



# Small eukaryotic phytoplankton communities in tropical waters off Brazil are dominated by symbioses between Haptophyta and nitrogen-fixing cyanobacteria

Catherine G erikas Ribeiro<sup>1,2</sup> · Adriana Lopes dos Santos<sup>1,3</sup> · Dominique Marie<sup>1</sup> · Frederico Pereira Brandini<sup>2</sup> · Daniel Vaultot<sup>1</sup>

Received: 22 May 2017 / Revised: 1 November 2017 / Accepted: 20 December 2017

  International Society for Microbial Ecology 2018

## Abstract

Symbioses between eukaryotic algae and nitrogen-fixing cyanobacteria have been recognized in recent years as a key source of new nitrogen in the oceans. We investigated the composition of the small photosynthetic eukaryote communities associated with nitrogen-fixing cyanobacteria in the Brazilian South Atlantic Bight using a combination of flow cytometry sorting and high throughput sequencing of two genes: the V4 region of 18S rRNA and *nifH*. Two distinct eukaryotic communities were often encountered, one dominated by the Mamiellophyceae *Bathycoccus* and *Ostreococcus*, and one dominated by a prymnesiophyte known to live in symbiosis with the UCYN-A1 nitrogen-fixing cyanobacterium. Among *nifH* sequences, those from UCYN-A1 were most abundant but three other UCYN-A clades (A2, A3, A4) were also found. Network analysis confirmed the relation between A1 and A2 clades and their hypothesized hosts and pointed out to the potential association between novel clade A4 with *Braarudosphaera bigelowii*, previously hypothesized to host A2.

## Introduction

Small photosynthetic eukaryotes [1] are key component of the biomass and primary production in marine ecosystems [2–4]. In coastal waters, these small photosynthetic eukaryotes are often dominated by green algae from Mamiellophyceae [5], whereas in oceanic waters they are much more diverse, including green algae (Mamiellophyceae,

prasinophytes clade VII) but also pelagophytes, chrysophytes, and prymnesiophytes [3, 6, 7]. While for some groups such as Mamiellophyceae, prasinophytes clade VII, or pelagophytes, numerous representatives have been isolated in culture [8], this is not the case for chrysophytes or prymnesiophytes. Initially, these small eukaryotes were thought to be purely phototrophs, but recently novel types of trophic modes and associations have been discovered. In particular, quite a few appear to be mixotrophs, such as *Micromonas* in polar regions [9], and a new type of symbiosis has been found between a small haptophyte alga and a nitrogen-fixing cyanobacterium UCYN-A [10].

Nitrogen (N<sub>2</sub>) fixation, performed by diazotrophs, is an anaerobic process catalyzed by the nitrogenase enzyme, which produces biologically available ammonium (NH<sub>4</sub><sup>+</sup>). It has a strong impact on primary productivity in past [11], present [12], and future oceans [13]. *Trichodesmium* sp. was initially identified as the most important marine diazotroph [14, 15]. However, recent discoveries have set the ground for a paradigm shift in our understanding of N<sub>2</sub> fixing players regarding their diversity, ecological strategies, geographic distribution, and impact on the primary productivity in the global ocean ([16–18] and references therein). Two decades ago, an uncultivated unicellular cyanobacterium UCYN-A was discovered by PCR

Catherine G erikas Ribeiro and Adriana Lopes dos Santos contributed equally to the paper.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s41396-018-0050-z>) contains supplementary material, which is available to authorized users.

✉ Daniel Vaultot  
vaultot@sb-roscoff.fr

- <sup>1</sup> Sorbonne Universit , CNRS, UMR7144, Station Biologique de Roscoff, Roscoff, France
- <sup>2</sup> Departamento de Oceanografia Biol gica, Instituto Oceanogr fico, Universidade de S o Paulo, S o Paulo, Brazil
- <sup>3</sup> GEMA Center for Genomics, Ecology & Environment, Universidad Mayor, Camino La Pir mide 5750, Huechuraba, Santiago, Chile

amplification of the nitrogenase gene *nifH* [19]. The genome of UCYN-A was found to be extremely streamlined, lacking the genes involved in the oxygen-evolving pathway, CO<sub>2</sub> fixation [20], and biosynthetic pathways of several amino acids and purines [21]. Such modifications in the genome, resembling the ones observed in cellular organelles, led to the hypothesis of a symbiotic partnership [21]. This was confirmed later by the identification of an association between UCYN-A and an apparently non-calcifying prymnesiophyte microalga [10], closely related to the coccolithophorid *Braarudosphaera bigelowii*, which bears pentagon-shaped coccoliths (pentaliths), and to the non-calcifying haptophyte *Chrysochromulina parkeae* [10]. Recent data suggest that this symbiosis could be responsible for a major part of the oceanic nitrogen fixation [22].

UCYN-A appears to be quite diversified with at least six clades (A1 to A6), which forms a monophyletic group with *Crocospaera* sp. and *Cyanothece* sp. [16, 23–25]. The hosts of the two major clades (A1 and A2) appear to have different cell sizes, to harbor a different number of symbionts [16, 26], and present distinct oceanic distributions, the former being less widespread but with higher relative abundance peaks than the latter [27].

Many open questions remain concerning these symbioses, including the ecological distribution of the hosts and their importance within the phytoplankton community. The South Atlantic Bight off Brazil is reported to have a high UCYN-A abundance in comparison to other oceanic regions [27] and also appears to be a hot spot of nitrogen fixation [28]. We investigated pico and nano-phytoplankton composition in this area using flow cytometry sorting [29] combined with parallel high throughput sequencing of the V4 region of the 18S rRNA and of a *nifH* gene fragment. Our data revealed two major types of small photosynthetic eukaryote communities, one dominated by green algae and the other dominated by the UCYN-A1 host. By analyzing the *nifH* diversity in the same samples, we recovered sequences from UCYN-A, confirming the importance of this symbiosis in oceanic regions off Brazil.

## Material and methods

### Sampling

Three transects (TR0, 1 and 2) were sampled off the coast of Brazil during the CARBOM V cruise onboard the R/V “Alpha Crucis” in November 2013 (Table 1; Fig. 1). TR0 was sampled mostly in surface, TR1 at the level of deep chlorophyll maximum (DCM) and TR2 at 3 depths. During the TR2 transect, a *Trichodesmium* sp. bloom was observed at Station 101, where additional samples were taken (Station Bloom). All samples were collected with 12 L Niskin

bottles attached to a CTD Teledyne model PS7000M (Teledyne Technologies Inc., CA, USA), except for surface samples from TR0 and Station Bloom, collected with a polycarbonate bucket. Samples (1.5 mL) for cell sorting by flow cytometry were collected into cryotubes with 10% DMSO (final concentration), flash-frozen in liquid nitrogen, and stored at –80 °C until analysis at the Station Biologique de Roscoff, France.

### Flow cytometry sorting

Samples were sorted using a FACSAria™ flow cytometer (Becton Dickinson, San Jose, CA), equipped with a laser emitting at 488 nm and a 70-mm nozzle. Emitted light was collected through the following set of filters: 488/10 band pass (BP) for side scatter, 576/26 BP for orange fluorescence, and 655 long pass for red fluorescence. Signal detection was triggered on chlorophyll fluorescence. Photosynthetic pico and nanoeukaryotes populations were selected based on light scatter, orange phycoerythrin, and red chlorophyll fluorescence as described previously [29]. Tris–HCl 50 mM, pH 8.0, NaCl 10 mM was used as sheath liquid. Sheath pressure was set at 70 PSI and nozzle frequency was 90.000 Hz with a deflection voltage of 6.000 V. Cells were sorted in purity mode and collected into Eppendorf tubes containing Tris-EDTA lysis buffer (Tris 10 mM, EDTA 1 mM, and 1.2% Triton, final concentration). In addition, sheath fluid samples were collected and analyzed as negative controls in all subsequent steps including sequencing, to test for contamination in the flow sorting process.

### PCR and Illumina run

DNA from sorted cells was extracted by one cycle of freezing and thawing in liquid nitrogen. Two genes, 18S rRNA and *nifH*, were amplified with nested PCR on the same sorted populations.

For eukaryotes, the V4 region of the 18S rRNA gene was targeted (about 380 bp). The first round of PCR amplification was done using the following 10 µL mix: 5 µL KAPA HiFi HotStart ReadyMix® 2×, 0.3 µM final concentration of primer 63F, 0.3 µM final concentration of primer 1818R (Supplementary Table 2, [30]), 1 µL of DNA and H<sub>2</sub>O. Thermal conditions were: 95 °C for 5 min, followed by 25 cycles of 98 °C for 20 s, 52 °C for 30 s, 72 °C for 90 s, and a final cycle of 72 °C for 5 min. For the second round: 12.5 µL KAPA HiFi HotStart ReadyMix® 2×, 0.3 µM final concentration of primer V4F\_illum, 0.3 µM final concentration of primer V4R\_illum (Supplementary Table 2, [31]), 2.5 µL of first round product and H<sub>2</sub>O for a 25 µL reaction. Thermal conditions were: 95 °C for 3 min, followed by 25 cycles of 98 °C for 20 s, 65 °C for 1 min, 72 °C

**Table 1** List of samples analyzed

Transect	Station	Depth (m)	Picoplankton sorted samples				Nanoplankton sorted samples				
			Sample code	Sorted cells #	18S sequence #	nifH sequence #	Sample code	Sorted cells #	18S sequence #	nifH sequence #	
0	6	45	1p	7651	19,466	137,117	1n	4845	95,054	<i>163</i>	
			2p	7343	107,644	113,897	2n	3258	45,111	<i>143</i>	
	19	0	3p	1005	134,873	92,500	3n	898	131,031	<i>116</i>	
			5p	793	112,590	25,341	5n	660	24,696	<i>77,360</i>	
			7p	907	22,348	26,969	7n	856	40,829	<i>7667</i>	
1	81	140	9p	3181	44,610	53	9n	1235	19,193	<i>34</i>	
			10p	3278	47,390	6241	10n	1232	53,230	<i>36</i>	
	85	110	11p	16,312	31,899	10,201	11n	1615			
	86	105	13p	6366	59,626	11,954	13n	1007	46,001	21,316	
	87	105	15p	6189	78,390	<i>1033</i>	15n	622	22,468	2678	
2	96	5	120p	1150	76,182	23,147	120n	75	70,455	<i>93</i>	
			30	121p	1737	71,785	23,706	121n	218	52,401	26,838
			50	122p	853	37,364	11,045	122n	234	78,740	15,543
	98	5	125p	3086	55,179	21,461	125n	1300	27,381	14,331	
			50	126p	1217	30,406	10,140	126n	782	65,714	16,929
			85	127p	3420			127n	226	60,610	11,493
	101	5	140p	500	46,569	12,301	140n	366	48,126	25,286	
			60	141p	1046	64,221	10,428	141n	485	30,081	21,302
			110	142p	641	89,797	17,156	142n	159	85,219	11,753
	106	5	155p	355	50,782	66,172	155n	18	54,162	20,674	
			60	156p	1800	43,917	16,093	156n	300	55,065	14,447
			100	157p	6910	51,848	15,204	157n	1152	29,078	15,532
	114	5	165p	728	48,514	39,918	165n	226	50,732	14,706	
			60	166p	660	62,897	28,107	166n	578	53,412	24,442
			80	167p	722	49,934	13,971	167n	390	31,424	20,616
Bloom	0	tri01p	1002	36,576	7772	tri01n	194	14,162	11,792		
		tri02p	744	46,889	17,259	tri02n	206	23,906	16,934		
		tri03p	600	55,630	15,152	tri03n	218	34,892	21,523		

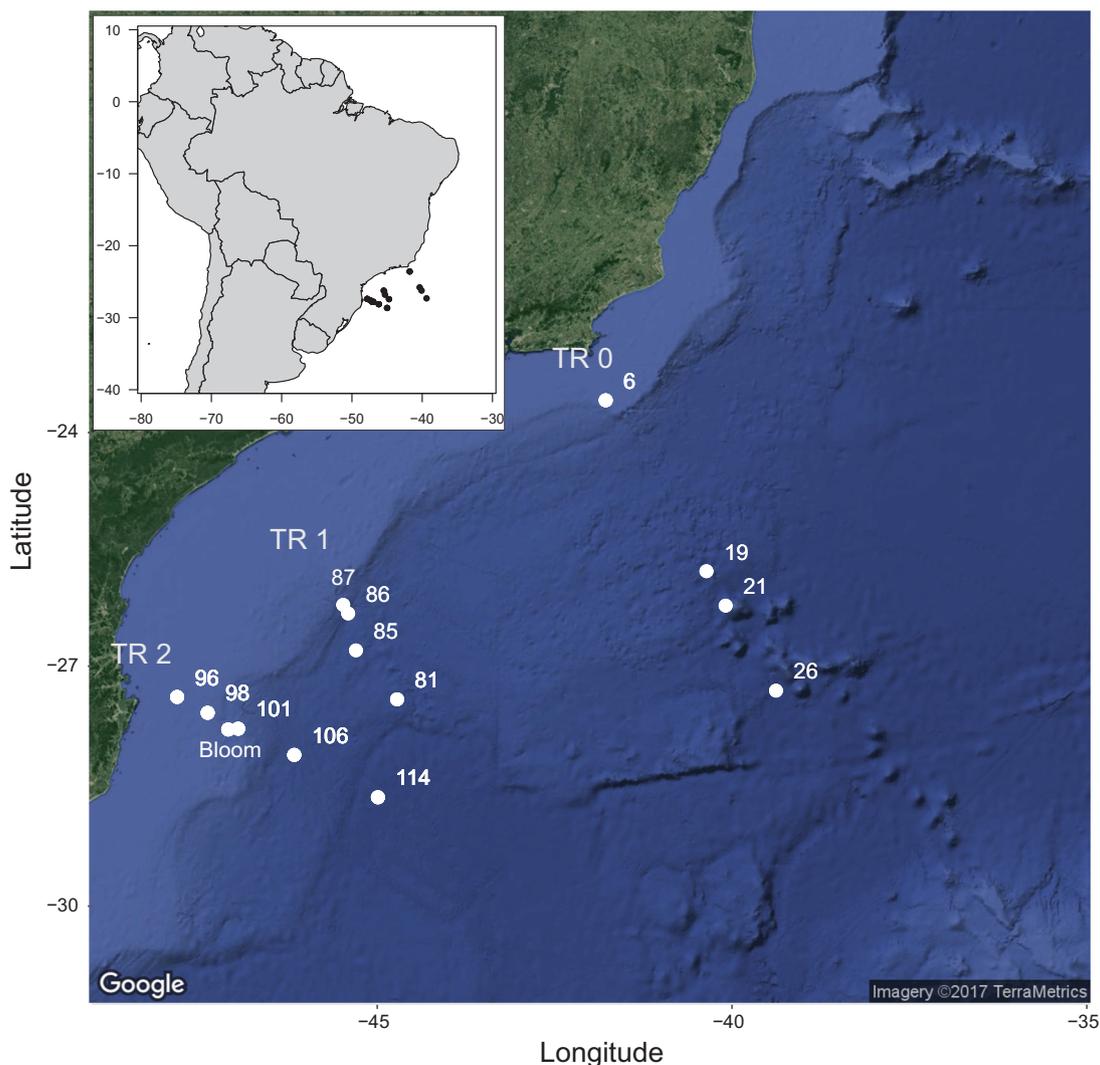
Samples corresponding to sorted photosynthetic picoeukaryote populations are labeled with p and those corresponding to sorted photosynthetic nanoeukaryote populations are labeled with n. Samples with <2000 sequences (in italics) were not considered in the analysis

for 90 s, and a final cycle of 72 °C for 5 min. One sample (127p, Table 1) failed to amplify the 18S gene. Negative controls and sheath fluid yielded either no amplification or faint bands which provided either no Illumina sequences or sequences from groups not considered in our analysis (Rhizaria, Mammalia) except for one Chrysophyceae clade C sequence that was also found in a few samples.

For the partial *nifH* gene amplification, the PCR reaction contained for the first round: 7.5 µL HotStartTaq® (Qiagen) buffer, 1.5 µL CoralLoad (Qiagen), 0.6 µM final concentration of primer nifH3, 0.6 µM final concentration of primer nifH4 [32], 2 µL of DNA and H<sub>2</sub>O for a 15 µL reaction. Thermal conditions were: 95 °C for 10 min, followed by 35 cycles of 95 °C for 1 min, 45 °C for 1 min, 72 °C for 1 min, and a final cycle of 72 °C for 10 min. For the

second round: 12.5 µL KAPA HiFi HotStart ReadyMix® 2×, 0.3 µM final concentration of primer nifH1\_illum, 0.3 µM final concentration of primer nifH2\_illum [32], 2.5 µL of first round product and H<sub>2</sub>O to a 25 µL reaction. Thermal conditions were: 95 °C for 5 min, followed by 25 cycles of 98 °C for 20 s, 54 °C for 1 min, 72 °C for 1 min, and a final cycle of 72 °C for 5 min. Second round PCRs were performed in triplicate and pooled together. Several samples had very low or no amplification of the *nifH* gene (Table 1). Negative controls did not yield any amplification.

Library was prepared as detailed on the Illumina® support website ([http://support.illumina.com/documents/documentation/chemistry\\_documentation/16s/16s-meta-genomic-library-prep-guide-15044223-b.pdf](http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-meta-genomic-library-prep-guide-15044223-b.pdf)) with final concentration of 1 nM. PCR products were checked initially



**Fig. 1** Map of stations

by gel electrophoresis and Bioanalyzer High Sensitivity DNA<sup>®</sup> kit. PCR cleanup and fragment size-selection was made with AMPure XP<sup>®</sup> (Beckman Coulter, Inc.). The Nextera XT Index Kit (Illumina<sup>®</sup>, FC-131-1002) was used to attach adapters that are complementary to the flow cell of the Illumina MiSeq<sup>®</sup> and to provide dual, unique indices for each sample. After another round of purification and size-checking validation, we performed quantification of the final products in triplicate with both Quant-iT<sup>™</sup> PicoGreen<sup>®</sup> dsDNA Assay Kit and qPCR, with KAPA<sup>®</sup> Library Quantification Kit. DNA concentration (nM) was computed as:

$$d = 10^6 \frac{D}{660 \times S}$$

where  $d$  is the DNA concentration in nM,  $D$  the DNA concentration in  $\text{ng } \mu\text{L}^{-1}$  and  $S$  the amplified fragment size.

Sequencing run was performed on a MiSeq system at the GenoMer platform at Station Biologique in Roscoff.

Sample concentration was normalized to have equal nM DNA concentration, pooled, denatured, diluted to 6 pM and loaded to the MiSeq run, along with 5% of denatured PhiX to prevent sequencing errors due to low-diversity libraries. For the sequencing step, we used MiSeq Reagent Kits v2<sup>®</sup>. The run was designed with Illumina Experiment Manager (<https://support.illumina.com/downloads/illumina-experiment-manager-v1-13.html>).

### Illumina data processing

Sequences were demultiplexed using the Illumina software. For each sample, the quality of sequences was first verified using the FastQC software [33]. Only sequences longer than 200 bp and with at least 75% of the bases with a quality score higher than 20 were kept. After this step, unpaired reads were also removed. The following steps were performed using Mothur version 1.38 [34]. Contigs were

assembled from forward and reverse reads for each sample. Contigs that contained ambiguities (N) were discarded. Amplicons for 18S rRNA and *nifH* contigs were extracted using the command `pcr.seqs`, allowing for two mismatches between the sequence and the primers (Supplementary Table 3). For each gene, sequences were dereplicated and singletons were removed. Unique sequences for 18S rRNA and *nifH* were aligned, respectively, to the Silva\_123 SEED available on Mothur website ([https://www.mothur.org/wiki/Silva\\_reference\\_files](https://www.mothur.org/wiki/Silva_reference_files)) and the aligned *nifH* database [35] available from <http://www.css.cornell.edu/faculty/buckley/nifh.htm>. Aligned unique sequences were pre-clustered allowing for two differences. Chimeras were removed with the UCHIME algorithm, as implemented in Mothur. At this stage, sequences from control samples (e.g., sheath fluid sort) and sequences corresponding to <10 reads were removed. The latter step makes all subsequent steps much faster to perform while removing a very small number of reads: for example for 18S rRNA, it decreases by sixfold the number of unique sequences while reducing the total number of reads by only 0.4% (Supplementary Table 3). OTUs were built at the 98% similarity level using Mothur average neighbor algorithm. The 274 OTUs for 18S rRNA were taxonomically assigned using the PR2 database [36] version 4.4 available on figshare repository ([https://figshare.com/articles/PR2\\_rRNA\\_gene\\_database/3803709/6](https://figshare.com/articles/PR2_rRNA_gene_database/3803709/6)) and on GitHub ([https://github.com/vaulot/pr2\\_database/releases](https://github.com/vaulot/pr2_database/releases)). This version of PR2 incorporates a revision of Haptophyta taxonomy following Edvardsen et al. [37, 38]. 18S rRNA and *nifH* OTUs were also searched against the GenBank database using BLAST plugin (November 2016) in Geneious® 10 [39]. One sample (11n, Table 1) was dominated by Metazoa sequences and was not considered further. For 18S rRNA, 16 OTUs (7.9 % of total reads) from Opisthokonta and Streptophyta were not further considered. Six groups (Dinophyta, Ochrophyta, Haptophyta, Cryptophyta, Chlorophyta, and Chlorarachniophyceae) were considered as autotrophic. Most analyses focus on these autotrophic groups to the exclusion of non-autotrophic groups such as Syndiniales. Autotrophic 18S rRNA or *nifH* OTUs were considered as “major” (Supplementary Table 4; Supplementary Table 5) if they contributed to a minimum of 20% of reads in at least one sample. This definition seems better than only considering the total contribution of a given OTU because some specific samples may have one dominant OTU that is not found in any other sample and which has therefore a low total abundance over the whole sample set. *nifH* sequences from major OTUs were aligned to related sequences obtained by BLAST as well as to some reference sequences using MAFFT [40]. Phylogenetic analysis was performed with FastTree [41]. Both programs were used as implemented in Geneious 10 [39] with default parameters.

Mothur script as well as OTU sequences for 18S rRNA and *nifH* as fasta files are available on GitHub ([https://github.com/vaulot/Ribeiro\\_CARBOM\\_ISME\\_2018](https://github.com/vaulot/Ribeiro_CARBOM_ISME_2018)). Raw FASTQ sequence files were deposited to GenBank under the project number PRJNA377956.

## Statistical analyses

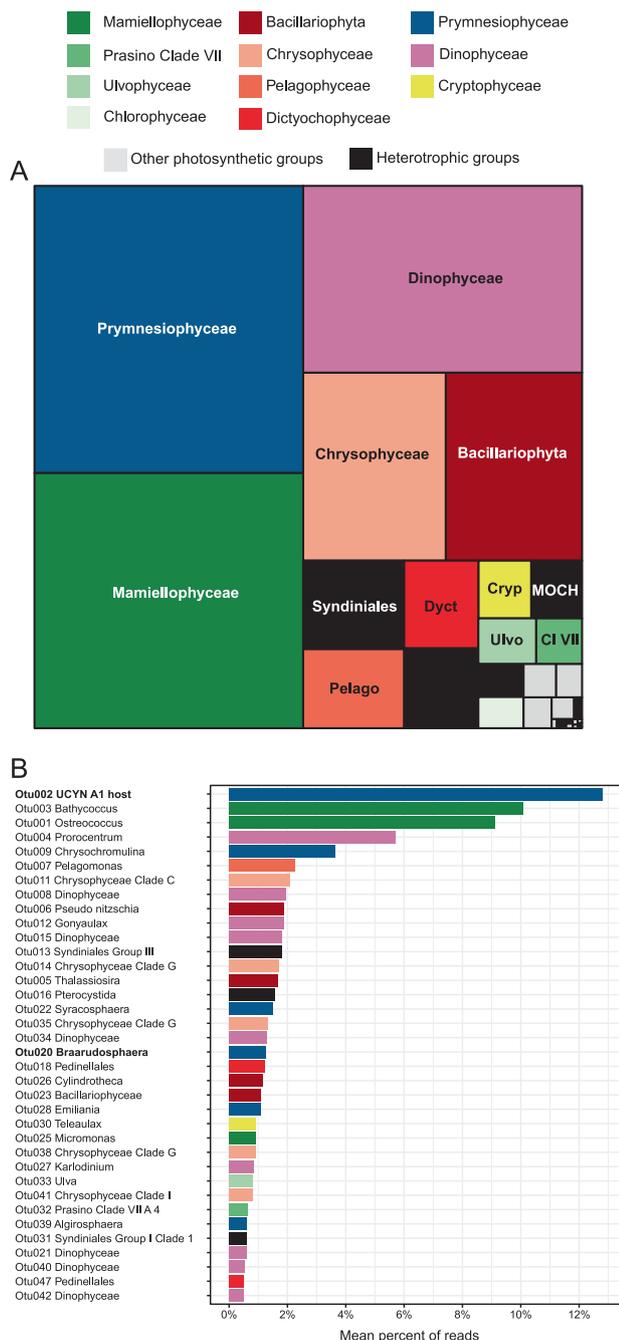
All statistical analyses and graphics were performed under R [42] using the following libraries: *ggplot2*, *ggmap*, *dplyr*, *treemap*, *vegan*. NMDS analysis of communities based on major 18S rRNA OTUs was performed using Bray–Curtis distance with *vegan* metaMDS function. Environmental parameters (Supplementary Material: [https://github.com/vaulot/Ribeiro\\_CARBOM\\_ISME\\_2018](https://github.com/vaulot/Ribeiro_CARBOM_ISME_2018)) were mapped using *vegan* `envfit` function. Network analysis was conducted using a matrix combining major 18S rRNA and *nifH* OTUs as rows, samples as columns, and read abundances as values. Correlations between OTUs and pseudo *p*-values were computed using SparCC [43]. Only correlations larger than 0.25 and with a pseudo *p*-value  $\leq 0.05$  were retained. Network between OTUs was constructed using the R package *igraph* and drawn using Gephi [44]. Scripts are available on GitHub ([https://github.com/vaulot/Ribeiro\\_CARBOM\\_ISME\\_2018](https://github.com/vaulot/Ribeiro_CARBOM_ISME_2018)).

## Results

The CARBOM V cruise took place during the austral spring of 2013 in a region located off South East Brazil and extending from the continental shelf to the 3510 m isobath (Fig. 1). Phytoplankton samples were preserved with DMSO and deep frozen (Table 1). Once back to the laboratory, we sorted small autotrophic eukaryotes using flow cytometry based on scatter and chlorophyll fluorescence properties. For each sorted population, we sequenced using high throughput techniques (Illumina) simultaneously for the V4 region of the 18S rRNA gene and a fragment of the *nifH* gene.

### 18S rRNA

The number of 18S rRNA reads varied between 14,000 and 135,000 depending on the sample (Table 1). Using a 98% similarity level, 258 OTUs were found (excluding opisthokonts and streptophytes) of which 53 could be attributed to heterotrophs, representing <7% of the reads. Within autotrophs, the major groups were Prymnesiophyceae, Mamiellophyceae, Dinophyceae, Bacillariophyta (diatoms), and Chrysophyceae, with minor contributions of Pelagophyceae and Dictyochophyceae (Fig. 2a). All abundant autotrophic OTUs matched with high similarity (from 99 to



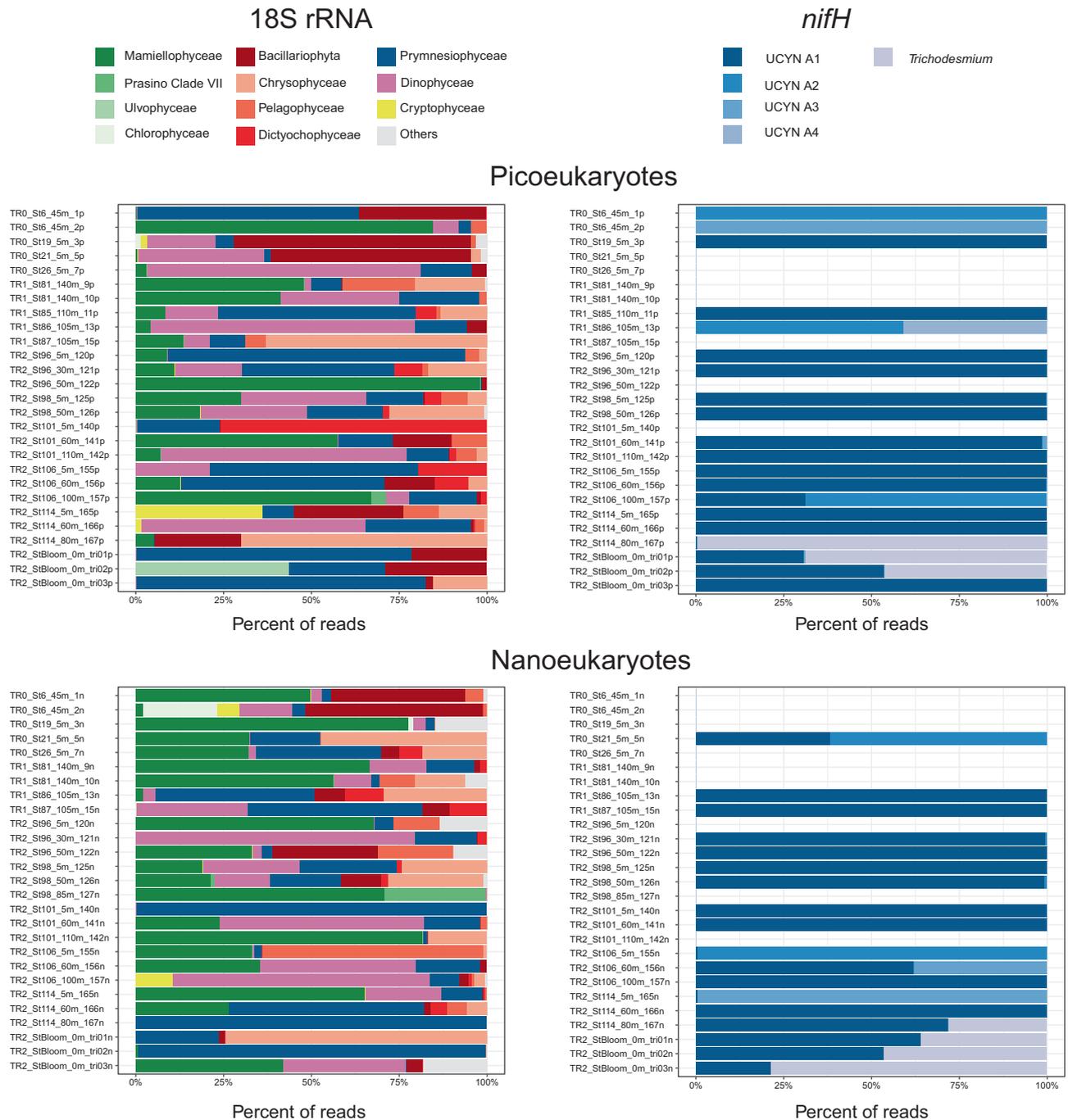
**Fig. 2** **a** Mean relative contribution of each class to 18S rRNA sequences for all sorted samples (pico- and nanoeukaryotes). MOCH: Marine Ochrophyta. **b** Mean relative contribution for major 18S rRNA OTUs clustered at 98% similarity including both autotrophic (colored bars) and heterotrophic (black bars) groups. Major OTUs are defined as those that contribute to more than 20 % of reads in at least one sample

100%) existing GenBank sequences but did not always correspond to sequences from cultures (Supplementary Table 4). The OTU that made the highest contribution (Fig. 2b, Otu002, 12.8% on average over all samples—note

that this was not the OTU with the highest number of reads because read number varied between samples) matched an environmental sequence (FJ537341, clone BIOSOPE T60.34) from an uncultivated Braarudosphaeraceae (Prymnesiophyceae) that has been demonstrated to host the nitrogen-fixing cyanobacteria UCYN-A1 [10]. An additional OTU (Otu020, 1.3% on average), with a lower average contribution, matched another Braarudosphaeraceae, *B. bigelowii*, which is known to host UCYN-A2 [45]. Two OTUs (Otu003 and Otu001) with high average contributions (10.1 and 9.1%, Fig. 2b) matched respectively two Mamiellophyceae, *Bathycoccus* [46], and the cultivated but yet undescribed clade B of *Ostreococcus* [47]. Other abundant OTUs (Fig. 2) were assigned to dinoflagellates (*Prorocentrum*, *Gonyaulax*, *Karlodinium*), Prymnesiophyceae (*Chrysochromulina*, *Syracosphaera*, *Emiliana*), diatoms (*Pseudo-nitzschia*, *Thalassiosira*, *Cylindrotheca*), uncultivated Chrysophyceae (clade G), Pelagophyceae (*Pelagomonas*), and other Mamiellophyceae (*Micromonas*). Within heterotrophs, the most abundant OTU (Otu013) was assigned to Syndiniales, known parasites of dinoflagellates [48].

The spatial distribution of the autotrophic groups for the different samples (Fig. 3) or along TR2 (Supplementary Figure 1) for which different depths have been sampled revealed that Prymnesiophyceae were in some cases more abundant in surface and Mamiellophyceae at depth (e.g., Stations 96 or 101). Some classes with overall relatively low abundance dominated some specific samples, as was the case for members of the division Ochrophyta: Pelagophyceae (nanoeukaryotes, St. 106, 5 m: 63% of reads), Chrysophyceae (picoeukaryotes, St. 87, 105 m and St. 114 80 m and nanoeukaryotes, St. Bloom 5 m, from 57 to 84%), and Dictyochophyceae (picoeukaryotes, St. 101, 5 m, 76%).

Samples were clustered based on the contribution of dominant autotrophic OTUs using Bray–Curtis dissimilarity (Fig. 4). Four major types of communities emerged (clusters A to D). Cluster A encompassed the largest number of samples (Fig. 4) and corresponded to communities dominated by Otu002, the uncultivated Braarudosphaeraceae known to host UCYN-A1. Some samples from this cluster also contained Otu003 (*Bathycoccus*) and some dinoflagellate OTUs. Samples from cluster B (Fig. 4) were dominated by Mamiellophyceae, with Otu001 (*Ostreococcus* clade B) and Otu003 (*Bathycoccus prasinos*) occurring together most of the time in quite similar proportions. In a few samples, *Ostreococcus* was replaced by *Micromonas* from clade A (Otu025). Samples from cluster C (Fig. 4) were composed by a mix of Prymnesiophyceae including the host of UCYN-A1 but also OTUs related to *Chrysochromulina*, *Syracosphaera*, *Emiliana* as well as



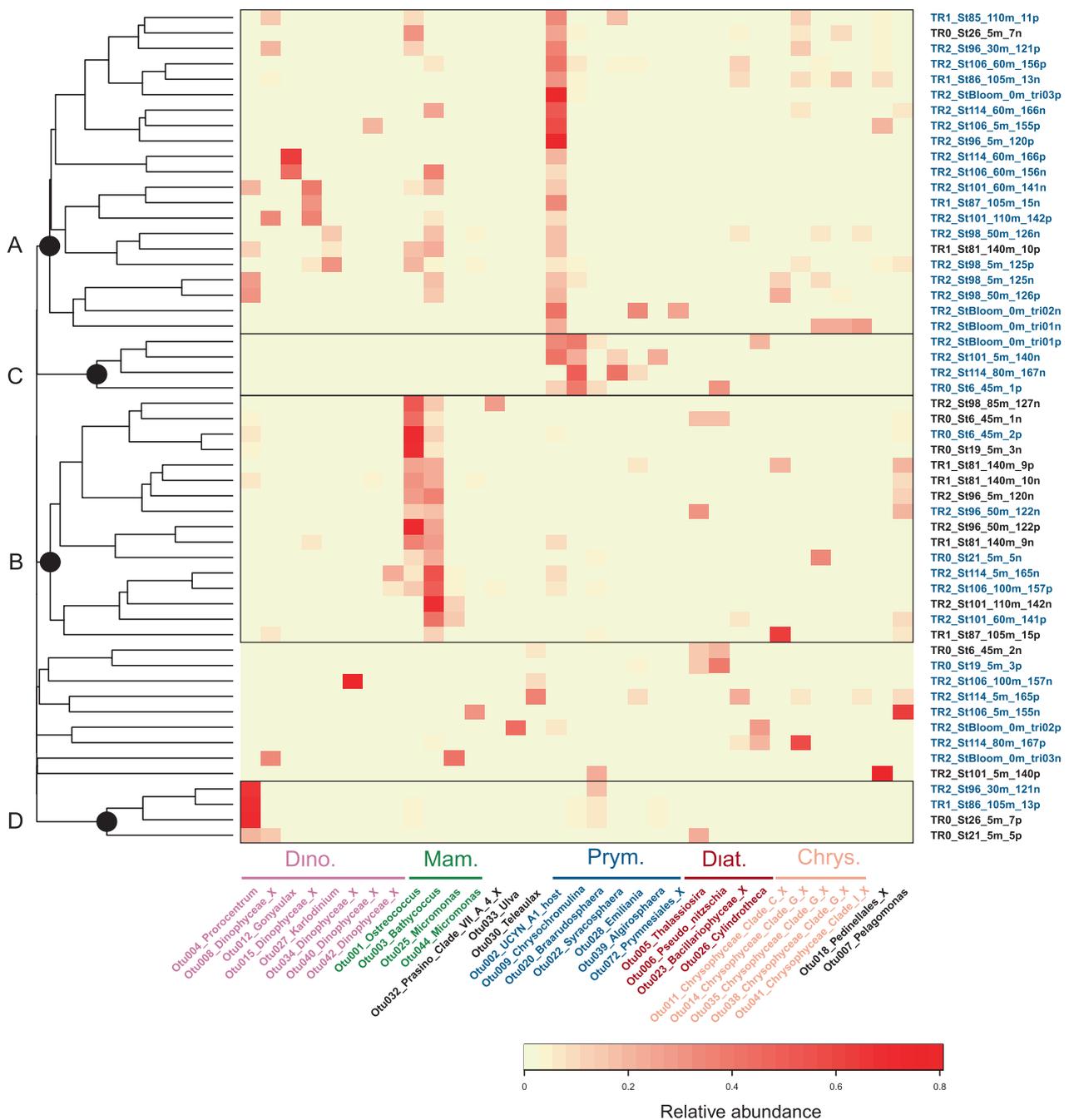
**Fig. 3** Relative read abundance for 18S rRNA autotrophic classes (left) and for cyanobacteria *nifH* OTUs (right) in each sorted sample for pico- (top) vs. nanoeukaryotes (bottom). Classes or OTUs that did not

contribute to at least 20% of sequences in at least one sample were regrouped into “Others”. Samples for which we could not amplify cyanobacteria *nifH* are displayed as empty (right side)

*Braarudosphaera*. Finally, cluster D (Fig. 4) was dominated by a dinoflagellate OTU related to *Prorocentrum* along with a smaller contribution of *Braarudosphaera*. The rest of the samples had all quite dissimilar communities without any other clear structure.

NMDS analysis of OTUs based on Bray–Curtis dissimilarity between samples onto which environmental variables

were mapped [49] revealed a clear sample clustering based on the type of community present in the sample (Fig. 5). Samples dominated by Mamiellophyceae (cluster B) were clearly separated from those dominated by UCYN-A1 haptophyte host (cluster A). The Mamiellophyceae community was linked to high chlorophyll fluorescence, higher N/P, larger depth, and lower temperature corresponding to



**Fig. 4** Heatmap for major 18S autotrophic OTUs (see Supplementary Table 4) ordered by taxonomy. Labels of samples corresponding to sorted picoeukaryotes and nanoeukaryotes are ending with p and n, respectively. Sample names colored in blue correspond to samples from which cyanobacteria *nifH* could be amplified. Samples were

clustered using Bray–Curtis dissimilarity. Four clusters can be defined: A correspond to samples dominated by UCYN-A1 host, B by the Mamiellophyceae *Ostreococcus* and *Bathycoccus*, C by an assemblage of Prymnesiophyceae and D by a dinoflagellate from the genus *Prorocentrum*

the interface between South Atlantic Current and Tropical waters [49], whereas the UCYN-A1 host community was linked to high temperature, low chlorophyll fluorescence, and low N/P corresponding to warm tropical waters. Cluster C corresponding to Prymnesiophyceae dominated samples was also associated to these tropical waters. Finally *Prorocentrum* dominated samples were from low salinity and

*Synechococcus*-rich samples corresponding to more coastal waters.

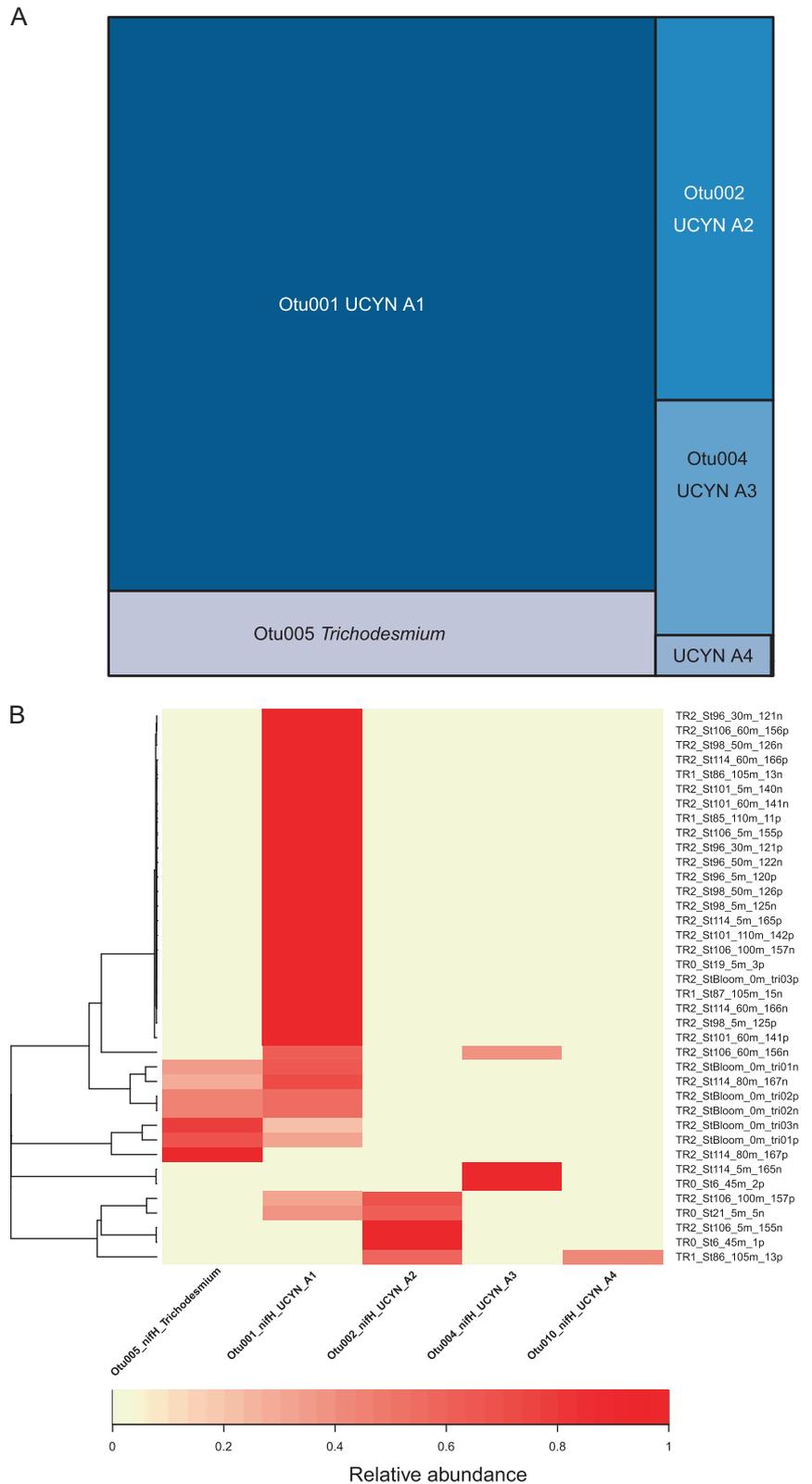
**nifH**

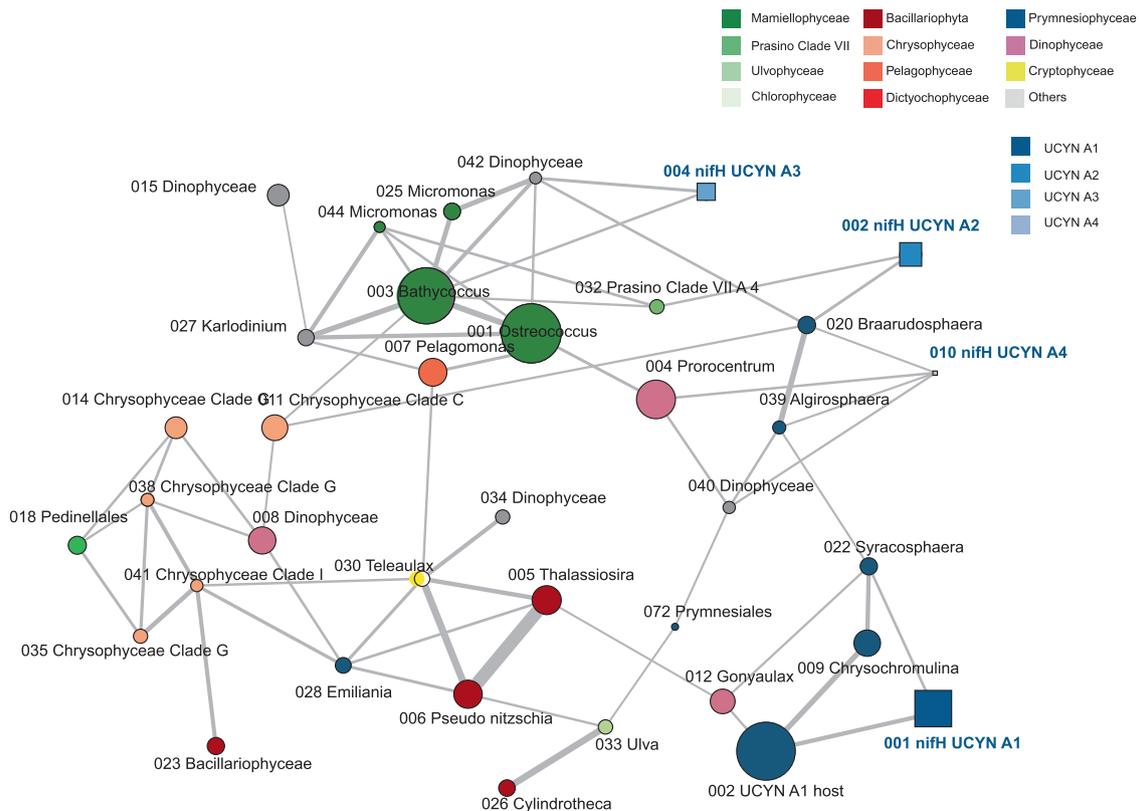
### nifH

The number of *nifH* reads per sorted sample varied from a few reads to 137,000. The number of OTUs (22 at



**Fig. 6 a** Mean relative contribution of cyanobacteria *nifH* OTUs clustered at 98% similarity for all sorted samples (both pico- and nanoeukaryotes), excluding only samples from which *nifH* could not be amplified. **b** Heatmap for cyanobacteria *nifH* OTUs (Supplementary Table 5). Labels of samples corresponding to sorted picoeukaryotes and nanoeukaryotes are ending with p and n, respectively. Samples were clustered using Bray–Curtis dissimilarity





**Fig. 7** Network analysis of the major autotrophic 18S rRNA (circles) and *nifH* (squares) OTUs (see Supplementary Table 4 and Supplementary Table 5) using SparCC correlation [43]. Colors of nodes correspond to taxonomic assignment. Size of nodes is proportional to

number of reads obtained. Width of edges is proportional to correlation between OTUs. Only correlations  $>0.20$  with pseudo  $p$ -values  $<0.05$  were considered

were also correlated to other 18S OTUs, for example A1 correlated to *Syracosphaera* or A2 to prasinophytes clade VII.

## Discussion

Phytoplankton communities in the euphotic zone of the Brazilian South Atlantic Bight are influenced by two main water masses: the warm and oligotrophic Tropical Water on the top 100 meters and the cold and nutrient-rich South Atlantic Central Water (SACW) below. In this region, photosynthetic pico and nanoeukaryotes average about 3000 cells per mL and 850 cells per mL, respectively [49]. To assess the taxonomy of these small eukaryotes, we analyzed populations sorted by flow cytometry based on their size and pigment fluorescence. This method has the advantage of targeting specific populations and allows to focus in components of the diversity that usually go unnoticed from filtered samples, due to the competition with more abundant or more easily amplified targets [29]. Almost all previous studies using this approach relied on sorting live samples [6, 51, 52], which requires the

availability of a sorting flow cytometer onboard ships and are therefore complex in terms of logistics. A recent study [53] used preserved samples but required both sample concentration and progressive freezing. We used instead a much simpler preservation method, relying on the addition of 10% DMSO to 1.5 mL of non-concentrated seawater samples, followed by a quick flash freezing. By eliminating sample filtration, the protocol used here requires  $<15$  min hands-on work after the water samples get onboard, therefore making sampling considerably easier in oceanographic cruises and minimizing changes in the plankton community.

Despite the small number of cells sorted in some samples (18–16,000), we could successfully PCR amplify and sequence both the 18S rRNA gene and the *nifH* gene on these samples. The vast majority of eukaryotes sequences were from autotrophs (93%). The major heterotrophic group was Syndiniales, which are known to infect dinoflagellates [48]. These parasites were probably sorted in association with the dinoflagellates that were present in our samples (e.g., Otu004 or Otu008). The successful targeting of the autotrophic population by flow cytometry sorting is also demonstrated by the low number of OTUs (258) for 18S

rRNA using a 98% similarity level, which is more than one order of magnitude less compared to what is found in studies using filtered material (e.g., [31, 54]). Likewise, the number of OTUs for *nifH* was very low (22) even compared to typical clone library studies (e.g., [55]), demonstrating that we probably only targeted bacteria that were in association with the sorted eukaryotes.

Heatmap and network analysis (Figs. 4 and 7) demonstrated that for some of the phylogenetic groups, 18S rRNA OTUs belonging to the same class had a tendency to co-occur, suggesting that population response to environmental forcing occurred often at the class level. This was the case for Mamiellophyceae, Prymnesiophyceae, and Chrysophyceae, which are well known contributors to phytoplankton in tropical oligotrophic and mesotrophic waters [3, 6].

The most prevalent guild, characteristic of warm tropical waters (Fig. 5), was dominated by the Prymnesiophyceae Otu002, which sequence is identical to that of the known host of UCYN-A1 (Fig. 4). This sequence has been shown to be abundant in specific oceanic areas, in particular off Brazil, where it can exceed 10 % of Prymnesiophyceae reads [27]. However, in our case, it could be much more dominant, representing up 81% of small photosynthetic eukaryotic reads (at St.96, 5 m). Another OTU (Otu020) matched *B. bigelowii*, which is hypothesized to be the host of UCYN-A2 [45], and is more widespread but less abundant than the host of UCYN-A1 [27]. Other relatively abundant genera included the coccolith-bearing *Algirosphaera*, *Syracosphaera*, and *Emiliania*. Although the latter genus is extremely widespread [56], data on the distribution of the two former ones are much more scarce, although they were among the most abundant OTUs in a seasonal study of Oslo fjord (Norway, [57]).

We observed four types of UCYN-A off Brazil belonging to the clades A1 to A4 [16], but none were related to the two recently defined clades A5 and A6 [25]. Although for all samples where UCYN-A was present, either the UCYN-A1 host or *Braarudosphaera bigelowii* were detected, these two OTUs were relatively abundant (in excess of 20% of the reads) in two samples where UCYN-A was not observed (St 26 and St 101 both at 5 m). This suggests that UCYN-A symbiosis may not be always obligatory for the host or that, in these two cases, UCYN-A cell dislodgement may have occurred during sampling or sorting, as observed previously [10, 27]. The most prevalent clade was A1, which is consistent with recent reports of this clade being the most abundant in metabarcoding surveys (e.g., [25]). Its co-occurrence (Fig. 7) with the prymnesiophyte 18S rRNA Otu002 reinforces the hypothesis of the specificity of this symbiotic relationship. Clade A2 was less frequent and less abundant, but was also clearly linked in our network analysis to its hypothesized host, *B. bigelowii*. It co-occurred with A1 in a few samples as observed elsewhere in coastal influenced pelagic waters [25]. In contrast to A1 and A2, virtually

nothing is known about the other clades. Off Brazil, A3 could co-occur with A1 as recently observed [25]. Network analysis using major OTUs suggested that A3 could be correlated to *Bathycoccus* and to a dinoflagellate while A4 could be similar to A2 also hosted by *B. bigelowii* (Fig. 7). However, as A3 and A4 were in relatively low abundance, they also could occur in symbiosis with a rare autotrophic eukaryote not taken into account in the network analysis.

We also detected some *nifH* sequences related to *T. erythreum* [58] which had been previously observed in this area [59]. These sequences originated from samples taken within a *Trichodesmium* bloom and at 80 m in the most offshore station of TR2. It is surprising that we could sort *Trichodesmium* cells because they possess phycoerythrin and our settings discriminated against phycoerythrin-containing cells. However, senescent cells of *Trichodesmium* display lower phycoerythrin fluorescence signal [60] and could have been sorted along with the eukaryotes. Both *Trichodesmium* and UCYN-A1 co-occurred in the same samples (Fig. 3), confirming that these two nitrogen-fixing cyanobacteria are not mutually exclusive, which has been observed elsewhere [55, 61].

The second type of community was dominated by Mamiellophyceae, either *Bathycoccus* or *Ostreococcus*, or both (Fig. 4). Mamiellophyceae are known to be ubiquitous in marine ecosystems but more prevalent in coastal waters, being replaced by prasinophytes clade VII in pelagic regions [62]. *B. prasinos* was first described from a deep Mediterranean Sea isolate [46] but since it has been observed in many areas from coastal waters to the open ocean [7, 63]. The existence of two different species of *Bathycoccus* has been hypothesized based on metagenomic data and strains available in cultures [7, 64]. Unfortunately, these two species have exactly the same 18S rRNA sequence, and therefore we were not able to differentiate them in the present study. In contrast, at least 4 clades can be distinguished by 18S rRNA for *Ostreococcus*, three corresponding to named species *O. tauri*, *O. "lucimarinus"* and *O. mediterraneus* and the last one, clade B [47, 65] also called OII [7], not yet described. The major *Ostreococcus* OTU in our study corresponds to clade B, which has been found in more oligotrophic regions than *O. "lucimarinus"* (e.g., [7]). For the third major Mamiellophyceae genus *Micromonas*, its two main OTUs were in relatively low abundance and related to the newly described species [66] *M. commoda* (clade A) and *M. bravo* (clade B), which are known to be widespread and abundant in coastal waters [67]. Network analysis demonstrated that these three genera (*Ostreococcus*, *Bathycoccus* and *Micromonas*), plus an OTU belonging to another green alga (prasinophytes clade VII A4, [62]), often co-occurred in our samples (Fig. 7).

A recent global study on UCYN-A symbioses [27] did not find a strong correlation between their distribution and

temperature, N/P ratios or chlorophyll *a*, suggesting their structuring factors may lie elsewhere. In the present work, NMDS analysis suggests that the alternation between communities dominated by UCYN-A1 host and Mamiellophyceae (Fig. 5) is linked to water masses with the former found in low chlorophyll tropical waters and the latter near the interface between warm tropical waters and colder South Atlantic waters where the chlorophyll maximum develops. Two key factors that could explain the dominance of the UCYN-A1 host community could be higher temperatures and lower N/P ratios, both favoring nitrogen-fixing cyanobacteria [17]. Tropical waters off Brazil have a very low N/P ratio (down to 0.4 in the present study) which could explain why this symbiosis is so prevalent. Another factor which could be important but was not measured in our study is iron, a key element for nitrogen-fixers [17]. By comparison, the “green” community was found in samples at depth with high chlorophyll fluorescence (Supplementary Figure 2), corresponding to the DCM at the interface between tropical water and SACW [49]. This environment matches the hypothesized habitat for *Ostreococcus* clade B [47, 68] and *Bathycoccus* [46, 64].

The lack of representative cultures from UCYN-A symbioses limits our comprehension of the factors structuring their populations and their competitive fitness in relation to other populations. Our study provides evidences that small phytoplankton communities in the region are highly structured and fall into a few specific types. It confirms the existing links between UCYN-A cyanobacteria bacteria and Prymnesiophyceae as well as the ecological importance of this association in the South Atlantic off Brazil. Moreover, our approach allows to formulate hypothesis about other UCYN-A clades for which host have not been characterized to this date.

**Acknowledgements** We thank the officers and crew of the R/V *Alpha Crucis* for logistical support, the chief scientist Ison Carlos A. da Silveira for the opportunity to participate in the cruise and for the CTD data and Mayza Pompeu for the analysis of inorganic nutrient data. Financial support for this work was provided by the CNRS Groupement de Recherche International (GDRI) “Diversity, Evolution and Biotechnology of Marine Algae”, the COFECUB-CAPES project “Pico Brás” (TE 871-15), FAPESP (*Fundação de Amparo à Pesquisa do Estado de São Paulo*—2012/04800-9 and 2014/15242-2) and FONDECYT grant PiSCOSouth (1171802). The GenoMer platform in Roscoff and C. Daguin are thanked for technical support with MiSeq sequencing. Access to bioinformatics resources was provided by the ABIMS platform (Station Biologique de Roscoff). We warmly thank Angélique Gobet for providing mothur analysis scripts, Hanna Farnelid for providing *nifH* sequences, and Jon Zehr for his thoughtful insights, especially concerning *nifH* contamination.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Vaulot D, Eikrem W, Viprey M, Moreau H. The diversity of small eukaryotic phytoplankton (<3 μm) in marine ecosystems. *FEMS Microbiol Rev.* 2008;32:795–820.
- Li WKW. Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr.* 1994;39:169–75.
- Rii YM, Duhamel S, Bidigare RR, Karl DM, Repeta DJ, Church MJ. Diversity and productivity of photosynthetic picoeukaryotes in biogeochemically distinct regions of the South East Pacific ocean. *Limnol Oceanogr.* 2016;61:806–24.
- Worden AZ, Nolan JK, Palenik B. Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. *Limnol Oceanogr.* 2004;49:168–79.
- Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N. A single species *Micromonas pusilla* (Prasinophyceae) dominates the eukaryotic picoplankton in the western English Channel. *Appl Environ Microbiol.* 2004;70:4064–72.
- Shi XL, Marie D, Jardillier L, Scanlan DJ, Vaulot D. Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific ocean. *PLoS ONE.* 2009;4:e7657.
- Simmons MP, Sudek S, Monier A, Limardo AJ, Jimenez V, Perle CR, et al. Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the eastern North Pacific Ocean. *Appl Environ Microbiol.* 2016;82:1693–705.
- Le Gall F, Rigaut-Jalabert F, Marie D, Garczarek L, Viprey M, Godet A, et al. Picoplankton diversity in the South-East Pacific ocean from cultures. *Biogeosciences.* 2008;5:203–14.
- McKie-Krisberg ZM, Sanders RW. Phagotrophy by the picoeukaryotic green alga *Micromonas*: implications for Arctic oceans. *ISME J.* 2014;10:1953–61.
- Thompson AW, Foster RA, Krupke A, Carter BJ, Musat N, Vaulot D, et al. Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science.* 2012;337:1546–50.
- Falkowski PG. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO<sub>2</sub> in the ocean. *Nature.* 1997;387:272–5.
- Karl D, Letellier R, Tupas L, Dore J, Christian J, Hebel D. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific ocean. *Nature.* 1997;388:533–8.
- Hutchins DA, Fu F-X, Zhang Y, Warner ME, Feng Y, Portune K, et al. CO<sub>2</sub> control of *Trichodesmium* N<sub>2</sub> fixation, photosynthesis, growth rates, and elemental ratios: implications for past, present, and future ocean biogeochemistry. *Limnol Oceanogr.* 2007;52:1293–304.
- Capone DG, Carpenter EJ. Nitrogen fixation in the marine environment. *Science.* 1982;217:1140–2.
- Carpenter EJ, Romans K. Major role of the cyanobacterium *Trichodesmium* in nutrient cycling in the North-Atlantic ocean. *Science.* 1991;254:1356–8.
- Farnelid H, Turk-Kubo K, Muñoz-Marín M, Zehr J. New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquat Microb Ecol.* 2016;77:125–38.
- Zehr JP. Nitrogen fixation by marine cyanobacteria. *Trends Microbiol.* 2011;19:162–73.
- Zehr JP, Shilova IN, Farnelid HM, Muñoz-Marín Carmen M, del C, Turk-Kubo KA. Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A. *Nat Microbiol.* 2016;2:16214.
- Zehr JP, Mellon MT, Zani S. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl Environ Microbiol.* 1998;64:3444–50.

20. Zehr JP, Bench SR, Carter BJ, Hewson I, Niazi F, Shi T, et al. Globally distributed uncultivated oceanic N<sub>2</sub>-fixing cyanobacteria lack oxygenic photosystem II. *Science*. 2008;322:1110–2.
21. Tripp HJ, Bench SR, Turk KA, Foster RA, Desany BA, Niazi F, et al. Metabolic streamlining in an open ocean nitrogen-fixing cyanobacterium. *Nature*. 2010;464:90–94.
22. Martínez-Pérez C, Mohr W, Löscher CR, Dekaezemacker J, Littmann S, Yilmaz P, et al. The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat Microbiol*. 2016;1:16163.
23. Bombar D, Heller P, Sanchez-Baracaldo P, Carter BJ, Zehr JP. Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J*. 2014;8:1–13.
24. Thompson A, Carter BJ, Turk-Kubo K, Malfatti F, Azam F, Zehr JP. Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ Microbiol*. 2014;16:3238–49.
25. Turk-Kubo KA, Farnelid HM, Shilova IN, Henke B, Zehr JP. Distinct ecological niches of marine symbiotic N<sub>2</sub>-fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa* sublineages. *J Phycol*. 2017;38:42–49.
26. Cornejo-Castillo FM, Cabello AM, Salazar G, Sánchez-Baracaldo P, Lima-Mendez G, Hingamp P, et al. Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat Commun*. 2016;7:11071.
27. Cabello AM, Cornejo-Castillo FM, Raho N, Blasco D, Vidal M, Audic S, et al. Global distribution and vertical patterns of a prymnesiophyte–cyanobacteria obligate symbiosis. *ISME J*. 2016;10:693–706.
28. Fernández-Castro B, Mouriño-Carballido B, Maraño E, Chouciño P, Gago J, Ramírez T, et al. Importance of salt fingering for new nitrogen supply in the oligotrophic ocean. *Nat Commun*. 2015;6:8002.
29. Marie D, Shi XL, Rigaut-Jalabert F, Vaultot D. Use of flow cytometric sorting to better assess the diversity of small photosynthetic eukaryotes in the English Channel. *FEMS Microbiol Ecol*. 2010;72:165–78.
30. Lepère C, Demura M, Kawachi M, Romac S, Probert I, Vaultot D. Whole Genome Amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microbiol Ecol*. 2011;76:516–23.
31. Piredda R, Tomasino MP, D'Erchia AM, Manzari C, Pesole G, Montresor M, et al. Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean long term ecological research site. *FEMS Microbiol Ecol*. 2017;93:fiw200.
32. Zehr JP, Turner PJ. Nitrogen fixation: Nitrogenase genes and gene expression. *Methods Microbiol*. 2001;30:271–286.
33. Andrew S. FastQC a quality control tool for high throughput sequence data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
34. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75:7537–41.
35. Gaby JC, Buckley DH. A comprehensive aligned *nifH* gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database*. 2014;2014:bau001.
36. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res*. 2013;41:D597–D604.
37. Edvardsen B, Egge ES, Vaultot D. Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding—a review. *Perspect Phycol*. 2016;3:77–91.
38. Edvardsen B, Eikrem W, Throndsen J, AS, Probert I, Medlin L. Ribosomal DNA phylogenies and a morphological revision provide the basis for a new taxonomy of Prymnesiales (Haptophyta). *Eur J Phycol*. 2011;46:202–28.
39. Kears M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
40. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059–66.
41. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5:e9490.
42. R Development Core Team. R: a language and environment for statistical computing. *R Found Stat Comput*. 2013;1:409.
43. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol*. 2012;8:e1002687.
44. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. 2009. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
45. Hagino K, Onuma R, Kawachi M, Horiguchi T. Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae). *PLoS ONE*. 2013;8:e81749.
46. Eikrem W, Throndsen J. The ultrastructure of *Bathycoccus* gen. nov. and *B. prasinos* sp. nov., a non-motile picoplanktonic alga (Chlorophyta, Prasinophyceae) from the Mediterranean and Atlantic. *Phycologia*. 1990;29:344–50.
47. Rodríguez F, Derelle E, Guillou L, Le Gall F, Vaultot D, Moreau H. Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ Microbiol*. 2005;7:853–9.
48. Chambouvet A, Morin P, Marie D, Guillou L. Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science*. 2008;322:1254–7.
49. Gérikas Ribeiro C, Lopes dos Santos A, Marie D, Helena Pellizari V, Pereira Brandini F, Vaultot D. Pico and nanoplankton abundance and carbon stocks along the Brazilian Bight. *PeerJ*. 2016;4:e2587.
50. Goto M, Ando S, Hachisuka Y, Yoneyama T. Contamination of diverse *nifH* and *nifH*-like DNA into commercial PCR primers. *FEMS Microbiol Lett*. 2005;246:33–38.
51. Farnelid HM, Turk-Kubo KA, Zehr JP. Identification of associations between bacterioplankton and photosynthetic picoeukaryotes in coastal waters. *Front Microbiol*. 2016;7:339. <https://doi.org/10.3389/fmicb.2016.00339>.
52. Balzano S, Marie D, Gourvil P, Vaultot D. Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *ISME J*. 2012;6:1480–98.
53. Kawachi M, Kataoka T, Sato M, Noël M-H, Kuwata A, Demura M, et al. Application of cryopreservation to genetic analyses of a photosynthetic picoeukaryote community. *Gene*. 2016;576:708–16.
54. Hu S, Campbell V, Connell P, Gellen AG, Liu Z, Terrado R, et al. Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. *FEMS Microb Ecol*. 2016;92:1–39.
55. Turk KA, Rees AP, Zehr JP, Pereira N, Swift P, Shelley R, et al. Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic. *ISME J*. 2011;5:1201–12.

56. Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, et al. Pan genome of the phytoplankton *Emiliana* underpins its global distribution. *Nature*. 2013;499:209–13.
57. Egge ES, Johannessen TV, Andersen T, Eikrem W, Bittner L, Larsen A, et al. Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high throughput sequencing. *Mol Ecol*. 2015;24:3026–42.
58. Hynes AM, Webb EA, Doney SC, Waterbury JB. Comparison of cultured *Trichodesmium* (Cyanophyceae) with species characterized from the field. *J Phycol*. 2012;48:196–210.
59. Carvalho M, Giancesella SMF, Saldanha-Corrêa FMP. *Trichodesmium erythraeum* bloom on the continental shelf off Santos, Southeast Brazil. *Braz J Oceanogr*. 2008;56:307–11.
60. Orcutt KM, Gundersen K. Another look at green *Trichodesmium* colonies. *Limnol Oceanogr*. 2008;53:2049–51.
61. Langlois RJ, Mills MM, Ridame C, Croot P, Laroche J. Diazotrophic bacteria respond to Saharan dust additions. *Mar Ecol Prog Ser*. 2012;470:1–14.
62. Lopes dos Santos A, Gourvil P, Tragin M, Noël M-H, Decelle J, Romac S, et al. Diversity and oceanic distribution of prasino-phytes clade VII, the dominant group of green algae in oceanic waters. *ISME J*. 2017;11:512–28.
63. Collado-Fabri S, Vaultot D, Ulloa O. Structure and seasonal dynamics of the eukaryotic picophytoplankton community in a wind-driven coastal upwelling ecosystem. *Limnol Oceanogr*. 2011;56:2334–46.
64. Vannier T, Leconte J, Seeleuthner Y, Mondy S, Pelletier E, Aury J-M, et al. Survey of the green picoalga *Bathycoccus* genomes in the global ocean. *Sci Rep*. 2016;6:37900.
65. Subirana L, Péquin B, Michely S, Escande ML, Meilland J, Derelle E, et al. Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). *Protist*. 2013;164:643–59.
66. Simon N, Foulon E, Grulois D, Six C, Desdevises Y, Latimier M, et al. Revision of the genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the type species *M. pusilla* (Butcher) Manton & Parke and of the species *M. commoda* van Baren, Bachy and Worden and description of two new species based on the 5 genetic and phenotypic characterization of cultured isolates. *Protist*. 2017;168:612–35.
67. Foulon E, Not F, Jalabert F, Cariou T, Massana R, Simon N. Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. *Environ Microbiol*. 2008;10:2433–43.
68. Limardo AJ, Sudek S, Choi CJ, Poirier C, Rii YM, Blum M, et al. Quantitative biogeography of picoprasinophytes establishes ecotype distributions and significant contributions to marine phytoplankton. *Environ Microbiol*. 2017;19:3219–34.