CHAPTER 11

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

Adriana Lopes dos Santos[a], Catherine Gérikas Ribeiro[b], Denise Ong[a], Laurence Garczarek[c], Xiao Li Shi[d], Scott D. Nodder[e], Daniel Vaulot[c,a], and Andres Gutiérrez-Rodríguez[e]

[a]Asian School of the Environment, Nanyang Technological University, Singapore
[b]GEMA Center for Genomics, Ecology & Environment, Faculty of Sciences, Universidad Mayor, Santiago, Chile
[c]UMR 7144, ECOMAP, CNRS, Sorbonne Université, CNRS, Station Biologique de Roscoff, Roscoff, France
[d]State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China
[e]National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand

11.1 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.

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). New information on the genetic diversity of small-sized phytoplankton groups is particularly important since these groups 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 has 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 techniques combined with improved taxonomically annotated reference sequence databases allow rapid and cost-effective assessment of the composition of plankton communities at a global level (de Vargas et al., 2015).

11.1.1 The concept of molecular markers

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

Carl Woese and George Fox in 1977 established a "third domain of life," called Archaebacteria, after examining the sequences of the 16S/18S ribosomal ribonucleic acid (rRNA) of plants, yeast, human cells, bacteria, and methanogenic archaea (Woese and Fox, 1977). 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, 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 (HGT), 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 (Kitahara and Miyazaki, 2013; Yabuki et al., 2014).

In 1977 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, whereas less-conserved regions would allow phylogenetic evaluations. The suggested approach of primer regions flanking target gene markers is still an important concept today.

11.1.2 The advent of environmental sequencing

Medlin et al. (1988)'s landmark paper proposed the first set of primers and the polymerase chain reaction (PCR) conditions for the amplification of small subunit (SSU) rRNA eukaryotic gene (18S rRNA). This allowed determination of the nearly complete 18S rRNA sequence from the marine diatom *Skeletonema costatum*, opening the door for using SSU rRNA gene as biodiversity marker in studies targeting the diversity of natural plankton communities, published 13 years later (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 has 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, now renamed Picozoa, Not et al., 2007a). Moreover, other genes besides the 18S rRNA gene 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).

11.1.3 The transition to high throughput sequencing (HTS)

About 15 years ago, Sanger sequencing was overtaken by next generation sequencing (NGS) or high throughput sequencing (HTS). 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 factors. This led to a rapid increase in the environmental studies of community structure.

The two main HTS technologies used for microbial diversity studies have been developed by two companies: 454 (Roche) 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 process. The use of the 454 system was first suggested for metabarcoding by Kysela et al. (2005): as 454 was not optimized 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 study the 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 generated 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).

Metabarcoding, which consists of amplifying a marker gene with PCR and sequencing the amplicon using HTS, is the focus of this chapter. It can be also referred to as "amplicon sequencing," "tag sequencing" and "HTS". Sometimes it is mistankenly cited as "metagenomics" which refers to sequencing unamplified DNA. 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 (Fig. 11.1).

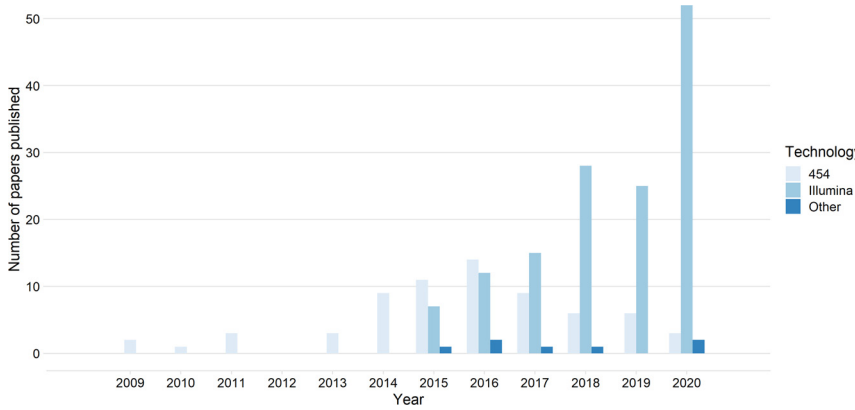

Figure 11.1 Papers using metabarcoding for aquatic protist studies published per year and technology (see Table S1 and https://github.com/vaulot/Chapter-2020-Lopes-Metabarcoding-phytoplankton).

11.2 The different steps of metabarcoding

11.2.1 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. 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 seagrasses (see below Section 11.4), all of which require specialized sampling techniques. The most common approach consists collecting sea or freshwater water with bottles (e.g., Niskin bottles) or with *in situ* pumps and 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; the choice may vary depending on the substrate considered, as well as the amount of material available. Most studies have focused 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) is then used as a proxy of

metabolic activity (Charvet et al., 2014; Egge et al., 2015; Hu et al., 2016; Massana et al., 2015; 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 factor of metabolic state or cell size) (Blazewicz et al., 2013).

11.2.2 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 targeted one used in eukaryotic plankton studies (Table 11.1). For prokaryotic phytoplankton (cyanobacteria), however, the equivalent 16S rRNA gene is not the most used (see Section 11.5). Though sequencing of the entire SSU rRNA is possible using Sanger sequencing, the sequence is too long for HTS. Small hypervariable regions of the 18S rRNA have been tested and selected over the years, based on their ability to distinguish different taxa. Hypervariable regions 4 (V4) and 9 (V9) are the 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 ~450 bp long, whereas V9 is located at the end of the gene with a ~150 bp length (Vaulot et al., 2021). Though both regions are able to provide a similar picture of eukaryotic phytoplankton diversity at higher taxonomic levels (division, class, see Tragin and Vaulot, 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), whereas 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 11.1).

Other marker genes have been used (Tables 11.1 and S1) to access eukaryotic phytoplankton diversity, for example, the plastid 16S rRNA (Trefault et al., 2021). The use of plastid 16S rRNA for metabarcoding has not been popularized, as had been earlier with clone libraries (e.g., Shi et al., 2011), but 16S plastid sequences have been obtained as a by-product of bacteria 16S metabarcoding studies (e.g., Needham and Fuhrman, 2016). One issue is the few reference sequences for plastidial 16S, despite efforts to increase the number of cultures sequenced (Decelle et al., 2015). The

Table 11.1 Most frequently used gene markers and 18S rRNA region in eukaryotic metabarcoding studies of aquatic phytoplankton with the number of papers (N) where used (Table S1).

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		

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). For example, Gran-Stadniczeñko et al. (2017) found more diversity using the D1–D2 region of 28S rRNA compared to the V4 region of 18S rRNA, but the lack of reference sequences remains a problem. Other marker genes that 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 must 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 2×300 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

Table 11.2 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	Amacher 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	López-García 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 et al. (2015)	10.3354/ame01740	7

TAReuk454FWD1 and TAReukREV3, as forward and reverse primers, respectively (Tables 11.2 and 11.3), which were among the first primers designed for metabarcoding (Stoeck et al., 2010). These primers target the V4 region, yielding a ~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 11.2) removes this bias, allowing more complete datasets (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 11.3 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 11.2 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

11.2.3 PCR and sequencing

Once DNA is extracted and the marker gene and the relevant primers selected, the next step involves one or two rounds of PCR, depending on the material available. Low amounts 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 sample "barcodes" that allow a large number of samples to be run together. Typically, for an Illumina MiSeq run, up to 300 samples can be multiplexed together, with 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.

11.2.4 Data processing

In this section, 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 the following: demultiplexing, quality filtering, removal of primers, assembly, clustering, removal of chimera, and taxonomic assignment. Several 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); whereas others 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 its initial versions being complicated to install, whereas dada2 running under R is becoming increasingly popular (Table 11.4). Another new option is PEMA, which groups a number of tools, such as VSEARCH and Swarm in a ready-to-use Docker container (Zafeiropoulos et al., 2020).

Table 11.4 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 a service, such as SILVAngs (https://ngs.arb-silva.de/), which is linked to the SILVA sequence database (see below) or FROGS (http://frogs.toulouse.inra.fr/). Such services suffer from having to upload your data on a web server, which can be slow, and from offering only fixed processing options (e.g., clustering or reference database).

Some web sites also facilitate downloading processed data from large scale projects such as Tara or Malaspina, (see below) and the Ocean Barcode Atlas (http://oba.mio.osupytheas.fr/ocean-atlas/, Vernette et al., 2021).

Read processing. Many samples are usually sequenced together, and 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 if this is not performed before, 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 for the adjustment of parameters using 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 artifacts at the latter processing stages.

Clustering. Highly similar sequences are grouped together and initial strategies defined as 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. Lastly, 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 datasets. 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 could 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. 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, whereas 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 the most comprehensive in terms of coverage. Its main handicap is the taxonomic annotation 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 (~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 taxonomic assignment can be done in several different ways. The best strategy is to use alignment-independent approaches, such as the naive Bayesian classifier, also known as the RDP (Ribossomal Database Project) 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 Erik, 2016). Other methods rely on alignments such as a LCAClassifier of Crest (Lanzén et al., 2012), which uses the aligned SILVA database as a reference, or pplacer (Matsen et al., 2010), which 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 or a reference database, but it is difficult to implement for a large number of OTU/ASV, and does not provide bootstrap values. 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 normalization. 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 can be taken (Love et al., 2014). Another concern is that the rRNA gene copy number varies between species. Though picoplankton typically have a few copies of rRNA operon, larger eukaryotes can have several thousands of 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. As the number of species for which the copy number is known remains very small, there are only few ways 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. One very useful R 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 (nonmetric multidimensional scaling), or to visualize networks. Another useful tool is the DESeq2 package (Love et al., 2014), which allows detecting 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. Researchers should be able to start from the original data and reproduce all steps to reach the final conclusions, following the FAIR Data Principles (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. Therefore it is imperative that raw sequences are deposited into public repositories, preferably the NCBI/EBI/DDBJ SRA (Sequence Read Archive), with standardized 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 also 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).

11.3 Protist metabarcoding studies in aquatic environments

To date, more than 240 studies have used metabarcoding to look at protist communities in aquatic systems (Table S1). The number of studies appears to increase exponentially in recent years (Fig. 11.1). A few of these studies have targeted specifically phytoplankton (e.g., using plastid 16S rRNA

primers), whereas most of them obtained data on the overall protist community since the most commonly used primer sets (Table 11.3) do not discriminate against heterotrophic groups, even if some of the studies focused on the sole analysis of phytoplankton groups. The majority of studies investigated marine systems, in particular oceanic and coastal (Table 11.5). Freshwater systems have been less intensively sampled, with the exception of lakes. Most aquatic studies have sampled the water and, much fewer, sediments (Table 11.5). Some key aquatic biotopes, such as coral reefs, have been scarcely sampled as have been epiphytic communities on algae or sea grasses.

Table 11.5 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 (Fig. 11.2). Metabarcoding data are available for two years (2014 and 2015), but sampling has resumed since 2018

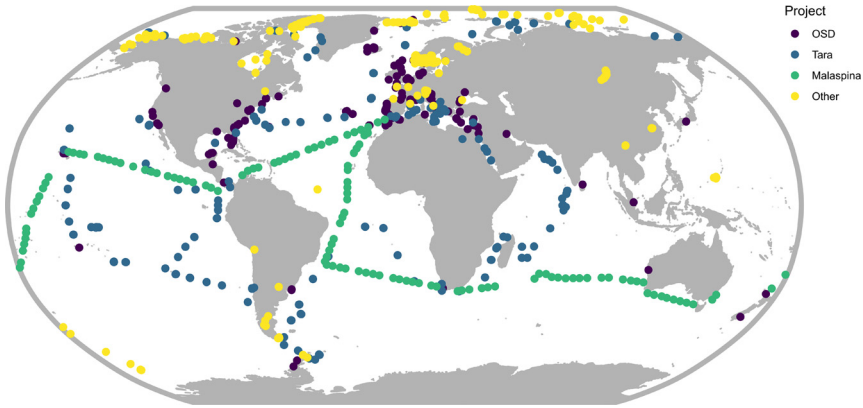

Figure 11.2 Map of stations investigated in a representative subset of 45 aquatic metabarcoding studies (see https://github.com/vaulot/Chapter-2020-Lopes-Metabarcoding-phytoplankton).

(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. Numerous papers have 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, Fig. 11.2), as well as 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) was acquired, as well as samples for microscopy. With respect to metabarcoding, one of the remarkable characteristics of the Tara expedition is the depth of sequencing achieved (typically one million of reads per sample), allowing sampling 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 as 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 (Lopes dos Santos et al., 2017; Ibarbalz et al., 2019). For more details on Tara *Oceans* expeditions, please refer to Chapter 15. The Malaspina ex-

pedition (2010–2011, Duarte, 2015) performed round-the-world sampling (Fig. 11.2) 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 localized studies allow inference of the ocean-wide distribution of newly described species (*Mantoniella beaufortii*, Yau et al., 2020), as well as of larger taxonomic groups (diatoms, Malviya et al., 2016).

11.3.1 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.

Metabarcoding has revealed the community progression is driven by drastic seasonal changes on solar energy input (Joli et al., 2017; Vick-Majors et al., 2014), and has demonstrated the interconnectivity between polar and subpolar regions (Sow et al., 2020), and between the Arctic and Antarctic domains (Segawa et al., 2018; Wolf et al., 2015). Metabarcoding also allows questions that microscopy-based methods could only address with difficulty: 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). Overall, many studies have collectively generated a solid knowledge base on polar protist diversity, unveiling phytoplankton community structure of the water (Duret et al., 2020; 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).

11.3.1.1 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 role in structuring plankton communities (Janout et al., 2016; Van Leeuwe et al., 2018), from light attenuation effects 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 and in melt pond water, under-ice water, and the deep-chlorophyll maximum using metabarcoding, and demonstrated that the sea ice, and specially multiyear ice, harbors a particularly high diversity of unique taxa. This study highlighted the specificity of the sympagic community dominated by diatoms, and a potential diversity loss in both sea ice and water column as multiyear 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 which are 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 the 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.

11.3.1.2 Impact of protist 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 beneath the melting sea ice for many kilometers from the ice edge into the pack ice (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. (2020) 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., 2020).

Microdiversity analysis of metabarcoding data (meaning, setting a threshold for diversity using a single nucleotide difference instead of fixed percentages of similarity, see Section 11.2.4) 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 subantarctic fronts (Sow et al., 2020). Tragin and Vaulot (2019) analyzed microdiversity within Mamiellophyceae members and reported a previously undetected *Micromonas* subarctic clade, which is phylogenetically close to the panarctic *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) (Fig. 11.3).

11.3.1.3 Bipolarity studies as a proof of concept for microbial dispersal theories

Although opposed to 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 interconnectivity paths and rates of evolutionary divergence (Darling et al., 2000; Sul et al., 2013). The use of metabarcoding enables a refined look at planktonic protists (including photosynthetic groups) 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.

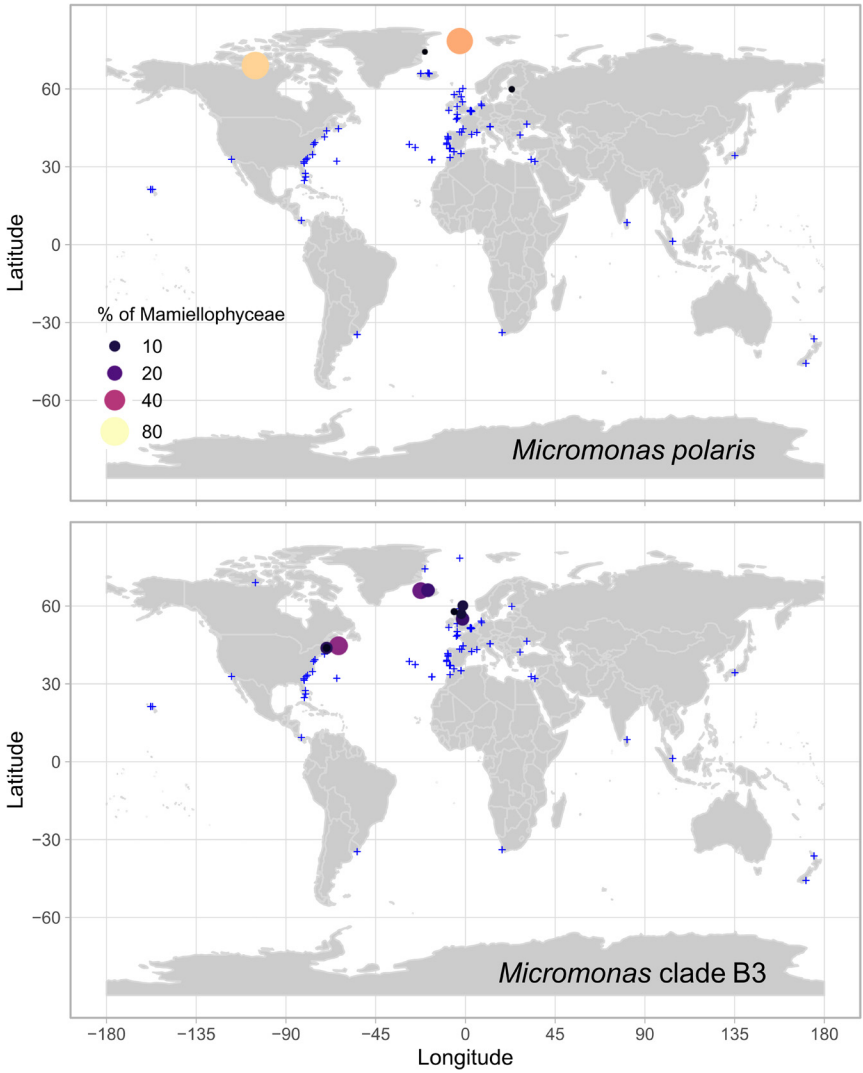

Figure 11.3 Distribution of the polar *Micromonas polaris* and the subarctic *Micromonas* B3 clade. Extracted from Tragin and Vaulot (2019).

(2018) analyzed distribution of red snow algae at 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".

11.3.1.4 Metabarcoding as a way to measure vulnerability of polar environments

Physicochemical 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 natural interactions can be rarely (if ever) duplicated in 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 of 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 ocean currents and their physicochemical parameters change. Moreover, metabarcoding monitoring can be useful at a much longer time scales, considering the paleo-oceanographic evolution of polar habitats. Pawłowska et al. (2020) investigated the microdiversity variability from Arctic planktonic foraminifera from the last 140,000 years using sediment ancient DNA (*sedaDNA*), and reported population-level responses to deglaciation and sea-ice retreat, due to Atlantic water inflow.

11.3.2 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 sinking particles; (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).

11.3.2.1 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). 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), however they

Table 11.6 Studies of vertical fluxes. na: not applicable; nd: not determined; metaB: metabarcoding; metaG: metagenomics; metaT: metatranscriptomics.

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 (Pras, Prym) in water column following deep mixing not reflected in sediment traps - stimulated production recycled locally
LeCleir et al. (2014)	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 confering 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 mitags	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 compositional differences between preserved and unpreserved traps with significant increase of heterotrophic protists (Stramenopiles-HNF-*Caecitellus* spp. and Phaeodaria) stress the importance of heterotrophic protist in remineralization and transformation of sinking particles
Duret et al. (2019)	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

continued on next page

Table 11.6 (*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
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.
Boeuf 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
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

are often not particularly resolutive for individual plankton studies (Nodder, 1997). Microscopy of the material collected in sediment traps has been used to characterize the biological components and to infer ecological processes driving POM vertical export. For example, 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; Scharek et al., 1999; Rynearson et al., 2013; Waite et al., 1992) as well as the contribution of other plankton groups such as coccolithophores (Honjo, 1976), radiolaria (Decelle et al., 2013; Gowing and Coale, 1989; Gowing, 1993; 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. Early applications of molecular approaches demonstrated the potential of genomic tools to characterize the biological composition of particles collected by sediment traps. Table 11.6 provides a summary of the studies that have coupled particle collection methods with molecular approaches to investigate the vertical fluxes of sinking particles. Sequencing of 18S rRNA clone libraries 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 the 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 commu-

nities across contrasting oceanographic regions (Boeuf et al., 2019; Duret et al., 2020; 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.

11.3.2.2 Phytoplankton and vertical export

Diatoms are arguably one of the most important contributors 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 fecal 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 contributions to export (Waite et al., 2000; Richardson, 2019).

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 (Fig. 11.4, 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 diatom contributions 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).

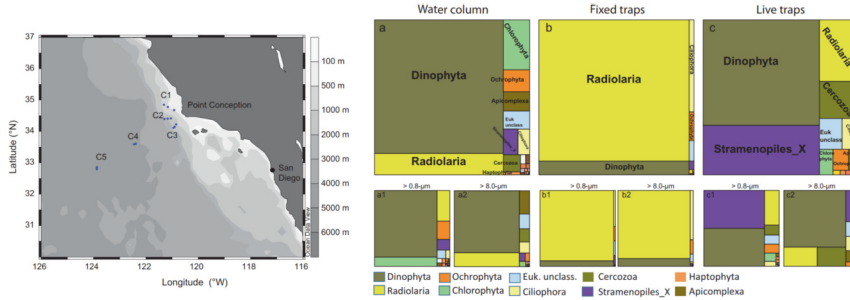

Figure 11.4 Protist community from the California current in the water column and in traps where material has been fixed or kept live (reprinted from Gutierrez-Rodriguez et al., 2019).

This study 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, whereas *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 a Jackson, 2007; Stukel et al., 2013; Waite et al., 2000). The formation of marine aggregates and trophic repackaging into fecal pellets, particularly by pelagic tunicates, such as 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. (2020) investigated the phytoplankton community composition of sinking and suspended particles sampled with the marine snow catcher at 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.

Metabarcoding of functional genes can also shed light on the contribution of specific groups to export. For example, Farnelid et al. (2019) applied 16S rRNA and nitrogenase gene (*nifH*) molecular barcodes to investigate the role of nitrogen-fixing organisms in the vertical export within the North Pacific subtropical gyre. This approach revealed for the first time the presence of a diverse suit of diazotrophs in sinking particles, indicating that new nitrogen could be directly exported vertically. In addition, the comparison between taxonomic affiliation of free-living and particle-associated N^2 fixers indicated that some taxa, assigned to larger size classes, were preferentially associated with particles, and hence present higher export potential.

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., 2020; 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 fecal 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, Acantharea and RAD-A, Preston et al., 2019). Similar analysis obtained from POM collected with moored traps (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.

Though metabarcoding can be only be taken, at best, as a semiquantitative 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. Coupled 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., ^{238}U–^{234}Th 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. It is also important to 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) and therefore can reflect different phytoplankton community composition. 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). Sampling biases between CTD-Niskin bottle and high volume *in situ* pump methods (Liu et al., 2009; Twining et al., 2015) highlighted 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. (Duret et al., 2020) for instance used the marine snow catcher that allows separate collection of sinking and suspended particles (Lampitt et al., 1993) in combination DNA metabarcoding analysis of different size fractions 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).

11.3.3 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 meso-zooplankton groups (copepods and cladocerans). Their comprehensive study revealed the shift from diatoms and dinoflagellate dominance in copepods diet during the

spring, whereas 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 nontargeted 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).

11.4 Marine picocyanobacteria

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

Marine picocyanobacteria of the genera *Prochlorococcus* and marine *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). They likely diverged from a common ancestor around 150 million years ago and, on the basis of 16S rRNA gene sequences, most members of the two groups would be considered to be the same species (Dufresne et al., 2005). The main difference between the two genera resides in their photosynthetic apparatus. Like most cyanobacteria, *Synechococcus* main light-harvesting antenna is the phycobilisome, which comprises phycobiliproteins (for example, phycoerythrin and phycocyanin), each binding to one or several light-harvesting chromophores, such as phycoerythrobilin and phycocyanobilin (Ting et al., 2002). *Prochlorococcus* (as *Prochloron* and *Prochlorothrix*) lacks phycobilisomes and its main light-harvesting antenna complex is composed of prochlorophyte chlorophyll-binding protein (Pcb), which binds divinyl chlorophyll *a* and *b* (Ting et al., 2002). Differences in cell size and photosynthetic pigments of these genera produce distinct flow cytometry profiles, which was used to reveal their abundance and ubiquity, 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, whereas *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 suggested the existence of two distinct *Prochlorococcus* ribotypes with different adaptations to light availability (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. Moore et al. (1998) demonstrated that co-occurring populations, displaying discrete distribution of Chl autofluorescence and light scatter, contain physiologically and genetically distinct types of *Prochlorococcus*, with different growth responses over a range of light intensities. Despite the complexity of their environment, these broad categories of *Prochlorococcus*, low- and high-light, shift their relative abundance throughout annual cycles in a predictable manner (Malmstrom et al., 2010; Zwirglmaier et al., 2007). Although the knowledge on *Prochlorococcus* diversity has increased with the availability of sequencing data and methods (Kashtan et al., 2014), the notion of the HL and LL groups that emerged from the analysis of 16S rRNA sequences combined with phenotypic and ecological data is still valid today (see Biller et al., 2015).

In contrast to *Prochlorococcus*, *Synechococcus* can be found in both marine and freshwater and is present in almost all marine environments, including high latitude and nutrient-rich coastal waters (Partensky et al., 1999). These differences might be explained by some of *Synechococcus* physiological traits, such as the capacity to tune its phycobilisome antenna systems to changing temperatures (Mackey et al., 2013; Pittera et al., 2014) and its higher growth rates (Chen et al., 2009). Marine *Synechococcus* have long been classified into marine clusters A, B, and C (MC-A, MC-B, and MC-C), based on the composition of their major light-harvesting pigments, their swimming motility (Waterbury et al., 1985), their salt requirement for growth, and their GC content (Waterbury and Rippka, 1989). *Synechococcus* 16 rRNA gene sequences from various marine isolates led to the identification of up to 10 clades within marine cluster-A (Fuller et al., 2003; Urbach et al., 1992, 1998; West and Scanlan, 1999), which was renamed subcluster (SC) 5.1 (Herdman et al., 2015).

Different approaches have been developed to assess the structure of marine picocyanobacteria communities based on 16S rRNA sequences. For instance, clade-specific oligonucleotides indicated that the dominance of specific *Synechococcus* clades could be correlated with physical features of the water column (Fuller et al., 2006, 2003, 2005). These studies and notably the study by Zwirglmaier et al. (2008) allowed a broad description of the main characteristics of four dominant *Synechococcus* clades and of their global biogeography. Clades I and IV were predominant in both coastal and cold, high latitude waters (above c. 30°N and below 30°S); clade II was found in coastal/continental shelf zones in tropical and subtropical waters; clade III was detected in a fairly narrow window of nitrate and phosphate concentrations, suggesting members of this clade are oligotrophs, whereas other clades, such as IX and X, were only detected at low abundance in the field (Zwirglmaier et al., 2007, 2008). These early studies targeted only *Synechococcus* clades established from isolated and sequenced strains, and therefore were likely missing the uncultivated diversity.

Several studies have focused on which variable region of the 16S rRNA gene is more suitable to describe bacterial diversity in natural assemblages, and although these comparisons led to contrasting results, the V4 and/or V4–V5 fragments were generally found to be the best for establishing phylogenetic relationships (Apprill et al., 2015; Caporaso et al., 2011; Parada et al., 2016). The variable region of the 16S rRNA gene that best reflects the full-length gene phylogeny may also depend on the tax onomic groups being studied. For picocyanobacteria, the V5, V6, and V7 hypervariable regions were shown to provide the highest taxonomic and phylogenetic resolution for the genera *Synechococcus*, *Prochlorococcus*, and *Cyanobium* (Huber et al., 2019).

The number of reads recovered from the cyanobacteria phylum using general 16S rRNA gene primers usually hardly exceeds 20% of the total read abundance, even when applying HTS methods, such as metagenomics and metabarcoding (Doherty et al., 2017; Sunagawa et al., 2015). Huber et al. (2019) developed new 16S rRNA primers targeting specifically picocyanobacteria, suitable to metabarcoding. By using these cyanobacteria-specific primers, most of the sequences obtained (97%) could be assigned to picocyanobacteria and used in phylogenetic reconstructions that featured topologies comparable to those generated when using the complete 16S rRNA gene sequence. Although they likely deserve to be further tested with more environmental samples and/or mock communities, these primers appear well-suited to study the diversity, community structure, and

dynamics of picocyanobacteria communities by 16S rRNA gene metabarcoding.

11.4.2 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 11.7). 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, such as *rpoC1* (Bergsland and Haselkorn, 1991; Tai and Palenik, 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

Table 11.7 Genes used for metabarcoding of *Synechococcus*.

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
petB	cytochrome b6	Mazard et al., 2012 Farrant et al., 2016
rbcL	large subunit of the RuBisCO (Ribulose 1,5-bisphosphate carboxylase-oxygenase)	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

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).

The different 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 sequences 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 11.4.3).

11.4.3 $_{mi}$tags as an alternative to picocyanobacteria metabarcoding

Metagenomics approaches are particularly suitable for diversity and functional studies of marine picocyanobacteria. Their relative high abundance and ubiquity *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). The availability of several complete or near-complete genomes 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 the mining off *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. The $_{mi}$tags approach is a non-PCR-based strategy to retrieve 16S rRNA sequences from metagenome datasets to access the taxonomic diversity and structure of prokaryotic communities (Logares et al., 2014). The $_{mi}$tags approach recovers sufficient reads for

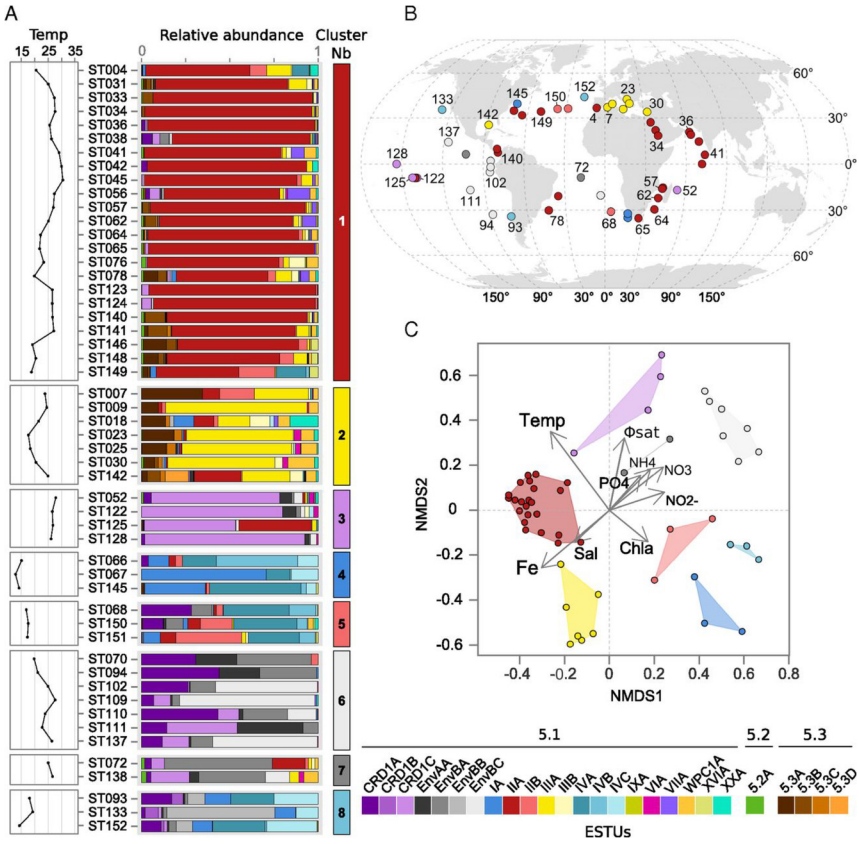

Figure 11.5 Biogeography of *Synechococcus* ESTUs in surface Tara *Oceans* metagenomes and relation to physico chemical parameters. (A) Histograms of the relative abundance of *Synechococcus* ESTUs at each station sorte by similarity. Left panels indicate seawater temperature (°C) at each station. (B) 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 physicochemical parameters. Reprinted from Farrant et al. (2016).

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 (Fig. 11.5). 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.

For example, within clade II, ESTU IIB was found to co-occur with clades I and IV in cold waters, whereas 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. 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, whereas 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 physicochemical 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 to analyze, especially for long-term monitoring or studies including large numbers of samples.

11.5 Future directions

The metabarcoding approach has 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 has revealed the existence of the "rare biosphere," where a large number of species are detected in very low abundance (Sogin et al., 2006). Though 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 (e.g., de Vargas et al., 2015).

Although metabarcoding is subject to various technical biases, currently 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 suggested 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 regions 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 has been applied to the characterization of food webs and predator-prey interactions, as mentioned in Section 11.4.3. 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, the 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.

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 with an open curated reference sequence database, is needed to advance our understanding of crucial ecosystem 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 the time of 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). The use of alternative long-read sequencing strategies (such as PacBio circular consensus sequencing) with specific pipelines was reported to reduce error rates near to zero (Callahan et al., 2019), and along with the continuous development of sequencing technologies, might be changing the long-read metabarcoding scenario soon. Moreover, new reference sequence databases containing longer gene fragments would need to be developed.

Acknowledgments

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 the

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 FONDECYT project 3190827.

Appendix 11.6 Supplementary material

Supplementary material related to this chapter can be found online at https://doi.org/10.1016/B978-0-12-822861-6.00020-0.

References

Ahlgren, N., Rocap, G., 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., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Amacher, J., Neuer, S., Anderson, I., Massana, R., 2009. Molecular approach to determine contributions of the protist community to particle flux. Deep-Sea Research. Part 1. Oceanographic Research Papers 56, 2206–2215.

Amacher, J., Neuer, S., Lomas, M., 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 2. Topical Studies in Oceanography 93, 71–83.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., Huse, S.M., 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., Kooistra, W.H., Levialdi Ghiron, J.H., Mann, D.G., Pröschold, T., Montresor, M., 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. Protist 158, 193–207.

Apprill, A., Mcnally, S., Parsons, R., Weber, L., 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquatic Microbial Ecology.

Arrigo, K.R., et al., 2014. Phytoplankton blooms beneath the sea ice in the Chukchi Sea. Deep-Sea Research. Part 2. Topical Studies in Oceanography 105, 1–16.

Bach, L.T., Stange, P., Taucher, J., Achterberg, E.P., Algueró-Muñiz, M., Horn, H., Esposito, M., Riebesell, U., 2019. The influence of plankton community structure on sinking velocity and remineralization rate of marine aggregates. Global Biogeochemical Cycles 33, 971–994.

Balzano, S., Abs, E., Leterme, S.C., 2015. Protist Diversity along a Salinity Gradient in a Coastal Lagoon. Aquatic Microbial Ecology 74 (3), 263–277.

Banse, K., 2013. Reflections about chance in my career, and on the top-down regulated world. Annual Review of Marine Science 5, 1–19.

Benner, I., Irwin, A.J., Finkel, Z.V., 2019. Capacity of the common Arctic picoeukaryote *Micromonas* to adapt to a warming ocean. Limnology and Oceanography Letters 5, 221–227.

Bergsland, K.J., Haselkorn, R., 1991. Evolutionary relationships among Eubacteria, Cyanobacteria, and Chloroplasts: evidence from the *rpoC1* gene of *Anabaena* sp. strain PCC 7120. Journal of Bacteriology 173, 3446–3455.

Beszteri, S., Thoms, S., Benes, V., Harms, L., Trimborn, S., 2018. The response of three Southern Ocean phytoplankton species to ocean acidification and light availability: a transcriptomic study. Protist 169, 958–975.

Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., Gorsky, G., Guidi, L., Kiko, R., Not, F., 2016. In situ imaging reveals the biomass of giant protists in the global ocean. Nature 532, 504–507.

Biers, E.J., Sun, S., Howard, E.C., 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., Berube, P.M., Lindell, D., Chisholm, S.W., 2015. *Prochlorococcus*: the structure and function of collective diversity. Nature Review, Microbiology 13, 13–27.

Bittner, L., Gobet, A., Audic, S., Romac, S., Egge, E.S., Santini, S., Ogata, H., Probert, I., Edvardsen, B., de Vargas, C., 2013. Diversity patterns of uncultured haptophytes unravelled by pyrosequencing in Naples Bay. Molecular Ecology 22, 87–101.

Blanchot, J., Rodier, M., 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 1. Oceanographic Research Papers 43, 877–895.

Blazewicz, S.J., Barnard, R.L., Daly, R.A., Firestone, M.K., 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. The ISME Journal 7, 2061–2068.

Boeuf, D., Edwards, B.R., Eppley, J.M., Hu, S.K., Poff, K.E., Romano, A.E., Caron, D.A., Karl, D.M., DeLong, E.F., 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.

Bower, S.M., Carnegie, R.B., Goh, B., Jones, S.R., Lowe, G.J., Mak, M.W., 2004. Preferential PCR amplification of parasitic protistan small subunit rDNA from metazoan tissues. Journal of Eukaryotic Microbiology 51, 325–332.

Boyd, P.W., Claustre, H., Levy, M., Siegel, D.A., Weber, T., 2019. Multi-faceted particle pumps drive carbon sequestration in the ocean. Nature 568, 327–335.

Boyd, P.W., Newton, P.P., 1999. Does planktonic community structure determine downward particulate organic carbon flux in different oceanic provinces? Deep-Sea Research. Part 1 46, 63–91.

Boyd, P.W., Strzepek, R., Takeda, S., Jackson, G., Wong, C.S., McKay, R.M., Law, C., Kiyosawa, H., Saito, H., Sherry, N., Johnson, K., Gower, J., Ramaiah, N., 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., Trull, T.W., 2007. Understanding the export of biogenic particles in oceanic waters: is there consensus? Progress in Oceanography 72, 276–312.

Bucklin, A., Allen, L., 2004. MtDNA sequencing from zooplankton after long-term preservation in buffered formalin. Molecular Phylogenetics and Evolution 30, 879–882.

Buesseler, K.O., Lamborg, C.H., Boyd, P.W., Lam, P.J., Trull, T.W., Bidigare, R.R., Bishop, J.K.B., Casciotti, K.L., Dehairs, F., Elskens, M., Honda, M., Karl, D.M., Siegel, D.A., Silver, M.W., Steinberg, D.K., Valdes, J., Van Mooy, B., Wilson, S., 2007. Revisiting carbon flux through the ocean's twilight zone. Science 316, 567–570.

Burd, A.B., Hansell, D.A., Steinberg, D.K., Anderson, T.R., Aristegui, J., Baltar, F., Beaupre, S.R., Buesseler, K.O., DeHairs, F., Jackson, G.A., Kadko, D.C., Koppelmann, R., Lampitt, R.S., Nagata, T., Reinthaler, T., Robinson, C., Robison, B.H., Tamburini, C., Tanaka, T., 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 2. Topical Studies in Oceanography 57, 1557–1571.
Cai, H., Wang, K., Huang, S., Jiao, N., Chen, F., 2010. Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. Applied and Environmental Microbiology 76, 2955–2960.
Calbet, A., Saiz, E., 2005. The ciliate-copepod link in marine ecosystems. Aquatic Microbial Ecology 38, 157–167.
Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11, 2639–2643.
Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from illumina amplicon data. Nature Methods 13, 581–583.
Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., McGill, S.K., Dougherty, M.K., 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Research 47 (18), e103.
Campbell, L., Vaulot, D., 1993. Photosynthetic picoplankton community structure in the sub-tropical North Pacific Ocean near Hawaii (station ALOHA). Deep-Sea Research 40, 2043–2060.
Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America.
Caron, D., Countway, P., 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. Aquatic Microbial Ecology 57, 227–238.
Caron, D.A., Countway, P.D., Jones, A.C., Kim, D.Y., Schnetzer, A., 2012. Marine protistan diversity. Annual Review of Marine Science 4, 467–493.
Carroll, E.L., Gallego, R., Sewell, M.A., Zeldis, J., Ranjard, L., Ross, H.A., Tooman, L.K., O'Rorke, R., Newcomb, R.D., Constantine, R., 2019. Multi-locus DNA metabarcoding of zooplankton communities and scat reveal trophic interactions of a generalist predator. Scientific Reports 9, 281.
Casey, J.M., Meyer, C.P., Morat, F., Brandl, S.J., Planes, S., Parravicini, V., 2019. Reconstructing hyperdiverse food webs: gut content metabarcoding as a tool to disentangle trophic interactions on coral reefs. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210X.13206.
Cavalier-Smith, T., Lewis, R., Chao, E.E., Oates, B., Bass, D., 2009. *Helkesimastix Marina* n. Sp. (Cercozoa: Sainouroidea Superfam. n.) a Gliding Zooflagellate of Novel Ultrastructure and Unusual Ciliary Behaviour. Protist 160 (3), 452–479.
Cavan, E.L., Henson, S.A., Belcher, A., Sanders, R., 2017. Role of zooplankton in determining the efficiency of the biological carbon pump. Biogeosciences 14, 177–186.
Chariton, A.A., Stephenson, S., Morgan, M.J., Steven, A.D., Colloff, M.J., Court, L.N., Hardy, C.M., 2015. Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. Environmental Pollution 203, 165–174.
Charvet, S., Vincent, W.F., Lovejoy, C., 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, B., Liu, H., Landry, M.R., Chen, M., Sun, J., Shek, L., Chen, X., Harrison, P.J., 2009. Estuarine nutrient loading affects phytoplankton growth and microzooplankton grazing at two contrasting sites in Hong Kong coastal waters. Marine Ecology. Progress Series 379.
Chen, F., Wang, K., Kan, J., Bachoon, D.S., Lu, J., Lau, S., Campbell, L., 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.
Chen, F., Wang, K., Kan, J., Suzuki, M.T., Wommack, K.E., 2006. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. Applied and Environmental Microbiology 72, 2239–2243.
Chénard, C., Suttle, C.A., 2008. Phylogenetic diversity of sequences of cyanophage photosynthetic gene *psbA* in marine and freshwaters. Applied and Environmental Microbiology 74, 5317–5324.
Cheung, M.K., Au, C.H., Chu, K.H., Kwan, H.S., Wong, C.K., 2010. Composition and genetic diversity of Picoeukaryotes in subtropical coastal waters as revealed by 454 Pyrosequencing. The ISME Journal 4 (8), 1053–1059.
Choi, D.H., Noh, J.H., 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., Li, W.K.W., Tremblay, J.-E., Carmack, E.C., Lovejoy, C., 2011. Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. PLoS ONE 6 (11), 1–12. https://doi.org/10.1371/journal.pone.0027492.
Comeau, A.M., Philippe, B., Thaler, M., Gosselin, M., Poulin, M., Lovejoy, C., 2013. Protists in Arctic drift and land-fast sea ice. Journal of Phycology 49, 229–240.
Cordier, T., Forster, D., Dufresne, Y., Martins, C.I., Stoeck, T., Pawlowski, J., 2018. Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. Molecular Ecology Resources 18, 1381–1391.
Craig, C., Kimmerer, W.J., Cohen, C.S., 2013. A DNA-based method for investigating feeding by copepod nauplii. Journal of Plankton Research 36, 271–275.
Darling, K.F., Wade, C.M., Stewart, I.A., Kroon, D., Dingle, R., Brown, A.J.L., 2000. Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. Nature 405, 43–47.
Davey, M.P., Norman, L., Sterk, P., Huete-Ortega, M., Bunbury, F., Loh, B.K.W., Stockton, S., Peck, L.S., Convey, P., Newsham, K.K., Smith, A.G., 2019. Snow algae communities in Antarctica: metabolic and taxonomic composition. New Phytologist 222, 1242–1255.
Dawson, S.C., Hagen, K.D., 2009. Mapping the protistan 'rare biosphere'. Journal of Biology.
De La Rocha, C.L., Passow, U., 2007. Factors influencing the sinking of POC and the efficiency of the biological carbon pump. Deep-Sea Research. Part 2. 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., Thomas, A.C., McInnes, J.C., Clarke, L.J., Vesterinen, E.J., Clare, E.L., Kartzinel, T.R., Eveson, J.P., 2019. Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? Molecular Ecology 28, 391–406.
Decelle, J., Martin, P., Paborstava, K., Pond, D.W., Tarling, G., Mahé, F., de Vargas, C., Lampitt, R., Not, F., 2013. Diversity, ecology and biogeochemistry of cyst-forming acantharia (radiolaria) in the oceans. PLoS ONE 8, e53598.
Decelle, J., Romac, S., Stern, R.F., Bendif, E.M., Zingone, A., Audic, S., Guiry, M.D., Guillou, L., Tessier, D., Le Gall, F., Gourvil, P., Lopes dos Santos, A., Probert, I., Vaulot, D., de Vargas, C., Christen, R., 2015. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. Molecular Ecology Resources 15, 1435–1445.
DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R., Chisholm, S.W., Karl, D.M., 2006.

Community genomics among stratified microbial assemblages in the ocean's interior. Science 311, 496–503.
Díez, B., Pedrós-Alió, C., Massana, R., 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., Yager, P.L., Moran, M.A., Coles, V.J., Fortunato, C.S., Krusche, A.V., Medeiros, P.M., Payet, J.P., Richey, J.E., Satinsky, B.M., Sawakuchi, H.O., Ward, N.D., Crump, B.C., 2017. Bacterial biogeography across the Amazon River-ocean continuum. Frontiers in Microbiology.
Doré, H., et al., 2020. Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria. Frontiers in Microbiology.
Duarte, C.M., 2015. Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition. Limnology and Oceanography Bulletin 24, 11–14.
Dufresne, A., Garczarek, L., Partensky, F., 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biology 6, R14.
Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., Paulsen, I.T., de Marsac, N.T., Wincker, P., Dossat, C., Ferriera, S., Johnson, J., Post, A.F., Hess, W.R., Partensky, F., 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biology 9, R90.
Dunthorn, M., Stoeck, T., Clamp, J., Warren, A., Mahé, F., 2014. Ciliates and the rare biosphere: a review. Journal of Eukaryotic Microbiology 61, 404–409.
Duret, M.T., Lampitt, R.S., Lam, P., 2019. Prokaryotic niche partitioning between suspended and sinking marine particles. Environmental Microbiology Reports 11 (3), 386–400.
Duret, M.T., Lampitt, R.S., Lam, P., 2020. 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, S49–S70. https://doi.org/10.1002/lno.11319.
Durkin, C.A., Van Mooy, B.A.S., Dyhrman, S.T., Buesseler, K.O., 2016. Sinking phytoplankton associated with carbon flux in the Atlantic Ocean. Limnology and Oceanography 61, 1172–1187.
Ebersbach, F., Assmy, P., Martin, P., Schulz, I., Wolzenburg, S., Nöthig, E.-M., 2014. Particle flux characterisation and sedimentation patterns of protistan plankton during the iron fertilisation experiment LOHAFEX in the Southern Ocean. Deep-Sea Research. Part 1. 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., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.
Edgcomb, V.P., Kysela, D.T., Teske, A., Gomez, A.d.V., Sogin, M.L., 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., Egge, E.S., Vaulot, D., 2016. Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding – a review. Perspectives in Phycology 3, 77–91.
Egge, E.S., Johannessen, T.V., Andersen, T., Eikrem, W., Bittner, L., Larsen, A., Sandaa, R.A., Edvardsen, B., 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. Meren Lab.. https://merenlab.org/2019/02/24/fantastic-data/.

Ernst, A., Becker, S., Wollenzien, U.I., Postius, C., 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.

Farnelid, H., Turk-Kubo, K., Ploug, H., Ossolinski, J.E., Collins, J.R., Van Mooy, B.A.S., Zehr, J.P., 2019. Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. The ISME Journal 13, 170–182.

Farrant, G.K., Doré, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M., Pitt, F.D., Wincker, P., Scanlan, D.J., Iudicone, D., Acinas, S.G., Garczarek, L., 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.

Ferris, M.J., Palenik, B., 1998. Niche adaptation in ocean cyanobacteria. Nature 396, 226–228.

Fitch, W.M., 1976. The molecular evolution of cytochrome c in eukaryotes. Journal of Molecular Evolution 8, 13–40.

Fitch, W.M., Margoliash, E., 1967. Construction of phylogenetic trees. Science 155, 279–284.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., Karl, D.M., Li, W.K., Lomas, M.W., Veneziano, D., Vera, C.S., Vrugt, J.A., Martiny, A.C., 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., Eppley, J.M., Samo, T.J., Karl, D.M., Delong, E.F., 2015. Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. Frontiers in Microbiology 6, 1–14.

Fryxell, G.A., Doucette, G.J., Hubbard, G.F., 1981. The genus *Thalassiosira*: the bipolar diatom *T. antarctica* comber. Botanica Marina 24, 321–336.

Fu, Y., Zheng, P., Zhang, X., Zhang, Q., Ji, D., 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., Campbell, C., Allen, D.J., Pitt, F.D., Le Gall, F., Vaulot, D., Scanlan, D.J., 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., Marie, D., Partensky, F., Vaulot, D., Post, A.F., Scanlan, D.J., 2003. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. Applied and Environmental Microbiology 69, 2430–2443.

Fuller, N.J., West, N.J., Marie, D., Yallop, M., Rivlin, T., Post, A.F., Scanlan, D.J., 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., Casamayor, E.O., Kirchman, D.L., Lovejoy, C., 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., Tamames, J., Linz, A.M., Pedrós-Alió, C., Puente-Sánchez, F., 2019. Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. The ISME Journal 13, 2969–2983.

Garczarek, L., Dufresne, A., Blot, N., Cockshutt, A.M., Peyrat, A., Campbell, D.A., Joubin, L., Six, C., 2008. Function and evolution of the *psbA* gene family in marine *Synechococcus*: *Synechococcus* sp. WH7803 as a case study. The ISME Journal.

Gérikas Ribeiro, C., Lopes dos Santos, A., Marie, D., Pereira Brandini, F., Vaulot, D., 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.

Giner, C.R., Pernice, M.C., Balagué, V., Duarte, C.M., Gasol, J.M., Logares, R., Massana, R., 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., Britschgi, T.B., Moyer, C.L., Field, K.G., 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature 345, 60–63.

Gong, W., Marchetti, A., 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., McPherson, J.D., Richard McCombie, W., 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 1. Oceanographic Research Papers 40, 517–545.

Gowing, M.M., Coale, S.L., 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-Stadniczeňko, S., Šupraha, L., Egge, E.D., Edvardsen, B., 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., Slack, G., Daniels, E.F., Selph, K.E., Palenik, B., Landry, M.R., 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., Stukel, M.R., Lopes dos Santos, A., Biard, T., Scharek, R., Vaulot, D., Landry, M.R., Not, F., 2019. High contribution of rhizaria (radiolaria) to vertical export in the California current ecosystem revealed by DNA metabarcoding. The ISME Journal 13, 964–976.

Hardge, K., Peeken, I., Neuhaus, S., Lange, B.A., Stock, A., Stoeck, T., Weinisch, L., Metfies, K., 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., Krull, E.S., Hartley, D.M., Oliver, R.L., 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., Gruber, N., Münnich, M., 2020. Sea-ice induced Southern Ocean subsurface warming and surface cooling in a warming climate. AGU Advances 1, e2019AV000132.

Henson, S., Le Moigne, F., Giering, S., 2019. Drivers of carbon export efficiency in the global ocean. Global Biogeochemical Cycles 33, 891–903.

Herdman, M., Castenholz, R.W., Waterbury, J.B., Rippka, R., 2015. Form-Synechococcus. Bergey's Manual of Systematics of Archaea and Bacteria, pp. 1–7.

Herndl, G.J., Reinthaler, T., 2013. Microbial control of the dark end of the biological pump. Nature Geoscience 6, 718–724.

Hess, W.R., Weihe, A., Loiseaux-de Goër, S., Partensky, F., Vaulot, D., 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., Eglinton, T., Taylor, C., Ulmer, K., Sievert, S., Bracher, A., German, C., Edgcomb, V., Francois, R., Iglesias-Rodriguez, M.D., Van Mooy, B., Rapeta, D., 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., Flintrop, C.M., Rost, B., 2018. The Arctic picoeukaryote *Micromonas pusilla* benefits synergistically from warming and ocean acidification. Biogeosciences 15, 4353–4365.

Horvat, C., Jones, D.R., Iams, S., Schroeder, D., Flocco, D., Feltham, D., 2017. The frequency and extent of sub-ice phytoplankton blooms in the Arctic Ocean. Science Advances 3, e1601191.

Hu, S.K., Campbell, V., Connell, P., Gellene, A.G., Liu, Z., Terrado, R., Caron, D.A., 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., Wilhelm, S.W., Harvey, H.R., Taylor, K., Jiao, N., Chen, F., 2012. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. The ISME Journal 6, 285–297.

Huber, P., Cornejo-Castillo, F.M., Ferrera, I., Sánchez, P., Logares, R., Metz, S., Balagué, V., Acinas, S.G., Gasol, J.M., Unrein, F., 2019. Primer design for an accurate view of picocyanobacterial community structure by using high-throughput sequencing. Applied and Environmental Microbiology 85, e02659.

Huse, S.M., Mark Welch, D.B., Voorhis, A., Shipunova, A., Morrison, H.G., Eren, A.M., Sogin, M.L., 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". United States National Marine Fisheries Service. Fishery Bulletin 70, 1053–1059.

Jain, R., Rivera, M.C., Lake, J.A., 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., Foster, R., Barbera, P., Czech, L., Kozlov, A., Stamatakis, A., Bending, G., Hilton, S., Bass, D., Burki, F., 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., Hölemann, J., Waite, A.M., Krumpen, T., von Appen, W.J., Martynov, F., 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., Zehr, J.P., Gibson, A., Campbell, L., 2006. Cyanobacterial assimilatory nitrate reductase gene diversity in coastal and oligotrophic marine environments. Environmental Microbiology 8, 2083–2095.

Jin, X., Gruber, N., Dunne, J.P., Sarmiento, J.L., Armstrong, R.A., 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., Monier, A., Logares, R., Lovejoy, C., 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., Konzack, M., Eisenhauer, N., Heintz-Buschart, A., 2020. The archives are half-empty: an assessment of the availability of microbial community sequencing data. Communications Biology 3, 474.

Kaczmarska, I., Mather, L., Aluddington, I., Muise, F., Ehrman, J.M., 2014. Cryptic diversity in a cosmopolitan diatom known as *Asterionellopsis glacialis* (Fragilariaceae): im-

plications for ecology, biogeography, and taxonomy. American Journal of Botany 101, 267–286.
Kalenitchenko, D., Joli, N., Potvin, M., Tremblay, J., Lovejoy, C., 2019. Biodiversity and species change in the Arctic Ocean: a view through the lens of nares strait. Frontiers in Marine Science 6, 1–17.
Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., Follows, M.J., Stepanauskas, R., Chisholm, S.W., 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science 344, 416–420.
Kaunisto, K.M., Roslin, T., Sääksjärvi, I.E., Vesterinen, E.J., 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., Franc, A., Rimet, F., Chaumeil, P., Frigerio, J.-M., Humbert, J.-F., Bouchez, A., 2014. A next-generation sequencing approach to river biomonitoring using benthic diatoms. Freshwater Science 33, 349–363.
Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G.M., Richardson, P., Chisholm, S.W., 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genetics 3, e231.
Kilias, E.S., Peeken, I., Metfies, K., 2014. Insight into protist diversity in Arctic sea ice and melt-pond aggregate obtained by pyrosequencing. Polar Research 33, 23466.
King, R.A., Read, D.S., Traugott, M., Symondson, W.O.C., 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., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. Journal of Molecular Evolution 31, 151–160.
Kitahara, K., Miyazaki, K., 2013. Revisiting bacterial phylogeny. Mobile Genetic Elements 3, e24210.
Knauer, G.A., Martin, J.H., Bruland, K.W., 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., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 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., Garcia-Robledo, C., Uriarte, M., Erickson, D.L., 2015. DNA barcodes for ecology, evolution, and conservation. Trends in Ecology & Evolution 30, 25–35.
Kysela, D.T., Palacios, C., Sogin, M.L., 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., Salter, I., Johns, D., 2009. Radiolaria: major exporters of organic carbon to the deep ocean. Global Biogeochemical Cycles 23, 1–9.
Lampitt, R.S., Wishner, K.F., Turley, C.M., Angel, M.V., 1993. Marine snow studies in the Northeast Atlantic-Ocean - distribution, composition and role as a food source for migrating plankton. Marine Biology 116 (4), 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., 1991. 16S/23S rRNA sequencing. In: Stackebrandt, E., Goodfellow, M. (Eds.), Nucleic Acid Techniques in Bacterial Systematics. John Wiley and Sons, New York, N.Y., pp. 115–175.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., Pace, N.R., 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., Jørgensen, S.L., Huson, D.H., Gorfer, M., Grindhaug, S.H., Jonassen, I., Øvreås, L., Urich, T., 2012. CREST - classification resources for environmental sequence tags. PLoS ONE 7.

Le Moigne, F.A.C., Pabortsava, K., Marcinko, C.L.J., Martin, P., Sanders, R.J., 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., DeBruyn, J.M., Maas, E.W., Boyd, P.W., Wilhelm, S.W., 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.

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J.M., Zuccarello, G.C., Clerck, O.D., 2014. DNA-based species delimitation in algae DNA. European Journal of Phycology 49, 179–196.

Leray, M., Meyer, C.P., Mills, S.C., 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., Søreide, J.E., Hessen, D.O., Falk-Petersen, S., Berge, J., 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.

Lin, S., 2011. Genomic understanding of dinoflagellates. Research in Microbiology 162, 551–569.

Lindell, D., Padan, E., Post, A.F., 1998. Regulation of ntcA expression and nitrite uptake in the marine *Synechococcus* sp. strain WH7803. Journal of Bacteriology 180, 1878–1886. https://doi.org/10.1128/JB.180.7.1878-1886.1998.

Litchman, E., Klausmeier, C.A., 2008. Trait-based community ecology of phytoplankton. Annual Review of Ecology, Evolution, and Systematics 39, 615–639.

Liu, Z., Cochran, J.K., Lee, C., Gasser, B., Miquel, J.C., Wakeham, S.G., 2009. Further investigations on why POC concentrations differ in samples collected by Niskin Bottle and in situ pump. Deep-Sea Research. Part 2. Topical Studies in Oceanography 56, 1558–1567.

Lockhart, P., Penny, D., 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 III, 499–505.

Logares, R., et al., 2014. Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24, 813–821.

Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.B., Rubinat-Ripoll, L., Mestre, M., Salazar, G., Ruiz-González, C., Sebastián, M., de Vargas, C., Acinas, S.G., Duarte, C.M., Gasol, J.M., Massana, R., 2020. Disentangling the mechanisms shaping the surface ocean microbiota. Microbiome 8, 55.

Logares, R., Mangot, J.-F., Massana, R., 2015. Rarity in aquatic microbes: placing protists on the map. Research in Microbiology 166, 831–841.

Lopes dos Santos, A., Pollina, T., Gourvil, P., Corre, E., Marie, D., Garrido, J.L., Rodríguez, F., Noël, M.-H., Vaulot, D., Eikrem, W., 2017. Chloropicophyceae, a new class of picophytoplanktonic prasinophytes. Scientific Reports 7, 14019.

López-García, P., Philippe, H., Gail, F., Moreira, D., 2003. Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. Proceedings of the National Academy of Sciences of the United States of America 100 (2), 697–702.

López-García, P., Rodriguez-Valera, F., Pedrós-Alió, C., Moreira, D., 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409, 603–607.
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550.
Lovejoy, C., Massana, R., Pedros-Alio, C., 2006. Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. Applied and Environmental Microbiology 72, 3085–3095.
Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct comparisons of illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS ONE 7, e30087.
Lutz, S., McCutcheon, J., McQuaid, J.B., Benning, L.G., 2018. The diversity of ice algal communities on the Greenland ice sheet as revealed by oligotyping. Microbial Genomics 4, 1–10.
Ma, Y., Jiao, N.Z., Zeng, Y.H., 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, K.R., Paytan, A., Caldeira, K., Grossman, A.R., Moran, D., Mcilvin, M., Saito, M.A., 2013. Effect of temperature on photosynthesis and growth in marine *Synechococcus* spp. Plant Physiology 163, 815–829.
Mahé, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M., 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.
Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., Chisholm, S.W., 2010. Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific Oceans. The ISME Journal 4, 1252–1264.
Maloy, A.P., Culloty, S.C., Slater, J.W., 2013. Dietary analysis of small planktonic consumers: a case study with marine bivalve larvae. Journal of Plankton Research 35, 866–876.
Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A., Bowler, C., 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., Philosof, A., Kirkup, B.C., Le Gall, F., Yogev, T., Berman-Frank, I., Polz, M.F., Vaulot, D., Béjà, O., 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., Cook, A., Millard, A., Bailey, S., Clokie, M., 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature 424, 741.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10.
Martin, P., Allen, J.T., Cooper, M.J., Johns, D.G., Lampitt, R.S., Sanders, R., Teagle, D.A.H., 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., Tai, A.P., Veneziano, D., Primeau, F., Chisholm, S.W., 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., Kodner, R.B., Armbrust, E.V., 2010. Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11, 538.

Mazard, S., Ostrowski, M., Partensky, F., Scanlan, D.J., 2012. Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. Environmental Microbiology 14, 372–386.

McDonnell, A.M., Boyd, P.W., Buesseler, K.O., 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., Holmes, S., 2013. Phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8, e61217.

Medlin, L., Elwood, H.J., Stickel, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71, 491–499.

Mella-Flores, D., Six, C., Ratin, M., Partensky, F., Boutte, C., Le Corguillé, G., Marie, D., Blot, N., Gourvil, P., Kolowrat, C., Garczarek, L., 2012. *Prochlorococcus* and *Synechococcus* have evolved different adaptive mechanisms to cope with light and Uv stress. Frontiers in Microbiology 3, 285.

Mestre, M., Ruiz-gonzález, C., Logares, R., Duarte, C.M., Gasol, J.M., 2018. Sinking particles promote vertical connectivity in the ocean microbiome. Proceedings of the National Academy of Sciences of the United States of America 115 (29), 6799–6807.

Metfies, K., Bauerfeind, E., Wolf, C., Sprong, P., Frickenhaus, S., Kaleschke, L., Nicolaus, A., Nöthig, E.-M., 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., Vartanian, J.P., Wain-Hobson, S., 1990. DNA recombination during PCR. Nucleic Acids Research 18, 1687–1691.

Michaels, A.F., Caron, D.A., Swanberg, N.R., Howse, F.A., Michaels, C.M., 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., Silver, M.W., 1988. Primary production, sinking fluxes and the microbial food web. Deep-Sea Research. Part A, Oceanographic Research Papers 35, 473–490.

Millard, A., Clokie, M.R., Shub, D.A., Mann, N.H., 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., Lovejoy, C., Orsini, L., Procaccini, G., Roy, S., 2003. Bipolar distribution of the cyst-forming dinoflagellate polarella glacialis. Polar Biology 26, 186–194.

Moon-van der Staay, S.Y., De Wachter, R., Vaulot, D., 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409, 607–610.

Moore, L.R., Rocap, G., Chisholm, S.W., 1998. Physiology and molecular phytogeny of coexisting *Prochlorococcus* ecotypes. Nature 393, 464–467.

Mordret, S., Piredda, R., Vaulot, D., Montresor, M., Kooistra, W.H.C.F., Sarno, D., 2018. Dinoref: a curated dinoflagellate (dinophyceae) reference database for the 18S rRNA gene. Molecular Ecology Resources 18, 974–987.

Muhling, M., Fuller, N.J., Millard, A., Somerfield, P.J., Marie, D., Wilson, W.H., Scanlan, D.J., Post, A.F., Joint, I., Mann, N.H., 2005. Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. Environmental Microbiology 7, 499–508.

Mühling, M., Fuller, N.J., Somerfield, P.J., Post, A.F., Wilson, W.H., Scanlan, D.J., Joint, I., Mann, N.H., 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., Tuji, A., Makino, W., Matsuzaki, S.-i.S., Nagata, N., Nakagawa, M., Noriko, Takamura, 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., Fuhrman, J.A., 2016. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nature Microbiology 1, 16005.
Needham, D.M., Sachdeva, R., Fuhrman, J.A., 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. The ISME Journal 11, 1614–1629.
Nelson, D.M., Tréguer, P., Brzezinski, M.A., Leynaert, A., Quéguiner, B., 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., Michel, C., Hille, K., Poulin, M., 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., Gausling, R., Azam, F., Heidelberg, J.F., Worden, A.Z., 2007b. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. Environmental Microbiology 9, 1233–1252.
Not, F., Valentin, K., Romari, K., Lovejoy, C., Massana, R., Töbe, K., Vaulot, D., Medlin, L.K., 2007a. Picobiliphytes, a new marine picoplanktonic algal group with unknown affinities to other eukaryotes. Science 315, 252–254.
Oksanen, A.J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., Hara, R.B.O., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., 2016. Package 'Vegan' (Version 2.4-0).
Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R., Stahl, D.A., 1986. Microbial ecology and evolution: a ribosomal RNA approach. Annual Review of Microbiology 40, 337–365.
Olsen, L.M., Laney, S.R., Duarte, P., Kauko, H.M., Fernández-méndez, M., Mundy, C.J., Rösel, A., Meyer, A., Itkin, P., Cohen, L., Peeken, I., Tatarek, A., Różańska-Pluta, M., Wiktor, J., Taskjelle, T., Pavlov, A.K., Hudson, S.R., Granskog, M.A., Hop, H., Assmy, P., 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., Vaulot, D., Chisholm, S.W., 1985. Marine phytoplankton distributions measured using shipboard flow cytometry. Deep-Sea Research. Part A, Oceanographic Research Papers 32, 1273–1280.
O'Rorke, R., Lavery, S.D., Wang, M., Nodder, S.D., Jeffs, A.G., 2013. Determining the diet of larvae of the red rock lobster (*Jasus edwardsii*) using high-throughput DNA sequencing techniques. Marine Biology 161, 551–563.
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., Foster, R.A., Jenkins, B.D., Montoya, J.P., Zehr, J.P., 2008. Phylogenetic diversity of cyanobacterial *narB* genes from various marine habitats. Environmental Microbiology 10, 3377–3387.
Paerl, R.W., Johnson, K.S., Welsh, R.M., Worden, A.Z., Chavez, F.P., Zehr, J.P., 2011. Differential distributions of *Synechococcus* subgroups across the California current system. Frontiers in Microbiology 2, 1–22.
Paerl, R.W., Turk, K.A., Beinart, R.A., Chavez, F.P., Zehr, J.P., 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.

Parada, A.E., Needham, D.M., Fuhrman, J.A., 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities. Time series and global field samples. Environmental Microbiology.

Partensky, F., Blanchot, J., Vaulot, D., 1999. Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. Marine Cyanobacteria. Monaco.

Pawłowska, J., Wollenburg, J.E., Zajaczkowski, M., Pawlowski, J., 2020. Planktonic foraminifera genomic variations reflect paleoceanographic changes in the Arctic: evidence from sedimentary ancient DNA. Scientific Reports 10, 15102.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., Wilding, T.A., 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., Esling, P., Lejzerowicz, F., Cordier, T., Visco, J., Martins, C., Kvalvik, A., Staven, K., Cedhagen, T., 2016. Benthic monitoring of salmon farms in Norway using foraminiferal metabarcoding. Aquaculture Environment Interactions 8, 371–386.

Pedrós-Alió, C., Acinas, S.G., Logares, R., Massana, R., 2018. Marine microbial diversity as seen by high-throughput sequencing. Microbial Ecology of the Oceans, 47–97.

Penno, S., Lindell, D., Post, A.F., 2006. Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. Environmental Microbiology.

Pérez-Cobas, A.E., Gomez-Valero, L., Buchrieser, C., 2020. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. Microbial Genomics, mgen000409.

Perovich, D.K., Richter-Menge, J.A., 2009. Loss of sea ice in the Arctic. Annual Review of Marine Science 1, 417–441.

Petchey, O.L., Gaston, K.J., 2006. Functional diversity: back to basics and looking forward. Ecology Letters 9, 741–758.

Peterson, M.L., Wakeham, S.G., Lee, C., Askea, M.A., Miquel, J.C., 2005. Novel techniques for collection of sinking particles in the ocean and determining their settling rates. Limnology and Oceanography: Methods 3, 520–532.

Piredda, R., Tomasino, M.P., D'Erchia, A.M., Manzari, C., Pesole, G., Montresor, M., Kooistra, W.H.C.F., Sarno, D., Zingone, A., 2017. Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean long term ecological research site. FEMS Microbiology, Ecology 93, fiw200.

Pittera, J., Humily, F., Thorel, M., Grulois, D., Garczarek, L., Six, C., 2014. Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. The ISME Journal.

Porazinska, D.L., Giblin-Davis, R.M., Sung, W., Thomas, W.K., 2012. The nature and frequency of chimeras in eukaryotic metagenetic samples. Journal of Nematology 44, 18–25.

Post, A.F., Penno, S., Zandbank, K., Paytan, A., Huse, S.M., Welch, D.M., 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., Durkin, C.A., Yamahara, K.M., 2019. DNA metabarcoding reveals organisms contributing to particulate matter flux to abyssal depths in the North East Pacific Ocean. Deep-Sea Research. Part 2. Topical Studies in Oceanography 173, 104708.

Preston, C.M., Durkin, C.A., Yamahara, K.M., 2020. DNA metabarcoding reveals organisms contributing to particulate matter flux to abyssal depths in the North East Pacific ocean. Deep-Sea Research. Part 2. Topical Studies in Oceanography 173, 104708.

Puigcorbé, V., Benitez-Nelson, C.R., Masqué, P., Verdeny, E., White, A.E., Popp, B.N., Prahl, F.G., Lam, P.J., 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., Ruiz-González, C., Masqué, P., Gasol, J.M., 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. https://doi.org/10.3389/fmicb.2020.01645.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41, D590.
Ramond, P., Sourisseau, M., Simon, N., Romac, S., Schmitt, S., Rigaut-Jalabert, F., Henry, N., de Vargas, C., Siano, R., 2019. Coupling between taxonomic and functional diversity in protistan coastal communities. Environmental Microbiology.
Ray, J.L., Skaar, K.S., Simonelli, P., Larsen, A., Sazhin, A., Jakobsen, H.H., Nejstgaard, J.C., Troedsson, C., 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., Lévy, M., McGillicuddy Jr., D.J., 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., a Jackson, G., 2007. Small phytoplankton and carbon export from the surface ocean. Science 315, 838–840.
Rocap, G., Distel, D.L., Waterbury, J.B., Chisholm, S.W., 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., Rocap, G., Salazar, G., Massana, R., 2013. Biogeography of the uncultured marine picoeukaryote MAST-4: temperature-driven distribution patterns. The ISME Journal 7, 1531–1543.
Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.
Rozema, P.D., Venables, H.J., van de Poll, W.H., Clarke, A., Meredith, M.P., Buma, A.G.J., 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., Austin, C.C., 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., Flickinger, S.A., Fontaine, D.N., 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.
Rynearson, T.A., Richardson, K., Lampitt, R.S., Sieracki, M.E., Poulton, A.J., Lyngsgaard, M.M., Perry, M.J., 2013. Major contribution of diatom resting spores to vertical flux in the sub-polar North Atlantic. Deep-Sea Research. Part 1. Oceanographic Research Papers 82, 60–71.
Samanta, B., Bhadury, P., 2016. A comprehensive framework for functional diversity patterns of marine chromophytic phytoplankton using rbcL phylogeny. Scientific Reports 6, 20783.
Santoferrara, L., Burki, F., Filker, S., Logares, R., Dunthorn, M., McManus, G.B., 2020. Perspectives from ten years of protist studies by high-throughput metabarcoding. Journal of Eukaryotic Microbiology 67, 612–622.

Sarthou, G., Timmermans, K.R., Blain, S., Tréguer, P., 2005. Growth physiology and fate of diatoms in the ocean: a review. Journal of Sea Research 53, 25–42.
Sautter, L.R., Thunell, R.C., 1989. Seasonal succession of planktonic foraminifera; results from a four-year time-series sediment trap experiment in the northeast Pacific. Journal of Foraminiferal Research 19, 253–267.
Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., Post, A.F., Hagemann, M., Paulsen, I., Partensky, F., 2009. Ecological genomics of marine picocyanobacteria. Microbiology and Molecular Biology Reviews 73, 249–299.
Scanlan, D.J., West, N.J., 2002. Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. FEMS Microbiology, Ecology 40, 1–12.
Schander, C., Kenneth, H.M., 2003. DNA, PCR and formalinized animal tissue – a short review and protocols. Organisms Diversity & Evolution 3, 195–205.
Scharek, R., Tupas, L.M., Karl, D.M., 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., Handelsman, J., 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Applied and Environmental Microbiology 71, 1501–1506.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75, 7537–7541.
Schneider, L., Anestis, K., Mansour, J., Anschütz, A., Gypens, N., Hansen, P., John, U., Klemm, K., Martin, J., Medic, N., Not, F., Stolte, W., 2020. A dataset on trophic modes of aquatic protists. Biodiversity Data Journal 8.
Schwarz, D., Spitzer, S.M., Thomas, A.C., Kohnert, C.M., Keates, T.R., Acevedo-Gutierrez, A., 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., Matsuzaki, R., Takeuchi, N., Akiyoshi, A., Navarro, F., Sugiyama, S., Yonezawa, T., Mori, H., 2018. Bipolar dispersal of red-snow algae. Nature Communications 9.
Serreze, M.C., Holland, M.M., Stroeve, J., 2007. Perspectives on the Arctic's shrinking sea-ice cover. Science 315, 1533–1536.
Shadwick, E.H., Trull, T.W., Thomas, H., Gibson, J.A.E., 2013. Vulnerability of polar oceans to anthropogenic acidification: comparison of Arctic and Antarctic seasonal cycles. Scientific Reports 3, 2339.
Shi, X.L., Lepère, C., Scanlan, D.J., Vaulot, D., 2011. Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. PLoS ONE 6, e18979.
Shiozaki, T., Itoh, F., Hirose, Y., Onodera, J., Kuwata, A., Harada, N., 2021. A DNA metabarcoding approach for recovering plankton communities from archived samples fixed in formalin. PLoS ONE 16, e0245936.
Sieburth, J.M., Smetacek, V., Lenz, J., 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., Buesseler, K.O., Doney, S.C., Sailley, S.F., Behrenfeld, M.J., Boyd, P.W., 2014. Global assessment of ocean carbon export by combining satellite observations and food-web models. Global Biogeochemical Cycles 28, 181–196.
Smetacek, V., et al., 2012. Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. Nature 487, 313–319.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 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.

Soto, D.F., Fuentes, R., Huovinen, P., Gómez, I., 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., Trull, T.W., Bodrossy, L., 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., Goebel, B.M., 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.

Steglich, C., Post, A.F., Hess, W.R., 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., Landry, M.R., 2017. Zooplankton and the ocean carbon cycle. Annual Review of Marine Science 9, 413–444.

Stemmann, L., Boss, E., 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., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.W., Richards, T.A., 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., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M.J., Chistoserdov, A., Orsi, W., Edgcomb, V.P., 2009. Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. BMC Biology 7, 72.

Stoeck, T., Taylor, G.T., Epstein, S.S., 2003. Novel eukaryotes from the permanently anoxic cariaco basin (Caribbean Sea). Applied and Environmental Microbiology 69, 5656–5663.

Stone, J.P., Steinberg, D.K., 2016. Salp contributions to vertical carbon flux in the Sargasso Sea. Deep-Sea Research. Part 1. Oceanographic Research Papers 113, 90–100.

Stroeve, J.C., Markus, T., Boisvert, L., Miller, J., Barrett, A., 2014. Changes in Arctic melt season and implications for sea ice loss. Geophysical Research Letters 41, 1216–1225.

Stukel, M.R., Aluwihare, L.I., Barbeau, K.A., Chekalyuk, A.M., Goericke, R., Miller, A.J., Ohman, M.D., Ruacho, A., Song, H., Stephens, B.M., Landry, M.R., 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., Décima, M., Selph, K.E., Taniguchi, D.A.A., Landry, M.R., 2013. The role of *Synechococcus* in vertical flux in the Costa Rica upwelling dome. Progress in Oceanography 112-113, 49–59.

Stukel, M.R., Landry, M.R., Benitez-Nelson, C.R., Goericke, R., 2011. Trophic cycling and carbon export relationships in the California current ecosystem. Limnology and Oceanography 56, 1866–1878.

Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zettler, L.A., Sogin, M.L., 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.

Tai, V., Palenik, B., 2009. Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. The ISME Journal.

Thompson, J.R., Hatfield, R.G., Batista, F.M., Bean, T.P., Fonseca, V.G., 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. Frontiers in Microbiology 11, 1–16.

Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., Martinez, N.D., Rantala, H., Romanuk, T.N., Stouffer, D.B., Tylianakis, J.M., 2012. Food webs: reconciling the structure and function of biodiversity. Trends in Ecology & Evolution 27, 689–697.

Ting, C.S., Rocap, G., King, J., Chisholm, S.W., 2002. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. Trends in Microbiology 10, 134–142.

Toledo, G., Palenik, B., 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., Vaulot, D., 2018. Green microalgae in marine coastal waters: the ocean sampling day (OSD) dataset. Scientific Reports 8, 14020.

Tragin, M., Vaulot, D., 2019. Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding. Scientific Reports 9, 5190.

Trefault, N., De la Iglesia, R., Moreno-Pino, M., Lopes dos Santos, A., Gérikas Ribeiro, C., Parada-Pozo, G., Cristi, A., Marie, D., Vaulot, D., 2021. Annual phytoplankton dynamics in coastal waters from Fildes Bay, Western Antarctic Peninsula. Scientific Reports 11, 1368.

Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L., Dugdale, R., Finkel, Z., Iudicone, D., Jahn, O., Guidi, L., Lasbleiz, M., Leblanc, K., Levy, M., Pondaven, P., 2018. Influence of diatom diversity on the ocean biological carbon pump. Nature Geoscience 11, 27–37.

Turner, J., Colwell, S.R., Marshall, G.J., Lachlan-Cope, T.A., Carleton, A.M., Jones, P.D., Lagun, V., Reid, P.A., Iagovkina, S., 2005. Antarctic climate change during the last 50 years. International Journal of Climatology 25, 279–294.

Turner, J.T., 2015. Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. Progress in Oceanography 130, 205–248.

Twining, B.S., Rauschenberg, S., Morton, P.L., Ohnemus, D.C., Lam, P.J., 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 2. Topical Studies in Oceanography 116, 273–282.

Urbach, E., Robertson, D.L., Chisholm, S.W., 1992. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. Nature 355, 267–270.

Urbach, E., Scanlan, D.J., Distel, D.L., Waterbury, J.B., Chisholm, S.W., 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., Tedesco, L., Arrigo, K.R., Assmy, P., Campbell, K., Meiners, K.M., Rintala, J.-M., Selz, V., Thomas, D.N., Stefels, J., Deming, J.W., Leeuwe, M.A.V., Tedesco, L., Arrigo, K.R., Assmy, P., Meiners, K.M., Rintala, J.-M., Selz, V., David, N., Stefels, J., 2018. Microalgal community structure and primary production in Arctic and Antarctic sea ice: a synthesis. Elementa: Science of the Anthropocene 6, 1–25.

Vaulot, D., Eikrem, W., Viprey, M., Moreau, H., 2008. The diversity of small eukaryotic phytoplankton (<3 μm) in marine ecosystems. FEMS Microbiology Reviews 32, 795–820.

Vaulot, D., Geisen, S., Mahé, F., Bass, D., 2021. Pr2-Primers: an 18S rRNA Primer Database for Protists. Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13465. In press.

Venter, J.C., et al., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.

Vernette, C., Henry, N., Lecubin, J., de Vargas, C., Hingamp, P., Lescot, M., 2021. The ocean barcode atlas: a web service to explore the biodiversity and biogeography of marine organisms. Molecular Ecology Resources 21 (4), 1347–1358.

Vestheim, H., Jarman, S.N., 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., Priscu, J.C., Amaral-Zettler, L.A., 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., Apothéloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillet, L., Pawlowski, J., 2015. Environmental monitoring: inferring the diatom index from next-generation sequencing data. Environmental Science & Technology 49, 7597–7605.

Volk, T., Hoffert, M.I., 1985. Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes. In: The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present. In: Geophysical Monograph Series, vol. 32, pp. 99–110.

Waite, A., Bienfang, P.K., Harrison, P.J., 1992. Spring bloom sedimentation in a subarctic ecosystem. Marine Biology 114, 131–138.

Waite, A.M., Safi, K.A., Hall, J.A., Nodder, S.D., 2000. Mass sedimentation of picoplankton embedded in organic aggregates. Limnology and Oceanography.

Wang, G.C.Y., Wang, Y., 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., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73, 5261–5267.

Waterbury, J., Rippka, R., 1989. The order Chroococcales. Bergey's Manual of Systematic Bacteriology 3, 1728–1746.

Waterbury, J.B., Willey, J.M., Franks, D.G., Valois, F.W., Watson, S.W., 1985. A cyanobacterium capable of swimming motility. Science 230, 74–76.

West, N.J., Scanlan, D.J., 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., Baker, C.C.M., Salzman, S.M., Martins, D.J., Pierce, N.E., 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., Fox, G.E., 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., Kilias, E., Metfies, K., 2015. Protists in the polar regions: comparing occurrence in the Arctic and Southern Oceans using pyrosequencing. Polar Research 34, 23225.

Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., Keeling, P.J., 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.

Xu, D., Kong, H., Yang, E.-j., Li, X., Jiao, N., Warren, A., Wang, Y., Lee, Y., Jung, J., Kang, S.-h., 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., Toyofuku, T., Takishita, K., 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., Lopes dos Santos, A., Eikrem, W., Gérikas Ribeiro, C., Gourvil, P., Balzano, S., Escande, M.-L., Moreau, H., Vaulot, D., 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., Viet, H.Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., Pavloudi, C., Pafilis, E., 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., Novotny, A., Winder, M., 2020. Reconstructing marine plankton food web interactions using DNA metabarcoding. Molecular Ecology 29, 3380–3395.

Zeidner, G., Preston, C.M., Delong, E.F., Massana, R., Post, A.F., Scanlan, D.J., Beja, O., 2003. Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses. Environmental Microbiology 5, 212–216.

Zeldis, J., James, M.R., Grieve, J., Richards, L., 2002. Omnivory by copepods in the New Zealand subtropical frontal zone. Journal of Plankton Research 24, 9–23.

Zhu, F., Massana, R., Not, F., Marie, D., Vaulot, D., 2005. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiology, Ecology 52, 79–92.

Zimmermann, J., Glöckner, G., Jahn, R., Enke, N., Gemeinholzer, B., 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Molecular Ecology Resources 15, 526–542.

Zuckerkandl, E., Pauling, L., 1965. Molecules as documents of evolutionary history. Journal of Theoretical Biology 8, 357–366.

Zwirglmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M., Zubkov, M.V., Scanlan, D.J., 2007. Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. Environmental Microbiology 9, 1278–1290.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczareck, L., Vaulot, D., Not, F., Massana, R., Ulloa, O., Scanlan, D.J., 2008. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages amongst oceanic biomes. Environmental Microbiology 10, 147–161.