Title: Primer design for an accurate view of picocyanobacterial community structure using high-throughput sequencing

Running title: Primer design for picocyanobacteria

Paula Huber\textsuperscript{a\!*}, Francisco M. Cornejo-Castillo\textsuperscript{b,c}, Isabel Ferrera\textsuperscript{b,d}, Pablo Sánchez\textsuperscript{b}, Ramiro Logares\textsuperscript{b}, Sebastián Metz\textsuperscript{a}, Vanessa Balague\textsuperscript{b}, Silvia G. Acinas\textsuperscript{b}, Josep M. Gasol\textsuperscript{b,e} and Fernando Unrein\textsuperscript{a}

\textsuperscript{a}Instituto de Investigaciones Biocientíficas-Instituto Tecnológico de Chascomús, UNSAM-CONICET, Chascomús, Buenos Aires, Argentina.

\textsuperscript{b}Department of Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC, Barcelona, Catalunya, Spain.

\textsuperscript{c}Ocean Sciences Department, University of California, Santa Cruz (UCSC), CA, USA.

\textsuperscript{d}Centro Oceanográfico de Málaga, Instituto Español de Oceanografía, Fuengirola, Málaga, Spain.

\textsuperscript{e}Centre for Marine Ecosystems Research, School of Science, Edith Cowan University, Joondalup, WA, Australia.

* To whom correspondence should be addressed. Tel/Fax +54 342 4511645/48; Email: mariapaulahuber@gmail.com; Address: Laboratorio de Plancton, Instituto Nacional de Limnologia. Ciudad Universitaria - Paraje El Polo, Santa Fe, CP. 3000. Santa Fe, Santa Fe, Argentina.

\textsuperscript{1} Present address: Instituto Nacional de Limnología (INALI), UNL-CONICET. Santa Fe, Santa Fe, Argentina.
ABSTRACT: High-throughput sequencing (HTS) of the 16S rRNA gene has been used successfully to describe the structure and dynamics of microbial communities. Picocyanobacteria are important members of bacterioplankton communities, and so far, they have predominantly been targeted using universal bacterial primers providing a limited resolution of the picocyanobacterial community structure and dynamics. To increase such resolution, the study of a particular target group is best approached with the use of specific primers. Here we aimed to design and evaluate specific primers for aquatic picocyanobacterial genera to be used with high throughput sequencing.

Since the various regions of the 16S rRNA gene have different degrees of conservation in different bacterial groups, we therefore first determined which hypervariable region of the 16S rRNA gene provides the highest taxonomic and phylogenetic resolution for genera *Synechococcus*, *Prochlorococcus* and *Cyanobium*. An in silico analysis showed that the V5-V7 hypervariable regions appear to be the most informative for this group. We then designed primers flanking these hypervariable regions and tested them in natural marine and freshwater communities. We successfully detected that most (97%) of the obtained reads could be assigned to picocyanobacterial genera. We defined operational taxonomic units as exact sequence variants (eOTUs), which allowed us to detect higher genetic diversity and infer ecologically-relevant information about picocyanobacterial community composition and dynamics in different aquatic systems. Our results open the door to future studies investigating picocyanobacterial diversity in aquatic systems.

Key Words: *Synechococcus*, *Cyanobium*, *Prochlorococcus*, 16S rRNA gene; primer design, HTS.
IMPORTANCE: The molecular diversity of the aquatic picocyanobacterial community cannot be accurately described using only the available universal 16S rRNA gene primers that target the whole bacterial and archaeal community. We show that the hypervariable regions V5-V7 of the 16S rRNA gene are better suited to study the diversity, community structure and dynamics of picocyanobacterial communities at a fine scale using Illumina MiSeq sequencing. Due to its variability, it allows reconstructing phylogenies featuring topologies comparable to those generated when using the complete 16S RNA gene sequence. Further, we successfully designed a new set of primers flanking the V5-V7 regions whose specificity for picocyanobacterial genera was tested in silico and validated in several freshwater and marine aquatic communities. This work represents a step forward for understanding the diversity and ecology of aquatic picocyanobacteria and sets the path for future studies on picocyanobacterial diversity.
Introduction

Picocyanobacteria are common components of plankton and, together with other groups of cyanobacteria, are the only prokaryotes capable of performing oxygenic photosynthesis. They have been present on Earth since 3.5 billons of years (1), and along their history they have evolved ecological strategies to adapt to various environmental conditions, from the poles to the tropics (2–7). Picocyanobacteria contribute substantially to the primary production of numerous aquatic systems and play a major role in aquatic ecosystem energy flow as they are at the base of food webs (8–11). Phenotypically, they are single or colonial rod-shaped cells ranging from 0.2 to 3 µm and, based on their phylogeny, belong to three genera: *Synechococcus*, *Prochlorococcus* and *Cyanobium* (12). *Prochlorococcus* is the sister group of the marine *Synechococcus* and has been divided into distinct clades with ecologically relevant differences in light adaptation (HL: high light or LL: low light intensities) and nutrient utilization (13, 14). *Synechococcus* is a polyphyletic genus containing marine and freshwater taxa (1, 15). Phylogenetic reconstructions resolved three sub-clusters (named 5.1, 5.2, and 5.3) with marine representatives (16–18), and at least 13 clades of non-marine picocyanobacteria (19). The 5.1 cluster encompasses most marine groups but some freshwater strains have also been affiliated to clusters 5.2 and 5.3 (15, 19). Less information is available about the *Cyanobium* clade. It is a cosmopolitan group represented by strains from systems with contrasting limnological characteristics (i.e from shallow hypertrophic to deep and oligotrophic lakes), and based on 16S rRNA gene phylogenetic reconstruction, is closely related to some freshwater *Synechococcus* clusters (19, 20).

As it occurs with most prokaryotes, picocyanobacteria lack obvious morphological differences and, consequently, the use of molecular techniques is essential to discriminate the distinct taxa (13, 21, 22). During the last decade, new molecular and bioinformatic techniques have changed our perspective of microbial diversity. Particularly, high-throughput sequencing (HTS) techniques applied to genetic markers such as the 16S rRNA gene, have allowed a better
understanding of the structure of bacterial communities and their response to environmental
changes (e.g. 23–27). To date, most researchers have used the 16S rRNA gene universal primers
targeting the whole bacterial community to explore bacterial diversity. However, in these works, the
number of reads recovered from Cyanobacteria phylum hardly exceed 20% of the total read
abundance (25, 27–31). Given that natural environments are composed of a few dominant taxa,
using universal primers may overlook the diversity and population genetic structure of less
abundant organisms even when using deep-sequencing. To increase such genetic resolution, the use
of specific primers is the best approach to the study a particular target group (32); and this was the
main goal of our work in order to get a more realistic picture on the diversity and structure of
picocyanobacteria in natural communities. Accurate information of picocyanobacterial genetic
diversity has been obtained using high resolution marker genes such as petB, rbcL, rpoC, cpcBA,
mpeBA, and mpeW genes and the 16S-23S rRNA internal transcribed spacer (ITS) (4, 6, 15, 17, 18,
33–36). However, this information is strongly biased towards marine picocyanobacterial strains,
leaving the freshwater species still under sampled on databases; thus, a comprehensive study of all
aquatic cyanobacterial using such functional markers is not currently possible. As alternative, using
the 16S rRNA gene as marker is still the best-suited approach.

In fact, a 16S rRNA gene primer pair specific for coccoid Cyanobacteria
CYA359F/CYA781(b)R (37) was developed and used in several DGGE (Denaturing Gradient Gel
Electrophoresis) fingerprinting studies in which its specificity was verified (38–41). Particularly,
Boutte et al. (42) observed that the primers designed by Nübel et al. (37) were the most specific for
coccoid Cyanobacteria. These primers flank the V2-V4 regions of the 16S rRNA gene,
comprising ca. 500 base pairs (bp), so that they could potentially be used in HTS. However, the
16S rRNA gene is a mosaic of conserved and variable regions with different degrees of
conservation in different prokaryotic groups (43) and, consequently, the choice of the region for
diversity studies requires careful consideration. The most used variable regions for the
characterization of prokaryotic communities are the V4, V3-V4, V4-V5 and V6 regions (e.g.(26,
44), but it has never been evaluated which hypervariable region is the most suitable for its use in picocyanobacterial-specific diversity studies.

With the aim of unveiling the hidden diversity within the picocyanobacteria and in order to establish a systematic approach based on 16S rRNA gene iTags that would facilitate evaluation of picocyanobacterial community structure, we: i) determined which hypervariable region of the 16S rRNA gene provides the highest taxonomic and phylogenetic information for genera Synechococcus, Prochlorococcus and Cyanobium, ii) designed specific 16S rRNA gene primers for these genera to be used in the Illumina MiSeq sequencing platform, and iii) evaluated the specificity and performance of the newly designed primers in marine (Blanes Bay, NW Mediterranean) and freshwater (Chascomús, Carpincho, Bragado and Monte lakes, the Pampa plain, Argentina) systems.

Results

Defining the most adequate hypervariable region.

To evaluate which 16S rRNA gene region is the most informative in picocyanobacteria we first created in silico libraries using public picocyanobacterial sequences. Two 16S rRNA gene libraries were constructed: Full Length Sequence (FLS) and Short Length Sequence (SLS) libraries, the first one contained almost the complete 16S rRNA gene (ca. 1400 bp). These FLS were then used as seed to construct the SLS libraries. Two types of SLS libraries were generated: one with 16S rRNA gene fragments containing the hypervariable regions V2-V4 (SLS_{V2-V4}) and another one with fragments containing the V5-V7 regions (SLS_{V5-V7}) (Table S1).

To test for different degrees of variability in the hypervariable regions for different clades, we organized the sequences into five phylogenetic groups: three contained sequences of Synechococcus (designated Syn and listed with consecutive numbers), one contained sequences of Cyanobium
(Cyab) and another one contained sequences of *Prochlorococcus* (Proch) (Table S1). Thus, a total
of six FLS libraries were constructed: one containing all the sequences (FLS-Pcy) and one for each
phylogenetic group (Table S1). The 16S rRNA gene fragments of SLS were extracted from each of
the 6 FLS libraries, resulting in a total of 12 SLS libraries (Table S1). Once the FLSs and SLSs
libraries were ready, we constructed maximum likelihood (ML) phylogenetic trees and analyzed (1)
the genetic variability and (2) the phylogenetic information contained in each of the V2-V4 and the
V5-V7 variable regions.

The genetic variability along the different 16S rRNA gene regions was explored through the
Faith phylogenetic diversity index (PFDaith) (45) and the Mean-Nearest-Taxon-Distance (MNTD)
(46). The PFDaith is defined as the sum of the minimum lengths of all branches in a phylogenetic
tree (45), while the MNTD calculates the phylogenetic distance between each taxon within a
community and its closest relative (46). Higher values of both metrics indicate greater phylogenetic
diversity but MNTD is more sensitive to variations of closely-related taxa than PFDaith (45). In this
study, the higher values of PFDaith were observed in the trees constructed with the Short Length
Sequence (SLS) libraries containing the V5-V7 regions (i.e. SLS$_{V5\text{-}V7}$) and were significantly higher
than those containing the V2-V4 regions (i.e. SLS$_{V2\text{-}V4}$) (Fig. 1a), indicating larger genetic
variability in the V5-V7 16S rRNA gene hypervariable regions for picocyanobacterial taxa. The
same pattern was observed when phylogenetic groups were analyzed separately (except for Syn2
group) (Fig.1a). The results obtained with the MNTD index, were in agreement with those obtained
with the PFDaith index (Fig.1b). The highest values were observed in the SLS$_{V5\text{-}V7}$ libraries, being
significantly higher for SLS-Pcy (all picocyanobacterial sequences) and SLS-Proch
(*Prochlorococcus* sequences). Thus, these results confirm that the fragment containing the V5-V7
hypervariable regions features more genetic variability than that of the V2-V4 hypervariable
regions.
To evaluate the phylogenetic information contained in the different regions, we compared the topology of the phylogenetic trees constructed from the Short and Full Length Sequence libraries. First, we calculated the patristic distances (i.e. the sum of branches lengths that link two tips in the tree, PDist) between all possible pairs of sequences in each constructed phylogeny and correlated each PDist of the SLS trees with those of the FLS trees (see methodology). The PDist correlations were always higher for FLS and SLS containing the V5-V7 hypervariable region (SLS_{V5-V7}) than for those in the SLS_{V2-V4}. This occurred when using all the sequences, and also when each phylogenetic group was studied separately (Table 1, Fig.2). We also calculated the Robinson Foulds (RF) and weighted Robinson Foulds (WRF) indices (47). Both metrics count the number of different bipartitions between two trees. The lower the values, the more similar the trees are with respect to their topology. The difference between the RF and WRF indices is that the latter takes into account the bootstrap support values of the bipartitions instead of looking only at their presence or absence, thus bifurcations with lower support are less penalized. In all cases the RF and WRF indices were significantly lower for the V5-V7 fragment than for the V2-V4 (Table 1), indicating that the number of different bipartitions between the FLS and SLS trees occurred to a lesser degree in FLS vs. SLS_{(V5-V7)} than in FLS vs. SLS_{(V2-V4)} trees (Table 1). In addition, the WRF values were much lower than the values of their respective RFs, indicating that the different bipartitions occurred in nodes with low bootstrap values. Although this was observed for both regions, the lowest values were between FLS and SLS_{(V5-V7)}. To further strengthen our results, we evaluated the congruency between FLS and SLS tree topologies by the F-measure defined as the harmonic mean of precision and recall (see methodology). The F values range from zero for fully dissimilar trees to a maximum value for fully similar trees, which value approaches one as the topologies of both trees become similar. For all compared trees, the SLS_{V5-V7} libraries showed high F-measure values compared to SLS_{V2-V4} libraries, supporting the results described above (Table 1).

Further, to test whether the topologies of the constructed SLS trees were not a methodological artifact generated by sequence length, we correlated the PDist of the three trees constructed from
Design of specific primers for picocyanobacteria. Based on the previous results, and in order to cover Cyanobacteria of the genera *Synechococcus*, *Prochlorococcus* and *Cyanobium*, we designed a set of primers flanking the hypervariable regions V5-V7 of the 16S rRNA gene (fragment size: 520 pb): Cya-771F (5'-AGGGGAGCGAAAGGATTA-3') and Cya-1294R (5'-GCCTACGATCTGAAGTGCAGC-3'). *In silico* analysis (Table 2) demonstrated a high specificity for the target sequences, i.e. more than 99% of the total sequences recovered by the primers Cya-771F/1294R belong to the three cyanobacterial picoplanktonic genera. Additionally, from 90.67% to 94.28% of total sequences within each genus (i.e. *Prochlorococcus*, *Synechococcus* and *Cyanobium*) could be recovered with our primer set, indicating a high coverage rate for these specific picocyanobacterial genera (Table 2). We also compared our results with the coverage range and specificity obtained with the primers proposed by Nübel et al. (37) specific for coccoid Cyanobacteria. When perfect matches were considered (i.e. 0 mismatch), our set of primers displayed higher specificity and coverage for the sequences of the targeted organisms (Table 2), while a higher number of sequences for other Cyanobacteria genera (e.g. *Leptolyngbyales*, *Nostocales*, *Oxyphotobacteria* and *Phormidiales*) were recovered using the Nübel et al. (37) primers. Furthermore, when 1 and 2 mismatches were allowed, both sets of tested primers increased their target group sequence coverage and decreased their specificity, as expected (Table S2 and Table S3). However, when considering 2 mismatches Nübel et al. (37) primers specificity declined from 99.7% to 97.1%, while the change in our primers was lower (from 99.9% to 99.2%) (Table S3). Finally, we calculated the coverage and specificity of two sets of universal prokaryotic primers broadly used in microbial communities studies: Bact_341F (5'-CCTACGGGNGGCWGGCAG-3') - Bact_805R (5'-GACTACHVGGGTATCTATCC-3') proposed by Herlemann et al. (48) and 515F-Y (5'-GTGYYCAGCMGCGCGTAA-3') - 926R (5'-GTYCAGCMGCGCGTAA-3') from Parada et al. (26). Regardless of the number of mismatches allowed (0,1 or 2), the coverage ranges for the
targeted organisms were similar to those obtained with the specific primers; however, the specificity values obtained with universal primers were, in both cases, considerably lower (Table S4 to S9 and Fig.S2).

**Testing the primers with environmental samples.** The specificity and usefulness of the designed primers were tested in both freshwater and marine samples using Illumina sequencing. We obtained a total of 21,477 and 56,610 high quality reads from the marine (Blanes Bay) and freshwater (Argentinian lakes) samples (see Material and Methods), respectively. All sequences from Blanes Bay were assigned either to the genera *Synechococcus* or *Prochlorococcus* and 96.3% of the reads retrieved from lakes belonged the genera *Synechococcus* (32%), *Cyanobium* (12%) and environmental clones closely related to them (51%) (Table S10). The remaining percentage was assigned to filamentous Cyanobacteria (2.3%), of the *Cylindrospermosis* and *Leptolythra* genera, as well as to heterotrophic bacteria (1.3%), particularly Planctomycetales and Verrumicrobiales. Rarefaction curves showed an asymptotic tendency only in two out of the ten samples (in Chascomús lake during the Spring of 2012 and in the Summer sample of Blanes Bay), while in the rest of the samples, picocyanobacterial diversity did not reach saturation with the sequencing effort used in this work (Fig.S3). A total of 113 zOTUs within the picocyanobacterial genera were defined, with 87 zOTUs being exclusive of freshwaters and 26 zOTUs being exclusive of the marine site (Blanes Bay).

The samples from Blanes Bay correspond to two different periods of the year (spring and summer) when peaks of picocyanobacteria are frequently registered. Out of 26 zOTUs, 24 belonged to *Synechococcus* and only 2 displayed high similarity with sequences of *Prochlorococcus* (Fig.3), which contributed with less than 2% to the total abundance in each sample (*Prochlorococcus* are more abundant in the Fall-Winter season in this site (49)). The taxonomic composition was very similar in both samples being only three zOTUs exclusive of one season (zOTU-91 and zOTU-117 from summer, and zOTU-101 from spring), and the community in summer showed a higher
To be submitted to *Applied and Environmental Microbiology*

diversity (Shannon = 2.4; PD\textsubscript{Faith} = 1.8) than in spring (Shannon = 1.5; PD\textsubscript{Faith} = 1.6). Despite community structure being significantly different between seasons (t-student, p<0.001), there was a single sequence (zOTU-9) that dominated in both seasons, contributing 63.4% of the total number of reads in spring and 30.6% in summer (Fig.3). All *Synechococcus* zOTUs belonged to the marine sub-cluster 5.1, clustering into five phylogenetic clades defined in the reference tree: I, IV, V, CRD1 and WPCI (Fig.3 and S3). In both samples the most abundant lineage was clade I, followed by clades WPCI and IV, which exhibited higher relative contribution in summer (Fig.4a). Regarding the *Prochlorococcus* zOTUs, both of them belonged to clade HLI and were closely related to the MED4 strain retrieved from the Mediterranean Sea (Fig.3; Fig.S4).

A total of 86 zOTUs were obtained from the five seasonal samples taken from Chascomús lake (CH). Richness varied between 66 (spring 2012) and 84 zOTUs (fall 2014), and the Shannon index varied little between samples (Table 2). However, the phylogenetic diversity (PD\textsubscript{Faith}) was notably lower in spring 2012 (Table 2). A cluster analysis revealed higher similarity in the picocyanobacterial community structure in samples within the same year independently of the season (Fig.5a). None of the environmental variables measured were significantly associated with community structure. In each sample, the abundant fraction of the communities was represented by 22 to 25 zOTUs (Table 2), accounting for more than 80% of total reads. Most zOTUs (67.8%) were found all year long. In particular, zOTU-2 was always present at >5% of relative abundance (Fig.5a). Most zOTUs clustered into 11 clades previously defined in the reference tree (Fig.S5 and 5a), while 4 new clades were formed exclusively by the newly obtained zOTUs (the new clades were named 1 to 4 in Fig.S5). zOTUs composing the core group (those that contributed with more than 5% of the reads) belonged to different phylogenetic clades, and their contribution changed markedly among samples (Fig.4b, 5a). For instance, clade P\textsubscript{1} was prevalent in spring-winter, clade P\textsubscript{1} in summer 2014 and clade A was more represented in 2014 (Fig.4b).
In the spatial comparison between the four Pampean lakes, zOTU richness varied between 78 in Chascomús and 59 in Bragado lakes. Bragado lake also had the lowest taxonomic and phylogenetic diversity (Table 2). Only 10 out of the total of 85 zOTUs were exclusive from any particular environment. The cluster analysis showed that the Monte and El Carpincho lakes presented a more similar community composition as compared to the other lakes (Fig.5b) and the structure of the picocyanobacterial community was highly correlated with the quality of light (expressed as the relation between Kd_red/Kd_green, Table S11) and total phosphorus concentration (Mantel’s test $r = 0.83$, $p<0.05$). The number of abundant zOTUs (i.e. relative abundances $>1\%$) varied between 13 (Bragado) and 30 (El Carpincho) and, in all cases, they contributed with more than 84% to the total reads of each lake (Table 2). The zOTUs were affiliated with the 13 clades previously described in the temporal study of Chascomús (Fig.S6). The dominant zOTUs of each environment belonged to different clades (Fig.5b, S5). The most notable case was clade 4, which was almost exclusively represented by sequences of Bragado (Fig.4c). Clade P\textsubscript{VII} was also more prevalent in this lake. Clade P\textsubscript{I} showed the highest relative abundance in Chascomús and El Carpincho lakes, and clade A was dominant in Monte Lake (Fig.4c).

**Discussion**

We found that the hypervariable regions V5-V7 of the 16S rRNA gene are the most appropriate for studying the diversity, structure and dynamics of picocyanobacterial communities using high-throughput sequencing. Specifically, we found that (1) the V5-V7 region features more variability in picocyanobacteria than other regions, and (2) that the phylogenies based on the V5-V7 region mirror to a larger extent the phylogenies obtained with the complete 16S rRNA gene. Additionally, we successfully designed and used a set of primers flanking the V5-V7 regions, which target almost exclusively *Synechococcus, Prochlorococcus* and *Cyanobium*, thus allowing the accurate characterization of natural picocyanobacterial communities.
We explored the taxonomic information contained in the hypervariable regions of the 16S rRNA gene using indices with different sensitivity to phylogenetic relatedness (i.e. PD\textsubscript{Faith} and MNTD). The results indicate that the V5-V7 region allows for a better discrimination of both, phylogenetically distant, as well as closely related picocyanobacteria (46). Which region is suitable for studying prokaryotic diversity has been mostly explored for entire bacterial communities and other taxonomic groups (26, 50–52), however, to our knowledge, this is the first time that it has been evaluated for picocyanobacterial genera. Reviewing the information present in the literature (26, 50–52), we detected contrasting results about the suitability of 16S rRNA gene hypervariable regions for the study of prokaryotic communities, which alerted us about the importance of selecting a particular hypervariable region of 16S rRNA according to the taxonomic level/group at which the results need to be inferred (53).

The study of phylogenetic relationships (i.e. patristic distance, RF, WRF and F-value) indicated that the information present in fragments of the 16S rRNA gene containing the V5-V7 region, allow the reproduction of the phylogenetic relationships obtained from the full-length 16S rRNA gene for the picocyanobacterial groups. In previous studies the V4 fragment was found to be the best for establishing phylogenetic relationships when the entire bacterial community was considered (26, 52, 54). However, this contrasts with our results, suggesting that the hypervariable region of the 16S rRNA gene that best reflects the full-length gene phylogeny depends on the taxonomic groups being studied. Additionally, we observed a high correlation between tree topologies constructed from the same SLS, which suggests that i) the difference in the correlations of SLS\textsubscript{V2-V4} and SLS\textsubscript{V5-V7} with the FLSs may be attributable to the phylogenetic information present in the fragments instead of the size of the sequences, and ii) the phylogenies obtained from partial sequences have a good reproducibility power.

The specificity of the new set of primers analyzed \textit{in silico} was validated in natural communities. Nevertheless, some discrepancies in the recovered taxa were detected, indicating the
importance performing primer tests on environmental samples. For example, some reads were
assigned to genera that appeared in the in silico analyzes allowing 1 or 2 mismatches (e.g.
*Planctomycetes* and *Verrucomicrobia*), while some genera typical of aquatic systems (e.g.
*Pseudanabaenales, Actinobacteria, Firmicutes*) recovered in the in silico analysis, did not appear in
the zOTU table.

On the other hand, by defining operational taxonomic units as exact sequence variants
(zOTUs) (55) we could achieve a fine-scale taxonomic resolution (55, 56) allowing us to detect a
larger genetic diversity among the picocyanobacteria as well as specific patterns of community
composition in different systems and seasons. In Blanes Bay, *Synechococcus* reads were much more
abundant and diverse than those of *Prochlorococcus*, as it was observed in the cell abundances
(Table S11). The number of operational taxonomic units defined here as zOTU, was higher than
those established in previously studies, where bacterial primers 341F (5’-
CCTACGGGNGGCWGCAG-3’)(48) and 806RB (5’-GGACTACNVGGGTWTCTAAAT-3’) (44)
were used to amplify partial 16S rRNA genes (i.e. 10 zOTU of *Synechococcus* and 1 zOTU of
*Prochlorococcus*, Ferrera & Gasol, pers. commun.). Picocyanobacterial community composition
differed between spring and summer, yet there was a unique zOTU that dominated in both seasons.
This supports observations from other sites that also revealed the presence of a single strain of
*Synechococcus* throughout the year (57, 58), and points towards the existence of a single strain
adapted to grow in different seasons characterized by quite different temperatures and nutrients (i.e.
49). When analyzing the phylogenetic structure of the communities, we observed that most
*Synechococcus* zOTUs correspond to clades typical of coastal zones: clades I, IV, V, and WPC1
(15, 18, 59) that tend to co-occur (e.g 4, 15, 29, 51). Yet, zOTU-17 was assigned to clade CRD1,
which has been described from the Costa Rica upwelling (61) as well as the Pacific, Indian and
Atlantic Oceans, but it is seemingly absent from the Mediterranean Sea (6). Regarding
*Prochlorococcus* zOTUs, both belonged to the HLI clade, a dominant clade in surface waters with
moderate stratification (17, 62–64).
In Chascomús lake there was a core of zOTUs (i.e. relative abundance >5%) that changed their relative contribution between seasons. This pattern is usually observed in temperate lakes regardless of their limnological characteristics (22, 65) and suggests different ecological roles within genetically close strains. However, as in previous reports (22, 65), we failed to relate the distinct niches among the dominant zOTUs with environmental factors such as nutrient concentrations, pH, underwater light or temperature (Table S11). This may suggest that biotic feedback mechanisms, such as competition, viral infection and prey-predator interactions are significant factors driving the temporal dynamics of the picocyanobacterial community, and further studies are needed to shed light on this issue. Regarding the spatial presence of the different zOTUs, each Pampean shallow lake was represented by a particular assemblage associated with local environmental factors. These results lend support to previous evidence about the importance of underwater light conditions and the availability of nutrients in governing freshwater picocyanobacterial distribution (35, 65–69).

Finally, the incorporation of zOTUs into the phylogenetic tree allowed the observation of a large number of coexisting clades with specific abundance dynamics changing over space and time. This has been well reported by others authors (20, 35, 70, 71) and seems to be a common feature in picocyanobacterial communities. What allows the coexistence of these very similar organisms? The great genetic microdiversity within picocyanobacterial genera may effectively be a good explanation. However, how environmental factors drive changes in clade abundances, and whether co-occurring phylotypes within the same and different clades affect each other in beneficial or detrimental ways are questions that still need to be posed.

In summary, we determined that the V5-V7 hypervariable region of the 16S rRNA gene provides more taxonomic and phylogenetic information for *Synechococcus*, *Prochlorococcus* and *Cyanobium* than other hypervariable regions, and designed a set of specific primers for these genera. The results obtained from marine and freshwater environments showed the usefulness of the
present approach and tools to explore the diversity and successional patterns in the picocyanobacterial assemblage via fine (i.e zOTU), or coarse-grained (clade and subclade), taxonomic resolutions.

Materials and Methods

Defining the most adequate hypervariable region

To determine the most informative hypervariable 16S rRNA gene region we followed the methodology proposed by Ghyselinck et al. (52). Genetic libraries were constructed using 1,334 sequences from representative picocyanobacterial clades selected from the public database SILVA 119 (https://www.arb-silva.de/download/arb-files/) (Table S1). The Full Length Sequence (FLS) libraries contained complete sequences with the information of all hypervariable regions (V1-V9).

To generated the Short Length Sequence (SLS) libraries we searched the V4 and V6 hypervariable regions from each FLS and trimmed out fragments of 500 bp in length containing either the V4 or the V6 region using V-Xtractor v 2.0 (72).

For each FLSs and SLSs libraries constructed, we aligned the sequences using MAFFT v.7 with default options (73) and constructed maximum likelihood (ML) phylogenetic trees using FastTree v.2.1.5 (74) with default options using the Geneious v 9.1. interface (75). A single ML tree was constructed for each of the 6 FLSs libraries and 3 trees for each of the 12 SLS libraries using the General Time Reversible (GTR) model with a gamma distributed rate of variation across sites.

To estimate the statistical support, 1,000 pseudo replicates were performed using the 'rapid bootstrapping' algorithm. Three sequences of Gloeobacter sp. were used as out-group. In those analyses in which rootless trees were required, these out-group sequences were discarded. The generated trees were used in downstream analyses.
Genetic variability was explored using the Faith phylogenetic diversity index (PD\text{Faith}) (45) and the Mean-Nearest-Taxon-Distance (MNTD) (46). PD\text{Faith} is defined as the sum of the phylogenetic branch lengths connecting all taxa present in a community; higher values of the PD\text{Faith} would indicate higher phylogenetic diversity and, therefore, larger genetic variability (45). The MNTD quantifies the phylogenetic distance between each taxon within a sample and its closest relative in the sample, returning an average (46). Thus, higher MNTD values indicate larger phylogenetic diversity among closely related taxa. To calculate PD\text{Faith} and MNTD values we used the package Picante (V 1.6) (76) within the R environment.

The phylogenetic information present in the different regions of the 16S rRNA gene was evaluated comparing the tree topology through the patristic distances (PD\text{ist}), the Robinson-Foulds distances (RF) and the Robinson-Foulds weighted distances (WRF) (47). The PD\text{ist} were calculated between all possible pairs of sequences in each phylogeny constructed, and were presented as distance matrices. Then, correlations between each SLS (in triplicate) and FLS distance matrices were tested using Mantel tests based on Pearson correlations. In order to visualize these data, the distances were normalized to a maximum value of one, and ordered according to the distance from the FLS tree at intervals of 0.01. For each interval we calculated the average of the FLS distance and corresponding averaged SLS distances and then plotted them in a graph. The standard deviations of the averaged SLS distances were used as error bars in the chart (54).

Similarly, the RF distances allowed us to compare topologies of phylogenetic trees counting the number of bipartitions that occur in one tree but not in the other, so a lower RF value indicates that both trees are topologically more similar. We also calculated the WRF (47). This measure, unlike the RF, takes into account the bootstrap value of each bipartition instead of considering presence/absence only. For example, if a bipartition is present in one tree but not in another and has a bootstrap value of 0.3, then WRF counts it as 0.3 instead of 1. Thus, by using the WRF, the bifurcations with low support are less penalized. When the WRF values are lower than the
respective RF values, it means that the differences between trees occurred in clades with low bootstrap values. The patristic distances and RF/WRF metrics were calculated in the R environment using the *Stats* and *Phangorn* V2.4 R packages (77).

Finally, to further strengthen our results, we compared the topologies between SLS and FLS trees using the precision and the recall metrics. These measures were defined as: precision = TP/(TP + FP) and recall = TP/(TP + FN), where true-positives (TP) are the set of bipartitions present in both the SLS tree and the FLS tree, FP are the number of bipartitions present in SLS tree but not in FLS tree and FN the number of bipartitions present in FLS tree but not in SLS tree (78). These metrics were calculated in the R environment using the *Phangorn* V2.4 R package (77). The results were expressed as F-measure, defined as the harmonic mean of precision and recall, and balances false positives and false negatives ($F = 2 \times (\text{precision} \times \text{recall})/(\text{precision} + \text{recall})$).

**Primer design**

We designed a set of specific primers for each of the *Synechococcus*, *Cyanobium* and *Prochlorococcus* genus. Using the FSL-Pcy library as reference, we searched possible primers with GENEIOUS v 9.1.5 (75), targeting a region that would start in the conservative area at the beginning of the V5 hypervariable region up to the end of the V7 hypervariable region, with an amplicon size of approximately 540 bp, suitable for Illumina MiSeq sequencing using 2x300 bp. Once the candidate primers were defined, we tested their specificity and coverage *in silico* against the SSU r132 RefNR database allowing 0, 1 and 2 mismatches using the Test Probe 1.0 feature from SILVA ([http://www.arb-silva.de/?id=650](http://www.arb-silva.de/?id=650)).

**Testing the primers with environmental samples**

**Field sampling, PCR and sequencing.** The specificity and usefulness of the designed primers were then tested in two sets of samples:
1- Two marine samples (Table S11) from the coastal Blanes Bay Microbial Observatory in the Mediterranean Sea (41° 39.90’N, 2° 48.03’ E) (49). We selected one sample taken in spring and one in summer 2013 (April and August) based on previous knowledge that during these seasons the abundance of *Synechococcus* reaches maximum values (up to $8 \times 10^4$ cell per milliliter) in this coastal station (e.g. (49, 57, 58)).

2- Eight freshwater samples belonging to four hypertrophic shallow lakes (Table S11) of the Salado River floodplain basin, located in the Pampa plain in Argentina (35° 41’S, 59° 35’O) (22, 69). The samples were selected in order to cover spatial and temporal variation. Three samples were taken during spring 2009 from the Bragado, El Carpincho and Monte lakes. In addition, five samples from Chascomús lake were collected between October 2012 and May 2014 (Table S11). Site features and sampling details are provided in detail in Fermani et al. (69) and Huber et al. (22).

Sample collection and DNA extractions are described in Ferrera et al. (58) for the marine samples, and in Huber et al. (22) for the freshwater samples. The volume of water processed is indicated in Table S11.

Sequencing was performed by the RTL Genomics division of Research and Testing Laboratories of the South Plains (Lubbock, Texas. http://rtlgenomics.com) using Illumina MiSeq 2 X 300 paired-end reads (sequencing runs 1 and 2, respectively), with a sequencing depth of approximately 20,000 sequences per sample. The V5-V7 region of the 16S rRNA gene was amplified using the primers Cya-771F and Cya-1294R designed in this work (see above). The PCR reaction conditions for these primers were as follows: 5-min pre-heating step at 94°C; 10 cycles of denaturation at 94°C for 1 min; an annealing touchdown procedure for 1 min in which the temperature decreased by 1°C every second cycle from 65 to 56°C; followed by 25 cycles of 1 min of denaturation at 94°C; annealing at 55°C for 1 min and extension at 72 °C for 3 min, with a final extension step of 10 min at 72°C. All sequences generated in this study can be accessed through
European Nucleotide Archive (ENA) under the Study PRJEB27291, accession ERR2639355-ERR2639364.

**Data analysis.** Raw sequences were processed using a modified version of the pipeline proposed by Logares (79) (https://github.com/ramalok). The reads were first analyzed for error correction using the algorithms based on Hamming graphs and Bayesian sub-clustering (BAYES HAMMER tool) (80) implemented in SPAdes v3.5.0 (81). Then, the forward and reverse sequences were assembled using the function `fastq_mergepairs` from USEARCH-v10 (82); the minimum overlap length was set to 20 bp and the assemblage sequences with less than 100 nucleotides were discarded, the rest of parameters were used as default. Quality check was performed by `fastq_filter` in USEARCH-v10 (82). Reads that passed the quality control were then analyzed using UNOISE2 (83) to define operational taxonomic units (OTUs) with no clustering (zero-radius OTUs [zOTUs] or amplicon sequence variants [ASVs]). This tool zOTUs provide a higher accuracy than OTUs by achieving single-nucleotide resolution after correcting for Illumina sequencing errors and chimeras (55, 56). Finally, taxonomy assignment of zOTUs was done by BLAST (84), using the SILVA database (SSURef 132 Nr99) as a reference, and the zOTU table was created with the function `otutab` in USEARCH-v10 (82).

zOTUs with a relative abundance below 1% per sample were considered representative of the "rare biosphere", while those that exceeded 1% were considered part of the "abundant fraction" (85). We constructed zOTU rarefaction curves for each environment, using vegan V2.5 to evaluate whether we saturated richness. The taxonomic diversity and richness of the picocyanobacterial communities were estimated with the Shannon (86) and Chao-1 (87) indices, respectively. Bray-Curtis dissimilarity matrices were constructed to analyze the similarities in the composition of the assemblages. The abundance of each zOTUs was previously transformed using Hellinger transformation as recommended by Legendre and Gallagher (88). We also calculated phylogenetic
diversity using the PD\textsubscript{Faith} index, based on maximum likelihood trees constructed using the RAxML v.8 program (89).

We also tested the environmental factors that determine community structure. For that purpose, we first selected possible drivers using the function BIOENV (90) within the R environment (91). After that, we constructed a Euclidian distance matrix with the selected variables and determined its correlation with the picocyanobacterial dissimilarity matrix using a Mantel test. The physical, chemical and biological variables that were used in this analysis are summarized in Table S11.

To infer the phylogenetic positions of the zOTUs, we incorporated the amplicon short sequences into freshwater and marine phylogenetic reference trees (RT). First we aligned the query sequences against the reference alignments using PaPaRa v2.5 (92) and then we placed the queries onto the trees using EPA (93) as implemented in RAxML (89). The RTs were constructed using sequences of the 16S rRNA gene (min. length 700 pb) from the NCBI public database assuring that the main clades of picocyanobacteria were represented (6, 18–20, 22, 94) (Table S12). The marine RT, included 94 sequences: 22 assigned to seven Prochlorococcus clades (HL and LL) and 72 to Synechococcus sub-clusters 5.2, 5.3 and eleven clades of sub-cluster 5.1. For the freshwater RT we selected 181 published sequences and also incorporated 82 unpublished sequences corresponding to environmental clones of the complete 16S rRNA gene retrieved from two Pampean lakes (see Supp. material). Twenty-two clades were defined; 13 have been described in previous studies and 9 were novel, being constituted mostly by taxa retrieved from Pampean systems (designated by the letter P and listed with Roman numerals i.e. PI to PIX); Synechococcus strain PCC 6301 was used as external group. The sequences were aligned using MAFFT v.7 and the RTs were constructed using the maximum likelihood method in RAxML v.8 (89) with the GTR+GAMMA model for freshwater taxa and the GTR+CAT+I model for marine taxa, considering in both cases 1,000 bootstraps. Trees were edited using ITOL (http://itol.embl.de) (95).
Acknowledgments

This study was supported by the Argentinian Network for the Assessment and Monitoring of Pampean shallow lakes (PAMPA2–CONICET), the ANPCyT (PICT-2014-1290 and PICT-2016-1079), and by grants MAGGY (CTM2017-87736-R) to Acinas SG and REMEI (CTM2015-70340-R) to Gasol JM from the Spanish Ministry of Economy and Competitiveness (MINECO). We thank Roberto U. Escaray and Jose Bustingorry (INTECH) for technical support and Dr. M. Devercelli for her advice on an early draft and for critical revision of the manuscript. We thank the members of the Laboratorio de Ecología y Fotobiología Acuática (IIB-INTECH) and the Departament de Biologia Marina i Oceanografia (ICM), and particularly the people involved in sampling the Blanes Bay Microbial Observatory for their support.

References


5. Wood SA, Maier MY, Puddick J, Pochon X, Zaiko A, Dietrich DR, Hamilton DP. 2017. Trophic state and geographic gradients influence planktonic cyanobacterial diversity and


Ban L, Pedro C, Casamayor EO, Scha H, Muyzer G, Schäfer H, Bañeras L, Pedrós-Alió C,


73. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:


Edgar RC. 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
sequencing. bioRxiv 081257.


To be submitted to *Applied and Environmental Microbiology*


Figures

**Figure 1:** Values of PD_{ran} (a) and MNTD (b) indices obtained from phylogenetic trees constructed from each short length sequence library (SLS): one containing all the sequences (Pey) and one for each phylogenetic group (Cyab = Cyanobium, Syn = Synechococcus, Proch = Prochlorococcus). Significant differences between the compared SLSs are indicated with an asterisk.

**Figure 2:** Patristic distances of FLS and SLS\textsubscript{(V2,V4)} trees (upper panel) and FLS and SLS\textsubscript{(V5,V7)} trees (lower panel). The distances were normalized to a maximum value of one, and ordered according to the distance from the FLS tree at intervals of 0.01. Each point represents the average of the FLS distance and corresponding averaged in SLS distances in each interval. The standard deviations of the averaged SLS distances were used as error bars on the chart.

**Figure 3:** Relative abundance of zOTUs determined for the two samples from Blanes Bay (left panel: Summer; right panel: Spring). For each zOTU, the Genbank accession number and the taxonomical and phylogenetic clade assignment is presented.

**Figure 4:** Bar graph showing the relative abundance of zOTUs per clades in a) Blanes Bay; b) Chascomús lake and c) Pampean lakes, based on the phylogenetic reconstruction shown in figures S4-S6.

**Figure 5:** Cluster analysis of zOTU abundance from a) Chascomús lake and b) the four Pampean shallow lakes studied. The heat-maps represent the relative abundance of the abundant zOTUs (i.e abundance ≥ 1%). The asterisks indicate the zOTUs with a relative abundance >5%. For each
zOTU the Genbank accession number and the taxonomical and phylogenetic clade assignment are presented.
### Table 1: Summary of measures and indices used to compare the phylogenetic information in trees constructed using full-length sequence (FLS) and short length sequence (SLS) libraries.

<table>
<thead>
<tr>
<th>Library</th>
<th>Region</th>
<th>Replicate of SLS</th>
<th>Patristic PC</th>
<th>RF</th>
<th>WRF</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLS-Pcy</td>
<td>SLS-Pcy</td>
<td>a</td>
<td>0.546</td>
<td>2460</td>
<td>46.777</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>(V2-V4)</td>
<td>b</td>
<td>0.544</td>
<td>2462</td>
<td>47.606</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.548</td>
<td>2433</td>
<td>46.880</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>SLS-Pcy</td>
<td>a</td>
<td>0.612</td>
<td>2402</td>
<td>44.327</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>(V5-V7)</td>
<td>b</td>
<td>0.544</td>
<td>2418</td>
<td>45.285</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.712</td>
<td>2412</td>
<td>43.970</td>
<td>0.089</td>
</tr>
<tr>
<td>FLS-Cyab</td>
<td>SLS-Cyab</td>
<td>a</td>
<td>0.442</td>
<td>132</td>
<td>1.500</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>(V2-V4)</td>
<td>b</td>
<td>0.542</td>
<td>133</td>
<td>1.580</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.426</td>
<td>134</td>
<td>1.600</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>SLS-Cyab</td>
<td>a</td>
<td>0.541</td>
<td>120</td>
<td>1.470</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>(V5-V7)</td>
<td>b</td>
<td>0.441</td>
<td>122</td>
<td>1.460</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.440</td>
<td>122</td>
<td>1.500</td>
<td>0.102</td>
</tr>
<tr>
<td>FLS-Syn1</td>
<td>SLS-Syn1</td>
<td>a</td>
<td>0.190</td>
<td>1116</td>
<td>9.191</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>(V2-V4)</td>
<td>b</td>
<td>0.197</td>
<td>1114</td>
<td>8.890</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.207</td>
<td>1110</td>
<td>8.862</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>SLS-Syn1</td>
<td>a</td>
<td>0.318</td>
<td>1104</td>
<td>8.620</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>(V5-V7)</td>
<td>b</td>
<td>0.311</td>
<td>1107</td>
<td>8.624</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.323</td>
<td>1102</td>
<td>8.590</td>
<td>0.054</td>
</tr>
<tr>
<td>FLS-Syn2</td>
<td>SLS-Syn2</td>
<td>a</td>
<td>0.712</td>
<td>418</td>
<td>6.967</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>(V2-V4)</td>
<td>b</td>
<td>0.6936</td>
<td>417</td>
<td>6.780</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.7121</td>
<td>420</td>
<td>6.770</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>SLS-Syn2</td>
<td>a</td>
<td>0.763</td>
<td>402</td>
<td>6.480</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(V5-V7)</td>
<td>b</td>
<td>0.7631</td>
<td>408</td>
<td>6.210</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.8031</td>
<td>404</td>
<td>6.490</td>
<td>0.098</td>
</tr>
<tr>
<td>FLS-Syn3</td>
<td>SLS-Syn3</td>
<td>a</td>
<td>0.456</td>
<td>46</td>
<td>0.360</td>
<td>0.281</td>
</tr>
<tr>
<td></td>
<td>(V2-V4)</td>
<td>b</td>
<td>0.569</td>
<td>43</td>
<td>0.340</td>
<td>0.317</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.490</td>
<td>45</td>
<td>0.330</td>
<td>0.285</td>
</tr>
<tr>
<td></td>
<td>SLS-Syn3</td>
<td>a</td>
<td>0.725</td>
<td>0</td>
<td>0.000</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>(V5-V7)</td>
<td>b</td>
<td>0.689</td>
<td>2</td>
<td>0.090</td>
<td>0.968</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.698</td>
<td>1</td>
<td>0.087</td>
<td>0.984</td>
</tr>
<tr>
<td>FLS-Proch</td>
<td>SLS-Proch</td>
<td>a</td>
<td>0.544</td>
<td>606</td>
<td>10.834</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>(V2-V4)</td>
<td>b</td>
<td>0.573</td>
<td>603</td>
<td>11.021</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.551</td>
<td>607</td>
<td>10.720</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>SLS-Proch</td>
<td>a</td>
<td>0.737</td>
<td>576</td>
<td>10.272</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>(V5-V7)</td>
<td>b</td>
<td>0.739</td>
<td>575</td>
<td>10.190</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c</td>
<td>0.737</td>
<td>577</td>
<td>10.212</td>
<td>0.124</td>
</tr>
</tbody>
</table>

a) Patristic PC: Pearson correlation of patristic distances assessed by Mantel Test. In all cases p<0.05.
b) RF: Robinson-Fould.
c) WRF: Weighted Robinson-Fould.
d) F-measure: 2 × ((precision × recall)/(precision+recall)).
Table 2. In silico evaluation of specificity and coverage of the primers pairs designed in this work compared with those designed by Nübel et al. (37) with zero mismatches.

<table>
<thead>
<tr>
<th>Taxonomy</th>
<th>this work</th>
<th>Nübel et al.</th>
<th>this work</th>
<th>Nübel et al.</th>
<th>this work</th>
<th>Nübel et al.</th>
<th>this work</th>
<th>Nübel et al.</th>
<th>this work</th>
<th>Nübel et al.</th>
<th>this work</th>
<th>Nübel et al.</th>
<th>this work</th>
<th>Nübel et al.</th>
<th>this work</th>
<th>Nübel et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Bacterio</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Bateria</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Bacteria</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 3: Number and abundance of zOTUs per sample and values of the different indices used for the characterization of the communities.

<table>
<thead>
<tr>
<th>System</th>
<th>Sample date</th>
<th>zOTUs</th>
<th>zOTUs &gt;1%&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reads &gt;1%&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Shannon</th>
<th>chao1</th>
<th>PD&lt;sub&gt;Faith&lt;/sub&gt; &lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marine samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blanes Bay</td>
<td>Spring (Apr-13)</td>
<td>24</td>
<td>10</td>
<td>96.5</td>
<td>3.44</td>
<td>21</td>
<td>1.6</td>
</tr>
<tr>
<td>Blanes Bay</td>
<td>Summer (Aug-13)</td>
<td>25</td>
<td>14</td>
<td>94.2</td>
<td>3.69</td>
<td>22</td>
<td>1.8</td>
</tr>
<tr>
<td>Chascomús</td>
<td>Spring (Oct-12)</td>
<td>66</td>
<td>24</td>
<td>86.5</td>
<td>3.32</td>
<td>71</td>
<td>0.28</td>
</tr>
<tr>
<td>Chascomús</td>
<td>Winter (Jun-13)</td>
<td>72</td>
<td>22</td>
<td>88.9</td>
<td>3.12</td>
<td>85</td>
<td>0.46</td>
</tr>
<tr>
<td>Chascomús</td>
<td>Spring (Nov-13)</td>
<td>76</td>
<td>24</td>
<td>84.5</td>
<td>3.37</td>
<td>87</td>
<td>0.43</td>
</tr>
<tr>
<td>Chascomús</td>
<td>Summer (Feb-14)</td>
<td>83</td>
<td>25</td>
<td>85.2</td>
<td>3.48</td>
<td>86</td>
<td>0.49</td>
</tr>
<tr>
<td>Chascomús</td>
<td>Autumn (May-14)</td>
<td>84</td>
<td>25</td>
<td>80.2</td>
<td>3.59</td>
<td>86</td>
<td>0.49</td>
</tr>
<tr>
<td>Freshwater</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>El Carpincho</td>
<td>Spring (Dic-09)</td>
<td>59</td>
<td>17</td>
<td>87.0</td>
<td>2.79</td>
<td>73</td>
<td>0.26</td>
</tr>
<tr>
<td>Bragado</td>
<td>Spring (Dic-09)</td>
<td>78</td>
<td>15</td>
<td>84.2</td>
<td>3.37</td>
<td>80</td>
<td>0.44</td>
</tr>
<tr>
<td>Monte</td>
<td>Spring (Dic-09)</td>
<td>68</td>
<td>14</td>
<td>91.3</td>
<td>3.39</td>
<td>71</td>
<td>0.45</td>
</tr>
<tr>
<td>Chascomús</td>
<td>Spring (Nov-13)</td>
<td>65</td>
<td>19</td>
<td>89.4</td>
<td>3.12</td>
<td>69</td>
<td>0.36</td>
</tr>
</tbody>
</table>

a) Number of zOTUs conforming the “abundant fraction” in each sample (i.e. abundance >1%)
b) Percentage contribution of zOTU abundant fraction to the total reads in each sample (i.e. abundance >1%)
c) Faith phylogenetic diversity index