ORIGINAL ARTICLE

Bacterial and eukaryotic microbial communities in urban water systems profiled via Illumina MiSeq platform

Adeline Su Yien Ting[1,4] · Muhammad Zarul Hanifah Md Zoqratt[1,2] · Hock Siew Tan[1] · Andreas Aditya Hermawan[3] · Amin Talei[3] · Soon Thiam Khu[3]

Abstract

Microbial communities from a lake and river flowing through a highly dense urbanized township in Malaysia were profiled by sequencing amplicons of the 16S V3–V4 and 18S V9 hypervariable rRNA gene regions via Illumina MiSeq. Results revealed that Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes were the dominant prokaryotic phyla; whereas, eukaryotic communities were predominantly of the SAR clade and Opisthokonta. The abundance of *Pseudomonas* and *Flavobacterium* in all sites suggested the possible presence of pathogens in the urban water systems, supported by the most probable number (MPN) values of more than 1600 per 100 mL. Urbanization could have impacted the microbial communities as transient communities (clinical, water-borne and opportunistic pathogens) coexisted with common indigenous aquatic communities (Cyanobacteria). It was concluded that in urban water systems, microbial communities vary in their abundance of microbial phyla detected along the water systems. The influences of urban land use and anthropogenic activities influenced the physicochemical properties and the microbial dynamics in the water systems.

Keywords 16S rRNA · 18S rRNA · Microbial community · Urban waters · Water quality

✉ Adeline Su Yien Ting
adeline.ting@monash.edu; adelsuyien@yahoo.com
https://scholar.google.com/citations?user=jJ7HgWIAAAAJ&hl=en&oi=ao

Muhammad Zarul Hanifah Md Zoqratt
Muhammad.ZarulHanifah@monash.edu

Hock Siew Tan
tan.hocksiew@monash.edu
https://scholar.google.com/citations?user=VvIzmIQAAAAJ&hl=en

Andreas Aditya Hermawan
andreas.aditya1@monash.edu
https://scholar.google.com/citations?user=o_dpUYoAAAAJ&hl=en

Amin Talei
amin.talei@monash.edu
https://scholar.google.com/citations?user=xQUS5BoAAAAJ&hl=en

Soon Thiam Khu
Soon.Thiam.Khu@monash.edu
https://scholar.google.com/citations?user=lmSRHSEAAAAJ&hl=en

1   School of Science, Monash University Malaysia, Jalan Lagoon Selatan, 47500 Bandar Sunway, Selangor, Malaysia

2   Genomics Facility, Tropical Medicine and Biology Platform, Monash University Malaysia, Petaling Jaya, Selangor, Malaysia

3   School of Engineering, Monash University Malaysia, Jalan Lagoon Selatan, 47500 Bandar Sunway, Selangor, Malaysia

4   School of Science, Monash University Malaysia, Jalan Lagoon Selatan, 46150 Bandar Sunway, Selangor Darul Ehsan, Malaysia

Introduction

Aquatic microbes are natural inhabitants of water systems. Common microbial communities in freshwater environments include members of Proteobacteria, Bacteroidetes,

Actinobacteria, Cyanobacteria, Verrucomicrobia and Planctomycetes (Zwart et al. 2002). They have major roles in various biochemical cycles; in nitrogen fixation, carbon recycling, and degradation of organic compounds in the aquatic ecosystem. In urban settings, the aquatic microbiome is speculated to be influenced by the urbanization process. Water pollution in urban water systems (e.g., lakes, rivers) is attributed to wastewater and industrial effluents (Sinang et al. 2015). Residential and industrial pollution flowing into freshwater ecosystems like artificial lakes can trigger the toxic cyanobacterial blooms (Sinang et al. 2015). Run-offs from agricultural sectors impede water quality as well, as the nutrient-rich input pollutes the water systems (Hafsi et al. 2016). This affects the aquatic ecosystem, decreasing water quality and possibly increasing risks of water-borne diseases that are hazardous to humans and other living organisms (Wang et al. 2011; Rashid et al. 2018). The change to water quality is expected to cause shifts in microbial communities (Ininbergs et al. 2015; Mittal et al. 2019), which may include the presence or possible proliferation of pathogenic strains (Noor et al. 2017). Microbial diversity changes as microbes adapt to the change in water quality. Microbes with determinative roles in various cycles are, therefore, influenced by the level of nutrients present in the aquatic system (Ininbergs et al. 2015). As such, river and lake water samples in urban areas are often used for the evaluation on the impact of anthropological activities to the environment (Cai et al. 2018; Jin et al. 2018).

In recent years, microbial communities have been characterized via molecular approaches, with the next-generation sequencing (NGS) technique emerging as a powerful tool for environmental microbiome research (Sogin et al. 2006; Jones et al. 2009; Mueller-Spitz et al. 2009). With NGS technology, based on either the whole-genome or 16S rRNA microbial sequences, various microbial communities have been successfully characterized (Ghai et al. 2011; Sunagawa et al. 2015). These microbial analyses have been extremely helpful in providing a greater insight in understanding aquatic communities. To date, the aquatic microbiome of pristine natural landscapes has been characterized more often than water systems in urban landscapes. Staley et al. (2013) detailed species richness along the Upper Mississippi river; whereas, Nakatsu et al. (2019) reported distinct microbial communities found in various sites in Lake Michigan. Medeiros et al. (2016) used whole-genome and amplicon sequence analyses to highlight that urbanized areas in Sao Pedro stream (Juiz de Fora, Brazil) were abundant in human pathogenic bacteria, contrasting with the ubiquitous environmental microbes found in non-urbanized areas. In South Africa, bacterial diversity of the water and sediment samples from the Apies River was affected by the settlements along the river (Abia et al. 2018). A similar observation was reported in China, whereby microbial communities

from the middle and downstream river revealed the dominance of Cyanobacteria and Thaumarchaeota, which were attributed to the input of untreated effluents (Bai et al. 2014). In all these reports, aquatic microbiome is observed to be evidently influenced by anthropogenic activities (i.e., settlements, effluents) although these sites are not exclusively from highly populated urban areas. These studies also revealed that point and non-point sources of pollution originating from municipal waste, agricultural runoff, industrial effluents, or fecal contamination, are known to contribute significantly in altering the microbial communities in the water systems. These changes are contributed by the nutrients, chemicals, pesticides, heavy metals, and presence of fecal coliforms, which flows into the water systems (Staley et al. 2013; Nakatsu et al. 2019). It is, therefore, evident that inputs from surrounding areas into the water systems influence the composition of the microbial communities.

In urban settings, lakes and rivers do exist and are often in close proximity to urban populations and inevitably receive inputs from neighboring sources as a consequence of land use. These aquatic water systems are, therefore, ideal for water analysis to determine the impact of anthropogenic influences and urbanization. In this study, the microbial communities of various urban water systems present in Bandar Sunway were profiled using the NGS technology. The urban water systems included in the analysis were the Kelana Jaya Lake, inlets into Kelana Jaya Lake and the outlet water from the lake, as well as multiple points from the Bandar Sunway river which flows through the various residential, industrial, commercial, recreational and construction areas in Bandar Sunway. It is postulated that the water quality and microbial composition differs in response to the anthropogenic activities and land use of the respective sites. Inputs such as those from commercial areas, construction sites, residential areas and golf courses, may influence the bacterial and eukaryotic communities due to the various levels of organic matter, suspended solids, nutrients, and dissolved oxygen, channeled into the sites. High numbers of coliforms, algae, cyanobacteria, or protists may then be found as a response to the inputs and activities in each site. The bacterial and eukaryotic communities found in the urban water systems and the association with physicochemical properties and site activities are, therefore, used to describe the impact of urbanization on the urban water systems.

The aim of this study is to characterize both prokaryotes and eukaryotes in urban water systems, so the impact of urbanization on the microbial communities can be understood. To achieve this, the water quality of all samples collected in this study was assessed using the Most Probable Number (MPN) test. Subsequently, microbial community analyses were performed using sequencing data generated from the Illumina MiSeq platform. The sequencing was carried out via amplification of the V3–V4 and V9

hypervariable region in the 16S and 18S rRNA genes to characterize the prokaryotic and eukaryotic communities, respectively. Results collected in this study provide valuable insights on the impact of urbanization on water systems in a highly urbanized setting (i.e., Bandar Sunway). This study is significant as it documents the bacterial and prokaryotic communities in urban water systems, which has not been fully explored. It also profiled the eukaryotic aquatic communities using MiSeq Illumina, which are not extensively studied compared to bacterial communities.

Materials and methods

Sampling sites

This study was carried out in an urbanized metropolitan city, Bandar Sunway, Selangor, Malaysia. This city is one of the largest population hubs in Malaysia, known as a center for education, tourism, high-quality living, and commercial activities. Several sites representing the water systems in Bandar Sunway were selected for this study (Fig. 1). The primary study site was Kelana Jaya Lake, which is bordered by

Fig. 1 Map of the sampling locations for 16S and 18S rRNA-based microbiome analysis. Information of these sampling sites are tabulated in Table 1
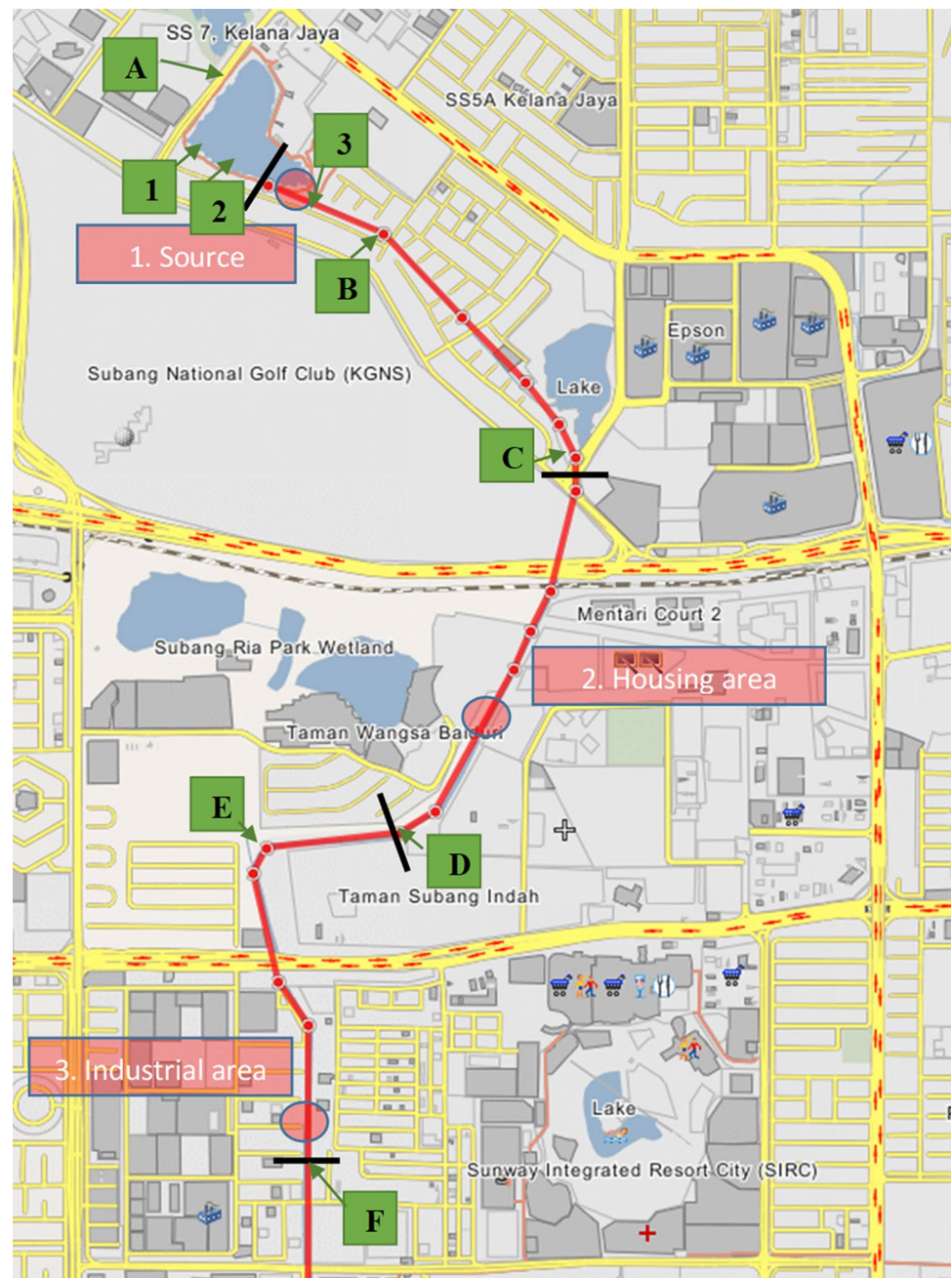
a commercial area and a recreational park. The samples were collected from the two inlets into the lake (Site 1, 2), the lake itself (Site A), and the outlet of the lake (Site 3) where the water from the lake flows downstream along the Bandar Sunway river before joining the Klang River. The first inlet into the lake (Site 1) receives wastewater from commercial areas, which flows into the lake; whereas, the second inlet (Site 2) carries wastewater originating from residential areas and a nearby golf course. Sampling point B (Site B) was located 300 m downstream from Kelana Jaya Lake, surrounded by residential area and the Subang National Golf Club. Site C was from a downstream point located between industrial and commercial areas. Sites D, E and F represent downstream rivers flowing through the Subang Ria Wetland Park and industrial area, residential area, and a construction site, respectively. The specific information of the sample sites is tabulated in Table 1. These nine sites were selected as they are accessible and the surrounding areas have different development and land use activities, reflecting the degree of anthropogenic influence to the water systems.

Sample collection and analytical methods

The water samples from the above-mentioned sites were collected twice, between May and June 2018. For each collection site, triplicates were collected in 1-L plastic bottles. The plastic bottles were washed and cleaned with detergent prior to collection of water samples. Water samples were then processed in the laboratory in which 150 mL were filtered using a glass filter funnel connected to a vacuum pump. The water samples were then measured for physical, chemical, and biological assessment according to the United States Environmental Protection Agency (USEPA) (US EPA 2018). The physicochemical profiles of the water samples collected were also determined based on pH (digital probe sensor, HACH SensION + PH31), total nitrogen (TN) content (HACH method 10071, HACH DR-6000 UV–Vis), total phosphorus (TP) content (HACH method 8190, HACH DR-6000 UV–Vis), chemical oxygen demand

(COD) (HACH method 8000, HACH DR-6000 UV–Vis), biochemical oxygen demand (BOD) (YSI-5000 DO instrument), and total suspended solids (TSS) (HACH method 8006, HACH DR-6000 UV–Vis). Readings for each parameter were recorded in triplicates. The Most Probable Number (MPN) technique was performed to determine the total coliform count in the water samples (Bartram and Pedley 1996). Samples were prepared in triplicates and coliform count were determined. The results were then compared with the MPN index (Malathy et al. 2017).

Microbial community analysis

The collected water samples from all nine sites were subjected to microbial community analysis. For this purpose, filtration was carried out by placing 0.22-µm membrane filters (Merck Millipore) in the filter funnel and filtering through 150 mL of the water samples. Once water samples were filtered, the membrane filter was collected and transferred into a centrifuge tube containing 30-mL absolute ethanol (C_2H_5OH), while the filtrate was discarded. After an overnight storage, the membrane filter was then transferred into a 2-mL PowerWater bead tube (Qiagen DNeasy PowerWater kit) containing 567 µL of ATL tissue lysis buffer and 63 µL of Proteinase K. DNA extraction from the membrane filter was carried out using Qiagen DNeasy Blood and Tissue kit (Qiagen, Germany) as per manufacturer's recommendation. The procedures were conducted using sterilized apparatus, either via autoclaving or UV rays, and in aseptic conditions where relevant.

Library preparation and Illumina MiSeq sequencing

Approximately 10 ng of DNA (measured with Nanodrop 2000 spectrophotometer) was used for PCR amplification and barcode incorporation as described previously (Watts et al. 2017). The hypervariable V3–V4 region of the 16S rRNA gene was amplified using forward primer and reverse primer as described by Klindworth et al. (2013). Amplification of

Table 1 GPS coordinates and description of the nine sampling sites downstream of Kelana Jaya lake along the Bandar Sunway river

Site	Area	GPS coordinate
1	Kelana Jaya Lake Inlet 1 (commercial area, shops, eateries)	N03.09405° E101.59554°
2	Kelana Jaya Lake Inlet 2 (residential area, golf course)	N03.09339° E101.59637°
A	Kelana Jaya Lake	N03.09647° E101.59753°
3	Kelana Jaya Lake outlet	N03.09296° E101.59907°
B	River near housing area and Subang National Golf Club	N03.09187° E101.60129°
C	River between Epson and Mentari Court (commercial and industrial area)	N03.08515° E101.60588°
D	River between Subang Ria Park Wetland and Industrial Area	N03.07637° E101.60093°
E	River channel between Taman Wangsa Baiduri and Taman Subang Indah (residential area)	N03.07876° E101.59735°
F	River near Sunway Construction stockyard (construction site)	N03.06774° E101.59885°

the hypervariable V9 region of the 18S rRNA gene was also carried out using the 1391f forward primer and EukBr reverse primer as described by Amaral-Zettler et al. (2009) and Stoeck et al. (2010). Primers were obtained from Integrated DNA Technologies Inc. Both sets of primers were incorporated with Illumina Nextera 5′ overhang adapter sequence. Successful PCR reaction was assessed using gel electrophoresis on 1% agarose gel, indicated by the presence of a single amplicon band for each sample and a positive control, and absence of that band in a negative control (deionised water). The constructed library of DNA samples were sent for 2×250 bp Illumina MiSeq sequencing platform (Illumina, San Diego, CA, USA) in Monash University Malaysia Genomics Facility. Sequence data can be accessed from NCBI database under BioProject PRJNA624028. The library preparation and sequencing procedures were carried out using sterilized apparatus, either via autoclaving or UV rays, and in aseptic conditions where relevant.

Bioinformatics analyses

The MiSeq sequencing data for 16S and 18S rRNA from all nine samples were pre-processed by trimming primer sequences using Cutadapt (Martin 2011) and merging paired-end reads using USearch (Edgar 2010) as described by Md Zoqratt et al. (2018). Sequence dereplication and denoising were done on DADA2 software using default parameters (Callahan et al. 2016). This is followed with taxonomic classification (using q2-classifier trained on SILVA 132 database), feature table normalization, alpha and beta diversity analyses, and principal coordinate analyses (PCoA), all which were performed with QIIME2 (Prusse et al. 2007; Bolyen et al. 2018). In this study, QIIME2 analysis for both 16S and 18S were performed in a similar manner with slight modifications for 16S analysis. SILVA 132 reference sequences were used to train the classifier, while the 18S reference sequences were left untrimmed. The chloroplast and mitochondrial contaminant sequences were filtered using QIIME2 (Bolyen et al. 2018).

The alpha diversity analyses based on Observed Features and Shannon indices were performed on the normalized feature table. For beta diversity, both 16S rRNA and 18S rRNA scatter plots were evaluated based on phylogenetic-based Unweighted UniFrac (Lozupone et al. 2010) and non-phylogenetic-based Jaccard diversity metrics (Baselga and Orme 2012).

Results and discussion

Physical, chemical and biological water quality indicators

A summary of the physical and chemical properties of the water samples, taken during May–June 2018, along with their statistical parameters are tabulated in Table 2. As it can be seen, the pH level was between 7 and 8 for all samples. The TN content was between 0.20 and 15.63 mg/L, with major contributors from samples from sites 2 (9.72 mg/L) and E (15.63 mg/L). Similarly, the TP content for samples from site 2 was found to be relatively higher (3 mg/L). This may be attributed to the nearby golf course area that utilizes fertilizer for their turf grass, resulting in the high phosphorus content. The 5-day biochemical oxygen demand (BOD_5) was higher for samples from site 1 (17.34 mg/L); whereas, COD levels were moderate throughout all sampling sites (mean of 42.32 mg/L), with the exception for samples from site F (255.67 mg/L). The TSS among all sampling points were considered low (6.33–69.67 mg/L). To better understand the impact of land use on water quality in the study site, an investigation was conducted to identify the land-use composition in the contributing catchment of each sampling point. Table 3 summarizes the results of this investigation and links them to the observed physical and chemical characteristics of water samples. As it can be seen, only a high percentage of industrial land use (site F, moderate for physical pollution) may cause an increase in physical pollution in water, while chemical pollution is mainly linked to high commercial or residential land uses (sites 1, 2, D and F, with 100, 26, 100 and 70% of commercial and residential land use, respectively) (Table 3).

All collected water samples produced MPN values exceeding the permissible safe levels, with MPN index values > 1600 (Table 4). This indicated that each water sample contained an estimated > 1600 coliforms per 100 mL. The water quality for all sites was reported as unsatisfactory with the high presence of coliforms. This observation

Table 2 Summary of statistical parameters for the physical and chemical water quality data

Parameters	Min	Max	Average	SD
pH[a]	7.03	7.85	7.42	0.19
TN (mg/L)	0.20	15.63	6.46	4.39
TP (mg/L)	0.64	5.62	2.11	1.25
TSS (mg/L)	6.33	69.67	18.52	10.82
COD (mg/L)	14.00	255.67	42.32	40.48
BOD_5	1.26	18.83	7.59	4.98

[a]No unit

Table 3 Land use and pollutant condition at each sampling point

Sampling point	Land use	Physical pollution	Chemical pollution	Catchment area (km^2)
1	70% commercial 30% residential	Low	High COD High BOD5	0.41
2	14% commercial 12% residential 74% golf course	Low	High TN High TP	0.54
3	Lake outlet	Low	Moderate	–
A	Lake	Very high	Very high	–
B	10% construction 20% commercial 25% residential 45% open area	Low	Moderate	0.11
C	10% construction 35% residential 55% golf course	Low	Moderate	1.10
D	100% residential	Low	High overall	0.21
E	20% commercial 80% residential	Low	Moderate	–
F	60% commercial 10% residential 30% industrial	Moderate	High overall	1.05

Table 4 Summary result of total coliform tests performed from May to June 2018 for the nine sites

	Month	MPN index per 100 mL	Grade	
Coliform test	May	> 1600	Class IV	Unsatisfactory
	June	> 1600	Class IV	Unsatisfactory

was consistent for all water samples tested throughout the sampling intervals from May to June 2018. This strongly indicated that water samples from the various sites in the urban water system in Bandar Sunway were of poor quality and may contain pathogens and fecal matter.

Relative abundance of microbial communities

The Amplicon Sequence Variant (ASV) for 16S rRNA microbiome analysis was constructed using an estimate of 26,249 merged mean reads per replicate. Rarefaction curve analysis indicated that approximately 10,000 reads were sufficient to capture the bacterial community structure (Fig. S1). Using the merged reads, an average of 428 ASVs per replicate for the nine samples were generated (Table 5). The mean numbers of ASVs for samples from sites 1, 2, A, 3, B, C, D, E and F were 415 ± 79, 526 ± 97, 198 ± 145, 620 ± 241, 416 ± 54, 385 ± 77, 505 ± 136, 422 ± 215, and 366 ± 169, respectively, with samples from sites 1 (415 ± 79), 2 (526 ± 97), 3 (620 ± 241), B (416 ± 54), D (505 ± 136) and E (422 ± 215) having higher ASV counts.

Based on the standard deviations, the ASV counts for all sampling points did not vary much during the different collection period except for samples from sites 3 and E. There were several ASV counts that were outliers (unusually high or low) when compared among the sampling time. For example, the outlier for samples from site A was during the second sampling (ASV 365), for samples from sites 3 (ASV 857) and E (670), the outliers were detected during the third sampling, and samples from site F (ASV 171) have an outlier during the first sampling (Table 5).

For the 18S rRNA microbiome analysis, the average merged reads were 35,924 that were used to generate an average of 438 ASVs per replicate. Similarly, the rarefaction curve analysis indicated that approximately 10,000 reads were sufficient to capture the eukaryotic community structure of Bandar Sunway freshwater environment (Fig. S2). The ASVs constructed for each site were as follows; 303 ± 256 (Site 1), 452 ± 181 (Site 2), 358 ± 77 (Site A), 571 ± 351 (Site 3), 333 ± 144 (Site B), 457 ± 87 (Site C), 803 ± 559 (Site D), 399 ± 264 (Site E), and 268 ± 108 (Site F) (Table 5). The lowest and highest ASVs recorded were for samples for sites F and D, respectively. For 18S rRNA-based microbiome analysis, several sampling sites had high variability among different sampling periods such as for sites 1, 3, D, and E with standard deviation values 256, 351, 559 and 264, respectively. There were more outliers in the ASV counts detected when compared to 16S rRNA. These include samples in site A (ASV 599) in the third sampling, site 2 (ASV 656) for the second sampling, and

Table 5 Illumina MiSeq sequencing data obtained for each replicate of the three sampling periods for the nine sites downstream of Kelana Jaya lake along the Bandar Sunway river

Sample	16S rRNA			18S rRNA			
	Reads	Merged reads	ASVs	Reads		Merged reads	ASVs
1	34,594	29,374	376	31,196		25,119	147
1	27,225	22,825	363	45,217		42,196	163
1	37,503	32,108	506	36,979		30,315	599
2	36,052	27,311	609	38,507		34,937	310
2	36,110	28,582	419	40,241		36,302	656
2	38,167	29,166	550	29,733		24,609	389
A	37,051	31,458	124	27,866		22,301	273
A	26,375	20,045	365	35,658		32,030	375
A	24,509	19,189	105	56,055		49,694	425
3	33,541	28,114	628	36,390		33,086	907
3	27,879	21,989	375	35,492		31,509	206
3	86,860	68,040	857	26,825		20,922	601
B	30,354	23,045	404	36,184		32,651	492
B	30,952	24,098	369	50,898		45,925	297
B	27,441	22,977	475	39,009		35,255	212
C	29,239	20,848	313	53,815		49,236	554
C	35,002	26,373	467	51,169		47,544	385
C	30,663	23,977	376	121,349		115,373	432
D	27,468	23,645	561	36,780		32,299	1449
D	33,617	26,204	605	38,407		35,655	476
D	28,958	22,418	350	25,624		18,879	485
E	25,872	20,363	304	60,856		57,331	702
E	29,390	22,742	292	31,467		25,924	214
E	24,544	19,228	670	25,197		19,758	282
F	29,730	25,784	171	34,668		27,241	148
F	31,848	27,058	475	18,064		16,170	359
F	28,328	21,773	451	34,568		27,697	296
Total	889,272	708,734	11,560		109,821	969,958	11,834

sites 3 (ASV 907), D (ASV 1449), and E (ASV 702) for the first sampling (Table 5).

Prokaryotic communities of the urban water systems (16S rRNA analysis)

There were four predominant bacterial phyla present in all samples from the sites, namely the Proteobacteria, Bacteroidetes, Actinobacteria, and Firmicutes (Fig. 2, Table S1). Proteobacteria was the most dominant phylum, present in reasonably high abundance in all sampled sites. The abundance of Proteobacteria was most evident in samples from sites F (river near construction site), D (river near wetland and industrial area) and 1 (Lake inlet from commercial area), and of a lesser degree in site A (Kelana Jaya Lake). The observations were suggestive that downstream river sites with neighboring commercial, industrial or construction activities contributed to the abundance of Proteobacteria. For Bacteroidetes, their relative abundance was generally higher in samples from sites C (river from industrial and commercial area), 2 (lake inlet 2 from residential area and golf course) and 3 (lake outlet), and lesser from sites 1 (inlet from commercial area), A (Kelana Jaya Lake) and F (river near construction site). The relative abundance of Actinobacteria was similar for samples in sites C, 3, B and 2, but the least in site 1 (lake inlet from commercial area). For Cyanobacteria, the relative abundance was highest in samples from site A (Kelana Jaya Lake), while samples from site 1 (Lake inlet 1 from commercial area), 2 (Lake inlet 2 from residential and golf course), D (site nearby wetland and industrial area) and F (nearby construction site) have the least abundance. This suggested that Cyanobacteria were more abundant in water systems that have slower flow rates (lake) when compared to inlet, outlet and downstream points. For samples from sites 1 and 2 (inlet sources to the Kelana Jaya Lake), where flow rate is greater, Cyanobacteria were found to be the least abundant. The other phyla such as Epsilonbacteraeota, Verrucomicrobia, Planctomycetes, Chloroflexi, and Patescibacteria, were present in the various samples, albeit with lower relative abundance (Fig. 2).
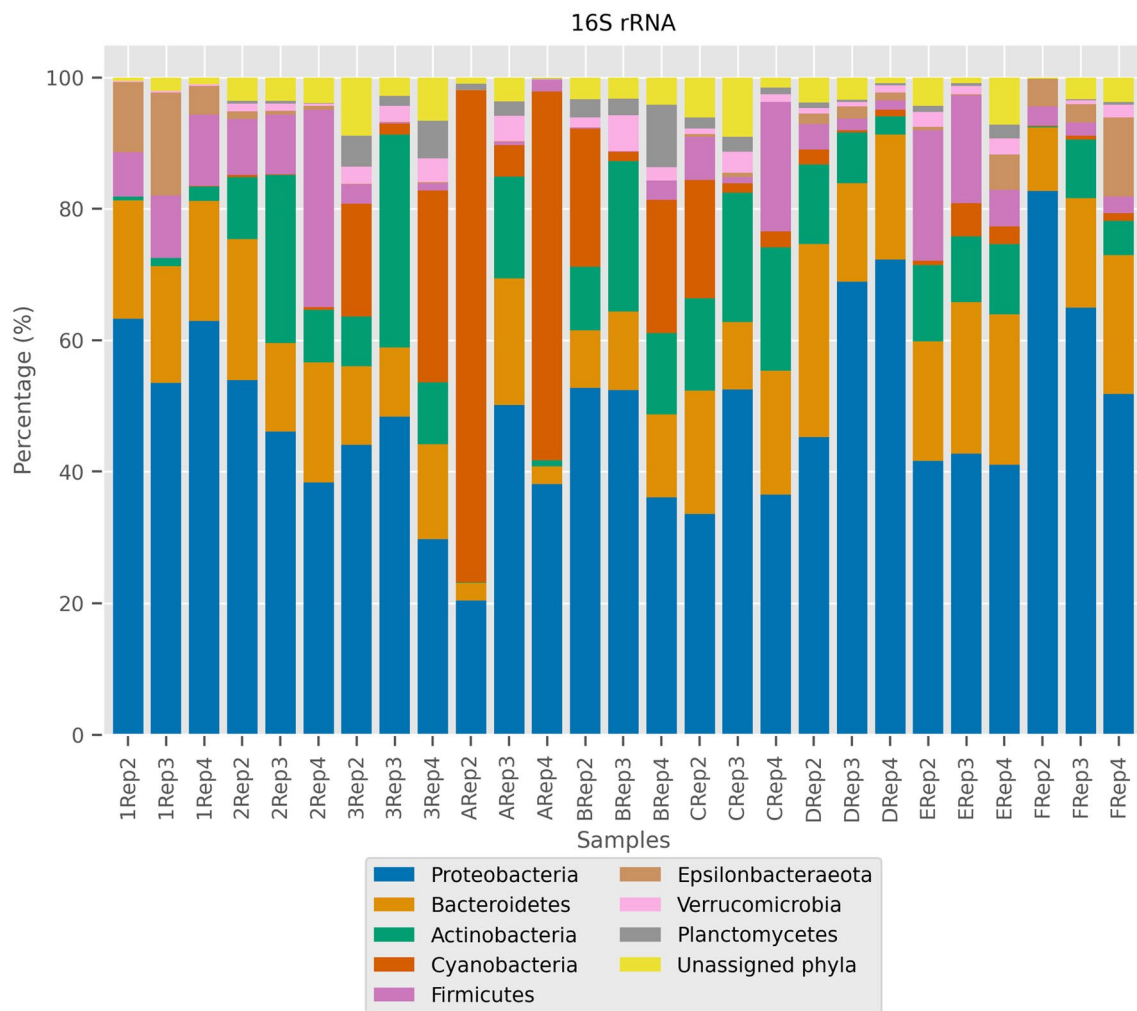
Fig. 2 Relative abundance of the microbial communities in Bandar Sunway lake and rivers grouped as 16S rRNA. Each taxonomic unit and representative color are listed in the figure legend on the top right

The microbiome analysis has, therefore, revealed that the 16S prokaryotic communities in the urban water systems of Bandar Sunway consisted primarily of Proteobacteria, Bacteroidetes, Actinobacteria, Cyanobacteria, and Firmicutes. Other groups such as Epsilonbacteraeota, Verrucomicrobia, Planctomycetes, Chloroflexi and Patescibacteria were detected only in certain sites such as in samples from sites 1 and F for Epsilonbacteraeota; sites A, 3, B, C and E for Planctomycetes; and sites 3, B, C, D and E for Chloroflexi.

The prevalence of bacterial genera in the 16S rRNA communities was further examined. Upon analysis, the prevalent bacterial genera were identified as *Flavobacterium* and *Pseudomonas* (more than 90% prevalence). *Flavobacterium* and *Pseudomonas* are members of the phylum Bacteroidetes and Proteobacteria, respectively (Fig. 3). Other relatively dominant genera are close to 90% prevalence include *Novosphingobium, Aeromonas,* and *Acinetobacter*, which all belong to the phylum Proteobacteria.

As observed in this study, the dynamics of the bacterial communities in the urban water systems differed along the sampling sites but were predominated by Proteobacteria. This did not come as a surprise as Proteobacteria is a large phylum that comprises of the coliforms (*Escherichia, Salmonella, Vibrio, Shigella, Pseudomonas*), food pathogens (*Helicobacter, Campylobacter*), and natural soil and environmental bacteria (*Rhizobium, Nitrosomonas, Agrobacterium, Desulfovibrio, Geobacter*, and *Acidithiobacillus*). Results here were consistent with the MPN results observed, confirming that total coliform count exceeding the safety level suggests the presence of fecal matter and possible pathogens. In this study, the ubiquity and abundance of Proteobacteria are evident in all samples, especially from samples in areas with high human activities (i.e., construction site, Subang wetland Park and industrial area, lake inlet from commercial areas). This suggested that urban water systems harbour both environmental bacteria as well as pathogens (coliforms,
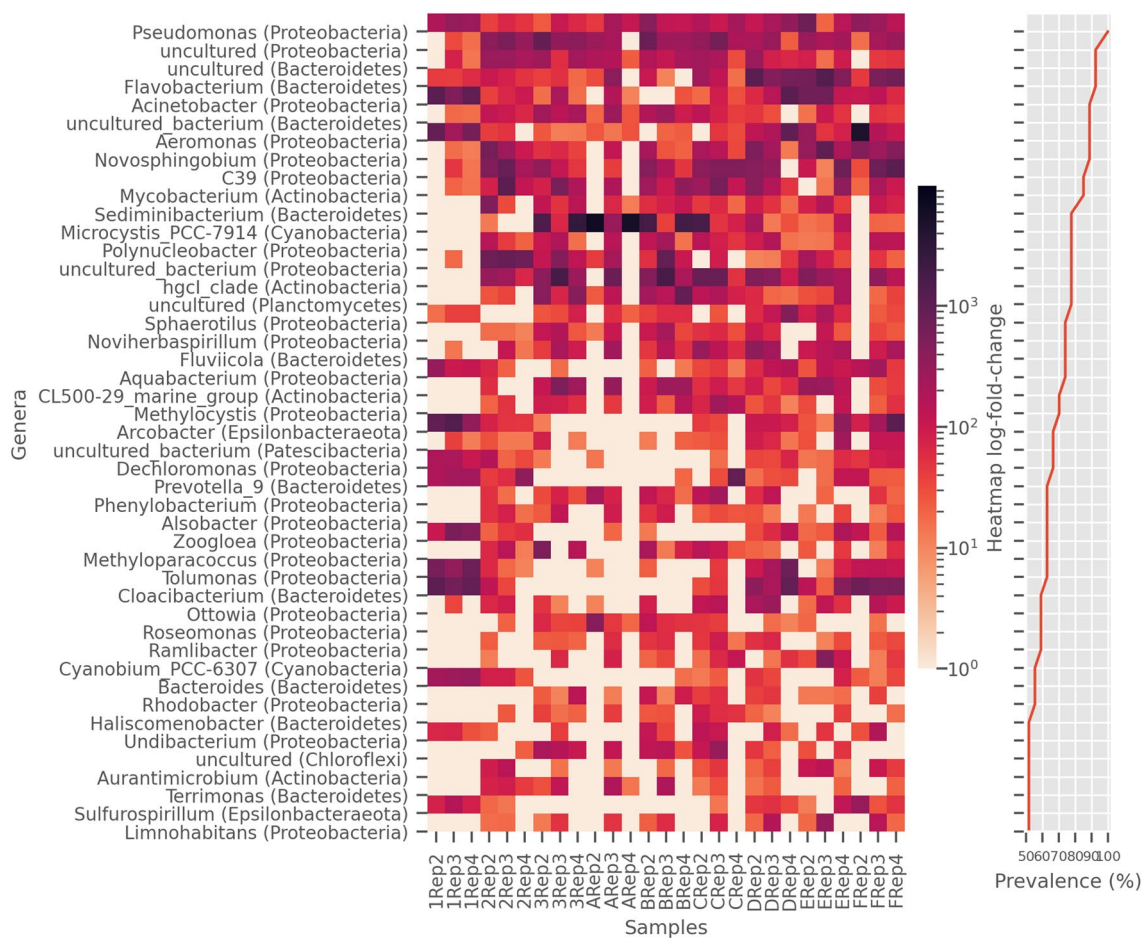
Fig. 3 Prevalence of different bacterial genera (minimum of 50%) across all samples. The phylum group of each bacterial genera are indicated in brackets

food pathogens), presumably from the input of untreated wastewater or polluted water into the natural water system. This observation is again aligned to the MPN results derived from the tested water samples, validating that the likelihood of pathogens in the water systems is high, as a consequence of the inputs from various sites. In addition, the prevalence of Proteobacteria is highly suggestive of the impact of human activities on the aquatic microbiome as Proteobacteria was notably less abundant in water samples from sites less exposed to human activities (i.e., lake, residential area). The negative impact of human activities and urbanization on the lakes and river sites was further evidenced by the prevalence of *Pseudomonas, Aeromonas, Aquabacterium,* and *Sphaerotilus* in the water samples. Their detection (from construction sites, eateries in commercial areas, industrial sites) indicated the possible presence of pathogens (Noor et al. 2017) or sludge (Suzuki et al. 2002), and especially with *Aeromonas* and *Aquabacterium* as causal agents of gastroenteritis (Janda and Abbott 1998; Dong et al. 2011). It is also possible that pathogens may originate from fecal matter

of animals (e.g., aquatic animals, terrestrial animals, birds) as animals are present in these areas (lakes, industrial sites, residential areas). Nevertheless, these are proposed theories as there are no published records on animal populations in sampled sites.

Other phyla indicative of the presence of possible pathogens include the detection of Bacteroidetes with high prevalence of *Flavobacteria, Bacteroides, Fluvicola, Cloacibacterium,* and *Sediminibacterium. Flavobacterium* is the most prevalent genus and is commonly present in soils and water. It is also a possible pathogen of freshwater fish diseases (Jin et al. 2018). The *Bacteroides* maintain beneficial relationships with the host when retained in the gut, but when in the environment, they can cause significant pathology, including bacteraemia and abscess formation (Zafar and Saier 2018). The Actinobacteria is another phylum, consisting of natural decomposers and nitrogen fixers, but some are pathogenic such as *Mycobacterium tuberculosis* (Ventura et al. 2007). The prevalence of *Mycobacterium* in the samples highlights the possible presence of pathogenic microbes in the water

systems. Firmicutes is another abundant phylum and this may be attributed to their resistance to desiccation (Galperin 2013). As expected, *Cyanobacteria* was the most abundant in the lake itself (site A), thriving in the lake where water is stagnant. This is commonly observed in urbanized lakes as there are higher levels of sediments or nutrients (Bai et al. 2014; Saleem et al. 2018; Lau et al. 2019). Continuous detection of *Cyanobacteria* at sites downstream from the lake, but in low abundance probably implied that the population of cyanobacteria decreases from enclosed water bodies to flowing water systems. *Microcystis*, the most abundant *Cyanobacteria* found in this study, grows well at higher temperatures and may out-compete diatoms and green algae, possibly leading to increased dominance and potential toxic blooming in aquatic environments (Almanza et al. 2016). In this study, Cyanobacteria were detected, but results did not discriminate if these were symbionts, free-living, or toxic cyanobacteria.

To summarize, the bacterial communities in urban waters systems were primarily from the phyla Proteobacteria, Bacteroidetes, Actinobacteria, and Cyanobacteria, which were also typically observed in various water sources: rivers, lakes, and streams (Staley et al. 2013; Medeiros et al. 2016; Noor et al. 2017; Cai et al. 2018; Lau et al. 2019; Mittal et al. 2019). These phyla are regularly clustered as ubiquitous bacterial communities in river and lake samples, and the origin of some of the bacteria from gut and feces may indicate contamination from human and urbanization activities. The poor water quality and possible contamination with feces is further strengthened with the detection of *Verrucomicrobia*, which is also present in human feces (Cai et al. 2018). In short, the association of multiple phyla with microbiota from both human and farm animal feces is a concern (Lee et al. 2011; Moon et al. 2018). Their higher abundance in downstream sites may imply that the increase in abundance of these groups result from urbanization activities along the river. These are evidence that downstream river points with neighboring commercial, industrial or construction activities, contribute to the abundance of Proteobacteria. As such, it is suggestive that the prevalence of Proteobacteria in the Bandar Sunway water system may be attributed to human and industrial activities. Although the prevalence of Proteobacteria is generally understood to expand and persist in various ecological niches (Zavarzin et al. 1991), results here are suggestive that Proteobacteria may form the core bacteria community which are consistent and prevalent throughout all sites in the water systems evaluated in Bandar Sunway. Observations here further strengthen the global richness of Proteobacteria as observed in other studies as well (Zavarzin et al. 1991; Louca et al. 2019). This finding contributes immensely to the understanding of aquatic microbiome, particularly from urban water systems. This may be one of the first few studies in Malaysia to document the microbiome

analysis from urban aquatic system in Malaysia as previous studies were more on freshwater lakes such as from the photic and aphotic zones of Temenggor Lake (Lau et al. 2019) and aquaculture farms in Temenggor Lake (Noor et al. 2017).

Eukaryotic communities of the urban water systems (18S rRNA analysis)

The eukaryotic communities were primarily of the phyla SAR and Opisthokonta (Fig. 4, Table S2). All the three SAR clades can be found in all the samples (Fig. S3). The relative abundance of SAR was evident in samples from sites E, C, F, 2, while lower abundance was found in sites 1 and D. This observation noted that most water samples from the downstream rivers (sites E, C, F) have more members from SAR, with the exception of site 2 (inlet from residential and golf course). The other dominant phylum, Opisthokonta, had higher relative abundance in samples from sites 3, B, C, and A, while the lowest relative abundance was from site 1. This observation suggested that Opisthokonta was most likely introduced from inlet 2 (site 2) into the Kelana Jaya Lake (site A), and gradually flowed downstream (sites 3, B, C). Other eukaryotic communities found in the water systems include Cryptophyceae, Excavata, Amoebozoa, Centrohelida, and Archaeplastida. Of these, Excavata was observed to be present in abundance in samples from sites B and 3, but extremely low in abundance in sites 1 and F. Amoebozoa was recovered from site A (Kelana Jaya Lake) but were negligible in other samples. It was also observed that a large proportion of the 18S rRNA amplicon sequencing reads were unassigned (Fig. 4, Table S2). This occurred in all nine samples where the unassigned group was relatively dominant in sites 1, D, F, 2 and E. Overall, the eukaryotic microbiome analysis has revealed that sites A (Kelana Jaya Lake) and 3 (outlet) were most diverse whereas site 1 (inlet from commercial area) was the least diverse.

The eukaryotic community analysis revealed that the largest taxa is the unassigned 18S rRNA, from commercial, industrial and construction sites. This is highly suggestive that water samples from these sites may have been contaminated with wastes or effluents of eukaryotic organisms or their debris/remnants. The lack of references in the database led to non-identification and the proposal that the taxa may not originate from microbes. There is therefore a need for the 18S rRNA reference database for microbiome analysis to be updated through future experiments prioritizing these communities (Santamaria et al. 2012). The next two prevalent taxa were SAR and Opisthokonta. The abundance of the SAR group may be attributed to the presence of algae, diatoms, planktons, dinoflagellates, and ciliates (Ortiz-Álvarez et al. 2018). Their high abundance in sites such as residential areas, golf course, commercial and construction sites,

Fig. 4 Relative abundance of the microbial communities in Bandar Sunway lake and rivers grouped as 18S rRNA. Each taxonomic unit and representative color are listed in the figure legend on the top right
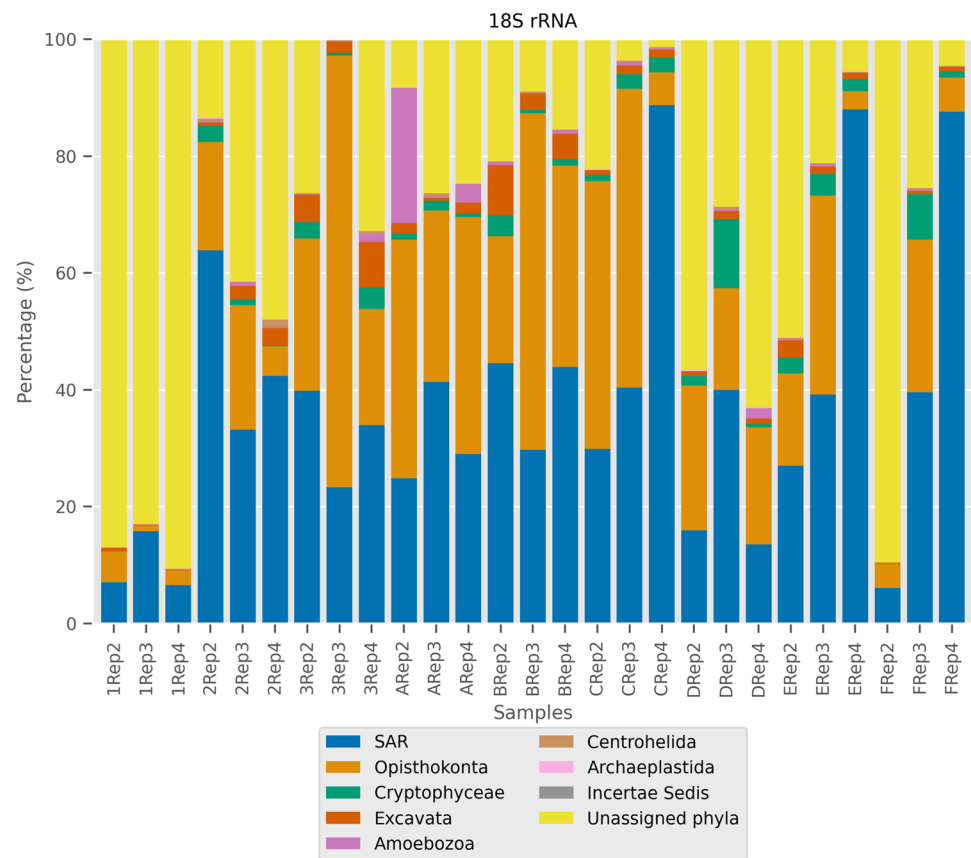

did not reveal any particular trend, except for the shared presence of human activities. For Opisthokonta, their relative abundance in lake outlet, lake, housing and golf course, and commercial area, may be attributed to the fact that both fungi-like Holomycota and animal-like Holozoa were profiled (Torruella et al. 2012). The abundance of Opisthokonta may be attributed to the various aquatic organisms such as chytrids and microsporidia, which may be naturally present in the water systems (Del Campo et al. 2015). Other eukaryotic phyla such as Cryptophyceae (e.g., algae *Cryptomonas, Rhodomonas*), Excavata (e.g., *Euglena*), Amoebozoa (e.g., Amoeboa), Centrohelida and Archaeplastida were also present in the water systems. They are typical aquatic eukaryotes, observed to be present in freshwater systems (Burki 2014; Simon et al. 2015; Wirth et al. 2019). This suggests that the urban water systems such as in Bandar Sunway have eukaryotic communities similar to freshwater systems, which comprise algae, aquatic fungi, protozoans, amoebozoans, planktons and aquatic parasites.

Diversity analyses of microbial communities in Bandar Sunway water systems

The diversity analyses of both the prokaryotic and eukaryotic microbiome were performed based on alpha and beta diversity. The alpha diversity analyses, based on Observed Features and Shannon indices, revealed no statistical differences observed for both communities. The alpha diversity for 16S rRNA showed p values of $p = 0.269$ and $p = 0.316$ for Observed Features and Shannon Index, respectively. For 18S rRNA analysis, $p = 0.450$ and $p = 0.507$ was recorded for Observed Features and Shannon Index, respectively (Fig. 5, Table S3). This was observed for both prokaryotic (16S rRNA) and eukaryotic (18S rRNA) communities. For beta diversity, the phylogenetic-based (unweighted UniFrac) and non-phylogenetic-based (Jaccard) methods were employed. Both 16S rRNA and 18 s rRNA scatter plots for Unweighted UniFrac and Jaccard revealed no clustering among samples (Fig. 6). As such, there were no significant differences in the species richness between different sampling sites from the various water systems, for both prokaryotic and eukaryotic communities (Fig. 6). This indicated that species richness of samples from different sites were similar to each other.

The alpha diversity analysis for prokaryotes confirmed that each site at different collection periods had similar bacterial diversity. This is presumably due to the fact that in urban settings, the driving pressure for change is rather subtle, because there are no monsoon seasons or drastic changes in weather. The bacterial communities were revealed to be relatively similar among all sample sites as no significant
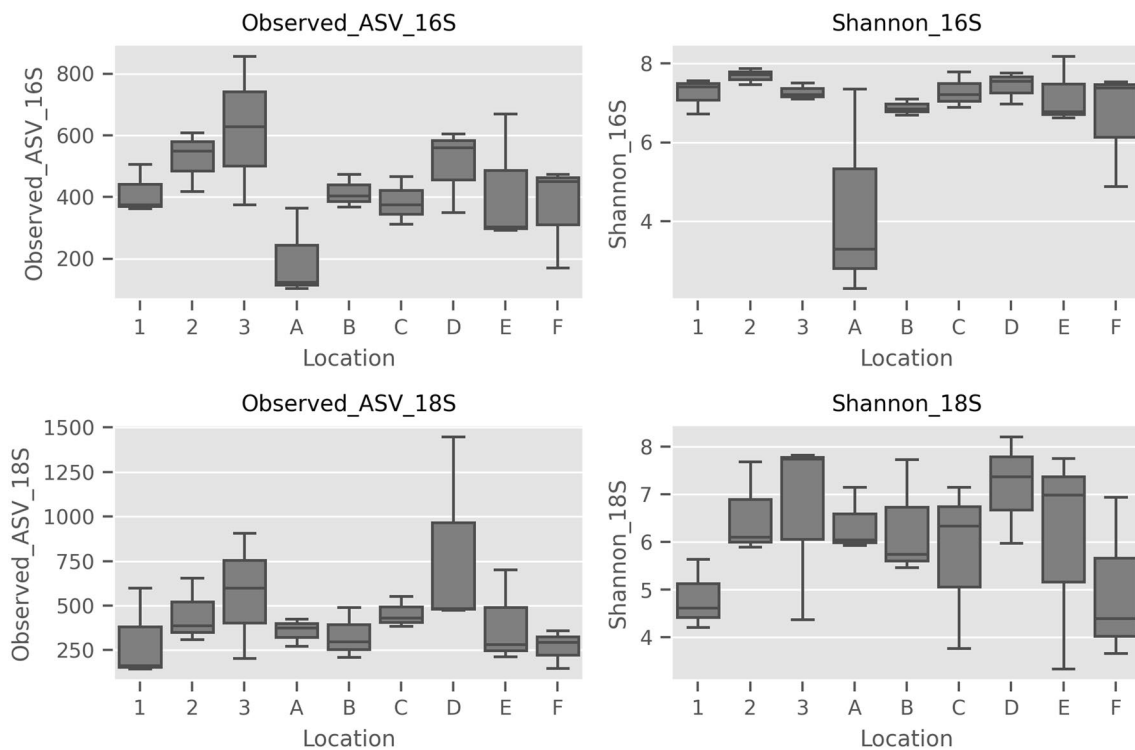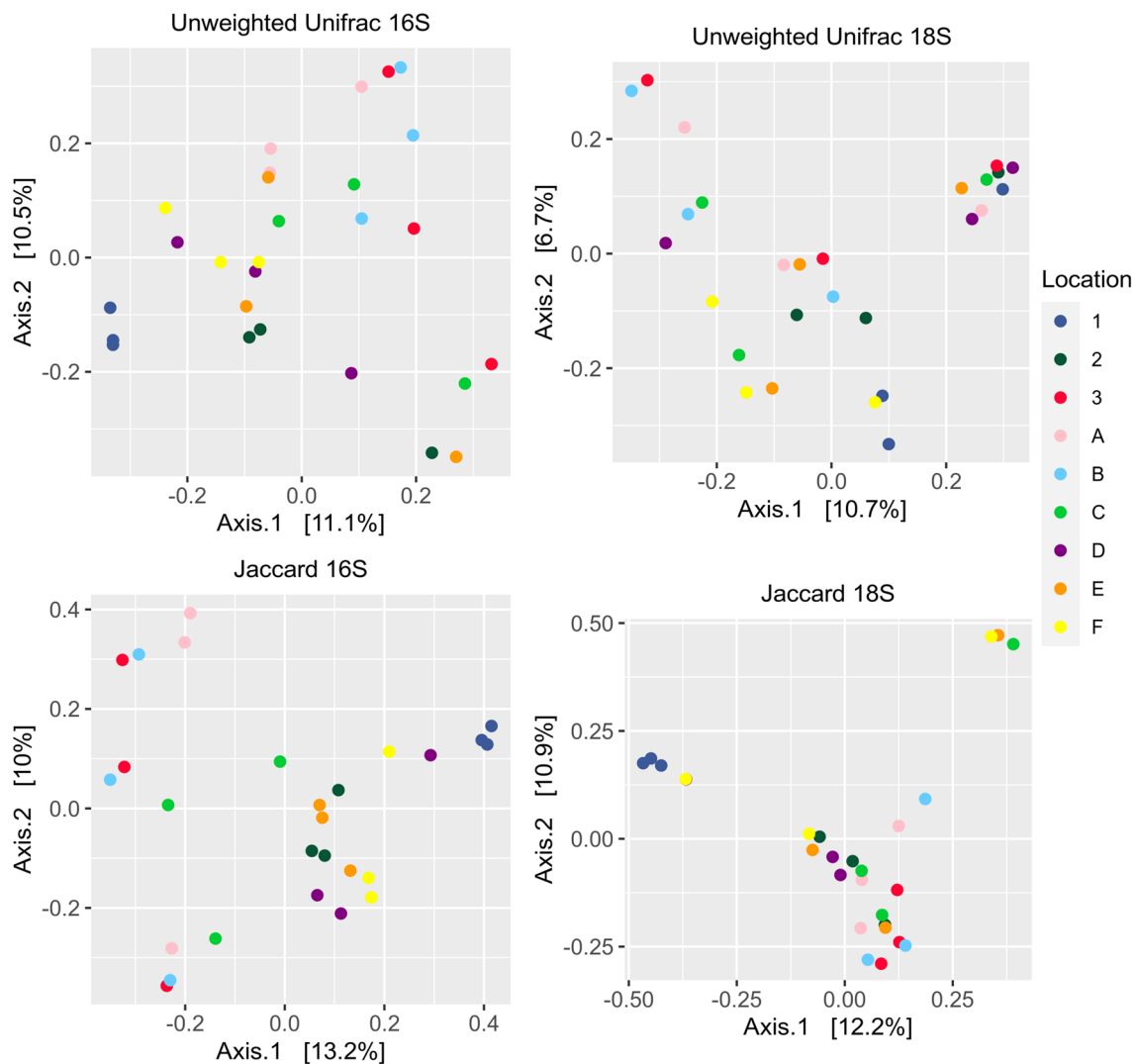
Fig. 5 Alpha diversity boxplots for 16S rRNA and 18S rRNA

differences were observed in the beta diversity analysis. Despite the lack of differences in the richness of species between the sites, the relative abundance of the same phyla, i.e., Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes, were observed and includes some other unique phyla (i.e., Planctomycetes and Chlorofexi). Similarly, the alpha diversity analysis of the eukaryotic community revealed that the samples collected at different sampling periods did not influence the overall eukaryotic diversity across all locations. Notably, earlier ASV counts indicated high variability between sampling periods for many sites. The beta diversity analysis also confirmed that across the different sampling sites, the species richness of the eukaryotic community remains the same. It, therefore, appears that eukaryotic communities in urban water systems are relatively consistent and the urbanization process did not cause any significant changes to the communities.

In short, this study has revealed the microbial communities in an urban (Bandar Sunway) water system, and the possible impact of human activities on the relative abundance of the various phyla. Combined with the land use and physicochemical parameters obtained, it is clear that neighboring commercial areas, residential, industrial areas and construction sites have somewhat a degree of influence on the quality of water that flows through the water system. The Kelana Jaya lake is a recreational area with human activities. But the discharge from inlet 1 and 2 carries possible wastes

that may affect the community structure in the lake. Inlet 1 and 2, with 26–100% of land use in commercial and residential purpose, may have channeled polluted water into the lake as high BOD5, COD, TN and TP were detected in the water samples. With high levels of TN and TP, it may have caused a bloom in the aquatic assemblages (Cyanobacteria, SAR, Opisthokonta and Amoebozoa). The contributing sedimentation of nutrients or other wastes has been known to contribute to eutrophic condition in the lakes (Menció and Mas-Pla 2010). Therefore, the prevalence of Cyanobacteria and Archaeplastida in the lake, which has very high levels of physical and chemical pollution, exemplifies the occurrence of eutrophication. Clearly, the land use and activities surrounding inlets 1 and 2 may have subsequently influenced the physicochemical characteristics and microbial communities in the lake as well (site A). As the water from the lake flows out from the outlet (site 3) to the river, the same community prevails as the water source carries with it the core bacterial communities (Proteobacteria, Bacteroidetes, Firmicutes). These core communities are consistent in their prevalence in water samples from sites with typical land use as residential or commercial purpose (sites B, C, D, E, F), suggesting that discharge from home, workplaces, shops, eateries, and industries, are often high in Proteobacteria, Bacteroidetes as well as SAR. This is often accompanied with moderate to high levels of chemical pollution detected in the water samples from the sites. The relative abundance,

Fig. 6 Beta diversity plots for 16S rRNA and 18S rRNA

however, may change due to gradual changes with input from neighboring sources, with clear evidence that high chemical pollution (TN, TP, COD, BOD5) would usually lead to higher Proteobacteria, Bacteroidetes, and SAR. Overall, the microbial community analysis of this study has highlighted the presence of both natural aquatic microbes as well as pathogenic contaminants due to urbanization in Bandar Sunway water systems. The land use influences the pollution levels in each site, and this has been shown to subsequently impact the microbial communities in the water system.

Conclusion

The 16S and 18S rRNA-based amplicon sequencing approach has successfully characterized the microbial communities of the urban water system in Bandar Sunway,

Malaysia. The abundant prokaryotic communities are Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes; whereas, the abundant eukaryotic communities are SAR group and Opisthokonta. The alpha diversity analysis has confirmed that microbial diversity in the urban water system is not influenced by different sampling periods. The beta diversity analysis has also concluded that the overall microbial community from inlets to lakes and rivers are similar regardless of surrounding activities. The neighboring land use and activities however, influenced the relative abundance of the various phyla. These activities determine the level of pollution (physicochemical), which gradually impacted the abundance of typical aquatic microbial communities, by causing algae and phytoplankton blooms (Cyanobacteria, SAR), potentially harboring common bacteria due to high total coliform counts (from MPN and community analysis of Proteobacteria, Bacteroidetes), or extraordinary recoveries

of Unassigned 18S rRNA taxa. The presence or increased abundance of potential harmful pathogens and toxic bloom in the water systems corresponds to the water quality and the possible deterioration of water quality due to urbanization in Bandar Sunway.

Author contributions ASYT: Conceptualization, methodology, data organization and interpretation, manuscript writing. HST: Initial mapping of sampling sites, data analysis and interpretation, manuscript review. MZHMZ: Bioinformatics analysis, data visualization and tabulation, manuscript review. AAH: Methodology, project administration, sample collection, investigation, manuscript review. AT: Conceptualization, methodology, resources, funding acquisition, project administration, manuscript review. STK: Conceptualization, methodology, resources, funding, manuscript review.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest in the publication.

Research involving human participants and/or animals This research does not involve human participants and/or animals, therefore no human and/or animal ethical approval is required.

Informed consent All authors consent to the submission of the manuscript to the journal for consideration. Author contribution is listed.

References

Abia ALK, Alisoltani A, Keshri J, Ubomba-Jaswa E (2018) Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. Sci Total Environ 616–617:326–334

Almanza V, Parra O, De M, Bicudo CE, Baeza C, Beltran J, Figueroa R, Urrutia R (2016) Occurrence of toxic blooms of Microcystis aeruginosa in a central Chilean (36° Lat. S) urban lake. Revista Chilena de Historia Nat 89:8

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS ONE 4:e6372

Bai Y, Qi W, Liang J, Qu J (2014) Using high-throughput sequencing to assess the impacts of treated and untreated wastewater discharge on prokaryotic communities in an urban river. Appl Microbiol Biotechnol 98:1841–1851

Bartram J, Pedley S (1996) Microbiological analyses. In: Bartram J, Ballance R (eds) Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programmes. E & FN Spon for UNEP and WHO, London, pp 237–262

Baselga A, Orme CDL (2012) Betapart: an R package for the study of beta diversity. Methods Ecol Evol 3:808–812

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C et al (2018) QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints 6:e27295v2

Burki F (2014) The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb Perspect Biol 6:a016147–a016147

Cai X, Yao L, Sheng Q, Jiang L, Dahlgren RA, Wang T (2018) Properties of bacterial communities attached to artificial substrates in a hypereutrophic urban river. AMB Express 8:22

Callahan B, McMurdie P, Rosen M, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583

Del Campo J, Mallo D, Massana R, de Vargas C, Richards TA, Ruiz-Trillo I (2015) Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. Environ Microbiol 17:3195–3207

Dong Q, Brulc JM, Iovieno A, Bates B, Garoutte A, Miller D, Revanna KV, Gao X, Antonopoulos DA, Slepak VZ, Shestopalov VI (2011) Diversity of bacteria at healthy human conjunctiva. Invest Ophthalmol Vis Sci 52:5408–5413

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19):2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Galperin MY (2013) Genome diversity of spore-forming Firmicutes. Microbiol Spectr. https://doi.org/10.1128/microbiolspectrum.TBS-0015-2012

Ghai R, Rodríguez-Valera F, McMahon KD, Toyama D, Rinke R, de Oliveira CZT, Wagner Garcia J, Pellon de Miranda F, Henrique-Silva F (2011) Metagenomics of the water column in the pristine upper course of the Amazon River. PLoS ONE 6:e23785

Hafsi R, Ouerdachi L, Kriker AEO, Boutaghane H (2016) Assessment of urbanization/impervious effects on water quality in the urban river Annaba (Eastern Algeria) using physicochemical parameters. Water Sci Technol 74:2051–2059

Ininbergs K, Bergman B, Larsson J, Ekman M (2015) Microbial metagenomics in the Baltic Sea: recent advancements and prospects for environmental monitoring. Ambio 44:439–450

Janda JM, Abbott SL (1998) Evolving concepts regarding the genus *Aeromonas*: an expanding panorama of species, disease presentations, and unanswered questions. Clin Infect Dis 27:332–344

Jin D, Kong X, Cui B, Jin S, Xie Y, Wang X, Deng Y (2018) Bacterial communities and potential waterborne pathogens within the typical urban surface waters. Sci Rep 8:13368

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J 3:442–453

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1–e1

Lau NS, Zarkasi KZ, Md Sah ASR, Shu-Chien AC (2019) Diversity and coding potential of the microbiota in the photic and aphotic zones of tropical man-made lake with intensive aquaculture activities: a case study on Temengor Lake, Malaysia. Microb Ecol 78:20–32

Lee JE, Lee S, Sung J, Ko G (2011) Analysis of human and animal fecal microbiota for microbial source tracking. ISME J 5:362–365

Louca S, Mazel F, Doebeli PLW (2019) A census-based estimate of Earth's bacterial and archaeal diversity. Plos Biol 17(2):e3000106

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2010) UniFrac: an effective distance metric for microbial community comparison. ISME J 5(2):169–172

Malathy BR, Sajeev SK, Thampy S, Guruvayurappan K, Ajitha PS (2017) Bacteriological analysis of drinking water by MPN method from Chennai, India. IOSR J Environ Sci 11(7):57–64

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17(1):10–12

Md Zoqratt MZH, Eng WWH, Thai BT, Austin CM, Gan HM (2018) Microbiome analysis of Pacific white shrimp gut and rearing water from Malaysia and Vietnam: implications for aquaculture research and management. PeerJ 6:e5826

Medeiros JD, Cantão ME, Cesar DE, Nicolás MF, Diniz CG, Silva VL, Vasconcelos ATRD, Coelho CM (2016) Comparative metagenome of a stream impacted by the urbanization phenomenon. Braz J Microbiol 47:835–845

Mencíó A, Mas-Pla J (2010) Influence of groundwater exploitation on the ecological status of streams in a Mediterranean system (Selva Basin, NE Spain). Ecol Indic 10:915–926

Mittal P, Prasoodanan PKV, Dhakan DB, Kumar S, Sharma VK (2019) Metagenome of a polluted river reveals a reservoir of metabolic and antibiotic resistance genes. Environ Microbiome 14:5

Moon CD, Young W, Maclean PH, Cookson AL, Bermingham EN (2018) Metagenomic insights into the roles of Proteobacteria in the gastrointestinal microbiomes of healthy dogs and cats. Microbiol Open 7:e00677–e00677

Mueller-Spitz SR, Goetz GW, McLellan SL (2009) Temporal and spatial variability in nearshore bacterioplankton communities of Lake Michigan. FEMS Microbiol Ecol 67:511–522

Nakatsu CH, Byappanahalli MN, Nevers MB (2019) Bacterial community 16S rRNA gene sequencing characterizes riverine microbial impact on Lake Michigan. Front Microbiol. https://doi.org/10.3389/fmicb.2019.00996

Noor MEM, Fadaak SNESJ, Isa MNM, Bakar MFA, Abdullah MDD (2017) Metagenomic insights into the diversity and functions of microbial assemblages in Tasik Kenyir ecosystem. bioRxiv 176453

Ortiz-Álvarez R, Triadó-Margarit X, Camarero L, Casamayor EO, Catalan J (2018) High planktonic diversity in mountain lakes contains similar contributions of autotrophic, heterotrophic and parasitic eukaryotic life forms. Sci Rep 8:4457

Prusse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNAsequence data compatible with ARB. Nucl Acids Res 35(21):7188–7196

Rashid M, Manzoor M, Mukhtar S (2018) Urbanization and its effects on water resources: an exploratory analysis. Asian J Water Environ Poll 15:67–74

Saleem F, Mustafa A, Kori J, Hussain M, Azim K (2018) Metagenomic characterization of bacterial communities in drinking water supply system of a Mega City. Microbial Ecol 76:899–910

Santamaria M, Fosso B, Consiglio A, De Caro G, Grillo G, Licciulli F, Liuni S, Marzano M, Alonso-Alemany D, Valiente G, Pesole G (2012) Reference databases for taxonomic assignment in metagenomics. Brief Bioinform 13:682–695

Simon M, Jardillier L, Deschamps P, Moreira D, Restoux G, Bertolino P, Lopez-Garcia P (2015) Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. Environ Microbiol 17:3610–3627

Sinang SC, Poh KB, Shamsudin S, Sinden A (2015) Preliminary assessment of Cyanobacteria diversity and toxic potential in ten freshwater lakes in Selangor, Malaysia. Bull Environ Contam Toxicol 95:542–547

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci 103:12115–12120

Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, Sadowsky MJ (2013) Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. J Appl Microbiol 115:1147–1158

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol 19:21–31

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P (2015) Structure and function of the global ocean microbiome. Science 348:1261359

Suzuki T, Kanagawa T, Kamagata Y (2002) Identification of a gene essential for sheathed structure formation in *Sphaerotilus natans*, a filamentous sheathed bacterium. Appl Environ Microbiol 68:365–371

Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, Shalchian-Tabrizi K, Ruiz-Trillo I (2012) Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Mol Biol Evol 29:531–544

United States Environmental Protection Agency (2018) Standard method for clean water act analytical method. Retrieved November, 11, 2020, from https://www.epa.gov/cwa-methods/approved-cwa-chemical-test-methods

Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen D (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev 71:495–548

Wang SY, Sudduth EB, Wallenstein MD, Wright JP, Bernhardt ES (2011) Watershed urbanization alters the composition and function of stream bacterial communities. PLoS ONE 6:e22972

Watts MP, Spurr LP, Gan HM, Moreau JW (2017) Characterization of an autotrophic bioreactor microbial consortium degrading thiocyanate. Appl Microbiol Biotechnol 101:5889–5901

Wirth C, Limberger R, Weisse T (2019) Temperature × light interaction and tolerance of high water temperature in the planktonic freshwater flagellates *Cryptomonas* (Cryptophyceae) and *Dinobryon* (Chrysophyceae). J Phycol 55:404–414

Zafar H, Saier MHJr, (2018) Comparative genomics of transport proteins in seven Bacteroides species. PLoS ONE 13:e0208151

Zavarzin GA, Stackebrandt E, Murray RGE (1991) A correlation of phylogenetic diversity in the Proteobacteria with the influences of ecological forces. Can J Microbiol 37(1):1–6

Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat Microbial Ecol 28:141–155