Phytoplankton diversity and ecology through the lens of high throughput sequencing technologies

Adriana Lopes dos Santos[1*], Catherine Gérikas Ribeiro[2], Denise Ong[1], Laurence Garczarek[3], Xiao Li Shi[4], Scott D. Nodder[5], Daniel Vaulot[3,1] and Andres Gutiérrez-Rodríguez[5].

[1] Asian School of the Environment, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

[2] GEMA Center for Genomics, Ecology & Environment, Faculty of Sciences, Universidad Mayor, Santiago, 8580745, Chile

[3] UMR 7144, ECOMAP, CNRS, Sorbonne Université, CNRS, Station Biologique de Roscoff, 29680 Roscoff, France

[4] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, 210008, China

[5] National Institute of Water and Atmospheric Research (NIWA), Hataitai, Wellington 6021, New Zealand

[*] Corresponding author:lopesas.ufrj@gmail.com

ORCID Numbers

- Adriana Lopes dos Santos: 0000-0002-0736-4937
- Catherine Gérikas Ribeiro: 0000-0003-0531-2313
- Denise Ong: 0000-0001-6053-6948
- Laurence Garczarek: 0000-0002-8191-8395
- Scott D. Nodder: 0000-0002-1963-8907
- Daniel Vaulot: 0000-0002-0717-5685
- Andres Gutiérrez-Rodríguez: 0000-0003-1274-3752

Abstract

Metabarcoding or high-throughput sequencing of a specific genetic marker is a powerful technique, widely used today, to analyze biodiversity across distinct environments and taxonomic groups. Plankton ecologists have benefited tremendously from the growing accumulation of metabarcoding studies. Novel biogeographic patterns have been established for cyanobacteria and microbial eukaryotes by the analysis of datasets from the Tara *Oceans* and Ocean Sampling Day projects. Novel lineages without cultured representatives have been uncovered. This chapter begins by going back to Carl Woese and George Fox who were the first to define the concept of "molecular marker". We then detail the different steps and choices that are involved in designing, performing and analysing a metabarcoding study. We are using a compilation of about 250 metabarcoding studies to present the major trends in terms of gene markers used and environments probed. We are then focusing on specific habitats and processes that have benefited from metabarcoding: the study of polar ecosystems, the functioning of the marine biological carbon pump, predator-prey interactions, and picoeukaryotic phytoplankton in highly urbanized lakes. An alternative approach to metabarcoding developed for marine picocyanobacteria is also briefly discussed. Finally, we offer some perspectives on emerging trends such as the use of metabarcodes combined with supervised machine learning for bio-monitoring, the link between metabarcoding and functional diversity in trait-based studies and the massive sequencing of long DNA fragments.

Keywords: high-throughput sequencing, gene marker, biological carbon pump, trophic-interactions, $_{mi}$tags, long-read amplicon sequencing.

1. Introduction

The study of taxonomic diversity and structure of biological communities is a cornerstone of ecology. Plankton communities are taxonomically and functionally diverse, with organisms covering a great range of cell sizes, morphologies and trophic modes (autotrophs, heterotrophs, mixotrophs), and belonging to archaea, bacteria, protists (microbial eukaryotes) and metazoans (Caron et al. 2012). In addition, plankton communities are extremely dynamic, with high growth and mortality rates. For example, phytoplankton growth rates are typically in excess of 1.0 day^{-1}, which is faster than those of archaea and bacteria (around 0.1 day^{-1}), although some of the most abundant bacterial species (from the SAR11 clade) can grow at rates similar to phytoplankton (Kirchman 2016).

The structural complexity and dynamic nature of plankton communities has historically precluded an accurate and comprehensive assessment of community structure and species diversity at temporal and spatial scales. For example, optical microscopy can be used to characterize the community composition down to species level, but requires a high level of taxonomic expertise, is time consuming, and often fails to identify cryptic species and to resolve smaller cells (<5 μm) that dominate phytoplankton communities in extensive areas of the ocean (Li et al. 1983). Pigment approaches when combined with standard algorithms such as CHEMTAX (Latasa 2007; Mackey et al. 1997) can partition quantitatively the phytoplankton biomass into chemo-taxonomic groups that correspond to broad taxonomic classes (e.g. Bacillariophyta), but do not resolve the organisms down to the species level. Similarly, flow cytometry (Marie et al. 1999) can provide accurate cell abundances of small phytoplankton ($\leqslant 20$ μm) but its taxonomic resolution is limited, failing to resolve the high diversity hidden within pico and nano-eukaryotic groups, only discriminating two important picocyanobacteria genera (*Synechococcus* and *Prochlorococcus*) and phycoerythrin-containing protists (cryptophytes).

During the last 30 years, the analysis of environmental diversity by molecular approaches has highlighted a massive unknown diversity, including entire lineages without any cultured representatives, for which only environmental sequences are available (Giovannoni et al. 1990). In particular we gained information on the genetic diversity of small-sized phytoplankton groups that lack distinctive morphological attributes (López-García et al. 2001; Moon-van der Staay et al. 2001; Vaulot et al. 2008). Morphological analyses combined with molecular approaches have demonstrated the existence of cryptic species even within well-known phytoplankton groups such as diatoms (Amato et al. 2007; Kaczmarska et al. 2014), powering the debate on phytoplankton species delineation (Leliaert et al. 2014). Currently, the advent of high-throughput sequencing (HTS) techniques combined with improved taxonomically-annotated reference sequence databases allows rapid and cost-effective assessment of the composition of plankton communities at a global level (de Vargas et al. 2015).

This chapter provides an overview of how metabarcoding has advanced our knowledge on phytoplankton diversity, community structure and ecosystem function, focusing on specific habitats or processes. In order to set up a common ground, we start by introducing the historical steps that led to the definition of molecular markers or barcodes, and by providing a condensed outline of steps and vocabulary within metabarcoding. The literature has witnessed a bloom of plankton metabarcoding studies in the last 10 years, in particular for its eukaryotic component. We have compiled data from about 250 studies (Table S1), and analyzed the major trends in terms of the marker used and the environment probed. We then focus on a few topics that have benefited from the advances in metabarcoding: the study of polar ecosystems, the functioning of the marine biological carbon pump, predator-prey interactions, and picoeukaryotic phytoplankton in lakes. We conclude by presenting alternative approaches to metabarcoding that have been developed for marine picocyanobacteria, and by reviewing what will the next steps be as new technologies allows the massive sequencing of longer DNA fragments.

2. The concept of molecular markers

"Science is impelled by technological advance and a guiding vision" - Carl Woese (2009)

Carl Woese and George Fox initiated a paradigm shift in microbiology when, in 1977, they established a 'third domain of life', they called Archaebacteria, after examining the sequences of the 16S/18S ribosomal ribonucleic acid (rRNA) of plants, yeast, human cells, bacteria and a group of microorganisms capable of growing by reducing carbon dioxide to methane, today known as methanogenic archaea (Woese and Fox 1977). The Sanger RNA sequencing method (Sanger et al. 1965) used by Woese and Fox consisted of applying T1 ribonuclease, which cuts RNA almost exclusively after guanosine residues, to digest radioactive rRNA of various known prokaryotic and eukaryotic cells. The products were then separated in two dimensions to reveal a chromatography pattern unique to the RNA molecule digested. A picture of the pattern obtained by exposing the chromatogram to an X-ray film provided the initial "fingerprint" of the organism. The more similar the fingerprints, the more closely related the organisms were. In addition, each spot corresponded to a specific sequence ending by G with no other Gs. The position of each spot indicated the length and number of uracils. The sequence of the spots could be resolved by treating the RNA from each spot with additional ribonucleases to produce smaller products that would again be separated on a gel and further re-digested. Ultimately, the complete nucleotide sequence information of a single spot would be assembled and the sequence of the RNA fragment deduced. This method would take 3 to 4 weeks per organism (Sapp and Fox 2013).

Comparative analysis of biomolecules to determine evolutionary relationships was already considered a powerful approach (Fitch and Margoliash 1967; Zuckerkandl and Pauling 1965) and had been applied to different groups of eukaryotes (vertebrates and invertebrates, Fitch 1976). However Woese and Fox were interested in determining "the relationships covering the entire spectrum of extant living systems" (Woese and Fox 1977) and therefore they needed what is today called a molecular marker. The molecule of choice was the rRNA, and the arguments they used to select rRNA still hold today (Olsen et al. 1986). Firstly, the rRNA is functionally and evolutionary homologous in all living organisms, and therefore a common denominator by which all organisms can be compared. Secondly, rRNA sequences have both highly conserved and highly variable regions, allowing the examination of both distant and close evolutionary relationships. Thirdly, in order to maintain its function, the structure of the sequences does not change rapidly, allowing homologous regions within the rRNA to be properly aligned and accurately compared. Lastly, the rRNA gene is presumably less susceptible to horizontal gene transfer as its product interacts with many molecules (also known as the complexity hypothesis, Jain et al. 1999). Hence, one would expect only evolutionary relationships to be reflected in rRNA sequences, although there are a few reported cases today of horizontal gene transfer for rRNA (HGT, Kitahara and Miyazaki 2013; Yabuki et al. 2014).

In 1977, only a few months after Woese and Fox published their work, Frederick Sanger (the author of Sanger RNA sequencing method) published a new method for determining nucleotide sequences of DNA molecules based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication (Sanger et al. 1977). The Sanger DNA sequencing method, as it is known today, became widely used in the next 40 years, allowing for more accurate and extensive comparisons of complete 16S/18S rRNA gene sequences. While the DNA Sanger sequencing process is fully automated today, it was initially done manually, and hence technically challenging and time consuming for long gene sequences. To reduce time and labour, David Lane and collaborators proposed that only homologous regions that were different enough to allow statistically significant comparisons should be sequenced (Lane et al. 1985). They noticed that conserved sequences were located next to less conserved regions. Conserved regions would serve as initiation sites for primer elongation sequencing techniques, while less conserved regions would allow phylogenetic evaluations. The suggested approach of primer regions flanking target gene markers is still an important concept today.

2.1 The advent of environmental sequencing

Although the description and classification of organisms with cultured representatives helped to determine many evolutionary relationships, it was suspected that much of the microbial diversity could not be grown in the lab. Stahl et al. (1985) were the first to obtain 5S rRNA sequences from an environmental sample to unveil the diversity of naturally occurring microbial populations. Even though the phylogenetic relationships among organisms from the environmental sample were roughly determined, the recovered dominant Archaea 5S RNA sequences were not closely related to any available sequence and therefore could not be accurately assigned. The authors remarked that "A larger sequence collection was needed to accurately place an organism of unknown phylogenetic affiliation". This study gave a first look at the possibility of unveiling uncultivated diversity from natural environments.

The use of the small subunit (SSU) rRNA gene was soon preferred to the 5S rRNA gene for phylogenetic analyses as longer sequences provided increased accuracy (Sogin and Gunderson 1987). Subsequent efforts of using Sanger sequencing were unsuccessful as the available methods for rRNA extraction from environmental samples did not yield enough RNA for sequencing. This was solved by using PCR amplification followed by sequencing, an approach largely used today. Medlin et al. (1988) landmark paper proposed the first set of primers and the polymerase chain reaction (PCR) conditions for the amplification of SSU eukaryotic rRNA (18S rRNA). This allowed to determine the nearly complete 18S rRNA sequence from the marine diatom *Skeletonema costatum* opening the door for using SSU rRNA gene as biodiversity marker for phytoplankton studies.

It took however thirteen more years before the first studies of the diversity of natural plankton communities based on 18S rRNA gene cloning and sequencing (Díez et al. 2001; López-García et al. 2001; Moon-van der Staay et al. 2001). These early studies revealed a widely unsuspected diversity among small planktonic eukaryotes including the discovery of new lineages within well known phytoplankton groups such as prasinophytes, haptophytes, dinoflagellates and stramenopiles. In subsequent years, a wide diversity of aquatic environments have been explored by the cloning approaches including deep sea ecosystems, anoxic environments or oligotrophic regions (Edgcomb et al. 2002; Lovejoy et al. 2006; Not et al. 2007b; Stoeck et al. 2003) leading to the discovery of other novel uncultivated groups (e.g. picobiliphytes Not et al. 2007a). Moreover besides the 18S rRNA gene, other genes have been used for example the 28S rRNA or the plastid 16S rRNA, *rbcL* or *psba* genes (e.g. Man-Aharonovich et al. 2010; Rodríguez-Martínez et al. 2013; Samanta and Bhadury 2016).

2.2 The transition to high throughput sequencing

About 15 years ago, Sanger sequencing was overtaken by the so-called Next Generation Sequencing (NGS) better referred as High Throughput Sequencing as by definition what belongs to the 'next generation' does not yet exists. HTS could provide millions of sequences of the target genetic marker without the need to construct clone libraries, which was one of the rate limiting factor. This led to a very rapid increase in the environmental studies of community structure.

The main HTS technologies used for microbial diversity studies have been mostly developed by two companies: 454 (initially an independent company, bought by Roche in 2007) and Illumina. HTS follows a sequencing by synthesis approach (Goodwin et al. 2016). As each nucleotide base complementary to the DNA fragment is added using a polymerase, a signal will identify what nucleotide was added to the elongating strand. This signal can be a fluorophore or a change in ionic concentration. Using this method, many different DNA fragments can be sequenced at one time, contributing to the massive parallelization and hence high throughput of the platform. 454 was first suggested for metabarcoding by Kysela et al. (2005): as 454 was not optimised to sequence many samples individually, these authors proposed tagging the sequences from different samples such that they could be sequenced together and separated later. Sogin et al. (2006) were the first to apply 454 technology to

sequence 16S rRNA gene amplicons to determine the bacteria and archaea diversity of marine samples. Soon after, the V9 region of the 18S rRNA gene was sequenced to the study diversity of marine protists (Amaral-Zettler et al. 2009; Stoeck et al. 2009). The 454 technology was initially more popular compared to Illumina because it had the capacity to generate longer reads, although with a higher error rate (Luo et al. 2012). However, Roche stopped production of 454 sequencing machines in late 2013, and discontinued its technical support in 2016 (Pedrós-Alió et al. 2018). Meanwhile, Illumina increased the length of the fragments it could sequence (up to 600 bp), and is now the preferred HTS technology (Jurburg et al. 2020).

There has been some confusion to the vocabulary applied to the HTS techniques. In addition to genome sequencing of cultured organismes, there are three ways HTS can be used: metagenomics, metatranscriptomics and metabarcoding (Pedrós-Alió et al. 2018). Metagenomics targets the sequence and reconstruction of full genomes from environmental microbial communities. Metatranscriptomics targets the transcribed genomics regions and expressed genes of natural communities by sequencing their mRNA pool. While metagenomics and metatranscriptomics were used early on to analyse the ecology of marine microbes (e.g. Venter et al. 2004), most recent plankton and phytoplankton studies use the third approach called metabarcoding which consists in amplifying a marker gene with PCR and sequencing the amplicon using HTS.

Metabarcoding, which is the focus of this chapter, can be found in the literature under an array of terms such as "amplicon sequencing", "tag sequencing", "HTS", as well as the misuse of the term "metagenomic" when dealing with the sequencing of PCR products. The methodological advantages of metabarcoding attracted an increasing number of researchers to apply this approach to microbial diversity studies (Santoferrara et al. 2020) in particular in aquatic ecosystems (Figure 1).

FIGURE 1 HERE

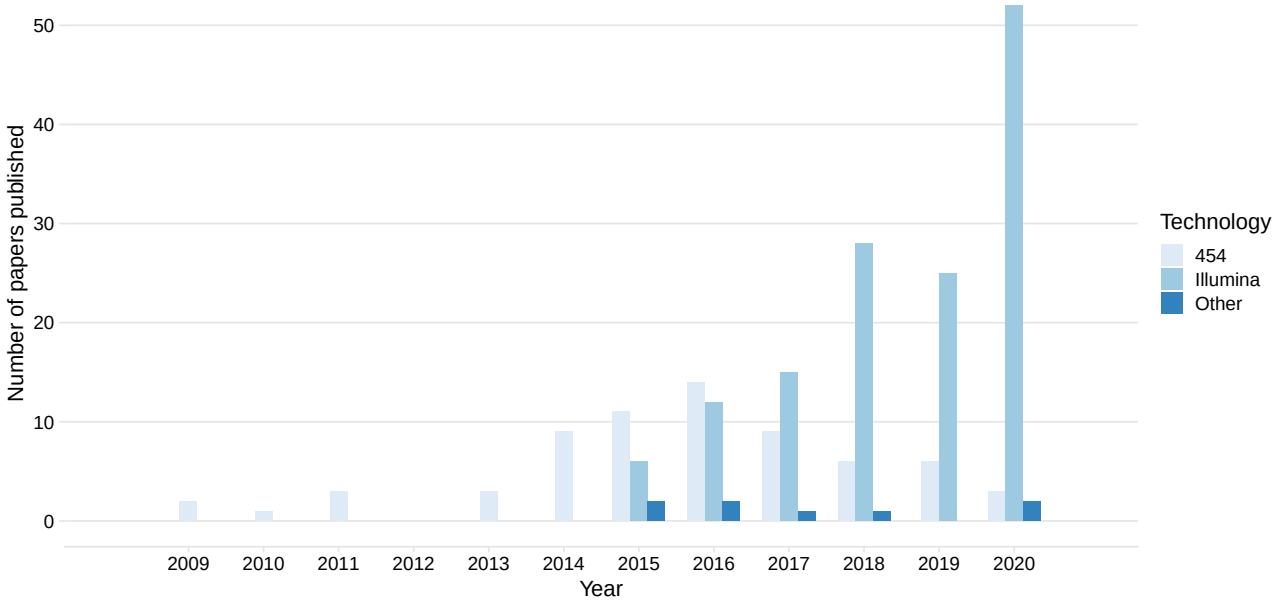

Figure 1: Papers using metabarcoding for aquatic protist studies published per year and technology (see Table S1).

3. The different steps of metabarcoding

We describe here the different steps of metabarcoding. A more comprehensive review on metabarcoding framework, including an overview of its limitation and potential bias for plankton community stud-

ies, can be found in Santoferrara (2019). A database of more than 240 studies from aquatic ecosystems to illustrate the major approaches applied for this type of environment was compiled for this chapter (Table S1 and https://github.com/vaulot/Chapter-2020-Lopes-Metabarcoding-phytoplankton).

3.1 Sampling and DNA extraction

Sampling of phytoplankton (or more generally of plankton) involves the collection of the cells that are free-floating in the water either by filtration, by using a plankton net or by separating individual cells by pipetting or flow cytometry sorting. Phytoplankton taxa can also be found on other substrates such as ice, sediment or even on the surface of other organisms such as coral or sea-grasses (see below Section 4) all of which require specialized sampling techniques. The most common approach consists in collecting water with bottles (e.g. Niskin bottles commonly used in oceanographic cruises) or with *in situ* pumps and then to pass it through a series of decreasing pore size filters (for example 20 μm, 2 μm, 0.2 μm) with the aim of separating the community into the plankton size classes proposed by Sieburth et al. (1978). The collected biomass is stored frozen or more rarely in ethanol or even formalin (Shiozaki et al. 2021).

Samples are then transported to the laboratory for acid nucleic (DNA and/or RNA) extraction. Several methods and commercial kits are available for this purpose which choice may vary depending on the substrate considered as well as the amount of material available. Most studies have been focusing on DNA. However, because phytoplankton can be found in a range of metabolic states (dormant, growing, decaying), a number of studies have sequenced the reverse transcribed RNA (cDNA) in addition to DNA. The ratio of RNA to DNA sequences (RNA:DNA) has been used as a proxy of metabolic activity (Charvet et al. 2014; Egge et al. 2015; Hu et al. 2016; Massana et al. 2015). Giner and collaborators suggested the use of cDNA templates rather than DNA as the most quantitative approach for picoeukaryotes assemblages (up to 3 μm in size) (Giner et al. 2020). Yet, evidence indicates that the general use of rRNA as an indicator of metabolic state of microbial communities has several limitations (e.g. rRNA copy numbers vary as a function of cell size) (Blazewicz et al. 2013).

3.2 Marker gene selection

The genetic markers used in a given study will vary depending on the target organisms and availability of reference sequences. The 18S rRNA gene is by far the most commonly used for eukaryotic phytoplankton (and plankton in general) (Table 1). In contrast, for prokaryotic phytoplankton (cyanobacteria), the equivalent 16S rRNA gene is not the most commonly used (see section 5). While sequencing of the entire SSU rRNA is possible using Sanger sequencing, the sequence is too long for HTS. Small hyper-variable regions of the 18S rRNA have been tested and selected over the years based on their ability to distinguish different taxa. Hyper-variable regions 4 (V4) and 9 (V9) are most often used to access the eukaryotic diversity in phytoplankton natural assemblages. The V4 region is located around 550 base pair (bp) from the beginning of the 18S rRNA gene (reference sequence *Saccharomyces cerevisiae*) and is $\simeq$ 450 bp long, while V9 is located at the end of the gene with a $\simeq$ 150 bp length (Vaulot et al. 2021). While both regions are able to provide similar pictures of eukaryotic phytoplankton diversity at higher taxonomic levels (division, class, see Tragin et al. 2018), at lower taxonomic levels (e.g. species) specific groups may be better resolved using a specific region. For example, V9 distinguish better between phytoplankton species of the class Chloropicophyceae (Lopes dos Santos et al. 2017), while it has poor resolution among the order Dolichomastigales (Class Mamiellophyceae, Tragin and Vaulot 2018). One drawback of the V9 region is the limited number of reference sequences in public databases. In aquatic metabarcoding studies, the V4 region is clearly the most often used, probably because of the better coverage of reference databases (Table 1).

TABLE 1 HERE

Table 1: Most frequently used gene markers and 18S rRNA region in eukaryotic metabarcoding studies of aquatic phytoplankton with the number of papers (N) where used (see Table S1 for the list of studies).

Marker gene	N	18S region	N
18S rRNA	221	V4	160
16S rRNA plastid	7	V9	34
28S rRNA	6	Other	27
ITS2	3		
ITS1	2		
rbcL plastid	2		
rRNA operon	2		
23S rRNA plastid	1		
coi mitochondrion	1		

Other marker genes have been used (Tables 1 and S1) to access eukaryotic phytoplankton diversity. Few metabarcoding studies have directly targeted plastid 16S rRNA (Trefault et al. 2021). However 16S plastid sequences have been analyzed as a by-product of bacteria 16S metabarcoding studies (e.g. Needham and Fuhrman 2016). One issue is the limited number of reference sequences available for plastidial 16S despite efforts to increase the number of cultures sequenced (Decelle et al. 2015). The D1-D2 region of the large subunit 28S rRNA gene has been used to determine diversity for Haptophyta (Bittner et al. 2013; Gran-Stadniczeñko et al. 2017). Although there are more reference sequences available for 18S rRNA (Edvardsen et al. 2016), the 28S rRNA has more variable regions, especially between closely related species of Haptophyta (Bittner et al. 2013; Liu et al. 2009). Gran-Stadniczeñko et al. (2017) found more diversity using the D1-D2 region of 28S rRNA compared to the V4 region of 18S rRNA. Other marker gene which have been used with the clone library approach prior to the development of HTS, such as the ITS (internally transcribed spacer of the rRNA operon), *psbA* or *rbcL* genes (Man-Aharonovich et al. 2010; Rodríguez-Martínez et al. 2013; Samanta and Bhadury 2016), have been very little used for phytoplankton metabarcoding either due to the absence of suitable short region, lack of general primers, or absence of reference sequences.

Once a marker gene has been selected, it is necessary to select suitable primers. For 18S rRNA, a wide range of primers and primer sets have been developed (for a review see Vaulot et al. 2021). Several criteria have to be met. First, the primers must have as little bias as possible, i.e. amplify all taxonomic groups with the same efficiency. Second, they must produce a fragment with a compatible size with the sequencing technology of choice. For example, a fragment should not surpass 550 bp when Illumina 2x300 kits are used, allowing for some overlap between the forward and reverse fragments sequenced. In studies of marine protists including phytoplankton, by far the most used primer pair is composed of TAReuk454FWD1 and TAReukREV3, as forward and reverse primers respectively (Tables 2 and 3) which were among the first primers designed for metabarcoding (Stoeck et al. 2010). These primers target the V4 region yielding a $\simeq$ 420 bp amplicon suitable for Illumina sequencing. One problem is that the reverse primer has one mismatch to Haptophyta, a very important phytoplankton group, at the 3' end of the primer which is particularly unfavorable, resulting in datasets where Haptophyta are absent. A small modification of this primer (V4 18S Next.Rev, Table 2) removes this bias (Piredda et al. 2017). For the V9 region two primer sets, 1380F/1510R and 1389F/1510R, are the ones most often used. Some primer sets are designed to target specific groups, enhancing for example the coverage of haptophytes or diatoms (Gran-Stadniczeñko et al. 2017; Rynearson et al. 2020).

TABLE 2 HERE

TABLE 3 HERE

Table 2: Eukaryotic 18S rRNA primers most often used in metabarcoding studies for aquatic phytoplankton and protists (see Table S1) with the number of papers (N) where used.

Name	Sequence	Region	Direction	Reference	DOI	N
TAReuk454FWD1	CCAGCASCYGCGGTAATTCC	V4	fwd	Stoeck et al (2010)	10.1111/j.1365-294X.2009.04480.x	83
1380F	CCCTGCCHTTTGTACACAC	V9	fwd	Amaral Zettler et al (2009)	10.1371/journal.pone.0006372	19
528F	GCGGTAATTCCAGCTCCAA	V4	fwd	Cheung et al. (2010)	10.1038/ismej.2010.26	18
E572F	CYGCGGTAATTCCAGCTC	V4	fwd	Comeau et al. (2011)	10.1371/journal.pone.0027492	16
1391F	GTACACACCGCCCGTC	V9	fwd	Lane (1991)		7
3NDF	GGCAAGTCTGGTGCCAG	V4	fwd	Cavalier-Smith et al. (2009)	10.1016/j.protis.2009.03.003	6
TAReukREV3	ACTTTCGTTCTTGATYRA	V4	rev	Stoeck et al (2010)	10.1111/j.1365-294X.2009.04480.x	61
1510R	CCTTCYGCAGGTTCACCTAC	V9	rev	Lopez-Garcia et al. (2003)	10.1073/pnas.0235779100	24
E1009R	AYGGTATCTRATCRTCTTYG	V4	rev	Comeau et al. (2011)	10.1371/journal.pone.0027492	16
V4 18S Next.Rev	ACTTTCGTTCTTGATYRATGA	V4	rev	Piredda et al. (2017)	10.1093/femsec/fiw200	13
EukB	TGATCCTTCTGCAGGTTCACCTAC	V9	rev	Medlin et al. (1988)	10.1016/0378-1119(88)90066-2	8
964iR	ACTTTCGTTCTTGATYRR	V4	rev	Balzano & Leterme (2015)	10.3354/ame01740	7

Table 3: 18S rRNA primer sets most often used in metabarcoding studies of aquatic phytoplankton and protists (see Table S1) with the number of papers (N) where used. Refer to Table 2 for sequence and reference of primers.

Primer fwd	Primer rev	Region	N
TAReuk454FWD1	TAReukREV3	V4	60
1380F	1510R	V9	19
E572F	E1009R	V4	15
TAReuk454FWD1	V4 18S Next.Rev	V4	13
1391F	EukB	V9	7
528F	1055R	V4	6
Other			101

3.3 PCR and sequencing

Once DNA is extracted and the marker gene and the relevant primers selected, the next step involves one or two round of PCR depending on the material available. Low amount of DNA for example from samples sorted by flow cytometry require usually 2 PCR rounds (Gérikas Ribeiro et al. 2018). As for any PCR, conditions (annealing temperature, primer concentration etc...) need to be optimized. Another round of PCR is then needed to add for each sample "barcodes" that will allow to run a large number of samples together. Typically for an Illumina MiSeq run, up to 300 of samples can be multiplexed together each sample producing 50 to 100 thousand sequences (reads). Higher throughput sequencers such as HiSeq can yield 20 to 50 times more sequences per sample.

3.4 Data processing

In what follows we will focus mostly on 18S rRNA metabarcodes, but the process would be similar for any other marker gene. HTS produces millions of raw sequences (reads) that need to be processed to produce meaningful data. In brief, the different steps are: demultiplexing, quality filtering, removal of primers, assembly, clustering, removal of chimera and taxonomic assignment. A number of bioinformatics programs have been developed to take care of these different steps (for a recent review Pérez-Cobas et al. 2020). Some can perform almost all necessary steps such as Qiime (Bolyen et al. 2019), mothur (Schloss et al. 2009), VSEARCH (Rognes et al. 2016) or dada2 (Callahan et al. 2016), while other perform only specific steps, such as cutadapt (Martin 2011) which focuses on primer removal. Qiime, a suite of programs written in Python is the most used to date, despite the fact that it was in its initial versions complicated to install, while dada2 running under R is becoming increasingly popular (Table 4). Another new option is PEMA, that regroups a number of tools such as VSEARCH and Swarm in a ready to use Docker container (Zafeiropoulos et al. 2020).

TABLE 4 HERE

Table 4: Software most often used to process sequence in metabarcoding studies of aquatic phytoplankton and protists (see Table S1) with the number of papers (N) where used.

Software	N
USEARCH	77
QIIME	68
mothur	43
swarm	25
dada2	19
Other	14

It is also possible to use web services to process datasets. VAMPS (https://vamps2.mbl.edu/ is one of the oldest such site and offers besides data processing the possibility to download and visualize public datasets (Huse et al. 2014). It is superseded by services like SILVAngs (https://ngs.arb-silva.de/ that is linked to the SILVA sequence database (see below) or FROGS (http://frogs.toulouse.inra.fr/). Drawbacks of using such services are linked to have to upload your data on a web server which can be slow and to be limited to fixed processing options (e.g. clustering or reference database).

Read processing. As mentioned earlier, many samples are usually sequenced together through the use of barcode indexing specific to each sample. Usually the demultiplexing step is performed by the sequencing software but programs such as cutadapt can also perform this step. It is always necessary to remove the primers which can be done with cutadapt, and eventually combined with demultiplexing. The next step is to remove reads that are either too short or of bad quality. Quality can be checked using for example the function *plotQualityProfile* in the R dada2 package, allowing to adjust parameters of the *FilterandTrim* function. When using Illumina technology, the next step is to build contigs or paired-end assemblies (Kozich et al. 2013) from the R1 (forward) and R2 (reverse) reads. Each of these steps may remove a significant number of reads, but it is always better to be more rigorous at this stage to prevent artefacts at latter processing stages.

Clustering. It is then necessary to regroup together sequences that are highly similar. Initially the strategy was to define so-called Operational Taxonomic Units (OTUs), where ideally each OTU would represent one microbial species (Pedrós-Alió et al. 2018). The conventional practice suggests that microbial organisms with a 97% similarity threshold (conversely a dissimilarity of 3%) are considered part of the same species (Schloss and Handelsman 2005; Stackebrandt and Goebel 1994) and this threshold has been widely used in metabarcoding studies (Table S1). However, this threshold is somewhat arbitrary, varying between the taxonomic groups that are being studied and the length and region of the marker gene used (Edgar 2018). Depending on the aim of the study, 97% can be too low and should be increased to 99% or 100% for example when looking at species distributions (Tragin and Vaulot 2019). The use of OTUs with a fixed similarity level has several drawbacks (Callahan et al. 2017). First, this clustering depends on the number of samples included in the analysis, and if the number of samples is changed, the analysis must be redone. Second, the sequences need to be aligned to a reference alignment, which is a very time consuming step as well as dependent on the quality of the reference alignment. Last, it is difficult to compare the results from different studies if they choose a different threshold. Therefore in recent years, more "natural" clustering algorithms have been developed, which includes swarm (Mahé et al. 2014) and dada2 (Callahan et al. 2016). These approaches are more computationally efficient and yield data that can be easily compared between data sets. The clusters produced by these methods are called either "swarms" or "ASVs" (Amplicon Sequence Variants). The processing software assigns sample reads to each cluster, forming what is known as an ASV or OTU table, with one line per OTU/ASV and one column per sample, the cells representing the number of reads for a given combination of OTU/ASV and sample.

One important source of error is linked to chimera formation during PCR amplification. Chimeric sequences are two or more fragments from distinct species that are combined during PCR to form a sequence which is further amplified and would be inaccurately detected as a novel species (Meyerhans et al. 1990). If undetected, the resulting diversity estimates will be falsely skewed (Porazinska et al. 2012; Wang and Wang 1997). A variety of tools have been developed to detect and remove these sequences such as UCHIME (Edgar et al. 2011) or the *removeBimeraDenovo* function in dada2. However, this process is often not satisfactory and chimera removal may have to be done by visual inspection of an alignment of OTU/ASV and reference sequences.

Taxonomic assignation. The taxonomic assignation of the OTU/ASV sequences is at the heart of the metabarcoding analysis. A poorly assigned dataset is as useful as a small umbrella during a tropical storm. OTU/ASV representative sequences are assigned according to their similarity to those from a reference database for which the taxonomy has been curated by experts. The database should cover the whole targeted taxonomic domain, e.g. eukaryotes when using the 18S rRNA gene. If some important groups are missing, the corresponding OTU/ASV will not be correctly annotated. The sequences from the reference database should also be correctly annotated: for example, if a reference sequence is annotated as a diatom while it is in reality a green alga, this will result in errors in the final community structure. For eukaryotes three general databases can been used. Genbank nr contains all published sequences and is therefore most comprehensive in terms of coverage. Its main handicap is the taxonomic annotation which is not curated and can be very minimal in particular for environmental sequences (e.g. many sequences are simply annotated as "eukaryotes"). SILVA (Quast et al. 2013) is better annotated than Genbank, and is periodically updated (https://www.arb-silva.de/). For protists however, the best choice is the Protist Ribosomal Reference database (PR^2, https://pr2-database.org) which contains fewer sequences ($\sim$ 180,000) but these sequences are periodically re-annotated by experts from each group using 8 taxonomic levels, from the kingdom to the species (Guillou et al. 2013). Some specialized databases have also been developed for specific taxonomic groups such as dinoflagellates (DINOREF, Mordret et al. 2018), or specific genes, such as the plastidial 16S rRNA (PHYTOREF, Decelle et al. 2015).

Once a reference database is selected, the assignment can be done several different ways. Probably the best strategy is to use alignment-independent approaches, such as the naive Bayesian classifier, also known as the RDP classifier (Wang et al. 2007). RDP is implemented in pipelines such as mothur or dada2, and provides bootstrap values for the confidence of the assignment at each taxonomic level. Similar approaches are used by the SINTAX algorithm (Edgar 2016), as implemented in VSEARCH or the *IdTaxa* function from the R DECIPHER package (Wright 2016). Other methods rely on alignments such as a LCAClassifier of Crest (Lanzén et al. 2012), that uses the aligned SILVA database as a reference, or pplacer (Matsen et al. 2010), that places sequences on a fixed reference tree. A simple BLAST (Altschul et al. 1990) search can also be used to obtain the closest matching sequences from either Genbank nr or a reference database, but it is difficult to implement for a large number of OTU/ASV and does not provide bootstrap values. Still it is useful to verify the results from other methods. If one is interested to investigate low taxonomic levels (e.g. genus or species), it is often wise to confirm OTU/ASV assignments with alignments and phylogenetic trees using closely related sequences obtained by BLAST.

Abundance normalisation. Metabarcoding data are used in general to calculate changes in relative taxa abundance between different samples. A first issue is that the total number of reads vary between samples. There are roughly two ways of normalizing samples (Pérez-Cobas et al. 2020). One approach is to rarefy all samples to the least abundant sample, which may lead to the loss of rare OTU/ASV using for example the *rarefy* function from the R vegan package (Oksanen et al. 2016). Alternatively, one can compute the relative abundance of a given OTU/ASV in each sample or normalize each sample to a fixed number of reads, e.g. the mean or median of all samples. More sophisticated approaches have been proposed (Love et al. 2014). Another concern is that the rRNA gene copy number varies between species. While picoplankton typically have a few copies of rRNA operon, larger eukaryotes can have

several thousands copies, especially among groups such as dinoflagellates (Gong and Marchetti 2019; Zhu et al. 2005). As such, eukaryotes with large number of rRNA gene copies will be over-represented in metabarcoding studies. However, as the number of species for which the copy number is known remains very small, it is nearly impossible to correct for this bias.

Downstream analyses. Once an OTU/ASV table has been constructed with the taxonomy of each sequence assigned and abundance normalized, many different analyses can be performed. In this respect R offers a range of tools that are very useful. One very useful package is phyloseq (McMurdie and Holmes 2013). It combines a large number of tools in a single package, providing ways to filter (for a sample type) and group the data (e.g. at higher taxonomic level), to display the data under a variety of graphic formats (barplots, heatmaps), to perform alpha and beta diversity analyses such as NMDS (non-metric multidimensional scaling), or to visualize networks. Another useful tool is the DESeq2 package (Love et al. 2014), which allows to detect OTU/ASV that are specific of a given set of samples.

Making data available: a critical but overlooked step. Data availability is key for scientific reproducibility. Some web sites offer to download processed data from large scale projects (e.g Tara or Malaspina, see below) such as the Ocean Barcode Atlas (http://oba.mio.osupytheas.fr/ocean-atlas/, Vernette et al. 2021). However, researchers should be able to start from the original data and reproduce all steps that allowed to reach the final conclusions, following the FAIR approach (Wilkinson et al. 2016). Moreover the primary data from any paper should be available to the community so they can be reused in new studies. For example, the ability to combine sequences from various metabarcoding datasets originating from different environments can give new insights into the distribution of specific taxa and reveal novel groups (e.g. Kuwata et al. 2018). Therefore it is imperative that raw sequences be deposited into public repositories, preferably the NCBI/EBI/DDBJ SRA (Sequence Read Archive), with standardised environmental metadata and technical information (Yilmaz et al. 2011). The accession number provided should be stated in the publication. Today very few journals request explicitly that authors provide primary sequence files and even when it is stated in the instructions to authors, this is only partially enforced. It is very revealing that only 80% of the HTS studies listed here (Table S1) have deposited raw sequences into public repositories, a situation that has been recently pointed out for 16S metabarcoding data (Jurburg et al. 2020). Scripts and pipelines should be provided and clearly documented. Derived data such as OTU/ASV tables and sequences with their assignation should also be freely available in open repositories (e.g. GitHub, Zenodo or Figshare). Statements such as "only made available upon request" should result in immediate refusal for serious journal to process submissions (Eren 2019).

4. Protist metabarcoding studies in aquatic environments

Up the time we wrote this chapter (January of 2021), more than 240 studies have used metabarcoding to look at protist communities in aquatic systems (Table S1) and the number appears to increase exponentially (Figure 1). Few of these studies have targeted specifically phytoplankton (e.g. using plastid 16S rRNA primers). Even when focusing on the sole analysis of phytoplankton groups, the majority of them obtained data on the overall protist community since the most commonly used primer sets (Table 3) do not discriminate against heterotrophic groups. Great part of studies investigated marine systems, in particular oceanic and coastal (Table 5). Freshwater systems have been less intensively sampled with the exception of lakes (see section 4.4). Most aquatic studies have sampled the water and much fewer sediments (Table 5). Some key aquatic biotopes such as coral reefs have been very little sampled as have been epiphytic communities on algae or sea grasses. In the future, it is likely that more effort will be devoted to ecosystem/substrate that have received less attention.

TABLE 5 HERE

Table 5: Ecosystems and substrates targeted by aquatic metabarcoding studies (see Table S1).

Ecosystem	N	Substrate	N
marine oceanic	106	water	206
marine coastal	91	sediments	29
freshwater lakes	23	ice	6
estuarine	11	biofilm	5
freshwater rivers	11	sorted phytoplankton	5
lagoons	6	coral	4
freshwater ponds	4	sediment trap	4
aquaculture	3	copepod gut content	2
freshwater gutters	1	periphyton	2
freshwater supply	1	ballast water	1
saltern	1	ice-algal aggregates	1
		macroalgae epiphytes	1
		rock	1
		seagrass	1
		sponge	1
		suspended and sinking particles	1
		water (experiment)	1
		zooplankton	1

In addition to these studies that targeted one or few sampling locations, three global surveys have been performed that contributed to popularize metabarcoding as a tool to study protist and phytoplankton communities. The Ocean Sampling Day network (Kopf et al. 2015) aimed at simultaneous sampling (on the day of the summer solstice in the northern hemisphere) coastal stations (Figure 2). Samples were taken in surface and not fractionated. Metabarcoding data are available for two years (2014 and 2015) but sampling has been conducted since 2018 (https://www.assembleplus.eu/research/ocean-sampling-day) as part of the ASSEMBLE Plus. Although initially heavily biased towards Europe, the network of stations has gradually expanded. Surprisingly quite few papers have been made use of this resource (e.g. Tragin and Vaulot 2018). The Tara *Oceans* launched in 2009 was much more ambitious as it covered all three major oceans (Atlantic, Pacific and Indian, Figure 2) as well the Arctic Ocean. Samples were taken at 2 depths (in general surface and deep-chlorophyll maximum) and separated into four size fractions (pico, nano, micro, meso). A large number of ancillary measurements (temperature, salinity, nutrients) were acquired as well as samples for microscopy. With respect to metabarcoding, one remarkable characteristics of the Tara expedition is the depth of sequencing achieved (typically one million of reads per sample) allowing to cover the "rare" biosphere, i.e. organisms in very low abundance (Logares et al. 2015). Moreover samples were also analyzed by metagenomics and metatranscriptomics approaches complementing the metabarcoding data. A final key element was the quick release of the primary and secondary data (raw and processed sequences) as well the associated environmental data. This led besides the initial papers (e.g. de Vargas et al. 2015) to a very large number of papers making use of the data, for example to look at the distribution of specific plankton groups (e.g. giant protists, Biard et al. 2016; Malviya et al. 2016) or at global diversity trends (dos Santos et al. 2017; Ibarbalz et al. 2019). For more details on Tara *Oceans* expeditions please refer to Chapter 3.2. The Malaspina expedition (2010-2011, Duarte 2015) performed round-the-world sampling (Figure 2) not only of surface but also of deep waters (Giner et al. 2020; Logares et al. 2020). Combining the data obtained from these expeditions with other more punctual studies allow to infer the ocean-wide distribution of newly described phytoplankton species (*Mantoniella beaufortii*, Yau et al. 2020) as well as of larger taxonomic groups (diatoms, Malviya et al. 2016).

In this section, we will focus on a few topics that have benefited from the advances in metabarcoding and for which a rich literature starts to blossom: the study of polar ecosystems, the functioning of the marine biological carbon pump and picoeukaryotic phytoplankton in lakes.

FIGURE 2 HERE

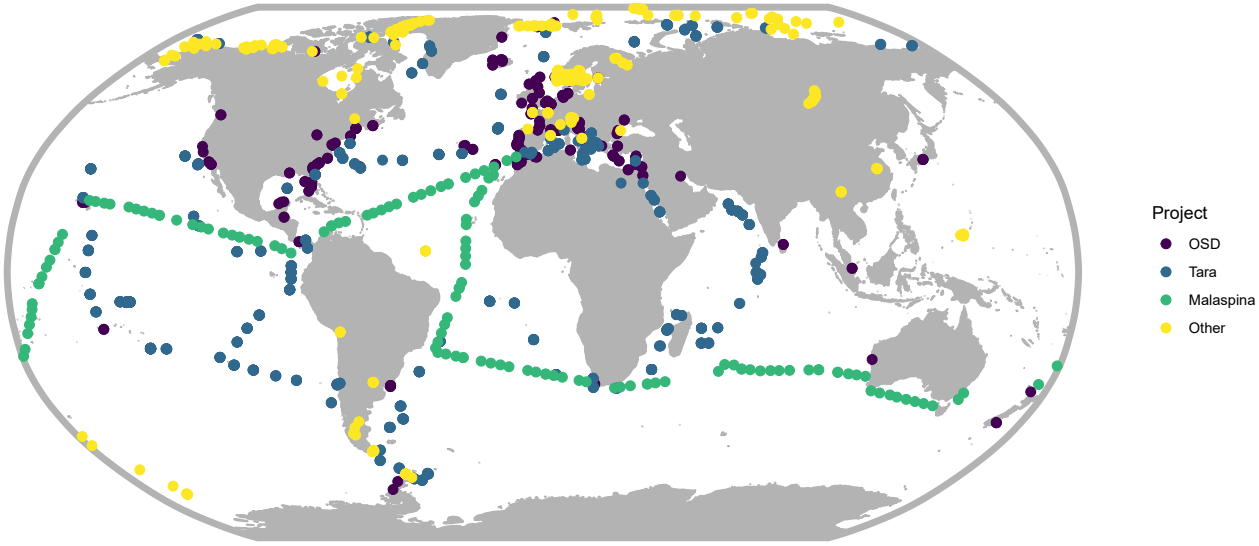

Figure 2: Map of stations investigated in a representative subset of 45 aquatic metabarcoding studies.

4.1 Arctic and Antarctic communities

The southern and northern poles are among the most environmentally susceptible regions to climate change (Haumann et al. 2020; Perovich and Richter-Menge 2009; Turner et al. 2005) and ocean acidification (Orr et al. 2005; Shadwick et al. 2013), and the least studied ecosystems on the planet, due to logistical constraints. Similarly to what happened in other marine and fresh water environments, the study of phytoplankton ecology on high latitudes gained resolution and speed with metabarcoding-based studies. Several studies have collectively generated a solid knowledge base on polar protist diversity, unveiling phytoplankton community structure of the water (Duret et al. 2019; Kalenitchenko et al. 2019), ice (Comeau et al. 2013; Hardge et al. 2017; Lutz et al. 2018) and snow (Davey et al. 2019; Soto et al. 2020). Metabarcoding revealed the community progression driven by drastic seasonal changes on solar energy input (Joli et al. 2017; Vick-Majors et al. 2014), and the interconnectivity between polar and sub-polar regions (Sow et al. 2020), and between the Arctic and Antarctic domains (Segawa et al. 2018; Wolf et al. 2015). Metabarcoding also allows to examine questions that microscopy-based methods could hardly address: for example, the seasonal progression during polar night of pico-sized algae for which morphological features are hard to distinguish (Joli et al. 2017), or the composition of the rare microbial biosphere and its ecological roles and biogeography (Galand et al. 2009).

4.1.1 The risk of habitat loss for ice-associated communities

Polar environments are subject to complex hydrographic, atmospheric and seasonal interactions, with prolonged periods of constant light or complete darkness, where ice dynamics plays a major community-structuring role (Janout et al. 2016; van Leeuwe et al. 2018), from light attenuation effect to changes in water stability patterns during melting (Rozema et al. 2017). Besides providing a complex matrix where different sympagic (ice-associated) algae communities thrive (Comeau et al. 2013), sea-ice has a recruiting/seeding role (Hardge et al. 2017; Niemi et al. 2011; Olsen et al. 2017) for the phytoplankton spring bloom in the water column. The question of how phytoplankton production, diversity and community structure will change due to climate-related loss of sea-ice has been increasingly addressed, especially in the Arctic (Horvat et al. 2017; Leu et al. 2011), where this phenomenon has been steeply accentuated in the last decades (Serreze et al. 2007; Stroeve et al. 2014).

Hardge et al. (2017) compared Arctic protist diversity on sea-ice, melt pond water, under-ice water and deep-chlorophyll maximum using metabarcoding, and demonstrated that the sea-ice, and specially multi-year ice, harbors a particularly high diversity of unique taxa, highlighting the specificity of the sympagic community dominated by diatoms, and a potential diversity loss in both sea-ice and water column as multi-year sea-ice decreases drastically in the region.

Due to the on-going reduction of sea-ice thickness and extent, melt ponds are increasingly becoming a common feature of Arctic's landscape, fueling a positive feedback on sea-ice retreat and enabling the formation of massive sub-ice blooms (Arrigo et al. 2014; Horvat et al. 2017). Using RNA-based metabarcoding, Xu et al. (2020) reported that melt ponds harbor a very different and less diverse active protist community compared to that of the sea water, dominated by ciliates, cercozoan and chrysophytes, and controlled by ecological drift rather than dispersal limitation. Kilias et al. (2014) observed that melt pond and bottom-ice protist communities were also dissimilar, with melt ponds being dominated by the genera *Carteria*, *Ochromonas* and *Dinobryon*. These studies indicate that, although subjected to variable degrees of community exchange, sea-ice, melt ponds and water column provide different community-structuring selective pressures on protists, and thus ecosystems shifts imposed by climate change will undoubtedly affect its diversity across polar environments.

4.1.2 Impact of polar phytoplankton diversity on ocean cycles and novel biogeographic patterns

As spring approaches and the light reaching the water column becomes sufficient to support the growth of autotrophic populations, a vast phytoplankton bloom is formed on the on the fringe of the melting sea-ice and many kilometers beneath it (Arrigo et al. 2014). Metabarcoding coupled with other methods can investigate the magnitude and structure of polar phytoplankton blooms and their impact on biogeochemical cycles. For example, Duret et al. (2019) investigated differences between eukaryotic plankton groups and the efficiency of carbon transfer via suspended/sinking particles in the Southern Ocean, using 18S rRNA metabarcoding data as a tool for identifying groups linked to a stronger biological carbon pump. *Chaetoceros*-enriched communities were prone to a more efficient carbon flux to deeper water layers, in contrast to *Phaeocystis*-enriched suspended particles, which tended to go through remineralization within the upper layers of the water column (Duret et al. 2019).

Microdiversity analysis of metabarcoding data (meaning setting a threshold for diversity using a single nucleotide difference instead of fixed percentages of similarity, see section 3.4) has enabled a higher degree of precision on determining ecotype biogeographical patterns within certain species. This approach has been increasingly recognized as ecologically robust, providing insights into evolutionary factors shaping microbial dynamics (Needham et al. 2017) and linked to functional stability within microbial communities (García-García et al. 2019). For example *Phaeocystis*, an important member of high-latitude environments, seems to display biogeographic patterns tightly linked to oceanographic boundaries, especially the polar and sub-antarctic fronts (Sow et al. 2020). Tragin & Vaulot (2019) analyzed microdiversity within Mamiellophyceae members and reported a previously undetected *Micromonas* subarctic clade, which is phylogenetically close to the pan-Arctic *Micromonas polaris*, and possibly has been masked until now by studies using clone libraries and similarity percentages of 98% or less. The *Micromonas* B3 clade has an overlapping distribution with *M. polaris* near 60°N latitude, but seems to have a markedly different distribution band towards lower latitudes down to 45°N, indicating a fairly distinct niche preference (Tragin and Vaulot 2019) (Figure 3) .

FIGURE 3 HERE

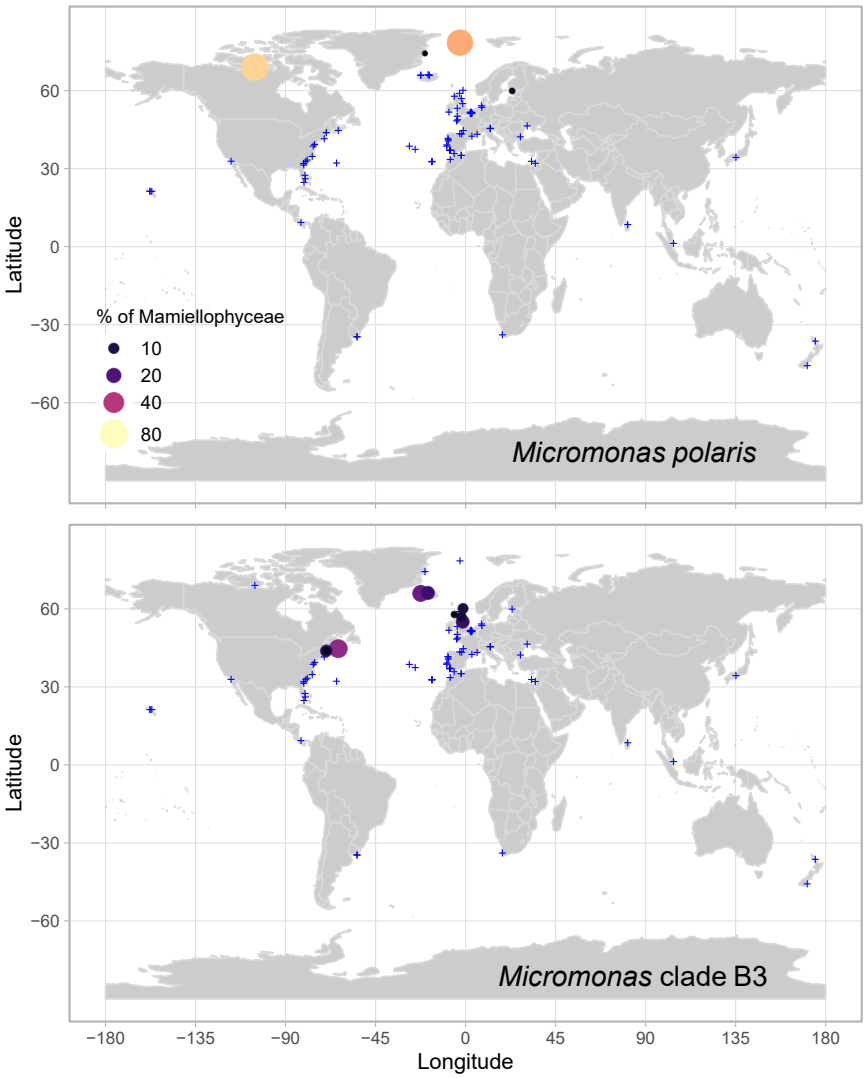

Figure 3: Distribution of the polar *Micromonas polaris* and the sub-arctic *Micromonas* B3 clade at the Ocean Sampling Day stations. Redrawn from Tragin & Vaulot (2019)

4.1.3 Bi-polarity studies as a proof of concept for microbial dispersal theories

Although opposed each other geographically, the North and the South poles share similarities such as extreme seasonal variation of solar energy input and a community phenology tightly linked to sea-ice dynamics. The existence of bipolar taxa evokes questions regarding inter-connectivity paths and rates of evolutionary divergence (Darling et al. 2000; Sul et al. 2013). The use of metabarcoding enables a refined look on planktonic protists (including photosynthetic) that thrive in both Arctic and Antarctic ecosystems, a comparison which was before mainly performed by morphological traits analysis (Fryxell et al. 1981; Montresor et al. 2003). Wolf et al. (2015), using the 18S rRNA V4 region and a 98% OTU similarity level, reported that more than 10% of the OTUs were shared between the two poles, most of them belonging to the rare biosphere, with the exception of two taxa: an unknown alveolate and *Micromonas*. Segawa et al. (2018) analyzed distribution of red snow algae on both poles using metabarcoding of the ITS (faster-evolving genomic region than the 18S rRNA) and showed a high degree of bipolarity, but also a pronounced endemism between Arctic and Antarctic ribotypes. These results question the Baas-Becking dispersion hypothesis that states: “Everything is everywhere, but the environment selects”.

4.1.4 Metabarcoding as a way to measure vulnerability of polar environments

Physico-chemical changes such as ocean warming and acidification have already been reported to affect polar plankton diversity based on experiments testing organism response to a given set of environmental parameters (Benner et al. 2019; Beszteri et al. 2018; Hoppe et al. 2018). However, the responses of the total community may vary from the results obtained in laboratory experiments, since the complexity of interactions can be rarely (if ever) duplicated on controlled environments. The use of metabarcodes to investigate plankton and phytoplankton diversity is an important tool also to analyze and predict their vulnerability to the ongoing drastic changes in these regions of the planet, and is complimentary to experiments on specific organisms, by taking a broader view on the complete community structure.

Metabarcoding can be used as a tool for assessing the dynamic response of communities to a fast-changing environment and a way to monitor fluctuations on both the abundant and the rare biosphere. For example, the ecotype boundaries described by Tragin and Vaulot (2019) and Sow et al. (2020) for Arctic *Micromonas* and Antarctic *Phaeocystis* respectively, can now be monitored as the oceans currents and their physico-chemical parameters change. Moreover, metabarcoding monitoring can be useful at a much wider time scale, comprising the paleo-oceanographic evolution of polar habitats. Pawlowska et al (2020) investigated the microdiversity variability from Arctic planktonic foraminifera from the last 140,000 years using sediment ancient DNA (*sed*aDNA), and reported population-level responses to deglaciation and sea-ice retreat due to Atlantic Water inflow.

4.2 The Biological Carbon Pump

The Biological Carbon Pump (BCP) refers to the process by which organic matter produced photosynthetically in the sunlit surface ocean is transported vertically downwards to depth through particle sinking, vertical mixing, and active transport by larger organisms (Boyd et al. 2019; Turner 2015; Volk and Hoffert 1985). The BCP is fundamental for the functioning of marine ecosystems by 1) removing carbon from the upper ocean contributing to long-term carbon sequestration; 2) providing remineralized sources of inorganic and organic nutrients at depth due to microbial decomposition of downward fluxes; 3) fueling mesopelagic, bathypelagic and benthic ecosystems that rely primarily on sinking particulate organic matter (POM) from the euphotic zone.

Despite its importance for marine ecosystem functioning and earth's climate, our understanding of the BCP, its variability and its role in the carbon cycling is still rudimentary (Falkowski et al. 1998; Honjo et al. 2014). The imbalance between vertical export and metabolic demand in the dark ocean accentuates the knowledge gap around the functioning of the BCP (Burd et al. 2010; Herndl and Reinthaler 2013), which is partly due to the sparsity of export measurements, the high temporal and spatial variability of the biological and physical processes driving the BCP (Boyd et al. 2019; Honjo et al. 2014), as well as the lack of information about the biological composition and ecological interactions that produce and transform sinking POM throughout the water column (Worden et al. 2015).

4.2.1 Molecular approaches applied to biological communities associated with sinking particles

Gravitational sinking is considered the dominant mechanism of POM vertical export, although physical mechanisms, by which POM is captured or drawn down into the deep ocean, as well as the active transport of organic materials by the vertical migration of larger animals can be equally significant (Boyd et al. 2019; Resplandy et al. 2019; Stukel et al. 2017). The importance of community composition and food web structure to the magnitude, strength and efficiency of the BCP is widely accepted

(Bach et al. 2019; Boyd and Newton 1999; De La Rocha and Passow 2007; Henson et al. 2019; Stukel et al. 2011). There is a growing body of evidence indicating that particle properties and processes (e.g. porosity and repackaging), which are intimately linked to plankton community composition and trophic structure, affect the sinking velocity and remineralization of particles (Bach et al. 2019; Cavan et al. 2017; Le Moigne et al. 2014; Puigcorbé et al. 2015; Siegel et al. 2014; Stemmann and Boss 2011). By altering the abundance, size and composition of sinking particles, phytoplankton (through the formation of marine aggregates and/or ballasting of POM), zooplankton (via consumption, transformation, and repackaging of POM) and bacteria (by the remineralization of POM) influence the temporal and spatial characteristics of vertical export.

Vertical export has been traditionally investigated with sediment traps designed to intercept and collect sinking particles (Boyd and Trull 2007; Buesseler et al. 2007; Knauer et al. 1979). Microscopy, flow cytometry, and pigment analysis of the material collected in sediment traps has been used to characterize the biological components and to infer ecological processes driving POM vertical export (Bauerfeind et al. 1997; Durkin et al. 2016; Nodder 1997; Silver and Gowing 1991; Stukel et al. 2013). Taxonomic identification of intact cells and resting spore composition in sediment trap samples has shown the importance of diatoms in sinking marine aggregates (Martin 2011; Rynearson et al. 2013; Scharek et al. 1999; Waite et al. 1992), with potential contribution of other plankton groups such as picophytoplankton (Waite et al. 2000), coccolithophores (Honjo 1976; Rigual Hernández et al. 2020), radiolaria (Decelle et al. 2013; Gowing 1993; Gowing and Coale 1989; Lampitt et al. 2009; Martin et al. 2010; Michaels et al. 1995), and foraminifera (Sautter and Thunell 1989; Schiebel 2002).

Taxonomic composition analysis of sinking particles based solely on morphological criteria is often challenging due to high levels of material degradation/transformation by microbial activity. Bulk and specific geochemical methods can be used to broadly infer the dominant contributors to export flux (e.g. biogenic silica fluxes are often related to diatoms; pigment or organic compound studies are used to infer functional groups, and/or ecological and metabolic pathways and linkages), however they are often not particularly resolutive for individual plankton studies (Nodder 1997). Largely due to these limitations, the contribution of specific plankton groups to export and transformation of particles through consumption and remineralization has remained elusive. The closing sentence of the Silver and Gowing (Silver and Gowing 1991) seminal paper abstract *"Clearly, reasonable inferences from traps require a better understanding of the nature of the 'particles', and particularly, the contribution of living organisms to trap collections" continue to apply.*

Early applications of molecular approaches demonstrated the potential of genomic tools to characterize the biological composition of particles collected by sediment traps. Table 6 provide a summary of the studies that have coupled particle collection methods with molecular approaches to investigate the vertical fluxes. Sequencing of 18S rRNA clone libraries built from water column and sediment trap samples showed for the first time significant differences between the eukaryotic plankton community associated with sinking particles and general water column communities in the eastern subtropical North East Atlantic (Amacher et al. 2009). A similar sampling strategy with DNA-based molecular fingerprinting of protist and cyanobacteria communities at the Bermuda Atlantic Time-Series (BATS) site in the western subtropical Atlantic showed a similar pattern of dissimilarity between water column and traps over a year-round sampling period, suggesting that not all phytoplankton groups contributed equally to vertical export (Amacher et al. 2013). Using metabarcoding, LeCleir et al (2014) showed significant differences in free-living and particle-associated bacterial communities of water column, sinking and incubated POM. Differences in diversity of particle-associated bacterial community after 24 hours of in situ incubation suggested that rapid changes may occur in communities associated with particles sinking from the euphotic zone. Results from these pioneering studies indicated that both the contribution of primary producers and heterotrophic bacteria to export production and remineralization of sinking particles differed between taxonomic groups.

More recent studies applying HTS to sinking POM have also reported significant differences between planktonic and particle-associated communities across contrasting oceanographic regions (Boeuf et al.

2019; Duret et al. 2019; Fontanez et al. 2015; Gutierrez-Rodriguez et al. 2019; Preston et al. 2019). The confirmation of these differences and the prevalence of certain taxa in sinking particles across a broad range of oceanographic conditions and regions further supports the distinctive contributions that specific taxonomic groups make to the processes driving or resulting from vertical export such as particle sinking, remineralization/transformation and, food web connectivity in the ocean.

TABLE 6 HERE

Table 6: Studies of vertical fluxes. na: not applicable; nd: not determined; metaB: metabarcoding; metaG: metagenomics; metaT: metatranscriptomics; HNF:heterotrophic nanoflagellates.

Study	Region	Surface Chl *a* range (mg/m^3)	Sampling Method	Depth of traps deployment (m)	Duration of deployment	Preservation	Molecular approach	Gene marker	Main Finding
Amacher et al., 2009	NASG-E - ESTOC	1.2 - 1.6	CTD-Niskin (euphotic) + VERTEX-type PITs	200, 500	26 h	Unpreserved brine	Clone Libraries	18S rDNA	Small rather than larger (diatom) phytoplankton taxa dominated sequences in trap material
Amacher et al., 2013	NASG-W - Bermuda Atlantic Time-series Study	0.1 - 0.5	CTD-Niskin (euphotic)+ VERTEX-type PITs	150	72 h	Formaldehyde (0.74%) and unpreserved brine	DGGE	18S & 16S rRNA	Small phytoplankton taxa present in trap samples. *Synechococcus* overrepresented in sediment trap samples. Increased abundance of small eukaryotes (Prasinophytes, Prymnesiophytes) in water column following deep mixing not reflected in sediment traps - stimulated production recycled locally.
LeCleir et al., 2013	Subtropical SW Pacific - FeCycle II	0.5 - 2.5	CTD-Niskin (euphotic) and RESPIRE traps	100, 120	24 h - 72 h	Unpreserved brine	MetaB	16S rRNA	Rapid colonization of sinking particles by bacteria in the first 24 h and little changes in the following 72 h. Sequences affiliated to *Roseobacter* enriched in traps relative to water samples conferring them a key role in degradation of sinking particles
Fontanez et al., 2015	NPSG- HOT	nd	CTD-Niskin (euphotic) and VERTEX-type PITs	150, 200, 300, 500	12 days	RNA Later and unpreserved brine	MetaG	18S & 16S rRNA $_{mi}$tags	Distinctive communities in live, poison and water column samples with live traps enriched in both bacterial and eukaryotic taxa associated with particle and dissolved organic matter degradation. Depth-related partitioning of bacterial taxa in sinking particles possibly related to colonization of particles by eukaryotic taxa
Gutierrez-Rodriguez et al., 2019	California Current Ecosystem	0.08 - 0.69	CTD-Niskin (euphotic) + VERTEX-type PITs	60, 100, 150	72 h	Formaldehyde (0.4%) & unpreserved brine	MetaB	18S rRNA	Rhizaria dominance in preserved traps across all trophic conditions suggest important role in export. Important taxon differences between preserved and unpreserved traps with significant increase of heterotrophic protists (Stramenopiles-HNF-*Caecithellus* spp. and Phaeodaria) stress the importance of heterotrophic protist in remineralization and transformation of sinking particles.
Duret et al., 2018	Southern Ocean	0.2 - 2.2	CTD + Marine Snow Catcher	Bottom of mixed layer & Mesopelagic (DCM +110 m)	na	na	MetaB	16S rRNA	Distinctive prokaryotic communities associated with sinking and suspended particles.Pseudomonadales, Vibrionales and Rhodobacterales enriched in sinking particles while Flavobacteriales were enriched in suspended particles
Mestre et al., 2018	Tropical and subtropical Atlantic, Pacific, Indian Oceans	nd	CTD-Niskin (surface to 4000 m)	na	na	na	MetaB	16S rRNA	Most abundant taxa in deep ocean are present in surface waters suggesting strong connectivity between both realms driven mainly through large size particles. Higher dissimilarity among free-living communities indicates stronger vertical partitioning.
Bouef et al., 2019	NPSG - HOT	nd	CTD + PARFLUX Deep sediment trap	4000	12 days/bottle	RNA Later	MetaB (MetaG & MetaT)	18S & 16S rRNA	Systematic dominance of Alteromonadales and Campylobacterales across 9-month survey in all MetaG, MetaT and MetaB libraries. More diverse contribution of eukaryotic groups with Syndiniales, Ciliates and Rhizaria dominating the protistan sequences

Table 6: *(continued)*

Study	Region	Surface Chl *a* range (mg/m^3)	Sampling Method	Depth of traps deployment (m)	Duration of deployment	Preservation	Molecular approach	Gene marker	Main Finding
Duret et al., 2020	Southern Ocean	0.2 - 2.2	CTD + Marine Snow Catcher	Bottom of mixed layer & Mesopelagic (DCM +110 m)	na	na	MetaB	18S rRNA	Diatoms were enriched in particles sinking into the upper mesopelagic while suspended particles here were enriched in Prymnesiophytes. Suggesting higher transfer efficiency for diatom-enriched particles. Diverse community of heterotrophic protist (Choanoflagellates, Radiolaria, dinoflagellates, ciliates and fungi) associated with particles. HNF and Acantharea were more abundant in sinking than suspended particles
Preston et al., 2020	California Current Ecosystem	nd	CTD + PARFLUX Deep sediment trap	3900, 3950	11 days/bottle	RNA Later	MetaB	18S rRNA	Diatoms associated with greater POC flux event to deep ocean, with co-occurring increases of zooplankton sequences suggesting their involvement in diatom-driven production transport. Small phytoplankton taxa were found to contribute substantially to deep export during the winter with metazoan consumers likely acting as vectors, while a diverse suit of Radiolarian taxa dominated the material reaching the traps frequently. Collodaria sequences dominated the community in POM collected between March and April (99%-65% of total reads), although their presence was not associated with increased POC fluxes

4.2.2 (Phyto)plankton and vertical export.

Diatoms are arguably one of the most important contributor to POC export accounting for nearly 40% globally (Jin et al. 2006; Nelson et al. 1995; Sarthou et al. 2005). Their relatively large cell size, silica-rich skeletons and bloom-forming capacity confer diatoms with high export potential (Tréguer et al. 2018). Visual and pigment analysis of the material collected by sediment traps have provided evidence of the important contribution made by diatoms to sinking POM across contrasting trophic conditions, seasons and latitudes, through direct sinking and aggregation, or mediated by zooplankton consumption and faecal pellet production (Boyd et al. 2005; Durkin et al. 2016; Ebersbach et al. 2014; Martin et al. 2010; Scharek et al. 1999; Smetacek et al. 2012). Yet, early molecular studies of sediment trap materials in the low productivity/oligotrophic waters of the North Atlantic Subtropical Gyre revealed lower than expected contributions of diatoms relative to small nanoflagellates and the cyanobacterium *Synechococcus* (Amacher et al. 2009, 2013), which is consistent with recent studies providing evidence of considerable small phytoplankton contribution to export.

In more productive systems such as the California Current upwelling, metabarcoding analysis of POM collected in short-term floating sediment traps deployed in the upper mesopelagic zone (< 500 m) identified diatoms and dinoflagellates as the dominant phytoplankton groups contributing to export production, despite the suppressed upwelling conditions during the survey period in summer (Figure 4 Gutierrez-Rodriguez et al. 2019). Similar analysis conducted on sinking particles collected in deep sediment traps (⩾4000 m) deployed in the same upwelling region over a 9-month period, found that diatoms contribution to total abundance of 18S rRNA reads was relatively low, although the highest POC flux event coincided with samples where diatom sequences in trap samples were higher (Preston et al. 2019).

FIGURE 4 HERE

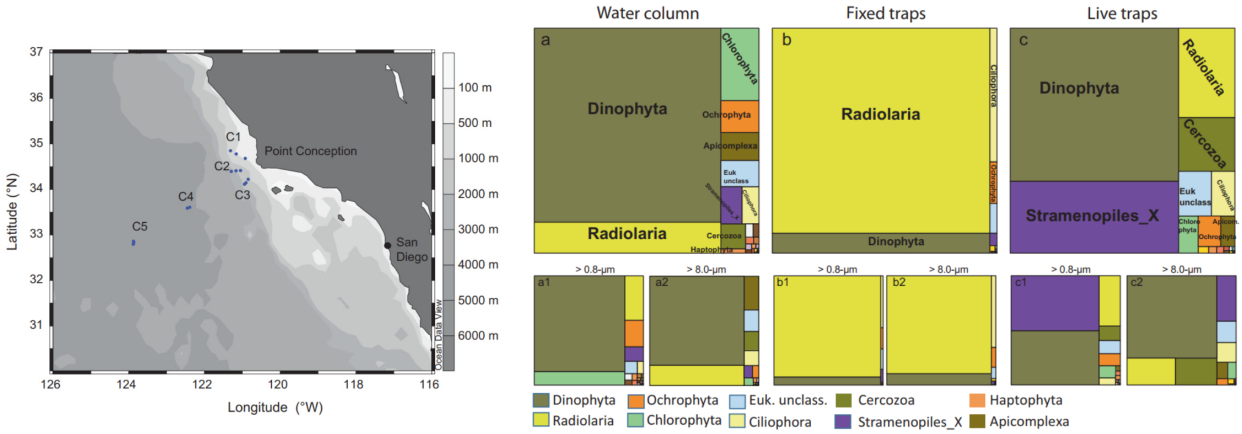

Figure 4: Protist community from the California Current in the water column and in traps where material has been fixed or kept live (reprinted from Gutierrez-Rodriguez et al. 2019).

These authors reported a succession in the diatom species associated with particles reaching the seafloor, with different species of *Chaetoceros* (*C. didymus*, *C. neogracile*, *C. socialis*) dominating in winter, followed by species belonging to *Minidiscus*, *Thalassiosira*, *Actinocyclus*, *Skeletonema* and *Ditylum* genera in spring, while *Thallassiosira aestivalis*, not previously detected in the study, appeared to dominate the diatom assemblage associated with the high POC flux observed in late June (early summer, Preston et al. 2019). The analysis of photosynthetic community composition from 16S rRNA gene metabarcoding provided evidence of the relative dominance of *Synechococcus* and Prasinophytes in sinking particles during early and late winter, respectively, consistent with molecular analysis of sinking particle at BATS (Amacher et al. 2013).

Although there is a general agreement that larger phytoplankton cells are considered to drive vertical

carbon export (Michaels and Silver 1988), the relative contribution of picophytoplankton to downward fluxes and efficiency of the BCP are still under debate (Henson et al. 2019; Puigcorbé et al. 2015; Richardson and Jackson 2007; Stukel et al. 2013; Waite et al. 2000). The formation of marine aggregates and trophic repackaging into faecal pellets, particularly by pelagic tunicates like salps and appendicularians (Stone and Steinberg 2016), are considered the main mechanisms by which small phytoplankton contribute to downward fluxes (Richardson 2019).

Duret et al. (2019) investigated the phytoplankton community composition of sinking and suspended particles sampled with the Marine Snow Catcher in four stations with different productivity in the Scotia Sea (Southern Ocean). The study showed that diatoms and prymnesiophytes were enriched in sinking and suspended particles, respectively. These results suggest that diatoms are more efficient in carbon transfer to the upper mesopelagic than prymnesiophytes in this area, since the latter are more easily disintegrated into suspended particles, therefore supporting the paradigm of higher export potential of larger (diatom) phytoplankton. Metabarcoding analysis of sinking particles, especially if combined with size fractionation and biogeochemical analyses, can provide valuable information in this regard.

Dinoflagellata and Radiolaria groups often dominate the protist community associated with sinking organic particles below the euphotic zone (Boeuf et al. 2019; Duret et al. 2019; Fontanez et al. 2015; Gutierrez-Rodriguez et al. 2019; Preston et al. 2019). Their high relative abundance revealed by molecular analysis is likely influenced, to some extent, by the high number of 18S rRNA gene copies characteristic of these groups (Biard et al. 2016; Lin 2011). However, their tendency to sink directly and to produce fast sinking faecal pellets cannot be ignored. For instance, the siliceous or strontium skeleton of polycystine radiolaria and acantharia, the large size (mm to cm) that solitary and colonial forms of some species can reach, and the sticky nature of the pseudopodia and mucilaginous matrix that these amoeboid organisms use as feeding structures, are attributes that are also consistent with a high export potential.

Metabarcoding analysis of 18S rRNA gene of sinking material reaching the seafloor (4000 m) of the California Current over extended periods of time showed the dominance of different radiolarian taxa (Collodaria, Spumellarida and Acantharea and RAD-A, Preston et al. 2019). Similar analysis obtained from POM collected with moored traps ($\geqslant$ 4000 m) deployed in the North Pacific Subtropical Gyre reported sporadic dominance of rhizarian that were affiliated mainly to Foraminifera and Phaeodaria (Boeuf et al. 2019). Whether the abundance of Rhizaria in trap-collected POM is due to direct sinking of mixotrophic species from the sunlit layer and enhanced vertical export, or due to consumption and remineralization of sinking particles by heterotrophic species that favor the attenuation of vertical fluxes, is not clear.

While metabarcoding can be only be taken, at best, as a semi-quantitative approach, these studies demonstrate that, when combined with targeted sampling strategies, it can provide valuable information about biological sources and ecological processes transforming sinking POM. Allied with elemental and size composition of particles collected with particle collection methods (e.g. sediment traps, *in situ* pumps), and indirect estimates of export (e.g. 238U-234Th disequilibria, oxygen and nutrient mass balances), this molecular approach can improve our understanding of the role of different plankton groups in carbon export. Any step towards improving the quantitative nature of metabarcoding analysis (e.g. reducing uncertainty around taxon-specific copy number variability) and integration with biogeochemical metrics (sequence to carbon conversion factors) will certainly enhance the value of metabarcoding approaches, not only for biogeochemical studies but also for diversity and ecological studies.

4.2.3 The importance of sampling strategy and particle collection methods

It is acknowledged that different methods commonly used to collect POM exported from the euphotic zone (e.g. sediment traps, in situ pumps, marine snow catcher, size fractionation of water bottle samples) are biased towards the different types of particles collected (Lee 2019; Peterson et al. 2005). VERTEX-type or Particle Interceptor Traps (PITs), for instance, are more efficient at capturing rapidly sinking particles than suspended particles and marine snow (McDonnell et al. 2015). Differences in POC and particulate trace-metal concentrations observed in CTD-Niskin bottle and high volume in situ pump samples suggested different sampling biases between these methods (Liu et al. 2009; Twining et al. 2015), highlighting the importance of combining different methods whenever possible, and the need to consider these biases when comparing and interpreting results from different studies (Puigcorbé et al. 2020). These methodological differences can also be used to our advantage. **Duret2020** for instance used the marine snow catcher that allows separate collection of sinking and suspended particles (Lampitt et al. 1993) in combination with size fractionation and DNA metabarcoding analysis to investigate differences in their biological sources and particle transformations in the Southern Ocean. These authors observed differences in the relative abundance of phytoplankton groups between sinking particles (enriched in diatoms) and suspended particles (enriched in prymnesiophytes).

Sediment traps are commonly filled with a brine mixed with a preservative (mainly formalin or mercuric chloride), to prevent further degradation of captured POM through the duration of the deployment (Knauer et al. 1984). However, particle solubilization is likely to be an issue, especially in moored, time-series sediment traps leading to the loss of more labile biological components of the downward flux (Antia 2005). These preservatives can also negatively affect the efficiency of DNA extraction and amplification, limiting the applicability of downstream molecular analysis (Gilbert et al. 2007). DNA has been successfully extracted and amplified from formalin-preserved sediment trap PIT tubes following short-term (days) deployments (Gutierrez-Rodriguez et al. 2019). The use of RNA-later as a preservative is likely to facilitate downstream molecular analysis of particles collected in VERTEX-type traps when used within long-term (month), moored traps (Preston et al. 2019), although the elevated cost of RNA-later compared to standards preservatives might be a limiting factor.

4.3 Predator-Prey interactions and trophic connectivity.

Functioning of ecosystems is maintained by the flow of material and energy through the food webs, which are a representation of trophic interactions between species (Thompson et al. 2012). In marine pelagic ecosystems, the bulk of organic material and energy is produced and consumed at the base of the food web, where the three domains of life constitute an intricate network of interactions. The microbial component of the food web comprises many trophic levels, through which over 95% of total primary and secondary production is processed before reaching larger metazoan zooplankton that are directly available as prey to higher trophic levels (Steinberg and Landry 2017). Despite their crucial role, our current knowledge of the structural and dynamic properties of planktonic food webs (i.e.'what eats what') remains fairly incomplete. This is due in part because of the extreme functional and taxonomic diverse nature of eukaryotic plankton (Caron et al. 2012; Mitra et al. 2016) and because of the high degree of omnivory and dynamic nature of predator-prey interactions (Banse 2013; Calbet and Saiz 2005; Isaacs 1972; Steinberg and Landry 2017; Zeldis et al. 2002). Conventional methods used for dietary and food web reconstruction (e.g. stable isotope analysis, visual inspection of stomach contents) are somewhat inadequate for such diverse and often unstructured plankton systems (Craig et al. 2013; Landry 2002; Maloy et al. 2013).

Metabarcoding represents an alternative that is rapidly gaining traction in trophic ecology due to its high sensitivity, taxonomic resolution and cost-effective nature (Clare 2014; King et al. 2008; Kress

et al. 2015; Majdi et al. 2018). When applied to gut, stomach or scat contents, metabarcoding can provide taxonomically comprehensive information on what a predator eats (Deagle et al. 2019), and allow for dietary richness estimates that illuminate niche partitioning from plankton to marine fishes, birds and mammals (Carroll et al. 2019; Casey et al. 2019; Craig et al. 2013; Kaunisto et al. 2017; Leray et al. 2015; Ray et al. 2016). Zamora-Terol et al. (2020) investigated the species-specific trophic links and seasonality in the Baltic Sea by applying 16S/18S rRNA metabarcoding to the gut content of several species of relevant metazooplankton groups (copepods and cladocerans). Their comprehensive study revealed the shift from diatoms and dinoflagellate dominance in copepods during the spring, while in the summer zooplankton including cladocerans showed a more diverse diet, dominated by cyanobacteria and heterotrophic protists. Their results indicated that copepods change their natural diet over seasons, and overall adapt their feeding strategies to the available prey spectrum.

Generating a good quality PCR product for the specific genomic region and group of interest remains a crucial step of the DNA metabarcoding workflow. This can be particularly challenging in the case of gut content analysis, where non-targeted DNA from the predator represents the majority of the originally extracted DNA material. To minimize the contribution of the predator's material to the PCR amplicon it is recommended to dissect out the predator gut/stomach where possible, and to adopt molecular strategies (e.g. protist-specific or blocking primers and restriction enzymes) that restrain the amplification of the predator's DNA (Bower et al. 2004; Craig et al. 2013; Maloy et al. 2013; O'Rorke et al. 2013; Vestheim and Jarman 2008).

4.4 Photosynthetic picoeukaryotes in Chinese lakes

Lakes, ponds, rivers and reservoirs harbor an array of micro-niches linked to nutrient levels, light availability, temperature and oxygen concentrations. These water bodies do not have a continuous geographical distribution which can also contribute to local endemism (Fernández et al. 2017). Compared to the marine environment, the diversity and ecology of freshwater phytoplankton (and plankton in general) have been far less investigated. Fernández et al. (2017) demonstrated that metabarcodes of phytoplanktonic cells sorted by flow cytometry can be a powerful strategy and an alternative to the filtered samples to reveal novel diversity among photosynthetic picoeukaryotes (PPE, $\leqslant 3\mu$m), a key component of primary production in aquatic ecosystems. Here we summarize the recent metabarcoding findings on picophytoplankton diversity obtained by the combination of flow cytometry sorting and metabarcoding from Chinese lakes ranging from oligotrophic to highly eutrophic status, present in highly urbanized areas (Li et al. 2017; Marie et al. 2010; Shi et al. 2009).

4.4.1 Lake Fuxian - oligotrophic

Fuxian is the second deepest lake in China and famous for its lost ancient city beneath its waters. With an average of total phosphorus and nitrogen concentrations around 0.015 and 0.30 mg/L^{-1}, lake Fuxian is a subtropical and oligotrophic freshwater lake. Picophytoplankton contribution to primary productivity on this lake can reach 66% (Shi et al. 2019). Fuxian PPEs were mainly composed of Chlorophyceae and Trebouxiophyceae in the spring, while in summer, they were replaced by Chrysophyceae and Prymnesiophyceae. Eustigmatophyceae and Chlorophyceae became the major PPEs in autumn, and Dinophyceae during the winter (Shi et al. 2019). The most abundant OTU in Lake Fuxian was affiliated with the chrysophyte *Dinobryon sociale* which has been reported to have high affinity for environment with low concentrations of inorganic phosphate and a capacity to absorb organically bound phosphate (Dokulil and Skolaut 1991; Lehman 1976). The Prymnesiophyceae *Chrysochromulina parva*, a common species in oligotrophic lakes, was also abundant. *Chrysochromulina parva* is a mixotrophic phytoplankton and potentially toxic species(Hansen et al. 1994). Chlorophyceae, including *Volvox aureus* and *Tetradesmus obliquus* were also abundant(Shi et al. 2019).

4.4.2 Lakes from the lower reaches of the Yangtze River - mesotrophic and eutrophic

PPEs community structure was investigated in more than 20 mesotrophic and eutrophic lakes along the middle-lower reaches of the Yangtze River in China (Shi et al. 2018b). The contribution of eukaryotes to total picophytoplankton abundance was generally higher in eutrophic lakes than in mesotrophic lakes. At the class level, PPEs were mainly dominated by three taxonomic groups: Cryptophyceae, Coscinodiscophyceae and Chlorophyceae. Shannon diversity was significantly higher in mesotrophic lakes than in eutrophic lakes. Coscinodiscophyceae dominated in the most eutrophic lakes, while Chrysophyceae, Dinophyceae and other classes of PPEs were more abundant in mesotrophic lakes. *Cyclotella atomus*, *Chlamydomonas* sp. and *Poterioochromonas malhamensis* tended to be more prevalent in eutrophic lakes.

4.4.3 Lake Poyang - eutrophic

Lake Poyang, the largest freshwater lake in China is connected to a river and undergoes dramatic intra-annual fluctuations of its level ranging from 7 m in the dry season to 19 m during the flood season, a difference of almost 12 m, resulting in the unique landscape that has been described as "flooding like the sea, drying like a thread". Total phosphorus and nitrogen concentrations are 0.13 and 1.72 mg/L^{-1}, respectively. A recent study showed that Chlorophyceae were dominating the PPE community (Shi et al. 2018a) especially in autumn. During spring, Mediophyceae, Trebouxiophyceae and Dinophyceae contributed to PPEs at certain sampling sites. Cryptophyceae were mainly observed during summer and winter, while Eustigmatophyceae were mainly observed during winter. The Chlorophyceae (order Sphaeropleales) *Mychonastes* was a dominant genus in Lake Poyang and accounted for approximately 66% of the PPE sequences. This genus was originally discovered in brackish waters of the Chesapeake Bay (Simpson and Van Valkenburg 1978) but this green algae has been also observed in intertidal environments (Margulis et al. 1988) and in another eutrophic lake (Hanagata et al. 1999).

4.4.4 Lake Taihu and Lake Chaohu, highly eutrophic

Lake Taihu and Lake Chaohu are the shallow lakes that are subjected since the 1990s to heavy eutrophication with regular cyanobacterial blooms covering large regions of the lake surface during the summers. The average total phosphorus and nitrogen concentrations are around 0.13 and 2.52 mg/L^{-1}, respectively. Chlorophyta and Stramenopiles dominated PPE diversity in the two lakes(Li et al. 2017). Metabarcoding demonstrated the presence of a close relative of the widespread marine PPE genus *Ostreococcus* (Tragin and Vaulot 2019) which was abundant especially in Lake Taihu in winter and spring. The sequence of another freshwater member of the family Bathycoccaceae (genus *Bathycoccus*) has also been observed in lake Musters (Argentinean Patagonia, Lara et al. 2017). These are the only two reports of Bathycoccaceae in freshwater which remain uncultivated.

Bacillariophyta and Chrysophyceae formed the two largest groups of autotrophic Stramenopiles. Despite the diversity of Chrysophyceae, they only accounted for a small proportion of reads in both lakes. Comparatively, Bacillariophyta were more abundant (27.4% and 20.6% of the reads in Lake Taihu and Lake Chaohu, respectively) but relatively less diverse. Most diatoms belonged to Coscinodiscophyceae and Bacillariophyceae. Other photosynthetic or mixotrophic groups included Cryptophyta, Euglenozoa, Dinophyceae and Prymnesiophyceae.

In Lake Chaohu, Chlorophyceae were mainly dominated in the Spring by two OTUs, affiliated to , and more diverse in other seasons. Similarly, in Lake Chaohu in autumn, three OTUs dominated, two bel-al specific peaks. Within Chlorophyceae an OTU affiliated to *Monoraphidium* sp. was most abundant in the spring in both lakes. An OTU matching an uncultured Chlorophyta sequence with 100% identity appeared in high abundance mainly in winter samples, especially in Lake Chaohu,

whereas OTUs affiliated to *Chlorella sorokiniana* (99% identity) and an uncultured Chlorophyta (100% identity) were found more prevalent in Lake Taihu in winter and spring. This seasonality also applied to many other OTUs, such as two Synurophyceae OTUs appearing in summer in Lake Taihu or an Eustigmatophyceae OTU found in winter in Lake Chaohu. An OTU affiliated to Prymnesiophyceae was dominant exclusively in winter in both lakes.

It had been suggested that seasonal changes in environmental variables of lake Chaohu are more important than trophic interactions for seasonal community succession(Shi et al. 2020). Temperature appeared to be the most important factor shaping PPEs community structure. Three temperature ranges could be defined harboring different PPE communities: high, intermediate, and low with boundaries temperatures > 21.8, 21.8-9.8 and < 9.8 °C, respectively. Cryptophyta were dominant at the intermediate temperatures level and Bacillariophyta at low ones. In comparison, for Chlorophyta PPEs, different orders corresponded to different temperature ranges.

The combination of flow cytometric sorting and metabarcoding also revealed the widespread existence of fungi attached to photosynthetic picoeukaryotes (PPE-attached fungi) in Lake Chaohu. PPE-attached fungal communities were mainly composed of Basidiomycota, Chytridiomycota and Ascomycota. Temperature, Si and PPE community structure were the most important driving factors for the temporal succession of PPE-attached fungal communities. The top 5 most abundant OTUs were affiliated to *Cladosporium endophytica*, unclassified Dothideomycetes, *Malassezia restricta*, *Tilletiopsis* sp. and *Wallemia sebi*. Phylogenetic molecular ecological networks indicated that the interaction between PPE-attached fungi and PPEs changed from antagonistic to cooperative with the decline in temperature. The most abundant OTU of PPE-attached fungi were affiliated with the *Cladosporium*, the most common saprophytic fungus. Whereas, the most fungal hub taxa were Chytridiomycota, the main parasite fungi of phytoplankton (Liu et al. 2020).

5. Marine Picocyanobacteria

5.1 Use of the universal marker gene, the 16S rRNA

Marine picocyanobacteria of the genera *Prochlorococcus* and *Synechococcus* are the two most abundant photosynthetic prokaryotes in the ocean, contributing up to 8 and 16 % of the net primary productivity, respectively (Flombaum et al. 2013; Scanlan et al. 2009). One of the main differences between the two genera resides in their photosynthetic apparatus. Like most cyanobacteria, *Synechococcus* main light-harvesting antenna is composed of phycobilisomes, made of a complex combination of phycobiliproteins (phycocyanin and phycoerythrins), each binding one to three chromophores (phycocyanobilin, phycoerythrobilin and/or phycourobilin), able to harvest different wavelengths (Six et al. 2007). Most *Prochlorococcus* lineages lacks phycobilisomes (like *Prochloron* and *Prochlorothrix*), although phycobilisomes were recently discovered in the basal lineages (Ulloa et al. 2021).*Prochlorococcus* main light-harvesting antenna complex is composed of prochlorophyte chlorophyll-binding proteins (Pcb), which binds divinyl chlorophylls *a* and *b* (Laroche et al. 1996). These differences in photosynthetic pigments and cell size between both genera cause distinct flow cytometry signatures, which were used early on to reveal their respective abundances and distributions, and highlighted some of the main ecological idiosyncrasies of each genus. When they co-occur, *Prochlorococcus* cell abundance is about one to two orders of magnitude larger than *Synechococcus*. *Prochlorococcus* is more abundant in warm (> 15 °C) oligotrophic waters, while *Synechococcus* dominates in coastal and more temperate or mesotrophic open ocean waters (Blanchot and Rodier 1996; Campbell and Vaulot 1993; Olson et al. 1985; Partensky et al. 1999).

Studies using 16S rRNA gene, clone libraries and Sanger sequencing brought the first insights into marine picocyanobacteria genetic diversity. In agreement with pigment data, these studies demon-

strated the existence of two distinct *Prochlorococcus* ribotypes with different adaptations to light availability (Moore et al. 1998; Scanlan and West 2002; Urbach et al. 1998): a "high light-adapted ecotype" (HL) containing sequences from *Prochlorococcus* cells collected in surface waters and a deeply branching lineage called "low light-adapted ecotype" (LL), including sequences from members isolated deeper in the water column. Similarly, *Synechococcus* 16S rRNA gene sequences from various marine isolates led to the identification of up to 10 clades within the Marine Cluster-A (Fuller et al. 2003; Urbach et al. 1992, 1998; West and Scanlan 1999), which dominated in oceanic waters and was later renamed subcluster (SC) 5.1 (Herdman et al. 2001). The use of this marker has notably allowed to start describing the vertical distribution of *Prochlorococcus* populations *in situ* (West et al. 2001; West and Scanlan 1999) and the global biogeography of the four dominant *Synechococcus* clades (I to IV) (Bouman et al. 2006; Fuller et al. 2006; Fuller et al. 2003, 2005; Zwirglmaier et al. 2007, 2008).

These early studies targeted only clades established from isolated strains and a few environmental sequences, therefore likely missed an important part of the uncultivated diversity. It is also important to note that the number of reads recovered from the Cyanobacteria phylum using general 16S rRNA gene primers usually hardly exceeds 20% of the total reads abundance even when applying HTS methods such as metagenomics and metabarcoding (Doherty et al. 2017; Sunagawa et al. 2015). Only recently that Huber et al. 2019 developed new 16S rRNA primers in the V5-V7 hypervariable regions targeting specifically picocyanobacteria and suitable to metabarcoding.

5.2 More resolutive markers

A general disadvantage of bacterial ribosome-based diversity analyses is the relatively low sequence divergence between closely related groups or species. For marine picocyanobacteria, the considerable genetic diversity observed within both *Prochlorococcus* and *Synechococcus* genera is only manifested through small-scale variations, about 2.8% and 4.5%, respectively (Doré et al. 2020). The well known high- and low-light adapted clades of *Prochlorococcus*, for example, differ by less than 3% of their 16S rRNA sequences (Biller et al. 2015). Moreover, the number of rRNA gene copies varies among bacteria, including picocyanobacteria. For example, two to three identical copies of the rRNA operon are present in most marine *Synechococcus* genomes (Dufresne et al. 2008; Farrant et al. 2016; Fuller et al. 2003), and two in some members of LL adapted *Prochlorococcus* clades, although most *Prochlorococcus* genomes have only one (Martiny et al. 2009).

These limitations led the scientific community to use other marker genes to analyze the genetic diversity of marine picocyanobacteria (Table 7). Generally these markers show more variability than the 16S rRNA gene, and thus provide more phylogenetic resolution over a short sequence stretch. For instance, the sequence identities between different picocyanobacteria clades are lower for ITS ($<$ 93%; Rocap et al. 2002), *petB* ($<$ 94 %: Mazard et al. 2012); *ntcA* (86%: Penno et al. 2006), *rpoC1* (82%: Toledo and Palenik 1997) than for 16S rRNA ($<$ 97%: Fuller et al. 2003; Moore et al. 1998). Furthermore, most protein encoding genes are single-copy, like *rpoC1* (Bergsland and Haselkorn, 1991; Tai and Panik 2009), *petB* (Mazard et al., 2012; Farrant et al., 2016), *ntcA* (Lindell et al., 1998; Penno et al., 2006), *rbcL* (Scanlan et al. 2009), and *narB* (Scanlan et al. 2009). The *psbA* multigene family is an exception to the list, since it can be found in a variable number (1 to 6) of copies per genome in marine *Synechococcus* and *Cyanobium*. It was shown to be subjected to intragenome homogenization (D1 protein of photosystem II, PSII), most probably mediated by gene conversion (Garczarek et al. 2008). However, numerous genes involved in carbon and energy metabolism (e.g. *psbA* or some subunits of the F0F1 ATP synthase) have been found to be present in cyanophages, which could bias the interpretation of diversity studies based on these markers (Chénard and Suttle 2008; Mann et al. 2003; Millard et al. 2004). Similarly, as mentioned before, a good molecular marker should also not be the subject to horizontal gene transfer which constitute a common, although not so frequent, gene gain mechanism in marine picocyanobacteria (Doré et al. 2020; Martiny et al. 2009). For most marker genes, lateral gene transfer has been checked by showing the congruence between gene marker and 16S rRNA phylogenies (Mazard et al. 2012; Urbach et al. 1998).

TABLE 7 HERE

Table 7: Genes used for metabarcoding of *Synechococcus*. RuBisCO: Ribulose 1,5-bisphosphate carboxylase-oxygenase.

Gene	Coding for	References
ITS	16S-23S rRNA internal transcribed spacer	Rocap et al. 2002 Ernst et al. 2003 Chen et al. 2006 Cai et al. 2010 Choi and Noh 2009 Ahlgren and Rocap 2012 Huang et al. 2012 Kashtan et al. 2014
rpoC1	gamma subunit of RNA polymerase	Palenik 1994 Toledo and Palenik 1997 Ferris and Palenik 1998 Toledo and Palenik 1997 Ma et al. 2004 Muhling et al. 2005 Tai and Palenik 2009 Gutiérrez-Rodríguez et al. 2014 Xia et al. 2019
petB	cytochrome b6	Mazard et al. 2012 Farrant et al. 2016
rbcL	large subunit of the RuBisCO	Chen et al. 2004 Chen et al. 2006 Paerl et al. 2012
narB	nitrate reductase	Paerl et al. 2008 Paerl et al. 2011 Jenkins et al. 2006
atpBE	intergenic region	Lockhart and Penny 1992
psbA	D1 protein of photosystem II reaction center	Kishino et al. 1990 Hess et al. 1995 Zeidner et al. 2003
cpeB	C-phycoerythrin class I, beta subunit	Steglich et al. 2003
ntcA	N-regulation gene	Penno et al. 2006 Post et al. 2011

Distinct marker genes also differ on how easily and reliably sequences can be accessed and/or aligned. Some of them, such as *petB* or *ntcA*, are highly conserved in length and sequences, allowing sequence to be automatically aligned (Mazard et al. 2012; Penno et al. 2006). In contrast, *rpoC1* and ITS gene sequences often require manual refinement of the alignment with the necessity to split the analysis per taxa and/or to suppress the most variable parts. For the ITS, the presence of inserts of variable size between members of *Synechococcus* SC 5.1, 5.2 and 5.3 makes it impossible to automatically align the sequences from different SC (Mella-Flores et al. 2012; Rocap et al. 2002). The higher the sequence variability of the gene marker is, the harder will be the identification of conserved primer binding areas leading to the design of degenerated primers. Finally, the use of new (or not widely used) markers also suffers from databases with a low number of reference sequences, which can lead to primer amplification biases. For example, metabarcoding *petB* primer set is heavily biased towards marine *Synechococcus* SC 5.1 (Mazard et al. 2012), although *petB* gene is an excellent marker to reliably assign both *Prochlorococcus* and *Synechococcus* with a high taxonomical resolution (Farrant et al. 2016). For *petB*, the $_{mi}$tags approach seem to be the alternative (see section 5.3).

5.3 $_{mi}$tags as an alternative to picocyanobacteria metabarcoding

Metagenomes, an approach devoid of PCR-amplification biases, appear particularly suitable to study genetic and functional diversity of marine picocyanobacteria. The high abundance and ubiquity of these microorganisms *in situ* compared to other taxa (Biller et al. 2015; Partensky et al. 1999) results in the dominance of their gene sequences in oceanic metagenomics datasets (DeLong et al. 2006; Rusch et al. 2007; Venter et al. 2004). In addition, numerous complete or near-complete picocyanobacterial genomes are available and can be used as references to annotate taxonomically and functionally environmental reads (Biller et al. 2015; Doré et al. 2020; Dufresne et al. 2008; Kettler et al. 2007; Scanlan et al. 2009). These characteristics allowed scientists to search *Prochlorococcus* and *Synechococcus* sequences in early environmental metagenomic libraries, such as the whole-genome shotgun datasets from the Sargasso Sea (Mühling et al. 2006; Venter et al. 2004), or the Global Ocean Sampling (GOS) (Biers et al. 2009; Huang et al. 2012; Rusch et al. 2007) projects.

With the improvement of HTS techniques and therefore increase of sequencing power, deeper sequenced metagenomes became available, (e.g. those generated during the Tara *Oceans* expeditions discussed in Chapter 3.2) (Karsenti et al. 2010; Pierella Karlusich et al. 2020) allowing to quantitatively recruit marker genes sequences from metagenomes to access the taxonomic diversity and structure of prokaryotic communities though an approach called $_{mi}$tags (Logares et al. 2014). First applied to 16S rRNA, $_{mi}$tags approach recovers sufficient reads for community taxonomic profiling as well as for richness, evenness, and beta diversity estimations. Using a reference database of the *petB* gene, Farrant et al. (2016) recruited reads out of 109 metagenomes from the Tara *Oceans* expedition (Figure 5). This work revealed novel genetic diversity within both *Prochlorococcus* and *Synechococcus*, even among the most abundant and well-characterized clades. In addition, 136 completely new *petB* sequences were successfully assembled from metagenomic reads, enriching the reference database. By combining the genetic information with environmental parameters, "ecologically significant taxonomic units (ESTUs)" were further defined representing organisms belonging to the same phylogenetic clade defined with *petB* and occupying a common ecological niche.

FIGURE 5 HERE

For example, within clade II, ESTU IIB was found to co-occur with clades I and IV in cold waters, while ESTU IIA dominated in warm, (sub)tropical waters. With respect to clade III, although the presence of the two ESTUs defined within this clade was confirmed in warm, oligotrophic waters, they seem to be particularly thriving in phosphorus depleted waters, a previously unknown characteristic for members of this clade. This was later confirmed by comparative genomics (Doré et al. 2020). The CRD1 and EnvB *Synechococcus* genotypes were shown to dominate the Pacific stations analyzed between 33 °S and 35 °N, and to be locally abundant in the South and North Atlantic as well as in the Indian ocean, all corresponding to Fe-limited ocean regions (Caputi et al. 2019; Farrant et al. 2016). Moreover, each of these clades could be split into three distinct ESTUs showing different distribution patterns: CRD1B and EnvBB thrive in cold waters, CRD1C and EnvBC in warm areas, while CRD1A and EnvBA were found in both types of ecological niches, suggesting that these latter populations are able to acclimate to a wide range of temperatures.

Although the metagenomic recruitment of marker genes has not been popular among studies focusing on eukaryotic phytoplankton (in particular due to the low representation of eukaryotic reads in environmental metagenome datasets), the $_{mi}$tags approach seem to constitute a powerful alternative to the metabarcoding for picocyanobacteria community allowing to have access to the still undescribed lineages and microdiversity within existing lineages and to outline the physico-chemical parameters driving the distribution of these lineages using the numerous metagenomics datasets now available. However, the recovery of sequences from low abundance components of the plankton community requires surveys with sequence depth coverage (i.e. the number of sequences obtained per sample) that are many orders of magnitude larger than those usually reported in the literature. Metagenomics thus remains complementary to metabarcoding approaches, which are still faster, less expensive, and easier

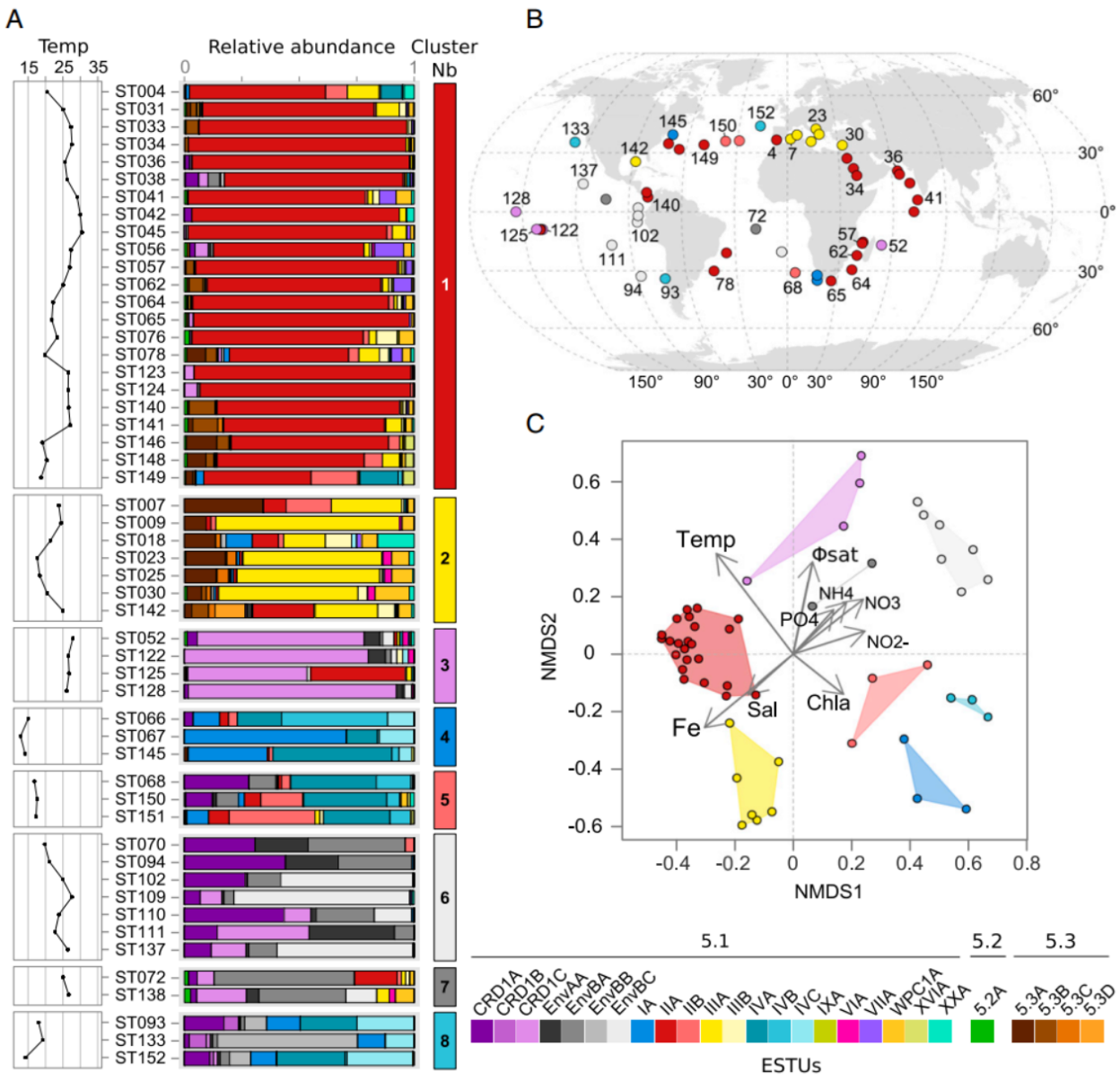

Figure 5: Biogeography of *Synechococcus* ESTUs in surface Tara Oceans metagenomes and relation to physico-chemical parameters. (A) Histograms of the relative abundance of *Synechococcus* ESTUs at each station sorted by similarity. Left panels indicate seawater temperature (°C) at each station. (B) Distribution of the ESTU assemblages, color-coded as in A, along the Tara Oceans transect. (C) NMDS analysis of stations according to Bray–Curtis distance between *Synechococcus* assemblages, with fitted physico-chemical parameters. Reprinted from Farrant et al. (2016).

to analyze, especially for long term monitoring or studies including large numbers of samples.

6. Future directions

Metabarcoding approach have brought an avalanche of exciting new data on the diversity of plankton. Key scientific findings were enabled by metabarcoding, as reviewed by Santoferrara (2020). For example, it revealed the existence of "rare biosphere", where a large number of species are detected in very low abundance (Sogin et al. 2006). While the "rare biosphere" is seemingly less important in ecosystem functioning when compared to abundant groups, it is hypothesized that they act as a "bank of redundant functions" that help to maintain a continuous ecosystem following shifts in environmental

conditions, making the ecosystem more resilient to change (Caron and Countway 2009; Dawson and Hagen 2009; Dunthorn et al. 2014). Metabarcoding also revealed novel and diverse lineages among well-known taxonomic groups (eg. de Vargas et al. 2015).

Although metabarcoding is subject to various technical biases, nowadays no other method can offer the analysis of a large numbers of samples with the same sensitivity, robustness and taxonomic resolution at a similar cost. These exceptional characteristics have led metabarcoding to be pointed and applied as a promising tool in biomonitoring programs, in both freshwater (Kermarrec et al. 2014; Visco et al. 2015; Zimmermann et al. 2015) and marine ecosystems (Chariton et al. 2015; Pawlowski et al. 2014, 2016). Cordier et al. (2018) went further and showed that metabarcoding combined with supervised machine learning (SML) can accurately predict biotic indices regardless of the taxonomic affiliation of the sequences. The predictive models were built using metabarcoding data from five different variable region within the ribosomal small subunit rRNA gene for monitoring the impact of salmon farming activities in Norway on the benthic communities (Cordier et al. 2018).

Because of the high sensitivity of metabarcoding (e.g. trace concentrations of DNA can be PCR amplified and sequence), this method have been applied to the characterization of food webs and predator-prey interactions, as mentioned in section 4.3. However, metabarcoding is a qualitative diversity tool and the nature of the sample as well as the several complex laboratory and bioinformatics steps of this technique can affect the final sequence abundance. The way read counts are used – i.e., frequency of occurrence of different prey taxa or their relative abundance – can influence derived dietary metrics (Deagle et al. 2019). The application of quantitative molecular techniques such as qPCR to assess absolute abundance of targeted groups can help ground-truth metabarcoding results (Schwarz et al. 2018). Similarly, combination of DNA sequencing with more quantitative approaches such as stable isotopes commonly used in trophic ecology has the potential to produce taxonomically resolved predator-prey interactions and improve dietary inferences from metabarcoding and associated material and energetic fluxes within the marine food web (Hardy et al. 2010; Nakamura et al. 2020; Whitaker et al. 2019).

The possibility of applying metabarcoding to archived samples also adds an historical angle that will allow ecosystem and biodiversity changes to be assessed through time (Metfies et al. 2017). The development of methods allowing for DNA extraction and amplification from preserved, sediment traps acquired over a time-series for example represents not only an affordable solution but also a significant step towards integration of molecular and biogeochemical measurements and opens the possibility for retrospective analyses of archived samples from long-term sampling programs. Shiozaki et al. (2021) have proposed that the DNA extraction of preserved plankton samples with a lysis solution of borate-NaOH buffer (pH 11), SDS and proteinase K can effectively remove the cross-links between acid nucleic and cellular proteins formed by formalin fixation. Studies on the DNA extraction of DNA formalin-fixed samples stored in museums and laboratories are scarce (Bucklin and Allen 2004; Ruane and Austin 2017; Schander and Kenneth 2003) and given the increasing use of molecular approaches and the variety of preservation methods used for plankton samples, further research on the effects of different preservatives on molecular and biogeochemical measurements is warranted.

Any DNA sequence without a proper database for comparison is just a string of letters with no meaning. Reference sequences databases linked to taxonomically identified specimens constitute the heart of metabarcoding application. Taxonomy is the language used to structure our "chaotic" natural world and the correct association of organisms names with genetic data is the keystone of modern ecological research, including plankton ecological research (e.g. the impact of biodiversity loss, identification of invasive species, description of biogeographic patterns, etc...). Founded almost 40 years ago, GenBank is perhaps the oldest public sequence database (Sayers et al. 2019). Although it offers a taxonomic framework which has recently been improved (Schoch et al. 2020), general databases such as GenBank struggle to keep pace with the increased flux of sequences and the evolving nature of microbial taxonomy. Because the presence of unassigned and/or incorrectly assigned sequences from phytoplankton groups (and microbes in general) can hinder the correct interpretation of metabarcod-

ing data, several efforts have been undertaken to provided taxonomically curated databases, e.g., for dinoflagellates (Mordret et al. 2018) or freshwater diatoms (Rimet et al. 2016, 2019). Living collections are priceless centralized storehouses of reference material, acting as "biological libraries". Large-scale sequencing projects focusing on extending the data from strains (including type strains) present in algal (microbial) culture collections, targeting different gene markers, will have a huge impact on the application of metabarcoding and our ability to understand and survey our biosphere.

The concept of functional diversity is generally centered around understanding communities and ecosystems from what function organisms can perform, rather than from their evolutionary history (Petchey and Gaston 2006). Several considerations should be addressed before applying functional diversity frameworks to the analysis of communities. For example, which functional information or traits about the organism are relevant and how different traits should be weighted according to their impact in the role of the organisms in the ecosystem. Given the complexity of plankton communities, both in terms of function and diversity, the understanding of the relationship between functional and taxonomic diversity remains a major challenge for plankton ecologists (Litchman and Klausmeier 2008). Few studies have attempted to correlate plankton molecular diversity with functional traits (Ramond et al. 2019; Schneider et al. 2020). Their results point to a promising field in plankton ecology. A global effort to create a consensus list of plankton traits combined to an open curated reference sequence database is in need to advance our understanding of crucial services provided to the biosphere by planktonic microbes.

Short amplicons produced by metabarcoding have a comparatively low phylogenetic signal. This can present a problem for identification, especially when an environmental sequence is distantly related to reference sequences. A new wave of sequencing technology is capable of long-read sequencing with high-throughput, minus the rigorous preparation required for Sanger sequencing. These technologies are known as single-molecule long-read sequencing or long-read amplicon sequencing. They might be confused with Illumina or 454, as they can also be referred to as high-throughput sequencing (HTS) or next generation sequencing (for the same reasons discussed above). Here, it will be called long-read amplicon sequencing. The two sequencing approaches that have been applied to microbes are from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (MinION), that can produce reads up to 20 Kb and 200 Kb respectively (Goodwin et al. 2016). With such technology, longer genetic markers can be targeted such as the ITS, large subunit (LSU) rRNA or the entire rRNA operon (which includes the SSU, ITS and LSU). These regions can span up to 4500 base pairs, which would greatly increase taxonomic resolution to uncover new taxonomic groups. Jamy et al. (2020) reported greater accuracy and sensitivity using a new "phylogeny-aware" approach, only possible with longer reads, compared to the similarity-based and phylogenetic placement-based methods for shorter reads. Longer reads also provide a new opportunity to directly investigate phylogenetic relationships, which would be especially beneficial for groups only known from environmental sequences and not cultures (Jamy et al. 2020). Long-read amplicon sequencing has only recently started to be applied to diversity studies of the marine microbial community (Fu et al. 2020; Thompson et al. 2020). There are still some limitations that need to be addressed. At moment we are writing this chapter, the error rate of long-read amplicon sequencing can be up to 12%, much higher than HTS at 0.1% (Goodwin et al. 2016). Moreover, new reference sequence databases containing longer gene fragments would need to be developed.

References cited

Ahlgren, N. and G. Rocap (2012). Diversity and Distribution of Marine *Synechococcus*: Multiple Gene Phylogenies for Consensus Classification and Development of qPCR Assays for Sensitive Measurement of Clades in the Ocean. *Frontiers in Microbiology* 3, 213.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215, 403–410.

Amacher, J., S. Neuer, I. Anderson, and R. Massana (2009). Molecular Approach to Determine Contributions of the Protist Community to Particle Flux. *Deep Sea Research Part I: Oceanographic Research Papers* 56, 2206–2215.

Amacher, J., S. Neuer, and M. Lomas (2013). DNA-Based Molecular Fingerprinting of Eukaryotic Protists and Cyanobacteria Contributing to Sinking Particle Flux at the Bermuda Atlantic Time-Series Study. *Deep Sea Research Part II: Topical Studies in Oceanography* 93, 71–83.

Amaral-Zettler, L. A., E. A. McCliment, H. W. Ducklow, and S. M. Huse (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE* 4, e6372.

Amato, A., W. H. Kooistra, J. H. Levialdi Ghiron, D. G. Mann, T. Pröschold, and M. Montresor (2007). Reproductive Isolation among Sympatric Cryptic Species in Marine Diatoms. *Protist* 158, 193–207.

Antia, A. N. (2005). Solubilization of Particles in Sediment Traps: Revising the Stoichiometry of Mixed Layer Export. *Biogeosciences* 2, 189–204.

Arrigo, K. R. et al. (2014). Phytoplankton Blooms beneath the Sea Ice in the Chukchi Sea. *Deep Sea Research Part II: Topical Studies in Oceanography* 105, 1–16.

Bach, L. T., P. Stange, J. Taucher, E. P. Achterberg, M. Algueró-Muñiz, H. Horn, M. Esposito, and U. Riebesell (2019). The Influence of Plankton Community Structure on Sinking Velocity and Remineralization Rate of Marine Aggregates. *Global Biogeochemical Cycles* 33, 971–994.

Banse, K. (2013). Reflections about Chance in My Career, and on the Top-down Regulated World. *Annual Review of Marine Science* 5, 1–19.

Bauerfeind, E., C. Garrity, M. Krumbholz, R. O. Ramseier, and M. Voss (1997). Seasonal Variability of Sediment Trap Collections in the Northeast Water Polynya. Part 2. Biochemical and Microscopic Composition of Sedimenting Matter. *Journal of Marine Systems* 10, 371–389.

Benner, I., A. J. Irwin, and Z. V. Finkel (2019). Capacity of the Common Arctic Picoeukaryote *Micromonas* to Adapt to a Warming Ocean. *Limnology and Oceanography Letters* 5, 221–227.

Beszteri, S., S. Thoms, V. Benes, L. Harms, and S. Trimborn (2018). The Response of Three Southern Ocean Phytoplankton Species to Ocean Acidification and Light Availability: A Transcriptomic Study. *Protist* 169, 958–975.

Biard, T., L. Stemmann, M. Picheral, N. Mayot, P. Vandromme, H. Hauss, G. Gorsky, L. Guidi, R. Kiko, and F. Not (2016). In Situ Imaging Reveals the Biomass of Giant Protists in the Global Ocean. *Nature* 532, 504–507.

Biers, E. J., S. Sun, and E. C. Howard (2009). Prokaryotic Genomes and Diversity in Surface Ocean Waters : Interrogating the Global Ocean Sampling Metagenome. *Applied and Environmental Microbiology* 75, 2221–2229.

Biller, S. J., P. M. Berube, D. Lindell, and S. W. Chisholm (2015). *Prochlorococcus*: The Structure and Function of Collective Diversity. *Nature Reviews Microbiology* 13, 13–27.

Bittner, L., A. Gobet, S. Audic, S. Romac, E. S. Egge, S. Santini, H. Ogata, I. Probert, B. Edvardsen, and C. de Vargas (2013). Diversity Patterns of Uncultured Haptophytes Unravelled by Pyrosequencing in Naples Bay. *Molecular Ecology* 22, 87–101.

Blanchot, J. and M. Rodier (1996). Picophytoplankton Abundance and Biomass in the Western Tropical Pacific Ocean during the 1992 El Niño Year: Results from Flow Cytometry. *Deep Sea Research Part I: Oceanographic Research Papers* 43, 877–895.

Blazewicz, S. J., R. L. Barnard, R. A. Daly, and M. K. Firestone (2013). Evaluating rRNA as an Indicator of Microbial Activity in Environmental Communities: Limitations and Uses. *The ISME Journal* 7, 2061–2068.

Boeuf, D., B. R. Edwards, J. M. Eppley, S. K. Hu, K. E. Poff, A. E. Romano, D. A. Caron, D. M. Karl, and E. F. DeLong (2019). Biological Composition and Microbial Dynamics of Sinking Particulate Organic Matter at Abyssal Depths in the Oligotrophic Open Ocean. *Proceedings of the National Academy of Sciences of the United States of America* 116, 11824–11832.

Bolyen, E. et al. (2019). Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2. *Nature Biotechnology* 37, 852–857.

Bouman, H. A., O. Ulloa, D. J. Scanlan, K. Zwirglmaier, W. K. Li, T. Platt, V. Stuart, R. Barlow, O. Leth, L. Clementson, V. Lutz, M. Fukasawa, S. Watanabe, and S. Sathyendranath (2006). Oceanographic Basis of the Global Surface Distribution of *Prochlorococcus* Ecotypes. *Science* 312, 918–921.

Bower, S. M., R. B. Carnegie, B. Goh, S. R. Jones, G. J. Lowe, and M. W. Mak (2004). Preferential PCR Amplification of Parasitic Protistan Small Subunit rDNA from Metazoan Tissues. *Journal of Eukaryotic Microbiology* 51, 325–332.

Boyd, P. W. and P. P. Newton (1999). Does Planktonic Community Structure Determine Downward Particulate Organic Carbon Flux in Different Oceanic Provinces? *Deep-Sea Research I* 46, 63–91.

Boyd, P. W., R. Strzepek, S. Takeda, G. Jackson, C. S. Wong, R. M. McKay, C. Law, H. Kiyosawa, H. Saito, N. Sherry, K. Johnson, J. Gower, and N. Ramaiah (2005). The Evolution and Termination of an Iron-Induced Mesoscale Bloom in the Northeast Subarctic Pacific. *Limnology and Oceanography* 50, 1872–1886.

Boyd, P. W. and T. W. Trull (2007). Understanding the Export of Biogenic Particles in Oceanic Waters: Is There Consensus? *Progress in Oceanography* 72, 276–312.

Boyd, P. W., H. Claustre, M. Levy, D. A. Siegel, and T. Weber (2019). Multi-Faceted Particle Pumps Drive Carbon Sequestration in the Ocean. *Nature* 568, 327–335.

Bucklin, A. and L. Allen (2004). MtDNA Sequencing from Zooplankton after Long-Term Preservation in Buffered Formalin. *Molecular Phylogenetics and Evolution* 30, 879–882.

Buesseler, K. O., C. H. Lamborg, P. W. Boyd, P. J. Lam, T. W. Trull, R. R. Bidigare, J. K. B. Bishop, K. L. Casciotti, F. Dehairs, M. Elskens, M. Honda, D. M. Karl, D. A. Siegel, M. W. Silver, D. K. Steinberg, J. Valdes, B. Van Mooy, and S. Wilson (2007). Revisiting Carbon Flux through the Ocean's Twilight Zone. *Science* 316, 567–570.

Burd, A. B., D. A. Hansell, D. K. Steinberg, T. R. Anderson, J. Aristegui, F. Baltar, S. R. Beaupre, K. O. Buesseler, F. DeHairs, G. A. Jackson, D. C. Kadko, R. Koppelmann, R. S. Lampitt, T. Nagata, T. Reinthaler, C. Robinson, B. H. Robison, C. Tamburini, and T. Tanaka (2010). Assessing the Apparent Imbalance between Geochemical and Biochemical Indicators of Meso- and Bathypelagic Biological Activity: What the @$#! Is Wrong with Present Calculations of Carbon Budgets? *Deep-Sea Research Part II-Topical Studies in Oceanography* 57, 1557–1571.

Cai, H., K. Wang, S. Huang, N. Jiao, and F. Chen (2010). Distinct Patterns of Picocyanobacterial Communities in Winter and Summer in the Chesapeake Bay. *Applied and Environmental Microbiology* 76, 2955–2960.

Calbet, A. and E. Saiz (2005). The Ciliate-Copepod Link in Marine Ecosystems. *Aquatic Microbial Ecology* 38, 157–167.

Callahan, B. J., P. J. McMurdie, and S. P. Holmes (2017). Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis. *The ISME Journal* 11, 2639–2643.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes (2016). DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nature Methods* 13, 581–583.

Campbell, L. and D. Vaulot (1993). Photosynthetic Picoplankton Community Structure in the Subtropical North Pacific Ocean near Hawaii (Station ALOHA). *Deep-Sea Research* 40, 2043–2060.

Caputi, L. et al. (2019). Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. *Global Biogeochemical Cycles* 33, 391–419.

Caron, D. and P. Countway (2009). Hypotheses on the Role of the Protistan Rare Biosphere in a Changing World. *Aquatic Microbial Ecology* 57, 227–238.

Caron, D. A., P. D. Countway, A. C. Jones, D. Y. Kim, and A. Schnetzer (2012). Marine Protistan Diversity. *Annual Review of Marine Science, Vol 4*. Vol. 4, 467–493.

Carroll, E. L., R. Gallego, M. A. Sewell, J. Zeldis, L. Ranjard, H. A. Ross, L. K. Tooman, R. O'Rorke, R. D. Newcomb, and R. Constantine (2019). Multi-Locus DNA Metabarcoding of Zooplankton Communities and Scat Reveal Trophic Interactions of a Generalist Predator. *Scientific Reports* 9, 281.

Casey, J. M., C. P. Meyer, F. Morat, S. J. Brandl, S. Planes, and V. Parravicini (2019). Reconstructing Hyperdiverse Food Webs: Gut Content Metabarcoding as a Tool to Disentangle Trophic Interactions on Coral Reefs. *Methods in Ecology and Evolution*, 2041–210X.13206.

Cavan, E. L., S. A. Henson, A. Belcher, and R. Sanders (2017). Role of Zooplankton in Determining the Efficiency of the Biological Carbon Pump. *Biogeosciences* 14, 177–186.

Chariton, A. A., S. Stephenson, M. J. Morgan, A. D. Steven, M. J. Colloff, L. N. Court, and C. M. Hardy (2015). Metabarcoding of Benthic Eukaryote Communities Predicts the Ecological Condition of Estuaries. *Environmental Pollution* 203, 165–174.

Charvet, S., W. F. Vincent, and C. Lovejoy (2014). Effects of Light and Prey Availability on Arctic Freshwater Protist Communities Examined by High-Throughput DNA and RNA Sequencing. *FEMS Microbiology Ecology* 88, 550–564.

Chen, F., K. Wang, J. Kan, M. T. Suzuki, and K. E. Wommack (2006). Diverse and Unique Picocyanobacteria in Chesapeake Bay, Revealed by 16S-23S rRNA Internal Transcribed Spacer Sequences. *Applied and Environmental Microbiology* 72, 2239–2243.

Chen, F., K. Wang, J. Kan, D. S. Bachoon, J. Lu, S. Lau, and L. Campbell (2004). Phylogenetic Diversity of *Synechococcus* in the Chesapeake Bay Revealed by Rilbulose-1,5-Bisphosphate Carboxylase-Oxygenase (RuBisCO) Large Subunit Gene (rbcL) Sequences. *Aquatic Microbial Ecology* 36, 153–164.

Chénard, C. and C. A. Suttle (2008). Phylogenetic Diversity of Sequences of Cyanophage Photosynthetic Gene *psbA* in Marine and Freshwaters. *Applied and Environmental Microbiology* 74, 5317–5324.

Choi, D. H. and J. H. Noh (2009). Phylogenetic Diversity of *Synechococcus* Strains Isolated from the East China Sea and the East Sea. *FEMS Microbiology Ecology* 69, 439–448.

Clare, E. L. (2014). Molecular Detection of Trophic Interactions: Emerging Trends, Distinct Advantages, Significant Considerations and Conservation Applications. *Evolutionary Applications* 7, 1144–1157.

Comeau, A. M., B. Philippe, M. Thaler, M. Gosselin, M. Poulin, and C. Lovejoy (2013). Protists in Arctic Drift and Land-Fast Sea Ice. *Journal of Phycology* 49, 229–240.

Cordier, T., D. Forster, Y. Dufresne, C. I. Martins, T. Stoeck, and J. Pawlowski (2018). Supervised Machine Learning Outperforms Taxonomy-Based Environmental DNA Metabarcoding Applied to Biomonitoring. *Molecular Ecology Resources* 18, 1381–1391.

Craig, C., W. J. Kimmerer, and C. S. Cohen (2013). A DNA-Based Method for Investigating Feeding by Copepod Nauplii. *Journal of Plankton Research* 36, 271–275.

Darling, K. F., C. M. Wade, I. A. Stewart, D. Kroon, R. Dingle, and A. J. L. Brown (2000). Molecular Evidence for Genetic Mixing of Arctic and Antarctic Subpolar Populations of Planktonic Foraminifers. *Nature* 405, 43–47.

Davey, M. P., L. Norman, P. Sterk, M. Huete-Ortega, F. Bunbury, B. K. W. Loh, S. Stockton, L. S. Peck, P. Convey, K. K. Newsham, and A. G. Smith (2019). Snow Algae Communities in Antarctica: Metabolic and Taxonomic Composition. *New Phytologist* 222, 1242–1255.

Dawson, S. C. and K. D. Hagen (2009). Mapping the Protistan 'Rare Biosphere'. *Journal of biology.*

De La Rocha, C. L. and U. Passow (2007). Factors Influencing the Sinking of POC and the Efficiency of the Biological Carbon Pump. *Deep Sea Research Part II: Topical Studies in Oceanography* 54, 639–658.

de Vargas, C. et al. (2015). Eukaryotic Plankton Diversity in the Sunlit Ocean. *Science* 348, 1261605.

Deagle, B. E., A. C. Thomas, J. C. McInnes, L. J. Clarke, E. J. Vesterinen, E. L. Clare, T. R. Kartzinel, and J. P. Eveson (2019). Counting with DNA in Metabarcoding Studies: How Should We Convert Sequence Reads to Dietary Data? *Molecular Ecology* 28, 391–406.

Decelle, J., P. Martin, K. Paborstava, D. W. Pond, G. Tarling, F. Mahé, C. de Vargas, R. Lampitt, and F. Not (2013). Diversity, Ecology and Biogeochemistry of Cyst-Forming Acantharia (Radiolaria) in the Oceans. *PLoS ONE* 8, e53598.

Decelle, J., S. Romac, R. F. Stern, E. M. Bendif, A. Zingone, S. Audic, M. D. Guiry, L. Guillou, D. Tessier, F. Le Gall, P. Gourvil, A. L. dos Santos, I. Probert, D. Vaulot, C. de Vargas, and R. Chris-

ten (2015). PhytoREF: A Reference Database of the Plastidial 16S rRNA Gene of Photosynthetic Eukaryotes with Curated Taxonomy. *Molecular Ecology Resources* 15, 1435–1445.

DeLong, E. F., C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N. U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, and D. M. Karl (2006). Community Genomics among Stratified Microbial Assemblages in the Ocean's Interior. *Science* 311, 496–503.

Díez, B., C. Pedrós-Alió, and R. Massana (2001). Study of Genetic Diversity of Eukaryotic Picoplankton in Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing. *Applied and Environmental Microbiology* 67, 2932–2941.

Doherty, M., P. L. Yager, M. A. Moran, V. J. Coles, C. S. Fortunato, A. V. Krusche, P. M. Medeiros, J. P. Payet, J. E. Richey, B. M. Satinsky, H. O. Sawakuchi, N. D. Ward, and B. C. Crump (2017). Bacterial Biogeography across the Amazon River-Ocean Continuum. *Frontiers in Microbiology*.

Dokulil, M. T. and C. Skolaut (1991). Aspects of Phytoplankton Seasonal Succession in Mondsee, Austria, with Particular Reference to the Ecology of *Dinobryon* Ehrenb. *Internationale Vereinigung für Theoretische und Angewandte Limnologie: Verhandlungen* 24, 968–973.

Doré, H. et al. (2020). Evolutionary Mechanisms of Long-Term Genome Diversification Associated with Niche Partitioning in Marine Picocyanobacteria. *Frontiers in Microbiology*.

dos Santos, A. L., P. Gourvil, M. Tragin, M.-H. Noel, J. Decelle, S. Romac, D. Vaulot, A. Lopes dos Santos, P. Gourvil, M. Tragin, M.-H. Noël, J. Decelle, S. Romac, and D. Vaulot (2017). Diversity and Oceanic Distribution of Prasinophytes Clade VII, the Dominant Group of Green Algae in Oceanic Waters. *The ISME Journal* 11, 512–528.

Duarte, C. M. (2015). Seafaring in the 21St Century: The Malaspina 2010 Circumnavigation Expedition. *Limnology and Oceanography Bulletin* 24, 11–14.

Dufresne, A., M. Ostrowski, D. J. Scanlan, L. Garczarek, S. Mazard, B. P. Palenik, I. T. Paulsen, N. T. de Marsac, P. Wincker, C. Dossat, S. Ferriera, J. Johnson, A. F. Post, W. R. Hess, and F. Partensky (2008). Unraveling the Genomic Mosaic of a Ubiquitous Genus of Marine Cyanobacteria. *Genome Biology* 9, R90.

Dunthorn, M., T. Stoeck, J. Clamp, A. Warren, and F. Mahé (2014). Ciliates and the Rare Biosphere: A Review. *Journal of Eukaryotic Microbiology* 61, 404–409.

Duret, M. T., R. S. Lampitt, and P. Lam (2019). Eukaryotic Influence on the Oceanic Biological Carbon Pump in the Scotia Sea as Revealed by 18S rRNA Gene Sequencing of Suspended and Sinking Particles. *Limnology and Oceanography* 65, lno.11319.

Durkin, C. A., B. A. S. Van Mooy, S. T. Dyhrman, and K. O. Buesseler (2016). Sinking Phytoplankton Associated with Carbon Flux in the Atlantic Ocean. *Limnology and Oceanography* 61, 1172–1187.

Ebersbach, F., P. Assmy, P. Martin, I. Schulz, S. Wolzenburg, and E.-M. Nöthig (2014). Particle Flux Characterisation and Sedimentation Patterns of Protistan Plankton during the Iron Fertilisation Experiment LOHAFEX in the Southern Ocean. *Deep Sea Research Part I: Oceanographic Research Papers* 89, 94–103.

Edgar, R. (2016). SINTAX: A Simple Non-Bayesian Taxonomy Classifier for 16S and ITS Sequences. *bioRxiv*, 74161.

Edgar, R. C. (2018). Updating the 97% Identity Threshold for 16S Ribosomal RNA OTUs. *Bioinformatics* 34, 2371–2375.

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight (2011). UCHIME Improves Sensitivity and Speed of Chimera Detection. *Bioinformatics* 27, 2194–2200.

Edgcomb, V. P., D. T. Kysela, A. Teske, A. d. V. Gomez, and M. L. Sogin (2002). Benthic Eukaryotic Diversity in the Guaymas Basin Hydrothermal Vent Environment. *Proceedings of the National Academy of Sciences of the United States of America* 99, 7658–7662.

Edvardsen, B., E. S. Egge, and D. Vaulot (2016). Diversity and Distribution of Haptophytes Revealed by Environmental Sequencing and Metabarcoding – a Review. *Perspectives in Phycology* 3, 77–91.

Egge, E. S., T. V. Johannessen, T. Andersen, W. Eikrem, L. Bittner, A. Larsen, R. A. Sandaa, and B. Edvardsen (2015). Seasonal Diversity and Dynamics of Haptophytes in the Skagerrak, Norway, Explored by High-Throughput Sequencing. *Molecular Ecology* 24, 3026–3042.

Eren, A. M. (2019). *Fantastic Data and How to Share Them: A Plea to Journal Editors and Reviewers (https://merenlab.org/2019/02/24/fantastic-data/)*.

Ernst, A., S. Becker, U. I. Wollenzien, and C. Postius (2003). Ecosystem-Dependent Adaptive Radiations of Picocyanobacteria Inferred from 16S rRNA and ITS-1 Sequence Analysis. *Microbiology* 149, 217–228.

Falkowski, P. G. et al. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* 281, 200–206.

Farrant, G. K., H. Doré, F. M. Cornejo-Castillo, F. Partensky, M. Ratin, M. Ostrowski, F. D. Pitt, P. Wincker, D. J. Scanlan, D. Iudicone, S. G. Acinas, and L. Garczarek (2016). Delineating Ecologically Significant Taxonomic Units from Global Patterns of Marine Picocyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* 113, E3365–E3374.

Fernández, L. D., C. E. Hernández, M. Romina Schiaffino, I. Izaguirre, and E. Lara (2017). Geographical Distance and Local Environmental Conditions Drive the Genetic Population Structure of a Freshwater Microalga (Bathycoccaceae; Chlorophyta) in Patagonian Lakes. *FEMS Microbiology Ecology* 93, fix125.

Ferris, M. J. and B. Palenik (1998). Niche Adaptation in Ocean Cyanobacteria. *Nature* 396, 226–228.

Fitch, W. M. and E. Margoliash (1967). Construction of Phylogenetic Trees. *Science* 155, 279–284.

Fitch, W. M. (1976). The Molecular Evolution of Cytochrome *c* in Eukaryotes. *Journal of Molecular Evolution* 8, 13–40.

Flombaum, P., J. L. Gallegos, R. A. Gordillo, J. Rincón, L. L. Zabala, N. Jiao, D. M. Karl, W. K. Li, M. W. Lomas, D. Veneziano, C. S. Vera, J. A. Vrugt, and A. C. Martiny (2013). Present and Future Global Distributions of the Marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America* 110, 9824–9829.

Fontanez, K. M., J. M. Eppley, T. J. Samo, D. M. Karl, and E. F. Delong (2015). Microbial Community Structure and Function on Sinking Particles in the North Pacific Subtropical Gyre. *Frontiers in Microbiology* 6, 1–14.

Fryxell, G. A., G. J. Doucette, and G. F. Hubbard (1981). The Genus *Thalassiosira*: The Bipolar Diatom *T. antarctica* Comber. *Botanica Marina* 24, 321–336.

Fu, Y., P. Zheng, X. Zhang, Q. Zhang, and D. Ji (2020). Protist Interactions and Seasonal Dynamics in the Coast of Yantai, Northern Yellow Sea of China as Revealed by Metabarcoding. *Journal of Ocean University of China* 19, 961–974.

Fuller, N. J., C. Campbell, D. J. Allen, F. D. Pitt, F. Le Gall, D. Vaulot, and D. J. Scanlan (2006). Analysis of Photosynthetic Picoeukaryote Diversity at Open Ocean Sites in the Arabian Sea Using a PCR Biased towards Marine Algal Plastids. *Aquatic Microbial Ecology* 43, 79–93.

Fuller, N. J., D. Marie, F. Partensky, D. Vaulot, A. F. Post, and D. J. Scanlan (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. *Applied and Environmental Microbiology* 69, 2430–2443.

Fuller, N. J., N. J. West, D. Marie, M. Yallop, T. Rivlin, A. F. Post, and D. J. Scanlan (2005). Dynamics of Community Structure and Phosphate Status of Picocyanobacterial Populations in the Gulf of Aqaba, Red Sea. *Limnology and Oceanography* 50, 363–375.

Galand, P. E., E. O. Casamayor, D. L. Kirchman, and C. Lovejoy (2009). Ecology of the Rare Microbial Biosphere of the Arctic Ocean. *Proceedings of the National Academy of Sciences of the United States of America* 106, 22427–22432.

García-García, N., J. Tamames, A. M. Linz, C. Pedrós-Alió, and F. Puente-Sánchez (2019). Microdiversity Ensures the Maintenance of Functional Microbial Communities under Changing Environmental Conditions. *The ISME Journal* 13, 2969–2983.

Garczarek, L., A. Dufresne, N. Blot, A. M. Cockshutt, A. Peyrat, D. A. Campbell, L. Joubin, and C. Six (2008). Function and Evolution of the *psbA* Gene Family in Marine *Synechococcus*: *Synechococcus* Sp. WH7803 as a Case Study. *The ISME Journal* 2, 937–953.

Gérikas Ribeiro, C., A. Lopes dos Santos, D. Marie, F. Pereira Brandini, and D. Vaulot (2018). Small Eukaryotic Phytoplankton Communities in Tropical Waters off Brazil Are Dominated by Symbioses between Haptophyta and Nitrogen-Fixing Cyanobacteria. *The ISME Journal* 12, 1360–1374.

Gilbert M P, T., T. Haselkorn, M. Bunce, J. J. Sanchez, S. B. Lucas, L. D. Jewell, E. Van Marck, and M. Worobey (2007). The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues–Which Methods Are Useful When? *PLoS ONE* 2, e537.

Giner, C. R., M. C. Pernice, V. Balagué, C. M. Duarte, J. M. Gasol, R. Logares, and R. Massana (2020). Marked Changes in Diversity and Relative Activity of Picoeukaryotes with Depth in the World Ocean. *The ISME Journal* 14, 437–449.

Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field (1990). Genetic Diversity in Sargasso Sea Bacterioplankton. *Nature* 345, 60–63.

Gong, W. and A. Marchetti (2019). Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. *Frontiers in Marine Science* 6, 219.

Goodwin, S., J. D. McPherson, and W. Richard McCombie (2016). Coming of Age: Ten Years of next-Generation Sequencing Technologies. *Nature Reviews Genetics* 17, 333–351.

Gowing, M. M. (1993). Seasonal Radiolarian Flux at the VERTEX North Pacific Time-Series Site. *Deep Sea Research Part I: Oceanographic Research Papers* 40, 517–545.

Gowing, M. M. and S. L. Coale (1989). Fluxes of Living Radiolarians and Their Skeletons along a Northeast Pacific Transect from Coastal Upwelling to Open Ocean Waters. *Deep Sea Research Part A. Oceanographic Research Papers* 36, 561–576.

Gran-Stadniczeñko, S., L. Šupraha, E. D. Egge, and B. Edvardsen (2017). Haptophyte Diversity and Vertical Distribution Explored by 18S and 28S Ribosomal RNA Gene Metabarcoding and Scanning Electron Microscopy. *Journal of Eukaryotic Microbiology* 64, 514–532.

Guillou, L. et al. (2013). The Protist Ribosomal Reference Database (PR^2): A Catalog of Unicellular Eukaryote Small Sub-Unit rRNA Sequences with Curated Taxonomy. *Nucleic Acids Research* 41, D597–D604.

Gutiérrez-Rodríguez, A., G. Slack, E. F. Daniels, K. E. Selph, B. Palenik, and M. R. Landry (2014). Fine Spatial Structure of Genetically Distinct Picocyanobacterial Populations across Environmental Gradients in the Costa Rica Dome. *Limnology and Oceanography* 59, 705–723.

Gutierrez-Rodriguez, A., M. R. Stukel, A. dos Santos, T. Biard, R. Scharek, D. Vaulot, M. R. Landry, and F. Not (2019). High Contribution of Rhizaria (Radiolaria) to Vertical Export in the California Current Ecosystem Revealed by DNA Metabarcoding. *The ISME journal* 13, 964–976.

Hanagata, N., N. Malinsky-Rushansky, and Z. Dubinsky (1999). Eukaryotic Picoplankton, *Mychonastes Homosphaera* (Chlorophyceae, Chlorophyta), in Lake Kinneret, Israel. *Phycological Research*.

Hansen, L. R., J. Kristiansen, and J. V. Rasmussen (1994). Potential Toxicity of the Freshwater *Chrysochromulina* Species *C. parva* (Prymnesiophyceae). *Hydrobiologia* 287, 157–159.

Hardge, K., I. Peeken, S. Neuhaus, B. A. Lange, A. Stock, T. Stoeck, L. Weinisch, and K. Metfies (2017). The Importance of Sea Ice for Exchange of Habitat-Specific Protist Communities in the Central Arctic Ocean. *Journal of Marine Systems* 165, 124–138.

Hardy, C. M., E. S. Krull, D. M. Hartley, and R. L. Oliver (2010). Carbon Source Accounting for Fish Using Combined DNA and Stable Isotope Analyses in a Regulated Lowland River Weir Pool. *Molecular Ecology* 19, 197–212.

Haumann, F. A., N. Gruber, and M. Münnich (2020). Sea-Ice Induced Southern Ocean Subsurface Warming and Surface Cooling in a Warming Climate. *AGU Advances* 1, e2019AV000132.

Henson, S., F. Le Moigne, and S. Giering (2019). Drivers of Carbon Export Efficiency in the Global Ocean. *Global Biogeochemical Cycles* 33, 891–903.

Herdman, M., R. W. Castenholz, J. B. Waterbury, and R. Rippka (2001). Form-Genus XIII. *Synechococcus. Bergey's Manual of Systematic Bacteriology*. 2d Ed. Vol. 1. New York, 508–512.

Herndl, G. J. and T. Reinthaler (2013). Microbial Control of the Dark End of the Biological Pump. *Nature geoscience* 6, 718–724.

Hess, W. R., A. Weihe, S. Loiseaux-de Goër, F. Partensky, and D. Vaulot (1995). Characterization of the Single *psbA* Gene of *Prochlorococcus marinus* CCMP 1375 (Prochlorophyta). *Plant Molecular Biology* 27, 1189–1196.

Honjo, S. (1976). Coccoliths: Production, Transportation and Sedimentation. *Marine Micropaleontology* 1, 65–79.

Honjo, S., T. Eglinton, C. Taylor, K. Ulmer, S. Sievert, A. Bracher, C. German, V. Edgcomb, R. Francois, M. D. Iglesias-Rodriguez, B. Van Mooy, and D. Rapeta (2014). Understanding the Role of the Biological Pump in the Global Carbon Cycle: An Imperative for Ocean Science. *Oceanography* 27, 10–16.

Hoppe, C. J. M., C. M. Flintrop, and B. Rost (2018). The Arctic Picoeukaryote *Micromonas pusilla* Benefits Synergistically from Warming and Ocean Acidification. *Biogeosciences* 15, 4353–4365.

Horvat, C., D. R. Jones, S. Iams, D. Schroeder, D. Flocco, and D. Feltham (2017). The Frequency and Extent of Sub-Ice Phytoplankton Blooms in the Arctic Ocean. *Science Advances* 3, e1601191.

Hu, S. K., V. Campbell, P. Connell, A. G. Gellene, Z. Liu, R. Terrado, and D. A. Caron (2016). Protistan Diversity and Activity Inferred from RNA and DNA at a Coastal Ocean Site in the Eastern North Pacific. *FEMS Microbiology Ecology* 92, fiw050.

Huang, S., S. W. Wilhelm, H. R. Harvey, K. Taylor, N. Jiao, and F. Chen (2012). Novel Lineages of *Prochlorococcus* and *Synechococcus* in the Global Oceans. *The ISME Journal* 6, 285–97.

Huber, P., F. M. Cornejo-Castillo, I. Ferrera, P. Sánchez, R. Logares, S. Metz, V. Balagué, S. G. Acinas, J. M. Gasol, and F. Unrein (2019). Primer Design for an Accurate View of Picocyanobacterial Community Structure by Using High-Throughput Sequencing. *Applied and Environmental Microbiology*.

Huse, S. M., D. B. Mark Welch, A. Voorhis, A. Shipunova, H. G. Morrison, A. M. Eren, and M. L. Sogin (2014). VAMPS: A Website for Visualization and Analysis of Microbial Population Structures. *BMC Bioinformatics* 15, 41.

Ibarbalz, F. M. et al. (2019). Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* 179, 1084–1097.e21.

Isaacs, J. D. (1972). Unstructured Marine Food Webs and "Pollutant Analogues". *U S Natl Mar Fish Serv Fish Bull* 70, 1053–1059.

Jain, R., M. C. Rivera, and J. A. Lake (1999). Horizontal Gene Transfer among Genomes: The Complexity Hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 96, 3801–3806.

Jamy, M., R. Foster, P. Barbera, L. Czech, A. Kozlov, A. Stamatakis, G. Bending, S. Hilton, D. Bass, and F. Burki (2020). Long-read Metabarcoding of the Eukaryotic rDNA Operon to Phylogenetically and Taxonomically Resolve Environmental Diversity. *Molecular Ecology Resources* 20, 429–443.

Janout, M. A., J. Hölemann, A. M. Waite, T. Krumpen, W. J. von Appen, and F. Martynov (2016). Sea-Ice Retreat Controls Timing of Summer Plankton Blooms in the Eastern Arctic Ocean. *Geophysical Research Letters* 43, 12, 493–12, 501.

Jenkins, B. D., J. P. Zehr, A. Gibson, and L. Campbell (2006). Cyanobacterial Assimilatory Nitrate Reductase Gene Diversity in Coastal and Oligotrophic Marine Environments. *Environmental Microbiology* 8, 2083–2095.

Jin, X., N. Gruber, J. P. Dunne, J. L. Sarmiento, and R. A. Armstrong (2006). Diagnosing the Contribution of Phytoplankton Functional Groups to the Production and Export of Particulate Organic Carbon, $CaCO_3$, and Opal from Global Nutrient and Alkalinity Distributions. *Global Biogeochemical Cycles* 20.

Joli, N., A. Monier, R. Logares, and C. Lovejoy (2017). Seasonal Patterns in Arctic Prasinophytes and Inferred Ecology of *Bathycoccus* Unveiled in an Arctic Winter Metagenome. *The ISME Journal* 11, 1372–1385.

Jurburg, S. D., M. Konzack, N. Eisenhauer, and A. Heintz-Buschart (2020). The Archives Are Half-Empty: An Assessment of the Availability of Microbial Community Sequencing Data. *Communications Biology* 3, 474.

Kaczmarska, I., L. Mather, I. Aluddington, F. Muise, and J. M. Ehrman (2014). Cryptic Diversity in a Cosmopolitan Diatom Known as *Asterionellopsis glacialis* (Fragilariaceae): Implications for Ecology, Biogeography, and Taxonomy. *American Journal of Botany* 101, 267–286.

Kalenitchenko, D., N. Joli, M. Potvin, J. éric Tremblay, and C. Lovejoy (2019). Biodiversity and Species Change in the Arctic Ocean: A View through the Lens of Nares Strait. *Frontiers in Marine Science* 6, 1–17.

Karsenti et al. (2010). A Holistic Approach to Marine Eco-Systems Biology. *PLoS biology* 9, e1001177.

Kashtan, N., S. E. Roggensack, S. Rodrigue, J. W. Thompson, S. J. Biller, A. Coe, H. Ding, P. Marttinen, R. R. Malmstrom, R. Stocker, M. J. Follows, R. Stepanauskas, and S. W. Chisholm (2014). Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild *Prochlorococcus*. *Science* 344, 416–20.

Kaunisto, K. M., T. Roslin, I. E. Sääksjärvi, and E. J. Vesterinen (2017). Pellets of Proof: First Glimpse of the Dietary Composition of Adult Odonates as Revealed by Metabarcoding of Feces. *Ecology and Evolution* 7, 8588–8598.

Kermarrec, L., A. Franc, F. Rimet, P. Chaumeil, J.-M. Frigerio, J.-F. Humbert, and A. Bouchez (2014). A Next-Generation Sequencing Approach to River Biomonitoring Using Benthic Diatoms. *Freshwater Science* 33, 349–363.

Kettler, G. C., A. C. Martiny, K. Huang, J. Zucker, M. L. Coleman, S. Rodrigue, F. Chen, A. Lapidus, S. Ferriera, J. Johnson, C. Steglich, G. M. Church, P. Richardson, and S. W. Chisholm (2007). Patterns and Implications of Gene Gain and Loss in the Evolution of *Prochlorococcus*. *PLoS Genetics* 3, e231.

Kilias, E. S., I. Peeken, and K. Metfies (2014). Insight into Protist Diversity in Arctic Sea Ice and Melt-Pond Aggregate Obtained by Pyrosequencing. *Polar Research* 33, 23466.

King, R. A., D. S. Read, M. Traugott, and W. O. C. Symondson (2008). Molecular Analysis of Predation: A Review of Best Practice for DNA-based Approaches. *Molecular ecology* 17, 947–963.

Kirchman, D. L. (2016). Growth Rates of Microbes in the Oceans. *Annual Review of Marine Science* 8, 285–309.

Kishino, H., T. Miyata, and M. Hasegawa (1990). Maximum Likelihood Inference of Protein Phylogeny and the Origin of Chloroplasts. *Journal of Molecular Evolution* 31, 151–160.

Kitahara, K. and K. Miyazaki (2013). Revisiting Bacterial Phylogeny. *Mobile Genetic Elements* 3, e24210.

Knauer, G. A., D. M. Karl, J. H. Martin, and C. N. Hunter (1984). In Situ Effects of Selected Preservatives on Total Carbon, Nitrogen and Metals Collected in Sediment Traps. *Journal of Marine Research* 42, 445–462.

Knauer, G. A., J. H. Martin, and K. W. Bruland (1979). Fluxes of Particulate Carbon, Nitrogen, and Phosphorus in the Upper Water Column of the Northeast Pacific. *Deep-Sea Research* 26, 97–108.

Kopf, A. et al. (2015). The Ocean Sampling Day Consortium. *GigaScience* 4, 27.

Kozich, J. J., S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the Miseq Illumina Sequencing Platform. *Applied and Environmental Microbiology* 79, 5112–5120.

Kress, W. J., C. Garcia-Robledo, M. Uriarte, and D. L. Erickson (2015). DNA Barcodes for Ecology, Evolution, and Conservation. *Trends in Ecology & Evolution* 30, 25–35.

Kuwata, A., K. Yamada, M. Ichinomiya, S. Yoshikawa, M. Tragin, D. Vaulot, and A. Lopes dos Santos (2018). Bolidophyceae, a Sister Picoplanktonic Group of Diatoms – a Review. *Frontiers in Marine Science* 5, 370.

Kysela, D. T., C. Palacios, and M. L. Sogin (2005). Serial Analysis of V6 Ribosomal Sequence Tags (SARST-V6): A Method for Efficient, High-Throughput Analysis of Microbial Community Composition. *Environmental Microbiology* 7, 356–364.

Lampitt, R. S., I. Salter, and D. Johns (2009). Radiolaria: Major Exporters of Organic Carbon to the Deep Ocean. *Global Biogeochemical Cycles* 23, 1–9.

Lampitt, R. S., K. F. Wishner, C. M. Turley, and M. V. Angel (1993). Marine Snow Studies in the Northeast Atlantic-Ocean - Distribution, Composition and Role as a Food Source for Migrating Plankton. *Marine Biology* 116, 689–702.

Landry, M. R. (2002). Integrating Classical and Microbial Food Web Concepts: Evolving Views from the Open-Ocean Tropical Pacific. *Hydrobiologia* 480, 29–39.

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace (1985). Rapid Determination of 16S Ribosomal RNA Sequences for Phylogenetic Analyses. *Proceedings of the National Academy of Sciences of the United States of America* 82, 6955–6959.

Lanzén, A., S. L. Jørgensen, D. H. Huson, M. Gorfer, S. H. Grindhaug, I. Jonassen, L. Øvreås, and T. Urich (2012). CREST - Classification Resources for Environmental Sequence Tags. *PLoS ONE* 7.

Lara, E., L. D. Fernández, M. R. Schiaffino, and I. Izaguirre (2017). First Freshwater Member Ever Reported for the Family Bathycoccaceae (Chlorophyta; Archaeplastida) from Argentinean Patagonia Revealed by Environmental DNA Survey. *European Journal of Protistology*.

Laroche, J., G. van der Staay, A. Ducret, R. Aebersold, R. Li, S. Golden, R. Hiller, P. Wrench, A. Larkum, and B. Green (1996). Independent Evolution of the Prochlorophyte and Green Plant Chlorophyll a/b Light-Harvesting Proteins. *Proceedings of the National Academy of Sciences of the United States of America* 93, 15244–15248.

Latasa, M. (2007). Improving Estimations of Phytoplankton Class Abundances Using CHEMTAX. *Marine Ecology Progress Series* 329, 13–21.

Le Moigne, F. A. C., K. Pabortsava, C. L. J. Marcinko, P. Martin, and R. J. Sanders (2014). Where Is Mineral Ballast Important for Surface Export of Particulate Organic Carbon in the Ocean? *Geophysical Research Letters* 41, 8460–8468.

LeCleir, G. R., J. M. DeBruyn, E. W. Maas, P. W. Boyd, and S. W. Wilhelm (2014). Temporal Changes in Particle-Associated Microbial Communities after Interception by Nonlethal Sediment Traps. *FEMS Microbiology Ecology* 87, 153–163.

Lee, C. (2019). Passing the Baton to the next Generation: A Few Problems That Need Solving. *Annual Review of Marine Science* 11, 1–13.

Lehman, J. T. (1976). Ecological and Nutritional Studies on *Dinobryon* Ehrenb.: Seasonal Periodicity and the Phosphate Toxicity Problem. *Limnology and Oceanography* 21, 646–658.

Leliaert, F., H. Verbruggen, P. Vanormelingen, F. Steen, J. M. López-Bautista, G. C. Zuccarello, and O. D. Clerck (2014). DNA-Based Species Delimitation in Algae DNA. *European Journal of Phycology* 49, 179–196.

Leray, M., C. P. Meyer, and S. C. Mills (2015). Metabarcoding Dietary Analysis of Coral Dwelling Predatory Fish Demonstrates the Minor Contribution of Coral Mutualists to Their Highly Partitioned, Generalist Diet. *PeerJ* 3, e1047.

Leu, E., J. E. Søreide, D. O. Hessen, S. Falk-Petersen, and J. Berge (2011). Consequences of Changing Sea-Ice Cover for Primary and Secondary Producers in the European Arctic Shelf Seas: Timing, Quantity, and Quality. *Progress in Oceanography* 90, 18–32.

Li, S. N., G. Bronner, C. Lépere, F. X. Kong, and X. L. Shi (2017). Temporal and Spatial Variations in the Composition of Freshwater Photosynthetic Picoeukaryotes Revealed by MiSeq Sequencing from Flow Cytometry Sorted Samples. *Environmental Microbiology* 19, 2286–2300.

Li, W. K. W., D. V. S. Rao, W. G. Harrison, J. C. Smith, J. J. Cullen, B. Irwin, and T. Platt (1983). Autotrophic Picoplankton in the Tropical Ocean. *Science* 219, 292–295.

Lin, S. (2011). Genomic Understanding of Dinoflagellates. *Research in microbiology* 162, 551–569.

Litchman, E. and C. A. Klausmeier (2008). Trait-Based Community Ecology of Phytoplankton. *Annual Review of Ecology, Evolution, and Systematics* 39, 615–639.

Liu, C., X. Shi, F. Wu, M. Zhang, G. Gao, and Q. Wu (2020). Temporal Patterns in the Interaction between Photosynthetic Picoeukaryotes and Their Attached Fungi in Lake Chaohu. *FEMS Microbiology Ecology* 96, fiaa123.

Liu, Z., J. K. Cochran, C. Lee, B. Gasser, J. C. Miquel, and S. G. Wakeham (2009). Further Investigations on Why POC Concentrations Differ in Samples Collected by Niskin Bottle and in Situ Pump. *Deep Sea Research Part II: Topical Studies in Oceanography* 56, 1558–1567.

Lockhart, P. and D. Penny (1992). The Problem of GC Content, Evolutionary Trees and the Origins of Chl-a/b Photosynthetic Organelles: Are the Prochlorophytes a Eubacterial Model for Higher Plant Photosynthesis. *Research in Photosynthesis Vol III*, 499–505.

Logares, R., I. M. Deutschmann, P. C. Junger, C. R. Giner, A. K. Krabberød, T. S. B. Schmidt, L. Rubinat-Ripoll, M. Mestre, G. Salazar, C. Ruiz-González, M. Sebastián, C. de Vargas, S. G. Acinas, C. M. Duarte, J. M. Gasol, and R. Massana (2020). Disentangling the Mechanisms Shaping the Surface Ocean Microbiota. *Microbiome* 8, 55.

Logares, R., J.-F. Mangot, and R. Massana (2015). Rarity in Aquatic Microbes: Placing Protists on the Map. *Research in Microbiology* 166, 831–841.

Logares, R. et al. (2014). Patterns of Rare and Abundant Marine Microbial Eukaryotes. *Current Biology* 24, 813–821.

Lopes dos Santos, A., T. Pollina, P. Gourvil, E. Corre, D. Marie, J. L. Garrido, F. Rodríguez, M.-H. Noël, D. Vaulot, and W. Eikrem (2017). Chloropicophyceae, a New Class of Picophytoplanktonic Prasinophytes. *Scientific Reports* 7, 14019.

López-García, P., F. Rodriguez-Valera, P.-A. C., and D. Moreira (2001). Unexpected Diversity of Small Eukaryotes in Deep-Sea Antarctic Plankton. *Nature* 409, 603–607.

Love, M. I., W. Huber, and S. Anders (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biology* 15, 550.

Lovejoy, C., R. Massana, and C. Pedros-Alio (2006). Diversity and Distribution of Marine Microbial Eukaryotes in the Arctic Ocean and Adjacent Seas. *Applied and Environmental Microbiology* 72, 3085–3095.

Luo, C., D. Tsementzi, N. Kyrpides, T. Read, and K. T. Konstantinidis (2012). Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample. *PLoS ONE* 7, e30087.

Lutz, S., J. McCutcheon, J. B. McQuaid, and L. G. Benning (2018). The Diversity of Ice Algal Communities on the Greenland Ice Sheet as Revealed by Oligotyping. *Microbial Genomics* 4, 1–10.

Ma, Y., N. Z. Jiao, and Y. H. Zeng (2004). Natural Community Structure of Cyanobacteria in the South China Sea as Revealed by *rpoC1* Gene Sequence Analysis. *Letters in Applied Microbiology* 39, 353–358.

Mackey, M. D., H. W. Higgins, D. J. Mackey, and S. W. Wright (1997). *CHEMTAX User's Manual: A Program for Estimating Class Abundances from Chemical Markers - Application to HPLC Measurements of Phytoplankton Pigments.* Tech. rep. Hobart, Tas.: CSIRO Marine Laboratories, 41.

Mahé, F., T. Rognes, C. Quince, C. de Vargas, and M. Dunthorn (2014). Swarm: Robust and Fast Clustering Method for Amplicon-Based Studies. *PeerJ* 2, e593.

Majdi, N. et al. (2018). There's No Harm in Having Too Much: A Comprehensive Toolbox of Methods in Trophic Ecology. *Food Webs* 17, e00100.

Maloy, A. P., S. C. Culloty, and J. W. Slater (2013). Dietary Analysis of Small Planktonic Consumers: A Case Study with Marine Bivalve Larvae. *Journal of Plankton Research* 35, 866–876.

Malviya, S., E. Scalco, S. Audic, F. Vincent, A. Veluchamy, J. Poulain, P. Wincker, D. Iudicone, C. de Vargas, L. Bittner, A. Zingone, and C. Bowler (2016). Insights into Global Diatom Distribution and Diversity in the World's Ocean. *Proceedings of the National Academy of Sciences of the United States of America* 113, E1516–E1525.

Man-Aharonovich, D., A. Philosof, B. C. Kirkup, F. Le Gall, T. Yogev, I. Berman-Frank, M. F. Polz, D. Vaulot, and O. Béjà (2010). Diversity of Active Marine Picoeukaryotes in the Eastern Mediterranean Sea Unveiled Using Photosystem-II *psbA* Transcripts. *The ISME journal* 4, 1044–1052.

Mann, N. H., A. Cook, A. Millard, S. Bailey, and M. Clokie (2003). Marine Ecosystems: Bacterial Photosynthesis Genes in a Virus. *Nature* 424, 741.

Margulis, L., G. Hinkle, H. McKhann, and B. Moynihan (1988). *Mychonastes desiccatus* Brown sp. nova (Chlorococcales, Chlorophyta)–an Intertidal Alga Forming Achlorophyllous Desiccation-Resistant Cysts. *Archiv fur Hydrobiologie Supplementband Algological studies* 78, 425–446.

Marie, D., C. P. D. Brussaard, F. Partensky, and D. Vaulot (1999). Flow Cytometric Analysis of Phytoplankton, Bacteria and Viruses. *Current Protocols in Cytometry.* Vol. 11.11. New York, 1–15.

Marie, D., X. L. Shi, F. Rigaut-Jalabert, and D. Vaulot (2010). Use of Flow Cytometric Sorting to Better Assess the Diversity of Small Photosynthetic Eukaryotes in the English Channel. *FEMS Microbiology Ecology* 72, 165–178.

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal* 17, 10.

Martin, P., J. T. Allen, M. J. Cooper, D. G. Johns, R. S. Lampitt, R. Sanders, and D. A. H. Teagle (2010). Sedimentation of Acantharian Cysts in the Iceland Basin: Strontium as a Ballast for Deep Ocean Particle Flux, and Implications for Acantharian Reproductive Strategies. *Limnology and Oceanography* 55, 604–614.

Martiny, A. C., A. P. Tai, D. Veneziano, F. Primeau, and S. W. Chisholm (2009). Taxonomic Resolution, Ecotypes and the Biogeography of *Prochlorococcus*. *Environmental Microbiology* 11, 823–832.

Massana, R. et al. (2015). Marine Protist Diversity in European Coastal Waters and Sediments as Revealed by High-Throughput Sequencing. *Environmental Microbiology* 17, 4035–4049.

Matsen, F. A., R. B. Kodner, and E. V. Armbrust (2010). Pplacer: Linear Time Maximum-Likelihood and Bayesian Phylogenetic Placement of Sequences onto a Fixed Reference Tree. *Bmc Bioinformatics* 11, 538.

Mazard, S., M. Ostrowski, F. Partensky, and D. J. Scanlan (2012). Multi-Locus Sequence Analysis, Taxonomic Resolution and Biogeography of Marine *Synechococcus*. *Environmental Microbiology* 14, 372–386.

McDonnell, A. M., P. W. Boyd, and K. O. Buesseler (2015). Effects of Sinking Velocities and Microbial Respiration Rates on the Attenuation of Particulate Carbon Fluxes through the Mesopelagic Zone. *Global Biogeochemical Cycles* 29, 175–193.

McMurdie, P. J. and S. Holmes (2013). Phyloseq: An r Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* 8, e61217.

Medlin, L., H. J. Elwood, S. Stickel, and M. L. Sogin (1988). The Characterization of Enzymatically Amplified Eukaryotic 16S-like rRNA-Coding Regions. *Gene* 71, 491–499.

Mella-Flores, D., C. Six, M. Ratin, F. Partensky, C. Boutte, G. Le Corguillé, D. Marie, N. Blot, P. Gourvil, C. Kolowrat, and L. Garczarek (2012). *Prochlorococcus* and *Synechococcus* Have Evolved Different Adaptive Mechanisms to Cope with Light and Uv Stress. *Frontiers in Microbiology* 3, 285.

Metfies, K., E. Bauerfeind, C. Wolf, P. Sprong, S. Frickenhaus, L. Kaleschke, A. Nicolaus, and E.-M. Nöthig (2017). Protist Communities in Moored Long-Term Sediment Traps (Fram Strait, Arctic)–Preservation with Mercury Chloride Allows for PCR-Based Molecular Genetic Analyses. *Frontiers in Marine Science* 4, 1–13.

Meyerhans, A., J. P. Vartanian, and S. Wain-Hobson (1990). DNA Recombination during PCR. *Nucleic Acids Research* 18, 1687–1691.

Michaels, A. F., D. A. Caron, N. R. Swanberg, F. A. Howse, and C. M. Michaels (1995). Planktonic Sarcodines (Acantaria, Radiolaria, Foraminifera) in Surface Waters near Bermuda: Abundance, Biomass and Vertical Flux. *Journal of Plankton Research* 17, 131–163.

Michaels, A. F. and M. W. Silver (1988). Primary Production, Sinking Fluxes and the Microbial Food Web. *Deep-Sea Research Part a-Oceanographic Research Papers* 35, 473–490.

Millard, A., M. R. Clokie, D. A. Shub, and N. H. Mann (2004). Genetic Organization of the *psbAD* Region in Phages Infecting Marine *Synechococcus* Strains. *Proceedings of the National Academy of Sciences of the United States of America* 101, 11007–11012.

Mitra, A. et al. (2016). Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition; Incorporation of Diverse Mixotrophic Strategies. *Protist* 167, 106–120.

Montresor, M., C. Lovejoy, L. Orsini, G. Procaccini, and S. Roy (2003). Bipolar Distribution of the Cyst-Forming Dinoflagellate Polarella Glacialis. *Polar Biology* 26, 186–194.

Moon-van der Staay, S. Y., R. De Wachter, and D. Vaulot (2001). Oceanic 18S rDNA Sequences from Picoplankton Reveal Unsuspected Eukaryotic Diversity. *Nature* 409, 607–610.

Moore, L. R., G. Rocap, and S. W. Chisholm (1998). Physiology and Molecular Phytogeny of Coexisting *Prochlorococcus* Ecotypes. *Nature* 393, 464–467.

Mordret, S., R. Piredda, D. Vaulot, M. Montresor, W. H. C. F. Kooistra, and D. Sarno (2018). Dinoref: A Curated Dinoflagellate (Dinophyceae) Reference Database for the 18S rRNA Gene. *Molecular Ecology Resources* 18, 974–987.

Muhling, M., N. J. Fuller, A. Millard, P. J. Somerfield, D. Marie, W. H. Wilson, D. J. Scanlan, A. F. Post, I. Joint, and N. H. Mann (2005). Genetic Diversity of Marine *Synechococcus* and Co-Occurring Cyanophage Communities: Evidence for Viral Control of Phytoplankton. *Environmental Microbiology* 7, 499–508.

Mühling, M., N. J. Fuller, P. J. Somerfield, A. F. Post, W. H. Wilson, D. J. Scanlan, I. Joint, and N. H. Mann (2006). High Resolution Genetic Diversity Studies of Marine *Synechococcus* Isolates Using *rpoC1*-Based Restriction Fragment Length Polymorphism. *Aquatic Microbial Ecology* 45, 263–275.

Nakamura, Y., A. Tuji, W. Makino, S.-i. S. Matsuzaki, N. Nagata, M. Nakagawa, and TakamuraNoriko (2020). Feeding Ecology of a Mysid Species, Neomysis Awatschensis in the Lake Kasumigaura: Combining Approach with Microscopy, Stable Isotope Analysis and DNA Metabarcoding. *Plankton and Benthos Research* 15, 44–54.

Needham, D. M., R. Sachdeva, and J. A. Fuhrman (2017). Ecological Dynamics and Co-Occurrence among Marine Phytoplankton, Bacteria and Myoviruses Shows Microdiversity Matters. *The ISME Journal* 11, 1614–1629.

Needham, D. M. and J. A. Fuhrman (2016). Pronounced Daily Succession of Phytoplankton, Archaea and Bacteria Following a Spring Bloom. *Nature Microbiology* 1, 16005.

Nelson, D. M., P. Tréguer, M. A. Brzezinski, A. Leynaert, and B. Quéguiner (1995). Production and Dissolution of Biogenic Silica in the Ocean: Revised Global Estimates, Comparison with Regional Data and Relationship to Biogenic Sedimentation. *Global Biogeochemical Cycles* 9, 359–372.

Niemi, A., C. Michel, K. Hille, and M. Poulin (2011). Protist Assemblages in Winter Sea Ice: Setting the Stage for the Spring Ice Algal Bloom. *Polar Biology* 34, 1803–1817.

Nodder, S. D. (1997). Short-Term Sediment Trap Fluxes from Chatham Rise, Southwest Pacific Ocean. *Limnology and Oceanography* 42, 777–783.

Not, F., K. Valentin, K. Romari, C. Lovejoy, R. Massana, K. Töbe, D. Vaulot, and L. K. Medlin (2007a). Picobiliphytes, a New Marine Picoplanktonic Algal Group with Unknown Affinities to Other Eukaryotes. *Science* 315, 252–254.

Not, F., R. Gausling, F. Azam, J. F. Heidelberg, and A. Z. Worden (2007b). Vertical Distribution of Picoeukaryotic Diversity in the Sargasso Sea. *Environmental Microbiology* 9, 1233–1252.

O'Rorke, R., S. D. Lavery, M. Wang, S. D. Nodder, and A. G. Jeffs (2013). Determining the Diet of Larvae of the Red Rock Lobster (*Jasus edwardsii*) Using High-Throughput DNA Sequencing Techniques. *Marine Biology* 161, 551–563.

Oksanen, A. J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. Mcglinn, P. R. Minchin, R. B. O. Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and E. Szoecs (2016). *Package ' Vegan '(Version 2.4-0)*.

Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl (1986). Microbial Ecology and Evolution: A Ribosomal RNA Approach. *Annual Review of Microbiology* 40, 337–365.

Olsen, L. M., S. R. Laney, P. Duarte, H. M. Kauko, M. Fernández-méndez, C. J. Mundy, A. Rösel, A. Meyer, P. Itkin, L. Cohen, I. Peeken, A. Tatarek, M. Różańska-Pluta, J. Wiktor, T. Taskjelle, A. K. Pavlov, S. R. Hudson, M. A. Granskog, H. Hop, and P. Assmy (2017). The Seeding of Ice Algal Blooms in Arctic Pack Ice: The Multiyear Ice Seed Repository Hypothesis. *Journal of Geophysical Research: Biogeosciences* 122, 1529–1548.

Olson, R. J., D. Vaulot, and S. W. Chisholm (1985). Marine Phytoplankton Distributions Measured Using Shipboard Flow Cytometry. *Deep Sea Research Part A, Oceanographic Research Papers* 32, 1273–1280.

Orr, J. C. et al. (2005). Anthropogenic Ocean Acidification over the Twenty-First Century and Its Impact on Calcifying Organisms. *Nature* 437, 681–686.

Paerl, R. W., R. A. Foster, B. D. Jenkins, J. P. Montoya, and J. P. Zehr (2008). Phylogenetic Diversity of Cyanobacterial *narB* Genes from Various Marine Habitats. *Environmental Microbiology* 10, 3377–3387.

Paerl, R. W., K. S. Johnson, R. M. Welsh, A. Z. Worden, F. P. Chavez, and J. P. Zehr (2011). Differential Distributions of *Synechococcus* Subgroups across the California Current System. *Frontiers in Microbiology* 2, 1–22.

Paerl, R. W., K. A. Turk, R. A. Beinart, F. P. Chavez, and J. P. Zehr (2012). Seasonal Change in the Abundance of *Synechococcus* and Multiple Distinct Phylotypes in Monterey Bay Determined by *rbcL* and *narB* Quantitative PCR. *Environmental Microbiology* 14, 580–593.

Palenik, B. (1994). Cyanobacterial Community Structure as Seen from RNA Polymerase Gene Sequence Analysis. *Applied and Environmental Microbiology* 60, 3212–3219.

Partensky, F., J. Blanchot, and D. Vaulot (1999). Differential Distribution and Ecology of *Prochlorococcus* and *Synechococcus* in Oceanic Waters: A Review. *Marine Cyanobacteria*. Monaco.

Pawłowska, J., J. E. Wollenburg, M. Zajaczkowski, and J. Pawlowski (2020). Planktonic Foraminifera Genomic Variations Reflect Paleoceanographic Changes in the Arctic: Evidence from Sedimentary Ancient DNA. *Scientific Reports* 10, 15102.

Pawlowski, J., P. Esling, F. Lejzerowicz, T. Cedhagen, and T. A. Wilding (2014). Environmental Monitoring through Protist Next-Generation Sequencing Metabarcoding: Assessing the Impact of Fish Farming on Benthic Foraminifera Communities. *Molecular Ecology Resources* 14, 1129–1140.

Pawlowski, J., P. Esling, F. Lejzerowicz, T. Cordier, J. Visco, C. Martins, A. Kvalvik, K. Staven, and T. Cedhagen (2016). Benthic Monitoring of Salmon Farms in Norway Using Foraminiferal Metabarcoding. *Aquaculture Environment Interactions* 8, 371–386.

Pedrós-Alió, C., S. G. Acinas, R. Logares, and R. Massana (2018). Marine Microbial Diversity as Seen by High-Throughput Sequencing. *Microbial Ecology of the Oceans*, 47–97.

Penno, S., D. Lindell, and A. F. Post (2006). Diversity of *Synechococcus* and *Prochlorococcus* Populations Determined from DNA Sequences of the N-Regulatory Gene *ntcA*. *Environmental Microbiology* 8, 1200–1211.

Pérez-Cobas, A. E., L. Gomez-Valero, and C. Buchrieser (2020). Metagenomic Approaches in Microbial Ecology: An Update on Whole-Genome and Marker Gene Sequencing Analyses. *Microbial Genomics*, mgen000409.

Perovich, D. K. and J. A. Richter-Menge (2009). Loss of Sea Ice in the Arctic. *Annual Review of Marine Science* 1, 417–441.

Petchey, O. L. and K. J. Gaston (2006). Functional Diversity: Back to Basics and Looking Forward. *Ecology Letters* 9, 741–758.

Peterson, M. L., S. G. Wakeham, C. Lee, M. A. Askea, and J. C. Miquel (2005). Novel Techniques for Collection of Sinking Particles in the Ocean and Determining Their Settling Rates. *Limnology and Oceanography: Methods* 3, 520–532.

Pierella Karlusich, J. J., F. M. Ibarbalz, and C. Bowler (2020). Phytoplankton in the Tara Ocean. *Annual Review of Marine Science* 12, 233–265.

Piredda, R., M. P. Tomasino, A. M. D'Erchia, C. Manzari, G. Pesole, M. Montresor, W. H. C. F. Kooistra, D. Sarno, and A. Zingone (2017). Diversity and Temporal Patterns of Planktonic Protist Assemblages at a Mediterranean Long Term Ecological Research Site. *FEMS Microbiology Ecology* 93, fiw200.

Porazinska, D. L., R. M. Giblin-Davis, W. Sung, and W. K. Thomas (2012). The Nature and Frequency of Chimeras in Eukaryotic Metagenetic Samples. *Journal of Nematology* 44, 18–25.

Post, A. F., S. Penno, K. Zandbank, A. Paytan, S. M. Huse, and D. M. Welch (2011). Long Term Seasonal Dynamics of *Synechococcus* Population Structure in the Gulf of Aqaba, Northern Red Sea. *Frontiers in Microbiology* 2, 131.

Preston, C. M., C. A. Durkin, and K. M. Yamahara (2019). DNA Metabarcoding Reveals Organisms Contributing to Particulate Matter Flux to Abyssal Depths in the North East Pacific Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography* 173, 104708.

Puigcorbé, V., C. R. Benitez-Nelson, P. Masqué, E. Verdeny, A. E. White, B. N. Popp, F. G. Prahl, and P. J. Lam (2015). Small Phytoplankton Drive High Summertime Carbon and Nutrient Export in the Gulf of California and Eastern Tropical North Pacific. *Global Biogeochemical Cycles* 29, 1309–1332.

Puigcorbé, V., C. Ruiz-González, P. Masqué, and J. M. Gasol (2020). Sampling Device-Dependence of Prokaryotic Community Structure on Marine Particles: Higher Diversity Recovered by *in situ* Pumps than by Oceanographic Bottles. *Frontiers in Microbiology*.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner (2013). The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Research* 41, D590.

Ramond, P., M. Sourisseau, N. Simon, S. Romac, S. Schmitt, F. Rigaut-Jalabert, N. Henry, C. de Vargas, and R. Siano (2019). Coupling between Taxonomic and Functional Diversity in Protistan Coastal Communities. *Environmental Microbiology*.

Ray, J. L., K. S. Skaar, P. Simonelli, A. Larsen, A. Sazhin, H. H. Jakobsen, J. C. Nejstgaard, and C. Troedsson (2016). Molecular Gut Content Analysis Demonstrates That Calanus Grazing on *Phaeocystis pouchetii* and *Phaeocystis pouchetii* Is Sensitive to Bloom Phase but Not Prey Density. *Marine Ecology Progress Series* 542, 63–77.

Resplandy, L., M. Lévy, and D. J. McGillicuddy Jr. (2019). Effects of Eddy-Driven Subduction on Ocean Biological Carbon Pump. *Global Biogeochemical Cycles* 33, 1071–1084.

Richardson, T. L. (2019). Mechanisms and Pathways of Small-Phytoplankton Export from the Surface Ocean. *Annual Review of Marine Science* 11, 57–74.

Richardson, T. L. and G. a Jackson (2007). Small Phytoplankton and Carbon Export from the Surface Ocean. *Science* 315, 838–40.

Rigual Hernández, A. S., T. W. Trull, S. D. Nodder, J. A. Flores, H. Bostock, F. Abrantes, R. S. Eriksen, F. J. Sierro, D. M. Davies, A.-M. Ballegeer, M. A. Fuertes, and L. C. Northcote (2020). Coccolithophore Biodiversity Controls Carbonate Export in the Southern Ocean. *Biogeosciences* 17, 245–263.

Rimet, F., P. Chaumeil, F. Keck, L. Kermarrec, V. Vasselon, M. Kahlert, A. Franc, and A. Bouchez (2016). R-Syst::Diatom: An Open-Access and Curated Barcode Database for Diatoms and Freshwater Monitoring. *Database-the Magazine of Electronic Database Reviews* 2016, baw016.

Rimet, F., E. Gusev, M. Kahlert, M. G. Kelly, M. Kulikovskiy, Y. Maltsev, D. G. Mann, M. Pfannkuchen, R. Trobajo, V. Vasselon, J. Zimmermann, and A. Bouchez (2019). Diat.Barcode, an Open-Access Curated Barcode Library for Diatoms. *Scientific Reports* 9, 15116.

Rocap, G., D. L. Distel, J. B. Waterbury, and S. W. Chisholm (2002). Resolution of *Prochlorococcus* and *Synechococcus* Ecotypes by Using 16S-23S Ribosomal DNA Internal Transcribed Spacer Sequences. *Applied and Environmental Microbiology* 68, 1180–1191.

Rodríguez-Martínez, R., G. Rocap, G. Salazar, and R. Massana (2013). Biogeography of the Uncultured Marine Picoeukaryote MAST-4: Temperature-Driven Distribution Patterns. *The ISME Journal* 7, 1531–1543.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé (2016). VSEARCH: A Versatile Open Source Tool for Metagenomics. *PeerJ* 4, e2584.

Rozema, P. D., H. J. Venables, W. H. van de Poll, A. Clarke, M. P. Meredith, and A. G. J. Buma (2017). Interannual Variability in Phytoplankton Biomass and Species Composition in Northern Marguerite Bay (West Antarctic Peninsula) Is Governed by Both Winter Sea Ice Cover and Summer Stratification. *Limnology and Oceanography* 62, 235–252.

Ruane, S. and C. C. Austin (2017). Phylogenomics Using Formalin-Fixed and 100+ Year-Old Intractable Natural History Specimens. *Molecular Ecology Resources* 17, 1003–1008.

Rusch, D. B. et al. (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS biology* 5, e77.

Rynearson, T. A., K. Richardson, R. S. Lampitt, M. E. Sieracki, A. J. Poulton, M. M. Lyngsgaard, and M. J. Perry (2013). Major Contribution of Diatom Resting Spores to Vertical Flux in the Sub-Polar North Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers* 82, 60–71.

Rynearson, T. A., S. A. Flickinger, and D. N. Fontaine (2020). Metabarcoding Reveals Temporal Patterns of Community Composition and Realized Thermal Niches of *Thalassiosira* spp. (Bacillariophyceae) from the Narragansett Bay Long-Term Plankton Time Series. *Biology* 9, 19.

Samanta, B. and P. Bhadury (2016). A Comprehensive Framework for Functional Diversity Patterns of Marine Chromophytic Phytoplankton Using *rbcL* Phylogeny. *Scientific Reports* 6, 20783.

Sanger, F., S. Nicklen, and A. R. Coulson (1977). DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74, 5463–5467.

Sanger, F., G. G. Brownlee, and B. G. Barrell (1965). A Two-Dimensional Fractionation Procedure for Radioactive Nucleotides. *Journal of Molecular Biology* 13, 373–398.

Santoferrara, L. F. (2019). Current Practice in Plankton Metabarcoding: Optimization and Error Management. *Journal of Plankton Research* 41, 571–582.

Santoferrara, L., F. Burki, S. Filker, R. Logares, M. Dunthorn, and G. B. McManus (2020). Perspectives from Ten Years of Protist Studies by High-Throughput Metabarcoding. *Journal of Eukaryotic Microbiology* 67, 612–622.

Sapp, J. and G. E. Fox (2013). The Singular Quest for a Universal Tree of Life. *Microbiology and Molecular Biology Reviews* 77, 541–550.

Sarthou, G., K. R. Timmermans, S. Blain, and P. Tréguer (2005). Growth Physiology and Fate of Diatoms in the Ocean: A Review. *Journal of Sea Research* 53, 25–42.

Sautter, L. R. and R. C. Thunell (1989). Seasonal Succession of Planktonic Foraminifera; Results from a Four-Year Time-Series Sediment Trap Experiment in the Northeast Pacific. *The Journal of Foraminiferal Research* 19, 253–267.

Sayers, E. W., M. Cavanaugh, K. Clark, J. Ostell, K. D. Pruitt, and I. Karsch-Mizrachi (2019). GenBank. *Nucleic Acids Research* 48, D84–D86.

Scanlan, D. J., M. Ostrowski, S. Mazard, A. Dufresne, L. Garczarek, W. R. Hess, A. F. Post, M. Hagemann, I. Paulsen, and F. Partensky (2009). Ecological Genomics of Marine Picocyanobacteria. *Microbiology and Molecular Biology Reviews* 73, 249–299.

Scanlan, D. J. and N. J. West (2002). Molecular Ecology of the Marine Cyanobacterial Genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiology Ecology* 40, 1–12.

Schander, C. and H. M. Kenneth (2003). DNA, PCR and Formalinized Animal Tissue – a Short Review and Protocols. *Organisms Diversity & Evolution* 3, 195–205.

Scharek, R., L. M. Tupas, and D. M. Karl (1999). Diatom Fluxes to the Deep Sea in the Oligotrophic North Pacific Gyre at Station ALOHA. *Marine Ecology Progress Series* 182, 55–67.

Schiebel, R. (2002). Planktic Foraminiferal Sedimentation and the Marine Calcite Budget. *Global Biogeochemical Cycles* 16, 3–1–3–21.

Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber (2009). Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* 75, 7537–7541.

Schloss, P. D. and J. Handelsman (2005). Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Applied and Environmental Microbiology* 71, 1501–1506.

Schneider, L., K. Anestis, J. Mansour, A. Anschütz, N. Gypens, P. Hansen, U. John, K. Klemm, J. Martin, N. Medic, F. Not, and W. Stolte (2020). A Dataset on Trophic Modes of Aquatic Protists. *Biodiversity Data Journal* 8.

Schoch, C. L., S. Ciufo, M. Domrachev, C. L. Hotton, S. Kannan, R. Khovanskaya, D. Leipe, R. Mcveigh, K. O'Neill, B. Robbertse, S. Sharma, V. Soussov, J. P. Sullivan, L. Sun, S. Turner, and I. Karsch-Mizrachi (2020). NCBI Taxonomy: A Comprehensive Update on Curation, Resources and Tools. *Database-the Magazine of Electronic Database Reviews* 2020.

Schwarz, D., S. M. Spitzer, A. C. Thomas, C. M. Kohnert, T. R. Keates, and A. Acevedo-Gutierrez (2018). Large-Scale Molecular Diet Analysis in a Generalist Marine Mammal Reveals Male Preference for Prey of Conservation Concern. *Ecology and Evolution* 8, 9889–9905.

Segawa, T., R. Matsuzaki, N. Takeuchi, A. Akiyoshi, F. Navarro, S. Sugiyama, T. Yonezawa, and H. Mori (2018). Bipolar Dispersal of Red-Snow Algae. *Nature Communications* 9.

Serreze, M. C., M. M. Holland, and J. Stroeve (2007). Perspectives on the Arctic's Shrinking Sea-Ice Cover. *Science* 315, 1533–1536.

Shadwick, E. H., T. W. Trull, H. Thomas, and J. A. E. Gibson (2013). Vulnerability of Polar Oceans to Anthropogenic Acidification: Comparison of Arctic and Antarctic Seasonal Cycles. *Scientific Reports* 3, 2339.

Shi, X. L., S. N. Li, M. Zhang, C. Q. Liu, and Q. L. Wu (2020). Temperature Mainly Determines the Temporal Succession of the Photosynthetic Picoeukaryote Community in Lake Chaohu, a Highly Eutrophic Shallow Lake. *Science of the Total Environment* 702, 134803.

Shi, X. L., D. Marie, L. Jardillier, D. J. Scanlan, and D. Vaulot (2009). Groups without Cultured Representatives Dominate Eukaryotic Picophytoplankton in the Oligotrophic South East Pacific Ocean. *PLoS One* 4, e7657.

Shi, X., S. Li, F. Fan, M. Zhang, Z. Yang, and Y. Yang (2018a). *Mychonastes* Dominates the Photosynthetic Picoeukaryotes in Lake Poyang, a River-Connected Lake. *FEMS Microbiology Ecology* 95, fiy211.

Shi, X., S. Li, H. Li, F. Chen, and Q. L. Wu (2019). The Community Structure of Picophytoplankton in Lake Fuxian, a Deep and Oligotrophic Mountain Lake. *Frontiers in Microbiology* 10, 2016.

Shi, X., S. Li, C. Liu, M. Zhang, and M. Liu (2018b). Community Structure of Photosynthetic Picoeukaryotes Differs in Lakes with Different Trophic Statuses along the Middle-Lower Reaches of the Yangtze River. *FEMS Microbiology Ecology* 94, fiy011.

Shiozaki, T., F. Itoh, Y. Hirose, J. Onodera, A. Kuwata, and N. Harada (2021). A DNA Metabarcoding Approach for Recovering Plankton Communities from Archived Samples Fixed in Formalin. *PLoS ONE* 16, e0245936.

Sieburth, J. M., V. Smetacek, and J. Lenz (1978). Pelagic Ecosystem Structure: Heterotrophic Compartments of the Plankton and Their Relationship to Plankton Size Fractions 1. *Limnology and Oceanography* 23, 1256–1263.

Siegel, D. A., K. O. Buesseler, S. C. Doney, S. F. Sailley, M. J. Behrenfeld, and P. W. Boyd (2014). Global Assessment of Ocean Carbon Export by Combining Satellite Observations and Food-Web Models. *Global Biogeochemical Cycles* 28, 181–196.

Silver, M. W. and M. M. Gowing (1991). The Particle-Flux - Origins and Biological Components. *Progress in Oceanography* 26, 75–113.

Simpson, P. D. and S. D. Van Valkenburg (1978). The Ultrastructure of *Mychonastes ruminatus* gen. & sp. nov., a New Member of the Chlorophyceae Isolated from Brackish Water. *British Phycological Journal*.

Six, C., J.-C. Thomas, L. Garczarek, M. Ostrowski, A. Dufresne, N. Blot, D. J. Scanlan, and F. Partensky (2007). Diversity and Evolution of Phycobilisomes in Marine *Synechococcus* Spp.: A Comparative Genomics Study. *Genome biology* 8, R259.

Smetacek, V. et al. (2012). Deep Carbon Export from a Southern Ocean Iron-Fertilized Diatom Bloom. *Nature* 487, 313–319.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl (2006). Microbial Diversity in the Deep Sea and the Underexplored "Rare Biosphere". *Proceedings of the National Academy of Sciences of the United States of America* 103, 12115–12120.

Sogin, M. L. and J. H. Gunderson (1987). Structural Diversity of Eukaryotic Small Subunit Ribosomal RNAs. *Annals of the New York Academy of Sciences* 503, 125–139.

Soto, D. F., R. Fuentes, P. Huovinen, and I. Gómez (2020). Microbial Composition and Photosynthesis in Antarctic Snow Algae Communities: Integrating Metabarcoding and Pulse Amplitude Modulation Fluorometry. *Algal Research* 45, 101738.

Sow, L. S. S., T. W. Trull, and L. Bodrossy (2020). Oceanographic Fronts Shape *Phaeocystis* Assemblages: A High-Resolution 18S rRNA Gene Survey from the Ice-Edge to the Equator of the South Pacific. *Frontiers in Microbiology* 11, 1847.

Stackebrandt, E. and B. M. Goebel (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology* 44, 846–849.

Stahl, D. A., D. J. Lane, G. J. Olsen, and N. R. Pace (1985). Characterization of a Yellowstone Hot Spring Microbial Community by 5S rRNA Sequences. *Applied and Environmental Microbiology* 49, 1379–1384.

Steglich, C., A. F. Post, and W. R. Hess (2003). Analysis of Natural Populations of *Prochlorococcus* spp. in the Northern Red Sea Using Phycoerythrin Gene Sequences. *Environmental Microbiology* 5, 681–690.

Steinberg, D. K. and M. R. Landry (2017). Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science* 9, 413–444.

Stemmann, L. and E. Boss (2011). Plankton and Particle Size and Packaging: From Determining Optical Properties to Driving the Biological Pump. *Annual Review of Marine Science* 4, 263–290.

Stoeck, T., D. Bass, M. Nebel, R. Christen, M. D. M. Jones, H. W. Breiner, and T. A. Richards (2010). Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water. *Molecular Ecology* 19, 21–31.

Stoeck, T., A. Behnke, R. Christen, L. Amaral-Zettler, M. J. Rodriguez-Mora, A. Chistoserdov, W. Orsi, and V. P. Edgcomb (2009). Massively Parallel Tag Sequencing Reveals the Complexity of Anaerobic Marine Protistan Communities. *BMC Biology* 7, 72.

Stoeck, T., G. T. Taylor, and S. S. Epstein (2003). Novel Eukaryotes from the Permanently Anoxic Cariaco Basin (Caribbean Sea). *Applied and Environmental Microbiology* 69, 5656–5663.

Stone, J. P. and D. K. Steinberg (2016). Salp Contributions to Vertical Carbon Flux in the Sargasso Sea. *Deep-Sea Research Part I: Oceanographic Research Papers* 113, 90–100.

Stroeve, J. C., T. Markus, L. Boisvert, J. Miller, and A. Barrett (2014). Changes in Arctic Melt Season and Implications for Sea Ice Loss. *Geophysical Research Letters* 41, 1216–1225.

Stukel, M. R., L. I. Aluwihare, K. A. Barbeau, A. M. Chekalyuk, R. Goericke, A. J. Miller, M. D. Ohman, A. Ruacho, H. Song, B. M. Stephens, and M. R. Landry (2017). Mesoscale Ocean Fronts Enhance Carbon Export Due to Gravitational Sinking and Subduction. *Proceedings of the National Academy of Sciences of the United States of America* 114, 1252–1257.

Stukel, M. R., M. Décima, K. E. Selph, D. A. A. Taniguchi, and M. R. Landry (2013). The Role of *Synechococcus* in Vertical Flux in the Costa Rica Upwelling Dome. *Progress in Oceanography* 112-113, 49–59.

Stukel, M. R., M. R. Landry, C. R. Benitez-Nelson, and R. Goericke (2011). Trophic Cycling and Carbon Export Relationships in the California Current Ecosystem. *Limnology and Oceanography* 56, 1866–1878.

Sul, W. J., T. A. Oliver, H. W. Ducklow, L. A. Amaral-Zettler, and M. L. Sogin (2013). Marine Bacteria Exhibit a Bipolar Distribution. *Proceedings of the National Academy of Sciences of the United States of America* 110, 2342–2347.

Sunagawa, S. et al. (2015). Structure and Function of the Global Ocean Microbiome. *Science* 348, 1261359.

Tai, V. and B. Palenik (2009). Temporal Variation of *Synechococcus* Clades at a Coastal Pacific Ocean Monitoring Site. *The ISME Journal*.

Thompson, J. R., R. G. Hatfield, F. M. Batista, T. P. Bean, and V. G. Fonseca (2020). The Application of Nanopore Sequencing Technology to the Study of Dinoflagellates : A Proof of Concept Study for Rapid Sequence-Based Discrimination of Potentially Harmful Algae. 11, 1–16.

Thompson, R. M., U. Brose, J. a Dunne, R. O. Hall, S. Hladyz, R. L. Kitching, N. D. Martinez, H. Rantala, T. N. Romanuk, D. B. Stouffer, and J. M. Tylianakis (2012). Food Webs: Reconciling the Structure and Function of Biodiversity. *Trends in Ecology & Evolution* 27, 689–697.

Toledo, G. and B. Palenik (1997). *Synechococcus* Diversity in the California Current as Seen by RNA Polymerase (*rpoC1*) Gene Sequences of Isolated Strains. *Applied and Environmental Microbiology* 63, 4298–4303.

Tragin, M. and D. Vaulot (2018). Green Microalgae in Marine Coastal Waters: The Ocean Sampling Day (OSD) Dataset. *Scientific Reports* 8, 14020.

— (2019). Novel Diversity within Marine Mamiellophyceae (Chlorophyta) Unveiled by Metabarcoding. *Scientific Reports* 9, 5190.

Tragin, M., A. Zingone, and D. Vaulot (2018). Comparison of Coastal Phytoplankton Composition Estimated from the V4 and V9 Regions of 18S rRNA Gene with a Focus on Photosynthetic Groups and Especially Chlorophyta. *Environmental Microbiology* 20, 1–41.

Trefault, N., R. De la Iglesia, M. Moreno-Pino, A. Lopes dos Santos, C. Gérikas Ribeiro, G. Parada-Pozo, A. Cristi, D. Marie, and D. Vaulot (2021). Annual Phytoplankton Dynamics in Coastal Waters from Fildes Bay, Western Antarctic Peninsula. *Scientific Reports* 11, 1368.

Tréguer, P., C. Bowler, B. Moriceau, S. Dutkiewicz, M. Gehlen, O. Aumont, L. Bittner, R. Dugdale, Z. Finkel, D. Iudicone, O. Jahn, L. Guidi, M. Lasbleiz, K. Leblanc, M. Levy, and P. Pondaven (2018). Influence of Diatom Diversity on the Ocean Biological Carbon Pump. *Nature Geoscience* 11, 27–37.

Turner, J. T. (2015). Zooplankton Fecal Pellets, Marine Snow, Phytodetritus and the Ocean's Biological Pump. *Progress in Oceanography* 130, 205–248.

Turner, J., S. R. Colwell, G. J. Marshall, T. A. Lachlan-Cope, A. M. Carleton, P. D. Jones, V. Lagun, P. A. Reid, and S. Iagovkina (2005). Antarctic Climate Change during the Last 50 Years. *International Journal of Climatology* 25, 279–294.

Twining, B. S., S. Rauschenberg, P. L. Morton, D. C. Ohnemus, and P. J. Lam (2015). Comparison of Particulate Trace Element Concentrations in the North Atlantic Ocean as Determined with Discrete Bottle Sampling and in Situ Pumping. *Deep Sea Research Part II: Topical Studies in Oceanography* 116, 273–282.

Ulloa, O., C. Henríquez-Castillo, S. Ramírez-Flandes, A. M. Plominsky, A. A. Murillo, C. Morgan-Lang, S. J. Hallam, and R. Stepanauskas (2021). The Cyanobacterium *Prochlorococcus* Has Divergent Light-Harvesting Antennae and May Have Evolved in a Low-Oxygen Ocean. *Proceedings of the National Academy of Sciences* 118.

Urbach, E., D. L. Robertson, and S. W. Chisholm (1992). Multiple Evolutionary Origins of Prochlorophytes within the Cyanobacterial Radiation. *Nature* 355, 267–270.

Urbach, E., D. J. Scanlan, D. L. Distel, J. B. Waterbury, and S. W. Chisholm (1998). Rapid Diversification of Marine Picophytoplankton with Dissimilar Light- Harvesting Structures Inferred from Sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *Journal of Molecular Evolution* 46, 188–201.

van Leeuwe, M. A., L. Tedesco, K. R. Arrigo, P. Assmy, K. Campbell, K. M. Meiners, J.-M. Rintala, V. Selz, D. N. Thomas, J. Stefels, J. W. Deming, M. A. V. Leeuwe, L. Tedesco, K. R. Arrigo, P. Assmy, K. M. Meiners, J.-M. Rintala, V. Selz, N. David, and J. Stefels (2018). Microalgal Community Structure and Primary Production in Arctic and Antarctic Sea Ice: A Synthesis. *Elementa: Science of the Anthropocene Sci Anth* 6, 1–25.

Vaulot, D., W. Eikrem, M. Viprey, and H. Moreau (2008). The Diversity of Small Eukaryotic Phytoplankton (< 3 Mm) in Marine Ecosystems. *FEMS Microbiology Reviews* 32, 795–820.

Vaulot, D., S. Geisen, F. Mahé, and D. Bass (2021). Pr2-Primers: An 18S rRNA Primer Database for Protists. *bioRxiv*, 2021.01.04.425170.

Venter, J. C. et al. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* 304, 66–74.

Vernette, C., N. Henry, J. Lecubin, C. de Vargas, P. Hingamp, and M. Lescot (2021). The Ocean Barcode Atlas: A Web Service to Explore the Biodiversity and Biogeography of Marine Organisms. *Molecular Ecology Resources* in press.

Vestheim, H. and S. N. Jarman (2008). Blocking Primers to Enhance PCR Amplification of Rare Sequences in Mixed Samples - a Case Study on Prey DNA in Antarctic Krill Stomachs. *Frontiers in Zoology* 5, 12.

Vick-Majors, T. J., J. C. Priscu, and L. A. Amaral-Zettler (2014). Modular Community Structure Suggests Metabolic Plasticity during the Transition to Polar Night in Ice-Covered Antarctic Lakes. *The ISME Journal* 8, 778–789.

Visco, J. A., L. Apothéloz-Perret-Gentil, A. Cordonier, P. Esling, L. Pillet, and J. Pawlowski (2015). Environmental Monitoring: Inferring the Diatom Index from next-Generation Sequencing Data. *Environmental Science & Technology* 49, 7597–7605.

Volk, T. and M. I. Hoffert (1985). Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-driven Atmospheric CO_2 Changes. *The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present*. Vol. 32. Geophysical Monograph Series, 99–110.

Waite, A. M., K. A. Safi, J. A. Hall, and S. D. Nodder (2000). Mass Sedimentation of Picoplankton Embedded in Organic Aggregates. *Limnology and Oceanography*.

Waite, A., P. K. Bienfang, and P. J. Harrison (1992). Spring Bloom Sedimentation in a Subarctic Ecosystem. *Marine Biology* 114, 131–138.

Wang, G. C. Y. and Y. Wang (1997). Frequency of Formation of Chimeric Molecules Is a Consequence of PCR Coamplification of 16S rRNA Genes from Mixed Bacterial Genomes. *Applied and Environmental Microbiology* 63, 4645–4650.

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole (2007). Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* 73, 5261–5267.

West, N. J., W. A. Schonhuber, N. J. Fuller, R. I. Amann, R. Rippka, A. F. Post, and D. J. Scanlan (2001). Closely Related Prochlorococcus Genotypes Show Remarkably Different Depth Distributions in Two Oceanic Regions as Revealed by in Situ Hybridization Using 16S rRNA-Targeted Oligonucleotides. *Microbiology Sgm* 147, 1731–1744.

West, N. J. and D. J. Scanlan (1999). Niche-Partitioning of *Prochlorococcus* Populations in a Stratified Water Column in the Eastern North Atlantic Ocean? *Applied and Environmental Microbiology*.

Whitaker, M. R. L., C. C. M. Baker, S. M. Salzman, D. J. Martins, and N. E. Pierce (2019). Combining Stable Isotope Analysis with DNA Metabarcoding Improves Inferences of Trophic Ecology. *PLOS ONE* 14, e0219070.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3, 1–9.

Woese, C. R. (2009). A New Biology for the 21st Century. *Frontiers in Ecology and the Environment* 7, 455.

Woese, C. R. and G. E. Fox (1977). Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* 74, 5088–5090.

Wolf, C., E. Kilias, and K. Metfies (2015). Protists in the Polar Regions: Comparing Occurrence in the Arctic and Southern Oceans Using Pyrosequencing. *Polar Research* 34, 23225.

Worden, A. Z., M. J. Follows, S. J. Giovannoni, S. Wilken, A. E. Zimmerman, and P. J. Keeling (2015). Rethinking the Marine Carbon Cycle: Factoring in the Multifarious Lifestyles of Microbes. *Science* 347, 1257594.

Wright Erik, S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in r. *The R Journal* 8, 352.

Xia, X., S. Cheung, H. Endo, K. Suzuki, and H. Liu (2019). Latitudinal and Vertical Variation of Synechococcus Assemblage Composition along 170 w Transect from the South Pacific to the Arctic Ocean. *Microbial Ecology* 77, 333–342.

Xu, D., H. Kong, E.-j. Yang, X. Li, N. Jiao, A. Warren, Y. Wang, Y. Lee, J. Jung, and S.-h. Kang (2020). Contrasting Community Composition of Active Microbial Eukaryotes in Melt Ponds and Sea Water of the Arctic Ocean Revealed by High Throughput Sequencing. *Frontiers in Microbiology* 11, 1–15.

Yabuki, A., T. Toyofuku, and K. Takishita (2014). Lateral Transfer of Eukaryotic Ribosomal RNA Genes: An Emerging Concern for Molecular Ecology of Microbial Eukaryotes. *The ISME Journal* 8, 1544–1547.

Yau, S., A. Lopes dos Santos, W. Eikrem, C. Gérikas Ribeiro, P. Gourvil, S. Balzano, M.-L. Escande, H. Moreau, and D. Vaulot (2020). *Mantoniella beaufortii* and *Mantoniella baffinensis* sp. nov. (Mamiellales, Mamiellophyceae), Two New Green Algal Species from the High Arctic. *Journal of Phycology* 56, 37–51.

Yilmaz, P. et al. (2011). Minimum Information about a Marker Gene Sequence (MIMARKS) and Minimum Information about Any (x) Sequence (MIxS) Specifications. *Nature Biotechnology* 29, 415–420.

Zafeiropoulos, H., H. Q. Viet, K. Vasileiadou, A. Potirakis, C. Arvanitidis, P. Topalis, C. Pavloudi, and E. Pafilis (2020). PEMA: A Flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S Ribosomal RNA, ITS, and COI Marker Genes. *GigaScience* 9, 1–12.

Zamora-Terol, S., A. Novotny, and M. Winder (2020). Reconstructing Marine Plankton Food Web Interactions Using DNA Metabarcoding. *Molecular Ecology* 29, 3380–3395.

Zeidner, G., C. M. Preston, E. F. Delong, R. Massana, A. F. Post, D. J. Scanlan, and O. Beja (2003). Molecular Diversity among Marine Picophytoplankton as Revealed by *psbA* Analyses. *Environmental Microbiology* 5, 212–216.

Zeldis, J., M. R. James, J. Grieve, and L. Richards (2002). Omnivory by Copepods in the New Zealand Subtropical Frontal Zone. *Journal of Plankton Research* 24, 9–23.

Zhu, F., R. Massana, F. Not, D. Marie, and D. Vaulot (2005). Mapping of Picoeucaryotes in Marine Ecosystems with Quantitative PCR of the 18S rRNA Gene. *FEMS Microbiology Ecology* 52, 79–92.

Zimmermann, J., G. Glöckner, R. Jahn, N. Enke, and B. Gemeinholzer (2015). Metabarcoding vs. Morphological Identification to Assess Diatom Diversity in Environmental Studies. *Molecular Ecology Resources* 15, 526–542.

Zuckerkandl, E. and L. Pauling (1965). Molecules as Documents of Evolutionary History. *Journal of Theoretical Biology* 8, 357–366.

Zwirglmaier, K., J. L. Heywood, K. Chamberlain, E. M. Woodward, M. V. Zubkov, and D. J. Scanlan (2007). Basin-Scale Distribution Patterns of Picocyanobacterial Lineages in the Atlantic Ocean. *Environmental Microbiology* 9, 1278–1290.

Zwirglmaier, K., L. Jardillier, M. Ostrowski, S. Mazard, L. Garczareck, D. Vaulot, F. Not, R. Massana, O. Ulloa, and D. J. Scanlan (2008). Global Phylogeography of Marine *Synechococcus* and *Prochlorococcus* Reveals a Distinct Partitioning of Lineages amongst Oceanic Biomes. *Environmental Microbiology* 10, 147–161.

Acknowledgements

We thank Fabrice Not for the suggestion to write this chapter. We thank publishers for allowing the inclusion of the previous published material. ALS was supported by Singapore Ministry of Education Academic Research Fund Tier 1 RG26/19. DV was supported by ANR contract PhytoPol (ANR-15-CE02-0007). CGR was supported by the FONDECYT project number 3190827.

List of Tables

List of Figures

Supplementary Material

Table S1: List of eukaryotic datasets and studies considered in this Chapter. The full list of references is available from GitHub repository.

Organelle	Gene	Region	Specificity	ID	Area	Ecosystem	Substrate	Bioproject	DOI
mitochondrion	coi			194	Balearic Islands	coastal	water, rock		10.1111/mec.15641
plastid	16S rRNA	V1-V2		119	North Atlantic Ocean	oceanic	water		10.1038/s41396-020-0636-0
plastid	16S rRNA	V1-V2		181	North Atlantic Ocean	oceanic	water	PRJNA489478	10.3389/fmicb.2020.542372
plastid	16S rRNA	V3		176	Portuguese Atlantic Port	coastal	biofilm	PRJEB33600	?10.3389/fmicb.2020.01938
plastid	16S rRNA	V3-V4		155	Baltic Sea	oceanic	water	PRJEB39191	10.1111/mec.15555
plastid	16S rRNA		photosynthetic eukaryotes	177	Fildes Bay, Antarctica	coastal	water	PRJNA280421	?10.1093/femsec/fiw088
plastid	16S rRNA			17	Southern Ocean	oceanic	water	PRJNA645261	10.1038/s41598-020-80568-8
plastid	16S rRNA			249	SPOT, North East Pacific Ocean	coastal	water	PRJEB10834	10.1038/nmicrobiol.2016.5
plastid	23S rRNA			253	Japan Sea off Korea, North West Pacific	coastal	water		10.7717/peerj.2115
plastid	rbcL		diatoms	229	Mayotte rivers	freshwater rivers	periphyton		10.1016/j.ecolind.2017.06.024
plastid	rbcL		diatoms	228	Swiss rivers	freshwater rivers	biofilm	PRJNA629002	10.1111/mec.15646
	18S rRNA	37f	foraminifera	278	Newfoundland shelf, NW Atlantic	oceanic	sediments	PRJNA668798	10.1038/s41598-020-77179-8
	18S rRNA	37f	foraminifera	279	Svalbard, NE Atlantic	coastal	sediments		10.1016/j.dib.2020.105553
	18S rRNA	37F-41F	foraminifera	148	Western Pacific Ocean	oceanic	sediments	PRJEB35877	10.1038/s41598-020-67221-0
	18S rRNA	V1-V3		232	Jeju Island, Korea	coastal	water	PRJNA298490	10.1016/j.bse.2015.10.002
	18S rRNA	V1-V3		162	Okinawa, Japan	coastal	water, sediments		10.1038/s41598-020-64858-9
	18S rRNA	V2-V3		118	Eastern English Channel	coastal	water	PRJNA240960	10.1371/journal.pone.0196987
	18S rRNA	V2-V3		264	Eastern English Channel	coastal	water	SRX768577	10.3389/fmars.2017.00416
	18S rRNA	V2-V3	ciliates	252	North West Atlantic Ocean	coastal	water	SRP064473	10.1038/ismej.2015.224
	18S rRNA	V2-V3		268	Yellow Sea, North West Pacific	coastal	water	SRP034609	10.1016/j.ecss.2017.03.019
	18S rRNA	V4		63	Araguaia River floodplain	freshwater rivers	water	PRJNA422037	10.1093/plankt/fbz019
	18S rRNA	V4		223	Arctic Ocean	coastal	water		10.1134/S0006297915110140
	18S rRNA	V4		6	Arctic Ocean	oceanic	ice	PRJEB7577	10.1080/09670262.2015.1077395
	18S rRNA	V4		38	Arctic Ocean	oceanic	ice	PRJNA368621	10.1007/s00248-017-1076-x
	18S rRNA	V4		5	Arctic Ocean	oceanic	water	PRJNA202104	10.1038/ismej.2014.197
	18S rRNA	V4		7	Arctic Ocean	oceanic	water	PRJNA262971	10.1111/jeu.12134
	18S rRNA	V4		8	Arctic Ocean	oceanic	water	SRA029114	10.1371/journal.pone.0027492
	18S rRNA	V4		9	Arctic Ocean	oceanic	water	PRJEB11449	10.1371/journal.pone.0148512
	18S rRNA	V4		37	Arctic Ocean	oceanic	water	PRJNA383398	10.1038/s41598-018-27705-6
	18S rRNA	V4		40	Arctic Ocean	oceanic	water	PRJNA243055	10.1128/AEM.02737-14
	18S rRNA	V4		41	Arctic Ocean	oceanic	water	PRJNA217438	10.1128/AEM.02737-14
	18S rRNA	V4		42	Arctic Ocean	oceanic	water	PRJEB24314	10.3389/fmars.2019.00479
	18S rRNA	V4		76	Arctic Ocean	oceanic	water	PRJNA561496	10.1038/s42003-020-0891-7
	18S rRNA	V4		105	Arctic Ocean	oceanic	water	PRJNA217362	10.1128/AEM.00247-19
	18S rRNA	V4		206	Arctic Ocean	oceanic	water	PRJEB9737	10.1016/j.cell.2019.10.008
	18S rRNA	V4		281	Arctic Ocean	oceanic	water	PRJNA384116	10.3389/fmicb.2017.01099
	18S rRNA	V4		39	Arctic Ocean	oceanic	water, ice, ice-algal aggregates	PRJEB23005	10.3389/fmicb.2018.01035.
	18S rRNA	V4		4	Arctic Ocean	oceanic	water, sediment trap	KT810188-KT818498	10.1128/AEM.03208-15
	18S rRNA	V4		250	Arctic Ocean and Nova Scotia shelf	oceanic	water	SRP058096, PRJNA283142, SRA05416018S	10.1093/plankt/fbw030
	18S rRNA	V4		13	Arctic Ocean, Fram Strait	oceanic	water		10.1111/jpy.12109
	18S rRNA	V4		156	Arctic Ocean, Fram Strait	oceanic	water	PRJEB26288	10.3389/fmars.2018.00429
	18S rRNA	V4		168	Arctic, Northern Atlantic	oceanic	water, ice		10.1093/plankt/fbw030
	18S rRNA	V4		213	Atlantic Ocean	coastal	water		10.1111/j.1365-294X.2009.04480.x
	18S rRNA	V4		151	Atlantic Ocean	oceanic	ballast water	PRJNA526939	10.1021/acs.est.0c01931
	18S rRNA	V4		20	Atlantic Ocean	oceanic	water	PRJNA497792	10.1111/jeu.12700
	18S rRNA	V4		212	Atlantic Ocean	oceanic	water		10.1038/s41598-021-82071-0
	18S rRNA	V4		134	Atlantic salmon farm, Isle of Lismore, Scotland	aquaculture	sediments	SUB4192838	10.1111/jeu.12670
	18S rRNA	V4		84	Australia	oceanic	water		10.1038/sdata.2018.130
	18S rRNA	V4		197	Austria Chile Ethiopia	freshwater lakes	water	PRJNA299108	10.1111/mec.13633
	18S rRNA	V4		248	Ballast water, Chesapeake Bay	coastal	water	PRJNA294684	10.1007/s00248-015-0684-6
	18S rRNA	V4		43	Baltic Sea	coastal	water	PRJEB23971	10.1002/lno.11177
	18S rRNA	V4	Haptophyta	44	Baltic Sea	coastal	water	PRJEB31858	10.1002/lno.11177
	18S rRNA	V4		101	Baltic Sea	oceanic	water	PRJNA361403	10.1016/j.ecss.2018.04.013
	18S rRNA	V4		154	Baltic Sea	oceanic	water	PRJEB39191	10.1111/mec.15555
	18S rRNA	V4		19	Baltic Sea	oceanic	water, ice	PRJEB21047	10.3354/meps12645
	18S rRNA	V4		287	Baltic Sea	oceanic	water, zooplankton	PRJEB39191	10.1111/mec.15555
	18S rRNA	V4		242	Barcelona, Mediterranean Sea	coastal	water	SRP079955	10.3389/fmars.2016.00147
	18S rRNA	V4		85	Bay of Banyuls-sur-Mer	oceanic	water	PRJNA449267	10.1038/s41396-018-0281-z
	18S rRNA	V4		99	Bay of Brest	coastal	water		10.1038/s41598-020-63326-8

Table S1: *(continued)*

Organelle	Gene	Region	Specificity	ID	Area	Ecosystem	Substrate	Bioproject	DOI
	18S rRNA	V4	Labyrinthulomycetes	270	Beaufort Inlet, NW Atlantic	coastal	water	PRJNA590600	10.1128/AEM.01652-20
	18S rRNA	V4		261	Bilge boat water, New Zealand	coastal	water		10.1371/journal.pone.0187636
	18S rRNA	V4		91	Black Sea	oceanic	water	DRA005367	10.1016/j.hal.2017.07.004
	18S rRNA	V4		190	Brazilian Atlantic Forest	freshwater rivers, freshwater lakes, lagoons	water	PRJEB37554	?10.1007/s00248-020-01612-8
	18S rRNA	V4		107	California coastal	coastal	water	PRJNA432755	10.1038/s41396-018-0172-3
	18S rRNA	V4		94	California Current Ecosystem	oceanic	water, sediment trap	PRJNA432581	10.1038/s41396-018-0322-7
	18S rRNA	V4		144	China Sea	coastal	water		10.1007/s00248-018-1235-8
	18S rRNA	V4	Labyrinthulomycetes	145	China Sea	coastal	water		10.1007/s00248-018-1235-8
	18S rRNA	V4		95	coast of Brasil	oceanic	sorted phytoplankton	PRJNA377956	10.1038/s41396-018-0050-z
	18S rRNA	V4		77	coast of California	coastal	water	PRJNA492502	10.1038/s41396-019-0472-2
	18S rRNA	V4		53	coast of Europe	coastal	water, sediments	PRJEB9133	10.1016/j.cub.2014.02.050
	18S rRNA	V4		111	Coorong Lagoon	estuarine	water		10.3354/ame01740
	18S rRNA	V4		231	Coorong Lagoon, Australia	lagoons	water		10.3354/ame01740
	18S rRNA	V4		251	Coral skeleton Australia and Papua New Guinea	coastal	coral	SRP073961	10.1038/srep31508
	18S rRNA	V4		244	East and South China Sea	oceanic	water	SRP070982, SRP071043	10.1111/1462-2920.13606
	18S rRNA	V4		139	East Atlantic Ocean	coastal, oceanic	water	PRJEB36099	10.5194/bg-17-2807-2020
	18S rRNA	V4		65	East China Sea	oceanic	water	PRJNA378896	10.1038/ismej.2017.183
	18S rRNA	V4	Haptophyta	89	East China Sea	oceanic	water	PRJNA436038	10.1007/s10872-019-00505-w
	18S rRNA	V4		123	East China Sea	oceanic	water	PRJNA418995	10.1016/j.hal.2020.101809
	18S rRNA	V4		237	East English Channel	coastal	water	SRP039908	10.1080/17451000.2015.1084425
	18S rRNA	V4		238	East English Channel	coastal	water		10.1016/j.pocean.2015.04.015
	18S rRNA	V4		106	Eastern Tropical South Pacific	oceanic	water	PRJNA263803	10.3389/fmicb.2014.00543
	18S rRNA	V4		28	English Channel	coastal	water		10.1038/s41396-020-0659-6
	18S rRNA	V4		29	English Channel	coastal	water		10.1038/s41396-020-0659-6
	18S rRNA	V4		196	Europe	freshwater lakes	water	PRJNA259710	10.1111/1462-2920.12591
	18S rRNA	V4		274	Finland	freshwater supply	water, biofilm	PRJNA509718	10.1186/s40168-019-0715-5
	18S rRNA	V4		64	Gulf of Gabes	coastal, oceanic	water	PRJEB84566	10.3354/ame01857
	18S rRNA	V4		265	HAM lagoon Magdalen Islands, NW Atlantic	lagoons	water	SRP043016	10.3354/ame01814
	18S rRNA	V4		67	Havre-ax-Maisons lagoon	coastal	water	PRJNA251749	10.3354/ame01814
	18S rRNA	V4		103	Isfjorden	oceanic	water, sediment trap		10.3354/ame01904
	18S rRNA	V4		198	Italy	freshwater lakes	water	PRJEB36925	10.3389/fmicb.2020.00789
	18S rRNA	V4		81	Jiaozhou Bay	estuarine	water	PRJNA577777	10.1016/j.hal.2020.101772
	18S rRNA	V4	biased against metazoans	72	Kerguelen Islands	oceanic	water		10.1016/j.protis.2019.125709
	18S rRNA	V4		108	Kuroshio Current	oceanic	water		10.1111/maec.12579
	18S rRNA	V4		147	Lagoon of Venice	coastal	water		10.4081/aiol.2020.8961
	18S rRNA	V4		233	Lake Bourget, France	freshwater lakes	sediments, water		10.1007/s00248-015-0627-2
	18S rRNA	V4		185	Lake Chaohu	freshwater lakes	water	PRJNA534176, PRJNA330896	?10.1016/j.scitotenv.2019.134803
	18S rRNA	V4		256	Lake Chaohu and Taihua, China	freshwater lakes	sorted phytoplankton	PRJNA330896	10.1111/1462-2920.13724
	18S rRNA	V4		183	Lake Fuxian	freshwater lakes	water	PRJNA534173	?0.3389/fmicb.2019.02016
	18S rRNA	V4		138	Lake Garda	freshwater lakes	water	PRJEB36925	10.3389/fmicb.2020.00789
	18S rRNA	V4		286	Lake Poyang, China	freshwater lakes	sorted phytoplankton		10.1093/femsec/fiy211
	18S rRNA	V4		178	Lakes in South America	freshwater lakes	water	PRJEB31039	10.1093/femsec/fiz038
	18S rRNA	V4		69	Mariana Trench	oceanic	water	PRJNA451086	10.1038/s41598-018-33790-4
	18S rRNA	V4		70	Mariana Trench	oceanic	water	PRJNA399026	10.3389/fmicb.2018.02023
	18S rRNA	V4		36	Mediterranean Sea	oceanic	water	PRJEB23788	10.1111/mec.14929
	18S rRNA	V4		49	Mediterranean Sea	oceanic	water	PRJEB24595	10.1093/femsec/fiw200
	18S rRNA	V4		97	Mulgol pond on Dokdo Iisland, Korea	freshwater ponds	water	PRJNA592034	10.1038/s41598-020-63561-z
	18S rRNA	V4	diatoms and raphidophytes	93	Narragansett Bay	estuarine	water		10.3390/biology9010019
	18S rRNA	V4		179	Neuse River Estuary	estuarine	water	PRJNA413761	10.1111/1462-2920.15221
	18S rRNA	V4	dinoflagellates	266	New Zealand	coastal	macroalgae epiphytes		10.1080/00288330.2017.1298632
	18S rRNA	V4		257	New Zealand	coastal	sediments	SRR5376949-SRR5376667	10.1016/j.marpolbul.2017.11.042
	18S rRNA	V4		164	New Zealand	oceanic	water	PRJEB37242	10.3354/ame01938
	18S rRNA	V4		260	New Zealand biofouling experiment	coastal	biofilm		10.1016/j.marenvres.2017.12.003
	18S rRNA	V4		61	Nordic Sea	oceanic	water	PRJNA503499	10.1016/j.ecolind.2019.105582
	18S rRNA	V4		71	north Atlantic Ocean	coastal	water		10.1111/1462-2920.14537
	18S rRNA	V4		102	North Atlantic Ocean	oceanic	copepod gut content		10.1016/j.dsr2.2020.104738
	18S rRNA	V4		79	North Atlantic Ocean	oceanic	water		10.1002/lno.11193
	18S rRNA	V4		282	North Atlantic Ocean	oceanic	water		10.1111/1462-2920.12431
	18S rRNA	V4		235	North East Pacific Ocean	oceanic	water		10.1093/femsec/fiv037

Table S1: *(continued)*

Organelle	Gene	Region	Specificity	ID	Area	Ecosystem	Substrate	Bioproject	DOI
	18S rRNA	V4		68	North East Pacific Ocean, Monterey Bay	oceanic	sediments	PRJNA521526	10.1038/s41396-019-0581-y
	18S rRNA	V4		122	North Equatorial Current	oceanic	water, sediments	SRP139064	10.1016/j.dsr.2020.103279
	18S rRNA	V4		73	North Pacific Subtropical Gyre	oceanic	water	PRJNA393172	10.3389/fmars.2018.00351
	18S rRNA	V4		56	North Sea	coastal	water	PRJEB37135	10.1093/plankt/fbaa017
	18S rRNA	V4		153	North Sea	coastal	water	PRJEB33076	10.1016/j.seares.2020.101914
	18S rRNA	V4		211	North Sea	coastal	water	PRJEB37135	10.1371/journal.pone.0244817
	18S rRNA	V4		100	Northern Adriatic Sea	oceanic	water	PRJNA576330	10.3389/fmicb.2019.02736
	18S rRNA	V4		182	NW Mediterranean coast	coastal	water	PRJEB23788, PRJEB38773, PRJEB38800, PRJEB38808	?10.1016/j.marpolbul.2020.111691
	18S rRNA	V4		1	Ocean survey	coastal, oceanic	water	PRJEB8682	10.1186/s13742-015-0066-5
	18S rRNA	V4		2	Ocean survey	coastal, oceanic	water		10.1186/s13742-015-0066-5
	18S rRNA	V4		3	Ocean survey	coastal, oceanic	water		10.1186/s13742-015-0066-5
	18S rRNA	V4		54	Ocean survey	oceanic	sediments	PRJEB33873	10.3389/fmars.2020.00234.
	18S rRNA	V4		34	Ocean survey	oceanic	water	PRJEB23771	10.1038/s41396-019-0506-9
	18S rRNA	V4		35	Ocean survey	oceanic	water	PRJEB23913	10.1186/s40168-020-00827-8
	18S rRNA	V4		205	Ocean survey	oceanic	water	PRJEB6610	10.1016/j.cell.2019.10.008
	18S rRNA	V4		240	Oil sands ponds, Canada	freshwater ponds	sediments	PRJNA313081	10.1111/jeu.12320
	18S rRNA	V4		191	Okinawa trough	oceanic	water	PRJNA546472	10.1007/s00248-020-01583-w
	18S rRNA	V4	Haptophyta	129	Oslofjorden, Skagerrak	coastal	water		10.1111/jeu.12388
	18S rRNA	V4	Haptophyta	236	Oslofjorden, Skagerrak	coastal	water	PRJEB5541	10.1111/mec.13160
	18S rRNA	V4		163	Panama	coastal	water	PRJNA507270	10.1007/s00338-020-01979-7
	18S rRNA	V4		269	Paris	freshwater gutters	water	PRJNA316490	10.1038/ismej.2017.166
	18S rRNA	V4		82	Pearl river to South China Sea	estuarine	water	PRJNA433777, PRJNA433778	10.1111/mec.14867
	18S rRNA	V4		60	Penghu archipelago, Taiwan	oceanic	water, sponge	PRJNA382576	10.1007/s13213-019-01476-5
	18S rRNA	V4		90	Red Sea	oceanic	water	PRJNA338026, PRJNA288901	10.1038/s41598-017-06928-z
	18S rRNA	V4		83	Red Sea coast	coastal	coral	PRJNA308164	10.1016/j.marenvres.2016.04.011
	18S rRNA	V4		57	Ria de Vigo	coastal	water	PRJEB23729	10.1111/1462-2920.14313
	18S rRNA	V4		230	Ria Formosa saltern, Portugam	saltern	water	SRP050177	10.1007/s00792-014-0713-2
	18S rRNA	V4		150	Saint-Charles River	freshwater rivers	water	PRJNA486319	10.3389/fmicb.2019.02359
	18S rRNA	V4		186	Saint-Charles River	freshwater rivers	water	PRJNA541322	10.3390/microorganisms8111631
	18S rRNA	V4		157	Salomon islands	lagoons	water		10.1111/1462-2920.13523
	18S rRNA	V4		74	San Pedro Ocean Time	oceanic	water		10.1038/s41396-018-0097-x
	18S rRNA	V4		87	Sargasso Sea	oceanic	water	PRJNA421139	10.1038/s41396-018-0163-4
	18S rRNA	V4		203	Scandinavia	freshwater lakes	water		10.1093/femsec/fiw231
	18S rRNA	V4		78	Scotia Sea	oceanic	suspended and sinking particles		10.1002/lno.11319
	18S rRNA	V4		110	Sendai Bay, Japan	coastal	sorted phytoplankton	DRA004439	10.1093/femsec/fiw229
	18S rRNA	V4		195	Siberia	freshwater lakes	water	PRJEB24415	10.3390/microorganisms8040543
	18S rRNA	V4		140	Skidaway River Estuary	estuarine	water	PRJNA575563	10.1128/mSphere.00209-20
	18S rRNA	V4		200	South America	freshwater rivers	water	PRJEB23471	10.3389/fevo.2018.00099
	18S rRNA	V4		283	South Atlantic Ocean	oceanic	water		10.5194/bg-15-5951-2018
	18S rRNA	V4		86	South China Sea	coastal	water	PRJNA506331	10.1016/j.pocean.2020.102309
	18S rRNA	V4		188	South China Sea	coastal	water	PRJNA628544	?10.1093/plankt/fbaa046
	18S rRNA	V4		58	South Pacific Ocean	oceanic	water	PRJNA385736	10.1073/pnas.1719335115
	18S rRNA	V4		92	Southeast Fujian, China	estuarine	water	PRJNA549238	10.3389/fmicb.2019.02640
	18S rRNA	V4		55	Southern Ocean	coastal, oceanic	water	PRJNA508517	10.1128/AEM.02634-18
	18S rRNA	V4		18	Southern Ocean	oceanic	sorted phytoplankton	PRJNA645244	10.1038/s41598-020-80568-8
	18S rRNA	V4		10	Southern Ocean	oceanic	water	PRJNA176875	10.1007/s00300-013-1438-x
	18S rRNA	V4		11	Southern Ocean	oceanic	water	PRJNA254097	10.1007/s00300-015-1815-8
	18S rRNA	V4		16	Southern Ocean	oceanic	water	PRJNA645244	10.1038/s41598-020-80568-8
	18S rRNA	V4		104	Southern Ocean	oceanic	water	PRJEB23910	10.1016/j.pocean.2018.10.008
	18S rRNA	V4		219	Southern Ocean	oceanic	water	SRA064723	10.1371/journal.pone.0113244
	18S rRNA	V4		280	Southern Ocean, Kerguelen region	oceanic	water		10.1016/j.dsr2.2018.12.003
	18S rRNA	V4		243	SPOT Station, North East Pacific	coastal	water	SRP070577	10.1093/femsec/fiw050
	18S rRNA	V4	diatoms	227	Swiss rivers	freshwater rivers	biofilm	PRJNA629002	10.1111/mec.15646
	18S rRNA	V4	diatoms	224	Switzerland waterways	freshwater rivers	periphyton		10.1021/es506158m
	18S rRNA	V4		66	Taiwan Strait	coastal	water	PRJNA560553	10.1016/j.scitotenv.2019.135753
	18S rRNA	V4		62	Thessaloniki Bay	coastal	water	PRJNA552665	10.3390/d12030114
	18S rRNA	V4		272	Ward Hunt lake, Arctic	freshwater lakes	water (experiment)	SAMN02258453	10.1111/1574-6941.12324
	18S rRNA	V4		180	Western Antarctic Peninsula	oceanic	water	PRJNA299401	10.1038/s41598-017-14109-1
	18S rRNA	V4		59	White Sea, Russia	oceanic	water	PRJNA368621	10.3390/d12030093
	18S rRNA	V4		184	Yangtze river	freshwater rivers	water		?10.1093/femsec/fiy011

Table S1: *(continued)*

Organelle	Gene	Region	Specificity	ID	Area	Ecosystem	Substrate	Bioproject	DOI
	18S rRNA	V4		146	Yantai coast, Yellow sea	coastal	water		10.1007/s11802-020-4461-x
	18S rRNA	V4		124	Yellow Sea	oceanic	water, sediments	PRJNA488669	10.1007/s00343-020-9234-2
	18S rRNA	V4-V5		75	Baltic Sea	oceanic	water	PRJEB12362	10.3389/fmicb.2016.00679
	18S rRNA	V4-V5		189	North West Pacific Ocean	oceanic	water	PRJNA476806	10.1016/j.scitotenv.2019.134289
	18S rRNA	V4-V5		267	Norwegian shelf, North Sea	coastal	sediments	PRJNA225939	10.1111/mec.13761
	18S rRNA	V5-V7		288	Zhoushan fishing ground, China Sea	oceanic	water	SRP104829.	10.1007/s11802-020-4327-2
	18S rRNA	V6-V8		160	Equatorial Pacific	coastal	water	PRJNA497851	10.1038/s41598-019-52648-x
	18S rRNA	V6-V8		116	Heard Island, near Antartica	freshwater ponds, lagoons, coastal	water	PRJNA335685	10.1038/srep44480
	18S rRNA	V6-V8		120	Sydney Harbour estuary	estuarine	water	PRJNA491799	10.1371/journal.pone.0209857
	18S rRNA	V7		171	Bering Sea	oceanic	copepod gut content	KC952737 - KC952871	10.1016/j.dsr2.2015.04.001
	18S rRNA	V7		117	East China	freshwater lakes	water	PRJNA506128	10.1186/s12302-020-00321-w
	18S rRNA	V7		174	Lake Bourget and Annecy	freshwater lakes	sediments		10.1111/1462-2920.13815
	18S rRNA	V7		241	Lake Bourget and Igaliku	freshwater lakes	sediments		10.1111/mec.13893
	18S rRNA	V7		166	Lakes in France	freshwater lakes	sediments	PRJEB35411	10.1038/s41467-020-17682-8
	18S rRNA	V7		172	West Antarctic Peninsula	oceanic	water, sediments	KR86490 - KR865949	10.1093/plankt/fbw005
	18S rRNA	V7-V9		226	Baltic Sea, Gulf of Finland	coastal	ice, water	PRJEB7625	10.1371/journal.pone.0130035
	18S rRNA	V7-V9		222	Pacific Ocean, Japan	coastal	water	DRA002426.	10.1016/j.gene.2015.10.026
	18S rRNA	V7-V9		254	Yatsushiro Sea, Japan, North West Pacific	coastal	water	DRA002425	10.1016/j.gene.2015.10.025
	18S rRNA	V8-V9		80	North Atlantic Ocean	oceanic	water		10.1002/lno.11193
	18S rRNA	V8-V9		158	Salomon islands	lagoons	water		10.1111/1462-2920.13523
	18S rRNA	V9		192	Alboran Sea	oceanic	water		10.1093/femsec/fiaa197
	18S rRNA	V9		50	Arctic Ocean	oceanic	water	PRJEB9737	10.1016/j.cell.2019.10.008
	18S rRNA	V9		165	Arctic, Adriatic sea, Cariaco basin	oceanic	water	PRJEB33135	10.1111/1462-2920.15190
	18S rRNA	V9		214	Atlantic Ocean, Framvaren fjord	coastal	water		10.1111/j.1365-294X.2009.04480.x
	18S rRNA	V9		216	Atlantic Ocean, Framvaren fjord and Cariaco trench	coastal, oceanic	water		10.1186/1741-7007-7-72
	18S rRNA	V9		135	Atlantic salmon farm, Isle of Lismore, Scotland	aquaculture	sediments	SUB4192838	10.1111/jeu.12670
	18S rRNA	V9		115	Canoe Cove	coastal	water	PRJNA328102	10.1038/s41467-017-02571-4
	18S rRNA	V9		275	Canoe Cove Nahant, NW Atlantic	coastal	water	SRR5177223-SRR5177507	10.1038/s41467-017-02571-4
	18S rRNA	V9		149	Eastern Mediterranean sea	oceanic	water	PRJEB26382	10.3354/ame01933
	18S rRNA	V9		221	English Channel	coastal	water	PRJEB5097	10.1111/jpy.12228
	18S rRNA	V9		202	Global	freshwater lakes	water	PRJEB41211	10.1016/j.envint.2020.106262
	18S rRNA	V9		218	Global	oceanic	water		10.1128/AEM.00057-14
	18S rRNA	V9		98	Mulgol pond on Dokdo Iisland, Korea	freshwater ponds	water	PRJNA592034	10.1038/s41598-020-63561-z
	18S rRNA	V9		113	North Atlantic and North Pacific Oceans	coastal	water, sediments	PRJNA362750	10.1111/1462-2920.13916
	18S rRNA	V9		112	North East Pacific Ocean	oceanic	water, sediment trap	PRJNA591905	10.1016/j.dsr2.2019.104708
	18S rRNA	V9		15	Ocean survey	oceanic	water	PRJEB6610	10.1126/science.1261605
	18S rRNA	V9		217	Pacific Ocean, Moorea reef	coastal	water		10.1038/ismej.2011.108
	18S rRNA	V9		245	Patagonia and Antarctic lakes	freshwater lakes	water		10.1111/1462-2920.13566
	18S rRNA	V9		114	Pearl river to South China Sea	estuarine, coastal	water	PRJNA430302	10.1016/j.ejop.2018.01.004
	18S rRNA	V9		246	Red Sea, KAUST mesocosms	coastal	water		10.1002/lno.10212
	18S rRNA	V9		247	Ross Sea	coastal, oceanic	water	PRJNA312643	10.1016/j.pocean.2016.10.003
	18S rRNA	V9		193	Shaying River	freshwater rivers	water		10.1111/gcb.15357
	18S rRNA	V9		167	Shidou and Bantou reservoirs	freshwater lakes	water	PRJNA415265	10.1016/j.watres.2020.116232
	18S rRNA	V9		263	South China Sea	oceanic	water	SRP104547	10.3389/fmicb.2017.01121
	18S rRNA	V9		169	South Pacific Ocean	oceanic	water	SRP140700	?10.1093/plankt/fbaa036
	18S rRNA	V9		143	Southern Ocean	oceanic	water	PRJEB16346	10.3389/fmicb.2018.01474
	18S rRNA	V9		220	Southern Ocean, Palmer station	coastal	water	SRP000903	10.3354/ame01703
	18S rRNA	V9		215	Southern Ocean, Palmer Station	coastal	water		10.1371/journal.pone.0006372
	18S rRNA	V9		173	Sydney Harbour	estuarine	water	PRJNA491799	10.1371/journal.pone.0209857
	18S rRNA	V9		121	Sydney Harbour estuary	estuarine	water	PRJNA491799	10.1371/journal.pone.0209857
	18S rRNA	V9		258	USA and China intertidal sands	coastal	sediments	PRJNA362750	10.1111/1462-2920.13916
	18S rRNA	V9		259	Xiamen island, China Sea	coastal	sediments, water	PRJNA342297	10.3389/fmicb.2017.01912
	18S rRNA	V9		239	Xiamen Island, China Sea	coastal	water	SRX651777	10.1371/journal.pone.0127721
	18S rRNA	V9		262	Xiamen reservoirs, China	freshwater lakes	water	SRP062446	10.1111/1755-0998.12652

Table S1: *(continued)*

Organelle	Gene	Region	Specificity	ID	Area	Ecosystem	Substrate	Bioproject	DOI
	28S rRNA	D1-D2		136	Atlantic salmon farm, Isle of Lismore, Scotland	aquaculture	sediments	SUB4192838	10.1111/jeu.12670
	28S rRNA	D1-D2	Haptophyta	131	Bay of Naples	coastal	water	PRJEB3337	10.1111/mec.12108
	28S rRNA	D1-D2		133	North West Mediterranean Sea	coastal	water	PRJEB18757	10.1016/j.hal.2017.06.003
	28S rRNA	D1-D2	Haptophyta	130	Oslofjorden, Skagerrak	coastal	water		10.1111/jeu.12388
	28S rRNA	D1-D2		142	West Greenland	coastal	water	PRJEB32397	10.3389/fmars.2020.00439
	28S rRNA	D1-D3		276	Nanao Island, China	coastal	water	SRP081548	10.1111/mec.14496
	ITS1		fungi	255	China Sea	coastal	sediments	SRP067807	10.1038/srep26528
	ITS1		fungi	271	Singapore and Malaysia coast	coastal	seagrass, sediments	PRJNA517736	10.1002/ece3.5631
	ITS2		Anthozoa and Demospongiae	175	Okinawa, Japan	coastal	water, sediments		10.1038/s41598-020-64858-9
	ITS2		Symbiodinium	234	Samoa coral	coastal	water, sediments, coral		10.1371/journal.pone.0145099
	ITS2		Symbiodinium	273	Virgin Islands	coastal	coral, sediments		10.7717/peerj.3472
	rRNA operon		fungi	141	Lake Stechlin	freshwater lakes	water, sediments	PRJNA437436	10.1111/1755-0998.12937
	rRNA operon		dinoflagellates	285	Unknown	coastal	water		10.3389/fmicb.2020.00844