Design and use of a new primer pair for the characterization of the cyanobacteria *Synechococcus* and *Prochlorococcus* communities targeting petB gene through metabarcoding approaches

Alexandra Coello-Camba*, Rubén Díaz-Rúa[1], Susana Agusti

Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

ARTICLE INFO

ABSTRACT

During the last years, the application of next-generation sequencing (NGS) technologies to search for specific genetic markers has become a crucial method for the characterization of microbial communities. Illumina MiSeq, likely the most widespread NGS platform for metabarcoding experiments and taxonomic classification, allows processing shorter reads than the classical SANGER sequencing method and therefore requires specific primer pairs that produce shorter amplicons. Specifically, for the analysis of the commonly studied *Prochlorococcus* and *Synechococcus* communities, the *petB* marker gene has recently stood out as able to provide deep coverage to determine the microdiversity of the community. However, current *petB* primer set produce a 597 bp amplicon that is not suitable for MiSeq chemistry.

Here, we designed and tested a *petB* primer pair that targets both *Prochlorococcus* and *Synechococcus* communities producing an appropriate amplicon to be used with state-of-the-art Illumina MiSeq. This new primer set allows the classification of both groups to a low taxonomic level and is therefore suitable for high throughput experiments using MiSeq technologies, therefore constituting a useful, novel tool to facilitate further studies on *Prochlorococcus* and *Synechococcus* communities.

- This work describes the *de novo* design of a *Prochlorococcus* and *Synechococcus*-specific *petB* primer pair, allowing the characterization of both populations to a low taxonomic level.
- This primer pair is suitable for widespread Illumina MiSeq sequencing technologies.
- *petB* was confirmed as an adequate target for the characterization of both picocyanobacteria.

Specifications table

Subject area:	Biochemistry, Genetics and Molecular Biology
More specific subject area:	*Picocyanobacteria,* petB, *MiSeq, primer design.*
Name of your method:	*New primer pair for Prochlorococcus and* Synechococcus *taxonomic classification with Illumina MiSeq*
Name and reference of original method:	*NA*
Resource availability:	*NA*

Method details

Background

Synechococcus and *Prochlorococcus* are small (< 2 μm) cyanobacteria very abundant in the oceans. These unicellular cyanobacteria are closely related to each other [1] and recent studies have described a high genotypic diversity for both groups [2,3]. In *Prochlorococcus,* up to twelve clades have been defined, distributed in two deeply branching clusters, high-light (HL) and low-light (LL) [2,4]. *Synechococcus* populations are genetically more diverse than those of *Prochlorococcus;* three different subclusters (5.1, 5.2, and 5.3), and at least 16 clades (I–XVI) have been distinguished so far [2,3,5].

During the last decades, massively parallel DNA sequencing has become routine in many microbiology laboratories, being cheaper and less time-consuming than culturing and classical sequencing methods. Amongst these new techniques, the use of Illumina MiSeq provides good coverage of short sequences and predominates over other next-generation methods [6].

These modern genomic methodologies require accurate DNA amplification of samples, during which the choice of adequate primer pairs is a critical step [7]. Typically, the sequence analysis of genes encoding small-subunit ribosomal RNA (16S rRNA), the 16S-23S internal transcribed spacer region (ITS), and other different single copy genes such as *ntcA* have been used for the identification and phylogenetic classification of cyanobacteria reaching a clade level [i. e. 4,8]. In 2012, Mazard *et al.* used a set of core loci to explore the genetic diversity within *Synechococcus*; amongst them, *petB* unveiled a great genetic diversity in this genus and has been used to classify *Synechococcus* communities to subclade level [9]. However, the length of the amplicon produced by the primer set defined by Mazard et al. (597 bp) is not suitable for Illumina MiSeq sequencing, as it requires a shorter fragment length [6,10]. Given the ubiquity of both *Prochlorococcus* and *Synechococcus*, and that these picocyanobacterial broadly share their oceanic niches [e.g. 2], there is a need for a suitable primer pair that not only allows detailed analysis of the microdiversity of their communities but that is also suitable for the most commonly available sequencing technologies.

Objective

To seize the full potential of *petB* as a target gene in *Synechococcus* and *Prochlorococcus* taxonomic studies, we aimed to design a new *petB* primer set producing amplicons suitable for Illumina MiSeq sequencing chemistry and able to identify both groups to the lowest taxonomic level possible. To do so, we compiled a complete database with available data on both *Prochlorococcus* and *Synechococcus petB* sequences from different sources and searched for *de novo* candidate primers with adequate amplicon sizes using computational tools. We tested them *in silico,* and then validated our results with picocyanobacteria cultures and natural samples collected from Red Sea waters.

PetB reference sequence database

The *petB* sequences were downloaded from the European Nucleotide Archive (www.ebi.ac.uk/ena). The database was composed of 547 references (128 *Prochlorococcus* and 419 *Synechococcus*) classified to clade or subclade level, when available (Table 1, Fig. 1). To perform the *de novo* primer design, we also added three different cyanobacteria species as outgroups: *Mastigocladus laminosus* and *Planktothrix agardhii*, and the unicellular eukaryote *Chattonella marina*. Once the primers were designed, these outgroup sequences were removed from the database.

In silico primer design and selection

A set of *de novo* candidate primers targeting the *petB* region was defined and analyzed using the PrimerProspector package (http://pprospector.sourceforge.net) [11] (Fig. 1). The primers were designed based on the aligned reference database with a 95% sensitivity threshold, selecting those candidates yielding amplicons with an estimated size within the 300–500 bp range. The candidate primers were then optimized for PCR reactions, and the overall database matches and taxonomic coverages were estimated, following the Primer Analysis pipeline.

Test of candidate primer pairs using reference cultures

Based on *in silico* estimations of amplicon lengths, primer weighted scores, and taxonomic coverages, the most appropriate primer pairs were selected and synthesized commercially by Sigma-Aldrich®.

The selected *de novo* primer pairs were tested together with the *petB* primer pair presented in Mazard *et al.* (2012), as a reference. To perform these primer tests, we used *Prochlorococcus* and *Synechococcus* cultures purchased from the Roscoff Culture Collection (RCC, Roscoff, France), selecting available strains with known taxonomic classification, preferably isolated from the Red Sea or adjacent waters (Fig. 1).

For DNA extraction from cell cultures, 50 mL of each culture were filtered through a polycarbonate filter, and DNA was extracted using the DNeasy® PowerWater® DNA Extraction kit (MoBio Laboratories, Inc., Carlsbad, CA) (Fig. 1). The DNA obtained was quantified with a Qubit® fluorimeter (Life Technologies, Carlsbad, CA).

PCR amplification was performed using an Eppendorf Mastercycler® Pro (Fig. 1). Four ng of DNA from each culture, Qiagen multiplex PCR master mix (QIAGEN, Valencia, CA), and a final primer concentration of 0.3 μM were used, to a final volume of 30 μL.

Table 1

petB reference database including 128 *Prochlorococcus* and 419 *Synechococcus* sequences, as compiled from Mazard *et al.* (2012) and Farrant *et al.* (2016). (Abbreviatures: Pro.: *Prochlorococcus*; Syn.: *Synechococcus*; Misc.: miscellaneous origin).

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
EnvTARA160	Pro.	HLI		N. Atlantic	43.7		5	KU377862
EnvTARA159	Pro.	HLI		N. Atlantic	43.7		5	KU377863
EnvTARA158	Pro.	HLI		N. Atlantic	43.7		5	KU377864
EnvTARA157	Pro.	HLI		N. Atlantic	36.2	-29.0	5	KU377865
EnvTARA156	Pro.	HLI		N. Atlantic	35.8	-37.2	40	KU377866
EnvTARA155	Pro.	HLI		N. Atlantic	35.9	-37.3	5	KU377867
EnvTARA154	Pro.	HLI		N. Pacific	35.4	-127.7	5	KU377868
EnvTARA153	Pro.	HLI		N. Pacific	35.4	-127.7	5	KU377869
EnvTARA152	Pro.	HLI		N. Pacific	35.4	-127.7	5	KU377870
EnvTARA151	Pro.	HLI		N. Pacific	35.4	-127.7	5	KU377871
EnvTARA150	Pro.	HLI		N. Pacific	35.4	-127.7	5	KU377872
EnvTARA149	Pro.	HLI		N. Pacific	35.4	-127.7	5	KU377873
EnvTARA142	Pro.	HLI		S. Pacific	-29.7	-101.2	5	KU377879
EnvTARA141	Pro.	HLI		S. Pacific	-29.7	-101.2	5	KU377880
EnvTARA140	Pro.	HLI		S. Pacific	-29.7	-101.2	5	KU377881
EnvTARA139	Pro.	HLI		S. Pacific	-29.7	-101.2	5	KU377882
EnvTARA138	Pro.	HLI		S. Pacific	-29.7	-101.2	5	KU377883
EnvTARA136	Pro.	HLI		S. Pacific	-32.8	-87.1	5	KU377884
EnvTARA135	Pro.	HLI		S. Pacific	-32.8	-87.1	5	KU377885
EnvTARA134	Pro.	HLI		S. Pacific	-32.8	-87.1	5	KU377886
EnvTARA133	Pro.	HLI		S. Atlantic	-31.0	4.7	50	KU377887
EnvTARA132	Pro.	HLI		S. Atlantic	-31.0	4.7	50	KU377888
EnvTARA131	Pro.	HLI		S. Atlantic	-31.0	4.7	5	KU377889
EnvTARA130	Pro.	HLI		N. Atlantic	36.6	-6.6	5	KU377890
EnvTARA129	Pro.	HLI		N. Atlantic	36.6	-6.6	5	KU377891
EnvTARA128	Pro.	HLI		N. Atlantic	36.6	-6.6	5	KU377892
EnvTARA127	Pro.	HLI		N. Atlantic	36.6	-6.6	5	KU377893
EnvTARA082	Pro.	HLI		S. Pacific	-32.8	-87.1	5	KU377935
EnvTARA081	Pro.	HLI		S. Pacific	-32.8	-87.1	5	KU377936
EnvTARA053	Pro.	HLI		S. Pacific	-23.2	-129.5	155	KU377956
EnvTARA027	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377965
EnvTARA026	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377966
EnvTARA025	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377967
EnvTARA023	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377968
EnvTARA022	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377969
EnvTARA021	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377970
EnvTARA020	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377971
EnvTARA019	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377972
EnvTARA018	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377973
EnvTARA017	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377974
EnvTARA016	Pro.	HLI		Mediterranean	35.8	14.3	60	KU377975
MIT9107	Pro.	HLII		Tropical Pacific	-15.0	-135.0	25	AF001490
EnvTARA147	Pro.	HLII		S. Pacific	-29.7	-101.2	5	KU377874
EnvTARA146	Pro.	HLII		S. Pacific	-29.7	-101.2	5	KU377875
EnvTARA145	Pro.	HLII		S. Pacific	-29.7	-101.2	5	KU377876
EnvTARA144	Pro.	HLII		S. Pacific	-29.7	-101.2	5	KU377877
EnvTARA143	Pro.	HLII		S. Pacific	-29.7	-101.2	5	KU377878
EnvTARA126	Pro.	HLII		Indian	-9.4	66.4	5	KU377894
EnvTARA125	Pro.	HLII		Indian	-9.4	66.4	5	KU377895
EnvTARA124	Pro.	HLII		Indian	-9.4	66.4	5	KU377896
EnvTARA123	Pro.	HLII		Indian	-9.4	66.4	5	KU377897
EnvTARA122	Pro.	HLII		Indian	-9.4	66.4	5	KU377898
EnvTARA121	Pro.	HLII		Indian	-9.4	66.4	5	KU377899
EnvTARA120	Pro.	HLII		Indian	-9.4	66.4	5	KU377900
EnvTARA119	Pro.	HLII		Indian	-9.4	66.4	5	KU377901
EnvTARA117	Pro.	HLII		S. Pacific	-23.2	-129.5	155	KU377903
EnvTARA116	Pro.	HLII		Indian	0.0	71.6	5	KU377904
EnvTARA115	Pro.	HLII		Indian	0.0	71.6	5	KU377905
EnvTARA113	Pro.	HLII		Indian	-17.3	42.3	66	KU377906
EnvTARA108	Pro.	HLII		N. Pacific	7.4	-79.3	5	KU377911
EnvTARA107	Pro.	HLII		N. Pacific	7.4	-79.3	5	KU377912
EnvTARA106	Pro.	HLII		N. Pacific	7.4	-79.3	5	KU377913
EnvTARA103	Pro.	HLII		Indian	18.6	66.5	25	KU377916
EnvTARA102	Pro.	HLII		Indian	18.6	66.5	25	KU377917
EnvTARA099	Pro.	HLII		Indian	19.0	64.5	25	KU377920
EnvTARA098	Pro.	HLII		N. Pacific	6.3	-103.0	60	KU377921
EnvTARA088	Pro.	HLII		Indian	6.0	73.9	80	KU377929

(*continued on next page*)

Table 1 (*continued*)

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
EnvTARA087	Pro.	HLII		Indian	14.6	70.0	60	KU377930
EnvTARA086	Pro.	HLII		Indian	14.6	70.0	60	KU377931
EnvTARA065	Pro.	HLII		S. Pacific	-23.2	-129.5	155	KU377945
EnvTARA063	Pro.	HLII		S. Pacific	-23.2	-129.5	155	KU377947
EnvTARA061	Pro.	HLII		S. Pacific	-23.2	-129.5	155	KU377949
EnvTARA060	Pro.	HLII		S. Pacific	-23.2	-129.5	155	KU377950
EnvTARA059	Pro.	HLII		S. Pacific	-23.2	-129.5	155	KU377951
EnvTARA084	Pro.	HLIII		Indian	-17.0	54.0	5	KU377933
EnvTARA083	Pro.	HLIII		Indian	-17.0	54.0	5	KU377934
EnvTARA080	Pro.	HLIV		Indian	-17.0	54.0	5	KU377937
EnvTARA079	Pro.	HLIV		Indian	-17.0	54.0	5	KU377938
EnvTARA078	Pro.	HLIV		Indian	-17.0	54.0	5	KU377939
EnvTARA077	Pro.	HLIV		Indian	-17.0	54.0	5	KU377940
EnvTARA073	Pro.	HLIV		Indian	-17.0	54.0	5	KU377941
EnvTARA070	Pro.	HLIV		Indian	-17.0	54.0	5	KU377942
EnvTARA069	Pro.	HLIV		Indian	-17.0	54.0	5	KU377943
EnvTARA068	Pro.	HLIV		Indian	-17.0	54.0	5	KU377944
EnvTARA032	Pro.	HLIV		Indian	-17.0	54.0	75	KU377961
NATL2A	Pro.	LLI	162					AF001489
EnvTARA118	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377902
EnvTARA112	Pro.	LLI		S. Pacific	-1.9	-84.6	50	KU377907
EnvTARA111	Pro.	LLI		S. Pacific	-1.9	-84.6	50	KU377908
EnvTARA110	Pro.	LLI		S. Pacific	-1.9	-84.6	50	KU377909
EnvTARA109	Pro.	LLI		S. Pacific	-1.9	-84.6	50	KU377910
EnvTARA105	Pro.	LLI		N. Pacific	31.5	-159.0	115	KU377914
EnvTARA104	Pro.	LLI		N. Pacific	31.5	-159.0	115	KU377915
EnvTARA101	Pro.	LLI		S. Atlantic	-8.7	-18.0	100	KU377918
EnvTARA100	Pro.	LLI		S. Atlantic	-8.7	-18.0	100	KU377919
EnvTARA097	Pro.	LLI		N. Pacific	6.3	-103.0	60	KU377922
EnvTARA096	Pro.	LLI		N. Pacific	6.3	-103.0	60	KU377923
EnvTARA093	Pro.	LLI		S. Pacific	-9.0	-139.1	115	KU377924
EnvTARA092	Pro.	LLI		S. Pacific	-9	-139.1	115	KU377925
EnvTARA091	Pro.	LLI		S. Pacific	-9	-139.1	115	KU377926
EnvTARA090	Pro.	LLI		S. Pacific	-9	-139.1	115	KU377927
EnvTARA085	Pro.	LLI		Indian	-17.0	54.0	75	KU377932
EnvTARA064	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377946
EnvTARA062	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377948
EnvTARA058	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377952
EnvTARA057	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377953
EnvTARA056	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377954
EnvTARA055	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377955
EnvTARA052	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377957
EnvTARA051	Pro.	LLI		S. Pacific	-23.2	-129.5	155	KU377958
EnvTARA034	Pro.	LLI		Indian	-17.0	54.0	75	KU377959
EnvTARA033	Pro.	LLI		Indian	-17.0	54.0	75	KU377960
EnvTARA031	Pro.	LLI		Indian	-17.0	54.0	75	KU377962
EnvTARA030	Pro.	LLI		Indian	-17.0	54.0	75	KU377963
EnvTARA029	Pro.	LLI		Indian	-17.0	54.0	75	KU377964
EnvTARA007	Pro.	LLII-III		Red Sea	23.4	37.2	80	KU377976
EnvTARA005	Pro.	LLII-III		Misc.				KU377977
AMT18025-A31	Pro.	LLIV		Atlantic	27.6	-37.3	19	JF307745
AMT18025-D44	Pro.	LLIV		Atlantic	27.6	-37.3	19	JF307754
AMT18025-D48	Pro.	LLIV		Atlantic	27.6	-37.3	19	JF307758
AMT18025-D49	Pro.	LLIV		Atlantic	27.6	-37.3	19	JF307759
AMT18070-D85	Pro.	LLIV		Atlantic	-16.6	-25.0	18	JF307769
EnvTARA089	Pro.	LLIV		Red Sea	23.4	37.2	80	KU377928
EnvTARA004	Pro.	LLIV		N. Pacific	14.2	-116.6	40	KU377978
EnvTARA003	Pro.	LLIV		N. Pacific	14.2	-116.6	40	KU377979
EnvTARA002	Pro.	LLIV		N. Pacific	14.2	-116.6	40	KU377980
EnvTARA001	Pro.	LLIV		N. Pacific	14.2	-116.6	40	KU377981
EnvTARA094	Pro.	LLIV		Red Sea	18.4	39.9	60	KU377990
EnvTARA049	Syn.	5.2		Mediterranean	39.4	19.4	5	KU377848
EnvTARA047	Syn.	5.2		Misc.			5	KU377850
EnvTARA045	Syn.	5.2		Misc.			5	KU377851
EnvTARA041	Syn.	5.2		Misc.			5	KU377852
EnvTARA039	Syn.	5.2		Misc.			5	KU377853
MINOS11	Syn.	5.2	2319	Mediterranean	34.0	18.0	20	KU377983
MITS9220	Syn.	CRD1	2571	Pacific		-140.0		JF307509
Biosope-45-B4-461	Syn.	CRD1	1016	Pacific	-9.1	-137.0	100	JF307512

Table 1 (*continued*)

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
Biosope-45-B5-463	Syn.	CRD1	1017	Pacific	-9.1	-137.0	100	JF307513
BIOS-H3-1	Syn.	CRD1	1018	Pacific	-9.1	-137.0	100	JF307514
BIOS-S15-1	Syn.	CRD1	1023	Pacific	-30.8	-95.4	100	JF307516
Biosope-48-B3Y	Syn.	CRD1	1027	Pacific	-9.1	-137.0	30	JF307518
BIOS-H3-2	Syn.	CRD1	1030	Pacific	-9.1	-137.0	30	JF307521
NOUM97010	Syn.	CRD1	66	Pacific		180.0	30	JF307546
AMT15006-B9a	Syn.	CRD1		Atlantic	47.9	-14.6	10	JF307631
AMT15006-B9b	Syn.	CRD1		Atlantic	47.9	-14.6	10	JF307632
AMT18012-A16	Syn.	CRD1		Atlantic	42.7	-22.2	12	JF307640
EnvTARA035	Syn.	CRD1		Indian	-17.0	54.0	75	KU377856
EnvTARA012	Syn.	CRD1		Indian	-17.0	54.0	5	KU377858
EnvTARA011	Syn.	CRD1		Indian	-17.0	54.0	5	KU377859
EnvTARA010	Syn.	CRD1		Indian	-17.0	54.0	5	KU377860
EnvTARA008	Syn.	CRD1		Indian	-17.0	54.0	5	KU377861
BIOS-U3-1	Syn.	CRD1	2533	Pacific	-34.0	-73.4	5	KU377986
BIOS-E4-1	Syn.	CRD1	2534	Pacific	-31.9	-91.4	40	KU377987
EnvTARA009	Syn.	CRD1		Indian	-17.0	54.0	5	KU377989
AMT23ST27-2	Syn.	CRD1		Atlantic	-3.7	-28.7	12.4	KU705417
NV36	Syn.	CRD1		Mediterranean	40.8	14.3		KU705460
AMT12-A6	Syn.	CRD1						KU937828
Biosope_199	Syn.	CRD1A	1026	Pacific	-34.0	-73.4	5	JF307517
Biosope_45 C4Y	Syn.	CRD1B	1020	Pacific	-9.1	-137.0	100	JF307515
Biosope_148 D3	Syn.	CRD1B	1028	Pacific	-31.9	-91.4	40	JF307519
Biosope_48 B6Y	Syn.	CRD1B	1031	Pacific	-9.1	-137.0	30	JF307522
AMT15006-B10a	Syn.	EnvA		Atlantic	47.9	-14.6	10	JF307612
AMT15006-B12a	Syn.	EnvA		Atlantic	47.9	-14.6	10	JF307616
AMT15006-B12b	Syn.	EnvA		Atlantic	47.9	-14.6	10	JF307617
AMT15006-B22	Syn.	EnvA		Atlantic	47.9	-14.6	10	JF307628
AMT18012-A2	Syn.	EnvA		Atlantic	42.7	-22.2	12	JF307642
AMT18025-A21	Syn.	EnvA		Atlantic	27.6	-37.3	19	JF307651
BATS12	Syn.	EnvA		Atlantic	31.7	-64.2	6	JF307712
BATS13	Syn.	EnvA		Atlantic	31.7	-64.2	6	JF307713
BATS2a	Syn.	EnvA		Atlantic	31.7	-64.2	6	JF307718
BATS23	Syn.	EnvA		Atlantic	31.7	-64.2	6	KU705447
AMT15006-B11a	Syn.	EnvB		Atlantic	47.9	-14.6	10	JF307613
AMT15006-B11b	Syn.	EnvB		Atlantic	47.9	-14.6	10	JF307614
AMT15006-B11c	Syn.	EnvB		Atlantic	47.9	-14.6	10	JF307615
AMT15006-B14	Syn.	EnvB		Atlantic	47.9	-14.6	10	JF307619
AMT15006-B17	Syn.	EnvB		Atlantic	47.9	-14.6	10	JF307622
AMT15006-B19	Syn.	EnvB		Atlantic	47.9	-14.6	10	JF307625
AMT15006-B7s	Syn.	EnvB		Atlantic	47.9	-14.6	10	JF307630
AMT18012-A1	Syn.	EnvB		Atlantic	42.7	-22.2	12	JF307633
AMT18012-A10	Syn.	EnvB		Atlantic	42.7	-22.2	12	JF307634
AMT18012-A17	Syn.	EnvB		Atlantic	42.7	-22.2	12	JF307641
EnvTARA038	Syn.	EnvB		N. Pacific	14.2	-116.6	5	KU377854
EnvTARA037	Syn.	EnvB		N. Pacific	14.2	-116.6	5	KU377855
EnvTARA015	Syn.	EnvB		Pacific			5	KU377857
CC9311	Syn.	IA	1086	Pacific	32.0	-124.5	95	CP000435
BIOS-U3-2	Syn.	IA	2532	Pacific	-33.9	-73.3	30	JF307511
RCC524	Syn.	IA	524					JF307542
WH8020	Syn.	IA	2437	Atlantic	38.7	-69.3	50	JF307571
Ellet15	Syn.	IA		Atlantic	57.6	-13.6	5	JF307732
Ellet4a	Syn.	IA		Atlantic	57.6	-13.6	5	JF307741
Ellet4b	Syn.	IA		Atlantic	57.6	-13.6	5	JF307742
MICROVIR-3CR-1	Syn.	IA	1579	North Sea	49.3	-3.3	10	KU377805
A1814-154m	Syn.	IA		Atlantic	38.9	-25.3	154	KU705404
AMT23P4-3	Syn.	IA		English Channel	50.6	-0.71	15	KU705408
AMT23ST46-1	Syn.	IA		Atlantic	-35.2	-38.3	10	KU705424
R3B	Syn.	IA		Atlantic	38.5	-25.3		KU705432
Ellet33	Syn.	IA		Atlantic	57.6	-13.6	5	KU705456
Ellet34	Syn.	IA		Atlantic	57.6	-13.6	5	KU705457
R10B	Syn.	IA		Atlantic	27.4	-37.0	55	KU937819
R5B	Syn.	IA		Atlantic	36.1	-27.7	55	KU937820
R1B	Syn.	IA		Atlantic	38.5	-25.3		KU937821
AMT23ST47-1	Syn.	IA		Atlantic	-36.0	-39.2	20	KU937825
AMT23ST01-5	Syn.	IA		Atlantic	44.0	-13.6	9.9	KU937827
ALMO3	Syn.	IB	2432	Mediterranean	36.2	-1.9		JF307500
ROS8604	Syn.	IB	2380	English Channel	48.7	-4.0		JF307527
PROS-9-1	Syn.	IB	328	Mediterranean	41.9	10.4	30	JF307534

Table 1 (*continued*)

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
SYN20	Syn.	IB	2035	North Sea	60.6	5.7		JF307568
WH8016	Syn.	IB	2535	Atlantic				JF307569
AMT15002-C1a	Syn.	IB		Atlantic	48.8	-7.9	25	JF307601
AMT15002-C1c	Syn.	IB		Atlantic	48.8	-7.9	25	JF307603
AMT15002-C2a	Syn.	IB		Atlantic	48.8	-7.9	25	JF307605
AMT15002-C2b	Syn.	IB		Atlantic	48.8	-7.9	25	JF307606
AMT15002-C7	Syn.	IB		Atlantic	48.8	-7.9	25	JF307611
AMT15006-B16	Syn.	IB		Atlantic	47.9	-14.6	10	JF307621
AMT15006-B21	Syn.	IB		Atlantic	47.9	-14.6	10	JF307627
Arctic732-1b	Syn.	IB		Arctic	72.1	20.0	15	JF307695
Arctic732-25b	Syn.	IB		Arctic	72.1	20.0	15	JF307697
Arctic732-2a	Syn.	IB		Arctic	72.1	20.0	15	JF307698
Arctic732-2b	Syn.	IB		Arctic	72.1	20.0	15	JF307699
Arctic732-35a	Syn.	IB		Arctic	72.1	20.0	15	JF307700
Arctic732-35b	Syn.	IB		Arctic	72.1	20.0	15	JF307701
Ellet06	Syn.	IB		Atlantic	57.6	-13.6	5	JF307723
Ellet08	Syn.	IB		Atlantic	57.6	-13.6	5	JF307725
Ellet09	Syn.	IB		Atlantic	57.6	-13.6	5	JF307726
Ellet11	Syn.	IB		Atlantic	57.6	-13.6	5	JF307728
Ellet12	Syn.	IB		Atlantic	57.6	-13.6	5	JF307729
Ellet16	Syn.	IB		Atlantic	57.6	-13.6	5	JF307733
Ellet1b	Syn.	IB		Atlantic	57.6	-13.6	5	JF307736
MVIR-1-1	Syn.	IB	1708	North Sea	48.8	-3.9	10	KF443067
MVIR-10-1	Syn.	IB	1688	North Sea	55.7	2.3	40	KF443068
MVIR-18-1	Syn.	IB	2385	North Sea	61.0	2.0	25	KF443072
MICROVIR-19CH-3	Syn.	IB	1604	English Channel	48.8	-3.9	10	KU377786
MICROVIR-18K-4-1	Syn.	IB	1647	North Sea	53.2	2.8	10	KU377787
MICROVIR-17CH-1	Syn.	IB	1675	North Sea	54.4	4.1	10	KU377789
MICROVIR-16CR-4-4	Syn.	IB	3012					KU377790
MICROVIR-15CR-3	Syn.	IB	1672	North Sea	59.2	0.7	20	KU377791
MICROVIR-14K-4-5	Syn.	IB	1599	North Sea	57.7	8.7	10	KU377792
MICROVIR-13CR-4-1	Syn.	IB	1584	North Sea	57.0	4.0	30	KU377793
MICROVIR-12CR-4-3	Syn.	IB	1670	North Sea	57.3	-0.3	10	KU377794
MICROVIR-12CH-4-1	Syn.	IB	1684	North Sea	57.3	-0.3	10	KU377795
MICROVIR-11CR-4-2	Syn.	IB	1586	North Sea	57.3	-0.3	35	KU377796
MICROVIR-9K-2	Syn.	IB	1643	North Sea	59.3	4.3	10	KU377797
MICROVIR-9CH-1	Syn.	IB	1638	North Sea	59.3	4.3	10	KU377798
MICROVIR-8CR-4-1	Syn.	IB	1636	North Sea	55.7	2.3	40	KU377800
MICROVIR-7CR-4-6	Syn.	IB	1591	North Sea	55.7	2.3	10	KU377801
MICROVIR-7CR-3	Syn.	IB	1589	North Sea	55.7	2.3	10	KU377802
MICROVIR-6CH-1	Syn.	IB	1653	North Sea	57.0	4.0	10	KU377803
MICROVIR-5CR-4-3	Syn.	IB	1574	North Sea	59.2	0.7	10	KU377804
MICROVIR-3CH-0	Syn.	IB	1630	North Sea	49.3	-3.3	10	KU377806
MICROVIR-2CR-4-3	Syn.	IB	1628	North Sea	54.4	4.1	30	KU377809
MICROVIR-1CH-4-2	Syn.	IB	1667	North Sea	59.3	4.3	50	KU377810
MICROVIR-1CH-2	Syn.	IB	1665	North Sea	59.3	4.3	50	KU377811
A18-98-78m	Syn.	IB		Atlantic	38.5	-38.3	78	KU705406
AMT23ST01-6	Syn.	IB		Atlantic	44.0	-13.6	9.9	KU705409
AMT23ST49-1	Syn.	IB		Atlantic	-38.8	-42.1	20	KU705425
Ellet20	Syn.	IB		Atlantic	57.6	-13.6	5	KU705450
Ellet31	Syn.	IB		Atlantic	57.6	-13.6	5	KU705455
Ellet35	Syn.	IB		Atlantic	57.6	-13.6	5	KU705458
Ellet10	Syn.	IC		Atlantic	57.6	-13.6	5	JF307727
Biosope-141-D	Syn.	IC	2568	Pacific	-31.9	-91.4	40	KU377847
Biosope-141-D	Syn.	IC	2568	Pacific	-31.9	0.0	40	KU377847
JCVI-1105112153936	Syn.	IIA						ER323992
A15-37	Syn.	IIA	2526	Atlantic	23.6	-20.0	10	JF307486
A15-38	Syn.	IIA		Atlantic	23.6	20.0	20	JF307487
TAK9802	Syn.	IIA	2528	Pacific	-14.5	-145.3	7	JF307523
M16.1	Syn.	IIA	791	Atlantic	27.7	-91.3	275	JF307548
RS9902	Syn.	IIA	2376	Red Sea	29.5	34.9	1	JF307553
RS9904	Syn.	IIA	543	Red Sea	29.5	34.9	10	JF307554
RS9907	Syn.	IIA	2382	Red Sea	29.5	34.9	10	JF307557
RS9911	Syn.	IIA	550	Red Sea	29.5	34.9	10	JF307561
RS9919	Syn.	IIA	558					JF307566
AMT15034-A23a	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307575
AMT15034-A23b	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307576

Table 1 (*continued*)

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
AMT15034-A24	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307577
AMT15034-A26	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307579
AMT15034-A30	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307583
AMT15034-A32	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307585
AMT15034-A33	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307586
AMT15034-C23a	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307588
AMT15034-C23b	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307589
AMT15034-C26a	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307590
AMT15034-C26b	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307591
AMT15034-C30	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307593
AMT15034-C36	Syn.	IIA		Atlantic	21.7	-17.8	2	JF307598
A15-44	Syn.	IIA	2527	Atlantic	21.7	-17.8	20	KF443066
CARBOM-02-Syn-04	Syn.	IIA						KU377812
AMT23ST25-1	Syn.	IIA		Atlantic	4.4	-29.4	10.3	KU705415
AMT23ST27-1	Syn.	IIA		Atlantic	-3.7	-28.7	12.4	KU705416
S59	Syn.	IIA		Atlantic	10.4	-32.4	15	KU705437
S60	Syn.	IIA		Atlantic	8.3	-31.0	28	KU705438
RS9908	Syn.	IIA	547	Gulf of Aqaba	29.5	34.9	10	JF307558
WH8109	Syn.	IIB	2033	Atlantic	39.5	-70.5		CP006882
AMT18012-A7	Syn.	IIB		Atlantic	42.7	-22.2	12	JF307646
AMT18025-A20	Syn.	IIB		Atlantic	27.6	-37.3	19	JF307650
AMT18025-A22	Syn.	IIB		Atlantic	27.6	-37.3	19	JF307652
AMT18025-A25	Syn.	IIB		Atlantic	27.6	-37.3	19	JF307655
AMT18025-A26	Syn.	IIB		Atlantic	27.6	-37.3	19	JF307656
AMT18025-A35	Syn.	IIB		Atlantic	27.6	-37.3	19	JF307664
AMT18070-A57	Syn.	IIB		Atlantic	-16.6	-25.0	18	JF307673
AMT18070-A60	Syn.	IIB		Atlantic	-16.6	-25.0	18	JF307676
AMT18070-A61	Syn.	IIB		Atlantic	-16.6	-25.0	18	JF307677
AMT18070-A69	Syn.	IIB		Atlantic	-16.6	-25.0	18	JF307685
AMT18070-A71	Syn.	IIB		Atlantic	-16.6	-25.0	18	JF307687
A14C38	Syn.	IIB						KU705398
A15-20	Syn.	IIB	1099	Atlantic	38.9	-21.4	110	JF307479
CC9605	Syn.	IIC	753	Pacific	30.4	-124.0	51	CP000110
A15-146	Syn.	IIC		Atlantic	40.0	10.0	50	JF307476
A15-19	Syn.	IIC	2525	Atlantic	38.9	-20.4	75	JF307478
A15-62	Syn.	IIC	2374	Atlantic	17.6	-21.0	15	JF307497
AMT15034-A28	Syn.	IIC		Atlantic	21.7	-17.8	2	JF307581
AMT15034-A31	Syn.	IIC		Atlantic	21.7	-17.8	2	JF307584
AMT15034-C34a	Syn.	IIC		Atlantic	21.7	-17.8	2	JF307595
AMT15034-C35	Syn.	IIC		Atlantic	21.7	-17.8	2	JF307597
AMT18012-A12	Syn.	IIC		Atlantic	42.7	-22.2	12	JF307636
AMT18070-A51a	Syn.	IIC		Atlantic	-16.6	-25.0	18	JF307666
AMT18070-A58	Syn.	IIC		Atlantic	-16.6	-25.0	18	JF307674
AMT18070-A65	Syn.	IIC		Atlantic	-16.6	-25.0	18	JF307681
AMT18070-A75	Syn.	IIC		Atlantic	-16.6	-25.0	18	JF307691
AMT23ST41-1	Syn.	IIC		Atlantic	-27.8	-30.9	20.2	KU705421
AMT23ST41-3	Syn.	IIC		Atlantic	-27.8	-30.9	20.2	KU705423
M1	Syn.	IIC		Atlantic	-23.6	-24.6	2	KU705430
S50	Syn.	IIC		Atlantic	19.3	-38.1	2	KU705436
AMT12-A14	Syn.	IIC		Atlantic				KU705440
AMT18070-A68	Syn.	IID		Atlantic	-16.6	-25.0	18	JF307684
AMT18070-A77	Syn.	IID		Atlantic	-16.6	-25.0	18	JF307693
PROS-3-1	Syn.	IIE	321	Mediterranean	38.0	3.8	5	JF307529
PROS-5-2	Syn.	IIE	374	Mediterranean	36.5	13.3	25	JF307537
BOUM84-1	Syn.	IIE	2455	Mediterranean	35.7	14.1	100	KU377830
AMT23P3-1	Syn.	IIE		English Channel	51.0	1.3	14	KU705407
A15-72	Syn.	IIF	1093	Atlantic	0.0			JF307498
AMT15034-A10s	Syn.	IIF		Atlantic	21.7	-17.8	2	JF307574
AMT18012-A8	Syn.	IIG		Atlantic	42.7	-22.2	12	JF307647
AMT18070-A55	Syn.	IIG		Atlantic	-16.6	-25.0	18	JF307671
AMT18070-A56	Syn.	IIG		Atlantic	-16.6	-25.0	18	JF307672
AMT18070-A64	Syn.	IIG		Atlantic	-16.6	-25.0	18	JF307680
AMT18070-A73	Syn.	IIG		Atlantic	-16.6	-25.0	18	JF307689
A15-147	Syn.	IIH		Atlantic	40.0	10.0		JF307477
PROS-U-1	Syn.	IIH	2369	Atlantic	30.1	-10.1	5	JF307538
AMT15034-A27	Syn.	IIH		Atlantic	21.7	-17.8	2	JF307580
AMT15034-C27	Syn.	IIH		Atlantic	21.7	-17.8	2	JF307592
AMT23ST14-1	Syn.	IIH		Atlantic	24.2	-26.3	14.1	KU705413
AMT23ST41-2	Syn.	IIH		Atlantic	-27.8	-30.9	20.2	KU705422

Table 1 (*continued*)

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
AMT2013-P209-F2	Syn.	IIH		English Channel	51.3	1	12	KU705428
AMT2013-P209-F4	Syn.	IIH		English Channel	51.3	1	13	KU705429
S68	Syn.	IIH		Atlantic	38.3	-38.3	9	KU705439
R1A	Syn.	IIH		Atlantic	38.5	-25.3	0	KU937822
PROSPOE 2-16	Syn.	IIH	442	Atlantic	30.1	-10.1	5	JF307540
A15-22	Syn.	III	1097	Atlantic	35.1	20.9	15	JF307480
A15-30	Syn.	III		Atlantic	31.3	-20.7	40	JF307484
A15-11	Syn.	IIIA	2569	Atlantic	46.0	-18.4	58	JF307472
A15-24	Syn.	IIIA	1087	Atlantic	35.1	-20.9	35	JF307481
RS9915	Syn.	IIIA	2553	Red Sea	29.5	34.9	10	JF307565
AMT18025-A19	Syn.	IIIA		Atlantic	27.6	-37.3	19	JF307649
AMT18025-A24	Syn.	IIIA		Atlantic	27.6	-37.3	19	JF307654
AMT18025-A33	Syn.	IIIA		Atlantic	27.6	-37.3	19	JF307662
AMT18025-A34	Syn.	IIIA		Atlantic	27.6	-37.3	19	JF307663
AMT18025-A36	Syn.	IIIA		Atlantic	27.6	-37.3	19	JF307665
BATS04	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307704
BATS05	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307705
BATS06	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307706
BATS08	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307708
BATS09	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307709
BATS10	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307710
BATS11	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307711
BATS14	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307714
BATS15	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307715
BATS3a	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307720
BATS3b	Syn.	IIIA		Atlantic	31.7	-64.2	6	JF307721
BOUM113-3	Syn.	IIIA	2416	Mediterranean	34.2	22.2	5	KU377813
BOUM112-1	Syn.	IIIA	2422	Mediterranean	38.1	10.2	5	KU377814
BOUM109-2	Syn.	IIIA	2425	Mediterranean	34.1	18.5	5	KU377815
BOUM107-16	Syn.	IIIA	2430	Mediterranean	34.1	18.5	5	KU377816
BOUM107-14	Syn.	IIIA	2429	Mediterranean	34.1	18.5	5	KU377817
BOUM104-5	Syn.	IIIA	2418	Mediterranean	33.9	26.8	5	KU377818
BOUM93-0	Syn.	IIIA	2466	Mediterranean	33.9	26.8	50	KU377821
BOUM91-4	Syn.	IIIA	2444	Mediterranean	34.2	22.2	50	KU377822
BOUM91-3	Syn.	IIIA	2463	Mediterranean	34.2	22.2	50	KU377823
BOUM91-1	Syn.	IIIA	2453	Mediterranean	34.2	22.2	50	KU377824
BOUM90-3	Syn.	IIIA	2446	Mediterranean	34.2	22.2	50	KU377825
BOUM85-0	Syn.	IIIA	2467	Mediterranean	33.6	32.6	100	KU377827
BOUM84-12	Syn.	IIIA	2450	Mediterranean	35.7	14.1	100	KU377828
BOUM84-5	Syn.	IIIA	2447	Mediterranean	35.7	14.1	100	KU377829
BOUM83-3	Syn.	IIIA	2462	Mediterranean	35.7	14.1	100	KU377831
BOUM81-5	Syn.	IIIA	2471	Mediterranean	38.1	10.2	60	KU377832
BOUM80-11	Syn.	IIIA	2440	Mediterranean	41.1	5.1	60	KU377834
BOUM79-2	Syn.	IIIA	2439	Mediterranean	41.1	5.1	60	KU377835
BOUM78-7	Syn.	IIIA	2461	Mediterranean	43.2	4.9	30	KU377836
BOUM78-1	Syn.	IIIA	2460	Mediterranean	43.2	4.9	30	KU377837
BOUM72-1	Syn.	IIIA	2458	Mediterranean	34.1	18.5	120	KU377838
BOUM70-7	Syn.	IIIA	2470	Mediterranean	34.1	18.5	120	KU377839
BOUM69-2	Syn.	IIIA	2448	Mediterranean	34.1	18.5	120	KU377840
BOUM68-1	Syn.	IIIA	2419	Mediterranean	39.1	5.3	12	KU377841
BOUM32-0	Syn.	IIIA	2465	Mediterranean	39.1	5.3	100	KU377843
BOUM29-3	Syn.	IIIA	2469	Mediterranean	39.1	5.3	100	KU377845
BOUM29-1	Syn.	IIIA	2468	Mediterranean	39.1	5.3	100	KU377846
BOUM118	Syn.	IIIA	2379	Mediterranean	33.6	32.6	5	KU377984
A18-40	Syn.	IIIA		Atlantic	10.6	-32.1	2	KU670815
A18-46.1	Syn.	IIIA		Atlantic	10.6	-32.1	15	KU670816
A18-16	Syn.	IIIA		Atlantic	36.1	-27.5	55	KU705399
A18-19b-surf	Syn.	IIIA		Atlantic	33.2	-30.5	2	KU705400
A18-22	Syn.	IIIA		Atlantic	30.2	-33.6	48	KU705401
A18-23-surf	Syn.	IIIA		Atlantic	27.6	-37.0	2	KU705402
A1814b-surf	Syn.	IIIA		Atlantic	38.9	-25.3	2	KU705405
AMT23ST08-1	Syn.	IIIA		Atlantic	35.0	-20.7	8.4	KU705410
AMT23ST09-1	Syn.	IIIA		Atlantic	33.9	-21.3	18.9	KU705411
AMT23ST15-1	Syn.	IIIA		Atlantic	23.1	-26.8	14.5	KU705414
R3A	Syn.	IIIA		Atlantic	38.5	-25.3	0	KU705431
R4	Syn.	IIIA		Atlantic	36.1	-27.7	55	KU705433
R5A	Syn.	IIIA		Atlantic	36.1	-27.7	55	KU705434
R11	Syn.	IIIA		Atlantic	27.4	-37.0	74	KU705435
BATS16	Syn.	IIIA		Atlantic	31.7	-64.2	6	KU705441

Table 1 (*continued*)

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
BATS20	Syn.	IIIA		Atlantic	31.7	-64.2	6	KU705445
BATS21	Syn.	IIIA		Atlantic	31.7	-64.2	6	KU705446
NV12	Syn.	IIIA		Mediterranean	40.8	14.3	0	KU705459
R12	Syn.	IIIA		Atlantic	27.4	-37.2	74	KU937818
AMT23ST07-3	Syn.	IIIA		Atlantic	38.2	-19.0	8.1	KU937826
A18-25b-55m	Syn.	IIIA		Atlantic	27.3	-37.0	55	KU937830
EUM14v	Syn.	IIIA	37	Atlantic	21.0	-31.1	105	JF307508
OLI31v	Syn.	IIIA	44	Pacific	-5.5	-150.0	70	JF307510
RS9905	Syn.	IIIA	544	Gulf of Aqaba	29.5	34.9	10	JF307555
WH8103	Syn.	IIIA	29	Sargasso Sea	28.5	-67.4		JF307573
A15-28	Syn.	IIIB	2556	Atlantic	31.3	-20.7	15	JF307482
AMT18025-A18	Syn.	IIIB		Atlantic	27.0	-37.3	19	JF307648
BATS1b	Syn.	IIIB		Atlantic	31.7	-64.2	6	JF307717
A15-29	Syn.	IIIB	1102	Atlantic	31.3	-20.7	15	JF307483
AMT18025-A29	Syn.	IIIC		Atlantic	27.6	-37.3	19	JF307659
AMT18025-A32	Syn.	IIIC		Atlantic	27.6	-37.3	19	JF307661
AMT18070-A53	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307669
AMT18070-A54	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307670
AMT18070-A62	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307678
AMT18070-A63	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307679
AMT18070-A66	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307682
AMT18070-A70	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307686
AMT18070-A72	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307688
AMT18070-A74	Syn.	IIIC		Atlantic	-16.6	-25.0	18	JF307690
AMT18070-D86	Syn.	IIIC		Atlantic	-16.6	-25.0	18	KU937824
AMT15034-C15s	Syn.	IVA		Atlantic	21.7	-17.8	2	JF307587
AMT15002-C1d	Syn.	IVA		Atlantic	48.8	-7.9	25	JF307604
AMT15002-C3	Syn.	IVA		Atlantic	48.8	-7.9	25	JF307607
AMT15002-C5b	Syn.	IVA		Atlantic	48.8	-7.9	25	JF307609
AMT15002-C6	Syn.	IVA		Atlantic	48.8	-7.9	25	JF307610
AMT15006-B15	Syn.	IVA		Atlantic	47.9	-14.6	10	JF307620
AMT15006-B18a	Syn.	IVA		Atlantic	47.9	-14.6	10	JF307623
AMT18012-A4	Syn.	IVA		Atlantic	42.7	-22.2	12	JF307643
Arctic732-1a	Syn.	IVA		Arctic	72.1	20.0	15	JF307694
Ellet07	Syn.	IVA		Atlantic	57.6	-13.6	5	JF307724
MVIR-7-1	Syn.	IVA	1648	North Sea	53.2	2.8	10	KF443073
MICROVIR-20CR-2	Syn.	IVA	1656	North Sea	50.2	0.3	10	KU377785
MICROVIR-17K-4-6	Syn.	IVA	1681	North Sea	54.4	4.1	10	KU377788
MICROVIR-8CR-4-5	Syn.	IVA	1637	North Sea	55.7	2.3	40	KU377799
MICROVIR-2K-2	Syn.	IVA	1633	North Sea	54.4	4.1	30	KU377807
MICROVIR-2CR-4-5	Syn.	IVA	1632	North Sea	54.4	4.1	30	KU377808
BOUM31-10	Syn.	IVA	2459	Mediterranean	39.1	5.3	100	KU377844
AMT23ST52-1	Syn.	IVA		Atlantic	-43.7	-47.7	11.8	KU705426
AMT23ST53-2	Syn.	IVA		Atlantic	-45.4	-49.7	8.5	KU705427
BL107	Syn.	IVA	515	Mediterranean	41.7		1800	
AMT15006-B13	Syn.	IVB		Atlantic	47.9	-14.6	10	JF307618
AMT15006-B18b	Syn.	IVB		Atlantic	47.9	-14.6	10	JF307624
AMT15006-B20	Syn.	IVB		Atlantic	47.9	-14.6	10	JF307626
Arctic732-8a	Syn.	IVB		Arctic	72.1	20.0	15	JF307702
Arctic732-8b	Syn.	IVB		Arctic	72.1	20.0	15	JF307703
Ellet05	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307722
Ellet13	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307730
Ellet14	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307731
Ellet17	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307734
Ellet2a	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307737
Ellet2b	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307738
Ellet3a	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307739
Ellet3b	Syn.	IVB		Atlantic	57.6	-13.6	5	JF307740
Ellet18	Syn.	IVB		Atlantic	57.6	-13.6	5	KU705448
Ellet19	Syn.	IVB		Atlantic	57.6	-13.6	5	KU705449
Ellet24	Syn.	IVB		Atlantic	57.6	-13.6	5	KU705451
Ellet25	Syn.	IVB		Atlantic	57.6	-13.6	5	KU705452
Ellet26	Syn.	IVB		Atlantic	57.6	-13.6	5	KU705453
Ellet29	Syn.	IVB		Atlantic	57.6	-13.6	5	KU705454
Ellet21	Syn.	IVB		Atlantic	57.6	-13.6	5	KU937823
RS9901	Syn.	IX	2529	Red Sea	29.5	34.9	1	JF307552
RS9916	Syn.	IX	555	Red Sea	29.5		10	
RS9921	Syn.	IX	559	Gulf of Aqaba	29.5	34.9	10	JF307567
WH7803	Syn.	V	28	Atlantic	33.8	-67.5	25	CT971583

Table 1 (*continued*)

Strain	Genus	Clade/Subclade	RCC	Origin	Latitude	Longitude	Depth	*petB* Accession
BMK-MC-1	Syn.	V	2438	Mediterranean	40.8	14.3	23	KU377985
A15-43	Syn.	V		Atlantic	21.7	17.8	40	JF307488
DIM(UW01)	Syn.	V	650		41.7	-3.6		JF307507
PROSOPE_77-2	Syn.	VI	329	Mediterranean	34.0	22.0	25	JF307471
PROS-5-3	Syn.	VIA	324	Mediterranean	36.5	13.3	5	JF307470
BL161	Syn.	VIA		Mediterranean	41.7	2.8		JF307502
BL164	Syn.	VIA		Mediterranean	41.7	2.8		JF307503
BL48	Syn.	VIA	511	Mediterranean	41.7	2.8		JF307505
BL8	Syn.	VIA		Mediterranean	41.7	2.8		JF307506
PROSOPE-53-19	Syn.	VIA	326	Mediterranean	36.5	13.3	5	JF307533
MEDNS5	Syn.	VIA	2368	Mediterranean	41.0	6.0	80	JF307539
PROS-8-2	Syn.	VIA	523	Mediterranean	39.1	14.1	70	JF307541
PROS-8-1	Syn.	VIA	527	Mediterranean	39.1	14.1	110	JF307544
PROS-M-1	Syn.	VIA	528	Mediterranean	34.0	22.0	50	JF307545
BOUM102-1	Syn.	VIA	2420	Mediterranean	33.9	31.9	5	KU377820
BOUM87-4	Syn.	VIA	2457	Mediterranean	33.6	32.6	100	KU377826
BOUM81-1	Syn.	VIA	2456	Mediterranean	38.1	10.2	60	KU377833
772	Syn.	VIA	329	Mediterranean	34.0	22.0	5	KU705397
PROSOPE_157-13	Syn.	VIA	320	Mediterranean	43.4	7.8		JF307528
PROSOPE_153-3	Syn.	VIA	325	Mediterranean	43.4	7.8	25	JF307532
PROS-D-1	Syn.	VIB	319	Mediterranean	43.4	7.8	15	JF307469
PROS-7-1	Syn.	VIB	2381	Mediterranean	37.4	15.6	5	JF307531
BL36	Syn.	VIB	508	Mediterranean	41.7	2.8	0	JF307504
PROSOPE_117-1	Syn.	VIB	322	Mediterranean	39.1	14.1	5	JF307530
RA000711-27-14	Syn.	VIB	359	English Channel	48.8	-4.0		JF307535
PROSOPE_25	Syn.	VIB	367	Mediterranean	38.0	3.8	25	JF307536
BL1237	Syn.	VIC		Mediterranean	41.7	2.8		JF307501
RCC525	Syn.	VIC	525					JF307543
BOUM104-4	Syn.	VIC	2417	Mediterranean	33.9	26.8	5	KU377819
M16.3	Syn.	VIC	792	Gulf of Mexico	27.7	-91.3	275	JF307549
WH8018	Syn.	VIC	2373	Atlantic	41.5	-70.7		JF307570
A15-60	Syn.	VII	2554	Atlantic	17.6	-21.0	10	JF307495
A15-74	Syn.	VII	1094	Atlantic	7.9	-23.3	25	JF307499
NOUM97012	Syn.	VII	67	Pacific	-22.3	166.3	80	JF307547
NOUM97013	Syn.	VII	2433	Pacific	-22.3	166.3		KU377982
A18-25c	Syn.	VII		Atlantic	27.3	-37.0	74	KU670814
AMT23ST11-2	Syn.	VII		Atlantic	30.4	-23.1	13.1	KU705412
JCVI110514332283b	Syn.	VIII						EK871779
JCVI-110516443717	Syn.	VIII		Pacific				ER072567.1
WH8101	Syn.	VIII	2555	Atlantic				JF307572
RS9906	Syn.	VIII	545	Gulf of Aqaba	29.5	34.9	10	JF307556
RS9909	Syn.	VIII	548	Gulf of Aqaba	29.5	34.9	10	JF307559
A15-127	Syn.	WPC1	2378	Atlantic	-31.1	-3.9	45	JF307473
AMT15034-A29	Syn.	WPC1		Atlantic	21.7	-17.8	2	JF307582
BATS07	Syn.	WPC1		Atlantic	31.7	-64.2	6	JF307707
A18-32b	Syn.	WPC1		Atlantic	22.3	-40.2	2	KU705403
BATS19	Syn.	WPC1		Atlantic	31.7	-64.2	6	KU705444
A18-25-surf	Syn.	WPC1		Atlantic	27.3	-37.0	2	KU937829
KORDI-49	Syn.	WPC1		Pacific	32.5		20	
EnvTARA048	Syn.	WPC1		Misc.			5	KU377849
AMT23ST31-2	Syn.	WPC1		Atlantic	-11.0		10.4	KU705418
AMT23ST31-3	Syn.	WPC1		Atlantic	-10.0		19.2	KU705419
AMT23ST31-5	Syn.	WPC1		Atlantic	-10.0		19.2	KU705420
A15-130	Syn.	WPC1		Atlantic	-34.5	-1.4	20	JF307474
PROS-3-2	Syn.	XVI	316	Mediterranean	38.0	3.8	110	JF307524
PROS-7-2	Syn.	XVI	2567	Mediterranean	37.4	15.6	90	JF307525
PROS-5-1	Syn.	XVI	318	Mediterranean	36.5	13.3	65	JF307526
BOUM33-22	Syn.	XVI	2427	Mediterranean	39.1	5.3	100	KU377842
M21B.3	Syn.	XVI	794	Gulf of Mexico	27.7	-91.3	275	JF307550
AMT18012-A11	Syn.	XX		Atlantic	42.7	-22.2	12	JF307635
AMT18012-A15	Syn.	XX		Atlantic	42.7	-22.2	12	JF307639
AMT18012-A5	Syn.	XX		Atlantic	42.7	-22.2	12	JF307644
AMT18025-A23	Syn.	XX		Atlantic	27.6	-37.3	19	JF307653
AMT18070-A67	Syn.	XX		Atlantic	-16.6	-25.0	18	JF307683
BATS2b	Syn.	XX		Atlantic	31.7	-64.2	6	JF307719
BATS17	Syn.	XX		Atlantic	31.7	-64.2	6	KU705442
BATS18	Syn.	XX		Atlantic	31.7	-64.2	6	KU705443
CC9616	Syn.	XX		Pacific				

PCR conditions consisted of an initial denaturation step at 94 °C for 15 min, 35 cycles of 30 s at 94 °C and 30 s at the different theoretical annealing temperatures calculated for each primer pair, 45 s at 72 °C, and a final elongation step of 5 min at 72 °C. The PCR products obtained were checked and quantified using Agilent TapeStation 4200 (Agilent Technologies, Santa Clara, USA). Subsequently, the amplicons were cleaned up using AMPure XP beads (Beckman Coulter, Brea, CA, USA), and sequenced by Sanger in both directions in our genomics facility.

Annealing temperature optimization for NGS primers

According to these results, we selected the primer pairs producing higher PCR product concentrations and strain coverages to run an annealing temperature optimization assay (Fig. 1). This test allowed us to determine the best PCR conditions to maximize the potential for amplification of the primer pairs selected. To run this optimization assay we purchased a new set of the selected primers with attached Illumina adaptor sequences (Forward overhang: 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG - *specific locus*, and Reverse overhang: 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG - *specific locus*), and used the RCC cultures less amplified on the previous step, covering 12 different annealing temperatures from 39.9 °C to 60.1 °C and using similar PCR conditions as those described above. PCR products were verified and quantified using Agilent TapeStation 4200 (Agilent Technologies, Santa Clara, USA).

Taxonomic characterization of reference cultures

After sequencing, the forward and reverse reads obtained for the RCC cultures were assembled, aligned to the reference database by MUSCLE [12], and manually trimmed with Geneious® 10.0.8 software [13]. Two phylogenetic trees were constructed using neighbor-joining clustering, with HKY as the substitution model and 1000 bootstrap replicates, and edited with iTOL (itol.embl.de) [14] (Fig. 2a and b). The sequences obtained from the RCC cultures were added to our reference database.

Analysis of a mock community and environmental samples

To compare and validate the accuracy of the amplification of the primer pairs, we prepared a mock community sample. The DNA obtained from each RCC culture was amplified, quantified, pooled in a mixture at the same concentration, and then sequenced.

Also, we collected natural Red Sea communities to test the performance of the primer pair selected. Seawater was sampled from 3 different depths at a coastal station nearby KAUST (22.28 °N, 38.85 °E), using 10 L Niskin bottles attached to a rosette-CTD system. Approximately 1 mL of sample was used to measure cyanobacteria cell abundances immediately after sampling, using a CyFlow® Space flow cytometer equipped with a blue laser beam (488nm).

For phylogenetic analysis, we filtered approximately 8 L through 3 μM and 0.2 μM Isopore™ membrane filters (142 mm), using a Millipore Masterflex peristaltic pump (Fig. 1). The 0.2 μM filters were kept carefully folded in 15 mL plastic tubes, and immediately frozen at -80 °C. DNA was extracted following the same method described for the RCC cultures, and amplified via PCR (final volume 30μl per sample) using Qiagen multiplex PCR master mix (QIAGEN, Valencia, CA), 3 μl gDNA as DNA template and a final primer concentration of 0.3 μM (Fig. 1). We followed the PCR conditions indicated above, with the optimized annealing temperature for the selected primer pair. Replicate PCR products were pooled and checked by gel electrophoresis (1.5% agarose).

Amplicons were cleaned using AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA), adding Nextera® XT Indexes via PCR, following the Illumina sequencing library preparation guide. Indexed amplicons were cleaned again by AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA), quantified using a Qubit® fluorimeter (Life Technologies, Carlsbad, CA), and then pooled in equimolar amounts.

A qPCR was performed for pool quantification using KAPA SYBR fast qPCR master mix and the pool size was cheeked on a Bioanalyzer (Agilent Technologies, Santa Clara, USA) (Fig. 1). A 6 pM pool was prepared for Illumina MiSeq platform sequencing using 2×300bp paired-end reads using MiSeq reagent Kit v3 (Illumina, Inc.) with 25% PhiX control. Pooled amplicons were sequenced at KAUST CORELab facilities.

Data analysis

Raw sequences were processed, analyzed, and filtered using QIIME 1.9.1 software [15] (Fig. 1). Primer sequences were removed from the ends of each sequence using *cutadapt* [16], and forward and reverse sequences were assembled using PEAR [17]. We excluded from the analysis those reads with a quality score lower than 30, with assembled lengths shorter than 250 bp or bigger than 450 bp. Filtered sequences were then dereplicated with the VSEARCH tool [18], and clustered by a 94% similarity, as determined by computing within-group distances using Mega 7.0.26 [19].

After excluding singletons and chimeras, the sequences were classified with MOTHUR [20] against our reference database (with RCC sequences included) (Fig. 1).

We also run DADA2 [21] to classify the reads obtained from natural samples to provide comparable results with this widely used pipeline.

We obtained a rarefaction curve of the number of groups observed and estimated Shannon indexes for TapeStation results by applying the *alpha_diversity.py* script on QIIME. Linear correlations were calculated with GraphPad Prism 7.0.

1. Compilation of a *Prochlorococcus* and *Synechococcus petB* reference database.

2. *In silico* design of primer pairs, optimization and preselection of candidate primers.

3. Candidate primers are tested with RCC cultures of *Prochlorococcus* and *Synechococcus:*
   - Strain coverage assessment.
   - Annealing temperature optimization.

4. The primer pair selected is tested again using natural samples from the Red Sea:
   - Samples are collected in 0.2 µm filters and kept frozen (-80°C). DNA is extracted from the filters with commercial kits.
   - PCR amplification setting the optimal annealing temperature determined previously.
   - Amplicons are cleaned, and indexed via PCR. After a second cleaning step, indexed amplicons are quantified.
   - A qPCR is performed for pool quantification and the pool size is checked by Bioanalyzer.
   - Illumina MiSeq sequencing. The raw sequences obtained are recovered for further bioinformatic processing.

5. Raw sequences are trimmed, quality-checked, dereplicated, singletons and chimeras are removed.

6. Processed *Prochlorococcus* and *Synechococcus* reads are classified to a low taxonomic level against the reference database.

Fig. 1. Graphical abstract summarizing the steps followed in this work.
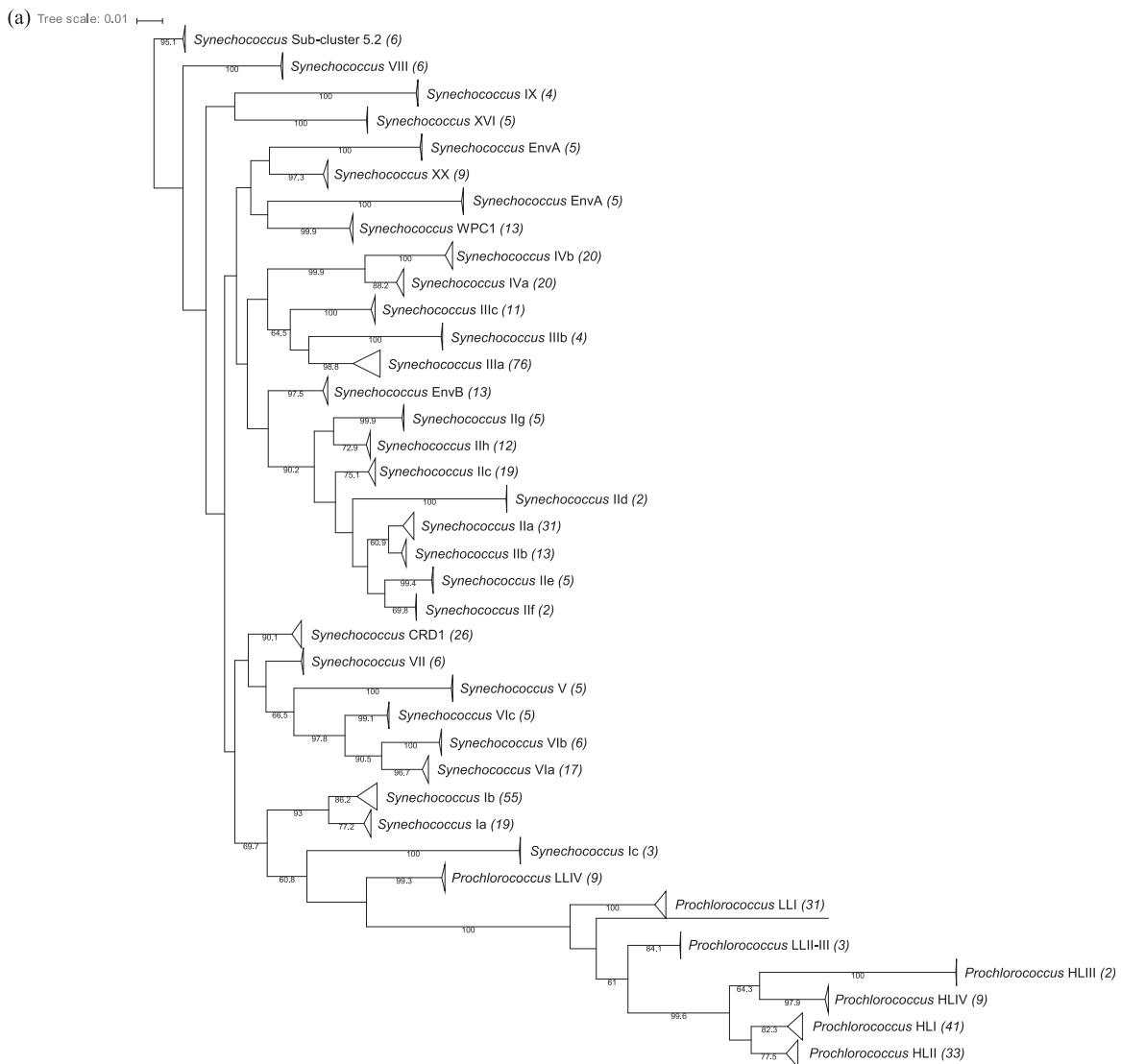
Fig. 2. Neighbor-Joining trees of the *Prochlorococcus* and *Synechococcus petB* sequences from the aligned reference database and RCC cultures. In (a) we used the whole *petB* sequences, and in (b) the *petB* sequences were trimmed to match the shorter sequences obtained using the 180f/587r primer pair. Values in parentheses indicate the total number of sequences included in the correspondent branch. Only bootstrap values >60% from 1000 replicates are shown. The scale bar indicates the number of nucleotide substitutions per site. Trees were created in Geneious® and edited using iTOL (https://itol.embl.de; Letunic and Bork, 2016).

Method validation

Using PrimerProspector, we obtained an initial set of 8 candidate primer pairs, with estimated amplicon sizes varying between 319 and 442 bp (Table S1). The efficiency of these primers was evaluated *in silico* following the Primer Analysis pipeline in Primer-Prospector. Overall database matches for each candidate primer were represented as weighted scores (Fig. S1). Amongst the forward and reverse primers, 180f and 587r showed, respectively, the lowest weighted scores, indicating better coverage across the reference database and a lower number of mismatches and gaps (Fig. S1).

The taxonomic coverage for *Synechococcus* at the genus level for all the primer pairs tested, was in all cases higher than 85% (Fig. S2a). For *Prochlorococcus,* this percentage was more variable, from 54% for the 180f/662r pair to 94% for the 207f/587r pair (Fig. S2a). At the lowest taxonomic level tested (clades or subclades, Fig. S2b), the amplification of the different groups varied depending on the primer pair used, as observed for *Synechococcus* IC, V, IX or *Prochlorococcus* HLIII and LLII-III (Fig. S2b). Only the candidate pairs with 180f as a forward primer showed *in silico* amplification of all the *Prochlorococcus* groups present in the database, with the exception of 180f/662r which did not amplify LLII-III (Fig. S2b). For *Synechococcus*, the best coverage was provided by the pairs 180f/587r and 180f/605r (Fig. S2b).
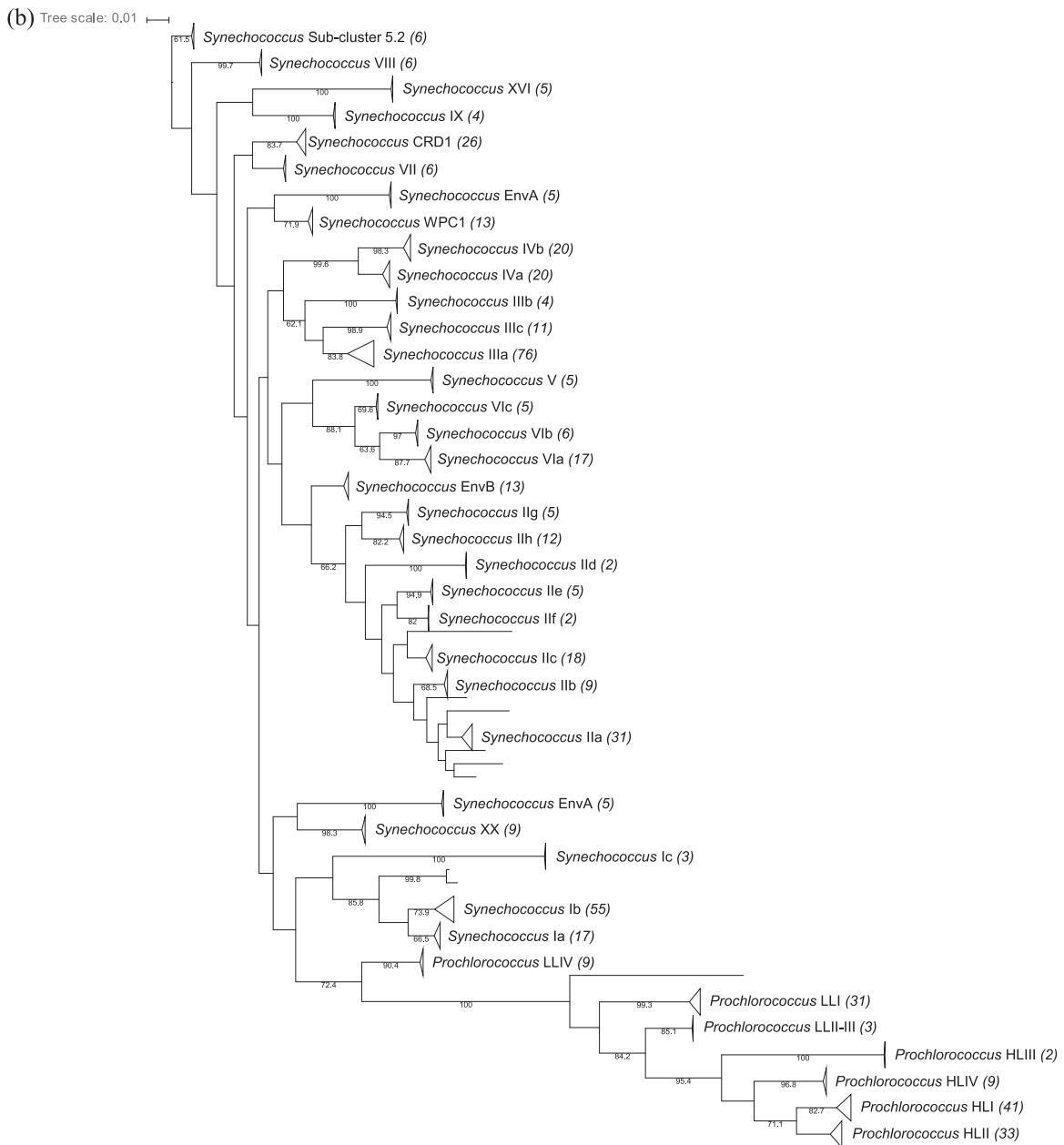
Fig. 2. Continued

Based on the results of the *in silico* tests, and considering the estimated length of the amplicons produced by each candidate primer pair, we selected 5 primer pairs to be commercially synthesized and tested using reference cultures (Table 2). These included the 180f/587r pair due to its lower weighted scores and higher taxonomic coverage, the 207f/662r and 180f/662r pairs, due to their higher amplicon size (potentially more taxonomic resolution), and the 180f/539r and 207f/587r pairs because these showed a good coverage for *Prochlorococcus*. We also used different online tools to estimate *in silico* the annealing temperatures for each primer pair and their stability in terms of homo- and heterodimer formation (Table 2). None of the primer pairs to be tested showed concerning dimerization values *in silico*.

To test the preselected primer pairs, we purchased 14 different cyanobacterial cultures from the RCC (Table 3). Five of the cultures corresponded to *Prochlorococcus* clades, including RCC408 which was classified as HL, likely HLII. Amongst the 9 *Synechococcus* cultures, 3 belonged to different subclades of *Synechococcus* II: IIa (RCC2384), IIe (RCC2455), and IIh (RCC2531) (Table 3).

We obtained higher PCR product concentrations for *Synechococcus* than for *Prochlorococcus* cultures (Fig. S3a). The primer pairs 207f/662r and 207f/587r showed the lowest amplification in general, while 180f/539r and 180f/662r led to higher amplicon con-

Table 2

List of candidate primer pairs after *in silico* selection, with their respective sequences: Theoretical annealing temperatures estimated *in silico* with the online calculator OligoCalc (http://basic.northwestern.edu/biotools/OligoCalc.html), with and without salt adjustment, GC content estimated with the same online tool, and homodimer delta and heterodimer maximum delta values estimated in silico with the OligoAnalyzer™ tool (https://eu.idtdna.com/pages/tools/oligoanalyzer). The degenerate bases included in the primer sequences correspond to: R: A or G; S: C or G; Y: C or T; W: A or T; D: A, G or T; M: A or C; K: G or T. Data estimated for the commonly used 16S primer pair has also been added to the table as a comparison.

Primer pair	Sequences	Olical (°C)		GC content (%)	Homodimer Delta G* (kcal/mol)	Heterodimer Maximum Delta G* (kcal/mol)
		Basic	Salt ajusted			
207f	GACRTCRSYWCCAARTACGT	47.7-55.9	54.3-62.5	40	-10.72	-8.135
662r	CCDCGRAGCAGYTCMACCAT	51.8-60	58.4-66.6	50	-13.41	
207f	GACRTCRSYWCCAARTACGT	47.7-55.9	54.3-62.5	50	-10.72	-11.13
587r	CARTARCCAACYTGRTCCCA	47.7-55.9	54.3-62.5	50	-9.21	
180f	CGTCTKGAAATYCARGACAT	45.6-51.8	52.3-58.4	42.5	-13.06	-7.81
662r	CCDCGRAGCAGYTCMACCAT	51.8-60	58.4-66.6	59.2	-13.41	
180f	CGTCTKGAAATYCARGACAT	45.6-51.8	52.3-58.4	42.5	-13.06	-8.38
587r	CARTARCCAACYTGRTCCCA	47.7-55.9	54.3-62.5	50	-9.21	
180f	CGTCTKGAAATYCARGACAT	45.6-51.8	52.3-58.4	42.5	-13.06	-5.02
539r	GMAACDGTKATCACDGCCAT	47.7-55.9	54.3-62.5	48.3	-20.09	
16S	*F 341-CCTACGGGNGGCWGCAG*	*54.3-56.7*	*59.8-61.8*	*73.5*	*-10.24*	*-22.22*
	R 785-GTGYCAGCMGCCGCGGTAA	*55.4-59.7*	*61.6-66.1*	*68.4*	*-16.5*	

* Gibbs free energy ΔG in primer design is the amount of energy needed for a primer to form a particular secondary structure. In general, ΔG represents the spontaneity of a reaction when held at a constant temperature and pressure. Structures with a higher ΔG (greater than 0, or positive ΔG) require an input of energy (heat) to form, so there is a lower likelihood they will form spontaneously without extra energy. Secondary structures with a lower ΔG (negative ΔG) will happen easily and spontaneously without additional energy. Very negative ΔG numbers indicate there's an affinity to form that structure and it will likely require a lot of heat to reverse the dimer back to linear form, thus more stable secondary structures (larger negative ΔG values) should be avoided.

Table 3

Reference numbers, strain names, taxonomic classification, and origin of the *Synechococcus* and *Prochlorococcus* cultures purchased from Roscoff Culture Collection. (*: Classified as HL, probably HLII).

Roscoff reference number	Strain name	Genus	Clade/ Subclade	Origin
RCC156	SS120-04/95	*Prochlorococcus*	LLI	Atlantic Ocean
RCC162	NATL2-M98	*Prochlorococcus*	LLII	Atlantic Ocean
RCC407	MIT9313	*Prochlorococcus*	LLIV	Atlantic Ocean
RCC408	REDSEA_12-2	*Prochlorococcus*	HL*	Indian Ocean
RCC3377	MIT9301	*Prochlorococcus*	HLII	Atlantic Ocean
RCC2319	MINSyn129-20m-11	*Synechococcus*	5.3	Mediterranean Sea
RCC2372	RS9905 Clonal	*Synechococcus*	IIIa	Red Sea
RCC2378	A15-127 Clonal	*Synechococcus*	WPC1	Atlantic Ocean
RCC2383	RS9909 Clonal	*Synechococcus*	VIII	Red Sea
RCC2384	RS9912 Clonal	*Synechococcus*	IIa	Red Sea
RCC2455	BOUM84-1	*Synechococcus*	IIe	Mediterranean Sea
RCC2529	RS9901 Clonal	*Synechococcus*	IX	Red Sea
RCC2531	M11.1 Clonal	*Synechococcus*	IIh	Atlantic Ocean
RCC2554	A15-60 Clonal	*Synechococcus*	VII	Atlantic Ocean

centrations only in *Synechococcus* (Fig. S3a). The primer pair 180f/587r showed the most even amplification, as indicated by a higher Shannon index (Fig. S3a).

According to these results, we selected the 180f/587r primer pair due to its better taxonomic coverage and higher amplification, and the 180f/662r pair because of the longer size of its amplicon, to perform an annealing temperature test. We used the five *Prochlorococcus* cultures and three of the *Synechococcus* cultures less amplified in the previous step (Fig. S3b) to optimize as possible the annealing performance.

The temperature tests showed that the higher PCR product concentrations for most of the cultures were found at 43.5 °C, and 41.7 °C for the 180f/587r and the 180f/662r primer pairs, respectively. The quality and concentration of the amplicons were analyzed by Tapestation, and no multi-band or non-specific amplification was observed (Fig. S3c and d).

The analysis of the mock community using the 180f/587r primer pair allowed the detection of all the groups present, with proportions varying from 0.03% (for *Prochlorococcus* LLIV) to 16.6 % (for *Prochlorococcus* HLII) (Fig. 3). The 180f/662r primer pair also detected the different groups at different proportions, except for *Prochlorococcus* LLII that was not detected (Fig. 3). This last pair produced a higher proportion of unclassified groups than the 180f/587r primer pair (Fig. 3).

As the 180f/587r primer pair showed the highest taxonomic coverage using reference cultures and reported all the *Prochlorococcus* and *Synechococcus* clades present in the mock sample, it was selected as the best candidate to be tested with natural communities.
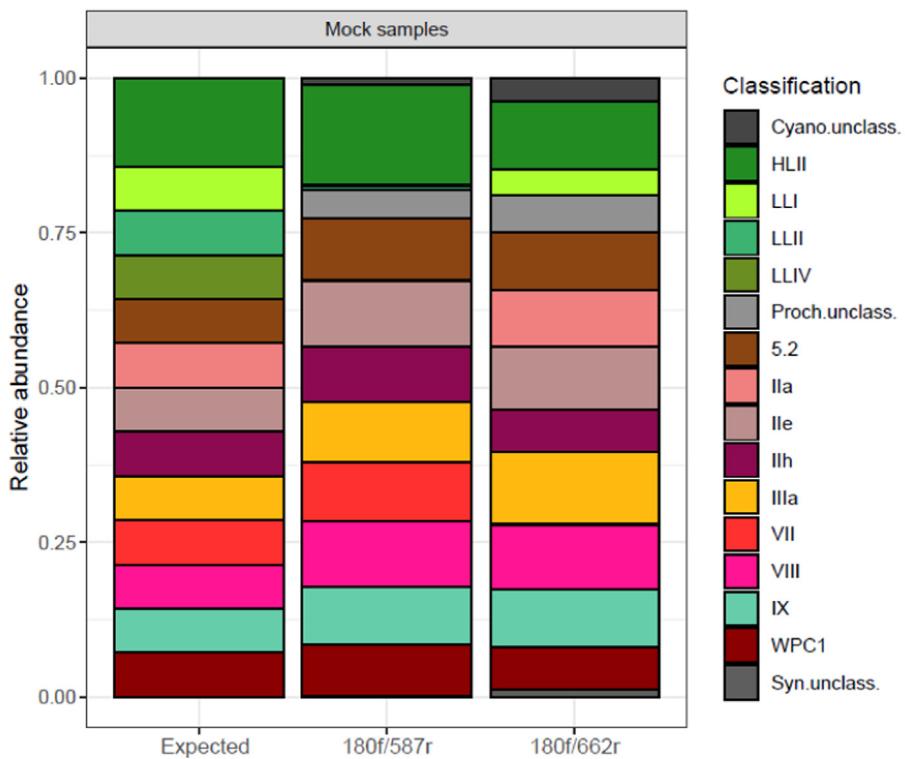
Fig. 3. Expected and actual proportions of the different RCC cultures in a mock sample, obtained with 180f/587r and 180f/662r primer pairs.
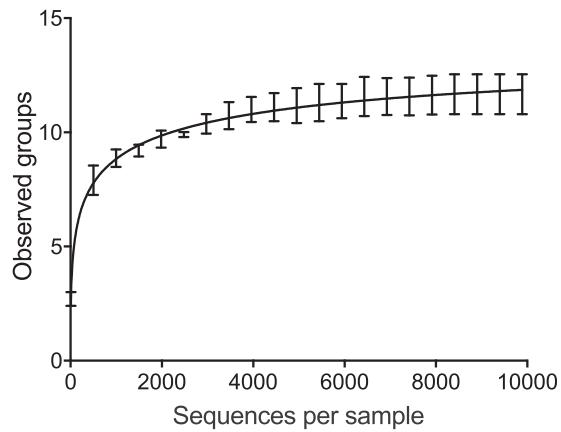

Fig. 4. Rarefaction curve for the number of observed groups (±S.E.) in the Red Sea natural samples.

The analysis of environmental samples showed a higher percentage of *Synechococcus* than *Prochlorococcus* reads in the most superficial sample, decreasing with depth (slope = -0.008, R^2=0.97). This agreed with the trend observed for the cell abundances determined with flow cytometry, where the relative abundance of *Synechococcus* decreased with depth (slope= -0.004, R^2=0.93), and *Prochlorococcus* predominated deeper in the water column.

The rarefaction curves assessing the clades or subclades (groups) detected for each sample tended towards saturation, approaching the expected maximum richness estimated by Chao1 (12 groups) (Fig. 4). The percentage of reads assigned to the different clades or subclades is represented in Fig. 5. For *Synechococcus*, a clear dominance of subclade IIa was observed in all sampling sites, followed far behind by IIc, IIb, and IIIa (Fig. 5). Amongst the classified *Prochlorococcus*, the majority of the sequences belonged to clade HLII, reaching the highest proportions in the shallower samples, followed by clade LLI with depth (Fig. 5). A high percentage of unclassified reads was found on the deepest sample. Many of these unclassified sequences were identified as *Prochlorococcus*, but could not be assigned to any of its clades (Fig. 5).
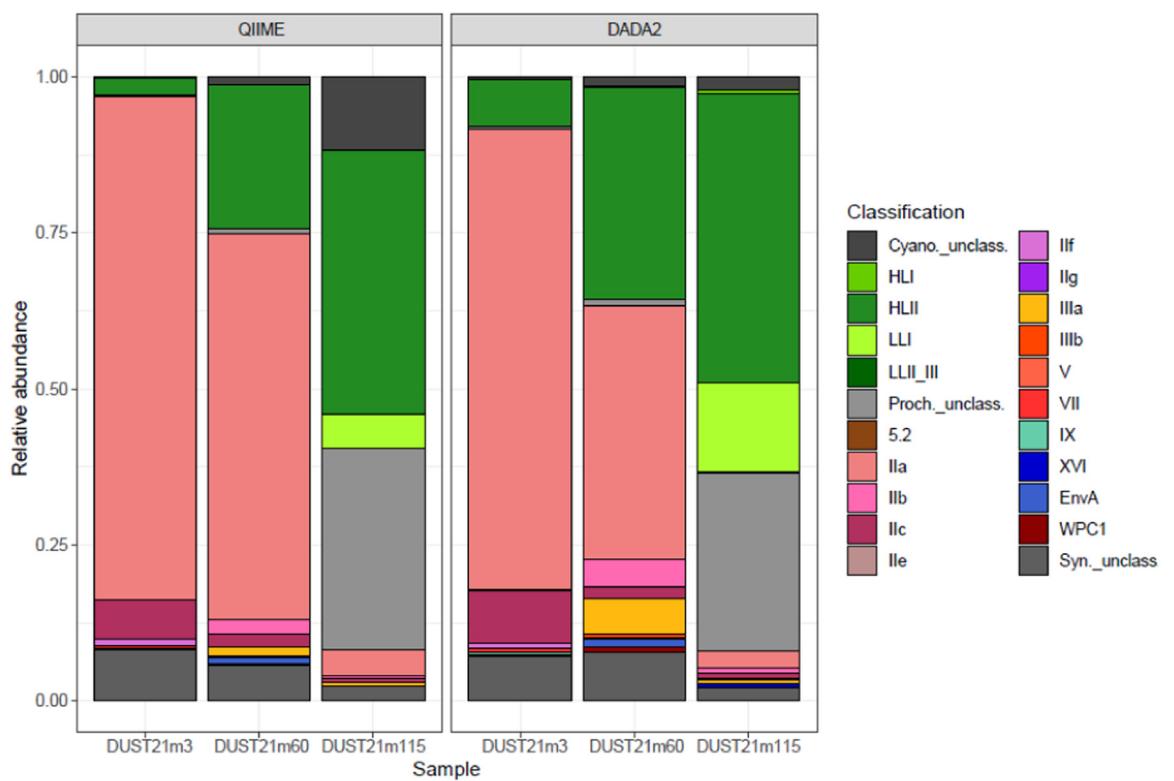
Fig. 5. Relative abundance (in percentage of reads) of the different taxa identified within *Synechococcus* and *Prochlorococcus* communities on three natural Red Sea samples using the 180f/587r primer pair, with QIIME and DADA2 pipelines.

Additional information

Our results indicated that the newly designed primer pair 180f/587r produced an amplicon suitable (367 bp) for Illumina MiSeq, and was able to amplify both *Prochlorococcus* and *Synechococcus petB* sequences, allowing a deep taxonomic identification to a subclade level.

The recent onset of high sequencing technologies has revolutionized the analysis of microbial communities. As DNA sequencing cost continues to decline, the interest in next-generation technologies for amplicon sequencing grows. Amongst them, Illumina stands out as the predominant platform for short-read sequencing providing great sequence coverage per run. It requires shorter amplicon targets than other common methods [6,10], making necessary the design of adequate, shorter primers.

These improvements in genetic analyses have significantly facilitated thorough studies of the composition and structure of different marine cyanobacteria communities [22]. The high ecological diversity of the ubiquitous cyanobacteria *Synechococcus* at the finest scale was unveiled by Mazard *et al.* (2012), using traditional Sanger sequencing and targeting the single-copy gene *petB*. Their *petB* primer set produced sequences 597 bp long and was highly specific for *Synechococcus*. Therefore, despite its good coverage it is not suitable for short-read MiSeq technology, and would only amplify *Prochlorococcus* under specific circumstances [3]. Following their pace, we contribute with the 180f/587r primer pair presented here, designed to produce shorter amplicons for MiSeq sequencing and targeting both *Prochlorococcus* and *Synechococcus* communities.

The 180f/587r primer pair defined here allowed us to cover the genotypic diversity of the *Synechococcus* and *Prochlorococcus* clades tested, as predicted *in silico*. These computer-based simulations can provide useful information on the theoretical taxonomy coverage and base mismatches produced by primer sequences, but the actual influence of sequencing processes and PCR amplification on primer performance has to be experimentally validated. The good taxonomic coverage of the 180f/587r primer pair was also observed in the mock sample, as it was able to identify all clades present in the mock community.

In addition to *in silico* tests and mock community analysis, it is important to evaluate primer pairs using natural samples to accurately evaluate their effectiveness [23]. In our Red Sea natural samples, the proportion of reads assigned to *Synechococcus* was higher towards the surface, with the opposite trend for *Prochlorococcus*. These observations mimicked the distribution of cell densities for the two groups determined with flow cytometry.

Using the 180f/587r primer pair, we observed a predominance of *Synechococcus* clade II, which usually thrives in warm tropical/ subtropical waters [22,24]. Specifically, subclade IIa was the most abundant, as previously observed in Red Sea samples [3,25]. Moreover, our new primer pair was exhaustive enough to also identify much less abundant subclades, also present in *Synechococcus* populations, such as IIc, IIb, and IIIa.

Regarding *Prochlorococcus*, clade HLII was the most abundant in our samples, in agreement with previous analyses on this sea [5,26]. Low-light adapted *Prochlorococcus* were also present, more abundant down the water column, and included LLI and LLII-III clades [26]. Our results also led to a high percentage of unidentified *Prochlorococcus*, likely because new clades are still being described [i. e. 2], and therefore their corresponding *petB* sequences are not yet available.

By applying the DADA2 pipeline we observed some differences in the detection of the least abundant groups. This pipeline allowed us to detect the presence of *Prochlorococcus* HLI and LLII-III in the shallowest and deepest samples, respectively, and *Synechococcus* IIg, IIIa, IIIb, V, and XVI, that were not described with the QIIME pipeline (Fig. 5). Other low-abundant clades as IIe and IX detected with QIIME were not detected with DADA2.

Conclusion

The newly designed 180f/587r primer pair presented here has proved to be suitable for widespread Illumina sequencing technologies and has successfully described to a low taxonomic level *Synechococcus* and *Prochlorococcus* populations, facilitating the genomic analysis of natural samples where both groups are present. Also, our results support the use of the *petB* gene as an adequate target for a deep taxonomic classification of both genera down to a clade or subclade level. However, the performance of our primer pair would substantially benefit from further sequencing efforts and a more complete *petB* database.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Alexandra Coello-Camba: Conceptualization, Resources, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Rubén Díaz-Rúa:** Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Susana Agusti:** Conceptualization, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2023.102444.

References

[1] E. Urbach, D. Robertson, S.W. Chisholm, Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation, Nature 355 (1992) 267–270, doi:10.1038/355267a0.

[2] S. Huang, S.W. Wilhelm, H.R. Harvey, K. Taylor, N. Jiao, F. Chen, Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans, ISME J. 6 (2012) 285, doi:10.1038/ismej.2011.106.

[3] S. Mazard, M. Ostrowski, F. Partensky, D.J. Scanlan, Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*, Environ. Microbiol. 14 (2012) 372–386, doi:10.1111/j.1462-2920.2011.02514.x.

[4] S.J. Biller, P.M. Berube, J.W. Berta-Thompson, L. Kelly, S.E. Roggensack, L. Awad, K.H. Roache-Johnson, H. Ding, S.J. Giovannoni, G. Rocap, L.R. Moore, S.W. Chisholm, Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*, Sci. Data 1 (2014) 140034, doi:10.1038/sdata.2014.34.

[5] N.J. Fuller, N.J. West, D. Marie, M. Yallop, T. Rivlin, A.F. Post, D.J. Scanlan, Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea, Limnol. Oceanogr. 50 (2005) 363–375, doi:10.4319/lo.2005.50.1.0363.

[6] C.W. Knetsch, E.M. van der Veer, C. Henkel, P. Taschner, E. van Pelt-Verkuil, W.B. van Leeuwen, R. te Witt, Sequencing, in: Molecular Diagnostics. part 2: Clinical, Veterinary, Agrobotanical and Food Safety Applications, Springer Nature, Singapore, 2019, pp. 339–360, doi:10.1007/978-981-13-1604-3.

[7] A. Klindworth, E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, F.O. Glöckner, Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic Acids Res. 41 (2013) e1-e1, doi:10.1093/nar/gks808.

[8] S. Penno, D. Lindell, A.F. Post, Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*, Environ. Microbiol. 8 (2006) 1200–1211, doi:10.1111/j.1462-2920.2006.01010.x.

[9] J. Pittera, F. Humily, M. Thorel, D. Grulois, L. Garczarek, C. Six, Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*, ISME J. 8 (2014) 1221, doi:10.1038/ismej.2013.228.

[10] N.A. Bokulich, D.A. Mills, Next-generation approaches to the microbial ecology of food fermentations, BMB Rep. 7 (2012) 377–389, doi:10.5483/BM-BRep.2012.45.7.148.

[11] W.A. Walters, J.G. Caporaso, C.L. Lauber, D. Berg-Lyons, N. Fierer, J.R. Knight, PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers, Bioinformatics 27 (2011) 1159–1161, doi:10.1093/bioinformatics/btr087.

[12] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, doi:10.1093/nar/gkh340.

[13] M. Kearse, R. Moir, A. Wilson. S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics 28 (2012) 1647–1649, doi:10.1093/bioinformatics/bts199.

[14] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation, Bioinformatics 23 (2007) 127–128, doi:10.1093/bioinformatics/btl529.

[15] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A. Gonzalez-Peña, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (2010) 335, doi:10.1038/nmeth.f.303.

[16] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet J. 17 (2011) 10–12, doi:10.14806/ej.17.1.200.

[17] J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: a fast and accurate Illumina Paired-End reAd merger, Bioinformatics 30 (2013) 614–620, doi:10.1093/bioinformatics/btt593.

[18] T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: a versatile open source tool for metagenomics, PeerJ 4 (2016) e2584, doi:10.7717/peerj.2584.

[19] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, Mol. Biol. Evol. 24 (2007) 1596–1599, doi:10.1093/molbev/msm092.

[20] P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn, C.F. Weber, Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, Appl. Environ. Microbiol. 75 (2009) 7537–7541, doi:10.1128/AEM.01541-09.

[21] B.J. Callahan, P.J. McMurdie, M.J. Rosen, A.W. Han, A.J. Johnson, S.P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data, Nat. Methods 13 (2016) 581–583, doi:10.1038/nmeth.3869.

[22] A.G. Kent, S.E. Baer, C. Mouginot, J.S. Huang, A.A. Larkin, M.W. Lomas, A.C. Martiny, Parallel phylogeography of *Prochlorococcus* and *Synechococcus*, ISME J. 13 (2019) 430, doi:10.1038/s41396-018-0287-6.

[23] A.E. Parada, D.M. Needham, J.A. Fuhrman, Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples, Environ. Microbiol. 18 (2016) 1403–1414, doi:10.1111/1462-2920.13023.

[24] A.F. Post, S. Penno, K. Zandbank, A. Paytan, S. Huse, D.M. Welch, Long term seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, Northern Red Sea, Front. Microbiol. 2 (2011) 131, doi:10.3389/fmicb.2011.00131.

[25] G.K. Farrant, H. Doré, F.M. Cornejo-Castillo, F. Partensky, M. Ratin, M. Ostrowski, F.D. Pitt, P. Wincker, D.J. Scanlan, D. Iudicone, S.G. Acinas, L. Garzarek, Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria, PNAS 113 (2016) E3365–E3374, doi:10.1073/pnas.1524865113.

[26] A.A. Shibl, L.R. Thompson, D.K. Ngugi, U. Stingl, Distribution and diversity of *Prochlorococcus* ecotypes in the Red Sea, FEMS Microbiol. Lett. 356 (2014) 118–126, doi:10.1111/1574-6968.12490.