Nested PCR Approach for petB Gene Metabarcoding of Marine Synechococcus Populations

Denise Rui Ying Ong, Andrés Gutiérrez-Rodríguez, Laurence Garczarek, Dominique Marie, Adriana Lopes dos Santos

▶ **To cite this version:**

Denise Rui Ying Ong, Andrés Gutiérrez-Rodríguez, Laurence Garczarek, Dominique Marie, Adriana Lopes dos Santos. Nested PCR Approach for petB Gene Metabarcoding of Marine Synechococcus Populations. Microbiology Spectrum, 2023, 11 (2), pp.04086-22. 10.1128/spectrum.04086-22 . hal-04245006

HAL Id: hal-04245006
https://cnrs.hal.science/hal-04245006

Submitted on 17 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nested PCR approach for *petB* gene metabarcoding of marine *Synechococcus*

Denise Rui Ying Ong [a,#], Andrés Gutiérrez-Rodríguez [b*], Laurence Garczarek [c], Dominique Marie [c], Adriana Lopes dos Santos [a]

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

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

[c] Sorbonne Université, CNRS, UMR 7144 (AD2M), Station Biologique de Roscoff (SBR), Roscoff, France;

[*] Present address: Instituto Español de Oceanografía, Centro Oceanográfico de Gijón, Avda. Príncipe de Asturias 70 BIS, 33212, Gijón, Spain

Compiled January 30, 2023

This is a draft manuscript, pre-submission

Address correspondence to Denise Rui Ying Ong, denise.ry.ong@gmail.com.

ABSTRACT

The molecular diversity of marine picocyanobacterial populations, an important component of phytoplankton communities, is better characterised using high-resolution marker genes compared to the 16S rRNA gene as it has greater sequence divergence to differentiate between closely related picocyanobacteria groups. Although specific ribosomal primers have been developed, another general disadvantage of bacterial ribosome-based diversity analyses is the variable number of rRNA gene copies. To overcome these issues, the single-copy *petB* gene, encoding the cytochrome b_6 subunit of the cytochrome b_6f complex, has been used as a high-resolution marker gene to characterise *Synechococcus* diversity. We have designed new primers targeting *petB* gene and proposed a nested polymerase chain reaction (PCR) method (Ong_2022) for metabarcoding of marine *Synechococcus* populations obtained by flow cytometry cell sorting. We evaluated the specificity and sensitivity of Ong_2022 against the standard amplification protocol (Mazard_2012) using filtered seawater samples. Ong_2022 approach was also tested on flow cytometry sorted *Synechococcus* populations. Samples (filtered and sorted) were obtained in the Southwest Pacific Ocean, from subtropical (ST) and subantarctic (SA) waters masses. The two PCR approaches on filtered recov-

ered the same dominant subclades Ia, Ib, IVa and IVb with small differences in their relative abundance across the distinct samples. For example, subclade IVa was the dominant in ST samples with the Mazard_2012 approach, while the same samples processed with Ong_2022 showed similar contribution of subclades IVa and Ib to total community. Ong_2022 approach generally captured a higher genetic diversity of *Synechococcus* subcluster 5.1 compared to Mazard_2012, while having a lower proportion of incorrectly assigned ASVs. All flow cytometry sorted *Synechococcus* samples could only be amplified by our nested approach. The taxonomic diversity obtained with our primers on both sample types was in agreement with clade distribution observed by previous studies that applied other marker genes or PCR-free metagenomic approaches in similar environmental conditions.

IMPORTANCE The *petB* gene has been proposed as a high-resolution marker gene to access the diversity of marine *Synechococcus*. A systematic metabarcoding approach based on the *petB* gene would improve the characterization/assessment of *Synechococcus* community structure in marine planktonic ecosystems. We have designed and tested specific primers to be applied in a nested polymerase chain reaction (PCR) protocol (Ong_2022) for metabarcoding of the *petB* gene. Ong_2022 protocol can be applied to samples with low DNA content, such as those obtained by flow cytometry cell sorting, granting the simultaneous assessment of the genetic diversity of *Synechococcus* populations with cellular properties and activities (e.g., nutrient cell ratios or carbon uptake rates). Our approach will allow future studies using flow cytometry to investigate the link between ecological traits and taxonomic diversity of marine *Synechococcus*.

KEYWORDS: marine picocyanobacteria, marine *Synechococcus*, *petB*, nested PCR, metabarcoding.

INTRODUCTION

Marine photosynthetic bacterioplankton is dominated by two genera of picocyanobacteria: *Prochlorococcus* and *Synechococcus*. *Synechococcus* is a polyphyletic genus comprising both marine and freshwater lineages (1). They contribute up to 16% of the net marine primary productivity (2, 3). Changes in the physico-chemical properties of marine waters induced by climate change are predicted to have an impact on the distribution of phytoplankton groups, including *Synechococcus* (4). Given their major role in the marine carbon cycling and ocean food web (5, 6), an extensive assessment of *Synechococcus* community genetic diversity at spatial and temporal scale as well as their ecological traits are critical to understand the future evolution of this group under the ongoing global change.

Whilst studies using the 16S rRNA gene have revealed some degree of diversity within *Prochlorococcus* and *Synechococcus* (7, 8, 9), the relatively low sequence divergence between closely related clades or species and the variability in copy number of rRNA genes (8, 10) have prompted the scientific community to use high-resolution marker genes to analyse the genetic diversity within these groups of picocyanobacteria (e.g. *ntcA* (11), *petB* (12), ITS (13) and *rpoC1* (14)). Although most of these gene markers have been amplified from natural *Synechococcus* populations by standard polymerase chain reaction (PCR) amplification, others like *rpoC1* requires a nested PCR approach (15, 16).

Nested and semi-nested PCR approaches involve two sequential amplification reactions with different pair of primers. These amplification methods are often applied to increase the sensitivity and/or specificity of the reaction. They are also particularly useful for samples with low nucleic acid concentrations (15, 17). For example, metabarcoding of flow cytometry (FCM) sorted phytoplankton populations typically require a nested or semi-nested PCR amplification approach because sorted cells are in general in very low abundance compared to those collected in filtered samples (18, 19).

Metabarcoding is a suitable technique to assess the taxonomic diversity of flow cytometry sorted populations with high sensitivity (e.g. trace concentrations of DNA can be PCR amplified and sequenced) and of many samples, including *Synechococcus*. Although alternative approaches of metagenome or whole genome sequencing

could similarly obtain the taxonomic diversity without metabarcoding limitations of primer and amplification biases, these approaches require a much larger sequencing depth and sample processing time which would increase the cost and time needed (20). Whole genome sequencing would instead be more advantageous when applied to functional diversity and population genetics (21, 22). Flow cytometric sorted *Synechococcus* cells have provided important information about their metabolism. These sorted populations have been used to measure cellular properties such as nutrient stoichiometric and isotopic composition (23, 24) and cellular activity such as glucose (25), phosphorus (26), nitrogen (27, 28) and CO_2 uptake rates (5, 25, 27, 29, 30, 31). However, only few of these studies combined quantitative cell measurements with a fine taxonomic identification of the sorted populations by molecular methods (clone library sequencing and fluorescence in-situ hybridisation only) (29, 30, 31).

The *petB* gene, encoding the cytochrome b_6 subunit of the cytochrome b_6f complex, possesses several features that make this locus a good candidate for a gene marker. The *petB* is a single-copy gene, therefore reducing amplification bias (12). It is highly conserved in length and sequence allowing the obtained reads to be easily aligned (12) and has a comprehensive reference nucleotide sequence database that encompass most of the genetic diversity identified within marine *Prochlorococcus* and *Synechococcus* (32). Few specific primers targeting *petB* gene amplification are currently available in the literature. Mazard et al. (12) developed the original primer set, petB-F and petB-R. Ohnemus et al. (33) modified the original pair by increasing their degeneracy (up to 32 for both forward and reverse primers) in order to expand the diversity coverage of marine picocyanobacteria groups. The annealing region of these primers set is very similar and the degeneracy too high, precluding their use in nested or semi-nested PCR (polymerase chain reaction) amplification approaches.

With the objective of establishing a new nested PCR methodology to survey the diversity of flow cytometry sorted *Synechococcus* populations, we have designed and tested a new pair of primers (petB-50F and petB-634R) targeting the *petB* gene locus. Our approach to assess the diversity of marine *Synechococcus* populations by metabarcoding of *petB* includes a nested-PCR amplification (Ong_2022) which combines the newly developed pair (petB-50F and petB-634R) with the original primer set (petB-F

and petB-R) established by Mazard et al. (12). We have evaluated the sensitivity and specificity of our nested PCR protocol against the standard PCR protocol with petB-F and petB-R (Mazard_2012) on *Synechococcus* populations recovered from filtered seawater samples and sorted by flow cytometry from the same water parcels as the filtered samples.

RESULTS

Primer design and PCR amplification. The design of the new primers followed three main criteria. First, the primers must contain none or a very low number of mismatches (≥ 3) across all *Synechococcus* reference sequences present in the database in order to avoid bias in the amplification. Second, the resulting amplicon sequence must contain enough nucleotide variation to differentiate between the clades and subclades and therefore provide good taxonomic resolution. Last, the amplicon length ideally should be less than 550 bp long, so sequencing could be performed with Illumina 2x300 bp chemistry.

While the first two benchmarks were fulfilled with the new primers set petb-50F and petB-634R, the final amplicon length criteria could not be achieved. In Ong_2022 protocol, the new primers combined with petB-F and petB-R from Mazard et al. (12) (petB-634R during the first round and petB-50F during the second round of amplification) produced an amplicon of 571 bp length (from 50-618 position of the *petB* gene *Synechococcus* sp. WH8109 reference, accession number CP00688) while the standard Mazard_2012 PCR generates a fragment of 597bp (Table 1). There was no overlap between the forward and reverse reads obtained with both standard and nested PCR. To solve this problem, during sequence processing, we have merged our reads by adding 10 degenerate base 'N at their 3 prime end. Amplicons without overlap have been successfully processed with DADA2 pipeline using this bioinformatic manoeuvre, which does not significantly influence k-mer based taxonomic classification such as the naive Bayesian classifier implemented in DADA2 (34, 35). Moreover, concatenated sequences (i.e. joined without overlap) for longer amplicons were shown to improve taxonomic recovery and classification over merged sequences (36).

A total of 71 DNA samples extracted from filtered seawater and 18 flow cytometry

sorted samples were obtained from subtropical (ST1 and ST2) and subantarctic waters (SA1, SA2 and SA3) along the Chatham rise (Figure S1, Table S1). Among the DNA extracted samples, an amplification product was obtained for all environmental samples using Ong_2022 nested PCR and one sample could not be amplified with the standard Mazard_2012. Hence, from this point, the comparison between Mazard_2012 and Ong_2022 approaches will only include the 70 samples successfully amplified by both methods (Table S1). The results from sorted samples were only obtained with Ong_2022 PCR since none of the sorted samples were amplified with success by standard PCR reaction.

Following sequence processing, the total number of reads retained in the final dataset was about 49% and 23% of the initial number of reads generated with Mazard_2012 and Ong_2022 respectively (Table 2). The chimera detection step removed 66% of reads obtained with Ong_2022 and 35% for Mazard_2012 after merging forward and reverse reads (Table 2). The resulting mean number of reads obtained with the standard PCR was almost 3 times higher than the mean number of reads obtained with the nested PCR. However, the number of amplicon sequence variants (ASVs) generated with the nested PCR was higher (6307) when compared with Mazard_2012 standard reaction (4716) (Table 2).

Community composition from filtered samples. Four major *Synechococcus* subclades that were present in almost all samples were Ia, Ib, IVa and IVb (Figure 1). While these subclades represented nearly 99% of the total number of reads obtained by both approaches, the percentage of reads per sample for each of the major subclades differed between the amplification methods: Mazard_2012 produced a higher proportion of reads from subclades IVa and IVb while Ong_2022 amplified a higher proportion of subclade Ia and Ib sequences (Figure 2A). Yet, patterns of community composition such as the higher relative abundance of clade IVa in ST compared to SA waters were captured by both approaches. Among the seven minor clades and subclades present in both amplifications (CRD1, EnvA, EnvB, UC-A, Ic, II-WPC2, V), the proportion of reads of clades CRD1, EnvA, EnvB and V were higher in Ong_2022 compared to Mazard_2012 (Figure 2B, Table S2). Six clades and subclades with low contribution to the total community were only amplified by either Ong_2022 (WPC1, IIh, IIe, VIb) or Mazard_2012

(IIb, VIc) PCR method (Figure 2B).

Paired samples Wilcoxon test indicated that the median percent of reads for each clade and subclade in Mazard_2012 was significantly different from Ong_2022, except for clade V present at very low proportion (Table S2). The mean percent of reads for subclades Ia and Ib with Ong_2022 were higher than with Mazard_2012 (8.51% and 46.73% vs. 3.47% and 31.04%, respectively). The mean percent of reads for subclades IVa and IVb in Ong_2022 were lower compared to Mazard_2012 (11.96% and 31.40% vs. 20.39% and 44.04%, respectively) (Table S2). Among clades and subclades with minor contribution of reads to the total community, the mean percentage of reads of clade UC-A and subclades Ic and II-WPC2 were higher in Mazard_2012 compared to Ong_2022 (Figure 2B, Table S2). Clade UC-A was detected in all samples across both amplification types, but the average percentage of reads was higher in Mazard_2012 (0.63%) compared to Ong_2022 (0.13%). Subclades II-WPC2 and Ic were present in higher proportion and detected in more samples with Mazard_2012 compared to Ong_2022. For subclade II-WPC2, the average percentage of reads in Mazard_2012 was 0.26% and found in 67 samples, compared to Ong_2022, 0.01% in 27 samples. The average percentage of reads of subclade Ic in Mazard_2012 was 0.12% and found in 51 samples, compared to Ong_2022 of 0.03% and found in 20 samples. The average percentage of reads of clades CRD1, EnvA, EnvB, and V were higher in Ong_2022 (0.93%, 0.18%, 0.11% and 0.0006%, respectively) compared to Mazard _2012 (0.02%, 0.001%, 0.03% and 0.0003%, respectively) (Figure 2B, Table S2). Clades CRD1, EnvA and V were present in more samples in Ong_2022 (35, 21 and 3 samples, respectively) compared to Mazard_2012 (24, 5 and 1 samples, respectively). Clade EnvB was detected in the same 23 samples for both amplification types. Clade WPC1 and subclades IIh, IIe and VIb were present only in Ong_2022, although they were found in low proportion (less than 0.21% of reads in a sample) and only in 1 to 4 samples each. Subclades VIc and IIb were present only in Mazard_2012, but in very low proportion (less than 0.10 % of reads in a sample) and only in 1 to 2 samples (Table S2).

Most clades and subclades had a higher number of ASVs with Ong_2022 compared to Mazard_2012 (Figure 2C). Subclades with a higher mean percent of reads also had a higher number of ASVs. The exceptions were clade UC-A, and subclades Ic and II-

WPC2 which had a higher number of ASVs relative to their percent of reads and a higher number of ASVs in Mazard_2012 (Table S2).

The relative abundance of the major subclades followed a similar pattern across amplification methods and cycles, with notable exception for subclades IVa and IVb in ST1 and SA2 cycles, respectively. Subclade IVa was the dominant in ST1 samples with the Mazard_2012 approach, while the same samples processed with Ong_2022 showed similar contribution of subclades IVa and Ib to total community (Figure 1, Table S3). In cycles ST2 and SA1, the proportions of subclade IVa decreases and subclade IVb increases in both methods. In SA2 and SA3, subclade IVb becomes dominant only in Mazard_2012 samples (Figure 1, Table S3). Clades CRD1, EnvA and EnvB were found in higher proportion in subantarctic oceanic cycles SA2 and SA3 compared to more coastal SA1 and oceanic subtropical cycles for both amplifications (Figure 1, Table S3). The mean percent of reads for each subclade in cycles ST1, ST2 and SA1 was less than 0.004% and 0.06% for Mazard_2012 and Ong_2022, respectively. The mean percent of reads increased in cycles SA2 and SA3 to 0.06% and 2.8% for CRD1, 0.003% and 0.58% for EnvA, and 0.09% and 0.35% for EnvB, for Mazard_2012 and Ong_2022, respectively.

ASV sequences similarity to reference database. Within each clade and subclade, we compared the nucleotide similarity of the ASVs obtained with Mazard_2012 and Ong_2022 and with the reference sequences in the database (32) (Figure 3 and Table S4). In general, similarity between the ASVs within each clade and subclade obtained with both PCR approaches was lower when compared with the reference database, showing that both primers sets were able to capture a higher diversity than the one available in the database (Figure 3). The only exceptions were clades CRD1, EnvA and EnvB, and subclade VIc (Figure 3 and Table S4). Amidst the clades and subclades amplified by both PCR approaches, similarity within the ASVs obtained with Ong_2022 (from 85.9% to 94.7%) were lower than with Mazard_2012 (from 92.3% to 99.8%) for all the subclades with two exceptions, clade UC-A (87.7% and 30.5% respectively) and subclade II-WPC2 (85.9% and 25.9%, respectively) (Figure 3 and Table S4), indicating that Ong_2022 captured a higher microdiversity than Mazard_2012.

To confirm the taxonomic assignment of the ASVs, the nucleotide sequences of ASVs and references were translated into protein sequences. For each clade and sub-

clade, we compared the translated ASVs from each PCR approaches (Mazard_2012 and Ong_2022) against the translated reference sequences. ASVs for which the protein sequence had 95% pairwise identity against translated reference sequences were regarded as correctly assigned. Overall, Ong_2022 had a higher proportion of total ASVs that were correctly assigned compared to Mazard_2012 (98.2% vs. 84.2%; Table S4). The percentage of correctly assigned ASVs in each clade and subclade was comparable between Mazard_2012 (from 96.2% to 100%) and Ong_2022 (from 90.5% to 100%) except for clade UC-A (no correctly assigned ASVs) and subclade II-WPC2 (2.4%) obtained with Mazard_2012 (Table S4). This result shows that several ASVs in Mazard_2012 assigned to UC-A and II-WPC2 did not represent *petB* gene sequences, hence the very low percentage of nucleotide similarity among the ASVs for these lineages as mentioned previously. These ASVs assigned to UC-A and II-WPC2 from Mazard_2012 amplification were further examined by BLAST against GenBank nucleotide database. More than half of these ASVs had no sequence match in the database, indicating that they could have resulted from spurious amplification or the current database did not have a corresponding reference sequence.

Community composition from flow cytometry sorted *Synechococcus* populations. We tested our nested PCR with *Synechococcus* cells sorted by flow cytometry from 18 water samples collected in five cycles (Table S1). The 998 ASVs obtained from sorted cells with Ong_2022 were assigned to fewer clades and subclades belonging from subcluster (SC) 5.1 compared to ASVs obtained from filtered samples (Figure 4, S5).

Similar to the results obtained with filtered samples, Ia, Ib, IVa and IVb were the major subclades present in all the samples, representing an average of 99.98% of total reads (Figure 4 and Table S5). Clades UC-A, CRD1 and EnvB and subclade II-WPC2 had low contribution to the total community (Figure 4 and Table S5). The mean percent of reads for subclades Ia and Ib (3.84% and 31.33% respectively) from sorted cells were lower than those obtained from filtered samples (8.51% and 46.73%, Table S2 and S5). In contrast, for subclades IVa and IVb, the mean percent of reads were higher (25.47% and 39.35% respectively) in sorted cells than filtered samples (11.96% and 31.40%, Table S2 and S5).

Among minor clades and subclades, clade UC-A was found in 11 out of 18 sorted samples with an average proportion reads of 0.01% in contrast with filtered samples where this clade represented 0.13% of reads and was present in all samples. Clades CRD1 and EnvB and subclade II-WPC2 were present in 1 or 2 sorted samples, while in filtered samples they were present in one-third to half of the samples. Clades CRD1 and EnvB, and subclade II-WPC2 had a lower average percent of reads (0.003%, 0.002% and 0.002%) compared to Ong_2022 (0.93%, 0.11% and 0.01%). The 7 minor clades and subclades (EnvA, WPC1, V, IIh, IIe, Ic and VIb) that were found in filtered samples were not present in sorted samples.

The average number of ASVs per sample obtained from sorted *Synechococcus* cells was almost 2.5 times lower compared to filtered samples (82 ± 35 ASVs and 217 ± 63 ASVs, respectively). Each clade and subclade also had a lower number of ASVs. The relative abundance of major subclades followed a similar pattern between sorted *Synechococcus* and filtered seawater samples, with the exceptions of subclades IVa and IVb (Figure 1 and 4). Subclade IVa and subclade IVb were dominant in ST and SA sorted *Synechococcus* populations, respectively. This pattern was also observed with Mazard_2012 approach on filtered samples but differ of Ong_2022 where the overall contribution of subclades IVa and IVb was similar to that of subclades Ia and Ib (Figure 1).

DISCUSSION

We have established a nested PCR protocol to assess the diversity of marine *Synechococcus* based on metabarcoding analysis of the *petB* gene marker. Our protocol combines a pair of newly developed primers (petB-50F and petB-634R) and the original primer set (petB-F and petB-R) established by Mazard et al. (12). We have compared the sensitivity and specificity of the new nested amplification protocol (Ong_2022) against the original standard PCR (Mazard_2012) on DNA extracted from filtered seawater samples and determined the performance of the proposed nested amplification protocol on flow cytometry sorted *Synechococcus* populations. Whilst both standard and nested PCR worked in nearly all filtered seawater samples, sorted *Synechococcus* samples could not be amplified with Mazard_2012 standard PCR likely due to low ini-

tial concentration of *Synechococcus* DNA. Nested PCR is often employed when DNA concentration is too low to be amplified by a standard PCR (37, 38), thus Ong_2022 approach could be an alternative method when the initial concentration of *Synechococcus* cells is expected to be low, especially in sorted samples.

None of the ASVs obtained by both methods on filtered samples have been assigned to *Prochlorococcus*, although *Prochlorococcus* cells were detected through flow cytometry (from 0.3 x 10^3 to 90.6 x 10^3 cells ml^{-1}) in subantarctic water samples (SA1, SA2 and SA3) (data not shown). Similar to our new primers, Mazard et al. (12) petB-F and petB-R are also biased toward *Synechococcus* (12). However, during their study, *Prochlorococcus petB* sequences were recovered at low frequency from samples where *Prochorococcus* cell abundances were at least 45 times higher compared to *Synechococcus*. In our study, *Synechococcus* cells were always in equivalent or higher abundance (1.2 to 45 times) than *Procholorococcus* which might explain the absence of ASVs assigned to the latter.

ASVs assigned to *Synechococcus* SC 5.2 and 5.3 were also not detected. Mazard et al. (12) (petB-F and petB-R) and our primers (petB-50F and petB-634R) are biased toward SC 5.1, which is represented in the database by the majority of the sequences (32). In addition, *Synechococcus* cells from SC 5.2 and 5.3 were probably absent in our samples. SC 5.2 representatives are mainly found in estuaries and river-influenced coastal waters (39, 40, 41, 42) and although SC 5.3 contain both freshwater and marine representatives (32, 43), the latter have been sporadically detected at high abundance only in some regions such as the Red Sea, Mediterranean Sea and northwestern Atlantic Ocean (32, 44, 45).

One of the main concerns with nested PCR is the generation of chimeric sequences that are not true members of the community (37). Indeed, based on chimera removal analysis, our nested PCR approach had twice the proportion of chimeras detected than the standard amplification. However, Ong_2022 had an overall higher proportion of correctly assigned ASVs compared to Mazard_2012, indicating that the nested PCR was more specific toward *Synechococcus*. While the use of higher number of cycles might increase the chimera generation and sensitivity (e.g., Ong_2022 generally increased the proportion of reads of low abundant clades and subclades) in nested approaches,

the improved specificity of Ong_2022 PCR is derived from the binding of two separate sets of oligonucleotides to the same target template (46). Additionally, Ong_2022 PCR captured a higher genetic diversity of *Synechococcus* SC 5.1 compared to Mazard_2012. Within common subclades amplified by both Mazard_2012 and Ong_2022, almost all subclades had a higher number of ASVs in the latter approach. These ASVs had lower nucleotide similarity within each subclades indicating that our approach was able to capture a higher microdiversity within *Synechococcus* subclades. This directly contrasts with previous studies, which reported that nested PCR approaches resulted in a lower number of lineages and operational taxonomic units compared to a standard PCR (37, 38).

The number and proportion of reads of minor clades and subclades decreased in sorted samples (11 vs. 4 clades and subclades, 1% vs. 0.2% of reads for filtered and sorted samples, respectively). Metz et al. (19) investigated the diversity of photosynthetic picoeukaryotes in lakes using a similar approach. They found that the richness of filtered samples was on average lower than in the sorted samples. One major difference between Metz et al. (19) sorting protocol and Ong_2022 is the number of cells sorted. In Metz et al. (19), the number of cells sorted was 22 to 50 times higher than in Ong_2022. Since microbial plankton communities contain a large number of taxa that are present in low abundance (47), a larger sample size would probably increase the number of taxa detected and improve the representation of rare taxa with Ong_2022 approach on sorted cells (48). However, the trade-off between flow cytometry sorting time and the detection of rare microbial groups needs to be considered when designing the study. Cell sorting by flow cytometry is a time-consuming process, and increasing the number of cells sorted for numerous natural marine samples might not be feasible.

The taxonomic diversity obtained with our primers agreed with the clade distribution observed in previous studies conducted with other marker genes and PCR free methods in geographical regions with similar environmental conditions. For example, subclades Ia, Ib, IVa and IVb were the major lineages observed in the datasets obtained from filtered samples (~99%) and sorted samples (99.8%) with nested and standard PCR. This agrees with previous reports, in which clades I and IV most often co-occur

in the open ocean at latitudes above 30°N and below 30°S (9, 12, 32, 16, 45, 49). However, the relative abundance of clades I and IV varied between filtered samples amplified with Mazard_2012 and Ong_2022 as well as in sorted samples. Filtered samples amplified with Mazard_2012 and sorted samples amplified with Ong_2022 followed a similar pattern, whereby clade IV dominated in subtropical and subantarctic waters. In contrast, the proportion of clades I and IV remained similar across sampling locations when filtered samples were amplified with Ong_2022. Prior studies conducted in surrounding areas report different proportions of clade I to IV. Both clades were detected in similar proportions along the coast of New Zealand at overlapping latitudes (45°S) by PCR independent method (42). Clade IV has been observed in higher proportion compared to clade I in the western and eastern South Pacific Ocean at slightly lower latitudes with warmer conditions (around 30°S compared to 42°S in this study) (9, 32, 45). Hence, we cannot ascertain which method is more closely reflecting the natural abundance of the dominant clades.

We observed that relative abundance of clades CRD1, EnvA and EnvB increased in high nutrient low chlorophyll (HNLC) subantarctic waters where iron is typically the proximate limiting nutrient (filtered samples amplified with both approaches) (50, 51). Clade CRD1 was reported to co-occur with clades I and IV in the southeastern Pacific Ocean and other cold and high latitude HNLC conditions (32, 16, 45). More recent reports suggested that EnvA and EnvB also correspond to lineages found in low iron conditions (32, 45, 52, 53). The Ong_2022 approach recorded subclades IIh and IIe in cycle ST1 at very low abundance. Although members from clade II are mostly warm thermal types that dominate in nutrient depleted tropical areas (3), subclades IIe and IIh were also reported at low relative abundances in cooler waters (14.1–17.5 °C) where iron was not limiting for phytoplankton growth (32). Surface waters at ST1 were at a similar temperature at the time of sampling (13 °C), which is likely within the temperature range where subclades IIe and IIh were detected. Subclades WPC1, VIb, VIc and V were detected in low concentrations, but their precise ecological niche and biogeographical distribution have not yet been established (32).

In summary, we have designed a nested PCR protocol with two new primer sequences to amplify the *petB* gene from marine *Synechococcus*. Results obtained from

filtered seawater samples suggest that the new nested PCR approach is specific of *Synechococcus* and captured a wider genetic diversity, especially for rare groups. The nested PCR protocol successfully amplified *Synechococcus* cells from flow cytometry sorted samples and recovered a composition for the dominant subclades similar to that of filtered samples. Our amplification protocol will improve our understanding *Synechococcus* communities by allowing the determination of *Synechococcus* diversity directly associated with quantitative measurements (e.g. nutrient cell ratios or carbon update rates) obtained on sorted samples.

MATERIALS AND METHODS

Sampling Location. Samples were collected during the Salp Particle expOrt and Oceanic Production (SalpPOOP) cruise (TAN1810, 21st October – 21st November 2018) conducted near the Chatham Rise (Aotearoa - New Zealand) on board the R/V Tangaroa. The Chatham Rise is a dynamic region where northward-moving subtropical (ST) water masses mix with southward moving subantarctic (SA) water masses to form the Subtropical Convergence Zone (54, 55, 56, 57). ST waters are warm, saline and macronutrient-depleted while SA waters are cool, less saline and high-nitrate, low chlorophyll, low-silicate (HNLC-LSi) waters where low iron and silicate are the primary limiting factors for phytoplankton growth and productivity (50, 58).

We adopted Lagrangian sampling strategy using a satellite-tracked 15 m drogue drifted array that allowed us to track and sample the same water parcel multiple times during 3 to 6 days (hereafter referred to as 'cycles') (59). Five cycles were conducted, two in subtropical waters (ST1 and ST2) and three in subantarctic waters (SA1, SA2 and SA3) (Figure S1). In each cycle, water column profiles of temperature, salinity, dissolved oxygen, photosynthetically active radiation (PAR) and fluorescence were measured 4–6 times daily by Seabird (SBE 911plus) Conductivity - Temperature – Depth (CTD) with a PAR sensor. The cycles were identified as subtropical or subantarctic based on physical measurements during sampling following Lüskow et al. (57) definition. Samples used for picocyanobacterial community analysis presented in this study were collected using a CTD rosette sampler equipped with 10 L Niskin bottles from six depths (5–70 m). For DNA analysis of filtered seawater samples, 1.5–2 L seawater

from all depths was filtered through a 0.22 µm pore-size Sterivex filter using a peristaltic pump (n samples = 71) (Table S1). For DNA analysis of flow cytometry sorted samples, 1.5 mL samples were collected from the surface mixed layer (SUR, 12 m) and the deep chlorophyll maximum (DCM, 40 m or 70 m) (Table S1), where they were preserved with a solution of DMSO and 1/100 pluronic acid mix at 10% final concentration, and flash-frozen in liquid nitrogen (n samples = 21). All samples were stored at -80 °C until laboratory processing.

Flow cytometry cell sorting. *Synechococcus* cells were sorted using FACSAria™ flow cytometer (Becton Dickinson, San Jose, CA), equipped with a laser emitting at 488 nm, 70 mm nozzle and the following filters: 488/10 band pass (BP) for size side scatter (SSC), 576/26 BP for orange fluorescence, and 655 long pass for red fluorescence. *Synechococcus* populations were detected by their signature in plots of chlorophyll-a red fluorescence versus size scatter (SSC) and discriminated from photosynthetic picoeukaryotes based on the orange fluorescence of phycoerythrin. Tris–HCl 50 mM, pH 8.0 and, NaCl 10 mM buffer filter-sterilized (0.22 µm) 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) as described previously (18). Sheath fluid samples collected during sorting were processed and analysed as negative controls in all subsequent steps, including sequencing, to test for contamination in the flow sorting process.

petB-50F and petB-634R primer design. The criteria for primer design were: (i) 3 or fewer mismatches in binding region, (ii) no nucleotide ambiguity or mismatches at the 5 and 3 prime ends, (iii) low degeneracy, (iv) binding region between 15 and 20 base pairs, (v) annealing temperature difference between the forward and reverse primers not greater than 4 °C, (vi) primer binding region should be absent of insertions and deletions among the different groups and (vii) no hairpins and self-dimerisation within primers sequence. Four hundred and seven *Synechococcus* nucleotide sequences of the *petB* gene from Farrant et al. (32) reference database were aligned using Multiple Alignment and Fast Fourier Transform (MAFFT) alignment Version 7.450 (60) in Geneious Prime Version 2019.2.3 (61) with default parameters and used to locate the

biding region attending the criteria described. The selected forward primer petB-50F is located at positions 50-66 with reference to the first nucleotide of the *petB* gene from *Synechococcus* sp. WH8109 (accession number CP00688) (Table 1). The selected reverse primer petB-634R is located at positions 615-634 with reference to the first nucleotide of the *petB* gene from *Synechococcus* sp. WH8109 (accession number CP00688) (Table 1). Primer petB-50F had no mismatches in the binding region for *Synechococcus* SC 5.1 reference sequences, 2 to 3 mismatches for SC 5.2 and 5.3, and excluded *Prochlorococcus* (≥ 3 mismatches) and other marine cyanobacteria (e.g., *Richelia* sp.). Primer petB-634R had mainly no mismatches in the binding region for *Synechococcus* SC 5.1 to 5.3, and excluded *Prochlorococcus* (≥ 3 mismatches). After the binding region was selected, the sequences were truncated to the length amplified by the new forward primer sequence and petB-R from Mazard et al. (12) and used to infer a maximum likelihood phylogenetic tree with FastTree Version 2.1.11 (62) in Geneious Prime with default parameters. The phylogenetic tree was used to determine if the shorter *petB* amplicon could distinguish between subclades. All alignments and phylogenetic trees can be found in https://github.com/deniseong/marine-Synechococcus-metaB.

DNA extraction and PCR. DNA from Sterivex filters was extracted using the Nucleospin Plant II DNA extraction kit from Macherey-Nagel with modification of manufacturer instructions. The modified protocol can be found on https://github.com/deniseong/marine-Synechococcus-metaB. DNA from flow cytometry sorted *Synechococcus* cells was obtained using three cycles of flash-freezing and thawing in liquid nitrogen (18). Mazard_2012 protocol refers to PCR amplification with petB-F (5'-TACGACTGGTTCCAGGAACG-3') and petB-R (5'-GAAGTGCATGAGCATGAA-3') primers (12), performed on extracted DNA from filtered seawater. PCR reactions were done using the following reaction: extracted DNA template with an average final concentration of 1.3 ng μL^{-1} (from 0.007 to 4.11 ng μL^{-1}), 12.5 µL KAPA HiFi HotStart ReadyMix 2X, 0.3 µM primers (petB-F and petB-R) final concentration, 0.1 to 0.4 µg μL^{-1} final concentration of bovine serum albumin (BSA) and H_2O to 25 µL final volume reaction. Thermal conditions were: 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s, and a final cycle of 72 °C for 6 min. All PCRs were performed in duplicates and pooled together. Ong_2022 protocol refers to a nested PCR performed on extracted DNA from either

filtered seawater or flow cytometry sorted *Synechococcus* cells. The first round of PCR amplification was done using the following reaction: extracted DNA template with an average final concentration of 0.54 ng μL^{-1} (from 0.015 to 2.055 ng μL^{-1}) or volume corresponding to approximately 160-400 sorted *Synechococcus* cells (4 µL maximum), 5 µL KAPA HiFi HotStart ReadyMix 2×, 0.3 µM final concentration of primer petB-F, 0.3 µM final concentration of primer petB-634R (5'-GCTTVCGRATCATCARGAAG-3'), 0.1 µg μL^{-1} final concentration of BSA (only for filtered samples) and H_2O for a 10 μL reaction. Thermal conditions were: 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 59 °C for 30 s, 72 °C for 45 s, and a final cycle of 72 °C for 6 min. For the second round of amplification the following conditions were used: 2.5 µL of first round product, 12.5 µL KAPA HiFi HotStart ReadyMix 2X, 0.3 µM final concentration of primer petB-50F (5'-TYCAGGACATYGCTGAY-3'), 0.3 µM final concentration of primer petB-R and H_2O for a 25 µL reaction. Thermal conditions were: 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s, and a final cycle of 72 °C for 6 min. All first and second round PCRS were performed in duplicates and pooled together. Samples were purified, barcoded, and sequenced by the GeT-PlaGe platform of GenoToul (INRAE Auzeville, France) using an Illumina MiSeq platform (2 x 300 bp).

***petB* rRNA gene amplicon analysis.** Sequences were processed on RStudio Version 1.4.1717 (63). Primer sequences were removed using Cutadapt Version 3.4 (64). Fastq files were trimmed and quality filtered using function 'filterandtrim' with the DADA2 R package Version 1.12 (65). Based on sequence quality score, reads obtained with Mazard_2012 and Ong_2022 were trimmed and filtered using the following options: truncLen = c(280, 280) and c(250, 280), minLen = c(280, 280) and minLen = c(250, 280) respectively with truncQ = 2, maxEE = c(10, 10) in both datasets. Forward and reverse sequences were dereplicated and grouped. To merge the forward and reverse reads, the 'mergePairs' function on DADA2 was used with the justConcatenate = TRUE option to add 10 degenerate base 'N' between forward and reverse reads. Chimeras were identified and removed using the function 'removeBimeraDenovo'. Taxonomy was obtained with the function 'assignTaxonomy'. ASVs were assigned against *petB* reference database from Farrant et al. (32) reformatted for use with DADA2. The following taxonomic nomenclature was maintained: three major

Synechococcus/Cyanobium lineages called subclusters (SC) 5.1 through 5.3; fourteen clades (I to IX, XVI, XX, UC-A, WPC1, CRD1, EnvA and EnvB) within SC 5.1 and 20 subclades (Ia to Ic, IIa to IIh and II-WPC2, IIIa to IIIc, IVa and IVb, and VIa to VIc). ASVs assigned as *Richelia* sp. were removed prior to normalising reads. At station SA1_5, sorted samples from two depths (12 m and 40 m) were sorted to two and three different total number of cells (Table S1). As the community was similar at each depth, only samples that had 10,000 cells sorted were used for analysis and for filtered samples, all depths were combined. Analysis was performed with R Phyloseq package Version 1.36.0 (66) and Geneious Prime Version 2019.2.3 (61). All sequence processing and analysis scripts along with the reformatted database can be found in https://github.com/deniseong/marine-Synechococcus-metaB.

Availability of data and materials. Raw sequencing data were deposited in NCBI Sequence Read Archive under the bioProject number PRJNA885274. Source code, DNA extraction protocol as well as all metadata generated are available on GitHub (https://github.com/deniseong/marine-Synechococcus-metaB). Raw data files from flow cytometry sorting is available on http://flowrepository.org/experiments/1773 (Repository ID: FR-FCM-Z5P8). Step-by-step protocol for DNA extraction and PCR amplification are published on protocols.io at DOI: 10.17504/protocols.io.5qpvorkj7v4o/v1.

Author contributions statement DRYO and ALS conceived the study. AGR organized the SalpPOOP voyage and sampling structure. AGR and ALS developed field protocol and processed the samples during the voyage. ALS and DM performed flow cytometry cell sorting. DRYO developed primers and processed the samples for sequencing. DRYO analysed data and drafted the manuscript. All authors were involved in data interpretation and critically revised the manuscript.

ACKNOWLEDGMENTS

We acknowledge the crew of RV *Tangaroa* for their efforts in facilitating the sampling throughout the TAN1810 voyage. We thank Florian Humily and Sophie Mazard for providing the primer design of petB-634R. We thank Daniel Vaulot (CNRS) for feedback on data processing and editing comments. We are grateful to Jaret Bilewitch and Debbie Hulston (National Institute of Water and Atmospheric Research, New Zealand) for

their laboratory work on the DNA extraction of filtered seawater samples and Debbie Hulston for coordinating the logistics of transporting the DNA extracted samples from New Zealand to Singapore. We thank Priscillia Gourvil (CNRS) for coordinating the logistics of transporting samples from France to Singapore. We are grateful to Karen Selph (University of Hawaii at Manoa, School of Ocean and Earth Science & Technology) for her comments on the abundance of *Prochlorococcus* cell obtained by flow cytometry in the samples collected during TAN1810. We thank Molly A. Moynihan (Marine Biological Laboratory) for laboratory assistance and guidance, as well as coordinating the logistics of Illumina sequencing. We thank Clarence Sim Wei Hung (NTU) for laboratory assistance.

ALS and DO were supported by the Singapore Ministry of Education, Academic Research Fund Tier 1 (RG26/19). LG was supported by the French "Agence Nationale de la Recherche" Programs CINNAMON (ANR-17-CE02-0014-01). AGR was supported by NIWA via the New Zealand Ministry of Business, Innovation and Employment's Strategic Science Investment Funding to the National Coasts Ocean Centre.

REFERENCES

1. **Dvořák P, Casamatta DA, Poulíčková A, Hašler P, Ondřej V, Sanges R**. 2014. *Synechococcus*: 3 Billion Years of Global Dominance. Mol Ecol 23 (22):5538–5551. doi:10.1111/mec.12948.
2. **Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, Karl DM, Li WKW, Lomas MW, Veneziano D, Vera CS, Vrugt JA, Martiny AC**. 2013. Present and Future Global Distributions of the Marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. Proc Natl Acad Sci 110 (24):9824–9829. doi:10.1073/pnas.1307701110.
3. **Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F**. 2009. Ecological Genomics of Marine Picocyanobacteria. Microbiol Mol Biol Rev 73 (2). doi:10.1128/mmbr.00035-08.
4. **Flombaum P, Martiny AC**. 2021. Diverse but Uncertain Responses of Picophytoplankton Lineages to Future Climate Change. Limnol Oceanogr 66 (12):4171–4181. doi:10.1002/lno.11951.
5. **Li WKW**. 1994. Primary Production of Prochlorophytes, Cyanobacteria, and Eucaryotic Ultraphytoplankton: Measurements from Flow Cytometric Sorting. Limnol Oceanogr 39 (1):169–175. doi:10.4319/lo.1994.39.1.0169.
6. **Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR, Coelho LP, Espinoza JCI, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G**. 2016. Plankton Networks Driving Carbon Export in the Oligotrophic Ocean. Nature 532 (7600):465–470. doi:10.1038/nature16942.
7. **Urbach E, Robertson DL, Chisholm SW**. 1992. Multiple Evolutionary Origins of Prochlorophytes within the Cyanobacterial Radiation. Nature 355 (6357):267–270. doi:10.1038/355267a0.
8. **Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ**. 2003. Clade-Specific 16S Ribosomal DNA Oligonucleotides Reveal the Predominance of a Single Marine *Synechococcus* Clade throughout a

Stratified Water Column in the Red Sea. Appl Environ Microbiol 69 (5):2430–2443. doi:10.1128/aem.69.5.2430-2443.2003.

9. **Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, Not F, Massana R, Ulloa O, Scanlan DJ**. 2008. Global Phylogeography of Marine *Synechococcus* and *Prochlorococcus* Reveals a Distinct Partitioning of Lineages among Oceanic Biomes. Environ Microbiol 10 (1):147–161. doi:10.1111/j.1462-2920.2007.01440.x.
10. **Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C, Ferriera S, Johnson J, Post AF, Hess WR, Partensky F**. 2008. Unraveling the Genomic Mosaic of a Ubiquitous Genus of Marine Cyanobacteria. Genome Biol 9 (5):R90. doi:10/bjx8nx.
11. **Penno S, Lindell D, Post AF**. 2006. Diversity of *Synechococcus* and *Prochlorococcus* Populations Determined from DNA Sequences of the N-regulatory Gene *ntcA*. Environ Microbiol 8 (7):1200–1211. doi:10.1111/j.1462-2920.2006.01010.x.
12. **Mazard S, Ostrowski M, Partensky F, Scanlan DJ**. 2012. Multi-Locus Sequence Analysis, Taxonomic Resolution and Biogeography of Marine *Synechococcus*. Environ Microbiol 14 (2):372–386. doi:10.1111/j.1462-2920.2011.02514.x.
13. **Rocap G, Distel DL, Waterbury JB, Chisholm SW**. 2002. Resolution of *Prochlorococcus* and *Synechococcus* Ecotypes by Using 16S-23S Ribosomal DNA Internal Transcribed Spacer Sequences. Appl Environ Microbiol 68 (3):1180–1191. doi:10.1128/AEM.68.3.1180-1191.2002.
14. **Palenik B**. 1994. Cyanobacterial Community Structure as Seen from RNA Polymerase Gene Sequence Analysis. Appl Environ Microbiol 60 (9):3212–3219. doi:10.1128/aem.60.9.3212-3219.1994.
15. **Mühling M, Fuller NJ, Somerfield PJ, Post AF, Wilson WH, Scanlan DJ, Joint I, Mann NH**. 2006. High Resolution Genetic Diversity Studies of Marine *Synechococcus* Isolates Using *rpoC1*-based Restriction Fragment Length Polymorphism. Aquat Microb Ecol 45 (3):263–275. doi:10/cctgt3.
16. **Xia X, Cheung S, Endo H, Suzuki K, Liu H**. 2019. Latitudinal and Vertical Variation of *Synechococcus* Assemblage Composition Along 170° W Transect From the South Pacific to the Arctic Ocean. Microb Ecol 77 (2):333–342. doi:10.1007/s00248-018-1308-8.
17. **Erlich HA, Gelfand D, Sninsky JJ**. 1991. Recent Advances in the Polymerase Chain Reaction. Science 252 (5013):1643–1651. doi:10.1126/science.2047872.
18. **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 J 12 (5):1360–1374. doi:10.1038/s41396-018-0050-z.
19. **Metz S, Lopes dos Santos A, Berman MC, Bigeard E, Licursi M, Not F, Lara E, Unrein F**. 2019. Diversity of Photosynthetic Picoeukaryotes in Eutrophic Shallow Lakes as Assessed by Combining Flow Cytometry Cell-Sorting and High Throughput Sequencing. FEMS Microbiol Ecol 95 (5):fiz038. doi:10/gj82wp.
20. **Santoferrara L, Burki F, Filker S, Logares R, Dunthorn M, McManus GB**. 2020. Perspectives from Ten Years of Protist Studies by High-Throughput Metabarcoding. J Eukaryot Microbiol 67 (5):612–622. doi:10.1111/jeu.12813.
21. **Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, Roggensack S, Berube PM, Henn MR, Chisholm SW**. 2013. Ecology of Uncultured *Prochlorococcus* Clades Revealed through Single-Cell Genomics and Biogeographic Analysis. The ISME J 7 (1):184–198. doi:10.1038/ismej.2012.89.
22. **Santoro AE, Kellom M, Laperriere SM**. 2019. Contributions of Single-Cell Genomics to Our Understanding of Planktonic Marine Archaea. Philos Trans Royal Soc B: Biol Sci 374 (1786):20190096. doi:10.1098/rstb.2019.0096.
23. **Lomas MW, Baer SE, Mouginot C, Terpis KX, Lomas DA, Altabet MA, Martiny AC**. 2021. Varying Influence of Phytoplankton Biodiversity and Stoichiometric Plasticity on Bulk Particulate Stoichiometry across Ocean Basins. Commun Earth & Environ 2 (1):1–10. doi:10.1038/s43247-021-00212-9.
24. **Hansman RL, Sessions AL**. 2016. Measuring the in Situ Carbon Isotopic Composition of Distinct Marine Plankton Populations Sorted by Flow Cytometry. Limnol Oceanogr Methods 14 (2):87–99. doi:10.1002/lom3.10073.
25. **Duhamel S, Van Wambeke F, Lefevre D, Benavides M, Bonnet S**. 2018. Mixotrophic Metabolism by Natural Communities of Unicellular Cyanobacteria in the Western Tropical South Pacific Ocean. Environ Microbiol 20 (8):2743–2756. doi:10.1111/1462-2920.14111.
26. **Casey JR, Lomas MW, Michelou VK, Dyhrman ST, Orchard ED,**

Ammerman JW, Sylvan JB. 2009. Phytoplankton Taxon-Specific Orthophosphate (Pi) and ATP Utilization in the Western Subtropical North Atlantic. Aquat Microb Ecol 58 (1):31–44. doi:10.3354/ame01348.

27. **Berthelot H, Duhamel S, L'Helguen S, Maguer JF, Wang S, Cetinić I, Cassar N**. 2019. NanoSIMS Single Cell Analyses Reveal the Contrasting Nitrogen Sources for Small Phytoplankton. The ISME J 13 (3):651–662. doi:10.1038/s41396-018-0285-8.
28. **Aldunate M, Henríquez-Castillo C, Ji Q, Lueders-Dumont J, Mulholland MR, Ward BB, von Dassow P, Ulloa O**. 2020. Nitrogen Assimilation in Picocyanobacteria Inhabiting the Oxygen-Deficient Waters of the Eastern Tropical North and South Pacific. Limnol Oceanogr 65 (2):437–453. doi:10.1002/lno.11315.
29. **Rii YM, Duhamel S, Bidigare RR, Karl DM, Repeta DJ, Church MJ**. 2016. Diversity and Productivity of Photosynthetic Picoeukaryotes in Biogeochemically Distinct Regions of the South East Pacific Ocean. Limnol Oceanogr 61 (3):806–824. doi:10/f8nf3k.
30. **Jardillier L, Zubkov MV, Pearman J, Scanlan DJ**. 2010. Significant CO2 Fixation by Small Prymnesiophytes in the Subtropical and Tropical Northeast Atlantic Ocean. The ISME J 4 (9):1180–1192. doi:10.1038/ismej.2010.36.
31. **Hartmann M, Gomez-Pereira P, Grob C, Ostrowski M, Scanlan DJ, Zubkov MV**. 2014. Efficient CO2 Fixation by Surface *Prochlorococcus* in the Atlantic Ocean. The ISME J 8 (11):2280–2289. doi:10.1038/ismej.2014.56.
32. **Farrant GK, Doré H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M, Pitt FD, Wincker P, Scanlan DJ, Iudicone D, Acinas SG, Garczarek L**. 2016. Delineating Ecologically Significant Taxonomic Units from Global Patterns of Marine Picocyanobacteria. Proc Natl Acad Sci 113 (24):E3365–E3374. doi:10.1073/pnas.1524865113.
33. **Ohnemus DC, Rauschenberg S, Krause JW, Brzezinski MA, Collier JL, Geraci-Yee S, Baines SB, Twining BS**. 2016. Silicon Content of Individual Cells of *Synechococcus* from the North Atlantic Ocean. Mar Chem 187:16–24. doi:10.1016/j.marchem.2016.10.003.
34. **Needham DM, Fuhrman JA**. 2016. Pronounced Daily Succession of Phytoplankton, Archaea and Bacteria Following a Spring Bloom. Nat Microbiol 1 (4):16005. doi:10.1038/nmicrobiol.2016.5.
35. **Jeraldo P, Kalari K, Chen X, Bhavsar J, Mangalam A, White B, Nelson H, Kocher JP, Chia N**. 2014. IM-TORNADO: A Tool for Comparison of 16S Reads from Paired-End Libraries. PLOS ONE 9 (12):e114804. doi:10.1371/journal.pone.0114804.
36. **Dacey DP, Chain FJJ**. 2021. Concatenation of Paired-End Reads Improves Taxonomic Classification of Amplicons for Profiling Microbial Communities. BMC Bioinform 22 (1):493. doi:10.1186/s12859-021-04410-2.
37. **Yu G, Fadrosh D, Goedert JJ, Ravel J, Goldstein AM**. 2015. Nested PCR Biases in Interpreting Microbial Community Structure in 16S rRNA Gene Sequence Datasets. PLOS ONE 10 (7):e0132253. doi:10.1371/journal.pone.0132253.
38. **Haro C, Anguita-Maeso M, Metsis M, Navas-Cortés JA, Landa BB**. 2021. Evaluation of Established Methods for DNA Extraction and Primer Pairs Targeting 16S rRNA Gene for Bacterial Microbiota Profiling of Olive Xylem Sap. Front Plant Sci 12:296. doi:10.3389/fpls.2021.640829.
39. **Jing H, Liu H, Suzuki K**. 2009. Phylogenetic Diversity of Marine *Synechococcus* Spp. in the Sea of Okhotsk. Aquat Microb Ecol 56:55–63. doi:10.3354/ame01316.
40. **Xia X, Vidyarathna NK, Palenik B, Lee P, Liu H**. 2015. Comparison of the Seasonal Variations of *Synechococcus* Assemblage Structures in Estuarine Waters and Coastal Waters of Hong Kong. Appl Environ Microbiol 81 (21):7644–7655. doi:10.1128/aem.01895-15.
41. **Xia X, Guo W, Tan S, Liu H**. 2017. *Synechococcus* Assemblages across the Salinity Gradient in a Salt Wedge Estuary. Front Microbiol 8. doi:10.3389/fmicb.2017.01254.
42. **Doré H, Leconte J, Guyet U, Breton S, Farrant GK, Demory D, Ratin M, Hoebeke M, Corre E, Pitt FD, Ostrowski M, Scanlan DJ, Partensky F, Six C, Garczarek L**. 2022. Global Phylogeography of Marine *Synechococcus* in Coastal Areas Reveals Strikingly Different Communities than in the Open Ocean. bioRxiv.
43. **Cabello-Yeves PJ, Picazo A, Camacho A, Callieri C, Rosselli R, Roda-Garcia JJ, Coutinho FH, Rodriguez-Valera F**. 2018. Ecological and Genomic Features of Two Widespread Freshwater Picocyanobacteria. Environ Microbiol 20 (10):3757–3771. doi:10.1111/1462-2920.14377.
44. **Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F**. 2012. Novel Lineages of *Prochlorococcus* and *Synechococcus* in the Global Oceans. The ISME J 6 (2):285–297. doi:10/cn7hc2.

45. **Sohm JA, Ahlgren NA, Thomson ZJ, Williams C, Moffett JW, Saito MA, Webb EA, Rocap G**. 2016. Co-Occurring *Synechococcus* Ecotypes Occupy Four Major Oceanic Regimes Defined by Temperature, Macronutrients and Iron. The ISME J 10 (2):333–345. doi: 10.1038/ismej.2015.115.

46. **Green MR, Sambrook J**. 2019. Nested Polymerase Chain Reaction (PCR). Cold Spring Harb Protoc 2019 (2):pdb.prot095182. doi: 10.1101/pdb.prot095182.

47. **Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ**. 2006. Microbial Diversity in the Deep Sea and the Underexplored "Rare Biosphere". Proc Natl Acad Sci United States Am 103 (32):12115–12120. doi:10/bwrmgj.

48. **Cermeño P, Teixeira IG, Branco M, Figueiras FG, Marañón E**. 2014. Sampling the Limits of Species Richness in Marine Phytoplankton Communities. J Plankton Res 36 (4):1135–1139. doi: 10.1093/plankt/fbu033.

49. **Kent AG, Baer SE, Mouginot C, Huang JS, Larkin AA, Lomas MW, Martiny AC**. 2019. Parallel Phylogeography of *Prochlorococcus* and *Synechococcus*. The ISME J 13 (2):430–441. doi:10.1038/s41396-018-0287-6.

50. **Boyd P, LaRoche J, Gall M, Frew R, McKay RML**. 1999. Role of Iron, Light, and Silicate in Controlling Algal Biomass in Subantarctic Waters SE of New Zealand. J Geophys Res Ocean 104 (C6):13395–13408. doi:10.1029/1999JC900009.

51. **Longhurst AR**. 2007. Ecological Geography of the Sea. Second ed Elsevier.

52. **Caputi L, Carradec Q, Eveillard D, Kirilovsky A, Pelletier E, Pierella Karlusich JJ, Rocha Jimenez Vieira F, Villar E, Chaffron S, Malviya S, Scalco E, Acinas SG, Alberti A, Aury JM, Benoiston AS, Bertrand A, Biard T, Bittner L, Boccara M, Brum JR, Brunet C, Busseni G, Carratalà A, Claustre H, Coelho LP, Colin S, D'Aniello S, Da Silva C, Del Core M, Doré H, Gasparini S, Kokoszka F, Jamet JL, Lejeusne C, Lepoivre C, Lescot M, Lima-Mendez G, Lombard F, Lukeš J, Maillet N, Madoui MA, Martinez E, Mazzocchi MG, Néou MB, Paz-Yepes J, Poulain J, Ramondenc S, Romagnan JB, Roux S, Salvagio Manta D, Sanges R, Speich S, Sprovieri M, Sunagawa S, Taillandier V, Tanaka A, Tirichine L, Trottier C, Uitz J, Veluchamy A, Veselá J, Vincent F, Yau S, Kandels-Lewis S, Searson S, Dimier C, Picheral M, Coordinators TO, Bork P, Boss E, de Vargas C, Follows MJ, Grimsley N, Guidi L, Hingamp P, Karsenti E, Sordino P, Stemmann L, Sullivan MB, Tagliabue A, Zingone A, Garczarek L, d'Ortenzio F, Testor P, Not F, d'Alcalà MR, Wincker P, Bowler C, Iudicone D**. 2019. Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. Glob Biogeochem Cycles 33 (3):391–419. doi:10/gftgkp.

53. **Ahlgren NA, Belisle BS, Lee MD**. 2020. Genomic Mosaicism Underlies the Adaptation of Marine *Synechococcus* Ecotypes to Distinct Oceanic Iron Niches. Environ Microbiol 22 (5):1801–1815. doi: 10.1111/1462-2920.14893.

54. **Heath RA**. 1983. Observations on Chatham Rise Currents. New Zealand J Mar Freshw Res 17 (3):321–330. doi: 10.1080/00288330.1983.9516006.

55. **Chiswell SM**. 1994. Acoustic Doppler Current Profiler Measurements over the Chatham Rise. New Zealand J Mar Freshw Res 28 (2):167–178. doi:10.1080/00288330.1994.9516605.

56. **Stramma L, Peterson RG, Tomczak M**. 1995. The South Pacific Current. J Phys Oceanogr 25 (1):77–91. doi:10.1175/1520-0485(1995)025<0077:TSPC>2.0.CO;2.

57. **Lüskow F, Pakhomov EA, Stukel MR, Décima M**. 2020. Biology of Salpa Thompsoni at the Chatham Rise, New Zealand: Demography, Growth, and Diel Vertical Migration. Mar Biol 167 (12):175. doi: 10.1007/s00227-020-03775-x.

58. **Hall JA, James MR, Bradford-Grieve JM**. 1999. Structure and Dynamics of the Pelagic Microbial Food Web of the Subtropical Convergence Region East of New Zealand. Aquat Microb Ecol 20 (1):95–105. doi: 10.3354/ame020095.

59. **Landry MR, Ohman MD, Goericke R, Stukel MR, Tsyrklevich K**. 2009. Lagrangian Studies of Phytoplankton Growth and Grazing Relationships in a Coastal Upwelling Ecosystem off Southern California. Prog Oceanogr 83 (1-4):208–216. doi:10.1016/j.pocean.2009.07.026.

60. **Katoh K, Standley DM**. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30 (4):772–780. doi:10.1093/molbev/mst010.

61. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A**. 2012. Geneious Basic: An Integrated

and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. Bioinformatics 28 (12):1647–1649. doi: 10.1093/bioinformatics/bts199.

62. **Price MN, Dehal PS, Arkin AP**. 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles Instead of a Distance Matrix. Mol Biol Evol 26 (7):1641–1650. doi:10.1093/molbev/msp077.
63. **RStudio Team**. 2021. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.
64. **Martin M**. 2011. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. EMBnet.journal 17 (1):10–12. doi: 10.14806/ej.17.1.200.
65. **Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP**. 2016. DADA2: High-resolution Sample Inference from Illumina Amplicon Data. Nat Methods 13 (7):581–583. doi: 10.1038/nmeth.3869.
66. **McMurdie PJ, Holmes S**. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8 (4):e61217. doi:10.1371/journal.pone.0061217.
67. **Amante C, Eakins BW**. ETOPO1 1 Arc-Minute Global Relief Model Procedures, Data Sources and Analysis .
68. **Vihtakari M**. 2022. ggOceanMaps: Plot Data on Oceanographic Maps using 'ggplot2'. https://CRAN.R-project.org/package=ggOceanMaps. R package version 1.2.6.

TABLES

TABLE 1 Primers sequence and PCR conditions used for amplification. Based on the *Synechococcus* sp. WH 8109 (CP006882), the *petB* gene is located between position 440,919 and 441,575 of the genome. The total length of the *petB* gene is 657bp. The amplicon length includes the primers. The position of the primer in *petB* gene is indicated in reference to the first nucleotide of the gene. Deg = Degeneracy, T(°C) = Annealing temperature

	PCR type	Round	Amplicon length	Primer name	Direction	Sequence	Gene position	GC%	Deg	T (°C)	Citation
Mazard_2012	Standard	1	597	petB-F	Forward	TACGACTGGTTCCAGGAACG	22–41	55	0	55	Mazard et al. (12)
				petB-R	Reverse	GAAGTGCATGAGCATGAA	601–618	44	0		Mazard et al. (12)
Ong_2022	Nested	1	613	petB-F	Forward	TACGACTGGTTCCAGGAACG	22–41	55	0	59	Mazard et al. (12)
				petB-634R	Reverse	GCTTVCGRATCATCARGA	615–634	47	12		This study
		2	569	petB-50F	Forward	CAGGACATYGCTGAY	50–66	50	8	55	This study
				petB-R	Reverse	GAAGTGCATGAGCATGAA	601–618	44	0		Mazard et al. (12)

TABLE 2 Number of reads and ASVs per sample (Mean ± SD) after each sequence processing step using DADA2 on filtered samples obtained with Mazard_2012 and Ong_2022 approaches.

	DADA2 steps		No. of reads/ASVs	
			Mazard_2012	Ong_2022
No. of reads per sample	1	Initial	57,321 ± 5,918	46,382 ± 4,614
	2	Remove primers	47,299 ± 5,718	34,524 ± 4,174
	3	Filter and trim	45,072 ± 6,006	32,274 ± 4,327
	4	Merged	43,653 ± 6,015	30,903 ± 4,402
	5	Remove chimera	28,237 ± 5,908	10,500 ± 2,843
No. of ASVs		Total	4,716	6,307
		Per sample	361 ± 64	217 ± 63

FIGURES

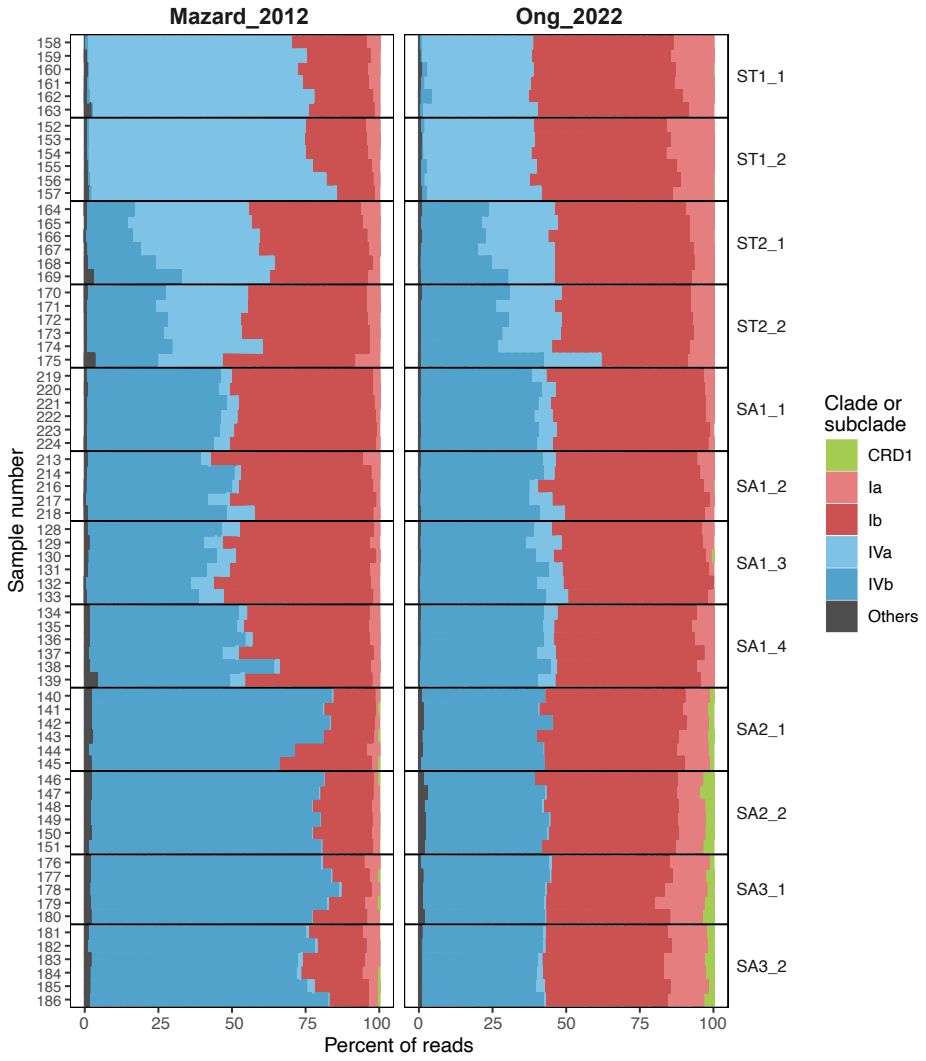

FIG 1 ***Synechococcus*** **taxonomic composition at clade and subclade level from filtered samples obtained with Mazard_2012 and Ong_2022 approaches. Samples were grouped by cycles and ordered across a spatial gradient, from subtropical (ST) to subantarctic (SA) cycles. Samples from all depths were combined.**

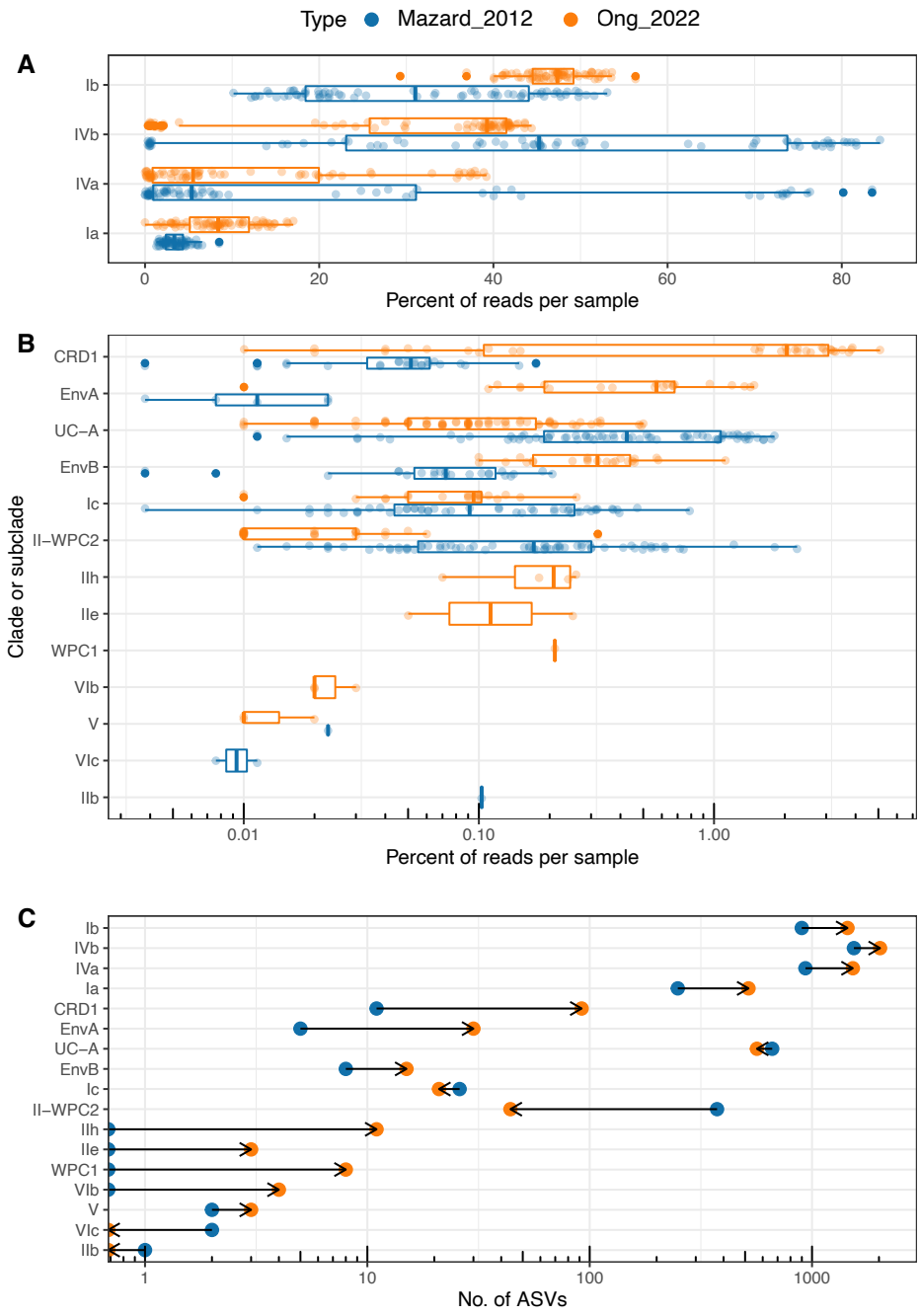

FIG 2 **Percentage of *Synechococcus* reads in each filtered sample for major subclades (A) and minor clades and subclades (B), and total number of ASVs (C) for each clade and subclade with Mazard_2012 and Ong_2022. Clades and subclades are arranged in descending order according to the mean percentage of reads from Ong_2022.**

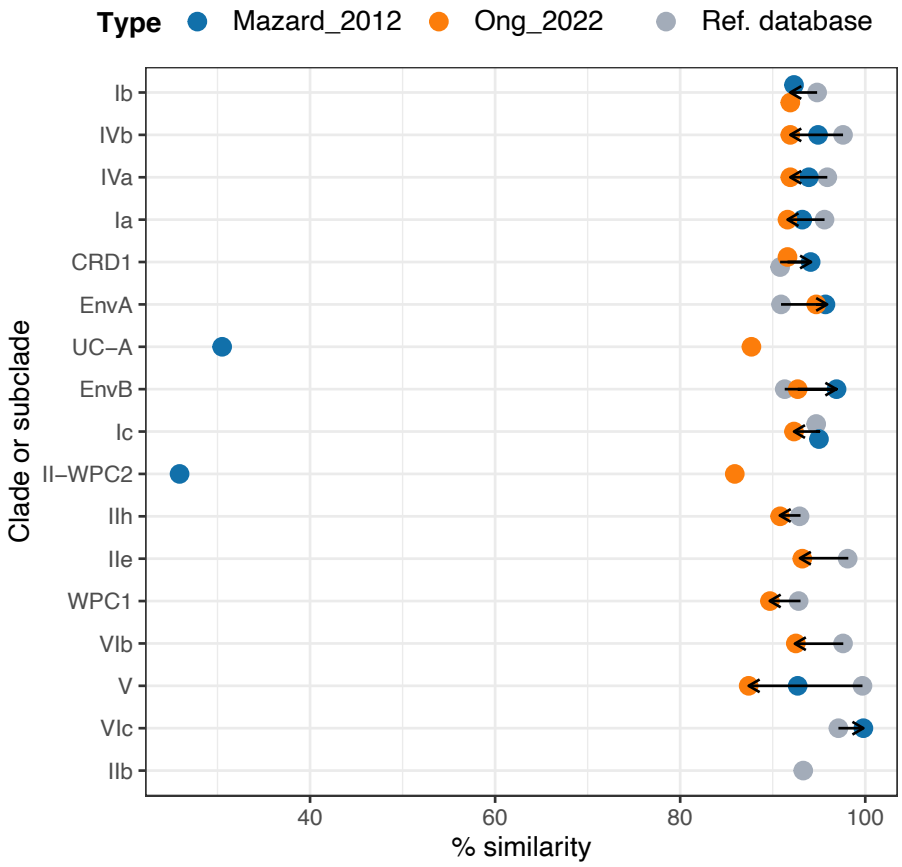

FIG 3 ASV nucleotide similarity within each clade and subclade obtained with Mazard_2012 and Ong_2022 approaches and reference sequence database.

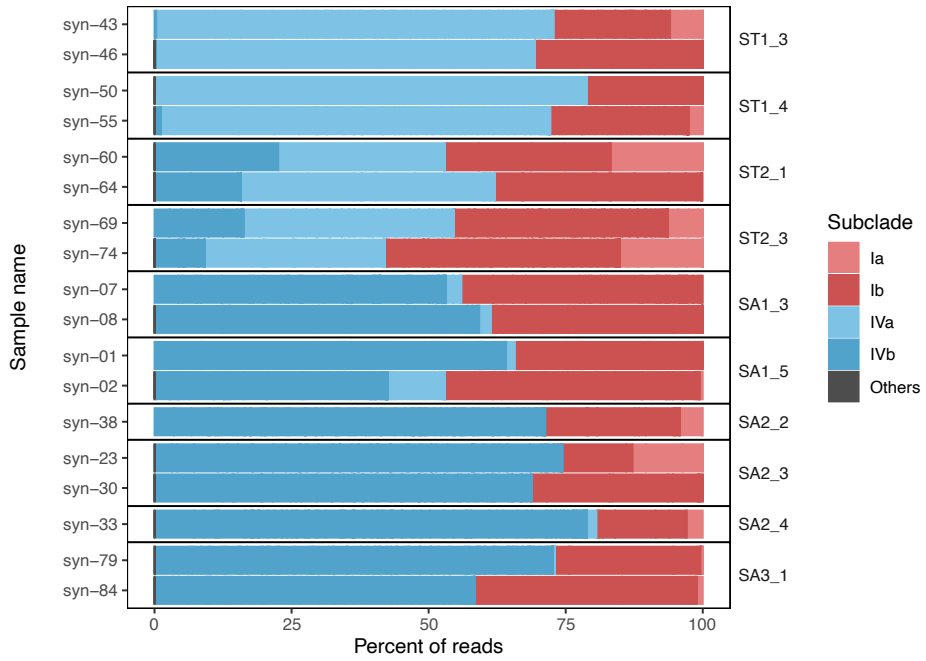

FIG 4 ***Synechococcus*** **taxonomic composition at clade and subclade level from flow cytometry sorted populations. Samples were grouped by cycles and ordered across a spatial gradient, from subtropical (ST) to subantarctic (SA) cycles. The SUR and DCM depths samples were combined.**

List of Tables

List of Figures

SUPPLEMENTAL MATERIAL

TABLE S1 Location and date of samples.

						Filtered seawater			Sorted *Synechococcus*	
Date	Latitude	Longitude	cycle_station	CTD	Depth (m)	Sample number	Mazard_2012	Ong_2022	Sample name	No. of cells
29/10/2018	-44.54	174.11	SA1_1	U9115	05	219	CTD-petB-219	CTD-NpetB-219		
29/10/2018	-44.54	174.11	SA1_1	U9115	12	220	CTD-petB-220	CTD-NpetB-220		
29/10/2018	-44.54	174.11	SA1_1	U9115	20	221	CTD-petB-221	CTD-NpetB-221		
29/10/2018	-44.54	174.11	SA1_1	U9115	30	222	CTD-petB-222	CTD-NpetB-222		
29/10/2018	-44.54	174.11	SA1_1	U9115	40	223	CTD-petB-223	CTD-NpetB-223		
29/10/2018	-44.54	174.11	SA1_1	U9115	50	224	CTD-petB-224	CTD-NpetB-224		
27/10/2018	-44.60	174.21	SA1_2	U9109	05	213	CTD-petB-213	CTD-NpetB-213		
27/10/2018	-44.60	174.21	SA1_2	U9109	12	214	CTD-petB-214	CTD-NpetB-214		
27/10/2018	-44.60	174.21	SA1_2	U9109	20	215		CTD-NpetB-215		
27/10/2018	-44.60	174.21	SA1_2	U9109	30	216	CTD-petB-216	CTD-NpetB-216		
27/10/2018	-44.60	174.21	SA1_2	U9109	40	217	CTD-petB-217	CTD-NpetB-217		
27/10/2018	-44.60	174.21	SA1_2	U9109	50	218	CTD-petB-128-1	CTD-NpetB-218		
26/10/2018	-44.60	174.21	SA1_3	U9106	05	128	CTD-petB-128	CTD-NpetB-128		
26/10/2018	-44.60	174.21	SA1_3	U9106	12	129	CTD-petB-129	CTD-NpetB-129	syn-07	10,000
26/10/2018	-44.60	174.21	SA1_3	U9106	20	130	CTD-petB-130	CTD-NpetB-130		
26/10/2018	-44.60	174.21	SA1_3	U9106	30	131	CTD-petB-131	CTD-NpetB-131		
26/10/2018	-44.60	174.21	SA1_3	U9106	40	132	CTD-petB-132	CTD-NpetB-132	syn-08	10,000
26/10/2018	-44.60	174.21	SA1_3	U9106	50	133	CTD-petB-133	CTD-NpetB-133		
31/10/2018	-44.64	174.27	SA1_4	U9122	05	134	CTD-petB-134	CTD-NpetB-134		
31/10/2018	-44.64	174.27	SA1_4	U9122	12	135	CTD-petB-135	CTD-NpetB-135		
31/10/2018	-44.64	174.27	SA1_4	U9122	20	136	CTD-petB-136	CTD-NpetB-136		
31/10/2018	-44.64	174.27	SA1_4	U9122	30	137	CTD-petB-137	CTD-NpetB-137		
31/10/2018	-44.64	174.27	SA1_4	U9122	40	138	CTD-petB-138	CTD-NpetB-138		
31/10/2018	-44.64	174.27	SA1_4	U9122	50	139	CTD-petB-139	CTD-NpetB-139		
25/10/2018	-44.60	174.51	SA1_5	U9103	12				syn-01-10	10,000
25/10/2018	-44.60	174.51	SA1_5	U9103	12				syn-01-20	20,000
25/10/2018	-44.60	174.51	SA1_5	U9103	40				syn-02-1	1,000
25/10/2018	-44.60	174.51	SA1_5	U9103	40				syn-02-5	5,000
25/10/2018	-44.60	174.51	SA1_5	U9103	40				syn-02-10	10,000
02/11/2018	-44.56	178.48	SA2_1	U9125	05	140	CTD-petB-140	CTD-NpetB-140		
02/11/2018	-44.56	178.48	SA2_1	U9125	12	141	CTD-petB-141	CTD-NpetB-141		
02/11/2018	-44.56	178.48	SA2_1	U9125	20	142	CTD-petB-142	CTD-NpetB-142		
02/11/2018	-44.56	178.48	SA2_1	U9125	30	143	CTD-petB-143	CTD-NpetB-143		
02/11/2018	-44.56	178.48	SA2_1	U9125	40	144	CTD-petB-144	CTD-NpetB-144		
02/11/2018	-44.56	178.48	SA2_1	U9125	60	145	CTD-petB-145	CTD-NpetB-145		
05/11/2018	-44.54	179.49	SA2_2	U9136	05	146	CTD-petB-146	CTD-NpetB-146		
05/11/2018	-44.54	179.49	SA2_2	U9136	12	147	CTD-petB-147	CTD-NpetB-147	syn-38	2,000
05/11/2018	-44.54	179.49	SA2_2	U9136	20	148	CTD-petB-148	CTD-NpetB-148		
05/11/2018	-44.54	179.49	SA2_2	U9136	30	149	CTD-petB-149	CTD-NpetB-149		

TABLE S1 *(continued)*

						Filtered seawater			Sorted *Synechococcus*	
Date	Latitude	Longitude	cycle_station	CTD	Depth (m)	Sample number	Mazard_2012	Ong_2022	Sample name	No. of cells
05/11/2018	-44.54	179.49	SA2_2	U9136	40	150	CTD-petB-150	CTD-NpetB-150		
05/11/2018	-44.54	179.49	SA2_2	U9136	60	151	CTD-petB-151	CTD-NpetB-151		
03/11/2018	-44.57	178.68	SA2_3	U9128	12				syn-23	2,000
03/11/2018	-44.57	178.68	SA2_3	U9128	40				syn-30	2,000
05/11/2018	-44.60	179.18	SA2_4	U9133	12				syn-33	2,000
16/11/2018	-45.56	179.52	SA3_1	U9161	05	176	CTD-petB-176	CTD-NpetB-176		
16/11/2018	-45.56	179.52	SA3_1	U9161	12	177	CTD-petB-177	CTD-NpetB-177	syn-79	2,000
16/11/2018	-45.56	179.52	SA3_1	U9161	25	178	CTD-petB-178	CTD-NpetB-178		
16/11/2018	-45.56	179.52	SA3_1	U9161	45	179	CTD-petB-179	CTD-NpetB-179		
16/11/2018	-45.56	179.52	SA3_1	U9161	70	180	CTD-petB-180	CTD-NpetB-180	syn-84	2,000
18/11/2018	-45.55	179.51	SA3_2	U9171	05	181	CTD-petB-181	CTD-NpetB-181		
18/11/2018	-45.55	179.51	SA3_2	U9171	12	182	CTD-petB-182	CTD-NpetB-182		
18/11/2018	-45.55	179.51	SA3_2	U9171	30	183	CTD-petB-183	CTD-NpetB-183		
18/11/2018	-45.55	179.51	SA3_2	U9171	50	184	CTD-petB-184	CTD-NpetB-184		
18/11/2018	-45.55	179.51	SA3_2	U9171	60	185	CTD-petB-185	CTD-NpetB-185		
18/11/2018	-45.55	179.51	SA3_2	U9171	70	186	CTD-petB-186	CTD-NpetB-186		
10/11/2018	-42.78	178.41	ST1_1	U9148	05	158	CTD-petB-158	CTD-NpetB-158		
10/11/2018	-42.78	178.41	ST1_1	U9148	12	159	CTD-petB-159	CTD-NpetB-159		
10/11/2018	-42.78	178.41	ST1_1	U9148	20	160	CTD-petB-160	CTD-NpetB-160		
10/11/2018	-42.78	178.41	ST1_1	U9148	30	161	CTD-petB-161	CTD-NpetB-161		
10/11/2018	-42.78	178.41	ST1_1	U9148	40	162	CTD-petB-162	CTD-NpetB-162		
10/11/2018	-42.78	178.41	ST1_1	U9148	50	163	CTD-petB-163	CTD-NpetB-163		
07/11/2018	-42.66	178.00	ST1_2	U1938	05	152	CTD-petB-152	CTD-NpetB-152		
07/11/2018	-42.66	178.00	ST1_2	U1938	12	153	CTD-petB-153	CTD-NpetB-153		
07/11/2018	-42.66	178.00	ST1_2	U1938	20	154	CTD-petB-154	CTD-NpetB-154		
07/11/2018	-42.66	178.00	ST1_2	U1938	25	155	CTD-petB-155	CTD-NpetB-155		
07/11/2018	-42.66	178.00	ST1_2	U1938	30	156	CTD-petB-156	CTD-NpetB-156		
07/11/2018	-42.66	178.00	ST1_2	U1938	40	157	CTD-petB-157	CTD-NpetB-157		
08/11/2018	-42.74	178.09	ST1_3	U1941	12				syn-43	2,000
08/11/2018	-42.74	178.09	ST1_3	U1941	25				syn-46	2,000
09/11/2018	-42.79	178.27	ST1_4	U9144	12				syn-50	2,000
09/11/2018	-42.79	178.27	ST1_4	U9144	40				syn-55	2,000
12/11/2018	-42.78	178.41	ST2_1	U9149	05	164	CTD-petB-164	CTD-NpetB-164		
12/11/2018	-42.78	178.41	ST2_1	U9149	12	165	CTD-petB-165	CTD-NpetB-165	syn-60	2,000
12/11/2018	-42.78	178.41	ST2_1	U9149	20	166	CTD-petB-166	CTD-NpetB-166		
12/11/2018	-42.78	178.41	ST2_1	U9149	30	167	CTD-petB-167	CTD-NpetB-167	syn-64	2,000
12/11/2018	-42.78	178.41	ST2_1	U9149	40	168	CTD-petB-168	CTD-NpetB-168		
12/11/2018	-42.78	178.41	ST2_1	U9149	50	169	CTD-petB-169	CTD-NpetB-169		
15/11/2018	-43.73	-179.82	ST2_2	U9159	05	170	CTD-petB-170	CTD-NpetB-170		
15/11/2018	-43.73	-179.82	ST2_2	U9159	12	171	CTD-petB-171	CTD-NpetB-171		

TABLE S1 *(continued)*

						Filtered seawater			Sorted *Synechococcus*	
Date	Latitude	Longitude	cycle_station	CTD	Depth (m)	Sample number	Mazard_2012	Ong_2022	Sample name	No. of cells
15/11/2018	-43.73	-179.82	ST2_2	U9159	20	172	CTD-petB-172	CTD-NpetB-172		
15/11/2018	-43.73	-179.82	ST2_2	U9159	30	173	CTD-petB-173	CTD-NpetB-173		
15/11/2018	-43.73	-179.82	ST2_2	U9159	40	174	CTD-petB-174	CTD-NpetB-174		
15/11/2018	-43.73	-179.82	ST2_2	U9159	50	175	CTD-petB-175	CTD-NpetB-175		
13/11/2018	-43.48	179.94	ST2_3	U9152	12				syn-69	2,000
13/11/2018	-43.48	179.94	ST2_3	U9152	40				syn-74	2,000

TABLE S2 Number of ASVs and percentage of reads at clade and subclade level for 70 distinct samples: mean, maximum, minimum and standard deviation, and the total number of ASVs using Mazard_2012 and Ong_2022. Significant differences of the percent of reads per clade or subclade between the two approaches were tested using a paired sample Wilcoxon test (P-value < 0.05). The rows are arranged in descending order according to the mean percentage of reads from Ong_2022.

			Mazard_2012					Ong_2022					
				% of reads					% of reads				
	Clade	Subclade	No. of ASVs	Mean	Max	Min	SD	No. of ASVs	Mean	Max	Min	SD	P-value
1	I	Ib	899	31.04	53.07	10.16	13.46	1444	46.73	56.32	29.30	4.33	4.07E-12
2	IV	IVb	1543	44.04	84.40	0.33	28.00	2023	31.40	44.39	0.35	14.99	2.33E-06
3	IV	IVa	934	20.39	83.48	0.00	27.60	1526	11.96	39.21	0.01	13.45	6.17E-03
4	I	Ia	249	3.47	8.52	1.32	1.42	518	8.51	17.03	0.00	4.26	3.30E-12
5	CRD1		11	0.02	0.18	0.00	0.03	92	0.93	5.08	0.00	1.40	5.28E-11
6	EnvA		5	9.78E-04	0.02	0.00	4.13E-03	30	0.18	1.48	0.00	0.37	1.67E-06
7	UC-A		662	0.63	1.81	0.01	0.52	565	0.13	0.50	9.98E-03	0.11	1.72E-09
8	EnvB		8	0.03	0.21	0.00	0.05	15	0.11	1.12	0.00	0.21	2.38E-07
9	I	Ic	26	0.12	0.79	0.00	0.15	21	0.03	0.26	0.00	0.05	2.57E-09
10	II	II-WPC2	374	0.26	2.26	0.00	0.38	44	0.01	0.32	0.00	0.04	1.20E-11
11	II	IIh	0					11	0.01	0.26	0.00	0.05	
12	II	IIe	0					3	4.30E-03	0.25	0.00	0.03	
13	WPC1		0					8	3.01E-03	0.21	0.00	0.03	
14	VI	VIb	0					4	1.00E-03	0.03	0.00	4.86E-03	
15	V		2	3.26E-04	0.02	0.00	2.73E-03	3	5.71E-04	0.02	0.00	2.89E-03	0.87
16	VI	VIc	2	2.72E-04	0.01	0.00	1.63E-03	0					
17	II	IIb	1	1.47E-03	0.10	0.00	0.01	0					

TABLE S3 The mean percentage of reads for each clade or subclade in each cycle, using Mazard_2012 and Ong_2022 amplification protocol. The rows are arranged in descending order according to the mean percentage of reads of each clade or subclade from Ong_2022.

			Cycle				
Clade	Subclade	Amplification protocol	ST1	ST2	SA1	SA2	SA3
I	Ib	Mazard_2012	20.62	38.76	45.92	19.05	15.96
		Ong_2022	48.00	44.53	50.31	46.08	40.98
IV	IVb	Mazard_2012	0.56	22.57	45.45	76.26	76.81
		Ong_2022	1.31	26.66	40.20	40.46	41.11
IV	IVa	Mazard_2012	74.72	33.13	5.10	0.33	1.06
		Ong_2022	36.98	20.54	5.68	0.39	1.06
I	Ia	Mazard_2012	3.45	4.74	2.76	2.51	4.67
		Ong_2022	13.37	8.00	3.71	8.93	13.33
CRD1		Mazard_2012	0.00	0.00	1.65E-04	0.04	0.07
		Ong_2022	0.03	7.52E-03	0.01	2.79	2.79
EnvA		Mazard_2012	0.00	0.00	0.00	2.85E-03	3.11E-03
		Ong_2022	0.00	8.35E-04	4.78E-03	0.76	0.29
UC-A		Mazard_2012	0.22	0.43	0.42	1.13	1.16
		Ong_2022	0.15	0.23	0.08	0.14	0.07
EnvB		Mazard_2012	0.00	0.00	0.00	0.07	0.11
		Ong_2022	0.00	0.00	0.00	0.34	0.35
I	Ic	Mazard_2012	0.26	5.07E-03	0.04	0.30	0.04
		Ong_2022	0.05	0.00	3.05E-03	0.09	3.64E-03
II	II-WPC2	Mazard_2012	0.17	0.36	0.31	0.30	0.13
		Ong_2022	5.85E-03	0.04	5.65E-03	6.67E-03	0.01
II	IIh	Mazard_2012	0.00	0.00	0.00	0.00	0.00
		Ong_2022	0.06	0.00	0.00	0.00	0.00
II	IIe	Mazard_2012	0.00	0.00	0.00	0.00	0.00
		Ong_2022	0.03	0.00	0.00	0.00	0.00
WPC1		Mazard_2012	0.00	0.00	0.00	0.00	0.00
		Ong_2022	0.02	0.00	0.00	0.00	0.00
VI	VIb	Mazard_2012	0.00	0.00	0.00	0.00	0.00
		Ong_2022	0.00	0.00	1.31E-03	1.66E-03	1.82E-03
V		Mazard_2012	0.00	0.00	9.93E-04	0.00	0.00
		Ong_2022	0.00	0.00	0.00	3.33E-03	0.00
VI	VIc	Mazard_2012	1.59E-03	0.00	0.00	0.00	0.00
		Ong_2022	0.00	0.00	0.00	0.00	0.00
II	IIb	Mazard_2012	0.00	0.00	4.47E-03	0.00	0.00
		Ong_2022	0.00	0.00	0.00	0.00	0.00

TABLE S4 Percent similarity of nucleotide sequences/ASVs and percent of correctly assigned ASVs in each clade or subclade from Mazard_2012 and Ong_2022 approaches on filtered samples and reference database.

		No. of	% similarity of nucleotide seq			% of ASVs correctly assigned	
Clade	Subclade	ref seq	Ref seq	Mazard_2012	Ong_2022	Mazard_2012	Ong_2022
I	Ib	55	94.8	92.3	91.9	98.0	99.6
IV	IVb	20	97.6	94.9	91.9	99.3	96.7
IV	IVa	21	95.9	93.9	91.9	99.5	98.4
I	Ia	19	95.6	93.2	91.6	99.2	98.5
CRD1		23	90.8	94.1	91.6	100	93.5
EnvA		10	90.9	95.7	94.7	100	100
UC-A		1		30.5	87.7	0	98.1
EnvB		13	91.3	96.9	92.7	100	100
I	Ic	18	94.7	95.0	92.3	96.2	90.5
II	II-WPC2	1		25.9	85.9	2.4	97.7
II	IIh	10	92.9		90.8		100
II	IIe	4	98.1		93.2		100
WPC1		11	92.8		89.7		100
VI	VIb	20	97.6		92.5		100
V		2	99.7	92.7	87.4	100	100
VI	VIc	4	97.1	99.8		100	
II	IIb	13	93.3	N/A		100	

TABLE S5 Percentage of reads at clade and subclade level across 18 samples: mean, maximum, minimum and standard deviation, and the total number of ASVs of sorted *Synechococcus* cells. The rows are arranged in descending order according to the mean percentage of reads.

				% of reads			
	Clade	Subclade	No.of ASVs	Mean	Max	Min	SD
1	IV	IVb	413	39.35	78.82	0.00	30.39
2	I	Ib	274	31.33	46.38	12.84	9.82
3	IV	IVa	244	25.47	78.76	0.00	30.04
4	I	Ia	37	3.84	16.84	0.00	5.55
5	UC-A		24	8.11E-03	0.05	0.00	0.01
6	CRD1		2	3.34E-03	0.05	0.00	0.01
7	II	II-WPC2	3	2.15E-03	0.04	0.00	0.01
8	EnvB		1	2.15E-03	0.04	0.00	0.01

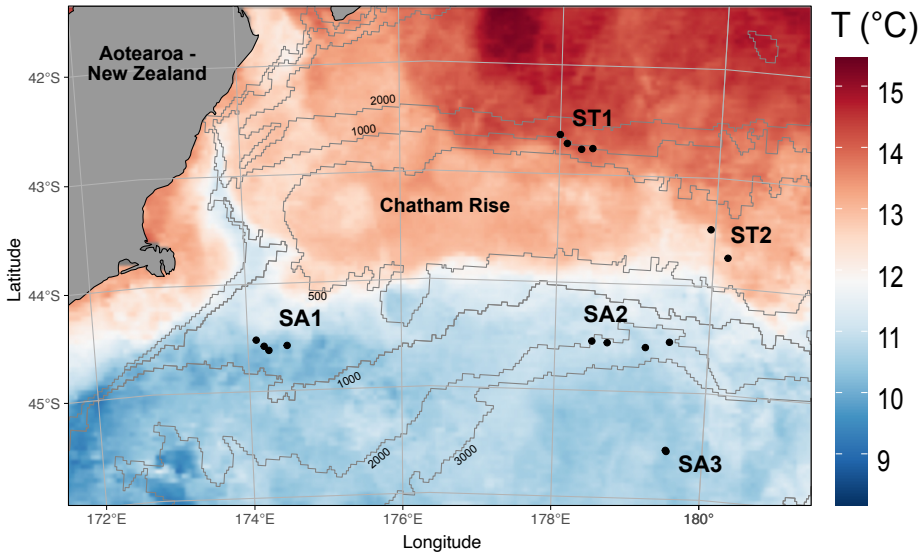

FIG S1 Study area with cycle locations at the Chatham Rise east of Aotearoa-New Zealand. Black dots represent the sampling location within each cycle. Bathymetric lines (m) are indicated. Bathymetry was obtained from Amante & Eakins (67) and plotted with ggOceanMaps (68). Sea surface temperature data (°C) was obtained from MODIS (NASA) and averaged over November 2018. ST and SA refers to subtropical and sub-antarctic respectively and numbers to the cycle

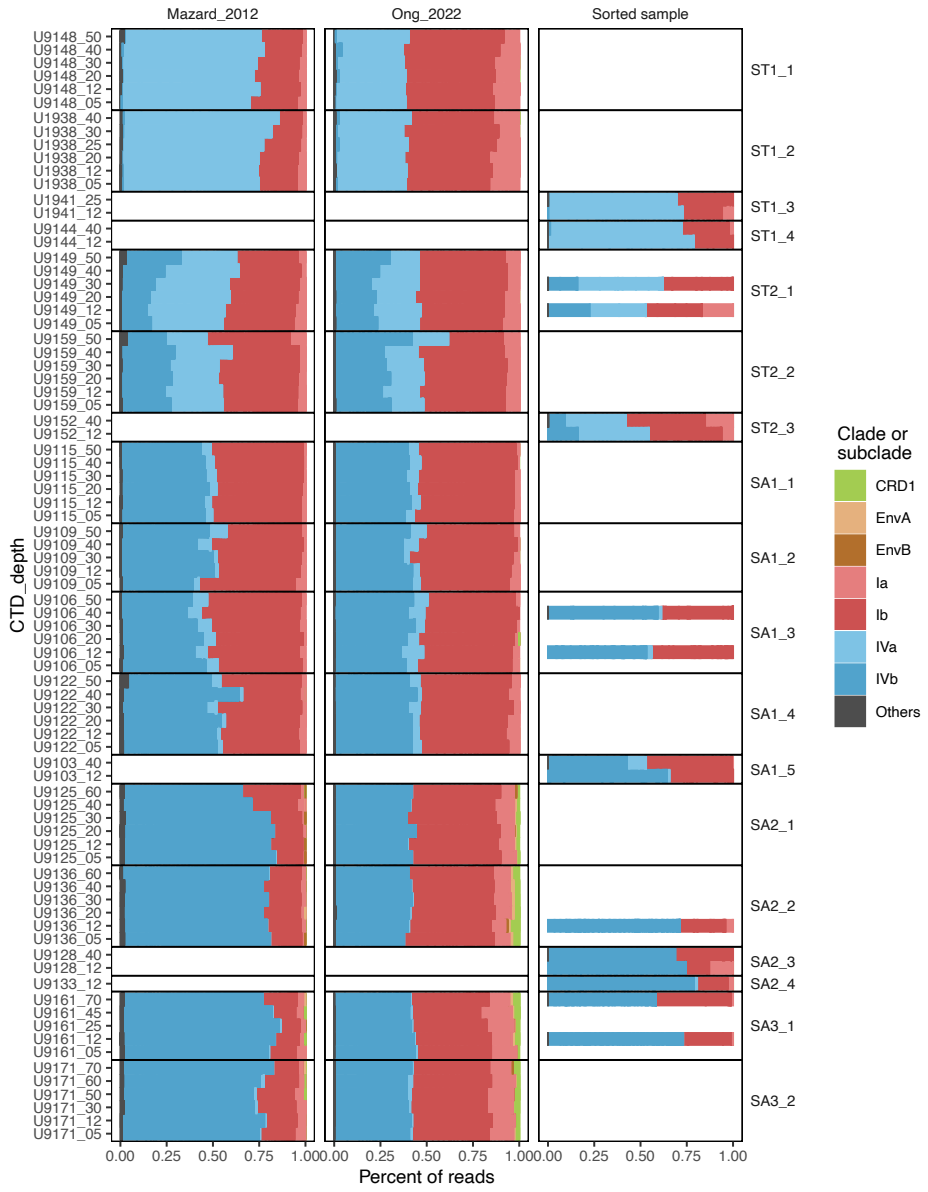

FIG S2 ***Synechococcus*** **taxonomic composition at clade and subclade level using Mazard_2012 and Ong_2022 protocol on filtered seawater samples and sorted** ***Synechococcus*** **cells. Samples were grouped by station. Each station was labelled as 'cycle'_'station' and ordered across a spatial gradient, from subtropical (ST) to subantarctic (SA) cycles.**