y., Gou, H. L., Gong, Y. H., Ji, Y. T., Su, X. L., Zhang, J., et al. (2019). Seasonal dynamics of the coastal bacterioplankton at intensive fish-farming areas of the Yellow Sea, China revealed by high-throughput sequencing. *Mar. Pollut. Bull.* 139, 366–375. doi: 10.1016/j.marpolbul.2018.12.052

Kalcheva, H. V., Nikolova, L. N., Terziyski, D. I., Stoeva, A. S., Pehlivanov, L. Z., and Kalchev, R. K. (2008). Relationships between bacterioplankton, zooplankton and environmental factors in fertilized and non-fertilized carp fishponds. *Acta Zool. Bulgarica Suppl.* 2, 175–184.

Kang, B., Liu, M., Huang, X. X., Li, J., Yan, Y. R., Han, C. C., et al. (2018). Fisheries in Chinese seas: what can we learn from controversial official fisheries statistics? *Rev. Fish Biol. Fish.* 28, 503–519. doi: 10.1007/s11160-018-9518-1

Keen, M. R., Schwarz, A. M., and Wini-Simeon, L. (2018). Towards defining the Blue Economy: practical lessons from pacific ocean governance. *Mar. Policy* 88, 333–341. doi: 10.1016/j.marpol.2017.03.002

Lanfranconi, M. P., Bosch, R., and Nogales, B. (2010). Short-term changes in the composition of active marine bacterial assemblages in response to diesel oil pollution. *Microb. Biotechnol.* 3, 607–621. doi: 10.1111/j.1751-7915.2010.00192.x

Larkin, M. J., Kulakov, L. A., and Allen, C. C. R. (2005). Biodegradation and Rhodococcus – masters of catabolic versatility. *Curr. Opin. Biotechnol.* 16, 282–290. doi: 10.1016/j.copbio.2005.04.007

Lee, Y. K., Ahn, C. Y., Kim, H. S., and Oh, H. M. (2010). Cyanobactericidal effect of *Rhodococcus* sp isolated from eutrophic lake on *Microcystis* sp. *Biotechnol. Lett.* 32, 1673–1678. doi: 10.1007/s10529-010-0350-5

Legendre, P., and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280. doi: 10.1007/s004420100716

Li, J., Cao, J. L., Wang, X., Liu, N., Wang, W. M., and Luo, Y. (2017). Acinetobacter pittii, an emerging new multi-drug resistant fish pathogen isolated from diseased blunt snout bream (*Megalobrama amblycephala* Yih) in China. *Appl. Microbiol. Biotechnol.* 101, 6459–6471. doi: 10.1007/s00253-017-8392-4

Li, Y. L., and Boyd, C. E. (2016). Laboratory tests of bacterial amendments for accelerating oxidation rates of ammonia, nitrite and organic matter in aquaculture pond water. *Aquaculture* 460, 45–58. doi: 10.1016/j.aquaculture.2016.03.050

Liu, C., Shen, W. L., Hou, C. C., Gao, X. M., Wang, Q. F., Wu, X. F., et al. (2019). Low temperature-induced variation in plasma biochemical indices and aquaglyceroporin gene expression in the large yellow croaker *Larimichthys crocea*. *Sci. Rep.* 9:2717. doi: 10.1038/s41598-018-37274-3

Liu, M., and de Mitcheson, Y. S. (2008). Profile of a fishery collapse: why mariculture failed to save the large yellow croaker. *Fish Fish.* 9, 219–242. doi: 10.1111/j.1467-2979.2008.00278.x

Louca, S., Parfrey, L. W., and Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science* 353, 1272–1277. doi: 10.1126/science.aaf4507

Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., et al. (2018). Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* 2, 936–943. doi: 10.1038/s41559-018-0519-1

Lozupone, C. A., and Knight, R. (2007). Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11436–11440. doi: 10.1073/pnas.0611525104

Lü, H., Xu, J., and Vander Haegen, G. (2008). Supplementing marine capture fisheries in the East China Sea: sea ranching of Prawn *Penaeus orientalis*, restocking of Large Yellow Croaker *Pseudosciaena crocea*, and cage culture. *Rev. Fish. Sci.* 16, 366–276. doi: 10.1080/10641260701678207

Manser, R., Gujer, W., and Siegrist, H. (2006). Decay processes of nitrifying bacteria in biological wastewater treatment systems. *Water Res.* 40, 2416–2426. doi: 10.1016/j.watres.2006.04.019

Mari-Almiralla, M., Cosgaya, C., Pons, M. J., Nemec, A., Ochoa, T. J., Ruiz, J., et al. (2019). Pathogenic Acinetobacter species including the novel *Acinetobacter dijkshoorniae* recovered from market meat in Peru. *Int. J. Food Microbiol.* 305:108248. doi: 10.1016/j.ijfoodmicro.2019.108248

Martínková, L., Uhnáková, B., Pátek, M., Nešvera, J., and Křen, V. (2009). Biodegradation potential of the genus *Rhodococcus*. *Environ. Int.* 35, 162–177. doi: 10.1016/j.envint.2008.07.018

Meot, A., Legendre, P., and Borcard, D. (1998). Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling framework. *Environ. Ecol. Stat.* 5, 1–27. doi: 10.1023/A:1009693501830

Müller, A. L., de Rezende, J. R., Hubert, C. R. J., Kjeldsen, K. U., Lagkouvardos, I., Berry, D., et al. (2014). Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J.* 8, 1153–1165. doi: 10.1038/ismej.2013.225

Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2015). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi: 10.1111/1462-2920.13023

Patel, V., Munot, H., Shouche, Y. S., and Madamwar, D. (2014). Response of bacterial community structure to seasonal fluctuation and anthropogenic pollution on coastal water of Alang–Sosiya ship breaking yard, Bhavnagar, India. *Bioresour. Technol.* 161, 362–370. doi: 10.1016/j.biortech.2014.03.033

Penuelas, J., Rico, L., Ogaya, R., Jump, A. S., and Terradas, J. (2012). Summer season and long-term drought increase the richness of bacteria and fungi in the foliar phyllosphere of *Quercus ilex* in a mixed Mediterranean forest. *Plant Biol.* 14, 565–575. doi: 10.1111/j.1438-8677.2011.00532.x

Pontarp, M., Canback, B., Tunlid, A., and Lundberg, P. (2012). Phylogenetic analysis suggests that habitat filtering is structuring marine bacterial communities across the globe. *Microb. Ecol.* 64, 8–17. doi: 10.1007/s00248-011-0005-7

Rowbotham, T. J., and Cross, T. (1977). Ecology of rhodococcus coprophilus and associated actinomycetes in fresh water and agricultural habitats. *J. Gen. Microbiol.* 100, 231–240. doi: 10.1099/00221287-100-2-231

Shurin, J. B., Cottenie, K., and Hillebrand, H. (2009). Spatial autocorrelation and dispersal limitation in freshwater organisms. *Oecologia* 159, 151–159. doi: 10.1007/s00442-008-1174-z

Stegen, J. C., Lin, X. J., Fredrickson, J. K., Chen, X. Y., Kennedy, D. W., Murray, C. J., et al. (2013). Quantifying community assembly processes and identifying features that impose them. *ISME J.* 7, 2069–2079. doi: 10.1038/ismej.2013.93

Stegen, J. C., Lin, X. J., Fredrickson, J. K., and Konopka, A. E. (2015). Estimating and mapping ecological processes influencing microbial community assembly. *Front. Microbiol.* 6:370. doi: 10.3389/fmicb.2015.00370

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359

Tamminen, M., Karkman, A., Corander, J., Paulin, L., and Virta, M. (2011). Differences in bacterial community composition in Baltic Sea sediment in response to fish farming. *Aquaculture* 313, 15–23. doi: 10.1016/j.aquaculture.2011.01.020

Topper, B., Larsen, A., Thingstad, T. F., Thyrhaug, R., and Sandaa, R. A. (2010). Bacterial community composition in an Arctic phytoplankton mesocosm bloom: the impact of silicate and glucose. *Polar Biol.* 33, 1557–1565. doi: 10.1007/s00300-010-0846-4

Tursi, A., Maiorano, P., Sion, L., and D'Onghia, G. (2015). Fishery resources: between ecology and economy. *Rendiconti Lincei Sci. Fisiche E Nat.* 26, 73–79. doi: 10.1007/s12210-014-0372-3

Verreydt, D., De Meester, L., Decaestecker, E., Villena, M.-J., Van Der Gucht, K., Vannormelingen, P., et al. (2012). Dispersal-mediated trophic interactions can generate apparent patterns of dispersal limitation in aquatic metacommunities. *Ecol. Lett.* 15, 218–226. doi: 10.1111/j.1461-0248.2011.01728.x

Walke, S., Kashte, S., Parwe, N., and Chavan, S. (2015). Opportunistic pathogens in the mucous of skin, gills, fins and mouth of Labeo Rohita. *Int. J. Fish. Aquat. Stud.* 3, 169–172.

Wang, K., Hu, H. J., Yan, H. Z., Hou, D. D., Wang, Y. T., Dong, P. S., et al. (2019). Archaeal biogeography and interactions with microbial community across complex subtropical coastal waters. *Mol. Ecol.* 28, 3101–3118. doi: 10.1111/mec.15105

Wang, K., Ye, X. S., Chen, H. P., Zhao, Q. F., Hu, C. J., He, J. Y., et al. (2015). Bacterial biogeography in the coastal waters of northern Zhejiang, East China Sea is highly controlled by spatially structured environmental gradients. *Environ. Microbiol.* 17, 3898–3913. doi: 10.1111/1462-2920.12884

Wang, L., Shi, X. F., Su, Y. Q., Meng, Z. N., and Lin, H. R. (2012). Loss of genetic diversity in the cultured stocks of the Large Yellow Croaker, *Larimichthys crocea*, revealed by microsatellites. *Int. J. Mol. Sci.* 13, 5584–5597. doi: 10.3390/ijms13055584

Wu, G. F., Liu, J., and Ye, Z. X. (2009). Characterization of phosphobacteria isolated from eutrophic aquatic ecosystems. *Microbiology* 78, 769–775. doi: 10.1134/S0026261709060149

Wu, W. X., Lu, H. P., Sastri, A., Yeh, Y. C., Gong, G. C., Chou, W. C., et al. (2018). Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME J.* 12, 485–494. doi: 10.1038/ismej.2017.183

Xia, L., Xiong, D. M., Gu, Z. M., Xu, Z., Chen, C. F., Xie, J., et al. (2008). Recovery of *Acinetobacter baumannii* from diseased channel catfish (*Ictalurus punctatus*) in China. *Aquaculture* 284, 285–288. doi: 10.1016/j.aquaculture.2008.07.038

Xiong, J. B., Chen, H. P., Hu, C. J., Ye, X. S., Kong, D. J., and Zhang, D. M. (2015). Evidence of bacterioplankton community adaptation in response to long-term mariculture disturbance. *Sci. Rep.* 5:15274. doi: 10.1038/srep15274

Xu, Z. L., and Chen, J. J. (2011). Analysis of migratory route of *Larimichthys crocea* in the East China Sea and Yellow Sea. *J. Fish. China* 35, 429–437.

Yan, D. B., Xia, P. H., Song, X., Lin, T., and Cao, H. P. (2019). Community structure and functional diversity of epiphytic bacteria and planktonic bacteria on submerged macrophytes in Caohai Lake, southwest of China. *Ann. Microbiol.* 69, 933–944. doi: 10.1007/s13213-019-01485-4

Yang, B. Y., Wang, C. C., Hu, H. H., Tu, Y. Q., Han, D., Zhu, X. M., et al. (2015). Repeated handling compromises the immune suppression and improves the

disease resistance in overwintering channel catfish (*Ictalurus punctatus*). *Fish Shellfish Immunol.* 47, 418–428. doi: 10.1016/j.fsi.2015.09.010

Yang, W., Zhu, J., Zheng, C., Qiu, H., Zheng, Z., and Lu, K. (2018). Succession of bacterioplankton community in intensive shrimp (*Litopenaeus vannamei*) aquaculture systems. *Aquaculture* 497, 200–213. doi: 10.1016/j.aquaculture.2018.07.053

Zhang, J. Y., Liu, Y. X., Zhang, N., Hu, B., Jin, T., Xu, H. R., et al. (2019). NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat. Biotechnol.* 37, 676–684. doi: 10.1038/s41587-019-0104-4

Zhou, J., and Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* 81:e00002-17. doi: 10.1128/MMBR.00002-17

Zhou, J. Z., Liu, W. Z., Deng, Y., Jiang, Y. H., Xue, K., He, Z. L., et al. (2013). Stochastic assembly leads to alternative communities with distinct functions in a bioreactor microbial community. *mBio* 4:e00584-12. doi: 10.1128/mBio.00584-12

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 September 2020
doi: 10.3389/fmicb.2020.542372

Seasonal and Geographical Transitions in Eukaryotic Phytoplankton Community Structure in the Atlantic and Pacific Oceans

Chang Jae Choi[1,2†], Valeria Jimenez[2†], David M. Needham[1,2], Camille Poirier[1,2], Charles Bachy[1,2], Harriet Alexander[2,3], Susanne Wilken[2,4], Francisco P. Chavez[2], Sebastian Sudek[2], Stephen J. Giovannoni[5] and Alexandra Z. Worden[1,2]*

[1] Ocean EcoSystems Biology Unit, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany, [2] Monterey Bay Aquarium Research Institute, Moss Landing, CA, United States, [3] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, United States, [4] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands, [5] Department of Microbiology, Oregon State University, Corvallis, OR, United States

OPEN ACCESS

Edited by:
Julie LaRoche,
Dalhousie University, Canada

Reviewed by:
Connie Lovejoy,
Laval University, Canada
Lisa R. Moore,
Macquarie University, Australia

***Correspondence:**
Alexandra Z. Worden
azworden@geomar.de

†Present address:
Chang Jae Choi,
Marine Science Institute,
The University of Texas at Austin,
Port Aransas, TX, United States
Valeria Jimenez,
Sorbonne Université, CNRS, Station Biologique de Roscoff, Roscoff, France

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: 12 March 2020
Accepted: 17 August 2020
Published: 30 September 2020

Citation:
Choi CJ, Jimenez V, Needham DM, Poirier C, Bachy C, Alexander H, Wilken S, Chavez FP, Sudek S, Giovannoni SJ and Worden AZ (2020) Seasonal and Geographical Transitions in Eukaryotic Phytoplankton Community Structure in the Atlantic and Pacific Oceans. Front. Microbiol. 11:542372. doi: 10.3389/fmicb.2020.542372

Much is known about how broad eukaryotic phytoplankton groups vary according to nutrient availability in marine ecosystems. However, genus- and species-level dynamics are generally unknown, although important given that adaptation and acclimation processes differentiate at these levels. We examined phytoplankton communities across seasonal cycles in the North Atlantic (BATS) and under different trophic conditions in the eastern North Pacific (ENP), using phylogenetic classification of plastid-encoded 16S rRNA amplicon sequence variants (ASVs) and other methodologies, including flow cytometric cell sorting. Prasinophytes dominated eukaryotic phytoplankton amplicons during the nutrient-rich deep-mixing winter period at BATS. During stratification ('summer') uncultured dictyochophytes formed ~35 ± 10% of all surface plastid amplicons and dominated those from stramenopile algae, whereas diatoms showed only minor, ephemeral contributions over the entire year. Uncultured dictyochophytes also comprised a major fraction of plastid amplicons in the oligotrophic ENP. Phylogenetic reconstructions of near-full length 16S rRNA sequences established 11 uncultured Dictyochophyte Environmental Clades (DEC). DEC-I and DEC-VI dominated surface dictyochophytes under stratification at BATS and in the ENP, and DEC-IV was also important in the latter. Additionally, although less common at BATS, *Florenciella*-related clades (FC) were prominent at depth in the ENP. In both ecosystems, pelagophytes contributed notably at depth, with PEC-VIII (Pelagophyte Environmental Clade) and (cultured) *Pelagomonas calceolata* being most important. Q-PCR confirmed the near absence of *P. calceolata* at the surface of the same oligotrophic sites where it reached ~1,500 18S rRNA gene copies ml^{-1} at the DCM. To further characterize phytoplankton present in our samples, we performed staining and at-sea single-cell sorting experiments. Sequencing results from these indicated several uncultured dictyochophyte clades are comprised of predatory mixotrophs. From an evolutionary perspective, these cells showed both conserved and unique features in the chloroplast genome. In ENP metatranscriptomes we observed high expression of

multiple chloroplast genes as well as expression of a selfish element (group II intron) in the *psaA* gene. Comparative analyses across the Pacific and Atlantic sites support the conclusion that predatory dictyochophytes thrive under low nutrient conditions. The observations that several uncultured dictyochophyte lineages are seemingly capable of photosynthesis and predation, raises questions about potential shifts in phytoplankton trophic roles associated with seasonality and long-term ocean change.

Keywords: dictyochophytes, phytoplankton diversity, time-series, single-cell genomics, chloroplast genome

INTRODUCTION

Open ocean ecosystems undergo seasonal changes that influence water column vertical structure and productivity, and these ecosystems are predicted to expand under future ocean conditions (Flombaum et al., 2013; Worden et al., 2015). The Bermuda Atlantic Time-series Study (BATS) is located in a seasonally oligotrophic subtropical gyre in the Sargasso Sea (Carlson et al., 1994; Steinberg et al., 2001; Lomas et al., 2013). Pronounced seasonality at BATS is reflected in winter deep-mixing and strong summer thermal stratification, making it an exceptional site for studying transitions in plankton communities associated with warming oceans and declining productivity (Carlson et al., 1994; Steinberg et al., 2001; Lomas et al., 2013). In average years at BATS, deep vertical mixing entrains nutrients into the photic zone during the winter months, supporting relatively high primary productivity in spring, followed by strong thermal stratification and lower summertime productivity. The seasonal dynamics of phytoplankton communities track these periods, transitioning from eukaryotic phytoplankton dominance in the winter and spring, to cyanobacterial dominance in summer and fall, based on high-performance liquid chromatography (HPLC) analysis of pigments and numerical abundance by flow cytometry (Durand et al., 2001; Lomas et al., 2010). Further, although eukaryotic phytoplankton number less than cyanobacteria in this subtropical gyre, they often comprise a significant proportion of phytoplankton biomass (Goericke, 1998; Durand et al., 2001; Cuvelier et al., 2010; Lomas et al., 2010).

While cyanobacterial communities have been well studied at BATS (Malmstrom et al., 2010), characterization of eukaryotic phytoplankton here and at other oligotrophic ocean settings has mainly focused on broad taxonomic categories. At BATS, an early feature of the winter/spring cycle of phytoplankton turnover is a prasinophyte bloom, based on terminal restriction fragment length polymorphism (T-RFLP) analysis and limited qPCR data (Treusch et al., 2012), while afterwards haptophytes and pelagophytes appear to rise in importance based on HPLC pigment analyses and limited molecular data (Goericke, 1998; Cuvelier et al., 2010). Apart from the haptophytes (Cuvelier et al., 2010), the specific taxa that comprise eukaryotic phytoplankton communities at BATS have not been characterized using high throughput molecular marker gene sequencing approaches that help to identify patterns at finer taxonomic scales.

Global scale efforts like TARA Oceans (de Vargas et al., 2015) have found a diversity of stramenopiles across the oceans. Stramenopiles include algae that are considered important in higher latitude and coastal systems, such as diatoms, which appear to bloom only rarely at BATS (Steinberg et al., 2001; Lomas et al., 2010). Apart from diatoms, other relatively well studied stramenopile algal classes include pelagophytes (Andersen et al., 1993, 1996) and chrysophytes (Hibberd, 1976; Andersen et al., 1999), and lesser known ones include the bolidophytes (Guillou et al., 1999) and dictyochophytes (Henriksen et al., 1993; Andersen, 2004). The ecology of the former two lineages has been studied with multiple methods in various ocean regions, including BATS. These studies utilized approaches such as HPLC and T-RFLP (Andersen et al., 1996; Goericke, 1998; Steinberg et al., 2001; Treusch et al., 2012). However, one serious issue is that some stramenopile classes, like the dictyochophytes, cannot be discriminated from diatoms and pelagophytes using HPLC (Guillou et al., 1999; Daugbjerg and Henriksen, 2001). Further, the paucity of sequence data from a broad range of characterized stramenopiles or taxonomically verified environmental sequences has restricted studies of diversity and distributions. Hence, although distributions of pelagophytes (and some other stramenopiles) have been described at BATS based on HPLC data (Steinberg et al., 2001), the seasonal patterns and depth distributions for specific pelagophyte taxa are as yet unknown. In the Mediterranean Sea and South Pacific Ocean, analyses of the plastid-derived *psbA* gene and rRNA genes have indicated the presence of uncultured pelagophytes (Man-Aharonovich et al., 2010; Shi et al., 2011), while targeted metagenomic approaches connected to traditional metagenomic analyses indicate that pelagophytes highly similar to *Pelagomonas calceolata* are distributed across subtropical surface ocean waters (Worden et al., 2012). Further, in the Pacific Ocean this species can dominate 0.1-20 μm size fractionated samples from the subsurface chlorophyll maximum (SCM), based on metagenomics analyses in coastal California (Dupont et al., 2015).

Perhaps one of the most enigmatic stramenopile algal classes is the dictyochophytes. They have few cultured representatives, little biogeographical data apart from TARA (de Vargas et al., 2015), and little seasonal data. This group contains the silicoflagellates, which can be microscopically identified by their distinctive siliceous skeletons, and have been studied mainly in coastal northern ecosystems and sediments or as paleoecological markers (Henriksen et al., 1993; Rigual-Hernandez et al., 2016; van de Poll et al., 2018). Indeed, there are few oceanic time-series analyses that tease apart the contributions of diatoms, chrysophytes, and pelagophytes, and more newly discovered

groups like bolidophytes (Goericke, 1998; Steinberg et al., 2001; Lomas et al., 2013), or those with few cultured representatives, like the dictyochophytes. Perhaps the best time-series molecular characterization has been performed at the San Pedro Ocean Time-series (SPOT) in the eastern North Pacific (ENP) off coastal USA where several dictyochophyte Amplicon Sequence Variants (ASVs) were reported as being common (Needham et al., 2018). The ENP differs dramatically from BATS, in part due to the paucity of phosphate in the latter as compared to the former (Coleman and Chisholm, 2010). In addition to SPOT, eukaryotic phytoplankton dynamics have been extensively studied in the ENP using classical approaches such as HPLC and microscopy as part of the California Cooperative Oceanic Fisheries Investigation (CalCOFI), which samples multiple transect 'lines' that run perpendicular to the California, United States coast on an approximately bimonthly basis (Collins et al., 2003; McClatchie, 2014). These transect lines typically cross the coastal zone, traverse productive upwelling waters and continue outward to the edge of the North Pacific Subtropical Gyre. Along the ENP's CalCOFI Line-67, which extends from Monterey Bay to 800 km offshore, cyanobacterial diversity has been relatively well described at the level of molecular diversity (Sudek et al., 2015), and quantitative data exists for some algal groups, based on methods such as qPCR (Simmons et al., 2016; Kolody et al., 2019), HPLC and microscopy (Paerl et al., 2011; Chavez et al., 2017). However, less is known about the general molecular diversity of eukaryotic algae in the different trophic regimes of the ENP.

Here, we examined eukaryotic phytoplankton community structure in the Sargasso Sea over seasonal cycles and compared this data to algal distributions along the ENP Line-67. V1-V2 16S rRNA gene amplicon data from BATS photic-zone profiles (Treusch et al., 2009) were analyzed using phylogenetic approaches (Matsen et al., 2010; Choi et al., 2017). At BATS, we found a dominance of non-diatom stramenopile-derived amplicons in the multi-year dataset. This prompted us to develop well-curated, full-length 16S reference alignments and phylogenetic reconstructions for the most highly represented stramenopile groups, specifically the pelagophytes and dictyochophytes. The reconstructions provided new insights into clade diversity within these groups, as well as spatiotemporal patterns of fine-scale phytoplankton diversity. Comparison to community structure at Line-67 mesotrophic and oligotrophic sites highlighted similarities in phytoplankton community composition, with parallels between the influences of the seasonal cycle of nutrient availability at BATS and variation in nutrients along Line-67. To evaluate amplicon-based relative abundance patterns, a qPCR primer probe set was developed and implemented to enumerate *Pelagomonas calceolata*, which was identified as being the most abundant pelagophyte in amplicon data. Finally, because a prior Pacific Ocean 18S rRNA stable isotope probing experiment indicated that bolidophytes and a *Florenciella*-like dictyochophyte may act as mixotrophic predators (Frias-Lopez et al., 2009), we also performed single-cell food-vacuole-staining / chlorophyll-based cell sorting (Wilken et al., 2019) in order to capture potential mixotrophs among the wild phytoplankton taxa that were abundant in our flow cytometric histograms. This led to sequencing of uncultured dictyochophytes that appear to be predatory mixotrophs (present in the field sort experiments) and rendered the complete chloroplast genome of one such species. Collectively, the similarities observed between the two ecosystems, BATS and the ENP, were found to be especially strong for predatory mixotrophic stramenopiles in surface oligotrophic environments, based on trends that emerged when sequence data was parsed at high taxonomic resolution and with the aid of single-cell sort data from wild dictyochophytes.

RESULTS

Eukaryotic Phytoplankton in the Western North Atlantic

A spring phytoplankton bloom and subsequent thermal stratification during summer were evident in chl *a* data (**Figures 1A,B**; Lomas et al., 2013). At BATS, the highest nutrient concentrations in surface waters occur during the period of Deep Mixing (DM), which helps initiate the bloom, while nutrient concentrations are low during the summer months (Steinberg et al., 2001). During summer, nitrate and phosphate concentrations in most samples were below detection limits, which are 30 and 10 nM, respectively. During month 0, defined as the month when deepest mixing occurs, eukaryotic phytoplankton contributed 56 ± 17% (**Figure 1C**) and 59 ± 24% (**Supplementary Figure S1A**) of total phytoplankton amplicons at the surface and at depth, respectively. More generally, in winter/spring (−1 to +5), eukaryotes contributed a higher proportion (44 ± 19%) of photosynthetic amplicons relative to months +6 to +10 (12 ± 8%, $p < 0.0001$, two-tailed Mann-Whitney *U*-test). During the latter strongly stratified summer months, *Prochlorococcus* was abundant (**Supplementary Table S1**), as observed in previous studies based on cell counts (Durand et al., 2001). Thus, eukaryotic phytoplankton amplicon contributions rivaled those of cyanobacteria (*Synechococcus* and *Prochlorococcus* together; **Figure 1C**) during the winter/spring months, but were much lower during the period of most intense stratification, as reported previously (Treusch et al., 2012).

Average contributions of broad eukaryotic phytoplankton groups, i.e., stramenopiles, haptophytes, prasinophytes (Viridiplantae) and cryptophytes, to total plastid amplicons varied over the year, especially at the surface (**Figure 1D** and **Supplementary Figure S1B**). To identify patterns related to seasonal changes, we performed hierarchical clustering on the relative abundances of groups in surface amplicon data. This resulted in three statistically supported groupings, 'Deep Mixing (DM)', 'Early Stratification' and 'Late Stratification', comprised of month −1 and 0, month +1 to +5 and month +6 to +10, respectively (**Figure 1E**). During DM, contributions were relatively even throughout the water column. Viridiplantae algae (largely prasinophytes) formed 55 ± 19% (surface) and 53 ± 25% (at depth) of the plastid amplicons and stramenopiles formed 23 ± 15% (surface) and 25 ± 17% (at depth). Haptophytes showed lower but more consistent relative contributions throughout the year, forming 23 ± 8% (23 ± 8%, surface; 22 ± 9%, depth) of plastid amplicons. During Early Stratification

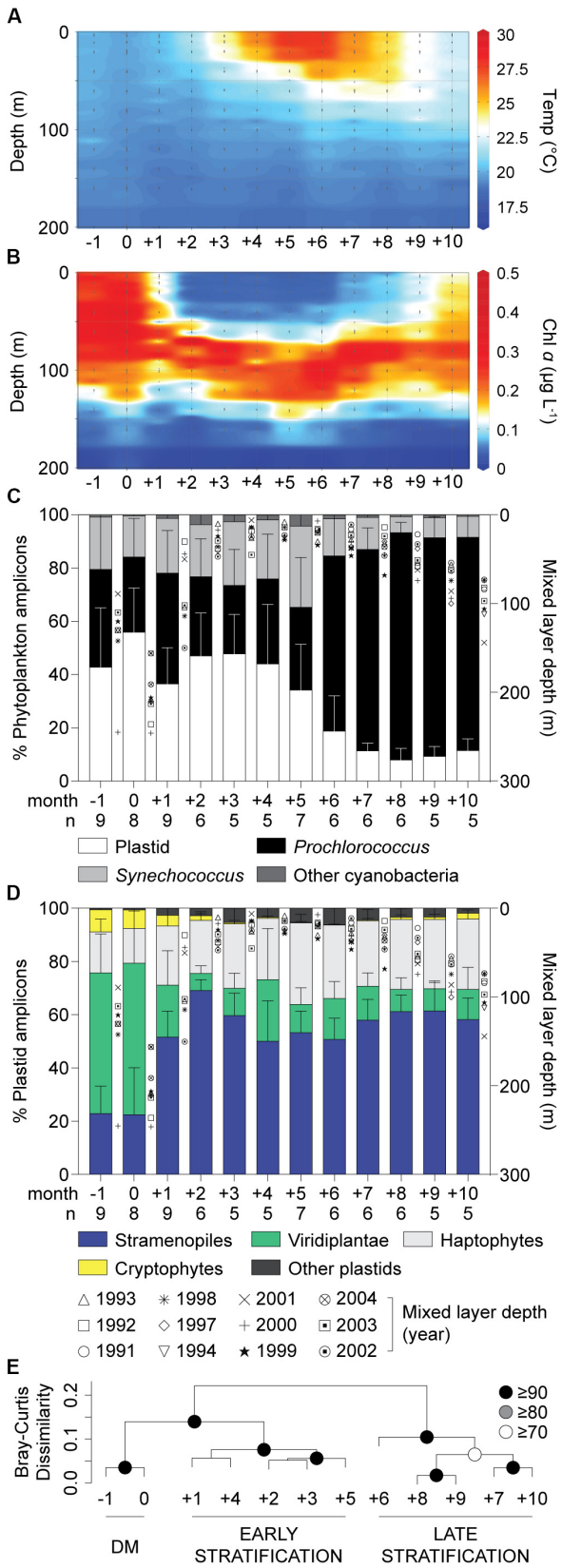


FIGURE 1 | Continued

FIGURE 1 | Characteristics of the water column and phytoplankton distributions in the northwestern Sargasso Sea. **(A)** Mean temperature and **(B)** chlorophyll *a* concentrations show the oceanographic details in this study site. **(C)** The relative abundance of overall phytoplankton 16S V1-V2 amplicons expressed as a percentage of amplicons phylogenetically assigned to plastids, *Prochlorococcus*, *Synechococcus* and other cyanobacterial groups and **(D)** major eukaryotic phytoplankton lineages illustrate broad spatiotemporal dynamics of phytoplankton diversity and distribution in surface waters. Integrated monthly data from 1991 to 2004 at the BATS site after adjusting to the month of the maximum mixed layer depth (month 0). **(E)** Hierarchical clustering analysis with Bray-Curtis dissimilarity based on the relative abundance of plastid amplicons after assignment to broad taxonomic groups alongside cyanobacterial amplicons in surface waters. Approximately unbiased (AU) probability values based on multiscale bootstrap resampling (10,000 replicates) were calculated and expressed as *p*-values (%). The annual transition from the deep mixing (DM) period to early (Early Stratification) and late stratified (Late Stratification) periods is indicated. Error bars represent the standard deviation. The number of samples used (with each having all data types) for developing unweighted means and pooled standard deviation is indicated by *n*.

(+1 to +5) the community shifted such that stramenopiles comprised about half of plastid amplicons at the surface (57 ± 12%) and developing DCM (53 ± 7%), and had similar contributions during Late Stratification (58 ± 9%, surface; 50 ± 10%, DCM; **Figure 1D**, **Supplementary Figure S1B**, **Supplementary Table S2**). These variations raised questions about whether the stramenopile contributions came from a few taxa, with changes in relative amplicon abundances being a byproduct of dynamics of other phytoplankton groups, or potentially shifts within stramenopile community structure. Therefore, we developed reference phylogenetic trees based on near full-length 16S rRNA gene sequences to further resolve stramenopile taxa.

Stramenopiles at BATS

Eleven photosynthetic classes were delineated in the near-full length 16S stramenopile plastid reference tree developed herein (**Supplementary Figure S2**). All 10 known photosynthetic stramenopile classes (Andersen, 2004) were recovered with statistical support alongside a new clade, containing only environmental sequences in a supported position adjacent to bolidophytes, that likely represents an unrecognized class.

Dictyochophytes dominated surface stramenopile amplicons during both the Early and Late Stratification periods, but showed lower, more variable percentages during DM (**Figure 2A**). Pelagophytes were also notable during these periods and other groups formed a smaller, more variable proportion of stramenopile amplicons. Hierarchical clustering using stramenopile classes alone delineated the DM clearly, as well as a transition into and out of the stratified period, bringing together months +1, +2, +8, +9 and +10 (termed Transitional), separate from months +3 to +7 (Maximum Stratified; **Supplementary Figure S3**). This grouping gained modest statistical support using pvclust, but not by similarity profile analysis (SIMPROF, $p > 0.05$). Thus, a slightly different seasonal organization of the stramenopile community (albeit with weak statistical support) was observed versus that based on all phytoplankton

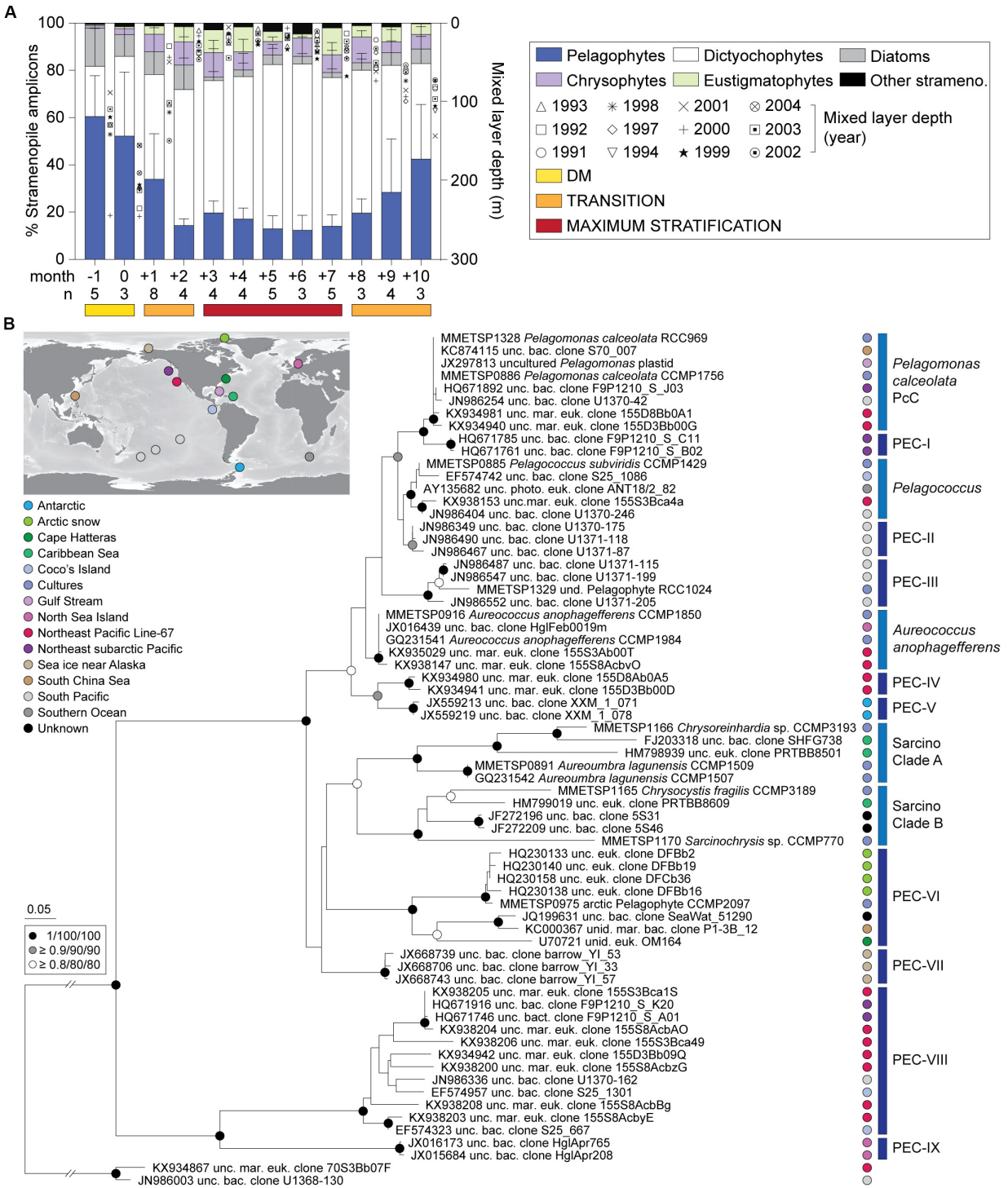


FIGURE 2 | Photosynthetic stramenopile class distributions in the northwestern Sargasso Sea and reference trees for finer-scale analyses. **(A)** Photosynthetic stramenopile distributions expressed as a percentage of amplicons phylogenetically assigned to stramenopile classes at BATS surface (0.3–7.6 m). Error bars represent the standard deviation and 'n' indicates the number of samples used for the analysis. Phylogenetic reference trees were developed to then resolve clades within the **(B)** pelagophytes and dictyochophytes (**Supplementary Figure S5**) using near full-length 16S rRNA gene sequences. For pelagophytes, sequences came from nine cultured species, 56 environmental clones and two dictyochophyte sequences as an outgroup. The tree was constructed with Maximum Likelihood inference (RAxML) under the gamma corrected GTR model of evolution with 1,000 bootstrap replicates. Additional phylogenetic reconstructions were performed with PhyML and MrBayes and the node statistical supports are indicated. Five known clades were resolved and previously unrecognized nine additional clades were identified Pelagophyte Environmental Clades I-IX (PEC-I to PEC-IX). These were primarily comprised of environmental sequences. Colors indicate oceanic region of origin.

amplicons (**Figure 1E**). The stramenopile-based sample clustering corresponded to distinct temperature ranges during DM (20 ± 1°C), Transitional (22 ± 2°C) and Maximum Stratified (26 ± 2°C; $p < 0.0001$, one-way ANOVA with Tukey's post-hoc test; **Supplementary Figure S3**).

Diatom relative abundances, as a fraction of stramenopile amplicons, were low and highly variable at the surface: 13 ± 13%, DM; 7 ± 10%, Transitional; and 3 ± 3%, Maximum Stratified period (**Figure 2A**) as well as at depth (**Supplementary Figure S1C**, **Supplementary Table S3**, and **Supplementary Data S1**). Other groups were typically < 5% of stramenopile amplicons, except chrysophytes (surface; 8 ± 5%, Transitional; 8 ± 5%, Maximum Stratified) and eustigmatophytes (surface; 5 ± 6%, Transitional; 7 ± 7%, Maximum Stratified). Overall, among all stramenopiles recovered, dictyochophytes and pelagophytes dominated amplicon contributions over all three seasonally influenced periods.

Dictyochophytes were most important in surface waters during low nutrient periods, specifically Transitional (51 ± 19%) and Maximum Stratified (64 ± 12%) (**Figure 2A** and **Supplementary Table S3**). They formed lower and more variable contributions during DM (28 ± 26%) and their contributions varied, similarly, at depth (40–80 m) during DM and in the DCM as the year progressed (**Supplementary Figure S1C**). Pelagophyte surface contributions trended opposite to dictyochophytes forming 56 ± 23% (DM) of stramenopile amplicons and dropping to 28 ± 20% and 15 ± 6% in the Transitional and Maximum Stratified periods, respectively ($p < 0.0001$, two-tailed Mann-Whitney *U*-test). Pelagophyte contributions at depth were more consistent year round (54 ± 23%; **Supplementary Figure S1C**).

These overall trends were upheld at multiple depths in the subset of samples having sufficient sequencing depth for plastid analysis (**Supplementary Figure S4A**). During DM, pelagophytes dominated stramenopile amplicons throughout the water column (63 ± 16%). The exception was from single samples (160 m, −1; 120 m, 0) when diatoms (e.g., *Chaetoceros*, *Thalassiosira* and unidentified species) were nearly as high, although low overall plastid amplicon numbers suggest few eukaryotic phytoplankton cells were present. Dictyochophytes were the dominant stramenopiles throughout the upper 80 m during the Transitional (43 ± 27%) and Maximum Stratified (59 ± 23%) periods, but generally decreased alongside increased pelagophyte relative abundances below 80 m.

An Expanded Phylogeny of Pelagophytes and Dictyochophytes

There is a growing literature on the ecological importance of pelagophytes in the ocean (Man-Aharonovich et al., 2010; Worden et al., 2012; Dupont et al., 2015), however, less is known about dictyochophyte distributions and diversity. We examined both these groups at higher phylogenetic resolution. Phylogenetic reconstruction of dictyochophyte 16S rRNA gene sequences delineated 20 clades, including 11 not previously reported and termed here Dictyochophyte Environmental Clades, DEC-I to DEC-VIII (**Supplementary Figure S5**) and *Florenciella* Clades FC-I to FC-III. All of these clades lack cultured representatives, but FC have high nucleotide identity to *Florenciella parvula,* which forms a separate clade (termed here FpC, *Florenciella parvula* Clade). DEC and FC contained sequences from surface ocean samples collected in the eastern North Pacific, South Pacific, South China Sea, and other habitats, such as the microbial mats from the hypersaline region in Guerrero Negro, Mexico (DEC-V) and the Puerto Rico Trench at 6000 m (DEC-VIII), indicating they are widely distributed and potentially are exported to deep waters.

The reconstruction of pelagophyte 16S rRNA gene sequences resolved the five known pelagophyte clades (Wetherbee et al., 2015), as well as nine undescribed pelagophyte clades, termed here Pelagophyte Environmental Clades, PEC-I to PEC-IX (**Figure 2B**). Some were solely comprised of sequences from a specific region/study, such as PEC-I (northeast subarctic Pacific Ocean), PEC-II (South Pacific), PEC-IV (ENP), PEC-V (Antarctic), PEC-VII (sea ice near Alaska) and PEC-IX (North Sea near Helgoland).

Diversity and Seasonality of Pelagophytes at BATS

We detected pelagophyte amplicons at BATS belonging to 13 of the 14 clades, including all the uncultured clades formed of full-length sequences from a single region (**Figures 2B**, **3A**). Most clades contributed < 1% of the total pelagophyte amplicons, apart from PEC-V (the one clade not detected in BATS data). By contrast, *P. calceolata* Clade (hereafter PcC) formed 56 ± 23% (surface) and 62 ± 14% (at depth) of pelagophyte amplicons during DM and at the DCM later in the year (59 ± 38%). PcC were low and variable at the surface during the Transitional period (22 ± 21%) and rare during the Maximum Stratified period (1 ± 5%).

Pelagophyte Environmental Clades-VIII contributed relatively high percentages of pelagophyte amplicons during DM at the surface (41 ± 23%) and more so during stratified periods (70 ± 19%, Transitional; 80 ± 16%, Maximum Stratified at the surface; **Figure 3A**). These trends were upheld when multiple depths were examined for months with deeper amplicon data (−1, 0, +4 and +6; **Supplementary Figure S4B** and **Supplementary Table S4**). Additionally, PEC-III and a group of basal sequences (BP-I) were also relatively abundant in Maximum Stratified surface waters (**Figure 3A**). Overall, PcC and uncultured PEC-VIII had the highest relative abundances among pelagophytes at BATS, but BP-I and PEC-III, also rose in prominence during intense stratification (**Figure 3A**).

Uncultured Dictyochophytes in the Sargasso Sea

A diverse set of uncultured dictyochophyte clades exhibited seasonal variations. Dictyochophytes formed 35 ± 10% of all plastid amplicons in BATS Maximum Stratified surface waters (**Figure 2A**). Assignment of amplicons to distinct clades showed that uncultured groups DEC-I and DEC-VI were prevalent during this period while FC-III appeared more important during DM throughout the water column (**Figure 3B**). DEC-I

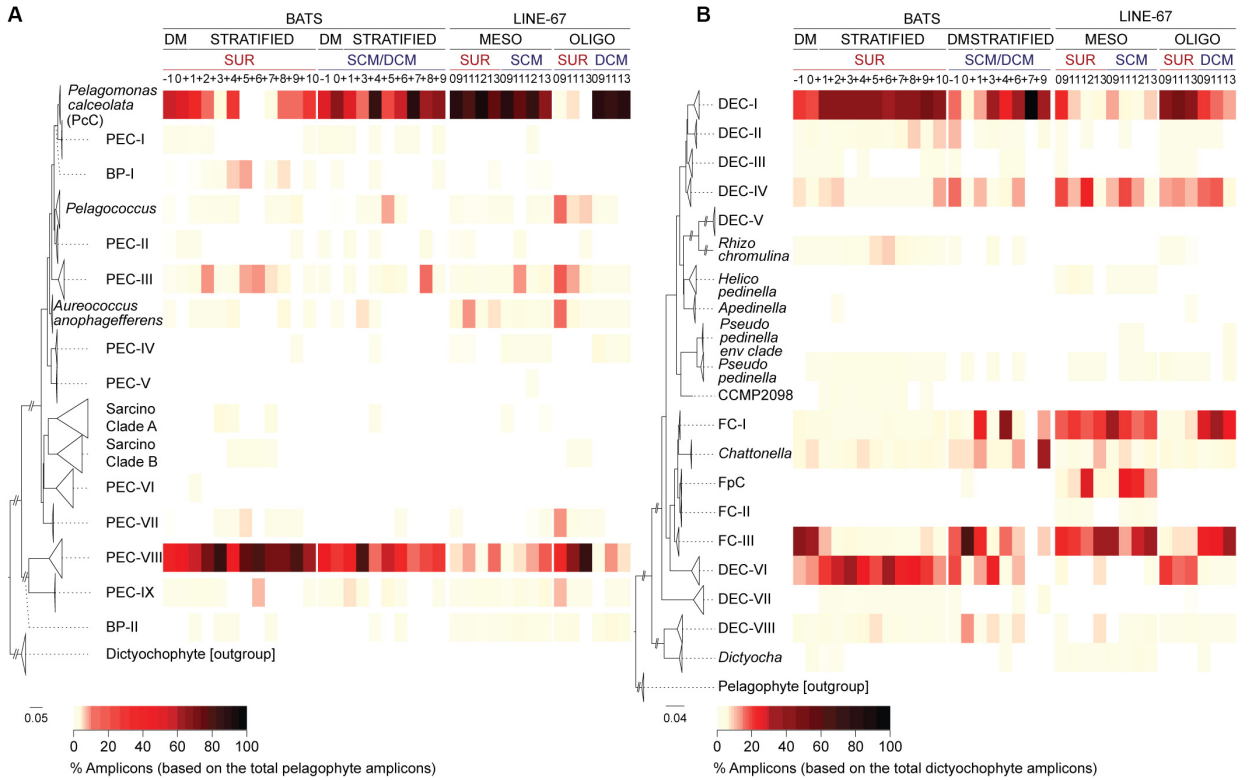


FIGURE 3 | Diversity and distributions of newly resolved dictyochophyte and pelagophyte clades at BATS and in the ENP. High-taxonomic resolution of **(A)** pelagophyte and **(B)** dictyochophyte distributions at the surface (SUR) and depth (SCM/DCM) at BATS and ENP. Reference trees to the left of each panel indicate clades. Heatmaps show the percentage of amplicons relative to those of pelagophytes **(A)** and dictyochophytes **(B)** during each BATS month or ENP samples (vertical columns; see **Figure 4** for additional details on ENP samples). Bootstrap support and phylogenetic inferences are as detailed in **Figure 2B** and **Supplementary Figure S5**. ENP samples are partitioned based on standing stock nutrient concentrations which were considered mesotrophic (MESO) or oligotrophic (OLIGO) (**Supplementary Table S5**).

contributed the most dictyochophyte amplicons at the surface during Transitional (43 ± 11%) and Maximum Stratified periods (39 ± 11%), and at the DCM during the latter (49 ± 25%). DEC-VI was also notable at the surface during Transitional (23 ± 14%) and Maximum Stratified (26 ± 15%), but less important at depth. Corresponding FC-III contributions were minor except during DM, when they formed 38 ± 27% (surface) and 39 ± 16% (at depth) of total dictyochophyte amplicons.

These patterns were generally maintained in 19 vertical profiles examined. During the DM period FC-III formed the greatest percentage of dictyochophyte amplicons throughout the photic zone (∼140 m; **Supplementary Figure S4C**). In contrast, during the Transitional and Maximum Stratified periods, DEC-I had the highest relative contributions at the surface, but at the DCM (ranging from 40–120 m) it was rivaled by other clades. Thus, uncultured environmental groups comprised most of the dictyochophyte amplicons at BATS (e.g., in surface waters 92 ± 4%, DM; 92 ± 5%, Transitional; 85 ± 4%, Maximum Stratified). Further, they contributed a large proportion of the total eukaryotic phytoplankton amplicons in surface waters, especially during Transitional (30 ± 14%) and Maximum Stratified (30 ± 10%) periods. This indicates that uncultured dictyochophytes form a large portion of the eukaryotic phytoplankton that persist at BATS during oligotrophic periods, when *Prochlorococcus* typically dominates cell counts.

Comparisons With ENP Communities

The Sargasso Sea results indicated the potential importance of diverse uncultured dictyochophyte clades under oligotrophic conditions and a phylogenetically narrower range of pelagophyte clades at the nutricline of stratified water columns. We therefore next examined ENP communities at two well characterized (Monier et al., 2012; Sudek et al., 2015; Simmons et al., 2016; Limardo et al., 2017) stratified oligotrophic sites (termed "OLIGO"), specifically Station 67-135 and 67-155, based on 3 surface and 3 DCM samples (**Figures 4A,B** and **Supplementary Table S5**). We also characterized samples from a more mesotrophic region (termed MESO), based on four surface and four SCM samples collected at Stations 67-60drift and 67-70. Here, we distinguish between periods or sites exhibiting a deep chlorophyll maximum (DCM), as seen in oligotrophic gyres, typically located between 80 to 130 m, depending on the season and system (Karl and Church, 2014), and subsurface chlorophyll maxima (SCM), that often occur in the upper 40 m of more nutrient-rich sites.

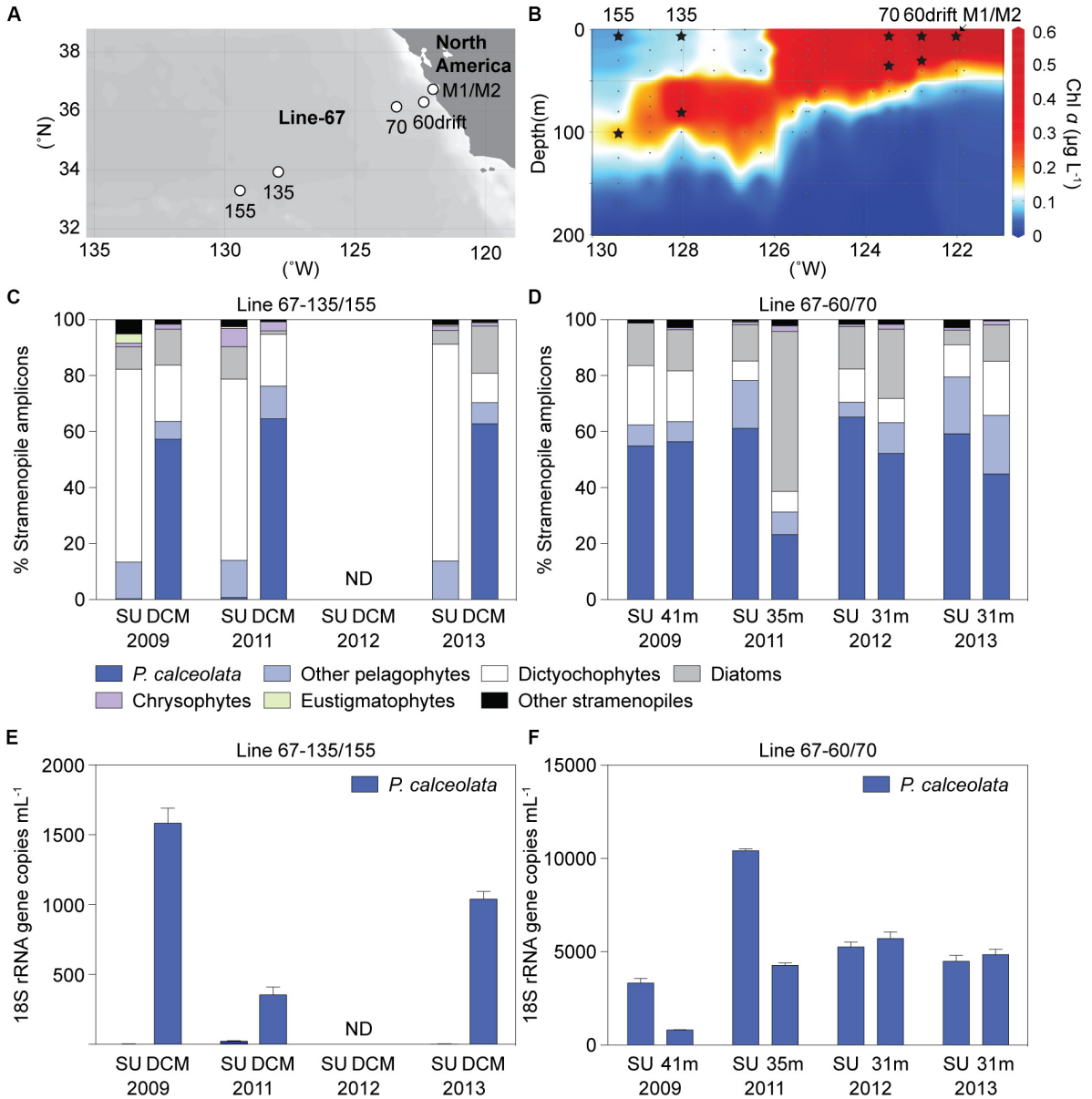


FIGURE 4 | ENP **(A)** stations and **(B)** chlorophyll *a* depth profile in 2009 with sampling for DNA indicated by stars (Stations 67-60drift and 67-70, mesotrophic sites; Stations 67-135 and 67-155, oligotrophic sites). Relative abundance of stramenopile V1-V2 16S rRNA gene amplicons assigned to the *Pelagomonas calceolata* Clade or other groups in Line-67 samples. **(C)** Taxon distributions at sites with well stratified water columns where nitrate was 0.02 and 0.08 μM (plus 1 sample < 30 nM detection limit) at the surface and 0.05 to 0.3 μM (0.17 ± 0.13 μM, $n = 3$) in the deep chlorophyll maximum (DCM). **(D)** Distributions in surface (SU, 2–14 m) or near the base of the photic zone (31-41 m) at mesotrophic sites. **(E,F)** *P. calceolata* 18S rRNA gene copies ml^{-1} enumerated by qPCR in the same samples as analyzed for panels C, D. Note the different scale on the *y*-axis in **(E)** and **(F)**. ND: no data. Error bars represent the standard deviation of three technical replicates.

Oligotrophic nitrate concentrations at the surface ranged from < 30 nM (detection limit) to 80 nM, and from 50 to 300 nM ($n = 3$) at the DCM, while phosphate was much higher (e.g., > 340 nM, **Supplementary Table S5**). Similar to BATS, *Prochlorococcus* dominated photosynthetic amplicons at ENP OLIGO sites, forming 92 ± 2% and 84 ± 6% at the surface and DCM phytoplankton amplicons, respectively (**Supplementary Figure S6A**). Eukaryotic phytoplankton amplicon contributions (7 ± 2%, surface; 15 ± 6%, DCM) followed and were higher than *Synechococcus* (0.9 ± 0.0%, surface; 0.2 ± 0.1%, DCM). Stramenopiles comprised 69 ± 5% and 57 ± 7% of plastid amplicons, at the surface and DCM, respectively, appearing to dominate eukaryotic phytoplankton amplicons in the oligotrophic stations (**Supplementary Figure S6C**).

At ENP MESO stations, eukaryotic phytoplankton amplicons were more abundant at the surface (59 ± 18%) and SCM (65 ± 21%) than cyanobacteria (**Supplementary Figure S6B**). In the four MESO SCM samples stramenopiles averaged 45 ± 8%

of total plastid amplicons. During the more nutrient rich ENP MESO sample periods (2011 and 2012, 2.05 ± 0.66 μM; surface NO_3^-) stramenopile contributions were 22 ± 9% in surface waters, similar to BATS during the DM period (23 ± 15%). In 2009 and 2013, when ENP MESO surface nitrate was lower (0.33 ± 0.10 μM), stramenopile contributions were stronger, more akin to oligotrophic waters (**Supplementary Figure S6D**).

Among stramenopiles, dictyochophyte amplicons were the most highly represented in OLIGO surface samples (70 ± 6%). They formed 13 ± 6% of stramenopile amplicons at the MESO surface and SCM, similar to the OLIGO DCM (16 ± 5%) (**Figures 4C,D**). Pelagophytes had the highest relative abundances within stramenopile amplicons in OLIGO DCM samples (70 ± 6%). They formed 73 ± 8% and 56 ± 17% of stramenopile amplicons at the MESO surface and SCM, respectively, while much lower in OLIGO surface samples (14 ± 0%).

ENP Dictyochophyte and Pelagophyte Diversity Patterns

We next evaluated the community structure of dictyochophytes and pelagophytes in the ENP. Among dictyochophytes, DEC-I (43 ± 3%) and DEC-VI (19 ± 3%) were notable in OLIGO surface samples (**Figure 3B**) as observed in BATS surface samples during the Transitional and Maximum Stratified periods. At the ENP SCM and DCM, FC-III and FC-I contributions were higher than either DEC-I and DEC-VI (**Figure 3B**; 29 ± 6%, FC-III; 28 ± 3%, FC-I). FC-III was also prominent in ENP mesotrophic waters, representing 24 ± 8% (surface) and 27 ± 10% (SCM) of total dictyochophyte amplicons (**Figure 3B**). Contributions from other *Florenciella* clades to dictyochophyte amplicons were also notable in ENP MESO samples, such as FC-I (19 ± 2%, surface; 22 ± 6%, SCM) as well as FpC (12 ± 11%, surface; 17 ± 9% SCM) which was not observed in ENP OLIGO samples. The latter two *Florenciella* clades were rare in BATS amplicon data. Another clade that appeared to have greater importance in the ENP relative to BATS was DEC-IV, which formed 15 ± 9% (surface) and 12 ± 6% (SCM) of MESO dictyochophyte amplicons.

With respect to pelagophytes, PcC was the most prominent clade in ENP mesotrophic waters, forming 83 ± 8% (surface) and 78 ± 9% (SCM) of the total pelagophyte amplicons (**Figure 3A**). It represented 24 ± 13% (surface) and 20 ± 8% (SCM) of total plastid amplicons. PEC-VIII formed only 8 ± 6% (surface) and 8 ± 7% (SCM) MESO pelagophyte amplicons, similar to OLIGO DCM (6 ± 3%) levels. PcC contributions to total pelagophyte amplicons were high (88 ± 3%) in OLIGO DCM samples but dropped to 2 ± 3% in surface waters, concurrent with an increase in PEC-VIII (63 ± 26%). Amplicons assigned to the clade containing *A. anophagefferens* and PEC-III were ≤ 10% on average in all ENP OLIGO and MESO samples. Other environmental pelagophyte clades were detected and contributed < 1% or, for *Pelagococcus*, PEC-VII and PEC-IX, averaged < 10% of total pelagophyte amplicons in oligotrophic surface waters.

To further validate results derived from amplicon relative abundance patterns we quantified abundance using qPCR (**Figures 4E,F**). Secondly, having noticed that dictyochophytes reach their highest relative abundances during the stratified conditions, when the oligotroph *Prochlorococcus* exhibits its greatest numerical abundances, we examined the potential trophic modes and evolution of uncultured dictyochophytes for clues to their success in nutrient-deplete systems (Durand et al., 2001; Malmstrom et al., 2010).

Relationship Between Relative Amplicon Abundance and Quantitative Analysis

Our analyses identified previously unrecognized environmental clades and provided insights into dynamics of phytoplankton community structure. However, patterns in relative amplicon abundance are strongly influenced by the interplay between a taxon's own abundance and changing abundances of other taxa. Moreover, differences in gene copy number between different phytoplankton groups remain unconstrained (Demir-Hilton et al., 2011; Limardo et al., 2017). Therefore, using PcC as a case study, we quantified *P. calceolata* abundance in Line-67 samples using an 18S rRNA qPCR primer-probe set developed herein (**Supplementary Tables S6, S7**). *P. calceolata* was selected because of its importance in samples with higher nutrient concentrations at both BATS and the ENP, and in other semi-quantitative studies (Worden et al., 2012; Dupont et al., 2015). The abundance of *P. calceolata* varied considerably in mesotrophic samples over the four years: 5,882 ± 2,855 (surface) and 3,917 ± 1,976 (SCM) 18S rRNA gene copies ml^{-1} (range from 792 to 10,596 18S rRNA gene copies ml^{-1}). It had low abundance in OLIGO surface samples (9 ± 11 copies ml^{-1}; range from below detection to 26 copies ml^{-1}; **Figures 4E,F**), in agreement with amplicon analyses. Its abundance was higher at the OLIGO DCM (992 ± 538 copies ml^{-1}; range 289 to 1,665 copies ml^{-1}), supporting its prominence in DCM amplicon data (**Figures 4C,D**).

Amplicon relative abundance data did sometimes mask differences observed in quantitative abundance data. For example in 2009, surface and SCM mesotrophic samples showed abundances of 3,334 ± 425 and 819 ± 24 18S rRNA gene copies ml^{-1}, respectively (**Figure 4F**; SU vs. 41 m), whereas relative abundance of *P. calceolata* in stramenopile amplicons was almost equal (55%, surface; 56%, SCM) between these samples (**Figure 4D**), and only moderately different relative to all plastid amplicons (40%, surface; 30% SCM). Additionally, while there was little difference in absolute abundance between 2011 and 2012 mesotrophic SCM samples (**Figure 4F**; 35 m in 2011 vs. 31 m in 2012), amplicon data suggested there were large changes (**Figure 4D**), which, in light of the qPCR data, now appear to have been driven by increased abundance of other phytoplankton taxa in the 2011 sample, reducing relative pelagophyte contributions.

Dictyochophyte Dynamics and Activity Expose a Predatory Mixotroph Signature

The distribution patterns of environmental dictyochophyte clades showed greater prominence during conditions of low nutrient availability where predatory mixotrophic strategy might

be advantageous (Edwards, 2019; Wilken et al., 2019). We therefore used single-cell sorting that targeted eukaryotic cells (Cuvelier et al., 2010) with stained food vacuoles to try to recover mixotrophic dictyochophytes in ENP waters (**Figures 4A,B**; Stations M1 and M2; **Supplementary Figure S7A**). Several uncultured dictyochophytes were recovered based on 16S V4 amplicon sequencing (**Supplementary Figure S7B**). Sequencing of single-cell sorts revealed a dictyochophyte representing DEC-IV (DEC-IV sort, 100% nt identity), a *Pseudopedinella*-like dictyochophyte (*Pseudopedinella*-like sort, 97% nt identity), two Pedinellales-like sorts and seven other more divergent dictyochophytes that had < 95% nucleotide identity to reference sequences, based on analysis of the V4 region of the 16S rRNA gene.

The seven more divergent sorted dictyochophytes formed two distinct environmental clades. One, DEC-IX comprised four sorted cells (**Supplementary Figure S7B**) for which we then sequenced the complete 16S rRNA gene in order to better resolve their phylogeny relative to other dictyochophytes. The reconstruction based on near-full length 16S sequences indicates that DEC-IX is basal to multiple other environmental sequences and clades (**Supplementary Figure S8**). The other three more divergent sorted cells grouped together, forming the distinct clade DEC-X (**Supplementary Figure S7B**). DEC-X cells were unlike other sorted cells in having minimal chl *a* fluorescence (i.e., near the baseline in the chl *a* channel), indicating they potentially had non-functional remnant plastids (achlorophyllous), or very low chlorophyll content, and hence likely require prey to grow (**Supplementary Figure S7A**).

We next constructed both full-length 18S rRNA gene and 18S V4 amplicon–based reconstructions for the dictyochophyte sorts (**Supplementary Figure S9**). All but one 18S amplicon were placed in environmental-only clades. The one exception (termed here CCMP2098-like sort 1) had 100% 18S rRNA gene identity to cultured Pedinellales isolates from the Arctic Ocean, i.e., CCMP2098, RCC2301 and RCC2286. The branching position of the CCMP2098 clade differed between the 18S and 16S rRNA gene reconstructions, but it was consistently separate from *Florenciella*. Unfortunately, incomplete taxon sampling of dictyochophyte (especially the lack of 16S and 18S rRNA genes from the same reference organisms) confounds comparisons of the tree topologies at finer scales.

Given the interesting phylogenetic position of the new DEC-IX clade, as well as the paucity of genomic data from dictyochophytes, we next performed metagenomic sequencing on cells that had both chlorophyll fluorescence and food vacuoles (based on LysoTracker and LysoSensor signals; **Supplementary Figure S7A**). We assembled DEC-IX's complete chloroplast genome, which has 100,783 bp (36.9% G+C) and encodes 113 proteins, 28 tRNAs, and an rRNA operon as well as other features (**Figure 5A**). Compared to *P. calceolata* (Worden et al., 2012), which was also abundant in our amplicon data, the DEC-IX chloroplast genome has two fewer protein encoding genes, 6 additional characterized genes (*psaE*, *psbY*, *ycf66*, *rpl12*, *rpl32* and *rps19*), and additional ORFs, ORF1 and ORF2, of unknown function (**Supplementary Data S2**).

The DEC-IX plastid genome also has a group II intron (2923 bp) inserted in the *psaA* gene (**Figure 5C**). The DEC-IX *psaA* gene appears to be intact with no sign of disruption by this intron invasion and was actively expressed in ENP metatranscriptomes (**Figure 5A**). Based on predicted secondary structure, it is a type IIB intron with six predicted RNA domains (DI-DVI) common to group II introns and contains the conserved specific junction sequences at the boundaries of the introns (5′-GUGYG. . .AY-3′) (Lambowitz and Zimmerly, 2011). An IEP (Intron Encoded Protein) of 593 amino acids is encoded within the loop of DIV, containing the characteristic domains for reverse transcriptase (RT), maturase (X), DNA binding (D) domain and the H-N-H endonuclease domain (En). Transcripts from *psaA* and the group II intron were identified in the seven ENP metatranscriptomes analyzed by mapping to the chloroplast genome generated herein (**Figure 5**). Otherwise, major components of PSII from both DEC-IX and *P. calceolata*, encoded by *psbA*, *psbB*, *psbC* and *psbD*, were most highly expressed along the transect, relative to the other chloroplast-encoded genes they contain, as were the RUBISCO small and large subunits *rbcS* and *rbcL* (**Figures 5A,B** and **Supplementary Data S3**). Furthermore, the two novel DEC-IX ORFs (ORF1, 276 nt and ORF2, 915 nt) were transcribed. Akin to amplicon relative abundance patterns, higher read numbers mapped to DEC-IX in surface samples, than in the single DCM metatranscriptome sequenced, in which *P. calceolata* recruited many transcripts (**Supplementary Data S3**).

DISCUSSION

Here, 16S rDNA-based metabarcoding methods were used to examine phytoplankton diversity in time-series (BATS) and transect (ENP) samples. We captured marked seasonal trend and contributions from photosynthetic stramenopiles, which accounted for up to 73% of the total eukaryotic phytoplankton amplicons in the samples. The high relative abundances of dictyochophytes in oligotrophic waters and pelagophytes in mesotrophic or seasonally nutrient rich waters that we report are a previously unrecognized feature of stramenopile ecology. We focused on the unexpected observation that dictyochophytes, a stramenopile class largely represented by silicoflagellates in culture (Eikrem et al., 2004), reached their highest relative abundances in the most oligotrophic samples. High precision in phylogenetic identification was achieved by building a well curated backbone tree from plastid sequences, enabling us to identify transitions among numerous uncultured clades that had not clearly emerged from previous studies.

Phytoplankton Successional Patterns

Our amplicon analyses demonstrated that prasinophytes bloom during the BATS DM period, consistent with results from an earlier T-RFLP study (Treusch et al., 2012). These blooms were followed by a shift to eukaryotic communities dominated by stramenopiles as the water column became stratified and nutrients decreased (**Figure 1D**). At the coarse level of phylogenetic supergroups, eukaryotic phytoplankton did

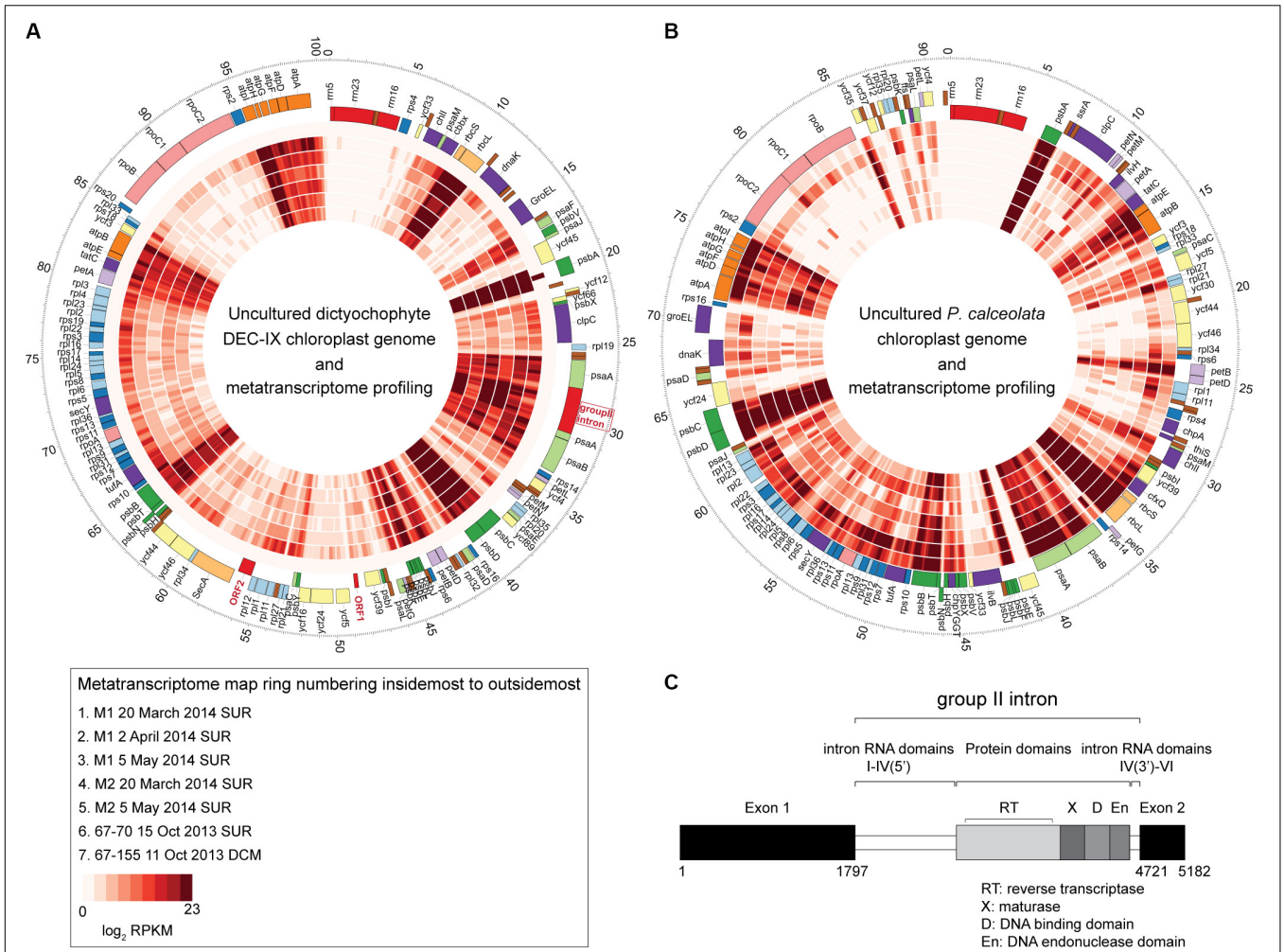


FIGURE 5 | The complete chloroplast genome of **(A)** uncultured dictyochophyte DEC-IX **(B)** and wild *P. calceolata* (Worden et al., 2012), and transcriptional activity in the ENP. Recruitment of metatranscriptomic reads (inner layers) at ≥ 99% nucleotide identity shown as mapped reads binned at 100 bp increments and log2 transformed for the heat plots which reflect the normalized transcripts. While the metatranscriptomes were unreplicated and therefore unsuitable for quantifying expression, qualitatively, similar patterns of expression were observed for genes present in both the DEC-IX and *P. calceolata* chloroplast genomes. All come from surface samples except the 67-155 metatranscriptome (100 m, DCM). Outermost ticks represent kilobase pairs (chloroplast genome size). **(C)** Structure of the *psaA* gene group II intron, showing upstream and downstream exons, intron domains and conserved IEP consisting of a reverse transcriptase (RT), a maturase domain (X), a DNA-binding domain (D) and an endonuclease domain (En).

not shift dramatically after DM, unlike cyanobacteria, where *Prochlorococcus* dominance was most pronounced late in the stratified period (**Figure 1C**). However, the high-resolution analysis of eukaryotic plastids captured previously unobserved seasonal shifts in the stramenopile community at BATS. During the DM period, pelagophytes, specifically PcC, were the dominant stramenopiles. In the transition to stratification, we observed reduced percentages of pelagophyte amplicons and increased contributions from dictyochophytes. During the stratified period, when stramenopiles were the most prominent eukaryotic phytoplankton, surface waters were dominated by dictyochophytes, while pelagophytes exhibited higher relative contributions in DCM communities (**Figure 2A** and **Supplementary Figure S1**). Diatoms, the most well-studied photosynthetic taxa among stramenopiles, rarely reached high relative abundances in our 10 year ∼monthly sample set and their variability (0 to 37% of stramenopile amplicons) appeared to be driven by punctuated, ephemeral blooms occurring or captured in some years and not others.

When the highly resolved genetic data was synopsized to provide a qualitative overview of the biogeography of the major phyla in the ENP, distinct spatial distributions were identified across environmental gradients, from oligotrophic ocean dominated by *Prochlorococcus*, to a mesotrophic region with increased signals from *Synechococcus* and eukaryotic phytoplankton. While not a focus in this paper, our primers clearly recovered diatoms as these phytoplankton were abundant in an ENP coastal sample (**Supplementary Figure S10**). Among eukaryotic phytoplankton at the mesotrophic and oligotrophic sites, stramenopiles and prasinophytes were

the two most abundant phyla (**Supplementary Figure S6**). The seasonal analysis of BATS revealed that large overall contributions of stramenopiles at the surface and DCM during stratified periods were driven by pelagophytes and dictyochophytes, paralleling features of community structure observed in OLIGO ENP samples. In contrast, the ENP mesotrophic region showed interannual variability at the surface, with stramenopile dominance during 2009 and 2013, versus prasinophyte dominance during 2011 and 2012. Mesotrophic ENP stations were dominated by pelagophyte amplicons throughout the photic zone (64 ± 15%), similar to what was observed during the DM period at BATS (60 ± 21%). Indeed, hierarchical clustering based on stramenopile classes using BATS and ENP surface amplicon data, for which we had the greatest statistical power, showed the BATS DM period clustering with mesotrophic ENP sites and the stratified periods with oligotrophic ENP sites (**Supplementary Figure S11**). Additionally, contributions at depth were remarkably stable over the regimes and sites studied, although stramenopiles in the ENP MESO SCM were notably higher than at BATS during DM period (**Supplementary Figure S6D**). We note that the ENP MESO SCM has measurable nutrients while at BATS during the DM period nutrients are frequently below detection. As a whole, stramenopiles contributed less to the overall community relative to other phytoplankton groups, particularly prasinophytes and cyanobacteria, in BATS DM and ENP MESO samples. However, collectively, the dominant stramenopile lineages across seasonal and combined regime averages for BATS, and the ENP OLIGO and MESO sites, were dictyochophytes and pelagophytes.

High-performance liquid chromatography-based studies have reported that pelagophytes and haptophytes are important at BATS, with abundance peaks occurring during DM and at the DCM (Andersen et al., 1996; Goericke, 1998; Haidar and Thierstein, 2001; Steinberg et al., 2001). Additionally, prasinophytes were shown by qPCR to have high numerical abundances during spring and even to bloom relative to other taxa during DM based on T-RFLP analyses (Treusch et al., 2012), similar to our observations. Silicoflagellates have also been reported at BATS, in a microscopy-based study on haptophytes (Haidar and Thierstein, 2001). Our studies resolve phytoplankton molecular diversity and seasonal patterns, and demonstrate a previously unrecognized nutrient-linked seasonal switch between pelagophytes and dictyochophytes.

Our findings demonstrate that dictyochophyte amplicons form the greatest percentage of those from eukaryotic phytoplankton, and dominate stramenopile amplicons in surface waters throughout the stratified period at BATS (**Figure 2A**) and in oligotrophic surface waters of the North Pacific Gyre (**Figure 4C**). These results are different from conclusions on important taxa in prior studies and it is possible that the seasonal importance of dictyochophytes observed herein has gone undetected in some studies, especially HPLC-based studies, because their pigments overlap with those of better-studied taxa, such as diatoms. Cultured dictyochophytes contain Chl *a* and *c*, and fucoxanthin as the main carotenoid, a carotenoid also found in diatoms, haptophytes and chrysophytes, alongside one acylfucoxanthin derivative, 19′-butanoyloxyfucoxanthin [19′-but], which is generally considered a specific marker for pelagophytes (Daugbjerg and Henriksen, 2001; Eikrem et al., 2004; Chang et al., 2012). Thus, cultured dictyochophytes exhibit pigments typically considered as markers for other, better-studied phytoplankton classes (Daugbjerg and Henriksen, 2001). Variations in pigment composition among the few cultured dictyochophyte groups further confound field sample interpretation. For example, *Dictyocha speculum* and *Vicicitus globosus* possess 19′-hexanoyloxyfucoxanthin [19′-hex], a major pigment in haptophytes (Daugbjerg and Henriksen, 2001; Chang et al., 2012), whereas pedinellids and *Rhizochromulina* have 19′-but and lack 19′-hex (Daugbjerg and Henriksen, 2001). Our analyses indicate that dictyochophytes likely contribute significantly to the oceanic 19′-but and 19′-hex pigment pool, and that their contributions were previously miss-assigned to pelagophytes and haptophytes. Indeed, as outlined above, prior HPLC studies at BATS indicated haptophytes and pelagophytes were the most abundant eukaryotic phytoplankton groups (Andersen et al., 1996; Goericke, 1998; Steinberg et al., 2001). Given that our primers clearly recovered pelagophytes and haptophytes, it seems likely that the importance of dictyochophytes in eukaryotic phytoplankton communities has been significantly underestimated. Collectively, these results shift our view of how different phytoplankton groups contribute at BATS, and in the ENP, and identify potential shifts in trophic modes among eukaryotic phytoplankton (see below).

Extensive Molecular Diversity of Photosynthetic Stramenopiles

In addition to the broad trends discussed above, phylogenetic reconstructions of near full-length 16S rRNA gene sequences revealed numerous previously unrecognized clades within both the pelagophyte (**Figure 2B**) and dictyochophyte (**Supplementary Figure S5**) classes. Moreover, we identified a novel eukaryotic (plastid-based) clade that is a sister of bolidophytes and may represent a new stramenopile class (**Supplementary Figure S2**).

High resolution phylogenetic comparisons of stramenopiles between BATS and the ENP indicated that several dictyochophytes clades, DEC-I, DEC-VI and FC-III, were prominent at BATS and ENP (**Figure 3B**). These groups also drove the trends seen for the overall class. Among dictyochophyte amplicons, DEC-I and DEC-VI were prevalent during the Transitional and Maximum Stratified periods in BATS surface waters and oligotrophic ENP surface waters, while FC-III was prominent during DM period and the DCM of the ENP oligotrophic site. Different patterns were observed for other clades, such as FC-I and FpC, which were barely detected at BATS and more abundant along ENP Line-67. Similarly, increased contributions from DEC-IV were found in ENP samples (**Figure 3B**). Among pelagophytes PcC showed high relative abundances at depth throughout the year at BATS and on Line-67 (mesotrophic and oligotrophc stations), whereas it had much lower relative abundances at the surface during the BATS stratified period and on Line-67 at the most oligotrophic

sites (**Figure 3A**). In contrast PEC-VIII showed high relative abundances among pelagophytes at BATS throughout the year regardless of depth or period, but in the ENP only exhibited high abundance at the surface of the oligotrophic site. Our results also indicate that current phytoplankton isolation enrichment methods are not suited to recovering these important taxa, as frequently discussed for heterotrophic protists (Keeling et al., 2014; Needham et al., 2019), or that insufficient effort has been made to recover cultures from oligotrophic sites. In fact, just one dictyochophyte clade observed in our samples is represented in culture, and it exhibited higher relative abundances in more nutrient-rich waters, whereas the key clades found in oligotrophic samples remain uncultured.

Distinct Pelagophyte Clades and Contributions During More Nutrient Rich Periods

The first described pelagophyte, *Pelagomonas*, is considered a cosmopolitan genus that is frequently isolated from sea water (Andersen et al., 1993; Andersen et al., 1996). Recently, metagenomic approaches revealed their presence in SCM communities, but not in the surface (Dupont et al., 2015). Our results clearly demonstrated that *P. calceolata* is among the major phytoplankton taxa in SCM/DCM communities, at least based on relative amplicon abundances. This corresponds well with results from a metagenomic mapping study using the *P. calceolata* chloroplast genome, which showed it being broadly distributed (Worden et al., 2012). Additional metatranscriptomic (Dupont et al., 2015; Kolody et al., 2019) and qPCR (herein) studies correspond relatively well with our amplicon-based inferences. Interestingly, a metatranscriptomic study in the oligotrophic North Pacific suggested that *P. calceolata* dominated nitrate assimilation in the SCM, but was experiencing nutrient stress in surface waters (Dupont et al., 2015). Here, during the BATS DM period, when vertical mixing brings nutrients to the surface, *P. calceolata* showed higher relative abundances at the surface than throughout the rest of the year (**Figure 3A**), supporting the idea that this species relies on nitrate as a primary nitrogen source. In terms of overall abundances, while *P. calceolata* was detected by qPCR and formed a notable percentage of stramenopile amplicons, other eukaryotic phytoplankton comprised the majority of plastid amplicons at BATS during DM and often in the ENP MESO samples (e.g., prasinophytes, up to 79% of plastid amplicons in each setting).

We identified an unknown pelagophyte clade, PEC-VIII, in ENP OLIGO and BATS stratified surface waters where nutrient availability was low relative to other periods and sites (**Figure 3A**). Reference sequences for PEC-VIII clade come from prior studies in the northeast subarctic Pacific, South Pacific, tropical Pacific near Cocos island, and the ENP (Allers et al., 2013; Yin et al., 2013; Choi et al., 2017), suggesting their broad distribution and potential importance in the phytoplankton communities of oligotrophic surface oceans. At BATS, PEC-VIII was a major contributor at the surface during the stratified period. Phagotrophy has not yet been observed in cultured pelagophytes (Stoecker et al., 2006), however, *Aureococcus* utilizes dissolved organic matter (as likely many phytoplankton do).

Dictyochophyte Importance in Oligotrophic Surface Oceans

Dictyochophytes have been reported in other molecular surveys of planktonic communities (Shi et al., 2011; de Vargas et al., 2015). A striking finding from our high-resolution phylogenetic analyses beyond the broad molecular diversity of dictyochophytes (**Supplementary Figure S5**), which has been noted in the Tara Oceans study (de Vargas et al., 2015), is the remarkably strong, reproducible patterns that aligns with low nutrient conditions or potentially co-associated factors (e.g., temperature, **Figure 3B**). In our samples dictyochophytes formed up to 81% and 77% of stramenopile amplicons in oligotrophic BATS and ENP surface waters, respectively, and, on average, 64 $\pm$ 12% during the Maximum Stratified period at BATS, and 70 $\pm$ 6% in the ENP OLIGO surface waters (**Figures 2A**, **4D**). Further, the majority of these amplicons came from uncultured dictyochophyte groups that we identify herein.

One factor that could drive plankton composition in oligotrophic waters is vitamin biosynthesis capabilities, since vitamins exhibit low concentrations (Sañudo-Wilhelmy et al., 2012). For example, eukaryotic phytoplankton taxa present at BATS exhibit different configurations of thiamin biosynthetic genes (McRose et al., 2014; Gutowska et al., 2017; Paerl et al., 2017), which led us to investigate available dictyochophyte transcriptome-derived genomic information (Keeling et al., 2014). *Pseudopedinella elastica* CCMP716 and the related undescribed species CCMP2098 have complete thiamin pathways (**Supplementary Table S8**) and phylogenetically group together with clade DEC-I, the main dictyochophyte clade in stratified waters (**Figure 3B**). The other dictyochophyte with a complete pathway, *D. speculum* CCMP1381, is never numerically abundant in our samples. The four *Florenciella* transcriptomes in the MMETSP dataset contain only one or no genes from the thiamin pathway (**Supplementary Table S8**). While absence in transcriptomes is not definitive, it is interesting to note that the environmental *Florenciella*-like clades are most abundant when nutrients are elevated (although vitamin concentrations are not known), such as in BATS surface waters during deep mixing, the DCM, and ENP mesotrophic waters (**Figure 3B**).

Biology and Evolution of Dictyochophytes

Most of the clades identified herein lack cultured representatives. Hence, reports of dictyochophyte biology, including reports of predatory, mixotrophic nutritional modes (Sekiguchi et al., 2003), come largely from cultured taxa that belong to phylogenetic clades different from those that dominated at BATS and the ENP (**Supplementary Figure S5**). Likewise, our knowledge of dictyochophyte pigment composition (Daugbjerg and Henriksen, 2001; Eikrem et al., 2004; Chang et al., 2012), and silicoflagellate cell structure (Henriksen et al., 1993), comes from cultured species. Silicoflagellates have served as paleoecological markers, due to their distinct silicate skeletons (Rigual-Hernandez et al.,

2016), and have been shown to be common in nutrient rich, high-latitude environments by HPLC and light microcopy (van de Poll et al., 2018). The Dictyochales order as a whole takes a basal position within the dictyochophyte phylogeny (**Supplementary Figure S5**). Further, *Florenciella* which holds a less basal position, encodes the well characterized *SIT* gene, a Si transporter, also present in diatoms and several other marine eukaryotic lineages (Marron et al., 2016) and is purportedly a mixotroph based on field SIP data (Frias-Lopez et al., 2009). Because the dominant clades observed at BATS and the ENP lack cultured representatives, it remains unclear whether the capacity to feed via phagocytosis extends across dictyochophyte diversity. However, as with diatoms (Durkin et al., 2016), which were low in relative abundance at our study sites, dictyochophyte silicification has been associated with efficient carbon export (Marron et al., 2016; Rigual-Hernandez et al., 2016).

Just one known predatory mixotrophic stramenopile had a sequenced chloroplast genome, the chrysophyte alga *Ochromonas* CCMP1393 (**Supplementary Figure S2**) (Sevcikova et al., 2015), until the recent addition of plastid genomes from four cultured dictyochophytes species *D. speculum*, *Rhizochromulina marina*, *F. parvula* and *P. elastica* (Han et al., 2019). We sequenced the complete chloroplast genome of newly identified uncultivated clade DEC-IX from a wild cell. The phylogenetic position of DEC-IX requires further evaluation using phylogenomic approaches, as the tree topology using near full length 16S rRNA gene sequences was supported primarily at terminal nodes (**Supplementary Figure S8**), and the use of missing positions to place 16S and 18S amplicons (**Supplementary Figures S7, S9**) also rendered largely unsupported nodes, apart from terminal nodes. The DEC-IX chloroplast genome (100,783 bp) is smaller than that of *Ochromonas* and several other stramenopiles, e.g., diatoms and eustigmatophytes (Sevcikova et al., 2015), and the four dictyochophytes, which range from 108,152 to 140,025 bp. However, it is larger than that of *P. calceolata* (91,306 bp), one of the major players in our study, and other pelagophytes (*A. anophagefferens*, 89,599 bp; *A. lagunensis*, 94,346 bp) (Ong et al., 2010). The numbers of protein encoding genes in the chloroplast genomes of DEC-IX and the four cultured dictyochophytes (Han et al., 2019) are similar. Curiously, both the *D. speculum* and DEC-IX plastid genomes lack the inverted repeats (IRs) present in other dictyochophyte plastid genomes (Han et al., 2019) and many other plastid-bearing taxa (Sabir et al., 2014). *D. speculum* however, has a distinct phylogenetic position from DEC-IX, based on 16S and 18S phylogenies (**Supplementary Figures S7–S9**).

The group II intron in the *psaA* gene of DEC-IX is relatively rare. *PsaA* encodes a protein critical for binding P700, the primary electron donor of photosystem I. The presence of a mobile genetic element, specifically a group II intron in a *psaA* gene, has previously been reported in the green alga *Chlamydomonas reinhardtii* (Perron et al., 1999), and, among stramenopiles, in a diatom, *Toxarium undulatum* (Ruck et al., 2017), and most recently *D. speculum*, but not *R. marina*, *F. parvula*, *P. elastica* or other taxa (Han et al., 2019). Group II introns have also been reported in a handful of other plastid genes in other taxa (Perrineau et al., 2015). The intron within the diatom *psaA* gene lacks the En domain that confers mobility, and the *C. reinhardtii* chloroplast genome lacks this domain as well. In contrast, both the *D. speculum* and DEC-IX *psaA* genes contain a fully functional mobility element. The field transcriptomes mapping to this intronic region indicates its expression levels are similar to *psaA* (**Figure 5A**).

Mixotrophy in Dictyochophytes

It is long known that some groups of phytoplankton engage in phagotrophy. These predatory mixotrophs prey upon other microbes and photosynthesize, an important factor because, depending on the extent of prey consumption and respiration versus photosynthesis, contributions to primary production are altered (Mitra et al., 2014; Flynn et al., 2019). Further, predatory mixotrophs are recognized as being quantitatively important in marine ecosystems (Worden et al., 2015; Ward and Follows, 2016; Edwards, 2019). Field experiments indicate that mixotrophs contribute close to 50% of total bacterivory at an oligotrophic site in the Mediterranean (Unrein et al., 2007) and in open ocean areas along a north-south transect in the Atlantic Ocean (Hartmann et al., 2012). The relative importance of predatory mixotrophs may vary with depth, for example a Sargasso Sea study found that 50% of photosynthetic nanoflagellates consumed prey in the surface mixed layer, while only 0.5% ingested prey in the DCM (Arenovski et al., 1995). Further, based on modeling studies, the combination of high light intensities and low availability of dissolved nutrients (as would occur in the surface) has been suggested to favor mixotrophs over photoautotrophic algae in oligotrophic surface waters (Duhamel et al., 2019; Edwards, 2019).

The phylogenetic affiliations of predatory mixotrophs remain poorly characterized, especially in offshore oligotrophic waters. Identification based on FISH showed high rates of bacterivory by haptophytes in the Mediterranean Sea, as well as cryptophytes and dinoflagellates (Unrein et al., 2014), the latter two being well recognized mixotrophic groups, including in oligotrophic ocean regions (Stoecker, 1999; Duhamel et al., 2019). By the same approach, haptophytes and chrysophytes (another stramenopile lineage) were found to feed on *Prochlorococcus* in the Atlantic subtropical gyres (Hartmann et al., 2013). Moreover, an obligate-predatory strategy has been reported for two cultured marine chrysophytes belonging to the *Ochromonas* genus, which, although actively photosynthesizing in the light, cannot achieve positive growth rates by photosynthesis alone, and one of which can live without light (Wilken et al., 2020).

Microscopy on field samples comes with its own set of challenges, including that FISH requires sequence information for the design of probes with specific targets. Our results provide target sequence information and an impetus to examine dictyochophytes in the field. This may help to fill in gaps, as in the Mediterranean study discussed above, where only about 50% of mixotrophic flagellates could be identified either morphologically or by the FISH-probes employed (Unrein et al., 2014). To date feeding by dictyochophytes has not been targeted in FISH studies, but situations in which they dominate amplicon reads from eukaryotic phytoplankton are ecologically similar to the oligotrophic surface waters of the Sargasso Sea (Arenovski et al.,

1995), where a high proportion of actively feeding photosynthetic flagellates had been observed and the dominant phytoplankton are cyanobacteria (**Figure 1C**). In combination with our observation that flow sorted phytoplankton cells containing acid vacuoles are substantially composed of photosynthetic dictyochophytes, this suggests that the success of at least several dictyochophyte clades at BATS involves predatory mixotrophy - and that this is likely a means by which summertime photosynthetic eukaryotes acquire nutrients and can co-exist with cyanobacteria in the upper water column. The higher amplicon abundances of dictyochophytes relative to diatoms and other eukaryotic phytoplankton outside the picoplankton size class, may further reflect a capacity to outcompete these other eukaryotic taxa for nutrients by circumventing reliance on transporter affinity and an optimized surface to volume ratio.

Various lines of evidence indicate that predatory mixotrophy is at least common within dictyochophytes. Members of the Pedinellales from freshwater and marine environments have been reported as predatory mixotrophs (Sekiguchi et al., 2003), and photosynthetic freshwater *Pseudopedinella* can be important bacterivores in lakes (Gerea et al., 2016). The purely heterotrophic dictyochophyte species *Pteridomonas danica* and *Ciliophrys infusionum* lost photosynthetic capacity independently from each other, although both retain a non-photosynthetic plastid (Sekiguchi et al., 2002). DEC-X had an ill-supported position in the phylogenetic reconstructions and potentially could be a heterotrophic predator, since, although it retains a plastid 16S rRNA gene and presumably a plastid genome, it lacked chlorophyll fluorescence and was selected by having a stained vacuole (**Supplementary Figure S7**). In contrast to DEC-X, DEC-IX clearly retained a fully functional plastid, showing both chlorophyll fluorescence and expression of key photosynthetic genes. Evidence that a marine *Florenciella* species (**Supplementary Figure S9**; 98.5 – 99.8% nt identity to *F. parvula*) could be a predatory mixotroph comes solely from 18S rRNA data from an on-deck SIP experiment in the Pacific Ocean (Frias-Lopez et al., 2009) during which these taxa incorporated ^{13}C from labeled cyanobacteria. Predatory mixotrophic trophic modes could thus be widely represented among diverse dictyochophytes, potentially including some of the environmental clades recovered here. The fact that 74 ± 15% of the dictyochophyte amplicons in our oligotrophic surface ocean samples do not have cultured counterparts calls for targeted efforts to characterize their physiology and ecological roles.

CONCLUSION

Our field studies show that similar eukaryotic phytoplankton community structure patterns can be observed along different zones of the ENP when compared to seasonal periods in the Sargasso Sea. Prasinophyte algae comprise the bulk of eukaryotic amplicons in mesotrophic conditions and the deep mixing period at BATS, but among stramenopiles, pelagophytes are also notable under these conditions. We observed considerable diversity in pelagophytes and found that one species in particular, *P. calceolata*, persisted at the DCM in oligotrophic regimes. Additionally, a previously undescribed environmental clade, PEC-VIII, was also abundant among stramenopile amplicons. However, it was dictyochophytes that dominated eukaryotic phytoplankton amplicons in those regions of the water column where nutrient availability was at its lowest, regions where prior studies indicate mixotrophy is most prevalent. The cell sorting experiments indicated that the taxa comprising natural dictyochophyte populations have food vacuoles (Wilken et al., 2019), implying that they are predatory mixotrophs. Of note, both diatoms and at least some dictyochophytes require Si, opening the potential for competitive exclusion between these taxa, and for both to be agents of biomass export. We propose that dictyochophytes acquire nutrients from prey, circumventing direct competition with *Prochlorococcus* and *Synechococcus* for inorganic nutrients. This would enable them to contribute significantly to net primary production in the oligotrophic regions without relying on uptake of scarce dissolved nutrients.

Our findings support the outcomes of models that a synergistic coupling exists between prey ingestion for nutrients, and photosynthetic acquisition of carbon and energy, which enables mixotrophs to be successful in stable oligotrophic gyres (Duhamel et al., 2019; Edwards, 2019). Mixotrophic behavior also has consequences when assimilated into global food web models, shifting biomass to larger size classes and thereby enhancing sinking carbon flux (Ward and Follows, 2016). Thus, although the amount of carbon traversing food webs via mixotrophic inputs and the distribution of mixotrophic behaviors across algal diversity remain sparsely explored, there are mounting reasons to regard mixotrophy as a widespread behavior that significantly alters carbon cycling. Our results, combined with outcomes from modeling studies (Ward and Follows, 2016; Edwards, 2019), indicate that the dynamics of these dictyochophytes and other mixotrophs are likely to be sensitive to environmental change connected to enhanced ocean stratification, including changes in the duration and intensity of light exposure.

MATERIALS AND METHODS

Oceanographic Sampling

BATS samples were collected between August 1991 to February 1994, and September 1997 to January 2004 as in Carlson et al. (2009). For ENP study sites, 15 samples were collected on four cruises, WFAD09, CANON11, C0912 and CN13ID, from 2009, 2011, 2012 and 2013, 8 sequenced for a prior cyanobacterial study (Sudek et al., 2015) and 7 sequenced herein (**Supplementary Table S5**). Samples were collected using Niskin bottles mounted on a rosette equipped with an SBE9 conductivity, temperature and depth (CTD) sensor (Sea-Bird Electronics). Samples for nucleic acid extraction were collected by filtering 500–2000 ml seawater through 0.2 μm pore size polyethersulfone membrane filters (Supor 200, Pall Gelman). Filters were placed into sterile cryovials, flash-frozen in liquid nitrogen and transferred to −80°C until further use. ENP RNA samples were collected by filtering 15–20 L of seawater onto a 3 μm pore-size, 142 mm diameter Versapor filter (Pall Corporation), storing at −80°C and extracting as described previously (Needham et al., 2019). These came from five 2014 samples, two previously sequenced (Needham et al., 2019). Others were processed and sequenced

herein, specifically from 20 March at 10 m (ERS3865411) and 20 m (ERS2592094), 5 May at 10 m (ERS3865412) and 20 m (ERS3865413) and 2 April at 20 m (ERS2592093). Samples from 2 April 2014 and 5 May 2014 were pre-filtered through a 30 μm mesh, while those from 20 Mar 2014 were not. In 2013, additional metatranscriptome samples were collected at 10 m at 67–70 and 100 m at Station 67-155 and pre-filtered through a 20 μm mesh before final filtration, freezing and processing. Nutrient data were acquired as described previously for BATS (Treusch et al., 2009) and the ENP (Sudek et al., 2015).

DNA Extraction

Briefly, 384 monthly BATS samples (from August 1991 to February 1994 and September 1997 to January 2004) were extracted as described previously in Treusch et al. (2009). The 15 ENP samples were extracted using the DNeasy Plant Mini Kit (Qiagen) with a modification of the manufacturer's protocol, as described previously (Demir-Hilton et al., 2011).

Pelagomonas calceolata CCMP1756 was acquired from the National Center for Marine Algae and Microbiota (NCMA) and grown on L1-Si medium at 21°C under a 14:10 h light:dark cycle with 100 μmol quanta m^{-2} s^{-1}. The culture was harvested during exponential growth by centrifugation at 8,000 × *g* for 10 min. Most of the supernatant was removed, the pellet resuspended in the remaining ∼1 ml supernatant, transferred to a microcentrifuge tube and pelleted again. DNA was extracted from pellets using the DNeasy Plant Mini Kit following the manufacturer's protocol.

16S rRNA Gene V1-V2 Amplicon PCR and Sequencing

BATS amplicons were generated as part of a prior study and ENP sequencing used the same 16S rRNA V1-V2 primers 27FB (5′-AGRGTTYGATYMTGGCTCAG-3′) and 338RPL (5′-GCWGCCWCCCGTAGGWGT-3′) as in Vergin et al. (2013), Sudek et al. (2015) using the 454 platform. Twelve samples per plate were sequenced using Roche/454 GS FLX Titanium platform (Roche 454 Life Sciences). Quality control (QC) of 454-pyrosequenced reads was performed using published methods (Hamady et al., 2008). This amounted to 907,022 ENP amplicons in total after QC, as well as 2,540,966 16S V1-V2 amplicons after QC from 384 BATS samples (latitude ranges from 31.164 to 31.906 and longitude from −64.679 to −63.773).

Reference Alignments and Trees for Amplicon Placement

16S rRNA V1-V2 amplicons were initially parsed using PhyloAssigner v6.166 (Vergin et al., 2013) with amplicons from either plastids or cyanobacteria first taxonomically placed using global cyanobacterial and plastid phylogenetic trees, as described in Sudek et al. (2015), Choi et al. (2017). Briefly, this method aligns amplicons to the unmasked alignment from the phylogenetic reconstruction based on full length sequences and requires statistical support for the placement to be considered valid (Matsen et al., 2010). Here, in addition to using the previously published PhyloAssigner alignments (Choi et al., 2017), we took a total of 270 high-quality, near full-length 16S rRNA gene sequences from 84 described stramenopile species and 182 environmental sequences from undescribed stramenopile taxa, including many from 2007 ENP samples (Choi et al., 2017), and aligned them using MAFFT v7.055b (Katoh and Standley, 2013) with default parameters. After manual curation this alignment was used to generate a high-quality stramenopile reference tree, which included four rhodophyte sequences as an outgroup. Positions with gaps were masked using Gblocks v0.91b (Castresana, 2000) and phylogenetic inferences were done using Maximum Likelihood inference implemented in RAxML v8.0.0 (Stamatakis, 2014) under the gamma corrected GTR model of evolution with 1,000 bootstrap replicates (-m GTRGAMMA -f a -# 1,000 parameters) based on the analysis of 1,033 homologous positions. Another maximum likelihood inference was made using PhyML v3.0.1 (Guindon et al., 2010) with the substitution model of gamma corrected GTR and 100 bootstrap replicates (-m GTR -f e -v e -c 8 -a e -b 100 -BEST -n_rand_starts 10). Additional phylogenetic reconstructions were performed in MrBayes v3.2.6 (Ronquist et al., 2012) with the parameters of lset nst = 6 rates = invgamma ncat = 6, and ngenval = 10,000,000 samplefreqval = 1,000 and tempval = 0.200, and the final tree figure was produced with assistance from FigTree v1.4.3 and topology from RAxML (**Supplementary Figure S2**). More refined specific reference alignments and phylogenetic trees were constructed for two stramenopile classes, pelagophytes and dictyochophytes (**Figure 2B** and **Supplementary Figure S5**) using the same approaches described above. Briefly, a total of 68 and 61 sequences were used for pelagophytes and dictyochophytes with 1,210 and 1,066 homologous positions, respectively. To improve the resolution of the trees, 16S rRNA gene reference sequences from the MMETSP datasets (10 for pelagophytes, 5 for dictyochophytes) were added.

In order to ascertain the accuracy of placement of amplicon sequences on the unmasked alignment by PhyloAssigner we performed quality control with a set of known sequences. Specifically, for both the pelagophyte and dictyochophyte reference trees, 10 sets of V1-V2 sequences were generated from 50 near-full length 16S rRNA gene sequences and tested by running them through PhyloAssigner, with the new reference trees, to verify correct classification. The placement results were 100% correct.

Data Set Description and Statistical Analysis

For BATS, 384 monthly samples collected over a 12 year period from 1991 to 2004 were aligned using the month of deepest mixing (month "0") as the reference point to establish months −1 to +10 (Carlson et al., 2009). Because our focus was on eukaryotic phytoplankton, samples with low amplicon sequencing depth (and poor statistical quality for high taxonomic resolution plastid analysis) were excluded, resulting in retention of 191 samples having > 1,000 total 16S V1-V2 rRNA amplicons and > 100 representing plastids. Of these, 77 were from the surface (0–10 m, 8,025 ± 4,896 16S amplicons $sample^{-1}$) and 27 from the SCM/DCM (40 – 120 m, 7,395 ± 3,528 16S amplicons

sample^{-1}). Eighty-seven came from other depths between 40 and 300 m (7,612 ± 4,592 16S amplicons sample^{-1}). For stramenopile analyses samples were excluded with < 100 for total stramenopile, or < 20 for an individual stramenopile class, amplicons.

Within samples, relative abundance for each phylogenetic subgroup was calculated based on the total number of amplicons from the respective higher taxonomic group (**Supplementary Figure S12**). Additionally, because *Prochlorococcus* and *Synechococcus* generally have one and two copies of the rRNA operon, respectively, the latter were divided by two prior to computing percentages as in (Sudek et al., 2015). At each depth averages and standard deviations (SD) of phylogenetic group abundances were computed per month (−1 to +10) at BATS (Carlson et al., 2009). Due to uneven sample numbers for each month (data S1), the unweighted mean, notated here as $\mu_{\bar{x}}$, equation (1), was calculated as

$$\mu_{\bar{x}} = \frac{\sum_i \mu_{\bar{x}_i}}{n} \quad (1)$$

and the pooled SD, notated here as $\sigma_{\bar{x}}$, equation (2), was calculated as

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_i \sigma^2_{\bar{x}_i}}{n} + \frac{\sum_{i<j} \left(\mu_{x_i} - \mu_{x_j}\right)^2}{n^2}} \quad (2)$$

where n is the number of means. Due to limited SCM/DCM data, averages and SDs at this depth were computed using monthly data with ≥ 3 samples that met amplicon cutoff criteria (months 0, DM; +4, +6 and +9 stratified).

Fifteen ENP samples were analyzed using the same PhyloAssigner approach. Overall ENP sequencing depth was greater, with on average 11,300 ± 6,894 total and 3,366 ± 2,641 plastid-derived amplicons recovered per sample excluding one sample with 63,526 total and 26,672 plastid-derived amplicons.

Hierarchical Clustering

Hierarchical clustering was carried out using log-transformed normalized unweighted means of relative amplicon abundances. The approximately unbiased *p*-values (%) as well as bootstrap probabilities were computed via multiscale bootstrap resampling with 10,000 replications using the R package Pvclust (Suzuki and Shimodaira, 2006), modified to allow Bray-Curtis similarities for distance calculations.

qPCR Primer Design, Testing and Implementation

To generate 18S rRNA gene insert-bearing plasmid standards for qPCR, the 18S rRNA gene was amplified from *P. calceolata* CCMP1756 using the primers 5′-ACCTGGTT GATCCTGCCAG-3′ and 5′-TGATCCTTCYGCAGGTTCAC-3′ (Moon-Van Der Staay et al., 2001). PCR cloning and plasmids purification was performed as in (Demir-Hilton et al., 2011). Clones were bidirectionally (Sanger) sequenced using Big Dye Terminator chemistry (Applied Biosystems) with plasmid primers M13F/M13R and the two internal primers 502f and 1174r (Worden, 2006).

To quantify *P. calceolata* abundance, a *P. calceolata*-specific TaqMan primer-probe set was developed (**Supplementary Table S6**) after retrieving cultured pelagophyte and environmental 18S rRNA gene sequences from GenBank. Primers specific to *P. calceolata* were designed manually and a *P. calceolata*-specific probe sequence was identified using Beacon Designer 8.14 (PREMIER Biosoft International). Melting temperature and secondary structures were checked using OligoAnalyzer 3.1 from IDT SciTools. The primer-probe set specificity for *P. calceolata* was validated using *Pelagococcus subviridis* and *Aureococcus anophagefferens* (**Supplementary Table S7**). qPCR including inhibition tests and analysis was performed as described previously (Demir-Hilton et al., 2011) for fourteen environmental samples from mesotrophic and oligotrophic ENP sites that were also amplicon sequenced.

Single-Cell Sorting

For flow sorting, ENP seawater was collected on 2 April 2014 (20 m) at Station M1 and 5 May 2014 (10 m) at M1 and M2. Seawater was run through a BD Influx flow cytometer with two lasers (488 nm and 355 nm excitation), operated using sterile nuclease-free 1 × PBS as sheath (Cuvelier et al., 2010). Prior to sorting, samples were pre-filtered through a 30 μm nylon mesh and concentrated 30–50 times by gravity filtering over a 0.8 μm Supor (Pall Gelman). Two stains, LysoTracker Green DND-26 (25 nM, final concentration) and LysoSensor Blue DND-167 (1 μM, final concentration) with emission collected through a 520/35 nm bandpass and a 435/30 nm bandpass, respectively, natural Forward Angle Light Scatter (FALS, a proxy for cell size) and chlorophyll fluorescence (692/40 nm bandpass) were used in a variety of gating scenarios. Stained samples were incubated for 10 to 20 min before being run. Unstained controls were run to discern positive signals from stains. 384-well plates were illuminated with UV for 2 min prior to performing the sorts and selected cells were then sorted using the Single-cell sorting mode from the BD FACS Sortware (software v1.0.0.650), ensuring that only one cell would be sorted into each well. The drop delay and sort quality were controlled by sorting known numbers of fluorescent beads and counting on a microscope. Negative and positive controls on each plate involved leaving a subset of wells empty and having a column of wells receive 20 cells, respectively. Plates were covered with foil and immediately frozen at −80°C.

Single-Cell Genome Amplification and Analysis

Whole genome amplification (Cuvelier et al., 2010) and rRNA amplicon sequence analysis of sorted dictyochophyte cells followed the methods detailed previously (Needham et al., 2019) with the latter using pooled triplicate reactions from the Illumina adapted TAReuk454FWD1/TAReukREV3 primers for the 18S rRNA V4 region (Stoeck et al., 2010). Sequences were *de novo* clustered at 99% nt identity by UCLUST. The dictyochophyte wells had only a single 18S amplicon type. Whole genome sequencing of two single cells was then performed via Illumina HiSeq (Illumina), generating 14,848,205 and 15,058,189 PE 2 × 251 bp reads. Analysis of genome sequences from

the dictyochophyte cells were quality controlled as previously detailed (Needham et al., 2019). For assembly of the uncultured dictyochophyte, paired and unpaired quality filtered reads from the two dictyochophyte-containing wells were initially assembled as individual single cell assemblies as well as a co-assembly of both samples using SPADES v 3.11.0 (Bankevich et al., 2012) as previously described (Cuvelier et al., 2010). Then, mapping of reads to the assembled contigs and identification of chloroplast genomes was performed in Anvi'o v3 as previously described (Eren et al., 2015), and enabled by identification of chloroplast rRNA genes with metaxa2 (Bengtsson-Palme et al., 2015). The chloroplast contigs from the two wells were determined to have > 99.9% nt identity to one another via the *ani.rb* function within the enveomics package (Rodriguez-R and Konstantinidis, 2016). A single well was chosen as a representative of the dictyochophyte chloroplast genome assembly and used for the downstream analysis of metatranscriptome mapping. The complete chloroplast genome was annotated using DOGMA (Wyman et al., 2004) and all intergenic spaces were manually scanned using BLAST for ORFs or additional proteins that might not be present in chloroplast genomes from other phytoplankton.

RNA Extraction and Sequencing

RNA was extracted using the TotallyRNA kit (Life Technologies) with the following initial steps to maximize lysis and accommodate the large filter: in a sterile petri dish 2 ml of lysis buffer from the kit was added to the frozen filter half. The filter was then cut into six pieces and the filter pieces along with buffer transferred into two 2 ml screw cap tubes pre-filled with ∼200 μl of a 1:1 mixture by volume of 0.1 mm and 0.5 mm diameter autoclaved glass beads (Biospec Products). After one minute of bead beating filter pieces and lysis buffer from the two tubes were recombined in a 15 ml screw cap tube and a further 3 ml lysis buffer added. The remainder of the extraction followed the manufacturer's instructions. DNA was digested using the TurboDNA-free kit (Life Technologies) following manufacturer's instructions. RNA integrity was evaluated on a Bioanalyzer (Agilent) and quantity determined on a Qubit fluorometer (Life Technologies). RNA yield ranged from 7.1 to 19.5 μg and after library construction was sequenced via Illumina HiSeq. Quality-filtered metatranscriptome reads (QC performed as above) were mapped to the uncultured dictyochophyte chloroplast genome assembly and *P. calceolata* chloroplast genome (accession JX297813) via bbmap.sh (v37.17) (Bushnell, 2014) at a sequence similarity cutoff of 0.99. Mapped reads were parsed via HTSeq-count (Anders et al., 2015), binned at 100 bp increments and placed on the circos (Krzywinski et al., 2009) plot based on the start position of their match in the uncultured dictyochophyte plastid genome.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in GenBank under accession numbers: MN275232, SRR7789640–SRR7789654, and ERS3865411–ERS3865415.

AUTHOR CONTRIBUTIONS

AZW and SJG designed the study. AZW designed ENP cruises in collaboration with FPC. VJ collected ENP samples, with contributions from SW and SS. SS and VJ performed amplicon sequencing on ENP samples. CJC performed phylogenetic analyses, with input from CB and HA. CJC analyzed amplicon data. CP and CB performed single-cell sorting and initial analyses. CJC and DMN performed metagenomics and transcriptomic analyses. CJC, SJG, and AZW wrote the manuscript with edits from all authors, especially SW and HA.

FUNDING

This research was funded by NSF Dimensions grants DEB-1639033 and DEB-1638928 to AZW and SJG, BIOSCOPE to SJG and the Gordon and Betty Moore Foundation GBMF3788 to AZW.

ACKNOWLEDGMENTS

We thank the team at BIOS, Craig Carlson and Rachel Parsons for BATS sample access. We also thank the captain(s) and crew of the R/V Western Flyer, Tim Pennington and Alyson Santoro (who was the Co-Chief Scientist of CN13ID). We are grateful to Bank Beszteri and Stefan Neuhaus for PhyloAssigner support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.542372/full#supplementary-material

REFERENCES

Allers, E., Wright, J. J., Konwar, K. M., Howes, C. G., Beneze, E., Hallam, S. J., et al. (2013). Diversity and population structure of marine group a bacteria in the Northeast subarctic Pacific Ocean. *ISME J.* 7, 256–268. doi: 10.1038/ismej.2012.108

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Andersen, R. A. (2004). Biology and systematics of heterokont and haptophyte algae. *Am. J. Bot.* 91, 1508–1522. doi: 10.3732/ajb.91.10.1508

Andersen, R. A., Bidigare, R. R., Keller, M. D., and Latasa, M. (1996). A comparison of HPLC pigment signatures and electron microscopic observations for oligotrophic waters of the North Atlantic and Pacific Oceans. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 43, 517–537. doi: 10.1016/0967-0645(95)00095-x

Andersen, R. A., Saunders, G. W., Paskind, M. P., and Sexton, J. P. (1993). Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the Pelagophyceae classis nov. *J. Phycol.* 29, 701–715. doi: 10.1111/j.0022-3646.1993.00701.x

Andersen, R. A., Van De Peer, Y., Potter, D., Sexton, J. P., Kawachi, M., and Lajeunesse, T. (1999). Phylogenetic analysis of the SSU rRNA from members of the Chrysophyceae. *Protist* 150, 71–84. doi: 10.1016/s1434-4610(99)70010-6

Arenovski, A. L., Lim, E. L., and Caron, D. A. (1995). Mixotrophic nanoplankton in oligotrophic surface waters of the Sargasso Sea may employ phagotrophy to obtain major nutrients. *J. Plankton Res.* 17, 801–820. doi: 10.1093/plankt/17.4.801

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bengtsson-Palme, J., Hartmann, M., Eriksson, K. M., Pal, C., Thorell, K., Larsson, D. G. J., et al. (2015). metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol. Ecol. Resour.* 15, 1403–1414. doi: 10.1111/1755-0998.12399

Bushnell, B. (2014). *BBMap: a Fast, Accurate, Splice-Aware Aligner*. Berkeley, CA: University of California.

Carlson, C. A., Ducklow, H. W., and Michaels, A. F. (1994). Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. *Nature* 371, 405–408. doi: 10.1038/371405a0

Carlson, C. A., Morris, R., Parsons, R., Treusch, A. H., Giovannoni, S. J., and Vergin, K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J.* 3, 283–295. doi: 10.1038/ismej.2008.117

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chang, F. H., McVeagh, M., Gall, M., and Smith, P. (2012). *Chattonella globosa* is a member of Dictyochophyceae: reassignment to *Vicicitus* gen. nov., based on molecular phylogeny, pigment composition, morphology and life history. *Phycologia* 51, 403–420. doi: 10.2216/10-104.1

Chavez, F. P., Pennington, J. T., Michisaki, R. P., Blum, M., Chavez, G. M., Friederich, J., et al. (2017). Climate variability and change: response of a coastal ocean ecosystem. *Oceanography* 30, 128–145.

Choi, C. J., Bachy, C., Jaeger, G. S., Poirier, C., Sudek, L., Sarma, V. V. S. S., et al. (2017). Newly discovered deep-branching marine plastid lineages are numerically rare but globally distributed. *Curr. Biol.* 27, R15–R16.

Coleman, M. L., and Chisholm, S. W. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18634–18639. doi: 10.1073/pnas.1009480107

Collins, C., Pennington, J., Castro, C., Rago, T., and Chavez, F. (2003). The California current system off Monterey, California: physical and biological coupling. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 50, 2389–2404. doi: 10.1016/s0967-0645(03)00134-6

Cuvelier, M. L., Allen, A. E., Monier, A., Mccrow, J. P., Messie, M., Tringe, S. G., et al. (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14679–14684. doi: 10.1073/pnas.1001665107

Daugbjerg, N., and Henriksen, P. (2001). Pigment composition and *rbcL* sequence data from the silicoflagellate *Dictyocha speculum*: a heterokont alga with pigments similar to some haptophytes. *J. Phycol.* 37, 1110–1120. doi: 10.1046/j.1529-8817.2001.01061.x

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605.

Demir-Hilton, E., Sudek, S., Cuvelier, M. L., Gentemann, C. L., Zehr, J. P., and Worden, A. Z. (2011). Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. *ISME J.* 5, 1095–1107. doi: 10.1038/ismej.2010.209

Duhamel, S., Kim, E., Sprung, B., and Anderson, O. R. (2019). Small pigmented eukaryotes play a major role in carbon cycling in the P-depleted western subtropical North Atlantic, which may be supported by mixotrophy. *Limnol. Oceanogr.* 64, 2424–2440. doi: 10.1002/lno.11193

Dupont, C. L., Mccrow, J. P., Valas, R., Moustafa, A., Walworth, N., Goodenough, U., et al. (2015). Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J.* 9, 1076–1092. doi: 10.1038/ismej.2014.198

Durand, M. D., Olson, R. J., and Chisholm, S. W. (2001). Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 48, 1983–2003. doi: 10.1016/s0967-0645(00)00166-1

Durkin, C. A., Van Mooy, B. A. S., Dyhrman, S. T., and Buesseler, K. O. (2016). Sinking phytoplankton associated with carbon flux in the Atlantic Ocean. *Limnol. Oceanogr.* 61, 1172–1187. doi: 10.1002/lno.10253

Edwards, K. F. (2019). Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proc. Natl. Acad. Sci. U.S.A.* 116, 6211–6220. doi: 10.1073/pnas.1814860116

Eikrem, W., Romari, K., Latasa, M., Le Gall, F., Throndsen, J., and Vaulot, D. (2004). *Florenciella parvula* gen. et sp nov (Dictyochophyceae, Heterokontophyta), a small flagellate isolated from the English Channel. *Phycologia* 43, 658–668. doi: 10.2216/i0031-8884-43-6-658.1

Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. doi: 10.7717/peerj.1319

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9824–9829.

Flynn, K. J., Mitra, A., Anestis, K., Anschütz, A. A., Calbet, A., Ferreira, G. D., et al. (2019). Mixotrophic protists and a new paradigm for marine ecology: where does plankton research go now? *J. Plankton Res.* 41, 375–391. doi: 10.1093/plankt/fbz026

Frias-Lopez, J., Thompson, A., Waldbauer, J., and Chisholm, S. W. (2009). Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environ. Microbiol.* 11, 512–525. doi: 10.1111/j.1462-2920.2008.01793.x

Gerea, M., Saad, J. F., Izaguirre, I., Queimaliños, C., Gasol, J. M., and Unrein, F. (2016). Presence, abundance and bacterivory of the mixotrophic algae *Pseudopedinella* (Dictyochophyceae) in freshwater environments. *Aquat. Microb. Ecol.* 76, 219–232. doi: 10.3354/ame01780

Goericke, R. (1998). Response of phytoplankton community structure and taxon-specific growth rates to seasonally varying physical forcing in the Sargasso Sea off Bermuda. *Limnol. Oceanogr.* 43, 921–935. doi: 10.4319/lo.1998.43.5.0921

Guillou, L., Chretiennot-Dinet, M.-J., Medlin, L., Claustre, H., Loiseaux-de Goer, S., and Vaulot, D. (1999). *Bolidomonas*: a new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *J. Phycol.* 35, 368–381. doi: 10.1046/j.1529-8817.1999.3520368.x

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Gutowska, M. A., Shome, B., Sudek, S., McRose, D. L., Hamilton, M., Giovannoni, S. J., et al. (2017). Globally important haptophyte algae use exogenous pyrimidine compounds more efficiently than thiamin. *mBio* 8:e01459-17.

Haidar, A. T., and Thierstein, H. R. (2001). Coccolithophore dynamics off Bermuda (N. Atlantic). *Deep Sea Res. Part II Top. Stud. Oceanogr.* 48, 1925–1956. doi: 10.1016/s0967-0645(00)00169-7

Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., and Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* 5, 235–237. doi: 10.1038/nmeth.1184

Han, K. Y., Maciszewski, K., Graf, L., Yang, J. H., Andersen, R. A., Karnkowska, A., et al. (2019). Dictyochophyceae plastid genomes reveal unusual variability in their organization. *J. Phycol.* 55, 1166–1180. doi: 10.1111/jpy.12904

Hartmann, M., Grob, C., Tarran, G. A., Martin, A. P., Burkill, P. H., Scanlan, D. J., et al. (2012). Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5756–5760. doi: 10.1073/pnas.1118179109

Hartmann, M., Zubkov, M. V., Scanlan, D. J., and Lepère, C. (2013). *In situ* interactions between photosynthetic picoeukaryotes and bacterioplankton in the Atlantic Ocean: evidence for mixotrophy. *Environ. Microbiol. Rep.* 5, 835–840. doi: 10.1111/1758-2229.12084

Henriksen, P., Knipschildt, F., Moestrup, Ø., and Thomsen, H. A. (1993). Autecology, life history and toxicology of the silicoflagellate *Dictyocha speculum* (Silicoflagellata, Dictyochophyceae). *Phycologia* 32, 29–39. doi: 10.2216/i0031-8884-32-1-29.1

Hibberd, D. (1976). The ultrastructure and taxonomy of the Chrysophyceae and Prymnesiophyceae (Haptophyceae): a survey with some new observations on the ultrastructure of the Chrysophyceae. *Bot. J. Linn. Soc.* 72, 55–80. doi: 10.1111/j.1095-8339.1976.tb01352.x

Karl, D. M., and Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* 12, 699–713. doi: 10.1038/nrmicro3333

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., et al. (2014). The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 12:e1001889. doi: 10.1371/journal.pbio.1001889

Kolody, B., Mccrow, J., Allen, L. Z., Aylward, F., Fontanez, K., Moustafa, A., et al. (2019). Diel transcriptional response of a California current plankton microbiome to light, low iron, and enduring viral infection. *ISME J.* 13, 2817–2833. doi: 10.1038/s41396-019-0472-2

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Lambowitz, A. M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3:a003616. doi: 10.1101/cshperspect.a003616

Limardo, A. J., Sudek, S., Choi, C. J., Poirier, C., Rii, Y. M., Blum, M., et al. (2017). Quantitative biogeography of picoprasinophytes establishes ecotype distributions and significant contributions to marine phytoplankton. *Environ. Microbiol.* 19, 3219–3234. doi: 10.1111/1462-2920.13812

Lomas, M. W., Bates, N. R., Johnson, R. J., Knap, A. H., Steinberg, D. K., and Carlson, C. A. (2013). Two decades and counting: 24-years of sustained open ocean biogeochemical measurements in the Sargasso Sea. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 93, 16–32. doi: 10.1016/j.dsr2.2013.01.008

Lomas, M. W., Steinberg, D. K., Dickey, T., Carlson, C. A., Nelson, N. B., Condon, R. H., et al. (2010). Increased ocean carbon export in the Sargasso Sea linked to climate variability is countered by its enhanced mesopelagic attenuation. *Biogeosciences* 7, 57–70. doi: 10.5194/bg-7-57-2010

Malmstrom, R. R., Coe, A., Kettler, G. C., Martiny, A. C., Frias-Lopez, J., Zinser, E. R., et al. (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J.* 4, 1252–1264. doi: 10.1038/ismej.2010.60

Man-Aharonovich, D., Philosof, A., Kirkup, B. C., Le Gall, F., Yogev, T., Berman-Frank, I., et al. (2010). Diversity of active marine picoeukaryotes in the Eastern Mediterranean Sea unveiled using photosystem-II *psbA* transcripts. *ISME J.* 4, 1044–1052. doi: 10.1038/ismej.2010.25

Marron, A. O., Ratcliffe, S., Wheeler, G. L., Goldstein, R. E., King, N., Not, F., et al. (2016). The evolution of silicon transport in eukaryotes. *Mol. Biol. Evol.* 33, 3226–3248. doi: 10.1093/molbev/msw209

Matsen, F. A., Kodner, R. B., and Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11:538. doi: 10.1186/1471-2105-11-538

McClatchie, S. (2014). *Regional Fisheries Oceanography of the California Current System*. Berlin: Springer.

McRose, D., Guo, J., Monier, A., Sudek, S., Wilken, S., Yan, S., et al. (2014). Alternatives to vitamin B-1 uptake revealed with discovery of riboswitches in multiple marine eukaryotic lineages. *ISME J.* 8, 2517–2529. doi: 10.1038/ismej.2014.146

Mitra, A., Flynn, K. J., Burkholder, J. M., Berge, T., Calbet, A., Raven, J. A., et al. (2014). The role of mixotrophic protists in the biological carbon pump. *Biogeosciences* 11, 995–1005. doi: 10.5194/bg-11-995-2014

Monier, A., Welsh, R. M., Gentemann, C., Weinstock, G., Sodergren, E., Armbrust, E. V., et al. (2012). Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. *Environ. Microbiol.* 14, 162–176. doi: 10.1111/j.1462-2920.2011.02576.x

Moon-Van Der Staay, S. Y., De Wachter, R., and Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409, 607–610. doi: 10.1038/35054541

Needham, D. M., Fichot, E. B., Wang, E., Berdjeb, L., Cram, J. A., Fichot, C. G., et al. (2018). Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. *ISME J.* 12, 2417–2432. doi: 10.1038/s41396-018-0169-y

Needham, D. M., Yoshizawa, S., Hosaka, T., Poirier, C., Choi, C. J., Hehenberger, E., et al. (2019). A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc. Natl. Acad. Sci. U.S.A.* 116, 20574–20583. doi: 10.1073/pnas.1907517116

Ong, H. C., Wilhelm, S. W., Gobler, C. J., Bullerjahn, G., Jacobs, M. A., Mckay, J., et al. (2010). Analyses of the complete chloroplast genome sequences of two members of the Pelagophyceae: *Aureococcus anophagefferens* CCMP1984 and *Aureoumbra lagunensis* CCMP1507. *J. Phycol.* 46, 602–615. doi: 10.1111/j.1529-8817.2010.00841.x

Paerl, R. W., Bouget, F.-Y., Lozano, J.-C., Vergé, V., Schatt, P., Allen, E. E., et al. (2017). Use of plankton-derived vitamin B1 precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. *ISME J.* 11, 753–765. doi: 10.1038/ismej.2016.145

Paerl, R. W., Johnson, K. S., Welsh, R. M., Worden, A. Z., Chavez, F. P., and Zehr, J. P. (2011). Differential distributions of *Synechococcus* subgroups across the California current system. *Front. Microbiol.* 2:59. doi: 10.3389/fmicb.2011.00059

Perrineau, M.-M., Price, D. C., Mohr, G., and Bhattacharya, D. (2015). Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. *PeerJ* 3:e1017. doi: 10.7717/peerj.1017

Perron, K., Goldschmidt-Clermont, M., and Rochaix, J. D. (1999). A factor related to pseudouridine synthases is required for chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. *EMBO J.* 18, 6481–6490. doi: 10.1093/emboj/18.22.6481

Rigual-Hernandez, A. S., Trull, T. W., Mccartney, K., Ballegeer, A. M., Lawler, K. A., Bray, S. G., et al. (2016). Indices based on silicoflagellate assemblages offer potential for paleo-reconstructions of the main oceanographic zones of the Southern Ocean. *Geo Mar. Lett.* 36, 271–280. doi: 10.1007/s00367-016-0444-8

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4:e1900v1.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Ruck, E. C., Linard, S. R., Nakov, T., Theriot, E. C., and Alverson, A. J. (2017). Hoarding and horizontal transfer led to an expanded gene and intron repertoire in the plastid genome of the diatom, *Toxarium undulatum* (Bacillariophyta). *Curr. Genet.* 63, 499–507. doi: 10.1007/s00294-016-0652-9

Sabir, J. S., Yu, M., Ashworth, M. P., Baeshen, N. A., Baeshen, M. N., Bahieldin, A., et al. (2014). Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. *PLoS One* 9:e107854. doi: 10.1371/journal.pone.0107854

Sañudo-Wilhelmy, S. A., Cutter, L. S., Durazo, R., Smail, E. A., Gómez-Consarnau, L., Webb, E. A., et al. (2012). Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14041–14045. doi: 10.1073/pnas.1208755109

Sekiguchi, H., Kawachi, M., Nakayama, T., and Inouye, I. (2003). A taxonomic re-evaluation of the Pedinellales (Dictyochophyceae), based on morphological, behavioural and molecular data. *Phycologia* 42, 165–182. doi: 10.2216/i0031-8884-42-2-165.1

Sekiguchi, H., Moriya, M., Nakayama, T., and Inouye, I. (2002). Vestigial chloroplasts in heterotrophic stramenopiles *Pteridomonas danica* and *Ciliophrys infusionum* (Dictyochophyceae). *Protist* 153, 157–167. doi: 10.1078/1434-4610-00094

Sevcikova, T., Horak, A., Klimes, V., Zbrankova, V., Demir-Hilton, E., Sudek, S., et al. (2015). Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci. Rep.* 5:10134.

Shi, X. L., Lepere, C., Scanlan, D. J., and Vaulot, D. (2011). Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. *PLoS One* 6:e18979. doi: 10.1371/journal.pone.0018979

Simmons, M. P., Sudek, S., Monier, A., Limardo, A. J., Jimenez, V., Perle, C. R., et al. (2016). Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the Eastern North Pacific Ocean. *Appl. Environ. Microbiol.* 82, 1693–1705. doi: 10.1128/aem.02730-15

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., and Knap, A. H. (2001). Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 48, 1405–1447. doi: 10.1016/s0967-0645(00)00148-x

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W., et al. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19, 21–31. doi: 10.1111/j.1365-294x.2009.04480.x

Stoecker, D., Tillmann, U., and Graneli, E. (2006). "Phagotrophy in harmful algae," in *Ecology of Harmful Algae*, Vol. 189, eds E. Granéli and J. T. Turner (Berlin: Springer), 177–187. doi: 10.1007/978-3-540-32210-8_14

Stoecker, D. K. (1999). Mixotrophy among dinoflagellates. *J. Eukaryot. Microbiol.* 46, 397–401. doi: 10.1111/j.1550-7408.1999.tb04619.x

Sudek, S., Everroad, R. C., Gehman, A. L. M., Smith, J. M., Poirier, C. L., Chavez, F. P., et al. (2015). Cyanobacterial distributions along a physico-chemical gradient in the Northeastern Pacific Ocean. *Environ. Microbiol.* 17, 3692–3707. doi: 10.1111/1462-2920.12742

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542. doi: 10.1093/bioinformatics/btl117

Treusch, A. H., Demir-Hilton, E., Vergin, K. L., Worden, A. Z., Carlson, C. A., Donatz, M. G., et al. (2012). Phytoplankton distribution patterns in the northwestern Sargasso Sea revealed by small subunit rRNA genes from plastids. *ISME J.* 6, 481–492. doi: 10.1038/ismej.2011.117

Treusch, A. H., Vergin, K. L., Finlay, L. A., Donatz, M. G., Burton, R. M., Carlson, C. A., et al. (2009). Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J.* 3, 1148–1163. doi: 10.1038/ismej.2009.60

Unrein, F., Gasol, J. M., Not, F., Forn, I., and Massana, R. (2014). Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. *ISME J.* 8, 164–176. doi: 10.1038/ismej.2013.132

Unrein, F., Massana, R., Alonso-Sáez, L., and Gasol, J. M. (2007). Significant year-round effect of small mixotrophic flagellates on bacterioplankton in an oligotrophic coastal system. *Limnol. Oceanogr.* 52, 456–469. doi: 10.4319/lo.2007.52.1.0456

van de Poll, W. H., Kulk, G., Rozema, P. D., Brussaard, C. P., Visser, R. J., and Buma, A. G. (2018). Contrasting glacial meltwater effects on post-bloom phytoplankton on temporal and spatial scales in Kongsfjorden, Spitsbergen. *Elem. Sci. Anth.* 6:50. doi: 10.1525/elementa.307

Vergin, K. L., Beszteri, B., Monier, A., Thrash, J. C., Temperton, B., Treusch, A. H., et al. (2013). High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J.* 7, 1322–1332. doi: 10.1038/ismej.2013.32

Ward, B. A., and Follows, M. J. (2016). Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2958–2963. doi: 10.1073/pnas.1517118113

Wetherbee, R., Gornik, S. G., Grant, B., and Waller, R. F. (2015). *Andersenia*, a genus of filamentous, sand-dwelling Pelagophyceae from southeastern Australia. *Phycologia* 54, 35–48. doi: 10.2216/14-107.1

Wilken, S., Choi, C. J., and Worden, A. Z. (2020). Contrasting mixotrophic lifestyles reveal different ecological niches in two closely related marine protists. *J. Phycol.* 56, 52–67. doi: 10.1111/jpy.12920

Wilken, S., Yung, C. C. M., Hamilton, M., Hoadley, K., Nzongo, J., Eckmann, C., et al. (2019). The need to account for cell biology in characterizing predatory mixotrophs in aquatic environments. *Philos. Trans. R. Soc. B* 374:20190090. doi: 10.1098/rstb.2019.0090

Worden, A. Z. (2006). Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat. Microb. Ecol.* 43, 165–175. doi: 10.3354/ame043165

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., and Keeling, P. J. (2015). Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* 347:1257594. doi: 10.1126/science.1257594

Worden, A. Z., Janouskovec, J., McRose, D., Engman, A., Welsh, R. M., Malfatti, S., et al. (2012). Global distribution of a wild alga revealed by targeted metagenomics. *Curr. Biol.* 22, R675–R677.

Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352

Yin, Q., Fu, B., Li, B., Shi, X., Inagaki, F., and Zhang, X. H. (2013). Spatial variations in microbial community composition in surface seawater from the ultra-oligotrophic center to rim of the South Pacific Gyre. *PLoS One* 8:e55148. doi: 10.1371/journal.pone.0055148

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 11 November 2020
doi: 10.3389/fmars.2020.589566

Seasonal and Spatial Variation of Pelagic Microbial Food Web Structure in a Semi-Enclosed Temperate Bay

Haibo Li[1,2,3], Xue Chen[1,2,3], Michel Denis[4], Yuan Zhao[1,2,3], Lingfeng Huang[5,6], Zengjie Jiang[7], Wuchang Zhang[1,2,3] and Tian Xiao[1,2,3]*

[1] CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, [2] Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [3] Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China, [4] Aix-Marseille University, Toulon University, CNRS, IRD, Mediterranean Institute of Oceanography UM 110, Marseille, France, [5] Key Laboratory of the Ministry of Education for Coastal and Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Xiamen, China, [6] Fujian Provincial Key Laboratory of Coastal Ecology and Environmental Studies, Xiamen University, Xiamen, China, [7] Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China

OPEN ACCESS

Edited by:
Gipsi Lima Mendez, Catholic University of Louvain, Belgium

Reviewed by:
Yong Jiang, Ocean University of China, China
L. Antonio Cuevas, Coastal Ecosystems and Global Environmental Change Lab (ECCA Lab), Chile

***Correspondence:**
Wuchang Zhang
wuchangzhang@qdio.ac.cn

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Marine Science

Received: *31 July 2020*
Accepted: *20 October 2020*
Published: *11 November 2020*

Citation:
Li H, Chen X, Denis M, Zhao Y, Huang L, Jiang Z, Zhang W and Xiao T (2020) Seasonal and Spatial Variation of Pelagic Microbial Food Web Structure in a Semi-Enclosed Temperate Bay. Front. Mar. Sci. 7:589566. doi: 10.3389/fmars.2020.589566

Microbial food web (MFW) in the seawater encompasses the smallest organisms: viruses, autotrophic prokaryotes and heterotrophic prokaryotes (HP), nanoflagellates, eukaryotic phytoplankton and ciliates. For many years, scientists investigated the MFW structure differences in distinct water masses. However, the MFW structure seasonal variation in coastal areas remains poorly documented. In this study, we report on the seasonal and spatial variations of the MFW structure in the temperate Sanggou Bay in four successive seasons, from spring to winter. With a temperature increase from 1.90 to 24.20°C, HP biomass increased from 3.77 to 135.77 μg C dm^{-3}, almost covering the whole variation range for the global ocean. The autotrophic (AUTO) components, including *Synechococcus*, phototrophic picoeukaryotes (PEUK) and pigmented nanoflagellates (PNF), exhibited biomass variation ranges as large as previously reported. The MFW structure seasonal variation was driven by MFW relative biomasses (biomass ratios of MFW components to HP). With the increase of HP biomass, PNF and PEUK relative biomasses increased more rapidly than those of other groups while that of ciliates slightly decreased. The HETE:AUTO (biomass ratio of heterotroph to autotroph organisms) decreased with temperature, it was <1 in summer and >1 in other seasons. Cluster analyses distinguished Inside Bay and Outside Bay on the basis of hydrological characteristics. Consistently, the two subdivisions of Sanggou Bay exhibited different MFW structures as well as distinct tintinnid communities. The main MFW structure difference between Inside and Outside Bay was the biomass ratios of AUTO components to HP. Our results showed that the variations of autotrophic component biomass ratios relative to HP were the main factor responsible for the MFW structure seasonal variation. The spatial difference in MFW structure as well as in tintinnid taxonomic composition between Inside and Outside Bay was linked to the semienclosed nature of the Bay that does not favor efficient mixing with outside Yellow Sea waters.

Keywords: microbial food web, structure, seasonal, spatial, variation, Sanggou Bay

INTRODUCTION

The marine planktonic microbial food web (MFW) is a vital component of the pelagic ecosystem (Pomeroy, 1974; Azam et al., 1983). It encompasses the smallest marine organisms: viruses, autotrophic (*Prochlorococcus* and *Synechococcus*) prokaryotes and heterotrophic prokaryotes (HP), nanoflagellates, eukaryotic phytoplankton and ciliates (Sherr and Sherr, 1988; Kirchman, 2008). Abundances and biomasses of the MFW components may vary by 3 orders of magnitude in response to complex physiochemical and biological changes in different oceanic habitats. Very often they did not change in the same direction (increase or decrease) or with the same amplitude. Therefore, the spatial and temporal variations of MFW in different ocean areas are critical to our understanding of the pelagic ecosystem dynamics.

The biomass and abundance relationships among different MFW components (Microbial Structure) were compared in previous studies. According to the involved organisms, Microbial Structure was divided into several levels. Because picophytoplankton in one sample could be simultaneously analyzed by flow cytometry, the relative contribution of different picophytoplankton components (*Synechococcus*, *Prochlorococcus* and phototrophic picoeukaryotes) to total picophytoplankton was most frequently studied and was named as picophytoplankton community structure (e.g., Van Dongen-Vogels et al., 2012). When heterotrophic bacteria were included, it was called microbial community structure (Bock et al., 2018) or picoplankton community structure (e.g., Zubkov et al., 1998, 2000; Otero-Ferrer et al., 2018). The term MFW structure was used when ciliates and flagellates were incorporated into the microbial community (Garrison et al., 2000), which was scarcely studied.

The Microbial Structure temporal variation in a given study site is poorly documented. At two deep (>1000 m) Adriatic Sea stations, picoplankton community was dominated by SYN in December, while HP and PEUK were dominant in April (Šantić et al., 2019). Picoplankton community structure was investigated over 6 months at a station off Baja California (>100 m depth). Picoplankton biomass was positively and significantly correlated with Chl *a* (Linacre et al., 2010). In the Arabian Sea, HP biomass was higher than other MFW components in all seasons (Garrison et al., 2000). Temperate coastal waters exhibited the largest seasonal temperature differences in the world ocean (Mackas et al., 2012) which might generate MFW structure seasonal variations. However, the seasonal variation of MFW in temperate coastal waters has rarely been studied.

In contrast with the seasonal variation, the Microbial Structure spatial variation received more attention. At large scale level, different provinces exhibited different picoplankton community structures along the Atlantic Meridional Transect (Zubkov et al., 1998, 2000) and between coastal water and open sea in the central Adriatic Sea (Šantić et al., 2014). At mesoscale, microbial community structure was compared inside and outside of mesoscale eddies (Fernandez et al., 2008). At smaller scale, upwelling conditions influenced the picoplankton community structure in oceans south of Australia (Van Dongen-Vogels et al., 2012) and in coastal China Sea (Wu et al., 2014). Stratification (Bouman et al., 2011; Mena et al., 2019) and fronts (Pan et al., 2006) also separated water masses with different Microbial Structure. The coast geomorphology increases the diversity of habitat (Pierrot-Bults and Angel, 2012). The shoreline curvature forms bays in many places around the world ocean, which trap seawater inside and generate different environmental conditions with outside of the bay. As a result, the MFW inside and outside of the bay might be different, but this remains to be substantiated.

Sanggou Bay located in the east part of Shandong Peninsula, is a semi-enclosed (mouth of 11.5 km, area of 144 km^2) Bay in the Yellow Sea (**Figure 1**). The maximum and average depths of the Bay are 21 and 7.5 m, respectively. The seasonal temperature variation of 2–26°C in Sanggou Bay (Kuang et al., 1996) is one of the largest in the world ocean (Mackas et al., 2012). In fact, only very limited regions of the ocean have annual temperature variation larger than 12°C (Mackas et al., 2012). Sanggou Bay is among the most heavily exploited bays in the world, with suspended aquaculture rafts covering nearly 2/3 of the area. Each raft holds up cultured plants (kelps, asparagus) or animals (oysters, scallops) over most of the year. The cultured organisms could influence the MFW through releasing dissolved organic matter and direct grazing on the MFW (Lu et al., 2015a,b; Zhao et al., 2016). As the Yellow Sea waters intrude into the Bay from the north and outflow from the south slowly, water masses should be different inside and outside of the Bay (Zeng et al., 2015).

In this study, we analyzed the biomass of different MFW components in the Sanggou Bay in four successive seasons (from spring to winter) in order to identify (1) the seasonal variation of the MFW structure; (2) the difference of MFW structure inside and outside the Bay. Our hypotheses are that (1) variation trends and amplitudes of MFW components in different seasons should be different, which will lead to the seasonal variation of MFW structure; (2) differences of water masses inside and outside of the Bay will induce MFW structure differences.

MATERIALS AND METHODS

Investigated Area and Sampling Strategy

The Sanggou Bay MFW was assessed in the area located in 37.02–37.15°N and 122.45–122.65°E (**Figure 1**). Four cruises were conducted on spring (April 23–24, 2011), summer (August 3–4, 2011), autumn (October 26–27, 2011), and winter (January 5–6, 2012), on board the R/V *Lurongyuyang-65577*. The same 19 stations were occupied and sampled at each cruise (**Figure 1**).

Surface seawater samples were collected by bucket at each station. Temperature and salinity of surface seawater were determined by a portable water quality analyzer YSI (Professional Plus made in United States) as soon as the seawaters were collected by dropping the probe into the bucket seawater. Different subsamples were collected for chlorophyll *a* (Chl *a*) concentration determination, abundances of picoplankton, nanoflagellates and ciliates counting. Picoplankton included HP, *Synechococcus* (SYN) and phototrophic picoeukaryotes (PEUK). Nanoflagellates (NF) were divided into pigmented NF (PNF) and heterotrophic NF (HNF). We defined SYN, PEUK and PNF as

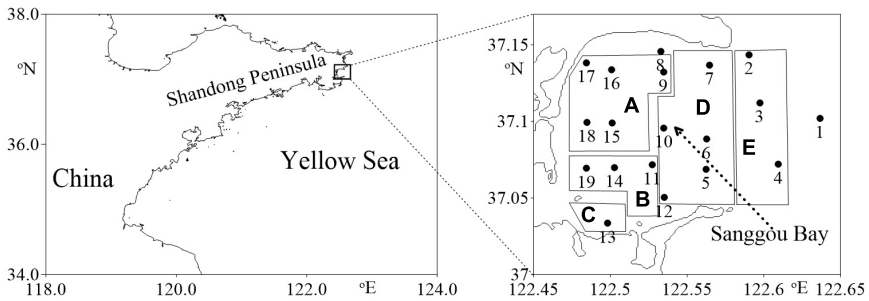


FIGURE 1 | Location of the sampling stations superimposed on the aquaculture area [**(A)** scallop monoculture area, **(B)** oyster monoculture area, **(C)** fish monoculture area, **(D)** kelp and bivalve polyculture area, **(E)** kelp *Saccharina japonica* monoculture area].

autotrophic organisms, while HP, HNF and ciliates were defined as heterotrophic organisms. ACLW-RS chlorophyll turbidity temperature sensor (ALEC Electronics Co., Ltd., Japan) with a precision of ± 0.1 μg dm^{-3} was used to determine the Chl *a* concentration. Detailed description of nutrient concentration determination was shown in Zhao et al. (2016).

Picoplankton

Picoplankton subsamples (5 cm^3) were fixed onboard with paraformaldehyde (final concentration 1%) immediately after collection and kept at room temperature for about 15 min, then freeze-trapped and stored in liquid nitrogen on the boat. Samples were stored at −80°C once in the laboratory where analyses were carried out within 3 months.

The seawater samples were thawed at room temperature before analysis. Picoplankton analyses were run with a FACS Vantage SE flow cytometer (Becton Dickinson) equipped with a water-cooled argon laser (488 nm, Coherent). To analyze autotrophic picoplankton, subsamples (1 cm^3) were initially supplemented with 1 mm^3 fluorescent beads suspension (2 μm, Polysciences) to be used as internal standard, and red fluorescence was set as the trigger signal to discard signals from HP and inorganic particles. SYN and PEUK were resolved on the basis of their side scatter and red fluorescence signals (Marie et al., 2000). For HP analysis, seawater subsamples (50 mm^3) were diluted five fold with TE buffer (Tris-EDTA, 100 mM Tris-Cl, 10 mM EDTA, pH = 8.0, Sigma, United States), then stained with the nucleic acid dye SYBR Green I (Molecular Probes, United States) (final dilution 10^{-4}, v/v) and let incubate 20 min in the dark before analysis. HP were resolved on the basis of their green fluorescence from staining and scatter properties (Marie et al., 2000). Data was collected and analyzed with CellQuest software (Version 3.3, Becton Dickinson). Biomass values of SYN, PEUK and HP were calculated by using the following conversion factors: 250 fg C cell^{-1} (Li et al., 1992), 1500 fg C cell^{-1} (Zubkov et al., 1998), 20 fg C cell^{-1} (Lee and Fuhrman, 1987), respectively.

Nanoflagellates

For the enumeration of NF cells, seawater samples were pre-filtered through a nylon mesh of 20 μm pore size by gravity, then fixed with cold glutaraldehyde (final concentration 0.5%, v/v). Subsamples (20 cm^3) were filtered onto 0.2 μm pore size black polycarbonate membrane filters at <100 mm Hg pressure. When 1 cm^3 of the sample remained in the funnel, turned off the vacuum pump and stained the sample with DAPI (final concentration 10 μg cm^{-3}) for 5 min, then turned on the pump again to let the residual liquid pass through the membrane filter completely. Then mounted the filter on a microscope slide, dropped a few drips of paraffin on the filter center and placed a coverslip on the top. Finally, stored the sample slide in the dark at −20°C immediately. NF cells were counted using epifluorescence microscopy (Leica DM4500B) at 1000× magnification. PNF and HNF were distinguished by the presence of red-autofluorescence in the former with a blue excitation filter set (Tsai et al., 2005). At least 30 fields of view were examined. Flagellates abundances were calculated from the average of cell counts made on duplicate samples. The length and width of a NF cell were measured on photomicrographs using the Leica DM 4500 self-carried software. At least 60 cells (PNF and HNF) were measured per sample. Cell volumes of NF individuals were estimated by assuming their nearest geometrical figures. The mean cell volumes were estimated and converted to carbon biomass using a conversion factor of 0.22 pg C μm^{-3} (Børsheim and Bratbak, 1987).

Ciliates

For ciliate counts, 1 dm^3 seawater samples were fixed with Lugol's solution (final concentration 1%). After 48 h settling, the supernatant was gently siphoned out to concentrate each ciliate sample to about 100 cm^3. A subsample of 20 or 25 cm^3 was settled in an Utermöhl counting chamber for at least 24 h, then examined using an inverted microscope (Olympus IX 71) at magnification of 100× or 200×. Aloricate ciliates and tintinnids were counted and their abundances were calculated respectively. Species of tintinnids were identified based on their morphological characteristics according to references (Kofoid and Campbell, 1929, 1939; Zhang et al., 2012). About 5–10 individuals of each species (a total of 75–150 individuals) were picked out randomly for calculating the biovolume. Ciliate dimensions including body length, oral diameter, etc., were measured and average biovolume of each taxon was estimated from appropriate geometric shapes. Biomass values of aloricate ciliates were calculated from their biovolume multiplied by a conversion factor (0.19 pg C μm^{-3})

(Putt and Stoecker, 1989). The tintinnid biomass was assumed to occupy 30% of the lorica volume (Gilron and Lynn, 1989).

Statistical Analysis

Correlation analysis was conducted using SPSS version 19. Spearman rank correlation analysis was used to estimate relationships between biological and environmental variables. Cluster analysis was performed using Primer 5.0 based on (i) environment variables (temperature, salinity), (ii) biomasses of six microbial groups, and (iii) abundances of different tintinnid species. Group-average linkage based on Bray-Curtis similarity matrix of fourth root transformed from original data was used.

RESULTS

Seasonal and Spatial Variation of Environmental Variables

Seasonally, temperature varied within a large range (**Table 1**) from 1.90°C in winter to 24.20°C in summer. Salinity fell in the range 26.17–31.57 over all four seasons, with the lowest salinity value observed in summer and the highest in winter. Over the four seasons, chlorophyll *a* (Chl *a*) concentration varied from 0.42 to 38.74 μg dm^{-3} (**Figure 2** and **Table 1**). Both the lowest and the highest concentration of dissolved inorganic nitrogen (DIN) occurred in autumn, which were 0.96 and 32.15 μM, respectively. Average DIN concentration was the lowest in summer and highest in autumn. PO_4^{3-} concentration fell in the range 0.00–0.23 μM over all four seasons, average concentration was the highest in spring and the lowest in winter (**Table 1**).

Temperature and salinity had obvious changes from outer part to inner part of the bay. The outer part had higher temperature in autumn and winter and vice versa in spring and summer (**Figure 2**). The inner part had lower salinity in summer, autumn and winter. The spatial salinity difference was very narrow in spring, autumn and winter (0.53, 0.74, and 0.25, respectively). In summer, salinity varied in the range 26.17–31.41 (**Table 1**). Similarly, Chl *a* concentrations showed inner-outer Bay gradients. In spring, Chl *a* concentration was high in both outer and inner parts of the bay with a narrow band of low values in between. In summer and autumn, Chl *a* concentration was very high in the inner part of the bay and sharply decreased toward outer part. In winter, it was higher in the outer part than in the inner part (**Figure 2**).

Seasonal Variation of Microbial Food Web Components

The six MFW component abundances and biomasses varied with seasons. HP abundances and biomasses were in the ranges 1.88–67.89 × 10^5 cells cm^{-3} and 3.77–135.77 μg C dm^{-3}, respectively (**Table 1**). Biomasses of all MFW components showed positive and significant correlations with temperature and Chl *a* (all $P < 0.01$; **Supplementary Table S1**). However, the amplitude of their increase was not proportional to that of temperature (**Figure 3**). Biomasses of most MFW components except SYN showed negative and significant correlations with salinity ($P < 0.01$; **Supplementary Table S1**).

The autotrophic (AUTO) and heterotrophic (HETE) biomasses were in the ranges 1.50–720.27 and 5.40–227.29 μg C dm^{-3}, respectively. The AUTO and HETE biomasses showed significant positive correlations with temperature (all $P < 0.01$; **Supplementary Table S1**), but AUTO biomass increased faster than that of HETE (**Figure 3**). Consequently, the biomass ratio HETE:AUTO showed significant and negative correlations with temperature ($P < 0.01$; **Supplementary Table S2**).

Totally 26 tintinnid species belonging to 8 genera were identified with 4, 14, 13, and 8 species in spring, summer, autumn and winter, respectively. Among them, only *Tintinnopsis beroidea* was observed in all four seasons, *T. acuminate* occurred in 3 seasons, 7 species occurred in 2 seasons. Total tintinnid abundance was higher in summer and autumn than in winter and spring (**Figure 4** and **Table 1**).

To assess the MFW structure, the biomass values of the MFW components [including *Synechococcus* (SYN), phototrophic picoeukaryotes (PEUK), pigmented nanoflagellates (PNF), heterotrophic nanoflagellates (HNF)] and ciliates were normalized by that of HP belonging to the same sample. Among them, PNF relative biomass had the largest variation range from 0.06 (St.7 in winter) to 4.20 (St. 18 in summer). SYN relative biomass had the narrowest variation range from 0.00 (St.3 in spring) to 0.64 (St.11 in autumn). Ciliate relative biomass was in the range 0.02 to 0.94, while PEUK relative biomass was in the range of 0.04–3.57 (**Figures 5**, **6**). The relative biomasses of most MFW components showed significant positive correlations with temperature, Chl *a* and HP biomass except ciliates which was significantly and negatively correlated with those environmental variables ($P < 0.01$; **Supplementary Table S2**). All autotrophic (AUTO) components relative biomasses increased with HP biomass at a higher rate than HNF and ciliates. The only exception was that ciliate relative biomass decreased when HP biomass increased (**Figure 5**).

The annual averaged MFW structure was thus HP: SYN: PEUK: PNF: HNF: ciliates = 1: 0.18: 1.40: 1.35: 0.44: 0.13 (**Figure 6**). The average relative biomass (ARB) seasonal variation trends of different organisms were distinct. ARBs of SYN and PEUK increased from spring to autumn and decreased in winter. ARBs of PNF and HNF increased from spring to summer, then decreased from summer to winter. ARB of ciliates decreased slightly from spring to summer, then increased until winter (**Figure 6**).

ARBs of all MFW components were <1 in spring and winter. ARBs of PEUK and PNF were >1, while SYN, HNF and ciliates were <1 in summer. In autumn, only ARB of PEUK was >1 (**Figure 6**). SYN ARB was lower than that of PEUK and PNF at all seasons, and was lower than that of HNF in spring, summer and winter. PEUK ARB was lower than that of PNF in spring and summer, but higher in autumn and winter. In spring and summer, PEUK ARB was higher than that of other organisms, whereas in autumn and winter, PNF ARB was the highest. PEUK and PNF ARBs were >1 over a year, while SYN, HNF and ciliate ARBs were <1 (**Figure 6**).

TABLE 1 | Maximum, minimum, and average values of temperature (T), salinity (S), Chlorophyll *a* concentration (Chl *a*), nutrient concentrations, microbial food web component abundances and biomasses in Sanggou Bay at four consecutive seasons.

		Spring		Summer		Autumn		Winter	
		Range	**Mean ± SD**	**Range**	**Mean ± SD**	**Range**	**Mean ± SD**	**Range**	**Mean ± SD**
T	°C	5.50 ~ 11.61	9.00 ± 2.12	18.50 ~ 24.20	21.36 ± 2.02	7.50 ~ 14.90	16.47 ± 0.79	1.90 ~ 5.60	3.76 ± 1.22
S		30.21 ~ 30.74	30.51 ± 0.12	26.17 ~ 31.41	29.39 ± 1.78	30.57 ~ 31.37	31.18 ± 0.23	31.32 ~ 31.57	31.52 ± 0.06
DIN	μM	2.38 ~ 10.79	6.24 ± 2.63	1.57 ~ 11.75	4.83 ± 2.69	0.96 ~ 32.15	10.44 ± 10.10	2.34 ~ 7.65	4.88 ± 1.72
PO_4^{3-}	μM	0.03 ~ 0.23	0.11 ± 0.07	0.02~0.10	0.04 ± 0.02	0.00 ~ 0.05	0.02 ± 0.02	0.00 ~ 0.06	0.02 ± 0.01
Chl *a*	μg dm^{-3}	0.70 ~ 2.70	1.27 ± 0.55	5.72 ~ 38.74	14.41 ± 9.74	0.67 ~ 19.62	6.49 ± 6.01	0.42 ~ 2.88	0.90 ± 0.55
SYN	Abundance (10^3 cells cm^{-3})	0.01 ~ 0.11	0.05 ± 0.03	21.31 ~ 70.64	33.20 ± 11.86	4.58 ~ 264.18	84.06 ± 78.59	0.09 ~ 0.83	0.51 ± 0.20
	Biomass (μg C dm^{-3})	0.00 ~ 0.03	0.01 ± 0.01	5.32 ~ 17.66	8.30 ± 2.97	1.14 ~ 66.04	21.01 ± 19.65	0.02 ~ 0.21	0.13 ± 0.05
PEUK	Abundance (10^3 cells cm^{-3})	0.40 ~ 24.51	9.15 ± 7.78	43.46 ~ 167.72	82.57 ± 30.82	2.51 ~ 244.69	57.42 ± 64.63	0.62 ~ 5.74	1.80 ± 1.41
	Biomass (μg C dm^{-3})	0.59 ~ 36.77	13.73 ± 11.67	65.18 ~ 251.59	123.85 ± 46.22	3.76 ~ 367.03	86.13 ± 99.94	0.92 ~ 8.60	2.70 ± 2.11
HP	Abundance (10^5 cells cm^{-3})	7.78 ~ 20.33	15.43 ± 4.06	13.43 ~ 67.89	40.77 ± 17.43	6.02 ~ 64.30	23.35 ± 16.80	1.88 ~ 4.08	3.00 ± 0.60
	Biomass (μg C dm^{-3})	15.56 ~ 40.65	30.87 ± 8.12	26.86 ~ 135.77	81.54 ± 34.81	12.03 ~ 128.60	46.71 ± 33.61	3.77 ~ 8.16	6.00 ± 1.20
HNF	Abundance (ind. cm^{-3})	746 ~ 2251	1430 ± 376	3251 ~ 35183	10841 ± 7886	549 ~ 7111	2298 ± 1620	171 ~ 831	418 ± 201
	Biomass (μg C dm^{-3})	3.19 ~ 10.69	5.26 ± 1.76	14.57 ~ 116.19	50.86 ± 28.07	1.83 ~ 33.90	13.75 ± 9.40	0.38 ~ 2.13	0.94 ± 0.41
PNF	Abundance (ind. cm^{-3})	289 ~ 15797	1951 ± 3764	3516 ~ 136442	26671 ± 30415	684 ~ 12782	5647 ± 4263	144 ~ 643	333 ± 137
	Biomass (μg C dm^{-3})	3.44 ~ 83.70	20.15 ± 19.84	19.42 ~ 512.52	166.08 ± 134.38	4.08 ~ 67.85	30.59 ± 18.56	0.38 ~ 8.43	2.48 ± 2.17
Tintinnids	Abundance (ind. dm^{-3})	0 ~ 380	95 ± 86	15 ~ 7144	2054 ± 2259	127 ~ 1760	602 ± 498	0.00 ~ 163	42.01 ± 51.06
	Biomass (μg C dm^{-3})	0.00 ~ 0.76	0.14 ± 0.17	0.02 ~ 9.11	2.43 ± 2.82	0.06 ~ 2.45	0.63 ± 0.62	0.00 ~ 0.38	0.07 ± 0.10
Total ciliates	Abundance (ind. dm^{-3})	500 ~ 61667	12461 ± 16542	1174 ~ 42315	9671 ± 11483	993 ~ 30038	10061 ± 9314	544 ~ 10730	3348 ± 2964
	Biomass (μg C dm^{-3})	0.76 ~ 7.12	3.71 ± 1.72	0.63 ~ 33.09	8.70 ± 8.23	0.78 ~ 20.70	7.20 ± 6.38	0.44 ~ 3.79	1.41 ± 0.91

DIN, dissolved inorganic nitrogen; SYN, Synechococcus; PEUK, phototrophic picoeukaryotes; HP, heterotrophic prokaryotes; HNF, heterotrophic nanoflagellates; PNF, pigmented nanoflagellates. SD, standard deviation.

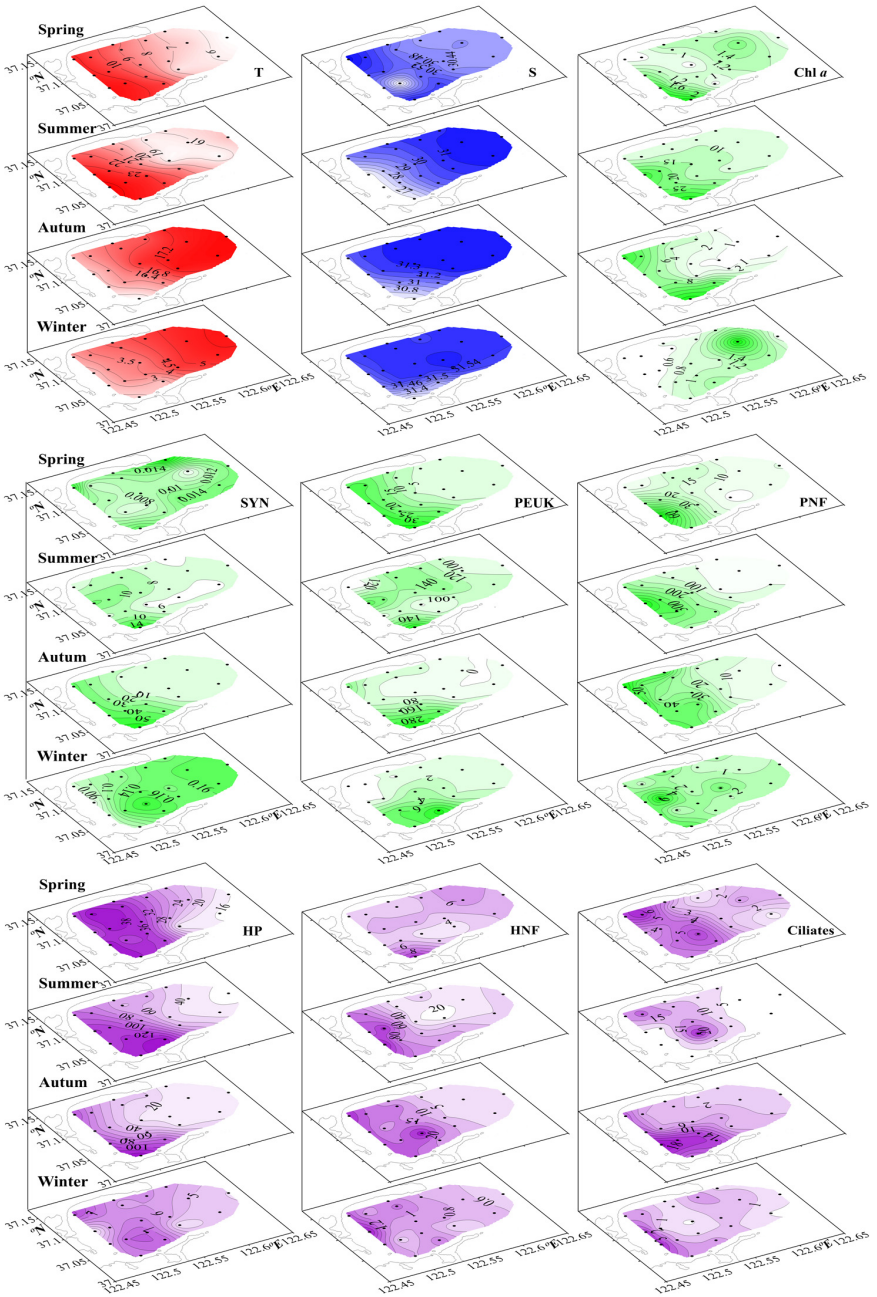


FIGURE 2 | Seasonal surface distribution of temperature (T, °C), salinity (S), Chlorophyll *a* concentration (Chl *a*, μg dm^{-3}), and microbial food wed component biomasses (μg C dm^{-3}) through four successive seasons, spring to winter. SYN, *Synechococcus*; PEUK, phototrophic picoeukaryotes; PNF, pigmented nanoflagellates; HP, heterotrophic prokaryotes; HNF, heterotrophic nanoflagellates.

The biomass ratio of HP:AUTO decreased with the increase of AUTO biomass. When the biomass ratio of HP:AUTO was about 1, the AUTO biomass was in the range of 6–35 μg C dm^{-3} (**Figure 5**). In winter and spring, the biomass ratio HETE:AUTO was >1, while it was <1 in summer. When the biomass ratio HETE:AUTO was about 1, the HP and AUTO biomasses were in the range 5–13 and 10–40 μg C dm^{-3}, respectively (**Figure 5**).

Biomass ratios of HP, HNF and ciliates to HETE were in the ranges of 0.40–0.86, 0.05–0.58, and 0.01–0.37, respectively (**Figure 7**). The relationships between biomass ratios of different HETE components to HETE were different: HP and ciliates were significantly and negatively correlated with HETE, while HNF was significantly and positively correlated with HETE (all $P < 0.01$; **Supplementary Table S2**). With the increase of preys

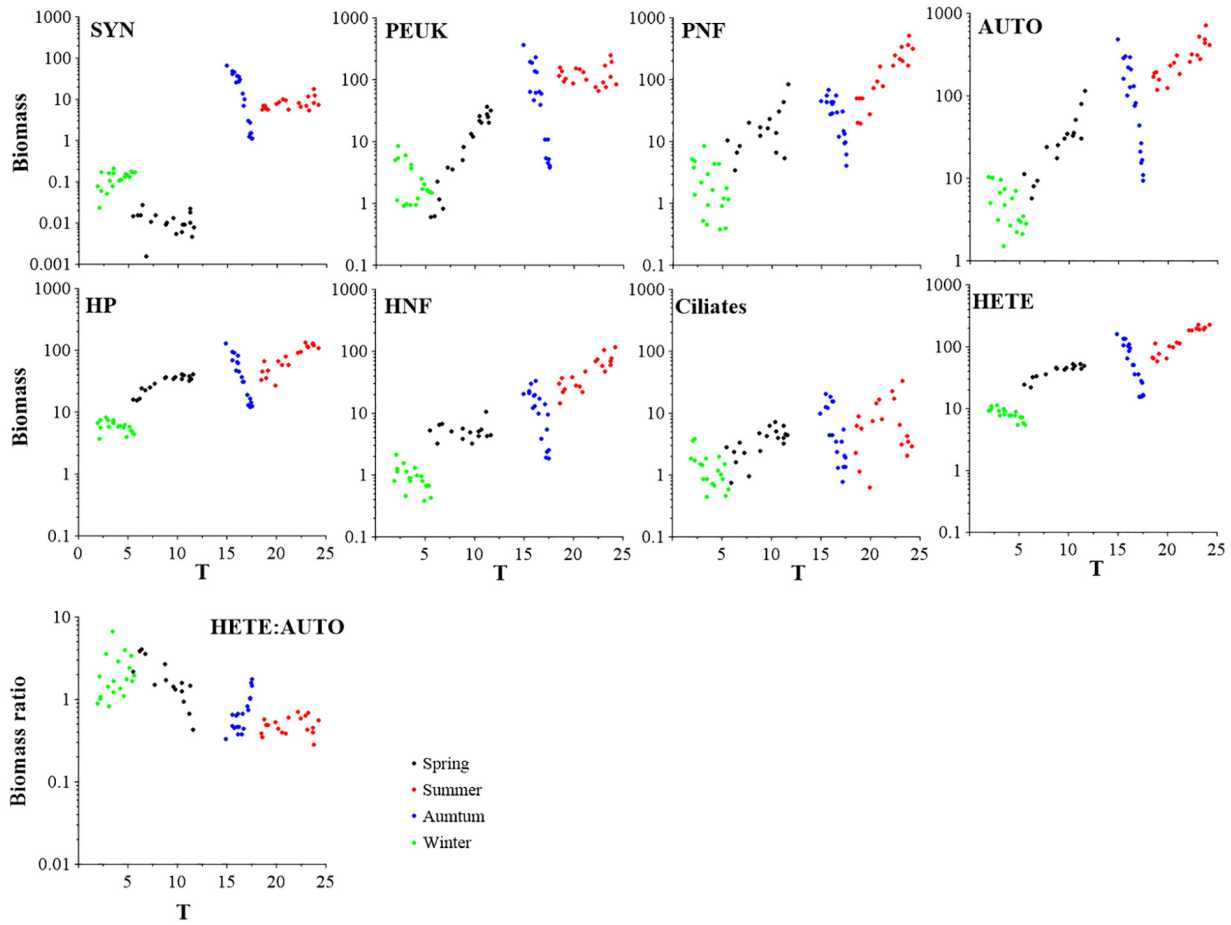


FIGURE 3 | Variations of different microbial food web component biomasses (μg C dm^{-3}) and biomass ratio of HETE to AUTO with respect to temperature increase. SYN, *Synechococcus*; PEUK, phototrophic picoeukaryotes; PNF, pigmented nanoflagellates; HP, heterotrophic prokaryotes; HNF, heterotrophic nanoflagellates; AUTO, autotrophic organisms; HETE, heterotrophic organisms.

(HP, SYN, PEUK, PNF, and HNF) biomass, biomass ratio of Ciliates:Preys decreased (**Figure 7**).

Spatial Variation of MFW Structure

Biomasses of MFW component showed inner-outer Bay gradients (**Figure 2**). For AUTO components, higher SYN biomass was observed in the outer part in spring and winter, and vice versa in summer and autumn. PEUK and PNF biomasses always exhibited higher values in the inner part (**Figure 2**). For HETE components, HP biomass showed higher values in the inner part than in the outer part at all seasons. In summer, autumn and winter, HNF biomass was higher in the inner part than in the outer part. In spring, HNF biomass reached higher values in both inner and outer parts with a narrow low-value band in between. The ciliate biomass exhibited higher values in the inner part over all seasons (**Figure 2**).

Although some tintinnid species occurred all over the bay (e.g., *T. rapa* in spring, *T. beroidea* in summer and autumn), the most conspicuous characteristic was that the inner part and outer parts of the bay had different species in every season. In spring, *T. acuminata* and *T. brasiliensis* occurred in the outer and inner parts of the bay, respectively. *T. beroidea* occurred in 3 stations in the outer and 2 stations in the inner part of bay, respectively. In summer, *T. kofoid* only occurred in the inner part, *Tintinnidium primitivum* mainly occurred in the outer part. *T. acuminate* was restricted to several stations in the north part of the bay. In autumn, *E. tubulosus* appeared in the inner part. *T. nana* and *T. primitivum* occurred in the outer part. In winter, high abundance of *T. nana* occurred in the outer part. *T. beroidea* occurred in 3 stations in the outer part and *T. acuminata* only occurred once in the outer part (**Figure 4**).

In each season, three distinct cluster analyses were run on three data sets of (i) environmental conditions including temperature and salinity (**Supplementary Figure S1**), (ii) biomasses of the MFW components (**Supplementary Figure S2**) and (iii) abundances of the tintinnid species (**Supplementary Figure S3**). All cluster analyses could divide the stations into similar Inner Bay cluster and Outer Bay cluster. Though the shape and position of the division lines between clusters were different

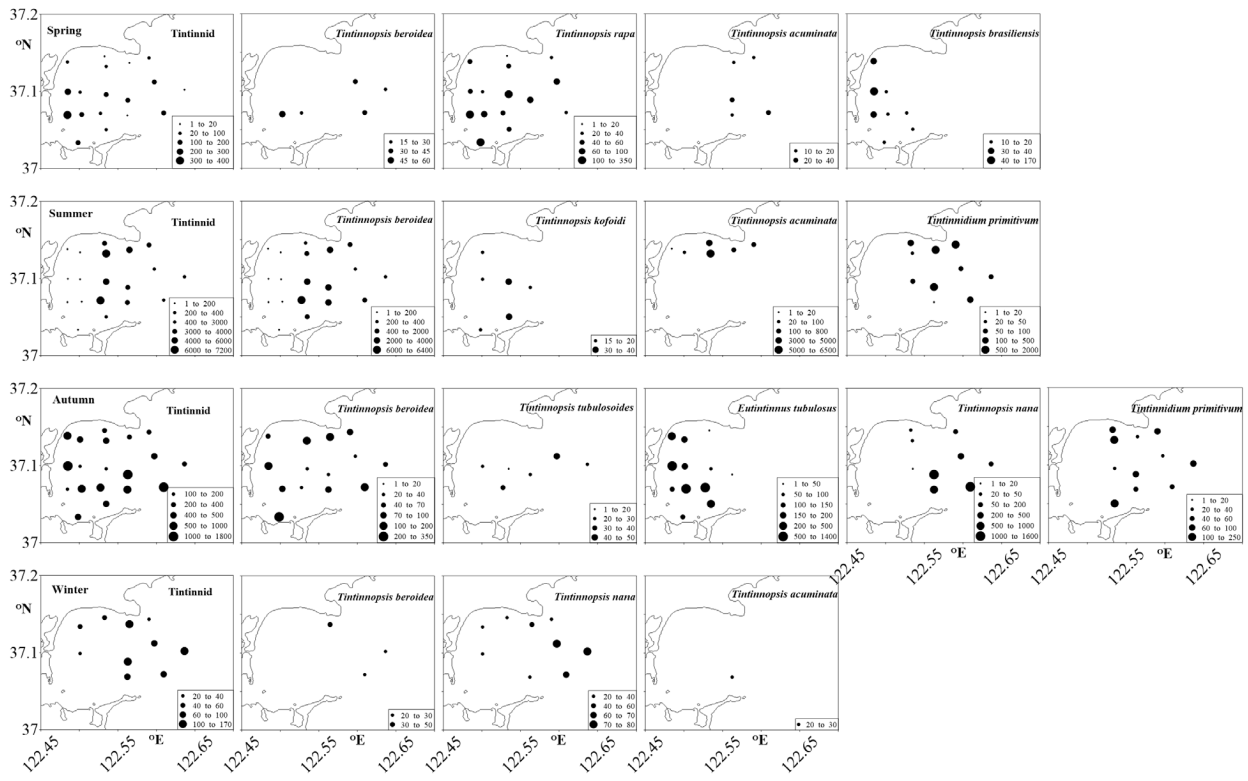


FIGURE 4 | Tintinnid and main species abundance (ind. dm^{-3}) surface distribution in four successive seasons, spring to winter.

from one season to another, they were generally similar to the nearest isothermal (**Figure 8**).

Differences of MFW structure between Inner Bay and Outer Bay were mainly caused by PEUK, PNF and HNF. In spring, relative biomass of PNF and PEUK were higher in Inner than Outer Bay. In summer, PEUK relative biomass was higher in Outer than Inner Bay. The HNF biomass was higher in Outer than Inner Bay in spring, but higher in Inner than Outer Bay in summer.

In autumn, the difference between Inner and Outer Bay was mainly due to SYN, PEUK and PNF. SYN and PEUK relative biomasses were higher in Inner than in Outer Bay, while PNF relative biomass was higher in Outer than Inner Bay. In winter, the difference between Inner and Outer Bay was mainly caused by PEUK and PNF, which also had higher relative biomasses in Inner than in Outer Bay (**Figure 9**).

DISCUSSION

Large Seasonal Variation Range of Environmental Variables and Microbial Food Web Components in the Sanggou Bay

Only data of surface waters were presented in this study. Most of the previous studies in the Sanggou Bay only sampled the surface layers because the bay is shallow and waters in the upper 10 m was well mixed in most seasons (Jihong Zhang, self-communication). The surface water temperature seasonal range of the Sanggou Bay was about 22°C, which was among the largest in the world ocean (Mackas et al., 2012). Salinity range was 26.17 and 31.57 in our study, this was a little larger than the range of 28.23–31.31 reported by Jiang et al. (2017) in the same Bay. The Sanggou Bay exhibited different trophic regimes, from oligotrophy to eutrophication with a large Chlorophyll *a* (Chl *a*) concentration-range from 0.42 to 38.74 $\mu g\ dm^{-3}$ in present study and 0.10–20.46 $\mu g\ dm^{-3}$ in Jiang et al. (2017). Phytoplankton with size fraction $< 20\ \mu m$ was the most important contributor to total Chl *a* in the Sanggou Bay, and contributed about 60.2, 83.0, 78.3, and 75.0% to total Chl *a* in spring, summer, autumn and winter, respectively (Jiang et al., 2017). The average concentration of DIN was highest in autumn in our study. This is consistent with Jiang et al. (2017) who observed the highest value of average DIN concentration in autumn. Seasonal variation of PO_4^{3-} concentration also presented the same trend in our study as in Jiang et al. (2017): decreasing from spring to winter.

These large gradients corresponded to a large range of HP abundance (1.88–67.89 $\times\ 10^5$ cells cm^{-3}) similar to the HP abundance ranges 1–152 and 1–100 $\times\ 10^5$ cells cm^{-3} in the world ocean, reported by Sanders et al. (1992) and Gasol (1994), respectively. The autotrophic (AUTO) organism biomass in our study, which was in the range of 1.50–720.27 $\mu g\ C\ dm^{-3}$,

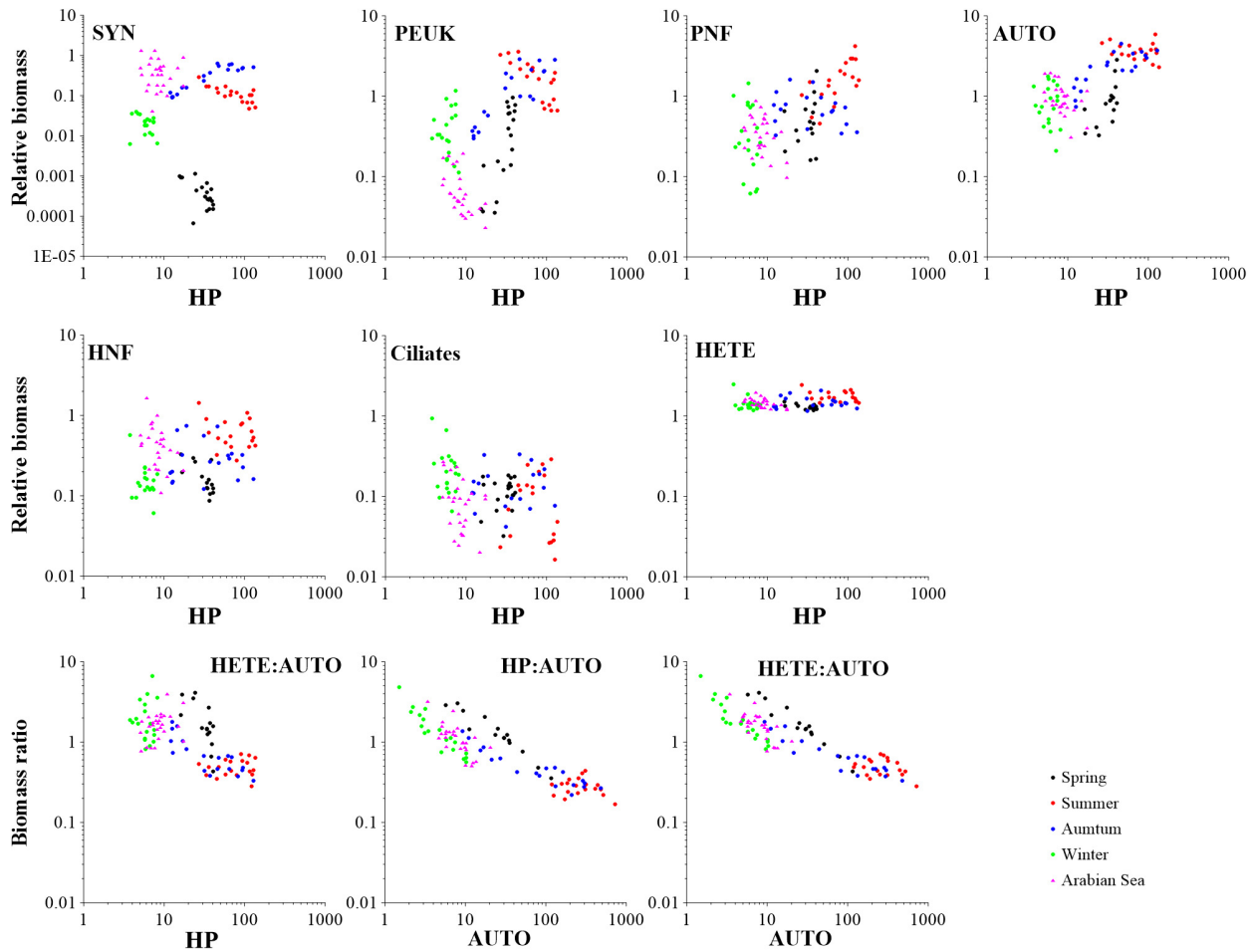


FIGURE 5 | Microbial food web component biomass variations relative to HP biomass (μg C dm^{-3}) changes. HETE to AUTO biomass ratio variations as a function of HP biomass increase and AUTO biomass (μg C dm^{-3}) in Sanggou Bay and Arabian Sea. The data of Arabian Sea were extracted from Garrison et al. (2000). HP, heterotrophic prokaryotes; SYN, *Synechococcus*; PEUK, phototrophic picoeukaryotes; PNF, pigmented nanoflagellates; HNF, heterotrophic nanoflagellates; AUTO, autotrophic organisms; HETE, heterotrophic organisms.

almost covered the full range of phytoplankton biomass in Herndl (1991), from 6.9 to 415.5 μg C dm^{-3}. These large ranges of biomass variations showed that the MFW components in Sanggou Bay experienced very large seasonal changes.

Relative Biomass Variation of Different MFW Components in Four Seasons

We found that relative biomass variations of different MFW components exhibited different trends with respect to HP biomass increase. The diverse variations of different groups relative to HP biomass express the seasonal variation of the MFW. To our knowledge, this finding was not reported previously.

The work of Garrison et al. (2000) was the only one trying to compare different MFW structures in four cruises in the Arabian Sea. The HP biomass in the Arabian Sea was in the range of 5.1–17.4 μg C dm^{-3} in the upper 100 m, which was very similar to that in spring and winter in Sanggou Bay (**Table 1**). The average relative biomasses (ARB) of each MFW component was <1 at all seasons in the Arabian Sea, which is equivalent to those in spring and winter in the Sanggou Bay (**Figure 6**). Therefore, the seasonal variation in Sanggou Bay is larger than that in the Arabian Sea. The MFW structure differences in Arabian Sea and Sanggou Bay may be caused by lower values and a smaller range of HP biomass in the Arabian Sea than in the Sanggou Bay.

ARB of PEUK was higher than ARB of SYN in all seasons in the Sanggou Bay, but it was the opposite in the Arabian Sea (Garrison et al., 2000) (**Figure 6**). This obvious difference between the MFW structure in Sanggou Bay and Arabian Sea may indicate different MFWs in coastal and open waters. Previous studies on picoplankton (picophytoplankton) community structure in coastal area were focused on spatial variation, but seasonal change was very seldom investigated. The dominance of PEUK in Sanggou Bay is consistent with previous studies in coastal areas (Pan et al., 2006; Linacre et al., 2010; Bock et al., 2018; Patten et al., 2018; Otero-Ferrer et al., 2018; Mena et al., 2019). In the deep waters of Adriatic Sea, SYN

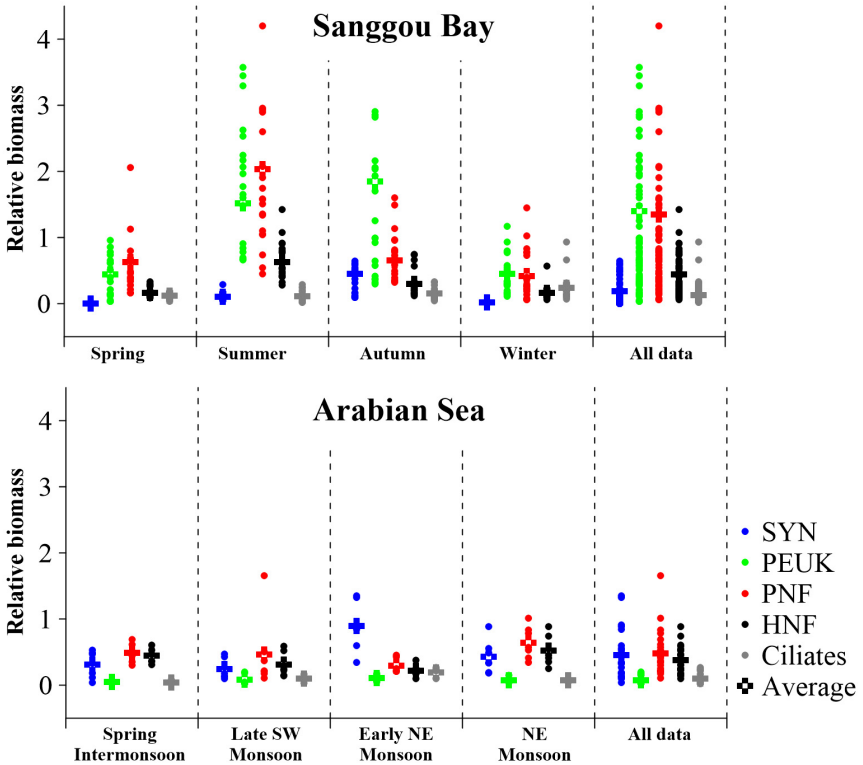


FIGURE 6 | Microbial food web structures in Sanggou Bay and Arabian Sea in four successive seasons and Monsoon stages, respectively. The data of Arabian Sea were extracted from Garrison et al. (2000). The relative biomasses of the MFW components were normalized by the HP biomass belonging to the same sample. SYN, *Synechococcus*; PEUK, phototrophic picoeukaryotes; PNF, pigmented nanoflagellates; HNF, heterotrophic nanoflagellates.

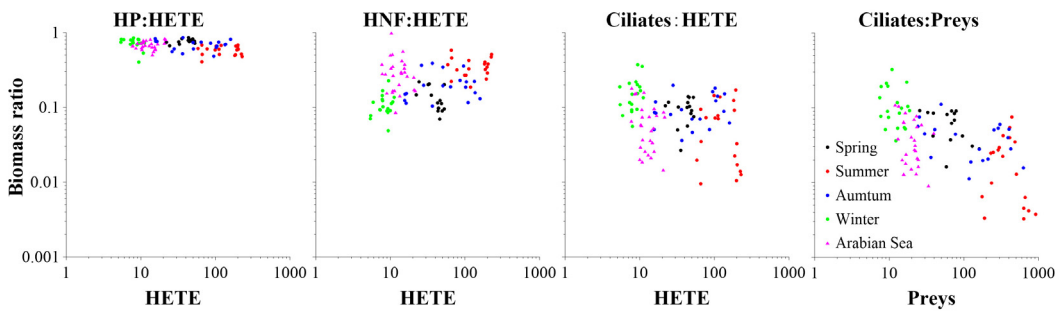


FIGURE 7 | Biomass ratio variations of HP, HNF and ciliates to HETE with the increase of HETE biomass (μg C dm^{-3}), and biomass ratio of ciliates to preys with the increase of prey biomass in Sanggou Bay and Arabian Sea. The data of Arabian Sea were extracted from Garrison et al. (2000). HP, heterotrophic prokaryotes; HNF, heterotrophic nanoflagellates; HETE, heterotrophic organisms.

dominated in December while HP and PEUK dominated in April (Šantić et al., 2019).

Relationships Between MFW Components

In our study, both AUTO and HETE biomasses increased with temperature, but with the increase of AUTO biomass, the biomass ratio of HETE:AUTO decreased. This was consistent with the work of Gasol et al. (1997) reporting that unproductive waters were characterized by high relative HETE biomass, whereas productive waters were distinguished by a smaller contribution of HETE to community biomass.

As the biomass ratio of HP:AUTO decreased with the increase of AUTO biomass, the AUTO biomass ranges where the biomass ratio of HP:AUTO was about 1 were considered as boundary below which a system can be considered oligotrophic enough for bacteria to dominate phytoplankton (Li et al., 1993). In our study, when the AUTO biomass was ranged in 6–35 μg C dm^{-3}, the biomass ratio of HP:AUTO was about 1. This was similar

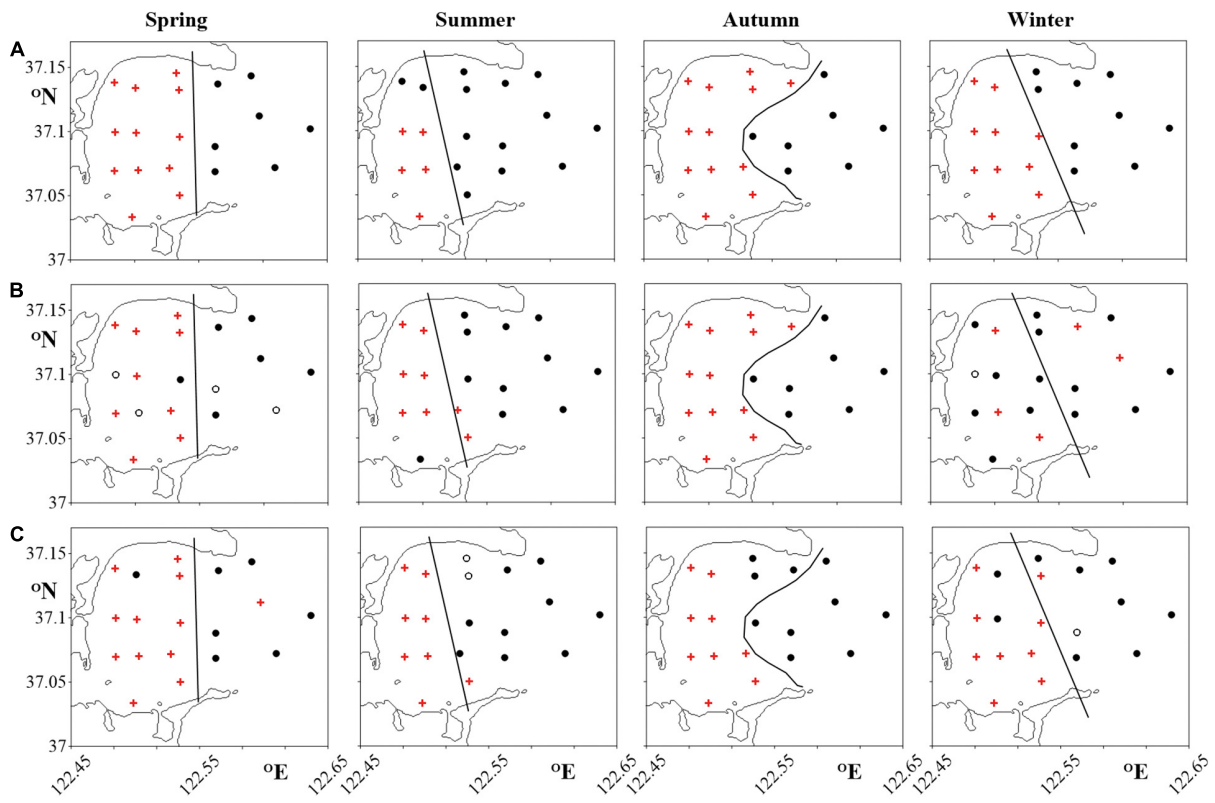


FIGURE 8 | Cluster analysis of stations according to **(A)** environmental conditions; **(B)** microbial food web component biomasses; **(C)** tintinnid abundances. •: Outer Bay cluster stations; +: Inner Bay cluster stations; ○: stations neither belonging to Inner Bay nor Outer Bay clusters. The solid lines show the dividing position between Inner Bay and Outer Bay in **Figure 9** mainly according to **(A)** in the four seasons.

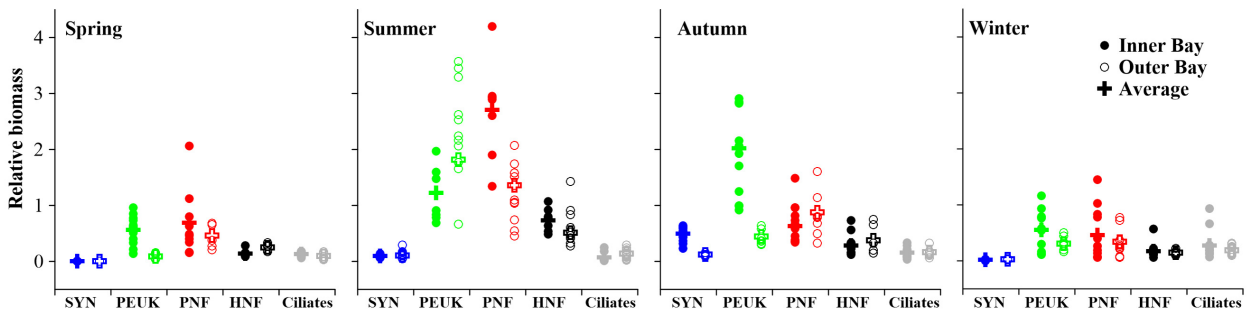


FIGURE 9 | Microbial food web structures in Inner and Outer Sanggou Bay in four successive seasons, spring to winter. The Inner and Outer Bay boundary were shown as solid lines in **Figure 8**. Relative biomasses of MFW components were normalized by heterotrophic prokaryotes biomass belonging to the same sample. SYN, *Synechococcus*; PEUK, phototrophic picoeukaryotes; PNF, pigmented nanoflagellates; HNF, heterotrophic nanoflagellates.

with phytoplankton biomass range of 10 and 20 μg C dm^{-3} in Li et al. (1993).

Although flagellates and ciliates were important components of MFW, their dynamics in relation to picoplankton were very rarely studied. We found that with the increase of the HETE biomass, biomass ratio of HP and ciliates to total HETE biomass decreased slightly, while HNF biomass ratio to total HETE increased slightly (**Figure 7**). In the absence of similar investigation in the literature, we could estimate the proportion of different HETE component biomasses in total HETE biomass from published figures. We thus found that the biomass ratio of HP:HETE in the present work (range 0.40–0.86) was similar to that in the northeastern Atlantic Ocean (range 0.70–0.87), as derived from the work of Maixandeau et al. (2005).

In our study, HP abundances and biomasses were in the ranges 1.88–67.89 × 10^5 cells cm^{-3} and 3.77–135.77 μg C

dm^{-3}, respectively (**Table 1**). Ciliate relative biomass was in the range 0.02–0.94, and its annual ARB was 0.16 (**Figures 5**, **6**). This was similar to that in the Lower St. Lawrence Estuary where HP abundance and biomass were in the ranges 1.2–67.9 × 10^5 cells cm^{-3} and 0.55–146.70 μg C dm^{-3}, respectively. Meanwhile, ciliate biomass represented about 22% of HP biomass (Sime-Ngando et al., 1995). The value in our study was a little higher than that in the Arabian Sea, where ciliate relative biomass was in the range 0.02–0.26 (average 0.10) (**Figures 5**, **6**) (Garrison et al., 2000).

With the increase of prey biomass, the biomass ratio of Ciliates:Preys decreased in our study (**Figure 7**). This is inconsistent with that in the deep east part of East China Sea where the ratio of ciliate biomass was larger under higher prey candidate conditions (Suzuki and Miyabe, 2007). The biomass range of ciliates (0.44–33.09 μg C dm^{-3}) in our study (**Table 1**) was similar to that in the deep east part of East China Sea (0.03–29.26 μg C dm^{-3}) (Suzuki and Miyabe, 2007). The study in the East China Sea was based on the biovolumes of ciliates and prey candidates, while our study was based on the biomasses of ciliates and prey candidates. This difference may alter the comparison, but does not affect the variation trends of both studies. In addition, the present study was based on surface data, while Suzuki and Miyabe (2007) addressed the vertical variation of ciliates and their prey candidates. This may be the main reason why the biomass ratio variation trends of ciliates to prey candidates appear different in our study and that in East China Sea.

Inner Bay and Outer Bay Difference of MFW Structure

More and more studies have indicated that MFW component distribution may be significantly influenced by water masses (Hashihama et al., 2008; Samo et al., 2012; Zhao et al., 2013; Bachy et al., 2014; Sun et al., 2020; Yang et al., 2020). Among all the seawater variables, temperature and salinity are the conservation properties used to distinguish water masses (Emery, 2001). Tintinnids were also used as water mass indicators (Modigh et al., 2003; Kim et al., 2012). In the present study, we divided the Sanggou Bay into the Inner and Outer Bay mainly according to cluster analysis based on temperature and salinity, as well as MFW component biomasses and tintinnid abundance (**Figure 8**).

Our results showed that in every season, the Inner and Outer Bay had different tintinnid communities (**Figure 8**). Similar to the division of tintinnid community, MFW structure also exhibited spatial partitioning: MFW structure was different in Inner and Outer Bay in every season (**Figure 8**), although the cluster analysis results based on different variables were not exactly coinciding. These divisions are generally corresponding to hydrobiological divisions (**Figure 8**).

In the Sanggou Bay, water intruded into the Bay mainly through north opening, while water was flowing out of the Bay through the opening south part (Zeng et al., 2015). The division into Inner and Outer Bay water masses showed obvious "north in, south out" pattern in summer, autumn and winter (**Figure 9**). This division pattern might be caused by exchange pattern between Inner Bay and Outer Bay. In addition to the water exchange between Inner and Outer parts of the Bay, the aquaculture activity also had influence on MFW components (Lu et al., 2015a,b). In our study, the division line between the Inner and Outer parts of the MFW in the four seasons did not match the division of aquaculture area. Therefore, the water exchange process between Inner and Outer parts of the Sanggou Bay would be the main factor influencing the MFW structure.

Tintinnids was the only MFW component that could be identified at the species level which was instrumental in distinguishing Inner Bay from Outer Bay. In the coastal area, neritic tintinnid was centered in waters at 30–50 m depth (Chen et al., 2018). Neritic species intruded from the coastal center area to the Sanggou Bay. Temperature and salinity had obvious influence on tintinnid community structure in coastal and estuarine waters (Feng et al., 2018; Li et al., 2019). As one important contributor of planktonic ciliate, aloricate ciliate distribution also showed obvious seasonal variations and was significantly influenced by water masses (Jiang et al., 2011; Yu et al., 2014). In the Jiaozhou Bay, northern China, species numbers, abundances and biomasses of both tintinnids and aloricate ciliates exhibited unimodal variations with peaks occurring in August (Jiang et al., 2011). In the Yellow Sea, ciliate (including tintinnid and aloricate ciliate) abundances and biomasses showed a clear annual cycle (Yu et al., 2014). Seasonal and spatial variations of aloricate ciliate community were not mentioned in our study. However, we speculate that aloricate ciliates might have similar seasonal and spatial variations as tintinnids.

CONCLUSION

Our study showed that MFW structure in a temperate coastal bay had different modes in different seasons. The different amplitudes of autotrophic component variations were the main reason for the seasonal change of MFW structure. Spatially, tintinnid community and MFW structure were different in Inner and Outer parts of the Sanggou Bay in every season. Those spatial differences were generated by differences in water masses in Inner and Outer parts of the Sanggou Bay.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

WZ and TX designed the experimental scheme and supervised the entire project. HL and XC integrated the all data and wrote the manuscript. MD and ZJ reviewed the manuscript. YZ and LH provided the data of picoplankton and nanoflagellate, respectively. All the authors approved the final manuscript.

FUNDING

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA19060201), Creative Team Project of the Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology (No. LMEES-CTSP-2018-1), CNRS-NSFC Joint Research Projects Program (No. 41711530149) and 2017-2019 Sino-French Cai Yuanpei Programme.

ACKNOWLEDGMENTS

We are grateful to the captain and crew members of the R/V *Lurongyuyang-65577* for their assistance during the cruises. We thank Dr. Jihong Zhang for providing the hydrological information of Sanggou Bay.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2020.589566/full#supplementary-material

Supplementary Figure 1 | Cluster analysis of stations according to temperature and salinity in four successive seasons, spring to winter.

Supplementary Figure 2 | Cluster analysis of stations according to microbial food web component biomasses in four successive seasons, spring to winter.

Supplementary Figure 3 | Cluster analysis of stations according to tintinnid abundance in four successive seasons, spring to winter.

REFERENCES

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyerreil, L. A., and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* 10, 257–263. doi: 10.3354/meps010257

Bachy, C., Moreira, D., Dolan, J. R., and Lopez-Garcia, P. (2014). Seasonal dynamics of free-living tintinnid ciliate communities revealed by environmental sequences from the North-West Mediterranean Sea. *FEMS Microbiol. Ecol.* 87, 330–342. doi: 10.1111/1574-6941.12224

Bock, N., Van Wambeke, F., Dion, M., and Duhamel, S. (2018). Microbial community structure in the western tropical South Pacific. *Biogeosciences* 15, 3909–3925. doi: 10.5194/bg-2017-562

Børsheim, K. Y., and Bratbak, G. (1987). Cell volume to cell carbon conversion factors for a bacterivorous *Monas* sp. enriched from seawater. *Mar. Ecol. Prog. Ser.* 36, 171–175. doi: 10.3354/meps036171

Bouman, H. A., Ulloa, O., Barlow, R., Li, W. K., Platt, T., Zwirglmaier, K., et al. (2011). Water-column stratification governs the community structure of subtropical marine picophytoplankton. *Environ. Microbiol. Rep.* 3, 473–482. doi: 10.1111/j.1758-2229.2011.00241.x

Chen, X., Li, H. B., Zhao, Y., Zhao, L., Dong, Y., Zhang, W. C., et al. (2018). Distribution of different biogeographical tintinnids in Yellow Sea and Bohai Sea. *J. Ocean Univ. China* 17, 371–384. doi: 10.1007/s11802-018-3482-1

Emery, W. J. (2001). "Water types and water masses," in *Encyclopedia of Ocean Sciences*, eds J. H. Steele, S. A. Thorpe, and K. K. Turekian (London: Academic Press), 3179–3187. doi: 10.1006/rwos.2001.0108

Feng, M. P., Wang, C. F., Zhang, W. C., Zhang, G. T., Xu, H. L., Zhao, Y., et al. (2018). Annual variation of species richness and lorica oral diameter characteristics of tintinnids in a semi-enclosed bay of western Pacific. *Estua. Coast. Shelf Sci.* 207, 164–174. doi: 10.1016/j.ecss.2018.04.003

Fernandez, C., Thyssen, M., and Denis, M. (2008). Microbial community structure along 18°W (39°N–44.5°N) in the NE Atlantic in late summer 2001 (POMME programme). *J. Mar. Syst.* 71, 46–62. doi: 10.1016/j.jmarsys.2007.06.003

Garrison, D. L., Gowing, M. M., Hughes, M. P., Campbell, L., Caron, D. A., Dennett, M. R., et al. (2000). Microbial food web structure in the Arabian Sea: a US JGOFS study. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 47, 1387–1422. doi: 10.1016/S0967-0645(99)00148-4

Gasol, J. M. (1994). A framework for the assessment of top-down vs bottom-up control of heterotrophic nanoflagellate abundance. *Mar. Ecol. Prog. Ser.* 113, 291–300. doi: 10.3354/meps113291

Gasol, J. M., Del Giorgio, P. A., and Duarte, C. M. (1997). Biomass distribution in marine planktonic communities. *Limnol. Oceanogr.* 42, 1353–1363. doi: 10.4319/lo.1997.42.6.1353

Gilron, G. L., and Lynn, D. H. (1989). Assuming a 50% cell occupancy of the lorica overestimates tintinnine ciliate biomass. *Mar. Biol.* 103, 413–416. doi: 10.1007/BF00397276

Hashihama, F., Horimoto, N., Kanda, J., Furuya, K., Ishimaru, T., and Saino, T. (2008). Temporal variation in phytoplankton composition related to water mass properties in the central part of Sagami Bay. *J. Oceanogr.* 64, 23–37. doi: 10.1007/s10872-008-0002-8

Herndl, G. J. (1991). Microbial biomass dynamics along a trophic gradient at the Atlantic Barrier-Reef off Belize (Central-America). *Mar. Ecol.* 12, 41–51. doi: 10.1111/j.1439-0485.1991.tb00082.x

Jiang, Y., Xu, H. L., Al-Rasheid, K. A. S., Warren, A., Hu, X. Z., and Song, W. B. (2011). Planktonic ciliate communities in a semi-enclosed bay of Yellow Sea, northern China: annual cycle. *J. Mar. Biol. Assoc. U.K.* 91, 97–105. doi: 10.1017/S002531541000175X

Jiang, Z. J., Du, M. R., Fang, J. H., Gao, Y. P., Li, J. Q., Zhao, L., et al. (2017). Size fraction of phytoplankton and the contribution of natural plankton to the carbon source of Zhikong scallop *Chlamys farreri* in mariculture ecosystem of the Sanggou Bay. *Acta Oceanol. Sin.* 36, 1–9. doi: 10.1007/s13131-017-0970-x

Kim, Y.-O., Shin, K., Jang, P.-G., Choi, H.-W., Noh, J.-H., Yang, E.-J., et al. (2012). Tintinnid species as biological indicators for monitoring intrusion of the warm oceanic waters into Korean coastal waters. *Ocean Sci. J.* 47, 161–172. doi: 10.1007/s12601-012-0016-4

Kirchman, D. L. (2008). *Microbial Ecology of the Oceans*. Hoboken: John Wiley & Sons, 593.

Kofoid, C. A., and Campbell, A. S. (1929). A conspectus of the marine fresh-water ciliate belonging to the suborder Tintinnoinea, with descriptions of new species principally from the Agassiz expedition to the eastern tropical Pacific 1904-1905. *Univ. Calif. Publ. Zool.* 34, 1–403.

Kofoid, C. A., and Campbell, A. S. (1939). The ciliata: the tintinnoinea. *Bull. Museum Comp. Zool.* 84, 1–473. doi: 10.1002/9783527678518.ehg2016001

Kuang, S. H., Fang, J. G., Sun, H. L., and Li, F. (1996). Seston dynamics in Sanggou Bay. *Mar. Fish. Res.* 2, 60–67.

Lee, S., and Fuhrman, J. A. (1987). Relationships between biovolume and biomass of naturally derived marine bacterioplankton. *Appl. Environ. Microbiol.* 53, 1298–1303. doi: 10.1002/bit.260290816

Li, H. B., Wang, C. F., Liang, C., Zhao, Y., Zhang, W. C., Grégori, G., et al. (2019). Diversity and distribution of tintinnid ciliates along salinity gradient in the Pearl River Estuary in southern China. *Estua. Coast. Shelf Sci.* 226:106268. doi: 10.1016/j.ecss.2019.106268

Li, W. K. W., Dickie, P. M., Harrison, W. G., and Irwin, B. D. (1993). Biomass and production of bacteria and phytoplankton during the spring bloom in the western North Atlantic Ocean. *Deep Sea Res. II Top. Stud. Oceanogr.* 40, 307–327. doi: 10.1016/0967-0645(93)90019-J

Li, W. K. W., Dickie, P. M., Irwin, B. D., and Wood, A. M. (1992). Biomass of bacteria, cyanobacteria, prochlorophytes and photosynthetic eukaryotes in the Sargasso Sea. *Deep Sea Res. A Oceanogr. Res. Pap.* 39, 501–519. doi: 10.1016/0198-0149(92)90085-8

Linacre, L. P., Landry, M. R., Lara-Lara, J. R., Hernandez-Ayon, J. M., and Bazan-Guzman, C. (2010). Picoplankton dynamics during contrasting seasonal oceanographic conditions at a coastal upwelling station off Northern Baja California, Mexico. *J. Plankton Res.* 32, 539–557. doi: 10.1093/plankt/fbp148

Lu, J. C., Huang, L. F., Luo, Y. R., Xiao, T., Jiang, Z. J., and Wu, L. N. (2015a). Effects of freshwater input and mariculture (bivalves and macroalgae) on spatial distribution of nanoflagellates in Sungo Bay, China. *Aquac. Environ. Interact.* 6, 191–203. doi: 10.3354/aei00124

Lu, J. C., Huang, L. F., Xiao, T., Jiang, Z. J., and Zhang, W. C. (2015b). The effects of Zhikong scallop (*Chlamys farreri*) on the microbial food web in a phosphorus-deficient mariculture system in Sanggou Bay, China. *Aquaculture* 448, 341–349. doi: 10.1016/j.aquaculture.2015.06.021

Mackas, D. L., Greve, W., Edwards, M., Chiba, S., Tadokoro, K., Eloire, D., et al. (2012). Changing zooplankton seasonality in a changing ocean: comparing time series of zooplankton phenology. *Prog. Oceanogr.* 97, 31–62. doi: 10.1016/j.pocean.2011.11.005

Maixandeau, A., Lefevre, D., Karayanni, H., Christaki, U., Van Wambeke, F., Thyssen, M., et al. (2005). Microbial community production, respiration, and structure of the microbial food web of an ecosystem in the northeastern Atlantic Ocean. *J. Geophys. Res. Oceans* 110:C07S17. doi: 10.1029/2005JC003234

Marie, D., Simon, N., Guillou, L., Partensky, F., and Vaulot, D. (2000). "Flow cytometry analysis of marine picoplankton," in *In Living Color: Protocols in Flow Cytometry and Cell Sorting*, eds R. A. Diamond and S. DeMaggio (New York, NY: Springer-Verlag), 421–454. doi: 10.1007/978-3-642-57049-0_34

Mena, C., Reglero, P., Hidalgo, M., Sintes, E., Santiago, R., Martin, M., et al. (2019). Phytoplankton community structure is driven by stratification in the oligotrophic Mediterranean Sea. *Front. Microbiol.* 10:1698. doi: 10.3389/fmicb.2019.01698

Modigh, M., Castaldo, S., Saggiomo, M., and Santarpia, I. (2003). Distribution of tintinnid species from 42°N to 43°S through the Indian Ocean. *Hydrobiologia* 503, 251–262. doi: 10.1023/B:HYDR.0000008477.38383.d6

Otero-Ferrer, J. L., Cermeno, P., Bode, A., Fernandez-Castro, B., Gasol, J. M., Moran, X. A. G., et al. (2018). Factors controlling the community structure of picoplankton in contrasting marine environments. *Biogeosciences* 15, 6199–6220. doi: 10.5194/bg-15-6199-2018

Pan, L. A., Zhang, J., Chen, Q., and Deng, B. (2006). Picoplankton community structure at a coastal front region in the northern part of the South China Sea. *J. Plankton Res.* 28, 337–343. doi: 10.1093/plankt/fbi117

Patten, N. L., Van Ruth, P. D., and Rodriguez, A. R. (2018). Spatial variability in picophytoplankton, bacteria and viruses in waters of the Great Australian Bight (southern Australia). *Deep Sea Res. II Top. Stud. Oceanogr.* 157, 46–57. doi: 10.1016/j.dsr2.2018.04.009

Pierrot-Bults, A. C., and Angel, M. V. (2012). Pelagic biodiversity and biogeography of the oceans. *Biol. Int.* 51, 9–35.

Pomeroy, L. R. (1974). The ocean's food web, a changing paradigm. *Bioscience* 24, 499–504. doi: 10.2307/1296885

Putt, M., and Stoecker, D. K. (1989). An experimentally determined carbon-volume ratio for marine oligotrichous ciliates from estuarine and coastal waters. *Limnol. Oceanogr.* 34, 1097–1103. doi: 10.4319/lo.1989.34.6.1097

Samo, T. J., Pedler, B. E., Ball, G. I., Pasulka, A. L., Taylor, A. G., Aluwihare, L. I., et al. (2012). Microbial distribution and activity across a water mass frontal zone in the California current ecosystem. *J. Plankton Res.* 34, 802–814. doi: 10.1093/plankt/fbs048

Sanders, R. W., Caron, D. A., and Berninger, U. G. (1992). Relationships between bacteria and heterotrophic nanoplankton in marine and fresh waters - an inter-ecosystem comparison. *Mar. Ecol. Prog. Ser.* 86, 1–14. doi: 10.3354/meps086001

Šantić, D., Kovačević, V., Bensi, M., Giani, M., Vrdoljak Tomaš, A., Ordulj, M., et al. (2019). Picoplankton distribution and activity in the deep waters of the southern Adriatic Sea. *Water* 11:1655. doi: 10.3390/w11081655

Šantić, D., Šestanović, S., Šolić, M., Krstulović, N., Kušpilić, G., Ordulj, M., et al. (2014). Dynamics of the picoplankton community from coastal waters to the open sea in the Central Adriatic. *Mediterranean Mar. Sci.* 15, 179–188. doi: 10.12681/mms.701

Sherr, E., and Sherr, B. (1988). Role of microbes in pelagic food webs - a revised concept. *Limnol. Oceanogr.* 33, 1225–1227. doi: 10.2307/2837257

Sime-Ngando, T., Gosselin, M., Roy, S., and Chanut, J. P. (1995). Significance of planktonic ciliated protozoa in the Lower St Lawrence Estuary: comparison with bacterial, phytoplankton, and particulate organic carbon. *Aquat. Microb. Ecol.* 9, 243–258. doi: 10.3354/ame009243

Sun, P., Wang, Y., Laws, E., and Huang, B. Q. (2020). Water mass-driven spatial effects and environmental heterogeneity shape microeukaryote biogeography in a subtropical, hydrographically complex ocean system - A case study of ciliates. *Sci. Tot. Environ.* 706:135753. doi: 10.1016/j.scitotenv.2019.135753

Suzuki, T., and Miyabe, C. (2007). Ecological balance between ciliate plankton and its prey candidates, pico- and nanoplankton, in the East China Sea. *Hydrobiologia* 586, 403–410. doi: 10.1007/s10750-007-0715-5

Tsai, A. Y., Chiang, K. P., Chang, J., and Gong, G. C. (2005). Seasonal diel variations of picoplankton and nanoplankton in a subtropical western Pacific coastal ecosystem. *Limnol. Oceanogr.* 50, 1221–1231. doi: 10.4319/lo.2005.50.4.1221

Van Dongen-Vogels, V., Seymour, J. R., Middleton, J. F., Mitchell, J. G., Seuront, L. J. E. C., and Science, S. (2012). Shifts in picophytoplankton community structure influenced by changing upwelling conditions. *Estua. Coast. Shelf Sci.* 109, 81–90. doi: 10.1016/j.ecss.2012.05.026

Wu, M. L., Wang, Y. S., Wang, D. X., and Dong, J. D. (2014). Effects of coastal upwelling on picophytoplankton distribution off the coast of Zhanjiang in South China Sea. *Oceanol. Hydrobiol. Stud.* 43, 283–291. doi: 10.2478/s13545-014-0143-x

Yang, J. P., Huang, S. X., Fan, W. X., Warren, A., Jiao, N. Z., and Xu, D. P. (2020). Spatial distribution patterns of planktonic ciliate communities in the East China Sea: potential indicators of water masses. *Mar. Pollut. Bull.* 156:111253. doi: 10.1016/j.marpolbul.2020.111253

Yu, Y., Zhang, W. C., Zhang, C. X., Zhou, F., Zhao, N., and Xiao, T. (2014). Basin-scale variation in planktonic ciliate distribution: a detailed temporal and spatial study of the Yellow Sea. *Mar. Biol. Res.* 10, 641–654. doi: 10.1080/17451000.2013.852683

Zeng, D. Y., Huang, D. J., Qiao, X. D., He, Y. Q., and Zhang, T. (2015). Effect of suspended kelp culture on water exchange as estimated by *in situ* current measurement in Sanggou Bay, China. *J. Mar. Syst.* 149, 14–24. doi: 10.1016/j.jmarsys.2015.04.002

Zhang, W. C., Feng, M. P., Zhang, C. X., and Xiao, T. (2012). *An Illustrated Guide to Contemporary Tintinnids in the World*. Beijing: Science Press.

Zhao, L., Zhao, Y., Xu, J. H., Zhang, W. C., Huang, L. F., Jiang, Z. J., et al. (2016). Distribution and seasonal variation of picoplankton in Sanggou Bay, China. *Aquac. Environ. Interact.* 8, 261–271. doi: 10.3354/aei00168

Zhao, L., Zhao, Y., Zhang, W. C., Zhou, F., Zhang, C. X., Ren, J. L., et al. (2013). Picoplankton distribution in different water masses of the East China Sea in autumn and winter. *Chinese J. Oceanol. Limnol.* 31, 247–266. doi: 10.1007/s00343-013-2085-3

Zubkov, M. V., Sleigh, M. A., Burkill, P. H., and Leakey, R. J. G. (2000). Picoplankton community structure on the Atlantic meridional transect: a comparison between seasons. *Prog. Oceanogr.* 45, 369–386. doi: 10.1016/s0079-6611(00)00008-2

Zubkov, M. V., Sleigh, M. A., Tarran, G. A., Burkill, P. H., and Leakey, R. J. G. (1998). Picoplanktonic community structure on an Atlantic transect from 50°N to 50°S. *Deep Sea Res. I Oceanogr. Res. Pap.* 45, 1339–1355. doi: 10.1016/S0967-0637(98)00015-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 January 2021
doi: 10.3389/fmicb.2020.599614

Patterns of Relative and Quantitative Abundances of Marine Bacteria in Surface Waters of the Subtropical Northwest Pacific Ocean Estimated With High-Throughput Quantification Sequencing

Jie Kong[1,2], *Xin Liu*[1,2], *Lei Wang*[3], *Hao Huang*[3], *Danyun Ou*[3], *Jiayu Guo*[1,2], *Edward A. Laws*[4] *and Bangqin Huang*[1,2]*

[1] State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China, [2] Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, College of the Environment and Ecology, Xiamen University, Xiamen, China, [3] Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, China, [4] Department of Environmental Sciences, College of the Coast and Environment, Louisiana State University, Baton Rouge, LA, United States

OPEN ACCESS

Edited by:
Gipsi Lima Mendez, Catholic University of Louvain, Belgium

Reviewed by:
Meinhard Simon, University of Oldenburg, Germany
David Anthony Pearce, Northumbria University, United Kingdom

***Correspondence:**
Bangqin Huang
bqhuang@xmu.edu.cn

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *27 August 2020*
Accepted: *21 December 2020*
Published: *21 January 2021*

Citation:
Kong J, Liu X, Wang L, Huang H, Ou D, Guo J, Laws EA and Huang B (2021) Patterns of Relative and Quantitative Abundances of Marine Bacteria in Surface Waters of the Subtropical Northwest Pacific Ocean Estimated With High-Throughput Quantification Sequencing. Front. Microbiol. 11:599614. doi: 10.3389/fmicb.2020.599614

Bacteria play a pivotal role in shaping ecosystems and contributing to elemental cycling and energy flow in the oceans. However, few studies have focused on bacteria at a trans-basin scale, and studies across the subtropical Northwest Pacific Ocean (NWPO), one of the largest biomes on Earth, have been especially lacking. Although the recently developed high-throughput quantitative sequencing methodology can simultaneously provide information on relative abundance, quantitative abundance, and taxonomic affiliations, it has not been thoroughly evaluated. We collected surface seawater samples for high-throughput, quantitative sequencing of 16S rRNA genes on a transect across the subtropical NWPO to elucidate the distribution of bacterial taxa, patterns of their community structure, and the factors that are potentially important regulators of that structure. We used the quantitative and relative abundances of bacterial taxa to test hypotheses related to their ecology. Total 16S rRNA gene copies ranged from 1.86×10^8 to 1.14×10^9 copies L^{-1}. Bacterial communities were distributed in distinct geographical patterns with spatially adjacent stations clustered together. Spatial considerations may be more important determinants of bacterial community structures than measured environmental variables. The quantitative and relative abundances of bacterial communities exhibited similar distribution patterns and potentially important determinants at the whole-community level, but inner-community connections and correlations with variables differed at subgroup levels. This study advanced understanding of the community structure and distribution patterns of marine bacteria as well as some potentially important determinants thereof in a subtropical oligotrophic ocean system. Results highlighted the importance of considering both the quantitative and relative abundances of members of marine bacterial communities.

Keywords: relative abundance, bacterial community structure, distribution patterns, ecology of marine bacteria, quantitative abundance

INTRODUCTION

Bacteria play a pivotal role in shaping ecosystems and contributing to the cycling of elements and the flow of energy in the oceans (Ducklow, 2000; Kirchman, 2016; Steinberg and Landry, 2017). Knowledge of the distribution patterns of marine bacteria, their community composition, and the factors that shape that composition in surface seawater is of great interest, especially because of the ongoing climate changes that are expected to cause warming and acidification of the surface waters of the ocean, shoaling of the surface mixed layer, and increased irradiance within that shallower mixed layer (Hutchins and Fu, 2017). With well-developed, high-throughput sequencing (HTS) technology, it is now technologically and economically feasible to finely resolve bacterial community composition. Sequences obtained from HTS can provide both genetic (taxonomic) information and relative abundance. This information has dramatically expanded our understanding of the composition and biogeography of bacterial communities and the mechanisms by which they are assembled (Sunagawa et al., 2015; Goodwin et al., 2016; Lindh et al., 2018). Even though HTS can provide only two basic attributes of a community on relative abundance and taxonomic affiliation; it cannot provide information on absolute abundance, which is needed for an accurate and comprehensive interpretation of the biological and ecological implications of bacterial community structure (Props et al., 2017; Vandeputte et al., 2017; Zhang et al., 2017; Yang et al., 2018). The addition of artificial standard spike-ins (i.e., internal standard sequences) before HTS (that is, high-throughput quantification sequencing, HTQS) makes it possible to estimate microbial abundances (Tourlousse et al., 2017; Yang et al., 2018). Although the effects of PCR bias cannot be ruled out in HTQS and would impact estimates of microbial abundances, it has been shown that estimates of microbial community structure and composition are not perturbed by artificial standard spike-ins, and even complex 16S rRNA gene pools do not affect quantification based on artificial standard spike-ins (Tourlousse et al., 2017). In this study, we considered that gene copy numbers estimated from HTQS were metrics of quantitative abundance rather than relative abundance and could provide, at least to some extent, information about all three community attributes (i.e., taxa, relative abundance, and quantitative abundance) and could thereby enhance the significance of HTS-based microbiome studies (Tourlousse et al., 2017; Wang et al., 2018; Yang et al., 2018; Jiang et al., 2019; Mou et al., 2020). However, to our knowledge, only one study has previously applied this improved HTQS method to an aquatic ecosystem. That study linked net community production and microbial community composition in the western North Atlantic and revealed that the microbial community that resulted from an algal bloom was associated with a regional peak of net community production (Wang et al., 2018). Moreover, little is known about the differences and similarities of the implications of relative and quantitative abundances in marine microbial ecology.

The subtropical NWPO is one of the largest biomes on Earth and is characterized by warm, nutrient-poor, low-biomass, stratified surface waters (Spalding et al., 2012; Tseng et al., 2016). Although the subtropical NWPO plays a significant role in moderating the global climate and biogeochemical cycles, it is undersampled and not well understood in terms of its ecosystem structure and functionality (Tseng et al., 2016; Karl and Church, 2017; Kavanaugh et al., 2018).

There have previously been many studies of marine bacterial communities over different geographic ranges with high sample numbers and/or great depth coverage. However, those studies have been conducted at only a few scattered stations or within only a few subdivisions of the North Pacific Ocean, and only a small number of samples have been collected from the subtropical NWPO (Fuhrman et al., 2008; Zinger et al., 2011; Sunagawa et al., 2015; Lindh, 2017; Shulse et al., 2017; Li et al., 2018; Lindh et al., 2018). Zinger et al. (2011) identified the global patterns of bacterial distributions across the world's oceans by analyzing 509 samples, encompassing snapshot locations in the North Pacific Ocean. Sunagawa et al. (2015) have studied the structure and function of the global ocean microbiome based on samples from 68 locations, including six from the eastern North Pacific Ocean. Li et al. (2018) have estimated bacterial diversity and nitrogen utilization in the northwestern Pacific Ocean, and Shulse et al. (2017) have collected samples from the Clarion-Clipperton Zone of the Eastern North Pacific to examine the diversity and composition of the microbial communities. The distribution patterns and mechanisms of assembly of communities are affected by spatial scale (Martiny et al., 2011; Shi et al., 2018). To our knowledge, however, there has been no trans-basin study of marine bacteria in the North Pacific Ocean.

In this study, we collected surface seawater samples for HTQS of the 16S rRNA gene along a transect across the subtropical NWPO. The objectives of the study were to (1) reveal bacterial community structure, distribution patterns, and potential determinants thereof, and (2), to compare the quantitative and relative bacterial abundances as indicators of the roles of bacterial communities in subtropical NWPO surface seawater.

MATERIALS AND METHODS

Sample Collection and Processing

Surface seawater from 20 stations located in the subtropical NWPO (**Figure 1**), ranging from 126.21°W to 158.77°E and from 9.31°N to 28.96°N, were collected with 12-L Niskin bottles deployed with a conductivity-temperature-depth (SBE911 plus CTD system) profiler or manually from July 16 to August 5, 2018, as a part of the Dayang 50 cruise (**Supplementary Table 1**). The map of sampling stations was constructed using the Ocean Data View version 4.7.10 (Schlitzer, 2015). Sea surface temperature was measured with an onboard thermosalinograph (SeaBird Electronics model SBE 45). Daily mean sea surface salinity data for each sampling location were extracted from the GLOBAL-REANALYSIS-PHY-001-30 reanalysis product (1/12° horizontal resolution) provided by the Copernicus Marine Environment Monitoring Service (http://marine.copernicus.eu/).

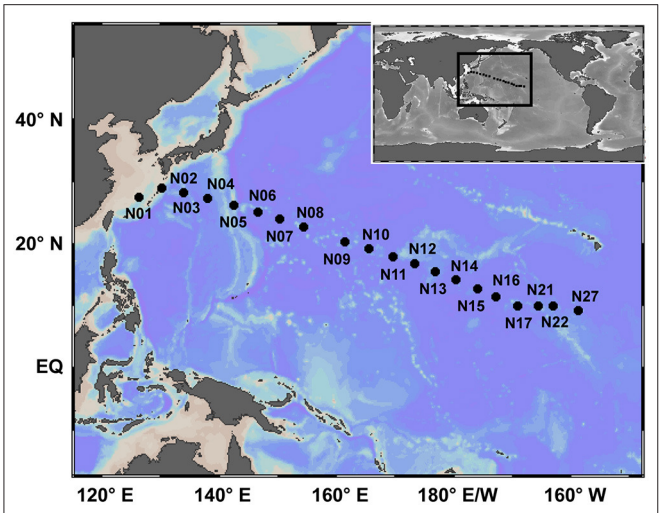


FIGURE 1 | Sampling locations in the Northwest Pacific Ocean. Surface seawater samples were collected from 20 stations (marked by black dots). The map with sampling locations was constructed using Ocean Data View version 4.7.10.

The seawater was passed through 200-μm nylon mesh to remove metazoans, and the cells from 4.25 to 7.3 liters of the filtrate were harvested through 47-mm-diameter polycarbonate filters (0.2-μm pore size, Millipore, USA). It was inevitable that bacteria-associated particles (> 200 μm) were also removed with the 200-μm nylon mesh. The filters were then stored at −80°C until DNA extraction. Duplicate seawater samples from each station were stored at −20°C in acid-washed, 250-mL polyethylene bottles, which were transported to the laboratory and analyzed (i) for dissolved reactive phosphorus (DRP) concentrations following the manual, solid phase extraction method described previously (Yuan et al., 2016), (ii) for NO_x (nitrate + nitrite) concentrations using a colorimetric method with a Flow Injection Analysis-Liquid Waveguide Capillary Cell system (Zhang, 2000) and (iii) for $Si(OH)_4$ concentrations using an AA3 AutoAnalyzer (Bran+Luebbe, Germany) following procedures described previously (Du et al., 2013).

For measurement of chlorophyll *a* (Chl *a*) concentrations, 4.3 liters of seawater from each station were filtered onto Whatman 25-mm-diameter GF/F filters and stored at −80°C. Chl *a* was extracted with N, N-dimethylformamide from filters, and measured by high-performance liquid chromatography (Liu et al., 2016). Seawater samples for enumeration of bacterial abundance were fixed with 50 % paraformaldehyde (Sangon Biotech, China) at a final concentration of 1 % (v/v), incubated in the dark for 10–15 min, and stored at −80°C. Bacteria were stained with 100× SYBR Green I (final concentration 1×) (Thermo Fisher Scientific, USA) were counted using a BD Accuri C6 flow cytometer (BD Biosciences, USA) after mixing with 1-μm yellow-green latex beads (Sigma, USA) (Marie et al., 1999).

High-Throughput Quantification Sequencing

To quantify taxa abundance across samples, HTQS, which has been described in pioneer studies (Tourlousse et al., 2017; Wang et al., 2018; Jiang et al., 2019; Mou et al., 2020), was applied and carried out mainly in Genesky Biotechnologies (China), with some modifications. Briefly, total genomic DNAs were extracted using a DNeasy PowerWater Kit (Qiagen, Germany) following the manufacturer's instructions from polycarbonate filters with concentrated cells. DNA quality and purity were evaluated by gel electrophoresis and NanoDrop 2000 (Thermo Fisher Scientific, USA), and the concentration was quantified with an Invitrogen Qubit 3.0 Spectrophotometer (Thermo Fisher Scientific, USA). Artificial standard spike-in sequences consisted of both conserved regions that were identical to natural 16S rRNA genes and artificial variable regions that were random sequences with about 40% G+C content and that shared negligible identity with sequences in the public databases. Artificial standard spike-ins were designed and synthesized in Genesky Biotechnologies (China). Nine artificial standard spike-ins were added to genomic DNAs for each sample at four different concentrations (10^3, 10^4, 10^5, and 10^6 copies of sequences for three, two, two, and two standard spike-ins, respectively), followed by amplification of the V3-V4 region of bacterial 16S rRNA gene with primers 341F and 805R in triplicate for 25 cycles (Herlemann et al., 2011). Amplicons were checked using gel electrophoresis and purified with Agencourt AMPure XP PCR Purification Beads (Beckman Coulter, USA). Purified amplicons of each sample were added with a sample-specific index sequence and then used to construct a library. The library was quantified, pooled, checked, and then sequenced using an Illumina Miseq Benchtop Sequencer (Illumina, USA) for 2 × 250 base pair (bp) paired-end reads.

Processing and Analysis of Sequencing Data

The raw paired-end reads with Q20 values ≥97.1% and Q30 values ≥94.37% were cleaned using Trim Galore v0.4.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), FLASH2 v2.2.00 (Magoc and Salzberg, 2011), mothur v.1.39.3 (Schloss et al., 2009), and Usearch v10 (Edgar, 2013), including quality checking, filtering, and assembly of data. First, Trim Galore was applied to trim adapter and bases with quality scores <20 at the end of the read and reads shorter than 100 bp were removed. Second, FLASH2 was used to merge paired-end reads, followed by the removal of low-quality sequences (Magoc and Salzberg, 2011). Third, mothur was applied to identify and remove primers from sequences and filter out sequences with N-base or homopolymer >6 bp. Finally, sequences with total base error rates larger than two or lengths shorter than 200 were removed using Usearch, resulting in clean sequences for further processing (Edgar, 2013). UPARSE (Edgar, 2013) implemented in Usearch v10 was performed for processing chimera removal, singleton removal, Operational Taxonomic Unit (OTU) clustering (97% similarity cut-off), and picking of representative sequences.

Representative sequences were then blasted against the Silva132 database (Yilmaz et al., 2014) for taxonomic assignment. OTUs assigned to spike-in sequences were identified, counted, and removed for each sample, and the relative and quantitative abundances of the remaining OTUs were then calculated. For relative abundance analyses, an OTU table without singleton and non-bacterial sequences (i.e., spike-in, archaea, and chloroplast) was randomly rarefied to the same sequence number of 85442 (the minimum number of sequences in the samples) for each sample. For analyses of quantitative abundances, a standard curve equation based on the 9 added standard spike-ins for each sample was first constructed following $y = ax + b$ (1), where y is the log-transformed number of spike-in OTU sequences, x represents log-transformed copies of the added spike-in, and the parameters a and b are fitting coefficients.

The quantitative abundance of each OTU in a sample was then determined as follows: $AB_{otu} = \frac{Y_{otu}}{R^* V}$ (2), where AB_{otu} is the quantitative abundance of OTUs in the unit of copies L^{-1}; Y_{otu} represents OTU abundance calculated from equation (1); R is the recovery ratio of DNA concentration (sequenced genomic DNA/total genomic DNA extracted in each sample), and V represents the volume of seawater filtered. Based on relative abundance, richness (observed OTU number), Chao 1, Shannon, and Faith's phylogenetic diversity (PD) (Faith, 1992) indices were calculated by applying command alpha_diversity.py in QIIME v.1.9.0 (Caporaso et al., 2010).

Statistical Analysis

All statistical analyses and figures were done with R (version 3.6.1) (R Core Team, 2016) unless otherwise mentioned. Spearman rank correlation tests were applied to pairs of variables, including biogeochemical factors and alpha diversity indices of bacteria, with *P*-values corrected using the false discovery rate (fdr) method. To evaluate inner-community connections, the function corr.test() (arguments: method = "spearman," adjust = "fdr") in package 'psych' (Revelle, 2017) were used to calculate the pairwise correlations for OTUs with relative abundance >0.1% (relative to total sequences), followed by depicturing significant correlations (adjusted $p < 0.05$ and $r \geq 0.6$) and analysis of network attributes through the software Gephi (Bastian et al., 2009). Pairwise Spearman correlation coefficients between relative and quantitative abundances were analyzed for OTUs recovered in all samples with *p*-values adjusted in the package "q-value" (Dabney et al., 2010). The procrustes analysis was applied to test the agreement between relative and quantitative abundances in presenting the distribution patterns of bacterial communities using the "vegan" package (Oksanen et al., 2016).

First, Principal Coordinates Analysis was performed separately on relative and quantitative abundances of bacterial communities to reduce their dimensionalities. Then, the Procrustes analysis was used to stretch and rotate the points in principal coordinates matrices using the function procrustes(). The statistical significance was measured with a Monte Carlo test, with M^2 indicating the goodness of fit. The Mantel test was applied to reveal the potential effects of factors such as environmental and geographic metrics on bacterial community composition based on Bray-Curtis distance.

To evaluate the relative effects of environmental and spatial variables in constructing the bacterial communities, a variation partitioning analysis (VPA) was performed based on redundancy analysis (RDA), as previously described (Borcard et al., 2011). A sparse partial least squares (sPLS) approach implemented in the R package "mixOmics" (Rohart et al., 2017) was performed to simultaneously select variables from bacterial subgroups (at phylum and order levels, respectively), and variables (i.e., environmental and geographic factors) that could identify certain bacterial subgroups having high correlations with specific parameters.

RESULTS

Environmental Characteristics

Twenty surface seawater samples were collected and measured for relevant environmental variables (**Supplementary Table 1**, **Supplementary Figure 1**) across the subtropical NWPO during the summer of 2018. Sea surface temperature averaged 29.11°C (range: 28.55–29.68°C) (**Supplementary Figure 1A**). Sea surface salinity averaged 34.36 (range: 33.66–35.12) (**Supplementary Figure 1B**). Concentrations of NO_x (nitrate + nitrite) averaged 0.033 μmol/L (range: 0.014–0.104 μmol/L) (**Supplementary Figure 1C**). Dissolved reactive phosphate (DRP) concentrations averaged 0.07 μmol/L (range: 0.01–0.15 μmol/L) (**Supplementary Figure 1D**). The geographical distribution of $Si(OH)_4$ concentrations, which averaged 1.08 μmol/L (range: 0.57–1.93 μmol/L) (**Supplementary Figure 1E**) was opposite that of the DRP concentrations. The $Si(OH)_4$ concentrations increased from west to east across the subtropical NWPO and decreased toward the north (**Supplementary Figures 1D,E**, **Figure 2A**). The concentrations of Chl *a* averaged 25.2 ng/L (range: 5.4–48.8 ng/L) (**Supplementary Figure 1F**). Bacterial abundances varied from 1.15×10^5 cells/ml to 1.68×10^5 cells/ml (**Supplementary Figure 1G**).

α-Diversity of Bacterial Communities

Rarefaction curves indicated that all amplicon samples were almost saturated with respect to the number of sequences. The implication of this was that the majority of bacterial communities in the amplicon samples collected from subtropical NWPO surface water were recovered at the current sequencing depth (**Supplementary Figure 2**). All samples rarefied to 85,442 sequences yielded a total of 1,192 OTUs at a cutoff value of 97% similarity. The samples contained 453–682 OTUs per sample (**Figure 2B**). The distribution patterns across the subtropical NWPO were consistent for richness based on the Chao 1, phylogenetic diversity (PD), and Shannon diversity indices. The Chao 1 ranged from 561 to 799 (average of 651), the PD index from 42 to 58 (average of 46), and the Shannon from 4.5 to 5.8 (average of 4.8). The richness, Chao 1, and phylogenetic indices peaked at station N01, whereas the maximum of Shannon index was at station N17 (**Figures 2B–D**).

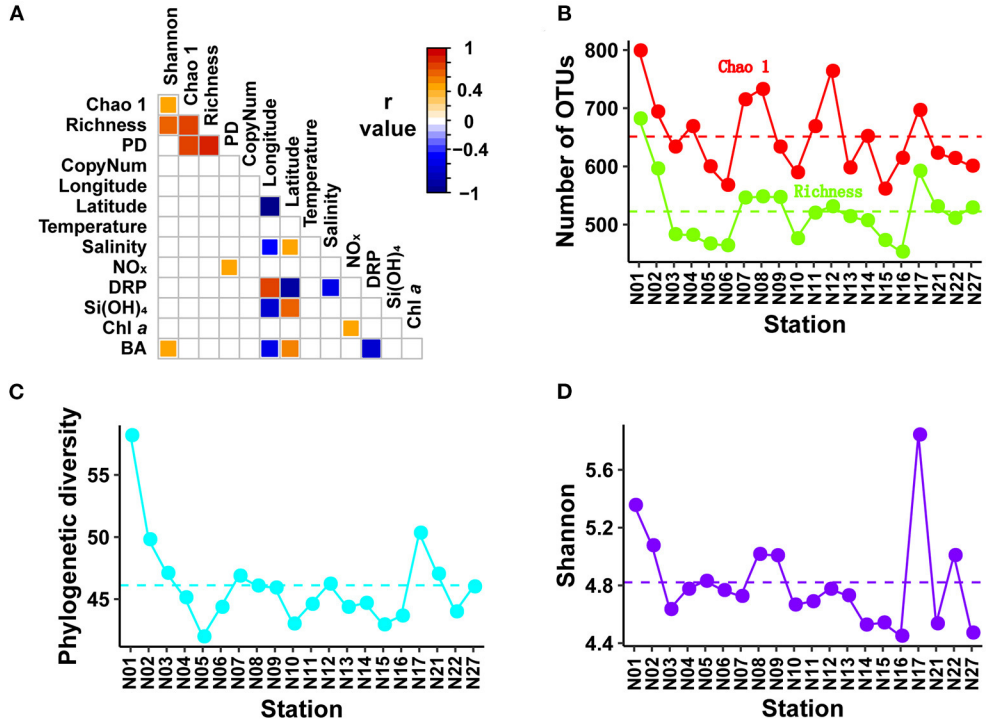


FIGURE 2 | Spearman correlation matrix and distribution patterns of bacterial α-diversity indices. The Spearman correlation matrix **(A)** showing the results of Spearman rank correlation tests among α-diversity indices, environmental variables, geographic factors, and copy number of 16S rRNA gene. The *p*-values were corrected using package "psych." Only adjusted $p < 0.05$ are shown in the Spearman correlation matrix with color indicating the *r* value. Distribution patterns of α-diversity indices including Chao1 and richness, PD, and Shannon are shown in **(B–D)**, respectively. PD, phylogenetic diversity; CopyNum, copy number of 16S rRNA gene; NO_x, nitrite + nitrate; DRP, dissolved reactive phosphorus; BA, bacterial abundance.

Structure and Distribution Patterns of Bacterial Communities

Strong linear relationships ($R^2 > 0.988$) were found between the added copy numbers of spike-in standards and the abundances of spike-in OTU sequences recovered from sequencing for all samples (**Supplementary Table 2**). The quantitative taxonomic abundances calculated from spike-in standards averaged 7.09×10^8 copies L^{-1} (range: 1.86×10^8-1.14×10^9 copies L^{-1}) for bacterial communities (**Figure 3B**). At the phylum level, both relative and quantitative sequence abundances derived from all samples were dominated by Proteobacteria (average about 65.4% and 4.6×10^8 copies L^{-1}, respectively), followed by Cyanobacteria (average about 22.0% and 1.9×10^8 copies L^{-1}, respectively). These two phyla accounted for more than 87% of the total relative abundance. The α-proteobacteria and γ-proteobacteria dominated within Proteobacteria, accounting for 48.7 and 15.3% of total sequences, respectively (**Supplementary Figure 3**).

Bacteroidetes and Actinobacteria were ubiquitous in the investigated region and present in high relative abundance at some stations (e.g., stations N17 and N22, **Supplementary Figure 3A**). At the order level, the SAR11 clade within the Proteobacteria dominated in both relative (average of 32.4%) and quantitative (4.1×10^7–4.0×10^8 copies L^{-1} with an average of 2.6×10^8 copies L^{-1}) sequence abundances at all stations and sequences assigned to Synechococcales (Cyanobacteria) accounted for 21.9% of total sequences (**Figure 3**).

The Spearman correlation coefficients between relative and quantitative abundances were <0.8 in 46.2% of the cases, and 7.3% of the Spearman correlation coefficients were not significant (adjusted $P > 0.05$; **Supplementary Figure 4**). Furthermore, the community connections revealed by relative and quantitative abundances were markedly different, and the latter showed more complex and tighter connections (**Figure 4**, **Table 1**). Although equal numbers of OTUs were included in the analyses, compared to relative abundances, the quantitative abundances were associated with: (i) twice as many connections, with much more positive (1005 vs. 151) and fewer negative (29 vs. 143) connections; (ii) a higher average degree of clustering and clustering coefficient; and (iii) shorter average path length and lower modularity (**Table 1**). Procrustes analysis showed remarkable agreement between relative and quantitative abundances in presenting the distribution patterns of bacterial communities with $M^2 = 0.017$ and $p = 0.001$. Moreover, it was noteworthy that bacterial communities, based on both relative and quantitative abundances presented distinct geographical patterns with spatially adjacent stations clustered together (**Figure 5**).

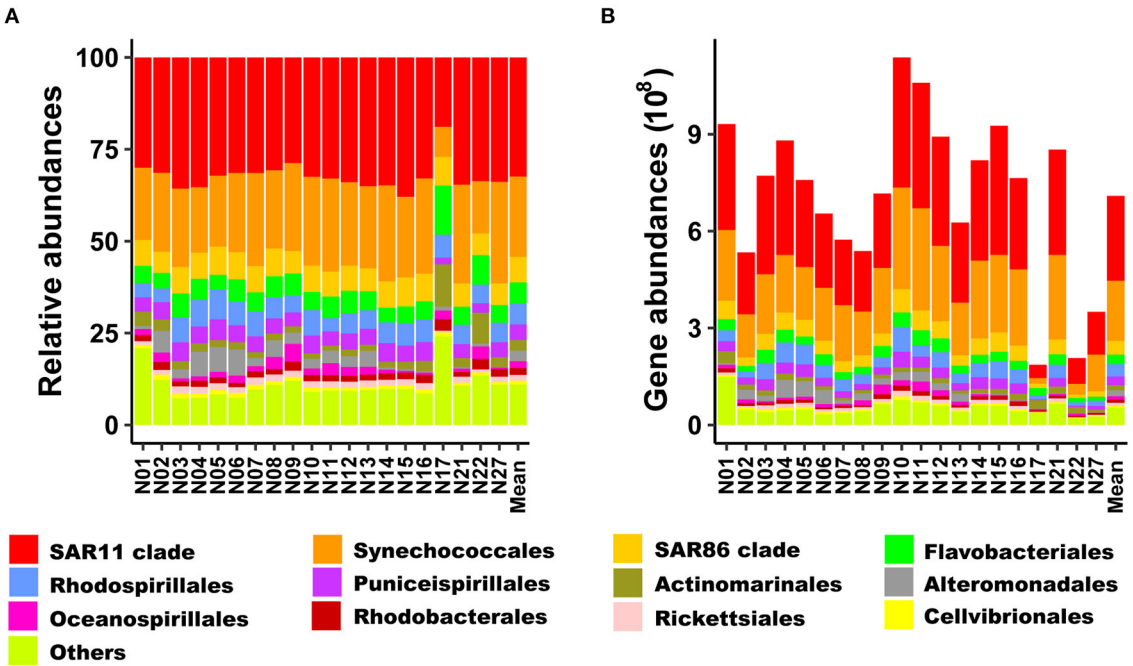


FIGURE 3 | Bacterial community composition of relative abundance and rRNA gene abundance at the order levels **(A,B)**, respectively. Only the most abundant orders (12) are shown. The other orders are presented as "Others".

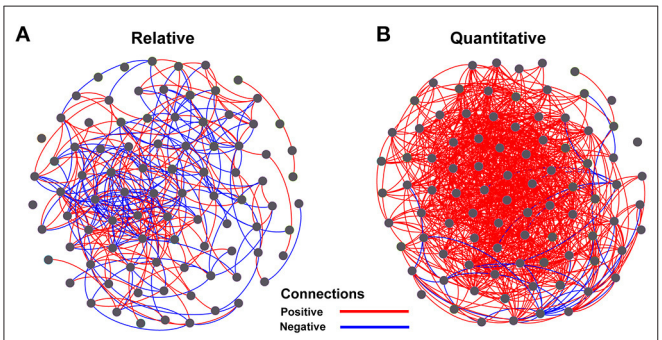


FIGURE 4 | Network analysis of relationships among OTUs for relative abundance **(A)** and quantitative abundance **(B)** of bacterial communities. The OTUs with relative abundances >0.1% of total sequences were included to calculate pairwise Spearman correlations using the function corr.test() in the package "psych," with *p*-values adjusted using the "fdr" method. Only significant correlations (adjusted $p < 0.05$ and $r \geq 0.6$) were used to construct the network using the software Gephi. Red and blue lines represent significant positive and negative correlations, respectively.

TABLE 1 | Network characteristics of bacterial community based on relative and quantitative abundances.

Attributes	Relative	Quantitative
Node	90	90
Average degree	6.5	23.0
Average clustering coefficient	0.49	0.63
Edge	294	1,034
Positive edge	151	1,005
Negative edge	143	29
Average path length	3.02	1.98
Modularity	11.45	0.26

Pairwise Spearman correlation coefficients for OTUs with relative abundance >0.1% of total sequences were calculated using the function corr.test() in the package "psych." The p-values were adjusted using the "fdr" method. Only significant correlations (adjusted p <0.05 and r ≥ 0.6) were used to analyze network attributes via the software Gephi.

Potentially Important Factors Influencing the Bacterial Communities

No significant relationship with measured environmental and geographic variables was found for alpha diversity indices or quantified rRNA gene abundances, except for bacterial abundances and NO_x concentrations, which were positively correlated with Shannon and PD indices, respectively (**Figure 2A**). The effects of environmental and geographic factors on community composition were evaluated using Mantel tests (**Table 2**), VPA (**Supplementary Figure 5**), and sPLS analyses (**Figure 6**, **Supplementary Figure 6**).

The variations of whole bacterial communities based on the Bray-Curtis distance were significantly correlated with salinity ($r = 0.396$, $P < 0.001$ and $r = 0.429$, $P < 0.001$ for relative and quantitative abundances, respectively), DRP concentrations ($r = 0.294$, $P = 0.002$ and $r = 0.322$, $P = 0.001$ for relative and quantitative abundances, respectively), $Si(OH)_4$ concentrations

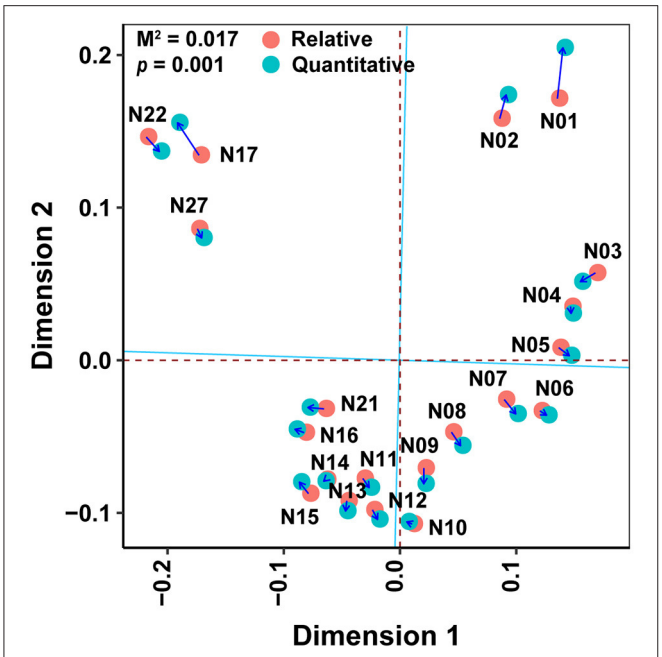


FIGURE 5 | Procrustes analysis of relative abundance against quantitative abundance in presenting bacterial distribution patterns obtained from Principal Coordinates Analysis. Principal Coordinates Analysis based on Bray–Curtis distance of bacterial communities. The M^2 that was approximated to zero indicated the remarkable agreement of comparison. The statistical significance was measured by the Monte Carlo test.

TABLE 2 | Mantel tests for the correlations between environmental variables, geographic factors, and bacterial composition based on Bray-Curtis distance for relative and quantitative abundances.

	Relative		Quantitative	
	r	*p*	*r*	*p*
Temperature	0.054	0.283	0.034	0.346
Salinity	**0.396**	**<0.001**	**0.429**	**<0.001**
NO_x	0.066	0.273	0.106	0.235
DRP	**0.294**	**0.002**	**0.322**	**0.001**
$Si(OH)_4$	**0.353**	**0.025**	**0.330**	**0.025**
Chl *a*	**0.367**	**0.014**	**0.291**	**0.035**
Bacteria	−0.102	0.719	−0.063	0.649
Envdist	**0.451**	**0.002**	**0.463**	**0.001**
Geodist	**0.738**	**<0.001**	**0.785**	**<0.001**

Significance tests were set at 9,999 permutations and p-values (<0.05) are in bold. NO_x, Nitrite + Nitrate; DRP, dissolved reactive phosphorus; BP, bacterial production; Bacteria, bacterial abundance; Envdist, pairwise Euclidean distances of environmental factors; Geodist, pairwise geographical distances based on Cartesian coordinates generated from longitude and latitude. Environmental variables were scaled to zero mean and unit variance.

($r = 0.353$, $P = 0.025$ and $r = 0.330$, $P = 0.025$ for relative and quantitative abundances, respectively), Chl *a* concentrations ($r = 0.367$, $P = 0.014$ and $r = 0.291$, $P = 0.035$ for relative and quantitative abundances, respectively), and geographic distances ($r = 0.738$, $P < 0.001$ and $r = 0.785$, $P < 0.001$ for relative and quantitative abundances, respectively) (**Table 2**). Furthermore, the VPA showed that purely geographic factors might play significant roles in shaping bacterial communities based on both relative (30.2%) and quantitative (31.9%) abundances, but purely measured environmental factors explained no significant percentage of community variance. A large percentage of community variance was unexplained based on relative (47.8%) and quantitative (47.0%) abundances (**Supplementary Figure 5**).

We used sPLS to identify and visualize significant correlations between subgroups of bacteria and each environmental parameter and geographic factor, and in this way, we were able to reveal many significant relationships (**Figure 6**, **Supplementary Figure 6**). In terms of both relative and quantitative abundances, groups such as Dadabacteria and Gammaproteobacteria were correlated with longitude and latitude, and bacteria recovered as phyla (Dadabacteria and Firmicutes) were correlated with DRP concentrations (**Supplementary Figure 6**). However, the relative or quantitative abundance of no phylum was found to be correlated with temperature, NO_x, or Chl *a* concentrations (**Supplementary Figure 6**). There were similar patterns at the order level for both relative and quantitative abundances. For example, the abundance of groups such as Alteromonadales, Bacillales, Betaproteobacteriales, cellvibrionales, Pirellulales, Pseudomonadales, and Verrucomicrobiales was correlated with longitude and latitude, whereas the abundances of lineages were uncorrelated with temperature and NO_x concentrations (**Figure 6**). Discrepancies of correlations with geographic factors, nutrients, and Chl *a* were also notably apparent for the relative and quantitative abundances of bacterial subgroups. For example, geographic factors correlated with the relative abundances of groups such as Ianctomycetes, Margulisbacteria, and Verrucomicrobia, but not for their quantitative abundances, whereas only the quantitative abundances of the orders Chitinophagales, Phycisphaerales, Planctomycetales, Puniceispirillales, and Opitutales were correlated with geographic variables (**Figure 6**, **Supplementary Figure 6**).

DISCUSSION

Bacterial Community Composition Across the Subtropical NWPO

Although HTQS does not preclude a PCR-bias effect, which would impact absolute microbial abundances, it could, to some extent, enhance the significance of HTS-based microbiome studies (Tourlousse et al., 2017; Wang et al., 2018; Yang et al., 2018; Jiang et al., 2019; Mou et al., 2020). In our study, the strong linear relationships observed between the added copy numbers of spike-in standards and the abundances of spike-in OTU sequences (**Supplementary Table 2**) suggested that HTQS might be a robust method for quantifying the sequence abundances. Our observations of total bacterial 16S rRNA abundances were within the range of bacterial 16S rRNA abundances ($1.78 \times 10^8 - 5.4 \times 10^9$ copies L^{-1}) reported in the surface seawater of the western North Atlantic, based on the HTQS method (Wang et al.,

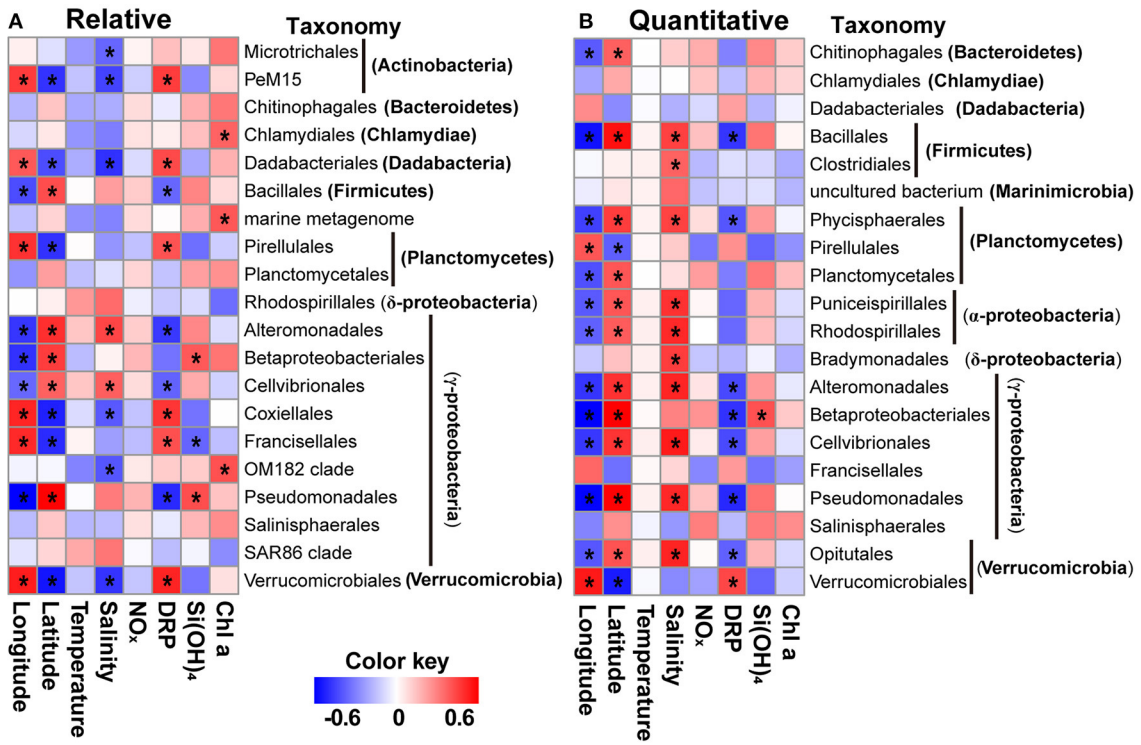


FIGURE 6 | Heatmaps showing significant correlations between specific orders **(A,B)** of bacteria and environmental variables (e.g., temperature, salinity, nutrients, and Chl *a*) and geographic factors for relative and quantitative abundances using a sparse partial least squares approach. The orders present in all samples were used. Significant correlations ($|r| > 0.5$) are labeled with asterisks. NO_x, nitrite + nitrate; DRP, dissolved reactive phosphorus.

2018) and consistent with bacterial 16S rRNA gene abundances based on real-time PCR in the surface seawater of the South Pacific Gyre (5.96×10^8–2.55×10^9 copies L^{-1}) (Yin et al., 2013). Quantitative abundances of the SAR11 16S rRNA gene in our samples were also in accordance with previous observations. For example, a mean quantitative abundance of 2.6×10^8 copies L^{-1} has been reported for SAR11 in surface seawater of the western North Atlantic based on the HTQS method (Wang et al., 2018), and SAR11 gene abundances estimated by quantitative PCR (qPCR) have been reported to fall in the range 0.3×10^8–6.3×10^8 copies L^{-1} at Station ALOHA (Eiler et al., 2009).

Our results were also consistent with the SAR11 cell abundances of 2×10^8 cells L^{-1} estimated from FISH in the Sargasso Sea (Morris et al., 2002), if there is one gene copy number per SAR11 cell (Giovannoni et al., 2005). It should be noted, however, that there was no correlation between bacterial abundance enumerated with flow cytometry and 16S rRNA gene abundance based on the HTQS method (**Figure 2A**). Because there are significant methodological differences between HTQS and flow cytometry, it was reasonable to anticipate that they do not provide estimates of the same characteristics of bacterial communities.

At the basin scale, the bacterial communities were dominated at the phylum level by Proteobacteria, followed by Cyanobacteria (**Supplementary Figure 3**). Furthermore, the SAR11 (Proteobacteria) and Synechococcales (Cyanobacteria) lineages were the predominant groups at the order level in our study (**Figure 3**).

With some exceptions, our results were generally consistent with the results of previous studies in oligotrophic surface seawater at both the global and local scales (Yin et al., 2013; Sunagawa et al., 2015; Lindh, 2017; Shulse et al., 2017; Li et al., 2018). Based on metagenomic data, for example, Sunagawa et al. (2015) found that typical members of the Proteobacteria (i.e., SAR11 and SAR86 clades) and Cyanobacteria were the dominant bacteria in samples collected across all oceanic provinces during the Tara Oceans expedition. Moreover, Proteobacteria and Cyanobacteria have generally been shown to be the predominant phyla in surface seawater based on either 16S rRNA gene clone libraries from the center to the edge of the South Pacific Gyre (Yin et al., 2013) or HTS results of the 16S rRNA gene (V3 region) in the NWPO (Li et al., 2018). However, Lindh (2017) found that Cyanobacteria dominated bacterial communities along a transect from Honolulu to Station ALOHA based on HTS of the V3–V4 region of the bacterial 16S rRNA gene, and Li et al. (2018) found that Bacteroidetes was the second most abundant phylum at several stations in the northwestern Pacific. These discrepancies could be partially attributed to the different methodologies used, such as PCR primers. Previous studies, for example, have revealed that different primer sets can result in biased diversity metrics for bacterial communities (Cai et al., 2013). The fact that the primers used by Lindh (2017) and by us were the same, but

that there were differences in the community compositions based on those primers indicates that factors other than primer sets contributed to the observed differences. Li et al. (2018) have shown that either Cyanobacteria or Proteobacteria can be the dominant bacterial phylum in the surface water of the NWPO. We, therefore, hypothesize that differences of the environmental factors associated with currents, water masses (Tseng et al., 2016), and/or physical processes such as eddies and upwelling might play a more important role than methodological differences (i.e., different primers) in explaining the differences of bacterial community composition in the different studies.

Potentially Important Factors That Influence Bacterial Communities

We identified environmental variables (i.e., salinity, nutrient, and Chl *a* concentrations) and geographic factors as potential determinants of bacterial communities (**Table 2**, **Supplementary Figure 5**). Moreover, these potential determinants were correlated to varying degrees with different bacterial lineages (**Figure 6**, **Supplementary Figure 6**). More specifically, Chl *a* concentrations were found to be potentially important factors that shaped whole bacterial communities (**Table 2**), but they were uncorrelated with specific lineages at the phylum level (**Supplementary Figure 5**). Similarly, Li et al. (2018) found that bacterial community structure is positively correlated only with temperature, whereas, at the phylum level, Proteobacteria, and Cyanobacteria are weakly correlated with both temperature and nutrients, and Chloroflexi are negatively correlated with ammonium concentrations in the surface water of the NWPO.

Partial least squares regression analyses revealed that latitude, longitude, and temperature are associated with specific groups of prokaryotic taxa in the western North Atlantic (Wang et al., 2018). In addition, Lindh et al. (2018) have shown that the effects of environmental and spatial factors vary in terms of their ability to explain whole bacterioplankton assemblages and subgroups thereof in the Clarion-Clippperton zone of the Pacific Ocean and along global ocean transects (i.e., the TARA and Malaspina expeditions). The implication is that the mechanisms that shape bacterioplankton communities differ between whole communities and subgroups. To comprehensively understand the biogeography of marine bacteria and the mechanisms that structure marine bacterial communities, the distinction between different taxonomic levels should be taken into consideration in future studies.

Our results revealed that spatial factors might play more important roles than environmental variables in regulating bacterial communities (**Table 2**, **Supplementary Figure 5**). Mantel tests revealed that spatial factors were more significantly correlated with bacterial communities than environmental variables (**Table 2**), and VPA showed that purely spatial factors could explain a large percentage of the variance of bacterial community composition, whereas the environmental metrics that we measured explained no significant percentage of the variance (**Supplementary Figure 5**). By contrast, Lindh et al. (2018) found that purely environmental conditions accounted for a significant percentage of the variance of epipelagic (5, 80, and 125 m) bacterioplankton composition in the Clarion-Clipperton Zone of the Eastern North Pacific. Based on rRNA gene data from the Tara oceans, the same study observed that environmental conditions might play a more important role than spatial differences in structuring epipelagic (0–200 m) bacterial assemblages. Water temperature was recognized to be the most important driver of the selection of surface prokaryotes sampled during the Malaspina and TARA Oceans expeditions (Logares et al., 2020). One possible explanation for these disparate results is that environmental heterogeneity was much higher in previous studies than in this study. For example, water temperatures ranged from 28.55 to 29.68°C in this study, but it fluctuated between 12.79 and 27.63°C in the epipelagic waters of the Clarion-Clipperton Zone during the study of Lindh et al. (2018) and from 15.7 to 29.3°C across the "Meta-119 Malaspina dataset" during the study of Logares et al. (2020).

Our results also revealed that environmental variables such as salinity, DRP, and $Si(OH)_4$ concentrations were correlated with spatial factors (**Figure 2A**), and that correlation could explain their shared effect on the variations of bacterial community structure (**Supplementary Figure 4**). It should be noted that there was a large percentage of community variance that was unexplained by the spatial and environmental variables in our study (**Supplementary Figure 5**). We speculate that this unexplained community variance might have been due to unmeasured but important environmental factors (e.g., wind speeds and water current velocities), ecological processes (e.g., dispersal and ecological drift), and/or biological variables, such as taxon-specific mortality by grazers, viral infection, and biotic interactions.

Comparing the Quantitative and Relative Abundances of Members of the Bacterial Communities

We evaluated community structure, distribution patterns, and the potentially important determinants of bacterial community composition in terms of both relative and quantitative abundances. The distribution patterns of the relative and quantitative abundances of whole communities of bacteria were similar (**Figure 5**) and were affected in similar ways by potential determinants of community structure (**Table 2**, **Supplementary Figure 5**). At the subgroup level, the distributions and potential determinants of relative and quantitative abundance differed (**Table 1**, **Figures 4**, **6**, **Supplementary Figures 4**, **6**). The attributes of networks constructed from relative and quantitative abundances were quite different (**Figure 4**, **Table 1**). Because positive connections might tend to destabilize bacterial communities, or vice versa (de Vries et al., 2018), our results may indicate that bacterial networks appear to be less stable when they are assessed in terms of quantitative abundance vs. relative abundance. Furthermore, the Spearman correlation coefficients between relative and quantitative abundances were <0.8 in 46.2% of cases, and 7.3% were not significant correlations (adjusted $P > 0.05$; **Supplementary Figure 4**). Hence there were significant

differences between relative and quantitative abundances at the subgroup level.

By combining the HTS approach with single-cell enumeration technology, Props et al. (2017) showed that there is no inevitable correlation between enrichment (increase in relative abundance) and outgrowth (increase in absolute abundance) of taxa. This result highlights the need to consider both relative and absolute abundances to present a comprehensive interpretation of ecological scenarios. The consideration of the potentially important factors that might shape bacterial community composition in this study was based on relative and quantitative abundances that revealed similar patterns for whole communities, but the patterns were dissimilar for specific subgroups (**Table 2, Figure 6**, **Supplementary Figures 4**, **6**). Previous studies have shown significantly different trends in the abundances of major groups within microbial communities, based on relative abundances generated using HTS and measured abundances estimated by direct or indirect measurements with adenosine triphosphate, flow cytometry, quantitative real-time PCR (qPCR), concentrations of phospholipid fatty acids or microbial carbon biomass (Zhang et al., 2017), and internal standard strains (Yang et al., 2018). In addition, Vandeputte et al. (2017) revealed that quantitative microbiome profiling plays a very important role in the analysis of relationships between species that occur together and the characterization of changes of pathogenic microorganisms through parallelization of HTS and flow cytometric enumeration technology. Collectively, the trends of the relative and quantitative abundances of marine bacteria might be significantly different at subgroup levels. A consideration of the quantitative and relative abundances of members of the bacterial community might therefore be pivotal in revealing important aspects of marine microbial ecology.

CONCLUSIONS

Although bacteria play a pivotal role in shaping ecosystems and contributing to the cycling of elements and flow of energy in the oceans, few studies have addressed the basin-scale distribution of marine bacteria based on the quantification of bacterial abundance. This study revealed that total 16S rRNA gene copies ranged from 1.86×10^8 to 1.14×10^9 copies L^{-1} in the subtropical NWPO. The spatial distributions of the bacterial communities were distinct, and geographic factors appeared to play important roles in structuring bacterial communities. Our analyses indicated that consideration of both the relative and quantitative abundances of bacteria in a community might help to reveal important aspects of marine microbial ecology. However, because our conclusions are based on analysis of only surface seawater sampled from the subtropical NWPO, further investigations are needed to extend this approach to subsurface waters (e.g., the deep chlorophyll maximum layer, twilight zone, and deep ocean), and/or to regions with gradients of environmental conditions (e.g., salinity and nutrient concentrations). Such studies would greatly enhance our knowledge of how ocean ecosystems work and how they may respond to climate change.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA614577.

AUTHOR CONTRIBUTIONS

BH and JK conceived and designed the study. JK and LW collected the samples. JK and JG conducted the experiments. JK analyzed the data. All authors wrote the manuscript.

FUNDING

This research was supported by a grant for the Monitoring and Protection of Ecology and Environment in the East Pacific Ocean (Grant No. DY135-E2-5-5), National Natural Science Foundation of China (Grant Nos. 41890803 and 41730533), and the Scientific Research Foundation of Third Institute of Oceanography, Ministry of Natural Resources (Grant No. 2018017).

ACKNOWLEDGMENTS

We would like to thank the crew of the R/V Xiangyanghong 03 for their assistance in sampling. We are also thankful to Lizhen Lin, Jixin Chen, and Lei Wang for their help with laboratory experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.599614/full#supplementary-material

REFERENCES

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM* 8, 361–362.

Borcard, D., Gillet, F., and Legendre, P. (2011). *Numerical Ecology with R*. New York, NY: Springer.

Cai, L., Ye, L., Tong, A. H., Lok, S., and Zhang, T. (2013). Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE* 8:e53649. doi: 10.1371/journal.pone.0053649

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Dabney, A., Storey, J. D., and Warnes, G. (2010). *Qvalue: Q-value Estimation for False Discovery Rate Control*. R package version 1.

de Vries, F. T., Griffiths, R. I., Bailey, M., Craig, H., Girlanda, M., Gweon, H. S., et al. (2018). Soil bacterial networks are less stable under drought than fungal networks. *Nat. Commun.* 9:3033. doi: 10.1038/s41467-018-05516-7

Du, C., Liu, Z., Dai, M., Kao, S. J., Cao, Z., Zhang, Y., et al. (2013). Impact of the Kuroshio intrusion on the nutrient inventory in the upper northern South China Sea: insights from an isopycnal mixing model. *Biogeosciences* 10, 6419–6432. doi: 10.5194/bg-10-6419-2013

Ducklow, H. (2000). Bacterial production and biomass in the oceans. *Microbial Ecol. Oceans* 1, 85–120.

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Eiler, A., Hayakawa, D. H., Church, M. J., Karl, D. M., and Rappe, M. S. (2009). Dynamics of the SAR11 bacterioplankton lineage in relation to environmental conditions in the oligotrophic North Pacific subtropical gyre. *Environ. Microbiol.* 11, 2291–2300. doi: 10.1111/j.1462-2920.2009.01954.x

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10. doi: 10.1016/0006-3207(92)91201-3

Fuhrman, J. A., Steele, J. A., Hewson, L., Schwalbach, M. S., Brown, M. V., Green, J. L., et al. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7774–7778. doi: 10.1073/pnas.0803070105

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309, 1242–1245. doi: 10.1126/science.1114057

Goodwin, S., Mcpherson, J. D., and Mccombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. doi: 10.1038/nrg.2016.49

Herlemann, D. P., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41

Hutchins, D. A., and Fu, F. (2017). Microorganisms and ocean global change. *Nat. Microbial.* 2:17058. doi: 10.1038/nmicrobiol.2017.58

Jiang, S. Q., Yu, Y. N., Gao, R. W., Wang, H., Zhang, J., Li, R., et al. (2019). High-throughput absolute quantification sequencing reveals the effect of different fertilizer applications on bacterial community in a tomato cultivated coastal saline soil. *Sci. Total Environ.* 687, 601–609. doi: 10.1016/j.scitotenv.2019.06.105

Karl, D. M., and Church, M. J. (2017). Ecosystem structure and dynamics in the North Pacific subtropical gyre: new views of an old ocean. *Ecosystems* 20, 433–457. doi: 10.1007/s10021-017-0117-0

Kavanaugh, M. T., Church, M. J., Davis, C. O., Karl, D. M., Letelier, R. M., and Doney, S. C. (2018). ALOHA from the edge: reconciling three decades of in situ eulerian observations and geographic variability in the North Pacific subtropical gyre. *Front. Mar. Sci.* 5:130. doi: 10.3389/fmars.2018.00130

Kirchman, D. L. (2016). Growth rates of microbes in the oceans. *Ann. Rev. Mar. Sic.* 8, 285–309. doi: 10.1146/annurev-marine-122414-033938

Li, Y. Y., Chen, X. H., Xie, Z. X., Li, D. X., Wu, P. F., Kong, L. F., et al. (2018). Bacterial diversity and nitrogen utilization strategies in the upper layer of the Northwestern Pacific Ocean. *Front. Microbial.* 9:797. doi: 10.3389/fmicb.2018.00797

Lindh, M. V. (2017). There and back again–unraveling mechanisms of bacterial biogeography in the North pacific subtropical gyre to and from station ALOHA. *bioRxiv* 141085. doi: 10.1101/141085

Lindh, M. V., Maillot, B. M., Smith, C. R., and Church, M. J. (2018). Habitat filtering of bacterioplankton communities above polymetallic nodule fields and sediments in the clarion-clipperton zone of the Pacific ocean. *Environ. Microbiol. Rep.* 10, 113–122. doi: 10.1111/1758-2229.12627

Liu, X., Xiao, W., Landry, M. R., Chiang, K. P., Wang, L., and Huang, B. (2016). Responses of phytoplankton communities to environmental variability in the East China sea. *Ecosystems* 19, 832–849. doi: 10.1007/s10021-016-9970-5

Logares, R., Deutschmann, I. M., Junger, P. C., Giner, C. R., Krabberod, A. K., Schmidt, T. S. B., et al. (2020). Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* 8:55. doi: 10.1186/s40168-020-00827-8

Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Marie, D., Partensky, F., Vaulot, D., and Brussaard, C. (1999). Enumeration of phytoplankton, bacteria, and viruses in marine samples. *Curr. Protoc. Cytometr.* 10:11. doi: 10.1002/0471142956.cy1111s10

Martiny, J. B., Eisen, J. A., Penn, K., Allison, S. D., and Horner-Devine, M. C. (2011). Drivers of bacterial beta-diversity depend on spatial scale. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7850–7854. doi: 10.1073/pnas.1016308108

Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., et al. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420, 806–810. doi: 10.1038/nature01240

Mou, J., Li, Q., Shi, W., Qi, X., Song, W., and Yang, J. (2020). Chain conformation, physicochemical properties of fucosylated chondroitin sulfate from sea cucumber Stichopus chloronotus and its *in vitro* fermentation by human gut microbiota. *Carbohydr. Polym.* 228:115359. doi: 10.1016/j.carbpol.2019.115359

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., et al. (2016). *vegan: Community Ecology Package, version 2.3–5.*

Props, R., Kerckhof, F. M., Rubbens, P., De Vrieze, J., Hernandez Sanabria, E., Waegeman, W., et al. (2017). Absolute quantification of microbial taxon abundances. *ISME J.* 11, 584–587. doi: 10.1038/ismej,.2016.117

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Revelle, W. R. (2017). *Psych: Procedures for Personality and Psychological Research*.

Rohart, F., Gautier, B., Singh, A., and Le Cao, K. A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol* 13, e1005752. doi: 10.1371/journal.pcbi.1005752

Schlitzer, R. (2015). *Ocean Data View*.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–7541. doi: 10.1128/AEM.01541-09

Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D., et al. (2018). Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain. *Microbiome* 6:27. doi: 10.1186/s40168-018-0409-4

Shulse, C. N., Maillot, B., Smith, C. R., and Church, M. J. (2017). Polymetallic nodules, sediments, and deep waters in the equatorial North Pacific exhibit highly diverse and distinct bacterial, archaeal, and microeukaryotic communities. *Microbiologyopen* 6:e00428. doi: 10.1002/mbo3.428

Spalding, M. D., Agostini, V. N., Rice, J., and Grant, S. M. (2012). Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. *Ocean Coast. Manag.* 60, 19–30. doi: 10.1016/j.ocecoaman.2011.12.016

Steinberg, D. K., and Landry, M. R. (2017). Zooplankton and the ocean carbon cycle. *Ann. Rev. Mar. Sic.* 9, 413–444. doi: 10.1146/annurev-marine-010814-015924

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359

Tourlousse, D. M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., and Sekiguchi, Y. (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res* 45:e23. doi: 10.1093/nar/gkw984

Tseng, Y.-H., Lin, H., Chen, H.-C., Thompson, K., Bentsen, M., Böning, C. W., et al. (2016). North and equatorial Pacific ocean circulation in the CORE-II hindcast simulations. *Ocean Model.* 104, 143–170. doi: 10.1016/j.ocemod.2016.06.003

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., and Valles-Colomer, M., Sabino, et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 551, 507–511. doi: 10.1038/nature24460

Wang, S., Lin, Y., Gifford, S., Eveleth, R., and Cassar, N. (2018). Linking patterns of net community production and marine microbial community structure in the western North Atlantic. *ISME J.* 2582–2595. doi: 10.1038/s41396-018-0163-4

Yang, L., Lou, J., Wang, H., Wu, L., and Xu, J. (2018). Use of an improved high-throughput absolute abundance quantification method to characterize soil bacterial community and dynamics. *Sci. Total Environ.* 633, 360–371. doi: 10.1016/j.scitotenv.2018.03.201

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* 42, D643–648. doi: 10.1093/nar/gkt1209

Yin, Q., Fu, B., Li, B., Shi, X., Inagaki, F., and Zhang, X. H. (2013). Spatial variations in microbial community composition in surface seawater from the ultra-oligotrophic center to rim of the South Pacific Gyre. *PLoS ONE* 8:e55148. doi: 10.1371/journal.pone.0055148

Yuan, Y., Wang, S., Yuan, D., and Ma, J. (2016). A simple and cost-effective manual solid phase extraction method for the determination of nanomolar dissolved reactive phosphorus in aqueous samples. *Limnol. Oceanogr.* 14, 79–86. doi: 10.1002/lom3.10072

Zhang, J. Z. (2000). Shipboard automated determination of trace concentrations of nitrite and nitrate in oligotrophic water by gas-segmented continuous flow analysis with a liquid waveguide capillary flow cell. *Deep Sea Res. Oceanogr. Res. Pap.* 47, 1157–1171. doi: 10.1016/S0967-0637(99)00085-0

Zhang, Z., Qu, Y., Li, S., Feng, K., Wang, S., Cai, W., et al. (2017). Soil bacterial quantification approaches coupling with relative abundances reflecting the changes of taxa. *Sci. Rep.* 7:4837. doi: 10.1038/s41598-017-05260-w

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B., et al. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS ONE* 6:e24570. doi: 10.1371/journal.pone.0024570

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 February 2021
doi: 10.3389/fgene.2021.586293

Dynamic Allocation of Carbon Storage and Nutrient-Dependent Exudation in a Revised Genome-Scale Model of *Prochlorococcus*

Shany Ofaim[1,2†], Snorre Sulheim[1,3,4†], Eivind Almaas[3,5], Daniel Sher[2] and Daniel Segrè[1,6,7,8]*

[1] Bioinformatics Program and Biological Design Center, Boston University, Boston, MA, United States, [2] Department of Marine Biology, University of Haifa, Haifa, Israel, [3] Department of Biotechnology and Food Science, NTNU – Norwegian University of Science and Technology, Trondheim, Norway, [4] Department of Biotechnology and Nanomedicine, SINTEF Industry, Trondheim, Norway, [5] K.G. Jebsen Center for Genetic Epidemiology, NTNU – Norwegian University of Science and Technology, Trondheim, Norway, [6] Department of Biomedical Engineering, Boston University, Boston, MA, United States, [7] Department of Physics, Boston University, Boston, MA, United States, [8] Department of Biology, Boston University, Boston, MA, United States

OPEN ACCESS

Edited by:
Karoline Faust,
KU Leuven, Belgium

Reviewed by:
Abhinav Achreja,
University of Michigan, United States
Adam Martiny,
University of California, Irvine, United States
Steffen Waldherr,
KU Leuven, Belgium

***Correspondence:**
Daniel Segrè
dsegre@bu.edu

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: 22 July 2020
Accepted: 14 January 2021
Published: 09 February 2021

Citation:
Ofaim S, Sulheim S, Almaas E, Sher D and Segrè D (2021) Dynamic Allocation of Carbon Storage and Nutrient-Dependent Exudation in a Revised Genome-Scale Model of Prochlorococcus. Front. Genet. 12:586293. doi: 10.3389/fgene.2021.586293

Microbial life in the oceans impacts the entire marine ecosystem, global biogeochemistry and climate. The marine cyanobacterium *Prochlorococcus*, an abundant component of this ecosystem, releases a significant fraction of the carbon fixed through photosynthesis, but the amount, timing and molecular composition of released carbon are still poorly understood. These depend on several factors, including nutrient availability, light intensity and glycogen storage. Here we combine multiple computational approaches to provide insight into carbon storage and exudation in *Prochlorococcus*. First, with the aid of a new algorithm for recursive filling of metabolic gaps (ReFill), and through substantial manual curation, we extended an existing genome-scale metabolic model of *Prochlorococcus* MED4. In this revised model (*i*SO595), we decoupled glycogen biosynthesis/degradation from growth, thus enabling dynamic allocation of carbon storage. In contrast to standard implementations of flux balance modeling, we made use of forced influx of carbon and light into the cell, to recapitulate overflow metabolism due to the decoupling of photosynthesis and carbon fixation from growth during nutrient limitation. By using random sampling in the ensuing flux space, we found that storage of glycogen or exudation of organic acids are favored when the growth is nitrogen limited, while exudation of amino acids becomes more likely when phosphate is the limiting resource. We next used COMETS to simulate day-night cycles and found that the model displays dynamic glycogen allocation and exudation of organic acids. The switch from photosynthesis and glycogen storage to glycogen depletion is associated with a redistribution of fluxes from the Entner–Doudoroff to

the Pentose Phosphate pathway. Finally, we show that specific gene knockouts in *i*SO595 exhibit dynamic anomalies compatible with experimental observations, further demonstrating the value of this model as a tool to probe the metabolic dynamic of *Prochlorococcus*.

Keywords: constraint-based reconstruction and analysis (COBRA), flux balance analysis (FBA), computation of microbial ecosystems in time and space (COMETS), cyanobacteria, exudation, gap-filling algorithm, photosynthesis

INTRODUCTION

Marine phytoplankton perform about one-half of the photosynthesis on Earth (Field et al., 1998). *Prochlorococcus* is one of the most abundant phytoplankton clades in the world's oceans and is estimated to produce about 4 Gt of organic carbon annually (Flombaum et al., 2013). As such, these clades play a key role in a variety of ecosystems (Partensky and Garczarek, 2010; Biller et al., 2015). Recent evolutionary studies suggested several evolved metabolic innovations contributing to high picocyanobacterial abundance in the harsh oligotrophic ocean waters, usually limited by several nutrients such as nitrogen, phosphorus, and iron. These innovations include a proteome that contains less nitrogen rich amino acids (Gilbert and Fagan, 2011), membranes that contain glyco- and sulfolipids rather than phospholipids (Van Mooy et al., 2006) and streamlining of the genome associated with outsourcing of important cellular functions to co-occurring organisms (Holtzendorff et al., 2008; Partensky and Garczarek, 2010; Morris et al., 2012; Ma L. et al., 2018; Braakman, 2019).

Another innovation employed by these organisms is an increased metabolic rate that in turn manifest in the exudation of organic compounds (Fogg et al., 1965; Mague et al., 1980; López-Sandoval et al., 2013; Braakman et al., 2017; Braakman, 2019; Moran and Durham, 2019). Typically, 2–25% of the carbon fixed by photosynthesis is released by exudation from the cell, although values as high as 90% have been reported (Bertilsson et al., 2005; López-Sandoval et al., 2013; Rothrosenberg et al., 2019; Szul et al., 2019). This exudation, combined with cell death, lytic viral infections, and grazing debris made by predators ("sloppy feeding"), makes dissolved organic matter of phytoplankton origin omnipresent in natural waters (Thornton, 2014). However, it is currently impossible to provide a universal chemical description of dissolved organic matter (Kujawinski, 2011; Arrieta et al., 2015; Moran et al., 2016), partly because the exuded organic compounds differ between strains and environmental conditions (Becker et al., 2014; Ma X. et al., 2018). Nevertheless, in general, phytoplankton exudate includes a small proportion of low-molecular weight compounds, such as organic acids, carbohydrates, and amino acids (Bertilsson et al., 2005), as well as a larger proportion of complex, high-molecular weight compounds (Kujawinski, 2011). Another strategy employed by these bacteria to manage their carbon budget is the internal storage of carbon in polymeric form, specifically, glycogen (Zinser et al., 2009; Reimers et al., 2017; Luan et al., 2019). The extent to which *Prochlorococcus,* in particular, also stores glycogen has recently been measured, showing increased glycogen pools (up to 40 fg cell^{-1}) in nitrogen-limited conditions compared to nitrogen-replete (Szul et al., 2019). Glycogen accumulates in the bacterial cell during the light hours and was recently suggested to have two primary roles; as energy storage in preparation for darkness and as a regulation strategy to manage high-light photosynthesis products (Welkie et al., 2019). The allocation of glycogen is suggested to be tightly associated with the overflow metabolism hypothesis and also known to be widely affected by nutrient limitations (Damrow et al., 2016; Cano et al., 2018; Forchhammer and Schwarz, 2019; Szul et al., 2019). Importantly, the carbon fixed and released by phytoplankton is then used by heterotrophic organisms as a source of energy, whereas the heterotrophic bacteria may recycle nutrient elements and support the growth of phytoplankton in other ways, as suggested by the Black Queen Hypothesis (Amin et al., 2012; Morris et al., 2012; Moran et al., 2016; Cirri and Pohnert, 2019; Moran and Durham, 2019). Thus, carbon fixation, storage and release are tightly intertwined with microbial interactions and microbial ecosystem dynamics.

Quantitative models at various scales have provided critical insights into how ocean microbial ecosystems function, and how they are related to broader biogeochemical cycles (Deutsch et al., 2007; Follows et al., 2007; Arteaga et al., 2016; Coles et al., 2017; Foster et al., 2018; Moradi et al., 2018; Nicholson et al., 2018; Braakman, 2019; Oschlies et al., 2019; Ward et al., 2019). Most of these models represent organisms in terms of simplified stoichiometric reactions converting elements into biomass, thus making it possible to incorporate biological processes into dynamic-coupled Earth System models (Follows et al., 2007; Reid, 2012). The exponential increase in genomic information on marine organisms provides an opportunity to seek methods to link such detailed genome-scale information to biochemical flows (Coles et al., 2017). In recent years, genome-scale metabolic models (GEMs), combined with linear programming, have made it possible to produce testable predictions of metabolic phenotypes of individual organisms or microbial communities (Gu et al., 2019). This computational framework is based on the identification of individual enzymes and transporters in an organism's genome, and on simplifying assumptions that bypass the need for kinetic parameters (Maarleveld et al., 2013; O'Brien et al., 2015; Casey et al., 2016; Kim et al., 2016; Reimers et al., 2017). While genome-scale modeling has proven to be a powerful approach in cyanobacterial model organisms such as *Synechocystis* sp. PCC 6803, *Synechococcus elongatus* PCC 7942 and *Prochlorococcus* MED4 (Kettler et al., 2007; Knoop et al., 2013; Broddrick et al., 2016; Casey et al., 2016; Yoshikawa et al., 2017), the exudation of organic compounds in

phototrophic organisms has not been studied in detail through Flux Balance Analysis (FBA) or similar methods (Varma and Palsson, 1994; Orth et al., 2010). On the other hand, several examples exist of FBA-based predictions of exudation-mediated interactions between different species, including those generated using the Computational of Microbial Ecosystems in Time and Space (COMETS) platform (Harcombe et al., 2014). In fact, FBA calculations also suggest that "costless" secretions (i.e., secretions that do not induce a fitness cost) might be quite common, and can support the growth of co-occurring organisms (Pacheco et al., 2018).

Experimental evidence and theoretical considerations indicate that *Prochlorococcus* exudes different metabolites in a way that strongly depends on environmental conditions (Dubinsky and Berman-Frank, 2001; Szul et al., 2019) as well as on the strain's genetic makeup (Becker et al., 2014; Roth-rosenberg et al., 2019). While GEMs can be used to predict these fluxes, they require modifications to deal with processes not usually considered in FBA, including: (a) the special nature of photon fluxes [which, unlike molecular fluxes, cannot easily be "shut off" at short time scales (Dubinsky and Berman-Frank, 2001)]; (b) the buffering role of intracellular storage molecules such as glycogen. The primary focus of this study is to obtain better knowledge of the potential metabolic effect of a combination of key nutrients (carbon, nitrogen, phosphorous, and light) and carbon fixation rate on the allocation (including storage and exudation) of carbon in *Prochlorococcus* using a revised genome-scale metabolic model (**Figure 1**). We start by describing model revisions and updates to capture the current, most complete metabolic knowledge available for *Prochlorococcus*. Next, we use a variety of FBA approaches to uncover the potential relationships between a set of key nutrients, carbon storage and exudates in static and dynamic (time dependent) settings. The implementation and use of these approaches improve our understanding of the intricate metabolic workings of *Prochlorococcus* and provide insights on its storage and exudation trends under different environmental conditions.

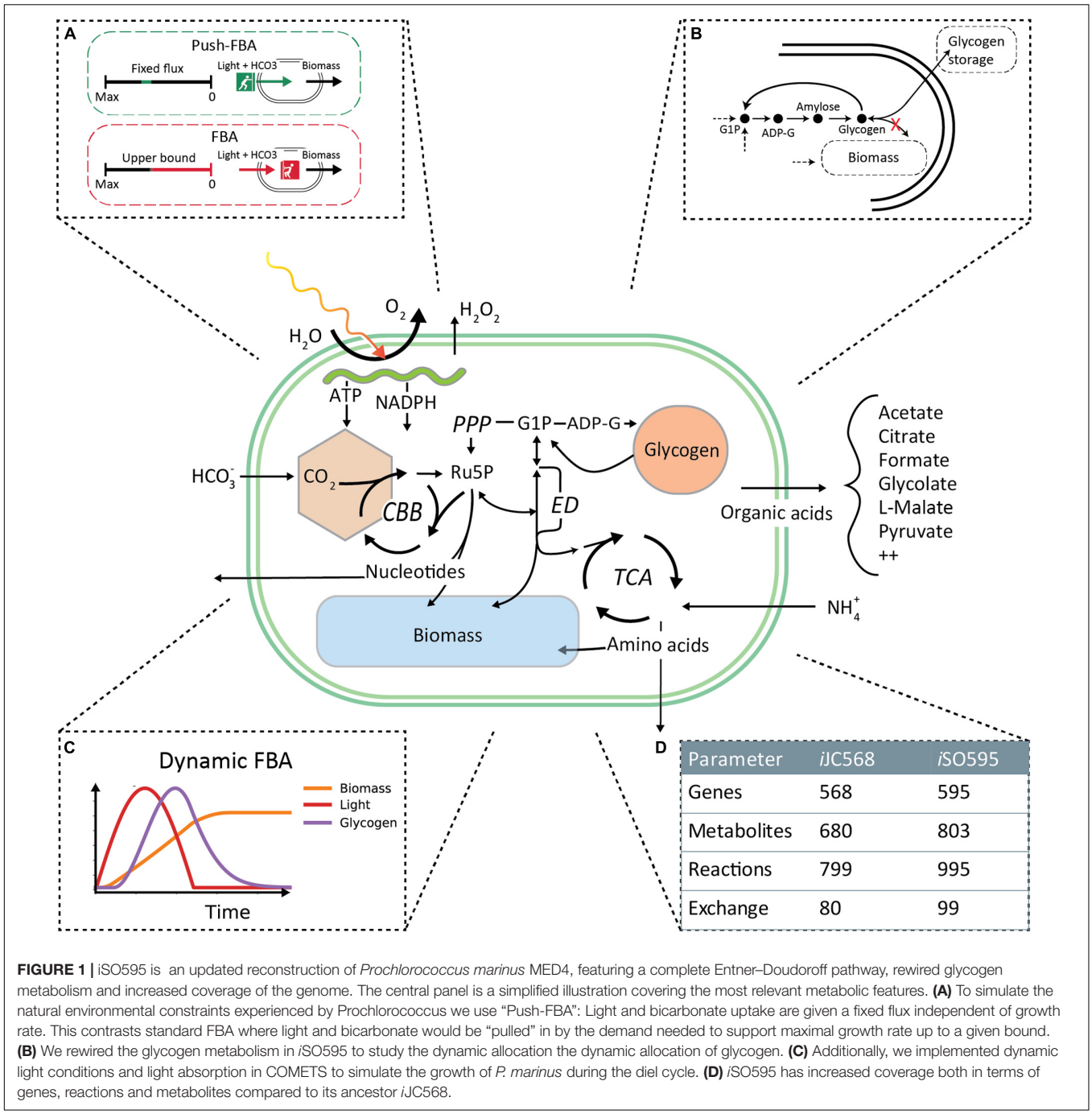
MATERIALS AND METHODS

Model Update and Curation

The *i*JC568 genome-scale reconstruction of *Prochlorococcus marinus subsp. pastoris str.* CMP1986 (referred to throughout the manuscript as MED4) as described by Casey et al. (2016), was used as the starting point for model enhancement. The update process started with an in-depth study of the reconstructed network and available knowledge not previously incorporated into the model of the organism. During this process, we ended up implementing the following specific steps of curation and update: (i) A key modification to the model was the decoupling between the glycogen storage flux and the biomass production. In standard stoichiometric reconstructions for FBA modeling (Thiele et al., 2011; Nogales et al., 2012; Feist et al., 2014; Broddrick et al., 2016; Monk et al., 2017; Kavvas et al., 2018), glycogen is listed as one of the biomass components, thus accounting for the carbon flux into storage. However, given the fixed stoichiometry of biomass composition, this classical implementation cannot account for the time-dependent storage and re-utilization of glycogen observed in picocyanobacteria. We thus removed the glycogen from the biomass function and streamlined the existing glycogen granule representation to a direct link between ADP-Glucose to the production of glycogen (**Figure 1B**). (ii) In addition to targeted refinement of selected reactions, we used the KEGG database (Kanehisa and Goto, 2000) to perform an extensive search for previously known but missing metabolic reaction annotations. Indeed, we found 354 reactions that could be potentially added to the existing network. To incorporate this knowledge, we developed a semi-automated algorithm (ReFill, described below). (iii) We coupled the implementation of the algorithm with several steps of manual curation. These included the addition of transports, such as that of hydrogen peroxide and ethanol, known to diffuse across the cell membrane (Seaver and Imlay, 2001; Noreña-Caro and Benton, 2018), and the addition of the complete Entner–Doudoroff pathway, that has recently been discovered in cyanobacteria (Chen et al., 2016). Additionally, we performed a BLAST search (**Supplementary Material 1**) (Altschul et al., 1990) from which we identified 6PG-dehydratase (EC: 4.2.1.12) encoded by PMM0774, thereby completing this pathway in the model reconstruction. (iv) The revised model was checked for redox and elemental balance. Since the biomass function was based on experimental data (Casey et al., 2016), it was not updated. In line with best practices, a *memote* quality assessment (**Supplementary Material 2**) (Lieven et al., 2020), as well as model files and a detailed changelog, are provided at https://github.com/segrelab/Prochlorococcus_Model. All reactions added to *i*JC658 to form *i*SO595 are found in **Supplementary Table 1** and modified reactions are found in **Supplementary Table 2**.

ReFill Algorithm

Following an extensive search of literature and the KEGG (Kanehisa and Goto, 2000), TransportDB (Elbourne et al., 2017) and Metabolights (Haug et al., 2013) databases, we found a large number of new or previously known but missing reaction, transporter, and metabolite annotations. Adding large amounts of data to an existing network might create new gaps and may give rise to new blocked reactions and orphan metabolites that in general reduce model quality and can convolute later curation efforts, quality control or assessment of model predictions. To add this knowledge to the network in a controlled approach, we developed the semi-automated recursive algorithm ReFill (Recursive Filler of metabolic gaps). The algorithm is based on the principle of using diverse information, such as enzyme and reaction annotations, and experimental data (such as metabolomics), to selectively increase the metabolic knowledge of an organism's existing curated genome-scale metabolic network. ReFill makes use of a repository of reactions, in this case KEGG reaction annotations for MED4 absent from the model, to construct all potential chains of reactions connecting two metabolites in the existing network. It systematically tests the potential of adding each new reaction and suggests adding it only if it can be a part of a chain in which all the metabolites are part of a path in the network (**Figure 2**). This prevents



Parameter	*i*JC568	*i*SO595
Genes	568	595
Metabolites	680	803
Reactions	799	995
Exchange	80	99

FIGURE 1 | iSO595 is an updated reconstruction of *Prochlorococcus marinus* MED4, featuring a complete Entner–Doudoroff pathway, rewired glycogen metabolism and increased coverage of the genome. The central panel is a simplified illustration covering the most relevant metabolic features. **(A)** To simulate the natural environmental constraints experienced by Prochlorococcus we use "Push-FBA": Light and bicarbonate uptake are given a fixed flux independent of growth rate. This contrasts standard FBA where light and bicarbonate would be "pulled" in by the demand needed to support maximal growth rate up to a given bound. **(B)** We rewired the glycogen metabolism in *i*SO595 to study the dynamic allocation the dynamic allocation of glycogen. **(C)** Additionally, we implemented dynamic light conditions and light absorption in COMETS to simulate the growth of *P. marinus* during the diel cycle. **(D)** *i*SO595 has increased coverage both in terms of genes, reactions and metabolites compared to its ancestor *i*JC568.

the creation of new orphan metabolites and potential blocked reactions. The algorithm starts by selecting a reaction from the repository. It then inspects each metabolite in the reaction for presence in the existing network. In case a metabolite is not present, the set of available reactions is scanned for other reactions using this metabolite as a substrate or product. If such a reaction is found, it is added to the chain of potential reactions. The algorithm then iteratively expands the chain until either the repository is exhausted or all the metabolites in the most recent reaction added are present in the network. After all the possible chains of new reactions are expanded, the algorithm examines the connectivity of all the metabolites in each chain (see example in **Figure 2B**). Following the manual addition of transporters found through TransportDB (Elbourne et al., 2017) and Metabolights (Haug et al., 2013) (Study MTBLS567), using the ReFill algorithm, we updated reactions that belong to several different pathways, including metabolism of cofactors and vitamins, carbohydrate metabolism, amino acid metabolism and nucleotide metabolism. A complete list of added reactions can be found in **Supplementary Table 1**. ReFill was coded in python

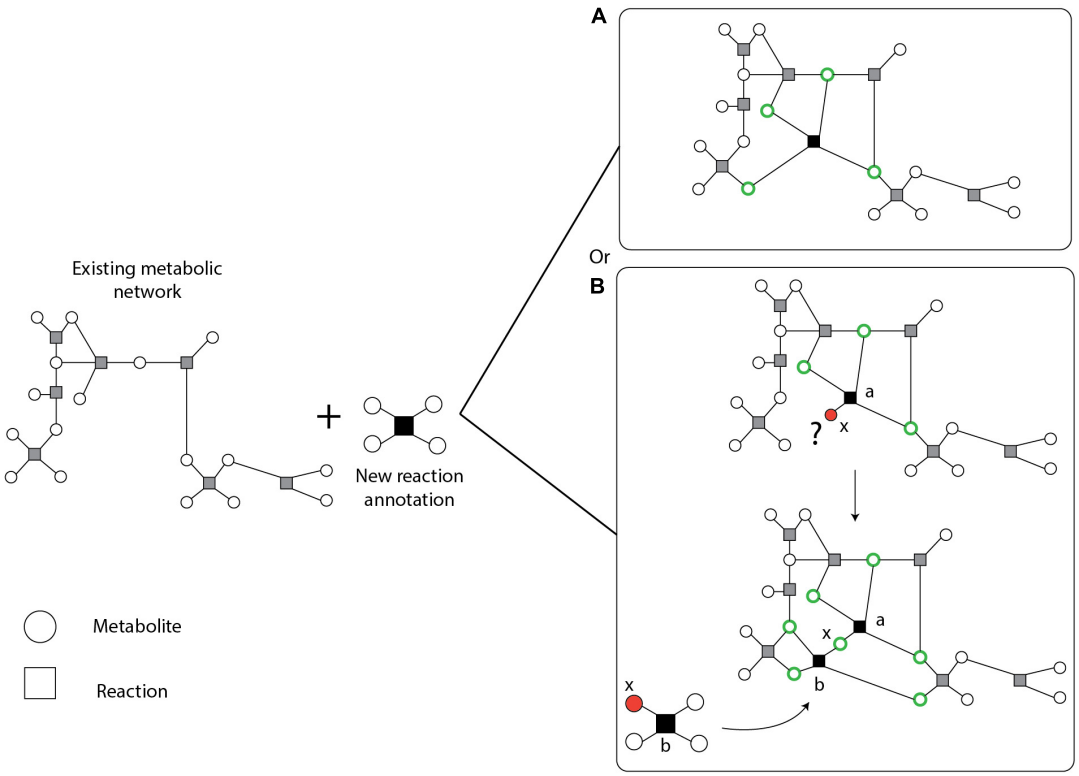


FIGURE 2 | Schematic description of the ReFill algorithm. New reaction annotations are added only if all the metabolites are connected to the existing network. A connected metabolite is denoted in a green circle. An orphan (unconnected metabolite) is denoted in a red circle. **(A)** A simple case in which all the metabolites comply with the selection rule. **(B)** In case one or more metabolites are not connected to the existing network, other new reactions may be added to complete the missing connections. For example, consider reaction *a* to be composed of two substrates and two products, one of which, *x* (a metabolite that did not exist in the initial network), is not connected to the existing network and is currently an orphan. After the expansion step, the algorithm identified one reaction, *b*, in which *x* is used as a substrate. All other metabolites in *b* exist in the network thus creating a path from reaction a through reaction *b* to the network.

3.7 and generates MATLAB-compatible files formatted to be used with the COBRA Toolbox (Heirendt et al., 2019), including a list of suggested reactions to add and their gene-reaction rules. Other outputs include the added reaction chains and possible metabolic circuits that can be formed by these additions.

Parameter Sampling

To study the effects of combinations of key nutrients on glycogen production and exudation in the *i*SO595 model we focused on four parameters representing the uptake fluxes of light, bicarbonate, ammonium, and phosphate. Light and inorganic carbon (bicarbonate) are the substrates for photosynthesis, whereas nitrogen and phosphorus limit the growth of *Prochlorococcus* in large regions of the world ocean (Davey et al., 2008; Moore et al., 2013; Saito et al., 2014), and nutrient limitation is likely to influence the exudation of fixed carbon (Dubinsky and Berman-Frank, 2001). We sampled 10,000 different environmental conditions by drawing random values from uniform distributions of these four parameters. The range of each parameter was based on physiologically relevant ranges we extracted from the literature and on the requirement that each range covers important phase transitions, such as nutrient and light limitations (**Supplementary Table 3**). Light flux was converted from micromole quanta m^{-2} sec^{-1} to mmol $gDW^{-1}h^{-1}$ similarly to Nogales et al. (2012) using 8% photosynthesis efficiency rate (Zinser et al., 2009). All uptake flux parameters are described in FBA-compatible units (mmol $gDW^{-1}h^{-1}$), while corresponding values in biogeochemistry relevant units are illustrated in **Supplementary Table 3**. As MED4 is a photoautotroph, it is exposed to a constant stream of light during daylight hours. The bacterium is then forced, or 'pushed,' to fix carbon even when there is not enough of other elements, such as nitrogen or phosphate, to combine the fixed carbon into biomass. To capture this phenomenon *in silico* we developed a 'push'-FBA framework where we fixed both upper and lower uptake rates of light and bicarbonate (**Figure 1A**). For the other sampled nutrients, ammonium and phosphate, we defined standard FBA bounds where the maximal uptake rate was set to the sampled value and the lower bound was set to zero. Note that we only considered uptake of sulfur in the form of sulfate (not hydrogen sulfide), and no upper limit was set for the uptake of sulfate because of its abundance in seawater. The maximum rate of RuBisCO (R00024) was fixed to 4.7 mmol gDW^{-1} h^{-1}, as previously reported (Casey et al., 2016). Before sampling we blocked a set of artificial exchange reactions that were added in the previous version of the model, most likely

to allow export of dead-end metabolites that would otherwise limit flux feasibility (**Supplementary Table 4**). Subsequently, we removed all unconditionally blocked reactions in the model to speed up computations. For each random sample, we first tested the model for feasibility using FBA (Varma and Palsson, 1994). If the solver returned a solution that was feasible and optimal, we further calculated optimal fluxes with parsimonious FBA (Lewis et al., 2010), and determined the range of possible fluxes at optimum with Flux Variability Analysis (FVA) (Gudmundsson and Thiele, 2010). Exchange fluxes from FBA, parsimonious FBA and FVA were recorded and used in subsequent analyses. All environmental sampling and calculations were performed using CobraPy (Ebrahim et al., 2013) and GUROBI 8.1.1 (Gurobi Optimization, Inc., Houston, TX, United States).

Statistical Analysis of the Sampled Spaces

We sampled 10,000 different environmental conditions based on the flux ranges described above, and analyzed the results of FBA optimization, with the goal of characterizing the distribution of, and correlation between, specific exchange (import/export) fluxes. To that end, we calculated Pearson correlations between exchange reaction fluxes in the sampling data using the python (version 3.7) Pandas package version 1.0.3 (McKinney, 2010). While negative values are normally used to define uptake in FBA, we converted them to positive values for the uptake of light, bicarbonate, phosphate, ammonium, and sulfate when calculating correlations to ease interpretation of the results. We also performed hierarchical clustering using the Nearest Point Algorithm in SciPy (Virtanen et al., 2020) to sort the order of the compounds in the correlation matrix.

We performed dimensionality reduction on normalized exchange reaction fluxes using the T-distributed Stochastic Neighbor Embedding (t-SNE) method (van der Maaten and Hinton, 2008) in Scikit-learn (Pedregosa et al., 2011) with perplexity of 50 and 3,000 iterations. The reaction fluxes were normalized to [−1,1] by dividing by the maximum absolute flux value of each reaction to ensure a consistent influence on the t-SNE results from the different exchange reactions. We considered other normalization schemes, in particular standardization, but found that it was preferable not to center the data to easily discriminate uptake and exudation without further modifications in subsequent data visualization. Finally, the t-SNE transformed data was clustered using HDBSCAN (McInnes et al., 2017) with a minimum cluster size of 200. Transport of inorganic ions, water, and protons were not considered when calculating correlations, dimensionality reduction or clustering. We also discarded transport reactions with no absolute flux value above 10^{-3} mmol gDW^{-1} h^{-1} in any of the environmental samples.

Dynamic Modeling of Light Absorption During the Diel Cycle in COMETS

Cyanobacteria follow a diel cycle. To capture this dynamic behavior, we extended the Computation Of Microbial Ecosystems in Time and Space (COMETS) platform (Harcombe et al., 2014; Dukovski et al., 2020), and developed a module for diurnal-cycle simulations allowing oscillations of light intensity and light absorption. Attenuation of light through each grid cell was modeled using the Beer–Lambert law, as described previously (Yang, 2011; Gomez et al., 2014):

$$I\,(t,\ z) = I_0(t)e^{-(a_w + a_{dw}X(t))z} \tag{1}$$

Here, $I(t,\ z)$ is the light irradiance given in mmol photons m^{-2} s^{-1}, t is the time, z is the depth (from the top of the grid cell), a_{dw} is the cell- and wavelength-specific absorption coefficient given in m^2 gDW^{-1}, a_{w} the absorption coefficient of pure water given in m^{-1}, $X(t)$ the biomass concentration in gDW m^{-3}, and $I_0\,(t)$ the time-dependent incident light irradiance at the top of the grid cell. In the current version, we simplified the process by assuming that the light irradiance is either monochromatic or a sum of the total light bandwidth, and the absorption coefficient should match the wavelength(s) of the light source. The total light attenuation (ΔI) through a grid cell of thickness Δz is then

$$\Delta I\,(t) = I\,(t,\ 0) - I\,(t,\ \Delta z) = I_0\left(1 - e^{-(a_w + a_{dw}X(t))\Delta z}\right) \tag{2}$$

The light absorbed by the cells is a fraction of the total light attenuation, i.e.,

$$I_{\mathrm{abs}}\,(t) = \frac{\Delta I(t)\cdot a_{\mathrm{dw}}X(t)}{a_{\mathrm{w}} + a_{\mathrm{dw}}X(t)}. \tag{3}$$

The total number of photons absorbed per dry cell weight [$\Phi(t)$] in mmol photons gDW^{-1} s^{-1} by the cells within a grid cell of thickness Δz, volume V, and surface area A is then

$$\Phi\,(\mathrm{t}) = \frac{I_{\mathrm{abs}}\,(t)\cdot A}{X(t)\cdot V} = \frac{I_0\,(t)}{\Delta z}\frac{a_{\mathrm{dw}}}{a_{\mathrm{w}} + a_{\mathrm{dw}}X(t)}\left(1 - e^{-(a_{\mathrm{w}} + a_{\mathrm{dw}}X(t))\Delta z}\right). \tag{4}$$

For all COMETS simulations presented here we have used monochromatic light at 680 nm with a calculated biomass-specific absorption coefficient a_{dw} as previously described (Morel and Bricaud, 1981; Bricaud et al., 2004). Briefly, the biomass-specific absorption is the weighted sum of the absorption coefficients of the light-absorbing pigments divinyl-chlorophyll A and B, since none of the other pigments in *Prochlorococcus* absorb light at 680 nm. Additionally, to account for the discrete distribution of chlorophyll into separate cells, the absorption coefficient is scaled by the packaging factor. All coefficients used to calculate light attenuation and absorption are provided in **Table 1**.

The changing light conditions throughout a diel cycle was modeled as

$$I_0\,(t) = A\max(\sin\,(\omega t)\,,\ 0), \tag{5}$$

where the angular frequency is $\omega = \frac{2\pi}{T}$.

Following the development of the diel cycle simulation capability in COMETS we set out to dynamically simulate the growth of MED4. Since the nutrient uptake follows Michaelis–Menten kinetics, we estimated the kinetic parameters V_{max} and K_{m} using a heuristic approach from experimental data

TABLE 1 | Coefficients and values used to calculate light absorption in COMETS.

Symbol	Description	Value	Unit	Reference
λ	Wavelength	680	nm	
a_w	Absorption coefficient of water at 680 nm	0.465	m^{-1}	Pope and Fry, 1997
a_{dw}	Biomass-specific absorption coefficient	0.285	m^2 gDW^{-1}	
$a_{dvchl-A}$	Absorption coefficient of divinyl-chlorophyll A at 680 nm	0.0184	m^2 (mg dvchl-A)$^{-1}$	Bricaud et al., 2004
$a_{dvchl-B}$	Absorption coefficient of divinyl-chlorophyll B at 680 nm	0.0018	m^2 (mg dvchl-B)$^{-1}$	Bricaud et al., 2004
$c_{dvchl-A}$	Amount of divinyl-chlorophyll A	0.0163	g dvchl-A gDW^{-1}	Casey et al., 2016
$c_{dvchl-B}$	Amount of divinyl-chlorophyll B	0.0013	g dvchl-B gDW^{-1}	Casey et al., 2016
d	Average diameter of MED4	0.6	μm	
n'	Imaginary part of the refractive index at 675 nm	0.01377		Stramski et al., 2001
Q^*	Packaging effect at 680 nm	0.945		

(Grossowicz et al., 2017), first by finding the range of possible parameter combinations corresponding to the gross growth rate of 0.5 d^{-1} (**Supplementary Figure 1A**), and secondly by comparing predicted growth and ammonium depletion with the experimental time-series cultivation data (**Supplementary Figures 1B,C**). The estimated parameters were used in the remaining dynamic FBA simulations in COMETS. Finally, to simulate the dynamic storage and consumption of glycogen we applied a multiple objective approach consisting of the following four steps: (1) Maximization of the flux through the non-growth associated maintenance reaction. Note that, this reaction has an upper bound of 1 mmol gDW^{-1} h^{-1} (Casey et al., 2016). In contrast to standard practice, where one uses a lower bound for the non-growth associated maintenance reaction, this method provides a more realistic scenario where the organism continues to consume resources trying to keep up cellular maintenance even at zero growth; (2) Maximization of growth; (3) Maximization of glycogen production (storage); and (4) Parsimonious objective which minimizes the sum of absolute fluxes. To simulate nitrogen-abundant and nitrogen-poor growth conditions, we used the PRO99 medium with standard (800 μMol) and reduced (100 μMol) ammonium concentration, as previously described (Grossowicz et al., 2017). Light availability was modeled as described in Equation 5, with an amplitude of 40 μmol Q m^{-2} s^{-1} and a period of 24 h. We also incorporated a death rate of 0.1 d^{-1}, similar to previous modeling efforts on *Prochlorococcus* (Grossowicz et al., 2017). All parameter values used in the COMETS simulations are given in **Supplementary Table 5**. All dynamic growth simulations were performed using COMETS v.2.7.4 with the Gurobi 8.1.1 solver, invoked using the associated MATLAB toolbox[1].

Simulating Growth of Knockout Mutants

Simulations of the knockout mutants where performed by constraining the flux to zero for the reactions catalyzed by the enzymes encoded by *glgC* (PMM0769) and *gnd* (PMM0770), respectively. For *glgC*, the reaction is glucose-1-phosphate adenylyltransferase (R00948) and for *gnd* the two reactions are $NADP^+$ and NAD^+ associated 6-phosphogluconate dehydrogenases (R01528 and R10221). We then used dynamic FBA in COMETS with PRO99 medium (Moore et al., 2007) with limited ammonium and diel light conditions to simulate growth over 7 days. The growth curves where qualitatively compared with experimental data from Shinde et al. (2020).

[1]https://github.com/segrelab/comets-toolbox

RESULTS AND DISCUSSION

Model Curation and Update

Prochlorococcus fixes carbon through photosynthesis during daytime. Fixed carbon that is neither used for cell growth nor stored in the form of glycogen is exuded. Here, we set out to study dynamic changes in the carbon allocation and storage mechanisms in MED4 using a genome-scale metabolic modeling approach. To that end, we first re-curated and updated the available *i*JC586 model (Casey et al., 2016), as described in detail in the "Materials and Methods" section. The update involved the development of a new semi-automatic algorithm (ReFill), which can be broadly applied to other reconstructions (see section Materials and Methods). Concurrently, we introduced a revised mechanism for carbon storage, effectively treating glycogen as an independent component of biomass. This dynamic implementation of glycogen storage, introduced here in dFBA, makes it possible for glycogen to be accumulated and depleted at variable rates (**Figure 1**), aligning with the overflow metabolism hypothesis (Szul et al., 2019; de Groot et al., 2020). Other key modifications induced by the ReFill algorithm and subsequent manual curation (see section Materials and Methods) include the completion of the Entner–Doudoroff (ED) pathway, recently discovered in cyanobacteria (Chen et al., 2016) and proposed as the primary *Prochlorococcus* glucose metabolism pathway under mixotrophic conditions (Biller et al., 2018; Muñoz-Marín et al., 2020). Additional revisions focused on the exudation of fixed carbon products from the cell and included various transports such as pyruvate, fumarate, citrate, ethanol, various nucleotides and hydrogen peroxide as well as metabolites found in both the endo- and exo- metabolome of *Prochlorococcus* (Metabolights study MTBLS567). The end product of our revision, reconstruction *i*SO595, has 595 genes, 802 metabolites and 994 reactions, i.e., 27 genes, 123 metabolites and 196 reactions more than the previous version, *i*JC568 (**Figure 1D**).

Carbon Fixation and Storage Are Affected by Nutrient Uptake Rate

Prochlorococcus thrive in oligotrophic environments (Johnson et al., 2006), where, in surface waters, its growth and carbon fixation rates are usually limited by the abundance of nitrogen, phosphate or iron (Krumhardt et al., 2013; Saito et al., 2014; Szul et al., 2019). Deeper in the water column *Prochlorococcus* growth becomes limited by light (Vaulot et al., 1995). We set out to explore the combined effect of different levels of light and nutrients on carbon fixation, storage and exudation. Similarly to Phenotypic Phase Plane analysis (Edwards et al., 2002), we sought a global perspective of metabolism in this multi-parameter spaces while explicitly taking into account the fact that the inflow of light and bicarbonate may not be easily controllable by the cell, and that *Prochlorococcus* may need to deal with excess amounts of fixed carbon. Thus, in contrast to normal FBA where the uptake of metabolites is constrained by an upper bound, we introduced a 'push- FBA' approach (**Figure 1A**), in which the influx of bicarbonate and light have a fixed imposed value (see section "Materials and Methods" and **Supplementary Table 3** for specific values used). This approach attempts to mimic implications of photosynthesis, in which light is the driving force. Once photons are absorbed by the chlorophyll in the photosynthetic reaction centers, most of the energy must be used to produce ATP and reducing power, otherwise it is dissipated in ways that may cause cell damage (Long et al., 1994). We note that this modeling approach over-simplifies the complex process of photosynthesis; for example, we do not account for the dynamics of photoprotective pigments, which allow some of the incident photons to be dissipated as heat. Indeed, the ratio of the photoprotective pigment zeaxanthin to divinyl *chlorophyll a* increases under nitrogen starvation, suggesting that, under these conditions, some of the photon flux may be diverted from the reaction centers (Steglich et al., 2001; Roth-rosenberg et al., 2019). Nevertheless, *Prochlorococcus* undergo photoinhibition at high light intensities (Moore et al., 1995; Mella-Flores et al., 2012), despite the presence of photoprotective pigments and other protection mechanisms such as cyclic electron flow [which is represented in the model (Casey et al., 2016)]. Thus, these mechanisms do not allow the cell to fully control the flux of photons through the photosystem and the resulting fluxes in ATP and reducing power, in a manner that is reflected in the push-FBA approach. This subtle difference in applied constraints has major effects on model predictions. While flux rearrangement is usually viewed as a consequence of environmental nutrient limitations, the results of this analysis show that a substantial rewiring of fluxes is caused by this imposed excess of fixed carbon as well.

To understand how different combinations of environmental parameters (availability of nitrogen, phosphate, light and bicarbonate) affect the way *Prochlorococcus* can manage its carbon budget, we implemented FBA under 10,000 randomly sampled growth environments. Overall, this sampling analysis demonstrated that the exudation of organic acids, amino-acids, and nucleobases/nucleosides, as well as the extent of glycogen storage, are strongly modulated by environmental factors (**Figure 3**). To observe the full range of possible optimal solutions per sample, we implemented and compared different flux balance analysis methods, including flux variability analysis (FVA) and parsimonious FBA (pFBA). These two methods provide complementary insight: FVA estimates the range of possible values for the flux of each reaction at the optimum, providing insight into the structure of the phenotypic space at maximal growth rate. In contrast, pFBA, by minimizing the sum of fluxes at optimality, generates flux predictions less likely to involve unrealistic loops, and thus potentially provides predictions closer to experimental values (Lewis et al., 2010). Together, these two FBA methods help analyze the solutions of our high-dimensionality dataset.

Our predictions simulate the metabolic effects and variability in glycogen production modulated by environmental constraints (**Figure 3**). Glycogen production was observed only above light levels of 50 mmol gDW^{-1} h^{-1} (corresponding to 7.5 micromole quanta m^{-2} sec^{-1}), and decreased as ammonium and phosphate concentrations increase. These observations do not contradict previous evidence showing increased glycogen accumulation in faster growing cyanobacteria (Zavřel et al., 2019), rather they align with previous studies finding that glycogen storage is enhanced in nutrient-limiting conditions (Monshupanee and Incharoensakdi, 2014; Szul et al., 2019). Interestingly, FVA consistently predicted the glycogen production range minimal value to be zero across all samples. This implies that glycogen storage is possible, but not necessary to achieve optimal growth in the feasible solution space. This was also the case in the more stringent pFBA analysis, indicating that while metabolism may be a strong modulator of glycogen metabolism, more types of regulation, not accounted for in FBA, are involved. One example of such regulation may be allosteric regulation of ADP-glucose pyrophosphorylase by 3-phosphoglycerate (Iglesias et al., 1991), possibly in combination with redox regulation (Díaz-Troya et al., 2014). Specific regulation aimed at tuning up glycogen storage may also occur at the transcriptional level, e.g., by multiple transcription factors previously suggested to be involved in the regulation of glycogen metabolism in fluctuating environments (Luan et al., 2019).

The range of possible rates of glycogen production (through FVA) displays a bell-shaped bicarbonate-dependent distribution, indicating low storage of glycogen (zero flux) under both low and high uptake rates of bicarbonate. When bicarbonate uptake rates are low, all available carbon is diverted into growth. The reduced glycogen storage at high bicarbonate uptake, when RuBisCO is saturated, seems to be caused by the increased ATP demand associated with the conversion of bicarbonate to exudation-products, since the onset and rate of change of this trade-off is modulated by the ATP availability, as demonstrated by phenotypic phase planes analysis (**Supplementary Figure 2**). This agrees with recent work suggesting that *Prochlorococcus* use available ATP to drive pathways to saturation by shifting reaction directions toward favoring dephosphorylation of ATP to ADP, disrupting the cellular ATP/ADP ratio and increasing the metabolic rate of the cell by pushing forward ATP consuming reactions, until it is restored. Together with organic carbon exudation this strategy allows for growth in lower nutrient concentrations (Braakman, 2019).

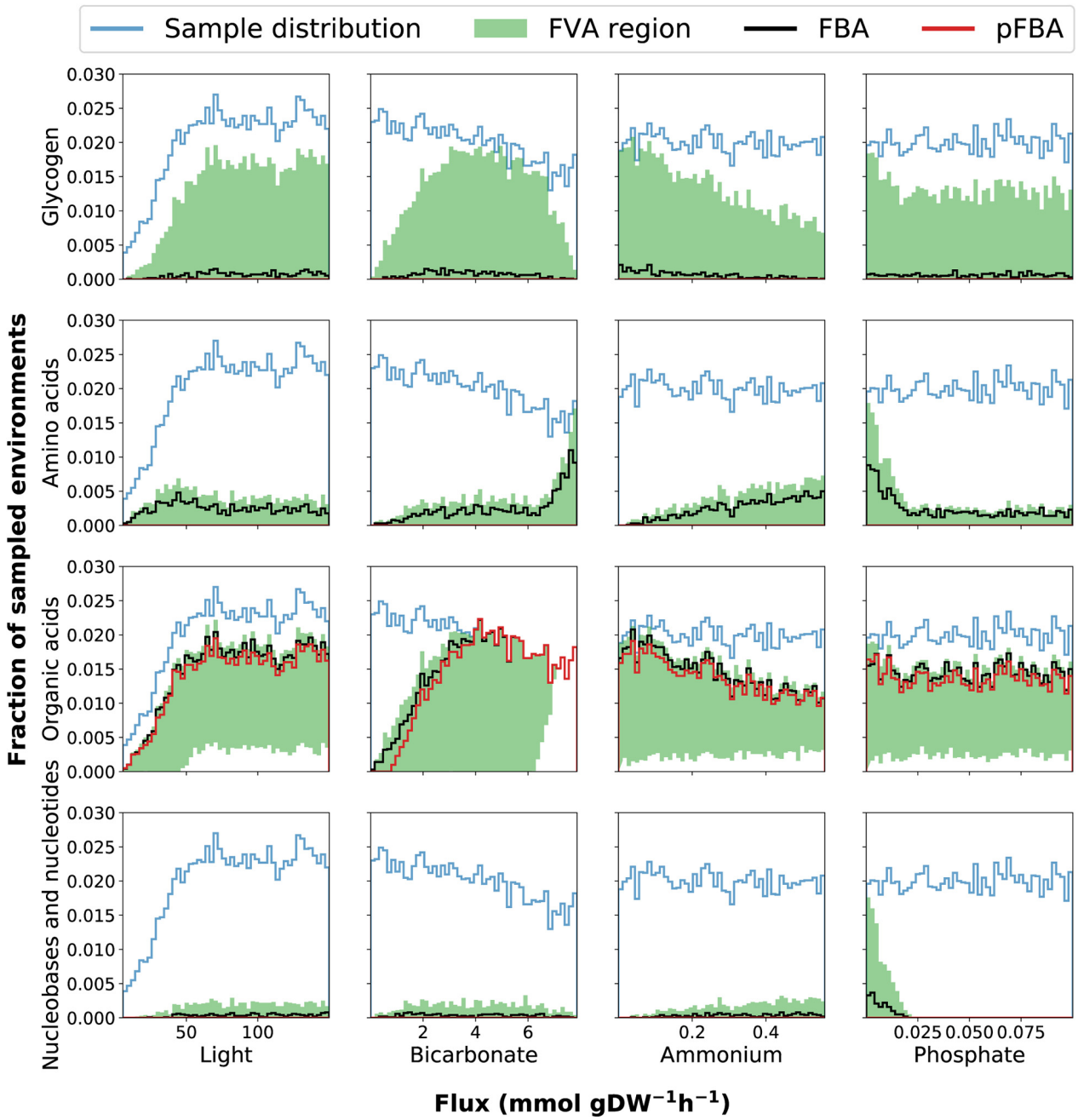


FIGURE 3 | Histograms of environmental sampling results provide insight into how the fixed uptake rates of light and bicarbonate and the upper bounds on ammonium and phosphate affect exudation of organic acids, nucleobases and nucleotides, amino acids as well as glycogen storage. The *y*-axis represents the fraction of all sampled environments yielding feasible models. Although we initially sample each parameter uniformly the final sample distribution is not uniform because some combinations of parameters represent infeasible phenotypes (no solution can satisfy the constraints). The final sample distribution for each parameter is therefore shown in the figure panels as a blue line. The black and red lines represent the histograms of samples where exudation is predicted by FBA and pFBA, respectively. The shaded green region represents the span between histograms of samples as predicted by FVA: the lower and upper bounds represent the number of samples where exudation is predicted using the minimum and maximum value of FVA, respectively. This FVA region covers the range of possible phenotypes. The lower bound of the FVA region displays the number of samples where a certain outcome is obligatory to maximize growth, while the upper bound of the FVA region displays the number of samples where the outcome is possible without reducing growth.

We next sought to explore the effect of combinations of key nutrients on storage and exudation patterns in our sampling spaces. To that end, we visualized the data using t-SNE clustering (**Figure 4A**). To explore the strongest trends, we chose to employ a high stringency approach and use only our set of pFBA results in this context. Due to the nature of pFBA,

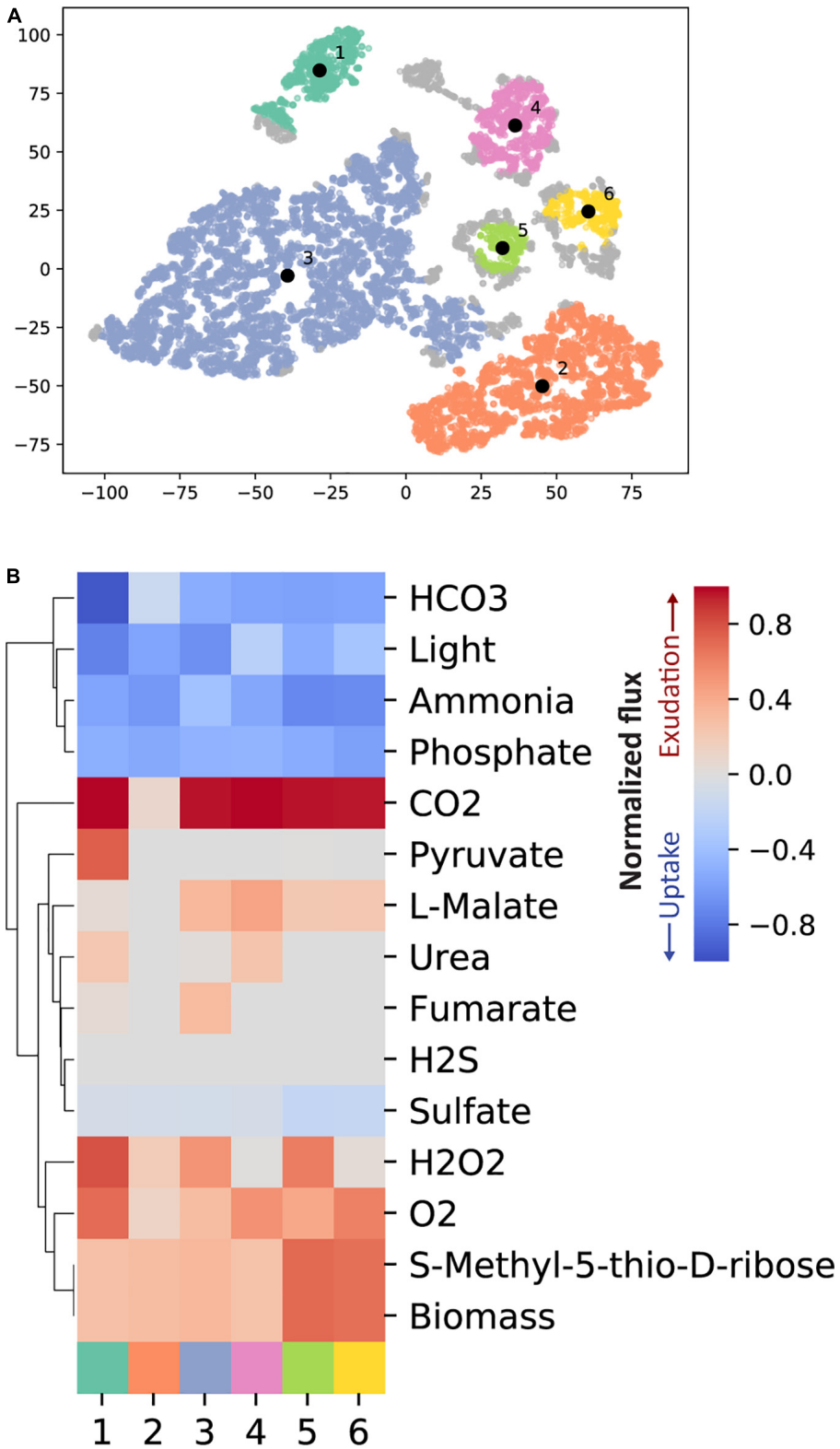


FIGURE 4 | T-SNE clustering identifies typical phenotypes from the pFBA results from the random samples. **(A)** The random samples are reduced into two dimensions with t-SNE. We have subsequently used HDBSCAN to cluster the data. HDBSCAN identified six disjoint clusters which represent different phenotypes. **(B)** For each of the six clusters the mean uptake or exudation across all samples within the respective cluster is shown. Only exchange reactions with an absolute flux above 1e-3 mmol gDW^{-1} h^{-1} in any of the random samples are included.

any exudation observed in this analysis could not be easily removed without imposing a cost on growth. We observed 6 typical phenotypes (clusters) rising out of the sampling spaces (**Figure 4**). These 6 phenotypes are characterized by subtle differences in combinations of environmental parameters, yielding significantly different exudation patterns. Generally, we observed the highest biomass value in phenotype 5, and the lowest in phenotype 4. All key nutrient uptake rates were highly variable (ranging from 33 to 44% variability). Phenotype 1 is characterized by high light, bicarbonate, a maximum RuBisCO flux (indicating maximal photosynthesis rate) but low nitrogen uptake. Additionally, we observed high exudation of pyruvate coming from the pentose phosphate and Entner–Doudoroff pathways. Both are alternative routes coming out of carbon fixation (Waldbauer et al., 2012; Chen et al., 2016). Together with a low biomass value, this phenotype might indicate a scenario of exudation due to overflow metabolism.

The two largest clusters (numbers 2 and 3, **Supplementary Figure 3**), tie together high and low light, carbon and nitrogen uptake rates, and different exudation patterns. Interestingly, phenotype 3 (high light) showed exudation of fumarate and malate while phenotype 2 (low light) did not. Recent work suggested that, in high light conditions, fumarate is generated through oxaloacetate and malate creating a broken acyclic form of the TCA cycle, while in the dark, fluxes are diverted into forming the cyclic form of it. This low light form of the TCA cycle is then active and works toward energy generation (Xiong et al., 2017). Similarly, we observed two forms of the TCA cycle in the high and low light phenotypes (2 and 3, respectively) with a difference in the direction of one reaction (KEGG R00342, **Supplementary Figure 3**). Phenotype 2, describing low-light conditions, showed the L-Malate/oxaloacetate balance to shift in favor of oxaloacetate, completing the route toward 2-Oxoglutarate, a key metabolite known to act as a starvation signal and modulator of the C/N balance in cyanobacteria (Domínguez-Martín et al., 2018; Zhang et al., 2018), and subsequently into energy generation. On the other hand, Phenotype 3, describing high light conditions, showed the L-Malate/oxaloacetate balance to shift in favor of L-Malate and away from the formation of 2-oxoglutarate. In both phenotypes fumarate is converted to L-Malate. While in Phenotype 2 it is fed into a semi-cyclic form of the TCA cycle, fumarate is partly exuded and partly converted to L-malate in phenotype 3, in agreement with overflow metabolism.

We observed a similar TCA cycle flux distribution in phenotype 4 as in phenotype 3, leading to high exudation of L-Malate. Interestingly, Phenotype 1 and 4 are comparable in all key nutrients except light (High in phenotype 1 and low in phenotype 4). As a result of an in-depth flux distribution analysis, we observed a reaction direction change in UDP-glucose:NAD+ 6-oxidoreductase [R00286, EC 1.1.1.22, PMM1261] between the two phenotypes. In phenotype 4 this reaction shifted toward the creation of UDP-glucose, a precursor for the production of glycogen (due to the high stringency of this analysis we did not observe the direct formation of glycogen). In phenotype 1, this reaction favored the formation of UDP-glucuronate which in turn was diverted into the formation of amino sugars. These phenotypes may correlate to the 12:00 (phenotype 1) and 16:00 (phenotype 4) scenarios described in Szul et al. (2019). Finally, Phenotypes 5 and 6 may represent a high-light nutrient-rich environment resulting in a high biomass value.

Nutrient Uptake Rates Modulate Exudation of Organic Compounds

The use of genome-scale metabolic models captures a comprehensive picture of the metabolic processes taking place in the cell, including those that lead to metabolite exudation. From the random sampling of environmental conditions, we identified conditions in which organic acids must be exuded. This was noticeable by a non-zero lower bound of the FVA region (**Figure 3**). Interestingly, organic acids were more likely to be exuded when the growth became limited by phosphate or nitrogen. Since *Prochlorococcus* is known to thrive in oligotrophic ocean gyres where nitrogen or phosphate is limited (Partensky et al., 1999; Flombaum et al., 2013), this represents a likely natural phenotype, and as such, supports previous findings (Bertilsson et al., 2005; Szul et al., 2019). Costly metabolites, essential for cell survival and growth, such as amino acids, nucleobases and nucleotides, tend to be exuded in nitrogen and carbon rich conditions and might be a result of overflow metabolism (Cano et al., 2018; Pacheco et al., 2018). To explore this phenomenon in further detail, we looked into exudation patterns of specific metabolites as a function of key nutrient limitations (**Figure 5**). Of the environmental factors, the uptake of nitrogen (ammonium) is a decisive factor differentiating between exudation of organic acids or amino acids. While it is positively correlated with the exudation of nitrogen-rich compounds such as amino acids, it is negatively correlated with exudation of organic acids and glycogen. Additionally, glycogen formation is positively correlated with the exudation of malate, citrate, fumarate, and succinate, which are most of the TCA cycle constituents. This is in line with previous findings suggesting the re-direction of carbon metabolism toward the formation of macromolecules (including glycogen) in nitrogen limiting conditions (Forchhammer and Selim, 2019; Szul et al., 2019). Thus, our reconstruction captured known possible aspects of the carbon/nitrogen balance in *Prochlorococcus*.

Finally, we observed a general pattern of strong positive correlations between amino acids, nucleobases, nucleosides, as well as a range of other compounds. In an interesting deviation from this general pattern, L-aspartate showed a decreased correlation with other exudates. L-aspartate, together with its role in protein nucleic acid biosynthesis, can serve as a precursor for nitrogen storage metabolites such as polyamines (Szul et al., 2019). Indeed, we observed a slightly stronger correlation between L-aspartate and the uptake of nitrogen compared to other amino acids. Finally, In contrast to other amino acids, L-aspartate is negatively correlated with light uptake and hydrogen peroxide exudation. Hydrogen peroxide is produced from L-aspartate and oxygen by L-aspartate oxidase [R00481, EC 1.4.3.16, PMM0100]. L-amino acid oxidases have been previously described in cyanobacteria and have been related to the use of amino acids as carbon sources (Campillo-Brocal et al., 2015). The production of hydrogen peroxide is also strongly correlated with

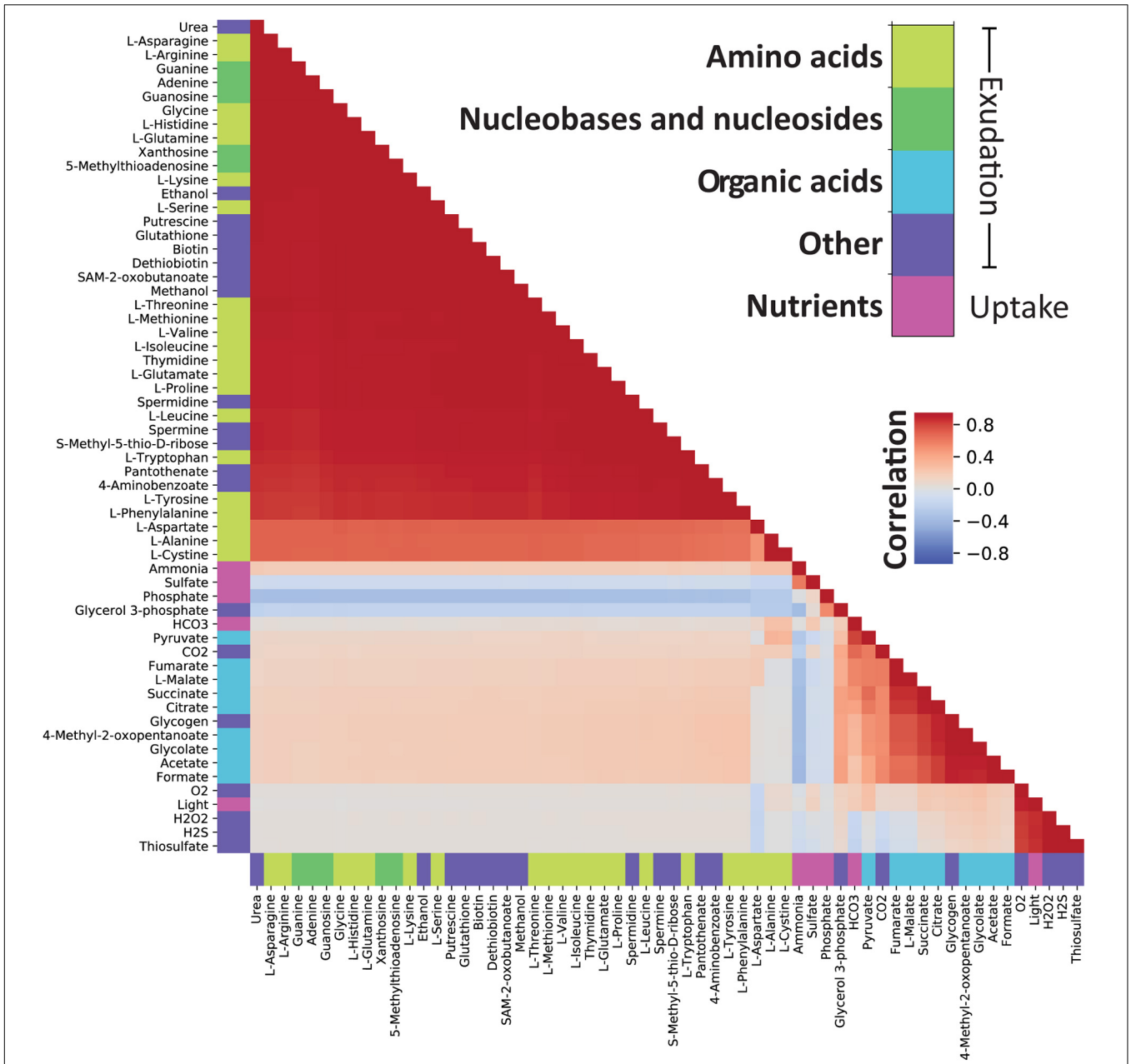


FIGURE 5 | Correlations between maximal FVA values demonstrate which compounds can be secreted in similar environmental conditions. There are two clearly correlated groups of compounds: The first group comprises nucleobases, nucleotides, and amino acids while the second contains organic acids, glycogen and glycerol 3-phosphate. From the correlation of each of these groups with the uptake bounds on ammonium and phosphate we observe that these factors determine which of the groups can be secreted. Note that the environmental constraints have been converted to positive values prior to calculating the correlation.

light, a result consistent with the expectation that reactive oxygen species are created during photosynthesis.

Dynamic Allocation of Carbon Storage

Nutrient and light limitations are well-known modulators of carbon storage in *Prochlorococcus* (Zinser et al., 2009; Szul et al., 2019). Recent work has suggested the storage of carbon to be one of the major metabolic tasks during the day-night cycle (Cano et al., 2018; Szul et al., 2019; Shinde et al., 2020). To explore time-modulated trade-offs and trends related to carbon storage, we performed *in silico* dynamic FBA diel-cycle simulations using the Computational of Microbial Ecosystems in Time and Space (COMETS) platform (Harcombe et al., 2014; Dukovski et al., 2020). COMETS is a population-based dynamic FBA implementation that can simulate growth of millions of cells, but it is important to note that the framework assumes continuous growth on a mesoscopic scale and does therefore not explicitly account for individual cells nor regulated cell

cycle events such as cell division. COMETS relies on uptake flux kinetic information such as K_m and V_{max} to simulate the spatial growth and exudation patterns of microbes in a simulated discretized time course. To improve the accuracy and biological relevance of our simulations we used kinetic constants either obtained from experimental measurements reported in the literature (Krumhardt et al., 2013; Hopkinson et al., 2014) (**Supplementary Table 3**) or from fitting model simulations to measured growth and depletion of ammonium rates (Grossowicz et al., 2017). We found K_m and V_{max} values of 0.39 mM and 0.9 mmol gDW^{-1} h^{-1} for the uptake of ammonium to best fit the experimental data (Grossowicz et al., 2017) (**Supplementary Figure 1**). Surprisingly, the estimated K_m value is 3 orders of magnitude larger than previous estimates (Marañón et al., 2013). This deviation might occur due to several reasons. First, our estimates are based on the assumption that growth is indeed limited by the availability of ammonium and that *Prochlorococcus* operates at a metabolic state close to optimal growth. Other limiting factors or non-optimal growth may lead to incorrect estimates. Nevertheless, it is challenging to fit K_m values accurately from batch cultivation data, as this parameter only becomes dominant in the short time-period immediately prior to nutrient depletion. Furthermore, the accuracy of the fitted K_m value can suffer from the rather high uncertainty in the measured ammonium concentrations, although not more than 2 orders of magnitude (**Supplementary Figure 1**). Finally, we raise the possibility that *Prochlorococcus* may possess several ammonium transporters with different affinity as previously observed in marine eukaryotic phytoplankton (McDonald et al., 2010) and cyanobacteria (Kashyap and Singh, 1985). To account for this uncertainty we assessed the sensitivity of our dFBA simulations to variation in the value of K_m, in combination with variation in the maximum uptake rate of ammonium (V_{max}), ammonium concentration and light intensity (**Supplementary Figure 4**). The parameters that dictate light absorption (**Table 1**) affect the number of available photons, so that by including a large span of light intensities in our sensitive analysis, we also cover their associated uncertainty. We find that ammonium concentration, kinetic coefficients for ammonium uptake and the availability of photons combined have a considerable impact on whether carbon is stored during daytime in our dynamic FBA simulations, underpinning the importance of accurate and context specific values for these parameters. This echoes the well-known modulation of carbon storage by nutrient and light limitations (Zinser et al., 2009; Szul et al., 2019). We note that, despite the potentially large impact of *Prochlorococcus* on marine nitrogen budgets, to the best of our knowledge there are currently no direct experimental measurements of the kinetics (K_m, V_{max}) of nitrogen uptake by *Prochlorococcus*.

Since the tight coupling between carbon and nitrogen metabolism in cyanobacteria is known to influence carbon allocation and storage (Zhang et al., 2018; Szul et al., 2019), it was chosen as a case study. As such, we focused in more detail on the dynamic changes in metabolism in nitrogen-abundant and nitrogen-poor media, as previously defined (Grossowicz et al., 2017). Specifically, we set out to explore glycogen production and consumption with COMETS in these conditions (**Figure 6**). We did not observe glycogen storage in nitrogen-abundant simulations, and therefore no growth nor cellular maintenance during nighttime. One explanation for this may arise from the limitations of the platform. First, the simulations performed in this work were performed in a modeling framework based on linear programming with ordered multi-objective optimization: (1) cellular maintenance; (2) growth; (3) glycogen storage. Thus, glycogen was only stored when there were excess energy and carbon available, which occurred when growth was nitrogen limited. Although some observer bias was introduced by assuming that *Prochlorococcus* is striving toward these cellular objectives, in this order, we found a reasonable conceptual alignment with previous work showing that bacterial metabolism balances a trade-off between maximal growth and the ability to adapt to changing conditions (Schuetz et al., 2012). However, we do note that one might obtain more nuanced results by taking into account suboptimal solutions (Segrè et al., 2002; Fischer and Sauer, 2005; Wintermute et al., 2013), and that real phenotypes may be in the continuum between the two extremes found here. Another limitation that might affect glycogen storage is the lack of regulatory mechanisms not usually accounted for in this version of dynamic FBA (Mahadevan et al., 2002). The addition of regulatory layers or more specifically tailored objective functions, such as global optimization over the entire diel cycle (Reimers et al., 2017), could lead to smaller but non-zero generation of glycogen also during nitrogen-rich conditions.

In agreement with previous work (Szul et al., 2019), under nitrogen-limiting conditions, glycogen accumulates throughout the day and is subsequently used to support respiration and growth during the night (**Figure 6A**). However, the predicted glycogen storage is simulated as not sufficient to support neither growth nor cellular maintenance throughout the night. This may contribute to the increased death rate during night time (Zinser et al., 2009; Ribalet et al., 2015). However, the rate of glycogen depletion is strongly affected by the associated kinetic parameters (**Supplementary Figure 5**), emphasizing the value of accurate kinetic coefficients for GlgP, the main contributing factor to glycogen catabolism in bacteria (Dauvillée et al., 2005; Alonso-Casajús et al., 2006; Fu and Xu, 2006), in future work. Furthermore, the rate of glycogen depletion might be modulated by transcriptional regulation. Previous work suggested that glycogen storages are not sustained beyond dawn, because the genes responsible for glycogen degradation are depleted during the first 5 h of darkness (Biller et al., 2018). Interestingly, the model predicts consumption of glycogen during dusk to increase growth when photosynthesis is declining (**Figure 6**), closely resembling observations in *Synechococcus,* in particular for the Δ*kaiC* mutant with a dysfunctional circadian clock (Diamond et al., 2015). The closer resemblance of the dysfunctional circadian clock phenotype might be a result from the limitations of the applied modeling framework that does not include regulatory mechanisms.

The switch from photosynthesis at daytime to glycogen consumption at nighttime is reflected in the metabolic shifts observed in key pathways (**Figure 6B**). Interestingly, we observed higher fluxes through the Calvin cycle in nitrogen-poor conditions. This difference may be caused by the increased

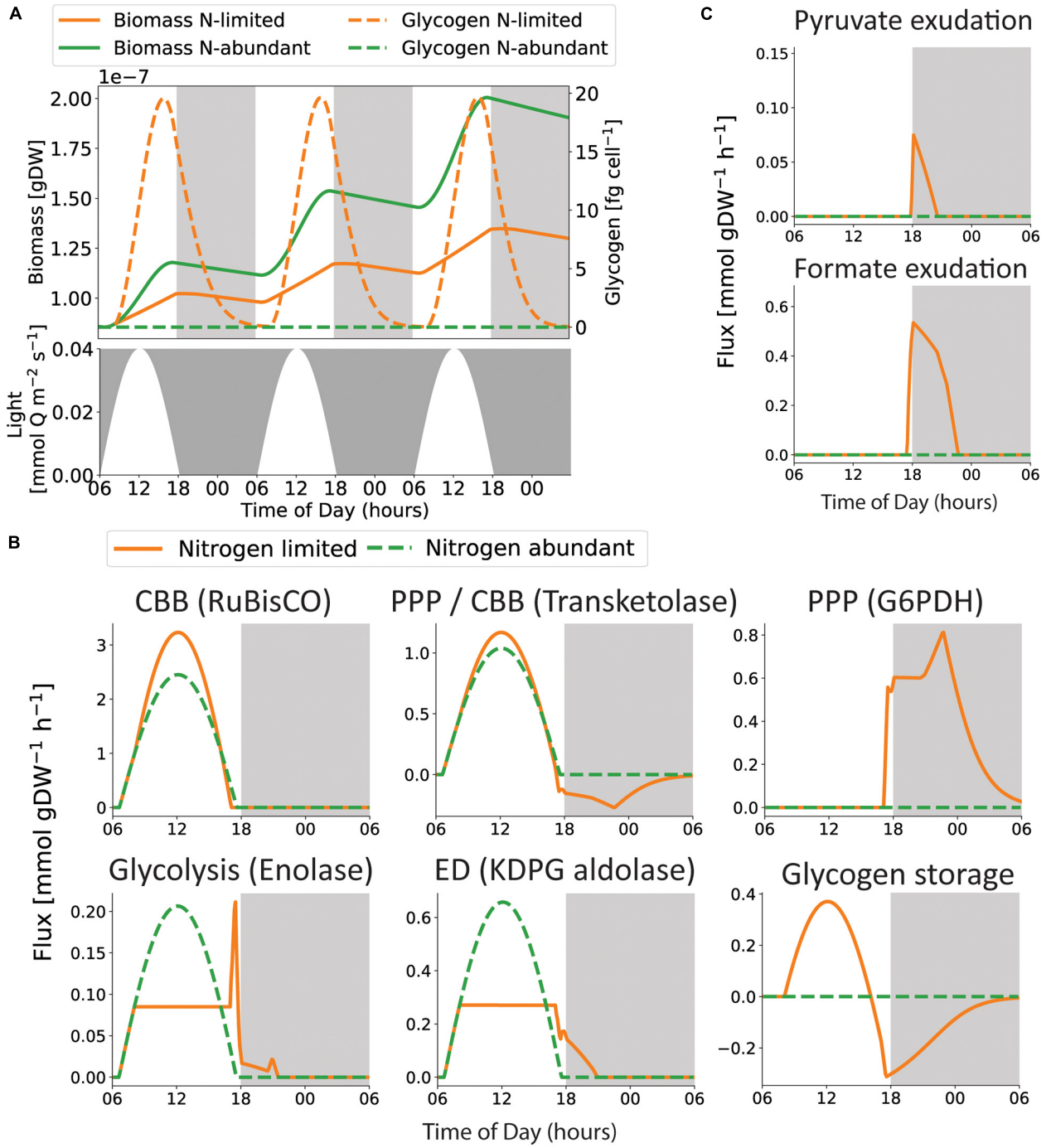


FIGURE 6 | Insight into metabolic rearrangements during the diel cycle. **(A)** Light irradiance, biomass and glycogen storage throughout the diel cycle. We observe that the largest accumulated growth is found in the nitrogen-abundant condition (green), but glycogen is only predicted to be stored in the nitrogen-poor condition (orange). **(B)** The flux distributions shifts when the metabolism switches from photosynthesis to glycogen catabolism, displayed by five reactions representing the Calvin Cycle (CBB), Glycogen metabolism, lower part of glycolysis, The Entner–Doudoroff (ED) and the Pentose Phosphate Pathway (PPP). **(C)** *i*SO595 predicts that the depletion of glycogen is accompanied by exudation of formate and pyruvate.

ATP demand necessary to support higher growth rates in nitrogen-abundant conditions. Additionally, our simulations predicted that the use of the Entner–Doudoroff pathway during photosynthesis creates precursor metabolites for growth during light hours, and a shift to the Pentose Phosphate Pathway (PPP) during nighttime. This trend might occur as an alternative for generating NADPH (**Supplementary Figure 6**). Upregulation of the PPP enzymes during dusk and the first half of the night time

was also observed in the proteome of *Prochlorococcus* (Waldbauer et al., 2012). Several enzymatic transformations participate in both the Calvin cycle and the PPP, although in opposite directions (Waldbauer et al., 2012). These transformations were captured in our simulations, specifically as demonstrated by transketolase (**Figure 6B**). Additionally, the consumption of glycogen during nighttime might lead to exudation of pyruvate and formate (**Figure 6C**). This prediction is supported by recent observations; formate is exuded during both nutrient-replete and phosphate-limited growth in *Prochlorococcus* strains MED4 and MIT9312 under constant light (Bertilsson et al., 2005), as well as when phosphonates are metabolized in *Prochlorococcus* strain MIT9301 (Sosa et al., 2019). Thus, *Prochlorococcus* are potential formate sources for heterotrophs. However, degradation of phosphonates yields formate as an immediate byproduct, and the current modeling framework is not suited to evaluate whether a equally high amount of intracellular formate is feasible during glycogen degradation, as intracellular metabolite concentrations are not readily represented in dFBA. Pyruvate exudation in *Prochlorococcus* is indicated from previous co-cultivations with SAR11 (Becker and Hogle, 2019), and from upregulation of genes encoding pyruvate kinase and a pyruvate efflux transporter during extended darkness (Biller et al., 2018). Furthermore, pyruvate is exuded when fixed carbon is consumed in the closely related strains *S. elongatus* PCC 7942 and *S.* sp. PCC 6803 (Carrieri et al., 2012; Benson et al., 2016).

The shift from photosynthesis and carbon fixation to glycogen catabolism is also associated with a switch in production and consumption of energetic cofactors (**Supplementary Figure 6**). Generation of ATP is performed concomitantly by ATP synthase in both the thylakoid membrane and the periplasmic membrane during photosynthesis. The periplasmic ATP synthase is first driven by reduced cofactors (NADPH) generated by the electron transport chain in the light-dependent part of photosynthesis (**Supplementary Figure 6**). ATP is consumed by two separate processes: growth- and maintenance-associated reactions reach a threshold once growth is limited by the nitrogen abundance, while the recycling of precursors for the Calvin cycle follows the shape of light absorption throughout the day. In agreement with previous work (Park and Choi, 2017), our model predicted higher rates of NADPH production than NADH.

Next, we explored the ability of our model to dynamically capture biologically relevant phenotypes by performing dynamic FBA simulations of knock-out mutants in *Prochlorococcus,* focusing on two gene deletions disrupting different parts of glycogen metabolism. Δ*glgC* breaks synthesis of ADP-glucose and thus the storage of glycogen and Δ*gnd*, knocking out 6-phosphogluconate dehydrogenase, a key reaction in the Pentose Phosphate pathway found to fuel the Calvin cycle with precursor metabolites during the onset of photosynthesis (Shinde et al., 2020). Our dynamic FBA simulations in COMETS (**Supplementary Figure 7**) showed similar growth between Δ*gnd* and the wild type and slightly lower growth for Δ*glgC*. We set out to compare these observations with available experimental data. Since genetic tools for the modification of *Prochlorococcus* are still lacking (Laurenceau et al., 2020), we chose data from the closely related cyanobacteria *Synechococcus* as recent work described the impact of Δ*glgC* and Δ*gnd* on its growth during diel cycles (Shinde et al., 2020). Indeed, we found very good agreement between measured and predicted growth for both the wild-type and Δ*glgC* mutant where glycogen storage is disrupted (Shinde et al., 2020) (**Supplementary Figure 7**). One of the notable limitations of dynamic FBA is the ability to quantify intermediates and precursor pools that might drive the initiation of a pathway. This comes mainly from the assumption of a quasi steady-state of intracellular metabolite pools at each time point. Although the comparison is strictly qualitative and concerns strains with known differences (Mary et al., 2004), these findings demonstrated the ability of our reconstruction to capture metabolic trends in response to genetic perturbations, indicating that *i*SO595 will be a valuable tool in future research of *Prochlorococcus.* Overall, our dynamic simulations display biological and physiological behaviors that are consistent with expectations, and at the same time provide valuable insight into the putative internal metabolic processes that might modulate the *Prochlorococcus* growth under environmental and genome-induced constraints.

CONCLUSION

Our study provides a detailed systematic view of the underlying metabolic trends modulating carbon storage and exudation in *Prochlorococcus*. *Prochlorococcus* is known to interact with other bacteria in its surroundings (Sher et al., 2011; Aharonovich and Sher, 2016; Biller et al., 2018; Hennon et al., 2018). It is currently impossible to predict the fluxes of organic matter (or of the myriad metabolites comprising it, such as amino acids, sugars, and organic acids) between phytoplankton and bacteria. Yet, quantifying such fluxes and predicting them from genomic surveys, as shown here, serves a number of roles: (1) It can provide experimentally testable and mechanistic hypotheses on inter-microbial exchanges and competition, (2) It has the potential to increase knowledge about the specific metabolites that may mediate these interactions; and (3) It would enable the construction of improved models of biogeochemical cycles which consider the diverse and powerful metabolic capabilities of the ocean microbiome.

Genome-scale metabolic-network reconstructions are powerful tools, but not without limitations. Mainly, the predictive accuracy rests on the quality and completeness of the metabolic network. The construction and curation of these metabolic networks depend heavily on data availability and annotation accuracy, which may be scarce for less studied organisms. Several methods have been developed to fill the gaps of incomplete network reconstructions. For example, FastGapFill incorporates missing knowledge from universal, non-organism specific data (Thiele et al., 2014), ModelSEED fills gaps through the use of thermodynamic parameters and FBA simulations to achieve minimal growth (Henry et al., 2010), and MENECO uses a topology graph based approach to look for minimal sets of metabolic reactions that support growth and the producibility of target metabolites (Prigent et al., 2017). In this work we used a novel semi-automated gap-filling method (ReFill) to

increase existing knowledge in the reconstruction by up to 25%. In contrast to other standard gap filling approaches, ReFill has the specific capability to add individual reactions through a recursive algorithm that guarantees complete connectivity to the existing network, incorporating the maximal possible amount of validated, organism specific metabolic annotations. However, this approach employs high stringency and thus adds limited amounts of knowledge. Considering that *Prochlorococcus* strains have some of the smallest known genomes among free-living organisms, a 25% increase in knowledge serves as a significant improvement in the predictive capacity of the model. However, reconstruction of high-quality genome-scale metabolic models is an iterative process, where new data, knowledge, and scope create opportunities for further model improvement. One example of this possibility is the CO_2 concentrating mechanisms in *Prochlorococcus.* This mechanism is known to be sustained by proton and ion gradients across the cell membrane at an energetic cost (Hopkinson et al., 2014; Burnap et al., 2015). However, the comprehensive knowledge and annotation of ion transporters necessary to model this mechanism are lacking, and are therefore not included in iSO595. With the advancement of data collection and annotation tools, together with the use of ReFill or similar algorithms, metabolic knowledge can be added to such reconstructions, improving their predictive abilities and mimicry of biological and physiological processes.

Other limitations of static and dynamic FBA simulations include the inability to represent metabolite concentrations and the lack of regulatory effects. Furthermore, since COMETS, like most other implementations of dFBA, simulates millions of asynchronously growing and dividing cells on the mesoscopic scale, cell cycle processes are not readily incorporated into this framework. Thus, future extensions to this work include the implementation of cell division in *Prochlorococcus*, known to occur in the afternoon (Vaulot et al., 1995). Another improvement would be an accurate representation of the costs associated with light damage and the production of protective pigments required to combat excessive light absorption. This could potentially be accounted for by extending the current *Prochlorococcus* GEM to a framework that includes macromolecular allocation, such as Resource Balance Analysis (Goelzer et al., 2011), conditional FBA (Rügen et al., 2015) or models of metabolism and macromolecular expression (ME models) (Thiele et al., 2012). Along these lines, one could relate mortality with an inability to maintain basic cellular functions, rather than a fixed death rate. However, the relationship between cell mortality and metabolism is not well constrained, and its representation in dFBA models is currently rudimentary. Future work is needed to better understand mortality and represent it in models of cell metabolism, ecosystems and biogeochemistry. Finally, our findings contribute to a growing body of work on the underlying metabolic mechanisms modulating the metabolic success of *Prochlorococcus*. The approaches shown here provide systematic insights corroborated in recent and well-known works and provide strong foundations for future studies of *Prochlorococcus* metabolism with particular interest in its interaction with other microorganisms and the effects of these on community composition and larger biogeochemical cycles.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: KEGG (through Python API): https://www.genome.jp/kegg/, TransportDB: http://www.membranetransport.org/transportDB2/index.html, Metabolights: https://www.ebi.ac.uk/metabolights/MTBLS567, and former Prochlorococcus model: https://msystems.asm.org/content/1/6/e00065-16.

AUTHOR CONTRIBUTIONS

DSh and DSe designed the study. SO, SS, and DSe developed the computational models and performed the computational analyses, with input from DSh. SO and SS wrote a first version of the manuscript. DSe, DSh, and EA oversaw the project and contributed to the final version of this manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by the Human Frontiers Science Program (grant RGP0020/2016) and the National Science Foundation (NSFOCE-BSF 1635070) to DSe and DSh. DSe also acknowledges support by the National Science Foundation (grant 1457695), the Directorates for Biological Sciences and Geosciences at the National Science Foundation and NASA (agreement nos. 80NSSC17K0295, 80NSSC17K0296 and 1724150) issued through the Astrobiology Program of the Science Mission Directorate, and the Boston University Interdisciplinary Biomedical Research Office. SS was funded by SINTEF, the Norwegian graduate research school in bioinformatics, biostatistics and systems biology (NORBIS) and by the INBioPharm project of the Centre for Digital Life Norway (Research Council of Norway grant no. 248885).

ACKNOWLEDGMENTS

We are grateful to members of the Segrè lab for constructive feedback on the manuscript, and to members of the labs of DS, Hans-Peter Grossart, and Maren Voss for helpful and fun discussions on marine microbes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.586293/full#supplementary-material

Supplementary Figure 1 | Estimation of kinetic parameters for the uptake of ammonium in *Prochlorococcus*. **(A)** All combinations of K_m and V_{max} along the red trajectory matches the observed gross growth rate of 0.5 d^{-1} (Grossowicz et al., 2017). However, when we compare the dynamics of cell density **(B)** and ammonium concentration **(C)** we find that the best overall prediction is achieved

using K_m = 0.39 mM and V_{max} = 0.9 mmol gDW^{-1} h^{-1} (marked by an orange dot in **A**).

Supplementary Figure 2 | ATP availability influence modulates the trade-off between glycogen storage and growth. Phenotypic phase planes (Edwards et al., 2002) illustrate the combined effect of glycogen storage and bicarbonate uptake on the maximal growth rate. Compared to the base model **(A)**, we observe how that the trade-off is strongly affected by modulated ATP availability, either from an artificial reaction providing extra ATP **(B)** or by increasing **(C)** or decreasing **(D)** the amount of available light. Increasing ATP allows more glycogen storage without reducing the growth rate.

Supplementary Figure 3 | TCA cycle flux diagram differences between the most common phenotypes. Flux diagrams of the TCA cycle in the most common phenotypes 2 (colored orange) and 3 (colored blue). Reactions are denoted by KEGG reaction ids. Reaction colors correspond to cluster colors presented in **Figure 4**.

Supplementary Figure 4 | Sensitivity analysis of dFBA simulations to variability in ammonium concentration, kinetic coefficients of ammonium uptake and light irradiance. **(A)** These phase diagrams display which combinations of K_m and V_{max}, describing the uptake of ammonium, that leads to glycogen accumulation (red area) at peak irradiance based on the light amplitude and ammonium concentration used to simulate nitrogen-abundant (left) and nitrogen-poor (right) conditions in **Figure 6**. In the orange area no glycogen accumulation is predicted as the growth is limited by the available light, rather than nitrogen. **(B)** These panels display similar phase diagrams as in **(A)**, but for different amplitudes of light irradiance, represented by black curves, and for different ammonium concentrations (indicated on top of each panel). The number written on each black curve represents the light irradiance amplitude in μmol Q m^{-2} s^{-1}. The ammonium concentration in the top left panel is equal to the concentration in our simulated nitrogen-poor conditions, and thus, the 40.0 40 μmol Q m^{-2} s^{-1} line in this panel is identical to the boundary between the two phases in the right panel in **(A)**. The range of ammonium concentrations is chosen so that it covers both our simulated environment and the ammonium concentration in oligotrophic oceans. The blue, green, and orange points display the combinations of K_m and V_{max} used/provided in this work and previous publications, respectively.

Supplementary Figure 5 | Sensitivity analysis of dFBA simulation to different parametrization of glycogen consumption. All panels display results obtained from dFBA simulations in COMETS with different combinations of K_m and V_{max}, describing the consumption of intracellular glycogen, with line colors and corresponding values as shown in the bottom right corner color matrix. The purple color corresponds to the values used to run the simulations shown in **Figure 6**. **(A)** Growth curves. **(B)** Accumulated glycogen per gram dry weight of biomass. **(C)** Predicted reaction fluxes for the same 8 reactions as shown in **Figure 6**.

Supplementary Figure 6 | The transition from daytime to nighttime is associated with a drastic change in the production and consumption of the energy-carrying cofactors. The figure panels show the major sources (left) and drains (right) of the cofactors ATP, NAPDH and NADH. The legend shows the reaction IDs used in iSO595. **(A)** ATP is produced by both the thylakoid (R00086th) and the periplasmic (R00086p) ATP synthase during daytime, but mostly by the periplasmic ATP synthase (respiration) during nighttime. ATP is consumed by reactions associated with growth (BIOMASS and BProtein), cellular maintenance (Maintenance) and storage of glycogen (R00948) in addition to reactions recycling precursors for the Calvin cycle (R01512 and R01523) and acetyl-CoA carboxylase (R00742). **(B)** NADPH is produced by ferredoxin reductase (fdr) during daytime and by the pentose phosphate pathway (R02736 and R01528) during nighttime. The NADPH is either used to drive the proton gradient across the periplasmic membrane (NADPHDHp) or in Gluconeogenesis (R01063) to refuel the Calvin cycle during photosynthesis. **(C)** NADH production is correlated with the growth rate and dominated by pyruvate dehydrogenase (R00209) during daytime and the glycine cleavage system (R01221) during nighttime. NADH is consumed by 6-phosphogluconate dehydrogenase (reverse, R10221) during daytime, NADH transhydrogenase (R00112) solely during dusk and concomitant with methylenetetrahydrofolate reductase (R07168) during nighttime.

Supplementary Figure 7 | Predicted growth curves show good agreement in a qualitative comparison with experimental growth of *Synechococcus*. To compare growth data we have overlaid growth curves predicted for the wild-type, the Δ*glgC*-mutant and the Δ*gnd*-mutant of *Prochlorococcus* with experimental OD measurements of *Synechococcus elongatus PCC 794* (Shinde et al., 2020). We find a very good agreement for the wild-type and Δ*glgC*-mutant, but not for the Δ*gnd*-mutant. The lower panel shows how the model predicts the allocation and consumption of each of the three strains.

Supplementary Table 1 | List of reactions added to iJC568 to form iSO595.

Supplementary Table 2 | List of reactions from iJC568 that are modified in iSO595.

Supplementary Table 3 | Parameter ranges used in the sampling of nutrient environments.

Supplementary Table 4 | List of blocked exchange reactions prior to sampling of nutrient environments.

Supplementary Table 5 | Parameter values used to run dynamic FBA in COMETS.

Supplementary Material 1 | Results from the BLAST-search used to identify 6PG-dehydratase (EC: 4.2.1.12) encoded by PMM0774 in *P. marinus* MED4.

Supplementary Material 2 | Memote snapshot report of *i*SO595.

REFERENCES

Aharonovich, D., and Sher, D. (2016). Transcriptional response of Prochlorococcus to co-culture with a marine Alteromonas: differences between strains and the involvement of putative infochemicals. *ISME J.* 10, 2892–2906. doi: 10.1038/ismej.2016.70

Alonso-Casajús, N., Dauvillée, D., Viale, A. M., Muñoz, F. J., Baroja-Fernández, E., Morán-Zorzano, M. T., et al. (2006). Glycogen phosphorylase, the product of the glgP gene, catalyzes glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli*. *J. Bacteriol.* 188, 5266–5272. doi: 10.1128/JB.01566-05

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

Amin, S. A., Parker, M. S., and Armbrust, E. V. (2012). Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* 76, 667–684. doi: 10.1128/MMBR.00007-12

Arrieta, J. M., Mayol, E., Hansman, R. L., and Herndl, G. J. (2015). Dilution limits dissolved organic carbon utilization in the deep ocean. *Science* 348, 331–334.

Arteaga, L., Pahlow, M., and Oschlies, A. (2016). Modeled Chl:C ratio and derived estimates of phytoplankton carbon biomass and its contribution to total particulate organic carbon in the global surface ocean. *Global Biogeochem. Cycles* 30, 1791–1810. doi: 10.1002/2016GB005458

Becker, J. W., Berube, P. M., Follett, C. L., Waterbury, J. B., Chisholm, S. W., Delong, E. F., et al. (2014). Closely related phytoplankton species produce similar suites of dissolved organic matter. *Front. Microbiol.* 5:111. doi: 10.3389/fmicb.2014.00111

Becker, J. W., and Hogle, S. L. (2019). Co-culture and biogeography of Prochlorococcus and SAR11. *ISME J.* 13, 1506–1519. doi: 10.1038/s41396-019-0365-4

Benson, P. J., Purcell-Meyerink, D., Hocart, C. H., Truong, T. T., James, G. O., Rourke, L., et al. (2016). Factors altering pyruvate excretion in a glycogen storage mutant of the cyanobacterium, synechococcus PCC7942. *Front. Microbiol.* 7:475. doi: 10.3389/fmicb.2016.00475

Bertilsson, S., Berglund, O., Pullin, M. J., and Chisholm, S. W. (2005). Release of dissolved organic matter by *Prochlorococcus*. *Vie et Milieu* 55, 225–231.

Biller, S. J., Berube, P. M., Lindell, D., and Chisholm, S. W. (2015). Prochlorococcus: the structure and function of collective diversity. *Nat. Rev. Microbiol.* 13, 13–27. doi: 10.1038/nrmicro3378

Biller, S. J., Coe, A., Roggensack, S. E., and Chisholm, W. (2018). Heterotroph interactions alter prochlorococcus transcriptome dynamics during extended periods of darkness. *mSystems* 3:e00040-18.

Braakman, R. (2019). Evolution of cellular metabolism and the rise of a globally productive biosphere. *Free Radic. Biol. Med.* 140, 172–187. doi: 10.1016/j.freeradbiomed.2019.05.004

Braakman, R., Follows, M. J., and Chisholm, S. W. (2017). Metabolic evolution and the self-organization of ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* 114, E3091–E3100. doi: 10.1073/pnas.1619573114

Bricaud, A., Claustre, H., Ras, J., and Oubelkheir, K. (2004). Natural variability of phytoplanktonic absorption in oceanic waters: influence of the size structure of algal populations. *J. Geophys. Res. Ocean* 109, 1–12. doi: 10.1029/2004JC002419

Broddrick, J. T., Rubin, B. E., Welkie, D. G., Du, N., Mih, N., Diamond, S., et al. (2016). Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. *Proc. Natl. Acad. Sci. U.S.A.* 113, E8344–E8353. doi: 10.1073/pnas.1613446113

Burnap, R., Hagemann, M., and Kaplan, A. (2015). Regulation of CO2 concentrating mechanism in cyanobacteria. *Life* 5, 348–371. doi: 10.3390/life5010348

Campillo-Brocal, J. C., Lucas-Elió, P., and Sanchez-Amat, A. (2015). Distribution in different organisms of amino acid oxidases with fad or a quinone as cofactor and their role as antimicrobial proteins in marine bacteria. *Mar. Drugs* 13, 7403–7418. doi: 10.3390/md13127073

Cano, M., Holland, S. C., Artier, J., Burnap, R. L., Ghirardi, M., Morgan, J. A., et al. (2018). Glycogen synthesis and metabolite overflow contribute to energy balancing in cyanobacteria. *Cell Rep.* 23, 667–672. doi: 10.1016/j.celrep.2018.03.083

Carrieri, D., Paddock, T., Maness, P.-C., Seibert, M., and Yu, J. (2012). Photocatalytic conversion of carbon dioxide to organic acids by a recombinant cyanobacterium incapable of glycogen storage. *Energy Environ. Sci.* 5:9457. doi: 10.1039/c2ee23181f

Casey, J. R., Mardinoglu, A., Nielsen, J., and Karl, D. M. (2016). Adaptive evolution of phosphorus metabolism in prochlorococcus. *mSystems* 1, 1–15. doi: 10.1128/mSystems.00065-16.Editor

Chen, X., Schreiber, K., Appel, J., Makowka, A., Fähnrich, B., Roettger, M., et al. (2016). The Entner-Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5441–5446. doi: 10.1073/pnas.1521916113

Cirri, E., and Pohnert, G. (2019). Algae–bacteria interactions that balance the planktonic microbiome. *New Phytol.* 223, 100–106. doi: 10.1111/nph.15765

Coles, V. J., Stukel, M. R., Brooks, M. T., Burd, A., Crump, B. C., Moran, M. A., et al. (2017). Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* 358, 1149–1154. doi: 10.1126/science.aan5712

Damrow, R., Maldener, I., and Zilliges, Y. (2016). The multiple functions of common microbial carbon polymers, glycogen and PHB, during stress responses in the non-diazotrophic cyanobacterium synechocystis sp. PCC 6803. *Front. Microbiol.* 7:966. doi: 10.3389/fmicb.2016.00966

Dauvillée, D., Kinderf, I. S., Li, Z., Kosar-Hashemi, B., Samuel, M. S., Rampling, L., et al. (2005). Role of the *Escherichia coli* glgX gene in glycogen metabolism. *J. Bacteriol.* 187, 1465–1473. doi: 10.1128/JB.187.4.1465-1473.2005

Davey, M., Tarran, G. A., Mills, M. M., Ridame, C., Geider, R. J., and LaRoche, J. (2008). Nutrient limitation of picophytoplankton photosynthesis and growth in the tropical North Atlantic. *Limnol. Oceanogr.* 53, 1722–1733. doi: 10.4319/lo.2008.53.5.1722

de Groot, D. H., Lischke, J., Muolo, R., Planqué, R., Bruggeman, F. J., and Teusink, B. (2020). The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cell. Mol. Life Sci.* 77, 441–453. doi: 10.1007/s00018-019-03380-2

Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N., and Dunne, J. P. (2007). Spatial coupling of nitrogen inputs and losses in the ocean. *Nature* 445, 163–167. doi: 10.1038/nature05392

Diamond, S., Jun, D., Rubin, B. E., and Golden, S. S. (2015). The circadian oscillator in Synechococcus elongatus controls metabolite partitioning during diurnal growth. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1916–E1925. doi: 10.1073/pnas.1504576112

Díaz-Troya, S., López-Maury, L., Sánchez-Riego, A. M., Roldán, M., and Florencio, F. J. (2014). Redox regulation of glycogen biosynthesis in the cyanobacterium synechocystis sp. PCC 6803: analysis of the AGP and glycogen synthases. *Mol. Plant* 7, 87–100. doi: 10.1093/mp/sst137

Domínguez-Martín, M. A., López-Lozano, A., Clavería-Gimeno, R., Velázquez-Campoy, A., Seidel, G., Burkovski, A., et al. (2018). Differential NtcA responsiveness to 2-oxoglutarate underlies the diversity of C/N balance regulation in Prochlorococcus. *Front. Microbiol.* 8:2641. doi: 10.3389/fmicb.2017.02641

Dubinsky, Z., and Berman-Frank, I. (2001). Uncoupling primary production from population growth in photosynthesizing organisms in aquatic ecosystems. *Aquatic Sci.* 63, 4–17. doi: 10.1007/PL00001343

Dukovski, I., Bajić, D., Chacón, J. M., Quintin, M., Vila, J. C., Sulheim, S., et al. (2020). *). Computation of Microbial Ecosystems in Time and Space (COMETS): An Open Source Collaborative Platform for Modeling Ecosystems Metabolism*. Available online at: http://arxiv.org/abs/2009.01734 (accessed December 17, 2020).

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74

Edwards, J. S., Ramakrishna, R., and Palsson, B. O. (2002). Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* 77, 27–36. doi: 10.1002/bit.10047

Elbourne, L. D. H., Tetu, S. G., Hassan, K. A., and Paulsen, I. T. (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* 45, D320–D324. doi: 10.1093/nar/gkw1068

Feist, A. M., Nagarajan, H., Rotaru, A. E., Tremblay, P. L., Zhang, T., Nevin, K. P., et al. (2014). Constraint-based modeling of carbon fixation and the energetics of electron transfer in geobacter metallireducens. *PLoS Comput. Biol.* 10:e1003575. doi: 10.1371/journal.pcbi.1003575

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237

Fischer, E., and Sauer, U. (2005). Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat. Genet.* 37, 636–640. doi: 10.1038/ng1555

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9824–9829. doi: 10.1073/pnas.1307701110

Fogg, G. E., Nalewajko, C., and Watt, W. D. (1965). Extracellular products of phytoplankton photosynthesis. *Proc. R. Soc. London. Ser. B. Biol. Sci.* 162, 517–534. doi: 10.1098/rspb.1965.0054

Follows, M. J., Dutkiewicz, S., Grant, S., and Chisholm, S. W. (2007). Emergent biogeography of microbial communities in a model ocean. *Science* 315, 1843–1846. doi: 10.1126/science.1138544

Forchhammer, K., and Schwarz, R. (2019). Nitrogen chlorosis in unicellular cyanobacteria – a developmental program for surviving nitrogen deprivation. *Environ. Microbiol.* 21, 1173–1184. doi: 10.1111/1462-2920.14447

Forchhammer, K., and Selim, K. A. (2019). Carbon/nitrogen homeostasis control in cyanobacteria. *FEMS Microbiol. Rev.* 44, 33–53. doi: 10.1093/femsre/fuz025

Foster, S. Q., Al-haj, A., Church, M. J., van Dijken, G. L., Dutkiewicz, S., Fulweiler, R. W., et al. (2018). Ecological control of nitrite in the upper ocean. *Nat. Commun.* 9:1206. doi: 10.1038/s41467-018-03553-w

Fu, J., and Xu, X. (2006). The functional divergence of two glgP homologues in Synechocystis sp. PCC 6803. *FEMS Microbiol. Lett.* 260, 201–209. doi: 10.1111/j.1574-6968.2006.00312.x

Gilbert, J. D. J., and Fagan, W. F. (2011). Contrasting mechanisms of proteomic nitrogen thrift in Prochlorococcus. *Mol. Ecol.* 20, 92–104. doi: 10.1111/j.1365-294X.2010.04914.x

Goelzer, A., Fromion, V., and Scorletti, G. (2011). Cell design in bacteria as a convex optimization problem. *Automatica* 47, 1210–1218. doi: 10.1016/j.automatica.2011.02.038

Gomez, J. A., Höffner, K., and Barton, P. I. (2014). DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis. *BMC Bioinform.* 15:409. doi: 10.1186/s12859-014-0409-8

Grossowicz, M., Roth-Rosenberg, D., Aharonovich, D., Silverman, J., Follows, M. J., and Sher, D. (2017). *Prochlorococcus* in the lab and in silico: the importance of representing exudation. *Limnol. Oceanogr.* 62, 818–835. doi: 10.1002/lno.10463

Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biol.* 20, 1–18. doi: 10.1186/s13059-019-1730-3

Gudmundsson, S., and Thiele, I. (2010). Computationally efficient flux variability analysis. *BMC Bioinform.* 11:489. doi: 10.1186/1471-2105-11-489

Harcombe, W. R., Riehl, W. J., Dukovski, I., Granger, B. R., Betts, A., Lang, A. H., et al. (2014). Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep.* 7, 1104–1115. doi: 10.1016/j.celrep.2014.03.070

Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., et al. (2013). MetaboLights–an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 41, 781–786. doi: 10.1093/nar/gks1004

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* 14, 639–702. doi: 10.1038/s41596-018-0098-2

Hennon, G. M., Morris, J. J., Haley, S. T., Zinser, E. R., Durrant, A. R., Entwistle, E., et al. (2018). The impact of elevated CO 2 on Prochlorococcus and microbial interactions with â € helper' bacterium Alteromonas. *ISME J.* 12, 520–531. doi: 10.1038/ismej.2017.189

Henry, C. S., Dejongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28, 977–982. doi: 10.1038/nbt.1672

Holtzendorff, J., Partensky, F., Mella, D., Lennon, J.-F., Hess, W. R., and Garczarek, L. (2008). Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium *Prochlorococcus marinus* PCC 9511. *J. Biol. Rhythms* 23, 187–199. doi: 10.1177/0748730408316040

Hopkinson, B. M., Young, J. N., Tansik, A. L., and Binder, B. J. (2014). The minimal CO2-concentrating mechanism of prochlorococcus spp. MED4 is effective and efficient. *Plant Physiol.* 166, 2205–2217. doi: 10.1104/pp.114.247049

Iglesias, A. A., Kakefuda, G., and Preiss, J. (1991). Regulatory and structural properties of the cyanobacterial ADPglucose pyrophosphorylases. *Plant Physiol.* 97, 1187–1195. doi: 10.1104/pp.97.3.1187

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., and Chisholm, S. W. (2006). Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. *Science* 311, 1737–1740. doi: 10.1126/science.1118052

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.

Kashyap, A. K., and Singh, D. P. (1985). Ammonium transport in unicellular cyanobacterium anacystis nidulans. *J. Plant Physiol.* 121, 319–330. doi: 10.1016/S0176-1617(85)80025-0

Kavvas, E. S., Seif, Y., Yurkovich, J. T., Norsigian, C., Poudel, S., Greenwald, W. W., et al. (2018). Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Syst. Biol.* 12:25. doi: 10.1186/s12918-018-0557-y

Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., et al. (2007). Patterns and implications of gene gain and loss in the evolution of prochlorococcus. *PLoS Genet.* 3:e231. doi: 10.1371/journal.pgen.0030231

Kim, J., Fabris, M., Baart, G., Kim, M. K., Goossens, A., Vyverman, W., et al. (2016). Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. *Plant J.* 85, 161–176. doi: 10.1111/tpj.13081

Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., et al. (2013). Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLoS Comput. Biol.* 9:e1003081. doi: 10.1371/journal.pcbi.1003081

Krumhardt, K. M., Callnan, K., Roache-Johnson, K., Swett, T., Robinson, D., Reistetter, E. N., et al. (2013). Effects of phosphorus starvation versus limitation on the marine cyanobacterium ProchlorococcusMED4 I: uptake physiology. *Environ. Microbiol.* 15, 2114–2128. doi: 10.1111/1462-2920.12079

Kujawinski, E. B. (2011). The impact of microbial metabolism on marine dissolved organic matter. *Ann. Rev. Mar. Sci.* 3, 567–599. doi: 10.1146/annurev-marine-120308-081003

Laurenceau, R., Bliem, C., Osburne, M. S., Becker, J. W., Biller, S. J., Cubillos-Ruiz, A., et al. (2020). Toward a genetic system in the marine cyanobacterium Prochlorococcus. *Access Microbiol.* 2:e000107. doi: 10.1099/acmi.0.000107

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., et al. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome−scale models. *Mol. Syst. Biol.* 6:390. doi: 10.1038/msb.2010.47

Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., et al. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* 38, 272–276. doi: 10.1038/s41587-020-0446-y

Long, S. P., Humphries, S., and Falkowski, P. G. (1994). Photoinhibition of photosynthesis in nature. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 45, 633–662. doi: 10.1146/annurev.pp.45.060194.003221

López-Sandoval, D. C., Rodríguez-Ramos, T., Cermeño, P., and Marañón, E. (2013). Exudation of organic carbon by marine phytoplankton: dependence on taxon and cell size. *Mar. Ecol. Prog. Ser.* 477, 53–60. doi: 10.3354/meps10174

Luan, G., Zhang, S., Wang, M., and Lu, X. (2019). Progress and perspective on cyanobacterial glycogen metabolism engineering. *Biotechnol. Adv.* 37, 771–786. doi: 10.1016/j.biotechadv.2019.04.005

Ma, L., Calfee, B. C., Morris, J. J., Johnson, Z. I., and Zinser, E. R. (2018). Degradation of hydrogen peroxide at the ocean's surface: the influence of the microbial community on the realized thermal niche of Prochlorococcus. *ISME J.* 12, 473–484. doi: 10.1038/ismej.2017.182

Ma, X., Coleman, M. L., and Waldbauer, J. R. (2018). Distinct molecular signatures in dissolved organic matter produced by viral lysis of marine cyanobacteria. *Environ. Microbiol.* 20, 3001–3011. doi: 10.1111/1462-2920.14338

Maarleveld, T. R., Khandelwal, R. A., Olivier, B. G., Teusink, B., and Bruggeman, F. J. (2013). Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnol. J.* 8, 997–1008. doi: 10.1002/biot.201200291

Mague, T. H., Friberg, E., Hughes, D. J., and Morris, I. (1980). Extracellular release of carbon by marine phytoplankton; a physiological approach. *Limnol. Oceanogr.* 25, 262–279. doi: 10.4319/lo.1980.25.2.0262

Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002). Dynamic Flux Balance Analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 83, 1331–1340. doi: 10.1016/S0006-3495(02)73903-9

Marañón, E., Cermeño, P., López-Sandoval, D. C., Rodríguez-Ramos, T., Sobrino, C., Huete-Ortega, M., et al. (2013). Unimodal size scaling of phytoplankton growth and the size dependence of nutrient uptake and use. *Ecol. Lett.* 16, 371–379. doi: 10.1111/ele.12052

Mary, I., Tu, C.-J., Grossman, A., and Vaulot, D. (2004). Effects of high light on transcripts of stress-associated genes for the cyanobacteria Synechocystis sp. PCC 6803 and Prochlorococcus MED4 and MIT9313. *Microbiology* 150, 1271–1281. doi: 10.1099/mic.0.27014-0

McDonald, S. M., Plant, J. N., and Worden, A. Z. (2010). The mixed lineage nature of nitrogen transport and assimilation in marine eukaryotic phytoplankton: a case study of Micromonas. *Mol. Biol. Evol.* 27, 2268–2283. doi: 10.1093/molbev/msq113

McInnes, L., Healy, J., and Astels, S. (2017). hdbscan: hierarchical density based clustering. *J. Open Source Softw.* 2:205. doi: 10.21105/joss.00205

McKinney, W. (2010). *pandas: a Foundational Python Library for Data Analysis and Statistics | R (Programming Language) | Database Index*. Available online at: https://www.scribd.com/document/71048089/pandas-a-Foundational-Python-Library-for-Data-Analysis-and-Statistics (accessed April 29, 2020).

Mella-Flores, D., Six, C., Ratin, M., Partensky, F., Boutte, C., Le Corguillé, G., et al. (2012). Prochlorococcus and synechococcus have evolved different adaptive mechanisms to cope with light and UV Stress. *Front. Microbiol.* 3:285. doi: 10.3389/fmicb.2012.00285

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., et al. (2017). iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* 35, 904–908. doi: 10.1038/nbt.3956

Monshupanee, T., and Incharoensakdi, A. (2014). Enhanced accumulation of glycogen, lipids and polyhydroxybutyrate under optimal nutrients and light intensities in the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Appl. Microbiol.* 116, 830–838. doi: 10.1111/jam.12409

Moore, C. M., Mills, M. M., Arrigo, K. R., Berman-Frank, I., Bopp, L., Boyd, P. W., et al. (2013). Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* 6, 701–710. doi: 10.1038/ngeo1765

Moore, L. R., Coe, A., Zinser, E. R., Saito, M. A., Sullivan, M. B., Lindell, D., et al. (2007). Culturing the marine cyanobacterium Prochlorococcus. *Limnol. Oceanogr. Methods* 5, 353–362. doi: 10.4319/lom.2007.5.353

Moore, L. R., Goericke, R., and Chisholm, S. W. (1995). Comparative physiology of Synechococcus and Prochlorococcus: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar. Ecol. Prog. Ser.* 116, 259–275.

Moradi, N., Liu, B., Iversen, M., Kuypers, M. M., Ploug, H., and Khalili, A. (2018). A new mathematical model to explore microbial processes and their constraints in phytoplankton colonies and sinking marine aggregates. *Sci. Adv.* 4, 1–10. doi: 10.1126/sciadv.aat1991

Moran, M. A., and Durham, B. P. (2019). Sulfur metabolites in the pelagic ocean. *Nat. Rev. Microbiol.* 17, 665–678. doi: 10.1038/s41579-019-0250-1

Moran, M. A., Kujawinski, E. B., Stubbins, A., Fatland, R., Aluwihare, L. I., Buchan, A., et al. (2016). Deciphering ocean carbon in a changing world. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3143–3151. doi: 10.1073/pnas.1514645113

Morel, A., and Bricaud, A. (1981). Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. *Deep Sea Res. Part A Oceanogr. Res. Pap.* 28, 1375–1393. doi: 10.1016/0198-0149(81)90039-X

Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The black queen hypothesis: evolution of dependencies through adaptive gene loss. *MBio* 3:e00036-12. doi: 10.1128/mBio.00036-12

Muñoz-Marín, M. C., Gómez-Baena, G., López-Lozano, A., Moreno-Cabezuelo, J. A., Díez, J., and García-Fernández, J. M. (2020). Mixotrophy in marine picocyanobacteria: use of organic compounds by Prochlorococcus and Synechococcus. *ISME J.* 14, 1065–1073. doi: 10.1038/s41396-020-0603-9

Nicholson, D. P., Stanley, R. H. R., and Doney, S. C. (2018). A Phytoplankton model for the allocation of gross photosynthetic energy including the trade-offs of diazotrophy. *J. Geophys. Res. Biogeosci.* 123, 1796–1816. doi: 10.1029/2017JG004263

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., and Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2678–2683. doi: 10.1073/pnas.1117907109

Noreña-Caro, D., and Benton, M. G. (2018). Cyanobacteria as photoautotrophic biofactories of high-value chemicals. *J. CO2 Util.* 28, 335–366. doi: 10.1016/j.jcou.2018.10.008

O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987. doi: 10.1016/j.cell.2015.05.019

Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614

Oschlies, A., Koeve, W., Landolfi, A., and Kähler, P. (2019). Loss of fixed nitrogen causes net oxygen gain in a warmer future ocean. *Nat. Commun.* 10, 1–7. doi: 10.1038/s41467-019-10813-w

Pacheco, A. R., Moel, M., and Segrè, D. (2018). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* 10:103.

Park, J., and Choi, Y. (2017). Cofactor engineering in cyanobacteria to overcome imbalance between NADPH and NADH: a mini review. *Front. Chem. Sci. Eng.* 11:66–71. doi: 10.1007/s11705-016-1591-1

Partensky, F., and Garczarek, L. (2010). *Prochlorococcus*: advantages and limits of minimalism. *Ann. Rev. Mar. Sci.* 2, 305–331. doi: 10.1146/annurev-marine-120308-081034

Partensky, F., Hess, W. R., and Vaulot, D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* 63, 106–127. doi: 10.1128/mmbr.63.1.106-127.1999

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res* 12, 2825–2830.

Pope, R. M., and Fry, E. S. (1997). Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. *Appl. Optics* 33, 8710–8723. doi: 10.1364/AO.36.008710

Prigent, S., Frioux, C., Dittami, S. M., Thiele, S., Larhlimi, A., Collet, G., et al. (2017). Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS Comput. Biol.* 13:e1005276. doi: 10.1371/journal.pcbi.1005276

Reid, A. (2012). *Incorporating Microbial Processes into Climate Change Models*. A Report by the American Academy of Microbiology. Washington, DC.

Reimers, A.-M., Knoop, H., Bockmayr, A., and Steuer, R. (2017). Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proc. Natl. Acad. Sci. U.S.A.* 114:201617508. doi: 10.1073/pnas.1617508114

Ribalet, F., Swalwell, J., Clayton, S., Jiménez, V., Sudek, S., Lin, Y., et al. (2015). Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8008–8012. doi: 10.1073/pnas.1424279112

Roth-rosenberg, D., Aharonovich, D., Omta, A.-W., Follows, M. J., and Sciences, P. (2019). Dynamic macromolecular composition and high exudation rates in Prochlorococcus. *bioRxiv* [Preprint]. doi: 10.1101/828897

Rügen, M., Bockmayr, A., and Steuer, R. (2015). Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. *Sci. Rep.* 5, 1–16. doi: 10.1038/srep15247

Saito, M. A., McIlvin, M. R., Moran, D. M., Goepfert, T. J., DiTullio, G. R., Post, A. F., et al. (2014). Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* 345, 1173–1177. doi: 10.1126/science.1256450

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., and Sauer, U. (2012). Multidimensional optimality of microbial metabolism. *Science* 336, 601–604. doi: 10.1126/science.1216882

Seaver, L. C., and Imlay, J. A. (2001). Hydrogen peroxide fluxes and compartmentalization inside growing *Escherichia coli*. *J. Bacteriol.* 183, 7182–7189. doi: 10.1128/JB.183.24.7182-7189.2001

Segrè, D., Vitkup, D., and Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15112–15117. doi: 10.1073/pnas.232349399

Sher, D., Thompson, J. W., Kashtan, N., Croal, L., and Chisholm, S. W. (2011). Response of Prochlorococcus ecotypes to co-culture with diverse marine bacteria. *ISME J.* 5, 1125–1132. doi: 10.1038/ismej.2011.1

Shinde, S., Zhang, X., Singapuri, S. P., Kalra, I., Liu, X., Morgan-Kiss, R. M., et al. (2020). Glycogen metabolism supports photosynthesis start through the oxidative pentose phosphate pathway in cyanobacteria. *Plant Physiol.* 182, 507–517. doi: 10.1104/pp.19.01184

Sosa, O. A., Casey, J. R., and Karl, D. M. (2019). Methylphosphonate oxidation in Prochlorococcus strain MIT9301 supports phosphate acquisition, formate excretion, and carbon assimilation into purines. *Appl. Environ. Microbiol.* 85:e00289-19. doi: 10.1128/AEM.00289-19

Steglich, C., Behrenfeld, M., Koblizek, M., Claustre, H., Penno, S., Prasil, O., et al. (2001). Nitrogen deprivation strongly affects Photosystem II but not phycoerythrin level in the divinyl-chlorophyll b-containing cyanobacterium Prochlorococcus marinus. *Biochim. Biophys. Acta Bioenerg.* 1503, 341–349. doi: 10.1016/S0005-2728(00)00211-5

Stramski, D., Bricaud, A., and Morel, A. (2001). Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. *Appl. Optics* 18, 2929–2945. doi: 10.1364/ao.40.002929

Szul, M. J., Dearth, S. P., Campagna, S. R., and Zinser, E. R. (2019). Carbon fate and flux in prochlorococcus under nitrogen limitation. *mSystems* 4:e00254-18. doi: 10.1128/msystems.00254-18

Thiele, I., Fleming, R. M. T., Que, R., Bordbar, A., Diep, D., and Palsson, B. O. (2012). Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PLoS One* 7:e45635. doi: 10.1371/journal.pone.0045635

Thiele, I., Hyduke, D. R., Steeb, B., Fankam, G., Allen, D. K., Bazzani, S., et al. (2011). A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella* Typhimurium LT2. *BMC Syst. Biol.* 5:8. doi: 10.1186/1752-0509-5-8

Thiele, I., Vlassis, N., and Fleming, R. M. T. (2014). FASTGAPFILL: efficient gap filling in metabolic networks. *Bioinformatics* 30, 2529–2531. doi: 10.1093/bioinformatics/btu321

Thornton, D. C. O. (2014). Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. *Eur. J. Phycol.* 49, 20–46. doi: 10.1080/09670262.2013.875596

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Van Mooy, B. A. S., Rocap, G., Fredricks, H. F., Evans, C. T., and Devol, A. H. (2006). Sulfolipids dramatically decrease phosphorus demand by

picocyanobacteria in oligotrophic marine environments. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8607–8612. doi: 10.1073/pnas.0600540103

Varma, A., and Palsson, B. O. (1994). Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology* 12, 994–998. doi: 10.1038/nbt1094-994

Vaulot, D., Marie, D., Olson, R. J., and Chisholm, S. W. (1995). Growth of Prochlorococcus, a photosynthetic prokaryote, in the equatorial Pacific Ocean. *Science* 268, 1480–1482. doi: 10.1126/science.268.5216.1480

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2

Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. (2012). Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS One* 7:e43432. doi: 10.1371/journal.pone.0043432

Ward, B. A., Collins, S., Dutkiewicz, S., Gibbs, S., Bown, P., Ridgwell, A., et al. (2019). Considering the role of adaptive evolution in models of the ocean and climate system. *J. Adv. Model. Earth Syst.* 11, 3343–3361. doi: 10.1029/2018MS001452

Welkie, D. G., Rubin, B. E., Diamond, S., Hood, R. D., Savage, D. F., and Golden, S. S. (2019). A hard day's night: cyanobacteria in diel cycles. *Trends Microbiol.* 27, 231–242.

Wintermute, E. H., Lieberman, T. D., and Silver, P. A. (2013). An objective function exploiting suboptimal solutions in metabolic networks. *BMC Syst. Biol.* 7:98. doi: 10.1186/1752-0509-7-98

Xiong, W., Cano, M., Wang, B., Douchi, D., and Yu, J. (2017). The plasticity of cyanobacterial carbon metabolism. *Curr. Opin. Chem. Biol.* 41, 12–19. doi: 10.1016/j.cbpa.2017.09.004

Yang, A. (2011). Modeling and evaluation of CO 2 supply and utilization in algal ponds. *Ind. Eng. Chem. Res.* 50, 11181–11192. doi: 10.1021/ie200723w

Yoshikawa, K., Toya, Y., and Shimizu, H. (2017). Metabolic engineering of *Synechocystis* sp. PCC 6803 for enhanced ethanol production based on flux balance analysis. *Bioprocess Biosyst. Eng.* 40, 791–796. doi: 10.1007/s00449-017-1744-8

Zavřel, T., Faizi, M., Loureiro, C., Poschmann, G., Stühler, K., Sinetova, M., et al. (2019). Quantitative insights into the cyanobacterial cell economy. *eLife* 8:e42508. doi: 10.7554/eLife.42508

Zhang, C. C., Zhou, C. Z., Burnap, R. L., and Peng, L. (2018). Carbon/nitrogen metabolic balance: lessons from cyanobacteria. *Trends Plant Sci.* 23, 1116–1130. doi: 10.1016/j.tplants.2018.09.008

Zinser, E. R., Lindell, D., Johnson, Z. I., Futschik, M. E., Steglich, C., Coleman, M. L., et al. (2009). Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, Prochlorococcus. *PLoS One* 4:e5135. doi: 10.1371/journal.pone.0005135

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 February 2021
doi: 10.3389/fmicb.2021.518865

Metagenomic Analysis Reveals Microbial Community Structure and Metabolic Potential for Nitrogen Acquisition in the Oligotrophic Surface Water of the Indian Ocean

Yayu Wang[1,2†], Shuilin Liao[1,3†], Yingbao Gai[4,5], Guilin Liu[6], Tao Jin[6], Huan Liu[1,7], Lone Gram[2], Mikael Lenz Strube[2], Guangyi Fan[6], Sunil Kumar Sahu[1,7], Shanshan Liu[6], Shuheng Gan[1], Zhangxian Xie[4], Lingfen Kong[4], Pengfan Zhang[1], Xin Liu[1,7] and Da-Zhi Wang[4]**

[1] BGI-Shenzhen, Shenzhen, China, [2] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark, [3] BGI Education Center, University of Chinese Academy of Sciences, Beijing, China, [4] State Key Laboratory of Marine Environmental Science, College of the Environment and Ecology, Xiamen University, Xiamen, China, [5] Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, China, [6] BGI-Qingdao, BGI-Shenzhen, Qingdao, China, [7] State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, China

OPEN ACCESS

Edited by:
Gipsi Lima Mendez,
Catholic University of Louvain, Belgium

Reviewed by:
Ahmed A. Shibl,
New York University Abu Dhabi, United Arab Emirates
Isabel Ferrera,
Instituto de Ciencias del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Spain

***Correspondence:**
Xin Liu
liuxin@genomics.cn
Da-Zhi Wang
dzwang@xmu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: 10 December 2019
Accepted: 25 January 2021
Published: 18 February 2021

Citation:
Wang Y, Liao S, Gai Y, Liu G, Jin T, Liu H, Gram L, Strube ML, Fan G, Sahu SK, Liu S, Gan S, Xie Z, Kong L, Zhang P, Liu X and Wang D-Z (2021) Metagenomic Analysis Reveals Microbial Community Structure and Metabolic Potential for Nitrogen Acquisition in the Oligotrophic Surface Water of the Indian Ocean. Front. Microbiol. 12:518865. doi: 10.3389/fmicb.2021.518865

Despite being the world's third largest ocean, the Indian Ocean is one of the least studied and understood with respect to microbial diversity as well as biogeochemical and ecological functions. In this study, we investigated the microbial community and its metabolic potential for nitrogen (N) acquisition in the oligotrophic surface waters of the Indian Ocean using a metagenomic approach. Proteobacteria and Cyanobacteria dominated the microbial community with an average 37.85 and 23.56% of relative abundance, respectively, followed by Bacteroidetes (3.73%), Actinobacteria (1.69%), Firmicutes (0.76%), Verrucomicrobia (0.36%), and Planctomycetes (0.31%). Overall, only 24.3% of functional genes were common among all sampling stations indicating a high level of gene diversity. However, the presence of 82.6% common KEGG Orthology (KOs) in all samples showed high functional redundancy across the Indian Ocean. Temperature, phosphate, silicate and pH were important environmental factors regulating the microbial distribution in the Indian Ocean. The cyanobacterial genus *Prochlorococcus* was abundant with an average 17.4% of relative abundance in the surface waters, and while 54 *Prochlorococcus* genomes were detected, 53 were grouped mainly within HLII clade. In total, 179 of 234 *Prochlorococcus* sequences extracted from the global ocean dataset were clustered into HL clades and exhibited less divergence, but 55 sequences of LL clades presented more divergence exhibiting different branch length. The genes encoding enzymes related to ammonia metabolism, such as urease, glutamate dehydrogenase, ammonia transporter, and nitrilase presented higher abundances than the genes involved in inorganic N assimilation in both microbial community and metagenomic *Prochlorococcus* population. Furthermore, genes associated with dissimilatory nitrate reduction, denitrification, nitrogen fixation, nitrification and anammox were absent in metagenome *Prochlorococcus* population,

i.e., nitrogenase and nitrate reductase. Notably, the *de novo* biosynthesis pathways of six different amino acids were incomplete in the metagenomic *Prochlorococcus* population and *Prochlorococcus* genomes, suggesting compensatory uptake of these amino acids from the environment. These results reveal the features of the taxonomic and functional structure of the Indian Ocean microbiome and their adaptive strategies to ambient N deficiency in the oligotrophic ocean.

Keywords: microbe, *Prochlorococcus*, metagenome, nitrogen metabolism, Indian Ocean

INTRODUCTION

The Indian Ocean is one of the largest oligotrophic water bodies, which covers approximately one-fifth of global ocean (Eakins and Sharman, 2010; Wei et al., 2019). As the warmest ocean on the earth, the Indian Ocean possesses unique biophysical properties that strongly influence the diversity and performance of its biota (Massana et al., 2011; Williamson et al., 2012; Díez et al., 2016; Thompson et al., 2017). However, the Indian Ocean is one of the least studied oceans regarding microbial community structure, functional capacity and the potential linkage of microbial taxa and environmental conditions, when compared with other oceans (Venter et al., 2004; Joint et al., 2011; Sjöstedt et al., 2014; Mende et al., 2017; Li Y. et al., 2018). A recent study on biodiversity and spatial distribution of bacteria in the water column of the eastern Indian Ocean has shown that Cyanobacteria and Actinobacteria are more predominant in the upper ocean while Alphaproteobacteria occur more frequently in the deeper layer (Wang et al., 2016). However, a study in the South Indian Ocean has indicated that Gammaproteobacteria dominate the microbial community and their potential functionality is shaped by the depth-related environmental parameters of the Agulhas Current (Phoma et al., 2018). Metagenomic analysis of the picocyanobacterial community in the equatorial waters of the Indian Ocean reveals that the genera *Prochlorococcus* and *Synechococcus* comprise 90% of the cyanobacterial reads (Díez et al., 2016). *Prochlorococcus* populations in the ocean have been defined with high light-adapted (hereafter HL) populations and low light-adapted (hereafter LL) populations, where *Prochlorococcus* HLIIA ecotype is abundant in the *Prochlorococcus* community in the Indian Ocean (Farrant et al., 2016). All together, these studies have provided a snapshot of the microbial diversity and geographical distribution in some specific areas of the Indian Ocean, but knowledge on the overall microbial taxonomic structure and functional capacity, and their influencing factors are still very limited.

Nitrogen (N) is an essential macronutrient limiting microbial growth and proliferation in the ocean, especially in the oligotrophic oceanic regimes (Zehr and Wad, 2002; Moore et al., 2013). Ambient N deficiency affects N assimilation, carbon fixation, photosynthesis, pigment and lipid accumulation of microbes (Schwarz and Forchhammer, 2005). Typically, a broad diversity of prokaryotes from Proteobacteria, Firmicutes, Verrucomicrobia, Planctomycetes, Acidobacteria, Chloroflexi, and Chlorobia can use dissolved inorganic N (DIN), such as nitrate, nitrite and ammonia, and these nutrients are sufficient to support microbial growth in coastal and upwelling areas (Li Y.-Y. et al., 2018; Pajares and Ramos, 2019). However, these N nutrients are extremely low in quantity throughout much of the surface oligotrophic ocean (Moore et al., 2013) and cannot support microbial growth. Instead, microbes have evolved diverse adaptive strategies to ambient N deficiency, for example, utilizing small but rapidly cycling dissolved organic nitrogen (DON), such as free amino acids, amines and urea as N source for cell growth and proliferation, indicating the essential roles of DON in maintaining microbial communities in the oligotrophic ocean (Wheeler and Kirchman, 1986; Zubkov et al., 2003; García-Fernández et al., 2004). However, little is known concerning nitrogen acquisition by microbes in the oligotrophic Indian Ocean (Kumar et al., 2009; Díez et al., 2016; Qian et al., 2018; Baer et al., 2019), which impedes our understanding of the mechanisms underlying the adaptive strategies of microbes to ambient N deficiency.

Metagenomic studies of the global ocean advance our understanding of diversity, evolution and functional potential of natural microbial communities (Sunagawa et al., 2015; Li Y. et al., 2018). Distribution of microbial diversity and biogeochemistry are structured largely by environmental gradients such as light, temperature, oxygen, salinity, and nutrients (Thompson et al., 2017). In this study, we applied a metagenomic approach to investigate microbial communities in the oligotrophic surface waters of the Indian Ocean, from the Andaman Sea in the east to the Red Sea in the west, and characterized their nitrogen acquisition strategies based on a 9.5 million (M) microbial gene set (**Figure 1** and **Supplementary Figure S1**). We paid particular attention to the predominant cyanobacterial genus *Prochlorococcus* and their nitrogen assimilation strategies due to its significant contribution to the stability, resilience and function of the marine ecosystem. This study expands our understanding of the microbial community and their metabolic potentials in the Indian Ocean, and provides fundamental metagenomic data for further microbial studies in the Indian Ocean, which is a significantly understudied realm.

MATERIALS AND METHODS

Sampling and DNA Extraction

A total of 17 surface seawater samples (S6–S38, 5 m depth) were collected from the outlet of profiling CTD (Seabird SBE21) during the 26th cruise of China Ocean Mineral Resources R

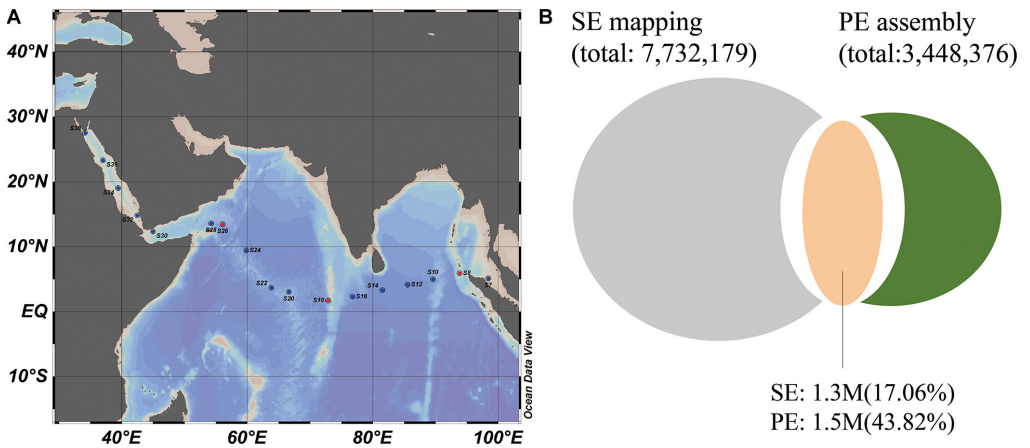


FIGURE 1 | Sampling locations and the microbial genes found in the Indian Ocean. **(A)** Sampling stations across the Indian Ocean and Red Sea. **(B)** The number of shared genes (orange) between the 17 SE sample gene set (gray) and three PE sample assembled gene set (green). 1.3 M and 1.7 M genes from two gene sets were shared, which occupied 17.06 and 48.3% of the gene numbers in two gene sets, respectively.

& D Association (COMRA[1]) on the R/V DayangYihao in May 2012. The sampling stations ranged across the Indian Ocean, from the Andaman Sea in the east to the Red Sea in the west (**Figure 1A**). Samples from the Red Sea were specifically referred to as S32, S34, S36, and S38. For each sample, more than 200 l of seawater was prefiltered using a GF/A glass fiber membrane (1.6 μm, 142 mm filter diameter, Whatman) with a Flojet Pump (04300242A) and then collected on a Supor polyethersulfone membrane (0.2 μm, 142 mm filter diameter, Pall). The membranes were placed in 50 ml tubes and immediately frozen in liquid nitrogen, and then stored at −80°C until further analysis. DNA extraction was performed using the PowerSoil DNA isolation kit (QIAGEN, United States) according to the instructions. Environmental parameters were also monitored, including temperature, salinity, pH and inorganic nutrients listed in **Supplementary Table S1**. Concentrations of phosphate, silicate, ammonium, nitrite and nitrate as well as chlorophyll-*a* concentration were measured in the laboratory according to the national standard protocol for investigation of chemical elements of seawater titled "Specifications for oceanographic survey-Part 4: Survey of chemical parameters in seawater" (GB/T 12763.4-2007, 2007).

DNA Sequencing on BGISEQ-500 and HiSeq4000

All 17 DNA samples were sequenced using a 50 bp single-ended (SE) sequencing strategy on a BGISEQ500 platform (Beijing Genomics Institute, Shenzhen). Firstly, the DNA sample was fragmented by ultrasound on Covaris E220 (Covaris, Brighton, United Kingdom). Then 400–700 bp DNA fragments were recovered to construct a library for each sample according to standard instructions (Fang et al., 2017). Three barcoded libraries were pooled together with equal amounts to make DNA nanoballs (DNB). Each DNB was loaded into one lane for sequencing on the BGISEQ-500 sequencer with the SE50 mode. The sequence images were base-called by the software Zebra call (base calling software developed for BGISEQ-500). A total of 12.81 billion raw reads were produced by the BGISEQ-500 platform, with a read length of 50 bp.

Since long and pair-end (PE) reads are more useful for metagenome assembly, three representative local samples were collected from the stations of S8 (E 93°49.228′, N 5°54.323′), S18 (E 72°52.169′, N 1°42.848′) and S26 (E 56°5.870′, N 13°26.297′) located at the east, central and west regions of the Indian Ocean, and were selected for in-depth sequencing using a 150 bp PE strategy on Hiseq4000 platform (Illumina). A total of 749.03 M reads were generated with an average 249.6 M reads for each sample.

Construction of the Gene Catalog

After a quality check and pre-processing by SOAP nuke (v1.5.3) (Chen et al., 2017), all clean reads of BGISEQ were aligned against the Ocean Microbiome Reference Gene Catalog (OM-RGC) (Sunagawa et al., 2013) using bowtie2.2.6 (Langmead and Salzberg, 2012) with parameters: bowtie2 –q –phred33 –sensitive –mixed –no-discordant -p 20 -k 200 -x. Then microbial genes mapped by unique reads were considered to be present in the samples. On average, 72.68% of BGISEQ SE reads were mapped to the OM-RGC, resulting in an average of 2.92 M genes in each sample (**Supplementary Table S2**). All the genes identified from 17 samples were merged together and then duplicates were removed, generating a 7.7 M unique gene set. We named this set the 17 SE sample gene set (**Supplementary Table S2**).

After quality control, the Illumina PE reads of three representative local samples (S8, S18, and S26) were *de novo* assembled into contigs using MegaHit (v1.0) with parameters: -t 20 -min-contig-len 500 –presets meta-large –min-count 5 (Li et al., 2016). A total of 23.04% of the Illumina PE reads were assembled into 3.26 M contigs (= 500bp). GeneMarkS (version

[1]http://www.comra.org/

2.7) (Besemer et al., 2001) was used to predict the genes using the parameter: gmhmmp -m MetaGeneMark_v1.mod in contigs obtained from each sample. The length of the predicted genes was no less than 150 bp. A non-redundant gene catalog (3.4M) was built by clustering all predicted genes using CDHIT (version 4.6.5) (Li and Godzik, 2006) with the following parameters: 95% identity and 90% coverage. The 3.4 M assembled gene set (called three PE sample gene set) was then compared with the 7.7 M gene set (17 SE sample gene set) to identify the overlapping genes using CDHIT with the same cutoff as above. As the shared genes accounted for a low proportion (**Figure 1B**) of the two gene sets, we merged these two gene sets together. We then removed the redundancy to generate a more complete gene set for the Indian ocean samples.

Taxonomic and Functional Annotation

Taxonomic and functional assignment of genes of the complete gene set was performed. Taxonomic assignments were performed by aligning protein sequence to the NCBI-NR database using DIAMOND (version 0.8.22) (Buchfink et al., 2015). The top alignment hits were retained according to the criteria of coverage $\geq$ 80%, identity $\geq$ 65%, and *e*-value $\leq$ 1e-5 as previous studies described (Qin et al., 2010; Li et al., 2014). Then, the taxonomic annotation of each gene was determined by the lowest common ancestor (LCA)-based algorithm implemented in MEGAN (Huson et al., 2007). Functional annotations were made by aligning the protein sequence against eggNOG (version 4.5) (Huerta-Cepas et al., 2016b) and KEGG (version.81) database by the highest scoring annotated hit(s) containing at least one high-scoring pair (HSP) scoring over 60 bits by DIAMOND.

Statistical Analyses

The relative abundances of genes were calculated based on the unique mapping reads as previously described (Qin et al., 2012). Firstly, the read counts from each gene were normalized by dividing with the gene length to calculate gene abundance. Then the relative abundance of each gene in samples were generated by dividing with the sum of abundance of all genes. Each KO abundance was generated by summing up the relative abundance of genes annotated to the same KO. The taxonomic profile was constructed using the same method. The meta-genes (unigenes) in the metagenomic gene set assigned to the genus of *Prochlorococcus* were extracted and the sum of their abundance in each sample was determined as the abundance of *Prochlorococcus*. The difference in abundance of *Prochlorococcus* between the central ocean and the coast samples, and the differences in dissolved inorganic nitrogen (DIN) concentration between the central ocean samples and coast samples were checked using *t*-test (basic function in R version 3.5.3). The clustering analysis based on physicochemical properties of all samples was performed using function pheatmap () in R package (version 3.5.3). To determine the shaping factors causing the variation in microbial communities among samples, the environmental variables were included in a redundancy analysis (RDA) model using the function rda () in "vegan" R-package, which was then reduced by the step-function in R according to the Akaike information criterion (AIC). The correlations of the relative abundance of *Prochlorococcus* with nitrate, nitrite and ammonium concentrations were separately calculated using Spearman's rank correlation test using the function cor. test () in R package (version 3.5.3).

Analysis of N Metabolism and Amino Acids Biosynthesis

The KOs involved in N metabolism were extracted from the merged gene set to analyze functional capacity of the Indian Ocean microbiome. The KOs involved in N metabolism and the annotated *Prochlorococcus* data were extracted to explore the N assimilatory pathways of the *Prochlorococcus* population in the Indian Ocean. The key enzymes were also manually annotated using the Pfam database to validate the existence of functional domain. All KOs in map01230-Biosynthesis of amino acids were extracted to explore biosynthesis reactions of amino acids. The KOs functionally annotated with "transporter" were extracted from the metagenome of *Prochlorococcus* population and the identified single *Prochlorococcus* genome respectively to predict the uptake model of amino acids from the environment. The single *Prochlorococcus* genome was identified based on 201 published *Prochlorococcus* genomes and the Indian Ocean metagenome sequencing data, and the detailed process was described below.

Phylogenetic Tree of *Prochlorococcus* Genomes

A total of 201 *Prochlorococcus* genomes including 27 known ecotype genomes (Biller et al., 2014) were downloaded from NCBI (**Supplementary Table S3**). The gene prediction of each genome was performed using GeneMarkS (version 2.7) (Besemer et al., 2001) with the parameter: gmhmmp -m Prochlorococcus_prefix.mod. All genes were merged together to construct a pangenome. Then the SE reads from each sample were mapped against the pangenome using bowtie2.2.6 to identify *Prochlorococcus* genes. If the portion of gene numbers mapped by reads was more than 50%, the Prochlorococcus strain was considered to be present in the sample. The abundance of each identified gene was calculated by dividing the mapping reads with the gene length. The abundance of the identified *Prochlorococcus* was represented by the average abundance of their identified genes. The same 40 single copy genes (SCGs) (Mende et al., 2013) from the 55 identified strains and 25 known ecotype strains were predicted by fetchMG (Kultima et al., 2012; Sunagawa et al., 2013). Then multiple alignments of 40 SCGs were performed using Mafft (Kazutaka and Daron, 2013). Here we selected 25 of 27 known ecotype strains as reference because we could not extract a full number of SCGs genes from the remaining two strains genomes. A phylogenic tree was constructed using ETE3 and the standard FastTree workflow (Price et al., 2010; Huerta-Cepas et al., 2016a).

To evaluate the diversity of *Prochlorococcus* population in the global ocean, 40 SCGs were extracted from 201 reference genomes, OM-RGC and the Indian Ocean gene set by fetchMG.

Among all the screened genes, the gene family COG0172 (*Seryl-tRNA synthetase*) showed the highest number, therefore, it was selected to construct the phylogenetic tree. Only the genes having a sequence length higher than 800 bp were retained in the tree.

RESULTS

Microbial Gene Capacity in the Indian Ocean

The locations of the 17 sampling stations along an east-west transect in the Indian Ocean (12°S, 96°E to 4°S, 39°E) are shown in **Figure 1A**. In total, 12.68 billion high quality SE reads were generated, resulting in an average of 37.16 Gb data per sample. Here, the microbial functional genes were identified by mapping these SE reads to the *Tara* Oceans gene set (OM-RGC, 40M genes). An average of 3 M genes per sample were identified with a 72.68% mapping rate. The reads of three PE sequenced samples were also mapped to OM-RGC and 94.90% of identified genes by PE reads were covered by SE reads (**Supplementary Figure S2**), indicating that the SE data was sufficient to map reference genes in this study. By merging all identified microbial genes of the 17 SE sequenced samples together, a 17 SE sample gene set was generated, comprising 7.7 M non-redundant genes (**Supplementary Table S2**).

To mine more specific microbial genes in the Indian Ocean surface water, a 3.4 M gene set based on *de novo* assembly was generated using three PE sequenced samples, which obtained 51.95–52.16% mapping rate in individual samples. The shared genes between the 17 SE sample gene set and three PE sample assembled gene set accounted for an average of 17.06 and 43.82% of gene abundance in the SE mapping gene set and PE assembly gene set (**Figure 1B**), suggesting that many novel genes could be explored by *de novo* assembly from local samples. So, a new complete microbial gene set (9.5M) from the Indian Ocean surface waters was generated by the combination of OM-RGC mapping and *de novo* assembly methods. Based on the new gene set, many more genes could be identified with higher reads mapping rate in each sample.

Taxonomic and Functional Variations in the Indian Ocean

A large proportion of genes (51.91%) in the complete gene set belong to bacteria, while 7.96%, 2.00% and 1.10% were annotated to virus, eukaryote and archaea, respectively (**Supplementary Figure S3A**). More than 69.51% of bacterial sequences could be annotated at the phylum level. Proteobacteria and Cyanobacteria dominated the microbial community by occupying an average 37.85 and 23.56% of total bacterial sequences of the Indian Ocean, respectively, followed by Bacteroidetes (3.73%), Actinobacteria (1.69%), Firmicutes (0.76%), Verrucomicrobia (0.36%), and Planctomycetes (0.31%) (**Supplementary Figure S3B**). Totally, 684 bacterial genera were found in the Proteobacteria phylum, of which the genus *Candidatus pelagibacter* was dominant with an average of 12.03% of relative abundance (**Supplementary Table S4**). The genera in the Bacteroidetes phylum presented similar relative abundances, such as the dominant genera *Fluviicola (1.53%* of relative abundance*)* and *Flavobacterium* (0.97% of relative abundance). The Actinobacteria were represented mainly by *Streptomyces* (5.41% of relative abundance) and *Candidatus actinomarina* (4.81% of relative abundance). The Firmicutes were mainly composed of *Clostridium* (8.80% of relative abundance) and *Bacillus* (8.83% of relative abundance). The genera *Coraliomargarita* (25.89% of relative abundance), *Pedosphaera* (11.05% of relative abundance) and *Verrucomicrobium* (5.24% of relative abundance) dominated the phylum of Verrucomicrobia. While *Rhodopirellula* (25.94% of relative abundance) and *Planctomyces* (12.13% of relative abundance) were the abundant groups of Planctomycetes. Among them, the photosynthetic cyanobacterial taxa such as *Prochlorococcus* and *Synechococcus* were the most abundant in each sample, comprising 86.11% of total cyanobacterial sequences (**Figure 2A** and **Supplementary Table S4**). Particularly, *Synechococcus* with higher relative abundance (average 25.9% of total bacterial sequence) predominated in the samples collected from S7, S8, and S32 stations close to the coastal areas with relatively high DIN concentration while *Prochlorococcus* dominated with an average 17.4% relative abundance in the middle area with low DIN concentration (**Supplementary Table S1**, *t*-test, P-value < 0.05) (**Figure 2A**). Besides, no obvious difference in microbial composition between the Red Sed and other locations of the Indian Ocean. Notably, the abundance of *Prochlorococcus* was negatively correlated with the abundance of *Synechococcus* across the transect (P-value = $2.2e^{-16}$, Spearman's rank correlation test with the correlation coefficient = −0.88). Although the DIN concentrations of the Indian Ocean surface waters varied among stations, it was much lower than the minimum detection limit and the N/P ratio was less than 8 (**Supplementary Table S1**), indicating that the Indian Ocean surface water belonged to an oligotrophic water body.

The clustering analysis based on physicochemical properties of all samples showed that the 17 samples were divided into three groups, the three Red Sea samples (S34, S36, and S38) were grouped together with the sample S30 from the west of the Indian Ocean, the Red Sea sample S32 was clustered with the samples S18, S22, S24, S26, which were located in the middle of the Indian Ocean, while the other remaining samples were grouped together (**Supplementary Figure S4**). These results demonstrated that the physicochemical properties of the west region of the Red sea was distinguishable from the Indian Ocean. RDA was performed to discern the relationship between microbial community structure and environmental parameters such as concentrations of phosphate, silicate, ammonium, nitrite, nitrate, as well as temperature, salinity and pH (**Supplementary Table S1**). A model describing the environmental parameters that were significantly correlated with microbial composition was selected based on Akaike information criterion, indicating that temperature, phosphate, silicate and pH rather than ammonium, nitrite, nitrate and salinity had significant effect on the microbial composition. The two axes could explain 40.5% of the overall variance in community structure with an adjusted R^2 (0.56), showing a remarkable correlation. The

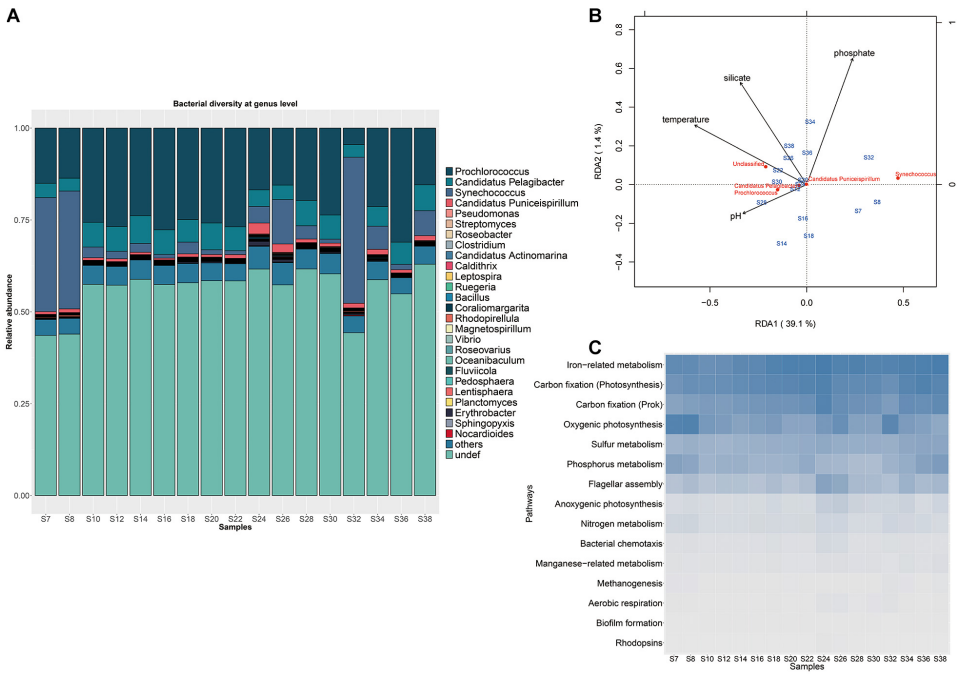


FIGURE 2 | Microbial communities in the surface waters of the Indian Ocean. **(A)** Microbial community structure with abundant *Prochlorococcus* and *Synechococcus*. The samples from stations S7, S8, and S32 near the coast with distinct taxonomic compositions distinct from the other samples. **(B)** Redundancy analysis of bacterial community structure and environmental variables. **(C)** Distributions of typical metabolism pathways among different sampling stations.

parameters of temperature, phosphate, silicate and pH drove the composition variance and separated samples into different groups, of which temperature was the most important and could explain the most variation in response variables (**Figure 2B**). It was found that the three western Red sea samples were grouped in the upper right of the RDA plot, and were mainly associated with concentrations of phosphate and silicate. But most of the Indian Ocean samples (except samples S7 and S8 near the coast) were in the lower left and were associated with temperature and pH. These results indicated that the variation among microbial communities in the Red Sea and in the Indian Ocean were driven by different environmental factors. In addition, the dominant genus *Prochlorococcus* was positively associated with temperature and pH in the multivariate analysis, while the genus *Synechococcus* was negatively associated with pH. Interestingly, the concentrations of ammonium, nitrate and nitrite were not associated with the variation of the whole microbial community. However, the abundance of *Prochlorococcus* presented a remarkably negative correlation with nitrite concentration (P-value < 0.01, Spearman's rank correlation test), but not with the concentrations of ammonium, nitrate, phosphate and other factors, indicating that nitrite impacted *Prochlorococcus* although it had no effect on the whole microbial composition of the Indian Ocean.

Variations of microbial functional composition were examined among stations. Overall, only 24.3% of genes but 82.6% of KOs were co-existed in all samples, indicating a high level of gene diversity and functional redundancy across the Indian Ocean. These shared KOs mainly contributed to metabolic function (carbohydrates, amino acids, nucleotide, cofactors and vitamins, lipid, energy metabolism), genetic information processing (replication and repair), secondary metabolite biosynthesis (glycan, terpenoids, and polyketides), and signal transduction, and the majority of them corresponded to housekeeping functions. The abundances of 15 typical pathways were calculated using a total abundance of marker KOs divided by the number of KOs as described in a previous study (Sunagawa et al., 2015). All samples showed a similar variation pattern in functional composition (**Figure 2C**). Function modules associated with iron-related metabolism, carbon fixation (photosynthesis), carbon fixation (Prok), oxygenic photosynthesis, sulfur and phosphorus metabolism with high abundances co-existed in all samples. In addition, other processes such as nitrogen metabolism, methanogenesis, manganese–related metabolism, flagellar assembly, bacterial chemotaxis, biofilm formation, aerobic respiration and anoxygenic photosynthesis with low abundances were prevalent in each sample.

Prochlorococcus Genomes in the Indian Ocean and the Global Ocean

Prochlorococcus an extensively studied photosynthetic bacterial taxon, dominated the microbial community of the Indian Ocean surface water with an average abundance of 14.5% among metagenome sequences (**Figure 2A**). Using the pangenome of the 201 published *Prochlorococcus* genomes as a reference (**Supplementary Table S3**), about 14.08% of reads could be

mapped to the *Prochlorococcus* gene set, which was very close to the abundance identified in the metagenome data, implying the integrity of the reference data set. Totally, 55 *Prochlorococcus* strains were detected, including 12 cultured strains and 43 uncultured strains. The detections of 55 strains implied the species diversity of *Prochlorococcus* in the surface water of the Indian Ocean. Of these, scB245a_518D8, scB241_528J8, MIT9301, MIT9302, GP2 and MIT9201 were most abundant *Prochlorococcus* strains among the samples (**Supplementary Figure S5**). Only 5 out of 12 cultured strains are within known HLII ecotype. All of 43 uncultured single-cell amplified genome and metagenome-assembled genomes have no ecotype information. To explore the evolutional relationship of these *Prochlorococcus* strains, phylogenetic clades of all these unknown strains (combined with 25 cultured strains with representative ecotypes) were explored using an alternative method based on 40 SCGs. The phylogenetic tree showed that all the identified genomes were divided into two major known ecotypes: HLI and HLII clades (**Figure 3A**), in accordance with their niches. Except for the strain UBA3999, which was finally validated to be affiliated to *Synechococcus* as only 70.81% average nucleotide identity (ANI) to *Prochlorococcus* strain MIT0701 and 30 conserved SCGs consistently annotated to *Synechococcus*, 53 of 54 *Prochlorococcus* strains in the Indian Ocean surface waters belonged to the HLII ecotype and only one strain belonged to the HLI ecotype. The ecotype-level community structure was similar across the Indian Ocean, but the relative abundances of *Prochlorococcus* strains varied among different stations. For example, the composition of *Prochlorococcus* strains in the Red Sea samples (S30, S32, S34, and S38) were clustered together and were more similar to the west locations of the Indian Ocean (**Supplementary Figure S5**).

To further characterize the diversity of *Prochlorococcus* in the global ocean, the 40 SCGs of *Prochlorococcus* genus were fetched from the Tara Ocean gene set, the PE sample assembled gene set and 201 *Prochlorococcus* strains. As the multiple sequence alignment was not suitable for the SCGs from metagenome data, a phylogenetic tree of *Prochlorococcus* was constructed based on one gene, namely COG0172 (*Seryl-tRNA synthetase*) as its highest number in different data sources. This gene is an evolutionarily conserved enzyme which catalyzes the formation of aminoacyl-tRNAs that is used as substrates for ribosomal protein biosynthesis. The *Prochlorococcus* sequences were divided into two main clades, HL and LL (**Figure 3B**). Of these, 179 of 234 sequences were clustered within the HL clade with similar phylogenic distance and 55 of sequences were clustered within LL clades with different branch distance. It suggested that the *Prochlorococcus* sequences in LL clades evolved more divergently, which might be attributed to the specific features of deep-sea environment.

Metabolic Potential for Nitrogen Assimilation

The relative abundance of ammonia input-related genes such as urease, glutamate dehydrogenase (*gdhA*) and ammonia transporter (*amt*) were predominant in each sample, followed by nitrilase, ferredoxin-nitrite reductase (*NirA*), cyanate lyase and formamidase (**Figures 4A,B**). Four metabolic reactions involved in nitrogen metabolism including dissimilatory nitrate reduction, assimilatory nitrate reductase, denitrification and nitrification were detected in our metagenomic dataset (**Figure 4C**). Although the assimilatory nitrate reduction (*narB*, *nasA*, *narsB, and nirA*) presented high abundance in nitrogen metabolism pathway, it contributed only a small portion to ammonia production. The functional genes related to nitrogen fixation and anammox were not detected in all samples (**Figure 4**).

To characterize nitrogen acquisition of the *Prochlorococcus* population in the Indian Ocean, all genes involved in nitrogen metabolism from *Prochlorococcus* genus were extracted from the metagenome. *Prochlorococcus* could potentially assimilate nitrogen for growth through six different routes (**Figure 4**). Three dominant routes were comprised of hydrolyzing urea by ureas, hydrolyzing nitrile to ammonia by nitrilase and converting glutamine into ammonia by *gdhA*. The fourth route was to directly transport extracellular ammonia into intracellular via transporter. The fifth route was to uptake extracellular nitrite and nitrate via transporters, and then catalyze the reduction of nitrite to ammonia by *NirA* and finally be converted into amino acids. The gene encoding nitrate reductase *(NarB*) involved in the reduction of nitrate to nitrite was not identified in the *Prochlorococcus* population. The nitrite could also be converted by oxidation of nitroalkanes using nitronate monooxygenase. The sixth route was to convert cyanate into ammonia by cyanate lyase. Only the metabolic reaction related to assimilatory nitrate reduction was detected in metagenomic *Prochlorococcus* population. The other key enzymes related to dissimilatory nitrate reduction, denitrification, nitrogen fixation, nitrification and anammox were not found (**Figure 4C**). Collectively, the *Prochlorococcus* population in the Indian Ocean could assimilate urea, ammonia, nitrite, nitroalkane, cyanate, and nitrile as nitrogen sources, but they lost the ability to assimilate nitrate and nitrogen due to the lack of *NarB* and *nifH*.

To validate the results from the metagenome, the abilities of nitrogen metabolism from 54 detected *Prochlorococcus* genomes (including six complete genomes) in the Indian Ocean were evaluated. A similar result was observed in those detected *Prochlorococcus* genomes that the enzyme *nifH* involved in nitrogen fixation was absent but the *NarB* catalyzing nitrate to nitrite was found in 14 strain genomes including two cultured strains (SB, MIT0604) and 12 single-cell amplified genomes (**Supplementary Table S5**). These results indicated that most of the *Prochlorococcus* strains lost the ability to reduce nitrate, but some of them still maintained the ability to assimilate nitrate and could utilize it as N source. Furthermore, the enzymes involved in assimilation of other nitrogen forms such as cyanate, nitrile, nitrite, and amino acids also differed among strains, suggesting their different survival strategies.

Amino Acid Biosynthesis of *Prochlorococcus* in the Indian Ocean

The amino acid biosynthesis model was predicted based on the functional enzymes of amino acid biosynthesis and amino acid transporter proteins in the metagenome of *Prochlorococcus*

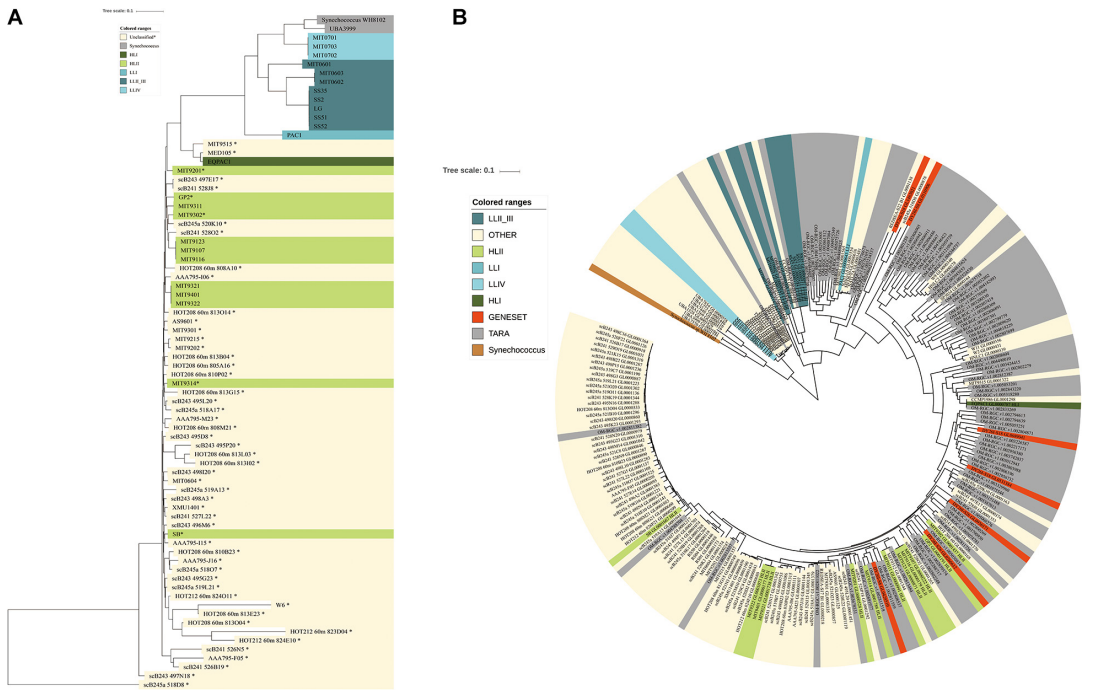


FIGURE 3 | Phylogenetic tree based on single copy genes from the 54 detected strains in the Indian Ocean and *Prochlorococcus* sequences from the global ocean, respectively. **(A)** Phylogenetic tree from the detected strains in our data set (names in bold with*) and representative *Prochlorococcus* genomes was constructed based on 40 SCGs. Phylogenetic clade affiliations are represented by different colors; HLII, light green; HLI, dark green; LLI, median blue; LLII_III, dark blue; LLIV, light blue; *Synechococcus*, gray; unclassified strains, yellow. **(B)** The Phylogenetic tree of *Prochlorococcus* populations in the global ocean based on gene sequences from COG0172. red, sequences identified in assembled gene set; gray, sequences from Tara Ocean gene set; yellow, sequences from *Prochlorococcus* genomes; reference sequences are colored according to their ecotypes.

population. Of the 230 functional KOs involved in amino acid biosynthesis, 96 were found in the *Prochlorococcus* population, which contributed to *de novo* biosynthesis of 14 amino acids such as valine, leucine, isoleucine, tryptophan, proline, threonine, cystenine, asparagine, glutamine, aspartic acid, glutamic acid, lysine, arginine, and histidine. However, reaction gaps for the biosynthesis of amino acids existed in six types of amino acids (**Table 1**). The enzyme *metC* that synthesizes precursor homocysteine of methionine and the *serB* that catalyzes phosphoserine to serine were missing. The enzyme *glyA* that functions bi-directionally transformation of serine and glycine was detected but the bidirection enzyme *glyA* for glycerine and threonine conversion was absent. The enzymes *asdA* and *ALT* for *de novo* biosynthesis of alanine were also absent. Furthermore, biosynthesis reactions of aromatic amino acids tyrosine and phenylalanine were incomplete due to the lack of the enzyme *tyrB* that catalyzes the precursor of tyrosine and converts phenylpyruvate into phenylalanine. Interestingly, the functional genes encoding two types of transporter proteins responsible for importing glycine and alanine into cells were found in the metagenome data of the Prochlorococcus population, suggesting that *Prochlorococcus* absorbed glycine and alanine from the environment. Thus, it could be postulated that serine might be achieved from glycine by *glyA*, and subsequently, converted to cysteine and methionine. The same observation was also found in the genomes of the 54 *Prochlorococcus* strains detected, which further validated the reaction gaps in amino acid biosynthesis of the *Prochlorococcus* population in the Indian Ocean.

DISCUSSION

Metagenomic studies of the global ocean advance our understanding of microbial diversity, community structure, functional potential, the environment influence and biotic interactions (Lima-Mendez et al., 2015; Sunagawa et al., 2015). In this study, we generated a 9.5 M gene set for surface microbes of the Indian Ocean combined with mapped genes from OM-RGC and *de novo* assembled genes from the local samples. Our results showed that most of the microbial genes of surface waters had been embraced by this gene set and could be captured by sample reads. Although the OM-RGC provides an ecosystem-wide data set, which enables ocean microbial genetic diversity accessible for various targeted studies, it does not cover all microbial functional genes owing to the limitation of samples and sampled regions. Therefore, it would be helpful to combine this data with a small set of assembly genes from local samples. Proteobacteria and Cyanobacteria were the abundant phyla in surface waters of the Indian Ocean, and photosynthetic cyanobacterial taxa such as *Prochlorococcus* and *Synechococcus* were dominant in all samples. These results are consistent with the previous studies (Díez et al., 2016; Wang et al., 2016;

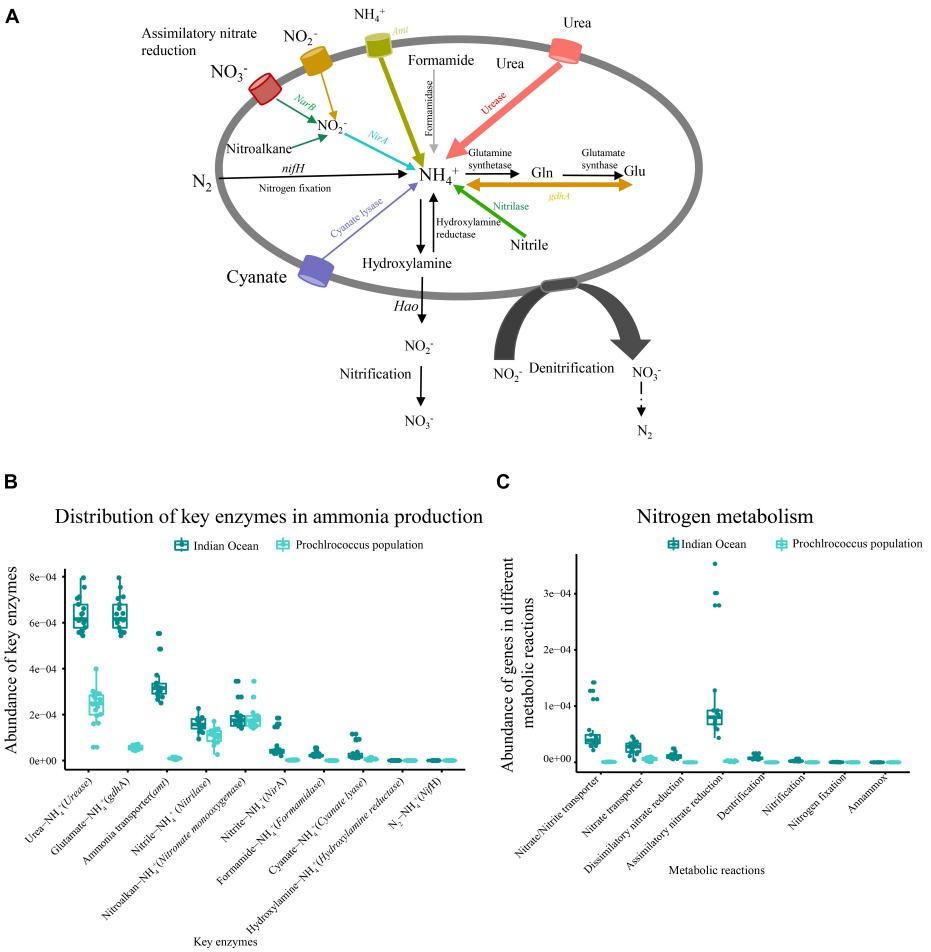


FIGURE 4 | Metabolic potential for N acquisition in the microbial community and metagenomic *Prochlorococcus* population of the Indian Ocean. **(A)** The generic nitrogen assimilation routes (colored line) in the metagenome and *Prochlorococcus* population of the Indian Ocean. **(B)** Abundances of key functional genes involved in ammonia production in the metagenome and *Prochlorococcus* population. **(C)** The functional models of an N metabolism pathway in the metagenome and *Prochlorococcus* population.

Phoma et al., 2018). However, variations at the sub-family level classification were observed among different regions of the Indian Ocean, for example, *Synechococcus* was abundant in the stations near the coast but *Prochlorococcus* dominated in the central area of the Indian Ocean, which might be caused by environmental variables. The effects of environmental factors on marine microbial communities have been widely studied (Coutinho et al., 2015). Composition, distribution and metabolic potential of microbes are structured, to a large extent, by environmental gradients in light, temperature, oxygen, salinity and nutrients (Thompson et al., 2017). Globally, temperature is the key environmental factor driving the geographical distribution of microbes in the ocean (Sunagawa et al., 2015). In this study, temperature, phosphate, silicate and pH were important environmental factors regulating microbial distribution in the Indian Ocean. However, the variation of microbial composition in the Red Sea was mainly associated with the concentrations of phosphate and silicate (**Figure 2B**). It should be noted that the physicochemical properties in the west of the Red Sea differed from the Indian Ocean, resulting in a slight difference in microbial composition. Investigation of microbial adaptation to environmental changes indicated that temperature explains most of the taxonomic and functional variations in the Red Sea, followed by nitrate, chlorophyll, phosphate, and salinity (Thompson et al., 2017). These highly shared KO functions across the samples might confer a buffering capacity for an ecosystem in scenarios of biodiversity loss. The analysis of typical pathways indicated that metabolic processes of microbial community mainly contributed to the production of a diverse array of nutritional substrates such as ferrous/ferric iron, carbohydrate, phosphate, and sulfate. Especially for carbon fixation and oxygenic photosynthesis, they enabled microbes to use organic compounds and molecular oxygen of photosynthetic origin in the oligotrophic oceans. Overall, the microbial gene capacity determines their adaptive ability to environmental variations in the ocean. A comprehensive microbial gene set from various ocean environments would facilitate to fully understand the connection between microbes and environmental variables.

TABLE 1 | Distribution patterns of key enzymes of six amino acids synthesis pathways in *Prochlorococcus* metagenomes from the Indian Ocean.

Amino acids	Key enzymes		Transporters
Met	*metC*	*metH*	–
	–	+	
Ser	*serC*	*serB*	–
	+	–	
Gly	*ltaE*	*glyA*	*betT, betS:*
	–	+	choline/glycine/proline betaine transport protein
Ala	*asdA*	*ALT*	alanine or glycine cation symporter
	–	–	
Tyr	*tyrB*	*tyrC*	–
	–	+	
Phe	*pheA2*	*tyrB*	–
	+	–	

metC, cysteine-S-conjugate beta-lyase; metH, 5-methyltetrahydrofolate–homocysteine methyltransferase; serC, phosphoserine aminotransferase; serB, phosphoserine phosphatase; ltaE, threonine aldolase; glyA, glycine hydroxymethyltransferase; asdA, aspartate 4-decarboxylase; ALT, alanine transaminase; tyrB, aromatic-amino-acid transaminase; tyrC, cyclohexadieny/prephenate dehydrogenase; tyrB, aromatic-amino-acid transaminase; pheA2, prephenate dehydratase. +, present; –, absent.

The picocyanobacterium *Prochlorococcus* was the most abundant photosynthetic phytoplankton in the oligotrophic oceans, and contributes to the stability, resilience and function of marine ecosystem with high species diversity and a wide range of environmental adaptability (Biller et al., 2015; Farrant et al., 2016; Kent et al., 2016). In this study, the major *Prochlorococcus* clades in the Indian Ocean belonged to HLII ecotypes, which is in accordance with the sample origins from surface water with high light intensity. A previous study also showed that most of the Indian Ocean *Prochlorococcus* sequences were distantly related to the HLII *Prochlorococcus marinus* spp. such as MIT9215 and MIT9301, which were also detected in our samples (Díez et al., 2016). Furthermore, HLII bins also dominate the *Prochlorococcus* communities of the surface waters of the Red Sea based on analysis of the *rpoC1* sequences (Shibl et al., 2014). But the HLI, HLII, and HLVI clades of *Prochlorococcus* dominate in the surface water of the North Pacific Ocean by clustering the ITS sequences (Larkin et al., 2016). Phylogenetic analysis of 504 core genes across global ocean samples indicate that the clades HLIII and HLIV significantly dominate in the Equatorial Pacific Ocean, and the clad HLI is abundant in the California Current and South Atlantic Ocean (Kent et al., 2016). Therefore, the ecotype structure of *Prochlorococcus* differs among the oceanic regions. In this study, we constructed a phylogenetic tree based on 40 SCGs. This alternative method has been widely used in the clade classification of bacterial genomes (Hug et al., 2016; Mende et al., 2017). The phylogenetic clades showed the same accuracy as the ITS sequence (Biller et al., 2014) when we examined them using 25 fully sequenced reference genomes with known ecotypes. We also examined the accuracy of COG0172 in distinguishing species difference among 25 known ecotype genomes before exploring the global distribution of *Prochlorococcus*. It was demonstrated to be sufficient to discern the phylogenetic relation among *Prochlorococcus* genomes. The *Prochlorococcus* sequences from the global ocean could be divided into two clades (Moore et al., 1998; Rocap et al., 2002). Our study indicated that the HL clade occupied most of the sequences and showed less divergence, but the LL clades showed more branches with various phylogenetic distance. Similar observations were found in Yan's study (Yan et al., 2018). At this point, most samples are sourced mainly from surface water and less from the deep ocean. The highly different phylogenetic distance of *Prochlorococcus* sequences from the deep ocean may reflect more diversity of *Prochlorococcus* genomes. These results illustrated that there was a considerable diversity of *Prochlorococcus* in the global ocean, but only a small portion can be cultured in the laboratory and a lot of unknown strains requires further excavation.

The RDA analysis showed that the *Prochlorococcus* abundance was positively associated with temperature and pH. Previous studies have shown that the HLI and HLII clades of *Prochlorococcus* present different optimal temperatures, indicating that temperature influences the distribution of *Prochlorococcus* (Zinser et al., 2007; Flombaum et al., 2013). The effect of pH on *Prochlorococcus* growth has been less studied, but it directly impacts the availability of bicarbonate for photosynthetic reaction, which may affect carbon fixation and cell growth. The abundance of *Prochlorococcus* was also significantly yet negatively correlated with nitrite concentration, suggesting the potential role of nitrite in *Prochlorococcus* growth.

The microbial functional genes involved in nitrogen metabolism indicated that bacteria in the oligotrophic ocean could assimilate N through different routes and they preferred urea, ammonia and nitrile, which need low energy to produce ammonia. The bacterial functional gene *nifH*, which is responsible for nitrogen fixation was absent in the surface water of the Indian Ocean and a similar phenomenon was also observed in the west Pacific Ocean (Li Y.-Y. et al., 2018), indicating that bacteria in the Indian Ocean surface waters lacked the ability to fix nitrogen. Ammonia assimilation of *Prochlorococcus* was explored in this study. The *Prochlorococcus* population can assimilate urea, ammonia, nitrite, nitroalkane, cyanate, and nitrile as nitrogen sources for cell growth, indicating that *Prochlorococcus* had evolved diverse adaptive strategies to ambient N deficiency in the oligotrophic Indian Ocean. The abundances of key enzymes involved in DON assimilation were higher than those involved in DIN assimilation in the *Prochlorococcus* population, implying that *Prochlorococcus* might prefer the organic N source in the oligotrophic Indian Ocean. The transcripts of key enzymes involved in the utilization of cyanate, urea and ammonia are also detected in the *Prochlorococcus* genomes under N stress condition (García-Fernández et al., 2004; Kamennaya and Post, 2011). To date, no *Prochlorococcus* isolate is able to utilize molecular N due to the lack of *nifH* genes, which concurs with our findings. The ability of nitrate assimilation is found only in a small portion of the *Prochlorococcus* genomes, such as SB and MIT0604, which is consistent with the findings of a recent study (Berube et al., 2019). They can grow on nitrate as the sole nitrogen source (Martiny et al., 2009; Berube et al., 2015). In our study, the gene encoding *NarB* catalyzing nitrate to nitrite was not detected in

the *Prochlorococcus* population although it was found in 14 of the identified 54 *Prochlorococcus* genomes, indicating the limitation in assembling full sequences from metagenome data and the importance of isolation and identification of single bacterial strain from the natural environment, which would help us to unravel more novel and divergent metabolic pathways. Our study elaborated nitrogen assimilation pathways in the *Prochlorococcus* population and *Prochlorococcus* genomes in the Indian Ocean, and identified the differences in utilization of inorganic and organic nitrogen sources among different strains. This finding enables the potential to discover more unknown and uncultured strains from oceans.

It has been known for decades that most of the free-living bacteria can synthesize 20 kinds of amino acids, but some bacteria still have gaps in their biosynthesis pathways of amino acids (Price et al., 2018). Our study showed that *Prochlorococcus* could not make all amino acids by themselves based on their genetic content and might need to uptake some of them from the ambient environment instead. The gaps in *de novo* synthesis pathways of six different amino acids implied a special model for amino acid utilization in the *Prochlorococcus* population. Moreover, the distribution pattern of key enzymes involved in N and amino acid metabolism pathways in the metagenome of the *Prochlorococcus* population was validated by the *Prochlorococcus* genomes identified, demonstrating the reliability of our results. In general, our findings illustrated the potential capacity of *Prochlorococcus* to assimilate nitrogen in the Indian Ocean, which implied the adaptation of *Prochlorococcus* to ambient N deficiency in the oligotrophic ocean.

CONCLUSION

The oligotrophic Indian Ocean is an understudied realm of the world's oceans. This study comprehensively analyzed the microbial diversity and metabolic potential for N acquisition in the surface waters of the oligotrophic Indian Ocean using a metagenomic approach. Proteobacteria and Cyanobacteria dominated the microbial community but the functional composition of microbes exhibited a high level of gene diversity and functional redundancy across the Indian Ocean from the east to the west. Environmental factors such as temperature, phosphate, silicate, and pH played important roles in regulating microbial distribution in the Indian Ocean. Ammonium was an important nitrogen source for microbial community, while bacterial functional gene *nifH* was absent in the surface water of the Indian Ocean, indicating weak nitrogen fixation in the Indian Ocean and other potential ammonium origins, which needs further study. The predominant cyanobacterial taxa *Prochlorococcus* presented high diversity but a simple ecotype. Moreover, the *Prochlorococcus* evolved diverse adaptive strategies to ambient N deficiency in the oligotrophic ocean. Interestingly, the gaps for specific amino acid biosynthesis pathways existed in *Prochlorococcus*, demonstrating that there could be some alternative ways to acquire some essential amino, and the potential roles of *Prochlorococcus* in the biogeochemical cycle of amino acids. Overall, this study facilitated our understanding of microbes in the oligotrophic Indian Ocean, and serves as an important resource for gene capacity of microbes in future studies.

DATA AVAILABILITY STATEMENT

The 17 fastq files of SE reads were deposited in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb with accession code CNP0000411. The 3 fastq files of PE reads were submitted to the NCBI SRA database with accession number PRJNA450884 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA450884).

AUTHOR CONTRIBUTIONS

D-ZW, Y-YW, TJ, and XL established the concept of the study. YBG, Z-XX, and L-FK collected and processed the samples. Y-YW, S-LL, G-LL, and S-HG performed the bioinformatics analyses. Y-YW, S-LL, and P-FZ performed the data analysis. Y-YW wrote the draft. D-ZW, HL, LG, M-LS, S-LL, S-KS, and G-YF revised and edited the manuscript. All the authors have discussed the results, read and approved the contents of the manuscript.

FUNDING

This work was supported by funding from the National Natural Science Foundation of China (41425021), the International S&T Cooperation (2016YFE0122000), Shenzhen Municipal Government of China (No. JCYJ2015015162041454), the Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011), and the Danish National Research Foundation (DNRF137).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.518865/full#supplementary-material

Supplementary Figure 1 | The workflow of experiment and data analysis. A total of 17 samples were collected in the Indian Ocean and metagenome sequencing was performed on BGISEQ and HiSeq platforms, separately.

Supplementary Figure 2 | Assessment of the efficiency of gene identification between SE reads mapping and PE reads mapping from the same three samples. Green and orange represents the novel genes in OM-RGC mapped by SE and PE reads, respectively. Yellow represents the genes shared by both.

Supplementary Figure 3 | Microbial taxonomic composition and predominant bacterial phylum in the Indian Ocean surface water. **(A)** Taxonomic composition. **(B)** The predominant bacterial phylum. Others represented bacterial abundance lower than 0.1%.

Supplementary Figure 4 | The heat map of physicochemical properties from oceanic samples.

Supplementary Figure 5 | The distribution pattern of the detected *Prochlorococcus* genomes among sampling stations of the Indian Ocean.

REFERENCES

Baer, S. E., Rauschenberg, S., Garcia, C. A., Garcia, N. S., Martiny, A. C., Twining, B. S., et al. (2019). Carbon and nitrogen productivity during spring in the oligotrophic Indian Ocean along the GO-SHIP IO9N transect. *Deep Sea Res. II Top. Stud. Oceanogr.* 161, 81–91. doi: 10.1016/j.dsr2.2018.11.008

Berube, P. M., Biller, S. J., Kent, A. G., Berta-Thompson, J. W., Roggensack, S. E., Roache-Johnson, K. H., et al. (2015). Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J.* 9:1195. doi: 10.1038/ismej.2014.211

Berube, P. M., Rasmussen, A., Braakman, R., Stepanauskas, R., and Chisholm, S. W. (2019). Emergence of trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. *Elife* 8:e41043.

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607

Biller, S. J., Berube, P. M., Berta-Thompson, J. W., Kelly, L., Roggensack, S. E., Awad, L., et al. (2014). Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. *Sci. Data* 1:140034.

Biller, S. J., Berube, P. M., Lindell, D., and Chisholm, S. W. (2015). *Prochlorococcus*: the structure and function of collective diversity. *Nat. Rev. Microbiol.* 13:13. doi: 10.1038/nrmicro3378

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., et al. (2017). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* 7:gix120.

Coutinho, F. H., Meirelles, P. M., Moreira, A. P. B., Paranhos, R. P., Dutilh, B. E., and Thompson, F. L. (2015). Niche distribution and influence of environmental parameters in marine microbial communities: a systematic review. *PeerJ.* 3:e1008. doi: 10.7717/peerj.1008

Díez, B., Nylander, J. A., Ininbergs, K., Dupont, C. L., Allen, A. E., Yooseph, S., et al. (2016). Metagenomic analysis of the Indian ocean picocyanobacterial community: structure, potential function and evolution. *PLoS One* 11:e0155757. doi: 10.1371/journal.pone.0155757

Eakins, B., and Sharman, G. (2010). *Volumes of the World's Oceans from ETOPO1*. Boulder, CO: NOAA National Geophysical Data Center.

Fang, C., Zhong, H., Lin, Y., Chen, B., Han, M., Ren, H., et al. (2017). Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *GigaScience* 7:gix133.

Farrant, G. K., Doré, H., Cornejo-Castillo, F. M., Partensky, F., Ratin, M., Ostrowski, M., et al. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3365–E3374.

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine *Cyanobacteria Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9824–9829.

García-Fernández, J. M., de Marsac, N. T., Diez, J. (2004). Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. *Microbiol. Mol. Biol. Rev.* 68, 630–638. doi: 10.1128/mmbr.68.4.630-638.2004

GB/T 12763.4-2007 (2007). *Specifications for Oceanographic Survey - Part 4: Survey of Chemical Parameters in Seawater*. Available online at: http://www.gb688.cn/bzgk/gb/newGbInfo?hcno=D7C5F44155DBE2C0BC40EA764A6BBF4A (accessed August 13, 2007).

Huerta-Cepas, J., Serra, F., and Bork, P. (2016a). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. doi: 10.1093/molbev/msw046

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016b). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1:16048.

Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107

Joint, I., Doney, S. C., and Karl, D. M. (2011). Will ocean acidification affect marine microbes? *ISME J.* 5, 1–7. doi: 10.1038/ismej.2010.79

Kamennaya, N. A., and Post, A. F. (2011). Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Appl Environ Microbiol.* 77, 291–301.

Kazutaka, K., and Daron, M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kent, A. G., Dupont, C. L., Yooseph, S., and Martiny, A. C. (2016). Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J.* 10, 1856–1865. doi: 10.1038/ismej.2015.265

Kultima, J. R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. R., et al. (2012). MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One* 7:e47656. doi: 10.1371/journal.pone.0047656

Kumar, S. P., Narvekar, J., Nuncio, M., Gauns, M., and Sardesai, S. (2009). "What drives the biological productivity of the northern Indian Ocean?," in . *Indian Ocean Biogeochemical Processes and Ecological Variability*, eds J. D. Wiggert, R. R. Hood, S. W. A. Naqvi, et al. (Washington, DC: American Geophysical Union), 33–56. doi: 10.1029/2008gm000757

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357. doi: 10.1038/nmeth.1923

Larkin, A. A., Blinebry, S. K., Howes, C., Lin, Y., Loftus, S. E., Schmaus, C. A., et al. (2016). Niche partitioning and biogeography of high light adapted *Prochlorococcus* across taxonomicranks in the North Pacific. *ISME J.* 10, 1555–1567. doi: 10.1038/ismej.2015.244

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., et al. (2016). MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11. doi: 10.1016/j.ymeth.2016.02.020

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., et al. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* 32:834.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Li, Y., Jing, H., Xia, X., Cheung, S., Suzuki, K., and Liu, H. (2018). Metagenomic insights into the microbial community and nutrient cycling in the western subarctic Pacific Ocean. *Front. Microbiol.* 9:623. doi: 10.3389/fmicb.2018.00623

Li, Y.-Y., Chen, X.-H., Xie, Z.-X., Li, D.-X., Wu, P.-F., Kong, L.-F., et al. (2018). Bacterial diversity and nitrogen utilization strategies in the upper layer of the northwestern Pacific Ocean. *Front. Microbiol.* 9:797. doi: 10.3389/fmicb.2018.00797

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015). Determinants of community structure in the global plankton interactome. *Science* 348:1262073.

Martiny, A. C., Kathuria, S., and Berube, P. M. (2009). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10787–10792. doi: 10.1073/pnas.0902532106

Massana, R., Pernice, M., Bunge, J. A., Del Campo, J. (2011). Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. *ISME J.* 5, 184–195 doi: 10.1038/ismej.2010.104

Mende, D. R., Bryant, J. A., Aylward, F. O., Eppley, J. M., Nielsen, T., Karl, D. M., et al. (2017). Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* 2:1367. doi: 10.1038/s41564-017-0008-3

Mende, D. R., Sunagawa, S., Zeller, G., Bork, P. (2013). Accurate and universal delineation of prokaryotic species. *Nat. Methods* 10, 881–884 doi: 10.1038/nmeth.2575

Moore, C. M., Mills, M. M., Arrigo, K. R., Berman-Frank, I., Bopp, L., Boyd, P. W., et al. (2013). Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* 6, 701–710

Moore, L. R., Rocap, G., Chisholm, S. W. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* 393, 464–467 doi: 10.1038/30965

Pajares, S., and Ramos, R. (2019). Processes and microorganisms involved in the marine nitrogen cycle: knowledge and gaps. *Front. Mar. Sci.* 6:739. doi: 10.3389/fmars.2019.00739

Phoma, S., Vikram, S., Jansson, J. K., Ansorge, I. J., Cown, D. A., van de Peer, Y., et al. (2018). Agulhas Current properties shape microbial community diversity and potential functionality. *Sci. Rep*. 8:10542.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Price, M. N., Zane, G. M., Kuehl, J. V., Melnyk, R. A., Wall, J. D., Deutschbauer, A. M., et al. (2018). Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. *PLoS Genet.* 14:e1007147. doi: 10.1371/journal.pgen.1007147

Qian, G., Wang, J., Kan, J., Zhang, X., Xia, Z., Zhang, X., et al. (2018). Diversity and distribution of anammox bacteria in water column and sediments of the eastern Indian Ocean. *Int. Biodeterior. Biodegr.* 133, 52–62. doi: 10.1016/j.ibiod.2018.05.015

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490:55.

Rocap, G., Distel, D. L., Waterbury, J. B., and Chisholm, S. W. (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* 68, 1180–1191. doi: 10.1128/aem.68.3.1180-1191.2002

Schwarz, R., and Forchhammer, K. (2005). Acclimation of unicellular cyanobacteria to macronutrient deficiency: emergence of a complex network of cellular reponse. *Microbiology* 151, 2503–2514. doi: 10.1099/mic.0.27883-0

Shibl, A. A., Thompson, L. R., Ngugi, D. K., and Stingl, U. (2014). Distribution of *Prochlorococcus* ecotypes in the red sea basin based on analyses of rpoC1 sequences. 2014. *FEMS Microbiol. Lett.* 356, 118–126.

Sjöstedt, J., Martiny, J. B., Munk, P., and Riemann, L. (2014). Abundance of broad bacterial taxa in the Sargasso Sea explained by environmental conditions but not water mass. *Appl. Environ. Microbiol.* 80, 2786–2795. doi: 10.1128/aem.00099-14

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359.

Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Kultima, J. R., et al. (2013). Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* 10, 1196–1199. doi: 10.1038/nmeth.2693

Thompson, L. R., Williams, G. J., Haroon, M. F., Shibl, A., Larsen, P., Shorenstein, J., et al. (2017). Metagenomic covariation along densely sampled environmental gradients in the Red Sea. *ISME J.* 11:138. doi: 10.1038/ismej.2016.99

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74. doi: 10.1126/science.1093857

Wang, J., Kan, J. J., Borecki, L., Zhang, X. D., Wang, D. X., and Sun, J. (2016). A snapshot on spatial and vertical distribution of bacterial community in the eastern Indian Ocean. *Acta Oceanol. Sin.* 6, 89–93.

Wei, Y., Sun, J., Zhang, X., Wang, J., and Huang, K. (2019). Picophytoplankton size and biomass around equatorial eastern Indian Ocean. *Microbiol. Open* 8:e00629. doi: 10.1002/mbo3.629

Wheeler, P. A., and Kirchman, D. L. (1986). Utilization of inorganic and organic nitrogen by bacteria in marine systems. *Limnol. Oceanogr.* 31, 998–1009. doi: 10.4319/lo.1986.31.5.0998

Williamson, S. J., Allen, L. Z., Lorenzi, H. A., Fadrosh, D. W., Brami, D., Thiagarajan, M., et al. (2012). Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS One* 7:e42047. doi: 10.1371/journal.pone.0042047

Yan, W., Wei, S., Wang, Q., Xiao, X., Zeng, Q., Jiao, N., et al. (2018). Genome rearrangement shapes *Prochlorococcus* ecological adaptation. *Appl. Environ. Microbiol.* 84:e1178-18.

Zehr, A. P., and Wad, B. B. (2002). Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl. Environ. Microbiol.* 68, 1015–1024. doi: 10.1128/aem.68.3.1015-1024.2002

Zinser, E. R., Johnson, Z. I., Coe, A., Karaca, E., Veneziano, D., Chisholm, S. (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol. Oceanogr.* 52, 2205–2220. doi: 10.4319/lo.2007.52.5.2205

Zubkov, M. V., Fuchs, B. M., Tarran, G. A., Burkill, P. H., and Amann, R. (2003). High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl. Environ. Microbiol.* 69, 1299–1304. doi: 10.1128/aem.69.2.1299-1304.2003

Conflict of Interest: Y-YW, S-LL, G-LL, TJ, HL, G-YF, S-SL, S-KS, S-HG, P-FZ, and XL was employed by the institute of BGI-Shenzhen.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 May 2021
doi: 10.3389/fmicb.2021.658803

Spatial Distribution of Arctic Bacterioplankton Abundance Is Linked to Distinct Water Masses and Summertime Phytoplankton Bloom Dynamics (Fram Strait, 79°N)

Magda G. Cardozo-Mino[1,2†], Eduard Fadeev[1,2†‡], Verena Salman-Carvalho[1,2‡] and Antje Boetius[1,2,3]*

[1] Max Planck Institute for Marine Microbiology, Bremen, Germany, [2] Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany, [3] MARUM, University of Bremen, Bremen, Germany

OPEN ACCESS

Edited by:
Gipsi Lima Mendez, Catholic University of Louvain, Belgium

Reviewed by:
Connie Lovejoy, Laval University, Canada
Klaus Jürgens, Leibniz Institute for Baltic Sea Research (LG), Germany

***Correspondence:**
Magda G. Cardozo-Mino
mcardozo@mpi-bremen.de

[†] These authors have contributed equally to this work

[‡] **Present address:**
Eduard Fadeev, Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria
Verena Salman-Carvalho, Microbiology Department, University of Massachusetts Amherst, Amherst, MA, United States

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: 26 January 2021
Accepted: 12 April 2021
Published: 10 May 2021

Citation:
Cardozo-Mino MG, Fadeev E, Salman-Carvalho V and Boetius A (2021) Spatial Distribution of Arctic Bacterioplankton Abundance Is Linked to Distinct Water Masses and Summertime Phytoplankton Bloom Dynamics (Fram Strait, 79°N). Front. Microbiol. 12:658803. doi: 10.3389/fmicb.2021.658803

The Arctic is impacted by climate warming faster than any other oceanic region on Earth. Assessing the baseline of microbial communities in this rapidly changing ecosystem is vital for understanding the implications of ocean warming and sea ice retreat on ecosystem functioning. Using CARD-FISH and semi-automated counting, we quantified 14 ecologically relevant taxonomic groups of bacterioplankton (*Bacteria* and *Archaea*) from surface (0–30 m) down to deep waters (2,500 m) in summer ice-covered and ice-free regions of the Fram Strait, the main gateway for Atlantic inflow into the Arctic Ocean. Cell abundances of the bacterioplankton communities in surface waters varied from 10^5 cells mL^{-1} in ice-covered regions to 10^6 cells mL^{-1} in the ice-free regions. Observations suggest that these were overall driven by variations in phytoplankton bloom conditions across the Strait. The bacterial groups *Bacteroidetes* and *Gammaproteobacteria* showed several-fold higher cell abundances under late phytoplankton bloom conditions of the ice-free regions. Other taxonomic groups, such as the *Rhodobacteraceae*, revealed a distinct association of cell abundances with the surface Atlantic waters. With increasing depth (>500 m), the total cell abundances of the bacterioplankton communities decreased by up to two orders of magnitude, while largely unknown taxonomic groups (e.g., SAR324 and SAR202 clades) maintained constant cell abundances throughout the entire water column (ca. 10^3 cells mL^{-1}). This suggests that these enigmatic groups may occupy a specific ecological niche in the entire water column. Our results provide the first quantitative spatial variations assessment of bacterioplankton in the summer ice-covered and ice-free Arctic water column, and suggest that further shift toward ice-free Arctic summers with longer phytoplankton blooms can lead to major changes in the associated standing stock of the bacterioplankton communities.

Keywords: Arctic Ocean, Fram Strait, bacterioplankton, CARD-FISH, water column

INTRODUCTION

Atmospheric and oceanic warming has a substantial impact on the Arctic Ocean already today (Dobricic et al., 2016; Sun et al., 2016; Dai et al., 2019). The strong decline in sea ice coverage (Peng and Meier, 2018; Dai et al., 2019) and heat transfer by the Atlantic water inflow (Beszczynska-Möller et al., 2012; Rudels et al., 2012; Walczowski et al., 2017) will affect stratification of the water column and can lead to an increase in upward mixing of the Atlantic core water, a process also termed "Atlantification" (Polyakov et al., 2017). The main inflow of Atlantic water into the Arctic Ocean occurs through the Fram Strait (Beszczynska-Möller et al., 2011), making it a sentinel region for observing the ongoing changes in the Arctic marine ecosystem (Soltwedel et al., 2005, 2016). The Fram Strait is also the main deep-water gateway between the Atlantic and the Arctic Ocean. It hosts two distinct hydrographic regimes; the West Spitsbergen Current (WSC) that carries relatively warm and saline Atlantic water northward along the Svalbard shelf (Beszczynska-Möller et al., 2012; von Appen et al., 2015), and the East Greenland Current (EGC) that transports cold polar water and sea ice southwards from the Arctic Ocean along the ice-covered Greenland shelf (de Steur et al., 2009; Wekerle et al., 2017).

Sea ice conditions have a strong impact on the seasonal ecological dynamics in Fram Strait and the whole Arctic Ocean (Wassmann and Reigstad, 2011), affecting light availability and stratification in the water column. The presence of sea ice and snow cover can suppress the seasonal phytoplankton bloom in the water column through light limitation (Mundy et al., 2005; Leu et al., 2011), or change its timing, e.g., by increasing stratification of the surface waters once the ice melts (Korhonen et al., 2013). Also, sea-ice algae can make up a significant proportion of the annual productivity (Leu et al., 2011; Boetius et al., 2013; Fernández-Méndez et al., 2014). Previous summer observations in the Fram Strait already suggested that total cell abundances and productivity of bacterioplankton communities in surface waters are driven by environmental parameters associated with phytoplankton bloom dynamics (Fadeev et al., 2018), such as the availability and composition of organic matter (Piontek et al., 2015; Engel et al., 2019), with differences between ice-covered and ice-free regions (Piontek et al., 2014; Fadeev et al., 2018).

Long-term summer observations in the region, conducted in the framework of the Long-Term Ecological Research site HAUSGARTEN, revealed strong ecological variations associated with the Atlantic Meridional Overturning Circulation (Soltwedel et al., 2016). Warming events during the past decades influenced seasonal phytoplankton blooms by causing a slow but continuous increase in biomass, and a shift from diatom- to flagellate-dominated communities (Nöthig et al., 2015; Engel et al., 2017; Basedow et al., 2018). It has been recently observed that phytoplankton blooms show an increasing partitioning of the produced organic carbon into the dissolved phase (Engel et al., 2019), which may result in a more active microbial loop in the upper ocean and less export of particulate matter (Vernet et al., 2017; Fadeev et al., 2020). In times of a rapidly changing Arctic ecosystem, investigating structure and dynamics of bacterioplankton communities remains a key component to the understanding of current changes in this environment. However, so far, an assessment of associated responses of the key bacterial taxa responsible for an increased recycling is missing, especially with regard to shifts in standing stocks.

To date, the majority of Arctic bacterioplankton studies are performed using high-throughput sequencing of the 16S rRNA gene, which cannot be directly converted to absolute standing stock abundances of specific taxonomic groups due to polymerase chain reaction (PCR) primers selection (Fadeev et al., 2021), as well as other quantitative biases (Gloor et al., 2017; Kumar et al., 2017; Piwosz et al., 2020). Here we used semi-automatic CAtalyzed Reporter Deposition-Fluorescence *In Situ* Hybridization (CARD-FISH; Pernthaler et al., 2002). The power of this technique lies in the ability to acquire absolute abundance of the targeted taxonomic groups free of compositional effect (Amann et al., 1990). Besides the ability to target and quantify specific taxonomic groups, the retrieval of a positive hybridization signal furthermore indicates that the analyzed cell was alive and active before fixation (Amann et al., 1990; DeLong et al., 1999). Automatization of the microscopic examination and counting procedure can reach a high-throughput standard (Schattenhofer et al., 2009; Teeling et al., 2012; Bižić-Ionescu et al., 2015; Bennke et al., 2016).

Using CARD-FISH and semi-automated cell counting, we quantified bacterial and archaeal cell abundances of 12 taxonomic groups selected based on a 16S rRNA gene survey of water column microbial communities during the same summer expedition in the Fram Strait (Fadeev et al., 2020) (**Supplementary Table 1**). Samples were collected from 11 stations at four different depths, targeting previously defined layers of the water column in the Fram Strait (Rudels et al., 2012): surface mixed layer (0–30 m; seasonally mixed layer of Atlantic and Arctic waters), epipelagic (100 m; mainly modified Atlantic water), deep mesopelagic (500–1,000 m; intermediate water), and bathypelagic (1,200–2,500 m; Eurasian Basin deep waters; **Table 1**). The main objective of this study was to assess the standing stocks of key taxonomic groups in the summer bacterioplankton across the Fram Strait. Using high-throughput cell counts data of bacterioplankton cell abundances we tested the following hypotheses: (1) in surface waters, the abundances of different bacterioplankton taxonomic groups are associated with phytoplankton bloom conditions, and, are linked to the abundances of specific phytoplankton populations; (2) water depth structures the bacterioplankton communities, and (3) differences between communities in ice-covered and ice-free regions decrease with increasing water depth.

RESULTS AND DISCUSSION

Hydrographic and Biogeochemical Conditions Across the Fram Strait

Based on the known hydrography of the Strait (Rudels et al., 2012) and the observed sea-ice conditions, we sampled three distinct regions of the Fram Strait (**Figure 1**): the ice-free eastern part of the Strait ("HG" stations) associated with the WSC

(Beszczynska-Möller et al., 2012), the ice-covered western part of the Strait ("EG" stations) associated with the EGC (de Steur et al., 2009), and the partially ice-covered north-eastern part of the Strait ("N" stations) that represents a highly productive ice-margin zone (Hebbeln and Wefer, 1991; Perrette et al., 2011).

At the time of sampling in June–July 2016, the low level of inorganic nutrients above the seasonal pycnocline, and the chlorophyll *a* concentrations, suggested a late stage of the phytoplankton bloom across the Strait (**Table 1**). Microscopic analyses of phyto- and protozooplankton communities previously conducted in representative stations of each region (LTER HAUSGARTEN stations EG1, EG4, N5, N4, HG4, and S3) at the chlorophyll *a* maximum 10–28 m depth), revealed that the communities at the time of sampling in the ice-covered EG and the ice-margin N stations had a higher abundance of diatoms, in contrast to the ice-free HG stations that had a higher abundance of *Phaeocystis* spp., (Fadeev et al., 2020). These locally defined conditions correspond to an interannual trend of distinct phytoplankton bloom conditions observed in the western ice-covered EGC and the eastern ice-free WSC (Nöthig et al., 2015; Fadeev et al., 2018).

Surface Water Bacterioplankton Communities Are Affected by Distinct Phytoplankton Bloom Conditions

Phytoplankton blooms in surface waters generally lead to an increased cell abundance of heterotrophic bacteria that are specialized in the degradation of organic matter from algal exudates and phytodetritus (Buchan et al., 2014; Teeling et al., 2016). Previous observations in Fram Strait revealed a strong influence of the summer phytoplankton bloom conditions on the composition and structure of bacterioplankton communities (Wilson et al., 2017; Müller et al., 2018), differing also between the ice-covered and ice-free regions of the Strait (Fadeev et al., 2018). We observed significantly higher total cell abundances of the bacterioplankton (i.e., all DAPI-stained bacterial and archaeal cells) in the surface water of the HG and N stations (6–17 $\times$ 10^5 cells mL^{-1}; **Supplementary Table 2**), as compared to the EG stations (3 $\times$ 10^5 cells mL^{-1}; Kruskal–Wallis test; $\chi^2 = 81.85$, $df = 2$, p-value < 0.01). The communities were dominated by bacterial cells that comprised 8–11 $\times$ 10^5 cells mL^{-1} in the HG and N stations, and 2 $\times$ 10^5 cells mL^{-1} in the EG stations (**Figure 2** and **Supplementary Table 2**). The bacterial communities exhibited high abundance of the classes *Bacteroidetes* (2.1 $\times$ 10^5 cells mL^{-1}) in the HG and N stations, followed by *Gammaproteobacteria* (from 1.6 to 2.1 $\times$ 10^5 cells mL^{-1}) and *Verrucomicrobia* (from 1.6 to 2.1 $\times$ 10^5 cells mL^{-1}), with a several-fold higher cell abundance, compared to the EG stations (where together they comprised between 0.1 and 0.3 $\times$ 10^5 cells mL^{-1}) (**Figure 3** and **Supplementary Table 3**). These taxonomic groups were previously suggested to be associated with the seasonal phytoplankton blooms in the region (Wilson et al., 2017; Fadeev et al., 2018). Previous molecular studies also have shown that various taxonomic groups had higher sequence proportion in surface waters of ice-covered, compared to ice-free, regions of the Fram Strait, and are likely associated with Arctic water masses and winter communities in the Fram Strait (Wilson et al., 2017; Fadeev et al., 2018, 2020; Müller et al., 2018). Our microscopy data showed that while *Thaumarchaeota* and the SAR202 clade had only little variations between the different regions, the class *Deltaproteobacteria* and the SAR324 clade exhibited much higher cell abundances in the ice-covered EG stations, as compared to the ice-free HG and ice-margin N stations (**Figure 4** and **Supplementary Table 3**). Hence, the observed patterns in surface water bacterioplankton communities seem to be driven by differences in environmental conditions across the Fram Strait.

In our study, the relative abundance of *Bacteroidetes, Gammaproteobacteria*, and *Verrucomicrobia* were consistent with in 16S rRNA gene observations of size-fractionated bacterioplankton communities (i.e., free-living and particle-associated) conducted during the same expedition (>75%) (Fadeev et al., 2020). However, other taxonomic groups (e.g., *Alteromonadaceae*) showed two- to threefold lower relative abundance in the molecular study (**Supplementary Figure 1**). These discrepancies can be explained by previously conducted direct methodological comparison between 16S rRNA gene observations and CARD-FISH counts (Fadeev et al., 2021), which suggested potential over-representation of the SAR11 clade in the microscopy counts that could affect the proportional representation of other taxonomic groups in the dataset. Alternatively, the potentially higher cellular activity (and thus higher ribosomal content) of phytoplankton bloom-associated taxonomic groups (e.g., *Bacteroidetes*) may have altered their representation in the PCR-based 16S rRNA gene dataset (Rosselli et al., 2016), and thus potentially lower sequence proportion of other taxonomic groups. The methodology applied in this study avoids this compositionality effect and allows for the direct determination of absolute cell abundances of each targeted taxonomic group.

To test the hypothesis that environmental conditions across the Strait shape bacterioplankton communities, we examined a set of key physicochemical environmental parameters that represent the distinct water masses (temperature and salinity) and the different phytoplankton bloom conditions (categorized based on chlorophyll *a* concentration and consumed inorganic nutrients) across the Strait. We did not find significant correlations between these physical and biogeochemical parameters, which suggest to some extent their independent effect on the bacterioplankton communities (**Supplementary Figure 2**). Based on this assumption we conducted specific correlation tests between each of these environmental parameters and cell abundances of various taxonomic groups (**Supplementary Table 4**). Cell abundances of *Verrucomicrobia* and its order *Opitutales*, as well as of the SAR11 clade and the family *Rhodobacteraceae* (both members of the class *Alphaproteobacteria*), showed significant positive correlations to water temperature (Pearson's correlation; $r > 0.5$, p-value < 0.05; **Supplementary Table 4**), suggesting an association with the warmer Atlantic waters of the eastern Fram Strait. The *Verrucomicrobia* has been previously shown to be a major polysaccharide-degrading bacterial taxonomic group in the north-western Svalbard fjord Smeerenburgfjord

TABLE 1 | Environmental parameters measured at different stations and microscopy counts (cells mL^{-1}) of diatoms and *Phaeocystis* spp.

Region	Station	PANGAEA Event ID	Water layer	Lat (°N)	Lon (°E)	Depth (m)	Temp (°C)	Sal	Chl a ($\mu g\ L^{-1}$)	NO_3 ($\mu mol\ L^{-1}$)	ΔNO_3 ($\mu mol\ L^{-1}$)	PO_4 ($\mu mol\ L^{-1}$)	ΔPO_4 ($\mu mol\ L^{-1}$)	N:P	NH_4 ($\mu mol\ L^{-1}$)	ΔNH_4 ($\mu mol\ L^{-1}$)	SiO_3 ($\mu mol\ L^{-1}$)	ΔSiO_3 ($\mu mol\ L^{-1}$)	Diatoms (pennate and centric)	*Phaeocystis* spp.
EGC	EG1	PS99/051-2	SUR	78.99	−5.42	13	−1.51	33.17	1.66	3.75 ± 0.01	3.55 ± 0.01	0.46 ± 0.00	0.13 ± 0.00	8.15	0 ± 0.00	0.00	4.50 ± 0.00	0.14 ± 0.02	148	163
EGC	EG1	PS99/051-2	EPI	78.99	−5.42	100	−1.11	34.23		8.44 ± 0.04		0.63 ± 0.00		13.40	0 ± 0.01		4.04 ± 0.00			
EGC	EG1	PS99/051-2	MES	78.99	−5.42	971	−0.14	34.89		12.62 ± 0.03		0.85 ± 0.00		14.85	0 ± 0.03		7.00 ± 0.04			
EGC	EG4	PS99/048-11	SUR	78.82	−2.73	24	−1.17	34.48	1.52	4.92 ± 0.02	6.34 ± 0.01	0.50 ± 0.00	0.35 ± 0.00	9.84	0 ± 0.02	0.08 ± 0.01	2.44 ± 0.14	2.38 ± 0.34	65	13178
EGC	EG4	PS99/048-11	EPI	78.82	−2.73	100	3.86	35.05		10.97 ± 0.07		0.81 ± 0.00		13.54	0 ± 0.02		4.83 ± 0.17			
EGC	EG4	PS99/048-1	MES	78.82	−2.73	1000	−0.24	34.91		13.77 ± 0.04		0.94 ± 0.02		14.65	0 ± 0.01		7.10 ± 0.01			
EGC	EG4	PS99/048-1	BAT	78.82	−2.73	2527	−0.76	34.91		14.97 ± 0.09		1.06 ± 0.01		14.12	0 ± 0.01		12.07 ± 0.19			
N	N3	PS99/054-1	SUR	79.58	5.17	34	3.28	34.40	0.95	8.09 ± 0.06	5.23 ± 0.01	0.72 ± 0.00	0.31 ± 0.00	11.24	1.29 ± 0.00	0.00	3.19 ± 0.05	1.02 ± 0.03		
N	N3	PS99/054-1	EPI	79.58	5.17	100	4.25	35.10		10.11 ± 0.04		0.78 ± 0.00		12.96	0.47 ± 0.01		3.29 ± 0.04			
N	N3	PS99/054-1	MES	79.58	5.17	1000	−0.33	34.10		13.51 ± 0.02		1.05 ± 0.00		12.87	0 ± 0.00		7.25 ± 0.00			
N	N3	PS99/054-1	BAT	79.58	5.17	2500	−0.72	34.92		14.81 ± 0.05		1.09 ± 0.00		13.59	0 ± 0.00		10.74 ± 0.03			
N	N4	PS99/055-1	SUR	79.74	4.51	22	2.66	33.99	2.21	3.33 ± 0.01	6.80 ± 0.01	0.53 ± 0.01	0.49 ± 0.01	6.28	0.29 ± 0.03	0.00	1.65 ± 0.00	2.06 ± 0.01	119	4047
N	N4	PS99/055-1	EPI	79.74	4.51	100	3.94	35.08		10.64 ± 0.00		1.00 ± 0.03		10.64	0.42 ± 0.01		3.90 ± 0.00			
N	N4	PS99/055-7	MES	79.74	4.51	1000	−0.41	34.91		13.96 ± 0.08		0.92 ± 0.00		15.17	0 ± 0.01		9.07 ± 0.19			
N	N4	PS99/055-7	BAT	79.74	4.51	2500	−0.74	34.92		14.47 ± 0.02		0.86 ± 0.00		16.83	0 ± 0.00		11.24 ± 0.14			
N	N5	PS99/053-2	SUR	79.92	3.06	19	0.75	33.59	7.40	0.97 ± 0.01	8.21 ± 0.01	0.51 ± 0.04	0.66 ± 0.02	1.90	0 ± 0.04	1.56 ± 0.01	0.71 ± 0.01	2.34 ± 0.04	14	9401
N	N5	PS99/053-2	EPI	79.92	3.06	100	4.27	35.10		10.05 ± 0.05		1.10 ± 0.06		9.14	0.42 ± 0.01		3.22 0.08			
N	N5	PS99/053-2	MES	79.92	3.06	1000	−0.23	34.91		13.04 ± 0.12		1.34 ± 0.01		9.73	0 ± 0.00		6.50 ± 0.02			
N	N5	PS99/053-2	BAT	79.92	3.06	2427	−0.74	34.92		14.29 ± 0.04		1.59 ± 0.00		8.99	0 ± 0.01		10.79 ± 0.01			
WSC	HG1	PS99/066-2	SUR	79.14	6.09	17	6.27	35.33	3.42											
WSC	HG1	PS99/066-5	EPI	79.14	6.09	100	4.38	35.09												
WSC	HG1	PS99/066-5	MES	79.14	6.09	500	1.46	34.95												
WSC	HG1	PS99/066-5	BAT	79.14	6.09	1253	−0.81	34.91												
WSC	HG2	PS99/057-1	SUR	79.13	4.91	22	2.30	34.90	2.23	6.15 ± 0.01	5.30 ± 0.01	0.89 ± 0.00	0.55 ± 0.02	6.91	0.94 ± 0.00	0.00	2.74 ± 0.02	1.36 ± 0.05		
WSC	HG2	PS99/057-1	EPI	79.13	4.91	100	3.60	35.04		10.84 ± 0.01		1.30 ± 0.01		8.34	0 ± 0.01		3.88 ± 0.11			
WSC	HG2	PS99/057-1	MES	79.13	4.91	1000	−0.58	34.91		14.15 ± 0.09		1.69 ± 0.01		8.37	0 ± 0.01		9.33 ± 0.06			
WSC	HG2	PS99/057-1	BAT	79.13	4.91	1492	−0.81	34.91		14.95 ± 0.02		1.78 ± 0.07		8.40	0 ± 0.00		12.35 ± 0.02			
WSC	HG4	PS99/042-11	SUR	79.07	4.19	28	0.41	34.44	3.54	5.79 ± 0.04	6.42 ± 0.04	0.66 ± 0.02	0.41 ± 0.02	8.77	0.56 ± 0.01	0.17 ± 0.01	2.66 ± 0.02	2.11 ± 0.03	29	4007
WSC	HG4	PS99/042-11	EPI	79.07	4.19	100	3.51	35.04		10.88 ± 0.07		0.87 ± 0.06		12.51	0 ± 0.01		4.16 ± 0.12			
WSC	HG4	PS99/042-1	MES	79.06	4.19	1000	−0.35	34.91		13.71 ± 0.04		0.96 ± 0.01		14.28	0 ± 0.01		7.70 ± 0.02			
WSC	HG4	PS99/042-1	BAT	79.06	4.19	2462	−0.73	34.92		14.67 ± 0.05		1.01 ± 0.00		14.52	0 ± 0.00		11.59 ± 0.03			
WSC	HG5	PS99/044-1	SUR	79.07	3.66	25	1.74	32.363	4.24	5.1 ± 0.03	7.08 ± 0.02	0.53 ± 0.00	0.37 ± 0.00	9.62	0.50 ± 0.02	0.00	3.19 ± 0.01	1.96 ± 0.06		
WSC	HG5	PS99/044-1	EPI	79.07	3.66	100	3.94	35.089		11.26 ± 0.03		0.80 ± 0.00		14.08	0 ± 0.01		4.75 ± 0.03			
WSC	HG5	PS99/044-1	MES	79.07	3.66	2600	−0.73	34.924												
WSC	HG5	PS99/044-1	BAT	79.07	3.66	3038	−0.70	34.925												
WSC	HG7	PS99/046-1	SUR	79.05	3.53	35	3.73	33.957	4.46	5.24 ± 0.01	2.61 ± 0.05	0.51 ± 0.00	0.09 ± 0.00	10.27	0.48 ± 0.00	0.00	2.84 ± 0.20	1.68 ± 0.08		
WSC	HG7	PS99/046-1	EPI	79.05	3.53	100	3.35	35.016		10.35 ± 0.04		0.74 ± 0.00		13.99	0.09 ± 0.00		4.21 ± 0.02			
WSC	HG7	PS99/046-1	MES	79.05	3.53	1005	−0.28	34.908												
WSC	HG7	PS99/046-1	BAT	79.05	3.53	3772	−0.63	34.924												
WSC	HG9	PS99/059-2	SUR	79.13	2.84	24	−1.24	35.089	1.90	1.74 ± 0.04	6.58 ± 0.01	0.67 ± 0.00	0.58 ± 0.13	2.60	0 ± 0.02	0.73 ± 0.21	1.72 ± 0.02	1.96 ± 0.01		
WSC	HG9	PS99/059-2	EPI	79.13	2.84	100	3.92	35.047		10.97 ± 0.01		1.12 ± 0.05		9.79	0.47 ± 0.28		3.70 ± 0.03			
WSC	HG9	PS99/059-2	MES	79.13	2.84	1000	−0.19	34.897		13.59 ± 0.02		1.17 ± 0.17		11.62	0.00 ± 0.17		6.82 ± 0.01			
WSC	HG9	PS99/059-2	BAT	79.13	2.84	2499	−0.72	34.919		15.23 ± 0.01		1.27 ± 0.03		11.99	0.00 ± 0.00		11.11 ± 0.03			

Chl. a, chlorophyll a; EGC, East Greenland current region; N, North region; WSC, West Spitsbergen current region; SUR, surface mixed water; EPI, epipelagic; MES, mesopelagic; BAT, bathypelagic zone; Lat, latitude; Lon, longitude; Temp, temperature; Sal, salinity.

Diatoms and Phaeocystis spp.

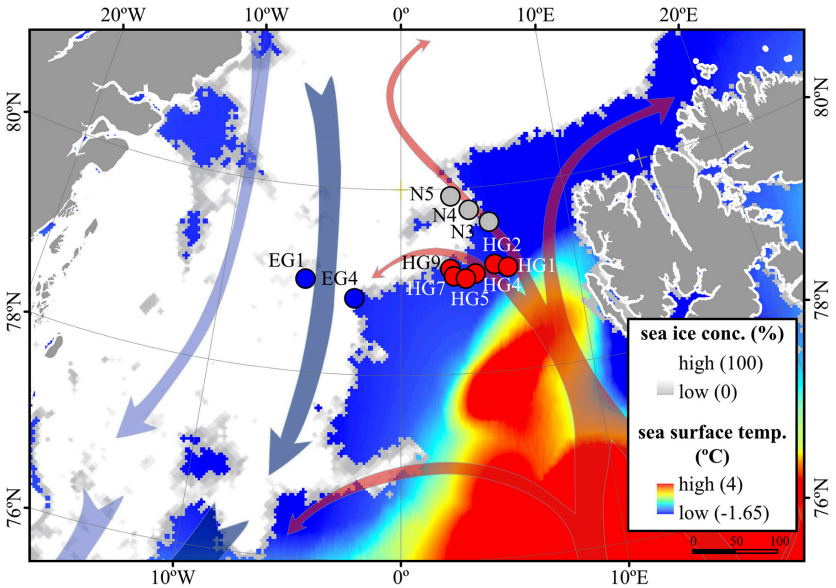


FIGURE 1 | Oceanographic overview of the Fram Strait, including the monthly mean of sea-ice cover and sea surface temperature during July 2016. The sea ice concentration is represented by inverted grayscale (gray = low, white = high). Arrows represent general directions of the West Spitsbergen Current (WSC) (in red) and the East Greenland Current (EGC) (in blue). Stations of water column sampling are indicated and colored according to their sea-ice conditions: ice-covered EGC (EG) stations–blue, ice-margin N stations–gray, ice-free WSC (HG) stations–red. The map was modified from (Fadeev et al., 2020).

(Cardman et al., 2014), and therefore may also be associated with the outflow from the Svalbard fjords (e.g., Kongsfjord) into the Atlantic waters of the WSC (Cottier et al., 2005) sampled for this study. The SAR11 clade and the *Rhodobacteraceae* have both been previously shown to correlate with temperature at high latitudes (Giebel et al., 2011; Tada et al., 2013), and are known to have distinct phylotypes in water masses with different temperatures (Selje et al., 2004; Sperling et al., 2012; Giovannoni, 2017). However, the *Rhodobacteraceae* are also known for their broad abilities in utilizing organic compounds (Buchan et al., 2014; Luo and Moran, 2014). Thus, one cannot rule out that their higher cell abundances in warmer waters of the HG and N stations are associated with the late stage of the phytoplankton bloom and their exudates. In addition, the SAR324 clade (*Deltaproteobacteria*) showed strong positive correlation with statistical significance to salinity (Pearson's correlation; $r > 0.5$, p-value < 0.05; **Supplementary Table 4**). During the summer, with increased melting of sea ice, a low-salinity water layer is formed in surface waters, and the strong stratification of this water layer enhances the development of the phytoplankton bloom (Fadeev et al., 2018). Consequently, the correlation of SAR324 with higher salinity suggests that their cell abundances are lower in surface waters where, in turn, we observe a strong phytoplankton bloom (e.g., in WSC).

The distinct surface water masses in the region differ not only in their physical but also in their biogeochemical characteristics (Wilson and Wallace, 1990; Fadeev et al., 2018), with higher concentrations of inorganic nitrogen and phosphate in the Atlantic, compared to the Arctic water masses. At the time of sampling, the typical Redfield ratio between inorganic nitrogen (mainly nitrate NO_3) and inorganic phosphate (PO_4) was below 16 (Redfield, 1963; Goldman et al., 1979). This suggests that surface waters across all three regions were nitrogen limited (**Table 1**) due to the progressing phytoplankton growth (Nöthig et al., 2015). In order to disentangle the effect of biological consumption of nutrients from water mass-specific nutrient signatures, we calculated the seasonal net consumption of inorganic nutrients, as the proxy for phytoplankton bloom conditions (**Table 1**). Consumed nitrate (ΔNO_3) and phosphate (ΔPO_4) revealed a very strong positive correlation with statistical significance (Pearson's correlation; $r = 0.86$, p-value < 0.05; **Supplementary Table 4**). The consumed silica (ΔSiO_3), used by diatoms, did not show a significant correlation to ΔPO_4 and ΔNO_3. This further supports the impact of different phytoplankton populations across the Strait (i.e., diatoms vs. *Phaeocystis*; Fadeev et al., 2020). Phytoplankton bloom-associated environmental parameters (chlorophyll *a* concentration and the consumed inorganic nutrients) revealed weaker relationships with cell abundances of different taxonomic groups (**Supplementary Table 4**). Furthermore, we did not observe significant positive correlations of the cell abundances of diatoms or *Phaeocystis* spp., with the quantified bacterioplankton taxa. This might be explained by time lags and local differences in the dynamic development of phytoplankton blooms across the entire Strait (Wilson et al., 2017; Fadeev et al., 2018).

Bacterioplankton Communities Strongly Change in Cell Abundance and Composition With Depth

The complexity of Fram Strait surface waters with different ice-coverages, a dynamic ice-melt water layer and mesoscale

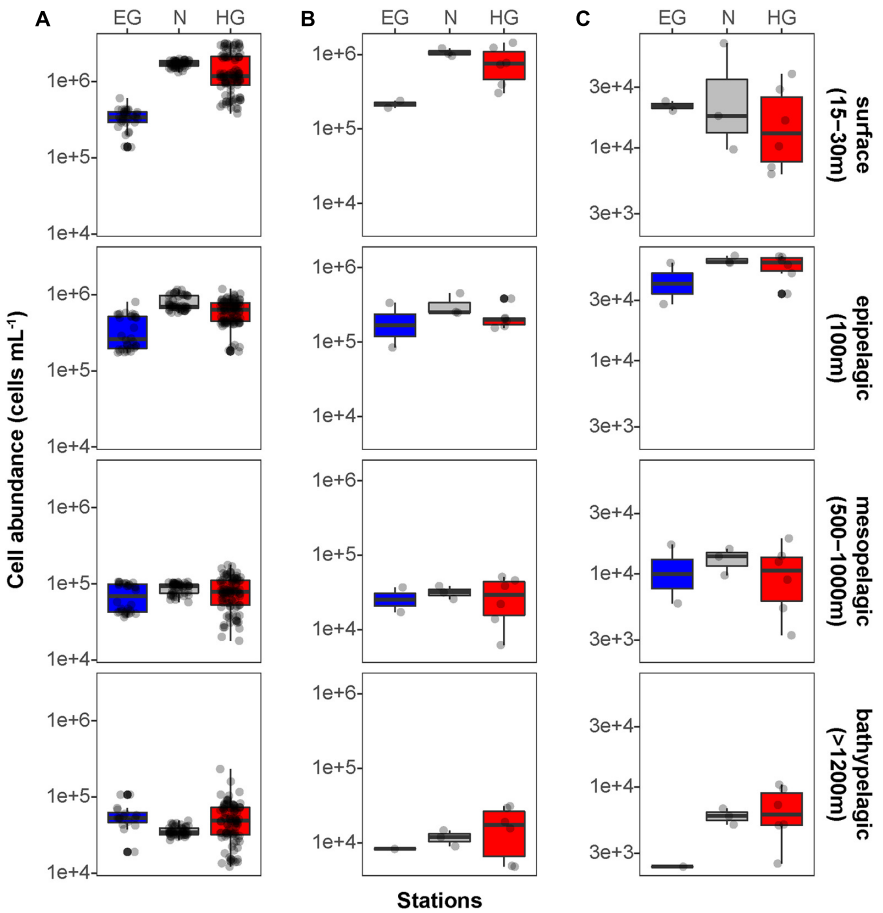


FIGURE 2 | Mean bacterioplankton cell abundances calculated in the different regions of the Fram Strait: Total bacterioplankton in panel **(A)**; *Bacteria* in panel **(B)**; and *Archaea* in panel **(C)**. Box plots were calculated based on cell abundance. Note the different scale of the cell abundances for *Archaea*. The different regions are indicated by color: ice-covered EGC–blue (EG stations), ice-margin N–gray (N stations), ice-free WSC–red (HG stations).

mixing events of Atlantic and Polar water masses by eddies (Wekerle et al., 2017), challenges the identification of specific associations between microbial cell abundances and environmental parameters. Some taxonomic groups (e.g., SAR11 clade) were potentially more influenced by the physical processes such as the presence of ice and distinct Arctic water masses (Kraemer et al., 2020). Likely the mixture of all these environmental variables shaped the observed bacterioplankton communities. We found that cell abundances of some taxonomic groups (e.g., *Gammaproteobacteria*) were higher in some stations with more advanced phytoplankton bloom conditions. However, as we have only limited observations of phytoplankton for this study, we cannot test previous hypotheses of direct associations between the abundances of specific phytoplankton groups and bacterioplankton taxa (Fadeev et al., 2018). Nonetheless, the here observed patterns could represent an enhanced growth of the bacterioplankton on algal exudates (Tada et al., 2011; Teeling et al., 2012). Alternatively, considering the advection of Atlantic waters (Wekerle et al., 2017), it is also possible that some of the observed trends represent lateral transport of phytoplankton or bacterioplankton, or both, from the southern part of the Strait.

In surface waters of all stations, ca. 60% of the total bacterioplankton community was covered by the *Bacteria*-specific probes (EUB388 I–III) and up to 8% was covered by the *Archaea*-specific probe (ARCH915; **Supplementary Table 2**). At depth (>100 m), the coverage of total cells by the *Bacteria*-specific probes strongly decreased to 16–40% of DAPI-stained cells (ANOVA; $F_3 = 15.39$, $p < 0.01$), while the coverage by the *Archaea*-specific probe significantly increased up to 17% of DAPI-stained cells (ANOVA; $F_3 = 34.31$, $p < 0.01$; **Supplementary Table 2**). A similar decrease in detectability of the *Bacteria* -specific probes was previously observed in other bacterioplankton microscopy studies (Karner et al., 2001; Herndl et al., 2005; Varela et al., 2008), and reasons may lie in a ribosomal nucleic acid concentration decrease within the bacterial cells (i.e., lower activity) toward the oligotrophic depths. In addition, there is a potential increase with greater water depths of microbial phylogenetic groups that are not captured by the currently existing probes (Hewson et al., 2006; Galand et al., 2009a; Agogué et al., 2011; Welch and Huse, 2011; Salazar et al., 2016).

We found that in all three regions, total cell abundances of the entire bacterioplankton community were highest at surface

with 10^5–10^6 cells mL^{-1}, and significantly decreased with depth down to 10^4 cells mL^{-1} at meso- and bathypelagic depths (**Figure 2A** and **Supplementary Table 3**; Kruskal–Wallis test; $\chi^2 = 554.39$, $df = 3$, p-value < 0.01). Members of the domain *Bacteria* dominated the communities throughout the entire water column, with highest cell abundances in surface waters (10^5–10^6 cells mL^{-1}), and significantly lower 10^4 cells mL^{-1} at depth (**Figure 2B**; Kruskal–Wallis test; $\chi^2 = 35.27$, $df = 3$, p-value < 0.01). Archaeal cells had an overall lower abundance than bacterial cells by an order of magnitude throughout the entire water column, ranging from 10^4 cells mL^{-1} at surface down to 10^3 cells mL^{-1} in bathypelagic waters (**Figure 2C**). However, unlike *Bacteria*, archaeal communities doubled their absolute cell abundances from ca. 3×10^4 cells mL^{-1} at surface to ca. 6×10^4 cells mL^{-1} at 100 m depth, followed by a significant decrease in cell abundance at meso- and bathypelagic depths (Kruskal–Wallis test; $\chi^2 = 29.04$, $df = 3$, p-value < 0.01). Compared to the stronger decline in bacterial cell numbers, this pattern mirrors the known global trend of relative archaeal enrichment in epipelagic waters (Karner et al., 2001; Herndl et al., 2005; Kirchman et al., 2007; Varela et al., 2008; Schattenhofer et al., 2009), and was also observed in other regions of the Arctic Ocean (Amano-Sato et al., 2013). Altogether, the here observed bacterioplankton cell abundances in surface waters were well within the range of previous observations in the Fram Strait waters, conducted by flow cytometry (Piontek et al., 2014; Fadeev et al., 2018; Engel et al., 2019). However, compared to recent CARD-FISH based observations in eastern Fram Strait (Quero et al., 2020), our cell abundances were consistently one order of magnitude lower along the entire water column. The discrepancy might be associated with methodological differences, such as shorter staining times and the usage of an automated over a manual counting approach in our study. Nevertheless, both studies showed a similar pattern of a strong decrease in bacterioplankton cell abundances with depth, which also matches observations in other oceanic regions

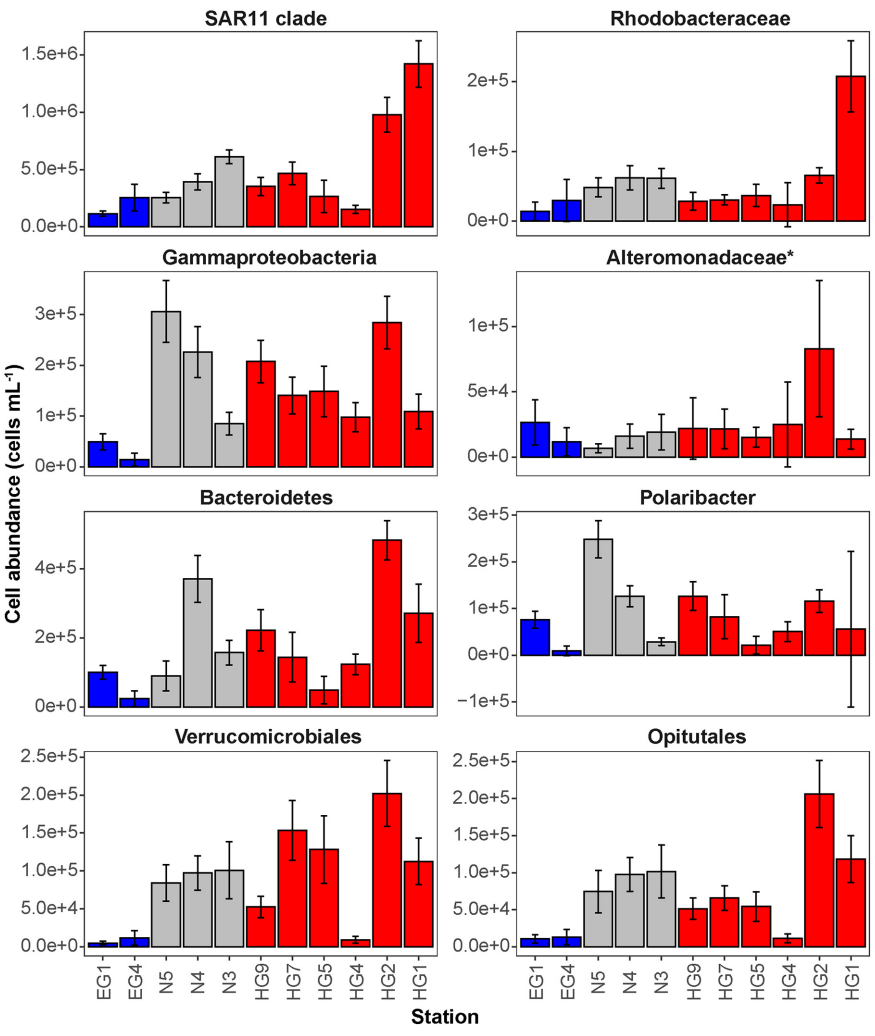


FIGURE 3 | Mean cell abundances of selected taxonomic groups at each station in surface (15–30 m) waters (cells mL^{-1}). The different regions are indicated by color: ice-covered EGC–blue, ice-margin N–gray, ice-free WSC–red.

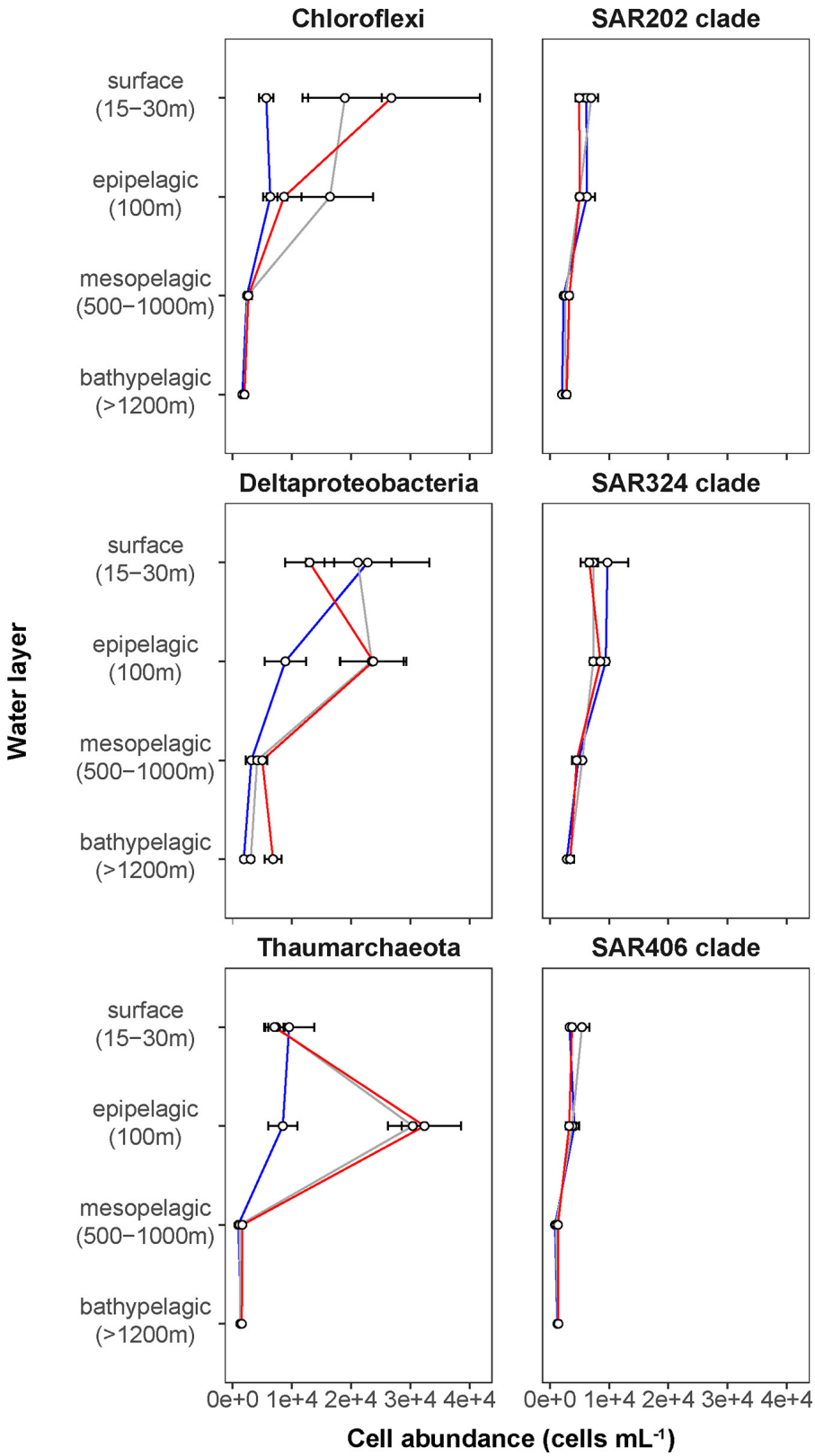


FIGURE 4 | Depth profiles of mean cell abundances of selected taxonomic groups (cells mL^{-1}) calculated for the different Fram Strait regions. The different regions are indicated by color: ice-covered EGC–blue (EG stations), ice-margin N–gray (N stations), ice-free WSC–red (HG stations).

(Karner et al., 2001; Church et al., 2003; Teira et al., 2004; Schattenhofer et al., 2009; Dobal-Amador et al., 2016).

Enigmatic Microbial Lineages Increase in Cell Abundance Toward the Deep Ocean

The deep waters of the Fram Strait basin (>500 m) have a rather homogeneous hydrography (von Appen et al., 2015), and are less affected by the seasonal dynamics that govern the surface layers (Wilson et al., 2017). Previous molecular observations of the deep water bacterioplankton communities showed high sequence abundances of largely unknown taxonomic groups, such as the SAR202 (class *Dehalococcoidia*), SAR324 (class *Deltaproteobacteria*), and SAR406 (phylum *Marinimicrobia*) (Wilson et al., 2017; Fadeev et al., 2020; Quero et al., 2020). There was also higher archaeal sequence abundance at depth, with the class *Nitrososphaeria* (i.e., *Thaumarchaeota*) reaching up to 15% of the sequences in mesopelagic waters (>200 m) (Wilson et al., 2017; Müller et al., 2018; Fadeev et al., 2020). However, it has also been recently shown that in ice-covered regions of the Strait surface-dominant taxonomic groups, such as *Gammaproteobacteria* and *Nitrososphaeria*, are exported via fast-sinking aggregates from surface to the deep ocean (>1,000 m), where they may realize an ecological niche (Fadeev et al., 2020). We observed that in all meso- and bathypelagic waters across all analyzed regions the total cell abundances of the bacterioplankton communities were in the range of 10^4 cells mL^{-1} (**Figure 2**), reflecting observations made in Arctic mesopelagic waters (Wells et al., 2006; Quero et al., 2020). Bacterial taxonomic groups that dominated the surface water communities (e.g., *Bacteroidetes*, *Gammaproteobacteria*, and *Verrucomicrobia*), in both ice-free and ice-covered regions of the Strait, decreased by two orders of magnitude in their cell abundances at meso- and bathypelagic depths (Kruskal–Wallis test; p-value < 0.01; **Figure 3** and **Supplementary Table 3**). This trend strongly correlated with the total bacterioplankton cell abundances along the general water column (Pearson's correlation; $r > 0.8$, p-value < 0.05 and **Supplementary Table 4**). In contrast, other bacterial groups, such as the SAR202 and SAR324 clades, proportionally increased in cell abundances with depth and maintained overall constant cell abundances of ca. 0.5×10^4 cells mL^{-1} until the deep basin (**Supplementary Table 3**). Previous molecular studies of bacterioplankton communities in the Fram Strait suggested a proportional increase of these largely understudied bacterial lineages in the deep ocean, which were previously found to be associated with winter (surface) bacterioplankton (Wilson et al., 2017; Fadeev et al., 2020). The cell abundances presented here indicate that their increasing proportional abundance at depth is due to stronger decrease in the cell abundances of other groups (**Figure 4** and **Supplementary Table 3**). Very little is currently known about these two taxonomic groups, but previous genetic observations suggest that they possess distinct metabolic capabilities, and may be involved in the degradation of recalcitrant organic matter (SAR202 clade; Landry et al., 2017; Colatriano et al., 2018; Saw et al., 2019), or, in sulfur oxidation (SAR324 clade; Swan et al., 2011; Sheik et al., 2014). Their homogeneous distribution from the stratified surface to the homogenous deep ocean suggests that through high functional plasticity these enigmatic bacterial groups fulfill various ecological niches throughout the water column (Saw et al., 2019; Wei et al., 2020), and thus may play important roles in oceanic nutrient cycling.

With depth, the decrease of archaeal cell abundances was less than that of members of the domain *Bacteria* (**Figure 4** and **Supplementary Table 3**), meaning that members of the *Archaea* were proportionally increasing in the total microbial deep-water communities. The *Thaumarchaeota* strongly correlated with the pattern of the archaeal cell abundances (Pearson's correlation; $r = 0.76$, p-value < 0.05; **Supplementary Table 4**), showing a two-fold increase in cell abundance from surface to epipelagic depth (100 m), followed by a substantial decrease toward meso- and bathypelagic waters (**Figure 4** and **Supplementary Table 3**). This two-fold increase toward the epipelagic depths corresponds to previous observations of *Thaumarchaeota* in the north Atlantic (Müller et al., 2018), and a further increase in their cell abundances at higher depths (>1,000 m) was also observed in other oceanic regions (Karner et al., 2001; Church et al., 2003; Herndl et al., 2005; Teira et al., 2006; Galand et al., 2009b). It has been shown in molecular studies that *Thaumarchaeota* comprise a large proportion of the bacterioplankton communities in the Fram Strait, especially in the epipelagic waters (Wilson et al., 2017; Müller et al., 2018; Fadeev et al., 2020). In our study, the *Thaumarchaeota* exhibited their highest cell abundances at 100 m in the ice-free HG, and at the ice-margin N stations (3×10^4 cells mL^{-1}), where they comprised half of the total archaeal community (**Figure 4** and **Supplementary Table 3**). The strong absolute decrease of *Thaumarchaeota* cell abundances toward the meso- and bathypelagic waters suggests a decrease in cell number or activity with depth (Herndl et al., 2005; Kirchman et al., 2007; Alonso-Sáez et al., 2012), and thus lower cell detectability. In deeper water layers, other pelagic archaeal groups, such as the phylum *Euryarchaeota* that was not quantified in this study, may increase in abundance and form the bulk of total archaeal cells here (Galand et al., 2010; Fadeev et al., 2020).

CONCLUSION

Using state-of-the-art semi-automatic microscopy cell counting, we quantified the absolute cell abundance of 12 key taxonomic groups in summer bacterioplankton communities of both ice-free and ice-covered regions of the Fram Strait. We found that in surface waters some taxonomic groups were associated with the distinct water masses of the Strait (e.g., *Rhodobacteraceae* with the Atlantic waters). Surface water bacterioplankton communities were dominated by *Gammaproteobacteria*, *Bacteroidetes*, and *Verrucomicrobia*, which corresponded with biogeochemical conditions in the ongoing seasonal phytoplankton bloom. This suggests that currently predicted longer seasonal phytoplankton blooms, as well as the increasing Atlantic influence on the Arctic Ocean (i.e., "Atlantification"), may have a strong impact on the composition and biogeographical distribution of certain bacterioplankton taxonomic groups in the surface Arctic waters.

This study also provides the first extensive quantification of bacterioplankton community standing stocks in the deep Arctic water column (>500 m). With depth, some taxonomic groups, such as the SAR202 clade, maintained similar abundances throughout the entire water column (2,500 m depth), where other taxa decline by several-fold. The observation of a homogenous abundance further supports the previously established hypothesis that through high functional plasticity these taxonomic groups are realizing various ecological niches throughout the entire water column.

Altogether, our quantitative data on cell abundances of ecologically relevant taxonomic bacterioplankton groups provide insights into factors structuring pelagic bacterioplankton communities from surface to the deep waters of the Arctic Ocean, and add to a baseline to better assess future changes in a rapidly warming region.

MATERIALS AND METHODS

Sampling and Environmental Data Collection

Sampling was carried out during the RV Polarstern expedition PS99.2 to the Long-Term Ecological Research (LTER) site HAUSGARTEN in Fram Strait (June 24th–July 16th, 2016). Sampling was carried out with 12 L Niskin bottles mounted on a CTD rosette (Sea-Bird Electronics Inc. SBE 911 plus probe) equipped with temperature and conductivity sensors, a pressure sensor, altimeter, and a chlorophyll fluorometer. In ice-covered regions the samples were collected through holes in the ice kept open by the research vessel. On board, the samples were fixed with formalin in a final concentration of 2% for 10–12 h, then filtered onto 0.2 μm polycarbonate Nucleopore Track-Etched filters (Whatman, Buckinghamshire, United Kingdom), and stored at −20°C for further analysis.

Hydrographic data of the seawater including temperature and salinity were retrieved from PANGEA (Schröder and Wisotzki, 2014), along with measured chlorophyll *a* concentration (Nöthig et al., 2018; Fadeev et al., 2020) (**Table 1**).

Relative abundance of relevant 16S rRNA as well as data on microscopic abundances of microbial eukaryotes in phytoplankton blooms of the sample location was obtained from (Fadeev et al., 2020).

Catalyzed Reporter Deposition-Fluorescence *in situ* Hybridization (CARD-FISH)

We quantified absolute cell abundances of 12 key bacterioplankton groups (**Supplementary Table 1**), members of the *Bacteria* and *Archaea*, based on their relatively high sequence abundance and recurrences in previous molecular studies of Arctic waters (Bowman et al., 2012; Wilson et al., 2017; Müller et al., 2018; Fadeev et al., 2020). The selected probes covered a variety of taxonomic entities to address standing stocks at different taxonomic levels. All probes were checked for specificity and coverage of their target groups against the SILVA database release 132 (Quast et al., 2013). CARD-FISH was applied based on the protocol established by (Pernthaler et al., 2002), using horseradish-peroxidase (HRP)–labeled oligonucleotide probes (Biomers.net, Ulm, Germany). All filters were embedded in 0.2% low-gelling-point agarose, and treated with 10 mg mL^{-1} lysozyme solution (Sigma-Aldrich Chemie GmbH, Hamburg, Germany) for 1 h at 37°C. Filters for enumerating *Archaea* and *Thaumarchaeota* were treated for an additional 30 min in 36 U mL^{-1} achromopeptidase (Sigma-Aldrich Chemie GmbH, Hamburg, Germany) and 15 μg mL^{-1} proteinase K at 37°C. Subsequently, endogenous peroxidases were inactivated by submerging the filter pieces in 0.15% H_2O_2 in methanol for 30 min before rinsing in Milli-Q water and dehydration in 96% ethanol. Then, the filters were covered in hybridization buffer and a probe concentration of 0.2 ng μL^{-1}. Hybridization was performed at 46°C for 2.5 h, followed by washing in pre-warmed washing buffer at 48°C for 10 min, and 15 min in 1x PBS. Signal amplification was carried out for 45 min at 46°C with amplification buffer containing either tyramide-bound Alexa 488 (1 μg/mL) or Alexa 594 (0.33 μg mL^{-1}). Afterward, the cells were counterstained in 1 μg/mL DAPI (4′,6-diamidino-2-phenylindole; Thermo Fisher Scientific GmbH, Bremen, Germany) for 10 min at 46°C. After rinsing with Milli-Q water and 96% ethanol, the filter pieces were embedded in a 4:1 mix of Citifluor (Citifluor Ltd., London, United Kingdom) and Vectashield (Vector Laboratories, Inc., Burlingame, CA, United States), and stored overnight at −20°C for later microscopy evaluation.

Automated Image Acquisition and Cell Counting

The filters were evaluated microscopically under a Zeiss Axio Imager.Z2 stand (Carl Zeiss MicroImaging GmbH, Jena, Germany), equipped with a multipurpose fully automated microscope imaging system (MPISYS), a Colibri LED light source illumination system, and a multi-filter set 62HE (Carl Zeiss MicroImaging GmbH, Jena, Germany). Pictures were taken via a cooled charged-coupled-device (CCD) camera (AxioCam MRm; Carl Zeiss AG, Oberkochen, Germany) with a 63 × oil objective, a numerical aperture of 1.4, and a pixel size of 0.1016 μm/pixel, coupled to the AxioVision SE64 Rel.4.9.1 software (Carl Zeiss AG, Oberkochen, Germany) as described by Bennke et al. (2016). Exposure times were adjusted after manual inspection with the AxioVision Rel.4.8 software coupled to the SamLoc 1.7 software (Zeder et al., 2011), which was also used to define the coordinates of the filters on the slides. For image acquisition, channels were defined with the MPISYS software, and a minimum of 55 fields of view with a minimum distance of 0.25 mm were acquired of each filter piece by recoding a z-stack of seven images in autofocus.

Cell enumeration was performed with the software Automated Cell Measuring and Enumeration Tool (ACMETool3, 2018-11-09; M. Zeder, Technobiology GmbH, Buchrain, Switzerland). Total bacterioplankton cells were determined as the total amount of DAPI-stained cells. Counts for each

taxonomic group included only cells that were simultaneously stained by DAPI and the taxa-specific FISH probe.

Calculation of Consumed Inorganic Nutrients

Following (Fadeev et al., 2018) the nutrient consumption (Δ) at each station was calculated by subtracting the mean value of all collected measurements above 50 m from the mean value of all collected measurements between 50 and 100 m (below the seasonal pycnocline).

Statistical Analyses

All statistical analyses and calculations in this study were performed using R (v4.0.2) (www.r-project.org) in RStudio (v1.3.1056), *i.e.,* statistical tests for normality, ANOVA and Kruskal–Wallis. *Post hoc* Wilcoxon test and Pearson's rank correlation coefficient were conducted with the R package "rstatix" (v0.6.0) (Kassambara, 2020). Plots were generated using the R package "ggplot2" (v3.3.2) (Wickham, 2016) and "tidyverse" (v1.3.0) (Wickham et al., 2019).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MC-M, EF, and VS-C designed and conducted the study, and wrote the manuscript with guidance from AB. MC-M performed the hybridizations, cell counting, data and statistical analysis with guidance from VS-C (probe selection, CARD-FISH application, and counting) and EF (data and statistical analysis). All authors critically revised the manuscript and gave their approval of the submitted version.

FUNDING

This project has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Program (FP7/2007–2013) research project ABYSS (Grant Agreement No. 294757) to AB. Additional funding came from the Helmholtz Association, specifically for the FRAM infrastructure, from the Max Planck Society for VS-C, from the Hector Fellow Academy for MC-M, and from the Austrian Science Fund (FWF) Grant No. M-2797 to EF.

ACKNOWLEDGMENTS

We thank the captain and crew of RV Polarstern expedition PS99.2, as well as the chief scientist Thomas Soltwedel for support with work at sea. We also thank Pier Offre for assistance in sampling, Greta Reintjes for designing the probe Opi346, Mareike Bach for technical support, Sinhue Torres-Valdes and Laura Wischnewski for conducting the inorganic nutrient measurements. We also thank Andreas Ellrott for the support with the automated microscope from the Max Planck Institute for Marine Microbiology. This work was conducted in the framework of the HGF Infrastructure Program FRAM of the Alfred-Wegener-Institute Helmholtz Center for Polar and Marine.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.658803/full#supplementary-material

REFERENCES

Agogué, H., Lamy, D., Neal, P. R., Sogin, M. L., and Herndl, G. J. (2011). Water mass-specificity of bacterial communities in the north atlantic revealed by massively parallel sequencing. *Mol. Ecol.* 20, 258–274. doi: 10.1111/j.1365-294X.2010.04932.x

Alonso-Sáez, L., Waller, A. S., Mende, D. R., Bakker, K., Farnelid, H., Yager, P. L., et al. (2012). Role for urea in nitrification by polar marine archaea. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17989–17994. doi: 10.1073/pnas.1201914109

Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., and Stahl, D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* 56, 1919–1925. doi: 10.1128/aem.56.6.1919-1925.1990

Amano-Sato, C., Akiyama, S., Uchida, M., Shimada, K., and Utsumi, M. (2013). Archaeal distribution and abundance in water masses of the arctic ocean, pacific sector. *Aquat. Microb. Ecol.* 69, 101–112. doi: 10.3354/ame01624

Basedow, S. L., Sundfjord, A., von Appen, W. J., Halvorsen, E., Kwasniewski, S., and Reigstad, M. (2018). Seasonal variation in transport of zooplankton into the arctic basin through the atlantic gateway, fram strait. *Front. Mar. Sci.* 5:194. doi: 10.3389/fmars.2018.00194

Bennke, C. M., Reintjes, G., Schattenhofer, M., Ellrott, A., Wulf, J., Zeder, M., et al. (2016). Modification of a high-throughput automatic microbial cell enumeration system for shipboard analyses. *Appl. Environ. Microbiol.* 82, 3289–3296. doi: 10.1128/AEM.03931-15

Beszczynska-Möller, A., Fahrbach, E., Schauer, U., and Hansen, E. (2012). Variability in atlantic water temperature and transport at the entrance to the arctic ocean, 19972010. *ICES J. Mar. Sci.* 69, 852–863. doi: 10.1093/icesjms/fss056

Beszczynska-Möller, A., Woodgate, R. A., Lee, C., Melling, H., and Karcher, M. (2011). A synthesis of exchanges through the main oceanic gateways to the arctic ocean. *Oceanography* 24, 83–99. doi: 10.5670/oceanog.2011.59

Bižić-Ionescu, M., Zeder, M., Ionescu, D., Orlić, S., Fuchs, B. M., Grossart, H. P., et al. (2015). Comparison of bacterial communities on limnic versus coastal marine particles reveals profound differences in colonization. *Environ. Microbiol.* 17, 3500–3514. doi: 10.1111/1462-2920.12466

Boetius, A., Albrecht, S., Bakker, K., Bienhold, C., Felden, J., Fernández-Méndez, M., et al. (2013). Export of algal biomass from the melting arctic sea ice. *Science* 339, 1430–1432. doi: 10.1126/science.1231346

Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., Rysgaard, S., and Sicheritz-Ponten, T. (2012). Microbial community structure of arctic multiyear sea ice

and surface seawater by 454 sequencing of the 16S RNA gene. *ISME J.* 6, 11–20. doi: 10.1038/ismej.2011.76

Buchan, A., LeCleir, G. R., Gulvik, C. A., and González, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* 12, 686–698. doi: 10.1038/nrmicro3326

Cardman, Z., Arnosti, C., Durbin, A., Ziervogel, K., Cox, C., Steen, A. D., et al. (2014). Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of svalbard. *Appl. Environ. Microbiol.* 80, 3749–3756. doi: 10.1128/AEM.00899-14

Church, M. J., DeLong, E. F., Ducklow, H. W., Karner, M. B., Preston, C. M., and Karl, D. M. (2003). Abundance and distribution of planktonic archaea and bacteria in the waters west of the antarctic peninsula. *Limnol. Oceanogr.* 48, 1893–1902. doi: 10.4319/lo.2003.48.5.1893

Colatriano, D., Tran, P. Q., Guéguen, C., Williams, W. J., Lovejoy, C., and Walsh, D. A. (2018). Genomic evidence for the degradation of terrestrial organic matter by pelagic arctic ocean chloroflexi bacteria. *Commun. Biol.* 1:90. doi: 10.1038/s42003-018-0086-7

Cottier, F., Tverberg, V., Inall, M., Svendsen, H., Nilsen, F., and Griffiths, C. (2005). Water mass modification in an arctic fjord through cross-shelf exchange: the seasonal hydrography of kongsfjorden, svalbard. *J. Geophys. Res. Ocean.* 110, 1–18. doi: 10.1029/2004JC002757

Dai, A., Luo, D., Song, M., and Liu, J. (2019). Arctic amplification is caused by sea-ice loss under increasing CO_2. *Nat. Commun.* 10:121. doi: 10.1038/s41467-018-07954-9

de Steur, L., Hansen, E., Gerdes, R., Karcher, M., Fahrbach, E., and Holfort, J. (2009). Freshwater fluxes in the east greenland current: a decade of observations. *Geophys. Res. Lett.* 36:L23611. doi: 10.1029/2009GL041278

DeLong, E. F., Taylor, L. T., Marsh, T. L., and Preston, C. M. (1999). Visualization and enumeration of marine planktonic archaea and bacteria by using polyribonucleotide probes and fluorescent in situ hybridization. *Appl. Environ. Microbiol.* 65, 5554–5563. doi: 10.1128/aem.65.12.5554-5563.1999

Dobal-Amador, V., Nieto-Cid, M., Guerrero-Feijoo, E., Hernando-Morales, V., Teira, E., and Varela-Rozados, M. M. (2016). Vertical stratification of bacterial communities driven by multiple environmental factors in the waters (0-5000 m) off the galician coast (NW Iberian margin). *Deep. Res. Part I Oceanogr. Res. Pap.* 114, 1–11. doi: 10.1016/j.dsr.2016.04.009

Dobricic, S., Vignati, E., and Russo, S. (2016). Large-scale atmospheric warming in winter and the arctic sea ice retreat. *J. Clim.* 29, 2869–2888. doi: 10.1175/JCLI-D-15-0417.1

Engel, A., Bracher, A., Dinter, T., Endres, S., Grosse, J., Metfies, K., et al. (2019). Inter-annual variability of organic carbon concentrations in the eastern fram strait during summer (2009-2017). *Front. Mar. Sci.* 6:187. doi: 10.3389/fmars.2019.00187

Engel, A., Piontek, J., Metfies, K., Endres, S., Sprong, P., Peeken, I., et al. (2017). Inter-annual variability of transparent exopolymer particles in the arctic ocean reveals high sensitivity to ecosystem changes. *Sci. Rep.* 7:4129. doi: 10.1038/s41598-017-04106-9

Fadeev, E., Cardozo-Mino, M. G., Rapp, J. Z., Bienhold, C., Salter, I., Salman-Carvalho, V., et al. (2021). Comparison of two 16S rRNA primers (V3–V4 and V4–V5) for studies of arctic microbial communities. *Front. Microbiol.* 12:283. doi: 10.3389/fmicb.2021.637526

Fadeev, E., Rogge, A., Ramondenc, S., Nöthig, E.-M., Wekerle, C., Bienhold, C., et al. (2020). Sea-ice retreat may decrease carbon export and vertical microbial connectivity in the eurasian arctic basins. *Nat. Res.* doi: 10.21203/rs.3.rs-101878/v1

Fadeev, E., Salter, I., Schourup-Kristensen, V., Nöthig, E. M., Metfies, K., Engel, A., et al. (2018). Microbial communities in the east and west fram strait during sea ice melting season. *Front. Mar. Sci.* 5:429. doi: 10.3389/fmars.2018.00429

Fernández-Méndez, M., Wenzhöfer, F., Peeken, I., Sørensen, H. L., Glud, R. N., and Boetius, A. (2014). Composition, buoyancy regulation and fate of ice algal aggregates in the central arctic ocean. *PLoS One* 9:e107452. doi: 10.1371/journal.pone.0107452

Galand, P. E., Casamayor, E. O., Kirchman, D. L., and Lovejoy, C. (2009a). Ecology of the rare microbial biosphere of the arctic ocean. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22427–22432. doi: 10.1073/pnas.0908284106

Galand, P. E., Casamayor, E. O., Kirchman, D. L., Potvin, M., and Lovejoy, C. (2009b). Unique archaeal assemblages in the arctic ocean unveiled by massively parallel tag sequencing. *ISME J.* 3, 860–869. doi: 10.1038/ismej.2009.23

Galand, P. E., Potvin, M., Casamayor, E. O., and Lovejoy, C. (2010). Hydrography shapes bacterial biogeography of the deep arctic ocean. *ISME J.* 4, 564–576. doi: 10.1038/ismej.2009.134

Giebel, H. A., Kalhoefer, D., Lemke, A., Thole, S., Gahl-Janssen, R., Simon, M., et al. (2011). Distribution of roseobacter RCA and SAR11 lineages in the north sea and characteristics of an abundant RCA isolate. *ISME J.* 5, 8–19. doi: 10.1038/ismej.2010.87

Giovannoni, S. J. (2017). SAR11 bacteria: the most abundant plankton in the oceans. *Ann. Rev. Mar. Sci.* 9, 231–255. doi: 10.1146/annurev-marine-010814-015934

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* 8:2224. doi: 10.3389/fmicb.2017.02224

Goldman, J. C., McCarthy, J. J., and Peavey, D. G. (1979). Growth rate influence on the chemical composition of phytoplankton in oceanic waters. *Nature* 279, 210–215. doi: 10.1038/279210a0

Hebbeln, D., and Wefer, G. (1991). Effects of ice coverage and ice-rafted material on sedimentation in the fram strait. *Nature* 350, 409–411. doi: 10.1038/350409a0

Herndl, G. J., Reinthaler, T., Teira, E., Van Aken, H., Veth, C., Pernthaler, A., et al. (2005). Contribution of archaea to total prokaryotic production in the deep atlantic ocean. *Appl. Environ. Microbiol.* 71, 2303–2309. doi: 10.1128/AEM.71.5.2303-2309.2005

Hewson, I., Steele, J. A., Capone, D. G., and Fuhrman, J. A. (2006). Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. *Limnol. Oceanogr.* 51, 1274–1283. doi: 10.4319/lo.2006.51.3.1274

Karner, M. B., Delong, E. F., and Karl, D. M. (2001). Archaeal dominance in the mesopelagic zone of the pacific ocean. *Nature* 409, 507–510. doi: 10.1038/35054051

Kassambara, A. (2020). *rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.5.0.999. R Packag. version 0.6.0*. Available Online at: https://rpkgs.datanovia.com/rstatix/.

Kirchman, D. L., Elifantz, H., Dittel, A. I., Malmstrom, R. R., and Cottrell, M. T. (2007). Standing stocks and activity of archaea and bacteria in the western arctic ocean. *Limnol. Oceanogr.* 52, 495–507. doi: 10.4319/lo.2007.52.2.0495

Korhonen, M., Rudels, B., Marnela, M., Wisotzki, A., and Zhao, J. (2013). Time and space variability of freshwater content, heat content and seasonal ice melt in the arctic ocean from 1991 to 2011. *Ocean Sci.* 9, 1015–1055. doi: 10.5194/os-9-1015-2013

Kraemer, S., Ramachandran, A., Colatriano, D., Lovejoy, C., and Walsh, D. A. (2020). Diversity and biogeography of SAR11 bacteria from the arctic ocean. *ISME J.* 14, 79–90. doi: 10.1038/s41396-019-0499-4

Kumar, M. S., Slud, E. V., Okrah, K., Hicks, S. C., Hannenhalli, S., and Bravo, H. C. (2017). Analysis and correction of compositional bias in sparse sequencing count data. *bioRxiv* 19:799. doi: 10.1101/142851

Landry, Z., Swa, B. K., Herndl, G. J., Stepanauskas, R., and Giovannoni, S. J. (2017). SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *mBio* 8, e413–e417. doi: 10.1128/mBio.00413-17

Leu, E., Søreide, J. E., Hessen, D. O., Falk-Petersen, S., and Berge, J. (2011). Consequences of changing sea-ice cover for primary and secondary producers in the european arctic shelf seas: timing, quantity, and quality. *Prog. Oceanogr.* 90, 18–32. doi: 10.1016/j.pocean.2011.02.004

Luo, H., and Moran, M. A. (2014). Evolutionary ecology of the marine roseobacter clade. *Microbiol. Mol. Biol. Rev.* 78, 1–16. doi: 10.1128/mmbr.88888-88

Müller, O., Wilson, B., Paulsen, M. L., Ruminska, A., Armo, H. R., Bratbak, G., et al. (2018). Spatiotemporal dynamics of ammonia-oxidizing thaumarchaeota in distinct arctic water masses. *Front. Microbiol.* 9:24. doi: 10.3389/fmicb.2018.00024

Mundy, C. J., Barber, D. G., and Michel, C. (2005). Variability of snow and ice thermal, physical and optical properties pertinent to sea ice algae biomass during spring. *J. Mar. Syst.* 58, 107–120. doi: 10.1016/j.jmarsys.2005.07.003

Nöthig, E. M., Bracher, A., Engel, A., Metfies, K., Niehoff, B., Peeken, I., et al. (2015). Summertime plankton ecology in fram strait - a compilation of long-and short-term observations. *Polar Res.* 34:23349. doi: 10.3402/polar.v34.23349

Nöthig, E.-M., Knüppel, N., and Lorenzen, C. (2018). *Chlorophyll a measured on water bottle samples during POLARSTERN cruise PS99.2 (ARK-XXX/1.2).*

Bremerhaven: Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research.

Peng, G., and Meier, W. N. (2018). Temporal and regional variability of arctic sea-ice coverage from satellite data. *Ann. Glaciol.* 59, 191–200. doi: 10.1017/aog.2017.32

Pernthaler, A., Pernthaler, J., and Amann, R. (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine Bacteria. *Appl. Environ. Microbiol.* 68, 3094–3101. doi: 10.1128/AEM.68.6.3094-3101.2002

Perrette, M., Yool, A., Quartly, G. D., and Popova, E. E. (2011). Near-ubiquity of ice-edge blooms in the arctic. *Biogeosciences* 8, 515–524. doi: 10.5194/bg-8-515-2011

Piontek, J., Sperling, M., Nöthig, E. M., and Engel, A. (2014). Regulation of bacterioplankton activity in fram strait (arctic ocean) during early summer: the role of organic matter supply and temperature. *J. Mar. Syst.* 132, 83–94. doi: 10.1016/j.jmarsys.2014.01.003

Piontek, J., Sperling, M., Nöthig, E. M., and Engel, A. (2015). Multiple environmental changes induce interactive effects on bacterial degradation activity in the arctic ocean. *Limnol. Oceanogr.* 60, 1392–1410. doi: 10.1002/lno.10112

Piwosz, K., Shabarova, T., Pernthaler, J., Posch, T., Šimek, K., Porcal, P., et al. (2020). Bacterial and eukaryotic small-subunit amplicon data do not provide a quantitative picture of microbial communities, but they are reliable in the context of ecological interpretations. *mSphere* 5, 1–14. doi: 10.1128/msphere.00052-20

Polyakov, I. V., Pnyushkov, A. V., Alkire, M. B., Ashik, I. M., Baumann, T. M., Carmack, E. C., et al. (2017). Greater role for atlantic inflows on sea-ice loss in the eurasian basin of the arctic ocean. *Science* 356, 285–291. doi: 10.1126/science.aai8204

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Quero, G. M., Celussi, M., Relitti, F., Kovačević, V., Del Negro, P., and Luna, G. M. (2020). Inorganic and organic carbon uptake processes and their connection to microbial diversity in meso- and bathypelagic arctic waters (eastern fram strait). *Microb. Ecol.* 79, 823–839. doi: 10.1007/s00248-019-01451-2

Redfield, A. C. (1963). "*The influence of organisms on the composition of seawater,*" *in The sea*. Hoboken, NJ: Wiley-Interscience, 26–77.

Rosselli, R., Romoli, O., Vitulo, N., Vezzi, A., Campanaro, S., De Pascale, F., et al. (2016). Direct 16S rRNA-seq from bacterial communities: a PCR-independent approach to simultaneously assess microbial diversity and functional activity potential of each taxon. *Sci. Rep.* 6, 1–12.

Rudels, B., Schauer, U., Björk, G., Korhonen, M., Pisarev, S., Rabe, B., et al. (2012). Observations of water masses and circulation in the eurasian basin of the arctic ocean from the 1990s to the late 2000s. *Ocean Sci. Discuss.* 9, 2695–2747. doi: 10.5194/osd-9-2695-2012

Salazar, G., Cornejo-Castillo, F. M., Benítez-Barrios, V., Fraile-Nuez, E., Álvarez-Salgado, X. A., Duarte, C. M., et al. (2016). Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J.* 10, 596–608. doi: 10.1038/ismej.2015.137

Saw, J. H. W., Nunoura, T., Hirai, M., Takaki, Y., Parsons, R., Michelsen, M., et al. (2019). Pangenomics reveal diversification of enzyme families and niche specialization in globally abundant SAR202 Bacteria. *bioRxiv* 11, e2919–e2975. doi: 10.1101/692848

Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., and Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the atlantic ocean. *Environ. Microbiol.* 11, 2078–2093. doi: 10.1111/j.1462-2920.2009.01929.x

Schröder, M., and Wisotzki, A. (2014). *Physical oceanography measured on water bottle samples during POLARSTERN cruise PS82 (ANT-XXIX/9)*. Bremerhaven: Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research.

Selje, N., Simon, M., and Brinkhoff, T. (2004). A newly discovered Roseobacter cluster in temperate and polar oceans. *Nature* 427, 445–448. doi: 10.1038/nature02272

Sheik, C. S., Jain, S., and Dick, G. J. (2014). Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ. Microbiol.* 16, 304–317. doi: 10.1111/1462-2920.12165

Soltwedel, T., Bauerfeind, E., Bergmann, M., Bracher, A., Budaeva, N., Busch, K., et al. (2016). Natural variability or anthropogenically-induced variation? insights from 15 years of multidisciplinary observations at the arctic marine LTER site HAUSGARTEN. *Ecol. Indic.* 65, 89–102. doi: 10.1016/j.ecolind.2015.10.001

Soltwedel, T., Bauerfeind, E., Bergmann, M., Budaeva, N., Hoste, E., Jaeckisch, N., et al. (2005). Hausgarten: Multidisciplinary investigations at a deep-Sea, long-term observatory in the arctic ocean. *Oceanography* 18, 46–61. doi: 10.5670/oceanog.2005.24

Sperling, M., Giebel, H. A., Rink, B., Grayek, S., Staneva, J., Stanev, E., et al. (2012). Differential effects of hydrographic and biogeochemical properties on the SAR11 clade and roseobacter RCA cluster in the north sea. *Aquat. Microb. Ecol.* 67, 25–34. doi: 10.3354/ame01580

Sun, L., Perlwitz, J., and Hoerling, M. (2016). What caused the recent "warm arctic, cold continents" trend pattern in winter temperatures? *Geophys. Res. Lett.* 43, 5345–5352. doi: 10.1002/2016GL069024

Swan, B. K., Martinez-Garcia, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamy, D., et al. (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333, 1296–1300. doi: 10.1126/science.1203690

Tada, Y., Makabe, R., Kasamatsu-Takazawa, N., Taniguchi, A., and Hamasaki, K. (2013). Growth and distribution patterns of roseobacter/rhodobacter, SAR11, and bacteroidetes lineages in the southern ocean. *Polar Biol.* 36, 691–704. doi: 10.1007/s00300-013-1294-8

Tada, Y., Taniguchi, A., Nagao, I., Miki, T., Uematsu, M., Tsuda, A., et al. (2011). Differing growth responses of major phylogenetic groups of marine bacteria to natural phytoplankton blooms in the western north pacific ocean. *Appl. Environ. Microbiol.* 77, 4055–4065. doi: 10.1128/AEM.02952-10

Teeling, H., Fuchs, B. M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C. M., et al. (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* 336, 608–611. doi: 10.1126/science.1218344

Teeling, H., Fuchs, B. M., Bennke, C. M., Krüger, K., Chafee, M., Kappelmann, L., et al. (2016). Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife* 5:e11888. doi: 10.7554/eLife.11888

Teira, E., Lebaron, P., Van Aken, H., and Herndl, G. J. (2006). Distribution and activity of bacteria and archaea in the deep water masses of the north atlantic. *Limnol. Oceanogr.* 51, 2131–2144. doi: 10.4319/lo.2006.51.5.2131

Teira, E., Reinthaler, T., Pernthaler, A., Pernthaler, J., and Herndl, G. J. (2004). Combining catalyzed reporter deposition-fluorescence in situ hybridization and microautoradiography to detect substrate utilization by bacteria and archaea in the deep ocean. *Appl. Environ. Microbiol.* 70, 4411–4414. doi: 10.1128/AEM.70.7.4411-4414.2004

Varela, M. M., Van Aken, H. M., and Herndl, G. J. (2008). Abundance and activity of chloroflexi-type SAR202 bacterioplankton in the meso- and bathypelagic waters of the (sub)tropical atlantic. *Environ. Microbiol.* 10, 1903–1911. doi: 10.1111/j.1462-2920.2008.01627.x

Vernet, M., Richardson, T. L., Metfies, K., Nöthig, E. M., and Peeken, I. (2017). Models of plankton community changes during a warm water anomaly in arctic waters show altered trophic pathways with minimal changes in carbon export. *Front. Mar. Sci.* 4:160. doi: 10.3389/fmars.2017.00160

von Appen, W. J., Schauer, U., Somavilla, R., Bauerfeind, E., and Beszczynska-Möller, A. (2015). Exchange of warming deep waters across fram strait. *Deep. Res. Part I Oceanogr. Res. Pap.* 103, 86–100. doi: 10.1016/j.dsr.2015.06.003

Walczowski, W., Beszczynska-Möller, A., Wieczorek, P., Merchel, M., and Grynczel, A. (2017). Oceanographic observations in the nordic sea and fram strait in 2016 under the IO PAN long-term monitoring program AREX. *Oceanologia* 59, 187–194. doi: 10.1016/j.oceano.2016.12.003

Wassmann, P., and Reigstad, M. (2011). Future arctic ocean seasonal ice zones and implications for pelagic-benthic coupling. *Oceanography* 24, 220–231. doi: 10.5670/oceanog.2011.74

Wei, Z.-F., Li, W.-L., Huang, J.-M., and Wang, Y. (2020). Metagenomic studies of SAR202 bacteria at the full-ocean depth in the mariana trench. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 165:103396. doi: 10.1016/j.dsr.2020.103396

Wekerle, C., Wang, Q., von Appen, W. J., Danilov, S., Schourup-Kristensen, V., and Jung, T. (2017). Eddy-resolving simulation of the atlantic water circulation

in the fram strait with focus on the seasonal cycle. *J. Geophys. Res. Ocean.* 122, 8385–8405. doi: 10.1002/2017JC012974

Welch, D. B. M., and Huse, S. M. (2011). Microbial diversity in the deep sea and the underexplored "rare biosphere.". *Handb. Mol. Microb. Ecol. II Metagen. Differ. Habitats* 103, 243–252. doi: 10.1002/9781118010549.ch24

Wells, L. E., Cordray, M., Bowerman, S., Miller, L. A., Vincent, W. F., and Deming, J. W. (2006). Archaea in particle-rich waters of the beaufort shelf and franklin bay, canadian arctic: clues to an allochthonous origin? *Limnol. Oceanogr.* 51, 47–59. doi: 10.4319/lo.2006.51.1.0047

Wickham, H. (2016). *"Getting started with ggplot2," in ggplot2*. New York City, NY: Springer, 11–31.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Wilson, B., Müller, O., Nordmann, E. L., Seuthe, L., Bratbak, G., and Øvreås, L. (2017). Changes in marine prokaryote composition with season and depth over an arctic polar year. *Front. Mar. Sci.* 4:95. doi: 10.3389/fmars.2017.00095

Wilson, C., and Wallace, D. W. R. (1990). Using the nutrient ratio NO/PO as a tracer of continental shelf waters in the central arctic ocean. *J. Geophys. Res.* 95:22193. doi: 10.1029/jc095ic12p22193

Zeder, M., Ellrott, A., and Amann, R. (2011). Automated sample area definition for high-throughput microscopy. *Cytom. Part A* 79, 306–310. doi: 10.1002/cyto.a.21034

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 May 2021
doi: 10.3389/fmicb.2021.612732

Temperature Responses of Heterotrophic Bacteria in Co-culture With a Red Sea *Synechococcus* Strain

*Abbrar Labban[1,2], Antonio S. Palacio[3], Francisca C. García[4], Ghaida Hadaidi[1], Mohd I. Ansari[1], Ángel López-Urrutia[5], Laura Alonso-Sáez[3], Pei-Ying Hong[2] and Xosé Anxelu G. Morán[1,5]**

[1] Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [2] Water Desalination and Reuse Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, [3] AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Sukarrieta, Spain, [4] Environment and Sustainability Institute, University of Exeter, Penryn, United Kingdom, [5] Centro Oceanográfico de Gijón/Xixón, Instituto Español de Oceanografía, Gijón, Spain

OPEN ACCESS

Edited by:
Gipsi Lima Mendez,
Catholic University of Louvain, Belgium

Reviewed by:
Eric A. Webb,
University of Southern California, United States
Malin Olofsson,
University of Georgia, United States

***Correspondence:**
Xosé Anxelu G. Morán
xelu.moran@kaust.edu.sa

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: 30 September 2020
Accepted: 29 March 2021
Published: 10 May 2021

Citation:
Labban A, Palacio AS, García FC, Hadaidi G, Ansari MI, López-Urrutia Á, Alonso-Sáez L, Hong P-Y and Morán XAG (2021) Temperature Responses of Heterotrophic Bacteria in Co-culture With a Red Sea Synechococcus Strain. Front. Microbiol. 12:612732. doi: 10.3389/fmicb.2021.612732

Interactions between autotrophic and heterotrophic bacteria are fundamental for marine biogeochemical cycling. How global warming will affect the dynamics of these essential microbial players is not fully understood. The aims of this study were to identify the major groups of heterotrophic bacteria present in a *Synechococcus* culture originally isolated from the Red Sea and assess their joint responses to experimental warming within the metabolic ecology framework. A co-culture of *Synechococcus* sp. RS9907 and their associated heterotrophic bacteria, after determining their taxonomic affiliation by 16S rRNA gene sequencing, was acclimated and maintained in the lab at different temperatures (24–34°C). The abundance and cellular properties of *Synechococcus* and the three dominant heterotrophic bacterial groups (pertaining to the genera *Paracoccus*, *Marinobacter,* and *Muricauda*) were monitored by flow cytometry. The activation energy of *Synechococcus*, which grew at 0.94–1.38 d^{-1}, was very similar (0.34 ± 0.02 eV) to the value hypothesized by the metabolic theory of ecology (MTE) for autotrophs (0.32 eV), while the values of the three heterotrophic bacteria ranged from 0.16 to 1.15 eV and were negatively correlated with their corresponding specific growth rates (2.38–24.4 d^{-1}). The corresponding carrying capacities did not always follow the inverse relationship with temperature predicted by MTE, nor did we observe a consistent response of bacterial cell size and temperature. Our results show that the responses to future ocean warming of autotrophic and heterotrophic bacteria in microbial consortia might not be well described by theoretical universal rules.

Keywords: *Synechococcus*, heterotrophic bacteria, metabolic ecology, temperature, growth rate, cell size, Red Sea

INTRODUTION

Bacterioplankton play significant roles in carbon and nutrient cycling in the oceans (Azam et al., 1983; Jiao et al., 2010; Kirchman, 2016). The interactions between autotrophic and heterotrophic bacteria are essential in marine food webs (Azam and Malfatti, 2007; Buchan et al., 2014), especially in oligotrophic tropical and subtropical regions. There, heterotrophic bacteria consume dissolved organic matter (DOM) ultimately fixed photosynthetically by cyanobacteria (mostly belonging to the genera *Prochlorococcus* and *Synechococcus*) and other small phytoplankton (Azam and Malfatti, 2007). In turn, heterotrophic bacteria facilitate the growth of primary producers including cyanobacteria by supplying fundamental micronutrients including vitamins (Hayashi et al., 2011), amino acids (Christie-Oleza et al., 2015b) or scavenging the reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2) from the environment (Morris et al., 2008, 2011), which can be formed photochemically in the oceans as a result of the photooxidation of DOM (Moffett and Zika, 1987; Moffett and Zajiriou, 1990). H_2O_2 can cause cell damage and death in the absence of protective enzymes such as catalase that can decompose it to water and oxygen (Imlay, 2003). Since *Prochlorococcus* lacks this enzyme (Morris et al., 2011) and *Synechococcus* has shown a weak catalase activity (Pena and Bullerjahn, 1995), specific groups of heterotrophic bacteria, by acting as H_2O_2 scavengers, can help the growth of both cyanobacteria.

Synechococcus is an ancient and genetically diverse genus (Scanlan et al., 2009), partly due to horizontal gene transfer events (Zhaxybayeva et al., 2006; Coutinho et al., 2016). The cell size of *Synechococcus* usually varies between 0.8 and 1.2 μm (Mella-Flores et al., 2012) and its genome size ranges between 2.2 and 2.9 Mb (Scanlan et al., 2009). Based on the pigment composition of phycobiliproteins in the photosynthetic antenna, all *Synechococcus* cells are composed of phycocyanin (PC), whereas some strains contain both PC and phycoerythrin (PE) (Liu et al., 2014), resulting in differently colored strains, ranging from blue-green to pink-orange (Wood et al., 1985; Six et al., 2007). A recent study suggests that marine representatives traditionally known as *Synechococcus* are actually comprised by three distinct taxa, suggesting the new names Ca. *Marinosynechococcus* and Ca. *Juxtasynechococcus* together with *Cyanobium* (Doré et al., 2020). Several *Synechococcus* strains have been found naturally associated with heterotrophic bacteria and are difficult to maintain in axenic conditions (Hayashi et al., 2011; Zheng et al., 2018). Axenic *Synechococcus* and *Prochlorococcus* cultures that were cultivated using antibiotic treatments were able to grow only with a high density of the inoculum (Morris et al., 2008, 2011; Hayashi et al., 2011), but they grew more efficiently when co-cultured with several heterotrophic bacteria (Morris et al., 2008; Hayashi et al., 2011; Zinser, 2018). This adds to the difficulty of obtaining axenic isolates of *Synechococcus* using the common culture isolation method (Zheng et al., 2018). The dominant groups of heterotrophic bacteria associated with different eutrophic and oligotrophic *Synechococcus* strains typically belong to Alphaproteobacteria, Gammaproteobacteria, and Flavobacteria (Zheng et al., 2018), some of the most abundant bacterioplankton classes in the ocean (Du et al., 2006; Willis et al., 2019).

Temperature plays a fundamental role in the physiology of microorganisms (Edwards and Richardson, 2004; Montoya and Raffaelli, 2010). Increasing temperature can impact planktonic microbes either directly by altering their metabolism (Sarmento et al., 2010) and individual size (Daufresne et al., 2009; Morán et al., 2010), or indirectly by shifting their distribution or through increased stratification resulting in reduced nutrient inputs into the upper layers (Behrenfeld et al., 2006; Keeling et al., 2010). Although warming will enhance metabolism in nutrient-sufficient conditions, there is an upper temperature limit that every organism can tolerate (Heinrich, 1977). Traditionally described using the Q_{10} coefficient (Eppley, 1972; Chen and Laws, 2017), another way of approaching the role of temperature in microbial plankton is through the framework of Metabolic Theory of Ecology (MTE) (Brown et al., 2004). Metabolism is obviously the result of a myriad of biochemical reactions (Morowitz et al., 2000; Brown et al., 2004), but some generalizations have been made. According to the MTE, the effect of temperature on metabolic rates such as specific growth rates can be described by the activation energy. For autotrophs, with metabolic rates essentially equaling the mean rate of photosynthetic reactions (Farquhar et al., 1980), an activation energy of ca. 0.32 eV is predicted (Allen et al., 2005; López-Urrutia et al., 2006). For heterotrophs, their metabolic rates approach the mean rate of respiration (del Giorgio and Duarte, 2002; López-Urrutia et al., 2006), resulting in an activation energy of ca. 0.65 eV (Gillooly et al., 2001). Another general ecological theory that has gained recent attention in the marine environment is the so-called temperature-size rule (TSR), which suggests an inverse relationship between ambient temperature and individual size (Atkinson, 1994; Atkinson and Sibly, 1997). Originally developed for metazoans (Azevedo et al., 2002), the decrease in individual size as a function of temperature has also been documented for picophytoplankton (Morán et al., 2010) and heterotrophic bacteria (Morán et al., 2015).

Global warming is impacting the structure and function of marine ecosystems worldwide. Although tropical planktonic bacteria encounter the highest natural temperatures in the world, there is no systematic information about how they cope with them or their potential adaptation to further warming. Understanding the physiological responses of a Red Sea *Synechococcus* strain and their associated heterotrophic bacteria to temperature will provide us with an estimation of how future warming may impact these crucial players in biogeochemical cycling in regions currently subject to lower temperature ranges. In this study, we used the non-axenic *Synechococcus* RS9907 culture previously acclimated to different temperatures spanning from 24 to 34°C. The strain RS9907 was originally isolated from the Gulf of Aqaba (29°28′N, 34°55′E) in the Red Sea (Fuller et al., 2003) and belongs to the clade II, the dominant in the Red Sea (Post et al., 2011; Coello-Camba et al., 2020). Although kept in culture for thousands of generations, disentangling how these Red Sea bacteria cope with high temperatures can help us predict future responses of bacterioplankton in other marine regions.

MATERIALS AND METHODS

Culture Conditions

The *Synechococcus* sp. RS9907 culture was obtained from the Roscoff Culture Collection (Roscoff, France), and grown in artificial seawater (PCR-S11-Red Sea medium prepared with a salinity value of 36) at 22°C under an irradiance of ca. 115 μmol photons m^{-2} S^{-1} in a 12:12 h light:dark cycle until the start of the experiments. Although according to Doré et al. (2020) we should rename it to *Marinosynechococcus*, we have opted to keep the traditional term until a wide consensus about renaming the group is reached.

Thermal Acclimation and Flow Cytometry

After acclimating the culture to the different temperatures for at least 8 generations as described in Palacio et al. (2020), 3 experiments were sequentially conducted targeting different groups and growth phases of the bacterial consortium (**Supplementary Table 1**). Co-culture true replicates were incubated in 160 mL vented cap flasks at 4–5 different temperatures ranging from 24 to 34°C in order to determine their thermal niches (Boyd et al., 2013). The first experiment (Experiment A), focused on monitoring the dynamics of RS9907 rather than the heterotrophs, and the cultures were incubated at 24, 28, 30, and 33°C. The sampling was done daily during the light cycle at 9:30 a.m. for a variable total duration depending on the actual experimental temperature (i.e., the warmer the shortest duration). Four replicates were sampled until *Synechococcus* entered the logarithmic growth phase (after 4–5 days) and from then onward separate duplicates were kept and monitored until the stationary phase was reached (7–13 days), in order to accurately estimate the carrying capacity (i.e., maximum abundance) of *Synechococcus*. The second experiment (Experiment B) lasted for 4 days and consisted of triplicates incubated at 24, 28, 30, and 33°C. No sampling was done in the first 2 days and samples were taken every 2 h from day 2 to 4 to finely document the exponential growth phase of the co-culture. In the third experiment (Experiment C), specifically aimed at targeting the dynamics of heterotrophic bacteria, In addition to the tested temperatures in experiments A and B, one set of triplicates was incubated at 34°C. Samples were collected every 2 h for 24 h and one more sample was taken after 48 h. During the acclimatization process, it became clear that any temperature above 30°C should be done with caution to avoid killing the cyanobacteria. We tried to determine the maximum temperature at which *Synechococcus* could grow. Since this study is more focused on experiment C, we acclimated the cultures at a temperature slightly higher than 33°C but still within the Red Sea's natural temperature range.

All samples (396 μl) were preserved in cryovials with glutaraldehyde (0.025% final concentration), incubated in the dark for 10 min to allow the cells to be fixed completely and then stored at –80°C until analysis (Palacio et al., 2020). Typically within 1 week from collection, samples were run using a BD FACSCanto II flow cytometer and the obtained data were analyzed using BD Paint-A-Gate software. The *Synechococcus* cells were easily detected in cytograms of red fluorescence (PerCP-Cy5-5, 498 nm) vs. right angle light or side scatter (SSC) signals and of red vs. orange fluorescence (PE, 433 nm) signals. After staining the samples with SYBRGreen (Marie et al., 1999), heterotrophic bacteria were identified and counted simultaneously using cytograms of green fluorescence (FITC, 360 nm) vs. red fluorescence signals in order to distinguish between autotrophs (naturally fluorescing in red) and heterotrophs, and also of green fluorescence vs. SSC signals. SSC is a flow cytometric variable common to all cells reflecting changes in individual size (Felip et al., 2007). The mean relative SSC (relSSC) values of the different flow cytometric groups to the SSC value of 1 μm fluorescent latex beads (Molecular Probes) were used to estimate their corresponding cell diameter (μm) using these calibrations: μm = 1.62 + 0.87 log relSSC for *Synechococcus* and μm = 0.91 + 0.34 log relSSC for the heterotrophic groups (Calvo-Díaz and Morán, 2006). Bacterial cell size was finally given as biovolume in μm^3 assuming spherical shape. We are aware that some of the bacteria present may well have been bacilli (rods) rather than cocci (spheres). Although the co-culture was not observed under the microscope to detect the exact cell shapes, since the analysis by flow cytometry is done cell by cell to a large total number (ca. 10,000 cells), it can be presumed that the relSSC value of a bacillus with the same width of a coccus' diameter (but a clearly longer length), would ultimately be higher.

DNA Extraction and 16S rRNA Gene Amplicon Sequencing

All DNA samples in Experiment C were collected after 48 h, a time at which the abundance of heterotrophic bacteria was higher than that of RS9907 at all tested temperatures (**Supplementary Table 2**). 130 mL of the *Synechococcus* RS9907 co-cultures in the available flasks were collected after 48 h, divided into three equal volumes and centrifuged at 4,000 rpm for 20 min at 4°C. The supernatant was removed and 15 mL was kept from each tube and mixed together in one tube. The 45 mL of mixed concentrated culture was centrifuged again at 4,000 rpm for 20 min at 4°C. The supernatant was removed and pellets were re-suspended and transferred to 2 mL tube. Samples were centrifuged at 14,000 rpm for 2 min at 4°C and the supernatant was discarded. DNA pellets were stored at –80°C until DNA extraction. DNA was extracted using DNeasy PowerSoil kit (QIAGEN, Germany) with a minor modification to the protocol by adding 15 μL of 100 mg/mL lysozyme to the pellets and incubation at 37°C (Ansari et al., 2015). PCR amplification of the variable regions 4 and 5 of the 16S rRNA gene was performed using the forward primer 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′) and reverse primer 926R (5′- CCGYCAATTYMTTTRAGTTT-3′) (Parada et al., 2015). PCR reaction mixture was prepared for all samples with 0.6 μl of each primer, 0.48 μl of Taq DNA polymerase HiFi (Quanta, BioSciences, Inc.^{TM}), and 15 μl of FailSafe^{TM} PCR 2X PreMix (Epicenter, WI, United States) in a total volume of 30 μl. Triplicate of 10 μl PCRs were performed with 25 cycles of thermal program following cycling conditions

of 926R primer (Parada et al., 2015). These triplicate samples were then pooled, and 2 μl was used on a 1% agarose gel to check the amplification. The remaining 28 μl was cleaned up using AMPure XP beads (Beckman Coulter). Indexing PCR was performed on the cleaned samples using Nextera XT index kit (Illumina, San Diego, CA, United States) following the manufacturer's protocol. The MiSeq spacer size was 412 bp and 16S rRNA gene library was sequenced using Illumina MiSeq platform in 2 × 300 bp paired-end mode with 7x depth of coverage. A 20% phiX was used as a control at the KAUST Bioscience Core Laboratory.

Sequence Data Processing and Analysis

Raw sequence data analysis was conducted in mothur (version 1.39.5) following Hadaidi et al. (2018) with some modifications. Sequence reads were joined into contigs using the "make.contigs" command, and contigs longer than 413 bp and ambiguous bases were removed from the analysis. All the sequences that occurred only once were subsequently excluded and SILVA database was used to align the remaining sequences. Sequences were then pre-clustered letting maximum 2 nt difference between the sequences and chimeras were removed applying "uchime." The sequences were assigned to Operational Taxonomic Units (OTUs) using the Greengenes database. Short-sequence reads did not allow the identification of species, hence our analysis was conducted at the genus level.

In order to explore the correspondence between the cytometric and taxonomic groups present in the co-culture, the abundance of the dominant taxa was estimated by multiplying their relative abundance in the amplicon sequencing dataset by the flow cytometry total bacterial abundance. The use of amplicon sequencing relative abundance multiplied by the abundance obtained by flow cytometry to provide an estimate of the abundance of the different OTUs is rather common (Props et al., 2017; Lin et al., 2019). Although prone to amplification biases, if only a few groups clearly stand out as the most abundant ones (as found in our co-culture), the potential error would be minimized. Since *Synechococcus* was unequivocally identified in the cytograms, its flow cytometric abundance was used as a control to match that based on percent reads.

Metabolic Theory of Ecology (MTE) and Temperature-Size Rule (TSR) Calculations

To estimate the specific growth rates (μ, d^{-1}) of *Synechococcus* and the heterotrophic bacterial groups distinguished by flow cytometry at the different temperatures of each of the three experiments, we determined the slope of ln-transformed abundance vs. time for the corresponding exponential phase of growth (Huete-Stauffer et al., 2015), which differed in timing for the different groups and experimental temperatures. The temperature responses of bacterial specific growth rates and carrying capacities (i.e., maximum abundances) were approached using the MTE central equation (Brown et al., 2004). Specifically, the μ thermal dependence was described by its activation energy (E_a). E_a was calculated in Arrhenius plots as the slope of ln-transformed specific growth rates vs. 1/kT (Huete-Stauffer et al., 2015), with k representing the Boltzmann's constant (8.62×10^{-5} eV K^{-1}), and T the temperature in Kelvin (Brown et al., 2004). The carrying capacity of each bacterial group was recorded as the maximum abundance that was actually reached at a given incubation temperature. In order to test the MTE that predicted inverse relationship between carrying capacity and temperature (Savage et al., 2004), we simply calculated the linear regression between both variables. The same method was used for testing the prediction of the TSR of a smaller individual size with higher temperature (Atkinson et al., 2003; Bernhardt et al., 2018). We represented the mean bacterial biovolume (see calculation details in the above section) over the exponential growth phase vs. experimental temperature and calculated the corresponding slope of the linear regression model.

RESULTS

Identification of *Synechococcus* sp. RS9907-Associated Heterotrophic Bacteria

To identify the heterotrophic bacteria associated to *Synechococcus* in the RS9907 co-culture, we analyzed 16S rRNA gene amplicon sequences. A total of 1396590 sequences with an average read length of 412 bp reads were obtained by amplicon sequencing targeting the 16S rRNA in the co-cultures acclimated at the five different temperatures (24, 28, 30, 33, and 34°C) after 48 h of incubation in experiment C (see above for details about the three experiments performed). After passing the quality filtering and exclusion of chimeras, 837308 sequences were retained. Sequences were then rarefied to 25,188 sequences per sample and their taxonomic affiliation was analyzed. At a 97% similarity cutoff, sequences were clustered into 49 Operational Taxonomic Units (OTUs). The three OTUs that had the highest relative abundances, excluding RS9907 (33.09% of the total) were affiliated to the genera *Marinobacter* (Gammaproteobacteria), *Muricauda* (Flavobacteriia), and *Paracoccus* (Alphaproteobacteria) (**Figure 1**). *Marinobacter* had a mean relative abundance of 61.96%, followed by *Muricauda* (4.32%), and *Paracoccus* (0.27%) (**Figure 1**), while the sum of the remaining 49 OTUs identified amounted to only 0.4% of the reads. The fifth most relatively abundant OTU (not shown) represented on average 0.2% of the reads. Even though the field is moving toward using amplicon sequence variants (ASV)-based sequencing analysis, using ASVs in a co-culture that did not contain a very diverse community and aiming at genus-level identification would not have been very different from our reported OTUs. According to Martinson et al. (2019), both OTU- and ASV-based analysis methods would likely have yielded nearly identical results in our study.

Three distinct heterotrophic bacterial groups were also consistently detected in the cytograms from the three experiments based on their relative nucleic acid content and cell size characteristics. The groups were initially named as 1, 2, and 3 ordered by decreasing values of green fluorescence

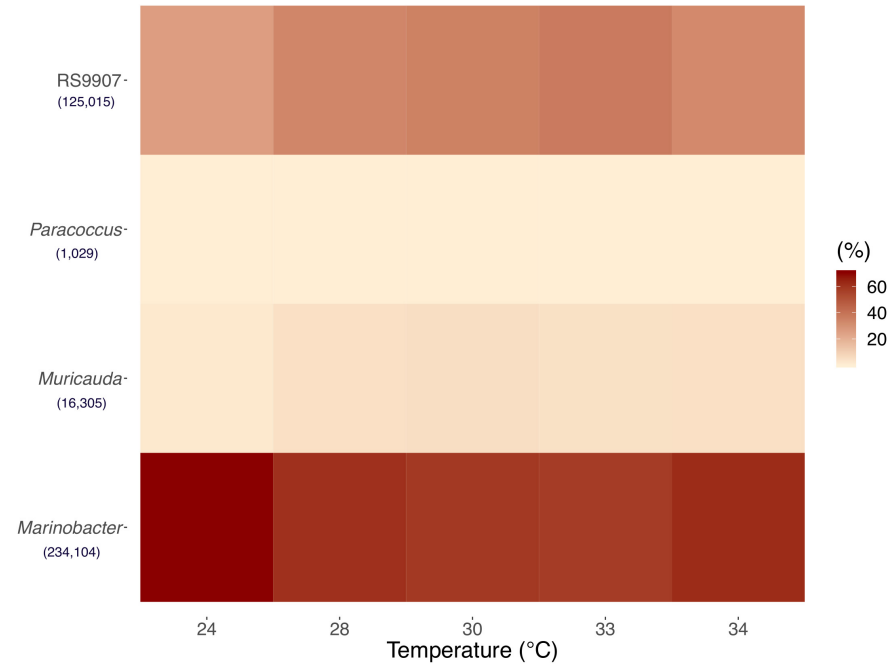


FIGURE 1 | Mean of triplicates 16S rRNA gene-based relative abundances of *Synechococcus* and the top three most abundant heterotrophic bacteria OTUs at the five incubation temperatures of Experiment C after 48 h. See the text for details. Numbers in brackets represent total OTUs reads.

and SSC (**Supplementary Figure 1**). After comparing their flow cytometry-based abundances (pooling all replicates and temperatures) together with that of RS9907 (easily distinguished because of its higher red fluorescence signal due to chlorophyll *a*) with the 16S rRNA gene-based abundance of the 4 dominant OTUs, we found significant correlations with *r*-values ranging from 0.6 to 0.9 over 3 orders of magnitude absolute abundances. Hereinafter, the flow cytometric groups 1, 2, and 3 will be referred to as *Paracoccus*, *Marinobacter,* and *Muricauda*, respectively (**Figure 2**).

Bacterial Dynamics and Cell Size

To document the dynamics of the RS9907 co-culture, three experiments were conducted at different bacterial growth phases and time points. Total heterotrophic bacterial abundance was slightly lower than that of *Synechococcus* at the onset of experiments A and C, with cyanobacteria ranging from 1.1 to 2.2 $\times$ 10^5 cells mL^{-1}. However, the total abundance of heterotrophic bacteria increased very fast (just after 1 day of incubation) and remained relatively constant at 4–7 $\times$ 10^6 cells mL^{-1} while *Synechococcus* increased in abundance over two orders of magnitude (**Figure 3A**). Consequently, the proportion of heterotrophic bacteria in the co-culture decreased steadily and similarly in the 3 experiments from ca. 90 to ca. 25% as the abundance of RS9907 increased to a maximum of ca. 1.6 $\times$ 10^7 cells mL^{-1} (**Figure 3B**). By plotting data from the three experiments corresponding to different stages in the RS9907 co-culture dynamics, it became clear that the importance of heterotrophic bacteria decreased with increasing abundance of *Synechococcus* once the cyanobacteria reached ca. 3.5 $\times$ 10^5 cells mL^{-1}.

The three experiments allowed us to detect highly variable but internally consistent responses of the three most abundant heterotrophic bacterial OTUs, with a different degree of temporal resolution (**Figure 4**). Experiment A and C were performed with lower initial abundance of both autotrophic and heterotrophic cells than experiment B (**Supplementary Table 1**). Experiment A began with low abundance (1.6 $\times$ 10^5 $\pm$ 32538 cells mL^{-1}) of *Synechococcus* (**Figure 4A**), experiment B started with higher initial abundances (2.5 $\times$ 10^6 $\pm$ 384751 cells mL^{-1}, **Figure 4B**) while experiment C was purportedly conducted with a lower initial abundance (9.2 $\times$ 10^4 $\pm$ 14,382 cells mL^{-1}, **Figure 4C**), in order to better track the fine-scale dynamics of heterotrophic bacteria. Although the initial abundance of RS9907 in experiment A (**Figure 4A**) and experiment C (**Figure 4C**) were closer, the daily sampling of experiment A did not allow us to adequately determine the response of the three heterotrophic OTUs. Experiment C was indeed the best suited to document the actual responses of the heterotrophic bacteria. In particular, *Paracoccus* showed very distinct dynamics compared to the other two groups: it invariably responded with a fast increase and decrease in the first 24 h (**Figures 4D–F**) at all temperatures (**Supplementary Figures 2D–F, 3D–F, 4D–F, 5B**). *Paracoccus* reached their maximum abundance after 1 day in experiment A, which was comparable to the abundance after 24 h in experiment C. Frequent sampling during the first 24 h in experiment C provided us with not only an assessment of the specific growth rate, but also a better estimation of the maximum abundance of *Paracoccus*.

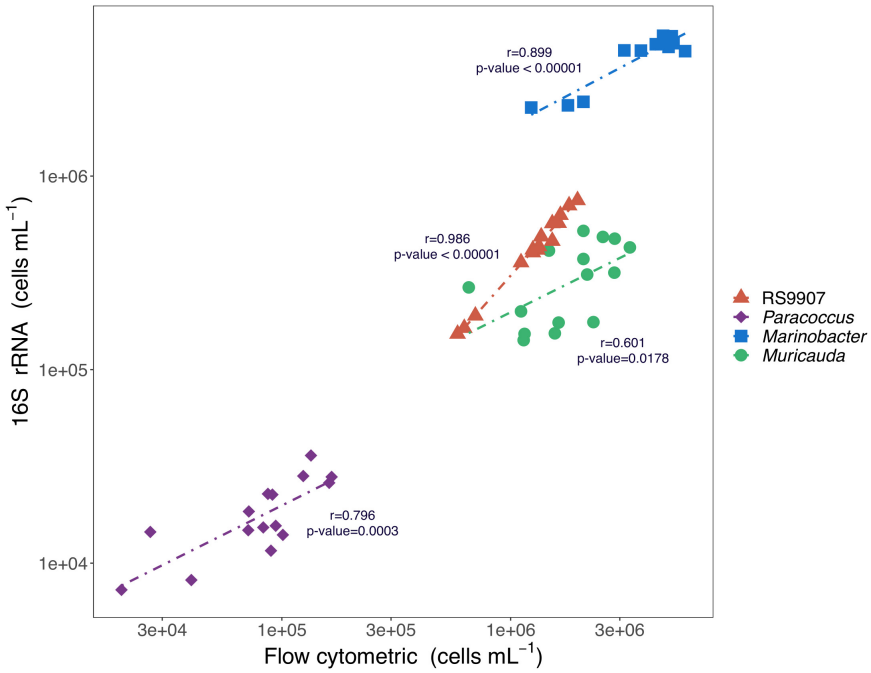


FIGURE 2 | Relationships between the 16S rRNA gene-based cell abundances (i.e., percent reads multiplied by total bacterial abundance as determined by flow cytometry) of *Synechococcus* RS9907, *Paracoccus*, *Marinobacter*, and *Muricauda* and the flow cytometric abundances of *Synechococcus* and the 3 major clusters of heterotrophic cells shown in **Supplementary Figure 1** for all replicates and temperatures after 48 h incubation of experiment C.

Marinobacter reached their carrying capacity of ca. 10^7 cells mL^{-1} after <1 day in experiments A and C and remained pretty constant hereinafter (**Figures 4G–I**) at all temperatures (**Supplementary Figures 2G–I, 3G–I, 4G–I, 5C**). Similarly, the every 2 h regular sampling of experiment C allowed us to measure the specific growth rate of *Marinobacter*. In contrast, *Muricauda* kept on growing during the sampled periods in the three experiments, following a response similar to that of *Synechococcus*, though with lower abundance (**Figures 4A–C,J–L**) at all temperatures (**Supplementary Figures 2A–C,J–L, 3A–C,J–L, 4A–C,J–L, 5A,D**). All other tested temperatures showed similar patterns to the bacterial dynamics found at 30°C (**Supplementary Figures 2–5**).

The differences in mean cell size of the 4 bacterial OTUs were almost consistent across the three experiments (**Supplementary Table 3**). Expectedly, *Synechococcus* were the largest bacteria, followed by *Paracoccus* and *Marinobacter*, while *Muricauda* showing the smallest cell size. The effects of temperature on the cell size of each group are discussed in the following section.

Temperature Responses

To determine the joint responses of the autotrophic and heterotrophic members of the co-culture to experimental warming using the MTE and TSR frameworks, we assessed the specific growth rates and cell sizes of the dominant members of the co-culture over a temperature range covering the natural variability found in Red Sea waters (24–34°C). In particular we wanted to test these hypotheses that (i) the specific growth rates of autotrophic and heterotrophic bacteria will have different temperature dependences (i.e., they differ in their corresponding MTE activation energies) and (ii) the cell size will decrease with warming in a similar fashion for all the bacterial groups identified. *Marinobacter* showed the highest specific growth rates of the four bacterial OTUs assessed in experiment C (**Figure 5**), with values exceeding 20 d^{-1} (19.6–24.4 d^{-1}), followed by *Paracoccus* (7.60–14.9 d^{-1}) and *Muricauda* (2.38–6.18 d^{-1}). The specific growth rates of the RS9907 strain (0.94–1.38 d^{-1}) were much lower than those of the three heterotrophs. Specific growth rates of all groups increased with temperature albeit with different patterns, resulting in different activation energies (see below). RS9907 showed the highest specific growth rates at 33°C, *Paracoccus* and *Marinobacter* grew best at 34°C and *Muricauda* showed a decrease in their specific growth rate at temperatures higher than 30°C (**Figure 5**). The specific growth rate of *Synechococcus* in experiment A (1.00–1.60 d^{-1}) was very similar to the values shown for experiment C and also followed a similar pattern with temperature. The corresponding activation energy (E_a) of RS9907 (0.34 ± 0.02 eV) was not significantly different from the value hypothesized for autotrophs by the MTE (0.32 eV, *t*-test, $p > 0.05$) while the E_a values of the three heterotrophs varied greatly, from 0.16 to 1.15 eV. Interestingly, the activation energies of *Paracoccus*, *Marinobacter,* and *Muricauda* were inversely correlated with their mean specific growth rates (**Figure 6**).

The decay phase of *Paracoccus* came just after a few hours of incubation during the first 24 h (the maximum abundance

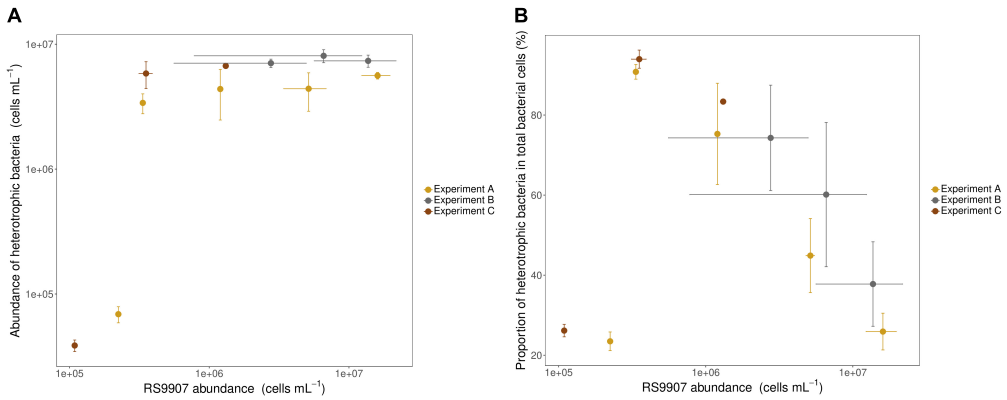


FIGURE 3 | (A) Absolute abundance of heterotrophic bacteria (i.e., sum of all heterotrophic cells counted by flow cytometry) and **(B)** proportion of heterotrophic bacteria to total bacterial cells vs. the abundance of RS9907 at 28°C in the three experiments. Each symbol corresponds to one daily measurement. Error bars represent the standard deviations of 4 replicates in experiment A and triplicates in experiments B and C.

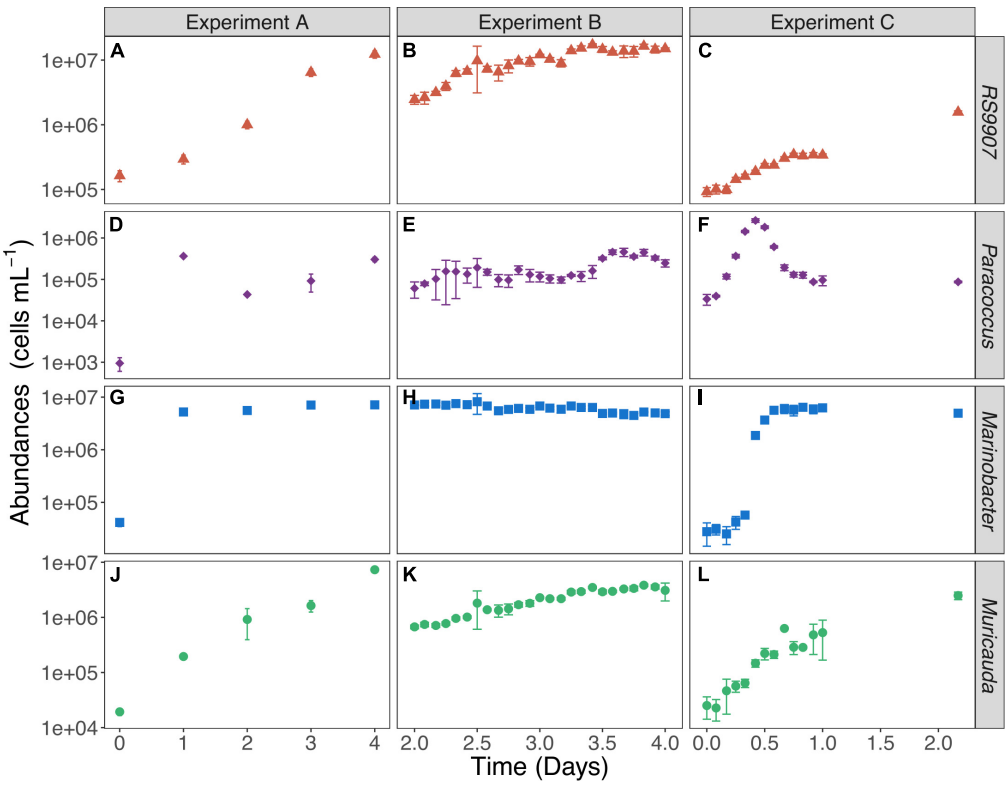


FIGURE 4 | Variations in the mean abundance of *Synechococcus* RS9907 **(A–C)** and the 3 heterotrophic bacteria [*Paracoccus* **(D–F)**, *Marinobacter* **(G–I)**, *Muricauda* **(J–L)**] at 30°C in the three experiments. Error bars represent the standard deviations of 4 replicates in experiment A and triplicates in experiments B and C.

was reached ca. half a day after the start of the incubation), allowing us to estimate an apparent mortality rate, which ranged from −1.98 to −12.01 d^{-1}. Similar to specific growth rates (**Figure 5B**), mean mortality rates were significantly and negatively correlated with increasing temperature in the 5 temperatures tested of experiment C (r^2 = −0.89, p = 0.001, n = 5, **Supplementary Figure 6**). Consequently, both mean specific growth and mortality rates of this group showed a strong correlation across the temperatures tested (r^2 = −0.96, p = 0.001, n = 5).

The carrying capacities (i.e., maximum abundances) of *Muricauda* and *Synechococcus* could only be accurately estimated from experiment A, since they both kept on growing in experiment C. The latter experiment permitted the determination

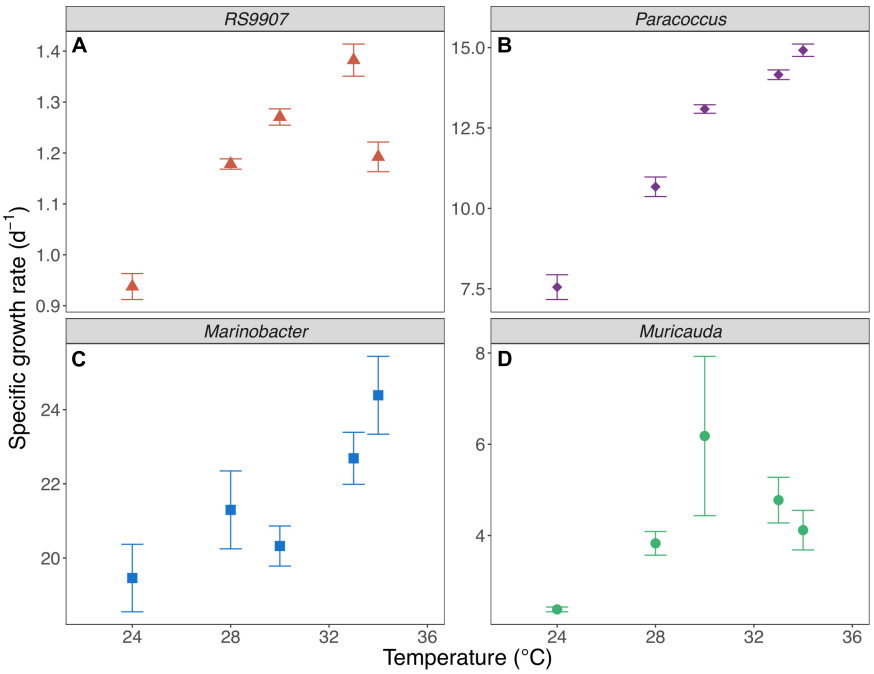


FIGURE 5 | Mean specific growth rates of *Synechococcus* RS9907 **(A)**, *Paracoccus* **(B)**, *Marinobacter* **(C)** and *Muricauda* **(D)** vs. incubation temperature in experiment C. Error bars represent standard deviations.

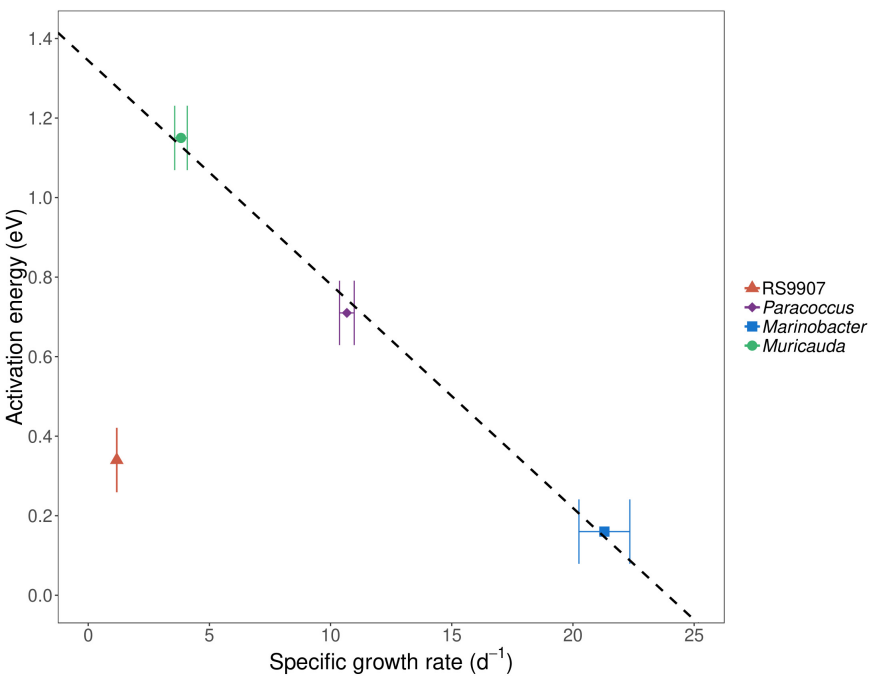


FIGURE 6 | Mean activation energy (E_a) estimated from specific growth rates for at least three different temperatures (see section Materials and Methods and **Figure 5** for details) vs. mean specific growth rate (μ) at 28°C of *Synechococcus* RS9907, *Paracoccus*, *Marinobacter*, and *Muricauda* in experiment C. Error bars represent standard deviations. The fitted line represents the ordinary least squares linear regression for heterotrophs (r^2 = 0.99, p = 0.04, n = 3). The activation energy of *Marinobacter* was estimated from specific growth rates at all tested temperatures (24–34°C) while *Muricauda* and *Paracoccus* were estimated including only temperatures from 24 to 30°C. For *Synechococcus,* the activation energy was calculated from specific growth rates from 24 to 33°C.

of the carrying capacities of *Paracoccus* and *Marinobacter*. The carrying capacities of *Synechococcus* (2.94×10^7–1.17×10^8 cells mL^{-1}) and *Paracoccus* (1.86–4.01 $\times 10^6$ cells mL^{-1}) were negatively correlated with temperature (r^2 = –0.69 and –0.41, p = 0.007 and 0.006, n = 8 and 15, respectively). On the contrary, a positive correlation was observed between *Marinobacter* carrying capacity (4.89–7.03 $\times 10^6$ cells mL^{-1}) and temperature (r^2 = 0.56, p = 0.003, n = 12, excluding 34°C values). The carrying capacities of *Muricauda* (3.16–7.32 $\times 10^6$ cells mL^{-1}) did not show any significant correlation with temperature (**Supplementary Figure 7A**).

The TSR was tested for the different bacterial groups with the mean cell size during the exponential growth phase. Interestingly, there was a strong positive relationship between the mean cell size of RS9907 during the exponential growth phase (0.39–0.60 μm^3) and temperature (r^2 = 0.88, p = 0.0001, n = 9, excluding > 30°C values), while none of the heterotrophic groups mean cell sizes bore any significant correlation with temperature (**Supplementary Figure 8**). However, we also estimated the corresponding cell size at maximum abundance of the 4 OTUs at each experimental temperature. *Marinobacter* cell size was the only bacterial group significantly correlated with temperature (r^2 = 0.59, p = 0.0005, n = 15), similarly to the result of RS9907 (**Supplementary Figure 7B**). The effect of temperature was consistent on the mean cell sizes across the three experiments (**Supplementary Table 3**). Mean cell sizes of *Synechococcus* also increased with temperature in experiments A, B, and C (r^2 = 0.98; 0.88; 0.59, p < 0.00001; < 0.00001; = 0.005, n = 16; 12; 15, respectively) with slightly different values while heterotrophic bacteria did not show any significant relationship with temperature.

DISCUSSION

Heterotrophic Bacteria in the *Synechococcus* sp. RS9907 Co-culture

The heterotrophic bacterial OTUs present in the RS9907 co-culture, making up to a mean of 66.69% of total reads when *Synechococcus* reached ca. 1 million cells mL^{-1} mostly belonged to Gammaproteobacteria and Alphaproteobacteria. Coincidentally, they are also the most abundant classes of heterotrophic bacterioplankton in the Red Sea, with Flavobacteriia typically found at lower abundances (Ngugi et al., 2012; Pearman et al., 2016, 2017), and only increasing in their importance in shallow waters (Ansari, personal communication), suggesting that their presence in the non-axenic culture dates back from the time of initial isolation.

The flow cytometric abundance of the 4 clusters we consistently observed (including *Synechococcus*) aligned independently with the read-based abundance of the top four dominant bacterial OTUs (**Figure 2**). Since these abundances spanned over three orders of magnitude, we can conclude that the affiliation of the top three most abundant heterotrophic bacterial OTUs to the genera *Marinobacter*, *Muricauda* and *Paracoccus* overcome any possible amplification bias of the 16S rRNA gene. *Marinobacter* spp. have been previously associated with dinoflagellate and coccolithophore (Amin et al., 2009) as well as diatom cultures (Amin et al., 2009, 2015). Several *Marinobacter* species have been reported to effectively promote the growth of different dinoflagellate strains (Amin et al., 2009; Bolch et al., 2017), partially due to the production of vibrioferrin, a siderophore secreted by the bacteria that stimulates algal absorption of iron (Amin et al., 2009). In return, not only in the decay phase but during exponential growth, phytoplankton naturally release DOM (Myklestad, 1995; Hansell, 2013), which can be utilized by the heterotrophic bacteria in the surrounding water or within the phycosphere, suggesting that this essential iron and fixed carbon mutualistic interaction may also occur between the *Marinobacter* species present in the RS9907 co-culture. Likewise, different *Muricauda* strains have also been found to promote the growth of microalgae, whose increase in cell density was accompanied by an increase in *Muricauda* cells number (Han et al., 2016). A strain of *Muricauda* was shown to contain appendages, vesicle-like structures with fibrillar-like structures on the surface, which are thought to be utilized to connect cells to one another (Bruns et al., 2001) and perhaps involved in interactions with the surrounding environment (Malfatti and Azam, 2009). Those might account for the positive correlation between the growth patterns of our *Muricauda* OTU and *Synechococcus*. Finally, a strain of *Paracoccus* has shown positive oxidase and catalase activities (Zheng et al., 2011) and a reduction in H_2O_2 has been reported in different *Prochlorococcus* strains when they were co-cultured with heterotrophic bacteria (Morris et al., 2008, 2011; Coe et al., 2016). Our *Paracoccus* member may play a role in scavenging the ROS possibly generated in the co-culture, since *Synechococcus* lacked its own catalase when we blasted the katG gene against the complete genome of the RS9907 strain. Remarkably, the same three genera of heterotrophic bacteria reported here were also found in *Synechococcus* sp. BL107 (Christie-Oleza et al., 2017), originally isolated from Blanes Bay, Mediterranean Sea (Dufresne et al., 2008). The heterotrophic bacterium presented in the WH7803 co-culture was found to rely on the exoproteome of the cyanobacterium, as the shotgun proteomics analysis showed that there were virtually no accumulated proteins in this co-culture (Christie-Oleza et al., 2015a). In contrast, the reduction of the accumulated organic matter generated by *Synechococcus* by one of the heterotrophic bacteria present was found to decelerate cyanobacteria mortality in the above-mentioned study (Christie-Oleza et al., 2017), suggesting that although no *Roseobacter* member was found among our dominant OTUs, this behavior might contribute to the apparent symbiotic relationship established between *Synechococcus* and the heterotrophic bacteria.

Bacterial Dynamics and Cell Size in Response to Temperature

Although when applied to prokaryotic plankton neither the MTE (Sinsabaugh and Shah, 2010; Huete-Stauffer et al., 2015; Arandia-Gorostidi et al., 2017; Morán et al., 2018) nor the TSR (Morán et al., 2015) are free from criticism, they are still valid frameworks to test general relationships of organisms in

response to temperature. We have used both the MTE and TSR principles in the controlled conditions of a co-culture in which the cyanobacterium and likely also the associated heterotrophs had been isolated from the Red Sea, one of the warmest marine basins on earth. According to MTE, the response to temperature can be assessed by the so-called activation energy (E_a). Simply put, the higher their E_a, the higher their temperature dependence of μ. In theory, in nutrient-sufficient conditions, temperature increases biochemical reaction rates exponentially, resulting in a similar exponential dependence for the integrated growth rates of autotrophs on one side and heterotrophs on the other one (Brown et al., 2004). Although μ steadily increased with warming in all four bacterial groups (**Figure 5**), the E_a values for heterotrophs were not fully consistent. The mean activation energy of *Synechococcus* in experiment C (very similar to the other two experiments estimates) was 0.34 eV ± 0.02, statistically indistinguishable from the value hypothesized for autotrophs by the MTE (0.32 eV). However, the μ E_a values for the three top most abundant heterotrophic bacteria OTUs (making on average 2/3 of the total bacterial counts) were notably more variable, resulting in a huge variation in the specific growth rate response to temperature (**Figure 5**). Although their mean activation energy (0.67 eV) also virtually equaled the corresponding MTE value for heterotrophs (0.65 eV), this value resulted from a similar E_a in the case of *Paracoccus* (0.71 ± 0.04 eV) but higher for *Muricauda* (1.15 ± 0.20 eV) and much lower value for *Marinobacter* (0.16 ± 0.03 eV). These values indicate that *Muricauda* μ was most affected by temperature while *Marinobacter* was the least. The negative relationship between the specific growth rates of the three heterotrophic bacterial groups (we chose to represent 28°C as an example, but it held for all other experimental temperatures) and their estimated activation energies was puzzling (**Figure 6**). One could argue that for temperature to have the predicted effect (i.e., E_a = 0.32 eV for *Synechococcus* and 0.65 eV for the heterotrophs), tropical bacteria should not be neither bottom-up nor top-down controlled (Morán et al., 2018). Theoretically, in the absence of predators, autotrophic and heterotrophic bacterial needs for carbon and all other nutrients required for growth should be met by the culture medium. It may have not been the case for *Marinobacter*, either because some key nutrient or vitamin was in short supply although its μ-values were very high. Indeed, the specific growth rates of the three heterotrophic bacteria assessed here were however much higher than the bulk values measured in coastal Red Sea waters, even for the presumably copiotrophic high nucleic acid (HNA) content group of cells (1.01–2.33 d^{-1}, Silva et al., 2018), suggesting that not only nutrient conditions affects E_a values (Arandia-Gorostidi et al., 2017; Morán et al., 2018), but that the temperature dependence is also species-specific and related to the identity of the heterotrophic bacteria living in close association with the same cyanobacterium.

An increase in temperature within the normal, physiological range or thermal niche will increase population growth rate when resources are abundant and densities are low (Brown et al., 2004; Sherman et al., 2016), while increasing temperature is hypothesized to decrease the carrying capacity (Brown et al., 2004; Bernhardt et al., 2018). The MTE thus predicts an inverse relationship between carrying capacity and temperature, due to the fact that higher temperature will raise the metabolic rate, so that each individual will use resources at a higher rate (Savage et al., 2004). However, our results showed that each bacterial group responded differently in terms of the temperature response of their carrying capacity. The negative effect of temperature on the maximum abundance of *Synechococcus* and *Paracoccus* followed perfectly the MTE prediction whereas *Marinobacter* showed a significant yet opposite trend (**Supplementary Figure 7A**). In previous studies with mixed heterotrophic bacterial assemblages in temperate waters, the carrying capacity was shown to increase rather than decrease with temperature, an observation linked to the nutrient availability along the seasonal cycle (Huete-Stauffer et al., 2015; Arandia-Gorostidi et al., 2017). Similar to the specific growth rates discussed above, species-specific responses appear as critical, preventing the lack of a coherent response shown by the three heterotrophic bacteria that dominated the RS9907 co-culture, even if we assumed nutrient-sufficient conditions.

The temperature-size rule (TSR, Savage et al., 2004) is another general ecological theory that only relatively recently has been tested with marine planktonic organisms (Daufresne et al., 2009; Morán et al., 2010, 2015; Forster et al., 2012; Huete-Stauffer et al., 2016). Cell size reduction could be an acclimation response to conditions that imply increasing population growth rate, and to compensate for the increasing demand for resources (Atkinson et al., 2003). It is generally thought that resources can be obtained more effectively in smaller cells, hence facilitating their growth with increasing temperatures (Atkinson and Sibly, 1997). An exception to the TSR may arise if resource limitations are removed or reduced as a consequence of warming (Atkinson et al., 2003). Remarkably, although a trend toward generalized larger sizes in the exponential phase with higher temperature was shown also by the three major heterotrophic bacteria OTUs (**Supplementary Figure 8**), *Synechococcus* was the only species in our co-culture showing a consistent, highly significant variation of bacterial cell size with temperature, but it was exactly the opposite of the TSR prediction. Similar to our study, a strong positive correlation was found between *in situ* temperature and the mean cell size of a natural population of *Synechococcus* characterized by low PE fluorescence in shallow Red Sea waters (**Supplementary Figure 9**). A recent study with data from the experiment labeled as A in this study provided an alternative explanation for the fact that RS9907 did not show the expected negative relationship between cell size and temperature at any time point during the diel cycle. Palacio et al. (2020) however demonstrated that the mean cell size of *Synechococcuss* new-born cells followed the predicted negative correlation with temperature suggested by TSR. Cyanobacteria cells became larger during the first period of light cycle due to the biomass accumulation induced by photosynthesis and became smaller during the dark cycle as a result of mother cells division (Palacio et al., 2020).

It remains to be explained, though, the positive association between environmental temperature and the size of the most abundant *Synechococcus* group detected by flow cytometry in the shallow waters of KAUST Harbor (**Supplementary Figure 9**), which was sampled always at the same time (ca. 9:00 a.m. local time) (Sabbagh et al., 2020). One potential explanation would involve the high salinities found in the Red Sea, which can induce the internal accumulation of osmolytes for protection, resulting in larger cell volumes (Brown, 1976). In other organisms such as the anemone *Aiptasia* (Gegner et al., 2017) it has been shown that this strategy to cope with high salinities can also been advantageous for high temperatures. A significant difference was determined in G1 phase (bacterial growth) and S phase (DNA synthesis) duration of a freshwater *Synechococcus* strain cell cycle among different salinity conditions (0, 10, 20, and 30 of salinity), but G2 phase (preparation for bacterial cell division) duration did not vary. Mean cell sizes of this freshwater *Synechococcus* strain after acclimating to higher salinity became larger, and the doubling time was also longer at the highest salinity (Bemal and Anil, 2018). These *Synechococcus* cells were found to investing more energy in producing exopolysaccharides than on cell division, resulting in reduced growth rate but larger cell size (Bemal and Anil, 2018).

In conclusion, the assessed *Synechococcus* strain and their associated heterotrophic bacteria responded to experimental warming differently, without any clear, common pattern for heterotrophs, suggesting that there will be no universal response of bacterioplankton assemblages to global warming. Overall, our results were poorly described by the MTE and TSR. Within the assessed temperature range (24–34°C), which is typical for a substantial portion of the Red Sea (Raitsos et al., 2013; Chaidez et al., 2017), increasing temperature resulted in higher specific growth rate and cell size for *Synechococcus,* reaching their maximum at 33°C, similar to other tropical and sup-tropical cyanobacteria strains that showed maximum growth rates at relatively higher temperatures (32–34°C) (Pittera et al., 2014). However, heterotrophic bacterial specific growth rates did also increase with warming, cell sizes did not change in a coherent way. Our data suggest that temperatures equal or higher than 34°C can already be detrimental for the cyanobacterium strain and one major heterotrophic bacteria player (*Muricauda*) in the RS9907 co-culture. The role of the three dominant heterotrophic bacteria in this cyanobacterial co-culture needs to be further analyzed and future metagenomic and metatranscriptomic studies may allow a deeper understanding on the metabolic interactions between autotrophic and heterotrophic bacteria in response to temperature.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

AL performed experiments A and C, analyzed data of all experiments, contributed in the experimental design of experiment C, and wrote the first draft of the manuscript. AP performed experiment A and contributed in the experimental design of experiment C. FG performed experiment B and contributed in editing the draft. GH contributed in editing the draft. MA collected time series data. ÁL-U designed experiment A. LA-S designed experiment A and contributed in editing the draft. P-YH supervised the study, contributed in editing the draft, and approved the final draft. XAGM designed experiment C, supervised the study, contributed in editing the draft, and approved the final draft. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by the King Abdullah University of Science and Technology (KAUST) through the research baseline funding provided to XAGM.

ACKNOWLEDGMENTS

We acknowledge the Spanish Ministry of Economy and Competitiveness (MINECO) to support LA-S with a Ramón y Cajal research contract (RYC-2012-11404), AP with a FPI Ph.D. fellowship and funding the Retos project TECCAM (CTM2014-58564-R). We would like to thank Dr. Tamara M. Huete-Stauffer for her assistance in the sampling and technical support and Eman Sabbagh for her assistance in the sampling. We would like also to thank KAUST Bioscience Core Lab (BCL) for sequencing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.612732/full#supplementary-material

REFERENCES

Allen, A. P., Gillooly, J. F., and Brown, J. H. (2005). Linking the global carbon cycle to individual metabolism. *Funct. Ecol.* 19, 202–213. doi: 10.1111/j.1365-2435.2005.00952.x

Amin, S. A., Green, D. H., Hart, M. C., Kupper, F. C., Sunda, W. G., and Carrano, C. J. (2009). Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17071–17076. doi: 10.1073/pnas.0905512106

Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., et al. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522, 98–101. doi: 10.1038/nature14488

Ansari, M. I., Harb, M., Jones, B., and Hong, P. Y. (2015). Molecular-based approaches to characterize coastal microbial community and their potential

relation to the trophic state of Red Sea. *Sci. Rep.* 5:9001. doi: 10.1038/srep09001

Arandia-Gorostidi, N., Huete-Stauffer, T. M., Alonso-Sáez, L., and Morán, X. A. G. (2017). Testing the metabolic theory of ecology with marine bacteria: different temperature sensitivity of major phylogenetic groups during the spring phytoplankton bloom. *Environ. Microbiol.* 19, 4493–4505. doi: 10.1111/1462-2920.13898

Atkinson, D. (1994). Temperature and organism size-A biological law for ectotherms? *Adv. Ecol. Res.* 25, 1–58. doi: 10.1016/s0065-2504(08)60212-3

Atkinson, D., and Sibly, R. M. (1997). Why are organisms usually bigger in colder environments? Making sense of a life history puzzle. *Trends Ecol. Evol.* 12, 235–239. doi: 10.1016/S0169-5347(97)01058-6

Atkinson, D., Ciotti, B. J., and Montagnes, D. J. (2003). Protists decrease in size linearly with temperature: ca. 2.5% degrees C(-1). *Proc. Biol. Sci.* 270, 2605–2611. doi: 10.1098/rspb.2003.2538

Azam, F., and Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* 5, 782–791. doi: 10.1038/nrmicro1747

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* 10, 257–263. doi: 10.3354/meps010257

Azevedo, R. B. R., French, V., and Partridge, L. (2002). Temperature modulates epidermal cell size in *Drosophila melanogaster*. *J. Insect Physiol.* 48, 231–237. doi: 10.1016/s0022-1910(01)00168-8

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., et al. (2006). Climate-driven trends in contemporary ocean productivity. *Nature* 444, 752–755. doi: 10.1038/nature05317

Bemal, S., and Anil, A. C. (2018). Effects of salinity on cellular growth and exopolysaccharide production of freshwater Synechococcus strain CCAP1405. *J. Plankton Res.* 40, 46–58. doi: 10.1093/plankt/fbx064

Bernhardt, J. R., Sunday, J. M., and O'Connor, M. I. (2018). Metabolic theory and the temperature-size rule explain the temperature dependence of population carrying capacity. *Am. Nat.* 192, 687–697. doi: 10.1086/700114

Bolch, C. J. S., Bejoy, T. A., and Green, D. H. (2017). Bacterial associates modify growth dynamics of the Dinoflagellate Gymnodinium catenatum. *Front. Microbiol.* 8:670. doi: 10.3389/fmicb.2017.00670

Boyd, P. W., Rynearson, T. A., Armstrong, E. A., Fu, F., Hayashi, K., Hu, Z., et al. (2013). Marine phytoplankton temperature versus growth responses from polar to tropical waters–outcome of a scientific community-wide study. *PLoS One* 8:e63091. doi: 10.1371/journal.pone.0063091

Brown, A. (1976). Microbial water stress. *Bacteriol. Rev.* 40:803. doi: 10.1128/mmbr.40.4.803-846.1976

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., and West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology* 85, 1771–1789. doi: 10.1890/03-9000

Bruns, A., Rohde, M., and Berthe-Corti, L. (2001). *Muricauda* ruestringensis gen. nov., sp. nov., a facultatively anaerobic, appendaged bacterium from German North Sea intertidal sediment. *Int. J. Syst. Evol. Microbiol.* 51, 1997–2006. doi: 10.1099/00207713-51-6-1997

Buchan, A., LeCleir, G. R., Gulvik, C. A., and Gonzalez, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* 12, 686–698. doi: 10.1038/nrmicro3326

Calvo-Díaz, A., and Morán, X. A. G. (2006). Seasonal dynamics of picoplankton in shelf waters of the southern Bay of Biscay. *Aquat. Microb. Ecol.* 42, 159–174. doi: 10.3354/ame042159

Chaidez, V., Dreano, D., Agusti, S., Duarte, C. M., and Hoteit, I. (2017). Decadal trends in Red Sea maximum surface temperature. *Sci. Rep.* 7:8144. doi: 10.1038/s41598-017-08146-z

Chen, B., and Laws, E. A. (2017). Is there a difference of temperature sensitivity between marine phytoplankton and heterotrophs? *Limnol. Oceanogr.* 62, 806–817. doi: 10.1002/lno.10462

Christie-Oleza, J. A., Armengaud, J., Guerin, P., and Scanlan, D. J. (2015a). Functional distinctness in the exoproteomes of marine *Synechococcus*. *Environ. Microbiol.* 17, 3781–3794. doi: 10.1111/1462-2920.12822

Christie-Oleza, J. A., Scanlan, D. J., and Armengaud, J. (2015b). "You produce while I clean up", a strategy revealed by exoproteomics during *Synechococcus-Roseobacter* interactions. *Proteomics* 15, 3454–3462. doi: 10.1002/pmic.201400562

Christie-Oleza, J. A., Sousoni, D., Lloyd, M., Armengaud, J., and Scanlan, D. J. (2017). Nutrient recycling facilitates long-term stability of marine microbial phototroph-heterotroph interactions. *Nat. Microbiol.* 2:17100. doi: 10.1038/nmicrobiol.2017.100

Coe, A., Ghizzoni, J., LeGault, K., Biller, S., Roggensack, S. E., and Chisholm, S. W. (2016). Survival of *Prochlorococcus* in extended darkness. *Limnol. Oceanogr.* 61, 1375–1388. doi: 10.1002/lno.10302

Coello-Camba, A., Diaz-Rua, R., Duarte, C. M., Irigoien, X., Pearman, J. K., Alam, I. S., et al. (2020). Picocyanobacteria community and cyanophage infection responses to nutrient enrichment in a mesocosms experiment in oligotrophic waters. *Front. Microbiol.* 11:1153. doi: 10.3389/fmicb.2020.01153

Coutinho, F., Tschoeke, D. A., Thompson, F., and Thompson, C. (2016). Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus*. *PeerJ* 4:e1522. doi: 10.7717/peerj.1522

Daufresne, M., Lengfellner, K., and Sommer, U. (2009). Global warming benefits the small in aquatic ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12788–12793. doi: 10.1073/pnas.0902080106

Doré, H., Farrant, G. K., Guyet, U., Haguait, J., Humily, F., Ratin, M., et al. (2020). Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria. *Front. Microbiol.* 11:567431. doi: 10.3389/fmicb.2020.567431

Du, H., Jiao, N., Hu, Y., and Zeng, Y. (2006). Diversity and distribution of pigmented heterotrophic bacteria in marine environments. *FEMS Microbiol. Ecol.* 57, 92–105. doi: 10.1111/j.1574-6941.2006.00090.x

Dufresne, A., Ostrowski, M., Scanlan, D. J., Garczarek, L., Mazard, S., Palenik, B. P., et al. (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 9:R90. doi: 10.1186/gb-2008-9-5-r90

Edwards, M., and Richardson, A. J. (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* 430, 881–884. doi: 10.1038/nature02808

Eppley, R. W. (1972). Temperature and phytoplankton growth in the sea. *Fish Bull.* 70, 1063–1085.

Farquhar, G. D., Caemmerer, S. V., and Berry, J. A. (1980). A biochemical model of photosynthetic CO 2 assimilation in leaves of C3 species. *Planta* 149, 78–90. doi: 10.1007/BF00386231

Felip, M., Andreatta, S., Sommaruga, R., Straskrabova, V., and Catalan, J. (2007). Suitability of flow cytometry for estimating bacterial biovolume in natural plankton samples: comparison with microscopy data. *Appl. Environ. Microbiol.* 73, 4508–4514. doi: 10.1128/AEM.00733-07

Forster, J., Hirst, A. G., and Atkinson, D. (2012). Warming-induced reductions in body size are greater in aquatic than terrestrial species. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19310–19314. doi: 10.1073/pnas.1210460109

Fuller, N. J., Marie, D., Partensky, F., Vaulot, D., Post, A. F., and Scanlan, D. J. (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl. Environ. Microbiol.* 69, 2430–2443. doi: 10.1128/aem.69.5.2430-2443.2003

Gegner, H. M., Ziegler, M., Radecker, N., Buitrago-Lopez, C., Aranda, M., and Voolstra, C. R. (2017). High salinity conveys thermotolerance in the coral model Aiptasia. *Biol. Open* 6, 1943–1948. doi: 10.1242/bio.028878

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., and Charnov, E. L. (2001). Effects of size and temperature on metabolic rate. *Science* 293, 2248–2251. doi: 10.1126/science.1061967

del Giorgio, P. A., and Duarte, C. M. (2002). Respiration in the open ocean. *Nature* 420, 379–384. doi: 10.1038/nature01165

Hadaidi, G., Ziegler, M., Shore-Maggio, A., Jensen, T., Aeby, G., and Voolstra, C. R. (2018). Ecological and molecular characterization of a coral black band disease outbreak in the Red Sea during a bleaching event. *PeerJ* 6:e5169. doi: 10.7717/peerj.5169

Han, J., Zhang, L., Wang, S., Yang, G., Zhao, L., and Pan, K. (2016). Co-culturing bacteria and microalgae in organic carbon containing medium. *J. Biol. Res.* 23:8. doi: 10.1186/s40709-016-0047-6

Hansell, D. A. (2013). Recalcitrant dissolved organic carbon fractions. *Ann. Rev. Mar. Sci.* 5, 421–445. doi: 10.1146/annurev-marine-120710-100757

Hayashi, S., Itoh, K., and Suyama, K. (2011). Growth of the cyanobacterium *Synechococcus* leopoliensis CCAP1405/1 on agar media in the presence of heterotrophic bacteria. *Microbes Environ.* 26, 120–127. doi: 10.1264/jsme2.me10193

Heinrich, B. (1977). Why have some animals evolved to regulate a high body temperature? *Am. Nat.* 111, 623–640. doi: 10.1002/ece3.5440

Huete-Stauffer, T. M., Arandia-Gorostidi, N., Alonso-Sáez, L., and Morán, X. A. (2016). Experimental warming decreases the average size and nucleic acid content of marine bacterial communities. *Front. Microbiol.* 7:730. doi: 10.3389/fmicb.2016.00730

Huete-Stauffer, T. M., Arandia-Gorostidi, N., Díaz-Pérez, L., and Morán, X. A. G. (2015). Temperature dependence of growth rates and carrying capacities of marine bacteria depart from metabolic theoretical predictions. *FEMS Microbiol. Ecol.* 91:fiv111. doi: 10.1093/femsec/fiv111

Imlay, J. A. (2003). Pathways of oxidative damage. *Annu. Rev. Microbiol.* 57, 395–418. doi: 10.1146/annurev.micro.57.030502.090938

Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., et al. (2010). Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* 8, 593–599. doi: 10.1038/nrmicro2386

Keeling, R. E., Kortzinger, A., and Gruber, N. (2010). Ocean deoxygenation in a warming world. *Ann. Rev. Mar. Sci.* 2, 199–229. doi: 10.1146/annurev.marine.010908.163855

Kirchman, D. L. (2016). Growth rates of microbes in the oceans. *Ann. Rev. Mar. Sci.* 8, 285–309. doi: 10.1146/annurev-marine-122414-033938

Lin, Y., Gifford, S., Ducklow, H., Schofield, O., and Cassara, N. (2019). Towards quantitative microbiome community profiling using internal standards. *Appl. Environ. Microbiol.* 85, e02634–18. doi: 10.1128/AEM

Liu, H., Jing, H., Wong, T. H., and Chen, B. (2014). Co-occurrence of phycocyanin- and phycoerythrin-rich *Synechococcus* in subtropical estuarine and coastal waters of Hong Kong. *Environ. Microbiol. Rep.* 6, 90–99. doi: 10.1111/1758-2229.12111

López-Urrutia, A., San Martin, E., Harris, R. P., and Irigoien, X. (2006). Scaling the metabolic balance of the oceans. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8739–8744. doi: 10.1073/pnas.0601137103

Malfatti, F., and Azam, F. (2009). Atomic force microscopy reveals microscale networks and possible symbioses among pelagic marine bacteria. *Aquat. Microb. Ecol.* 58, 1–14. doi: 10.3354/ame01355

Marie, D., Partensky, F., Vaulot, D., and Brussaard, C. (1999). Enumeration of phytoplankton, bacteria, and viruses in marine samples. *Curr. Protoc. Cytom* 10, 11.11.1–11.11.15.

Martinson, J. N. V., Pinkham, N. V., Peters, G. W., Cho, H., Heng, J., Rauch, M., et al. (2019). Rethinking gut microbiome residency and the *Enterobacteriaceae* in healthy human adults. *ISME J.* 13, 2306–2318. doi: 10.1038/s41396-019-0435-7

Mella-Flores, D., Six, C., Ratin, M., Partensky, F., Boutte, C., Le Corguille, G., et al. (2012). *Prochlorococcus* and *Synechococcus* have evolved different adaptive mechanisms to cope with light and UV stress. *Front. Microbiol.* 3:285. doi: 10.3389/fmicb.2012.00285

Moffett, J. W., and Zajiriou, O. C. (1990). An investigation of hydrogen peroxide chemistry in surface waters of Vineyard sound with H,180, and l8O2. *Limnol. Oceanogr.* 35, 1221–1229. doi: 10.4319/lo.1990.35.6.1221

Moffett, J. W., and Zika, R. G. (1987). Reaction kinetics of hydrogen peroxide with copper and iron in seawater. *Environ. Sci. Technol.* 21, 804–810. doi: 10.1021/es00162a012

Montoya, J. M., and Raffaelli, D. (2010). Climate change, biotic interactions and ecosystem services. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2013–2018. doi: 10.1098/rstb.2010.0114

Morán, X. A. G., Calvo-Díaz, A., Arandia-Gorostidi, N., and Huete-Stauffer, T. M. (2018). Temperature sensitivities of microbial plankton net growth rates are seasonally coherent and linked to nutrient availability. *Environ. Microbiol.* 20, 3798–3810. doi: 10.1111/1462-2920.14393

Morán, X. A. G., López-Urrutia, Á, Calvo-Díaz, A., and Li, W. K. W. (2010). Increasing importance of small phytoplankton in a warmer ocean. *Glob. Chang. Biol.* 16, 1137–1144. doi: 10.1111/j.1365-2486.2009.01960.x

Morán, X. A. G., Alonso-Sáez, L., Nogueira, E., Ducklow, H. W., González, N., López-Urrutia, A., et al. (2015). More, smaller bacteria in response to ocean's warming? *Proc. Biol. Sci.* 282:20150371. doi: 10.1098/rspb.2015.0371

Morowitz, H. J., Kostelnik, J. D., Yang, J., and Cody, G. D. (2000). The origin of intermediary metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7704–7708. doi: 10.1073/pnas.110153997

Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M., and Zinser, E. R. (2011). Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PLoS One* 6:e16805. doi: 10.1371/journal.pone.0016805

Morris, J. J., Kirkegaard, R., Szul, M. J., Johnson, Z. I., and Zinser, E. R. (2008). Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by "helper" heterotrophic bacteria. *Appl. Environ. Microbiol.* 74, 4530–4534. doi: 10.1128/AEM.02479-07

Myklestad, S. M. (1995). Release of extracellular products by phytoplankton with special emphasis on polysaccharides. *Sci. Tot. Environ.* 165, 155–164. doi: 10.1016/0048-9697(95)04549-G

Ngugi, D. K., Antunes, A., Brune, A., and Stingl, U. (2012). Biogeography of pelagic bacterioplankton across an antagonistic temperature-salinity gradient in the Red Sea. *Mol. Ecol.* 21, 388–405. doi: 10.1111/j.1365-294X.2011.05378.x

Palacio, A. S., Cabello, A. M., García, F. C., Labban, A., Morán, X. A. G., Garczarek, L., et al. (2020). Changes in population age-structure obscure the temperature-size rule in marine cyanobacteria. *Front. Microbiol.* 11:2059. doi: 10.3389/fmicb.2020.02059

Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2015). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi: 10.1111/1462-2920.13023

Pearman, J. K., Ellis, J., Irigoien, X., Sarma, Y. V. B., Jones, B. H., and Carvalho, S. (2017). Microbial planktonic communities in the Red Sea: high levels of spatial and temporal variability shaped by nutrient availability and turbulence. *Sci. Rep.* 7:6611. doi: 10.1038/s41598-017-06928-z

Pearman, J. K., Kurten, S., Sarma, Y. V., Jones, B. H., and Carvalho, S. (2016). Biodiversity patterns of plankton assemblages at the extremes of the Red Sea. *FEMS Microbiol. Ecol.* 92:fiw002. doi: 10.1093/femsec/fiw002

Pena, M. M. O., and Bullerjahn, G. S. (1995). The DpsA protein of Synechococcus sp. strain PCC7942 Is a DNA-binding hemoprotein. linkage of the Dps and bacterioferritin protein families. *J. Biol. Chem.* 270, 22478–22482. doi: 10.1074/jbc.270.38.22478

Pittera, J., Humily, F., Thorel, M., Grulois, D., Garczarek, L., and Six, C. (2014). Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J.* 8, 1221–1236. doi: 10.1038/ismej.2013.228

Post, A. F., Penno, S., Zandbank, K., Paytan, A., Huse, S. M., and Welch, D. M. (2011). Long term seasonal dynamics of *synechococcus* population structure in the gulf of aqaba, northern red sea. *Front. Microbiol.* 2:131. doi: 10.3389/fmicb.2011.00131

Props, R., Kerckhof, F. M., Rubbens, P., De Vrieze, J., Hernandez Sanabria, E., Waegeman, W., et al. (2017). Absolute quantification of microbial taxon abundances. *ISME J.* 11, 584–587. doi: 10.1038/ismej.2016.117

Raitsos, D. E., Pradhan, Y., Brewin, R. J., Stenchikov, G., and Hoteit, I. (2013). Remote sensing the phytoplankton seasonal succession of the Red Sea. *PLoS One* 8:e64909. doi: 10.1371/journal.pone.0064909

Sabbagh, E. I., Huete-Stauffer, T. M., Calleja, M. L. L., Silva, L., Viegas, M., and Morán, X. A. G. (2020). Weekly variations of viruses and heterotrophic nanoflagellates and their potential impact on bacterioplankton in shallow waters of the central Red Sea. *FEMS Microbiol. Ecol.* 96:fiaa033. doi: 10.1093/femsec/fiaa033

Sarmento, H., Montoya, J. M., Vázquez-Domínguez, E., Vaqué, D., and Gasol, J. M. (2010). Warming effects on marine microbial food web processes: how far can we go when it comes to predictions? *Philos. Trans. R Soc. Lond. B Biol. Sci.* 365, 2137–2149. doi: 10.1098/rstb.2010.0045

Savage, V. M., Gillooly, J. F., Brown, J. H., West, G. B., and Charnov, E. L. (2004). Effects of body size and temperature on population growth. *Am. Nat.* 163, 429–441. doi: 10.1086/381872

Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* 73, 249–299. doi: 10.1128/MMBR.00035-08

Sherman, E., Moore, J. K., Primeau, F., and Tanouye, D. (2016). Temperature influence on phytoplankton community growth rates. *Global Biogeochem Cycles* 30, 550–559. doi: 10.1002/2015gb005272

Silva, L., Calleja, M. L., Huete-Stauffer, T. M., Ivetic, S., Ansari, M. I., Viegas, M., et al. (2018). Low abundances but high growth rates of coastal heterotrophic

bacteria in the red sea. *Front. Microbiol.* 9:3244. doi: 10.3389/fmicb.2018.03244

Sinsabaugh, R. I., and Shah, J. J. F. (2010). Integrating resource utilization and temperature in metabolic scaling of riverine bacterial production. *Ecology* 91, 1455–1465. doi: 10.1890/08-2192.1

Six, C., Thomas, J. C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., et al. (2007). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol.* 8:R259. doi: 10.1186/gb-2007-8-12-r259

Willis, C., Desai, D., and LaRoche, J. (2019). Influence of 16S rRNA variable region on perceived diversity of marine microbial communities of the Northern North Atlantic. *FEMS Microbiol. Lett.* 366:fnz152. doi: 10.1093/femsle/fnz152

Wood, A. M., Horan, P. K., Muirhead, K., Phinney, D. A., Yentsch, C. M., and Waterbury, J. B. (1985). Discrimination between types of pigments in marine *Synechococcus* spp. by scanning spectroscopy, epifluorescence microscopy, and flow cytometry. *Limnol. Oceanogr.* 30, 1303–1315. doi: 10.4319/lo.1985.30.6.1303

Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., and Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* 16, 1099–1108. doi: 10.1101/gr.5322306

Zheng, Q., Wang, Y., Chen, C., Wang, Y., Xia, X., Fu, Y., et al. (2011). *Paracoccus* beibuensis sp. nov., isolated from the South China Sea. *Curr. Microbiol.* 62, 710–714. doi: 10.1007/s00284-010-9768-1

Zheng, Q., Wang, Y., Xie, R., Lang, A. S., Liu, Y., Lu, J., et al. (2018). Dynamics of heterotrophic bacterial assemblages within *Synechococcus* cultures. *Appl. Environ. Microbiol.* 84, e01517–17. doi: 10.1128/AEM.01517-17

Zinser, E. R. (2018). Cross-protection from hydrogen peroxide by helper microbes: the impacts on the cyanobacterium *Prochlorococcus* and other beneficiaries in marine communities. *Environ. Microbiol. Rep.* 10, 399–411. doi: 10.1111/1758-2229.12625

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 29 September 2021
doi: 10.3389/fmicb.2021.682651

Diel-Regulated Transcriptional Cascades of Microbial Eukaryotes in the North Pacific Subtropical Gyre

*Ryan D. Groussman[1], Sacha N. Coesel[1], Bryndan P. Durham[1,2] and E. Virginia Armbrust[1]**

[1] School of Oceanography, University of Washington, Seattle, WA, United States, [2] Department of Biology, Genetics Institute, University of Florida, Gainesville, FL, United States

Open-ocean surface waters host a diverse community of single-celled eukaryotic plankton (protists) consisting of phototrophs, heterotrophs, and mixotrophs. The productivity and biomass of these organisms oscillate over diel cycles, and yet the underlying transcriptional processes are known for few members of the community. Here, we examined a 4-day diel time series of transcriptional abundance profiles for the protist community (0.2–100 μm in cell size) in the North Pacific Subtropical Gyre near Station ALOHA. *De novo* assembly of poly-A+ selected metatranscriptomes yielded over 30 million contigs with taxonomic and functional annotations assigned to 54 and 25% of translated contigs, respectively. The completeness of the resulting environmental eukaryotic taxonomic bins was assessed, and 48 genera were further evaluated for diel patterns in transcript abundances. These environmental transcriptome bins maintained reproducible temporal partitioning of total gene family abundances, with haptophyte and ochrophyte genera generally showing the greatest diel partitioning of their transcriptomes. The haptophyte *Phaeocystis* demonstrated the highest proportion of transcript diel periodicity, while most other protists had intermediate levels of periodicity regardless of their trophic status. Dinoflagellates, except for the parasitoid genus *Amoebophrya*, exhibit the fewest diel oscillations of transcript abundances. Diel-regulated gene families were enriched in key metabolic pathways; photosynthesis, carbon fixation, and fatty acid biosynthesis gene families had peak times concentrated around dawn, while gene families involved in protein turnover (proteasome and protein processing) are most active during the high intensity daylight hours. TCA cycle, oxidative phosphorylation and fatty acid degradation predominantly peaked near dusk. We identified temporal pathway enrichments unique to certain taxa, including assimilatory sulfate reduction at dawn in dictyophytes and signaling pathways at early evening in haptophytes, pointing to possible taxon-specific channels of carbon and nutrients through the microbial community. These results illustrate the synchrony of transcriptional regulation to the diel cycle and how the protist community of the North Pacific Subtropical Gyre structures their transcriptomes to guide the daily flux of matter and energy through the gyre ecosystem.

Keywords: metatranscriptome, phytoplankton, biological oceanography, environmental genomics, protists

OPEN ACCESS

Edited by:
Rachel Ann Foster, Stockholm University, Sweden

Reviewed by:
Aditi Sengupta, California Lutheran University, United States
Rachael Marie Morgan-Kiss, Miami University, United States
Emma Rocke, University of Cape Town, South Africa

****Correspondence:***
E. Virginia Armbrust
armbrust@uw.edu

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *18 March 2021*
Accepted: *11 August 2021*
Published: *29 September 2021*

Citation:
Groussman RD, Coesel SN, Durham BP and Armbrust EV (2021) Diel-Regulated Transcriptional Cascades of Microbial Eukaryotes in the North Pacific Subtropical Gyre. Front. Microbiol. 12:682651. doi: 10.3389/fmicb.2021.682651

INTRODUCTION

The North Pacific Subtropical Gyre (NPSG) is a warm and oligotrophic (nitrogen-limited) ecosystem that hosts a diverse community of phototrophic, heterotrophic, and mixotrophic microbial eukaryotes (protists) spanning over three orders of magnitude in cell sizes. The phylogenetically diverse eukaryotic phytoplankton (phototrophs) represents nearly half of the phytoplankton biomass of the NPSG and are composed primarily of organisms derived through secondary or tertiary endosymbiosis such as dinoflagellates, haptophytes, and ochrophytes (photosynthetic stramenopiles) (Alexander et al., 2015; Hu et al., 2018). Strictly phototrophic members of the NPSG eukaryotic phytoplankton consist of haptophytes including *Emiliania huxleyi* (Hernández et al., 2020) and diatoms such as *Rhizosolenia* and *Hemiaulus* (Villareal et al., 1993), two genera found in symbioses with nitrogen fixing cyanobacteria that bloom during sporadic injections of nutrients into the surface waters. The strictly heterotrophic (zooplankton) component of the protist community is dominated by the Alveolata supergroup (including dinoflagellates), as well as stramenopiles and Rhizaria (Rii, 2016; Hu et al., 2018). Many eukaryotic lineages within the NPSG have mixotrophic life strategies, adjusting their relative balance of photosynthesis and phagotrophy to changing light and nutrient conditions (Mitra et al., 2016; Lambert et al., 2021). In addition, mixotrophs can be distinguished between constitutive (vertical inheritance of plastids) and non-constitutive (kleptoplastic, or acquisition of plastids from prey) (Mitra et al., 2016). Members of haptophyte, ochrophyte, and dinoflagellate lineages are constitutive mixotrophs (Faure et al., 2019), with evidence that they can graze on picocyanobacteria in the NPSG (Frias-Lopez et al., 2009). In the well-lit and low nutrient conditions of the NPSG, mixotrophy may be advantageous (Rothhaupt, 1996), and the gene family abundance profiles of many environmental protist species suggest widespread mixotrophy (Lambert et al., 2021). Some ciliates, such as *Strombidium*, are non-constitutive mixotrophs that retain the plastid of their consumed prey (Stoecker et al., 2009; Faure et al., 2019).

The daily cycles of light and darkness synchronize cell growth and division across the diverse members of the microbial communities within the sunlit waters of the NPSG. For example, the phototrophic cyanobacteria *Prochlorococcus* displays reproducible increases in mean cell diameter over the day as cells fix carbon into biomass, followed by decreases during the night as cells undergo cell division (Vaulot and Marie, 1999; Ribalet et al., 2015). From day to day there is little variation in the total cellular abundance of *Prochlorococcus*, indicating that growth rates roughly equal loss rates in an ecological system finely tuned to the daily solar cycle (Ribalet et al., 2015). The tight coupling of growth and division to the light/dark cycle is also clear within small eukaryotic protists (<10 μm size fraction), although the different species underlying the observed changes in biomass cannot be distinguished based on the optical measures (Freitas et al., 2020). Heterotrophic bacteria also display oscillating waves of species-specific transcriptional patterns over the diel cycle (Ottesen et al., 2014), potentially due to the daily cycle of phytoplankton production of organic matter.

Recent 'omics analyses have enhanced our understanding of the metabolic cascades resulting from synchrony to the daily light cycle in the NPSG. For example, transcript abundances in the haptophyte phototroph *E. huxleyi* underwent daily oscillations, with transcripts associated with carbon fixation and lipid synthesis proteins reaching a peak in abundance near dawn and at mid-day or dusk for those genes encoding respiration and lipid degradation, suggesting a cycle of energy-store biosynthesis and consumption (Hernández et al., 2020). A similar pattern was observed in the more nutrient-rich waters of the California Current upwelling, with oscillations in the abundance of diatom and green algae photosynthesis-related transcripts during the morning and cell division-related transcripts during the night (Kolody et al., 2019). Organisms in the NPSG are equipped with a diversified repertoire of photoreceptors (Coesel et al., 2021), which may allow them to sense and regulate their daily responses to changes in light conditions. Responses to the light cycle include the biosynthesis of energy-rich triacylglycerols during the day and their consumption during the night (Becker et al., 2018) and the replenishment of pigments during the night to compensate for photodegradation during the day (Becker et al., 2020). Moreover, a diel structuring of cross-kingdom interactions was demonstrated by the species-specific exchange of phytoplankton-produced organic sulfonated compounds, produced primarily by haptophytes and diatoms and consumed by heterotrophic bacteria (Durham et al., 2019).

Here, we examined diel transcriptional patterns across the microbial eukaryotes (protists) (0.2–100 μm in size) of the NPSG to evaluate how metabolic pathways may be synchronized across the microbial eukaryote community and the varied trophic states that comprise it. We hypothesized that eukaryotic gene expression would be strongly patterned by evolutionary lineage and trophic level, with phototrophs and mixotrophs expected to demonstrate the highest degree of diel periodicity in their transcriptome and, by inference, their metabolism. Analyses centered on the direct annotation of metatranscriptome-assembled contigs allowed us to investigate large-scale transcriptional patterns in abundant eukaryotic taxa and to identify functional metabolic processes operating on diel cycles.

MATERIALS AND METHODS

Cruise and Sample Collection

Duplicate samples for eukaryotic metatranscriptomes were collected from 15-m depth every 4 h (06:00, 10:00, 14:00, 18:00, 22:00, and 02:00 HST) over the course of 4 days on the R/V *Kilo Moana* cruise KM1513 (July and August 2015) approximately 100 km NE of Station ALOHA in the North Pacific Subtropical Gyre (see Wilson et al., 2017 for additional cruise details). For each of the 48 samples, 7 L of seawater was pre-filtered through a 100-μm nylon mesh and collected onto a 142-mm 0.2-μm polycarbonate filter using a peristaltic pump. Filters were immediately flash-frozen in liquid nitrogen and subsequently stored at −80°C until further processing.

Filters were extracted using the ToTALLY RNA Kit (Invitrogen, Carlsbad, CA, United States) with some modifications. Briefly, frozen filters were added to 50-ml falcon tubes containing 5 ml of denaturation solution and extraction beads (125-µl 100-µm zirconia beads, 125-µl 500-µm zirconia beads, and 250-µl 425–600-µm silica glass beads). A set of 14 internal mRNA standards was added to the extraction buffer for each sample to generate quantitative transcript inventories; these standards were synthesized as previously described (Satinsky et al., 2013), with the exception that eight standards were synthesized with poly(A) tails to mimic eukaryotic mRNAs. Total extracted RNA was treated with DNase I (Ambion, Grand Island, NY, United States) and purified with DNase inactivation reagent (Ambion). Eukaryotic mRNAs were poly(A)-selected, sheared to ∼225-bp fragments, and used to construct TruSeq cDNA libraries according to the Illumina TruSeq® RNA Sample Preparation v2 Guide for paired-end (2 × 150 bp) sequencing using the Illumina NextSeq 500 sequencing platform.

DNA Sequence Quality Control and *de novo* Assembly

Raw Illumina sequence reads were quality controlled with trimmomatic v0.36 (Bolger et al., 2014, parameters: *MAXINFO:135:0.5, LEADING:3, TRAILING:3, MINLEN:60,* and *AVGQUAL:20*). A total of 2,426,923,906 merged paired-end sequences were generated with a median length of ∼240 bp. Sequences were pooled for each of the 24 sampling times, and the 24 pooled samples were assembled using the Trinity *de novo* transcriptome assembler v2.3.2 (Grabherr et al., 2011, parameters: –normalize_reads –min_kmer_cov 2 –min_contig_length 300) on the Pittsburgh Supercomputing Center's Bridges Large Memory system. Trinity assembly yielded 52,489,585 total contigs from all 24 assemblies.

Quality Control of Assemblies, Translation, and Longest Frame Selection

The raw assemblies were quality controlled with Transrate v1.0.3 (Smith-Unna et al., 2016) to check contigs for chimeras, structural errors, and base errors, using their paired-end assembly method (parameters: –assembly $sample.fasta –left $left_reads –right $right_reads). A total of 31,284,431 contigs (59.6% of the raw pool) passed the optimized assembly score threshold and were retained for further analysis. The quality-controlled contigs were translated in six frames with transeq vEMBOSS:6.6.0.059 (Rice et al., 2000) using the Standard Genetic Code. The longest (or multiple frames if of equal lengths) open reading frame from each contig was retained for downstream analyses. A total of 32,536,410 translated frames that passed these criteria were retained.

Clustering

The 24 peptide assemblies were merged and clustered at the 99% identity threshold level with linclust within the MMseqs2 package (version 31e25cb081a874f225d443eec307a6254f06a291, Steinegger and Söding, 2018, –min-seq-id 0.99). A total of 30,015,008 peptide sequences (92% of input sequences) were retained as cluster representatives and used for further analysis.

Annotation

Following clustering at 99% identity, representative contigs were annotated for taxonomy and function. Contigs were annotated against the curated MarineRefII reference database[1] of 641 marine eukaryotes and prokaryotes, including the MMETSP transcriptomes (Keeling et al., 2014), and supplemented with available marine animal, fungal, choanoflagellate, and viral sequences (Coesel et al., 2021). Assembled contigs were aligned to the reference database using DIAMOND v 0.9.18 (Buchfink et al., 2014, parameters: –sensitive -b 65 -c 1 -e 1e-5 –top 10 -f 100). The Lowest Common Ancestor (LCA) was estimated using the LCA algorithm in DIAMOND in conjunction with NCBI taxonomy. The frame with the lowest (best) *e*-value was retained if multiple frames of a contig received annotations. A total of 15,302,768 contigs were assigned an NCBI tax_id (51.0%). Contigs assigned to the same NCBI taxon or daughter nodes were defined as members of the same environmental taxon "bin." To determine the putative function of each contig, we used hmmsearch, from HMMER 3.1b2 (Eddy, 2011, parameters: -T 30 –incT 30), to assign KEGG Orthology IDs (KOs) to contigs using 22,247 hmm profiles from KOfam ver. 2019-03-20 (Aramaki et al., 2020). The profile with the highest bitscore was retained for those contigs that mapped to more than one KOfam profile. If multiple frames of a contig received annotations, the frame with the highest annotation bitscore was retained. A total of 7,707,191 contigs were assigned an KEGG KO (25.7%).

Quantification and Normalization of Abundance

The clustered contig representatives were quantified by alignment of their nucleotide sequences against the paired-end reads using kallisto v0.43.1 (Bray et al., 2016, parameters: quant –rf-stranded -b 40). We normalized contigs abundance to fragments per kilobase of transcript per million total reads (FPKM), using total read values mapped to each taxonomic bin, rather than the total library size, as the denominator, "M." FPKM values for contigs with the same taxonomy and functional annotation combinations (NCBI tax_id and KEGG KO) were summed. Environmental taxa bins were integrated down taxonomic levels, summing abundance values from lower-ranking nodes in the NCBI taxonomy.

NMDS Ordination

Non-metric multidimensional scaling (NMDS) ordination was used to reduce taxonomic, temporal, functional, and abundance information into three-dimensional space. The input was a matrix of observations consisting of each of the 48 genus-level taxa for each of the 24 time points (1,152 total observations). The features for each observation were the *in silico* normalized counts for the 6,925 KOfams present in >5% of the observations. The features within observations were normalized such that the

[1] http://roseobase.org/data/

row sums equal 1. The metaMDS function in the R package "vegan" version 2.5-5 (Oksanen et al., 2019, parameters: $k = 3$, trymax = 100) was used to compute the Bray–Curtis distance matrix and find a solution between runs. A solution for this ordination with two dimensions was not achievable. The three-dimensional NMDS ordination results were visualized with the R package "plotly" (Sievert, 2018). NMDS ordinations were also individually generated for the 48 genera. The individual ordinations were resolved in two dimensions (parameters: $k = 2$). Mean stress of 48 NMDS = 0.127 ± 0.028 stdev.

Determining Significant Diel Periodicity

Significant periodicity of gene family abundances for each genus was determined with the Rhythmicity Analysis Incorporating Non-parametric Methods (RAIN) package implemented in R (Thaben and Westermark, 2014). The p-values from RAIN analyses were ranked and corrected at an FDR < 0.05 using the Benjamini–Hochberg false-discovery rate method (Benjamini and Hochberg, 1995), as implemented in Ottesen et al. (2014).

Enrichment of Significantly Diel Gene Families in KEGG Pathways

The KEGG pathways and their associated knums (gene family identifiers representing a KOfam) were parsed using the R package KEGGREST (Tenenbaum, 2016) to access the KEGG database. Only pathways with associated knums were used. A contingency matrix for each genus-pathway-time combination (a total of 124,704 combinations of contingency matrices for genus-level analysis, and 106,518 for order-level analysis) was constructed for the test and populated with the appropriate counts: "knum in pathway and is diel," "knum is in pathway and is not diel," "knum is not in pathway and is diel," and "knum is not in pathway and is not diel." The "diel" status of each knum was determined from the RAIN results, above. We removed contingency matrices with no identified knums in the pathway, reducing our number of tests to 108,288 for the genus-level analysis and 93,114 for the order-level analysis. We used Fisher's Exact Test on these matrices to determine enrichment, combined with a Benjamini–Hochberg multiple comparison correction and false discovery rate of less than 5% (genera-level adjusted maximum p-value of <3.412673e-05, order-level adjusted maximum p-value of 5.792821e-05) (Benjamini and Hochberg, 1995), both executed within R.

RESULTS

Taxonomic and Functional Composition of Environmental Transcriptome Bins

We examined the transcriptional profiles of eukaryotic microbes (protists) over the diel cycle by collecting size-fractionated (0.2–100 μm) RNA samples every 4 h (at 06:00, 10:00, 14:00, 18:00, 22:00, and 02:00 HST) over 4 consecutive days in the oligotrophic North Pacific Subtropical Gyre (NPSG), ∼100 km NE of Station ALOHA. Local sunrise was at ∼06:00 and sunset was at ∼18:00 HST. Surface illumination intensities reached over 2,000 μmol m^{-2} s^{-1} between 10:00 and 14:00 HST over the 4-day sampling period, during which the picoeukaryotic phytoplankton grew and divided on an oscillating daily basis, as estimated from continuous flow cytometry measurements (Coesel et al., 2021). Water column properties during the cruise confirm a warm (26.6–27.04°C with some diel variability) and nitrogen-deplete environment (nitrate + nitrite 8 ± 4 nmol L^{-1}) (Wilson et al., 2017). Casts were taken at 15 m depth, with the research vessel (KM1513) tracking a Lagrangian drogue placed at 15 m to allow repeat sampling of the planktonic community from the same water mass. Illumina deep sequencing of libraries created from poly-A+ selected RNA yielded a total of ∼2.4 billion transcript fragments (merged paired-end reads) with an average merged length of ∼240 bp (**Table 1**). *De novo* assembly of the metatranscriptome sequences generated about 52 million nucleotide contigs. Subsequent quality control, translation, frame selection, and clustering at 99% amino acid identity (**Table 1**) yielded 30 million amino acid sequences (hereafter referred to as "contigs"), with an N50 of 423 base pairs (141 amino acid residues, **Supplementary Figure 1** and **Table 1**).

The translated contigs were annotated in two ways. First, taxonomic identity and rank within the NCBI taxonomic framework were determined by mapping contigs against a reference database of 18.5 million predicted protein sequences from 554 marine eukaryotes, bacteria, archaea, and viruses (Coesel et al., 2021) and estimating the Lowest Common Ancestor (LCA) of matches. Placement to any taxonomic level was possible for 16 million contigs (51% of total; **Figure 1A**, **Table 1**, and **Supplementary Figure 2**); the remaining 49% received no taxonomic annotation. Bacteria or Archaea were assigned to 55,099 and 740 contigs, respectively; these contigs were not considered further in this study. Second, potential function was assigned by mapping the contigs against the Kyoto Encyclopedia of Gene and Genomes (KEGG) database of orthogenes (KOfams) using HMM profiles (Aramaki et al., 2020). A total of 7.7 million contigs (25.7%) were assigned to a putative KOfam and the associated KO term, with 13,765 total KOfams identified among all contigs (**Figure 1A** and **Table 1**). Approximately 5.2 million contigs were annotated to a genus rank or lower (**Figure 1A** and **Table 1**), with a

TABLE 1 | Sequence data metrics.

Sequence data	Count	Type
Merged read pairs	2,426,923,906	Nt
Raw contigs	52,489,585	Nt
QCed contigs (transrate)	31,284,431	Nt
Translated, clustered contigs	30,015,008	Aa
Contigs w/any taxonomy	16,061,543	Aa
Contigs w/any KEGG function	7,707,191	Aa
Genus-level taxonomy	5,181,384	Aa
Genus-level w/KEGG function	1,390,232	Aa

Volume of sequence data by short reads, assembled contigs, and annotated contigs. Type refers to nucleotide ("nt") or translated amino acid ("aa") sequence. Genus-level taxonomy and functional counts include contigs assigned to nested daughter taxonomies.

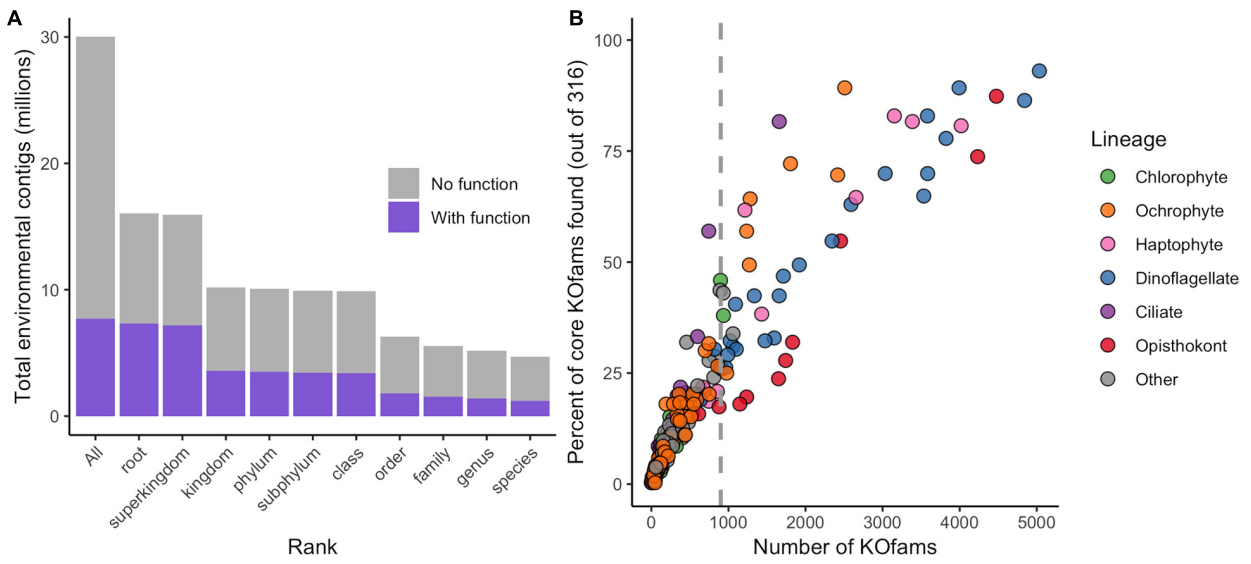


FIGURE 1 | Annotation of metatranscriptome assemblies by taxonomy, function, and abundance. **(A)** Cumulative number of contigs placed at each rank or lower. "All" includes taxonomically unassigned contigs. **(B)** Completeness of genus-level taxonomic bins based on total number of KOfams and percent of the 316 "core KOfams." Circles represent genus-level taxonomic bins, colored by lineage as in legend. Dashed line represents cutoff criteria of 900 detected KOs for subsequent analyses.

subset of 1.4 million contigs that also received a putative KEGG function. This set of 1.4 million contigs with putative genus-level taxonomy and function were grouped based on their assigned genus-level taxonomy to create environmental taxonomic bins and used for subsequent analyses. Two hundred thirty-one environmental eukaryotic genera bins were represented by at least one functional KOfam assignment (**Figure 1B**). A low proportion of reads aligned to reference sequences of metazoans (opisthokonts) larger in size than 100 μm, potentially reflecting varied biological sources including sloughed cells, gametes, and other life cycle stages.

We developed a metric of transcriptome completeness to identify a subset of well-represented environmental genera from this initial set of taxonomic bins for further analyses. We first estimated a minimum number of KOfams expected within marine eukaryotes by *de novo* mapping the proteomes of our 366 reference marine eukaryotes against 22,247 KOfam hmm profiles (ver. 2019-03-20; Aramaki et al., 2020). Three hundred fifty-five of the reference protists (over 96% of eukaryotic reference taxa) each contained 900 or more KOfams; only parasitic protists with a reduced gene content possessed significantly fewer KOfams (**Supplementary Figure 3** and **Supplementary Data Sheet 1**). A set of 316 KOfams were defined as "core KOfams" due to their presence in ∼95% of reference marine eukaryotes (**Supplementary Data Sheet 2**). The completeness of a given genus-level environmental taxonomic bin was estimated based on the percentage of core KOfams identified in the environmental bins (**Figure 1B** and **Supplementary Figure 3**). We constrained further analysis to 48 environmental genera bins that each had >900 detected KOfams. These 48 genera were grouped by eight higher-level lineages (**Supplementary Data Sheet 3**): dinoflagellates (n = 23), opisthokonts (n = 8), ochrophytes (photosynthetic stramenopiles, n = 7), haptophytes (n = 6), ciliates (n = 1), chlorophytes (n = 1), and two "other" lineages of kinetoplastids and oomycetes. These 48 genera corresponded to genera previously detected in the NPSG through metabarcoding (Hu et al., 2018) and collectively had an average of 54% (170 of 316) core KOfams positively identified (**Supplementary Data Sheet 3**). The highest proportion of core KOfams were detected in the taxonomic bin most closely related to the dinoflagellate *Karlodinium* (93%). Confidence values (*e*-value) for taxonomic assignments were assessed for all 48 genera, with the majority of contigs assigned to each bin receiving the highest-possible confidence value of 0 (**Supplementary Figure 4**). Some genera displayed a shallower distribution of *e*-value placements, reflecting a more likely entrainment of distantly related taxa to genus representatives. This is most noticeable in multicellular metazoans with single genus reference representatives (e.g., *Octopus*, *Salmo*, and *Orcinus*) and some dinoflagellates (*Amphidinium*, *Kryptoperidinium*, and *Lingulodinium*).

Dimensionality Reduction of Environmental Transcriptome Bins

Non-metric multidimensional scaling (NMDS) of Bray–Curtis dissimilarity was used to compare the similarity of transcript abundances for thousands of gene families across the 48 genus-level environmental bins over the 24 time points (**Figure 2A**). The input matrix consisted of each of the 48 genus-level environmental bins for each of the 24 time points (1,152 total observations) and the row-normalized number of transcripts associated with KOfams present in >5% of observations (6,925 total features). Several patterns emerged from the

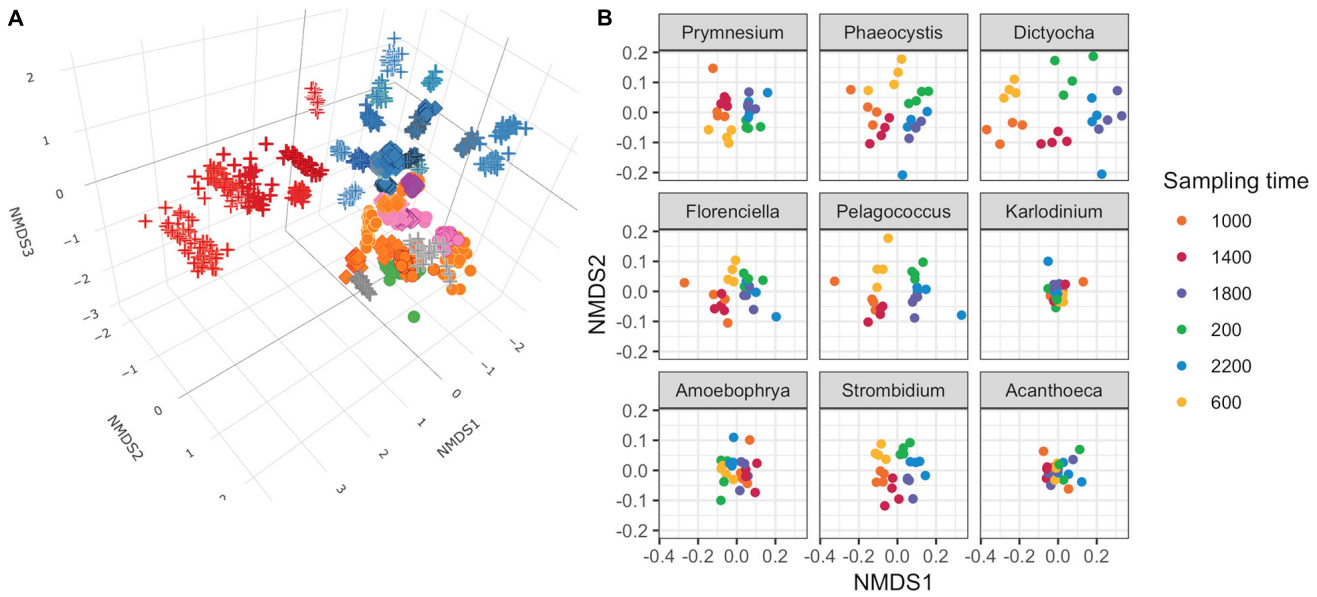


FIGURE 2 | Non-metric multidimensional scaling (NMDS) ordinations of transcript abundances from 48 genus-level environmental bins. **(A)** Three-dimensional ordination for 48 genus-level taxonomic bins meeting completeness criteria. Stress = 0.160. Each point is a discrete time point for each genus based on the mean transcript abundances for each KOfam from duplicate samples. Shapes indicate assumed trophic level: circle, phototroph; diamond, mixotroph; cross, heterotroph. Inner color corresponds to Lineage in legend from **Figure 1B**; outer color is unique to genera within a lineage. **(B)** NMDS ordination performed independently on gene families belonging to nine representative protist genera. Mean stress of nine independent NMDS = 0.114 ± 0.025 stdev.

resulting 3D NMDS ordination. First, the observations clustered together according to genus designation rather than time point, indicating that each taxonomic bin displayed a relatively distinct transcriptional fingerprint irrespective of the time of sampling. Second, phylogenetically related genera clustered near one another. One notable exception to this pattern was the environmental bin corresponding to the dinoflagellate genus *Amoebophrya*, a highly-derived genus of dinoflagellates with a parasitoid life cycle (Guillou et al., 2008). Third, both the dinoflagellate genera and opisthokont genera (seven metazoa and one choanoflagellate) clustered apart from other protists. Among the remaining protists, known phototrophic eukaryotes (diatoms, chlorophytes, and some haptophytes), heterotrophs, and potential mixotrophs formed distinct clusters, with the positioning of putative mixotrophs between the heterotroph and autotroph clusters reflecting their mixed metabolism.

We conducted independent NMDS ordinations on each of the 48 environmental genera bins to evaluate whether temporal partitioning of transcript abundances resolved within each environmental bin. Individual NMDS ordinations for the 48 genera were generated (mean stress = 0.114 ± 0.025 stdev) (**Supplementary Figure 5**). We highlighted a select subset of representative genera as representatives of the NPSG protist community in **Figure 2B**. We focused on nine representative environmental protist genera, based on their high sequence coverage, a high proportion of total and core KOfams, and their representation of different trophic states and evolutionary lineages (**Table 2** and **Supplementary Data Sheet 3**). Haptophytes with contrasting trophic modes were represented by the genus *Prymnesium* (order Prymnesiales), which contains known mixotrophic species (Faure et al., 2019) and the genus *Phaeocystis* (order Phaeocystales), which is considered strictly photosynthetic and can exist as a free-living flagellate or in a colonial form (Rousseau et al., 2007). Ochrophytes were represented by the silicoflagellate genus *Dictyocha* (order Dictyochales), the genus *Florenciella* (order Florenciellales), and the pico-eukaryotic (<2 μm cell size diameter) genus *Pelagococcus* (order Pelagomonadales). Both *Dictyocha* and *Florenciella* include known mixotrophic species (Quéguiner, 2016; Li et al., 2021), whereas *Pelagococcus* is thought to be strictly photosynthetic (Lewin et al., 1977). Dinoflagellates were represented by the genus *Karlodinium* (order Gymnodiniales), which contains mixotrophic species, similar to many of the dinoflagellate genera observed in this study (Faure et al., 2019). Isotopically labeled grazing experiments on *Prochlorococcus* and *Synechococcus* in the NPSG identified prymnesiophytes (Prymnesiophyceae), dictyophytes (Dictyochophyceae), and dinoflagellates (Dinoflagellata) as grazers of picocyanobacteria (Frias-Lopez et al., 2009). We also highlighted *Amoebophrya* (order Syndiniales), a parasitic dinoflagellate that infects eukaryotic host cells, because of its distinct lifestyle and because of its notable abundance in this environment as evidenced by the rRNA/rDNA libraries from the NPSG (Hu et al., 2018). The ciliate genus *Strombidium* (order Oligotrichida) can live heterotrophically as well as by non-constitutive mixotrophy through retention of plastids from engulfed prey (Stoecker et al., 2009). The choanoflagellate genus *Acanthoeca* (order Acanthoecida) is thought to be an obligate heterotroph.

The temporal partitioning of transcript abundances in the individual NMDS ordinations (**Figure 2B** and

TABLE 2 | General features of nine representative genera.

Genus	Order	Class	Putative trophic mode	References
Prymnesium	Prymnesiales	Haptophyta	Mixotrophic	Faure et al., 2019
Phaeocystis	Phaeocystales	Haptophyta	Phototrophic	Rousseau et al., 2007
Florenciella	Florenciellales	Ochrophyta	Mixotrophic	Quéguiner, 2016; Li et al., 2021
Dictyocha	Dictyochales	Ochrophyta	Photo- or mixotrophic	Quéguiner, 2016
Pelagococcus	Pelagomonadales	Ochrophyta	Phototrophic	Lewin et al., 1977
Karlodinium	Gymnodiniales	Dinophyceae	Mixotrophic	Faure et al., 2019
Amoebophrya	Syndiniales	Dinophyceae	Parasitic	Guillou et al., 2008
Acanthoeca	Acanthoecida	Choanoflagellata	Heterotrophic	
Strombidium	Oligotrichida	Spirotrichea	Hetero- or mixotrophic	Faure et al., 2019

Taxonomic levels are from the NCBI taxonomic framework.

Supplementary Figure 5) showed that sampling time was an important driver for the transcript ordination for a majority of genera, with samples clustered by collection time and organized in a clock-like fashion. Whereas most genera showed some form of diel partitioning, the transcriptional profiles of dinoflagellates such as *Karlodinium* and *Amoebophrya*, the choanoflagellate *Acanthoeca* (**Figure 2B** and **Supplementary Figure 5**), and the heterotrophic protists kinetoplastid *Neobodo* and the stramenopile *Phytophthora* (**Supplementary Figure 5**) were not distinguished by sampling time, indicating that not all organisms entrain their transcriptional activity to the diel cycle.

Diel Signatures of Environmental Transcriptome Bins

We determined the proportion of oscillating transcripts within each of the 48 environmental genera. The diel periodicity of transcript abundances across the 48 environmental genera was tested for a combined total of 103,904 KOfam gene families (RAIN analyses, maximum p-value of 0.0044, FDR < 0.05) (**Figure 3A** and **Supplementary Data Sheet 4**). Statistically significant diel periodicity in transcript abundance was detected for 9,153 gene families, with peaks in abundance assigned to one of the six sampling times: 06:00, 10:00, 14:00, 18:00, 22:00, and 02:00 HST (**Supplementary Figure 6** and **Supplementary Data Sheet 4**). Haptophytes displayed the highest proportions of diel-oscillating transcript abundances for different gene families (**Figure 3A** and **Supplementary Data Sheet 4**). About 34% of the *Phaeocystis* and about 18% of *Prymnesium* gene families underwent significant oscillations in transcript abundance. Most ochrophyte environmental genera displayed diel periodicity in transcript abundance in at least 15% of their gene families, with the highest proportion (21.5%) observed in the environmental silicoflagellate *Dictyocha* (**Figure 3A** and **Supplementary Data Sheet 3**). The dinoflagellates displayed a low proportion (average of less than 3%) of gene families with diel transcript abundance patterns, as did the purely heterotrophic opisthokonts (**Figure 3A**). The heterotrophic environmental stramenopile *Phytophthora* displayed a comparably low (3.2%) proportion of diel gene families. Three other heterotrophic organisms stood out in this analysis. The environmental genera of *Lepeophtheirus* (copepod), *Acanthoeca* (choanoflagellate), and *Strombidium* (ciliate) each displayed relatively high proportions of diel oscillations in transcript abundance across gene families (23, 16, and 20%, respectively), comparable to the haptophytes and non-diatom ochrophytes (**Figure 3A** and **Supplementary Data Sheet 3**). Thus, the extent of diel periodicity was not directly correlated with trophic mode and appeared instead to be taxa specific.

The transcription profiles of diel oscillating gene families (KOfams) were evaluated for the nine representative genera (**Figures 3A,B**) along with the remaining 41 environmental taxa (**Supplementary Figures 6, 7**). The representative environmental haptophyte *Phaeocystis* showed a transcriptional minimum at 10:00 (**Figure 3B**), coinciding with environmental light intensities reaching 2,000 μmol/m/s at the sea surface. Similar mid-morning minima are seen in *Acanthoeca* (**Figure 3B**) and other opisthokonts (**Supplementary Figures 6, 7**). Transcriptional patterns of the ochrophytes *Florenciella*, *Dictyocha*, and *Pelagococcus*, and the ciliate environmental genus *Strombidium* are more equally distributed across the 24-h diel cycle (**Figure 3B**). The dinoflagellate *Karlodinium* and most (21 of 23) other dinoflagellate environmental genera (**Supplementary Figures 6, 7**) displayed relatively minimal distinction of transcript abundances by time (**Figure 3B**), whereas the parasitic dinoflagellate *Amoebophrya* (**Figure 3B**) maintained a relatively high proportion (7.5%) of diel oscillating transcript abundances (**Supplementary Data Sheet 3**). We retrieved sufficient sequences with similarity to representative multicellular metazoan (animal) genera in our reference database (*Capitella*, *Orcinus*, *Octopus*, *Salmo*, *Lepeophtheirus*, *Nematostella*, and *Oikopleura*) to pass our cut-off criteria. Each of these environmental "genera" displayed some temporal partitioning (**Figure 3A** and **Supplementary Figures, 5–7**), particularly those identified as *Capitella* (annelid) and *Octopus* (mollusk). We constrained further analysis to protistan genera, recognizing that many metazoan taxa undergo substantial vertical migration over diurnal cycles. Diel vertical migration has also been observed in some species of motile phytoplankton (Shikata et al., 2015), but we assume that the relatively minor swimming speed of migrating protists is not a significant factor in the well-mixed surface layer of the NPSG. Overall, haptophyte and ochrophyte genera tended to display the highest proportions of diel oscillating gene families, regardless of

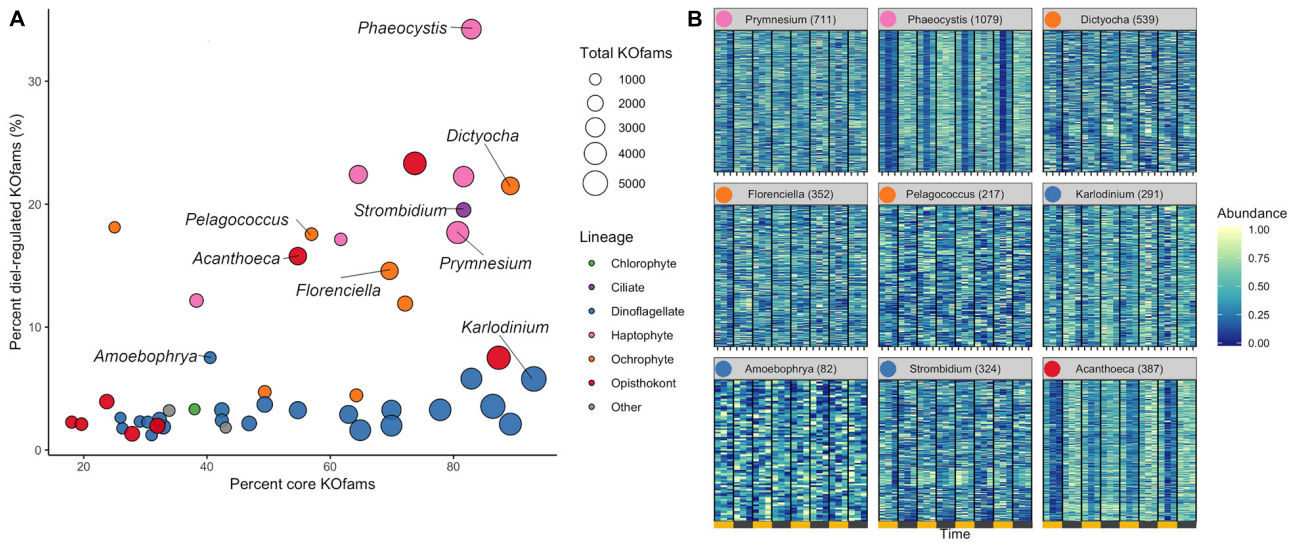


FIGURE 3 | Diel periodicity in environmental genus bins. **(A)** Estimated completeness and diel periodicity of gene family transcript abundance in 48 genera from **Figure 1B**. Each circle represents one environmental genus bin. The area is proportional to the total number of KOfams identified in the genus. The nine representative genera (**Figure 2B**) are labeled. **(B)** Normalized abundance heat map of significantly periodic gene families from nine representative protist genera. Yellow and gray bars denote light (06:00, 10:00, and 14:00 HST) and dark (18:00, 22:00, and 02:00 HST) periods, respectively. Numbers in parentheses indicate the number of significantly periodic gene families (FDR < 0.05). Each row corresponds to a gene family, ordered by hierarchical clustering of abundance patterns. Color corresponds to row-normalized abundance values for each gene family. Colored dot by genus name corresponds to Lineage from **(A)**. The order and presence of gene families is not maintained between facets.

whether the examined genera were primarily mixotrophic or photoautotrophic.

KEGG Pathways With Diel-Oscillating Transcript Levels

The observed variation in diel transcriptional patterns between environmental protistan taxa suggested targeted allocation of transcriptional resources to different functional processes over the diel cycle. We sought to identify specific pathways with strong temporal partitioning by using Fisher's Exact Test. We determined whether particular KEGG pathways were enriched in diel-oscillating gene families at each of the six time points by conducting tests on all unique taxa-time-pathway combinations. A total of 110,754 genus-time-pathway tests identified 78 significant taxa-time-pathway combination enrichments (BH < 0.05; **Table 3** and **Supplementary Data Sheet 5**); these enrichments represent 28 total enriched KEGG pathways out of 430 pathways tested. These pathways encompass varied metabolic pathways including central carbon metabolism, lipid biosynthesis and degradation, protein biosynthesis and turnover, organellar processes, and signaling. We focused on KEGG pathways enriched in at least two of the nine representative environmental genera: "Photosynthesis," "Carbon fixation in photosynthetic organisms," "Porphyrin and chlorophyll metabolism," "Proteasome," "Protein processing in endoplasmic reticulum," "TCA cycle," "Circadian entrainment," "Oxidative phosphorylation," and "Ribosome" (**Figure 4A**).

As expected, morning was characterized by enrichments in photosynthesis-related pathways. At dawn (06:00 HST), "Photosynthesis" and "Carbon fixation" pathways were enriched for diel oscillating gene families in phototrophs, including those with mixotrophic capabilities, with "Photosynthesis" the most frequently enriched pathway. A majority of gene families in the "Photosynthesis" pathway displayed peak transcript abundances at dawn (**Figure 4A**). At 10:00, "Porphyrin and chlorophyll metabolism" was the only pathway enriched in photosynthetic and mixotrophic lineages. The paucity of enrichments at 10:00 was consistent with the observed minimum in transcript peak across most environmental genera (**Supplementary Figures 6, 7**). One of the few whole pathway enrichments observed in dinoflagellate lineages was "Photosynthesis" in *Alexandrium* (**Supplementary Data Sheet 5**) at 10:00. Within the phagocytic taxa, *Strombidium* displayed significant enrichments in the "Ribosome" pathway at dawn (**Figure 4A**), similar to those seen in the sea squirt *Oikopleura* (**Supplementary Data Sheet 5**).

Later in the day, organisms shifted to protein processing and respiration-based pathways. The "Proteasome," "Protein processing in endoplasmic reticulum," and "TCA cycle" pathways were enriched at 14:00, with the latter pathway enriched in five haptophyte genera (**Figure 4A** and **Supplementary Data Sheet 5**). The haptophyte and ochrophyte genera had transcript abundance peaks in "Proteasome"-associated gene families in the afternoon (**Figure 4A**) whereas *Amoebophyra* and *Strombidium* displayed transcript abundance peaks in "TCA cycle"-associated gene families at this time point. By dusk (18:00), pathways involved in energy-yielding processes became transcriptionally prominent and included enrichments in "Oxidative phosphorylation" and additional enrichments in the TCA cycle. At this time the greatest number of pathway

TABLE 3 | KEGG pathway enrichment analysis at the Genus level.

Class/*genus*	Peak (HST)	KEGG pathway
Haptophyceae		
Prymnesium	06:00	*Carbon fixation*, glycolysis/gluconeogenesis, *photosynthesis*, pentose phosphate pathway
	14:00	*Proteasome, protein processing in ER*, antigen processing and presentation
	18:00	*TCA cycle, oxidative phosphorylation*, thermogenesis
	22:00	Circadian entrainment
	02:00	*Ribosome*
Phaeocystis	06:00	*Photosynthesis, carbon fixation*
	14:00	*Protein processing in ER*
	18:00	*TCA cycle, proteasome*, thermogenesis
	02:00	Lysosome
Dictyochophyceae		
Dictyocha	06:00	*Photosynthesis, carbon fixation*, glycolysis/gluconeogenesis
	18:00	Thermogenesis
	22:00	*Ribosome*
Florenciella	–	*No enriched pathways at FDR < 0.05*
Pelagococcus	18:00	Fatty acid degradation
	02:00	*Ribosome*
Dinophyceae		
Karlodinium	–	*No enriched pathways at FDR < 0.05*
Amoebophrya	14:00	*TCA cycle, oxidative phosphorylation*
Spirotrichea		
Strombidium	06:00	*Ribosome*
	10:00	Biosynthesis of unsaturated fatty acids
	14:00	*Carbon fixation, TCA cycle*, glyoxylate and dicarboxylate metabolism
Choanoflagellata		
Acanthoeca	02:00	Amoebiasis

Italics denote representative pathways plotted in ***Figure 4A****. Human Disease pathways not shown.*

enrichments was detected, although the timing of transcript abundance peaks for individual gene families varied across genera. The ochrophytes displayed peaks in transcript abundance for individual TCA-associated gene families throughout the day, while the haptophytes also displayed peaks for TCA-associated genes before and at dawn (**Figure 4A**). "Oxidative phosphorylation" peak times also occurred across other time points in ochrophytes, but primarily between dusk to dawn in *Prymnesium* and *Phaeocystis*, afternoon in *Amoebophrya*, and before noon in *Strombidium* (**Figure 4A**).

The night timepoints (22:00 and 02:00) were characterized by enrichments in "Circadian entrainment," "Oxidative phosphorylation," and "Ribosome" pathways in most of the nine genera. *Pelagococcus* did not display enrichment in Circadian entrainment at any time point, was enriched in "Oxidative phosphorylation" during the day rather than at night and was enriched in the "Ribosome" pathway at the 02:00 time point.

Additional pathways were enriched in specific genera (**Supplementary Data Sheet 5**). The greatest number of pathway enrichments at the genus level was found in the haptophyte *Prymnesium* (16), followed by other haptophytes and ochrophyte genera (38 and 17 pathway-time enrichments in total, respectively). "Fatty acid biosynthesis" and "Fatty acid degradation" were enriched in a subset of the genus-level analyses. For example, biosynthesis of unsaturated fatty acids was enriched in the ciliate *Strombidium* at 10:00 (**Supplementary Data Sheet 5**), alluding to a build-up of energy storage reserves during the day. Enrichment of the "Thermogenesis" pathway in several genera (*Phaeocystis*, *Prymnesium*, and *Dictyocha*) was driven by peaks in transcripts encoding mitochondrial-targeted proteins. In addition, a subset of enriched pathways at this time point is characterized as Human Diseases in KEGG; these pathways contain ubiquitous gene families such as calmodulin, calcium channels, cytochrome oxidase, ATPase, and some components of the TCA cycle (**Supplementary Data Sheet 5**).

Five pathway enrichments were detected in Dinoflagellate genera with the most in *Amoebophrya* ("TCA cycle" and "Oxidative phosphorylation" were both enriched at 14:00); the other dinoflagellates had either one or no enrichments. The choanoflagellate genus *Acanthoeca* had few significant peak times in the most commonly enriched diel pathways (**Figure 4A**), even though nearly 16% of *Acanthoeca* gene families displayed diel oscillations in transcript abundances, a value comparable to *Prymnesium* (18%) (**Figure 3A** and **Supplementary Data Sheets 3, 5**). The only significant enrichment attributed to *Acanthoeca* was the disease pathway "Amoebiasis," at 02:00.

The striking similarity in the overall patterns of the representative ochrophyte and haptophyte genera (**Figure 4A**) prompted an examination of transcription abundances for select gene families across the diel cycle, including those involved in "Photosynthesis" and "Chlorophyll metabolism," the "TCA cycle," "Ribosome," and "Proteasome" pathways (**Figure 4B**). We focused on those gene families with transcript abundances that oscillated over the diel cycle and were detected in at least three genera from the representative ochrophytes (*Dictyocha*, *Florenciella*, and *Pelagococcus*) or haptophytes (*Phaeocystis* and *Prymnesium*). Each of the photosynthesis gene family transcript abundance patterns was remarkably consistent across genera. In general, transcript abundance was highest at dawn with a decline through the day culminating in a dusk minimum. Gene family transcripts in the TCA cycle pathway were also tightly correlated with an inverse transcriptional pattern to photosynthesis-related transcripts, with sharp peaks at the dusk time point and 02:00 or 06:00 minima. Two gene families involved in either protein synthesis (small subunit ribosomal protein S1, rpsA) or protein degradation (proteasome activator subunit 3, PSME3) had generally opposing phases, with rpsA transcripts at higher abundances in the dark until dusk and PSME3 transcripts with higher daytime abundances. These results suggested that the ochrophyte and haptophyte phytoplankton lineages maintained similar diel regulation of genes within these pathways.

To identify potentially distinguishing features of the major phytoplankton lineages, we further examined pathway enrichments at the class level corresponding to Haptophyceae, Dictyophyceae, Pelagophyceae, and Dinophyceae (**Table 4**). These four classes are inclusive of 6, 23, 1, and 3 of the 48 genera meeting cutoff criteria, respectively. A total of 45 KEGG pathway enrichments over the diel cycle were detected for Haptophyceae;

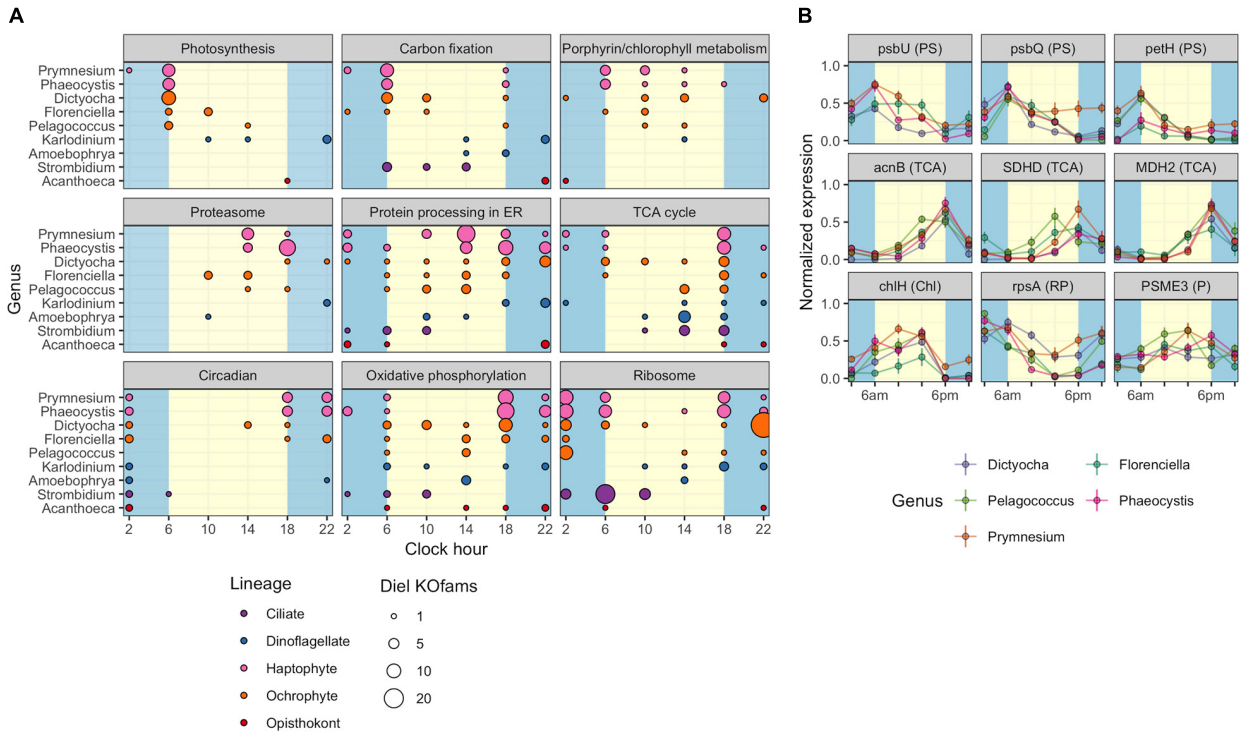


FIGURE 4 | Diel periodicity of metabolic pathways and gene families in select genera. **(A)** Peak times for gene families in select KEGG pathways for nine genera. Circles indicate peak timing for KOfams with significant periodicity (FDR < 0.05), with circle size scaled by the number of KOfams peaking at a given time. Blue and yellow shading denote night and day periods, respectively. **(B)** Temporal abundance of nine gene families from pathways in **(A)**. Haptophytes and ochrophytes from **(A)** are shown. Transcript abundances are min–max normalized. Lines connect the mean expression level across all identical sampling times (two replicates and 4 days). Vertical bars are standard error (n = 8). Blue and yellow shading denote night and day periods, respectively. Full protein names are psbU, photosystem II PsbU protein; psbQ, photosystem II oxygen-evolving enhancer protein 3; petH, ferredoxin-NADP+ reductase; acnB, aconitate hydratase 2; SDHD, succinate dehydrogenase membrane anchor subunit; MDH2, malate dehydrogenase; chlH, magnesium chelatase subunit H; rpsA, small subunit ribosomal protein S1; PSME3, proteasome activator subunit 3. Letters in parentheses next to the protein symbol indicate KEGG pathways: PS, Photosynthesis metabolism; TCA, TCA Cycle; N, Chl, Porphyrin and chlorophyll metabolism; RP, Ribosomal protein; P, Proteasome.

Dictyophyceae and Pelagophyceae displayed both lower overall numbers of KOfams and sequence coverage with detection of 21 and 13 enrichments, respectively. Few diel enrichments were detected within Dinophyceae.

These analyses uncovered a number of distinguishing features of specific groups (**Table 4**). Specific enrichment of "Fructose and mannose metabolism" pathways in Haptophyceae may reflect an enhanced sensitivity for detecting enrichments in this well-covered class of organisms as this pathway is directly linked to glycolysis. The specific enrichment in "Pyruvate metabolism" in Dictyophyceae may reflect a similar routing of fixed carbon toward lipids. Haptophyceae was also specifically enriched at dawn in multiple amino acid metabolic pathways ("Alanine aspartate and glutamate metabolism," "Phenylalanine, tyrosine and tryptophan biosynthesis," and "Cysteine and methionine metabolism") as well as "Biotin metabolism." Other pathways with significant temporal enrichment only in Haptophyceae include "Aldosterone synthesis and secretion"; "Cellular senescence pathway"; "Valine, leucine, and isoleucine biosynthesis"; and "Protein export" pathways.

The Dictyophyceae specifically displayed dawn enrichments in both "Sulfur metabolism" and "Glycine, serine and threonine metabolism," two pathways linked via the metabolite homoserine (**Supplementary Figures 8, 9**). Dictyophyceae also had a unique enrichment in the "Photosynthesis—antenna proteins" pathway. The morning-peaking gene families in the sulfur metabolism pathway form the majority of complete sulfate assimilation pathways in Dictyophyceae. Specific enrichment of "Antigen processing and presentation" and "Cardiac muscle contraction" pathways in the afternoon and at dusk, respectively, may reflect enrichments in general ATPase and transport functions. Enrichments identified only in Pelagophyceae were limited to "Olfactory transduction" and "Melanogenesis," both at 22:00. Surprisingly, the Dinophyceae were enriched in "Carbon fixation in photosynthetic organisms" and "Photosynthesis" pathways during the night, whereas all other classes were enriched in these pathways in the early morning. Dinophyceae also specifically displayed enrichment in the "Lysosome" pathway at dusk (18:00), suggesting a dominant signal from heterotrophic or mixotrophic dinoflagellates, with nighttime partitioning of gene transcription likely associated with digestion of engulfed prey (**Supplementary Figure 10**). We note that aside from the "Drug metabolism" pathway, all enrichments identified in Dinophyceae are organelle localized (plastid or lysosome).

TABLE 4 | KEGG pathway enrichment analysis at the class level.

Class	Peak (HST)	KEGG pathway
Haptophyceae		
	06:00	*Carbon fixation, photosynthesis*, fatty acid biosynthesis, glycolysis/gluconeogenesis, pentose phosphate pathway, p**henylalanine tyrosine and tryptophan biosynthesis**, **cysteine and methionine metabolism**, **alanine aspartate and glutamate metabolism, biotin metabolism, tropane piperidine and pyridine alkaloid biosynthesis**, **fructose and mannose metabolism**
	10:00	*Porphyrin and chlorophyll metabolism*
	14:00	*Protein processing in ER, proteasome*, **RNA transport**, spliceosome
	18:00	Spliceosome, **DNA replication,** *TCA cycle, oxidative phosphorylation***, cell cycle,** Synaptic vesicle cycle, **nucleotide excision repair**, *ribosome*, **meiosis**, fatty acid degradation, **mismatch repair,** thermogenesis, dopaminergic synapse
	22:00	**Calcium signaling pathway, cAMP signaling pathway**, thermogenesis, **Fc epsilon RI signaling pathway***, circadian entrainment*, **natural killer cell mediated cytotoxicity**, oxytocin signaling pathway, **renin secretion**, **aldosterone synthesis and secretion**, **phospholipase D signaling pathway**, **MAPK signaling pathway**
	02:00	**Valine leucine and isoleucine biosynthesis**, **protein export**
Dictyochophyceae		
	06:00	*Carbon fixatio*n, glycolysis/gluconeogenesis, *photosynthesis*, **glycine serine and threonine metabolism**, pentose phosphate pathway, **sulfur metabolism**, **pyruvate metabolism**, **bacterial secretion system**
	10:00	*Porphyrin and chlorophyll metabolism, proteasome*, *photosynthesis*, **photosynthesis – antenna proteins**
	14:00	*Protein processing in ER,* **antigen processing and presentation**
	18:00	*TCA cycle*, **cardiac muscle contraction**, thermogenesis, *oxidative phosphorylation*
	22:00	*Ribosome*, **salivary secretion**, **NOD-like receptor signaling pathway**
	02:00	*No enriched pathways at FDR < 0.05*
Pelagophyceae		
	06:00	*Carbon fixation*, *photosynthesis*, fatty acid biosynthesis,
	10:00	*Porphyrin and chlorophyll metabolism*
	14:00	*Protein processing in ER*
	18:00	*Ribosome*, *proteasome, TCA cycle*
	22:00	Dopaminergic synapse, **olfactory transduction**, oxytocin signaling pathway, **melanogenesis**, *circadian entrainment*
	02:00	*No enriched pathways at FDR < 0.05*
Dinophyceae		
	06:00	*No enriched pathways at FDR < 0.05*
	10:00	*No enriched pathways at FDR < 0.05*
	14:00	**Drug metabolism – other enzymes**
	18:00	*Carbon fixation*, **lysosome**
	22:00	Synaptic vesicle cycle, **collecting duct acid secretion**, *oxidative phosphorylation, photosynthesis*
	02:00	*Carbon fixation*

*Bold, pathways specific to each class. Italics denote representative pathways plotted in **Figure 4A**. Human Disease pathways not shown. Pathways for each time per taxa are listed by decreasing significance order (FDR < 0.05).*

DISCUSSION

Microbial eukaryotes perform vital functions in the NPSG ecosystem, including phototrophy, heterotrophy, mixotrophy, and parasitism. The generation of time-, function-, and taxonomy-resolved environmental transcriptome bins in this study produced insights about community composition and abundance, the connection between transcriptome composition and trophic state, the degree of diel regulation utilized by these environmental taxa, and the timing of functional processes throughout the diel cycle. Because of the high amount of functional, temporal, and taxonomic resolution provided by the annotated metatranscriptomes, we have utilized an analytical hierarchy in this study: we begin with broad survey of all environmental bins, then narrow our global functional analysis to the most complete environmental genera, and further focus our examination of metabolism to abundant representative genera and prominent pathways. This includes the smaller sized genera (<7 μm diameter) that make significant contributions to daily productivity in the mixed layer of the NPSG (Freitas et al., 2020). Although there may be subtle variations in environmental factors contributing to the transcriptional differences between genera, our assumption is that diel cycles are the critical driver of microbial life in the NPSG, and we have constrained the majority of our statistical analysis to diel periodicity of gene families and the timing of pathways.

The environmental genera in this study can be roughly categorized as having either high, intermediate, or low levels of diel transcriptional regulation. The haptophyte *Phaeocystis* has the singular distinction of having the highest proportion of diel gene families. Although *Phaeocystis* was not the only pure phototroph captured by our study, it was the most abundant obligate phototroph among the complete genera bins, as such it is difficult to conclude from this study alone whether similarly high levels of diel regulation are a common phototroph strategy. Furthermore, though we chose to constrain our analysis to the genus-level or higher, there may be strain-level differences in diel regulation magnitude throughout the cosmopolitan *Phaeocystis*.

Organisms of comparable sequencing depth and completeness show similar levels of diel regulation in the "intermediate" range (∼12–22% of gene families being diel-oscillating) despite differences in their trophic state and evolutionary lineage. This includes non-dinoflagellate mixotrophic genera (including the haptophytes *Prymnesium* and *Chrysochromulina*, the dictyophyte

Florenciella, and the non-constitutive mixotrophic ciliate *Strombidium*), the representative copepod genus *Lepeophtheirus*, and the choanoflagellate genus *Acanthoeca.* Diel transcriptome structuring in these genera may reflect physiological attunement to the diel cycle; in copepods, the high diel periodicity could be linked to diel vertical migration, but also reflects confounding issues from sub-group migration in and out of the mixed layer over diel cycles. The repeated observation of an "intermediate" level of diel regulation across disparate trophic modes and evolutionary lineages suggests that transcript synchronization to diel cycles is a common and advantageous strategy in the NPSG.

The "low" range of diel transcriptomes bin regulation (∼8% or less of gene families) is predominantly occupied by dinoflagellates, including mixotrophic and heterotrophic genera. The low degree of regulation in in dinoflagellates is consistent with studies showing a relatively dampened transcriptional response to environmental stimuli (Lin, 2011). Other strictly heterotrophic genera also had a low fraction of diel-regulated gene families, including the metazoan (animal) groups aside from *Lepeophtheirus* and heterotrophic protists other than *Acanthoeca*.

The coordination of protists to the diel period is also apparent in the distribution of diel transcript peak times. The highest proportions of transcript peak times occur at dusk, followed by dawn, underscoring the considerable metabolic re-arrangement of cells between light and dark periods. This has been observed previously in laboratory cultures: in the diatom *Thalassiosira pseudonana*, more gene transcript abundances peaked at dusk rather than dawn (Ashworth et al., 2013). The relative lack of gene family peak times at 10:00 HST, in particular, could be attributed to photo-protective purposes, minimizing transcription before the noon irradiation peak. This mid-morning minimum is also seen in prokaryotes: an earlier metatranscriptome study of the NPSG noted that the transcriptional minimum occurred closer to noon in contrast to MED4 culture transcriptomes (Ottesen et al., 2014). This "mid-day depression" could be a response to the detrimental effect of high UV radiation at mid-day in the surface layer and has been attributed to reduced growth, reduced DNA synthesis, and photochemical quenching in the picoplankton of the equatorial Pacific (Vaulot and Marie, 1999). The mid-day depression in transcription could also play a role in anti-viral defense, in that it restricts the replication of viral transcripts that are themselves tightly coordinated to diel cycles (Waldbauer et al., 2012).

Functional analysis of diel-regulated gene families allowed us to infer how protist lineages allocate transcriptional resources to functional processes over the diel cycle. Some metabolic features of phototrophs appear to be common strategies, such as dawn peaks in photosynthesis and energy storage pathways, and up-regulation of DNA synthesis and cell cycle elements at dusk. This strategy has been observed in diatoms and haptophytes previously. In the diatom *T. pseudonana*, genes encoding cell division, DNA replication and repair, carbon metabolism, and oxidative phosphorylation enzymes are highly expressed at dusk, while at dawn transcripts involved in Photosynthesis, Carbon fixation, and Ribosomes were higher (Ashworth et al., 2013); this general phototrophic strategy was also observed in metatranscripts from the calcifying haptophyte *E. huxleyi* (Hernández et al., 2020). Transcriptional evidence for the cycling of carbon is consistent with *in situ* field measurements of triacylglycerol in the NPSG that show increasing concentrations in cellular biomass through the day and decreases after nightfall (Becker et al., 2018). In particular, the striking similarity in transcriptional patterns between haptophytes and ochrophytes, both members of the Chromealveolate supergroup that separated an estimated 1 billion years ago (Yoon et al., 2004), implies selective pressure to conserve diel regulation of key cellular processes.

Taxa-specific metabolic features may be important in understanding the fate of carbon in the surface layer as a function of community composition. At the class-level, the haptophytes (Haptophyceae) constitute a "best case scenario" for diel pathway enrichments, as a combination of high diel regulation and population abundance contribute to sufficiently deep sequencing of haptophyte transcripts. The coordination of protein biosynthesis and turnover pathways to the diel cycle in haptophytes highlights the importance of these processes in diel cycles. The focused peaks of these pathway peaks in the afternoon suggests a large-scale proteomic remodeling prior to the evening metabolic switch, possibly involving the tagging and degradation of photosynthesis-related proteins in the dark hours when they are no longer useful. Intensive recycling of the proteome in the N-limited gyre could be an advantageous strategy to alleviate nitrogen stress, allowing nitrogen to be recycled between alternating proteomic regimes in a manner similar to the "hot-bunking" of iron atoms between anti-phase metalloenzymes in *Crocosphaera* (Saito et al., 2011). The functional pathway enrichments specific to dictyophytes, such as "Sulfur metabolism"; "Pyruvate metabolism"; and "Glycine, serine, and threonine metabolism" are evidence of additional lineage-specific temporal partitioning of metabolic processes. The "Sulfur metabolism" pathway includes sulfate assimilation, which is dominated by morning transcriptional peaks in this group. The connection of the "Sulfur metabolism" pathway to "Glycine, serine, and threonine Metabolism" through the metabolite homoserine highlights important connections between diel-enriched metabolic features. Many protists participate in diel cycling of sulfonated compounds (Durham et al., 2019), and these results hint at a possibly unexplored role of dictyophytes in sulfur cycling in the oceans. Future studies into the apparent metabolic differences between major protist lineages would benefit from metabolite- or protein-level data and fine-scale targeted investigation of specific metabolic pathways.

This study lends new perspective into dinoflagellate genetic regulation, which has significantly diverged from the transcriptional-level control utilized by most other eukaryotic lineages. The low proportion of diel gene families in dinoflagellate taxonomic bins is consistent with other studies showing a loss of transcriptional regulation in dinoflagellates (Kojima et al., 2011; Lin, 2011). The presence of 5′ spliced leaders, a transcript modification observed extensively in all major dinoflagellate orders, is a dinoflagellate adaptation that has been invoked as an alternative mechanism of gene regulation (Zhang et al., 2007). As expected from their low level of diel regulation,

the dinoflagellates did not return many significant pathway enrichments at any phylogenetic level of investigation. The few enriched dinoflagellate pathways are intriguing, as they contrast the generally minimum transcriptional regulation in dinoflagellates. At the class level, we observed organelle-targeted enrichments in three pathways ("Photosynthesis," "Carbon fixation," and "Lysosome"), and it would be interesting to investigate whether the gene families involved use 5′ spliced leaders in a similar manner as cytosolic transcripts. Along with the enrichments, we found the timing of the "Photosynthesis" and "Carbon fixation" pathways (evening) to be surprising, as all other photosynthetic groups maintained these pathways with predominantly morning (06:00) peaks. Proteomic studies have been helpful in describing the altered function of environmental dinoflagellates across spatial gradients (Cohen et al., 2021) and future proteomic studies of dinoflagellates with diel temporal resolution would be useful in determining the translational offset of dinoflagellate proteins from their transcript peaks over these cycles. Within the dinoflagellates, the notable outlier was Syndiniales (represented here by genus *Amoebophyra*), an order of obligate parasitoids that infect dinoflagellates and other marine organisms (Guillou et al., 2008). This lineage branched off from other dinoflagellates early in their evolution, before the transcriptional adaptations that characterize more-derived dinoflagellates. Conservation of diel periodicity in *Amoebophyra* may provide hints into the ancestral transcriptional regulation of dinoflagellates that appear to have lost the course of their evolutionary history in most other genera.

The proportion of taxonomy and function-calling for the assembled contigs in our dataset is comparable to the Tara Oceans Initiative eukaryotic gene catalog, which assigned about 25% of putative coding frames a functional annotation (Pfam) and about ∼50% of the contigs a taxonomy at any level (Carradec et al., 2018). Despite some differences in our reference databases and separate functional databases, we see nearly identical annotation results in our data, with 54 and 25.7% of contigs receiving confident taxonomical and functional (KEGG Orthology) annotations, respectively. Much remains to be discovered in the "microbial dark matter" of sequence data that has no match to extant taxonomic or functional databases. As novel organisms continue to be sequenced and functional annotation databases are expanded in the future, the annotation of the raw transcript data generated by this study can continue to be improved and exploited to gain further scientific insight. Regardless, the thousands of currently catalogd gene family profiles provide us rich detail into the presence and timing of known metabolic pathways.

This study reveals the diel transcriptional dynamics of eukaryotic protists in the surface layers of the NPSG, elucidating common patterns and striking differences in the transcriptional phenotypes of the most abundant protists in the surface community. Taking all of these results in aggregate, a picture emerges of a eukaryotic community tightly orchestrated to the daily rhythms of sunlight, with phototrophic organisms structuring their transcriptomes around the clock to harness and store solar energy during the day, and to replicate, divide, and possibly exchange signaling molecules at night. Mixotrophs are abundant and vital members of the protist community, and employ most of the core metabolic strategies as phototrophs. Understanding the most prevalent metabolic strategies employed by microbial eukaryotes in conjunction with the differences that distinguish them is critical to furthering our understanding of how carbon, nutrients, and energy flow through the surface ocean ecosystem.

DATA AVAILABILITY STATEMENT

KM1513 cruise information, plots, and associated environmental data for the HOE Legacy II cruise can be found online at http://hahana.soest.hawaii.edu/hoelegacy/hoelegacy.html. Raw metatranscriptome short-read sequence data is available in the NCBI Sequence Read Archive under BioProject ID PRJNA492142. Assembled contigs are deposited in Zenodo (https://doi.org/10.5281/zenodo.5009803). Code associated with this project is available on Github (https://github.com/armbrustlab/diel_eukaryotes).

AUTHOR CONTRIBUTIONS

RG, SC, BD, and EA conceived and designed the research project. RG and SC conducted the metatranscriptomic analyses. All authors contributed to data interpretation and the writing of the manuscript.

FUNDING

This work was supported by a grant from the Simons Foundation (SCOPE Award 329108 to EA) and XSEDE Grant Allocation (OCE160019 to RG).

ACKNOWLEDGMENTS

We thank the crew and science party of the R/V *Kilo Moana* on cruise KM1513.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.682651/full#supplementary-material

Supplementary Figure 1 | Sequence length distribution of ∼25 million quality-controlled Trinity-assembled contigs. Contigs less than 1,000 nucleotides in length are shown; approximately 482,000 contigs (1.9% of total) are longer than 1,000 nucleotides and range from 1,001 to 18,833 nucleotides.

Supplementary Figure 2 | Taxonomic assignments by primary Linnean ranks. Assignments totaling less than 2.5% of total placements for each rank are aggregated into "Other." Not all taxa have higher-level rank assignments in the NCBI taxonomy; phylum and class rank assignments do not exist for many lineages and were not manually assigned.

Supplementary Figure 3 | Frequency of unique KEGG ontologies (KOs) in eukaryotic MarineRefII reference taxa. *x*-axis: number of unique KOs found in each

taxa after mapping to KEGG's KOfam library of HMMer profiles. *y*-axis: frequency of occurrence. Dashed vertical line: minimum threshold of 900 KOs required for determination of "core KOs" present in >95% of reference taxa.

Supplementary Figure 4 | *E*-value distributions for DIAMOND taxonomy assignments for environmental genera bins. Counts include assignments directly at the genus level and lower nodes (e.g., species under a genus).

Supplementary Figure 5 | Non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis from row-normalized KOfam counts. NMDS ordination performed independently on gene families belonging to each of 48 environmental genera that met completeness criteria. Mean stress of 48 NMDS = 0.127 ± 0.028 stdev.

Supplementary Figure 6 | Gene family transcript peak times for 48 protist genera across the diel cycle. Gene families are significantly periodic with an FDR < 0.05. *x*-axis: clock hour in 24-h day. *y*-axis: percentage of significantly diel KOs with a peak at this time.

Supplementary Figure 7 | Normalized abundance heat map of significantly periodic gene families from 48 protist genera. Yellow and gray bars denote light (06:00, 10:00, and 14:00 HST) and dark (18:00, 22:00, and 02:00 HST) periods, respectively. Each row corresponds to a gene family, ordered by hierarchical clustering of abundance patterns. Color corresponds to row-normalized abundance values for each gene family. Colored dot by genus name corresponds to Lineage from **Figure 3A**. The order and presence of gene families is not maintained between facets.

Supplementary Figure 8 | Peak times for sulfur metabolism transcripts in dictyophytes. Most enzymes in this pathway peak at dawn, including the sulfate to sulfite component of assimilatory sulfate reduction. Metabolic maps were generated using KEGG Color Mapper (https://www.kegg.jp/kegg/mapper/color.html). Enzymes with significant periodicity are colored according to their peak time as determined through RAIN analysis of transcript abundance. Warm colors indicate peaks in dawn/daylight hours: yellow, 06:00 HST; orange, 10:00; red, 14:00. Cool colors indicate peaks in dusk/night hours: purple, 18:00 HST; blue, 22:00; green, 02:00.

Supplementary Figure 9 | Peak times for glycine, serine, and threonine metabolism transcripts in dictyophytes. Dictyophytes maintain peak times for most of the enzymes in this pathway throughout daylight hours. Metabolic maps were generated using KEGG Color Mapper (https://www.kegg.jp/kegg/mapper/color.html). Enzymes with significant periodicity are colored according to their peak time as determined through RAIN analysis of transcript abundance. Warm colors indicate peaks in dawn/daylight hours: yellow, 06:00 HST; orange, 10:00; red, 14:00. Cool colors indicate peaks in dusk/night hours: purple, 18:00 HST; blue, 22:00; green, 02:00.

Supplementary Figure 10 | Peak times for lysosome transcripts in dinoflagellates. Most periodic lysosomal enzymes have peaks in the afternoon (14:00), dusk (18:00), and early evening (22:00). Metabolic maps were generated using KEGG Color Mapper (https://www.kegg.jp/kegg/mapper/color.html). Enzymes with significant periodicity are colored according to their peak time as determined through RAIN analysis of transcript abundance. Warm colors indicate peaks in dawn/daylight hours: yellow, 06:00 HST; orange, 10:00; red, 14:00. Cool colors indicate peaks in dusk/night hours: purple, 18:00 HST; blue, 22:00; green, 02:00.

REFERENCES

Alexander, H., Rouco, M., Haley, S. T., Wilson, S. T., Karl, D. M., and Dyhrman, S. T. (2015). Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl. Acad. Sci. USA* 112, E5972–E5979.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252.

Ashworth, J., Coesel, S., Lee, A., Armbrust, E. V., Orellana, M. V., and Baliga, N. S. (2013). Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. *Proc. Natl. Acad. Sci. USA* 110, 7518–7523. doi: 10.1073/pnas.1300962110

Becker, K. W., Collins, J. R., Durham, B. P., Groussman, R. D., White, A. E., Fredricks, H. F., et al. (2018). Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean. *Nat. Commun.* 9:5179.

Becker, K. W., Harke, M. J., Mende, D. R., Muratore, D., Weitz, J. S., DeLong, E. F., et al. (2020). Combined pigment and metatranscriptomic analysis reveals highly synchronized diel patterns of phenotypic light response across domains in the open oligotrophic ocean. *ISME J.* 15, 520–533. doi: 10.1038/s41396-020-00793-x

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12, 59–60.

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018). A global ocean atlas of eukaryotic genes. *Nat. Commun.* 9:373.

Coesel, S. N., Durham, B. P., Groussman, R. D., Hu, S. K., Caron, D. A., Morales, R. L., et al. (2021). Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities. *Proc. Natl. Acad. Sci. USA* 118:e2011038118. doi: 10.1073/pnas.2011038118

Cohen, N. R., McIlvin, M. R., Moran, D. M., Held, N. A., Saunders, J. K., Hawco, N. J., et al. (2021). Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean. *Nat. Microbiol.* 6, 173–186. doi: 10.1038/s41564-020-00814-7

Durham, B. P., Boysen, A. K., Carlson, L. T., Groussman, R. D., Heal, K. R., Cain, K. R., et al. (2019). Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nat. Microbiol.* 4, 1706–1715. doi: 10.1038/s41564-019-0507-5

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Faure, E., Not, F., Benoiston, A. S., Labadie, K., Bittner, L., and Ayata, S. D. (2019). Mixotrophic protists display contrasted biogeographies in the global ocean. *ISME J.* 13, 1072–1083. doi: 10.1038/s41396-018-0340-5

Freitas, F. H., Dugenne, M., Ribalet, F., Hynes, A., Barone, B., Karl, D. M., et al. (2020). Diel variability of bulk optical properties associated with the growth and division of small phytoplankton in the North Pacific Subtropical Gyre. *Appl. Opt.* 59, 6702–6716. doi: 10.1364/ao.394123

Frias-Lopez, J., Thompson, A., Waldbauer, J., and Chisholm, S. W. (2009). Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environ. Microbiol.* 11, 512–525. doi: 10.1111/j.1462-2920.2008.01793.x

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R. M., Kirkham, A. R., Massana, R., et al. (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (*Alveolata*). *Environ. Microbiol.* 10, 3349–3365. doi: 10.1111/j.1462-2920.2008.01731.x

Hernández, M. L., Hennon, G. M. M., Harke, M. J., Frischkorn, K. R., Haley, S. T., and Dyhrman, S. T. (2020). Transcriptional patterns of *Emiliania huxleyi* in the North Pacific Subtropical Gyre reveal the daily rhythms of its metabolic potential. *Environ. Microbiol.* 22, 381–396. doi: 10.1111/1462-2920.14855

Hu, S. K., Connell, P. E., Mesrop, L. Y., and Caron, D. A. (2018). A hard day's night: diel shifts in microbial eukaryotic activity in the North Pacific Subtropical Gyre. *Front. Mar. Sci.* 5:351. doi: 10.3389/fmars.2018.00351

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in

the oceans through transcriptome sequencing. *PLoS Biol.* 12:e1001889. doi: 10.1371/journal.pbio.1001889

Kojima, S., Shingle, D. L., and Green, C. B. (2011). Post-transcriptional control of circadian rhythms. *J. Cell Sci.* 124, 311–320. doi: 10.1242/jcs.065771

Kolody, B. C., McCrow, J. P., Allen, L. Z., Aylward, F. O., Fontanez, K. M., Moustafa, A., et al. (2019). Diel transcriptional response of a California Current plankton microbiome to light, low iron, and enduring viral infection. *ISME J.* 13, 2817–2833. doi: 10.1038/s41396-019-0472-2

Lambert, B. S., Groussman, R. D., Schatz, M. J., Coesel, S. N., Durham, B. P., Alverson, A. J., et al. (2021). The dynamic trophic architecture of open-ocean protist communities revealed through machine-guided metatranscriptomics. *bioRxiv* doi: 10.1101/2021.01.15.426851 [Preprint].

Lewin, J., Norris, R. E., Jeffrey, S. W., and Pearson, B. E. (1977). An aberrant chrysophycean alga Pelagococcus subviridis gen. Nov. et sp. nov. from the North Pacific Ocean 1, 2. *J. Phycol.* 13, 259–266. doi: 10.1111/j.1529-8817.1977.tb02925.x

Li, Q., Edwards, K. F., Schvarcz, C. R., Selph, K. E., and Steward, G. F. (2021). Plasticity in the grazing ecophysiology of *Florenciella* (Dichtyochophyceae), a mixotrophic nanoflagellate that consumes *Prochlorococcus* and other bacteria. *Limnol. Oceanogr.* 66, 47–60. doi: 10.1002/lno.11585

Lin, S. (2011). Genomic understanding of dinoflagellates. *Res. Microbiol.* 162, 551–569. doi: 10.1016/j.resmic.2011.04.006

Mitra, A., Flynn, K. J., Tillmann, U., Raven, J. A., Caron, D., Stoecker, D. K., et al. (2016). Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: incorporation of diverse mixotrophic strategies. *Protist* 167, 106–120. doi: 10.1016/j.protis.2016.01.003

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2019). *Vegan: Community Ecology Package. R package version 2.5-5*.

Ottesen, E. A., Young, C. R., Gifford, S. M., Eppley, J. M., Marin, R., Schuster, S. C., et al. (2014). Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* 345, 207–212. doi: 10.1126/science.1252476

Quéguiner, B. (2016). *The Biogeochemical Cycle of Silicon in the Ocean*. Hoboken, NJ: John Wiley & Sons.

Ribalet, F., Swalwell, J., Clayton, S., Jiménez, V., Sudek, S., Lin, Y., et al. (2015). Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proc. Natl. Acad. Sci. USA* 112, 8008–8012.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/s0168-9525(00)02024-2

Rii, Y. M. (2016). *Ecology of Photosynthetic Picoeukaryotes in the Oligotrophic Ocean: Diversity, Activity, and Dynamics*. Doctoral dissertation. Honolulu, HI: University of Hawai'i at Manoa.

Rothhaupt, K. O. (1996). Laboratory experiments with a mixotrophic chrysophyte and obligately phagotrophic and photographic competitors. *Ecology* 77, 716–724. doi: 10.2307/2265496

Rousseau, V., Chrétiennot-Dinet, M. J., Jacobsen, A., Verity, P., and Whipple, S. (2007). The life cycle of *Phaeocystis*: state of knowledge and presumptive role in ecology. *Biogeochemistry* 83, 29–47. doi: 10.1007/s10533-007-9085-3

Saito, M. A., Bertrand, E. M., Dutkiewicz, S., Bulygin, V. V., Moran, D. M., Monteiro, F. M., et al. (2011). Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocosphaera watsonii*. *Proc. Natl. Acad. Sci. USA* 108, 2184–2189. doi: 10.1073/pnas.1006943108

Satinsky, B. M., Gifford, S. M., Crump, B. C., and Moran, M. A. (2013). Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods Enzymol.* 531, 237–250. doi: 10.1016/b978-0-12-407863-5.00012-5

Shikata, T., Matsunaga, S., Nishide, H., Sakamoto, S., Onistuka, G., and Yamaguchi, M. (2015). Diurnal vertical migration rhythms and their photoresponse in four phytoflagellates causing harmful algal blooms. *Limnol. Oceanogr.* 60, 1251–1264.

Sievert, C. (2018). *plotly for R. R package version 4.9.0.* Available online at: https://plotly-r.com

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., and Kelly, S. (2016). TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* 26, 1134–1144. doi: 10.1101/gr.196469.115

Steinegger, M., and Söding, J. (2018). Clustering huge protein sequence sets in linear time. *Nat. Commun.* 9:2542.

Stoecker, D. K., Johnson, M. D., de Vargas, C., and Not, F. (2009). Acquired phototrophy in aquatic protists. *Aquat. Microb. Ecol.* 57, 279–310. doi: 10.3354/ame01340

Tenenbaum, D. (2016). *KEGGREST: Client-side REST access to KEGG. R package version 1.12.3*.

Thaben, P. F., and Westermark, P. O. (2014). Detecting rhythms in time series with RAIN. *J. Biol. Rhythms* 29, 391–400. doi: 10.1177/0748730414553029

Vaulot, D., and Marie, D. (1999). Diel variability of photosynthetic picoplankton in the equatorial Pacific. *J. Geophys. Res. Oceans* 104, 3297–3310. doi: 10.1029/98jc01333

Villareal, T. A., Altabet, M. A., and Culver-Rymsza, K. (1993). Nitrogen transport by vertically migrating diatom mats in the North Pacific Ocean. *Nature* 363, 709–712. doi: 10.1038/363709a0

Waldbauer, J. R., Rodrigue, S., Coleman, M. L., and Chisholm, S. W. (2012). Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PloS One* 7:e43432. doi: 10.1371/journal.pone.0043432

Wilson, S. T., Aylward, F. O., Ribalet, F., Barone, B., Casey, J. R., Connell, P. E., et al. (2017). Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocosphaera*. *Nat. Microbiol.* 2, 1–9.

Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21, 809–818. doi: 10.1093/molbev/msh075

Zhang, H., Hou, Y., Miranda, L., Campbell, D. A., Sturm, N. R., Gaasterland, T., et al. (2007). Spliced leader RNA trans-splicing in dinoflagellates. *Proc. Natl. Acad. Sci. USA* 104, 4618–4623. doi: 10.1073/pnas.0700258104

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership