

Article

Not peer-reviewed version

---

# The selectivity of 16S primers influences the results of the diversity of cyanobacteria from small water bodies

---

[Łukasz Łach](#) , Nataliia Khomutovska , [Jan Kwiatowski](#) , [Iwona Jasser](#) \*

Posted Date: 18 April 2024

doi: 10.20944/preprints202404.1223.v1

Keywords: cyanobacteria; NGS; microbial mats; novel primer pairs V3-V4; novel primer pairs V4-V6; novel primer pairs V6



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# The Selectivity of 16S Primers Influences the Results of the Diversity of Cyanobacteria from Small Water Bodies

Łukasz Łach <sup>1</sup>, Nataliia Khomutovska <sup>1,2</sup>, Jan Kwiatowski <sup>1</sup> and Iwona Jasser <sup>1,\*</sup>

<sup>1</sup> Institute of Environmental Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland; l.lach@uw.edu.pl

<sup>2</sup> Department of Plant Protection Biology, Swedish University of Agricultural Sciences, 23422 Lomma, Sweden; nataliia.khomutovska@slu.se

\* Correspondence: i.jasser@uw.edu.pl

**Abstract:** The majority of investigations on microbial communities from various environments are presently built on culture-independent methods. Many studies point to the pivotal, selective role of primers targeting hypervariable regions of 16S rRNA in the metabarcoding of bacteria, including cyanobacterial communities. The selectivity of primers designed to amplify targeted regions of the 16S rRNA gene, which has been highlighted by many authors, limited effective amplification. Moreover, the type and specificity of the studied material can also negatively influence the results of 16S metabarcoding. Most of the studies of cyanobacterial communities have been performed for planktonic microbial communities that are often represented by common, well-studied species. In this study, we present the results of 16S metabarcoding analysis using three primer pairs - two already well-known and a third designed in this study - that amplify divergent regions of the 16S rRNA gene (V3-V4, V4-V6, V6) for benthic, microbial mat-forming cyanobacteria communities. Such communities can be a source of toxigenic cyanobacterial taxa and should be monitored with adequate primers. The comparison of three primer pairs suggested that those designed within the present study describe the structure and composition of highly heterogeneous cyanobacterial mats' communities better than the others

**Keywords:** cyanobacteria; NGS; microbial mats; novel primer pairs; V3-V4; V4-V6; V6

## 1. Introduction

Advanced sequencing technologies and metagenomic-based techniques, which appeared during the last decade have created great opportunities for the exploration of the taxonomic and functional diversity of microorganisms on Earth. However, many authors point out that primer specificity and sensitivity but universality on the other hand is key to the effective amplification of targeted Bacteria and Archaea [1–9]. Our studies of benthic [10,11] and soil-inhabiting [12–15] as well as endolithic microbial communities [10,11,16,17] in mountain desert ecosystems, indicated that when using universal primers designed by Klindworth and co-authors [2] targeting the V3-V4 region of 16S rRNA primers some information concerning representative genera of cyanobacteria occurring in these communities can be lost. To our knowledge, up to now most of the primers targeting the 16S rRNA gene and the hypervariable regions have been tested to investigate cyanobacteria from pelagic habitats selecting phytoplanktonic, often bloom-forming, and potentially toxic cyanobacteria [10–13,17–21]. Besides studies of planktonic cyanobacteria, however, investigations of benthic cyanobacteria become more relevant due to the presence of potentially toxic cyanobacteria [22,23].

Microbial mats are often dominated by cyanobacteria that thrive in water in a benthic community, at the border between aquatic and terrestrial environments, and even in arid places. Cyanobacteria prevailing in such communities may produce cyanotoxins, thus posing a health risk

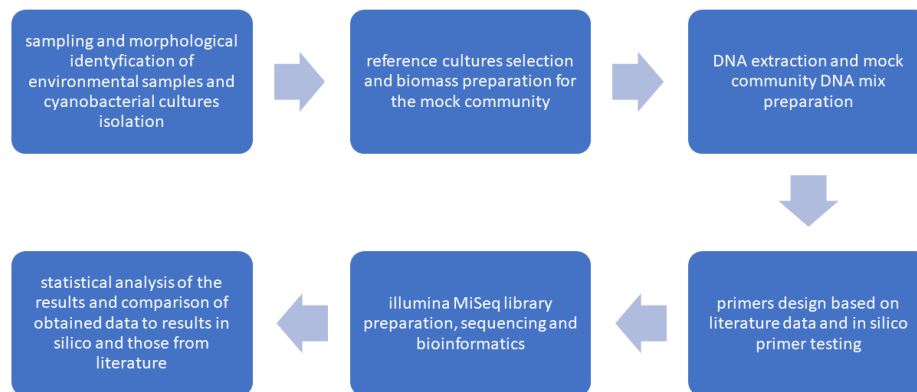
for humans and animals using that water [14,18,22–25]. Advanced molecular techniques based on good quality long reads are still costly, and the analyses take a longer time, while amplicon-based identification is much faster belonging to low-cost methods. Apart from the advantages, there are some challenges associated with fast amplicon-based analysis. One of them is the presence of the chemical substances (i.e., humic compounds and exopolymers produced by microorganisms) in the environmental samples that can be considered the inhibitors influencing DNA extraction. The presence of the inhibitors can also influence the genes' amplification, thus complicating metabarcoding-based studies. Microbial mats, soil-inhabiting, and endolithic communities, which are characterized by a much higher density of the environment than planktonic samples, may potentially accumulate these compounds which may, even more, hinder the molecular analyses. For this reason, special reagents and DNA extraction kits are suggested for such analyses [26]. However, these communities are also largely inhabited by different sets of organisms than planktonic environments and it is important to find primers that will amplify products in such difficult circumstances. As has been reported, microbial mats from Eastern Pamir are characterized by vast morphological and taxonomic diversities [11]; thus, the present study was performed to test and optimize amplicon-based identification of cyanobacteria in exemplary natural microbial mats' communities growing under heterogeneous conditions in the extreme environment such as the cold mountain desert of Eastern Pamir. Additionally, a mock community composed of various benthic cyanobacteria isolated from the studied environment as well as some planktonic, well-described cyanobacteria were chosen for the study.

We evaluated the amplification of three primer pairs; primers designed by Klindworth and co-authors (2013) targeting the V3-V4 hypervariable region of the 16S rRNA gene, primers designed by Lee et al. (2017) [27], targeting the V6 region of 16S rRNA gene and primers designed in the present study. Klindworth and co-authors [2] described primer pairs, which are specific to the Bacteria domain and are widely used to investigate bacterial phyla present in different environments [11,27]. The primers proposed by Lee et al. (2017) [27] were characterized by the authors as cyanobacteria-specific and were used for water quality monitoring in lakes and ponds [27] focusing mainly on representatives of heterocystous cyanobacteria (Nostocales). The third primer pair was designed for this study and was supposed to target a wide range of cyanobacteria taxa, including specific genera and species occurring in microbial mats in small water bodies in highly elevated areas with limited water resources. Thus, the aims of the present study were to search for optimal primers to study cyanobacteria from microbial mats that can be a source of toxins in small water bodies. To do so we aimed to: 1) design a new primer pair (C\_773F - C\_1083R) and 2) test *in silico* and *in vitro* three primer sets: newly designed primers, the primers that were described in the literature as cyanobacteria-specific (V6 - 1328F - 1664R) as well as those universal bacterial primers used in our previous studies (V3-V4- S-D-Bact-0341-b-S-17- S-D-Bact-0785-a-A-21).

## 2. Materials and Methods

### 2.1. Study Design

The presented experiment consisted of six consecutive steps from sampling and morphological identification of taxa, through the preparation of mock community (MIX) and isolation of DNA, primers testing *in silico*, and amplification of given 16S rDNA regions in studied microbial mats, followed by bioinformatic and statistical analyses (Figure 1)



**Figure 1.** Study design.

The in silico and in vitro experiments were performed to test the specificity of three pairs of primers: S-D-Bact-0341-b-S-17 - S-D-Bact-0785-a-A-21 (V3-V4) [2], 1328F - 1664R (V6) [27] and C\_773F - C\_1083R (V4-V6) (this paper) in order to explore the diversity of benthic microbial communities mainly represented by cyanobacteria according to our previous study [10]. Primer names, target regions, sequences, and references are given in Table 1. To make it easier to follow, in the rest of the text, we use the targeted regions instead of long primer names. Each primer pair was applied to investigate the diversity and relative abundance of bacterial communities, focusing mainly on Cyanobacteria.

**Table 1.** Primers use this work.

Primer name	Direction	Target region	Position	Sequence (5' to 3')	Reference
S-D-Bact-0341-b-S-17	F	V3-V4	341	CCTACGGGNGGCWGCAG	[2]
S-D-Bact-0785-a-A-21	R		785	GACTACHVGGGTATCTAATCC	
C_773F	F	V4-V6	773	AAWGGGATTAGATACCCCWGT	present study
C_1086R	R		1086	RGGKTGCGCTCGTTGCGGGA	
1328F	F	V6	870	GCTAACGCGTTAAGTATCCCGCCTG	[27]
1664R	R		1160	GTCTCTCTAGAGTGCCCACTTAATG	

## 2.2. Sampling and Samples

For the analysis, we used a mock community as well as three environmental samples of microbial mats collected in August 2017 in Eastern Pamir from three water bodies with different types of mats. They were: (i) an unlayered type of mat from a hot spring used for bathing - Cyx 15, (ii) a multilayer soft mat from a thermokarst pond potentially used as a source of drinking water for wild and domestic animals - Cyx8 and (iii) shallow pool with an unlayered mat beneath the soil - Cyx9 (Table 2). The types of mats are described in Khomutovska et al. [11] and Sandzewicz et al. [12].

**Mock community design.** The mock community (further addressed also as “MIX”) consists of the DNA isolated from fifteen cyanobacterial strains including those from culture collection (one

SAG), two planktonic culture collections of Institute of Oceanography (University of Gdansk), as well as strains isolated from environmental samples collected in Eastern Pamir (Table S1). The cultures from Pamir included representative cyanobacterial genera that were common in studied microbial mats.

Table 2. Environmental samples in Eastern Pamir.

Sampl e ID	Location	EC, uS	Macroscopic characteristic of mats	Morphology-based identification of mat-forming microorganisms (optical microscope)
Cyx8	Sassykkul (Bulunkul subregion)	347	Multilayer soft mat, high content of EPS	<i>Leptolyngbya</i> sp., <i>Chlorogloea</i> sp., <i>Nostoc punctiforme</i> , green algae ( <i>Haematococcus pluvialis</i> )
Cyx9	Alichur (Bulunkul subregion)	272 0	Unlayered beneath soil - Mat dominated by filamentous cyanobacteria covered by a mineral soil layer	<i>Calothrix</i> spp., <i>Nostoc commune</i> , <i>Chlorogloea</i> sp., <i>Phormidium</i> sp., <i>Leptolyngbya</i> sp., green algae ( <i>Haematococcus pluvialis</i> )
Cyx15 a	hot spring near Gunt river (Bulunkul subregion)	179 1	Unlayered type - mat dominated by filamentous cyanobacteria, lower content of EPS compared to Cyx8	<i>Hilbrichtia pamiria</i> gen. sp. nov., <i>Oscillatoria</i> sp.,

2.2. Environmental Samples

The set of three environmental samples was selected based on the macroscopic features, and the morphology of microbial mat and taxonomic composition of communities was investigated based on optical microscopy analysis Nikon ECLIPSE Ni with digital camera Nikon DS-Fi1c. As mentioned above they represented three types of mats: multilayer soft, unlayered beneath the soil and unlayered [12]. The samples contained different representatives of Nostocales, Oscillatoriales, Synechococcales, and Chroococcales. The genera identified in the studied environmental samples are listed below. Investigated microbial mats differed in morphology, the content of EPS (extracellular polymeric substances), and secondary metabolites (cyanotoxins) [12,28]. While in Cyx9 and Cyx15 no cyanotoxins were detected, the Cyx8 mat was characterized by the presence of dmMC-LR (m/z 981.5).

2.3. DNA Extraction

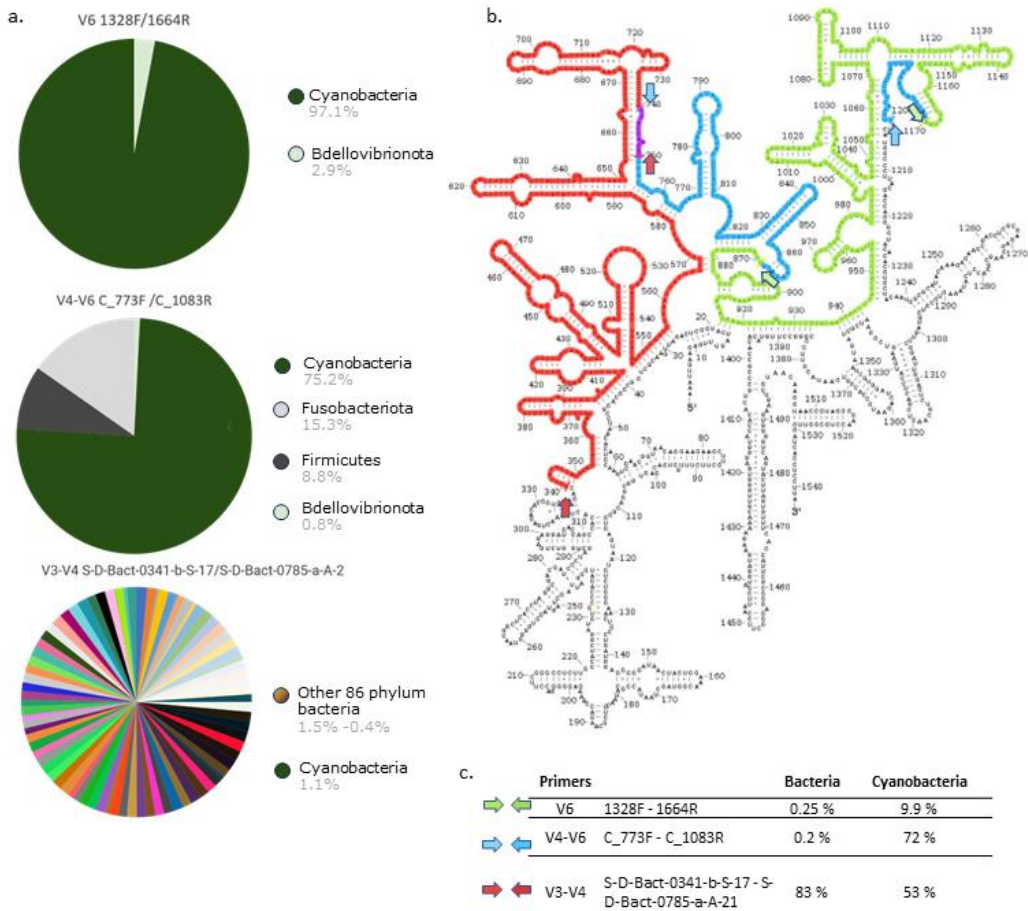
The cultures for the mock community were grown in liquid WC [29] and BG11[30] culture medium. The biomass of the cultures was centrifuged, next the pellet was discarded; about 200 mg of wet biomass was homogenized and frozen. From 150 to 250 mg of dried microbial mats were homogenized in sterile tubes. For each sample, the DNA extraction was performed at least 3 times, and then the subsamples were mixed into one sample. For the DNA extraction of both cultures and environmental samples, the Soil DNA Purification Kit (GeneMATRIX, EURx Ltd., Gdansk, Poland) was used. The genomic DNA was isolated according to the manufacturers’ protocol of EURx (Gdansk, Poland). The concentration and quality of the DNA were checked using a Hybrid Multi-Mode Reader (Synergy H1, BioTek, Bad Friedrichshall, Germany).

2.4. Primer Design and Characterization

The analysis included the whole set of cyanobacterial 16S rDNA sequences (352) publicly available at the National Center for Biotechnology Information (NCBI) website and the Silva database. The sequences from these two databases were compared using FSA [31] and the most conserved fragment DNA for cyanobacteria was identified (Figure 2). Cyanobacteria-specific primers



were designed using the Prime Design DNASTAR Lasergene v. 15, and AliView [32] tools. In silico analysis was performed using SILVAngs TestPrimer1.0 tool/pipeline.



**Figure 2.** A. Results of in silico amplification with test primers at phylum level according to Silva database; B. 16S rRNA *E. coli* with marked fragments amplified by the tested primers; C. The primers used in the study and percentage of bacterial and cyanobacterial species they detected in Silva database; color arrows shows forward and reverse primer on the hypervariable region on the 16S rRNA.

2.5. Sequencing and Bioinformatics

The next-generation sequencing was performed by Genomed Joint-Stock company (<http://www.genomed.pl/index.php/en>) using the Illumina MiSeq platform and V2 MiSeq Reagent Kit. The demultiplexed paired-end reads were analyzed in QIIME 2 (2020.11) using the following plugins: “dada2” for sequencing filtering and clustering of the representative sequences, “demux” for the visualization of the quality of reads, “classify-sklearn” to classify representative sequences with SILVA 138 classifier (Table S2).

The Sanger sequencing 16S rDNA reads from the mock community were joined into contigs using SeqMan software (Lasergene v. 14.00, DNASTAR).

2.6. Statistical and Phylogenetic Analysis

The statistical analysis was performed using R packages 4.0 such as “vegan”, “tidyverse”, “pheatmap”, “devtools” and “factoextra” [R core]. The non-metric multidimensional analysis was carried out based on the Bray-Curtis dissimilarity matrix calculated for studied communities. The heatmaps were constructed based on the Spearman rank correlation matrix and Euclidean distance phylogenetic analysis and visualization.

Phylogenetic placement of cyanobacterial ASVs was conducted to verify the taxonomic position of the most abundant taxa of Cyanobacteria. The query sequences were aligned to reference sequences using the Cydrasil database [33]. The sequences were mapped on Cydrasil-ML-tree-bs 1000, and the Algaebase-based nomenclature was used [33,34].

### 3. Results

#### 3.1. *In Silico* Testing of Designed and Commonly Used Primers

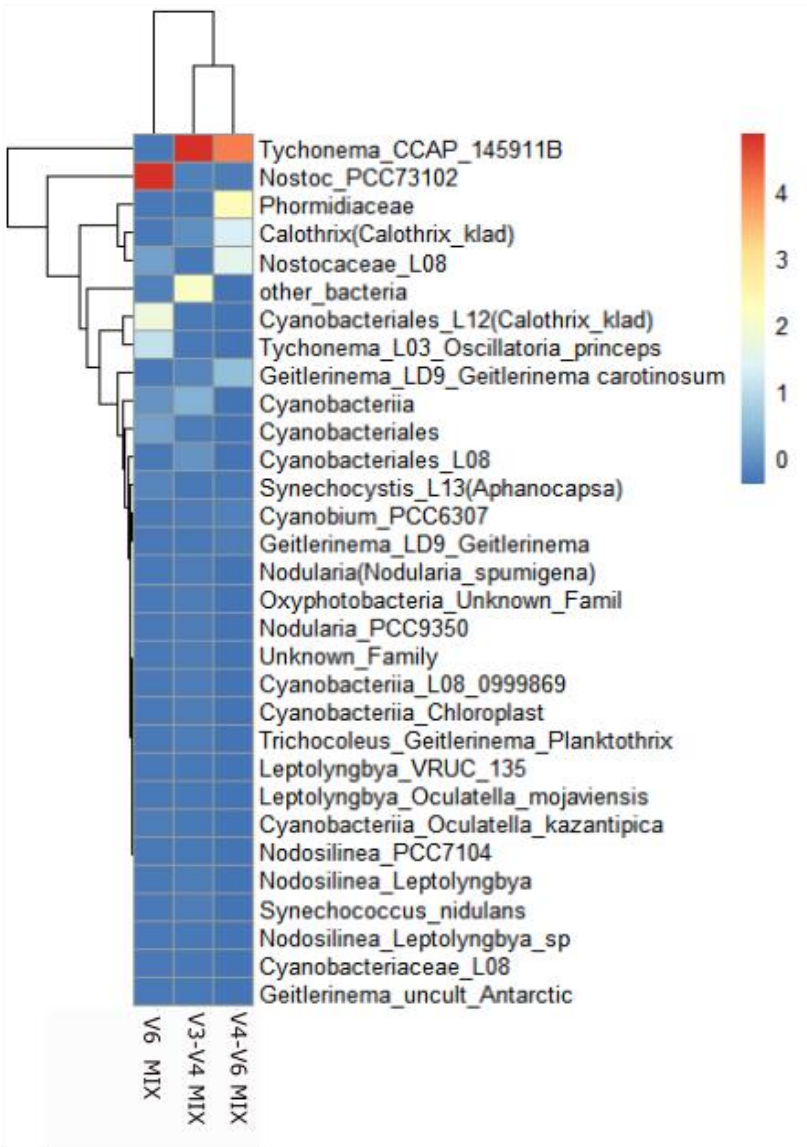
The result of the *in silico*, TestPrimer-based analysis indicated that the primers targeting V4-V6 region amplified mostly cyanobacterial sequences with the bacterial amplicons accounting only for about 25% of all sequences (Figure 2). The amplified sequences covered 72% of cyanobacteria and about 0.2% of other bacteria deposited in the Silva database. Amplifying the V3-V4 region, resulted in a Cyanobacteria share of about 1,1 % in the whole obtained sequences which covered 53% of cyanobacterial taxa in the Silva database. Additionally, the V4-V6 primer pair detected only 3 other phyla of the Bacteria domain (i.e., Bdellovibrionota, Firmicutes, Fusobacteriota) while the pair of V3-V4 primers detected representatives of 86 bacterial phyla, among which are Sericitochromatia, Vampirivibrionia, Firmicutes, Chloroflexi as well as some Archaea and eukaryotic algae (based on chloroplast DNA). The third pair of the tested primers, targeting the V6 region [24] identified only 10% of cyanobacterial sequences present in the Silva database although the second bacterial phylum, Bdellovibrionota accounted only for about 3% of all amplified sequences (Figure 2).

#### 2.2. *In Vitro* Analysis

The QIIME2-based analysis of the reads revealed significant differences in the numbers of reads obtained for the samples (Table S1). Those differences were observed in the reads results for divergent primer pairs within the same sample as well as between the samples. The highest number of reads within all sample types were obtained for V4-V6 primer pairs (designed in this study), while the lowest number of reads in most cases were for V3-V4 primer pairs. Also, a high number of reads was obtained for primers amplifying the V6 region that was designed by Lee [27]. The tested primer pairs showed different results of reads that passed quality filters (Table S2.) which varied for the particular communities. The result of V6 amplicons showed the highest percentage of reads that passed quality filters for studied communities, from 90.3% (84 433) to 91.6% (83 944) in various samples, although the total number of reads that passed quality filters was, in most cases, the highest for the primers designed within the present study. The results of amplification of the V3-V4 region demonstrated the highest differences in the numbers of reads that passed quality filters. For this last primer pair, we obtained the lowest number of reads for particular samples.

Concerning the cyanobacterial reads, the results of V4-V6 and V6 primer pairs show that amplified communities were mostly represented by Cyanobacteria. Cyanobacterial reads accounted for 100% and 97% of all amplified ASVs for the V4-V6 and V6 regions respectively. The results of V3-V4 amplification revealed that the lowest number of cyanobacterial reads accounted for 4.33%, while the maximum reached only 67% (Table S3).

The results of the mock community amplification showed that the highest number of cyanobacterial reads was obtained for the V3-V4 and the V4-V6 amplicons. The V6 regions that were described as cyanobacteria-specific gave the lowest number of reads. The general picture for all studied primers could be described as follows: the largest numbers of cyanobacterial ASVs come from V4-V6 primers amplification analysis and a little less from the V6 amplification (Figure 3., Table S2).



**Figure 3.** The results of the NGS analysis at the species level (after mapping on a Cydrasil tree) obtained for the mock community (MIX) using different primer pairs. Scaled dataset.

Next, we analyzed the query (amplicon) cyanobacterial sequences using references of full-length 16S rRNA sequences and the reference phylogenetic tree. This analysis helped to verify the cyanobacterial sequences obtained for studied primers and samples and to identify them up to the genus (names prefixed on heatmap). Comparing results for the mock community (MIX) demonstrated that the primer pairs V4-V6 and V3-V4 identified identical genera i.e., *Tychonema*, *Calothrix*, and *Geitlerinema*, while V6 primers prevailed in the determination of distinct representatives of *Nostoc*, *Calothrix* and *Tychonema* (Figure 3). Moreover, in detailed analysis with mapping obtained sequences on reference cyanobacterial phylogenetic tree (Cydrasil tree), the V4-V6 demonstrated more precise identification, because within these amplicons it was possible to identify the highest numbers of reads up to genus or family level (Table S4). Additionally, although the number of identified cyanobacterial sequences in the case of V6 (about 72 000) was slightly higher than in the case of V4-V6 (about 60 000), half of the V6 reads were assigned to single sequence of *Nostoc* (over 36 000 reads).

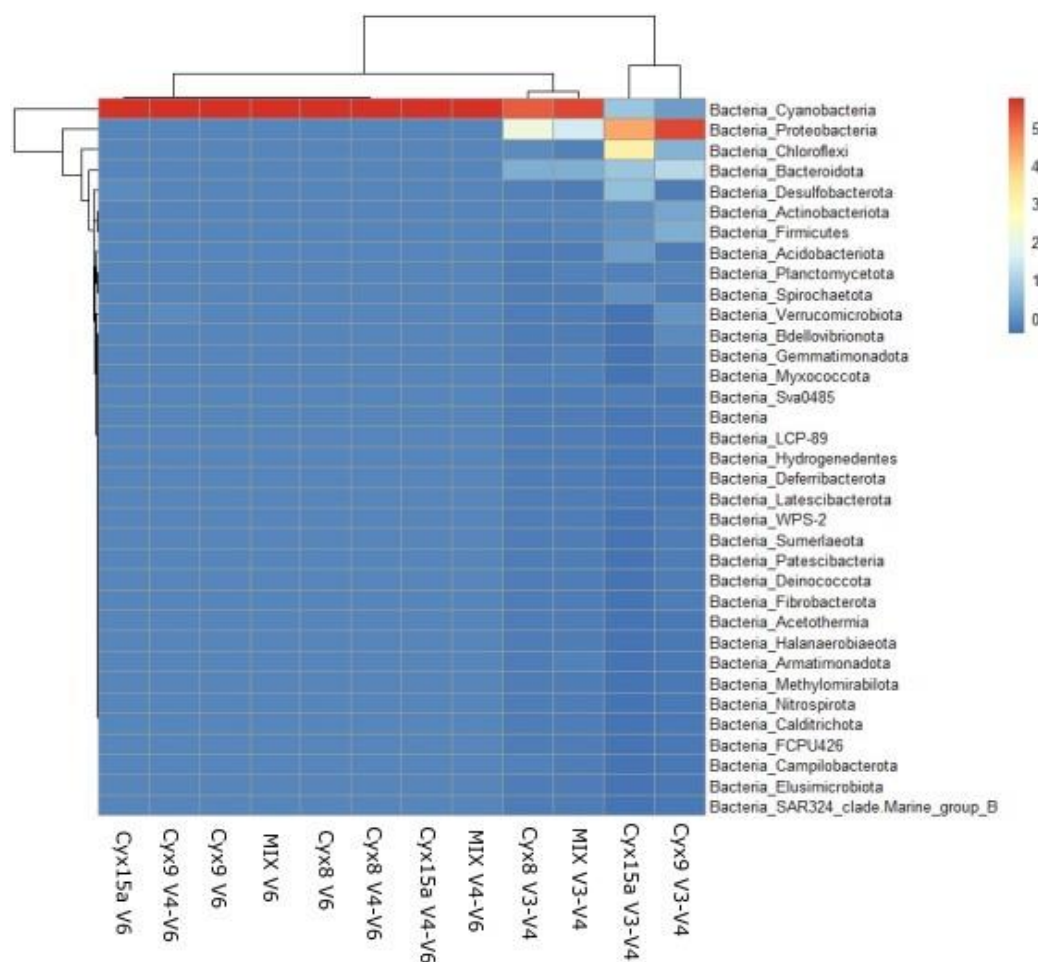
Mapping the results of mock community on the phylogenetic Cydrasil tree allowed also for identification of the most strains used for the mock community after amplification with V4-V6 primers. Our primers showed a greater number of ASVs in specific cyanobacteria and gave no count of other bacterial readings in contrary to the other two primers (V3-V4 and V6 regions). The mapping



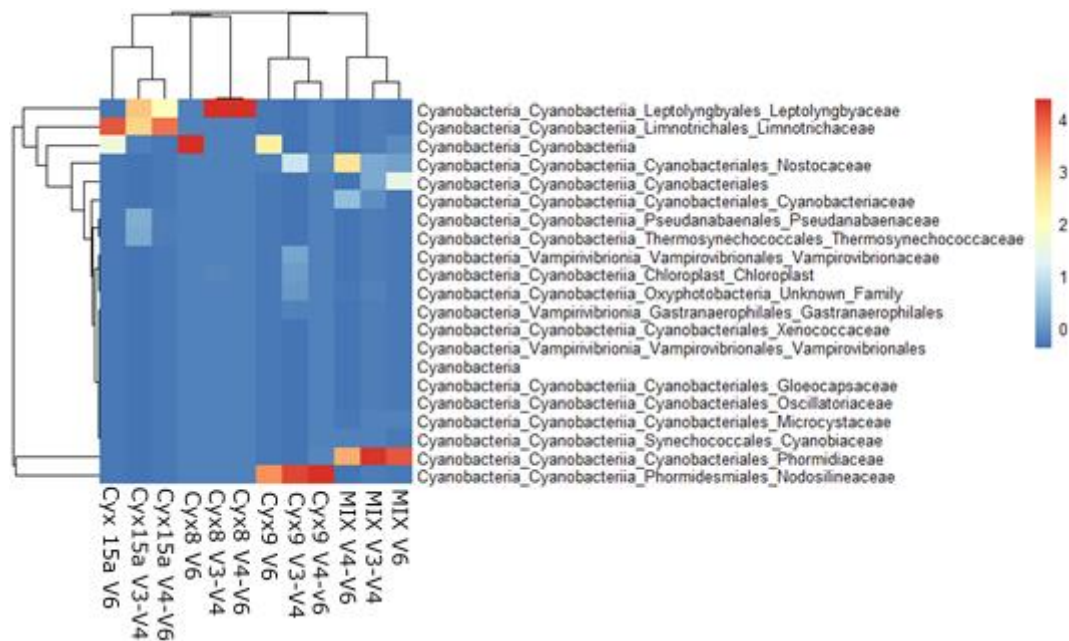
of query sequences performed for the V6 region revealed that it was hard to compare this analysis to the 16S sequencing results of the L08 strain and unequivocally identify the presence of this strain in the MIX sample.

The structure of bacterial communities at the phylum level demonstrated distinct differences in the sequences amplified with V3-V4 primer pairs and the other two primer pairs, targeting cyanobacteria. The sample Cyx15a and Cyx9 grouped together demonstrating the prevalence of Pseudomonadota (Proteobacteria) Chloroflexi, Bacteroidota, Desulfobacteriota, and Actinomycetota (Actinobacteriota). The V3-V4 amplicons of the Cxy8 and MIX sample formed another clade characterized by the dominance of Cyanobacteria, followed by Pseudomonadota (Proteobacteria) and Bacteroidota. The sequences amplified with V4-V6 and V6 primer pairs grouped together with Cyanobacteria phylum determining the structure of the amplified communities and no visible variability within other phyla (Figure 4.).

The structure of the cyanobacterial community at the family level was grouped according to the mat type, not a primer (Figure 5), though in most cases the highest similarity was between the amplicons of V3-V4 and V4-V6 regions. In the Cyx15 sample, Leptolyngbyaceae and Limnospiraceae families dominated, in Cyx8 Leptolyngbyaceae, Cyx9 Nodosilineaceae and in the mock community the structure was the most diverse, though with sequences identified as Phormidiaceae dominating the results. The results of all samples amplification with the V6 primers were characterized by lower specificity with a higher share of sequences identified only to the class or order level (Cyanobacteriia and Cyanobacteriales).

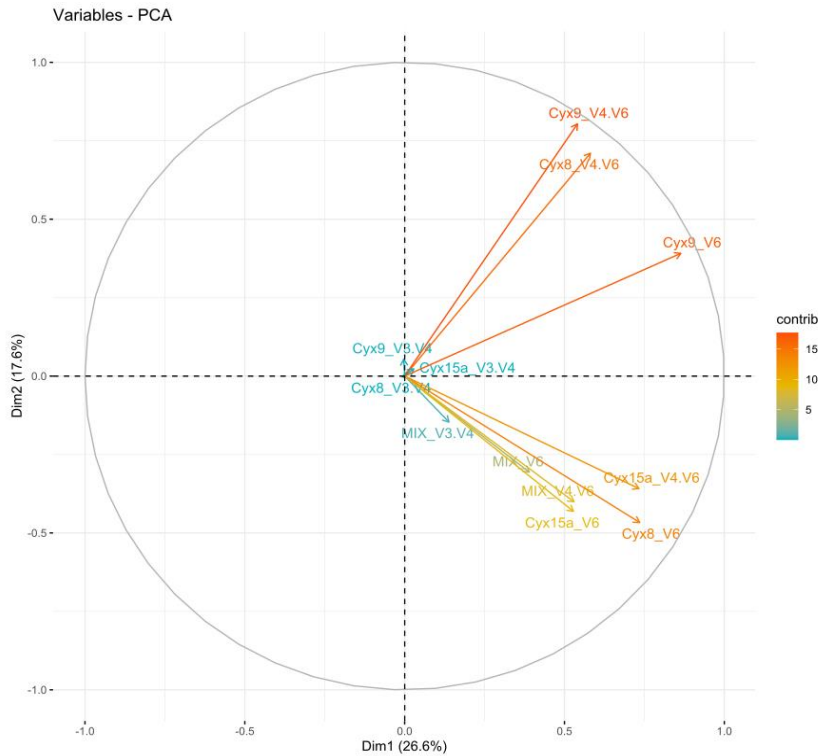


**Figure 4.** Heatmap performed for the environmental samples and mock community (MIX) at phylum level for total bacterial community.

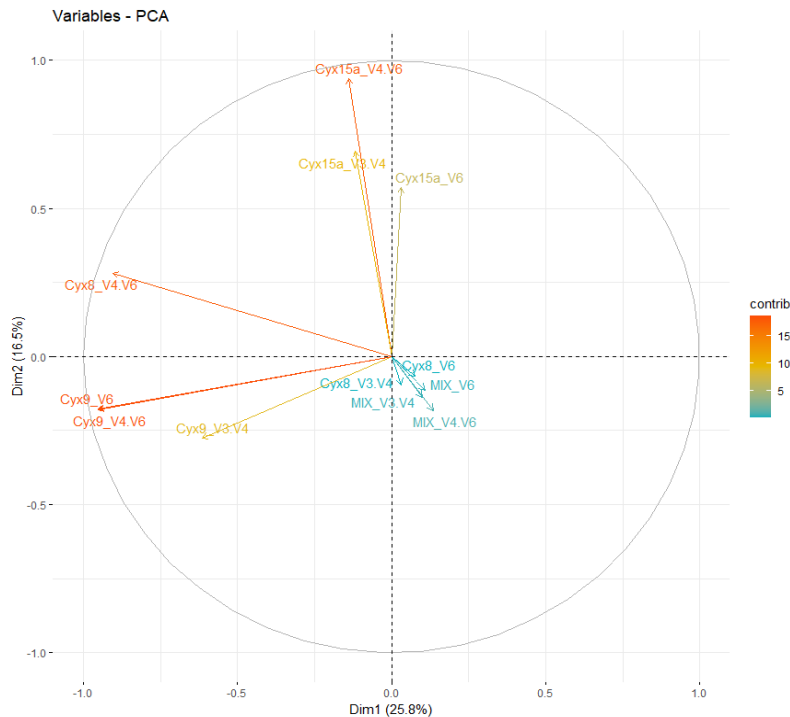


**Figure 5.** Heatmap performed for the environmental samples and mock community (MIX) at family level representing the cyanobacterial community. Scaled dataset.

The ASVs results of the amplification with three different primers' pairs were investigated further using the principal coordinate analysis (PCA) analysis (Figures 6 and 7). The results showed variable clustering for the whole bacterial communities. The amplicons of V3-V4 of all samples clustered together in the middle of the PCA, exhibiting very low contribution to the DIM1 and DIM2. The amplicons of V4-V6 primer pairs of all samples had a higher contribution than the corresponding amplicons of V6 primer pairs. In the case of the MIX sample the results of all three primers were close to each other (Figure 6). The PCA analysis of Cyanobacteria phylum (at the ASV level) demonstrated that the majority of samples were grouped according to the type of environmental samples (Figure 7). Similarly, as in the results of the whole bacterial community we observed that in the case of cyanobacterial ASVs the highest contribution to the placement of various samples mostly had the amplicons of the V4-V6 primers. The one exception was the Cyx9 sample when the amplicon V6 had a similar contribution as the V4-V6. We also noticed that the MIX sample, in contrast to the analysis of the whole bacteria community, had the lowest contribution to the placement of the results. Again, there was one exception – in the Cyx8 sample the results of V3-V4 and V6 amplification were grouped together, with the MIX displaying very low contribution to the placement of sequenced communities. However, the sequences amplified with the V4-V6 primer pair were placed in the opposite quarter with the highest contribution.



**Figure 6.** Principal component analysis (PCA) performed for whole bacterial communities at the ASV level.



**Figure 7.** Principal component analysis (PCA) was performed for cyanobacterial communities at the ASV level.

The highest parameters of the diversity metrics were obtained using V3-V4 primers (Table 3). The sample Cyx9 was characterized by the highest Shannon index (7.83) for primers V3-V4, while the other two primer pairs scored much lower values. However, the V4-V6 amplicons exhibited a higher Shannon index than V6 amplification. The same pattern was present in other analyzed samples. Verifying the diversity for cyanobacteria phylum amplification with the V3-V4 primer pair again

resulted in the highest value of the Shannon diversity index, with the V4-V6 having the second highest value. The differences between the V3-V4 and the other two primers were less pronounced. The number of cyanobacterial families identified with the V6 primer pair was generally lower (with one exception) than in the case of the other two primer pairs. The results concerning observed features of cyanobacteria at the ASV level were much more variable and did not repeat previous patterns.

**Table 3.** Diversity parameters calculated for Bacterial (B) and Cyanobacterial (C) communities.

Sample ID	Shannon_B	Faith's D_B	Observed_features/ASVs	Pielou's evenness_B	Shannon_C	Pielou's evenness_C	Nr of Cyanobacterial families	Observed ASVs_C
Cyx15a_V3-V4	6.51	21.91	184	0.86	1.41	0.68	5	8
<b>Cyx15a_V4-V6</b>	<b>1.68</b>	<b>3.96</b>	<b>9</b>	<b>0.53</b>	<b>1.16</b>	<b>0.56</b>	<b>5</b>	<b>8</b>
Cyx15a_V6	1.18	4.85	22	0.26	0.67	0.32	3	8
Cyx8_V3-V4	5.15	15.03	133	0.73	2.18	0.83	3	14
<b>Cyx8_V4-V6</b>	<b>1.22</b>	<b>4.00</b>	<b>8</b>	<b>0.41</b>	<b>0.84</b>	<b>0.38</b>	<b>3</b>	<b>9</b>
Cyx8_V6	0.41	3.98	14	0.11	0.26	0.11	3	11
Cyx9_V3-V4	7.83	37.86	453	0.89	2.41	0.77	5	23
<b>Cyx9_V4-V6</b>	<b>2.54</b>	<b>5.57</b>	<b>53</b>	<b>0.44</b>	<b>1.67</b>	<b>0.43</b>	<b>10</b>	<b>48</b>
Cyx9_V6	2.39	7.13	84	0.37	1.51	0.39	8	46
Mix_V3-V4	5.64	14.45	132	0.80	2.75	0.74	10	41
<b>Mix_V4-V6</b>	<b>2.85</b>	<b>4.27</b>	<b>23</b>	<b>0.63</b>	<b>2.01</b>	<b>0.64</b>	<b>9</b>	<b>23</b>
Mix_V6	2.56	4.74	29	0.53	1.74	0.56	7	23

4. Discussion

The results of the present study revealed that the primers designed within the present study amplifying fragment V4-V6 of 16S allowed for more precise identification of cyanobacteria in benthic microbial communities, compared to previously described primers [2,27] which were focused on planktonic communities. Additionally, the primers were tested experimentally with microbial mat samples from high altitude, arid environments such as small water bodies, that are less studied than the commonly investigated planktonic communities. Such types of samples are often claimed to be sources of novel microorganisms (including species of Cyanobacteria). Thus, using amplicon-based methods it is important to investigate the taxonomic composition of the microorganism at a higher level as it is often not possible to identify the species. Also, other researchers confirmed recently that there is a need to use different fragments of hypervariable 16S rRNA to study bacteria and archaea from less known environments [35]. The present study presents that the V4-V6 region of 16S rRNA seems to be a good solution to investigate diverse microbial benthic communities giving great resolution for Cyanobacteria, although not providing reliable results of the rest of the bacterial community. In turn the V3-V4 primer pair is more universal for all phyla of Bacteria including also some Archaea and eukaryotic algae (chloroplast DNA) but still provides considerably broad results for cyanobacteria. However, both results of the in silico as well as in vitro analyses showed that the V4-V6 primer pair allows for the detection of more cyanobacterial taxa (including representatives of



orders Cyanobacteriales, Gloeobacterales, Leptolyngbyales, Phormidesmiales, Synechococcales, Thermosynechococcales) compared to the other studied primer pairs, such as V6, and V3-V4. The analysis of the MIX sample using mapping on the phylogenetic tree also demonstrated that the V4-V6 primers resulted in a much higher number of identified various ASVs than the amplicons of the other two primer pairs. The high number of V6 amplicons was actually misleading as *Nostoc* accounted for half of the reads. Interestingly also we were not able in the V6 amplicon to identify the strain L08, *Hillbrichtia pamiria* gen. nov. sp. nov. The reason for this can be the novelty of the L08 strain and the specificity of the V6 region concerning the identification of cyanobacterial orders/families. The PCA analysis of Cyanobacteria phylum (at the ASV level).

On the other hand, for the studies of the whole bacterial communities (including Cyanobacteria) the use of the V3-V4 primers would be advisable (Table 3.). Described comparison was made for the MIX sample, in which we used “unialgal” cultures of Cyanobacteria and although they were not axenic, the contribution of Bacteria was generally lower than in natural mat samples. Using primers designed within the present study (V4-V6) we obtained much higher number of reads of cyanobacterial amplicons, that were assigned to various taxonomic levels on the reference phylogenetic tree, without having amplicons of other bacteria as in the case of V3-V4 (Table S4). Such specificity of V4-V6 primers facilitates bioinformatics analyses. The primer pair amplifying the V6 region of the 16S rRNA gene gives fewer bacterial sequences than V3-V4, but less varied results of Cyanobacteria than V4-V6. This can be explained by their higher specificity than in the case of primers designed within this study. Our results revealed that microbial communities in which the dominated Nostocaceae genera can be explored more precisely while applying V6 primers as was noted by previous authors [27]. However, this comes at the expense of data on other cyanobacterial taxa, which have fewer reads, as more than half of the reads of the V6 amplicons we assigned to *Nostoc*. Also using V6 primers resulted with higher percentage of sequences identified to a class or possibly an order level only. We are aware the differences in the number of reads obtained for the individual taxa of microorganisms during the DNA amplification can be also related to the presence of different inhibitory substances in the analyzed environmental samples, which was also pointed out by other authors [9]. The environmental DNA can become fragmented during the isolation resulting in short sequences that in some cases can be identified up to order level only [36–40]. All these factors may negatively affect the results of the analyses and make their comparison difficult.

The PCA analysis of the whole Bacteria domain demonstrated that from among the three primer pairs used, the highest contribution to the PCA results of given samples had amplicons of the V4-V6 primers. This was further confirmed in the analysis of cyanobacterial reads only, where results showed an even more pronounced contribution of V4-V6. An exception was observed in samples of Cyx9 mat, in which V6 amplicons influenced the placement in a similar degree to V4-V6. The differences in the placement between various samples in PCA of Bacteria and Cyanobacteria amplicons are of course connected with the composition and structure of analyzed communities and the differences in the contribution of other bacterial phyla than Cyanobacteria. For example, in the case of the mock community, the results of V3-V4 amplification were close to the other two primer pairs. That's possible because cyanobacterial strains used for the mock community (MIX), although not axenic, did not contain as much bacteria as the environmental samples. It was confirmed in the PCA for cyanobacteria only, in which results of all three primer pairs amplifications of MIX are close together. Regarding the Cyx15 sample, it grouped along the Y axis on the PCA of Cyanobacteria with distinctly the highest contribution of V4-V6 and more similar results in the case of V3-V4 and V6. The results may reflect the presence and predominance of *Hillbrichtia pamiria* gen. nov. sp. nov. in this mat [13]. It seems that the amplicons of both primer pairs, V3-V4 – the less specific as well V6 - the more specific towards Nostocales can be difficult to identify and may be classified to various taxonomic groups.

The primer pairs designed within the present study proved to amplify well and allowed for the identification, compared with other primers, of the most cyanobacterial taxa at the order level. However, our results also revealed that metabarcoding is still a limited method to study the composition and structure of cyanobacterial communities. Although the pair of primers designed in

this study scored high, we are currently unable to determine unequivocally that it provides the optimal results for environmental analyses. Both the V4-V6 and V3-V4 primer pairs produced a large, though different, diversity, with the latter yielding the highest results. That may be due to chimeras and unspecific amplification of V3-V4 primer pair. Additionally, the V4-V6 primer pairs gave much fewer chimeras, and the higher number of cyanobacterial reads identified to family or even species. The results of the V4-V6-based amplification also gave higher diversity parameters than the results of V6 amplification, though the values were more comparable.

## 5. Conclusions

In our research, we have shown that choosing a suitable primer can be crucial for expected results. Primers amplifying the V3-V4 region of 16S rRNA can be applied to investigate the structure of the entire bacterial community. However, if a study aims to obtain as much information as possible about Cyanobacteria, including potentially toxigenic ones and from specific environment such as benthic communities, we suggest using V4-V6 primers, which are more specific for cyanobacteria than V3-V4 and less specific to a single taxonomic group, as is in the case of V6.

**Supplementary Materials:** The following supporting information can be downloaded at: [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), **Table S1.** Strains used for mock community, concentration of isolated DNA, and percentage contribution of used DNA.; **Table S2.** The reads obtained for studied samples before and after filtering.; **Table S3.** The number of bacterial and cyanobacterial ASVs with the percentage contribution of cyanobacterial ASVs using studied primer pairs.; **Table S4.** The results of the amplicon sequencing of the mock community (MIX). The taxa names of the ASVs were verified based on the Cydrasill package with our matched 16S sequences from the mock community. The interactive tree is available under the link iTOL, [https://itol.embl.de/personal\\_page.cgi](https://itol.embl.de/personal_page.cgi) login: CyanoMIX, password: CyanoMIX2021.

**Author Contributions:** Conceptualization, J.K. and Ł.Ł.; methodology, Ł.Ł. and N.K.; software, Ł.Ł.; N.K.; validation, Ł.Ł., N.K.; formal analysis, N.K.; investigation, Ł.Ł.; resources, Ł.Ł. I.J.; data curation, Ł.Ł., N.K.; writing—original draft preparation, Ł.Ł.; writing—review and editing, N.K., I.J.; visualization, Ł.Ł. and N.K.; supervision, I.J.; project administration, I.J.; funding acquisition, I.J. All authors have read and agreed to the published version of the manuscript..

**Funding:** This research was funded by the National Science Centre (Grant 2015/19/B/NZ9/00473).

**Acknowledgments:** The authors would like to thank Małgorzata Sandzewicz for support during the creation of this work as well as Małgorzata Suska-Malawska and colleagues from Department of Ecology and Environmental conservation for help in field work in Pair Mountains.

## References

1. Bolhuis, H.; Stal, L.J. Analysis of Bacterial and Archaeal Diversity in Coastal Microbial Mats Using Massive Parallel 16S rRNA Gene Tag Sequencing. *The ISME Journal* **2011**, *5*, 1701–1712, doi:10.1038/ismej.2011.52.
2. Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glöckner, F.O. Evaluation of General 16S Ribosomal RNA Gene PCR Primers for Classical and Next-Generation Sequencing-Based Diversity Studies. *Nucleic Acids Research* **2013**, *41*, 1–11, doi:10.1093/nar/gks808.
3. Winand, R.; Bogaerts, B.; Hoffman, S.; Lefevre, L.; Delvoye, M.; Van Braekel, J.; Fu, Q.; Roosens, N.H.; Cj, S.; Keersmaecker, D.; et al. Targeting the 16S rRNA Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies. *International Journal of Molecular Sciences Article* **2019**, doi:10.3390/ijms21010298.
4. Bahram, M.; Anslan, S.; Hildebrand, F.; Bork, P.; Tedersoo, L. Brief Report Newly Designed 16S rRNA Metabarcoding Primers Amplify Diverse and Novel Archaeal Taxa from the Environment. **2018**, doi:10.1111/1758-2229.12684.
5. Casero, M.C.; Velázquez, D.; Medina-Cobo, M.; Quesada, A.; Cirés, S. Unmasking the Identity of Toxigenic Cyanobacteria Driving a Multi-Toxin Bloom by High-Throughput Sequencing of Cyanotoxins Genes and 16S rRNA Metabarcoding. *Science of The Total Environment* **2019**, *665*, 367–378, doi:https://doi.org/10.1016/j.scitotenv.2019.02.083.
6. Abdala Asbun, A.; Besseling, M.A.; Balzano, S.; van Bleijswijk, J.D.L.; Witte, H.J.; Villanueva, L.; Engelmann, J.C. Cascabel: A Scalable and Versatile Amplicon Sequence Data Analysis Pipeline Delivering Reproducible and Documented Results. *Frontiers in Genetics* **2020**, *11*, 1–14, doi:10.3389/fgene.2020.489357.

7. Pérez Gallego, R.; Bale, N.J.; Sinninghe Damste, J.S.; Villanueva, L. Developing a Genetic Approach to Target Cyanobacterial Producers of Heterocyte Glycolipids in the Environment. *Frontiers in Microbiology* **2023**, *14*, 1–14, doi:10.3389/fmicb.2023.1257040.
8. Kirk Harris, J.; Gregory Caporaso, J.; Walker, J.J.; Spear, J.R.; Gold, N.J.; Robertson, C.E.; Hugenholtz, P.; Goodrich, J.; McDonald, D.; Knights, D.; et al. Phylogenetic Stratigraphy in the Guerrero Negro Hypersaline Microbial Mat. *The ISME Journal* **2013**, *7*, 50–60, doi:10.1038/ismej.2012.79.
9. Amaral-Zettler, L.A.; McCliment, E.A.; Ducklow, H.W.; Huse, S.M. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE* **2009**, *4*, 1–9, doi:10.1371/journal.pone.0006372.
10. Jasser, I.; Kostrzewska-Szlakowska, I.; Kwiatowski, J.; Navruzshoev, D.; Suska-Malawska, M.; Khomutovska, N. Morphological and Molecular Diversity of Benthic Cyanobacteria Communities Versus Environmental Conditions in Shallow, High Mountain Water Bodies in Eastern Pamir Mountains (Tajikistan). *Polish Journal of Ecology* **2020**, *67*, 286–304, doi:10.3161/15052249PJE2019.67.4.002.
11. Khomutovska, N.; De Los Ríos, A.; Jasser, I. Microorganisms Diversity and Colonization Strategies of Endolithic Cyanobacteria in the Cold Mountain Desert of Pamir. **2020**, doi:10.3390/microorganisms9010006.
12. Sandzewicz, M.; Khomutovska, N.; Łach, Ł.; Kwiatowski, J.; Niyatbekov, T.; Suska-Malawska, M.; Jasser, I. Salinity Matters the Most: How Environmental Factors Shape the Diversity and Structure of Cyanobacterial Mat Communities in High Altitude Arid Ecosystems. *Frontiers in Microbiology* **2023**, *14*, 1–17, doi:10.3389/fmicb.2023.1108694.
13. Jasser, I.; Panou, M.; Khomutovska, N.; Sandzewicz, M.; Panteris, E.; Niyatbekov, T.; Łach, Ł.; Kwiatowski, J.; Kokociński, M.; Gkelis, S. Cyanobacteria in Hot Pursuit: Characterization of Cyanobacteria Strains, Including Novel Taxa, Isolated from Geothermal Habitats from Different Ecoregions of the World. *Molecular Phylogenetics and Evolution* **2022**, *170*, 107454, doi:https://doi.org/10.1016/j.ympev.2022.107454.
14. Kleinteich, J.; Hildebrand, F.; Wood, S.A.; Ciršs, S.; Agha, R.; Quesada, A.; Pearce, D.A.; Convey, P.; Küpper, F.C.; Dietrich, D.R. Diversity of Toxin and Non-Toxin Containing Cyanobacterial Mats of Meltwater Ponds on the Antarctic Peninsula: A Pyrosequencing Approach. *Antarctic Science* **2014**, *26*, 521–532, doi:10.1017/S0954102014000145.
15. Rasuk, M.C.; Fernández, A.B.; Kurth, D.; Contreras, M.; Novoa, F.; Poiré, D.; Farías, M.E. Bacterial Diversity in Microbial Mats and Sediments from the Atacama Desert. *Microbial Ecology* **2016**, *71*, 44–56, doi:10.1007/s00248-015-0649-9.
16. Khomutovska, N.; Jerzak, M.; Kostrzewska-Szlakowska, I.; Kwiatowski, J.; Suska-Malawska, M.; Syczewski, M.; Jasser, I. Life in Extreme Habitats: Diversity of Endolithic Microorganisms from Cold Desert Ecosystems of Eastern Pamir. *Polish Journal of Ecology* **2017**, *65*, 303–319, doi:10.3161/15052249PJE2017.65.4.001.
17. Khomutovska, N.; De Los Ríos, A.; Syczewski, M.D.; Jasser, I. Connectivity of Edaphic and Endolithic Microbial Niches in Cold Mountain Desert of Eastern Pamir (Tajikistan). **2021**, doi:10.3390/biology10040314.
18. Gkelis, S.; Panou, M.; Konstantinou, D.; Apostolidis, P.; Kasampali, A.; Papadimitriou, S.; Kati, D.; Di Lorenzo, G.M.; Ioakeim, S.; Zervou, S.K.; et al. Diversity, Cyanotoxin Production, and Bioactivities of Cyanobacteria Isolated from Freshwaters of Greece. *Toxins* **2019**, *11*, doi:10.3390/toxins11080436.
19. Ar, K.; Gkelis, S.; Vardaka, E.; Moustaka-gouni, M. Limnological Morphological and Molecular Analysis of Bloom-Forming Cyanobacteria in Two Eutrophic, Shallow Mediterranean Lakes. **2011**, *41*, 167–173, doi:10.1016/j.limno.2010.10.003.
20. Codd, G.; Bell, S.; Kaya, K.; Ward, C.; Beattie, K.; Metcalf, J. Cyanobacterial Toxins, Exposure Routes and Human Health. *European Journal of Phycology* **1999**, *34*, 405–415, doi:10.1080/09670269910001736462.
21. Teikari, J.; Baunach, M.; Dittmann, E. Cyanobacterial Genome Sequencing, Annotation, and Bioinformatics. *Methods in molecular biology (Clifton, N.J.)* **2022**, *2489*, 269–287, doi:10.1007/978-1-0716-2273-5\_14.
22. Quiblier, C.; Wood, S.; Echenique Subiabre, I.; Heath, M.; Villeneuve, A.; Humbert, J.-F. A Review of Current Knowledge on Toxic Benthic Freshwater Cyanobacteria--Ecology, Toxin Production and Risk Management. *Water research* **2013**, *47*, 5464–5479, doi:10.1016/J.WATRES.2013.06.042.
23. Brasell, K.A.; Heath, M.W.; Ryan, K.G.; Wood, S.A. Successional Change in Microbial Communities of Benthic Phormidium-Dominated Biofilms. *Microbial Ecology* **2015**, *69*, 254–266, doi:10.1007/s00248-014-0538-7.
24. Kleinteich, J.; Wood, S.A.; Puddick, J.; Schleheck, D.; Küpper, F.C.; Dietrich, D. Potent Toxins in Arctic Environments – Presence of Saxitoxins and an Unusual Microcystin Variant in Arctic Freshwater Ecosystems. *Chemico-Biological Interactions* **2013**, *206*, 423–431, doi:10.1016/J.CBI.2013.04.011.
25. Tikhonova, I.; Kuzmin, A.; Deeva, D.; Sorokovikova, E.; Potapov, S.; Lomakina, A.; Belykh, O. Cyanobacteria Nostoc Punctiforme from Abyssal Benthos of Lake Baikal: Unique Ecology and Metabolic Potential. *Indian Journal of Microbiology* **2017**, *57*, 422–426, doi:10.1007/s12088-017-0673-1.
26. Gaget, V.; Keulen, A.; Lau, M.; Monis, P.; Brookes, J.D. DNA Extraction from Benthic Cyanobacteria: Comparative Assessment and Optimization. **2016**, doi:10.1111/jam.13332.

27. Lee, E.; Khurana, M.S.; Whiteley, A.S.; Monis, P.T.; Bath, A.; Gordon, C.; Ryan, U.M.; Paparini, A. Novel Primer Sets for next Generation Sequencing-Based Analyses of Water Quality. *PLoS ONE* **2017**, *12*, 1–17, doi:10.1371/journal.pone.0170008.
28. Khomutovska, N.; Sandzewicz, M.; Łach, Ł.; Suska-Malawska, M.; Chmielewska, M.; Mazur-Marzec, H.; Cegłowska, M.; Niyatbekov, T.; Wood, S.A.; Puddick, J.; et al. Limited Microcystin, Anatoxin and Cylindrospermopsin Production by Cyanobacteria from Microbial Mats in Cold Deserts. *Toxins* **2020**, *12*, doi:10.3390/toxins12040244.
29. Guillard, R.R.L. and Lorenzen, C.J. Yellow-Green Algae with Chlorophyllidec. *Yellow-Green Algae with Chlorophyllidec. Journal of Phycology*, *8*, 10–14. 1972.
30. Rippka, R.; Deruelles, J.; Waterbury, J.B. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Journal of General Microbiology* **1979**, *111*, 1–61, doi:10.1099/00221287-111-1-1.
31. Bradley, R.K.; Roberts, A.; Smoot, M.; Juvekar, S.; Do, J. Fast Statistical Alignment. *PLoS Comput Biol* **2009**, *5*, 1000392, doi:10.1371/journal.pcbi.1000392.
32. Larsson, A. AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets. *BIOINFORMATICS APPLICATIONS* **2014**, *30*, 3276–3278, doi:10.1093/bioinformatics/btu531.
33. Roush, D.; Giraldo-Silva, A.; Garcia-Pichel, F. Cydrasil 3, a Curated 16S rRNA Gene Reference Package and Web App for Cyanobacterial Phylogenetic Placement., doi:10.1038/s41597-021-01015-5.
34. Guiry, M.D. & Guiry, G.M. No Title Available online: <https://www.algaebase.org>.
35. Fadeev, E.; Cardozo-mino, M.G.; Rapp, J.Z.; Bienhold, C.; Salter, I.; Salman-carvalho, V.; Molari, M.; Tegetmeyer, H.E.; Buttigieg, P.L.; Boetius, A. Comparison of Two 16S rRNA Studies of Arctic Microbial Communities. **2021**, *12*, 1–11, doi:10.3389/fmicb.2021.637526.
36. Sabat, A.J.; Van Zanten, E.; Akkerboom, V.; Wisselink, G.; Van Slochteren, K.; De Boer, R.F.; Hendrix, R.; Friedrich, A.W.; Rossen, J.W.A.; Kooistra-Smid, A.M.D. Targeted Next-Generation Sequencing of the 16S–23S rRNA Region for Culture-Independent Bacterial Identification-Increased Discrimination of Closely Related Species., doi:10.1038/s41598-017-03458-6.
37. Komarek, J. Cyanobacterial Taxonomy: Current Problems and Prospects for the Integration of Traditional and Molecular Approaches. *Algae* **2006**, *21*, 349–375, doi:10.4490/algae.2006.21.4.349.
38. Li, X.; Huo, S.; Xi, B. Updating the Resolution for 16S rRNA OTUs Clustering Reveals the Cryptic Cyanobacterial Genus and Species. *Ecological Indicators* **2020**, *117*, doi:10.1016/j.ecolind.2020.106695.
39. Li, X.C.; Huo, S.; Zhang, J.; Ma, C.; Xiao, Z.; Zhang, H.; Xi, B.; Xia, X. Metabarcoding Reveals a More Complex Cyanobacterial Community than Morphological Identification. *Ecological Indicators* **2019**, *107*, 105653, doi:10.1016/j.ECOLIND.2019.105653.
40. Zou, S.; Smith, L. Comprehensive Primer Sets and Cost Ecient Multiplex PCR-Based EDNA Sequencing for Community Dynamics of Cyanobacteria, Eukaryotic Phytoplankton and Zooplankton in Lake. *Research Square* **2020**, 1–29.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.