Ecological and genomic features of two widespread freshwater picocyanobacteria

Pedro J. Cabello-Yeves[1*], Antonio Picazo[2], Antonio Camacho[2], Cristiana Callieri[3], Riccardo Rosselli[1], Juan J. Roda-Garcia[1], Felipe H. Coutinho[1], Francisco Rodriguez-Valera[1]

[1] Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, Campus de San Juan, San Juan de Alicante, 03550 Alicante, Spain. [2] Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, E-46980 Paterna, Valencia, Spain. [3] Institute of Ecosystem Study – CNR, Largo Tonolli 50, 28922 Verbania, Italy

*Correspondence: Pedro J. Cabello-Yeves, E-mail: pedrito91vlc@gmail.com

Running title: Genomes of widespread freshwater picocyanobacteria

Summary:

We present two genomes of widespread freshwater picocyanobacteria isolated by extinction dilution from a Spanish oligotrophic reservoir. Based on microscopy and genomic properties, both picocyanobacteria were tentatively designated *Synechococcus lacustris* Tous, formerly described as a metagenome assembled genome (MAG) from the same habitat, and *Cyanobium usitatum* Tous, described here for the first time. Both strains were purified in unicyanobacterial cultures, and their genomes were sequenced. They are broadly distributed in freshwater systems; the first seems to be a specialist on temperate reservoirs (Tous, Amadorio, Dexter, Lake Lanier, Sparkling), and the second appears to also be abundant in cold environments including ice-covered lakes such as Lake Baikal, Lake Erie or the brackish Baltic Sea. Having complete genomes provided access to the flexible genome that does not assemble in MAGs. We found several genomic islands in both genomes, within which there were genes for nitrogen acquisition, transporters for a wide set of compounds and biosynthesis of phycobilisomes in both strains. Some of these regions of low coverage in metagenomes also included antimicrobial compounds, transposases and phage defence systems, including a novel type III CRISPR-Cas phage defence system that was only detected in *Synechococcus lacustris* Tous.

Originality and significance:

Genomic and ecological features of freshwater picocyanobacteria from the genera *Synechococcus* and *Cyanobium* remain obscure compared to the well-studied marine representatives. A particular knowledge gap is the biogeographic distribution of freshwater picocyanobacteria across different lakes, reservoirs and brackish systems. There is also limited information on the genomic repertoire, pan-genome and genomic islands present in freshwater representatives. Culture-based genomics, genetic and metagenomic approaches are all well-recognized methods for the study of picocyanobacteria, but culture provides more insightful information on the flexible genome of different strains, revealing features involved in niche occupancy across different ecosystems. To fill this gap in knowledge, we isolated by extinction dilution and subsequently sequenced two picocyanobacteria from an oligotrophic

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.14377

Spanish reservoir. Both genomes were retrieved from unicyanobacterial cultures. Through metagenomics, these strains were found to be widespread in various freshwater and brackish systems, colonizing different geographic areas. Different phage defence systems and genes involved in niche colonization were encoded in genomic islands. This study sheds light on broadly distributed freshwater picocyanobacteria at the genomic and ecological levels and opens up new perspectives on the study of these poorly understood freshwater counterparts.

Introduction:

Picocyanobacteria are, along with the other members of the cyanobacterial phylum, the only prokaryotes able to perform oxygenic photosynthesis, and are of paramount importance to the primary production of marine, brackish and freshwater ecosystems (Stockner, 1988; Stockner *et al.*, 2000; Callieri and Stockner, 2002; Scanlan and West, 2002; Camacho *et al.*, 2003). Freshwater picocyanobacteria include the genera *Cyanobium* and *Synechococcus* as the most representative members of the picocyanobacterial clade (Sanchez-Baracaldo *et al.*, 2005). Traditionally they have been described as non-bloom formers, non-toxic and present either as single or colonial cocci or rod-shaped cells in the < 3 µm size range (Callieri *et al.*, 2012). They have also evolved strategies to adapt to various environmental conditions allowing them to proliferate successfully from the poles to the tropics (Callieri, 2017). In the oceans, microdiverse *Synechococcus* lineages inhabit different regions of the globe across large spatial scales (Zwirglmaier *et al.*, 2008; Farrant *et al.*, 2016), allowing this genus to occupy virtually all ocean waters. One important picocyanobacterial phenotypic characteristic is their accessory pigment composition (presence of various combinations of phycoerythrin and phycocyanin in their phycobilisomes, hereafter PBS), which makes them efficient light harvesters, photosynthesizing at different depths and across a range of light intensities (Vörös *et al.*, 1998; Stomp *et al.*, 2007; Grébert *et al.*, 2018).

Genetic studies based on the 16S rRNA and cytochrome b6 (*petB*) genes have revealed the presence of two major marine sub-clusters 5.1A/B (Scanlan *et al.*, 2009; Mazard *et al.*, 2012), whilst approximately 13 non-marine sub-clusters of freshwater, brackish and euryhaline origin have been detected (Callieri *et al.*, 2013). While sub-clusters 5.1A/B are restricted to marine representatives (except for the halotolerant clade VIII), sub-clusters 5.2 and 5.3 encompass marine, euryhaline and freshwater species (Cabello-Yeves *et al.*, 2017), which widens the previously known polyphyletic nature of picocyanobacteria from the *Synechococcus*/*Cyanobium* genus (Shih *et al.*, 2013). Evolutionary reconstruction models based on phylogenomics, Bayesian relaxed molecular clocks or Bayesian stochastic character mapping analysis revealed a possible transition of *Synechococcus* species from freshwater to brackish and marine environments, which is also consistent with the greater diversity of freshwater species compared to their marine representatives (Blank and Sanchez-Baracaldo, 2010; Sánchez-Baracaldo, 2015).

Complete genome studies on several marine *Synechococcus* strains revealed a core genome containing approximately 1500 protein-encoding genes (Dufresne *et al.*, 2008). The same study shed light on the flexible genome of marine picocyanobacteria, showing genomic islands and hypervariable regions containing genes involved in LPS synthesis, ABC transporters, transcriptional regulators, DNA mobility or phage defence systems (Dufresne *et al.*, 2008). One

of these islands contains the genes for PBS synthesis and regulation, a feature that allows *Synechococcus* to adapt to different light conditions, with some strains being chromatic adaptors (Six *et al.*, 2007) that are ecologically abundant (Grébert *et al.*, 2018). Distribution patterns of different marine *Synechococcus* ecotypes revealed open ocean, coastal and opportunistic specialists (Dufresne *et al.*, 2008; Zwirglmaier *et al.*, 2008; Sohm *et al.*, 2016). Overall, marine *Synechococcus* are well studied at the ecological, taxonomical, phenotypic and genomic levels. However, the enormous diversity of the freshwater genera *Synechococcus* and *Cyanobium* is still unexplored at the genomic level. So far, only *Synechococcus elongatus* (Holtman *et al.*, 2005) and a small number of metagenome assembled genomes (MAGs) and draft genomes from different freshwater lakes and reservoirs have been deposited in databases (Bhaya *et al.*, 2007; Callieri *et al.*, 2013; Shih *et al.*, 2013; Guimarães *et al.*, 2015; Cabello-Yeves *et al.*, 2017, 2018; Di Cesare *et al.*, 2018). The scarcity of metagenomic studies on tropical, temperate, and cold lakes and reservoirs ranging from oligotrophic to eutrophic status, although recently improved, also hampers evaluation of the distribution patterns of freshwater, brackish and euryhaline strains.

Here, we present the ecological and genomic characterization of two picocyanobacterial strains both isolated from Tous reservoir (Spain), an oligotrophic freshwater system. The strains were characterized as *Synechococcus lacustris* Tous (Cabello-Yeves *et al.*, 2017), previously described as an MAG from the same habitat, and the other strain is proposed here as a member of the *Synechococcaceae* family but from a separate genus: *Cyanobium usitatum* Tous*.* Both picocyanobacteria have a coccoid morphology and a similar genome size, but are affiliated with different sub-clusters within the genera *Synechococcus/Cyanobium*. They are geographically widely distributed across different freshwater and brackish ecosystems.

Results and discussion:

General features of the two novel freshwater picocyanobacteria

This work is a continuation of a metagenomics study in which *Ca.* Synechococcus lacustris MAGs from Tous reservoir and Lake Lanier were described (Cabello-Yeves *et al.*, 2017). The predicted pigmentation of *S.lacustris* was pink according to the presence of phycoerythrin-I (PE-I) and phycocyanin (PC) subunits in its PBS operon. This picocyanobacterium was proven to be very abundant in Tous reservoir by FISH and metagenomic fragment recruitment analysis. Therefore, we developed cultures by extinction dilution to retrieve pink cultures from the same habitat. The cultures were screened by PCR of the MAG *rpoB* and *recA* genes (see experimental procedures and supplementary text) and one culture, *Synechococcus lacustris* Tous, was 98.76 % identical (seen by ANI of the whole genome) to the MAG (Table S1). Another culture was fully sequenced and revealed a novel isolate, tentatively named *Cyanobium usitatum* Tous, which also had high identity with small contigs reconstructed from Tous metagenome. Again, primers were developed to identify each microbe (see experimental procedures and supplementary text). Both cultures are unicyanobacterial (containing a single species of picocyanobacterium). The main features of the two sequenced strains are shown in Table 1, and the comparison of the genomes of *S. lacustris* determined from the MAG (Cabello-Yeves *et al.*, 2017) and the isolate is summarized in Table S1. The description of the

novel species is provided in the supplementary text. Both strains were isolated from the same 12 m sample from the oligotrophic Tous reservoir. As depicted in Fig. 1, both isolates show a similar coccoid shaped cell structure and pink pigmentation. A parallel culture of *S. lacustris* Tous lost its phycoerytrin (PE-I, spectrophotometrically measured) and pink pigmentation, developing a light green one (Fig. S1). Both genomes present a similar genome size but they differ in cell size (Table 1). They clearly belong to different lineages within the current genera *Synechococcus*/*Cyanobium* and display a 10 % difference in GC content. Concatenated protein based phylogenomics (Fig. 2) shows that *S. lacustris* Tous falls into a new freshwater branch within sub-cluster 5.3. Average nucleotide identity (ANI) and average amino acid Identity (AAI) values among all 5.3 and 5.2 sub-cluster members and this new freshwater representative are lower than 70 % (Fig. S2). *C. usitatum* Tous affiliates with polyphyletic sub-cluster 5.2, which comprises marine, brackish, euryhaline and freshwater species. This strain forms a branch with closely related *Cyanobium* MAGs (at 93 % ANI) from multiple cold habitats such as the eutrophic Baltic Sea (Hugerth *et al.*, 2015) and ultraoligotrophic and seasonally ice-covered Lake Baikal (Cabello-Yeves *et al.*, 2018). ANI and AAI values observed within members of this branch do not exceed 82 % ANI and 84 % AAI, the closest isolate being the pink-pigmented *Synechococcus* sp. 8F6, obtained from Lake Alchichica (Mexico). All genomes from sub-cluster 5.2 share at least 72 % ANI and 67-70 % AAI (Fig. S1). Marine genomes from sub-cluster 5.3 have ANIs and AAIs ranging from 72-99 %, but the freshwater *S. lacustris* Tous has < 70 % ANI and AAI compared with all members of sub-clusters 5.2 and 5.3.

General genomic and metabolic features

The main metabolic and genomic differences as derived from the genomes of both picocyanobacteria are highlighted in Table 2. Both strains use urea, ammonia and nitrate/nitrite as their main N sources, as we detected transporters for all of these compounds. *S. lacustris* Tous contains some metabolic features absent in *C. usitatum*, such as its ability to hydrolyse cyanate through cyanate hydratase or to acquire thiosulfate via the CysAUW transporter present in some freshwater cyanobacteria but absent from marine representatives (Laudenbach and Grossman, 1991; Cabello-Yeves *et al.*, 2017). The presence of a type II pink pigmentation and a PBS operon organization resembling the marine strain *Synechococcus* sp. WH7805 isolated from the Sargasso Sea (Six *et al.*, 2007; Dufresne *et al.*, 2008) are also distinguishing features of this strain (Fig. S3). As mentioned above, one of our cultures of *S. lacustris* Tous lost PE-I and pink pigmentation and displayed a light green colour instead. In contrast, *C. usitatum* presents type IIB pink pigmentation (Fig. S3), first observed in Baltic Sea picocyanobacteria (Larsson *et al.*, 2014) and later confirmed in several previously characterized freshwater strains from different origins (Callieri *et al.*, 2013). The presence of multiple copies of genes encoding phycocyanin subunits observed in some of the genomes containing type IIB pigmentation (including *C. usitatum* Tous) may lead to unique light absorption properties and different light adaptations of these strains. Another interesting feature found in the *C. usitatum* genome is hydrogen synthesis and the apparent regulation of its redox state inside the cell by a bidirectional hydrogenase (HoxHYUFE) and NiFe hydrogenase maturation factor (*hyp*), functions that have been previously observed in other cyanobacteria (Tamagnini *et al.*, 2002; Ludwig *et al.*, 2006; Di Cesare *et al.*, 2018). The abovementioned cyanate hydratase, the PBS clusters and the hydrogenases are all present in genomic islands (see below).

With regard to antimicrobial compounds, it must be noted that different systems such as nif-11 leader peptides, SunT bacteriocin/lantibiotic exporters or a NHLP bacteriocin secretion system protein were detected in both genomes, as previously seen in marine representatives (Scanlan *et al.*, 2009; Wang *et al.*, 2011). These genes are also associated with genomic islands. *C. usitatum* contains a type I polyketide synthase which could confer special properties to this strain leading to niche occupation and competition versus other bacteria or avoidance of grazing by other organisms such as ciliates (Christaki *et al.*, 1999; Scanlan *et al.*, 2009). Overall, we detected more transporters in *C. usitatum* than in *S.lacustris* (160 versus 140).

Phages are one of the major evolutionary driving forces for microorganisms on Earth. In oligotrophic and open ocean environments dominated by *Synechococcus* and *Prochlorococcus* constant infection by phages occurs (Suttle and Chan, 1993; Sullivan *et al.*, 2003). However, the absence of CRISPR Cas systems is notable in the vast majority of marine picocyanobacterial genomes sequenced thus far, except for *Synechococcus* sp. WH8020 and WH8016, which are the only (marine) ecotypes from the *Pro/Syn* subclade displaying them. In contrast, most of the remaining cyanobacteria, from filamentous to unicellular and bloom-forming, independently of their origin, contain CRISPR Cas systems (Cai *et al.*, 2013). Apparently, streamlined and relatively small genomes from the *Pro/Syn* subclade tend to contain other phage defence strategies such as modifications on the LPS and the outer membrane, toxin-antitoxin or restriction modification systems, all of which involve lower genetic load (Cai *et al.*, 2013). Moreover, CRISPR Cas systems had not yet been seen in their freshwater picocyanobacterial counterparts. However, in the case of *S. lacustris* Tous, a type III CRISPR Cas system was detected (Fig. S4). We found two clusters of 8 and 18 CRISPR repeats in this genome, together with 3 Csx1 proteins (type III-U system), Cas7/Cmr4, Cas5/Cmr3, Cas10/Cmr2, Cas1, Cas2 and another RAMP superfamily protein. It seems that other DNA/RNA metabolism (DNA methylase or DNA/RNA helicase, nuclease) and phage defence proteins (abortive infection phage resistance, AIPR) tend to accumulate in the proximity of the CRISPR Cas system of this strain. Most of these proteins show their highest similarities with the abovementioned marine *Synechococcus* WH8020 and WH8016, but it seems that this is a novel form of type III CRISPR Cas system. We observed prokaryotic defence system genes (mainly toxin-antitoxin, restriction and modification and DNA phosphothiolation systems) in both genomes, but *S. lacustris* Tous exhibited a higher number compared to *C. usitatum* Tous (50 versus 28). Paradoxically we noted the presence of several cyanophage contigs in *S. lacustris* Tous, which were tested for host-phage recognition homology with the CRISPR system of this picocyanobacterium (Fig. S4). At least 4 phage contigs gave positive matches to CRISPR spacers at very high identities, which make us hypothesize that they are infecting relatives of this genome which were not protected. It is possible that other clones of the same species but devoid of a CRISPR system were present in the enrichment culture at the beginning and were removed by phage infection. These viral scaffolds are also abundant in the different small fraction (0.1 – 5 µm) metagenomic datasets from Tous (at > 20 RPKG values, reaching up to 90 RPKG), which fits with them being abundant phages from a widespread *Synechococcus* ecotype.

Distribution and abundance at different brackish and freshwater ecosystems

Previously, (Cabello-Yeves *et al.*, 2017), we used double labelling DOPE-FISH (Behnam *et al.*, 2012) to show that *S. lacustris* could be the most abundant picocyanobacterium (representing

up to 88 % of the total picocyanobacterial) in the oligotrophic Tous reservoir during the winter mixing period (February 2015 and February 2016), when autotrophic picoplankton was detected in numbers up to 10^4 cells/mL. Here, we have again used flow cytometry combined with metagenomic recruitments to determine the abundance of the two new microbes in Tous reservoir. Flow cytometry was applied to a sample taken in autumn (October 2017) (Fig. S5) when the abundance of autotrophic picoplankton was higher reaching up to 10^5 cells/mL (a total of 6 % of the total bacteria) at the 10 m layer. This proportion was higher than that one observed in surface (1.55 %) and deeper layers (2.29 %). A general overview of the presence of each strain in different available metagenomes from freshwater lakes, reservoirs and brackish systems (see Supplementary DataSheet 1) is shown in Fig. 3. As depicted by reads per kilobase of genome per gigabase of metagenome (RPKG) values (Fig. 3) and recruitment plots (Fig. 4), *S. lacustris* Tous was present at very high proportions (from 85 to 120 RPKG, 8.5 to 12x coverage) in all Tous reservoir metagenomic datasets (winter and summer), whilst *C. usitatum* Tous appeared to be at lower proportions at the Deep Chlorophyll Maximum (DCM) layer of the summer period in Tous (< 5 RPKG), but was present in winter between 20 and 50 RPKG. It must be noted that MAG fragments from these two picocyanobacteria were binned by metagenomic approaches carried out in the same freshwater reservoir (Cabello-Yeves *et al.*, 2017). Taking these RPKG values together with DOPE-FISH and flow cytometry data, it seems that *S. lacustris* Tous and *C. usitatum* Tous, which are type II and IIB (respectively) pink pigmented strains, are representatives of two major picocyanobacterial ecotypes in the oligotrophic Tous reservoir.

The pattern observed by the metagenomic fragment recruitment in other temperate lakes and reservoirs (Fig. 3 and Fig. 4) points towards a high abundance of *S. lacustris* in temperate freshwater ecosystems, especially in the Lanier and Tous reservoirs, as previously described (Cabello-Yeves *et al.*, 2017). The datasets from the Sparkling Lake, Dexter and Amadorio reservoirs also showed the presence of this microbe (from 5 to 20 RPKG). No significant presence of it was obtained in cold or tropical freshwater ecosystems, which leads us to hypothesize that this strain could be better adapted to temperate conditions resembling those of Dexter, Tous, Amadorio, Lanier reservoirs and Sparkling Lake, all of which range from oligotrophic to mesotrophic status. Contrastingly, *C. usitatum* appears to be present not only in temperate reservoirs such as Tous or Dexter (absent in Lake Lanier or Amadorio), but also in cold perennially ice-covered lakes such as Lake Baikal (Cabello-Yeves *et al.*, 2018) or Lake Erie. We also found significant presence of this picocyanobacterium in samples of low salinity taken at 2 m depth in the Baltic Sea (at N 56°55.851, E 17°03.640, 10 km off the coast of Öland, Sweden) (Hugerth *et al.*, 2015). In all of these cases we observed values between 5 and 50 RPKG. Relatives of *C. usitatum* (ANI between 85 and 95 %), which could be tropical species, seem to be very abundant in Amazonian lakes Ananá and Mancapuru (Toyama *et al.*, 2016) (Fig. S6). These results show that both strains are globally distributed at least in oligo-mesotrophic systems although *C. usitatum* appears to be more versatile, including adaptations to brackish, cold and temperate environments.

None of these strains was detected significantly above the species level of 95 % ANI in other tropical environments such as Lake Gatun or Amazon lakes (see exceptions above), or in North-American (such as Mendota, Trout Bog, Yellowstone, Damariscotta) and European (Swedish, Finish) lakes. Other members of the sub-cluster 5.2, *S.elongatus* and the PCC

freshwater clade, were not significantly detected in the metagenomes where *S.lacustris* and *C.usitatum* presence was assessed. It appears that we stumbled upon microbes that are widespread in freshwater environments, being as relatively abundant as known ecologically relevant streamlined microbes such as *Polynucleobacter* (Hahn *et al.*, 2015), LD12 *Alphaproteobacteria* (Salcher *et al.*, 2011) or acI *Actinobacteria* (Neuenschwander *et al.*, 2017).
.

Salt tolerance of the novel strains

To confirm whether our strains could grow at different salt concentrations we designed an experiment with different salinities (see experimental procedures). As seen by the growth curves (Fig. 5), good growth was only observed at low salinities of 0.6, 1.3 and 3 g/L of NaCl. Both strains grew better at 0.6 and 1.3 g/L, which correspond to the salinities present in BG11 medium diluted two fold and non-diluted BG11 medium, respectively. *C. usitatum* Tous grew optimally at 0.6 g/L, and similarly at 1.3 g/L. *S. lacustris* Tous grew better at 1.3 g/L. Hence, their presence in several freshwater systems and their optimal growth in typical freshwater media support their freshwater specialization.

As noted above, *C. usitatum* was detected significantly by recruitment in Baltic Sea surface samples taken at low salinities of 5-7 ppt (Hugerth *et al.*, 2015), which correspond to a brackish system more similar to a freshwater body than to the ocean salinity-wise. The strain grew at 3 g/L, a salinity comparable to the one observed in the abovementioned Baltic Sea samples. However, the growth of this strain was not optimal at 3 g/L and decreased significantly at salinity ranges of 6-50 g/L after the first 10 days. The presence of some electrochemical potential-driven channels, transporters for polar amino acids and proline/betaine, glutamate:Na+ and glucose/mannose:H+ symporters or an aquaporin Z detected in the genome of this strain could help to explain their presence in low salinity regions (5 ppt) in the Baltic Sea. Unexpectedly, *S. lacustris* Tous appears to tolerate mid-range salinities of 6-12 g/L, although it grows slowly in these ranges. However, this strain was not detected in cold brackish systems such as the Baltic Sea, but it was in other temperate freshwater systems. As could be expected, none of these strains appear to tolerate salinity ranges resembling marine (30-36 g/L) or hypersaline (50 g/L) environments.

Pan-genome of the novel strains and other freshwater, marine and brackish representatives

We analysed through reciprocal gene homology identification (Contreras-Moreira and Vinuesa, 2013) the pan-genome of the novel strains with other representatives from sub-clusters 5.2 and 5.3 (see Supplementary Datasheet 1) to explore the diversity of this polyphyletic group (Fig. S7). In this comparison we used pure cultures, draft genomes and MAGs. We obtained the % of genes that belonged to soft core and strict core (Kaas *et al.*, 2012), and flexible genome separated into shell and cloud (Wolf *et al.*, 2012). An overview of the different proteins divided by category and genome is shown in Supplementary Datasheet 2. This comparison showed an average of 25 % of the genes belonging to the core genome, a much lower % of shared genomic content relative to previous comparisons of marine strains from sub-cluster 5.1 (Dufresne *et al.*, 2008), which confirms that members of the sub-clusters 5.2 and 5.3 share

fewer core genes than the marine and halotolerant members from sub-cluster 5.1. The genes associated with the core genome, as could be expected, corresponded to housekeeping genes such as fundamental photosystem components and the majority of ribosomal proteins, DNA polymerases, translation factors and central metabolism. As expected, MAGs presented a higher number of genes (20 %) constituting the core genome than pure cultures. This is a common observation within MAGs, which are typically binned for their core genome due to the under-representation of flexible genes in metagenomics reads.

The flexible genome represented up to the 70 % of the picocyanobacteria genomes (Fig. S7). Some of these genes are associated with genomic islands and represent differential adaptations of the multiple clones present in the species (Coleman *et al.*, 2006; Rodriguez-Valera *et al.*, 2009). Genetic functions encoded by the flexible genome included auxiliary metabolic genes which could be involved in niche colonization. Among these, complete clusters of ABC-transporters involved in solutes uptake, represented the most common elements associated with the flexible genome of all studied representatives. We detected transport of cobalt, phosphate and sulphonates among many others. Among the flexible genes (category cloud), nitrogenase *nif* genes were found in the planktonic picocyanobacterium *Vulcanococcus limneticus* LL (Di Cesare *et al.*, 2018). Other genes associated with the flexible genome encoded many hypothetical proteins, transposases and insertion elements, photosystem subunits, accessory pigments (PBS biosynthesis, allophycocyanin, phycobilin lyases) or chlorophyll biosynthesis. Finally, other flexible components were associated with defence mechanisms aimed at preventing phage infection, such as glycosyl transferase components of the O-chain of the LPS, restriction enzymes and the CRISPR-Cas found in the *S. lacustris* Tous isolate.

Genomic islands of the novel freshwater genomes

Considering that the genomes are complete but not closed in one single DNA contig, the study of genomic islands by position was done by making a single concatemer of each picocyanobacterial genome. To show the regions apparently rare in the environment, with low coverage and accounting for a low percentage of the bacterial population we used two metagenomics samples that were different from the isolation origin where the two picocyanobacteria highly recruit and are clearly present: the Baltic Sea metagenomes (Hugerth *et al.*, 2015) for *C. usitatum* and Lake Lanier datasets (Oh *et al.*, 2011) for *S. lacustris* Tous. The different genomic islands (Rodriguez-Valera *et al.*, 2009) and the genes present inside them for each concatenated picocyanobacterial genome are shown in Fig. S8.

Similar to the information available regarding marine *Synechococcus* (Dufresne *et al.*, 2008; Scanlan *et al.*, 2009), most of the regions of low coverage for our freshwater strains correspond to cell wall biogenesis and LPS biosynthesis, DNA metabolism, phage mobile genetic elements, transposases, integrases, recombinases, endonucleases or the phycobilisome (PBS) clusters (Six *et al.*, 2007). The type IIB PBS operon found in *C. usitatum* Tous appears to be in a genomic island in the Baltic Sea, but it was well covered in Tous. Paradoxically, metagenomes and previous studies showed that this is the dominant pigment type in the Baltic Sea (Larsson *et al.*, 2014). Along the same lines, very close relatives to *C.*

usitatum Tous containing this type IIB pigmentation have been found (Hugerth *et al.*, 2015). In contrast, the type II PBS operon in *S. lacustris* Tous does not fall inside regions of lower coverage in other tested freshwater datasets, and it appears to be a very abundant pigmentation type in freshwater systems such as Tous or Lanier, where picocyanobacterial species such as these clearly dominate the ecosystem.

As depicted in Fig. S8, antimicrobial peptides such as a type I polyketide synthase 2, a NHLP bacteriocin system secretion protein, SunT bacteriocin/lantibiotic exporter and ABC multidrug transporters are frequently associated with genomic islands in *C. usitatum*, while nif11 leader peptide and multidrug ABC transporters are located in regions of low coverage in *S. lacustris* Tous. The presence of nitrogen metabolism genes such as nitrate transporters, nitrite and nitrate reductases and urease in genomic islands in both strains is noteworthy, as this informs the nitrogen metabolism variants present in the Lake Lanier and Baltic Sea environments, where a high clonal diversity of *S. lacustris* and *C. usitatum* species was respectively observed above 95 % ANI (Fig. 4 and Fig. S8) . Moreover, molybdopterin and molybdenum cofactor biosynthesis genes, which were previously related to nitrate and nitrite assimilation processes (Scanlan *et al.*, 2009), also appear to be present in genomic islands in both strains. The cyanate hydratase found in *S. lacustris* Tous was also encoded in a low coverage region.

Another genomic island is associated with a *Synechococcus/Prochlorococcus* conserved hypothetical protein and a high light inducible protein encoded in a long region of low coverage that appears to be absent from the Baltic Sea environment in the *C. usitatum* genome. We also observed many transporters (K, Co, Cd, Zn, tricarboxylates, ATPases, Ca/Na, amino acid, sugars) and DNA metabolism genes (*rpoD*, nucleases, helicases, topoisomerases, hydrolases, kinases) encoded in genomic islands of both strains. One of the copies (of the four present in the genome) of the gene magnesium-protoporphyrin IX monomethyl ester cyclase (chlorophyll biosynthesis) was also encoded in a genomic island in *S. lacustris* Tous.

The CRISPR Cas system of *S. lacustris* is also present in a region of low coverage in Tous metagenomes, indicating that the mechanism is not present in all *S. lacustris* strains in the reservoir (ca. 20 %, seen by RPKG). However, the entire system was totally absent from other metagenomic datasets where *S. lacustris* Tous is very abundant such as Lake Lanier, illustrating that this CRISPR/Cas variant might be endemic to some freshwater ecosystems. It is possible that other CRISPR/Cas systems are present at other locations, or that they are replaced by other defence mechanisms involving less genetic expense, such as simple modifications of the outer membrane and restriction-modification systems, which were also found in regions of low coverage and hence are associated with genomic islands.

Conclusions:

The two picocyanobacteria retrieved in this work are widespread, and potentially abundant in different water bodies. They exhibit a different distribution pattern in freshwater and brackish ecosystems; it appears that *C. usitatum* Tous is a more widely distributed freshwater picocyanobacterium, abundant in cold waters such as the periodically ice-covered Lake Baikal and Lake Erie as well as in temperate reservoirs such as Dexter (USA) and Tous (Spain) and is

even widespread in low salinity, brackish cold ecosystems such as the Baltic Sea. Even tropical ecosystems such as Amazonian lakes indicated the presence of close relatives to this microbe. Contrastingly, *S. lacustris* Tous appears to be more dominant in temperate ecosystems such as Tous and Amadorio reservoirs (Spain), and lakes Lanier, Sparkling and Dexter (USA). The latter exhibits a wider set of tools for phage defence, including a novel type III CRISPR Cas system. The high diversity of genes encoded in genomic islands helps us to understand how the population diversity of these picocyanobacteria can be maintained in nature. Among these islands, fundamental genes for niche colonization of these strains were detected, such as those related to nitrogen metabolism, phage defence systems, metal, ion and nutrient transporters or PBS and other pigments.

The genomes derived from cultures are much more complete than MAGs. In our case the *S. lacustris* MAG contained only approximately 70 % of the actual genome in spite of values of completeness of 91 %. Specifically, MAGs are very poor predictors of the flexible genome, that is usually incomplete and assembles very poorly if at all. This emphasizes the importance of obtaining enrichment cultures and isolates (even if not pure cultures). Isolation and genome sequencing of new freshwater picocyanobacterial genomes, which undoubtedly are underway, will improve the classification and reveal the enormous genomic diversity of the freshwater representatives, opening new perspectives for evolutionary models of picocyanobacteria and global niche colonization, thus adding new information on their possible transition from freshwater to brackish and marine habitats or vice versa.

Experimental procedures:

Isolation of picocyanobacteria

S. lacustris Tous and *C. usitatum* Tous were isolated from a 12 m deep sample from Tous reservoir (39°14' N 0°65' W) during the winter water column mixing period (February 2016). The limnological features of Tous reservoir during sampling are given in Cabello-Yeves et al. (2017). Strains were cultured in BG11 medium (Rippka *et al.*, 1979) diluted two fold, at 14 °C with an irradiance of 10-20 µmol photons m^{-2} s^{-1} supplied by Gro-Lux lamps under light/dark cycles of 14:10 h. We followed the dilution to the extinction methodology. To do this, 9 mL of BG11 medium diluted two fold was prepared and sterilized in test tubes. One mL of water from the Tous sample was inoculated into five replicates of the medium. Each of the replicates was diluted in 10-fold increments from 10^0 to 10^{-4} mL. Growth of the strains was observed in 10^{-3} dilution.

Screening by specific PCR

Based on the genome sequence of the *S. lacustris* MAG (Cabello-Yeves *et al.*, 2017), we designed specific PCR primers of the *rpoB* and *recA* genes, which allowed us to screen *S. lacustris* species among different pink cultures from Tous. We used the *recA* gene as a target for the forward primer SynT-RecA-F (5'-TCGGGTTGTGGAGGTTTACG-3´) and SynT-RecA-R (5´-TGCTCAAGCCAAGTGATCGT-3´) as the reverse primer. We used the *rpoB* gene as a target for forward primer SynT-RpoB-F (5´-GGAAAGCCTTGGCAACTTGG-3´) and reverse primer SynT-

RpoB-R (5´-AATTCCTGCAGGGTGTAGGC-3´). Details of the PCR amplification step are summarized in the supplementary text. These specific primers confirmed the presence of *S. lacustris*, but were not specific for the other pink strain *C.usitatum*. Nonetheless, both strains were sequenced (see below).

DNA extraction, sequencing, assembling and annotation of the genomes

DNA of both cultures was extracted with an E.Z.N.A.® Soil DNA Kit (Omega Bio Tek). The *S. lacustris* Tous culture was sequenced with an Illumina MiSeq (Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana) with a Nextera®XT library preparation, obtaining approximately 0.5 million reads with an average read length of 127 bp. The *C. usitatum* culture was sequenced with an Ilummina HiSeq 3000 (Oklahoma Medical Research Foundation) with a Kapa DNA library preparation protocol, yielding a total of 39 million reads, with an average read length of 150 bp.

Paired-end reads were assembled with SPAdes (Bankevich *et al.*, 2012) both in meta and default mode with careful, only-assembler and default k-mer parameters. We performed mapping of reads to the genomes with BWA (Li and Durbin, 2009) and reassembled the genomes with SPAdes to improve contig length and remove redundancy. Additionally, we also used WiseScaffolder (Farrant *et al.*, 2015) to obtain larger contigs of our two strains (see Table S2). Nonetheless, genomes could not be closed but generated 79 contigs for *S. lacustris* and 22 for *C. usitatum*. Gene prediction was conducted with PRODIGAL (Hyatt *et al.*, 2010). Annotation of CDSs was done with the BLAST (Altschul *et al.*, 1997), RAST (Overbeek *et al.*, 2013), KEGG (KO), Blast Koala (Kanehisa and Goto, 2000; Kanehisa *et al.*, 2004, 2016), COG (Tatusov *et al.*, 2001) and TIGR (Haft *et al.*, 2001) databases, while tRNAs were detected with tRNAscan (Lowe and Eddy, 1997) and 16S rRNAs were determined with ssu-align (Nawrocki, 2009). Protein-specific hits and domains were predicted with CDD-SPARCLE (Marchler-Bauer *et al.*, 2016).

Specific rpoC2 primers for S. lacustris and C. usitatum detection

After sequencing their genomes, we designed specific primers for the *rpoC2* gene, which could be used for the detection of the two species in cultures. We used the forward primer RpoC2-SynLac-F (5´ GCAGCGTAAGGCTAAGGGAA 3´) and the reverse primer RpoC2-SynLac-R (5´ CGTATTCCTCCACAGGACGG 3´) for the detection of *S. lacustris* Tous, and we used the forward primer RpoC2-CyanUsit-F (5´ CCGCCTAGTTAGCGAGTACG 3´) and the reverse primer RpoC2-CyanUsit-R (5´ CACTGGCCGTCCTCAATCAT 3´) for the detection of *C. usitatum*. Details of the PCR amplification step are summarized in the supplementary text.

Phylogenomics, ANI, AAI and pan-genome analysis of Synechococcus/Cyanobium

A concatenated protein based phylogeny with 244 universal markers was made with the tool PhyloPhlAn (Segata *et al.*, 2013). We used all *Synechococcus* and *Cyanobium* isolates, either draft or complete genomes and MAGs available from NCBI on 1st January 2018. We used 9 *Prochlorococcus* genomes to root the phylogeny. Average Nucleotide Identity (ANI) between

them was also calculated as previously described (Konstantinidis and Tiedje, 2005; Goris *et al.*, 2007).

AAI was determined with the package compareM (https://github.com/dparks1134/CompareM). The pan-genome analysis of *Synechococcus* and *Cyanobium* genomes was conducted with the software Get-Homologous (Contreras-Moreira and Vinuesa, 2013) with the -M option that uses the orthoMCL algorithm to infer homology. According to reciprocal homology, genes belonging to the core genome were further divided in strict-core and soft-core (Kaas *et al.*, 2012), while the flexible genome was separated into shell and cloud (Wolf *et al.*, 2012). A summary with the general features of the different picocyanobacteria used in this work is presented in Supplementary Datasheet 1.

Abundance of the novel picocyanobacteria on different freshwater and brackish metagenomic datasets

Recruitment plots and RPKG values (expressed as reads per Kb of genome per Gb of metagenome) were obtained from a set of > 300 brackish and freshwater metagenomic datasets obtained from 35 different geographic sites from NCBI and JGI, as previously described (Cabello-Yeves *et al.*, 2018). A threshold value of > 95 % of identity and > 50 bp of alignment length was used to count a hit. The metagenomic data sets that gave positive hits for the presence of our strains are publicly available:, including for the Amadorio (Ghai *et al.*, 2014) and Tous reservoirs (Cabello-Yeves *et al.*, 2017), Lake Lanier (Oh *et al.*, 2011), Dexter Reservoir (PRJNA312985), Sparkling Lake (SRR6476262, SRR6476264), Amazonian Lakes (Toyama *et al.*, 2016), Lake Erie (SRR6201979, SRR6201526 and PRJNA288501), Lake Baikal (Cabello-Yeves *et al.*, 2018), and the Baltic Sea (Hugerth *et al.*, 2015). Additional features of the datasets used are described in Supplementary Datasheet 1.

Flow cytometry counts and epifluorescence microscopy of the novel picocyanobacteria

Water samples were fixed in situ with a paraformaldehyde:glutaraldehyde solution to reach a final concentration in the sample of 1 %:0.05 % (w/v) (Marie *et al.*, 1997). For flow cytometry counting of the picocyanobacteria and heterotrophic bacterioplankton, a Coulter Cytomics FC500 flow cytometer equipped with an argon laser (488 excitation), a red emitting diode (635 excitation), and five filters for fluorescent emission (FL1-FL5) was used. Flow Cytometer FC500 was equipped with five photomultipliers with a spectral sensitivity from 185 nm-900 nm collecting the fluorescence signals centred at FL1 (525 nm/40), FL2 (575/40), FL3 (620/20), FL4 (675/40) and FL5 (755/40). Picocyanobacteria counting was performed with discrimination by FL4, whereas for bacterioplankton identification, discrimination was made by fluorescence in FL1 after 1h SybrGreen-I staining (Veldhuis and Kraay, 2000). Analyses were run for 120 s at a flow rate of 128 μL min^{-1}. Abundance of each population was calculated according to the formula: $N = (n \cdot 1{,}000)/q \cdot t$, where q is the flow rate (microliter per minute), t is the length (minutes) of the data acquisition, n is the number of events counted by the flow cytometer, and N is the number of cells per mL. Flow rate was obtained gravimetrically considering the processed volume. Data were collected with the Beckman Coulter acquisition software "CXP Version 2.2 Acquisition", then analysed with the Beckman Coulter analysis software for analysis "CXP Version 2.2 Analysis."

Microscopy pictures were obtained from cultures for each picocyanobacteria with a Zeiss Axioplan microscope equipped with an HBO 100 W lamp, a Neofluar 100 x objective 1.25 x additional magnification and filter sets for blue (BP450-490, FT510, LP520) and green light excitation (LP510-560, FT580, LP590). Additionally, we used an inverted Zeiss III RS epifluorescence microscope (600 X, resolution 0.03 unit/pixel) using G365 as the exciting filter and LP420 as a suppression filter for blue light, as well as G546 as the exciting filter and an LP590 suppression filter for green light.

Salt tolerance experiments

Salinity tests were performed with *S. lacustris* Tous and *C. usitatum* Tous strains. Ten cultures were performed in triplicate with salinities of 0.6, 1.3, 3, 6, 12, 18, 24, 30, 36 and 50 g/L of NaCl in BG11 media. Sterile glass tubes were used to perform the growth experiments. We used an inoculum of $2.94 \cdot 10^5$ cells/mL for *S.lacustris* and $2.77 \cdot 10^6$ cells/mL for *C.usitatum*. Strains under different salinities were incubated at 14 °C with an irradiance of 16 µmol photons m^{-2} s^{-1} supplied by Gro-Lux lamps under light/dark cycles of 14:10 h. Measurements were taken with a Hitachi F7000 fluorescence spectrophotometer at 0, 2, 4, 10, 18, 24 and 31 days after inoculation. *S. lacustris* measurements were taken with an excitation scan between 550-700 nm, taking the emission at 680 nm, which corresponds to the signal of chlorophyll-*a*. *C.usitatum* measurements were taken with the same excitation scan, but obtaining the emission at 575 nm, which corresponds to the signal of phycoerythrin.

Genomes accessibility

The two picocyanobacterial genomes have been deposited in the NCBI-GenBank databases under the bioproject accession number PRJNA435996. *S. lacustris* Tous is under GenBank identifier PXVC00000000 and *C. usitatum* Tous is under identifier PXXO00000000.

Acknowledgements:

FR-V was supported by grant “VIREVO” CGL2016-76273-P [AEI/FEDER, EU], (cofunded with FEDER funds); Acciones de dinamización “REDES DE EXCELENCIA” CONSOLIDER-CGL2015-71523-REDC from the Spanish Ministerio de Economía, Industria y Competitividad and PROMETEO II/2014/012 “AQUAMET” from Generalitat Valenciana. AC and AP were also supported by the project CLIMAWET, CGL2015-69557-R”, granted to AC and funded by the Spanish Ministry of Economy and Competitiveness, and by the European Union through the European Fund for Regional Development "One way to make Europe”.

Conflict of interest:

The authors declare no conflict of interest.

References:

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477.

Behnam, F., Vilcinskas, A., Wagner, M., and Stoecker, K. (2012) A straightforward DOPE (double labeling of oligonucleotide probes)-FISH (fluorescence in situ hybridization) method for simultaneous multicolor detection of six microbial populations. *Appl. Environ. Microbiol.* **78**: 5138–5142.

Bhaya, D., Grossman, A.R., Steunou, A.-S., Khuri, N., Cohan, F.M., Hamamura, N., et al. (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* **1**: 703–713.

Blank, C.E. and Sanchez-Baracaldo, P. (2010) Timing of morphological and ecological innovations in the cyanobacteria–a key to understanding the rise in atmospheric oxygen. *Geobiology* **8**: 1–23.

Cabello-Yeves, P.J., Haro-Moreno, J.M., Martin-Cuadrado, A.-B., Ghai, R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2017) Novel Synechococcus Genomes Reconstructed from Freshwater Reservoirs. *Front. Microbiol.* **8**: 1151.

Cabello-Yeves, P.J., Zemskaya, T.I., Rosselli, R., Coutinho, F.H., Zakharenko, A.S., Blinov, V. V, and Rodriguez-Valera, F. (2018) Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. *Appl. Environ. Microbiol.* **84**: e02132-17.

Cai, F., Axen, S.D., and Kerfeld, C.A. (2013) Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. *RNA Biol.* **10**: 687–693.

Callieri, C. (2017) Synechococcus plasticity under environmental changes. *FEMS Microbiol. Lett.* **364**: fnx229.

Callieri, C., Coci, M., Corno, G., Macek, M., Modenutti, B., Balseiro, E., and Bertoni, R. (2013) Phylogenetic diversity of nonmarine picocyanobacteria. *FEMS Microbiol. Ecol.* **85**: 293–301.

Callieri, C., Cronberg, G., and Stockner, J.G. (2012) Freshwater picocyanobacteria: single cells, microcolonies and colonial forms. In, *Ecology of Cyanobacteria II*. Springer, pp. 229–269.

Callieri, C. and Stockner, J.G. (2002) Freshwater autotrophic picoplankton: a review. *J. Limnol.* **61**: 1–14.

Camacho, A., Miracle, M.R., and Vicente, E. (2003) Which factors determine the abundance and distribution of picocyanobacteria in inland waters? A comparison among different types of lakes and ponds. *Arch. für Hydrobiol.* **157**: 321–338.

Di Cesare, A., Cabello-Yeves, P.J., Chrismas, N.A.M., Sánchez-Baracaldo, P., Salcher, M.M., and Callieri, C. (2018) Genome analysis of the freshwater planktonic Vulcanococcus limneticus sp. nov. reveals horizontal transfer of nitrogenase operon and alternative pathways of nitrogen utilization. *BMC Genomics* **19**:.

Christaki, U., Jacquet, S., Dolan, J.R., Vaulot, D., and Rassoulzadegan, F. (1999) Growth and grazing on Prochlorococcus and Synechococcus by two marine ciliates. *Limnol. Oceanogr.* **44**: 52–61.

Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., DeLong, E.F., and Chisholm, S.W. (2006) Genomic islands and the ecology and evolution of Prochlorococcus. *Science (80-.).* **311**: 1768–1770.

Contreras-Moreira, B. and Vinuesa, P. (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* **79**: 7696–7701.

Dufresne, A., Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P., et al. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: R90.

Farrant, G.K., Doré, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M., et al. (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl. Acad. Sci.* **113**: E3365–E3374.

Farrant, G.K., Hoebeke, M., Partensky, F., Andres, G., Corre, E., and Garczarek, L. (2015) WiseScaffolder: an algorithm for the semi-automatic scaffolding of Next Generation Sequencing data. *BMC Bioinformatics* **16**: 281.

Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2014) Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Mol. Ecol.* **23**: 6073–6090.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**: 81–91.

Grébert, T., Doré, H., Partensky, F., Farrant, G.K., Boss, E.S., Picheral, M., et al. (2018) Light color acclimation is a key process in the global ocean distribution of Synechococcus cyanobacteria. *Proc. Natl. Acad. Sci.* 201717069.

Guimarães, P.I., Leão, T.F., de Melo, A.G.C., Ramos, R.T.J., Silva, A., Fiore, M.F., and Schneider, M.P.C. (2015) Draft genome sequence of the picocyanobacterium Synechococcus sp. strain GFB01, isolated from a freshwater lagoon in the Brazilian Amazon. *Genome Announc.* **3**: e00876-15.

Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T., and White, O. (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**: 41–43.

Hahn, M.W., Koll, U., Jezberová, J., and Camacho, A. (2015) Global phylogeography of pelagic Polynucleobacter bacteria: restricted geographic distribution of subgroups, isolation by distance and influence of climate. *Environ. Microbiol.* **17**: 829–840.

Holtman, C.K., Chen, Y., Sandoval, P., Gonzales, A., Nalty, M.S., Thomas, T.L., et al. (2005) High-throughput functional analysis of the Synechococcus elongatus PCC 7942 genome. *DNA Res.* **12**: 103–115.

Hugerth, L.W., Larsson, J., Alneberg, J., Lindh, M. V, Legrand, C., Pinhassi, J., and Andersson, A.F. (2015) Metagenome-assembled genomes uncover a global brackish microbiome.

Genome Biol. **16**: 1–18.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 1.

Kaas, R.S., Friis, C., Ussery, D.W., and Aarestrup, F.M. (2012) Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. *BMC Genomics* **13**: 577.

Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**: 27–30.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**: D277–D280.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**: 726–731.

Konstantinidis, K.T. and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 2567–2572.

Larsson, J., Celepli, N., Ininbergs, K., Dupont, C.L., Yooseph, S., Bergman, B., and Ekman, M. (2014) Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J.* **8**: 1892–1903.

Laudenbach, D.E. and Grossman, A.R. (1991) Characterization and mutagenesis of sulfur-regulated genes in a cyanobacterium: evidence for function in sulfate transport. *J. Bacteriol.* **173**: 2739–2750.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955–964.

Ludwig, M., Schulz-Friedrich, R., and Appel, J. (2006) Occurrence of hydrogenases in cyanobacteria and anoxygenic photosynthetic bacteria: implications for the phylogenetic origin of cyanobacterial and algal hydrogenases. *J. Mol. Evol.* **63**: 758–768.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., et al. (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**: D200–D203.

Marie, D., Partensky, F., Jacquet, S., and Vaulot, D. (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl. Environ. Microbiol.* **63**: 186–193.

Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D.J. (2012) Multi-locus sequence analysis, taxonomic resolution and biogeography of marine Synechococcus. *Environ. Microbiol.* **14**: 372–386.

Nawrocki, E.P. (2009) Structural RNA homology search and alignment using covariance models. Washington University in St. Louis.

Neuenschwander, S.M., Ghai, R., Pernthaler, J., and Salcher, M.M. (2017) Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J. https://doi.org/10.1038/ismej.2017.156.*

Oh, S., Caro-Quintero, A., Tsementzi, D., DeLeon-Rodriguez, N., Luo, C., Poretsky, R., and Konstantinidis, K.T. (2011) Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl. Environ. Microbiol.* **77**: 6000–6011.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., et al. (2013) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**: D206–D214.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* **111**: 1–61.

Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pašić, L., Thingstad, T.F., Rohwer, F., and Mira, A. (2009) Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**: 828–836.

Salcher, M.M., Pernthaler, J., and Posch, T. (2011) Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria 'that rule the waves'(LD12). *ISME J.* **5**: 1242.

Sánchez-Baracaldo, P. (2015) Origin of marine planktonic cyanobacteria. *Sci. Rep.* **5**: 17418.

Sanchez-Baracaldo, P., Hayes, P.K., and Blank, C.E. (2005) Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology* **3**: 145–165.

Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., et al. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* **73**: 249–299.

Scanlan, D.J. and West, N.J. (2002) Molecular ecology of the marine cyanobacterial genera Prochlorococcus and Synechococcus. *FEMS Microbiol. Ecol.* **40**: 1–12.

Segata, N., Börnigen, D., Morgan, X.C., and Huttenhower, C. (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**: 2304.

Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., et al. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci.* **110**: 1053–1058.

Six, C., Thomas, J.-C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., et al. (2007) Diversity and evolution of phycobilisomes in marine Synechococcus spp.: a comparative genomics study. *Genome Biol.* **8**: 1.

Sohm, J.A., Ahlgren, N.A., Thomson, Z.J., Williams, C., Moffett, J.W., Saito, M.A., et al. (2016) Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J.* **10**: 333–345.

Stockner, J., Callieri, C., and Cronberg, G. (2000) Picoplankton and other non-bloom-forming cyanobacteria in lakes. In, *The ecology of cyanobacteria*. Springer, pp. 195–231.

Stockner, J.G. (1988) Phototrophic picoplankton: an overview from marine and freshwater ecosystems. *Limnol. Oceanogr.* **33**: 765–775.

Stomp, M., Huisman, J., Vörös, L., Pick, F.R., Laamanen, M., Haverkamp, T., and Stal, L.J. (2007) Colourful coexistence of red and green picocyanobacteria in lakes and seas. *Ecol. Lett.* **10**: 290–298.

Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature* **424**: 1047–1051.

Suttle, C.A. and Chan, A.M. (1993) Marine cyanophages infecting oceanic and coastal strains of Synechococcus: abundance, morphology, cross-infectivity and growth characteristics. *Mar. Ecol. Prog. Ser.* 99–109.

Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., Wünschiers, R., and Lindblad, P. (2002) Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiol. Mol. Biol. Rev.* **66**: 1–20.

Tatusov, R.L., Natale, D.A., Garkavtsev, I. V, Tatusova, T.A., Shankavaram, U.T., Rao, B.S., et al. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**: 22–28.

Toyama, D., Kishi, L.T., Santos-Júnior, C.D., Soares-Costa, A., de Oliveira, T.C.S., de Miranda, F.P., and Henrique-Silva, F. (2016) Metagenomics analysis of microorganisms in freshwater lakes of the Amazon Basin. *Genome Announc.* **4**: e01440-16.

Veldhuis, M.J.W. and Kraay, G.W. (2000) Application of flow cytometry in marine phytoplankton research: current applications and future perspectives. *Sci. Mar.* **64**: 121–134.

Vörös, L., Callieri, C., Katalin, V., and Bertoni, R. (1998) Freshwater picocyanobacteria along a trophic gradient and light quality range. In, *Phytoplankton and Trophic Gradients*. Springer, pp. 117–125.

Wang, H., Fewer, D.P., and Sivonen, K. (2011) Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. *PLoS One* **6**: e22384.

Wolf, Y.I., Makarova, K.S., Yutin, N., and Koonin, E. V (2012) Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7**: 46.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., et al. (2008) Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. *Environ. Microbiol.* **10**: 147–161.

Figure legends:

Fig. 1. Epifluorescence microscopy of (A) *Cyanobium usitatum* Tous and (B) *Synechococcus lacustris* Tous. The top pictures were taken with a Zeiss III RS epifluorescence microscope at 600x and green excitation. The bottom pictures were taken with a Zeiss Axioplan at 1250x and blue excitation.

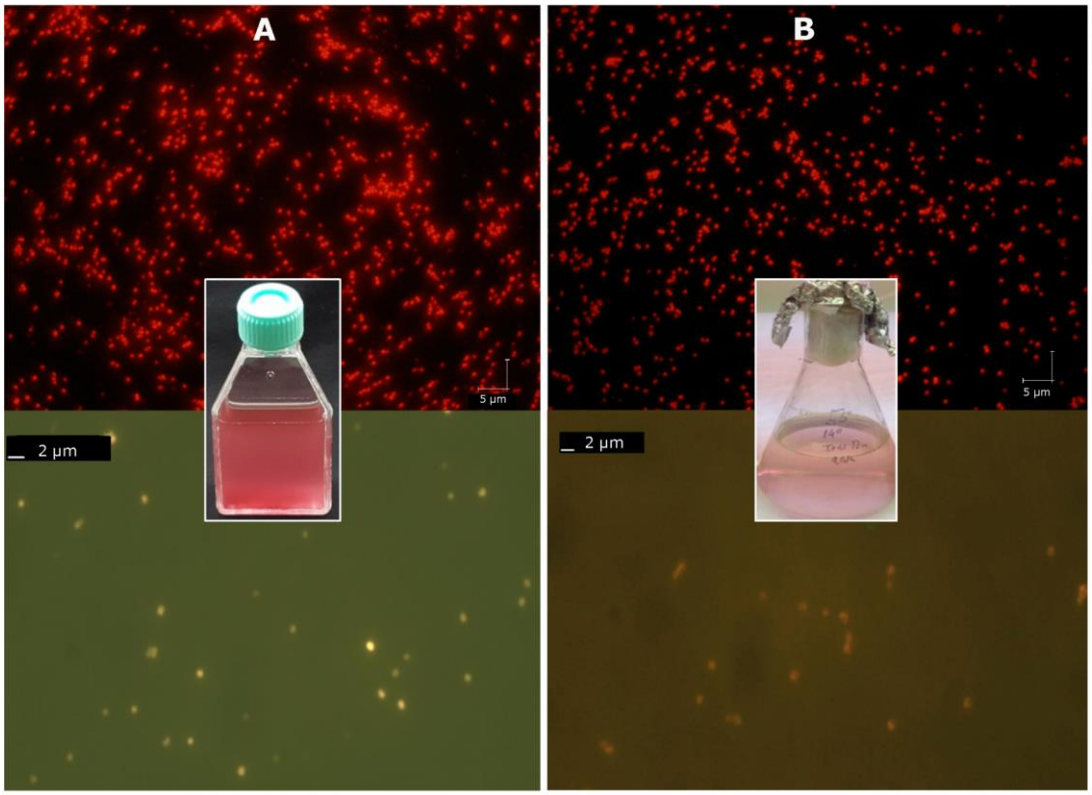
Fig. 2. Concatenated protein based phylogenomics of the *Synechococcus/Cyanobium* genera with 244 universal markers (PhyloPhlAn) encompassing marine, euryhaline, brackish and freshwater representatives from all known sub-clusters and clades. Nine Prochlorococcus genomes were also used. Average Nucleotide Identity (ANI) values of all members from sub-clusters 5.3 and 5.2 are also provided.

Fig. 3. Presence of *Cyanobium usitatum* Tous and *Synechococcus lacustris* Tous isolates on different worldwide freshwater/brackish datasets. Heatmap represents the RPKG values (expressed as reads per Kb of genome per Gb of metagenome) on different metagenomics datasets at > 95 % identity and 50 bp of alignment lengths.

Fig. 4. Recruitment plots of the novel picocyanobacteria against different freshwater and brackish datasets. Left blue coloured plots represent the recruitment of *C. usitatum* Tous and right black coloured plots represent the recruitment of *S. lacustris* Tous on different metagenomes.

Fig. 5. Growth curves of *S. lacustris* Tous and *C. usitatum* Tous at different salinities. The X axis represents the number of days and the Y axis represents arbitrary units of fluorescence (A.U.F). Salinity growth curves are colour-coded.

Fig. 1



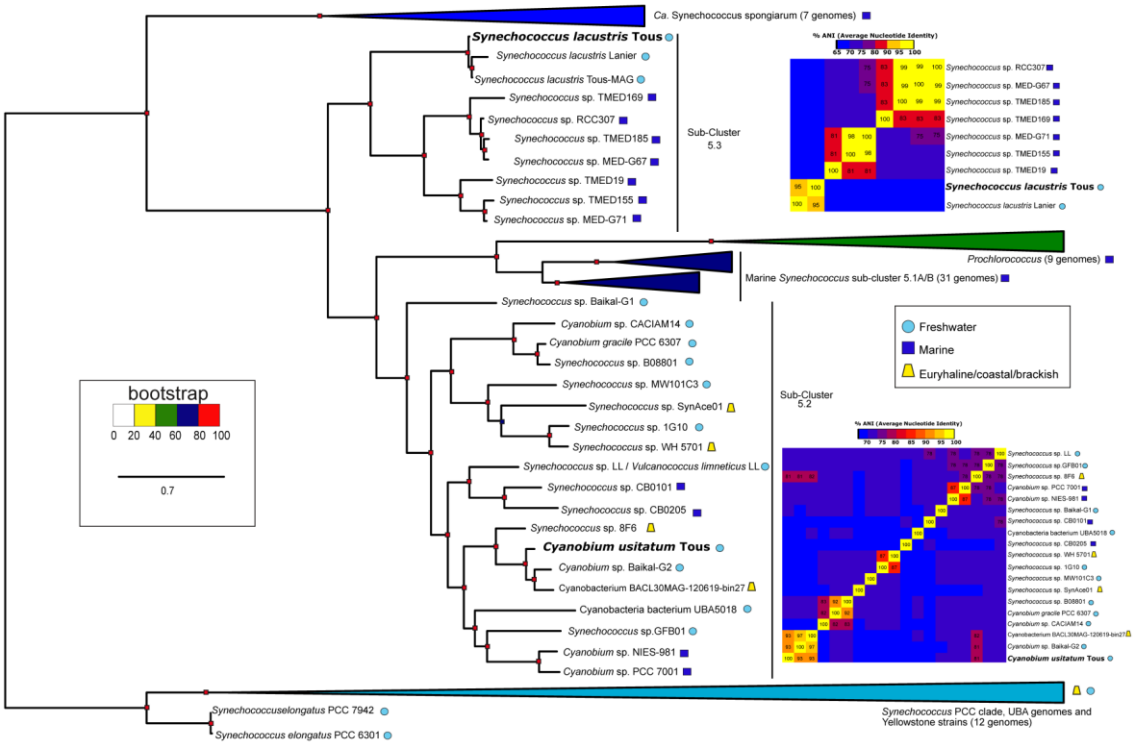


Fig. 3

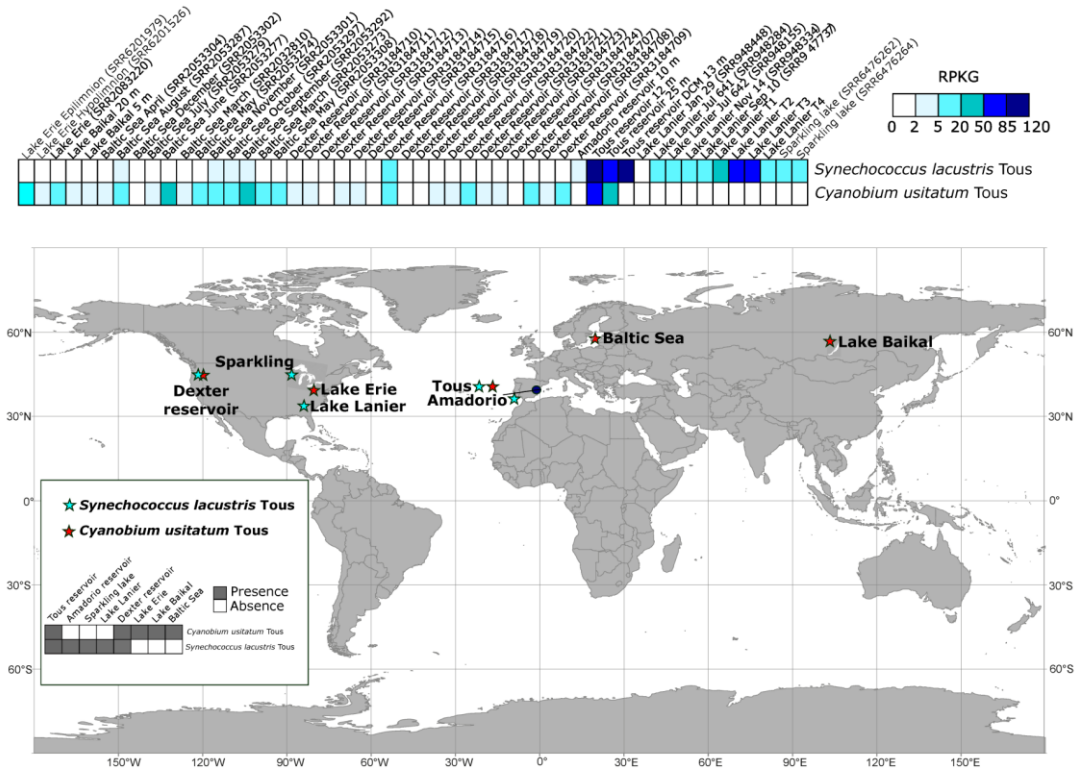


Fig. 4

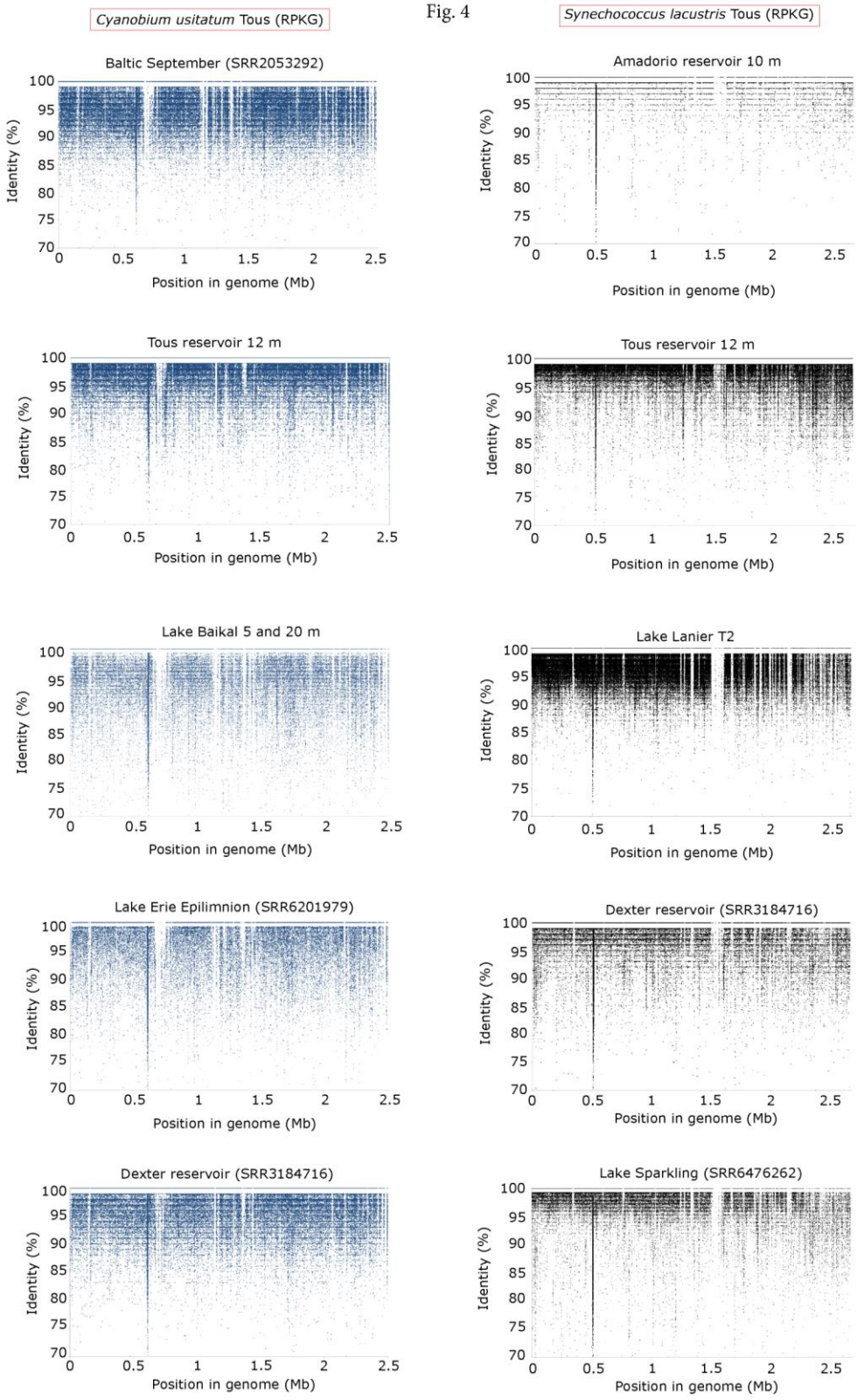


Fig. 5

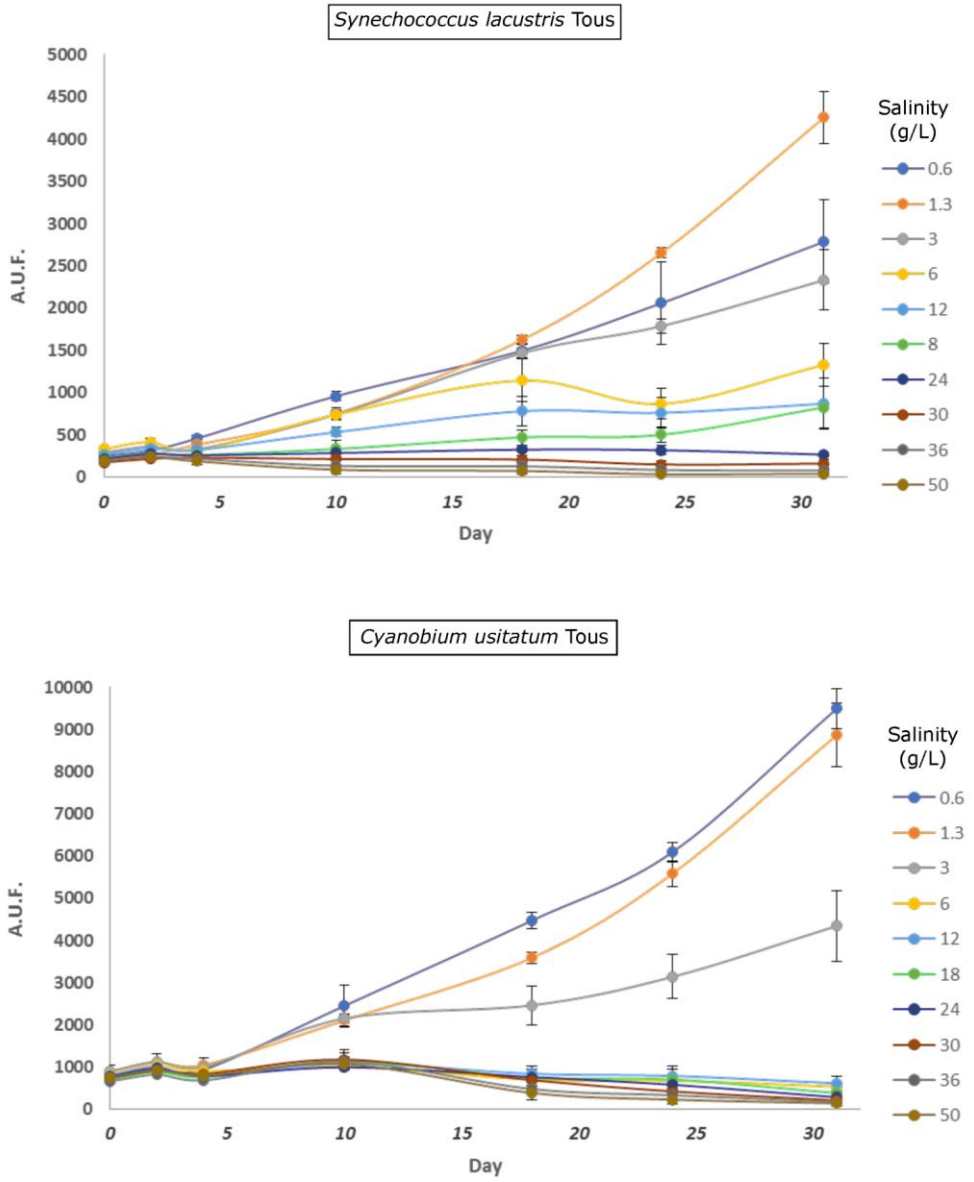


Table 1. Summary statistics and characteristics of the isolated picocyanobacteria.

Features	*Synechococcus lacustris* Tous	*Cyanobium usitatum* Tous
Cell shape	Coccoid	Coccoid
Cell size (µm)	1.14 ± 0.13	0.70 ± 0.07
Culture colour	Pink	Pink/red

Genome Size (bp)	2668880	2515235
GC %	51.392	62.584
nº CDS	3002	2698
Median Intergenic Spacer (bp)	25	31
Completeness (contamination)*	98.75 (1.88)	99.76 (0.35)
Current State (nº contigs)	WGS (303)	WGS (25)
Isolation	Tous reservoir (39.14 N 0.65 W)	Tous reservoir (39.14 N 0.65 W)
Depth	12 m	12 m
Type Pigment	Type II	Type IIB
Sub-cluster	5.3	5.2
Distribution	Temperate (Tous, Amadorio and Dexter reservoirs, Lake Lanier)	Temperate (Tous and Dexter reservoirs), Cold (Baltic Sea, Lake Baikal, Lake Erie)

*Estimated with 555 Synechococcus genes from CheckM pachage

Table 2. Comparison of the main metabolic features of the isolated picocyanobacteria. Features highlighted in bold are unique to this species compared to the other one.

	Synechococcus lacustris Tous	*Cyanobium usitatum* Tous
Phage defence systems	**CRISPR-Cas**, transposases, Restriction Modification, toxin-antitoxin, DNA phosphothiolation systems	Transposases, Restriction Modification, toxin-antitoxin, DNA phosphothiolation systems
N sources (transporters)	ammonia, nitrate/nitrite, **cyanate**, urea	ammonia, nitrate/nitrite, urea
Phosphorous metabolism	PstS, PstACB, **PhoU**, PhoRB, PPiK, Ppx, IP, Thba, AP	PstS, PstACB, PhoRB, PPiK, Ppx, IP, Thba, AP
Secretion systems	type IV pilus	type IV pilus, **type I system for autoaggregation**
Electron donating reactions	Respiratory dehydrogenases	Respiratory dehydrogenases, **bidirectional hydrogenase (Hox)**
Sulfur utilization	Thioredoxin-disulfide reductase, **thiosulfate Cys ABC transporter**	Thioredoxin-disulfide reductase
Antimicrobial peptides / proteins	NHLP bacteriocin system secretion protein, nif11 leader peptide, SunT bacteriocin/lantibiotic exporter and ABC multidrug tansporters	**Polyketide synthase type I,** NHLP bacteriocin system secretion protein, nif11 leader peptide, SunT bacteriocin/lantibiotic exporter and ABC multidrug tansporters
Cofactors, vitamins, pigments biosynthesis	**Type II PBS**, pyridoxin, thiamin, cobalamin, chlorophyll biotin, coenzyme B12, folate, riboflavin, molybdenum, ubiquinone, pterin, **menaquinone and phylloquinone**, NAD/NADP, lipoic acid, coenzyme A	**Type IIB PBS**, pyridoxin, thiamin, cobalamin, chlorophyll, biotin, coenzyme B12, folate, riboflavin, molybdenum, ubiquinone, pterin, NAD/NADP, lipoic acid, coenzyme A, **Coenzyme F420**
Unique Transporters	**Cellullose synthase, thiosulfate, Zn/Mn, cation:H^+ antiporter, LPS, indigoidine (blue pigment) exporter**	**aquaporin Z, glutamate:Na+ symporter, proline/betaine, biopolymer, Mg2+, xanthine/uracil, glucose/mannose:H+ symporter**

Supplementary text

Description of *Cyanobium usitatum* and *Synechococcus lacustris* species.

PCR details of picocyanobacterial screening

Supplementary figures and table legends:

Table S1. Comparison of genomic features determined for *Synechococcus lacustris* MAG (Cabello-Yeves et al, 2017) and *Synechococcus lacustris* Tous isolate.

Table S2. Summary statistics of the two picocyanobacterial genomes reassembled with BWA+SPAdes and WiseScaffolder pipelines.

Fig. S1. Epifluorescence microscopy of a parallel culture of *S. lacustris* Tous which lost phycoerythrin and pink pigmentation. Pictures were taken with a Zeiss Axioplan at 1250x and green excitation.

Fig. S2. (A) Average Amino acid Identity (AAI) and (B) Average Nucleotide Identity (ANI) among marine, freshwater and euryhaline *Synechococcus* and *Cyanobium* species from sub-clusters 5.2 and 5.3.

Fig. S3. Phycobilisome (PBS) operon comparison among different marine, coastal and freshwater *Synechococcus* and *Cyanobium* species. Comparison made with TBLASTX with >150 bp of alignment length and > 30 % of similarity.

Fig. S4. CRISPR Cas system of *Synechococcus lacustris* Tous and host-phage homology matches between CRISPRs and spacers. Comparison made with BLASTN at > 85 % of CRISPR-spacer identity.

Fig. S5. Flow cytometry counts of autotrophic picoplankton (APP) and heterotrophic picoplankton (HPP) in Tous reservoir (October 2, 2017) along its depth.

Fig. S6. Recruitment plots of *C. usitatum* on Lake Ananá and Mancapuru Great Lake Amazon datasets.

Fig. S7. Representation of the genomic repertoires of *Synechococcus* spp. and *Cyanobium* spp. belonging to sub-clusters 5.2 and 5.3. The percentages of genes belonging to each of the four categories in which core and flexible genomes are divided are shown. Genomes are coded according to isolation source origin, with MAGs in blue.

Fig. S8. Genomic islands inferred from recruitment plots of *C. usitatum* Tous (left) on a Baltic Sea metagenome (SRR2053292) and *S. lacustris* Tous (right) on a Lake Lanier T2 metagenome. The Y axis represents the % identity of the reads to the metagenome. The X axis represents the position of each artificial concatemer (genome) ordered by contig length.

Supplementary Datasheet 1. General features of the different picocyanobacteria and metagenomics datasets used in this work.

Supplementary Datasheet 2. Pan-genome of the different picocyanobacteria from sub-clusters 5.2 and 5.3. Each genome presents four gene categories (strict-core, soft-core, shell and cloud).