Sys Rev Pharm 2020;11(11):1107-1114
A multifaceted review journal in the field of pharmacy

Next-Generation Sequencing Technologies for Environmental DNA as an Efficient Bio Indicator for Bacterial Biodiversity in Tigris River, Iraq

Fikrat M Hassan[1], Warqaa Y. Salih[2*] and Halah H Al-Haideri[1]

[1]College of Science for Woman, Department of Biology, University of Baghdad, Baghdad, Iraq
[2] Institutes of Genetic Engineering and Biotechnology, University of Baghdad, Baghdad, Iraq

Corresponding Author: Warqaa Y. Salih
Email: warqaayehia@gmail.com

ABSTRACT
Access to multiple, large sequence data volumes, as well as further improvements in read length and sequence platform efficiency, lead to better sample diversity. In order to assess the ability of Next Generation Sequence (NGS) approach to establish and evaluate the environmental molecular diagnostics, directly monitor bulk bacterial communities, the Illumina system of bacterial 16S rRNA gene amplicon was investigated in total bacterial communities in water samples. The dominant bacterial species can be identified through this method. NGS data will also allow the identification of drivers of the composition of bacterial communities during water treatment and distribution in combination with metadata, demonstrating the power of this approach to track bacterial regrowth and contamination in technical systems.

Keywords: DNA extraction; microbial source tracking; molecular diagnostics; Next Generation Sequencing; water quality assessment

Correspondence:
Warqaa Y. Salih
Institutes of Genetic Engineering and Biotechnology, University of Baghdad, Baghdad, Iraq
Email: warqaayehia@gmail.com

INTRODUCTION
Freshwater surface ecosystems, including lakes, rivers, and streams, play a major role in nutrient recycling biogeochemical cycling, and energy flows. Different micro-organisms play a critical role in the functioning of these cycles in these ecosystems, which are associated with water's physicochemical parameters [1].

We focus on Tigris River in Baghdad, Iraq, the second longest river in Western Asia, a length 1,800 km^2, originates in Armenian Highlands in Turkey, source of water for supplies drinking water, agriculture, industry, domestic water supply and for disposal of sewage [Alrubayi *et al*., 2011; Al-Ansari, 2016]. A better understanding of the range of microbial population characteristics are distribution, wealth and structure, etc. The analysis of environmental DNA (eDNA) through using specific gene markers such as species-specific DNA barcodes has been a key application of next generation sequencing technologies in ecological and environmental research. A newly developed method is DNA metabarcoding, via next-generation sequencing (NGS), enables the rapid identification of species in environmental samples, uses short gene sequences [3]. High-throughput NGS methods recently developed, such as MiSeq pyrosequencing of the 16S rRNA gene, have provided more detailed descriptions of bacterial populations in different environments due to the increased number of sequence reads that can be obtained [4].A few studies reported the importance to applied the NGS techn to study the biodiversity in Tigris river (Abed *et al*., 2018;Al-Raqi *et al*., 2018;Al-Meshhdany and Hassan, 2020).

The objective of this study was to examine and compare the microbial diversity and abundance of five locations through Tigris River within Baghdad city, using the Illumina MiSeq system. Furthermore, the presence of potential pathogens in these water samples was also evaluated. Our results provide more detailed information on the characteristics of urban surface water microbial communities, as well as for risk management of potential bacterial pathogens in urban surface waters.

MATERIALS AND METHODS
Site description
From November 2018 to July 2019, water samples with three replicates were collected from five sites through the Tigris river within Baghdad city, where site one is located at Al-Muthanna Bridge upstream this site is situated in the north of the Tigris River (latitude 33°25'41.85″ N, longitude 44°20'49.63″E). Two active sites were located in Al-Sarafiya and Al-Shuhadaa Bridge through Baghdad City's midstream, greatly populated area, where residential houses, hospitals, government institutions (latitude 33°2112.99″ N and 33°2019.99″N) (longitude 44°22'28.77″ E and 44°23'19.91″ E) , respectively. Site 4 contains important facilities in Baghdad, is one of the residential areas, located in Al-Jadriya (latitude 33°16'58.35″ N and longitude 44°22'31.87″ E) is famous for the existence of orchards. Site 5 (downstream) one of the significant agricultural, includes a group of science and commercial centers located in Al-Zafraniya area (latitude 33°17'25.44″ N, longitude 44°26'58.23″ E). The Global Positioning System (GPS) was used to determine the positions of the study sites.

Sample collection
Collection of water samples was performed from the surface layer with a depth of 20-30 cm Stopper-fitted clean polyethylene bottles (2 liter) were employed to collect the samples following rinsing for few times with river water to assess the physiochemical properties. Analyses of parameters were performed directly after sampling. Temperature, electrical conductivity, pH of the samples was determined on-site using a Multi-parameter water quality meter (Crison mm40). Other analyses

included assessment of turbidity, total nitrogen, total phosphorus, which were performed according to standard methods [5].

For bacterial isolation, 5 g of sediments sample were collected from the riverbank by using a clean and dry sterile spatula in a clean polythene bag, and to reduce the numbers of microorganisms, decimal places were used [6].

DNA extraction, Manipulation, and PCR

DNA was extracted using a Wizard genomic DNA purification kit (Promega, USA) according to the manufacturer's instructions. The concentration of extracted DNA was measured using QuantiFluor® (Promega, USA) prior to subject in PCR. PCR was performed using 5x FIREPol® Master Mix in order to target the 16S rRNA gene in selected samples. In general, 1 µl (1 mM) of each forwarded and reversed universe primers was used in PCR master mix with 5 µl of DNA template. The volume was completed with up to 20µl using nuclease-free water. The amplification reactions were performed based on the program described in table (1). PCR products were electrophoresed on 1 % agarose gel and visualized under UV light [7]. The product was stored at -20˚ C for further use.

Table1: Illustrates the PCR reaction program

PCR Program			
Steps	**°C**	**m: s**	**Cycle**
Initial Denaturation	95	05:00	1
Denaturation	95	00:30	30
Annealing	60	00:30	
Extension	72	00:30	
Final extension	72	07:00	1
Hold	10	10:00	

PCR products of all isolated bacteria were sent for Sanger sequencing using Illumina platform by Next Generation Sequencing (NGS) workflow, which includes 4 basic steps, by Macrogen Corporation laboratories in Korea, as follow:

1) Samples were prepared for library construction and DNA/RNA extraction from the sample.
2) Library Construction: Preparation of the sequencing library was performed by randomly fragmenting the DNA samples, followed by PCR amplification and gel purification.
3) Sequencing for cluster generation and Bridge amplification of the fragments was then performed into distinct clonal clusters. After completion of cluster generation, the templates were ready to be sequenced.
4) Raw sequencing data were converted into raw data for the analysis.

Data Statistics of NGS

This study was designed to assess the suitability of a metagenome amplicon sequencing approach. 16S rRNA sequencing was applied for targeting community structure of bacteria in Illumina MiSeq is an integrated machine that is able to amplify the clones, sequence genomic DNA, and analyze data in one run, ensuring base calling, alignment, variant calling, and reporting. For each sample, the values were calculated as follows:

bp: Total number of bases sequenced and total number of reads.
GC (%): Guanine + Cytosine content.
AT (%): Adenine + Thymine content.
Q20 (%): Ratio of bases that have phred quality score of over 20.
Q30 (%): Ratio of bases that have phred quality score of over 30.

RESULTS

Physical and chemical analysis

The temperature of sampling sites ranged from 14 to 45 °C. The highest values of electrical conductivity and salinity was recorded in site 1. Total nitrogen was ranged 0.5 -1.75 mg L^{-1}.

The lowest value of total phosphorus was observed at site 1, and the highest at site 3. The lowest and highest values of turbidity were recorded at site 1 and ranged 15- 456 NTU. The physical and chemical properties of 15 water samples are listed in (table 2).

Diversity analysis using Operational Taxonomic Units (OTUs)

The high-resolution insight into community composition provided by deep amplicon sequencing makes it a valuable tool for monitoring of microbial communities in nature as well as technical aquatic systems [8].

In total, Illumina MiSeq sequencing generated total number of bases sequenced (85.413.665 bp), and total number of reads (391,994) referred sequences for the 15 samples retained for Illumina paired-end sequencing, this value refers to the sum of read 1 and read 2. Guanine cytosine (GC %) content (51.865). Adenine thyamine (AT%) content (48.13). High quality 16S rRNA gene sequences with an average Q20(%). Ratio of bases that have phred quality score of over 20 recorded (97.45) and Q30(%) ratio of bases that have phred quality score of over 30 recorded (91.97) (Figure 1).

Table2: Represents the physical and chemical properties of water samples for all selected sites alongside Tigris River.

Physical and chemical parameters											
	Units	**Site 1**		**Site 2**		**Site 3**		**Site 4**		**Site 5**	
Parameters		**Min**	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
Electrical Conductivity	**µs. cm^{-1}**	653	1337	1123	505	1124	505	1050	630	1033	404
Mean		1023.44		830.9		820.3		883.3		823.8	
Salinity	**%**	0.64	0.84	0.31	0.68	0.72	0.31	0.67	0.38	0.66	0.25
Mean		0.4		0.52		0.51		0.59		0.52	
Water temperature	˚C	13	39	12	39	13	39	13	40	13	37
Mean		21.83		22.48		22.28		23.39		22.6	
Air temperature	˚C	17	34	17.1	45	16	45	14	44	17	45
Mean		27.11		28.12		28.02		28		28.4	
Turbidity	**NTU**	13.3	456	16	493	20	346.7	26.49	212.3	19.2	299.3
Mean		115.8		127.18		150		56.83		72.7	
Total Nitrogen	**ppm**	0.5	2	0.98	1.3	0.6	0.95	0.5	0.92	0.62	1.75
Mean		0.92		1		0.93		0.82		0.94	
Total phosphate	**ppm**	0	0.34	0.14	0.22	0.02	0.77	0.045	0.73	0.028	0.46
Mean		0.11		0.08		0.19		0.16		0.15	

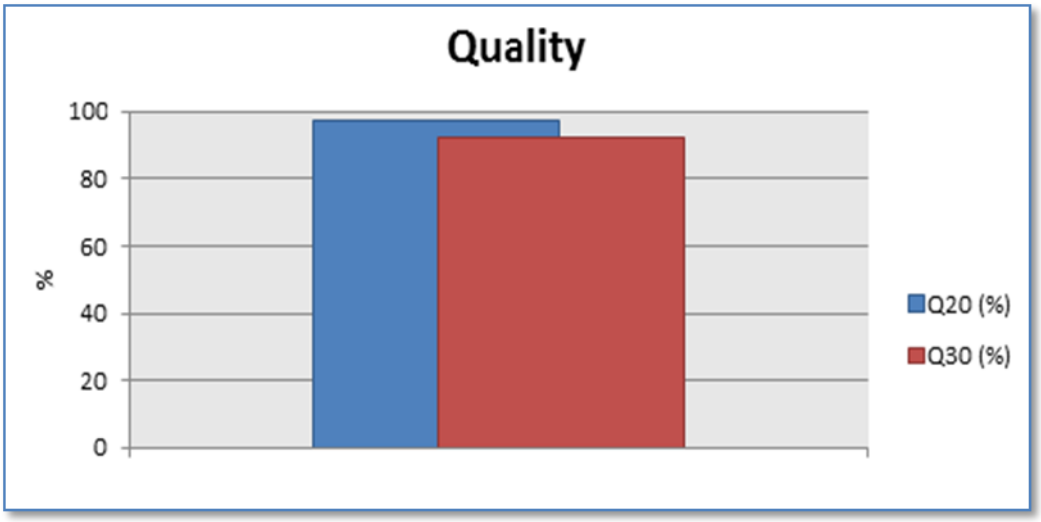

Figure 1: Ratio of bases that have Phred quality score of over 20 and 30

Sediments Samples OTUs were classified to (12) bacterial phyla the most predominant phyla being *Proteobacteria*, *Bacteroidetes,* and *Actinobacteria*. *Proteobacteria* and *Bacteroidetes* were the most commonly detected phyla in all the samples, accounting for (22.1 with abundance ratio 57.27 %) and (14.93 with abundance ratio 38.74 %) of sequences, respectively. The bacterial phyla abundances and accounting found in the different sample types investigated average relative abundance of Phylum in bacterial community composition of in this study (Table 3, Figure 2).

Table 3: An accounting of bacterial phyla during the period of study

Phylum	**Abundance Count**
Acidobacteria	34
Actinobacteria	1,247
Armatimonadetes	6
Bacteroidetes	14,927
Chloroflexi	3
Deinococcus-Thermus	9
Firmicutes	78
Fusobacteria	6
Planctomycetes	9

Proteobacteria	22,068
Synergistetes	3
Verrucomicrobia	35

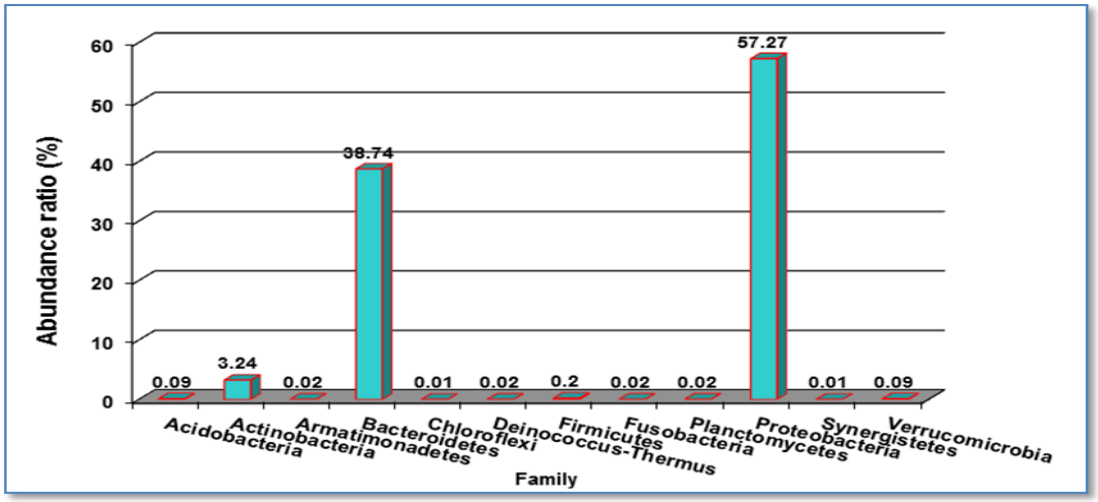

Figure 2: Family level distribution of groups in bacteria that present in Tigris River with abundance %.

According to 295 genera, which identified by NGS platform, most predominant species level 14 Outgroup taxonomic Units (OTUs) were observed, among them the species *Flavobacterium granuli* (19.27%), *Acinetobacter movanagherensis* (16.88 %), *Streptomyces olivoviridis* (0.87%). In less abundance *Anaerolinea thermolimosa* and *Cloacibacillus porcorum* were recorded (0.01) and some unclassified species were found in abundance (0.27) (Figure 3).

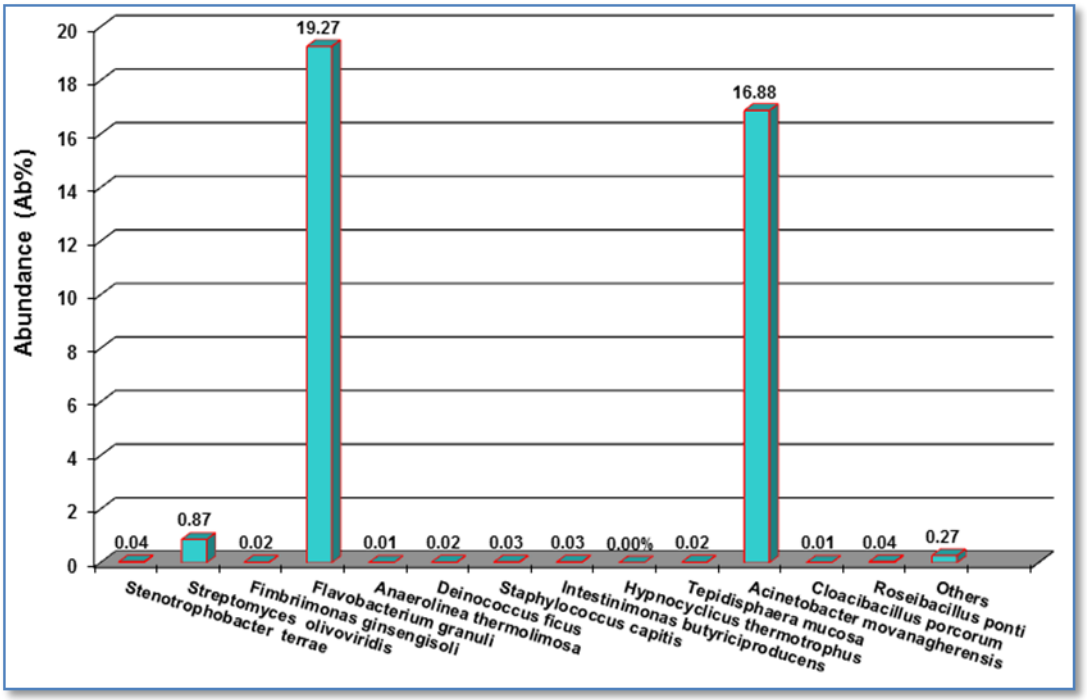

Figure 3: Species level distribution of bacterial populations present in Tigris River according to abundance (Ab%).

Identification of bacteria by Sequencing

Identification of bacterial isolates was confirmed by sequence-based phylogenetic tree (aligned sequences was conducted using MEGA 6 program) structuring analysis using 16S ribosomal RNA (16S rRNA) gene sequencing ; the resulted PCR products were subsequently sequenced to obtain DNA sequences, and 436 bp, 419 bp and 459 base pair (bp) product were obtained. For *Pseudomonas marginalis, Flavobacterium reichenbachii, and Acinetobacter calcoaceticus*, respectively in agarose gel electrophoresed for PCR product (Figure 19).

The amplicon was aligned online using BLAST at the NCBI. The 16S rRNA sequence of *P. marginalis*, *F. reichenbachii*, and *A. calcoaceticus* showed 100%, 96% and 95% homology with the NCBI sequence database with accession number and type of substitution, location

of nucleotide MK883114.1, KX809748.1 and AB680365.1, respectively (Table 4, Figures 4, 5, 6).

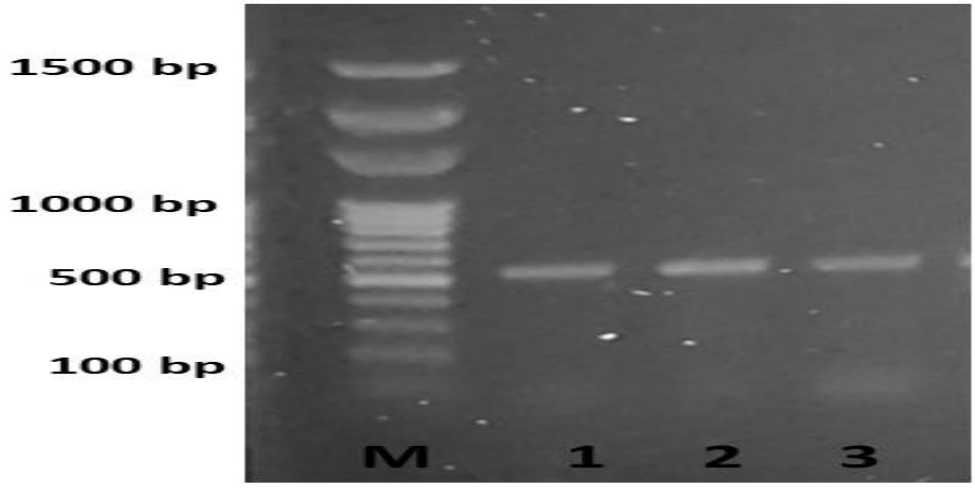

Figure 4: PCR product of 16S rRNA genes of unknown bacteria was analyzed on 1% agarose gel electrophoresis. The gene of interest was PCR amplified and analyzed by 1 % agarose gel. Lane 1, 2, and 3 refer to the samples of bacterial isolates as indicated by S1, S2, and S3. Lane L: DNA ladder (100bp).

Table 4: Types of polymorphism of 16S rRNA gene from isolated bacteria.

No. of sample	**Type of substitution**	**Location**	**Nucleotide**	**Sequence ID**	**Score**	**Identities**	Source
1	---------			**ID: MK883114.1**	787	100%	*Pseudomonas marginalis*
2	Transition	411	**C>T**	**ID: KX809748.1**	698	96%	*Flavobacterium reichenbachii*
	Transvertion	425	**A>C**				
	Transition	426	**C>T**				
	Transvertion	427	**T>G**				
	Transvertion	428	**C>G**				
	Transvertion	430	**A>T**				
	Transvertion	435	**T>A**				
	Transvertion	437	**G>C**				
	Transvertion	438	**A>C**				
	Transition	439	**G>A**				
	Transvertion	440	**T>G**				
	Transition	455	**G>A**				
	Transition	612	**G>A**				
3	Transvertion	346	**C>G**	**ID: AB680365.1**	734	95%	*Acinetobacter calcoaceticus*
	Transvertion	351	**G>C**				
	Transvertion	389	**A>T**				
	Transvertion	421	**C>A**				
	Transvertion	422	**T>G**				
	Transvertion	423	**A>C**				
	Transition	424	**C>T**				
	Transition	442	**G>A**				
	Transvertion	443	**A>T**				
	Transvertion	444	**T>G**				
	Transvertion	445	**A>C**				
	Transvertion	446	**G>T**				
	Transition	514	**G>A**				
	Transition	554	**C>T**				
	Transition	557	**C>T**				
	Transition	558	**T>C**				
	Transition	582	**C>T**				
	Transition	594	**G>A**				
	Transition	617	**A>G**				
	Transition	618	**G>A**				
	Transition	**723**	G>A				

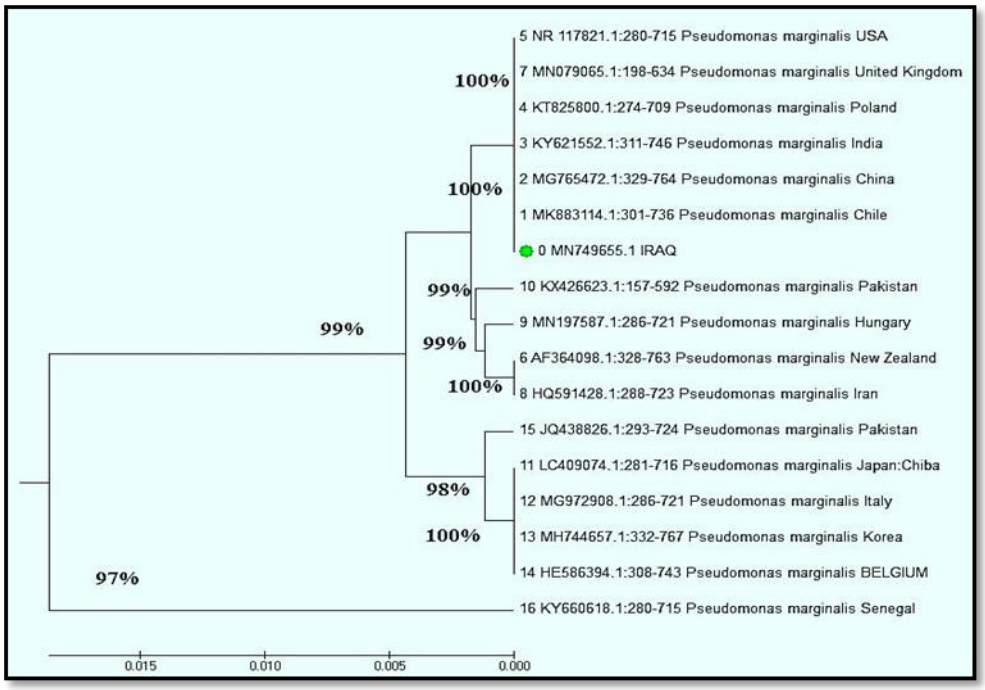

Figure 5: Phylogenetic tree of *P. marginalis* based on 16S rRNA gene sequences conferred by GeneBank database, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 436 bp of corresponding position of 16S rRNA gene sequence. MEGA 6 program was used for phylogenetic tree.

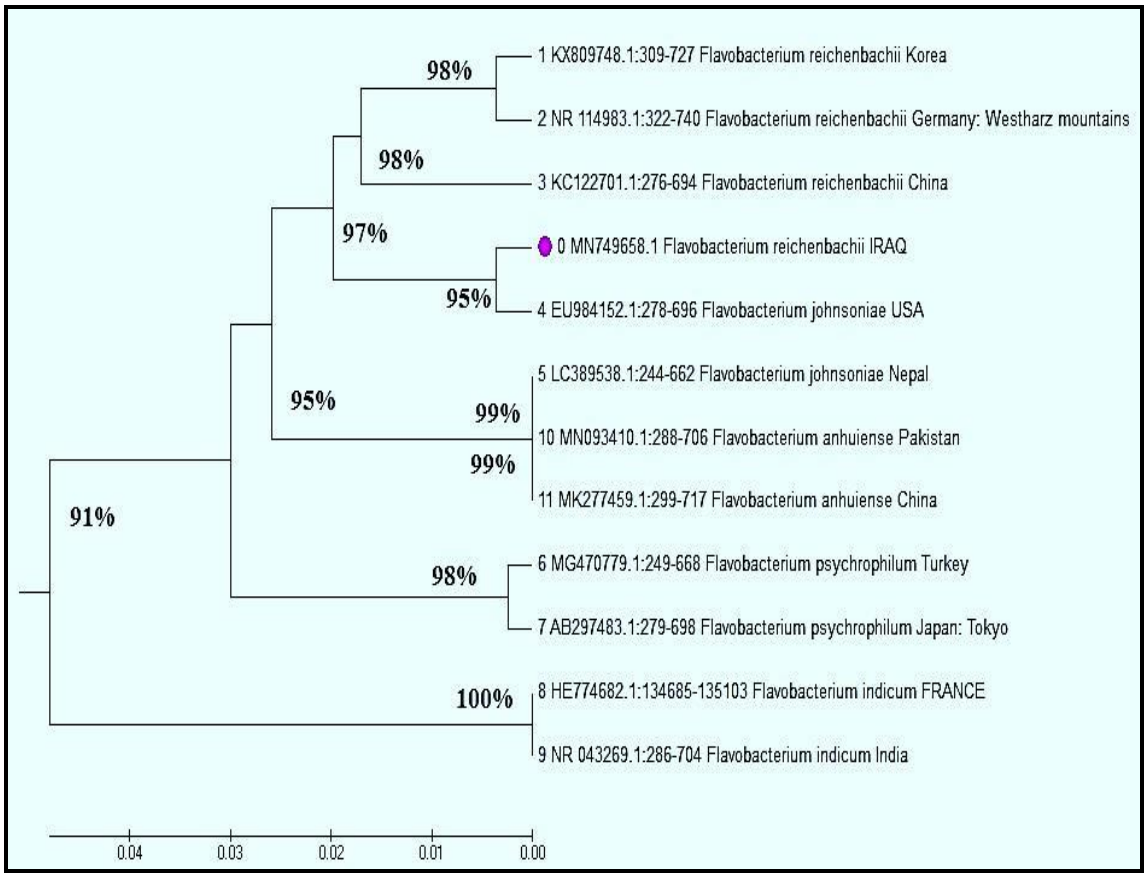

Figure 6: Phylogenetic tree of *F. reichenbachii* based on 16S rRNA gene sequences conferred by GeneBank data base, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 419 bp of corresponding position of 16S rRNA gene sequence. MEGA 6 program was used for phylogenetic tree.

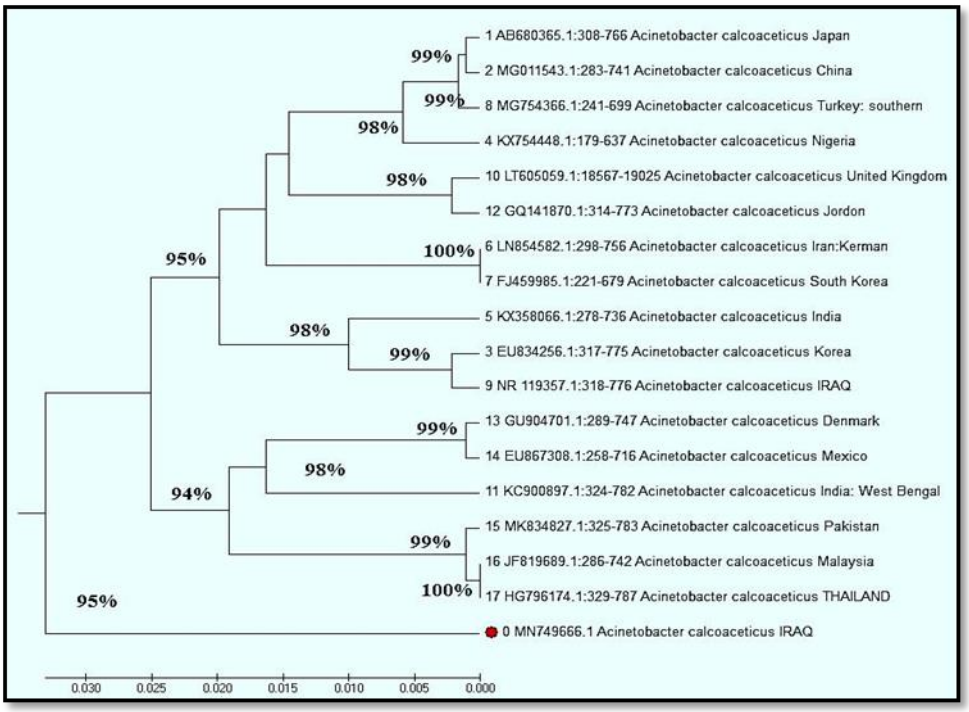

Figure 6: Phylogenetic tree of *A. calcoaceticus* based on 16S rRNA gene sequences conferred by Gene Bank database, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 459 bp of corresponding position of 18S rRNA gene sequence. MEGA 6 program was used for phylogenetic tree.

A total of 295 species were identified by NGS platform according to 16S rRNA analysis. Three bacterial species were added to the bacterial flora in Iraq and registered in NCBI as follows:

1. *A. calcoaceticus* (accession number MN 749666.1).
2. *P. marginalis* (accession number MN 749655.1).
3. *F. reichenbachii* (accession number MN 749658.1).

DISCUSSION

This study set out to assess the suitability of an amplicon sequencing approach targeting bacterial 16S rRNA genes using Illumina platform for the evaluation and development of molecular biological methods in water quality is testing as well as a direct tool for monitoring water quality. Surface water, consisting of both natural and man-made water bodies (rivers, streams, lakes, wetlands, parks, etc.) plays a vital role in urban ecosystem services [9], and remains an important source of drinking water, irrigation and recreation, etc. Changes in urban surface water can result in microbial community fluctuation, with implications for water quality and contamination of water-borne pathogen. To make water quality management plans, it is important to analyze and understand the microbial ecology of urban surface water from different regions. In our study, high-throughput amplicon sequencing was used to characterize bacterial community structure and diversity of 15 surface water and sediments samples collected from various locations in the city of Baghdad. These results suggest that the bacterial community composition of Tigris River was significantly different from each other.

The dominant phyla also differed among the sediment's samples. *Proteobacteria* (22- 63%) and *Bacteroidetes* (14-74%) were the dominant phyla, which is in agreement with the findings of Bai *et al*.2014, who examined the bacterial community structures in wastewater treatment bioreactors via high-throughput sequencing. However, *Actinobacteria* (41.17% of total sequences) and *Proteobacteria* (31.80%) were the dominant phyla. In 2016, Zhong *et al*. [10] reported that prokaryotic communities ' composition and diversity were strongly linked to salinity concentrations in Tibetan Plateau lakes. Ibekwe *et al* [11], stated that several primary physical and chemical water variables, including NO2, pH, and NO3, significantly influenced the microbial community structure of an urban river. In determining Ganjiang River's bacterial population composition, temperature was found to be the most influential factor [12]. In contrast to these studies, we found turbidity, and electrical conductivity signifcantly afected bacterial community, this may be because the 15 water samples showed frequently differences in values between sites.

Surface waters are highly polluted with pathogens in man y countries, especially in the developing world, leading to numerous outbreaks of waterborne disease [13].

However, few studies concerning waterborne pathogens in urban surface waters have utilized high throughput sequencing. In the present study, it was found that potentially pathogenic bacteria were ubiquitous across the sampled surface waters. Among the potential waterborne pathogens, *Pseudomonas,* were the most prevalent at all sites; followed by *Acinetobacter*, *Flavobacterium*, *Streptomyces, Sphingomonas*, *Arenimonas*, *Arcobacter*, *Legionella*, *Clostridium*, *Corynebacterium*. Definitions of how these species are potentially pathogenic include the following examples: in the United States, 21 percent of waterborne outbreaks caused by bacteria were recorded between 2007 and 2009, the leading etiological pathogenic bacteria being *E. coli* O157: H7, *Pseudomonas* spp, *Shigella sonnei*. Members of the genus Aeromonas are common aquatic microorganisms associated with human diseases [14].

CONCLUSION

The use of eDNA in combination with NGS to detect multiple species simultaneously is a powerful tool for monitoring species diversity in water bodies and will allow more accurate estimates of species diversity rather than targeted surveillance of one or a handful of species. NGS is an area of growing interest to regulatory bodies and ecologists and will open up new avenues of ecosystem monitoring by allowing the species richness of aquatic environments to be quantified. Microbial community characteristics of 15 surface water samples in Tigris River within Baghdad city were studied using 16S rRNA Metagenome amplicon sequencing. *Bacteroidetes*, *Proteobacteria* and *Actinobacteria* were the predominant phyla across all the samples. Principal coordinate analysis indicated significant differences in surface water bacterial community composition. Meanwhile, our results demonstrated the presence of potential waterborne pathogens in all of the surface waters sampled and that Pseudomonas are ubiquitous bacteria in a wide variety of surface waters.

REFERENCES

1. Wu CH, Sercu B, Van De Werfhorst LC, Wong J, DeSantis TZ, Brodie EL, Hazan TC, Holden P A, Andersen GA (2010) Characterization of coastal urban watershed bacterial communities leads to alternative community-based indicators. Plos ONE 5:e11285 https://doi.org/10.1371/journal.pone.0011285
2. Al-Ansari N, (2016) Hydro-politics of the Tigris and Euphrates basins. Engineering, 8: 140-172. DOI: 10.4236/eng.2016.83015
3. Alrubayi, A. A., Almukhtar, E A., Alwaily A J. (2011) Effects of Organic Pollution on some Fish and Benthic Macroinvertebrate Groups in Rivers Tigris and Diyala at Baghdad Area. Baghdad Science Journal, 8(1), 462-470. https://doi.org/10.21123/bsj.2011.8.1.462-470
4. Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A (2017) Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshwater Science 36: 162–177. https://doi.org/10.1086/690649
5. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, Sodergren E (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nature Communications, 10:1-11. DOI: 10.1038/s41467-019-13036-1
6. Alpha American Public Health Association (2017) Standard methods for the examination of water and wastewater. 23 ed., Washington, Dc, USA .pp :1200.
7. Krieg NR, Holt JG. Bergey's manual of systematic bacteriology, vol 1. Williams and Wilkins. Baltimore London. 1984.
8. Maniatis T, Pritsch E, Sambrook J (1982) Molecular Cloning: A Laboratory Manual. New York (NY).
9. Kusstatscher P, Zachow C, Harms K, Maier J, Eigner H, Berg G, Cernava T(2019) Microbiome-driven identification of microbial indicators for postharvest diseases of sugar beets. Microbiome 7:112. https://doi.org/10.1186/s40168-019-0728-0
10. Garcia X, Barcelo D, Comas J, Corominas L, Hadjimichael A, Page TJ, Acuna V (2016) Placing ecosystem services at the heart of urban water systems management. Sci Total Environ 563–564: 1078–1085 (2016). DOI: 10.1016/j.scitotenv.2016.05.010
11. Zhong Z, Liu Y, Miao L, Wang F, Chu L, Wang J, Lui Z (2016) Prokaryotic community structure driven by salinity and ionic concentrations in plateau lakes of the Tibetan plateau. Appl Environ Microbiol 82:1846–1858. http://dx.doi.org/10.1128/AEM.03332-15
12. Ibekw AM, Ma J, Murinda SE (2016) Bacterial community composition and structure in an urban river impacted by different pollutant sources. Sci Total Environ 566–567:1176–1185 (. doi: 10.1016/j.scitotenv
13. Wang P, Chen B, Yuan RQ, Li CQ, Li Y (2016) Characteristics of aquatic bacterial community and the influencing factors in an urban river. Sci Total Environ 569-570: 382–389 DOI: 10.1016/j.scitotenv
14. Pate CB, Shanke R, Gupta VK, Upadhyay RS (2016) Q-pcr based culture-independent enumeration and detection of enterobacter: an emerging environmental human pathogen in riverine systems and potable water. Front Microbiol 172 00172. https://doi.org/10.3389/fmicb
15. Hlavsa MC, Roberts VA, Anderson AR, Hill VR, Kahler AM, Orr M, Garrison LE, *et al* (2011) Surveillance for waterborne disease outbreaks and other health events associated with recreational water-united states, 2007–08. Mmwr Surveill Summ 60: 1–32.