

RESEARCH ARTICLE

Novel Primer Sets for Next Generation Sequencing-Based Analyses of Water Quality

Elvina Lee¹, Maninder S. Khurana¹, Andrew S. Whiteley², Paul T. Monis³, Andrew Bath⁴, Cameron Gordon⁴, Una M. Ryan¹, Andrea Paparini^{1*}

1 Vector- and Water-Borne Pathogen Research Laboratory, School of Veterinary and Life Sciences, Murdoch University, Perth, Western Australia, Australia, **2** School of Earth and Environment, The University of Western Australia, Crawley, Western Australia, Australia, **3** Australian Water Quality Centre, South Australian Water Corporation, Adelaide, South Australia, Australia, **4** Drinking Water Quality Branch, Water Corporation, Leederville, Western Australia, Australia

* a.paparini@murdoch.edu.au



OPEN ACCESS

Citation: Lee E, Khurana MS, Whiteley AS, Monis PT, Bath A, Gordon C, et al. (2017) Novel Primer Sets for Next Generation Sequencing-Based Analyses of Water Quality. PLoS ONE 12(1): e0170008. doi:10.1371/journal.pone.0170008

Editor: George-John Nychas, Agricultural University of Athens, GREECE

Received: July 26, 2016

Accepted: December 26, 2016

Published: January 24, 2017

Copyright: © 2017 Lee et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: BioProject ID: PRJNA330773, including NGS data and metadata (SRA) and samples data and metadata (BIOSAMPLE).

Funding: Resources of the Pawsey Supercomputing Centre (BLAST on MAGNUS) were used during some steps of the data analysis under an existing project (The microbiome of Australia's source water). Partial financial support for this study was indirectly provided by Murdoch University, the Australian Research Council (LP100100160), Bayer Healthcare, Bayer Australia

Abstract

Next generation sequencing (NGS) has rapidly become an invaluable tool for the detection, identification and relative quantification of environmental microorganisms. Here, we demonstrate two new 16S rDNA primer sets, which are compatible with NGS approaches and are primarily for use in water quality studies. Compared to 16S rRNA gene based universal primers, *in silico* and experimental analyses demonstrated that the new primers showed increased specificity for the Cyanobacteria and Proteobacteria phyla, allowing increased sensitivity for the detection, identification and relative quantification of toxic bloom-forming microalgae, microbial water quality bioindicators and common pathogens. Significantly, Cyanobacterial and Proteobacterial sequences accounted for ca. 95% of all sequences obtained within NGS runs (when compared to ca. 50% with standard universal NGS primers), providing higher sensitivity and greater phylogenetic resolution of key water quality microbial groups. The increased selectivity of the new primers allow the parallel sequencing of more samples through reduced sequence retrieval levels required to detect target groups, potentially reducing NGS costs by 50% but still guaranteeing optimal coverage and species discrimination.

Introduction

The growing accessibility of next generation DNA sequencing (NGS) methods has greatly advanced our understanding of microbial diversity in medical and environmental science [1–5]. Refinement of platforms, protocols and reagents for parallel high-throughput DNA sequencing technologies allows profiling of complex microbial communities at ever-increasing resolution [6, 7]. Being culture independent, NGS has the advantage of allowing direct analysis of communities as they exist under *in situ* conditions, including their genes, transcripts, proteins, and metabolites and how their reciprocal interactions impact their distribution patterns [6, 8, 9]. Microbiomes can be audited with or without the preliminary use of polymerase chain reaction (PCR) amplification of DNA; the sequences, from either selected loci, or entire

Ltd., Western Australia Water Corporation, WaterRA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Partial financial support for this study was indirectly provided by Murdoch University, the Australian Research Council (LP100100160), Bayer Healthcare, Bayer Australia Ltd., Western Australia Water Corporation, WaterRA. The funders had no role in selecting the information to publish, study design, data-collection, data-analysis, data-interpretation, decision to publish or drawing conclusions. The above partnerships do not alter our adherence to all PLOS ONE policies on sharing data and materials. The authors declare that no competing interests exist.

genomes [10, 11] are then taxonomically or functionally classified based on the similarity to known entries from existing databases [7].

The success and resultant quality of PCR-based sequencing can be affected by factors such as the amplicon targeted (locus and region), thermal cycling conditions used, sequencing method, bioinformatics pipeline etc. [6, 7, 12–14]; yet, the most critical step in amplification-dependent metagenomics studies still appears to be amplification region targeted and the choice of primers [15–18]. Compared to other loci, the hypervariability and multi-copy nature of the small ribosomal subunit RNA (16S rDNA) gene, coupled with the availability of growing sequence information, confer higher detectability and allow taxonomic classification of bacteria and archaea, potentially to species level [19].

Previous research has concentrated on the design of 16S rDNA primers that amplify all taxa with equal efficiency [16, 20, 21] to guarantee broad taxonomic representation and preserve the relative proportions originally present in the community [17, 20, 22–24]. 16S rDNA-targeted primers with broad taxonomic coverage (known as “universal” primers) are routinely employed for exhaustive molecular surveys of bacterial and archaeal communities. At inadequate sequencing depths and in the presence of strong dominance, however, universal primers are expected to generate a large proportion of the sequences from the most abundant taxa, while rarer taxa (possibly including pathogens, bioindicators or other informative target groups) may remain undetected [25].

In the context of the water quality, the utilisation of NGS has allowed for the discovery of new indicators of water contamination (as reviewed in [26]). Similarly, cyanobacteria are a global concern for the water industry, due to the presence of bloom forming and/or secondary metabolite producing strains that are a risk to public health or adversely affect the aesthetic quality of the water [27–29]. Molecular detection, quantification and identification of cyanobacteria are relatively common [8, 30–33], but the numerous primer sets currently available generate amplicons with lengths incompatible with NGS, and/or are specific for only a few members of this phylum [34–40]. To overcome these limitations, the present study sought to design taxon-specific primers targeting the cyanobacteria 16S rRNA gene for use with NGS. These new primers were tested on both cultures and environmental water samples; the results were compared against a highly validated universal primer set, to determine the potential utility of the novel primers from the current study, in the context of water quality monitoring.

Materials and Methods

Cyanobacteria cultures, sampling and DNA extraction

Two types of water sources were used in this study: Cyanobacterial cultures isolated in a previous study [33], and heterogeneous environmental samples obtained from a variety of public sites and unrestricted ecosystems around Perth, WA (Table 1). These water samples were obtained with the knowledge of the relevant municipal authorities. Field campaigns were conducted in open spaces accessible to the public and no permissions were required. Field campaigns did not involve any interaction with endangered or protected species or ecosystems. Urban freshwater samples (1 L) were collected in the winter of 2014 (Piney Lakes Reserve) or summer 2015. Grab samples (1 L) were collected in the winters of 2014 and 2015, from the final effluents of three wastewater treatment plants (WWTP) in Western Australia. These WWTP collection sites had previously been selected and sampled with the assistance and permission from the managing authority (WaterCorporation of WA). to cover different construction design, treatment technologies, and operating regimes. Extractions of DNA from Cyanobacterial cultures ($n = 6$) were performed as previously described [41]. Whole DNA was extracted from environmental water samples using the MOBIO PowerWater Sterivex DNA

Table 1. Environmental water samples collection sites.

Site	Approx. GPS Location	Sample type
Blue Gum Lake	-32.0408, 115.8467	Urban freshwater
Booragoon Lake	-32.0443, 115.8415	Urban freshwater
Frederick Baldwin Park Lake	-32.0594, 115.8131	Urban freshwater
Marmion Reserve Lake	-32.0390, 115.8114	Urban freshwater
Murdoch University Chinese Garden Koi Pond	-32.0667, 115.8324	Urban freshwater
Piney Lakes Reserve Lake	-32.0497, 115.8381	Urban freshwater
Alfred Cove Drain	-32.0291, 115.8126	Urban runoff
WWTP1	North West region of WA*	Wastewater treatment plant final effluent
WWTP2	South-West region of WA*	Wastewater treatment plant final effluent
WWTP3	South-West region of WA*	Wastewater treatment plant final effluent

*exact location is confidential.

doi:10.1371/journal.pone.0170008.t001

Isolation kit (MOBIO, USA) according to the manufacturer’s instructions (n = 10). Extraction blanks (n = 2) were always included in every batch of DNA extractions; these consisted of mock-extractions from only fresh reagents (for extractions from cell cultures) or a fresh Steri-vex filter (for extractions from water samples). All reactions were carried out using DNA-free reagents and consumables, with amplification and sequencing reactions set up in a dedicated laboratory area, physically isolated from post-amplification areas and free of PCR products.

Primer design and *in silico* validation

Cyanobacteria-specific primers separately targeting the V3 and V6 regions (Locations indicated in S1 Fig) of the 16S rDNA were designed using the Primer 3 [42] add-on in Geneious Pro 8.0.4 [43]. The primers were designed from an alignment of 6,513 cyanobacteria 16S rDNA sequences retrieved from NCBI GenBank (April 2014), with lengths ranging from 144 to 1,481 bp. The alignment was constructed in Geneious Pro 8.0.4 [43], using the MAFFT plugin [44] after discarding non-overlapping sequences. Primers were designed using target consensus sequences generated for a range of stringency levels (e.g., 40% -75% consensus). The sequences and regions targeted by the primers used in this study are shown in Table 2. *In silico* tests of primer specificity were conducted using NCBI BLAST, TestPrime 1.0 (<http://www.arb-silva.de/search/testprime/>) and ProbeMatch (<https://rdp.cme.msu.edu/probematch>) and the SILVA and Ribosomal Database Project (RDP) databases. The settings chosen for the SILVA analysis by TestPrime 1.0 were: Maximum number of mismatches: 1; length of

Table 2. List of primers used in the present study, region targeted and amplification properties.

Primer name	Sequence (5'-3')	Location on <i>E.coli</i> K-12 (NR102804)	Target region	Annealing temperature	Approximate amplicon size	Reference
293F	AGCCACACTGGGRCTGAGA	312–331	V3	50°C	255 bp	This study
751R	TGCGGACGCTTTACGCCCA	572–590				
515F	GTGCCAGCMGCCGCGGTAA	523–541	V4	55°C	253 bp	[20]
806R	GGA CTACHVGGGTWTCTAAT	796–816				
1328F	GCTAACGCGTTAAGTATCCCGCCTGG	870–896	V6	55°C	298 bp	This study
1664R	GTCTCTCTAGAGTGCCCAACTTAATG	1166–1185				

doi:10.1371/journal.pone.0170008.t002

0-mismatch zone at 3' end: 5 bases; SILVA Database: SSURef-122 NR. One and two differences were allowed for the RDP analysis by ProbeMatch.

Library preparations and NGS

During the present study, the designed primers were used together with a universal primer pair, targeting the V4 hypervariable region of the 16S rRNA gene, amplified with the modified versions [45] of the 515F_806R primers [20], following the protocol version 6_15 (<http://www.earthmicrobiome.org>). For the generated primers, thermal cycling conditions were optimized in 25 μ L reactions containing 1U of PerfectTaq DNA polymerase (5 Prime, Germany), 2 mM $MgCl_2$, 200 μ M of each deoxynucleotide triphosphate (dNTP, Promega), 0.2 μ M of each primer and 1 μ L (approx. 30ng) DNA template obtained either during a previous study [33]. Gradient PCR protocols were implemented in a Veriti 96-well Thermal Cycler (Applied Biosystems) with annealing temperatures ranging from 50°C to 60°C. Amplicons were visualised on a 2% agarose gel.

With fusion primers (IDT, USA), three separate amplification reactions were carried out per primer, for each sample; reaction mixtures were the same as above, but also contained 0.01 mg BSA (Fisher Biotech, Australia) and 3.3 μ M SYTO 9 (Thermo Fisher Scientific, USA). No-template-controls (NTC) and extraction blanks were included in every experiment. Thermal cycling involved an initial denaturation at 95°C for 5 mins, followed by 45 cycles of denaturation at 95°C for 30 s, annealing at the appropriate temperature for 30 s, and extension at 72°C for 45 s and a final extension at 72°C for 7 min, which was followed by a melting curve analysis in a Step-One real-time qPCR instrument (Applied Biosystems, USA). Chosen annealing temperatures were 50°C and 55°C for 293F_751R and 1328F_1664R, respectively.

Multiplexing was based on six-to-eight bp multiplex identifier (MID) sequences (Thermo Fisher Scientific, USA); a unique combination of forward and reverse MID sequences was used for each sample.

To obtain the final library, amplicons from all samples and controls were pooled in equimolar amounts, based on the Ct values obtained from each sample by qPCR-based relative quantifications. This was purified twice using 1.2 volumes of Agencourt Ampure XP beads (Agilent Technologies, USA) and quantified by qPCR using a known concentration of a serially diluted 152 bp synthetic oligonucleotide as a standard. For qPCR-based absolute quantifications, reactions contained approximately 50 pg of library, Power SYBR Green Master Mix (Thermo Fisher Scientific, USA), 0.4 μ M of each primer targeting the A and P1 Ion Torrent adapters (IDT, USA); an annealing temperature of 60°C was chosen.

Emulsion PCR and enrichment were performed according to the manufacturer's recommendations on the One-Touch 2 and One-Touch ES instruments (Thermo Fisher Scientific, USA). Sequencing was performed on an Ion Torrent PGM (Thermo Fisher Scientific, USA) using 400 bp chemistry and 316-V2 semiconductor chips, following the manufacturer's recommendations.

Bioinformatics analysis

Sequences were first processed in Geneious Pro 8.0.4 [43] by retaining only reads with perfect matches with the forward- and reverse-primers and MID sequences (no mismatches allowed). Sequences were then de-multiplexed into individual samples based on their unique combination of MID sequences. Primers were trimmed from each read, and sequences were filtered based on length by discarding all reads shorter than 200 bp. All sequences obtained with the same primer set, from all samples, were renamed and grouped, prior to further analysis. At this stage, there were 381,504, 950,040 and 284,297 sequences for 293F_751R, 515F_806R and

1328F_1664R, respectively. The reads were then quality-filtered using USEARCH [46], allowing only reads with a < 1% error rate to remain and singletons were removed on a per-sample basis. In order to have equal sampling depth for each primer set, every sample was then sub-sampled to an equal number of sequences in QIIME [47].

Operational taxonomic units (OTUs) were selected by clustering sequences at 97% similarity with the UPARSE algorithm [48], and bacterial genera present were identified. OTUs were checked against the ChimeraSlayer Gold reference database with the UCHIME algorithm [49] to ensure OTUs were not the result of chimeric reads. Taxonomy was assigned to OTUs against the GreenGenes 16S rRNA gene database (August 2013 release) [50] in QIIME 1.9.0 [47] using the UCLUST algorithm [46] with default parameters. Sequences identified in extraction blanks and NTCs were bioinformatically removed from each sample in which they appeared, to eliminate background contamination. Diversity analyses were then performed, with the number of sequences analysed for each primer set determined by the sample with the least number of sequences (for rarefaction).

16S rDNA sequences from the most abundant phyla obtained by all three primer sets (i.e., from Cyanobacteria and Proteobacteria), were compared against the NCBI GenBank nucleotide database, using the BLAST algorithm to obtain increased taxonomic resolution. The results were then visualised using the MEtaGenome ANalyzer (MEGAN) version 5 [51] (default settings except *Min Score* = 150 and *Min Complexity* = 0.3).

Results

In silico PCR

The results from both the TestPrime and ProbeMatch *in silico* analyses indicated that, compared to the universal primers pair (515F_806R) [20], the novel primers 293F_751R and 1328F_1664R targeted at least three times more Cyanobacterial sequences (Table 3). The TestPrime analyses showed that Cyanobacterial sequences formed 15.0% and 95.0% of the total sequences amplified by the 293F_751R and 1328F_1664R primer pairs respectively, as compared to the universal 515F_806R for which Cyanobacterial sequences comprised only 2.0%. This increased specificity for Cyanobacterial sequences was also seen in the analyses performed using ProbeMatch (Table 3).

Based on *in silico* PCR analyses and compared to primers 515F_806R [20], the novel primers 293F_751R and 1328F_1664R showed higher taxonomic specificity by targeting a narrower range of phyla available in the SILVA database (Fig 1). The 293F_751R and 1328F_1664R primer pairs, provided matches for 19 and 7 main phyla respectively, compared to 50 phyla for 515F_806R.

The phylum Cyanobacteria included the largest fractions of the predicted target sequences of the two novel primer sets, compared to the universal primers: 15.0% (293F_751R) and 95.3% (1328F_1664R) compared to 2.1% (515F_806R). Primers 293F_751R appeared also highly specific for Proteobacteria, as compared to 515F_806R (82.7% vs. 38.2% of the predicted target sequences).

Table 3. *In silico* analysis showing the fraction (%) of Cyanobacterial sequences, from the sequences present in the SILVA and RDP databases, that are amplified by the primer pairs 293F_751R, 515F_806R and 1328F_664R.

Primer pair	SILVA	RDP	
		1 difference allowed	2 differences allowed
293F_751R	15.0%	23.0%	12.0%
515F_806R	2.0%	4.0%	4.0%
1328F_664R	95.0%	95.0%	60.0%

doi:10.1371/journal.pone.0170008.t003

293F_751R

515F_806R

1328F_1664R

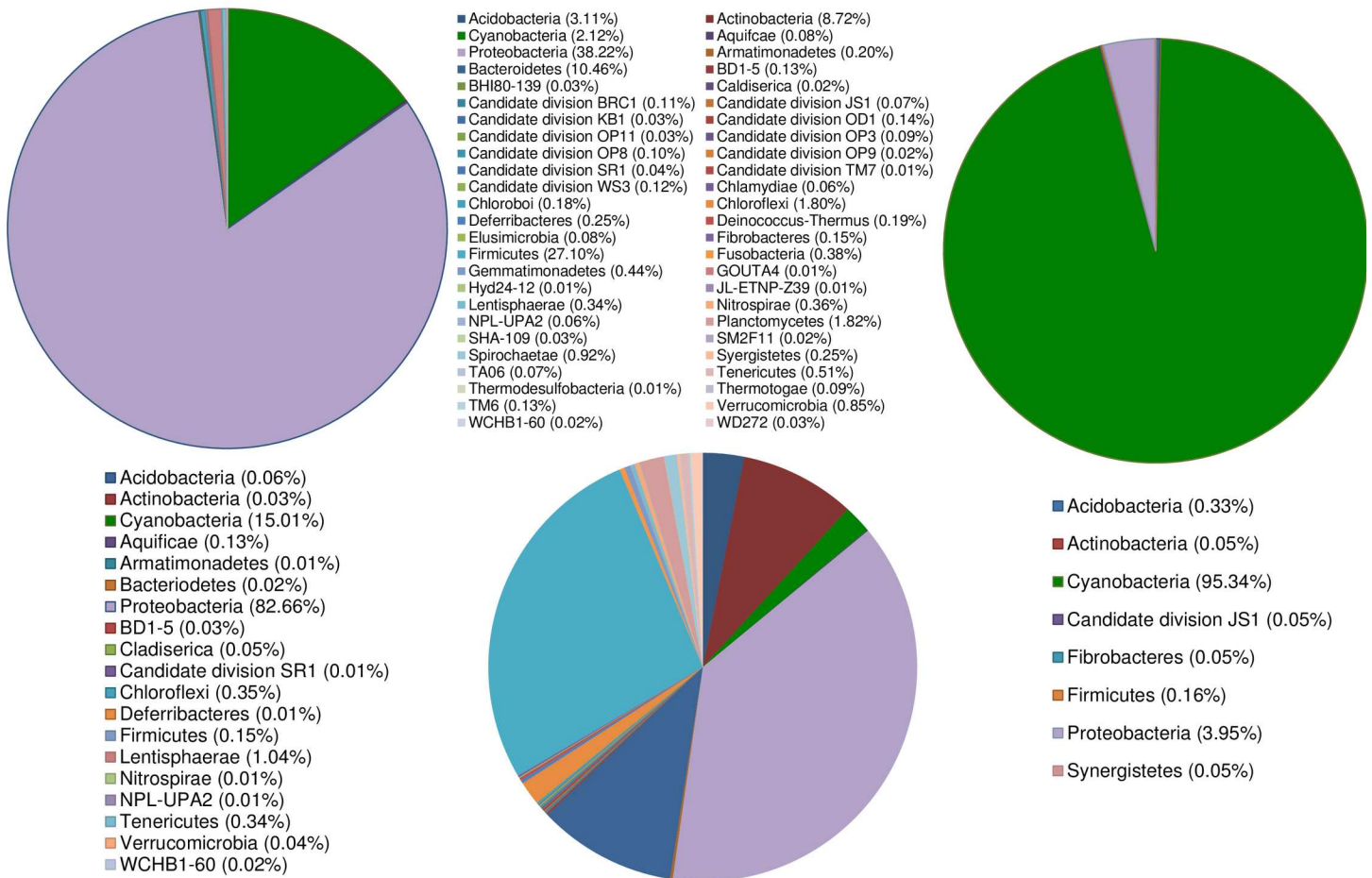


Fig 1. Pie charts showing the relative proportion of phyla amplified by each primer set used in the present study, as predicted by *in silico* PCR tests. *In silico* tests were performed against the SILVA database. For clarity, phyla forming less than 0.01% of the predicted amplifications were omitted.

doi:10.1371/journal.pone.0170008.g001

Community analysis by NGS

After *in silico* PCR, the novel primers were then tested by NGS on several environmental samples (Table 1) to confirm their utility and to determine optimal amplification conditions. The sequence and final optimal amplification conditions used for the primers designed in this study are given in Table 2. After quality filtering, a total of 339,008, 852,293 and 237,860 sequences remained for the 293F_751R, 515F_806R, and 1328F_1664R primer pairs respectively (BioProject ID: PRJNA330773). Subsampling resulted in the selection of 27,325 sequences per sample, from each of the three primer sets (Rarefaction analysis showing sampling depth shown in S2 Fig). Sample sequences were pooled by primer pair, which allowed for data comparisons across primers. In agreement with the *in silico* analyses, after exclusion of background contaminations and OTUs, which were either unassigned or classified only to the bacterial root, samples amplified with the universal 515F_806R primers had the greatest diversity of genetic variants within each sample (alpha diversity), while those with 1328F_1664R had the smallest (average of 44 OTUs vs. 23 OTUs respectively) (Fig 2). The increased specificity of the novel primers is also reflected in the number of bacterial phyla obtained by NGS (Fig 3). The two novel primer sets seem to perform similarly to each other, especially compared to what was predicted *in silico*.

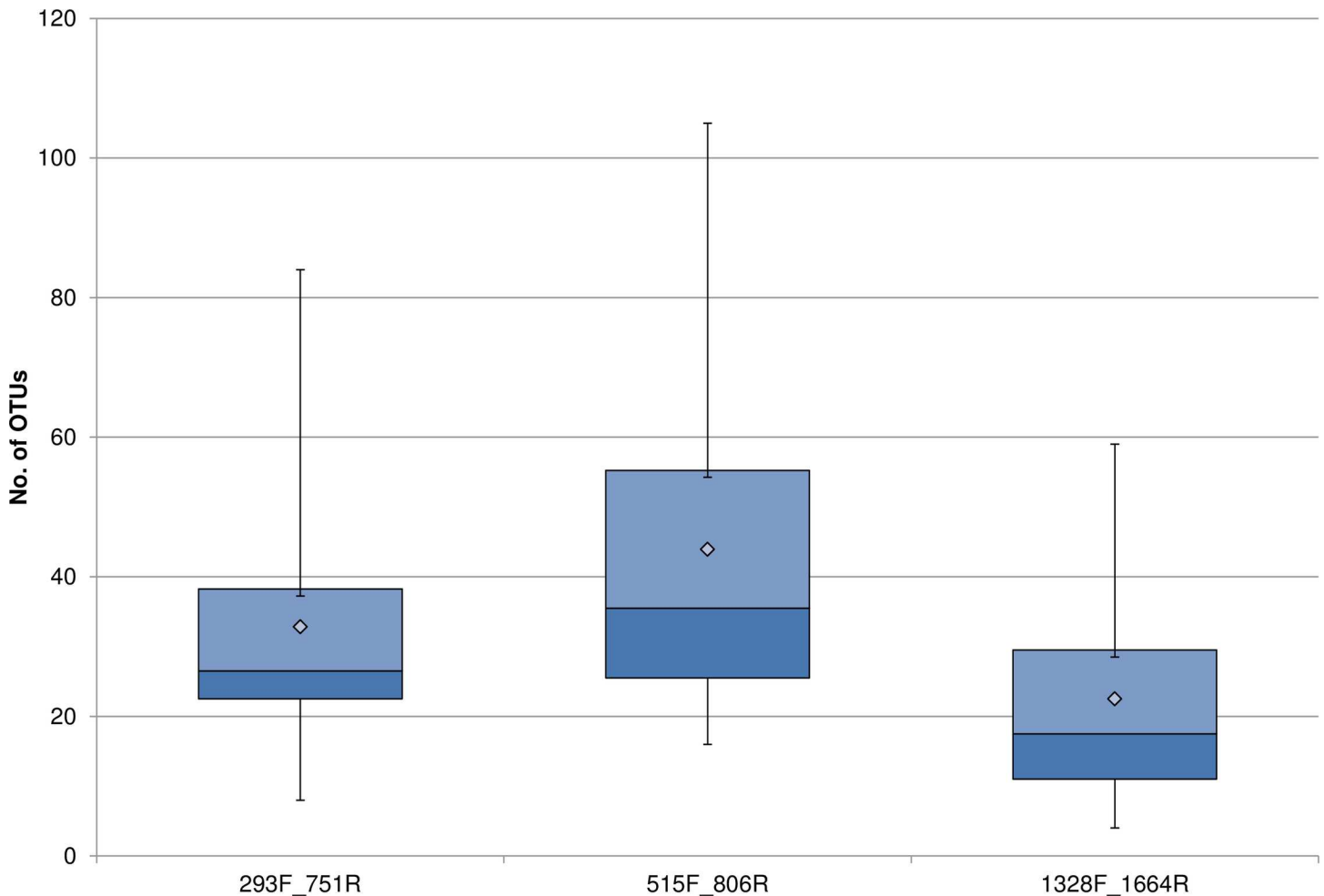


Fig 2. Box plot showing the alpha diversity values observed from all samples amplified by each primer set. Grey diamonds represent the average number of OTUs observed by each primer pair; boxes give minimum, first quartile, median, third quartile, and maximum, for each primer pair.

doi:10.1371/journal.pone.0170008.g002

The 293F_751R, 515F_806R, and 1328F_1664R primers detected a total of 8, 14 and 7 bacterial phyla respectively and also successfully amplified members of the phylum Proteobacteria. Altogether, Proteobacteria and Cyanobacteria formed approximately 95.0% of the subsampled sequences obtained by both the 293F_751R and 1328F_1664R primer pairs, as opposed to the universal 16S rDNA primers, where these two groups comprised less than half (48.7%) of the sequences (Fig 3). Other common environmental phyla were also detected and formed various proportions of the subsampled data for each primer set (e.g., Actinobacteria, Verrucomicrobia, Firmicutes, Bacteroidetes etc.). These bacterial groups, however, were only represented as minor fractions ($\leq 2.6\%$) of the taxa audited by the 293F_751R and 1328F_1664R primer pairs. In contrast, the broader specificity of the universal 16S rDNA primers 515F_806R [20] allowed greater amplification of sequences from these phyla, which formed at least 5.3% relative abundance (Fig 3, detailed results in S1 File).

The potential of the primers to taxonomically identify sequences to a lower (genus) level was compared using rank abundance curves (Fig 4). For the phylum Cyanobacteria, the primer set 293F_751R had the largest number of successfully assigned organisms, followed by 1328F_1664R and the universal 16S rDNA primers 515F_806R. For the identification of

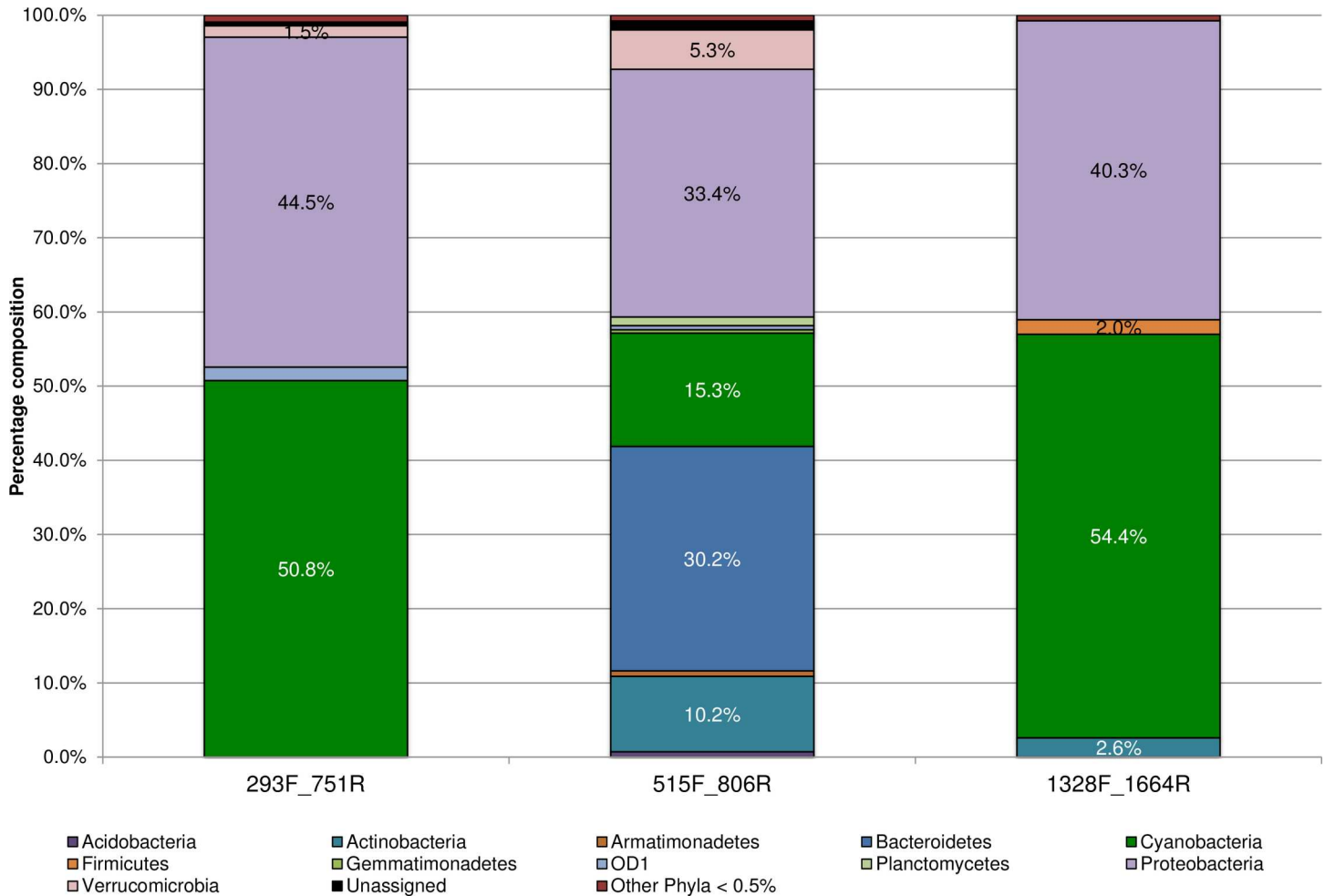


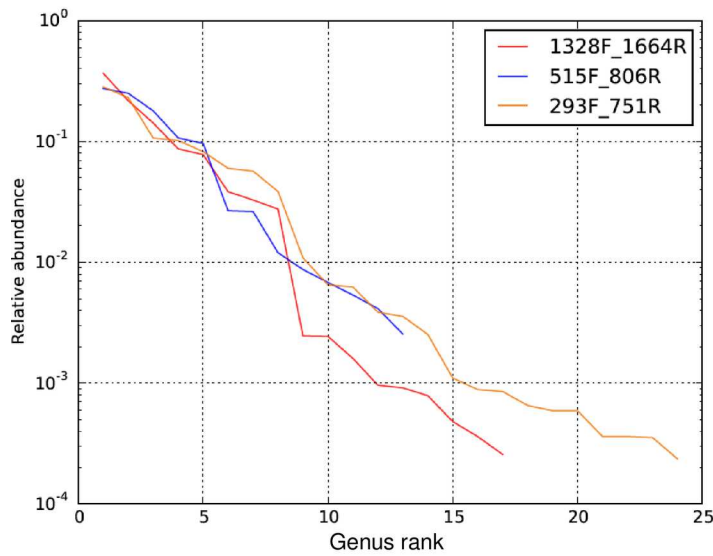
Fig 3. Percentage composition of sequences belonging to the indicated phyla, amplified by each primer set used in the present study. Subsampling was carried out to allow equal sequencing depth across all samples (~27,325 reads/sample) and to facilitate comparisons. For clarity, phyla forming < 0.5% abundance for each primer pair were combined.

doi:10.1371/journal.pone.0170008.g003

members of the Proteobacteria phylum, the 293F_751R and the universal 16S rDNA primers performed similarly (Fig 4).

Comparison of the results for each phyla showed that each of the primer sets appeared to preferentially amplify different orders from within the Cyanobacteria and Proteobacteria phyla. For Cyanobacteria, the 293F_751R primer set amplified all (9/9) of the 4C0d-2/Chloroplast (class), all (5/5) of the Pseudanabaenales (order), and 3/5 of the Oscillatoriophycideae (class) (1 Oscillatoriales and 2 Chroococcales (class)) observed. For 1328F_1664R, only Cryptophyta (order) was amplified from the 4C0d-2/Chloroplast class; 4/6 of the Nostocales (order), 4/5 of the Pseudanabaenales (order), all the Chroococcales (order) (3/3) and all of the Synechococcales (order) (2/2) observed were amplified. In comparison, the universal 16S rDNA primers amplified the smallest range of Cyanobacteria (cf. Fig 4, S2 File), amplifying 5/9 of 4C0d-2/chloroplast (class), 3/6 of the Nostocales and 2/5 of the Pseudanabaenales class/orders, a single genus from the Oscillatoriophycideae (class) was amplified, while no amplification was obtained for the Synechococcales (order) and Chroococcales (order). For the Proteobacteria, a similar pattern was also observed with the different primer sets having different degrees of amplification success for the various Proteobacterial orders (as illustrated in S3 File).

Cyanobacteria



Proteobacteria

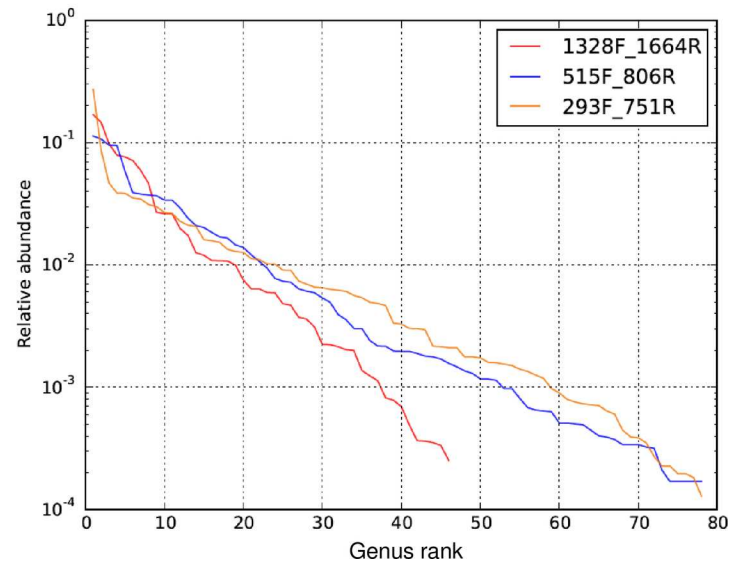


Fig 4. Comparison of taxonomic coverage of Cyanobacteria and Proteobacteria by primer sets.

doi:10.1371/journal.pone.0170008.g004

BLAST-searches were performed to further test the ability of the primers for detecting taxa of potential health and environmental concern. Human, animal and plant pathogens, as well as toxic bloom-forming Cyanobacterial species were detected by all primers in the water samples collected during the present experiments (Table 4). For instance, the 293F_751R primer pair demonstrated the potential to identify the presence of pathogenic species or genera such as *Neisseria canis*, *Coxiella burnetii*, the causative agent for Q fever [52], and *Ralstonia* sp., *Burkholderia* sp. (genera including plant and human pathogens). This demonstrates the potential usefulness of the novel primers in future DNA-based applications towards water quality monitoring.

Table 4. Illustrative comparison of taxa of potential health and environmental concern as detected by the various primer pairs, in the water samples collected during the present study. Classifications are given according to the lowest-level taxonomic assignments obtained, by bioinformatics analyses. Listed genera include those containing medically important organisms of humans and animals and toxic bloom-forming microalgae.

Phylum	Proteobacteria			Cyanobacteria		
	293F_751R	515F_806R	1328F_1664R	293F_751R	515F_806R	1328F_1664R
	<i>Alcaligenes</i> sp.	<i>Acrobacter</i> sp.		<i>Calothrix</i> sp.	<i>Calothrix</i> sp.	<i>Anabaena oscillarioides</i>
	<i>Aeromonas</i> sp.	<i>Coxiella burnetii</i>		<i>Limnothrix planktonica</i>	<i>Planktothrix</i> sp.	<i>Aphanizomenon gracile</i>
	<i>Burkholderia</i> sp.	<i>Pasteurella testudinis</i>		<i>Phormidium</i> sp.	<i>Trichormus variabilis</i>	<i>Limnothrix</i> sp.
	<i>Coxiella burnetii</i>	<i>Pseudomonas</i> sp.		<i>Planktothrix</i> sp.		<i>Nostoc linckia</i>
	<i>Endozoicomonas elysicola</i>	<i>Rickettsiales</i> sp.		<i>Trichormus variabilis</i>		<i>Phormidium</i> sp.
	<i>Neisseria canis</i>					<i>Pseudanabaena gelata</i>
	<i>Ralstonia</i> sp.					<i>Pseudanabaena limnetica</i>
						<i>Sphaerospermopsis aphanizomenoides</i>

doi:10.1371/journal.pone.0170008.t004

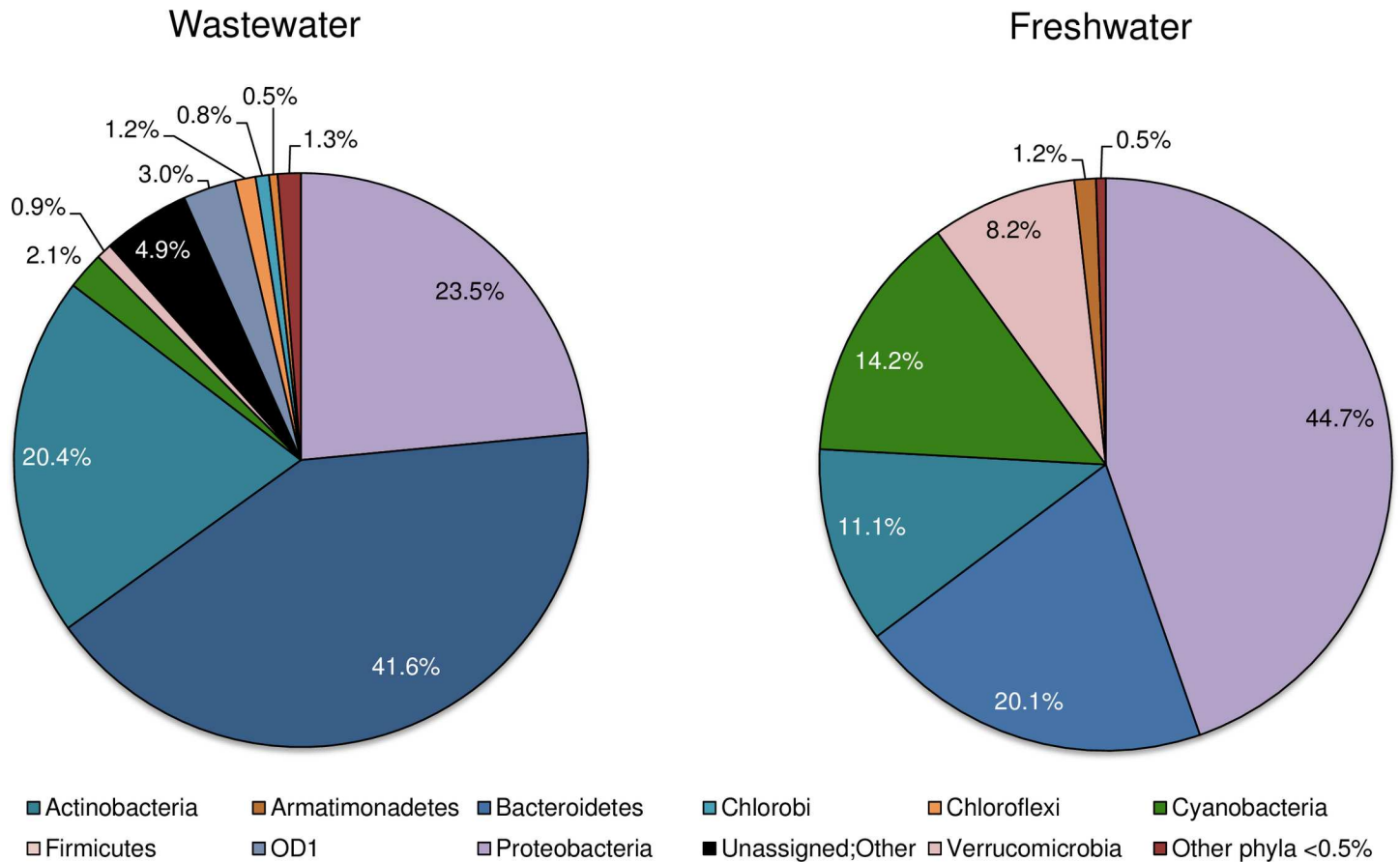


Fig 5. Composition of the main microbial phyla detected, by the universal primer set 515F_806R, in the wastewater and freshwater samples analysed during the study. For clarity, phyla forming < 0.5% abundance for each primer pair were combined.

doi:10.1371/journal.pone.0170008.g005

Microbial community in environmental water samples

For this study, a diverse range of environmental water sources, supposedly harbouring different microbial communities was chosen. To ensure that the assemblages were indeed different, an analysis of the results obtained by the universal primer set was performed. To this end, all the Bacterial and Archaeal phyla detected by the primers 515F_806R were pooled by water source (wastewater and freshwater) and compared (Fig 5; S4 File).

Overall, a greater range of the major bacterial phyla (n = 9) was observed in the wastewater samples, compared to freshwater (n = 5). Proteobacteria formed the largest proportion of bacterial phyla (34.1% on average, between the two water sources) but their average relative abundance clearly differed by water source, ranging from 44.7% in freshwater to 23.5% in wastewater. Bacteroidetes (30.8% on average, between the two water sources) and Actinobacteria (15.7%) were the second and third most abundant phyla, and were about twofold more abundant in wastewater than in freshwater. Interestingly the universal primer set detected Cyanobacteria at an average relative abundance of only 8.1%. This phylum was the fourth most abundant phylum overall, with values ranging from 14.2% in freshwater to only 2.1% in wastewater.

The fluctuations in the average relative abundance of the various taxa, detected by the universal primer set 515F_806R, confirm the dissimilarities in microbial composition, across the water samples chosen for the study.

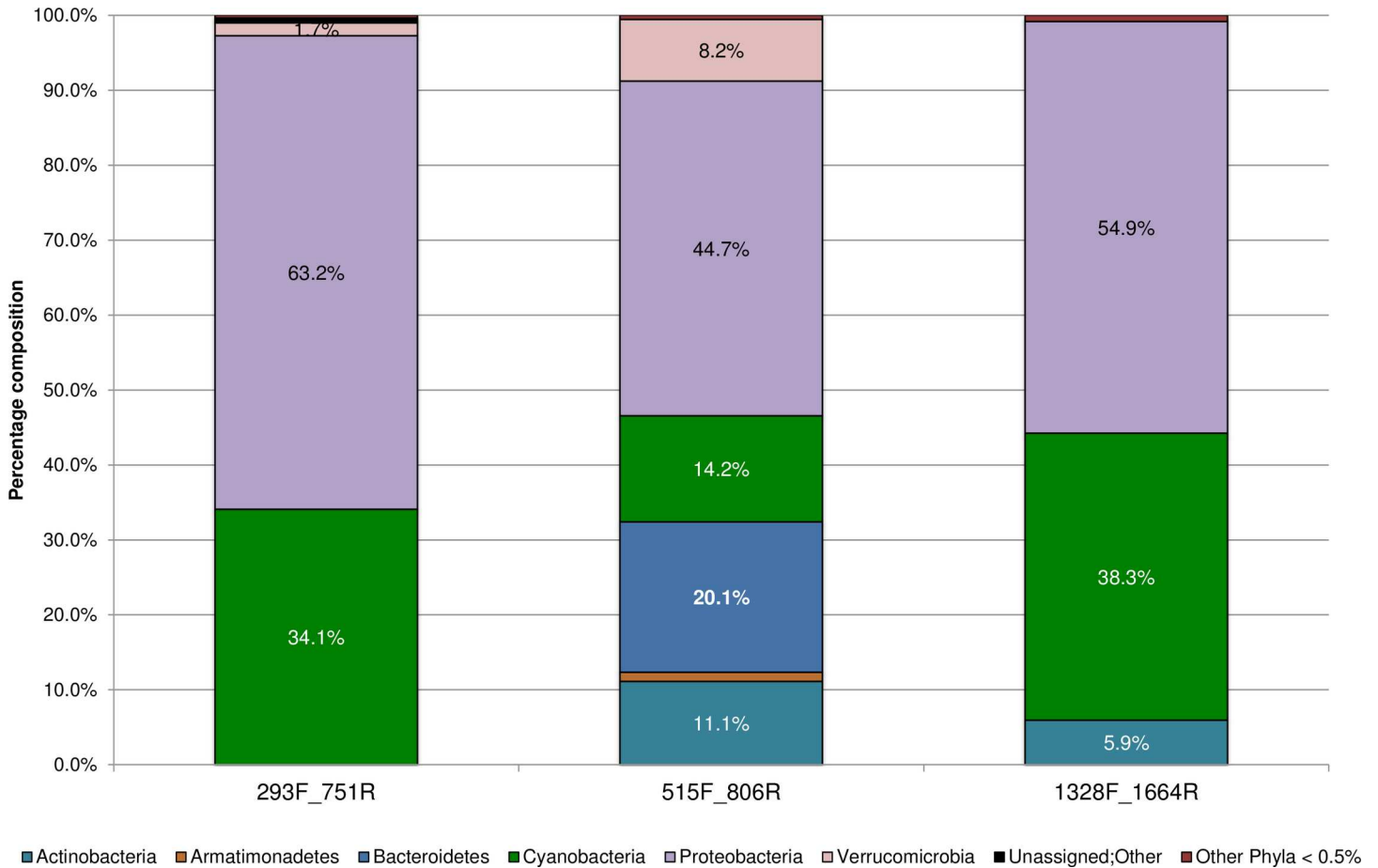


Fig 6. Composition of the main microbial phyla detected, by each primer set, in the freshwater samples analysed during the study. For clarity, phyla forming < 0.5% abundance for each primer pair were combined.

doi:10.1371/journal.pone.0170008.g006

Comparison of primer performance with freshwater samples

Due to the critical impact of bacterial contaminations on source water, a dedicated analysis was performed on the NGS data collected from freshwater samples only (Fig 6). Proteobacteria formed the largest proportion ($\geq 44.7\%$) of the bacterial phyla sequenced by all primer pairs (Fig 6). The specificity of the novel primers is clear, with Cyanobacteria and Proteobacteria combined forming more than 95.3% of the total fraction of sequences generated by the 1328F_1664R and 293F_751R primers. This is in comparison to the universal 16S rDNA primers where reads from these two phyla combined formed only 58.9% of the total (Actinobacteria, Bacteroidetes and Verrucomicrobia formed the major proportions of the remaining 41.1% of sequences). Apart from Verrucomicrobia (1.7% of the sequences generated by the 293F_751R primers) and Actinobacteria (5.9% of the sequences generated by the 1328F_1664R primers), all other phyla formed less than 1% of the total sequences generated, from freshwater, by the novel primers from the present study (Fig 6).

Discussion

This study sought to design Cyanobacteria-specific 16S rDNA primers for use with NGS platforms, and to allow for rapid detection and identification of Cyanobacterial strains from a

variety of environmental water samples. To this end, two different primer sets targeting different regions (V3, and V6 regions) of the 16S rDNA gene were designed and tested to compare their utility. This was performed in order to overcome the limitations posed by the use of universal 16S rDNA primers that were naturally designed to amplify across as broad a range as possible of bacterial and archaeal taxa.

The abundance of sequence data makes the 16S rDNA locus ideal for the design of phylum-specific primers and for the successive taxonomic assignment. *In silico*- and laboratory-tests performed during the present study confirmed that the novel primers have specificity for Cyanobacteria amplifying at least three times more Cyanobacterial sequences than the universal 16S rDNA primers (Fig 1, Table 3). Increased specificity for this phylum of algae is useful in water monitoring programs. For example, when there is the need to detect the presence of problematic Cyanobacteria present at low levels within complex multi-taxa aquatic communities whilst avoiding other non-problematic taxa. In these situations, the ability to enrich for members of a specific phylum of interest is clearly advantageous and we demonstrate this approach here together with the recovery of a broader range of Cyanobacterial orders when compared to the use of universal primers (Fig 4).

In addition to the preferential amplification of Cyanobacteria specific sequences, the specificity of the primers designed in this study was also directed toward the detection of Proteobacteria. Although the primers showed bias towards Proteobacteria sequences, they produced similar Proteobacterial profiles to that of the universal primers (Fig 4). This bias for Proteobacterial sequences was evident based upon the SILVA TestPrime results (Fig 1, esp. for 293F_751R), but was not observed in our initial primer specificity tests, or, when the primers were tested on Cyanobacterial cultures (i.e. sequencing results of the primers did not produce mixed signals, data not shown). However, amplification of Proteobacterial sequences by the 293F-751R and 1328F-1664R primer sets does not detract from their increased selectivity for Cyanobacterial sequences. This is best demonstrated in Fig 3 whereby these primers amplified more than twice as many Cyanobacterial sequences as compared to the universal 515F_806R primer set.

Proteobacteria are the largest and most phenotypically diverse bacterial division, accounting for the majority of known Gram-negative bacteria [53–55]. They form a complex phylum, arbitrarily divided into subdivisions based on their rRNA gene sequences, with pleomorphic members with various physiological and phenotypic characteristics, including numerous known human and animal pathogens [53, 55–57]. As such, the potential also exists to utilise these novel primers as a tool for detecting potentially pathogenic Proteobacteria from mixed communities.

For this study, the freshwater samples were obtained mainly from lakes/ ponds (lentic water systems). Future studies testing an even wider diversity of water sources (e.g. marine samples) or mock bacterial communities can be performed to better determine the applicability and limitations of these primers.

In specific applications where Cyanobacteria and Proteobacteria are considered the target groups, the increased specificity of the primers designed in this study allows for the pooling more samples in the same NGS run, while still guaranteeing adequate sequencing depth and spectrum of target taxa. This has obvious cost benefits. When combined with the faster sample processing and increased taxonomic resolution, this makes NGS cheaper, and faster, than current traditional morphology-based taxonomic methods of bioassessment [14]. The ever reducing cost per base of sequencing [3, 58], together with the reproducibility and potential for standardisation [8] make NGS a potential tool for biomonitoring [12, 19, 26, 59]. When Cyanobacteria and Proteobacteria are important target taxa to monitor, the increased selectivity of the novel primers should allow parallel sequencing of at least twice the number of samples, per

each run, compared to the universal primers. This should allow a reduction of the NGS costs, yet guaranteeing optimal coverage and species discrimination.

Supporting Information

S1 Fig. Cyanobacteria targeted primers as mapped onto *Escherichia coli*. Screen capture from Geneious (Biomatters, NZ) showing cyanobacteria 16S rDNA consensus sequence mapped onto *E. coli* K-12 substrain MG1655 (NR10284) Annotations indicate the positions of the hypervariable regions (blue boxes), and primer binding sites. Black arrows indicate positions of reference primers, green/grey arrows indicate the positions of primers designed in this study.

(DOCX)

S2 Fig. Rarefaction plots showing OTUs observed by primer set. Rarefaction of sampled OTUs observed by the primer sets used in the study. A sampling depth of 205 taxonomic counts per sample was used to ensure inclusion of all samples used in the study.

(DOCX)

S1 File. Detailed relative abundance of phyla detected by each primer set used in the study.

(XLSX)

S2 File. Detailed relative abundance of cyanobacterial sequences detected by each primer set used in the study. Class and order, together with the total number of genera detected by all primer sets are indicated in the left-most column.

(XLSX)

S3 File. Detailed relative abundance of proteobacterial sequences detected by each primer set used in the study. Class and order, together with the total number of genera detected by all primer sets are indicated in the left-most column.

(XLSX)

S4 File. Relative abundance of phyla in the environmental water samples as amplified by each primer set used in the study. Breakdown of the different phyla detected in each water source by each primer set used in the study.

(XLSX)

Acknowledgments

Authors' gratitude goes to the West Coast Microbiome Network (www.wcmicrobiome.org), Dr David Berryman, Ms Frances Brigg, Mr Alex Gofton (Murdoch University) for assistance. Resources of the Pawsey Supercomputing Centre (BLAST on MAGNUS) were used during some steps of the data analysis under an existing project (The microbiome of Australia's source water). Partial financial support for this study was indirectly provided by Murdoch University, the Australian Research Council (LP100100160), Bayer Healthcare, Bayer Australia Ltd., Western Australia Water Corporation, WaterRA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: AP.

Formal analysis: AP EL.

Funding acquisition: AP UR.

Investigation: AP EL MK.

Methodology: AP EL MK.

Resources: AP EL UR CG AB.

Supervision: AP UR.

Validation: AP EL.

Visualization: AP EL.

Writing – original draft: AP EL.

Writing – review & editing: AP EL MK AW PM AB CG UR.

References

1. Kolmakova OV, Gladyshev MI, Rozanov AS, Peltek SE, Trusova MY. Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing. *FEMS Microbiol Ecol.* 2014; 89(2):442–50. doi: [10.1111/1574-6941.12355](https://doi.org/10.1111/1574-6941.12355) PMID: [24848736](https://pubmed.ncbi.nlm.nih.gov/24848736/)
2. Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG, et al. An Expanded Genomic Representation of the Phylum Cyanobacteria. *Genome Biol Evol.* 2014; 6(5):1031–45. doi: [10.1093/gbe/evu073](https://doi.org/10.1093/gbe/evu073) PMID: [24709563](https://pubmed.ncbi.nlm.nih.gov/24709563/)
3. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A.* 2011; 108(Supplement 1):4516–22.
4. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A.* 2013; 110(3):1053–8. doi: [10.1073/pnas.1217107110](https://doi.org/10.1073/pnas.1217107110) PMID: [23277585](https://pubmed.ncbi.nlm.nih.gov/23277585/)
5. Holscher HD, Caporaso JG, Hooda S, Brulc JM, Fahey GC, Swanson KS. Fiber supplementation influences phylogenetic structure and functional capacity of the human intestinal microbiome: follow-up of a randomized controlled trial. *Am J Clin Nutr.* 2015; 101(1):55–64. doi: [10.3945/ajcn.114.092064](https://doi.org/10.3945/ajcn.114.092064) PMID: [25527750](https://pubmed.ncbi.nlm.nih.gov/25527750/)
6. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 2010; 38(22):e200–e. doi: [10.1093/nar/gkq873](https://doi.org/10.1093/nar/gkq873) PMID: [20880993](https://pubmed.ncbi.nlm.nih.gov/20880993/)
7. Mizrahi-Man O, Davenport ER, Gilad Y. Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *PLoS ONE.* 2013; 8(1):e53608. doi: [10.1371/journal.pone.0053608](https://doi.org/10.1371/journal.pone.0053608) PMID: [23308262](https://pubmed.ncbi.nlm.nih.gov/23308262/)
8. Eiler A, Drakare S, Bertilsson S, Pernthaler J, Peura S, Rofner C, et al. Unveiling Distribution Patterns of Freshwater Phytoplankton by a Next Generation Sequencing Based Approach. *PLoS ONE.* 2013; 8(1):e53516. doi: [10.1371/journal.pone.0053516](https://doi.org/10.1371/journal.pone.0053516) PMID: [23349714](https://pubmed.ncbi.nlm.nih.gov/23349714/)
9. Su C, Lei L, Duan Y, Zhang KQ, Yang J. Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Appl Microbiol Biotechnol.* 2012; 93(3):993–1003. Epub 2011/12/23. doi: [10.1007/s00253-011-3800-7](https://doi.org/10.1007/s00253-011-3800-7) PMID: [22189863](https://pubmed.ncbi.nlm.nih.gov/22189863/)
10. Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, et al. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Mol Ecol.* 2010; 19:4–20. doi: [10.1111/j.1365-294X.2009.04473.x](https://doi.org/10.1111/j.1365-294X.2009.04473.x) PMID: [20331766](https://pubmed.ncbi.nlm.nih.gov/20331766/)
11. Machida RJ, Knowlton N. PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. *PLoS ONE.* 2012; 7(9):e46180. doi: [10.1371/journal.pone.0046180](https://doi.org/10.1371/journal.pone.0046180) PMID: [23049971](https://pubmed.ncbi.nlm.nih.gov/23049971/)
12. Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C. Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLoS ONE.* 2014; 9(2):e87624. doi: [10.1371/journal.pone.0087624](https://doi.org/10.1371/journal.pone.0087624) PMID: [24516555](https://pubmed.ncbi.nlm.nih.gov/24516555/)
13. Lanzén A, Jørgensen SL, Bengtsson MM, Jonassen I, Øvreås L, Urich T. Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiol Ecol.* 2011; 77(3):577–89. doi: [10.1111/j.1574-6941.2011.01138.x](https://doi.org/10.1111/j.1574-6941.2011.01138.x) PMID: [21627670](https://pubmed.ncbi.nlm.nih.gov/21627670/)
14. Paparini A, Gofton A, Yang RC, White N, Bunce M, Ryan UM. Comparison of Sanger and next generation sequencing performance for genotyping *Cryptosporidium* isolates at the 18S rRNA and actin loci. *Exp Parasitol.* 2015; 151:21–7. doi: [10.1016/j.exppara.2015.02.001](https://doi.org/10.1016/j.exppara.2015.02.001) PMID: [25662433](https://pubmed.ncbi.nlm.nih.gov/25662433/)

15. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE*. 2011; 6(12):e27310. doi: [10.1371/journal.pone.0027310](https://doi.org/10.1371/journal.pone.0027310) PMID: [22194782](https://pubmed.ncbi.nlm.nih.gov/22194782/)
16. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013; 41(1):e1. doi: [10.1093/nar/gks808](https://doi.org/10.1093/nar/gks808) PMID: [22933715](https://pubmed.ncbi.nlm.nih.gov/22933715/)
17. Milani C, Lugli GA, Turrioni F, Mancabelli L, Duranti S, Viappiani A, et al. Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *FEMS Microbiol Ecol*. 2014; 90(2):493–503. doi: [10.1111/1574-6941.12410](https://doi.org/10.1111/1574-6941.12410) PMID: [25117972](https://pubmed.ncbi.nlm.nih.gov/25117972/)
18. Hong S, Bunge J, Leslin C, Jeon S, Epstein SS. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J*. 2009; 3(12):1365–73. doi: [10.1038/ismej.2009.89](https://doi.org/10.1038/ismej.2009.89) PMID: [19693101](https://pubmed.ncbi.nlm.nih.gov/19693101/)
19. Keramarrec L, Franc A, Rimet F, Chaumeil P, Humbert JF, Bouchez A. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. *Mol Ecol Resour*. 2013; 13(4):607–19. doi: [10.1111/1755-0998.12105](https://doi.org/10.1111/1755-0998.12105) PMID: [23590277](https://pubmed.ncbi.nlm.nih.gov/23590277/)
20. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012; 6(8):1621–4. doi: [10.1038/ismej.2012.8](https://doi.org/10.1038/ismej.2012.8) PMID: [22402401](https://pubmed.ncbi.nlm.nih.gov/22402401/)
21. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing. *PLoS ONE*. 2014; 9(8).
22. Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, et al. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods*. 2012; 91(1):80–8. doi: [10.1016/j.mimet.2012.07.008](https://doi.org/10.1016/j.mimet.2012.07.008) PMID: [22849830](https://pubmed.ncbi.nlm.nih.gov/22849830/)
23. Sinclair L, Osman OA, Bertilsson S, Eiler A. Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform. *PLoS ONE*. 2015; 10(2):e0116955. doi: [10.1371/journal.pone.0116955](https://doi.org/10.1371/journal.pone.0116955) PMID: [25647581](https://pubmed.ncbi.nlm.nih.gov/25647581/)
24. Menchaca AC, Visi DK, Strey OF, Teel PD, Kalinowski K, Allen MS, et al. Preliminary Assessment of Microbiome Changes Following Blood-Feeding and Survivorship in the *Amblyomma americanum* Nymph-to-Adult Transition using Semiconductor Sequencing. *PLoS ONE*. 2013; 8(6):e67129. doi: [10.1371/journal.pone.0067129](https://doi.org/10.1371/journal.pone.0067129) PMID: [23826210](https://pubmed.ncbi.nlm.nih.gov/23826210/)
25. Gofton AW, Oskam CL, Lo N, Beninati T, Wei H, McCarl V, et al. Inhibition of the endosymbiont "Candidatus Midichloria mitochondrii" during 16S rRNA gene profiling reveals potential pathogens in Ixodes ticks from Australia. *Parasit Vectors*. 2015; 8:345. Epub 2015/06/26. PubMed Central PMCID: [PMCPMC4493822](https://pubmed.ncbi.nlm.nih.gov/PMCPMC4493822/). doi: [10.1186/s13071-015-0958-3](https://doi.org/10.1186/s13071-015-0958-3) PMID: [26108374](https://pubmed.ncbi.nlm.nih.gov/26108374/)
26. Tan B, Ng CM, Nshimiyimana JP, Loh L-L, Gin KY-H, Thompson JR. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. *Front Microbiol*. 2015; 6.
27. Welker M, von Dohren H. Cyanobacterial peptides—Nature's own combinatorial biosynthesis. *FEMS Microbiol Rev*. 2006; 30(4):530–63. doi: [10.1111/j.1574-6976.2006.00022.x](https://doi.org/10.1111/j.1574-6976.2006.00022.x) PMID: [16774586](https://pubmed.ncbi.nlm.nih.gov/16774586/)
28. Humpage A, Falconer I, Bernard C, Froschio S, Fabbro L. Toxicity of the cyanobacterium *Limnospira AC0243* to male Balb/c mice. *Water Res*. 2012; 46(5):1576–83. doi: [10.1016/j.watres.2011.11.019](https://doi.org/10.1016/j.watres.2011.11.019) PMID: [22119368](https://pubmed.ncbi.nlm.nih.gov/22119368/)
29. Paerl HW, Otten TG. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microb Ecol*. 2013; 65(4):995–1010. doi: [10.1007/s00248-012-0159-y](https://doi.org/10.1007/s00248-012-0159-y) PMID: [23314096](https://pubmed.ncbi.nlm.nih.gov/23314096/)
30. Valério E, Chambel L, Paulino S, Faria N, Pereira P, Tenreiro R. Molecular identification, typing and traceability of cyanobacteria from freshwater reservoirs. *Microbiology*. 2009; 155(2):642–56.
31. Dall'agnol LT, Ghilardi-Junior R, McCulloch JA, Schneider H, Schneider MPC, Silva A. Phylogenetic and gene trees of *Synechococcus*: Choice of the right marker to evaluate the population diversity in the Tucuruí Hydroelectric Power Station Reservoir in Brazilian Amazonia. *J Plankton Res*. 2012; 34(3):245–57.
32. Wood SA, Smith KF, Banks JC, Tremblay LA, Rhodes L, Mountfort D, et al. Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. *New Zeal J Mar Fresh*. 2013; 47(1):90–119.
33. Lee E, Ryan UM, Monis P, McGregor GB, Bath A, Gordon C, et al. Polyphasic identification of cyanobacterial isolates from Australia. *Water Res*. 2014; 59:248–61. doi: [10.1016/j.watres.2014.04.023](https://doi.org/10.1016/j.watres.2014.04.023) PMID: [24810741](https://pubmed.ncbi.nlm.nih.gov/24810741/)
34. Kataoka T, Homma T, Nakano S, Hodoki Y, Ohbayashi K, Kondo R. PCR primers for selective detection of intra-species variations in the bloom-forming cyanobacterium, *Microcystis*. *Harmful Algae*. 2013; 23:46–54.

35. Rasmussen JP, Giglio S, Monis PT, Campbell RJ, Saint CP. Development and field testing of a real-time PCR assay for cylindrospermopsin-producing cyanobacteria. *J Appl Microbiol*. 2008; 104(5):1503–15. doi: [10.1111/j.1365-2672.2007.03676.x](https://doi.org/10.1111/j.1365-2672.2007.03676.x) PMID: [18179541](https://pubmed.ncbi.nlm.nih.gov/18179541/)
36. Al-Tebrineh J, Gehringer MM, Akcaalan R, Neilan BA. A new quantitative PCR assay for the detection of hepatotoxic cyanobacteria. *Toxicol*. 2011; 57(4):546–54. doi: [10.1016/j.toxicol.2010.12.018](https://doi.org/10.1016/j.toxicol.2010.12.018) PMID: [21194539](https://pubmed.ncbi.nlm.nih.gov/21194539/)
37. Fergusson KM, Saint CP. Molecular phylogeny of *Anabaena circinalis* and its identification in environmental samples by PCR. *Appl Environ Microbiol*. 2000; 66(9):4145–8. PMID: [10966445](https://pubmed.ncbi.nlm.nih.gov/10966445/)
38. Neilan BA, Jacobs D, Goodman AE. Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. *Appl Environ Microbiol*. 1995; 61(11):3875–83. PMID: [8526499](https://pubmed.ncbi.nlm.nih.gov/8526499/)
39. Man-Aharonovich D, Kress N, Zeev EB, Berman-Frank I, Béjà O. Molecular ecology of nifH genes and transcripts in the eastern Mediterranean Sea. *Environ Microbiol*. 2007; 9(9):2354–63. doi: [10.1111/j.1462-2920.2007.01353.x](https://doi.org/10.1111/j.1462-2920.2007.01353.x) PMID: [17686031](https://pubmed.ncbi.nlm.nih.gov/17686031/)
40. Manen J-F, Falquet J. The cpcB-cpcA locus as a tool for the genetic characterization of the genus *Arthrospira* (Cyanobacteria): evidence for horizontal transfer. *Int J Syst Evol Microbiol*. 2002; 52(3):861–7.
41. Gaget V, Gribaldo S, de Marsac NT. An rpoB signature sequence provides unique resolution for the molecular typing of cyanobacteria. *Int J Syst Evol Microbiol*. 2011; 61:170–83. doi: [10.1099/ijs.0.019018-0](https://doi.org/10.1099/ijs.0.019018-0) PMID: [20190018](https://pubmed.ncbi.nlm.nih.gov/20190018/)
42. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res*. 2012; 40(15).
43. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12):1647–9. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199) PMID: [22543367](https://pubmed.ncbi.nlm.nih.gov/22543367/)
44. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30(4):772–80. Epub 2013/01/19. PubMed Central PMCID: [PMC3603318](https://pubmed.ncbi.nlm.nih.gov/PMC3603318/). doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010) PMID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/)
45. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol*. 2015; 75(2):129–37.
46. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010; 26(19):2460–1. doi: [10.1093/bioinformatics/btq461](https://doi.org/10.1093/bioinformatics/btq461) PMID: [20709691](https://pubmed.ncbi.nlm.nih.gov/20709691/)
47. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010; 7(5):335–6. doi: [10.1038/nmeth.f.303](https://doi.org/10.1038/nmeth.f.303) PMID: [20383131](https://pubmed.ncbi.nlm.nih.gov/20383131/)
48. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013; 10(10):996–8. <http://www.nature.com/nmeth/journal/v10/n10/abs/nmeth.2604.html> - supplementary-information. doi: [10.1038/nmeth.2604](https://doi.org/10.1038/nmeth.2604) PMID: [23955772](https://pubmed.ncbi.nlm.nih.gov/23955772/)
49. Edgar RC, Haas BJ, Clemente JC, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011; 27(16):2194–21200. doi: [10.1093/bioinformatics/btr381](https://doi.org/10.1093/bioinformatics/btr381) PMID: [21700674](https://pubmed.ncbi.nlm.nih.gov/21700674/)
50. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol*. 2006; 72(7):5069–72. doi: [10.1128/AEM.03006-05](https://doi.org/10.1128/AEM.03006-05) PMID: [16820507](https://pubmed.ncbi.nlm.nih.gov/16820507/)
51. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res*. 2007; 17(3):377–86. doi: [10.1101/gr.5969107](https://doi.org/10.1101/gr.5969107) PMID: [17255551](https://pubmed.ncbi.nlm.nih.gov/17255551/)
52. Tissot-Dupont H, Raoult D. Q Fever. *Infect Dis Clin North Am*. 2008; 22(3):505–14. doi: [10.1016/j.idc.2008.03.002](https://doi.org/10.1016/j.idc.2008.03.002) PMID: [18755387](https://pubmed.ncbi.nlm.nih.gov/18755387/)
53. Gupta RS. The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev*. 2000; 24(4):367–402. PMID: [10978543](https://pubmed.ncbi.nlm.nih.gov/10978543/)
54. Imhoff JF. The Anoxygenic Phototrophic Purple Bacteria. In: Boone DR, Castenholz RW, Garrity GM, editors. *Bergey's Manual of Systematic Bacteriology*. vol. 1. 2 ed. New York: Springer; 2001. p. 631–7.
55. Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. *The Prokaryotes*. 4th ed. Berlin, Heidelberg: Springer 2014.
56. Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res*. 2014; 42(Database issue):D613–D6. doi: [10.1093/nar/gkt1111](https://doi.org/10.1093/nar/gkt1111) PMID: [24243842](https://pubmed.ncbi.nlm.nih.gov/24243842/)

57. Stackebrandt E, Murray RGE, Trüper HG. Proteobacteria classis nov., a Name for the Phylogenetic Taxon That Includes the "Purple Bacteria and Their Relatives". *Int J Syst Evol Microbiol.* 1988; 38(3):321–5.
58. Hajibabaei M. The golden age of DNA metasystematics. *Trends Genet.* 2012; 28(11):535–7. doi: [10.1016/j.tig.2012.08.001](https://doi.org/10.1016/j.tig.2012.08.001) PMID: [22951138](https://pubmed.ncbi.nlm.nih.gov/22951138/)
59. Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol Evol.* 2012; 27(4):233–43. doi: [10.1016/j.tree.2011.11.010](https://doi.org/10.1016/j.tree.2011.11.010) PMID: [22244672](https://pubmed.ncbi.nlm.nih.gov/22244672/)