Original Research Article

Next-generation sequencing for investigating the diversity of microorganisms and pathogenic bacteria in a water source

Yingliang Ge[1,2] and Shuili Yu[1]

[1]School of Municipal and Environmental Engineering, Harbin Institute of Technology, Harbin, Heilongjiang Province 150090, [2]School of Food Science and Technology, Harbin University, Harbin, Heilongjiang Province 150080, China

*For correspondence: **Email:** shuiliyuhu@163.com*

Abstract

Purpose: To employ next-generation sequencing (NGS) to investigate the diversity of microorganisms and pathogenic bacteria from a water source in Tai Lake, China, in winter.
Methods: Water samples from the same source were collected over a period of 3 months (December 2013 to February 2014), and their physicochemical characteristics were determined. The DNA of the samples were extracted and amplified by polymerase chain reaction (PCR). The PCR products were sequenced by Miseq PE300 pyro sequencing platform. The results for 16S rDNA were analysed using visualization software Gephi, and the 16S rDNA gene pool of known pathogenic bacteria was established.
Results: A total of 144,292 16S rDNA gene sequences were obtained and ranked by RDP classifier. The average length of the sequences was 395.66 bp. They revealed 580 operational taxonomic units (OTUs) classified into 16 phyla. A full length of 16S rDNA gene database of common pathogenic bacteria was established. After blasting, 17 species of pathogenic bacteria were found. The most abundant potential human pathogenic bacteria were affiliated to B. tribocorum. Most environmental factors had significant impact on pathogenic bacteria.
Conclusion: These results indicate that NGS can be used for the simultaneous detection of most recognized water-borne pathogenic bacteria. Variations in microorganisms in water source at different periods in winter can provide insight into the diversity of microorganisms in the water.

Keywords: Next-generation sequencing, Pathogenic bacteria, Diversity, 16S rDNA gene

INTRODUCTION

Water-borne diseases are illnesses caused by intake of water that harbors pathogenic microorganisms [1]. In developing countries, thousands of people die from water-borne diseases every year [2]. Water-borne diseases are caused by water-borne pathogens, which are recognized as the microbial risk in drinking water [3]. In addition to pathogens, drinking water may also harbor various animal and plant pathogenic microorganisms [4], which survive even after water processing [5]. These microorganisms persist in drinking water, thereby posing danger to people's health.

Several advances have been made in analysis of drinking water in an effort to make it safe for drinking. For instance, indicator organisms such as total coliforms, faecal coliforms and *Escherichia coli*, have been

used for nearly a century in monitoring the quality of drinking water [6]. However, each indicator organism has its own limitations, and sometimes it may not be effectively detected in the water [7]. Pure culture method has also been used to detect pathogenic microorganisms. However, a large number of microorganisms and pathogens cannot be detected by pure culture because although they are viable, they are non-culturable [8]. In addition, water-borne microorganisms are characterized by low abundance, diversity, and complexity, which make them difficult to detect. Therefore, it is necessary to explore novel methods for detecting pathogenic microorganisms in drinking water.

Molecular techniques such as PCR, quantitative real-time PCR and DNA microarray technology have been explored for pathogen identification and quantification in water [9]. However, these molecular methods rely on specific oligonucleotide probes, and the known sequences of the microbes detected [10]. Furthermore, the molecular techniques cannot detect enough pathogenic bacteria at a go, and are also unable to identify diversity and abundance of pathogenic bacteria in water. Recent technological advances in next-generation sequencing (NGS) offer better prospects for detection of pathogenic microorganisms and investigating their diversity [11]. NGS can quickly generate huge amounts of DNA reads, and the technique is affordable [12].

In this study, the pathogens in a water source from East Tai Lake was analyzed in winter by Illumina Miseq sequencing of the PCR products of the V4+V5 region of the bacterial 16S rDNA gene. The sequences obtained were assigned to taxonomic ranks.

EXPERIMENTAL

Study site and sample collection

Water samples were collected over a period of three months in winter (Dec 2013 to Feb 2014) from the same place (31º0'8.88"N, 120º27'21.15"E) of the water source at a biologically activated carbon drinking-water treatment plant in Wu Jiang, at 8:30 every day. The samples were appropriately labelled with Dec 13, Jan 14 and Feb 14. For each sample, 10 L water was obtained under the surface of the lake (about 1.2 m) near the water inlet. The samples were put into sterilized glass bottles and stored in foam insulation box with frozen ice packs prior to transportation to the laboratory. The microorganisms were enriched from 2 L water sample by 0.22 μm tangential flow filter (TFF; Pellicon XL 50 cm^2, Milipore) on sterile console at 4 ^{o}C, centrifuged at 5000 rpm for 5 min, and stored at -80 ^{o}C.

Water quality analysis

Physico-chemical characteristics of water, such as dissolved oxygen, temperature, pH, electrical conductivity and total dissolved solids were measured using portable rapid detection (HACH HQ30D and HANNA HI98130). Turbidity, total organic carbon, total nitrogen, total phosphate, and colony lump-sum were determined according to the standard methods.

DNA extraction and PCR amplification

Soil DNA kit (OMEGA, USA) was used to extract DNA under the guidance of the instructions in the kit. The treatment replicates were pooled, and the extracted DNA was diluted in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and stored at -20 °C. A Thermo NanoDrop 2000 spectrophotometer was used to determine the quantity and quality of the extracted DNA prior to 1 % agarose gel electrophoresis.

The samples were selected for microbial community and pathogen analysis by 16S rDNA tag pyro sequencing, and in each sample, an aliquot of the extracted DNA was used as a template for amplification. The 16S rDNA genes were amplified by PCR using primers with universal 16S primer sequences 515F (GTGCCAGCMGCCGCGG) and 907R (CCGTCAATTCMTTTRAGTTT), targeting the V4+V5 hyper variable regions (about 392 bp) of the 16S rDNA gene. Amplifications were carried out in a total volume of 20 μL, containing 5×FastPfu buffer, 2.5 mM dNTPs, 5 μM forward primer, 5 μM reverse primer, FastPfu polymerase (TransGen AP221-02, China), and 10 ng template DNA. One step PCR was performed using a mixture of TransStart Fastpfu DNA polymerase and FastPfu DNA polymerase under the following thermal profile: 95 ^{o}C for 2 min, then 27 cycles of 95 ^{o}C for 30 s, 55 ^{o}C for 30 s, and 72 ^{o}C for 30 s, followed by one cycle of 72 ^{o}C for 5 min and 10 ^{o}C hold. The PCR products were put together by sample label, and subjected to 2 % agarose gel electrophoresis. Then the products were isolated with

AxyPrep DNA nucleic acid purification kits (AXYGEN, USA), and quantified using a QuantiFluor™-ST (Promega, USA).

High-throughput pyro sequencing and sequence diversity analysis

Pyro sequencing was used to analyse bacterial diversity of the source water. An equal amount of the PCR product was mixed in a single tube and run on a Miseq PE300 pyro sequencing platform (Majorbio BioPharm, Shanghai). The sequences of double-ended DNA from Miseq PE300 pyro sequencing were merged into one sequence based on the overlap relationships between PE reads, using Seqprep (https://github.com/jstjohn/SeqPrep). The quality of reads and the merged effect were inspected.

The minimum length of overlap was 15 bp, the maximum error ratio was 0.02, and the minimum matching rate was 0.9. The sequences were subsequently assigned to each sample with a bar code. The length distribution of valid sequences (301-500 bp) was 99.99 %, and the average length was 395.66 bp. At 97 % similarity level, the trimmed and unique 16S rDNA sequences were lumped into the operational taxonomic units (OTUs) [13] by applying RDP pyro sequencing pipeline. The results were stored at the short reads archive (SRA) database of National Centre for Biotechnology Information Search database (NCBI, http://www.ncbi.nlm.nih.gov) (accession number: SRP068068).

Visual analysis of the data of microbial diversity

The result of 16S rDNA blast was analysed using the visualization software Gephi (https://gephi.org) [14,15]. After high-throughput pyro sequencing and sequence diversity analysis, a csv file was built, which could be received by Gephi base on the genera (source column), environment (target column) and abundance (the number of 16s rDNA column). Gephi can intuitively reveal patterns and trends, and highlight outliers without learning graph theory. The csv files were uploaded to the free cloud space (http://pan.baidu.com/s/1pKgzpof, share password: 8tub) and was freely accessed. The relationships between microorganisms and environment were shown by a graph layout algorithm Force Atlas 2 [16].

Establishment of 16S rDNA gene pool of pathogenic bacteria

16S rDNA is highly conserved and specific, and has become a means of pathogen detection. The 16S rDNA gene pool of the known pathogenic bacteria was established based on the virulence factor database (VFDB, http://www.mgc.ac.cn/VFs/main.htm) and NCBI database. Then the gene pool was uploaded to free cloud space (http://pan.baidu.com/s/1skgvvDr, share password: 784u), named pathogen_16s_rrna.fasta. The result of Miseq PE300 high-throughput pyro sequencing was blasted with the gene pool using NCBI's BLASTN [17]. The sequences that were highly similar (> 97 %) to the sequences of 16S rDNA for standard pathogenic bacteria, with length > 395 bp and scores > 360 were identified from the blast result.

Relationship between pathogenic bacteria and environment factor

The influence of environmental factors on the abundance of pathogenic bacteria was investigated using RDA. This method can identify almost all biological information of pathogenic microorganisms and find out the environmental factors that influence them. The angle formed by the line connecting the dot (pathogenic bacteria) and the origin and the parameter rays shows the correlation between pathogenic microorganisms and environment factors.

RESULTS

Water quality

The physico-chemical parameters of the water source at the same site in three winter months are shown in Table 1 and Table 2. Results on Table 1 show that physical quality parameters were relatively stable except temperature and turbidity. The pH values ranged from 7.77 ± 0.05 to 7.84 ± 0.07.

From Dec. 2013 to Feb. 2014, the temperature decreased from $10.09 \pm 2.15\ ^{\circ}C$ to $7.04 \pm 2.12\ ^{\circ}C$, while dissolved oxygen concentrations increased from 11.15 ± 0.71 to 11.58 ± 0.79 mg/L. Mean values of turbidity, dissolved oxygen and conductivity in Feb. 2014 were higher than that in the other two periods.

Table 1: Physical quality parameters of the three water samples

Sample	DO (mg/L)	pH	T(℃)	Cond (µS/cm)	NTU
Dec 2013	11.15 ± 0.71	7.84 ± 0.07	10.09 ± 2.15	551.13 ± 16.39	4.68 ± 1.54
Jan 2014	11.49 ± 0.42	7.79 ± 0.07	8.26 ± 1.18	552.29 ± 7.66	4.27 ± 1.22
Feb 2014	11.58 ± 0.79	7.77 ± 0.05	7.04 ± 2.12	564.11 ± 7.96	7.73 ± 3.07

DO = dissolved oxygen; Cond = conductivity; NTU=Turbidity

Table 2: Biological and chemical water quality parameters of three samples

Sample	COD (mg/L)	TN (mg/L)	TP (mg/L)	AN (mg/L)	A (10^4/L)
Dec 2013	3.70 ± 0.23	0.54 ± 0.10	0.012 ± 0.001	0.20 ± 0.04	39.13 ± 5.95
Jan 2014	3.43 ± 0.16	0.53 ± 0.09	0.014 ± 0.004	0.21 ± 0.05	34.74 ± 5.11
Feb 2014	3.55 ± 0.16	0.53 ± 0.10	0.020 ± 0.005	0.21 ± 0.04	44.54 ± 5.87

COD = chemical oxygen demand; TN = total nitrogen; TP = total phosphorus; AN = ammonia nitrogen; A = algae density

The chemical oxygen demand, total nitrogen and ammonia nitrogen were unstable in the three months. The algae density did not change significantly, but mean values of total phosphorus increased significantly (Table 2), implying that the organic content of the water increased with less precipitation in winter. The high turbidity may protect microorganisms in the water.

Microbial diversity and richness

To give a comprehensive insight into the microbial communities in the water source, 16S rDNA gene sequences of bacteria were amplified using universal primer pairs 515F/907R. A total of 144,292 high-quality bacterial sequence reads (≥200 bp, average length of about 395.66 bp) were generated from the PCR amplicons by Miseq PE300 pyro sequencing.

NGS results showed that the microbial communities included 16 different phyla: *Verrucomicrobia*, *Spirochaetae*, *Proteobacteria*, *Planctomycetes*, *Nitrospirae*, *Lentisphaerae*, *Gemmatimonadetes*, *Fusobacteria*, *Firmicutes*, *Elusimicrobia*, *Cyanobacteria*, *Chloroflexi*, *Chlorobi*, *Bacteroidetes*, *Actinobacteria*, and *Acidobacteria.* A few sequences could not be classified, such as BD1-5 and TM6. At the OTU level, the differences in relative abundances between samples are shown in Figure 1A. A total of 215 genera were obtained, and the relative abundance of microbial community was different in the three samples. The relative abundance of microbes in the Dec 13 sample was similar to that in the Jan 14 sample, but was different from that in Feb 14 sample (Figure 1B). *Chloroplast norank*, *hgcI clade*, *Candidatus planktophila*, *Arcicella*, *Flavobacterium*, and *Methylotenera* were the most

dominant bacteria in the three samples (Figure 1C).

Visualized microbial diversity

The relationships between samples and bacterial survival were generated by the visual program Gephi, the visual figure upload of google drive, and the share URL: https://drive.google.com/file/d/0B9IKtmqDKU JKMHpENm9yLXdhblU/view?usp=sharing. There were 215 genera in the three samples, among which only 16 were in Feb 14 sample, 11 were in Jan 14 sample, 9 were in both Dec 13 and Feb 14 samples; 8 were in Dec 13 and Jan 14 samples, and 33 were both in Jan 14 and Feb 14 samples. The Gephi visual picture reflects the presence and belonging of a microorganism more intuitionally than heatmap picture.

Verified pathogen sequence

A pathogenic bacterial gene pool including 468 genera and 1608 integral 16S rDNA sequences was built for 16S rDNA blasting. These sequences were long enough (> 390 bp) to avoid errors caused by short sequences. The sequences obtained in this study were blasted by the pathogenic 16S rDNA gene pool generated in this study. The most similar sequences with the standards were selected (score > 340, similarity > 97 %, E-value < 10^{-5}, and length > 395 bp) A total of 17 species of pathogenic bacteria, were obtained after blasting with 16S rDNA gene database (Table 3). These include *Brucella suis*, *Clostridium botulinum B str.*, *Legionella pneumophila Paris*, *Clostridium beijerinckii*, *Bartonella tribocorum*, *Salmonella, Enterica (serovarty phimurium)*, *Bacillus thuringiensis, Serovar konkukian str.*, *S. haemolyticus*, *M. smegmatis*, *P. fluorescens*, *S. agalactiae*, *Y.*

pseudotuberculosis, *L. welshimeri*, *Bordetella parapertussis*, *P. fluorescens* and *P. syringae pv*. In sequences of NGS, 28,604 sequences may belong to the pathogens, accounting for about 19.82 % of total sequences. The abundance of the pathogenic bacteria and

their similarities are shown in Figure 2. *B. suis*, *B. tribocorum* and *S. thermophilus* were the most abundant species of pathogenic bacteria in the water samples.
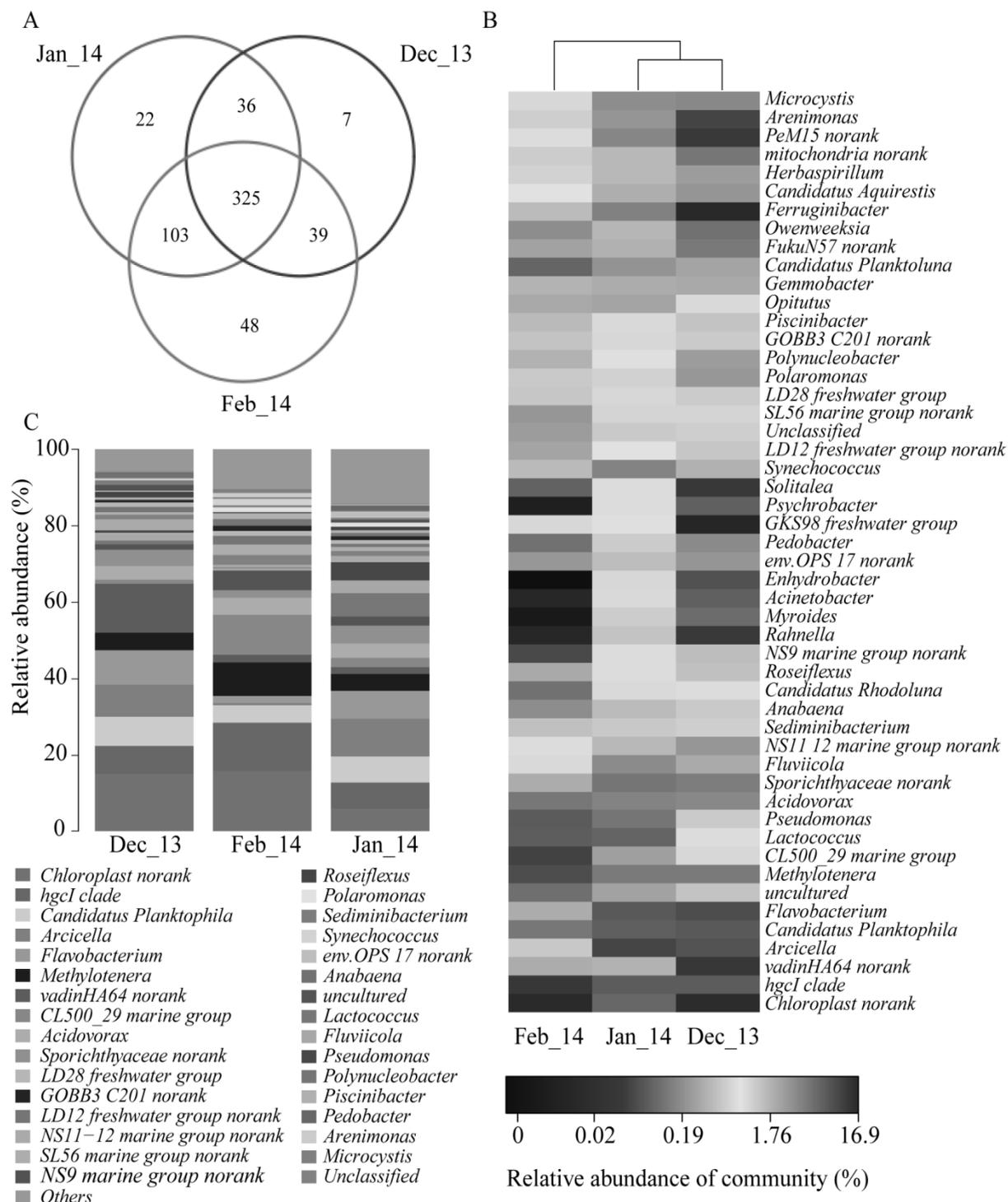

Figure 1: Microbial diversity and abundance in water source from Tai Lake in winter. (A) Venn diagram; (B) relative abundance of microbial community, and (C) heatmap showed the diversity and abundance of microorganisms in the water, as well as the differences among the three water samples

Table 3: Possible pathogenic bacteria after blasting by established 16s rDNA gene database

Species	Similarity (%)	length of sequence consistent with blast sequence		Pathogenic
		Q-Begin	*Q-End*	
P. syringae pv. tomato	99	1	396	plant pathogen
P. fluorescens	99	1	396	human pathogen
Bordetella parapertussis	97	1	395	human and animal pathogen
L. welshimeri	97	1	397	human and animal pathogen
Y. pseudotuberculosis	99	1	396	human and animal pathogen
S. agalactiae	100	1	396	human and animal pathogen
S. thermophiles	100	1	396	potentially pathogenic bacteria
E. faecalis	97	1	397	human and animal pathogen
M. smegmatis	98	1	398	opportunistic pathogen
S. haemolyticus	100	1	397	opportunistic pathogen
Bacillus thuringiensis serovar konkukian str.	100	1	397	opportunistic pathogen
S. enterica (serovar typhimurium)	98	1	396	human and animal pathogen
Bartonella tribocorum	97	1	395	human and animal pathogen
C. beijerinckii	99	1	396	opportunistic pathogen
L. pneumophila Paris	97	1	397	human and animal pathogen
C. botulinum B str.	97	1	396	human and animal pathogen
B. suis	99	1	395	human and animal pathogen

Relationship between pathogenic bacteria and environment factor

Most environmental factors, including dissolved oxygen, pH, temperature, total phosphate, and algae density, had significant impact on pathogenic bacteria (Table 4). As shown in Figure 3, algae density was positively correlated with conductivity, and only a few pathogenic microorganisms were negatively correlated with the algae density and conductivity, such as *C. botulinum B str* and *Bartonella tribocorum*. Total phosphate was negatively correlated with temperature, and dissolved oxygen was negatively correlated with pH.

DISCUSSION

The present study analysed the physico-chemical parameters as well as the biodiversity of water source from Tai Lake in winter. The results showed that the bacterial species were relatively stable and most of the bacteria were detected simultaneously in three water samples. The abundance of pathogenic bacteria changed with environment factors. The physico-chemical parameters investigated in this research were used to assess the quality of water from Tai Lake. From Dec. 2013 to Feb. 2014, the pH values were stable, ranging from 7.77 ± 0.05 to 7.84 ± 0.07, which are in accordance with the pH of natural waters. Natural waters have pH values between 6.5 and 8.5, which make them useful as drinking water [17]. The temperature of the water source decreased from Dec. 2013 to Feb. 2014, in keeping with changes in weather temperature. However these changes do not affect the quality of drinking water. It was also found that with decrease in temperature, dissolved oxygen concentration increased. Higher concentrations of dissolved oxygen promote the growth of aerobic organisms, such as algae. Moreover, the instabilities in turbidity and chemical oxygen demand may be affected by the dredging of the Tai Lake.

NGS analysis identified 144,292 high-quality bacterial sequence reads. After chimera check and removal of obviously erroneous reads, the validated bacterial reads were assigned into genus level taxonomic ranks with the RDP Classifiers. The 16S rDNA sequence in bacteria genome is highly conserved. The RDP Classifiers were used to assign gene sequences from the water samples, and affiliate them to 16 phyla and 215 genera. The relative abundance of microbial communities was different in the three samples, which may be due to the environment factors. Microbial diversity was shown in the Gephi visual picture which reflects the presence and classification of a microorganism more intuitionally than heatmap picture. Besides, Gephi can also reflect the relationship between microorganisms and environment. Gephi visual picture is considered suitable for the display of multiple samples of microbial species in the study of biodiversity [14].
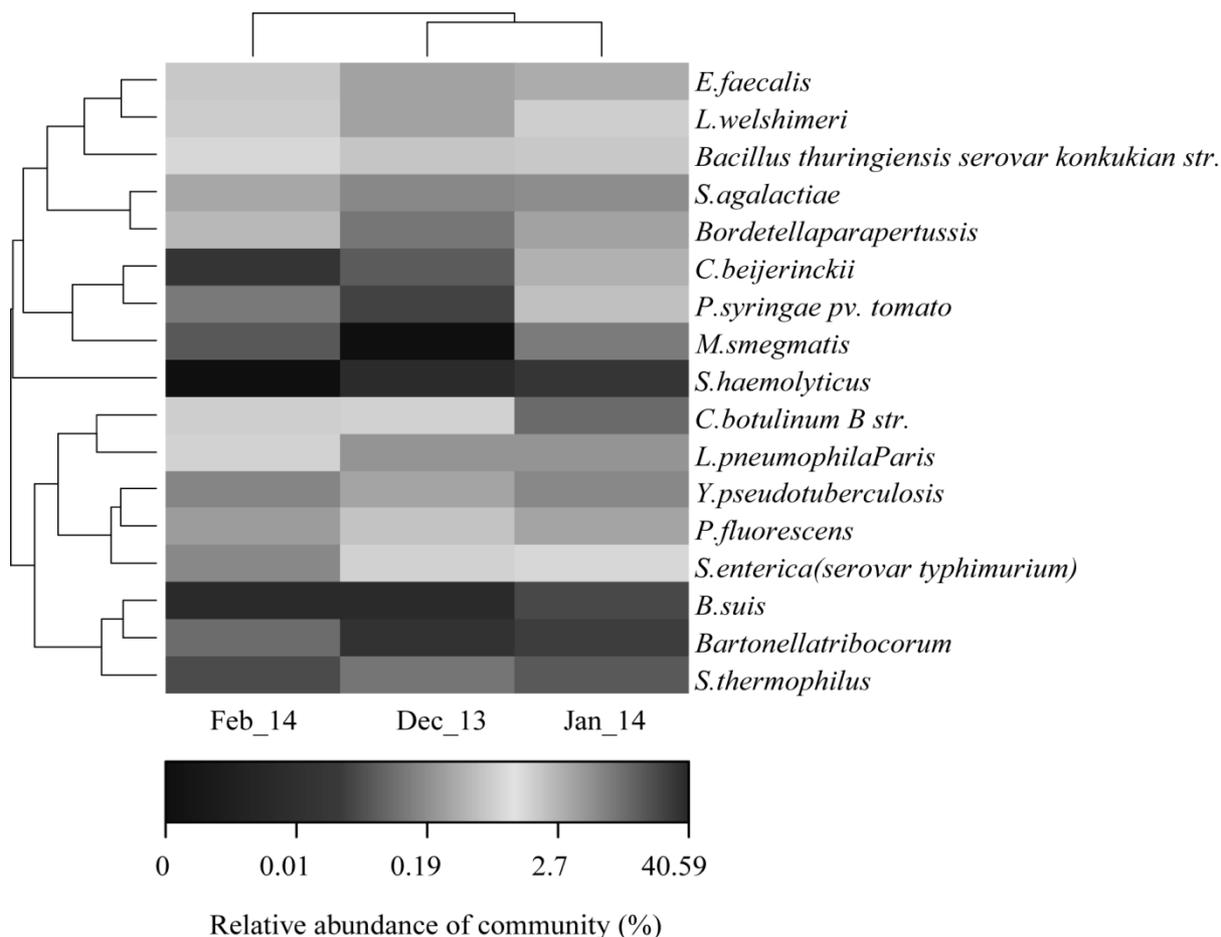

Figure 2: Heatmap of pathogenic bacteria in source water. The species were different in the three water samples

Table 4: Permutation test data for RDA

Environment factor	RDA1	RDA2	r^2	Pr (> r)	Significance
Dissolved oxygen	0.98156	-0.19114	1	0.001	***
pH	-0.99384	0.11081	1	0.001	***
Temperature	-0.99989	-0.01508	1	0.001	***
Conductivity	0.87187	0.48973	1	0.509	
Total phosphate	0.91609	0.40096	1	0.001	***
Ammonia nitrogen	0.99918	0.04061	1	0.181	
Algae density	0.43579	0.90005	1	0.001	***

P-values based on 0.999 permutations. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
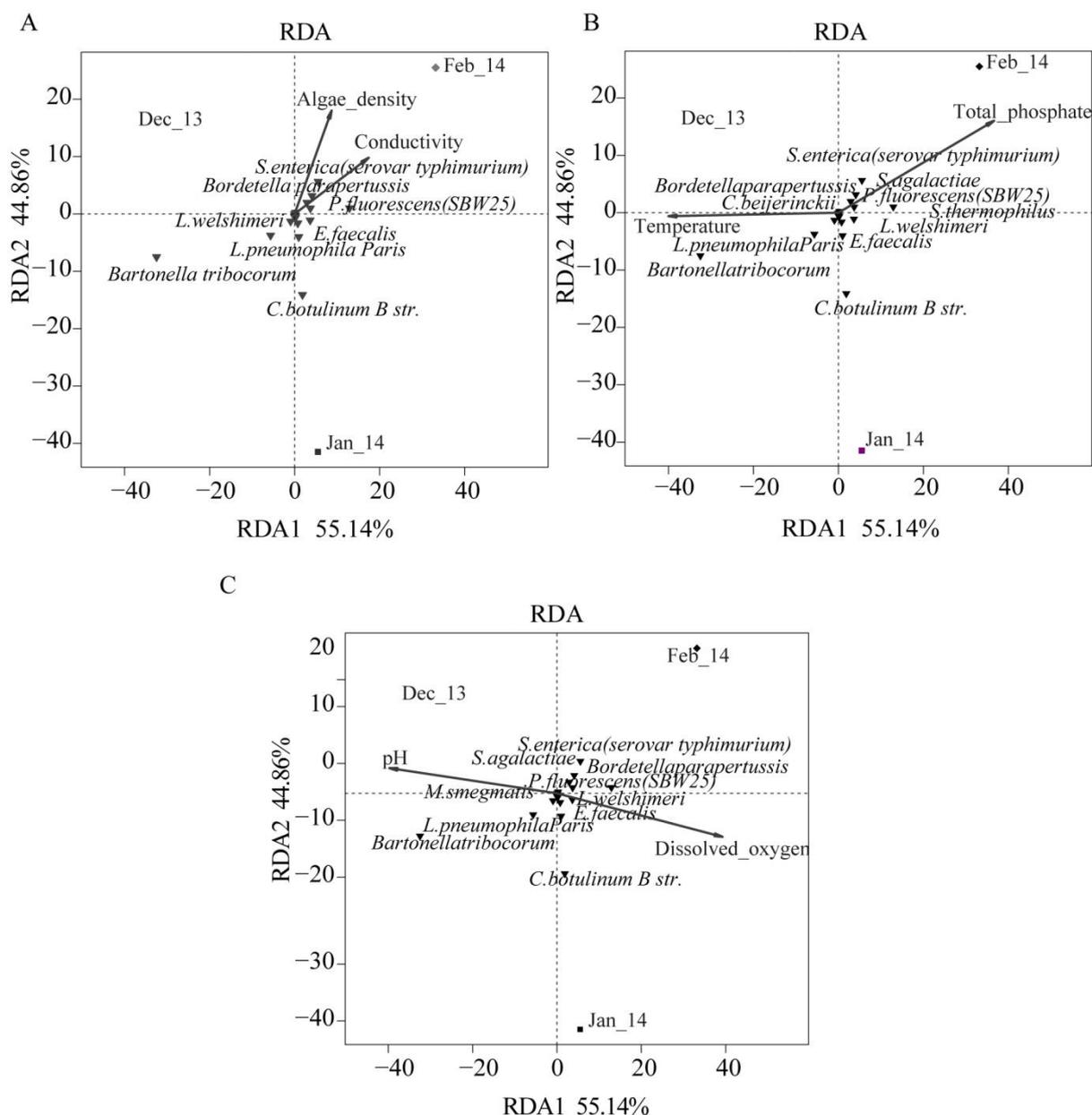
Figure 3: RDA graph reflecting the relationships between the pathogenic bacteria and environment. (A) relationships between algae density, conductivity and pathogenic bacteria; (B) relationships between total phosphate, temperature and pathogenic bacteria; (C) relationships between dissolved oxygen, pH and pathogenic bacteria.

Based on the Illumina Miseq PE300 pyro sequencing platform, so many sequences were obtained, making it difficult to identify the potentially pathogenic bacterial sequences. Although many pathogenic bacteria databases have been established, like NMPDR [18] , EuPathDB [19] and VFDB [20], they are difficult to blast on the result of 16S rDNA sequences of NGS.

Therefore, 16S rDNA gene pool of pathogenic bacteria was built for blasting the 16S rDNA sequences got from sequencing. This revealed the pathogenic bacteria at the species level in the source water samples.

After analysis, *B. suis*, *B. tribocorum* and *S. thermophilus* appeared to be the most abundant species of pathogenic bacteria in the water samples. *B. suis* is a common pathogenic bacteria. Some lesions caused by *B. suis* include lesions in reproductive organs and joints. In particular, several species may lead to severe diseases in humans [21]. *B. tribocorum* has been isolated from rodents in different parts of the world in recent years. However, there is currently no evidence that *B. tribocorum* has a zoonotic potential [22]. *S. thermophilus* is a major dairy starter used for the manufacture of yoghurt and cheese [23].

It is difficult to understand the relationships between pathogenic bacteria and environmental factors due to the complexity and variety of environment factors. In the present study, the influence of environment factors on the abundance of pathogenic bacteria was investigated using the RDA. The abundance of some 16S rDNA sequences of pathogenic bacteria, such as *B. tribocorum* and *C. botulinum B str* changed with environmental factors. Total phosphate was negatively correlated with temperature. It has been demonstrated that total phosphate is associated with the eutrophication of water which is beneficial to the growth of pathogens [24].

In addition, it is well known that with decreases in temperature, the concentration of dissolved oxygen increases, which promotes the growth of pathogens because most pathogens are aerobic. Therefore, reduction in emissions of nitrogen and phosphorus, and control of eutrophication of water bodies in source water may be effective ways of reducing the risk of pathogenic microorganism leaking into drinking water.

CONCLUSION

Comprehensive the 16S rDNA sequences of bacteria and pathogens in water samples from Tai lake in winter, through NGS indicate that the pathogenic bacterial species were relatively stable in the lake in winter, suggesting a potential microbial risk in the water source. The abundance of pathogenic bacteria correlates with environmental factors. The methods used in this study to detect the potential pathogenic bacteria in the water source are more comprehensive than those used in previous studies. Thus, the results afford a reliable insight into the identity of pathogenic bacteria in this water source.

ASSOCIATED CONTENT

Supporting information: The additional information include the list of the common pathogenic bacteria of humans, animals and plants; the 16S rDNA gene database of common pathogenic bacteria with fasta file and the data source of Gephi with csv file. These materials are available free of charge via the internet at yun.baidu.com and drive.google.com with the share-password.

DECLARATIONS

Conflict of Interest

No conflict of interest associated with this work.

Contribution of Authors

The authors declare that this work was done by the authors named in this article and all liabilities pertaining to claims relating to the content of this article will be borne by them.

REFERENCES

1. *Santamaría J, Toranzos GA. Enteric pathogens and soil: a short review. Int Microbiol 2003; 6(1): 5-9.*
2. *Montgomery MA, Elimelech M. Water and sanitation in developing countries: including health in the equation. Environ Sci Technol 2007; 41(1): 17-24.*
3. *Riley MR, Gerba CP, Elimelech M. Biological approaches for addressing the grand challenge of providing access to clean drinking water. J Biol Eng 2011; 5(1): 1-10.*
4. *Sobsey MD, Pillai SD. Where future emerging pathogens will come from and what approaches can be used to find them, besides VFARs. J Water Health 2009; 7 Suppl 1(4): 75-93.*
5. *Propato M, Uber JG. Vulnerability of water distribution systems to pathogen intrusion: how effective is a disinfectant residual? Environ Sci Technol 2004; 38(13): 3713-3722.*
6. *Tallon P, Magajna B, Lofranco C, Leung KT. Microbial indicators of faecal contamination in water: a current perspective. Water Air Soil Pollut 2005; 166(1): 139-166.*

7. *Payment P, Locas A. Pathogens in water: value and limits of correlation with microbial indicators. Ground Water 2011; 49(1): 4-7.*

8. *Ramamurthy T, Ghosh A, Pazhani GP, Shinoda S. Current perspectives on viable but non-culturable (VBNC) pathogenic bacteria. Front Public Health 2014; 2:103.doi: 10.3389/fpubh.2014.00103.*

9. *Shaw JLA, Monis P, Weyrich LS, Sawade E, Drikas M, Cooper AJ. Using amplicon sequencing to characterize and monitor bacterial diversity in drinking water distribution systems. Appl Environ Microbiol 2015; 81(18): 6463-6473.*

10. *Ye L, Zhang T. Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. Environ Sci Technol 2011; 45(17): 7173-7179.*

11. *Costa D, Mercier A, Gravouil K, Lesobre J, Delafont V, Bousseau A, Verdon J, Imbert C. Pyrosequencing analysis of bacterial diversity in dental unit waterlines. Water Res 2015; 81(14): 223-231.*

12. *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): e116955.*

13. *Patin NV, Kunin V, Lidström U, Ashby MN. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. Microb Ecol 2013; 65(3): 709-719.*

14. *Jacomy M, Bastian M, Heymann S, "Gephi: an open source software for exploring and manipulating networks," in Proceedings of the Third International ICWSM Conference, (San Jose, California, USA, 2009), pp 361-364.*

15. *Kauffman J, Kittas A, Bennett L, Tsoka S. DyCoNet: a gephi plugin for community detection in dynamic complex networks. Plos One 2014; 9(7): e101357.*

16. *Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. Plos One 2014; 9(6): e98679.*

17. *Tepe Y. Some physicochemical characteristics of Yarseli Lake, Hatay, Turkey. Tur J Fish Aquat Sc 2005; 5(s2-3): 35-42.*

18. *Mcneil LK, Reich C, Aziz RK, Bartels D, Cohoon M, Disz T, Edwards RA, Gerdes S, Hwang K, Kubal M. The national microbial pathogen database resource (NMPDR): a genomics platform based on subsystem annotation. Nucleic Acids Res 2007; 35(suppl_1): 347-353.*

19. *Aurrecoechea C, Brestelli J, Brunk BP, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M. EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res 2010; 38(suppl_1): 415-419.*

20. *Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. Nucleic Acids Res 2016; 44(D1): 694-697.*

21. *Fretin D, Whatmore AM, Dahouk SA, Neubauer H, Garin-Bastuji B, Albert D, Hessche MV, Ménart M, Godfroid J, Walravens K. Brucella suis identification and biovar typing by real-time PCR. Vet Microbiol 2008; 131(3-4): 376-385.*

22. *Harrus S, Bargal GK, Golan A, Elazarivolcani R, Kosoy MY, Morick D, Avidor B, Baneth G. Isolation and genetic characterization of a Bartonella strain closely related to Bartonella tribocorum and Bartonella elizabethae in Israeli commensal rats. Am J Trop Med Hyg 2009; 81(1): 55-58.*

23. *Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblondbourget N, Decaris B, Bolotin A, Delorme C, Dusko ES. New insights in the molecular biology and physiology of Streptococcus thermophilus revealed by comparative genomics. Fems Microbiol Rev 2005; 29(3): 435-463.*

24. *Wu RSS. Eutrophication, water borne pathogens and xenobiotic compounds: environmental risks and challenges. Mar Pollut Bull 1999; 39(1-12): 11-22.*