

Nested PCR Approach for *petB* Gene Metabarcoding of Marine *Synechococcus* Populations

🐵 Denise Rui Ying Ong, ª 💿 Andrés Gutiérrez-Rodríguez, 🏼 🕲 Laurence Garczarek, 🕻 Dominique Marie, 🕻 💿 Adriana Lopes dos Santos ª

^aAsian School of the Environment, Nanyang Technological University, Singapore ^bNational Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand ^cSorbonne Université, CNRS, UMR 7144 (AD2M), Station Biologique de Roscoff, Roscoff, France

Microbiology Spectrum

AMERICAN SOCIETY FOR MICROBIOLOGY

ABSTRACT The molecular diversity of marine picocyanobacterial populations, an important component of phytoplankton communities, is better characterized using high-resolution marker genes than the 16S rRNA gene as they have 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 characterize Synechococcus diversity. We have designed new primers targeting the petB gene and proposed a nested PCR method (termed 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 (termed Mazard_2012) using filtered seawater samples. The 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) water masses. The two PCR approaches using filtered samples recovered the same dominant subclades, Ia, Ib, IVa, and IVb, with small differences in relative abundance across the distinct samples. For example, subclade IVa was dominant in ST samples with the Mazard_2012 approach, while the same samples processed with Ong_2022 showed similar contributions of subclades IVa and Ib to the total community. The Ong_2022 approach generally captured a higher genetic diversity of Synechococcus subcluster 5.1 than the Mazard_2012 approach while having a lower proportion of incorrectly assigned amplicon sequence variants (ASVs). All flow cytometry-sorted Synechococcus samples could be amplified only by our nested approach. The taxonomic diversity obtained with our primers on both sample types was in agreement with the clade distribution observed by previous studies that applied other marker genes or PCR-free metagenomic approaches under similar environmental conditions.

IMPORTANCE The *petB* gene has been proposed as a high-resolution marker gene to access the diversity of marine *Synechococcus* populations. A systematic metabarcoding approach based on the *petB* gene would improve the characterization/assessment of the *Synechococcus* community structure in marine planktonic ecosystems. We have designed and tested specific primers to be applied in a nested PCR protocol (Ong_2022) for metabarcoding the *petB* gene. The Ong_2022 protocol can be applied to samples with low DNA content, such as those obtained by flow cytometry cell sorting, allowing the simultaneous assessment of the genetic diversity of *Synechococcus* populations and 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

Editor Blaire Steven, Connecticut Agricultural Experiment Station

Copyright © 2023 Ong et al. This is an openaccess article distributed under the terms of the Creative Commons Attribution 4.0 International license.

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

*Present address: Andrés Gutiérrez-Rodríguez, Instituto Español de Oceanografía, Centro Oceanográfico de Gijón, Gijón, Spain.

The authors declare no conflict of interest. Received 14 October 2022 Accepted 30 January 2023 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 physicochemical properties of marine waters induced by climate change are predicted to have an impact on the distribution of phytoplankton groups, including *Synechococcus* (4). Given the major role of these phytoplankton groups in marine carbon cycling and the ocean food web (5, 6), an extensive assessment of *Synechococcus* community genetic diversity at spatial and temporal scales as well as the ecological traits of this community is critical to understand the future evolution of this group under the ongoing global change.

While studies using the 16S rRNA gene have revealed some degree of diversity within *Prochlorococcus* and *Synechococcus* populations (7–9), 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 analyze the genetic diversity within these groups of picocyanobacteria (e.g., *ntcA* [11], *petB* [12], internal transcribed spacer [ITS] [13], and *rpoC1* [14]). Although most of these gene markers have been amplified from natural *Synechococcus* populations by standard PCR amplification, others like *rpoC1* require a nested PCR approach (15, 16).

Nested and semi-nested PCR approaches involve two sequential amplification reactions with different pairs 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 for 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 cytometry-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 compositions (23, 24), and cellular activities, such as glucose (25), phosphorus (26), nitrogen (27, 28) and CO₂ uptake rates (5, 25, 27, 29–31). However, only a 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* hybridization only) (29–31).

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. *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 encompasses most of the genetic diversity identified within marine *Prochlorococcus* and *Synechococcus* populations (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 (with degeneracy of up to 32 for both forward and reverse primers) in order to expand the diversity coverage of marrine picocyanobacterial groups. The annealing regions of these primer sets are very similar and the degeneracy too high, precluding their use in nested or semi-nested PCR 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

			Amplicon	Primer			Gene				
Method	PCR type	Round	length (bp)	name	Direction	Sequence	positions	GC%	Deg	T (°C)	Reference
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 1 Primers sequence and PCR conditions used for amplification^a

^aNested PCR approach for *petB* gene metabarcoding of marine *Synechococcus* populations. Based on *Synechococcus* sp. strain WH 8109 (CP006882), the *petB* gene is located between positions 440919 and 441575 of the genome. The total length of the *petB* gene is 657 bp. 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.

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 *petB* includes a nested PCR amplification method (termed 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 those of the standard PCR protocol with petB-F and petB-R (termed Mazard_2012) on *Synechococcus* populations recovered from filtered seawater samples and populations 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, so sequencing could be performed with Illumina 2×300 bp chemistry.

While the first two benchmarks were fulfilled with the new primer set, petB-50F and petB-634R, the final amplicon length criterion could not be achieved. In the 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 with a length of 571 bp (from position 50 to 618 of the *petB* gene of the *Synechococcus* sp. strain WH8109 reference; accession number CP006882) while the standard Mazard_2012 PCR generates a fragment of 597 bp (Table 1). There was no overlap between the forward and reverse reads obtained with both standard and nested PCRs. To solve this problem, during sequence processing, we merged our reads by adding 10 degenerate base 'N' (for any one base) at their 3' end. Amplicons without overlap have been successfully processed with the DADA2 pipeline using this bioinformatic maneuver, which does not significantly influence k-mer-based taxonomic classification such as the na-ive 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 (SA1, SA2, and SA3) waters along the Chatham Rise (see Fig. S1 and Table S1 in the supplemental material). 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 by the standard Mazard_2012 method. Hence, from this point, the comparison between Mazard_2012 and Ong_2022 approaches included only the 70 samples successfully amplified by both methods (Table S1). The results from sorted samples were obtained only with the Ong_2022 PCR since none of the sorted samples were amplified with success by standard PCR.

Following sequence processing, the total numbers of reads retained in the final data

		Value (mean ± SD) for:				
Parameter	DADA2 step	Mazard_2012	Ong_2022			
No. of reads/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. Merge	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			

TABLE 2 Number of reads and ASVs per sample after each sequence processing step using

 DADA2 on filtered samples obtained by the Mazard 2012 and Ong 2022 approaches

set were 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% obtained with Mazard_2012 after forward and reverse reads were merged (Table 2). The resulting mean number of reads obtained with the standard PCR was almost three 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 (6,307) than that generated with the Mazard_2012 standard reaction (4,716) (Table 2).

Community composition from filtered samples. Four major *Synechococcus* subclades that were present in almost all samples were Ia, Ib, IVa, and IVb (Fig. 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 (Fig. 2A). Yet, patterns of community composition, such as the higher relative abundance of clade IVa in ST than in 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, and V), the proportions of reads of clades CRD1, EnvA, EnvB, and V were higher in Ong_2022 than in Mazard_2012 (Fig. 2B; Table S2). Six clades and subclades with low contributions to the total community were amplified only by either the Ong_2022 (WPC1, IIh, Ile, and VIb) or the Mazard_2012 (IIb and VIc) PCR method (Fig. 2B).

A paired sample Wilcoxon test indicated that the median percentage of reads for each clade and subclade in Mazard 2012 was significantly different from that in Ong 2022, except for clade V, which was present at a very low proportion (Table S2). The mean percentages of reads for subclades la and lb with Ong_2022 were higher than those with Mazard_2012 (8.51% and 46.73% versus 3.47% and 31.04%, respectively). The mean percentages of reads for subclades IVa and IVb in Ong 2022 were lower than those in Mazard 2012 (11.96% and 31.40% versus 20.39% and 44.04%, respectively) (Table S2). Among clades and subclades with minor contributions of reads to the total community, the mean percentages of reads of clade UC-A and subclades Ic and II-WPC2 were higher in Mazard 2012 than in Ong 2022 (Fig. 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%) than in Ong_2022 (0.13%). Subclades II-WPC2 and Ic were present in higher proportions and detected in more samples with Mazard_2012 than with 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, with an average percentage of reads of 0.01% and found 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, with an average of 0.03% and found in 20 samples. The average percentages of reads of clades CRD1, EnvA, EnvB, and V were higher in Ong_2022 (0.93%, 0.18%, 0.11%, and 0.0006%, respectively) than in Mazard_2012 (0.02%, 0.001%, 0.03%, and 0.0003%, respectively) (Fig. 2B; Table S2). Clades CRD1, EnvA, and V were present in more samples in Ong_2022 (35,



FIG 1 Synechococcus taxonomic composition at clade and subclade level from filtered samples obtained with the 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.

21, and 3 samples, respectively) than in 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 proportions (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 proportions (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 in Ong_2022 than in Mazard_2012 (Fig. 2C). Subclades with a higher mean percentage 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 percentage of reads and a higher number of ASVs in Mazard_2012 (Table S2).



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 obtained by the Mazard_2012 and Ong_2022 methods. Clades and subclades are arranged in descending order according to the mean percentage of reads with the Ong_2022 method.

The relative abundance of the major subclades followed a similar pattern across amplification methods and cycles, with notable exceptions for subclades IVa and IVb in ST1 and SA2 cycles, respectively. Subclade IVa was dominant in ST1 samples with the Mazard_2012 approach, while the same samples processed with Ong_2022 showed similar contributions of subclades IVa and Ib to the total community (Fig. 1; Table S3). In cycles ST2 and SA1, the proportions of subclade IVa decreased and subclade IVb increased in both methods. In SA2



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

and SA3, subclade IVb became dominant only in Mazard_2012 samples (Fig. 1; Table S3). Clades CRD1, EnvA, and EnvB were found in higher proportions in subantarctic oceanic cycles SA2 and SA3 than in more coastal SA1 and oceanic subtropical cycles for both amplifications (Fig. 1; Table S3). The mean percentages of reads for each subclade in cycles ST1, ST2, and SA1 were less than 0.004% and 0.06% for Mazard_2012 and Ong_2022, respectively. The mean percentages of reads 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 sequence similarity to reference database. Within each clade and subclade, we compared the nucleotide similarities of the ASVs obtained with Mazard_2012 and Ong_2022 and with the reference sequences in the database (32) (Fig. 3; Table S4). In general, similarity between the ASVs within each clade and subclade obtained with both PCR approaches was lower than the reference database, showing that both primer sets were able to capture a higher diversity than the one available in the database (Fig. 3). The only exceptions were clades CRD1, EnvA, and EnvB and subclade VIc (Fig. 3; Table S4). Amid the clades and subclades amplified by both PCR approaches, percentages of similarity within the ASVs obtained with Ong_2022 (from 85.9% to 94.7%) were lower than those obtained 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) (Fig. 3; 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 subclade, we compared the translated ASVs from each PCR approach (Mazard_2012 and Ong_2022) against those of 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 than did Mazard_2012 (98.2% versus 84.2%) (Table S4). The percentages of correctly assigned ASVs in each clade and subclade were 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,

10.1128/spectrum.04086-22



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

and 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 the 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 that the current database did not have a corresponding reference sequence.

Community composition from flow cytometry-sorted Synechococcus popula-tions. 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 to subcluster (SC) 5.1 than the ASVs obtained from filtered samples (Fig. 4; Fig. 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 (Fig. 4; Table S5). Clades UC-A, CRD1, and EnvB and subclade II-WPC2 had low contributions to the total community (Fig. 4; Table S5). The mean percentages 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%) (Tables S2 and S5). In contrast, for subclades IVa and IVb, the mean percentages of reads were higher (25.47% and 39.35%, respectively) in sorted cells than in filtered samples (11.96% and 31.40%) (Table S2 and S5).

Among minor clades and subclades, clade UC-A was found in 11 of 18 sorted samples with an average proportion of reads of 0.01% in contrast with filtered samples, for which 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 one-half of the samples. Clades CRD1 and EnvB and subclade II-WPC2 had lower average percentages of reads (0.003%, 0.002%, and 0.002%) than those in Ong_2022 (0.93%, 0.11%, and 0.01%). The 7 minor clades and subclades (EnvA, WPC1, V, IIh, Ile, 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 than that from 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 in sorted *Synechococcus* and filtered seawater samples, with the exceptions of subclades IVa and IVb (Fig. 1 and 4). Subclade IVa and subclade IVb were dominant in ST and SA sorted *Synechococcus* populations, respectively. This pattern was also observed with the Mazard_2012 approach on filtered samples but differed with the Ong_2022 approach, where the overall contribution of subclades IVa and IVb was similar to that of subclades Ia and Ib (Fig. 1).

DISCUSSION

We have established a nested PCR protocol to assess the diversity of marine *Synechococcus* populations 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 those of the original standard PCR (Mazard_2012) on DNA extracted from filtered seawater samples and determined the performance of the proposed nested amplification protocol with flow cytometry-sorted *Synechococcus* populations. While both standard and nested PCRs worked with nearly all filtered seawater samples, sorted *Synechococcus* samples could not be amplified with the Mazard_2012 standard PCR, likely due to a low initial concentration of *Synechococcus* DNA. Nested PCR is often employed when the DNA concentration is too low to be amplified by a standard PCR (37, 38), and thus the 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 the two methods using filtered samples have been assigned to *Prochlorococcus*, although *Prochlorococcus* cells were detected by flow cytometry (from 0.3×10^3 to 90.6×10^3 cells mL⁻¹) in subantarctic water samples (SA1, SA2, and SA3) (data not shown). Similar to our new primers, petB-F and petB-R of Mazard et al. (12) are also biased toward *Synechococcus* (12). However, during their study, *Prochlorococcus* cell abundances were recovered at low frequency from samples where *Prochorococcus* cell abundances were at least 45 times higher than those of *Synechococcus*. In our study, *Synechococcus* cells were always at the same 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 SC 5.3 were also not detected. The primers of 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 a majority of the sequences (32). In addition, *Synechococcus* cells from SC 5.2 and SC 5.3 were probably absent in our samples. SC 5.2 representatives are mainly found in estuaries and river-influenced coastal waters (39–42), and although SC 5.3 contains 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, the Mediterranean Sea, and the 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 that detected by the standard amplification. However, Ong_2022 had an overall higher proportion of correctly assigned ASVs than Mazard_2012, indicating that the nested PCR was more specific for *Synechococcus*. While the use of a higher number of cycles might increase the chimera generation and sensitivity (e.g., Ong_2022 generally increased the proportion of reads of low-abundance 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 than Mazard_2012 PCR. Within common subclades amplified by both Mazard_2012 and Ong_2022 PCRs, almost all subclades had a

higher number of ASVs in the latter approach. These ASVs had lower nucleotide similarity within each subclade, 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 than a standard PCR (37, 38).

The number and proportion of reads of minor clades and subclades decreased in sorted samples (11 versus 4 clades and subclades, 1% versus 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 that in the sorted samples. One major difference between the sorting protocols of Metz et al. (19) and Ong_2022 is the number of cells sorted. In the Metz et al. (19) protocol, the number of cells sorted was 22 to 50 times higher than that in the Ong_2022 protocol. 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 the Ong_2022 approach for 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 data sets obtained from filtered samples (\sim 99%) and sorted samples (99.8%) with nested and standard PCRs. 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, 16, 32, 45, 49). However, the relative abundance of clades I and IV varied between filtered samples amplified by Mazard_2012 and Ong_2022 PCRs and 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 clades I to IV. Both clades were detected in similar proportions along the coast of New Zealand at overlapping latitudes (45°S) by a PCR-independent method (42). Clade IV has been observed in a higher proportion than 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 more closely reflects the natural abundance of the dominant clades.

We observed that relative abundances 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 by both approaches) (50, 51). Clade CRD1 was reported to cooccur with clades I and IV in the southeastern Pacific Ocean and other cold and high-latitude HNLC conditions (16, 32, 45). More recent reports suggested that EnvA and EnvB also correspond to lineages found under low-iron conditions (32, 45, 52, 53). The Ong_2022 approach recorded subclades IIh and Ile in cycle ST1 at a very low abundance. Although members from clade II are mostly warm thermal types that dominate in nutrient-depleted tropical areas (3), subclades Ile and IIh were also reported at low relative abundances in cooler waters (14.1 to 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 Ile and IIh were detected. Subclades WPC1, VIb, VIc, and V were detected with low relative abundances, 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* populations. Results obtained

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

MATERIALS AND METHODS

Sampling location. Samples were collected during the Salp Particle expOrt and Oceanic Production (SalpPOOP) cruise (TAN1810, 21 October to 21 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 (54–57). ST waters are warm, saline, and macronutrient depleted, while SA waters are cool, less-saline, and high-nitrate, low-chlorophyll, and low-silicate (HNLC-LSi) waters where low iron and silicate are the primary limiting factors for phytoplankton growth and productivity (50, 58).

We adopted a 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 over 3 to 6 days (here 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) (see Fig. S1 in the supplemental material). In each cycle, water column profiles of temperature, salinity, dissolved oxygen, photosynthetically active radiation (PAR), and fluorescence were measured 4 to 6 times daily by a Seabird (SBE 911plus) conductivity-temperature-depth (CTD) instrument with a PAR sensor. The cycles were identified as subtropical or subantarctic based on physical measurements during sampling in accordance with the definitions of Lüskow et al. (57). 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 to 70 m). For DNA analysis of filtered seawater samples, 1.5 to 2 L of seawater from all depths was filtered through a $0.22-\mu$ m-pore-size Sterivex filter using a peristaltic pump (number of 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 or 70 m) (Table S1), where they were preserved with a solution of dimethyl sulfoxide (DMSO) and 1/100 Pluronic acid mix at a 10% final concentration and flash-frozen in liquid nitrogen (number of samples = 21). All samples were stored at -80°C until laboratory processing.

Flow cytometry cell sorting. *Synechococcus* cells were sorted using a FACSAria flow cytometer (Becton, Dickinson, San Jose, CA), equipped with a laser emitting at 488 nm, a 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 SSC and discriminated from photosynthetic picoeukaryotes based on the orange fluorescence of phycoerythrin. Filter-sterilized (0.22- μ m pore size) Tris-HCI (50 mM, pH 8.0)–NaCI (10 mM) buffer was used as the sheath liquid. Sheath pressure was set at 70 lb/in², 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 analyzed 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 as follows: (i) 3 or fewer mismatches in the binding region, (ii) no nucleotide ambiguity or mismatches at the 5' and 3' ends, (iii) low degeneracy, (iv) binding region between 15 and 20 bp, (v) annealing temperature difference between the forward and reverse primers not greater than 4°C, (vi) insertions and deletions among the different groups should be absent from the primer binding region, and (vii) no hairpins and self-dimerization within the primer sequence. Four hundred seven Synechococcus nucleotide sequences of the petB gene from the 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 binding region conforming to the criteria described. The selected forward primer petB-50F is located at positions 50 to 66 with reference to the first nucleotide of the petB gene from Synechococcus sp. WH8109 (accession number CP006882) (Table 1). The selected reverse primer, petB-634R, is located at positions 615 to 634 with reference to the first nucleotide of the petB gene from Synechococcus sp. WH8109 (accession number CP006882) (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 SC 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 SC 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 of 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 the manufacturer's instructions. The modified protocol can be found at https://github.com/deniseong/marine-Synechococcus-metaB. DNA from flow cytometrysorted Synechococcus cells was obtained using three cycles of flash-freezing and thawing in liquid nitrogen (18). The 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. PCRs were done using the following reaction mixture: extracted DNA template with an average final concentration of 1.3 ng μ L⁻¹ (from 0.007 to 4.11 ng μ L⁻¹), 12.5 μ L KAPA HiFi hotstart readymix 2×, 0.3 μ M primers (petB-F and petB-R) (final concentration), 0.1 to 0.4 μ g μ L⁻¹ final concentration of bovine serum albumin (BSA), and H_2O to a 25- μ L final reaction volume. Thermal conditions were as follows: 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s, and a final cycle of 72°C for 6 min. All PCRs were performed in duplicate, and results were pooled. The 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 μ L⁻¹ (from 0.015 to 2.055 ng μ L⁻¹) or a volume corresponding to approximately 160 to 400 sorted Synechococcus cells (4 μ L maximum), 5 μ L KAPA HiFi hotstart readymix 2×, 0.3 μ M final concentration of primer petB-F, 0.3 µM final concentration of primer petB-634R (5'-GCTTVCGRAT CATCARGAAG-3'), 0.1 μ g μ L⁻¹ final concentration of BSA (only for filtered samples), and H₂O for a 10- μ L reaction volume. Thermal conditions were as follows: 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 59°C for 30 s, and 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 μ L of first-round product, 12.5 μ L KAPA HiFi hotstart readymix 2×, 0.3 µM final concentration of primer petB-50F (5'- TYCAGGACATYGCTGAY-3'), 0.3 µM final concentration of primer petB-R, and H₂O for a 25-µL reaction volume. Thermal conditions were as follows: 94°C for 5 min, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s, and a final cycle of 72°C for 6 min. All firstand second-round PCRs were performed in duplicate, and results were pooled. Samples were purified, barcoded, and sequenced by the GeT-PlaGe platform of GenoToul (INRAE, Auzeville, France) using an Illumina MiSeq platform (2 \times 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 the function 'filterandtrim' with the DADA2 R package version 1.12 (65). Based on sequence quality score, reads obtained with the Mazard_2012 and Ong_2022 methods 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 and maxEE = c(10, 10) in both data sets. 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' (for any one base) 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 the petB reference database of 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, 14 clades (I to IX, XVI, XX, UC-A, WPC1, CRD1, EnvA, and EnvB) within SC 5.1, and 20 subclades (la to Ic, Ila to Ilh, and Il-WPC2, Illa to Illc, IVa and IVb, and VIa to VIc). ASVs assigned as Richelia sp. were removed prior to normalization of reads. At station SA1_5, sorted samples from two depths (12 m and 40 m) were sorted to two and three different total numbers of cells (Table S1). As the communities were 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 of scripts along with the reformatted database can be found at https://github.com/ deniseong/marine-Synechococcus-metaB.

Data availability. Raw sequencing data were deposited in the NCBI Sequence Read Archive under BioProject number PRJNA885274. Source code, DNA extraction protocol, and all metadata generated are available on GitHub (https://github.com/deniseong/marine-Synechococcus-metaB). Raw data files from flow cyyometry sorting are available at http://flowrepository.org/id/FR-FCM-Z5P8 (repository ID, FR-FCM-Z5P8). Step-by-step protocols for DNA extraction and PCR amplification are published on protocols.io at https://doi.org/10.17504/protocols.io.5qpvorkj7v4o/v1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only. SUPPLEMENTAL FILE 1, PDF file, 0.8 MB. SUPPLEMENTAL FILE 2, XLSX file, 0.02 MB. SUPPLEMENTAL FILE 3, XLSX file, 0.01 MB. SUPPLEMENTAL FILE 4, XLSX file, 0.01 MB. SUPPLEMENTAL FILE 5, XLSX file, 0.01 MB. SUPPLEMENTAL FILE 6, XLSX file, 0.01 MB.

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* cells 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.

A.L.D.S. and D.R.Y.O. were supported by the Singapore Ministry of Education, Academic Research Fund Tier 1 (RG26/19). L.G. was supported by the French "Agence Nationale de la Recherche" Program CINNAMON (ANR-17-CE02-0014-01). A.G.-R. 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.

D.R.Y.O. and A.L.D.S. conceived the study. A.G.-R. organized the SalpPOOP voyage and sampling structure. A.G.-R. and A.L.D.S. developed the field protocol and processed the samples during the voyage. A.L.D.S. and D.M. performed flow cytometry cell sorting. D.R.Y.O. developed primers and processed the samples for sequencing. D.R.Y.O. analyzed data and drafted the manuscript. All authors were involved in data interpretation and critical revision of the manuscript.

REFERENCES

- Dvořák P, Casamatta DA, Poulíčková A, Hašler P, Ondřej V, Sanges R. 2014. Synechococcus: 3 billion years of global dominance. Mol Ecol 23:5538–5551. https://doi.org/10.1111/mec.12948.
- 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 U S A 110:9824–9829. https://doi.org/ 10.1073/pnas.1307701110.
- 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 73:249–299. https:// doi.org/10.1128/MMBR.00035-08.
- Flombaum P, Martiny AC. 2021. Diverse but uncertain responses of picophytoplankton lineages to future climate change. Limnol Oceanogr 66: 4171–4181. https://doi.org/10.1002/lno.11951.
- Li WKW. 1994. Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. Limnol Oceanogr 39:169–175. https://doi.org/10.4319/lo.1994.39.1.0169.
- 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, Tara Oceans Coordinators. 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature 532:465–470. https://doi.org/10.1038/nature16942.
- Urbach E, Robertson DL, Chisholm SW. 1992. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. Nature 355:267–270. https://doi.org/10.1038/355267a0.
- Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. 2003. Cladespecific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. Appl Environ Microbiol 69:2430–2443. https://doi.org/10 .1128/AEM.69.5.2430-2443.2003.
- ZwirgImaier 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 10:147–161. https://doi.org/10 .1111/j.1462-2920.2007.01440.x.

- 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:R90. https://doi .org/10.1186/gb-2008-9-5-r90.
- 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 8:1200–1211. https://doi.org/10.1111/j .1462-2920.2006.01010.x.
- Mazard S, Ostrowski M, Partensky F, Scanlan DJ. 2012. Multilocus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. Environ Microbiol 14:372–386. https://doi.org/10.1111/j.1462-2920.2011.02514.x.
- Rocap G, Distel DL, Waterbury JB, Chisholm SW. 2002. Resolution of *Pro-chlorococcus* and *Synechococcus* ecotypes by using 16S-23S Ribosomal DNA internal transcribed spacer sequences. Appl Environ Microbiol 68: 1180–1191. https://doi.org/10.1128/AEM.68.3.1180-1191.2002.
- Palenik B. 1994. Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. Appl Environ Microbiol 60:3212–3219. https:// doi.org/10.1128/aem.60.9.3212-3219.1994.
- 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 45:263–275. https://doi.org/10.3354/ame045263.
- 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 77:333–342. https://doi .org/10.1007/s00248-018-1308-8.
- Erlich HA, Gelfand D, Sninsky JJ. 1991. Recent advances in the polymerase chain reaction. Science 252:1643–1651. https://doi.org/10.1126/science .2047872.
- 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. ISME J 12:1360–1374. https://doi.org/10.1038/ s41396-018-0050-z.
- 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:fiz038. https://doi .org/10.1093/femsec/fiz038.

- 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 67:612–622. https://doi.org/10.1111/jeu.12813.
- 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. ISME J 7:184–198. https://doi.org/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 R Soc Lond B Biol Sci 374:20190096. https://doi.org/10.1098/rstb .2019.0096.
- 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 2:143. https://doi.org/10.1038/s43247-021-00212-9.
- Hansman RL, Sessions AL. 2016. Measuring the in situ carbon isotopic composition of distinct marine plankton populations sorted by flow cytometry. Limnol Oceanogr Methods 14:87–99. https://doi.org/10.1002/ lom3.10073.
- 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 20:2743–2756. https://doi.org/10.1111/1462-2920.14111.
- 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 subtrop- ical North Atlantic. Aquat Microb Ecol 58:31–44. https://doi.org/10.3354/ame01348.
- Berthelot H, Duhamel S, L'Helguen S, Maguer J-F, Wang S, Cetinić I, Cassar N. 2019. NanoSIMS single cell analyses reveal the contrasting nitrogen sources for small phytoplankton. ISME J 13:651–662. https://doi.org/10 .1038/s41396-018-0285-8.
- Aldunate M, Henríquez-Castillo C, Ji Q, Lueders-Dumont J, Mulholland MR, Ward BB, Dassow P, Ulloa O. 2020. Nitrogen assimilation in picocyanobacteria inhabiting the oxygen-deficient waters of the eastern tropical North and South Pacific. Limnol Oceanogr 65:437–453. https://doi.org/10 .1002/lno.11315.
- 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 61:806–824. https://doi.org/10.1002/lno.10255.
- Jardillier L, Zubkov MV, Pearman J, Scanlan DJ. 2010. Significant CO₂ fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. ISME J 4:1180–1192. https://doi.org/10.1038/ismej.2010.36.
- Hartmann M, Gomez-Pereira P, Grob C, Ostrowski M, Scanlan DJ, Zubkov MV. 2014. Efficient CO₂ fixation by surface *Prochlorococcus* in the Atlantic Ocean. ISME J 8:2280–2289. https://doi.org/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 U S A 113: E3365–E3374. https://doi.org/10.1073/pnas.1524865113.
- Ohnemus DC, Rauschenberg S, Krause JW, Brzezinski MA, Collier JL, Geraci-Yee S, Baines SB, Twining BS. 2016. Silicon content of individual cells of *Syne-chococcus* from the North Atlantic Ocean. Mar Chem 187:16–24. https://doi .org/10.1016/j.marchem.2016.10.003.
- Needham DM, Fuhrman JA. 2016. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat Microbiol 1: 16005. https://doi.org/10.1038/nmicrobiol.2016.5.
- 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:e114804. https://doi.org/10.1371/ journal.pone.0114804.
- Dacey DP, Chain FJJ. 2021. Concatenation of paired-end reads improves taxonomic classification of amplicons for profiling microbial communities. BMC Bioinform 22:493. https://doi.org/10.1186/s12859-021-04410-2.
- 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:e0132253. https://doi.org/10.1371/journal.pone.0132253.
- 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 12:640829. https://doi.org/10.3389/fpls.2021.640829.

- Jing H, Liu H, Suzuki K. 2009. Phylogenetic diversity of marine Synechococcus spp. in the Sea of Okhotsk. Aquat Microb Ecol 56:55–63. https://doi .org/10.3354/ame01316.
- 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: 7644–7655. https://doi.org/10.1128/AEM.01895-15.
- Xia X, Guo W, Tan S, Liu H. 2017. Synechococcus assemblages across the salinity gradient in a salt wedge estuary. Front Microbiol 8:1254. https:// doi.org/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. https://doi.org/10.1101/2022.03.07.483242.
- 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 20:3757–3771. https://doi.org/10.1111/1462-2920.14377.
- Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. 2012. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. ISME J 6:285–297. https://doi.org/10.1038/ismej.2011.106.
- 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. ISME J 10:333–345. https://doi.org/10.1038/ismej.2015.115.
- Green MR, Sambrook J. 2019. Nested polymerase chain reaction (PCR). Cold Spring Harb Protoc 2019:095182. https://doi.org/10.1101/pdb.prot095182.
- 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 U S A 103:12115–12120. https:// doi.org/10.1073/pnas.0605127103.
- Cermeño P, Teixeira IG, Branco M, Figueiras FG, Marañón E. 2014. Sampling the limits of species richness in marine phytoplankton communities. J Plankton Res 36:1135–1139. https://doi.org/10.1093/plankt/fbu033.
- Kent AG, Baer SE, Mouginot C, Huang JS, Larkin AA, Lomas MW, Martiny AC. 2019. Parallel phylogeography of *Prochlorococcus* and *Synechococcus*. ISME J 13:430–441. https://doi.org/10.1038/s41396-018-0287-6.
- 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 104:13395–13408. https://doi.org/10.1029/1999JC900009.
- 51. Longhurst AR. 2007. Ecological geography of the sea, 2nd ed. Academic Press. Burlington, MA, USA.
- 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, et al. 2019. Community-level responses to iron availability in open ocean plankton ecosystems. Glob Biogeochem Cycles 33: 391–419. https://doi.org/10.1029/2018GB006022.
- Ahlgren NA, Belisle BS, Lee MD. 2020. Genomic mosaicism underlies the adaptation of marine *Synechococcus* ecotypes to distinct oceanic iron niches. Environ Microbiol 22:1801–1815. https://doi.org/10.1111/ 1462-2920.14893.
- Heath RA. 1983. Observations on Chatham Rise currents. New Zealand J Mar Freshw Res 17:321–330. https://doi.org/10.1080/00288330.1983.9516006.
- Chiswell SM. 1994. Acoustic Doppler current profiler measurements over the Chatham Rise. New Zealand J Mar Freshw Res 28:167–178. https://doi .org/10.1080/00288330.1994.9516605.
- 56. Stramma L, Peterson RG, Tomczak M. 1995. The South Pacific current. J Phys Oceanogr 25:77–91. https://doi.org/10.1175/1520-0485(1995)025 %3C0077:TSPC%3E2.0.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 167:175. https://doi.org/10.1007/s00227 -020-03775-x.
- 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 20:95–105. https://doi.org/10 .3354/ame020095.

- 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 83:208–216. https://doi.org/10.1016/j.pocean.2009.07.026.
- Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
- 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 28:1647–1649. https://doi.org/10.1093/bioinformatics/ bts199.
- Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26: 1641–1650. https://doi.org/10.1093/molbev/msp077.
- 63. RStudio Team. 2021. RStudio: integrated development environment for R. RStudio, PBC, Boston, MA.
- 64. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.
- McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8: e61217. https://doi.org/10.1371/journal.pone.0061217.