Unlocking the genomic taxonomy of the *Prochlorococcus* collective

Diogo Tschoeke[1#], Livia Vidal[1#], Mariana Campeão[1], Vinícius W. Salazar[1], Jean Swings[1,2], Fabiano Thompson[1*], Cristiane Thompson[1*]

[1]Laboratory of Microbiology. SAGE-COPPE and Institute of Biology. Federal University of Rio de Janeiro. Rio de Janeiro. Brazil. Av. Carlos Chagas Fo 373, CEP 21941-902, RJ, Brazil. [2]Laboratory of Microbiology, Ghent University, Gent, Belgium.

*Corresponding authors:

E-mail: fabianothompson1@gmail.com , thompsoncristiane@gmail.com

Phone no.: +5521981041035, +552139386567

#These authors contributed equally.

ABSTRACT

Prochlorococcus is the most abundant photosynthetic prokaryote on our planet. The extensive ecological literature on the *Prochlorococcus* collective (PC) is based on the assumption that it comprises one single genus comprising the species *Prochlorococcus marinus*, containing itself a collective of ecotypes. Ecologists adopt the distributed genome hypothesis of an open pan-genome to explain the observed genomic diversity and evolution patterns of the ecotypes within PC. Novel genomic data for the PC prompted us to revisit this group, applying the current methods used in genomic taxonomy. As a result, we were able to distinguish the five genera: *Prochlorococcus*, *Eurycolium*, *Prolificoccus*, *Thaumococcus* and *Riococcus*. The novel genera have distinct genomic and ecological attributes.

KEYWORDS: *Prochlorococcus* collective, ecogenomics, genomic taxonomy, evolution

INTRODUCTION

Prochlorococcus cells are the smallest cells that emit red chlorophyll fluorescence, which is how they were discovered approximately 30 years ago [1, 2]. It is the smallest and most abundant photosynthetic microbe on Earth. The oxytrophic free-living marine *Prochlorococcus marinus* contain divinyl chlorophyll a (chl a2) and both monovinyl and divinyl chlorophyll b (chl b2) as their major photosynthetic pigments rather than chlorophyll a and phycobiliproteins, which are typical of cyanobacteria, allowing them to adapt to different light intensities [1, 3–6], particularly to capture the blue wavelengths prevailing in oligotrophic waters [7]. *Prochlorococcus* strains have been cultured and are available in culture collections, but the limited number of cultures does not reflect the true phenotypic and genomic diversity found in the ocean [8].

Prochlorococcus is a smaller, differently pigmented variant of marine *Synechococcus* [2]. The pioneering work on the eco-genomics of *Prochlorococcus* was performed by Chisholm's group at MIT [2, 9–14]. "When I look at our emerald green laboratory cultures of captive *Prochlorococcus*, my mind quickly turns to their wild cousins, drifting freely in the world oceans. I am reminded that while they carry out a respectable fraction of the photosynthesis on our planet, they escaped our attention" [15]. The taxonomy of the PC has been neglected in the last three decades. The designation PC was first coined by Kashtan et al. [11]. This collective comprises the genera, *Prochlorococcus, Prolificoccus, Eurycolium,* and *Thaumococcus* [16–18], a group of picocyanobacteria that radiated from *Synechococcus* millions of years ago. We preferer to use the neutral term "collective" here instead of "community", "complex" or any other term.

"Speciation" within PC.

Several authors have commented on speciation within this group of picocyaonobacteria. Berube et al. [19] argue that speciation is likely driven by a combination of vertical inheritance and gene loss and is rarely due to horizontal gene acquisition driven by non-homologous recombination. Speciation and niche partitioning involve brief periods of enhanced drift and (rapid) change that have facilitated, e.g., the genomic rearrangement of nitrate assimilation genes. Thompson et al. [20] demonstrated that *Prochlorococcus* from the Red Sea is not genetically isolated, despite its endemic genomic adaptations to the unique combination of high solar irradiance, high temperature, high salinity, and low nutrient levels in this sea. The authors consider orthologous groups of genes found in Red Sea *Prochlorococcus* to be cosmopolitan. Species exist as clusters with genetic and ecological similarity, and speciation seems mainly to be driven by natural selection,

regardless of the balance between horizontal and vertical descent. “If complete boundaries to gene flow take some time to emerge, we can think of gene sets rather than whole genomes as the units that inhabit ecological niches. If the gene flow boundaries never emerge, speciation does not occur (i.e. we are left with one species, not two) and this corresponds to the gene ecology model” of Shapiro et al. or to Cohan’s stable ecotype model of speciation [21–24]. Shapiro [25] further developed the ideas that “no genome is an island” and that natural genetic engineering responds to the ecological challenges that arise in the life histories of all evolving organisms.

Speciation within the PC may be influenced by co-evolution with *Pelagibacter*, a member of the SAR 11 clade [26–28]. The role of a cryptic nitrogen cycle has been elucidated recently [3]: *Pelagibacter* mainly uses nucleotide (purine) catabolism to obtain reducing power and releases urea, while *Prochlorococcus* utilizes this urea. The authors postulate the existence of a cryptic nitrogen cycle that transfers reducing power from *Prochlorococcus* to *Pelagibacter* and returns nitrogen to *Prochlorococcus*. This positive feedback loop would strengthen the co-evolution between these two abundant oceanic microbes.

Which forces have shaped the massive cell size/genome reductions and gigantic free-living populations of the PC during its millions years of evolution? The streamlining theory is generally cited to explain the massive genomic DNA losses and cell size reduction and the accompanying AT increase [29]. Low genetic drift and strong purifying selection have been postulated. The major driving force for genome reduction seems to be the substantial economy in energy and materials that has occurred, as the AT increase economizes nitrogen [30].

The PC is cited as an example of the evolution of dependencies of free-living organisms through adaptive gene losses, better known as the Black Queen hypothesis [31, 32], summarized by Shapiro (2019) [25] as "no genome is an island", to stress continuous biosphere communications during evolution.

Most literature on PC is based on the assumption that it comprises a single genus, in which the species *Prochlorococcus marinus* is within a "collective" of ecotypes. The main reason for this concept is that 16S rRNA gene sequence similarities within this group are >97%. To describe the observed diversity in the PC, ecologists adopt the distributed genome hypothesis of an open bacterial pan-genome (supra-genome) [12, 14]. In this hypothesis, the pan-genome is the full complement of genes that each member of the population contributes to and draws genes from. No single isolate contains the full complement of genes, which contributes to high diversity and provides selective advantages [33]. Described in its extreme form, the pan genome of the bacterial domain is of infinite size, and the bacteria as a whole present an open pan-genome (supragenome), in which the constant "rain" of genetic material in genomes from a cloud of frequently transferred genes enhances the chance of survival of a species [34]. For scientific progress, the use of these metaphors is necessary to clarify ideas and formulate new hypotheses. The PC case illustrates the continuously growing gap between prokaryotic taxonomy as it is practised today and in-depth eco-genomic studies. Genomic clusters are referred to as ecotypes, ecospecies, sub-ecotypes, clades, populations, subpopulations, OTUs or CTUs. This informal nomenclature finds its way into databases, adding further nomenclatural chaos to "terrae incognitae" [35, 36] comprising these picocyanobacteria.

In their theory of self-amplification and self-organization of the biosphere, Braakman et al. [3], hypothesize that *Prochlorococcus* clades living in surficial oceanic waters adapted to high light and low nutrient levels have evolved from more ancient *Prochlorococcus s*ubpopulations (~800 million years ago) inhabiting deeper oceanic areas, with lower light and higher nutrient concentrations. They refer to these evolutionary paths as "niche constructing adaptive radiations". Sun & Blanchard [37] hypothesise three phases in the evolutionary history of the PC: genome reduction, genome diversification and ecological niche specialization, all due to different evolutionary forces. The PC has adapted to live under different light intensities, sea water temperatures and nutrient levels, and the abundance of the different ecotypes varies markedly along gradients of these parameters [38–44]. However, only high light (HL)-adapted and low light (LL)-adapted ecotypes were initially distinguished [45]. At present, twelve phylogenetic ecotypes are delineated: six high light (HL)-adapted ecotypes and six low light (LL)-adapted ecotypes [46]. These ecotypes are based on the phylogeny of the 16S/23S rRNA internal transcribed spacer region, differences in behaviour towards light, whole-genome analysis, and physiological properties, combined with their environmental distribution [42, 44, 47]. Not only light adaptation but also temperature adaptation has been recognized in the two dominant near-surface HL ecotypes, the HL I ecotype *Eurycolium* eMED4 and the HL II ecotype *Eurycolium* eMIT9312, which both exhibit growth maxima at approximately 25 °C, but *Eurycolium* eMED4 appears to be adapted to thrive at much lower temperatures (approx. 14 °C), whereas *Eurycolium* eMIT9312 is able to grow at 30 °C [48]. These ecotypes appear to have nearly identical genomes [9], in which the major differences are restricted to five genomic islands, but show marked phenotypic differences.

The relative abundance of cells belonging to the different ecotypes has been estimated in several oceanic regions, revealing patterns that agree for the most part with the HL/LL phenotypes. HL-adapted cells dominate the surface mixed layer, and LL-adapted cells most often dominate deeper waters [46, 49–52]. By combining the physiological features of isolates and ecotype abundance studies in the ocean, it was shown that temperature, in addition to light, is an important determinant of the abundance of these ecotypes in oceans [53, 54]. There is also tremendous variability within ecotypes, particularly with regard to nutrient acquisition [55], susceptibility to predation or phage infection [56], and modes of interaction with other members of the community [57], including interactions with *Alteromonas* for protection against ROS [58, 59]. *Prochlorococcus* and *Synechococcus* can use different organic compounds containing key elements to survive in oligotrophic oceans, and can also take up glucose and use it as a source of carbon and energy[60].

Antisense RNA in *Prochlorococcus* is said to protect a set of host mRNAs and viral transcripts from RNAse degradation during phage infection [61]. In addition, CRISPR-Cas (clustered regularly interspaced short palindromic repeats, CRISPR-associated genes), an extremely adaptable defense system, is widely distributed in cyanobacteria [62]. The CRISPR-Cas system is found in the majority of cyanobacteria except for *Synechococcus* and *Prochlorococcus*, in which cyanophages are a known force shaping evolution [62]. However, more recently, CRISP-Cas systems have been found in novel *Synechococcus* genomes, indicating the possibility of additional mechanisms for genome plasticity [63].

Kashtan et al. [11] studied the genomic diversity of single *Prochlorococcus* cells in samples collected in the Bermuda-Atlantic Time-Series Study (BATS) site in summer, spring, and winter, and hundreds of new *Prochlorococcus* "subpopulations" with distinct genomic backbones were found. These subpopulations are estimated to have diverged millions of years ago, suggesting ancient, stable niche partitioning followed by selection (not by genetic drift). The authors suggested that the global PC may be conceived as an assortment of thousands of species. They argued that such a large set of coexisting species with distinct genomic backbones is a characteristic feature of free-living prokaryotes with very large population sizes living in highly mixed habitats. In another study by these authors focused on the Pacific Ocean, they concluded that the Atlantic and Pacific Oceans are occupied by different PC populations that experienced separate evolutionary paths for a few million years [13].

Phylogeography seems to be echoed in the extensive genome diversity of the PC identified in 226 metagenomic samples from the GOS (Global Ocean Sampling) expedition [44]. In another paper, Kent et al. [64] studied the co-occurrence of *Prochlorococcus* and *Synechococcus* in 339 samples from three cruises covering equatorial, subtropical gyre, and colder nutrient-rich mid-latitude regions in the Atlantic and Pacific using the phylogenetic marker gene *rpoC*1, a single-copy gene encoding the γ subunit of RNA polymerase. They hypothesized that where *Prochlorococcus* and *Synechococcus* co-occur, parallel evolutionary diversification occurs, caused by the same stressors. The authors distinguish three major matching surface ecotypes: *Prochlorococcus* HLIII/IV (corresponding to *Synechococcus* CRD1); HLII (*Synechococcus* clade II/III) and HLI (*Synechococcus* I/IV). This parallel

phylogeography extends from the ecotype level further into finer genomic networked diversity.

Principles of genomic prokaryote taxonomy

From 1970 to the present, the need to classify, name and identify prokaryotes has been based on a pragmatic polyphasic consensus approach [65]. This non-theory-based approach aims to integrate phenotypic, genotypic, and phylogenetic and ecological characteristics to establish stable and informative classification systems. The central unit is the prokaryotic species. In this procedure, the definition of prokaryotic species is based mainly on DNA–DNA hybridization (DDH) and 16S rRNA sequence similarity: a species is defined as a monophyletic group of strains sharing high phenotypic and genomic similarity (>70% DDH, >98.3% 16S rRNA sequence similarity) [66–70].

Prokaryote taxonomy is becoming more genome based, considering the genomes themselves as the ultimate end products of microbial evolution [68, 71, 72]. Such an approach offers the opportunity to obtain a more natural taxonomic system and allows the definition of consistent and stable monophyletic species and genera with maximal utility for its users (see 47, 48, 78-98). Cohan has extensively developed the ecotype or ecospecies theory for prokaryotes, in which "ecotypes are the most newly divergent populations that are distinct ecologically from one another [21, 73]. In this theory, the splitting of one ecotype into two is the fundamental diversity-creating process of speciation in bacteria. In his stable ecotype model, ecotypes (ecospecies) are long lived, and different ecotypes coexist indefinitely. They are cohesive by virtue of periodic selection events that purge the ecotype of sequence diversity. In this case, the ecotype

has all the characteristics of a species [23, 73]. Ecotypes have been discerned among the PC collective.

The place of the PC in marine microbial taxonomy

The main motivation for the present study arose from the consideration that the taxonomy of *Prochlorococcus* has remained relatively neglected and unresolved compared to the tremendous advances in marine microbial taxonomy achieved for other groups, e.g., *Alteromonas* (N=38 species) [74], *Colwellia* (N=21 species) [75], *Endozoicomonas* (N=9 species) [76], *Marinobacter* (N=45 species) [77], *Marinobacterium* (N=17 species) [78], *Marinomonas* (N=30 species) [79], *Pseudoalteromonas* (N=53 species) [80], *Pseudovibrio* (N=6 species) [81], *Roseobacter* (N=5 species) [82], *Ruegeria* (N=22 species) [83], *Shewanella* (N=67 species) [84], *Salinispora* (N=3 species) [85], *Streptomyces* (N=848 species) [86], *Thalassomonas* (N=9 species) [87], and the vibrio clade (N=199 species, including the genera *Vibrio*, *Photobacterium, Enterovibrio, Aliivibrio, Thaumasiovibrio,* and *Grimontia*) [88]. Even marine prokaryotes living in extreme oceanic environments such as the OMZ, sulfate-reducing bacteria, N cycle bacteria, and deep sea hydrothermal vent sulfur-oxidizing *Sulfurovum* (N=4 species) [89] and *Sulfurimonas* (N=4 species) [90] have been the subject of advanced taxonomic classification. Nevertheless, even in the well-studied Gram-negative *Roseobacter* [91] and Gram-positive *Streptomyces* [92] genera, the paraphyletic nature of these genera leads to noncoherent assignment of scientific names and causes confusion when interpreting their ecology or evolutionary biology.

As new habitats and niches are found for *Prochlorococcus*, it is important to establish a solid taxonomic system to contribute to the interpretation of the ecology of the PC.

Novel PC genomic data allowed us to address the hypothesis that novel genera and species are present within the PC. In this review, we revisit the current taxonomy of the PC and propose the creation of a new genus to encompass the clearly recognized clusters. The novel genera have distinct genomic and ecological attributes.

METHODS

Genome quality assessment

A total of 208 genomes and the corresponding metadata were downloaded from NCBI GenBank [93]. Contigs or scaffold files were input into Prodigal software v2.6.3 [94] for gene and protein prediction. To check quality of the genomes, completeness, contamination and GC%, the genomes were submitted to CheckM v1.0.11 [95] using default parameters for this estimation. CheckM uses a broader set of marker genes specific to the position of a genome within a reference genome tree. The completeness is measured using a set of marker genes that are expected to be present. This expectation can be defined along an evolutionary gradient, ranging from highly conserved genes to species-specific genes. The genomes used are of high quality and have ≥ 85% of completeness, ≥ 1Mb of size and ≤ 5% of contamination.

Average amino acid identity analyses (AAI)

The average amino acid identity was calculated as described previously for all 208 genomes [96]. All PC genomes were used to calculate the amino acid average identity using the *CompareM* tool (https://github.com/dparks1134/CompareM). The result was used to plot a dendrogram. The *CompareM* output was converted to a dissimilarity square matrix. The *spatial.distance.squareform* function from *scipy* package v1.1.0 was used in order to obtain a condensed list from the lower triangular elements of the matrix.

A complete linkage clustering was performed using *cluster.hierarchy.linkage* function from *scipy* with the output of *squareform* and cityblock metric distance. The figure that represents the linkage matrix was plotted with *cluster.hierarchy.dendrogram* function from *scipy*. The intraspecies limit is assumed as ≥95%, whereas genera delimitation is assumed as ≥70%. Genera were delineated only for those lineages exhibiting at least 3 representative genome sequences.

Phylogenomic reconstruction

The genome sequences of five representative PCs corresponding to the type species of the five genera were selected based on AAI analysis. The MLSA approach was based on the concatenated sequences of four house-keeping genes (*gyrB, pyrH, recA, rpoB*) with intraspecies relatedness ranging from >95 to 100 % MLSA[16]. This was achieved by using tBLASTn [97]. A nucleotide alignment was then made for each of the four genes using MAFFT [98], and these alignments concatenated using MAFFT with default parameters and a custom Perl script. The tree was built from the concatenated alignment using FastTree 2 [99] with default parameters and rooted on *S. elongatus* PCC 6301 using ETE 3 [100]. To build the core genome tree, we used the GToTree package [101] with default parameters. The 208 selected genomes were searched against a Hidden Markov Model of 249 Cyanobacterial marker genes using HMMER3 [102]. A concatenated protein alignment from the 249 genes was constructed using Muscle [103] and subsequently trimmed using TrimAl [104] (Gutierrez et al. 2009). The alignment was then used to construct a tree using FastTree 2 [99] with default parameters and the pairwise distance matrix using MEGA 6.0 [105]. The tree was

rooted on S. elongatus PCC 6301 using ETE 3 [100]. Part of the processing was done with GNU Parallel [106].

In silico phenotype

The phenotype was predicted utilizing the sequences of the nitrogen (*narB, nirA, ntcA, glnB, amt, cynS, ureA, glnA, glsF, gdhA*), phosphorous (*phoB, ptrA, phoR*), arsenate (*arsC*) and iron (*piuC, som*) related genes of *Prochlorococcus* obtained from UniProt and GenBank as queries for a local blastp [107] search against all PC ORF-predicted genomes. The results were parsed with Python scripts, and the sequences corresponding to the blastp hits were extracted from the respective genomes. The extracted sequences were used to perform a blastp search against the nr database to confirm the functional annotation. Manual verification of these results was performed. The confirmed results were used to build a matrix of presence/absence for the genes related to the indicated proteins in each PC genome.

RESULTS AND DISCUSSION

A proposal of a new taxonogenomic framework for the PC

The availability of *Prochlorococcus* genomes labelled *Prochlorococcus* in public databases prompted us to test the hypothesis that the PC comprises a greater number of stable genera and species than has been previously recognized (Table 1 and S1). We first collected all the publicly available genomes designated as *Prochlorococcus* and applied the theory-free genomic methodologies and principles discussed previously. The *Prochlorococcus* genomes came from samples distributed around the world, especially in the Pacific and Atlantic oceans (Table S1). First, the GC% content of the genomes was determined. A phylogenomic framework was established using genomic signatures,

i.e., average amino acid identity (AAI), MLSA and core genome-based tree. We then classified the genomes into different genera and species using relevant AAI and phenotypic information obtained directly from the genomes.

GC content reveals great genomic diversity in the PC

One of the most striking findings of our study came from the analysis of GC content, which ranged from 30% to 50.7% (Figure 1 and Table 1). The distribution of the GC values of all the genomes comprised five groups: i. the majority (75% of the genomes) exhibited 30 to 33% GC, comprising the genus *Eurycolium*; ii. a second group (12% of the genomes) exhibited 34 to 35% GC, comprising the genera *Prolificoccus*; iii. a third group *Prochlorococcus* (4.8% of the genomes) presented 36 - 37% GC; iv. a fourth group *Riococcus* (1.4% of the genomes) exhibited 37 to 38% GC; and v. the fifth group (6.7% of the genomes) exhibited 50 to 50.7% GC, which was characteristic of the genera *Thaumococcus*. The average genome sizes of the first four GC groups were similar, whereas the genome sizes of the fifth GC group were almost two-fold larger. This GC analysis provides a clear indication of the taxonomic width of the PC. This diversification is the result of millions of years of evolution. Mutational bias is dominated by GC > AT transitions even in bacteria with a high % GC content [108, 109]. This means that there must be some pressure to maintain high % GC contents in those prokaryotes that exhibit high % GC values. Why this is so remains a mystery, but Hershberg & Petrov noted that selective forces consistently operate both genome-wide and over long periods of time [109]. The nature of such selection remains obscure. If natural selection plays such a strong role in determining % GC content, it suggests that there are no truly neutrally evolving sites.

Average amino acid identity of PC

The species and genus delimitations were based on clusters with three or more genomes that shared ≥95% or ≥70% amino acid average identity (AAI), respectively. Here, we revisit the current taxonomy of the PC and propose the creation of a new genus *Riococcus* and 113 novel species (N=122 species) (Table 1 and S1). We confirmed the previous proposal of Walter et al. [18], and we now expand the classification of PC (Figure 2). *Eurycolium* (N=97) and *Prolificoccus* (N=16), were the genera with the greatest numbers of species followed by *Prochlorococcus* (N=4), *Riococcus* (N=3) and *Thaumococcus* (N=2) (Figure 2, Table 1 and S1). The AAI among PC members varied considerably, between 57 and 99.9%. The closest genus to *Prochlorococcus marinus* CCMP1375^{T} was *Riococcus* MIT9211^{T} having 69.4% AAI. The percentages of AAI variability between genera and species were 57 to 69.9% and 70 to 94%, respectively. The percentages of AAI variability within genera and species were >70 to 94% and 95 to 99.9%, respectively.

Phylogenomic framework for the PC

To establish the phylogenetic structure of the five PC genera, phylogenetic analysis was performed using a four-locus based tree (MLSA) and 249 core protein sequences. Bootstrap analysis indicated that most of the branches were highly significant. These analyses have a much higher resolution than the 16S rRNA phylogeny. The PC harbours very similar 16S rRNA gene sequences with similarity ranging from 97 to 99% [2]. The MLSA and the core protein tree showed the relationship of the five genera (Figure 3). The MLSA and the core protein sequence identities among PC members varied considerably, from 49 to 100% and from 56 to 100%, respectively. The type species of the genus *Prochlorococcus*, *P. marinus* CCMP1375^{T}, showed identity maxima of 80% and 70% MLSA and core protein sequence similarity, respectively with

all other genera. The variability between species was < 95, while that within species ranged from 96 to 100% similarity. The phylogenetic analysis based on MLSA and core protein sequences (Figure 3) demonstrated congruence between the genetic and ecological features of the five novel genera. For instance, *Eurycolium* comprises high-light (HLI/II/VI) ecotypes with 30-33% GC contents, whereas the genera *Prolificoccus* and *Prochlorococcus* comprise low-light (LLI and LLII/III, respectively) ecotypes with 34-35% and 36-37% GC contents, respectively. *Riococcus* and *Thaumococcus* consists of low-light ecotypes (LLII/III and LLIV, respectively) with 37 to 38% and 50 to 50.7% GC contents, respectively.

***In silico* phenotype prediction**

Useful phenotypic features that may be used for the identification and classification of the PC were identified, including nitrogen, phosphorous, arsenate, and iron utilization (Table 2). The genus *Prochlorococcus* does not present some nitrogen-related genes (*narB*, *nirA*, *cynS*, *ureA*, and *gdhA*) and phosphorous-related genes (*phoB, ptrA and phoR*) but some genomes do present other nitrogen-related genes (*ntcA*, *glnB*, *amt*, *glnA*, *glsF*), an arsenate-related gene (*arsC*) and iron-related genes (*piuC* and *som*). *Prochlorococcus* is differentiated from its closest genus *Riococcus* by the presence of the *arsC* gene. All genera present the gene *Pro1404* that confers the capability for glucose uptake in PC [60].

Overall, we demonstrated that the methods used (AAI, core protein sequences and MLSA) correlated and provided significant taxonomic resolution for differentiation of genera and species in PC (Figure 2 and 3). The proposed genera are in complete agreement with previous literature and support previous theories on the evolution of

Prochlorococcus [11, 13, 14, 19, 45, 47, 50]. The PC genomic signatures were more similar between closely related species than distantly related species. The analysis performed in this study establishes a new taxonomic framework for PC.

Challenges in the taxonomy of the Cyanobacteria phylum and the PC

Through evolution, Cyanobacteria became one of the most diverse and widely distributed prokaryote groups, occupying many niches within terrestrial, planktonic, and benthic habitats. Its long history has resulted in the evolution of broad heterogeneity, comprising unicellular and multicellular, photosynthetic and non-photosynthetic, free-living, symbiotic, toxic and predatory organisms [110–113]. Cyanobacteria have been named according to the Botanical Code [114]. The inclusion of Cyanobacteria in taxonomic schemes of Bacteria was only proposed in 1978 by Stanier et al [115] ; over time, bacterial taxonomic names have come into conflict with the botanical nomenclature [116, 117]. More than two decades passed before the Note to General Consideration 5 (1999) was published indicating that Cyanobacteria should be included under the rules of the International Committee on Systematic Bacteriology (ICSB)/International Committee on Systematic of Prokaryotes (ICSP) [118–120]. Taxon nomenclature within this group has long been a topic of discussion, but there is currently no consensus [121–124]. As a result, more than 50 genera of Cyanobacteria have been described since 2000, and many of them remain unrecognized in the List of Prokaryotic Names with Standing in Nomenclature (LPSN, http://www.bacterio.net) [125] and in databases (e.g., NCBI) or have been the target of high criticism [126, 127].

Cyanobacterial taxonomy is based on morphologic traits and may not reflect the results of phylogenetic analyses [110, 128–132]. There is a predominance unrelated

Cyanobacteria morphologically assembled into polyphyletic species, genera and higher taxonomic categories, which will require revisions in the future [133]. The polyphyly is indicative of the taxonomic mislabelling of many taxa. The analysis of 16S rRNA gene sequences is useful for charting and characterizing microbial communities [134] but lacks sensitivity to evolutionary changes that occur in association with ecological dynamics, in which microbial diversity is organized by physicochemical parameters [135, 136]. Hence, the processes that shape cyanobacterial communities over space and time are less well known. A recent study proposed that there should be 170 genera of Cyanobacteria based only on 16S rRNA sequences [134]. Farrant et al. delineated 121 ecologically significant taxonomic units (ESTUs) of *Prochlorococcus* and 15 ESTUs *Synechococcus* in the global ocean using single-copy *petB* sequences (encoding cytochrome b6) and environmental cues [137]. Although it was assumed that all the disclosed *Prochlorococcus* belonged to a single genus, it was not clear how many phylogenetic groups (distinct genera) corresponded to these 121 *Prochlorococcus*. Our study suggests that PC comprises at least five genera and at least 122 species. The 121 ESTUS obtained correspond to our five genera and 121species.

Unlocking the genomic taxonomy of the PC

Despite the astonishing advances in understanding the ecology of the PC, its taxonomic structure has remained puzzling until recently. The *Prochlorococcus* collective is thought to present a high degree of panmixis due to horizontal gene transfer and is composed of one genus and one species comprising two subspecies, *P. marinus* subsp *marinus* and *P. marinus* subsp *pastoris* [47, 138–141]. We remark that at least 12 stable ecotypes have been delineated in ecological studies. On the other hand, the taxonomic heterogeneity among the named HL and LL ecotypes was noted by Thompson et al.,

and the current ecotypes were split into 10 species based on species definition as a group of strains that share >95% DNA identity in MLSA, >95% AAI, >70% identity of GGD and a maximum GC% range 2%. Strains of the same species form monophyletic groups on the basis of MLSA [16]. Walter et al. [18] further split the *Prochlorococcus* collective into three new genera: (1) the genus *Eurycolium* (type species *Eurycolium pastoris* with the type strain MED4^T (= CCMP 1986^T), *E. chisholmii* with its type strain AS9601^T, *E. ponticus* with its type strain MIT9301^T, *E. nereus* with its type strain MIT9202^T), *E. neptunius* with its type strain MIT9312^T, *E. ponticus* with its type strain MIT 9301^T, and *E. tetisii* with its type strain MIT9515^T; (2) the genus *Prolificoccus*, with *Prolificoccus proteus* as the type species and NATL2A^T as its type strain; and (3) the genus *Thaumococcus*, with *Thaumococcus swingsii* as the type species and the type strain MIT9313^T. The current (type) genus *Prochlorococcus* contains two species: the type species *Prochlorococcus marinus* and its type strain SS120^T (= CCMP 1375^T) and *P. ceticus* with its type strain MIT9211^T. The genus *Eurycolium* seems to form a large diverse eco-genomic group related to high temperature and oligotrophic environments, whereas the genera *Prochlorococcus* and *Prolificoccus* seem to be abundant at low temperature; *Thaumococcus* appears to thrive at low temperature and in copiotrophic environments [18].

On the basis of the eco-genomic and in silico phenotype evidence gathered amassed here, we propose five new genera within the PC. The taxonomic structure of the PC now comprises the following type species and type strains or type sequences: (1) the genus *Prochlorococcus* (Pro'chlo ro coc"cus. Gr. pref. pro, before (primitive); Gr. adj chloros, green; M.L. masc n. coccus, berry. Primitive green sphere) (the type genus). The type species *Prochlorococcus marinus* (ma'ri"nus. marinus L. adj., marine) with its

type strain CCMP1375^T (=SS120^T). The genome of this strain presents a GC content of 36,4%; (2) the genus *Prolificoccus* (Pro.li.fi.co.ccus. L. prolificus, productive, abundant, numerous; *Prolificoccus*, referring to an abundant coccus). The type species *Prolificoccus proteus* (L. n. proteus, ofprotos, "first", is an early sea god, the old man of the sea) with its type strain NATL2A^T. The genome of this strain presents a GC content of 35,1%; (3) the genus *Eurycolium* (Eur.y.co.lium. Gr. adj. eury, wide, broad; L. cole, inhabit; Gr. ium, quality or relationship, *Eurycolium*, referring to the spread inhabiting trait in marine habitats). The types species *Eurycolium pastoris* (pa.sto'ris. Latinized form of Pasteur) with its type strain CCMP1986^T (=MED4^T). The genome of this strain presents a GC content of 30,8%; (4) the genus *Thaumococcus* (Thau.mo.co.ccus. Referring to Thaumas, Greek god of the wonders of the sea; N.L. masc. n. coccus (from Gr. Masc. n. kokkos, grain, seed, kernel); N.L masc. n. *Thaumococcus*, referring to a coccus living in the sea). The type species *Thaumococcus swingsii* (L. gen. masc. n. *swingsii*, of Swings, in honor of the Belgian microbiologist Jean Swings) with its type strain MIT 9313^T. The genome of this strain presents a GC content of 50,7%; and (5) the genus *Riococcus* (Rio.coc'cus. N.L. masc. n. Rio, of Rio de Janeiro, a city in Brazil; N.L. masc. n. kokkos, a grain; N.L. masc. n. *Riococcus*, referring the city of Rio de Janeiro). The type species *Riococcus ceticus (*L. n. *ceticus*, of Cetus, denotes a large fish) with its type strain MIT9211^T. The genome of this strain presents a GC content of 38%; We suggest that a PC species may be defined as a group of strains that share >95% DNA identity in core protein sequences, and >95% AAI. Strains of the same species will form monophyletic groups on the basis of core protein sequences.

CONCLUSIONS

The PC represents an excellent model for further theory development in ecology, evolution, phylogeny and taxonomy. The PC refers to a group of picocyanobacteria that radiated from *Synechococcus* during millions of years of evolution in the oceans. The taxonogenomic analysis of 208 genomes revealed at least 5 new genera in addition to the original genus, *Prochlorococcus* [142]. As we delineated genera only for those groups with at least three available genome sequences, we anticipate that new genera will be included in our new proposal when more representative genomes are added to groups with only 1 or 2 genomes. On average, 20% of the genes of PC genomes have unknown functions and await further characterization. The pan-genome of the PC is estimated to include over 80000 genes [14, 46].

Because of the universal occurrence of the PC in the oceans, it may serve as a basis for monitoring the health of the oceans and climate change. Ecological transitions and changes in the oceans trigger natural genetic engineering and genome changes on a massive scale [25, 143]. The PC has evolved in multiple ecologic niches in the global ocean and participates in relevant biogeochemical cycles, such as the recently postulated parasitic arsenic cycle between *Eurycolium* and *Pelagibacter ubique* [144, 145]. We demonstrated here that the current PC is a diverse group of genera that are clearly distinguishable by eco-genomics. Ecological studies on the PC may need to consider the diversity of genera and species when novel ecologic and evolutionary hypotheses are tested.

ACKNOWLEDGMENTS

The authors thank CNPq, CAPES and FAPERJ.

REFERENCES

1. Chisholm SW, Frankel SL, Goericke R, et al (1992) Prochiorococcus marinus nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. Arch Microbiol 297–300
2. Chisholm SW (2017) Prochlorococcus. Curr Biol 27:R447--R448. https://doi.org/10.1016/j.cub.2017.02.043
3. Braakman R, Follows MJ, Chisholm SW (2017) Metabolic evolution and the self-organization of ecosystems. Proc Natl Acad Sci 114:E3091--E3100. https://doi.org/10.1073/pnas.1619573114
4. Ting CS, Ramsey ME, Wang YL, et al (2009) Minimal genomes, maximal productivity: comparative genomics of the photosystem and light-harvesting complexes in the marine cyanobacterium, Prochlorococcus. Photosynth Res 101:1–19
5. Scanlan DJ, Ostrowski M, Mazard S, et al (2009) Ecological Genomics of Marine Picocyanobacteria. Microbiol Mol Biol Rev 73:249–299. https://doi.org/10.1128/MMBR.00035-08
6. Ralf G, Repeta DJ (1992) The pigments of Prochlorococcus marinus: The presence of divinylchlorophyll a and b in a marine procaryote. Limnol Oceanogr 37:425–433
7. Partensky F, Garczarek L (2010) Prochlorococcus: advantages and limits of minimalism. Ann Rev Mar Sci 2:305–331
8. Mühling M (2012) On the culture-independent assessment of the diversity and distribution of Prochlorococcus. Environ Microbiol 14:567–579
9. Coleman ML, Sullivan MB, Martiny AC, et al (2006) Genomic Islands and the Ecology and Evolution of Prochlorococcus. Science (80-) 311:1768–1770. https://doi.org/10.1126/science.1122050

10. Coleman ML, Chisholm SW (2007) Code and context : Prochlorococcus as a model for cross-scale biology. 15:. https://doi.org/10.1016/j.tim.2007.07.001

11. Kashtan N, Roggensack SE, Rodrigue S, et al (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science (80-) 344:416–420. https://doi.org/10.1126/science.1248575

12. Biller SJ, Berube PM, Berta-Thompson JW, et al (2014) Genomes of diverse isolates of the marine cyanobacterium Prochlorococcus. Sci Data 1:140034

13. Kashtan N, Roggensack SE, Berta-Thompson JW, et al (2017) Fundamental differences in diversity and genomic population structure between Atlantic and Pacific Prochlorococcus. ISME J 11:1997–2011. https://doi.org/10.1038/ismej.2017.64

14. Berube PM, Biller SJ, Hackl T, et al (2018) Data descriptor: Single cell genomes of Prochlorococcus, Synechococcus, and sympatric microbes from diverse marine environments. Sci Data 5:1–11. https://doi.org/10.1038/sdata.2018.154

15. Chisholm SW (2012) The Life and Times of the Ocean's Smallest Photosynthetic Cell. In: Kolter, Roberto and Maloy S (ed) Microbes and evolution: the world that Darwin Never Saw. ASM Press, pp 165–172

16. Thompson CC, Silva GGZ, Vieira NM, et al (2013) Genomic Taxonomy of the Genus Prochlorococcus. Microb Ecol 66:752–762. https://doi.org/10.1007/s00248-013-0270-8

17. Coutinho F, Tschoeke DA, Thompson F, Thompson C (2016) Comparative genomics of Synechococcus and proposal of the new genus Parasynechococcus. PeerJ 4:e1522. https://doi.org/10.7717/peerj.1522

18. Walter JM, Coutinho FH, Dutilh BE, et al (2017) Ecogenomics and taxonomy of Cyanobacteria phylum. Front Microbiol 8:.

https://doi.org/10.3389/fmicb.2017.02132

19. Berube PM, Rasmussen A, Braakman R, et al (2019) Emergence of trait variability through the lens of nitrogen assimilation in Prochlorococcus. Elife 8:1–28. https://doi.org/10.7554/elife.41043

20. Thompson LR, Haroon MF, Shibl AA, et al (2019) Red Sea SAR11 and Prochlorococcus Single-cell Genomes Reflect Globally Distributed Pangenomes. bioRxiv 549816. https://doi.org/https://doi.org/10.1101/549816

21. Cohan FM (2002) What are bacterial species? Annu Rev Microbiol 56:457–487

22. Cohan FM (2016) Dispatches Bacterial Speciation : Genetic Sweeps in Bacterial Species. Curr Biol 26:R112–R115. https://doi.org/10.1016/j.cub.2015.10.022

23. Cohan FM, Kopac SM (2017) A Theory-Based Pragmatism for Discovering and Classifying Newly Divergent Species of Bacterial Pathogens. Genet Evol Infect Dis 25

24. Shapiro BJ, Leducq JB, Mallet J (2016) What Is Speciation? PLoS Genet 12:1–15. https://doi.org/10.1371/journal.pgen.1005860

25. Shapiro JA (2019) No genome is an island: toward a 21st century agenda for evolution. Ann N Y Acad Sci nyas.14044. https://doi.org/10.1111/nyas.14044

26. Alves Junior N, Meirelles PM, de Oliveira Santos E, et al (2014) Microbial community diversity and physical–chemical features of the Southwestern Atlantic Ocean. Arch Microbiol 197:165–179. https://doi.org/10.1007/s00203-014-1035-6

27. Armengol L, Calbet A, Franchy G, Rodrígu A (2019) Planktonic food web structure and trophic transfer efficiency along a productivity gradient in the tropical and subtropical Atlantic Ocean. Sci Rep 1–19. https://doi.org/10.1038/s41598-019-38507-9

28. Sunagawa S, Coelho LP, Chaffron S, et al (2015) Structure and Function of the Global Ocean Microbiome. Science (80-) 348:

29. Giovannoni SJ, Cameron Thrash J, Temperton B (2014) Implications of streamlining theory for microbial ecology. ISME J 8:1553–1565. https://doi.org/10.1038/ismej.2014.60

30. Dufresne A, Garczarek L, Partensky F (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biol 6:R14

31. Morris JJ, Lenski RE, Zinser ER (2012) The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio 3:e00036----12

32. Mas A, Jamshidi S, Lagadeuc Y, et al (2016) Beyond the black queen hypothesis. ISME J 10:2085

33. Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11:472–477

34. Lapierre P, Gogarten JP (2009) Estimating the size of the bacterial pan-genome. Trends Genet 25:107–110

35. Rosselló-Móra R, Whitman WB (2019) Dialogue on the nomenclature and classification of prokaryotes. Syst Appl Microbiol 42:5–14. https://doi.org/10.1016/j.syapm.2018.07.002

36. De Vos P, Thompson F, Thompson C, Swings J (2017) A Flavor of Prokaryotic Taxonomy: Systematics Revisited. In: Microbial Resources: From Functional Existence in Nature to Applications. Elsevier Inc., pp 29–44

37. Sun Z, Blanchard JL (2014) Strong genome-wide selection early in the evolution of Prochlorococcus resulted in a reduced genome through the loss of a large number of small effect genes. PLoS One 9:e88837

38. Johnson ZI, Zinser ER, Coe A, et al (2006) Niche partitioning among

Prochlorococcus ecotypes along ocean-scale environmental gradients. Science (80-) 311:1737–1740

39. Garczarek L, Dufresne A, Rousvoal S, et al (2007) High vertical and low horizontal diversity of Prochlorococcus ecotypes in the Mediterranean Sea in summer. FEMS Microbiol Ecol 60:189–206. https://doi.org/10.1111/j.1574-6941.2007.00297.x

40. Rusch DB, Martiny AC, Dupont CL, et al (2010) Characterization of Prochlorococcus clades from iron-depleted oceanic regions. Proc Natl Acad Sci 107:16184–16189

41. Larkin AA, Blinebry SK, Howes C, et al (2016) Niche partitioning and biogeography of high light adapted Prochlorococcus across taxonomic ranks in the North Pacific. ISME J 10:1555

42. West NJ, Lebaron P, Strutton PG, Suzuki MT (2011) A novel clade of Prochlorococcus found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. ISME J 5:933

43. Huang S, Wilhelm SW, Harvey HR, et al (2012) Novel lineages of prochlorococcus and synechococcus in the global oceans. ISME J 6:285–297. https://doi.org/10.1038/ismej.2011.106

44. Kent AG, Dupont CL, Yooseph S, Martiny AC (2016) Global biogeography of Prochlorococcus genome diversity in the surface ocean. ISME J 10:1856–1865. https://doi.org/10.1038/ismej.2015.265

45. Moore LR, Chisholm SW (1999) Photophysiology of the marine cyanobacterium Prochlorococcus: ecotypic differences among cultured isolates. Limnol Oceanogr 44:628–638

46. Biller SJ, Berube PM, Lindell D, Chisholm SW (2015) Prochlorococcus: The

structure and function of collective diversity. Nat Rev Microbiol 13:13–27. https://doi.org/10.1038/nrmicro3378

47. Rocap G, Larimer FW, Lamerdin J, et al (2003) Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. Nature 424:1042

48. Chandler JW, Lin Y, Gainer PJ, et al (2016) Variable but persistent coexistence of Prochlorococcus ecotypes along temperature gradients in the ocean's surface mixed layer. Environ Microbiol Rep 8:272–284

49. Binder BJ, Chisholm SW, Olson RJ, et al (1996) Dynamics of picophytoplankton, ultraphytoplankton and bacteria in the central equatorial Pacific. Deep Sea Res Part II Top Stud Oceanogr 43:907–931

50. Ferris MJ, Palenik B (1998) Niche adaptation in ocean cyanobacteria. Nature 396:226

51. García-Fernández JM, Hess WR, Houmard J, Partensky F (1998) Expression of thepsbAGene in the Marine OxyphotobacteriaProchlorococcusspp. Arch Biochem Biophys 359:17–23

52. Fuhrman JA, McCallum K, Davis AA (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. Appl Environ Microbiol 59:1294–302

53. Martiny AC, Coleman ML, Chisholm SW (2006) Phosphate acquisition genes in Prochlorococcus ecotypes: evidence for genome-wide adaptation. Proc Natl Acad Sci 103:12552–12557

54. Martiny AC, Huang Y, Li W (2009) Occurrence of phosphate acquisition genes in Prochlorococcus cells from different ocean regions. Environ Microbiol 11:1340–1347

55. Moore LR, Post AF, Rocap G, Chisholm SW (2002) Utilization of different

nitrogen sources by the marine cyanobacteria Prochlorococcus and Synechococcus. Limnol Oceanogr 47:989–996

56. Sullivan MB, Waterbury JB, Chisholm SW (2003) Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. Nature 424:1047

57. Sher D, Thompson JW, Kashtan N, et al (2011) Response of Prochlorococcus ecotypes to co-culture with diverse marine bacteria. ISME J 5:1125

58. Biller SJ, Coe A, Chisholm SW (2016) Torn apart and reunited: impact of a heterotroph on the transcriptome of Prochlorococcus. ISME J 10:2831

59. Biller SJ, Berube PM, Dooley K, et al (2018) Data descriptor: Marine microbial metagenomes sampled across space and time. Sci Data 5:1–7. https://doi.org/10.1038/sdata.2018.176

60. Muñoz-marín MC, Gómez-baena G, López-lozano A, et al (2020) Mixotrophy in marine picocyanobacteria : use of organic compounds by Prochlorococcus and Synechococcus. ISME J. https://doi.org/10.1038/s41396-020-0603-9

61. Georg J, Hess WR (2018) Widespread Antisense Transcription in Prokaryotes. Microbiol Spectr 1–20. https://doi.org/10.1128/microbiolspec.RWR-0029-2018.Correspondence

62. Cai F, Axen SD, Kerfeld CA (2013) Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. RNA Biol 10:687–693. https://doi.org/10.4161/rna.24571

63. Burgsdorf I, Slaby BM, Handley KM, et al (2015) Lifestyle Evolution in Cyanobacterial Symbionts of Sponges. MBio 6:1–14. https://doi.org/10.1128/mbio.00391-15

64. Kent AG, Baer SE, Mouginot C, et al (2019) Parallel phylogeography of Prochlorococcus and Synechococcus. ISME J. https://doi.org/10.1038/s41396-

018-0287-6

65. Vandamme P, Pot B, Gillis M, et al (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Mol Biol Rev 60:407–438

66. Wayne LG, Brenner DJ, Colwell RR, et al (1987) Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. Int J Syst Evol Microbiol 37:463–464. https://doi.org/10.1099/00207713-37-4-463

67. Stackebrandt E, Frederiksen W, Garrity GM, et al (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52:1043–1047

68. Gevers D, Cohan FM, Lawrence JG, et al (2005) Defining prokaryotic species Reevaluating prokaryotic species. Nat Rev Microbiol 3:733–739

69. Thompson CC, Chimetto L, Edwards RA, et al (2013) Microbial genomic taxonomy. BMC Genomics 14:913. https://doi.org/10.1186/1471-2164-14-913

70. Thompson CC, Amaral GR, Campeão M, et al (2015) Microbial taxonomy in the post-genomic era: rebuilding from scratch? Arch Microbiol 197:359–370

71. Castelle CJ, Banfield JF (2018) Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell 172:1181–1197

72. Parks DH, Chuvochina M, Waite DW, et al (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229

73. Cohan FM (2016) Bacterial speciation: Genetic sweeps in bacterial species. Curr Biol 26:R112--R115. https://doi.org/10.1016/j.cub.2015.10.022

74. Baumann L, Baumann P, Mandel M, et al (1972) Taxonomy of Aerobic Marine. Microbiology 110:402–429

75. Deming JW, Somers LK, Straube WL, et al (1988) Isolation of an Obligately

Barophilic Bacterium and Description of a New Genus, Colwellia gen. nov. Syst Appl Microbiol 10:152–160. https://doi.org/10.1016/S0723-2020(88)80030-4

76. Kurahashi M, Yokota A (2007) Endozoicomonas elysicola gen. nov., sp. nov., a γ-proteobacterium isolated from the sea slug Elysia ornata. Syst Appl Microbiol 30:202–206. https://doi.org/10.1016/j.syapm.2006.07.003

77. GAUTHIER MJ, LAFAY B, CHRISTEN BR, et al (1992) Marinobacter hydrocarbonoclasticus gen. nov,, sp. nov., a New, Extremely Halotolerant, Hydrocarbon-Degrading Marine Bacterium. Int J Syst Bacteriol 42:568–576. https://doi.org/10.1373/49.4.644

78. GONZALEZ JM, MAYER F, MORAN MA, et al (2009) Microbulbifer hydrolyticus gen. nov., sp. nov., and Marinobacterium georgiense gen. nov., sp. nov., Two Marine Bacteria from a Lignin-Rich Pulp Mill Waste Enrichment Community. Int J Syst Bacteriol 47:369–376. https://doi.org/10.1099/00207713-47-2-369

79. LANDSCHOOT A VAN, LEY J DE (1983) Intra- and Intergeneric Similarities of the rRNA Cistrons of Alteromonas, Murimmonas (gen. nov.) and Some Other Gram-negative Bacteria. J Gen Microbiol 129:3057–3074

80. Gauthier G, Gauthier M, R C (1995) Phylogenetic Analysis of the Genera Alteromonas, Shewanella, and Moritella Using Genes Coding for Small-Subunit rRNA Sequences and Division of the Genus Alteromonas into Two Genera, Alteromonas (Emended) and Pseudoalteromonas gen. nov., and Proposal of Tw. Int J Syst Bacteriol 45:755–761

81. Shieh WY, Lin Y Te, Jean WD (2004) Pseudovibrio denitrificans gen. nov., sp. nov., a marine, facultatively anaerobic, fermentative bacterium capable of denitrification. Int J Syst Evol Microbiol 54:2307–2312.

https://doi.org/10.1099/ijs.0.63107-0

82. Shiba T (1991) Roseobacter litoralis gen. nov., sp. nov., and Roseobacter denitrificans sp. nov., Aerobic Pink-Pigmented Bacteria which Contain Bacteriochlorophyll a. Syst Appl Microbiol 14:140–145. https://doi.org/10.1016/S0723-2020(11)80292-4

83. Uchino Y, Hirata A, Yokota A, Sugiyama J (2005) Reclassification of marine Agrobacterium species: Proposals of Stappia stellulata gen. nov., comb. nov., Stappia aggregata sp. nov., nom. rev., Ruegeria atlantica gen. nov., comb. nov., Ruegeria gelatinovora comb. nov., Ruegeria algicola comb. nov., and A. J Gen Appl Microbiol 44:201–210. https://doi.org/10.2323/jgam.44.201

84. MacDonell MT, Colwell RR (1985) Phylogeny of the Vibrionaceae, and Recommendation for Two New Genera, Listonella and Shewanella. Syst Appl Microbiol 6:171–182. https://doi.org/10.1016/S0723-2020(85)80051-5

85. Maldonado LA, Fenical W, Jensen PR, et al (2005) Salinispora arenicola gen. nov., sp. nov. and Salinispora tropica sp. nov., obligate marine actinomycetes belonging to the family Micromonosporaceae. Int J Syst Evol Microbiol 55:1759–1766. https://doi.org/10.1099/ijs.0.63625-0

86. Waksman SA, Henrici AT (1895) The nomenclature and classification of the actinomycetes. J Am Med Assoc XXIV:452. https://doi.org/10.1001/jama.1895.02430120030007

87. Macia MC, Ludwig W, Schleifer KH, et al (2001) Thalassomonas viridans gen . nov ., sp . nov ., a novel marine γ -proteobacterium. Int J Syst Evol Microbiol 51:1283–1289

88. Pacini F (1854) Osservazione microscopiche e deduzioni patologiche sul cholera asiatico. Gaz Medicale Ital Toscano Firenze 6:405–4012

89. Inagaki F, Takai K, Nealson KH, Horikoshi K (2004) Sulfurovum lithotrophicum gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the E-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. Int J Syst Evol Microbiol 54:1477–1482. https://doi.org/10.1099/ijs.0.03042-0

90. Inagaki F, Takai K, Kobayashi H, et al (2003) Sulfurimonas autrotrophica gen. nov., sp. nov., a novel sulfur-oxidizing E-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. Int J Syst Evol Microbiol 53:1801–1805. https://doi.org/10.1099/ijs.0.02682-0

91. Luo H, Moran MA (2014) Evolutionary Ecology of the Marine Roseobacter Clade. Microbiol Mol Biol Rev 78:573–587. https://doi.org/10.1128/MMBR.00020-14

92. Barka EA, Vatsa P, Sanchez L, et al (2016) Taxonomy, Physiology, and Natural Products of Actinobacteria. Am Soc Microbiol 80:1–43. https://doi.org/10.1128/MMBR.00019-15.Address

93. Clark K, Karsch-Mizrachi I, Lipman DJ, et al (2015) GenBank. Nucleic Acids Res 44:D67----D72

94. Hyatt D, Chen G-L, LoCascio PF, et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119

95. Parks DH, Imelfort M, Skennerton CT, et al (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055

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

97. Camacho C, Coulouris G, Avagyan V, et al (2009) BLAST+: architecture and

applications. BMC Bioinformatics 10:421

98. Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol. https://doi.org/10.1007/978-1-59745-251-9_3

99. Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One. https://doi.org/10.1371/journal.pone.0009490

100. Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol. https://doi.org/10.1093/molbev/msw046

101. Lee MD, Ponty Y (2019) GToTree: A user-friendly workflow for phylogenomics. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz188

102. Eddy SR (2011) Accelerated profile HMM searches. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1002195

103. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. https://doi.org/10.1093/nar/gkh340

104. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. https://doi.org/10.1093/bioinformatics/btp348

105. Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. https://doi.org/10.1093/molbev/mst197

106. Tange O (2011) GNU Parallel: the command-line power tool. ;login USENIX Mag. https://doi.org/10.5281/zenodo.16303

107. Altschul SF, Madden TL, Schäffer AA, et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids

Res 25:3389–3402

108. Hildebrand F, Meyer A, Eyre-Walker A (2010) Evidence of selection upon genomic GC-content in bacteria. PLoS Genet 6:e1001107

109. Hershberg R, Petrov DA (2010) Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet 6:e1001115

110. Schirrmeister BE, Antonelli A, Bagheri HC (2011) The origin of multicellularity in cyanobacteria. BMC Evol Biol 11:45

111. Di Rienzi SC, Sharon I, Wrighton KC, et al (2013) The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. Elife 2:e01102

112. Soo RM, Skennerton CT, Sekiguchi Y, et al (2014) An expanded genomic representation of the phylum Cyanobacteria. Genome Biol Evol 6:1031–1045

113. Soo RM (2015) In search of non-photosynthetic Cyanobacteria. Univ Queensl

114. Kauff F, Büdel B (2010) Phylogeny of cyanobacteria: an overview. In: Progress in Botany 72. Springer, pp 209–224

115. STANIER RY, SISTROM WR, HANSEN TA, et al (1978) Proposal to Place the Nomenclature of the Cyanobacteria (Blue-Green Algae) Under the Rules of the International Code of Nomenclature of Bacteria. Int J Syst Evol Microbiol 28:335–336. https://doi.org/10.1099/00207713-28-2-335

116. Oren A (2004) A proposal for further integration of the cyanobacteria under the Bacteriological Code. Int J Syst Evol Microbiol 54:1895–1902. https://doi.org/10.1099/ijs.0.03008-0

117. Oren A, Garrity GM (2014) Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. Antonie Van Leeuwenhoek 106:43–56

118. Tindall BJ (1999) Note: Proposals to update and make changes to the Bacteriological Code. Int J Syst Evol Microbiol 49:1309–1312

119. De Vos P, Trüper HG (2000) Judicial commission of the international committee on systematic bacteriology; IXth international (IUMS) congress of bacteriology and applied microbiology. Int J Syst Evol Microbiol 50:2239–2244

120. Labeda DP (2000) International committee on systematic bacteriology; IXth international (IUMS) congress of bacteriology and applied microbiology. Int J Syst Evol Microbiol 50:2245–2247

121. Hoffmann L, Komárek J, Kaštovsky J (2005) System of cyanoprokaryotes (cyanobacteria)--state in 2004. Arch Hydrobiol Suppl Algol Stud 117:95–115

122. Oren A, Tindall BJ (2005) Nomenclature of the cyanophyta/cyanobacteria/cyanoprokaryotes under the International Code of Nomenclature of Prokaryotes. Arch Hydrobiol Suppl Algol Stud 117:39–52

123. Oren A, Komárek J, Hoffmann L (2009) Nomenclature of the Cyanophyta/Cyanobacteria/Cyanoprokaryotes--what has happened since IAC Luxembourg? Arch Hydrobiol Suppl Algol Stud 130:17–26

124. Oren A, Ventura S (2017) The current status of cyanobacterial nomenclature under the “prokaryotic” and the “botanical” code. Antonie Van Leeuwenhoek 110:1257–1269

125. Parte AC (2014) LPSN--list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 42:D613--6. https://doi.org/10.1093/nar/gkt1111

126. 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. I. Syst

Appl Microbiol 38:141–158

127. Gaget V, Welker M, Rippka R, de Marsac Tandeau N (2015) Response to:" Comments on:" 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, Planktothri. Syst Appl Microbiol 38:368–370

128. Boone DR, Castenholz RW (2001) The Archaea and the Deeply Branching and Phototrophic Bacteria. In: Garrity G, Boone DR, Castenholz RW (eds) Bergey's manual of systematic bacteriology volume one. Springer-Verlag, New York, pp 473–487

129. Rippka R, Deruelles J, Waterbury JB, et al (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology 111:1–61

130. Gugger MF, Hoffmann L (2004) Polyphyly of true branching cyanobacteria (Stigonematales). Int J Syst Evol Microbiol 54:349–357

131. Hugenholtz P, Skarshewski A, Parks DH (2016) Genome-based microbial taxonomy coming of age. Cold Spring Harb Perspect Biol 8:a018085

132. Shih PM, Wu D, Latifi A, et al (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci 110:1053–1058. https://doi.org/10.1073/pnas.1217107110

133. Komarek J, Kastovsky J, Mares J, Johansen JR (2014) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014 , using a polyphasic approach. Preslia 86:295–335

134. Kozlov AM, Zhang J, Yilmaz P, et al (2016) Phylogeny-aware identification and correction of taxonomically mislabeled sequences. Nucleic Acids Res 44:5022–

5033

135. Choudoir MJ, Campbell AN, Buckley DH (2012) Grappling with Proteus: population level approaches to understanding microbial diversity. Front Microbiol 3:336

136. Becraft ED, Wood JM, Rusch DB, et al (2015) The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of Synechococcus inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. Front Microbiol 6:590. https://doi.org/10.3389/fmicb.2015.00590

137. Farrant GK, Doré H, Cornejo-Castillo FM, et al (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. Proc Natl Acad Sci 113:E3365--E3374. https://doi.org/10.1073/pnas.1524865113

138. Dufresne A, Salanoubat M, Partensky F, et al (2003) Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. Proc Natl Acad Sci 100:10020–10025

139. Kettler GC, Martiny AC, Huang K, et al (2007) Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. PLoS Genet 3:2515–2528. https://doi.org/10.1371/journal.pgen.0030231

140. Zinser ER, Lindell D, Johnson ZI, et al (2009) Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, Prochlorococcus. PLoS One 4:e5135

141. García-Fernandez JM, Marsac NT, Diez J (2004) Streamlined Regulation and Gene Loss as Adaptive Mechanisms in Prochlorococcus for Optimized Nitrogen Utilization in Oligotrophic Environments. Microbiol Mol Biol Rev 68:630–638. https://doi.org/10.1128/MMBR.68.4.630

142. Chisholm SW, Frankel SL, Goericke R, et al (1992) Prochlorococcus marinus nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. Arch Microbiol 157:297–300

143. Voigt K, Sharma CM, Mitschke J, et al (2014) Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. ISME J 8:2056–2068. https://doi.org/10.1038/ismej.2014.57

144. Giovannoni SJ, Halsey KH, Saw J, et al (2019) A Parasitic Arsenic Cycle That Shuttles Energy from Phytoplankton to Heterotrophic Bacterioplankton. MBio 10.

145. Saunders JK, Rocap G (2016) Genomic potential for arsenic efflux and methylation varies among global Prochlorococcus populations. ISME J 10:197–209. https://doi.org/10.1038/ismej.2015.85

Figures

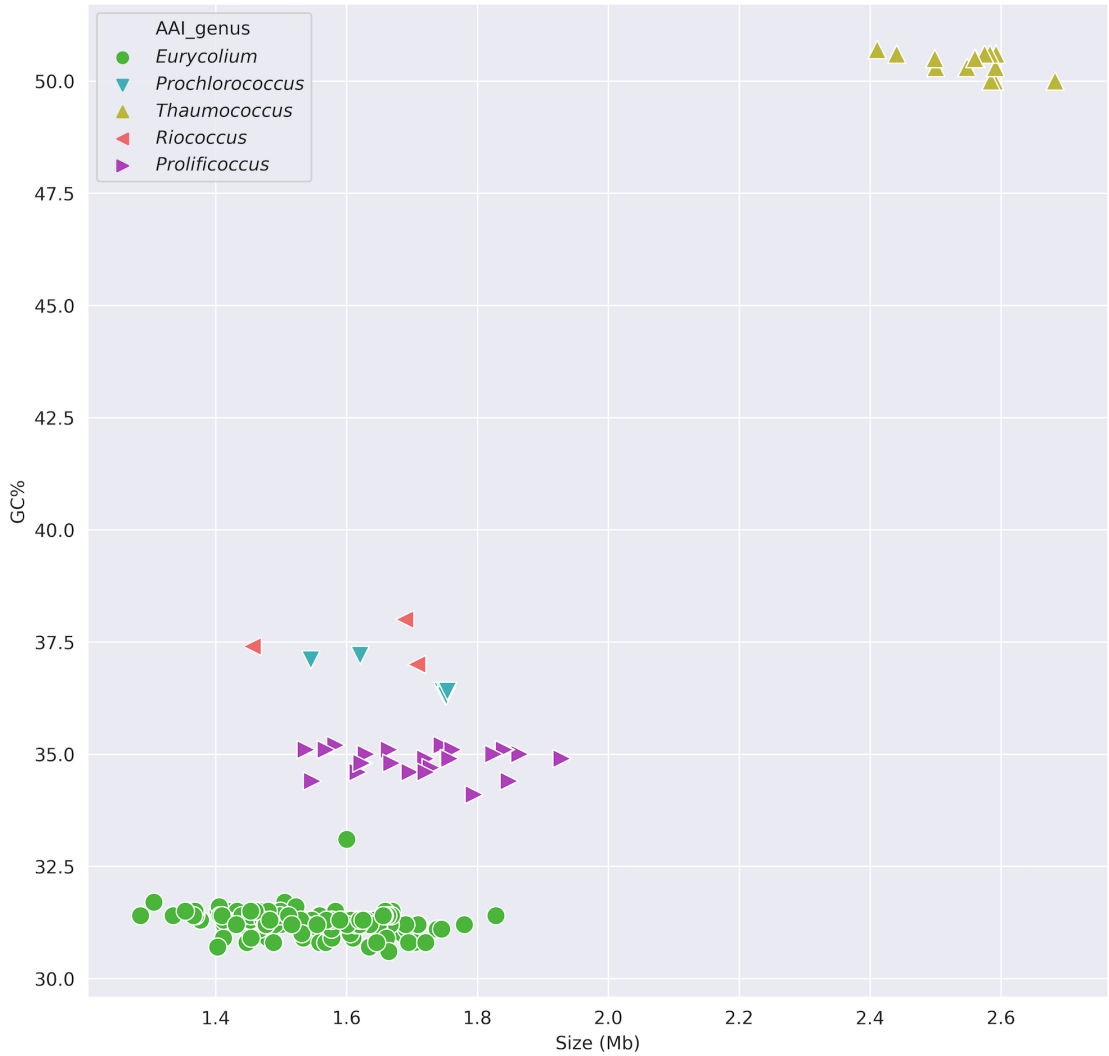

Figure 1. Genome clustering in five groups based on and GC content and genome size. The colors (green, red, blue, purple and yellow) represent the five groups. The five different symbols represent the genera where each genome were classified.

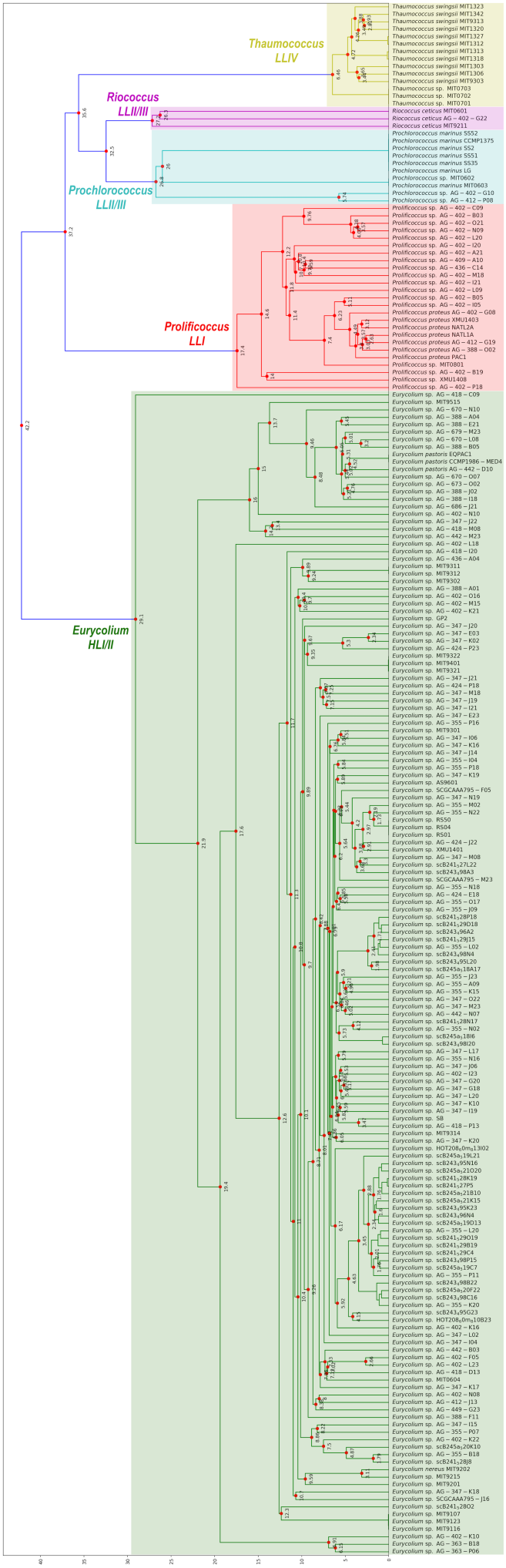
Thaumococcus
LLIV
Riococcus
LLII/III
Prochlorococcus
LLII/III
Prolificoccus
LLI
Eurycolium
HLI/II
Thaumococcus swingsii MIT1323
Thaumococcus swingsii MIT1342
Thaumococcus swingsii MIT9313
Thaumococcus swingsii MIT1320
Thaumococcus swingsii MIT1327
Thaumococcus swingsii MIT1312
Thaumococcus swingsii MIT1313
Thaumococcus swingsii MIT1318
Thaumococcus swingsii MIT1303
Thaumococcus swingsii MIT1306
Thaumococcus swingsii MIT9303
Thaumococcus sp. MIT0703
Thaumococcus sp. MIT0702
Thaumococcus sp. MIT0701
Riococcus ceticus MIT0601
Riococcus ceticus AG – 402 – G22
Riococcus ceticus MIT9211
Prochlorococcus marinus SS52
Prochlorococcus marinus CCMP1375
Prochlorococcus marinus SS2
Prochlorococcus marinus SS51
Prochlorococcus marinus SS35
Prochlorococcus marinus LG
Prochlorococcus sp. MIT0602
Prochlorococcus marinus MIT0603
Prochlorococcus sp. AG – 402 – G10
Prochlorococcus sp. AG – 412 – P08
Prolificoccus sp. AG – 402 – C09
Prolificoccus sp. AG – 402 – B03
Prolificoccus sp. AG – 402 – O21
Prolificoccus sp. AG – 402 – N09
Prolificoccus sp. AG – 402 – L20
Prolificoccus sp. AG – 402 – I20
Prolificoccus sp. AG – 402 – A21
Prolificoccus sp. AG – 409 – A10
Prolificoccus sp. AG – 436 – C14
Prolificoccus sp. AG – 402 – M18
Prolificoccus sp. AG – 402 – I21
Prolificoccus sp. AG – 402 – L09
Prolificoccus sp. AG – 402 – B05
Prolificoccus sp. AG – 402 – I05
Prolificoccus proteus AG – 402 – G08
Prolificoccus proteus XMU1403
Prolificoccus proteus NATL2A
Prolificoccus proteus NATL1A
Prolificoccus proteus AG – 412 – G19
Prolificoccus proteus AG – 388 – O02
Prolificoccus proteus PAC1
Prolificoccus sp. MIT0801
Prolificoccus sp. AG – 402 – B19
Prolificoccus sp. XMU1408
Prolificoccus sp. AG – 402 – P18
Eurycolium sp. AG – 418 – C09
Eurycolium sp. MIT9515
Eurycolium sp. AG – 670 – N10
Eurycolium sp. AG – 388 – A04
Eurycolium sp. AG – 388 – E21
Eurycolium sp. AG – 679 – M23
Eurycolium sp. AG – 670 – L08
Eurycolium sp. AG – 388 – B05
Eurycolium pastoris EOPAC1
Eurycolium pastoris CCMP1986 – MED4
Eurycolium pastoris AG – 442 – D10
Eurycolium sp. AG – 670 – O07
Eurycolium sp. AG – 673 – O02
Eurycolium sp. AG – 388 – J02
Eurycolium sp. AG – 388 – I18
Eurycolium sp. AG – 686 – J21
Eurycolium sp. AG – 402 – N10
Eurycolium sp. AG – 347 – J22
Eurycolium sp. AG – 418 – M08
Eurycolium sp. AG – 442 – M23
Eurycolium sp. AG – 402 – L18
Eurycolium sp. AG – 418 – I20
Eurycolium sp. AG – 436 – A04
Eurycolium sp. MIT9311
Eurycolium sp. MIT9312
Eurycolium sp. MIT9302
Eurycolium sp. AG – 388 – A01
Eurycolium sp. AG – 402 – O16
Eurycolium sp. AG – 402 – M15
Eurycolium sp. AG – 402 – K21
Eurycolium sp. GP2
Eurycolium sp. AG – 347 – J20
Eurycolium sp. AG – 347 – E03
Eurycolium sp. AG – 347 – K02
Eurycolium sp. AG – 424 – P23
Eurycolium sp. MIT9322
Eurycolium sp. MIT9401
Eurycolium sp. MIT9321
Eurycolium sp. AG – 347 – J21
Eurycolium sp. AG – 424 – P18
Eurycolium sp. AG – 347 – M18
Eurycolium sp. AG – 347 – J19
Eurycolium sp. AG – 347 – I21
Eurycolium sp. AG – 347 – E23
Eurycolium sp. AG – 355 – P16
Eurycolium sp. MIT9301
Eurycolium sp. AG – 347 – I06
Eurycolium sp. AG – 347 – K16
Eurycolium sp. AG – 347 – J14
Eurycolium sp. AG – 355 – I04
Eurycolium sp. AG – 355 – P18
Eurycolium sp. AG – 347 – K19
Eurycolium sp. AS9601
Eurycolium sp. SCGCAAA795 – F05
Eurycolium sp. AG – 347 – N19
Eurycolium sp. AG – 355 – M02
Eurycolium sp. AG – 355 – N22
Eurycolium sp. RS50
Eurycolium sp. RS04
Eurycolium sp. RS01
Eurycolium sp. AG – 424 – J22
Eurycolium sp. XMU1401
Eurycolium sp. AG – 347 – M08
Eurycolium sp. scB241$_2$7L22
Eurycolium sp. scB243$_1$98A3
Eurycolium sp. SCGCAAA795 – M23
Eurycolium sp. AG – 355 – N18
Eurycolium sp. AG – 424 – E18
Eurycolium sp. AG – 355 – O17
Eurycolium sp. AG – 355 – J09
Eurycolium sp. scB241$_2$8P18
Eurycolium sp. scB241$_2$9O18
Eurycolium sp. scB243$_1$96A2
Eurycolium sp. scB241$_2$9J15
Eurycolium sp. AG – 355 – L02
Eurycolium sp. scB243$_1$98N4
Eurycolium sp. scB243$_1$95L20
Eurycolium sp. scB245a$_1$18A17
Eurycolium sp. AG – 355 – J23
Eurycolium sp. AG – 355 – A09
Eurycolium sp. AG – 355 – K15
Eurycolium sp. AG – 347 – O22
Eurycolium sp. AG – 347 – M23
Eurycolium sp. AG – 442 – N07
Eurycolium sp. scB241$_2$8N17
Eurycolium sp. AG – 355 – N02
Eurycolium sp. scB245a$_1$18I6
Eurycolium sp. scB243$_1$98I20
Eurycolium sp. AG – 347 – L17
Eurycolium sp. AG – 355 – N16
Eurycolium sp. AG – 347 – J06
Eurycolium sp. AG – 402 – I23
Eurycolium sp. AG – 347 – G20
Eurycolium sp. AG – 347 – G18
Eurycolium sp. AG – 347 – L20
Eurycolium sp. AG – 347 – K10
Eurycolium sp. AG – 347 – I19
Eurycolium sp. S8
Eurycolium sp. AG – 418 – P13
Eurycolium sp. MIT9314
Eurycolium sp. AG – 347 – K20
Eurycolium sp. HOT208$_0$0m$_1$3I02
Eurycolium sp. scB245a$_1$19L21
Eurycolium sp. scB243$_1$95N16
Eurycolium sp. scB245a$_2$21O20
Eurycolium sp. scB241$_2$8K19
Eurycolium sp. scB241$_2$7P5
Eurycolium sp. scB245a$_2$21B10
Eurycolium sp. scB245a$_2$21K15
Eurycolium sp. scB243$_1$95K23
Eurycolium sp. scB243$_1$96N4
Eurycolium sp. scB245a$_1$19D13
Eurycolium sp. AG – 355 – L20
Eurycolium sp. scB241$_2$9O19
Eurycolium sp. scB241$_2$9B19
Eurycolium sp. scB241$_2$9C4
Eurycolium sp. scB243$_1$98P15
Eurycolium sp. scB245a$_1$19C7
Eurycolium sp. AG – 355 – P11
Eurycolium sp. scB243$_1$98B22
Eurycolium sp. scB245a$_2$20F22
Eurycolium sp. scB243$_1$98C16
Eurycolium sp. AG – 355 – K20
Eurycolium sp. scB243$_1$95G23
Eurycolium sp. HOT208$_0$0m$_1$10B23
Eurycolium sp. AG – 402 – K16
Eurycolium sp. AG – 347 – L02
Eurycolium sp. AG – 347 – I04
Eurycolium sp. AG – 442 – B03
Eurycolium sp. AG – 402 – F05
Eurycolium sp. AG – 402 – L23
Eurycolium sp. AG – 418 – D13
Eurycolium sp. MIT0604
Eurycolium sp. AG – 347 – K17
Eurycolium sp. AG – 402 – N08
Eurycolium sp. AG – 412 – J13
Eurycolium sp. AG – 449 – G23
Eurycolium sp. AG – 388 – F11
Eurycolium sp. AG – 347 – I15
Eurycolium sp. AG – 355 – P07
Eurycolium sp. AG – 402 – K22
Eurycolium sp. scB245a$_2$20K10
Eurycolium sp. AG – 355 – B18
Eurycolium sp. scB241$_2$8J8
Eurycolium nereus MIT9202
Eurycolium sp. MIT9215
Eurycolium sp. MIT9201
Eurycolium sp. AG – 347 – K18
Eurycolium sp. SCGCAAA795 – J16
Eurycolium sp. scB241$_2$8O2
Eurycolium sp. MIT9107
Eurycolium sp. MIT9123
Eurycolium sp. MIT9116
Eurycolium sp. AG – 402 – K10
Eurycolium sp. AG – 363 – B18
Eurycolium sp. AG – 363 – P06
40
35
30
25
20
15
10
5
0

Figure 2. Hierarchical clustering based on average amino acid identity (AAI) analysis of the 208 PC genomes. The intraspecies limit is assumed as ≥95%, whereas genera delimitation is assumed as ≥70%. The colors (green, red, blue, purple and yellow) represent the five genera.

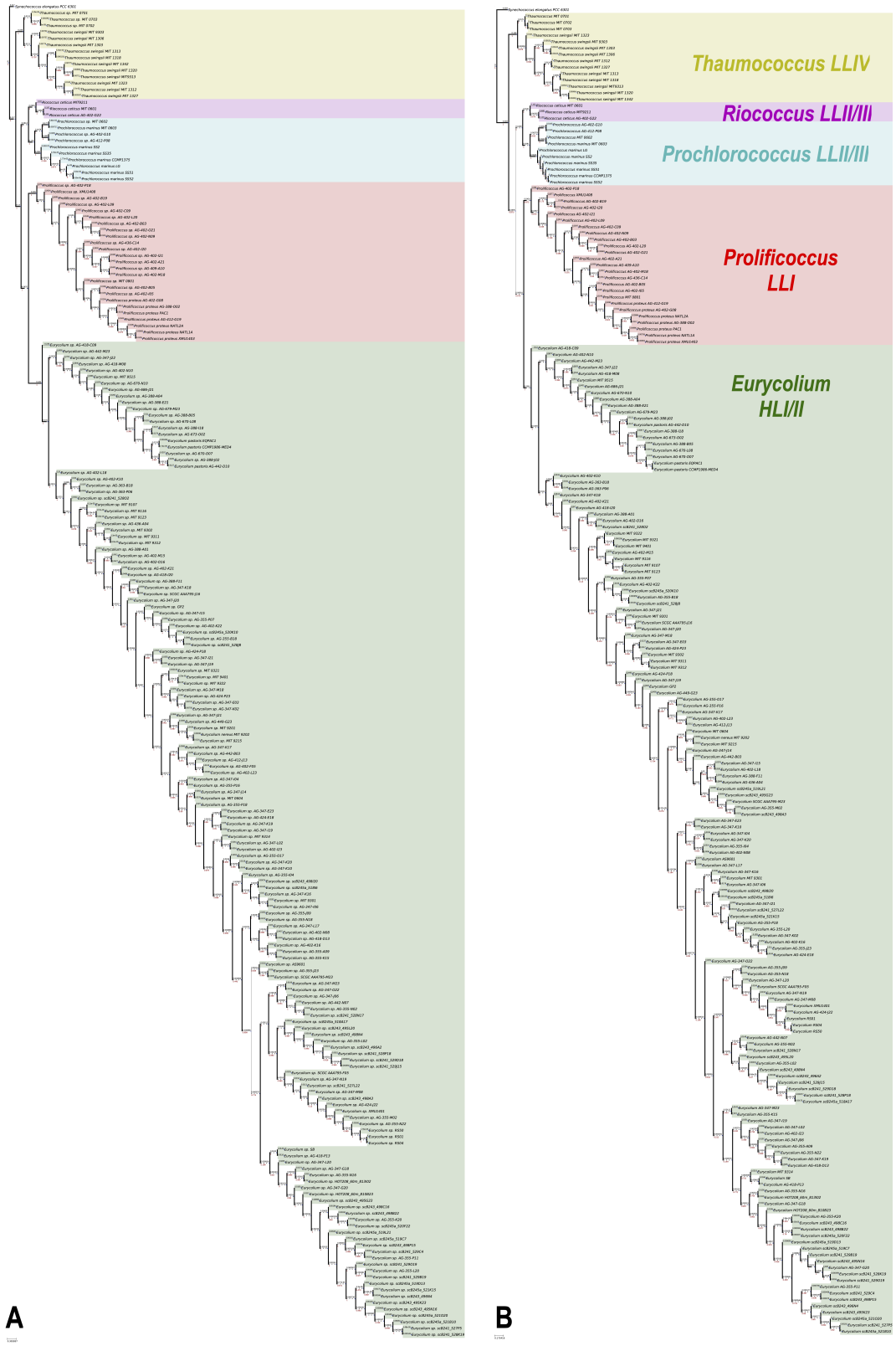

Figure 3. Phylogenetic trees using (A) 249 core protein sequences and (B) four house-keeping genes (*gyrB, pyrH, recA, rpoB*) based on Maximum Likelihood method.

Bootstrap test was conducted using 1,000 replicates. The tree outgroup is *Synechococcus elongatus* PCC 6301.

Tables

Genus (No. of species)	No. of genomes	Type species	Type strain	Range % mol GC	Average length (Mbp) [a]
Eurycolium (97)	156	*Eurycolium pastoris* •	CCMP1986[T]	30-33	1.5±0.1
Prochlorococcus (04)	10	*Prochlorococcus marinus*	CCMP 1375[T]	36-37	1.7±0.07
Prolificoccus (16)	25	*Prolificoccus proteus*•	NATL2A[T]	34-55	1.7±0.1
Thaumococcus (02)	14	*Thaumococcus swingsii*•	MIT 9313[T]	50-50.7	2.5±0.06
Riococcus gen. nov. (03)	03	*Riococcus ceticus* sp. nov.	MIT9211 [T]	37-38	1.6±0.1

Table 1. Proposal of a new PC taxonomic classification. t• New taxonomic classification proposed by Walter et al. (2017).

a: Average genome size was calculated using all genomes from each genus. The standard deviation was represented by ± symbol.

Genus	Ecotype[a]	*narB*	*nirA*	*ntcA*	*glnB*	*amt*	*cynS*	*ureA*	*glnA*	*glsF*	*gdhA*	*phoB*	*ptrA*	*phoR*	*arsC*	*piuC*	*som*
Eurycolium	HLI/II/VI	v	v	+	v	+	v	v	+	+	v	v	-	-	v	-	+
Prochlorococcus	LLII/III	-	-	v	v	v	-	-	v	v	-	-	-	-	v	v	v
Prolificoccus	LLI	v	v	v	v	v	v	v	v	v	-	-	-	-	v	v	v
Thaumococcus.	LLIV	-	v	v	v	v	-	v	v	v	v	-	v	v	v	v	v
Riococcus	LLII/III	-	-	v	v	v	-	-	v	v	-	-	-	-	-	v	v

Table 2. Predicted phenotype of new proposed genera of PC.

"+" Presence, "-" absent, "V" variable. Nitrogen-related genes: narB, nirA, ntcA, glnB, amt, cynS, ureA, glnA, glsF, gdhA. Phosphorous-related genes: phoB, ptrA, phoR. Arsenate-related gene: arsC. Iron-related genes: piuC, som.

a: Data extracted from Berube et al (2018,2019).