Arch Microbiol (2016) 198:973–986
DOI 10.1007/s00203-016-1256-y

ORIGINAL PAPER

Proposal of fifteen new species of *Parasynechococcus* based on genomic, physiological and ecological features

F. H. Coutinho[1,2] · B. E. Dutilh[3] · C. C. Thompson[1] · F. L. Thompson[1,4,5]

Received: 19 January 2016 / Revised: 23 May 2016 / Accepted: 7 June 2016 / Published online: 23 June 2016

Abstract Members of the recently proposed genus *Parasynechococcus* (*Cyanobacteria*) are extremely abundant throughout the global ocean and contribute significantly to global primary productivity. However, the taxonomy of these organisms remains poorly characterized. The aim of this study was to propose a new taxonomic framework for *Parasynechococcus* based on a genomic taxonomy approach that incorporates genomic, physiological and ecological data. Through in silico DNA–DNA hybridization, average amino acid identity, dinucleotide signatures and phylogenetic reconstruction, a total of 15 species of *Parasynechococcus* could be delineated. Each species was then described on the basis of their gene content, light and nutrient utilization strategies, geographical distribution patterns throughout the oceans and response to environmental parameters.

Keywords Parasynechococcus · Genomic taxonomy · *Cyanobacteria* · Pan-genome

Communicated by Erko Stackebrandt.

Electronic supplementary material The online version of this article (doi:10.1007/s00203-016-1256-y) contains supplementary material, which is available to authorized users.

✉ C. C. Thompson
thompsoncristiane@gmail.com

✉ F. L. Thompson
fabianothompson1@gmail.com

1 Instituto de Biologia (IB), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

2 Centre for Molecular and Biomolecular Informatics (CMBI), Radboud Institute for Molecular Life Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands

3 Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, The Netherlands

4 COPPE, SAGE, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

5 CCS, IB, BIOMAR, Laboratório de Microbiologia, Cidade Universitária, Av. Carlos Chagas Filho. S/N, BLOCO A (Anexo) A3 - sl 102, Rio de Janeiro, RJ 21941-599, Brazil

Introduction

The *Parasynechococcus* genus has recently been described on the basis of genomic taxonomy (Coutinho et al. 2016). The new genus encompasses marine strains of *Cyanobacteria* formerly known as *Synechococcus*, closely related to the genus *Prochlorococcus*. *Parasynechococcus* is a dominant member of marine microbial communities and is among the most important contributors of global primary productivity (Flombaum et al. 2013; Coutinho et al. 2015). These organisms have developed diverse strategies for nutrient uptake (Palenik et al. 2003), production of secondary metabolites (Micallef et al. 2015) and light harvesting and utilization (Six et al. 2007a, b). Nevertheless, no species have been described within this genus; thus, the taxonomic framework of this taxon remains uncharacterized.

The very concept of bacterial species is an evolving topic. Classical polyphasic approaches based solely on morphological and biochemical traits coupled with DNA–DNA hybridization and 16S rRNA sequencing cannot fully describe the growing diversity of microorganisms. Genome-based analysis allows for the description and comparison of both cultured (Al-Saari et al. 2015; Appolinario et al. 2016) and uncultured microorganisms, from which complete genomes can now be obtained without the need for cultivation (Iverson et al. 2012; Hugerth et al. 2015; Brown et al. 2015). Furthermore, genomic analysis

has been successful in the identification of cryptic species, when organisms cannot be differentiated on the basis of their morphological or metabolic features alone (Luo et al. 2011; Hahn et al. 2016; Zangenah et al. 2016). Therefore, genomic taxonomy has become fundamental for the description of novel microbial taxa (Coenye et al. 2005; Gevers et al. 2005; Thompson et al. 2013; Kim et al. 2014; Varghese et al. 2015).

Large scale sequencing expanded the knowledge regarding the genetic diversity within *Cyanobacteria*, leading to a revision of the taxonomy of the entire phylum (Shih et al. 2013). Historically, the taxonomic framework of *Cyanobacteria* has been hampered by the lack of consensus regarding the naming of members of this taxon that oscillates between the botanical and bacteriological nomenclature codes (Gaget et al. 2015). Recent polyphasic studies have tried to combine morphology and genomic-based classifications into a unified classification system for *Cyanobacteria* (Komárek et al. 2014; Komárek 2016). However, in order to study the diversity of *Cyanobacteria* present in the environment, novel classification schemes are needed that also incorporate ecological information. In the present study, we sought to delineate new species within the genera *Parasynechococcus* on the basis of genomic taxonomy, coupled with the analysis of physiological and ecological traits of these organisms to describe fifteen species within this genus.

Methods

Comparative genomics

Complete genomes of 15 *Parasynechococcus* strains were retrieved from GenBank for analysis. These strains were chosen because they have had their pan-genome previously characterized and taxonomy redefined on the basis of genomic traits (Coutinho et al. 2016). Three metrics were applied to quantify the degree of genetic similarity between genomes: dinucleotide signature (Karlin et al. 1997); in silico DNA–DNA hybridization (Meier-Kolthoff et al. 2013); and average amino acid identity (Konstantinidis and Tiedje 2005). These analyses were carried out through GenTaxo as previously described (Coutinho et al. 2016). Phylogenomic reconstruction was performed using a set of concatenated alignments of 607 orthologous groups determined as the core genome of *Parasynechococcus*, *Prochlorococcus* and *Synechococcus* (Coutinho et al. 2016). Phylogenetic tree was inferred through neighbor joining using Tajima–Nei distances. Bootstrap test was conducted in 1000 replications.

Environmental distribution and influence of environmental parameters

To determine the environmental distribution profiles of the 15 analyzed genomes across marine habitats, raw reads from 191 microbial fraction Tara Oceans metagenomes (Sunagawa et al. 2015) were mapped to a database of *Parasynechococcus* genomes through Bowtie2 (Langmead and Salzberg 2012), setting the—very-sensitive-local and—options. Counts of ambiguous reads (that can be mapped to more than one region or genome) were corrected as described in (Iverson et al. 2012). Spearman correlation scores were calculated between the abundance of each genome across samples and measured environmental parameters. Tara Oceans metagenomes were compared on the basis of their relative abundance of *Parasynechococcus* genomes through non-metric multidimensional scaling (NMDS).

Results and discussion

Parasynechococcus genomes exhibited levels of AAI below 95 %, DDH below 70 % and dinucleotide signature distances above 0.05 (Fig. 1; Table S1). According to these criteria, each genome represented a new species within the genus, as evidenced the levels of genomic similarity drastically below the species cutoff established for the aforementioned metrics. These species formed a monophyletic group that had *Prochlorococcus* as a sister clade (Fig. 2) (Coutinho et al. 2016). Over the next sections, the 15 proposed species are described according to their genomic, physiological and ecological features.

Proposal of a new taxonomic structure and species definition for *Parasynechococcus*

The newly identified species were named according to their geographical distribution profile (Figure S1) as: *Parasynechococcus africanus* CC9605^T, *Parasynechococcus antarcticus* WH5701^T, *Parasynechococcus benguelii CC9311*, *Parasynechococcus chillensis CC9902*, *Parasynechococcus equatorialis* RS9917^T, *Parasynechococcus gyrus* RS9916^T, *Parasynechococcus indicus* CB0205^T, *Parasynechococcus marearabicus* WH8109^T, *Parasynechococcus marenigrum* WH8102^T, *Parasynechococcus mediterranei* RCC307^T, *Parasynechococcus nordiatlanticus* BL107^T, *Parasynechococcus pacificus* WH7803^T, *Parasynechococcus subtropicalis* WH7805^T, *Parasynechococcus sudiatlanticus* CB0101^T, *Parasynechococcus sudipacificus* WH8016^T.

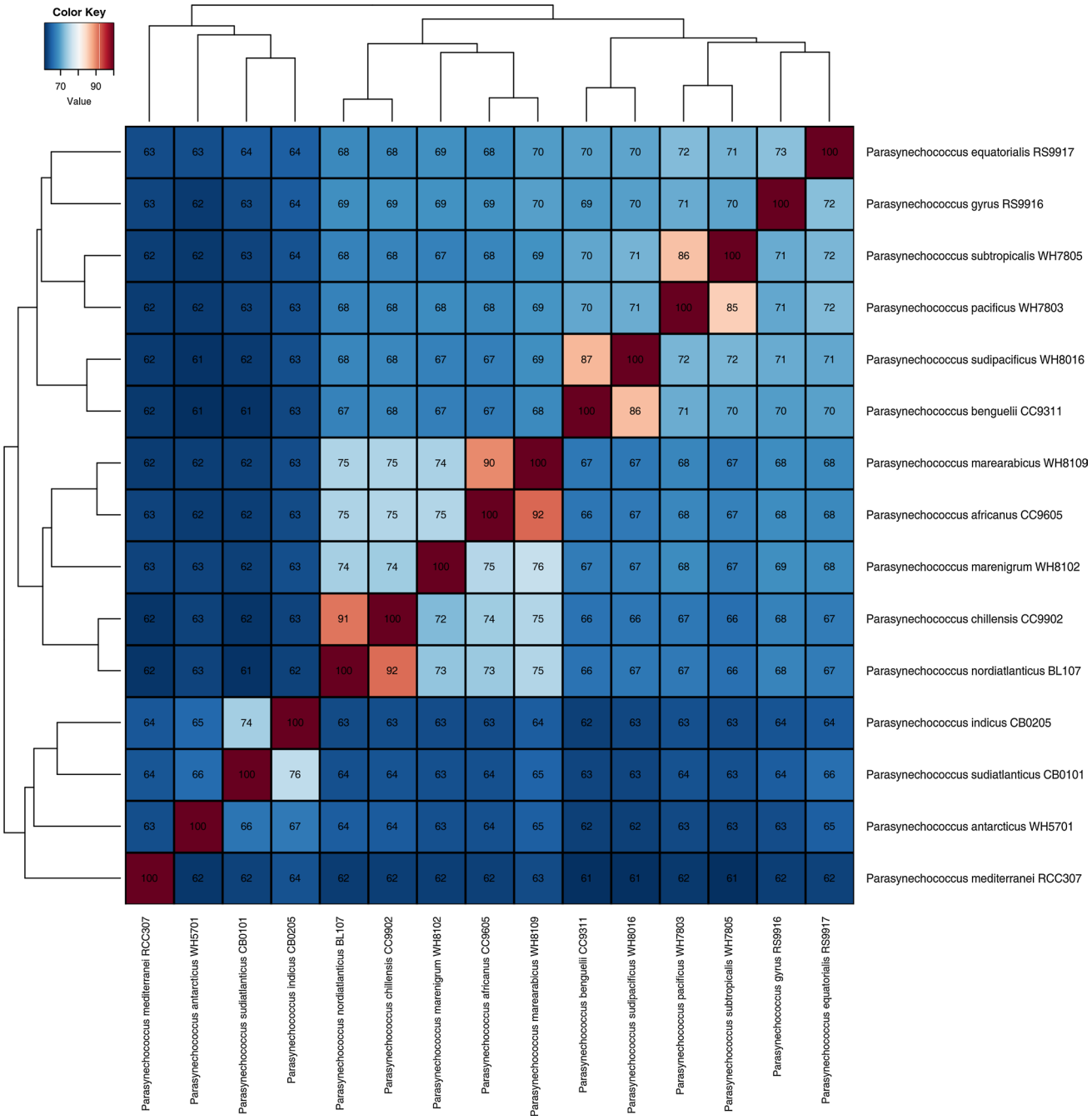

Fig. 1 Heatmap displaying the average amino acid identity levels between *Parasynechococcus* genomes. Hierarchical clustering was performed based on Manhattan distances

Genomic features

Analysis of the pan-genome profile of *Parasynechococcus* revealed 3714 orthologous groups (OGs) that were shared by two or more of the 15 strains (Table S1) (Coutinho et al. 2016). The proposed species can be discerned according to the presence/absence profile or number of copies of these OGs. Furthermore, each species harbored a set of 241–680 exclusive OGs (Table S2), not detected in any other of the analyzed genomes, that add up to a total of 6523 orthologous groups. The majority of those encode hypothetical proteins of unknown function, nevertheless, those that do have an annotated function are genes involved in diverse physiological processes, such as nutrient uptake and utilization, carbohydrate metabolism and photosynthesis. Thus, the pan-genome profile of *Parasynechococcus* represents

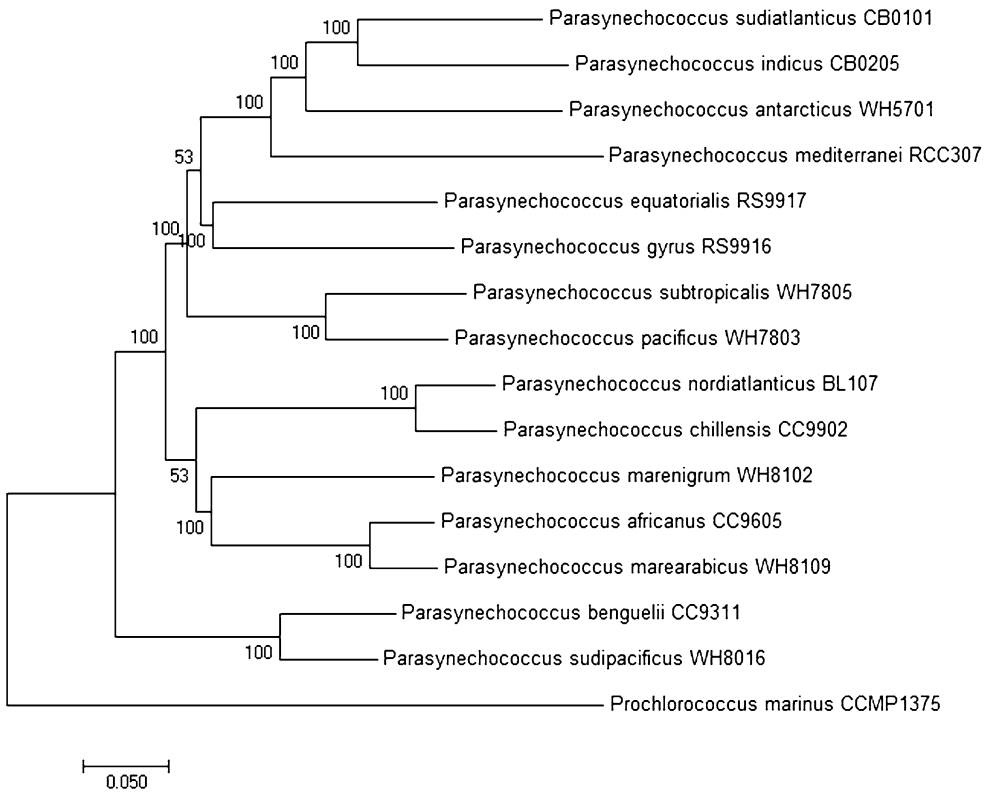

Fig. 2 Neighbor-joining phylogenetic tree of 15 *Parasynechococcus* species based on the concatenated alignments of 607 orthologous groups. *Prochlorococcus marinus* CCMP1375 was used as the root

an extensive catalog of more than ten thousand genetic features that can be used to discern between the fifteen proposed species.

Diagnostic physiological features

To our knowledge, no study has characterized the physiological features of all the analyzed strains through a consistent methodology. Therefore, most of the information regarding the physiology of these organisms remains patchy, usually focused on just a subset of these strains. One study compared strains of *Parasynechococcus* regarding their phycobilisome structure and composition (Six et al. 2007b). *Cyanobacteria* make use of these structures to harvest light, and phycobilisome composition determines the range of wavelengths that can be utilized by these organisms and gives their cells their characteristic coloring. Three major types of phycobilisome pigmentation were identified among those species (Table 1). Another important physiological feature previously investigated in *Parasynechococcus* strains was their ability to produce bacteriocins (Wang et al. 2011). Gene clusters encoding these secondary metabolites that have antimicrobial activities have been shown to be diverse and widespread in *Cyanobacteria* (Micallef et al. 2015). These gene clusters are present in the proposed species in different combinations (Table 1). Those physiological features can also be used to discern between species.

Ecological features

Analyzing the abundance of *Parasynechococcus* genomes across 191 TARA Oceans metagenomes allowed us to determine the geographical distribution patterns of the 15 strains across the global ocean (Figure S2). Coupling this information with the available metadata revealed how the abundance of each genome correlated with changes in the physical, chemical and biological parameters measured at the sampling sites (Fig. 3). Next, genomes were clustered based on their correlation profiles, revealing three major groups of organisms (Table 1): Members of Ecological cluster A were *Parasynechococcus sudiatlanticus* CB0101, *Parasynechococcus indicus* CB0205 and *Parasynechococcus antarcticus* WH5701. This cluster was characterized by positive correlation values to minimum generation time, SI, $NO_2 + NO_3$, PO_4, nitrate and depth; and negative correlations with longitude, salinity, latitude, NO_2, oxygen, temperature, abundance of autotrophs, bacteria, heterotrophs, and picoeukaryotes. Cluster B was formed by strains *Parasynechococcus*

Table 1 Species description of *Parasynechococcus* strains based on genetic, phenotypic and ecological traits

Genus	Species	Strain	Source	Genome size (Mbp)	# Genes	%GC	Layer preference	Eco-logical cluster	Phycobiliosome pigementation type	Bacteriocin gene clusters	# Marker ogs
Parasynechococcus	nordiatlan-ticus	BL107	Blanes Bay, Mediterra-nean Sea, 1800 m	2.29	2655	54.2	DCM > SRF > MIX > MES	B	3d	I, U	308
Parasynechococcus	sudiatlanticus	CB0101	Chesapeake Bay	2.69	2881	64.2	MES > DCM > SRF > MIX	A	ND	ND	388
Parasynechococcus	indicus	CB0205	Chesapeake Bay	2.43	2661	63	MES > DCM > SRF > MIX	A	ND	ND	322
Parasynechococcus	benguelii	CC9311	California current, Pacific (coastal), 95 m	2.61	3164	52.4	DCM > SRF > MIX > MES	B	3d	I, U	680
Parasynechococcus	africanus	CC9605	California current, Pacific (oligotrophic), 51 m	2.51	3016	59.2	SRF > DCM > MIX > MES	B	3c	I, V	468
Parasynechococcus	chillensis	CC9902	California current, Pacific (oligotrophic), 5 m	2.23	2635	54.2	DCM > SRF > MES > MIX	B	3d	I	310
Parasynechococcus	mediterranei	RCC307	Mediterranean Sea, 15 m	2.22	3064	60.8	DCM > SRF > MES > MIX	C	3b	I	420
Parasynechococcus	gyrus	RS9916	Gulf of Aqaba, Red Sea, 10 m	2.66	2927	59.8	SRF > DCM > MIX > MES	C	3d	I, IV, V	484
Parasynechococcus	equatorialis	RS9917	Gulf of Aqaba, Red Sea, 10 m	2.58	2719	64.4	SRF > DCM > MES > MIX	C	1	I	268
Parasynechococcus	antarcticus	WH5701	Long Island Sound, USA	3.28	3185	65.4	MES > SRF > DCM > MIX	A	1	I	593
Parasynechococcus	pacificus	WH7803	Sargasso Sea, 25 m	2.37	2660	60.2	DCM > SRF > MES > MIX	C	3a	ND	241
Parasynechococcus	subtropicalis	WH7805	Sargasso Sea	2.63	2931	57.6	SRF > DCM > MIX > MES	B	2	I, U	452
Parasynechococcus	sudipacificus	WH8016	Woods Hole, MA, USA	2.69	3184	54.1	SRF > DCM > MIX > MES	B	ND	ND	625
Parasynechococcus	marenigrum	WH8102	Sargasso Sea	2.43	2752	59.4	SRF > DCM > MES > MIX	B	3c	I, U	365
Parasynechococcus	marearabicus	WH8109	Sargasso Sea	2.12	2510	60.1	SRF > DCM > MIX > MES	B	ND	ND	244

nordiatlanticus BL107, *Parasynechococcus benguelii* CC9311, *Parasynechococcus africanus* CC9605, *Parasynechococcus chillensis* CC9902, *Parasynechococcus subtropicalis* WH7805, *Parasynechococcus sudipacificus* WH8016, *Parasynechococcus marenigrum* WH8102, *Parasynechococcus marearabicus* WH8109. This cluster, exhibited the opposite trend observed for cluster A, was characterized by negative correlations to minimum generation time, SI, $NO_2 + NO_3$, PO_4, nitrate and depth and positive correlations with longitude, salinity, latitude, NO_2, oxygen, temperature abundance of autotrophs, bacteria, heterotrophs and picoeukaryotes. Cluster C was made of *Parasynechococcus mediterranei* RCC307, *Parasynechococcus gyrus* RS9916, *Parasynechococcus equatorialis* RS9917 and *Parasynechococcus pacificus* WH7803. This cluster was characterized by negative correlations to minimum generation time, SI, $NO_2 + NO_3$, PO_4, nitrate and depth, weak correlations with longitude, salinity, latitude, NO_2 and positive correlations with oxygen, temperature and abundance of autotrophs, bacteria, heterotrophs and picoeukaryotes, but weaker than those observed for members of cluster B. Measuring the median abundance of each genome in Tara Oceans samples across each depth layer revealed that members of the proposed ecological clusters showed consistent patterns regarding their layer preferences (Table 1). Members of cluster A were more abundant at mesophotic layer, while members of clusters B and C had a preference for the surface and DCM layers.

Species composition of *Parasynechococcus* communities in the analyzed marine metagenomes was compared through non-metric multidimensional scaling (Fig. 4). The positioning of samples across PC1 and PC2 was correlated with measurements of environmental parameters available for each site. Significant correlations were detected between levels of P and N sources and positioning of samples across PC1 and PC2 (Table S4), suggesting the importance of nutrient abundance in determining *Parasynechococcus* community composition throughout the oceans. Therefore, the pan-genome of *Parasynechococcus* was scanned for differences in OGs involved with nutrient uptake and utilization that could help explain the differences observed in the environmental distribution profile of those strains. The occurrence of OGs associated with the ability of these organisms to harvest and utilize different sources of nitrogen, phosphorus and iron is described in Table 2. Each genome harbored a diverse and unique array of features to uptake and metabolize these nutrients. The combination of these features shed light into the partitioning of these resources among members of this taxon and represent an additional set of features to discern between the proposed species.

Species description

Gathering the information of the 15 *Parasynechococcus* species allowed us to define each of them on the basis of genomic, physiological and ecological traits. These features (summarized by Tables 1, 2, S2, S3) represent thousands of traits that can be used to distinguish between the new species, as well as affiliating new strains to the proposed taxa. Pan-genome composition of *Parasynechococcus* revealed diverse combinations of genes associated with nutrient uptake and utilization within this group. These genes influence the capacity of these organisms to occupy different ecological niches across marine habitats. Therefore, we propose the use of these features for the genome-based taxonomy of *Parasynechococcus* species.

Description of *Parasynechococcus chilensis* (chi.len.sis. NL. Referring to the increased abundance of this organism at the Chile–Peru coastal province): The phycobilisome pigmentation of this strain belongs to class 3d. Bacteriocin gene clusters detected in this genome belong to class I. Type strain is CC9902^T, which has a genome of 2.23 Mbp with a GC content of 54.2 % encoding 2635 genes that include 310 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus chillensis* CC9902 is determined by the presence of the following genes: Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN*, *NifS*, *NifU*, Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC*, *UreD*, *UreE*, *UreF*, *UreG*, *UrtA*, *UrtB*, *UrtC*, *UrtD*, *UrtE*, *phnE*, *phoH*, *pstA*, *pstB*, *pstC*, *pstS*.

Description of *Parasynechococcus pacificus* (pa.ci.fi.cus L. *masc. adj.* Referring to the increased abundance of this organism at the Pacific Ocean provinces): The phycobilisome pigmentation of this strain belongs to class 3a. Bacteriocin gene clusters have not been investigated in this genome. Type strain is WH7803^T, which has a genome of 2.37 Mbp with a GC content of 60.2 % encoding 2660 genes that include 241 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus pacificus* WH7803 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN, NifS*, *NifU*, Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *UrtA*, *phnE*, *phoB*, *phoH*, *phoR*, *pstA*, *pstB*, *pstC*, *pstS*.

Description of *Parasynechococcus indicus* (in.di.cus NL. Referring to the increased abundance of this organism at the Indian Ocean provinces): The phycobilisome pigmentation of this strain has not been characterized.

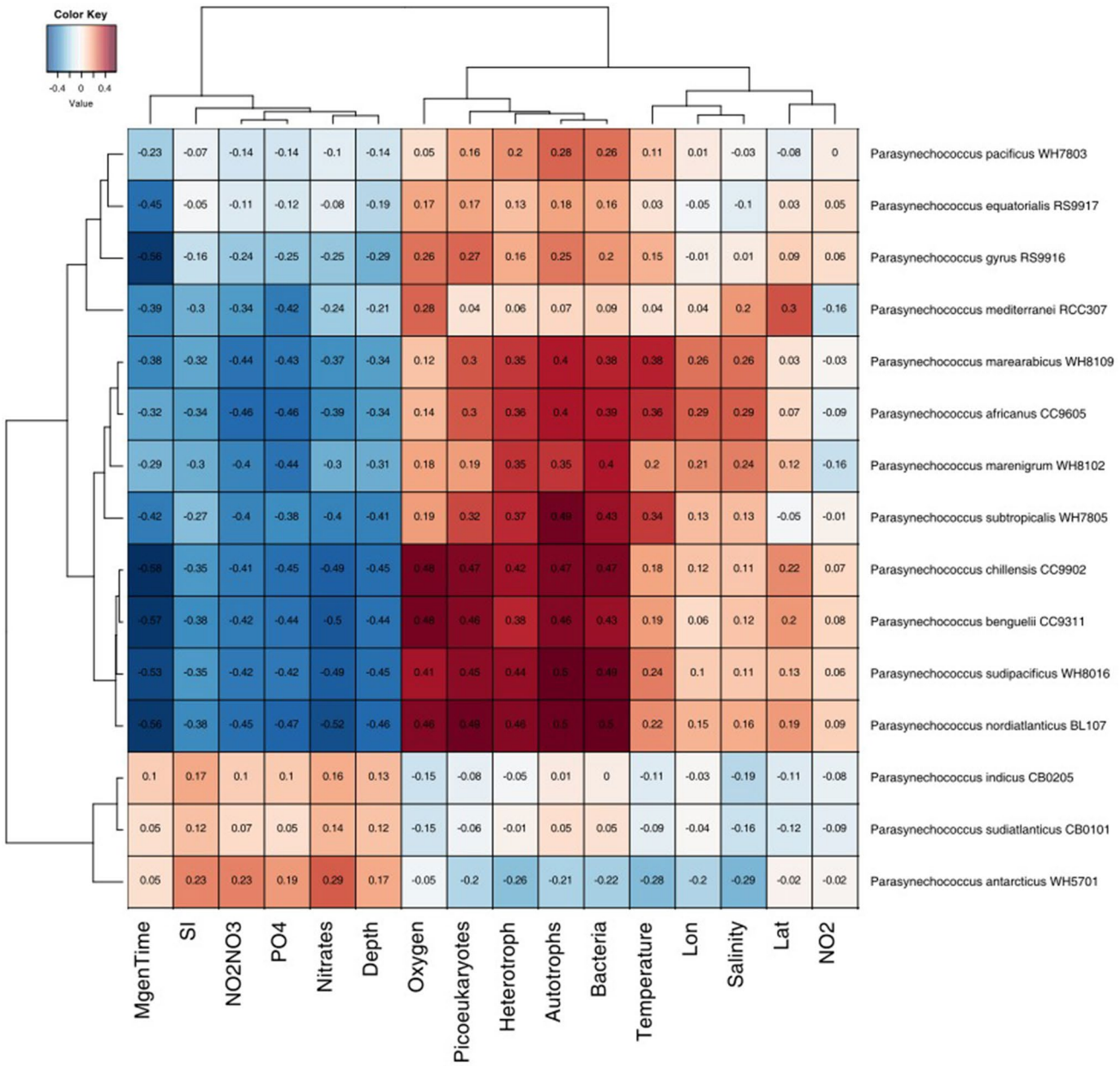

Fig. 3 Heatmap displaying Spearman correlation scores between the abundance of *Parasynechococcus* species and the measured environmental parameters at Tara Oceans metagenomes. Hierarchical clustering was performed based on Manhattan distances

Bacteriocin gene clusters have not been investigated in this genome. Type strain is CB0205^{T}, which has a genome of 2.43 Mbp with a GC content of 63 % encoding 2661 genes that include 322 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus indicus* CB0205 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN*, *NifS*, *NifU*, Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, Protein PII, *UreD*, *UreE*, *UreF*, *UreG*, *UrtA*, *UrtB*, *UrtC*, *UrtD*, *UrtE*, *phnE*, *phoB*, *phoH*, *phoR*, *pstA*, *pstB*, *pstC*, *pstS*.

Description of *Parasynechococcus mediterranei* (me.di.ter.ra.nei L. *masc. adj.* Referring to the increased abundance of this organism at the Mediterranean province): The phycobilisome pigmentation of this strain belongs to class 3b. Bacteriocin gene clusters detected in this genome belong to class I. Type strain is RCC307^{T}, which has a genome of 2.22 Mbp with a GC content of 60.8 % encoding 3064 genes that include 420 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus mediterranei*

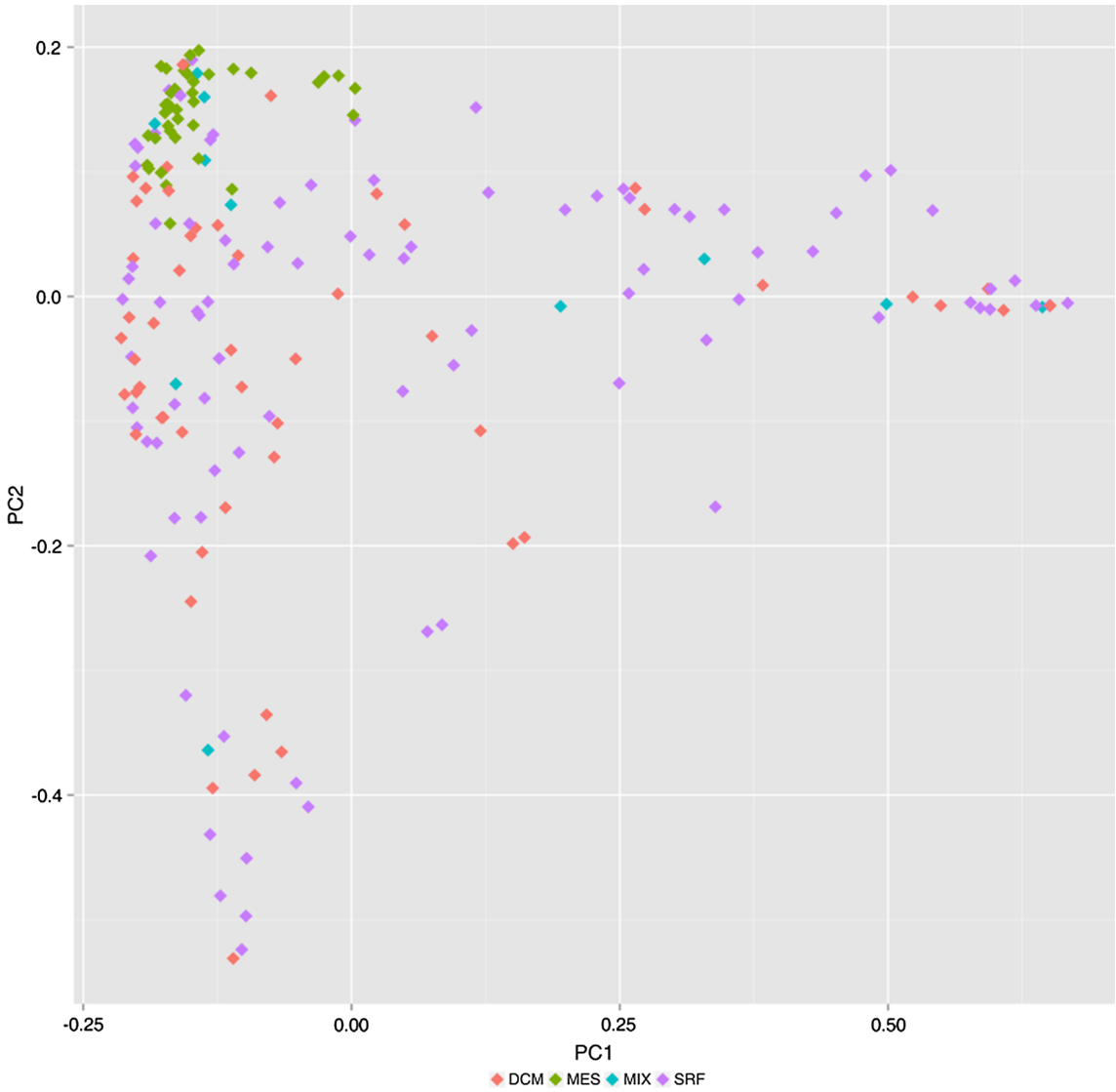

Fig. 4 NMDS of Tara Oceans metagenomes based on relative abundances of *Parasynechococcus* strains. Samples are *color coded* according to the oceanic province to which they belong (color figure online)

RCC307 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN*, *NifS*, *NifU*, Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, Protein PII, *UreD*, *UreE*, *UreF*, *UreG*, *UrtA*, *UrtB*, *UrtC*, *UrtD*, *UrtE*, *phnE*, *phoB*, *phoH*, *phoR*, *phoU*, *pstA*, *pstB*, *pstC*, *pstS*.

Description of *Parasynechococcus equatorialis* (e.qua.tor.ia.lis L. *masc. adj.* Referring to the increased abundance of this organism at the Equatorial province): The phycobilisome pigmentation of this strain belongs to class 1. Bacteriocin gene clusters detected in this genome belong to class I. Type strain is RS9917^{T}, which has a genome of 2.58 Mbp with a GC content of 64.4 % that encodes 2719 genes that include 268 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus equatorialis* RS9917 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, Ferrous iron transport protein B, *GlnN*, *NifS*, *NifU*, Nitrite transporter formate/nitrite family, *UreD*, *UreE*, *UreF*, *UreG*, *UrtA*, *UrtB*, *UrtC*, *UrtD*, *UrtE*, *phnE*, *phoB*, *phoH*, *phoR*, *pstA*, *pstB*, *pstC*, *pstS*.

Description of *Parasynechococcus sudiatlanticus* (su.di.a.tlan.ti.cus NL. Referring to the increased abundance of this organism at the South Atlantic province): The phycobilisome pigmentation of this strain has not been characterized. Bacteriocin gene clusters have not been investigated in this genome. Type strain is CB0101^{T}, which has a genome of 2.69 Mbp with a GC content of 64.2 % that

Table 2 Genes involved in nitrogen, phosphorus and iron uptake and metabolism across *Parasynechococcus* genomes

	Parasynechococcus nordiatlanticus BL107	Parasynechococcus sudiatlanticus CB0101	Parasynechococcus indicus CB0205	Parasynechococcus benguelii CC9311	Parasynechococcus africanus CC9605	Parasynechococcus chillensis CC9902	Parasynechococcus mediterranei RCC307	Parasynechococcus gyrus RS9916	Parasynechococcus equatorialis RS9917	Parasynechococcus antarcticus WH5701	Parasynechococcus pacificus WH7803	Parasynechococcus subtropicalis WH7805	Parasynechococcus sudipacificus. WH8016	Parasynechococcus marenigrum WH8102	Parasynechococcus marearabicus WH8109
Nitrogen															
Ammonium transporter	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GlnN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NifB	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
NifE	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
NifH	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
NifO	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
NifS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NifT	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
NifU	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NifW	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
NifX	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
NifZ	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Nitrate ABC transporter	−	−	−	−	−	−	−	−	−	+	−	−	−	+	−
Nitrate/Nitrite transporter	+	+	+	+	+	+	+	+	−	−	+	+	+	+	+
Nitrate/Sulfonate/Bicarbonate transport system	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Nitrite transporter formate/nitrite family	+	+	+	+	+	+	+	+	+	−	+	+	+	−	+

Table 2 continued

	Parasynechococcus nordiatlanticus BL107	Parasynechococcus sudiatlanticus CB0101	Parasynechococcus indicus CB0205	Parasynechococcus benguelii CC9311	Parasynechococcus africanus CC9605	Parasynechococcus chillensis CC9902	Parasynechococcus mediterranei RCC307	Parasynechococcus gyrus RS9916	Parasynechococcus equatorialis RS9917	Parasynechococcus antarcticus WH5701	Parasynechococcus pacificus WH7803	Parasynechococcus subtropicalis WH7805	Parasynechococcus sudipacificus. WH8016	Parasynechococcus marenigrum WH8102	Parasynechococcus marearabicus WH8109
Protein PII	−	+	+	−	−	−	+	−	−	+	−	−	−	−	−
UreD	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
UreE	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
UreF	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
UreG	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
UrtA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UrtB	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
UrtC	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
UrtD	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
UrtE	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+
Phosphorus															
phnE	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
phnF	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
phnG	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
phnH	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
phnI	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
phnJ	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
phnK	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
phnL	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
phoB	+	+	+	−	+	−	+	+	+	+	+	+	+	+	+
phoH	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
phoR	−	+	+	−	+	−	+	+	+	+	+	+	−	+	+
phoU	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−
pstA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pstB	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pstC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pstS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Iron															
EfeB	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
EfeO	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−

Table 2 continued

	Parasynechococcus nordiatlanticus BL107	Parasynechococcus sudiatlanticus CB0101	Parasynechococcus indicus CB0205	Parasynechococcus benguelii CC9311	Parasynechococcus africanus CC9605	Parasynechococcus chillensis CC9902	Parasynechococcus mediterranei RCC307	Parasynechococcus gyrus RS9916	Parasynechococcus equatorialis RS9917	Parasynechococcus antarcticus WH5701	Parasynechococcus pacificus WH7803	Parasynechococcus subtropicalis WH7805	Parasynechococcus sudipacificus. WH8016	Parasynechococcus marenigrum WH8102	Parasynechococcus marearabicus WH8109
Ferric iron ABC transporter, ATP-binding protein	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+
Ferric iron ABC transporter, iron-binding protein	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ferric iron ABC transporter, permease protein	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ferrous iron transport protein A	−	+	−	+	−	−	−	−	−	−	−	−	−	−	−
Ferrous iron transport protein B	−	+	−	+	−	−	−	−	+	+	−	−	−	−	−
PiuB	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
PiuC	+	−	−	+	+	+	−	+	−	+	−	+	+	−	+

encodes 2881 genes that include 388 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus sudiatlanticus* CB0101 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, Ferrous iron transport protein A, Ferrous iron transport protein B, *GlnN*, *NifS*, *NifU*, Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, Protein PII, *UreD*, *UreE*, *UreF*, *UreG*, *UrtA*, *UrtB*, *UrtC*, *UrtD*, *UrtE*, *phnE*, *phoB*, *phoH*, *phoR*, *pstA*, *pstB*, *pstC, pstS.*

Description of *Parasynechococcus subtropicalis* (sub. tro.pi.ca.lis NL. Referring to the increased abundance of this organism at Subtropical provinces): The phycobilisome pigmentation of this strain belongs to class 2. Bacteriocin gene clusters detected in this genome belong to classes I, U. Type strain is WH7805^T, which has a genome of 2.63 Mbp with a GC content of 57.6 % that encodes 2931 genes that include 452 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus subtropicalis* WH7805 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN*, *NifS*, *NifU*, Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC*, *UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, phoR, pstA, pstB, pstC, pstS.*

Description of *Parasynechococcus marenigrum* (su. di.a.tlan.ti.cus NL. Referring to the increased abundance of this organism at the Black Sea province): The phycobilisome pigmentation of this strain belongs to class 3c. Bacteriocin gene clusters detected in this genome belong to classes I, U. Type strain is WH8102^T, which has a genome of 2.43 Mbp with a GC content of 59.4 % that encodes 2752 genes that include 365 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus marenigrum* WH8102 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN, NifS, NifU,* Nitrate ABC transporter, Nitrate/Nitrite transporter, *UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, phoR, pstA, pstB, pstC, pstS*.

Description of *Parasynechococcus marearabicus* (ma. re.a.ra.bi.cus NL. Referring to the increased abundance of this organism at the Arabian Sea province): The phycobilisome pigmentation of this strain has not been characterized. Bacteriocin gene clusters have not been investigated in this genome. Type strain is WH8109^T, which has a genome of 2.12 Mbp with a GC content of 60.1 % that encodes 2510 genes that include 244 diagnostic orthologous groups. Taxonomic affiliation to Parasynechococcus marearabicus WH8109 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN, NifS, NifU,* Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC, UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, phoR, pstA, pstB, pstC, pstS.*

Description of Parasynechococcus gyrus (gyr.us NL. Referring to the increased abundance of this organism at the Gyral provinces): The phycobilisome pigmentation of this strain belongs to class 3d. Bacteriocin gene clusters detected in this genome belong to classes I, IV, V. Type strain is RS9916^T, which has a genome of 2.66 Mbp with a GC content of 59.8 % that encodes 2927 genes that include 484 diagnostic orthologous groups. Taxonomic affiliation to Parasynechococcus gyrus RS9916 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN, NifS, NifU,* Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC, UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, phoR, pstA, pstB, pstC, pstS.*

Description of Parasynechococcus benguelii (ben.gue.lii NL. Referring to the increased abundance of this organism at the Benguela current coastal province): The phycobilisome pigmentation of this strain belongs to class 3d. Bacteriocin gene clusters detected in this genome belong to classes I, U. Type strain is CC9311^T, which has a genome of 2.61 Mbp with a GC content of 52.4 % that encodes 3164 genes that include 680 diagnostic orthologous groups. Taxonomic affiliation to Parasynechococcus benguelii CC9311 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, Ferrous iron transport protein A, Ferrous iron transport protein B, *GlnN, NifS, NifU,* Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC, UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoH, pstA, pstB, pstC, pstS.*

Description of *Parasynechococcus nordiatlanticus* (nor. di.a.tlan.ti.cus NL. Referring to the increased abundance of this organism at North Atlantic provinces): The phycobilisome pigmentation of this strain belongs to class 3d. Bacteriocin gene clusters detected in this genome belong to classes I, U. Type strain is BL107^T, which has a genome of 2.29 Mbp with a GC content of 54.2 % that encodes 2655 genes that include 308 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus nordiatlanticus* BL107 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein,

Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN, NifS, NifU,* Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC, UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, pstA, pstB, pstC, pstS.*

Description of *Parasynechococcus antarcticus* (ant.arc.ti.cus L. *masc. adj.* Referring to the increased abundance of this organism at the Antarctic province): The phycobilisome pigmentation of this strain belongs to class 1. Bacteriocin gene clusters detected in this genome belong to class I. Type strain is WH5701^T, which has a genome of 3.28 Mbp with a GC content of 65.4 % that encodes 3185 genes that include 593 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus antarcticus* WH5701 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, Ferrous iron transport protein B, *GlnN, NifS, NifU,* Nitrate ABC transporter, PiuC, Protein PII, *UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, phoR, pstA, pstB, pstC, pstS.*

Description of *Parasynechococcus sudipacificus* (su.di.pa.ci.fi.cus NL. Referring to the increased abundance of this organism at South Pacific provinces): The phycobilisome pigmentation of this strain has not been characterized. Bacteriocin gene clusters have not been investigated in this genome. Type strain is WH8016^T, which has a genome of 2.69 Mbp with a GC content of 54.1 % that encodes 3184 genes that include 625 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus sudipacificus* WH8016 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN, NifS, NifU,* Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC, UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, pstA, pstB, pstC, pstS.*

Description of *Parasynechococcus africanus* (af.ri.ca.nus L. *masc. adj.* Referring to the increased abundance of this organism at the Eastern Africa Coastal province): The phycobilisome pigmentation of this strain belongs to class 3c. Bacteriocin gene clusters detected in this genome belong to classes I, V. Type strain is CC9605^T, which has a genome of 2.51 Mbp with a GC content of 59.2 % that encodes 3016 genes that include 468 diagnostic orthologous groups. Taxonomic affiliation to *Parasynechococcus africanus* CC9605 can be defined by the presence of Ammonium transporter, Ferric iron ABC transporter: ATP-binding protein, Ferric iron ABC transporter: iron-binding protein, Ferric iron ABC transporter: permease protein, *GlnN, NifS, NifU,* Nitrate/Nitrite transporter, Nitrite transporter formate/nitrite family, *PiuC, UreD, UreE, UreF, UreG, UrtA, UrtB, UrtC, UrtD, UrtE, phnE, phoB, phoH, phoR, pstA, pstB, pstC, pstS.*

Conclusions

Genomics has emerged as a powerful tool to characterize and compare microorganisms. We suggest that future affiliation of new species or strains to *Parasynechococcus* should be made on the basis of a complete genome analysis, as genetic features are less subjected to the biases and inconsistencies yielded by morphological or metabolic traits that have hampered microbial taxonomy for so long. The present study makes use of previously published datasets to assess the species diversity within *Parasynechococcus*. Furthermore, we described their distribution patterns throughout the global oceans and how the abundance of each strain is influenced by environmental parameters. Thus, a new taxonomic framework is provided for the genera *Parasynechococcus*, through a polyphasic approach that incorporates information on features that range from the genetic to the ecological level.

Acknowledgments The authors thank CNPq, CAPES and FAPERJ for support.

References

Al-Saari N, Gao F, Rohul AAKM, Sato K, Sato K, Mino S, Suda W, Oshima K, Hattori M, Ohkuma M, Meirelles PM, Thompson FL, Thompson C, Filho GMA, Gomez-Gil B, Sawabe T, Sawabe T (2015) Advanced microbial taxonomy combined with genome-based-approaches reveals that Vibrio astriarenae sp. Nov., an agarolytic marine bacterium, forms a new clade in Vibrionaceae. PLoS One 10:1–17

Appolinario LR, Tschoeke D, Venas T, Campeão ME, Amaral G, Leomil L, Oliveira L, Otsuki RK, Swings J, Thompson FL, Thompson CC (2016) Description of Endozoicomonas arenosclerae sp nov using a genomic taxonomy approach. Antonie van Leeuwenhoek J Microb 109(3):431–438

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF (2015) Unusual biology across a group comprising more than 15 % of domain Bacteria. Nature 523:208–211

Coenye T, Gevers D, Van de Peer Y, Vandamme P, Swings J (2005) Towards a prokaryotic genomic taxonomy. FEMS Microbiol Rev 29:147–167

Coutinho FH, Meirelles PM, Moreira APB, Paranhos RP, Dutilh BE, Thompson FL (2015) Niche distribution and influence of environmental parameters in marine microbial communities: a systematic review. PeerJ 3:e1008

Coutinho F, Tschoeke DA, Thompson F, Thomson C (2016) Comparative genomics of Synechococcus and proposal of the new genus Parasynechococcus. PeerJ 4:e1522

Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, Karl D, Li W, Lomas M, Veneziano D, Vera C, Vrugt JA, Martiny

AC (2013) Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. Proc Nat Acad Sci 110:9824–9829

Gaget V, Welker M, Rippka R, de Marsac NT (2015) A polyphasic approach leading to the revision of the genus Planktothrix (Cyanobacteria) and its type species, P. agardhii, and proposal for integrating the emended valid botanical taxa, as well as three new species, Planktothrix paucivesiculata sp. nov. Syst Appl Microbiol 38:141–158

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J (2005) Re-evaluating prokaryotic species. Nat Rev Microbiol 3:733–739

Hahn MW, Jezberova J, Koll U, Saueressig-Beck T, Schmidt J (2016) Complete ecological isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene sequences. ISME J. doi:10.1038/ismej.2015.237.

Hugerth LW, Larsson J, Alneberg J, Lindh MV, Legrand C, Pinhassi J, Andersson AF (2015) Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biology 16(279):1–18

Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV (2012) Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science 335:587–590

Karlin S, Mrazek J, Campbel AM (1997) Compositional biases of bacterial genomes and evolutionary implications. J Bacteriol 179:3899–3913

Kim M, Oh HS, Park SC, Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351

Komárek J (2016) Review of the cyanobacterial genera implying planktic species after recent taxonomic revisions according to polyphasic methods: state as of 2014. Hydrobiologia 764(1):259–270. doi:10.1007/s10750-015-2242-0

Komárek J, Kaštovský J, Mareš J, Johansen JR (2014) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia 86:295–335

Konstantinidis KT, Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258–6264

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT (2011) Genome sequencing of environmental Escherichia coli expands understanding of the ecology and speciation of the model bacterial species. Proc Natl Acad Sci USA 108:7200–7205

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform 14:60

Micallef ML, D'Agostino PM, Al-Sinawi B, Neilan BA, Moffitt MC (2015) Exploring cyanobacterial genomes for natural product biosynthesis pathways. Mar Genom 21:1–12

Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA, Waterbury J (2003) The genome of a motile marine Synechococcus. Nature 424:1037–1042

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Nat Acad Sci USA 110:1053–1058

Six C, Finkel ZV, Irwin AJ, Campbell DA (2007a) Light variability illuminates niche-partitioning among marine Picocyanobacteria. PLoS One 2:e1341

Six C, Thomas J-C, Garczarek L, Ostrowski M, Dufresne A, Blot N, Scanlan DJ, Partensky F (2007b) Diversity and evolution of phycobilisomes in marine Synechococcus spp.: a comparative genomics study. Genome Biol 8:R259

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-castillo FM, Costea PI, Cruaud C, Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C (2015) Structure and function of the global ocean microbiome. Science 348:1–10

Thompson CC, Silva GG, Vieira NM, Edwards R, Vicente AC, Thompson FL (2013) Genomic taxonomy of the genus prochlorococcus. Microb Ecol 66:752–762

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A (2015) Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771

Wang H, Fewer DP, Sivonen K (2011) Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. PLoS One 6

Zangenah S, Abbasi N, Andersson AF, Bergman P (2016) Whole genome sequencing identifies a novel species of the genus Capnocytophaga isolated from dog and cat bite wounds in humans. Sci Rep 6:22919