**Next-generation sequencing reveals fecal contamination and potentially pathogenic bacteria in a major inflow river of Taihu Lake**

**Kiran Kumar Vadde a, Qiaoli Feng a, Jianjun Wang b, c, Alan J. McCarthy d and Raju Sekar a, \***

a *Department of Biological Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, China.*

b*State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China.*

c *University of Chinese Academy of Sciences, Beijing, China*

d *Microbiology Research Group, Institute of Integrative Biology, University of Liverpool, UK.*

\*Corresponding author.   
Department of Biological Sciences,

Xi’an Jiaotong-Liverpool University,

Suzhou 215123, Jiangsu Province, China

*E-mail address:* [Sekar.Raju@xjtlu.edu.cn](mailto:Sekar.Raju@xjtlu.edu.cn) (R. Sekar)

**A B S T R A C T**

Taihu Lake is one of the largest freshwater lakes in China and serves as an important source for drinking water. This lake is suffering from eutrophication, cyanobacterial blooms and fecal pollution, and the inflow Tiaoxi River is one of the main contributors. The goal here was to characterize the bacterial community structure of Tiaoxi River water by next-generation sequencing (NGS), paying attention to bacteria that are either fecal-associated or pathogenic, and to examine the relationship between environmental parameters and bacterial community structure. Water samples collected from 15 locations in three seasons, and fecal samples collected from different hosts and wastewater samples were used for bacterial community analysis. The phyla Proteobacteria, Actinobacteria, Bacteroidetes, and Cyanobacteria were predominant in most of the water samples tested. In fecal samples, Bacteroidetes, Firmicutes, and Proteobacteria were abundant, while wastewater samples were dominated by Proteobacteria, Bacteroidetes, Acidobacteria, and Chloroflexi. The cluster analysis and principal coordinate analysis indicated that bacterial community structure was significantly different between water, fecal and sewage samples. Shared OTUs between water samples and chicken, pig, and human fecal samples ranged from 4.5 to 9.8% indicating the presence of avian, pig and human fecal contamination in Tiaoxi River. At genus level, five bacterial genera of fecal origin and sequences of seven potential pathogens were detected in many locations and their presence was correlated well with the land use pattern. The sequencing data revealed that *Faecalibacterium* could be a potential target for human-associated microbial source-tracking qPCR assays. Our results suggest that pH, conductivity, and temperature were the main environmental factors in shaping the bacterial community based on redundancy analysis. Overall, NGS is a valuable tool for preliminary investigation of environmental samples to identify the potential human health risk, providing specific information about fecal and potentially pathogenic bacteria that can be followed up by specific methods.

**Main findings of the work:**

The results from this study indicate that NGS is a valuable tool for preliminary screening of fecal pollution and potential pathogens in lakes and rivers in order to identify the human health risk.

**Highlights**

* Next-generation sequencing (NGS) is applied to investigate fecal contamination and pathogens.
* Host associated fecal and potential pathogenic bacteria are present in Tiaoxi River.
* A number of bacterial genera in feces are suitable for development of fecal contamination markers.
* NGS is effective for analyzing the aquatic environment for fecal contamination and pathogens.

**Keywords:** Tiaoxi River; Taihu watershed; Fecal contamination; Bacterial pathogens; Next-generation sequencing.

**1. Introduction**

Waterborne diseases cause about 2.2 million deaths annually, with the majority occurring in children under the age of 5 years (WHO, 2015). In China, rapid urbanization, industrialization, and socio-economic development have led to a high degree of pollution in lakes and rivers, and this is regarded as a major challenging environmental issue (Jiang, 2009). Although surface water quality in China has improved rapidly in recent years, it has been reported that nearly 200 million people are still using unsafe water sources and approximately 60,000 people die every year due to water pollution related diseases (Han et al., 2016; Jiang, 2015). As diarrheal diseases are primarily caused by contamination of water sources with human or animal feces (WHO, 2017), monitoring of fecal contamination and pathogens in waters used for human consumption and recreational activities has become mandatory.

Most conventional fecal monitoring studies rely on the enumeration of fecal indicator bacteria (FIB), and this is considered as the “gold standard” to assess microbiological water quality and pathogen presence in environmental waters (Savichtcheva and Okabe, 2006). However, conventional FIB enumeration methods do not determine the origin of the fecal source, and previous reports have shown little or no correlation between FIB and the presence of pathogenic organisms (Shahryari et al., 2014). Therefore, several qualitative and quantitative “Microbial Source Tracking (MST)” methods have been developed to overcome this limitation, and MST methods have focused on determining the origin of the fecal sources (Green et al., 2014; Kildare et al., 2007; Mieszkin et al., 2009). In addition, several quantitative polymerase chain reaction (qPCR) methods have been widely used to determine the presence and abundance of pathogens (Ahmed et al., 2009; Oster et al., 2014). However, these methods can identify only the targeted pathogens that are specifically chosen for monitoring. As a result, it is a challenge to monitor a wide-range of pathogens in a watershed with culture-based or culture-independent qPCR-based methods. Next-generation sequencing (NGS) methods targeting the 16S rRNA gene have gained attention in order to explore the diversity of bacterial communities and their influence on microbial water quality (Ibekwe et al., 2013; Staley et al., 2013), potentially overcoming the limitations of culture and PCR-based detection protocols. Although community-based NGS methods (OTU comparison between fecal sources and environmental samples) have been proposed for microbial source tracking (Jeong et al., 2011; Unno et al., 2010), they are considered as a qualitative method for assessing fecal pollution since the OTU comparison results show discrepancies depending on the nature of the fecal OTU library applied (Boehm et al., 2013). However, a Bayesian algorithm based NGS method (SourceTracker) has been developed to predict the quantitative presence of fecal contamination (Knights et al., 2011). Recent reports indicated lower confidence in quantification results and also spatiotemporal limitations of the SourceTracker method indicating the necessity for optimization and validation prior to application in a new geographical area (Ahmed et al., 2015; Brown et al., 2017; Staley et al., 2018; Unno et al., 2018). The NGS based microbial community analysis approach is still considered as a valuable tool for preliminary investigation of water samples to assess public health risk associated with fecal contamination or pathogens (Tan et al., 2015), though such studies are very limited. The NGS method can evaluate bacterial diversity (including fecal and pathogenic bacteria) of water or other environmental samples and their relative abundance, providing valuable information to prioritize specific exposure assessment of suitable targets using quantitative (qPCR) methods (Tan et al., 2015). Most of the NGS studies on monitoring microbial communities in water samples have relied on targeting hypervariable regions of the 16S rRNA gene (Guo et al., 2013). Although 454-pyrosequencing was the commonly used sequencing platform in earlier studies, it has now been discontinued and is superseded by the sequencing of much larger libraries on the Illumina platform to yield community 16S rRNA based sequence datasets that are orders of magnitude larger and much more informative (Loman et al., 2012; Newton et al., 2015; Sinclair et al., 2015). The commonly used open source software packages for sequence analysis such as QIIME and MOTHUR have also been upgraded to analyze “pair-end” sequence data produced by Illumina sequencers, improving the performance (Caporaso et al., 2010; Kozich et al., 2013).

Taihu Lake is one of the top five largest freshwater lakes in China and serves as an important source of drinking water in addition to providing fisheries, transportation and flood protection (Chen et al., 2016). Currently, Taihu lake is connected to more than 200 rivers and tributaries, though the main inflow river is limited to thirteen rivers (Qin et al., 2007). Previous reports indicated that the inflow rivers were contributing to eutrophication, cyanobacterial blooms and fecal pollution, particularly Tiaoxi River (Hagedorn and Liang, 2011; Vadde et al., 2018). Although previous studies have addressed the microbial community composition in water and sediment samples of Taihu Lake (Cai et al., 2013; Paerl et al., 2011; Wilhelm et al., 2011; Zhao et al., 2017; Zheng et al., 2017), these have not focused on fecal bacteria and pathogens in the Tiaoxi River. Here, the bacterial community structure and relative abundance of fecal bacteria and pathogens in Tiaoxi River water are determined by the interrogation of large 16S rRNA gene sequence datasets generated by NGS on the Illumina platform. The specific objectives were to i) study the spatial and temporal variations in bacterial diversity in Tiaoxi River water, ii) determine the relative abundance of bacteria and potential pathogens of fecal origin and iii) assess the influence of environmental factors on bacterial diversity.

**2. Methods**

*2.1. Collection of water and fecal samples*

Twenty-five sampling locations were initially selected, covering domestic, agricultural and industrial areas of Tiaoxi river (Taihu watershed) (Fig.1). Based on the detailed physico-chemical and microbiological characterization of the water, 15 locations were identified as pollution hotspots in our earlier study (Vadde et al., 2018) and water samples collected from those locations were used for bacterial community analysis. The details of land use type around the selected sampling locations are provided in Vadde et al. (2018). Sampling at these locations was carried out on three occasions: autumn 2014, winter and summer 2015. Five liters of surface water was collected in sterile polypropylene containers from each location and brought to the laboratory on ice. The samples were processed within 8 hours of collection. Water samples (250 ml) were filtered through 0.22-µm polycarbonate membrane filters (Millipore, UK) and filters were stored at -20oC prior to DNA extraction.

A total of 120 fresh individual fecal samples from a range of hosts comprising chicken (CK), cow (CW), dog (DG), duck (DU), goose (GO), human (HU), and pigs (PI) in the vicinity of Tiaoxi river and primary effluents (n=6) from a wastewater treatment plant (WWTP) were also analysed. Fresh human fecal samples were provided by healthy volunteers (n=12) aged between 16 and 40 years and the samples were collected in sterile containers by the volunteers. Ethical approval for handling fecal samples in this study was acquired from Xi’an Jiaotong- Liverpool University Research Ethics Committee who also approved the safety guidelines and handling procedures. Individual fecal samples from animal hosts representing pigs (n=28), chicken (n=23), dog (n=21), duck (n=13), goose (n=11) and cow (n=12) were collected from farms located near the Tiaoxi River. All the samples were brought to the laboratory on ice and were stored at -20oC within 6 hours of sample collection. Approximately 0.5 g of different individual fecal samples of a host were pooled together to form a composite sample (in duplicate) used for DNA extractions. Each composite fecal sample was prepared from at least five different individuals of respective host fecal samples (Dubinsky et al., 2016). Primary effluents (500 mL; n=6) from a WWTP located in Suzhou were collected and brought to the laboratory on ice. Biomass from primary influents was collected by centrifugation (4000xg for 10min at 4oC) and the DNA was extracted immediately.

*2.2. Physico-chemical and microbiological analyses*

The physico-chemical and microbiological analyses of the water samples collected from 25 sampling locations along the Tiaoxi river and the results have been reported in our previous study (Vadde et al. 2018).

*2.3. Extraction of genomic DNA from water and fecal samples*

Genomic DNA was extracted from water samples (membrane filters) using the PowerSoil DNA Isolation Kit (MoBio Inc., Carlsbad, CA), following the manufacturer’s instructions. Fecal DNA was extracted in duplicate for each composite fecal/sewage sample using PowerFecal® DNA isolation kit (MoBio, Carlsbad, CA USA) as per the manufacturer’s instructions. 0.25 grams of composite feces or biomass from primary effluent samples were used for DNA extraction and the extraction was carried out for each type of fecal source separately to avoid any cross-contamination with other hosts. The DNA concentrations and quality were assessed using a NanoDrop ND 2000UV spectrophotometer (Thermo Fisher Scientific., Vienna, Austria) and the DNA extracts were stored at -20oC until further analysis.

*2.4. Bacterial community analysis by Illumina MiSeq Sequencing (NGS)*

The diversity and composition of bacterial communities in water and fecal samples was investigated by NGS using the Illumina MiseqPE250 platform. The NGS was carried out at the Shanghai Majorbio Pharmaceutical Technology Limited, China. The hypervariable region (V4) of bacterial 16S rRNA genes present in water and fecal DNA samples was amplified by PCR using an universal primer set described previously by Kozich et al. (2013). The barcoded 515F (5’-GTGCCAGCMGCCGCGGTAA-3’) and 806R (5’-GGACTACHVGGGTWTCTAAT-3’) primer set was used for amplification. Triplicate PCR reactions were carried out for each sample with each reaction mixture comprising 4 μL of 5 × FastPfu Buffer, 0.4 μL of FastPfu Polymerase, 2 μL of 2.5 mM dNTPs, 0.8 μL of each primer (5 μM), 10 ng of template DNA and reaction volume was made up to 20 μL with nuclease-free water. PCR amplification conditions used were as follows: 95 °C for 2 min, followed by 30 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s with a final extension at 72 °C for 10 min to ensure complete amplification. The amplified PCR products that were separated on 2% agarose gels were extracted and purified using the Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) following the manufacturer instructions and quantified using QuantiFluor™ -ST (Promega, USA). Purified amplicons were pooled in equimolar concentration and paired-end sequenced (2 × 250) on an Illumina MiSeqPE250 platform according to the standard protocols.

*2.5. Sequence data processing and analysis*

Raw data sequence files from Illumina were de-multiplexed, poor quality sequences were removed, and barcode, adaptor, and primer sequences were trimmed off using QIIME (version 1.17). UPARSE（version 7.1 http://drive5.com/uparse/) was used to cluster Operational Taxonomic Units (OTUs) with 97% similarity cutoff and chimeric sequences were removed using UCHIME to produce high-quality sequences. These sequences (ca. 250bp) were aligned with sequences in SILVA 16S rRNA database and RDP Classifier (<http://rdp.cme.msu.edu/>) was used for phylogenetic affiliation of each high-quality sequence. The sequence data were normalized and lowest sequence reads of 18640 were used for each sample for further comparison between samples (Ahmed et al., 2015). The sequences obtained in this study were submitted to the National Center for Biotechnological Information (NCBI) Short Read Archive (SRA) database under the succession numbers SAMN09469451 to SAMN09469511.

*2.6. Statistical analysis*

The species diversity and richness of bacterial communities within each sample (alpha diversity) were determined by Shannon (H’) and Simpson (D) diversity indices, abundance-based coverage estimator (ACE), and Chao1 richness estimator using MOTHUR (<http://www.mothur.org>) (Schloss, 2009). Principle Coordinate Analysis (PCoA) and cluster analysis were carried out to compare the bacterial diversity between different samples (beta diversity) by using QIIME (Caporaso et al., 2010; Lozupone and Knight, 2005). The relationship between the environmental parameters and bacterial community in the samples was performed by redundancy analysis (RDA) with the R language vegan package (http://www.R-project.org/ 2013).

**3. Results**

*3.1. Assessment of physico-chemical and microbiological parameters*

The detailed results of physico-chemical and microbiological parameters measured in various locations in Tiaoxi river were reported in our earlier study (Vadde et al. 2018). Among the tested parameters, pH and conductivity were found to be within the acceptable limits set by the Ministry of Environmental Protection (MEP), China (MEP, 2016), however TN was significantly higher than acceptable levels in all the locations and across the three seasons. TP, NH4-N, and NO2-N exceeded the acceptable levels in some locations and the influence of these parameters on bacterial community structure and diversity are discussed in detail in section 3.5. According to the MEP, the fecal coliforms (FC) are included in total coliforms (TC) and the national standards for TC are 10000/1L (1000/100mL). As there were no specific guidelines for FC according to MEP, they were compared to USEPA standards in the current study (USEPA, 2012). The spatial and temporal variation in FC counts across different locations of Tiaoxi River showed that acceptable limits (250 CFU/100mL) were exceeded in 15 locations indicating potential fecal contamination at these locations. Therefore, water samples collected from these 15 locations were selected for NGS in order to characterize the bacterial community and determine relative abundance of bacteria including fecal-associated and potentially pathogenic bacteria.

*3.2. Bacterial diversity and community composition in water and fecal samples*

In total, 61 samples (45 water samples collected from 15 locations in three seasons and 14 host-specific fecal and 2 wastewater samples) were used for sequencing and in-depth monitoring of bacterial community composition by NGS. A total of 1,694,935 bacterial 16S rRNA reads were generated from water, wastewater and fecal samples with sequence libraries of size ranging from 18640 to 37367 reads. The rarefaction curves based on the total number of OTUs with normalized sequence reads (n=18640 sequence reads per sample) are presented in Fig. S1. The average number of OTUs, species diversity (Shannon and Simpson) indices and richness (ACE and Chao1) values for bacterial diversity observed in water, wastewater and fecal samples are given in Table S1.1. For water samples (n=45), a total of 39192 OTUs were generated ranging from 564 to 1292 OTUs. The highest diversity and species richness were observed at location 20 (WIW20) and 21 (WIW21) during winter and the lowest at locations 2 (WIW02) and 3 (WIW03) as indicated by OTU richness, Chao and Shannon indices (Table S1.2). Similarly, for fecal samples (n=14), 5254 OTUs were generated ranging from 138 to 640 OTUs and the lowest species diversity and richness was observed for duck, human, dog, and chicken fecal samples compared to pig, goose, and cow fecal samples (Table S1.1). A total of 2360 OTUs were generated for wastewater samples (n=2) with an average of 1180, and these samples had the highest diversity relative to fecal and water samples (Table S1.1).

The RDP classifier categorized all the OTUs of water into 20 bacterial phyla, however, their relative abundance varied with the type of sample (Fig. 2). Proteobacteria, Actinobacteria, Bacteroidetes*,* and Cyanobacteria were the predominant phyla accounting for ca. 90% of total relative abundance in all locations except in locations 2 (WIW02) and 3 (WIW03) samples collected in the winter season. In these two samples, Proteobacteria, Firmicutes, Bacteroidetes*,* and Cyanobacteria were abundant. In Tiaoxi river water samples, 180 different genera were identified, although their relative abundance varied e.g. members of Cyanobacteria and Actinobacteria(hgcI\_clade) showed higher relative abundance in all the locations of Tiaoxi River water, except location 2 (WIW02) and 3 (WIW03) samples collected in the winter season (Tables S2.1 to S2.3). Bacterial genera such as *Microcystis, Flavobacterium, Sediminibacterium,* and *Fluviicola* were also relatively abundant in the water samples, however, the focus here is on bacteria that are fecal-associated and/or potentially pathogenic. In fecal samples, only 16 phyla were observed of which Bacteroidetes, Firmicutes, and Proteobacteria had the highest relative abundance together constituting 90% of the dataset (Fig. 2). At the genus level, 143 genera were observed but *Prevotella,* *Bacteroides,* and *Lactobacillus* were the predominant genera in fecal samples (Table S2.4). In wastewater samples, 20 different phyla were observed but the phylum Proteobacteria had the highest relative abundance followed by Bacteroidetes, Chloroflexi, and Acidobacteria.Genera of the Proteobacteria such as *Dechloromonas* and *Arcobacter* were relatively abundant in wastewater samples (Table S2.4).

*3.3. Comparison of bacterial community structure between water and fecal samples*

Hierarchical cluster analysis was performed to identify the similarities between different water and fecal samples (β-diversity). When water samples were compared by season, those from location 2 (WIW02) and 3 (WIW03) collected during winter clustered separately from other water samples indicating that the bacterial composition was different in these samples (Fig. 3A). As stated previously, the phylum and genus level bacterial composition at these locations was significantly different from those of other water samples (Fig. 2 and Table S2.1 to S2.3). The β-diversity analyzed by PCoA also revealed that these two samples clustered separately from other water samples (Fig. 4A). The results of β-diversity analysis also indicated that significant seasonal variation was apparent for water samples that formed distinct seasonal clusters, though few autumn and summer water samples were closely related (Fig. 3A and 4A). When comparison was carried out between water, fecal and wastewater samples, fecal samples clustered separately demonstrating that the bacterial composition of fecal samples is distinct from water and wastewater samples (Fig.3B-D & Fig.4B-D). However, WIW02 and WIW03 water samples clustered close to wastewater (WW) and fecal samples, indicating the strong presence of fecal related microbiota in these samples.

Venn diagram analysis was performed with unambiguous OTUs for better analysis of shared and specific OTUs present in water, wastewater and fecal samples (Fig. 5A-D). When three season water samples were analyzed, the results indicated that 42.7% of OTUs were shared between water samples, however, winter season water samples (WIW) had the highest number of specific OTUs (22.1%) compared to autumn (AUW) (8.4%) and summer (SUW) (7.8%) (Fig. 5A); 39.1%, 41.6%, and 38.7% of OTUs from autumn, winter, and summer season water samples were shared with the OTUs of total fecal samples (Fig.5B), indicating fecal contamination of the watercourse. Similarly, when OTUs from individual fecal sources were compared with total water samples, the results showed that 4.5% from human, 7.1% from pig and 9.8% from chicken fecal samples were shared with OTUs of total water samples (Fig. 5C). The analysis was also performed with fecal and wastewater samples and the results revealed that 14%, 40% and 57% of OTUs were host specific for human, chicken and pig fecal samples (Fig. 5D). These specific OTUs could be useful as potential targets in developing host-specific fecal indicator bacteria or markers for future studies. The wastewater samples showed shared OTUs (17.9%) with the three fecal samples (Fig. 5D).

*3.4. Relative abundance of fecal-associated or potentially pathogenic bacterial genera in Tiaoxi River water samples*

Although 180 different genera were recognized in Tiaoxi River water samples, this study was mainly focused on the genera that are associated with fecal sources and/or potential pathogens in nature. The fecal-associated genera detected in Tiaoxi River water samples are *Bacteroides*, *Prevotella*, *Blautia*, *Faecalibacterium*, *Dorea,* and *Macellibacteroides*. *Bacteroides* and *Prevotella*, which are Gram-negative obligate anaerobic bacteria that are in the human and animal gut (Layton et al., 2006; Lee et al., 2011), were the most frequently detected fecal associated bacteria (39 and 35 samples out of 45 samples tested) in Tiaoxi River water samples (Fig. 6A &B). The relative abundance of *Bacteroides* was comparatively high in the winter season and the highest was observed at WIW02 (2.6%) and WIW03 samples (2.4%) (Fig. 6A), while the highest relative abundance of *Prevotella* was observed at location 16 during the autumn season (Fig.6B). Similarly, *Blautia* and *Faecalibacterium,* whichareobligate anaerobic Gram-positive bacteriathat are primarily present in mammalian gut (Garcia-Mazcorro et al., 2012; Miquel et al., 2013), were detected in several samples (26 and 19 out of 45 samples tested) of Tiaoxi River water and the highest relative abundance of each detected in WIW02 (0.10 and 0.038%) and WIW03 (0.11 and 0.028%) samples (Fig. 6C&D). The genus *Dorea*, which comprises Gram-positive obligate anaerobic bacteria present mainly in human feces (Taras et al., 2002), was detected in 14 out of 45 water samples of Tiaoxi River and the highest relative abundance was detected at WIW02 (0.06%) and WIW03 (0.04%) water samples (Fig. 7A). Additionally, a few more human-associated fecal bacteria such as *Parabacteroides* and *Bifidobacterium* were also detected with low relative abundance at different locations of Tiaoxi River (Tables S2.1 to S2.3) indicating potential human contamination at these locations. Finally, *Macellibacteroides* that belongs to *Bacteroidales* and comprises strict anaerobic Gram-positive bacteria mostly isolated from WWTPs (Jabari et al., 2012), were also detected in many locations (27 out of 45 samples) (Fig. 7B). However, *Macellibacteroides* has been detected in goose fecal samples of the current study (Table S2.4) and more frequently detected at locations 4, 5, 6 and 16 of Tiaoxi River in all seasons. Furthermore, the sequence data from fecal sources in this study revealed that the genus *Faecalibacterium* is abundantly present in human fecal samples, as opposed to *Bacteroides* that are abundant in both humans and animals, suggesting that *Faecalibacterium* would be a useful target for designing human-specific MST markers.

The sequence data were further analyzed to assess the presence of potentially pathogenic bacterial genera in Tiaoxi River water samples by comparing with known pathogenic bacteria from the database of Pathosystems Resource Integration Center (PATRIC) (Wattam et al., 2014). In total, 14 potentially pathogenic bacterial genera were detected in the water samples (Table S2.1 to S2.3). However, only seven (*Acinetobacter, Aeromonas, Arcobacter, Brevundimonas, Enterococcus, Escherichia-Shigella,* and *Streptococcus*) showed relative abundance greater than 0.1 at different locations, and these data are presented in Table 1. Although the sequences were not reliably classified to species level, the highest relative abundance of *Acinetobacter* was observed at location 2 and *Aeromonas* at location 5 and as specified earlier, these locations have higher levels of human and animal associated fecal bacterial genera. The genus *Arcobacter* detected in this study showed 97% identity to *A. cryaerophilus,* which is an emerging enteropathogen for humans and animals (Fernandez et al., 2015)*. Arcobacter* was one of the most frequently (9 out of 15 locations with above 0.1% relative abundance) detected pathogens in Tiaoxi River and the highest relative abundance (1.3%) was observed at location 12. *Brevundimonas* was the most commonly detected potential pathogen in Tiaoxi River. It was found with higher relative abundance (>0.10) at 10 monitoring locations and the highest relative abundance (3.9%) was detected at location-1 (Taihu Lake) during the winter season. The sequences of this genus were identified as *Brevundimonas alba, B. bullata* and *B. vesicularis*, of which the latter is an opportunistic pathogen associated with bacteremia (Ryan and Pembroke, 2018; Zhang et al., 2012). Although *Enterococcus and Escherichia-Shigella* were detected with high relative abundance at several locations,and *Streptococcus* in a few locations, they were not reliably classified to species level with 97% similarity cutoff. However, all the seven potential pathogenic bacterial genera with relative abundance >0.1 were detected at location 3 and six of these with higher abundance were observed at location 2 and 4 emphasizing that these locations could constitute a human health risk.

*3.5. Relationship between environmental parameters and bacterial community composition*

RDA analysis was carried out to determine the influence of environmental parameters (pH, temperature, conductivity, TN, TP, NO2-N and NH4-N) on the bacterial community composition observed at different locations and seasons. For autumn 2014, RDA1 and RDA2 explained 26.92% and 14.72% of the total variation respectively and the RDA biplot indicated that pH, temperature, conductivity, TP, and TN had major influence on the bacterial community composition (Fig. 8A). In particular, bacterial community composition at locations 3 and 12 was mainly influenced by pH, conductivity, and TP. In winter 2015, the pattern was different and the bacterial community composition was mainly influenced by pH and conductivity. The RDA1 and RDA2 contributed 44.7% and 16.4% of the total variation during winter 2015 (Fig. 8B). The RDA biplot pattern was entirely different for samples collected during summer 2015 (Fig. 8C). The RDA1 and RDA2 contributed 41.9% and 10.2% of total variation in the community composition, with temperature, conductivity, pH, NH4-N, and NO2-N as the major influencing environmental factors. The overall analysis indicated that pH, conductivity and temperature were the main environmental factors that showed strong influence on the bacterial community composition, although some of the nutrients showed seasonal variation.

**4. Discussion**

Recent advances in NGS technologies coupled with reduced cost has enabled the application of microbial community analysis for monitoring of microbial quality and diversity in the aquatic environment and profiling the microbiota associated with fecal samples (Ley et al., 2008; Marti et al., 2017; Newton et al., 2011; Vierheilig et al., 2015). In this study, bacterial communities in water, fecal and wastewater samples were studied by Illumina sequencing by targeting the V4 hypervariable region of the 16S rRNA genes. The average read length generated by Illumina sequencing after trimming was about 250bps and reads were processed for assigning taxonomy. The diversity indices for bacterial communities associated with chicken, dog, duck, and human fecal samples were lower compared to other fecal, wastewater and Tiaoxi River samples (Table S1.1) due to the abundance of a relatively small number of taxa (Jeong et al., 2011). In chicken and duck fecal samples, Proteobacteria and Firmicutes were dominant and represented >90% of taxa; in dog and human fecal samples, Bacteroidetes and Firmicutes accounted for 90% of bacterial taxa. Ley et al. (2008) also reported the dominance of these bacterial taxa in fecal samples, with carnivores and omnivores (human, dog, duck, and chicken) having less diverse gut microbiomes compared to herbivores. The high diversity values observed in wastewater samples here are consistent with previous studies (Newton et al., 2013; Shanks et al., 2013) as diverse bacteria from human feces and other environmental sources are released into wastewater (Newton et al., 2013). Moreover, the wastewater contains high amount of nutrients, which support the growth of diverse microbial populations (Arfken et al., 2015; McLellan et al., 2013). Although the hierarchal cluster analysis results indicated that human fecal samples are closely related to dog fecal samples (Fig. 3B-D), the Venn diagram analysis of human fecal samples with pig and chicken fecal samples (as they are common livestock in the study area) revealed that human fecal samples harbor microbes that are closely related to those in pig feces (Fig.5D). This could be due to more shared OTUs from the genera *Bacteroides* and *Prevotella* present in human and pig fecal samples (Dick et al., 2005).

Many species of *Bacteroides, Prevotella,* and *Blautia* are abundant in the mammalian gut and have been advocated as fecal indicators (Newton et al., 2011; Savichtcheva and Okabe, 2006). Newton et al. (2011) studied the bacterial community in sewage samples by NGS based 16S rRNA gene sequencing, and the results revealed the presence of the human fecal-associated bacterial genus *Blautia* in the samples. Species of *Bacteroides* have host specificity and limited survival in the environment, making them ideal indicators of recent fecal contamination and they are often used for microbial source tracking (MST) (Bernhard and Field, 2000; Layton et al., 2006). In the present study, the *Bacteroides* detected in Tiaoxi River water samples were classified to species level; *Bacteroides plebeius, B. propionicifaciens, B. massiliensis, B. graminisolvens, B. nordii, B. stercoris, B. caccae,* and *B. paurosaccharolyticus*. *B. caccae* and *B. plebeius* have all been isolated from human feces (Kitahara et al., 2005; Wei et al., 2001) and their presence is therefore indicative. These species were frequently detected with high relative abundance (OTUs) at location 2 and 3 of Tiaoxi River (Tables S3.1 to S3.3) and the highest was observed in the samples collected during winter (WIW02 and WIW03 samples). During winter sampling (Jan/Feb 2015), the rainfall was relatively low (only 66 mm precipitation, NBSC (2016)) and fecal contamination at these locations could be due to direct discharge of sewer and septic waste rather than runoff (Ohad et al., 2015). The samples collected at location 3 were close to a fishing village where people live on boats without proper sanitation facilities (Vadde et al., 2019) and the high relative abundance of *Bacteroides* could be associated with the entry of sewage at this location. Based on land use pattern, the presence of *Bacteroides* with higher relative abundance at location 2 could be associated with the transport of these bacteria from location 3 (Marti et al., 2013). The presence of *Parabacteroides merdae* at locations 2 and 3 also confirms potential human fecal contamination at these locations (McLuskey et al., 2016). In the present study, *B. propionicifaciens* was found only in goose and duck fecal samples and its frequent detection at location 6 of the Tiaoxi River water samples indicates avian fecal contamination. The presence of *B. massiliensis*, *B. nordii* and *B. stercoris* with high relative abundance at location 2 and 3 points to potential human risks, as these species are associated with anaerobic bacteremia and abdominal infections (Fenner et al., 2005; Otte et al., 2017; Song et al., 2004). *B. graminisolvens* and *B. paurosaccharolyticus* were frequently detected at location 4, 5, 15 and 16 which are located close to WWTPs (Zheng et al., 2017). Therefore, these data can be added to previous studies, which reported that these *Bacteroides* species originate from effluents of WWTPs (Nishiyama et al., 2009; Ueki et al., 2011), although we can now add that avian (goose) fecal contamination is an additional source. Overall, the species level identification of *Bacteroides* matches with the land use pattern in the current study.

The remaining fecal associated bacterial genera (*Prevotella, Blautia, Faecalibacterium,* and *Dorea)* detected in this study were not reliably classified to species level due to the limitations of using short reads of 16S rRNA generated by NGS (Nguyen et al., 2016). However, *Prevotella* was detected at 7 locations (location 4, 5, 6, 12, 14, 16 and 21) in three seasons (Fig. 3.5B) indicating that these locations are potentially polluted with human or animal fecal contamination (Lee et al., 2011). Locations 4, 5 and 16 are situated near WWTPs (Zheng et al., 2017) and at location 6 and 21, several household backyard and commercial farms for poultry and pigs were observed during sampling. Locations 12 and 14 are urban or suburban areas and the presence of *Prevotella* at these locations could be due to sewage entry or transport of bacteria from upstream locations such as location 13 where poultry and pigs farms are located (Vadde et al., 2018). While studying the fecal contaminations in an urban River, Newton et al. (2011) noticed an increase in fecal indicators and human-associated sequences after heavy rainfall and triggering of combined sewer overflows (CSOs). The fecal indicator bacteria *Blautia, Faecalibacterium* and *Dorea* were higher at locations 2 and 3 during the winter season (Figs. 6C, D & 7A), highlighting the presence of fecal contamination at these locations again. Overall, the presence of fecal indicator bacteria at different locations of Tiaoxi River water emphasizes the need for application of MST techniques to determine the source of fecal pollution at Tiaoxi River. Additionally, analysis of the sequence data of fecal samples from this study suggests that *Faecalibacterium* could be a useful target for designing human-specific MST markers. A recent study (Sun et al., 2016) also provided initial evidence to support the use of *Faecalibacterium* as a human-specific MST marker. Therefore, development of MST qPCR assay targeting this genus could provide a more specific human-associated MST assay. Furthermore, *Macellibacteroides* species, which were isolated from WWTPs in previous studies (Jabari et al., 2012), were also identified in goose fecal samples (Table S2.4), indicating geese as an additional source of these genera.

As fecal indicator bacteria were observed at several locations, further analysis was performed to detect potentially pathogenic bacteria, and *Acinetobacter, Aeromonas, Arcobacter, Brevundimonas, Enterococcus, Escherichia-Shigella,* and *Streptococcus* were accordingly found to be present in Tiaoxi River water (Table 1; Tables S3.1 to 3.4). Several species of *Acinetobacter* and *Aeromonas* are considered as opportunistic pathogens causing nosocomial infections and gastroenteritis (Khosravi et al., 2015; Laukova et al., 2018). However, *Acinetobacter* has more than 20 species, only three of which are considered as pathogens (Shamsizadeh et al., 2017); *Aeromonas* has 21 species of which three are mainly considered as pathogens (Janda and Abbott, 2010). Therefore, the sequence data were carefully examined again and only two could be identified to species level: *A. cryaerophilus* and *B. vesicularis*. The former causes acute to chronic diarrhea in humans and was frequently detected at locations 12 and 16. In the case of *A.* *cryaerophilus*, they are often confused with *C. jejuni* and therefore unambiguous monitoring methods such as virulence gene detection or quantification of these bacteria are required for accurate identification (Figueras et al., 2014). With respect to *B. vesicularis,* very few OTUs were observed in Tiaoxi River water (Tables S3.1 to S3.3) and the remaining potentially pathogenic genera such as *Acinetobacter, Aeromonas, Enterococcus, Escherichia-Shigella*, and *Streptococcus* were not classified to species level at 97% similarity. However, most of these potential pathogenic bacterial genera were detected with higher relative abundance (>0.1) at locations 2, 3 and 4, which were highly contaminated with feces. Therefore, NGS has proved to be very useful for preliminary assessment of a wide-range of fecal and potentially pathogenic bacteria present in environmental samples. Application of this technology enables prioritization of samples for further in-depth quantification of MST markers and genes of bacterial pathogens by methods such as qPCR in order to establish the risk to human health.

**5. Conclusions**

A total of 20 different phyla were observed in most of the water samples of Tiaoxi River and wastewater samples, while only 16 phyla were detected in fecal samples. Hierarchical cluster analysis and PCoA performed for fecal, wastewater and water samples showed that fecal and wastewater samples clustered separately from water. Venn diagrams revealed that chicken fecal samples (9.8%) shared the highest number of OTUs with total water samples, followed by pig (7.1%), and human samples (4.5%) indicating the presence of avian, pig and human fecal contamination in Tiaoxi River. Presence of five bacterial genera associated with fecal sources (*Bacteroides, Prevotella, Blautia, Faecalibacterium*, and *Dorea*) at several locations indicates human and animal fecal contamination in these locations. Similarly, the presence of seven potentially pathogenic genera (*Acinetobacter, Aeromonas, Arcobacter, Brevundimonas, Enterococcus, Escherichia-Shigella,* and *Streptococcus*) at relative abundance >0.1 in several locations of Tiaoxi river indicates potential health risk. Overall, the results indicate that 16S rRNA genes targeted NGS is a valuable tool to screen for a wide variety of bacteria including pathogens and those of fecal origin as an initial step to identify human health risk and to prioritize sites for further assessment using more specific methods.

**Conflicts of Interests**

The authors declare no conflicts of interest.

**Acknowledgments**

The authors acknowledge the Natural Science Foundation of Jiangsu Province (Grant No. BK20141211), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No. 13KJB180022) and Key Program Special Fund in Xi’an Jiaotong-Liverpool University (XJTLU; Grant No. KSF-E-20) for financial support. We would like to thank the Department of Biological Sciences, XJTLU and Suzhou Institute for Advanced Study, University of Science and Technology (USTC) for providing facilities. Financial support to KKV was provided through the Postgraduate Research Scholarship (PGRS-12-01-08) awarded by XJTLU. Ethical approval for this study was provided by XJTLU Research Ethics Committee (EXT-16-07-01).

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/

**References:**

Ahmed, W., Sawant, S., Huygens, F., Goonetilleke, A., Gardner, T., 2009. Prevalence and occurrence of zoonotic bacterial pathogens in surface waters determined by quantitative PCR. Water Res 43, 4918-4928.

Ahmed, W., Staley, C., Sadowsky, M.J., Gyawali, P., Sidhu, J.P., Palmer, A., Beale, D.J., Toze, S., 2015. Toolbox Approaches Using Molecular Markers and 16S rRNA Gene Amplicon Data Sets for Identification of Fecal Pollution in Surface Water. Appl Environ Microbiol 81, 7067-7077.

Arfken, A.M., Song, B., Mallin, M.A., 2015. Assessing hog lagoon waste contamination in the Cape Fear Watershed using Bacteroidetes 16S rRNA gene pyrosequencing. Applied Microbiology and Biotechnology 99, 7283-7293.

Bernhard, A.E., Field, K.G., 2000. A PCR assay To discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. Appl Environ Microbiol 66, 4571-4574.

Boehm, A.B., Van De Werfhorst, L.C., Griffith, J.F., Holden, P.A., Jay, J.A., Shanks, O.C., Wang, D., Weisberg, S.B., 2013. Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. Water Res 47, 6812-6828.

Brown, C.M., Staley, C., Wang, P., Dalzell, B., Chun, C.L., Sadowsky, M.J., 2017. A High-Throughput DNA-Sequencing Approach for Determining Sources of Fecal Bacteria in a Lake Superior Estuary. Environ Sci Technol 51, 8263-8271.

Cai, H.Y., Yan, Z.S., Wang, A.J., Krumholz, L.R., Jiang, H.L., 2013. Analysis of the attached microbial community on mucilaginous cyanobacterial aggregates in the eutrophic Lake Taihu reveals the importance of Planctomycetes. Microbial Ecology 66, 73-83.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7, 335-336.

Chen, Y., Zhao, K., Wu, Y., Gao, S., Cao, W., Bo, Y., Shang, Z., Wu, J., Zhou, F., 2016. Spatio-temporal patterns and source identification of water pollution in lake Taihu (China). Water 8, 86.

Dick, L.K., Bernhard, A.E., Brodeur, T.J., Santo Domingo, J.W., Simpson, J.M., Walters, S.P., Field, K.G., 2005. Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. Appl Environ Microbiol 71, 3184-3191.

Dubinsky, E.A., Butkus, S.R., Andersen, G.L., 2016. Microbial source tracking in impaired watersheds using PhyloChip and machine-learning classification. Water Res 105, 56-64.

Fenner, L., Roux, V., Mallet, M.N., Raoult, D., 2005. Bacteroides massiliensis sp. nov., isolated from blood culture of a newborn. Int J Syst Evol Microbiol 55, 1335-1337.

Fernandez, H., Villanueva, M.P., Mansilla, I., Gonzalez, M., Latif, F., 2015. Arcobacter butzleri and A. cryaerophilus in human, animals and food sources, in southern Chile. Braz J Microbiol 46, 145-147.

Figueras, M.J., Levican, A., Pujol, I., Ballester, F., Rabada Quilez, M.J., Gomez-Bertomeu, F., 2014. A severe case of persistent diarrhoea associated with Arcobacter cryaerophilus but attributed to Campylobacter sp. and a review of the clinical incidence of Arcobacter spp. New Microbes New Infect 2, 31-37.

Garcia-Mazcorro, J.F., Dowd, S.E., Poulsen, J., Steiner, J.M., Suchodolski, J.S., 2012. Abundance and short-term temporal variability of fecal microbiota in healthy dogs. Microbiologyopen 1, 340-347.

Green, H.C., Haugland, R.A., Varma, M., Millen, H.T., Borchardt, M.A., Field, K.G., Walters, W.A., Knight, R., Sivaganesan, M., Kelty, C.A., Shanks, O.C., 2014. Improved HF183 Quantitative Real-Time PCR Assay for Characterization of Human Fecal Pollution in Ambient Surface Water Samples. Applied and Environmental Microbiology 80, 3086-3094.

Guo, F., Ju, F., Cai, L., Zhang, T., 2013. Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. Plos One 8, e76185.

Hagedorn, C., Liang, X., 2011. Current and future trends in fecal source tracking and deployment in the Lake Taihu Region of China. Phys Chem Earth 36, 352-359.

Han, D., Currell, M.J., Cao, G., 2016. Deep challenges for China's war on water pollution. Environ Pollut 218, 1222-1233.

Ibekwe, A.M., Leddy, M., Murinda, S.E., 2013. Potential Human Pathogenic Bacteria in a Mixed Urban Watershed as Revealed by Pyrosequencing. PLoS One 8.

Jabari, L., Gannoun, H., Cayol, J.L., Hedi, A., Sakamoto, M., Falsen, E., Ohkuma, M., Hamdi, M., Fauque, G., Ollivier, B., Fardeau, M.L., 2012. Macellibacteroides fermentans gen. nov., sp. nov., a member of the family Porphyromonadaceae isolated from an upflow anaerobic filter treating abattoir wastewaters. Int J Syst Evol Microbiol 62, 2522-2527.

Janda, J.M., Abbott, S.L., 2010. The genus Aeromonas: taxonomy, pathogenicity, and infection. Clinical Microbiology Reviews 23, 35-73.

Jeong, J.Y., Park, H.D., Lee, K.H., Weon, H.Y., Ka, J.O., 2011. Microbial community analysis and identification of alternative host-specific fecal indicators in fecal and river water samples using pyrosequencing. J Microbiol 49, 585-594.

Jiang, Y., 2009. China's water scarcity. Journal of Environmental Management 90, 3185-3196.

Jiang, Y., 2015. China’s water security: Current status, emerging challenges and future prospects. Environmental Science & Policy 54, 106–125.

Kildare, B.J., Leutenegger, C.M., McSwain, B.S., Bambic, D.G., Rajal, V.B., Wuertz, S., 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. Water Res 41, 3701-3715.

Kitahara, M., Sakamoto, M., Ike, M., Sakata, S., Benno, Y., 2005. Bacteroides plebeius sp. nov. and Bacteroides coprocola sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 55, 2143-2147.

Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., Kelley, S.T., 2011. Bayesian community-wide culture-independent microbial source tracking. Nat Methods 8, 761-763.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79, 5112-5120.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R., Sayler, G., 2006. Development of Bacteroides 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol 72, 4214-4224.

Lee, J.E., Lee, S., Sung, J., Ko, G., 2011. Analysis of human and animal fecal microbiota for microbial source tracking. ISME J 5, 362-365.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008. Evolution of mammals and their gut microbes. Science 320, 1647-1651.

Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., Pallen, M.J., 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30, 434-439.

Lozupone, C., Knight, R., 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71, 8228-8235.

Marti, R., Gannon, V.P.J., Jokinen, C., Lanthier, M., Lapen, D.R., Neumann, N.F., Ruecker, N.J., Scott, A., Wilkes, G., Zhang, Y., Topp, E., 2013. Quantitative multi-year elucidation of fecal sources of waterborne pathogen contamination in the South Nation River basin using Bacteroidales microbial source tracking markers. Water Research 47, 2315-2324.

Marti, R., Ribun, S., Aubin, J.B., Colinon, C., Petit, S., Marjolet, L., Gourmelon, M., Schmitt, L., Breil, P., Cottet, M., Cournoyer, B., 2017. Human-Driven Microbiological Contamination of Benthic and Hyporheic Sediments of an Intermittent Peri-Urban River Assessed from MST and 16S rRNA Genetic Structure Analyses. Frontiers in Microbiology 8, 19.

McLellan, S.L., Newton, R.J., Vandewalle, J.L., Shanks, O.C., Huse, S.M., Eren, A.M., Sogin, M.L., 2013. Sewage reflects the distribution of human faecal Lachnospiraceae. Environmental Microbiology 15, 2213-2227.

McLuskey, K., Grewal, J.S., Das, D., Godzik, A., Lesley, S.A., Deacon, A.M., Coombs, G.H., Elsliger, M.A., Wilson, I.A., Mottram, J.C., 2016. Crystal Structure and Activity Studies of the C11 Cysteine Peptidase from Parabacteroides merdae in the Human Gut Microbiome. J Biol Chem 291, 9482-9491.

MEP, 2016. Ministry of Environmental Protection of the People's Republic of China. China environmental Bulletin in 2015., China Environmental Science Press.

Mieszkin, S., Furet, J.P., Corthier, G., Gourmelon, M., 2009. Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. Appl Environ Microbiol 75, 3045-3054.

Miquel, S., Martin, R., Rossi, O., Bermudez-Humaran, L.G., Chatel, J.M., Sokol, H., Thomas, M., Wells, J.M., Langella, P., 2013. Faecalibacterium prausnitzii and human intestinal health. Curr Opin Microbiol 16, 255-261.

NBSC, 2016. National Bureau of Statistics of China.

Newton, R.J., Bootsma, M.J., Morrison, H.G., Sogin, M.L., McLellan, S.L., 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized coast of Lake Michigan. Microbial Ecology 65, 1011-1023.

Newton, R.J., McLellan, S.L., Dila, D.K., Vineis, J.H., Morrison, H.G., Eren, A.M., Sogin, M.L., 2015. Sewage reflects the microbiomes of human populations. Mbio 6, e02574.

Newton, R.J., Vandewalle, J.L., Borchardt, M.A., Gorelick, M.H., McLellan, S.L., 2011. Lachnospiraceae and Bacteroidales alternative fecal indicators reveal chronic human sewage contamination in an urban harbor. Appl Environ Microbiol 77, 6972-6981.

Nguyen, N.P., Warnow, T., Pop, M., White, B., 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. NPJ Biofilms Microbiomes 2, 16004.

Nishiyama, T., Ueki, A., Kaku, N., Watanabe, K., Ueki, K., 2009. Bacteroides graminisolvens sp. nov., a xylanolytic anaerobe isolated from a methanogenic reactor treating cattle waste. Int J Syst Evol Microbiol 59, 1901-1907.

Ohad, S., Vaizel-Ohayon, D., Rom, M., Guttman, J., Berger, D., Kravitz, V., Pilo, S., Huberman, Z., Kashi, Y., Rorman, E., 2015. Microbial Source Tracking in Adjacent Karst Springs. Appl Environ Microbiol 81, 5037-5047.

Oster, R.J., Wijesinghe, R.U., Haack, S.K., Fogarty, L.R., Tucker, T.R., Riley, S.C., 2014. Bacterial pathogen gene abundance and relation to recreational water quality at seven Great Lakes beaches. Environ Sci Technol 48, 14148-14157.

Otte, E., Nielsen, H.L., Hasman, H., Fuglsang-Damgaard, D., 2017. First report of metronidazole resistant, nimD-positive, Bacteroides stercoris isolated from an abdominal abscess in a 70-year-old woman. Anaerobe 43, 91-93.

Paerl, H.W., Xu, H., McCarthy, M.J., Zhu, G., Qin, B., Li, Y., Gardner, W.S., 2011. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy. Water Res 45, 1973-1983.

Qin, B., Xu, P., Wu, Q., Luo, L., Zhang, Y., 2007. Environmental issues of Lake Taihu, China. Hydrobiologia 581, 3-14.

Ryan, M.P., Pembroke, J.T., 2018. Brevundimonas spp: Emerging global opportunistic pathogens. Virulence 9, 480-493.

Savichtcheva, O., Okabe, S., 2006. Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Res 40, 2463-2476.

Schloss, P.D., 2009. A high-throughput DNA sequence aligner for microbial ecology studies. Plos One 4, e8230.

Shahryari, A., Nikaeen, M., Khiadani Hajian, M., Nabavi, F., Hatamzadeh, M., Hassanzadeh, A., 2014. Applicability of universal Bacteroidales genetic marker for microbial monitoring of drinking water sources in comparison to conventional indicators. Environmental Monitoring and Assessment 186, 7055-7062.

Shamsizadeh, Z., Nikaeen, M., Nasr Esfahani, B., Mirhoseini, S.H., Hatamzadeh, M., Hassanzadeh, A., 2017. Detection of antibiotic resistant Acinetobacter baumannii in various hospital environments: potential sources for transmission of Acinetobacter infections. Environ Health Prev Med 22, 44.

Shanks, O.C., Newton, R.J., Kelty, C.A., Huse, S.M., Sogin, M.L., McLellan, S.L., 2013. Comparison of the microbial community structures of untreated wastewaters from different geographic locales. Appl Environ Microbiol 79, 2906-2913.

Sinclair, L., Osman, O.A., Bertilsson, S., Eiler, A., 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the illumina platform. Plos One 10, e0116955.

Song, Y.L., Liu, C.X., McTeague, M., Finegold, S.M., 2004. "Bacteroides nordii" sp. nov. and "Bacteroides salyersae" sp. nov. isolated from clinical specimens of human intestinal origin. J Clin Microbiol 42, 5565-5570.

Staley, C., Kaiser, T., Lobos, A., Ahmed, W., Harwood, V.J., Brown, C.M., Sadowsky, M.J., 2018. Application of SourceTracker for Accurate Identification of Fecal Pollution in Recreational Freshwater: A Double-Blinded Study. Environ Sci Technol 52, 4207-4217.

Staley, Z.R., Chase, E., Mitraki, C., Crisman, T.L., Harwood, V.J., 2013. Microbial water quality in freshwater lakes with different land use. Journal of Applied Microbiology 115, 1240-1250.

Tan, B., Ng, C., Nshimyimana, J.P., Loh, L.L., Gin, K.Y., Thompson, J.R., 2015. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. Frontiers in Microbiology 6, 1027.

Taras, D., Simmering, R., Collins, M.D., Lawson, P.A., Blaut, M., 2002. Reclassification of Eubacterium formicigenerans Holdeman and Moore 1974 as Dorea formicigenerans gen. nov., comb. nov., and description of Dorea longicatena sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 52, 423-428.

Ueki, A., Abe, K., Ohtaki, Y., Kaku, N., Watanabe, K., Ueki, K., 2011. Bacteroides paurosaccharolyticus sp. nov., isolated from a methanogenic reactor treating waste from cattle farms. Int J Syst Evol Microbiol 61, 448-453.

Unno, T., Jang, J., Han, D., Kim, J.H., Sadowsky, M.J., Kim, O.S., Chun, J., Hur, H.G., 2010. Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. Environ Sci Technol 44, 7777-7782.

Unno, T., Staley, C., Brown, C.M., Han, D., Sadowksy, M.J., Hur, H.G., 2018. Fecal Pollution: New Trends and Challenges in Microbial Source Tracking Using Next-Generation-Sequencing. Environmental Microbiology.

USEPA, 2012. United States Environmental Protection Agency. Water: Monitoring & Analysis., p. <https://archive.epa.gov/water/archive/web/html/vms59.html>.

Vadde, K., Wang, J., Cao, L., Yuan, T., McCarthy, A., Sekar, R., 2018. Assessment of Water Quality and Identification of Pollution Risk Locations in Tiaoxi River (Taihu Watershed), China. Water 10, 183.

Vadde, K.K., McCarthy, A.J., Rong, R., Sekar, R., 2019. Quantification of Microbial Source Tracking and Pathogenic Bacterial Markers in Water and Sediments of Tiaoxi River (Taihu Watershed). Frontiers in Microbiology 10.

Vierheilig, J., Savio, D., Ley, R.E., Mach, R.L., Farnleitner, A.H., Reischer, G.H., 2015. Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. Water Science and Technology 72, 1962-1972.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E.K., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M., Schulman, J., Stevens, R.L., Sullivan, D.E., Vonstein, V., Warren, A., Will, R., Wilson, M.J., Yoo, H.S., Zhang, C., Zhang, Y., Sobral, B.W., 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42, D581-591.

Wei, B., Dalwadi, H., Gordon, L.K., Landers, C., Bruckner, D., Targan, S.R., Braun, J., 2001. Molecular cloning of a Bacteroides caccae TonB-linked outer membrane protein identified by an inflammatory bowel disease marker antibody. Infect Immun 69, 6044-6054.

WHO, 2015. World Health Statistics, World Health Organization, Geneva, Switzerland.

WHO, 2017. World Health Organization: Diarrhoeal disease fact sheets.

Wilhelm, S.W., Farnsley, S.E., LeCleir, G.R., Layton, A.C., Satchwell, M.F., DeBruyn, J.M., Boyer, G.L., Zhu, G., Paerl, H.W., 2011. The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. Harmful Algae 10, 207-215.

Zhang, C.C., Hsu, H.J., Li, C.M., 2012. Brevundimonas vesicularis bacteremia resistant to trimethoprim-sulfamethoxazole and ceftazidime in a tertiary hospital in southern Taiwan. J Microbiol Immunol Infect 45, 448-452.

Zhao, D., Cao, X., Huang, R., Zeng, J., Shen, F., Xu, H., Wang, S., He, X., Yu, Z., 2017. The heterogeneity of composition and assembly processes of the microbial community between different nutrient loading lake zones in Taihu Lake. Appl Microbiol Biotechnol 101, 5913-5923.

Zheng, J., Gao, R., Wei, Y., Chen, T., Fan, J., Zhou, Z., Makimilua, T.B., Jiao, Y., Chen, H., 2017. High-throughput profiling and analysis of antibiotic resistance genes in East Tiaoxi River, China. Environ Pollut 230, 648-654.

**Figure Legends**

**Fig. 1.** Map of sampling locations in Tiaoxi river (Taihu watershed).

**Fig. 2.** Relative abundance of different bacterial phyla in Tiaoxi River water (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water), fecal (CK-Chicken; CW-Cow; DG-Dog; DU-Duck; GO-Goose; HU-Human; PG-Pig) and wastewater (WW) samples.

**Fig. 3.** Cluster analysis showing similarity in bacterial diversity in water samples collected in three seasons (A) and in combinations of diversity in water, fecal and wastewater samples (B, C, D). (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; CK-Chicken; CW-Cow; DG-Dog; DU-Duck; GO-Goose; HU-Human; PG-Pig; WW-Wastewater).

**Fig. 4.** Visualization of β-diversity among water, fecal and wastewater samples by principal coordinate analysis (PCoA). (A) Comparison between autumn, winter and summer water samples (B) Comparison of water samples (autumn and winter) with all fecal and wastewater samples (C) Comparison of water samples (autumn and summer) with all fecal and wastewater samples, (D) Comparison of water samples (winter and summer) with all fecal and wastewater samples (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; FEC-Fecal samples; CK-Chicken; CW-Cow; DG-Dog; DU-Duck; GO-Goose; HU-Human; PG-Pig; WW-Wastewater).

**Fig. 5.** Venn diagrams showing shared and unique OTUs identified when the sequences were compared between Autumn, Winter and Summer water samples (A), between water samples collected during three seasons and pooled fecal samples (B) and between pooled water samples and individual fecal samples (human, pig, and chicken) (C) and between fecal and wastewater samples (D) (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; WW-Wastewater; CK-Chicken; HU-Human; PG-Pig).

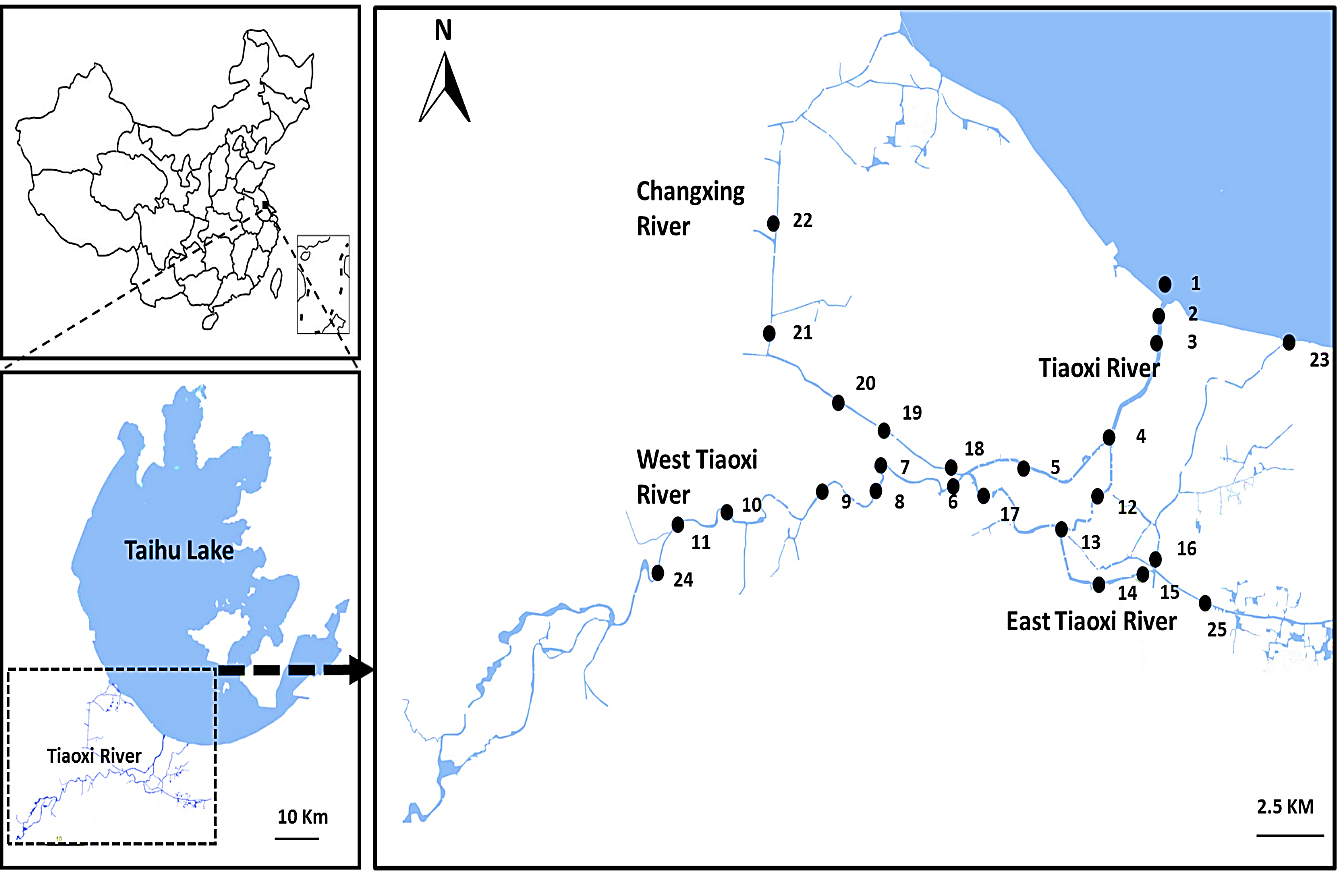
**Fig. 6.** Relative abundance of fecal-associated genera (A) *Bacteroides*, (B) *Prevotella*, (C) *Blautia* (D) *Faecalibacterium* observed at different sampling locations in Tiaoxi River.

**Fig. 7.** Relative abundance of fecal and sewage associated genera: (A) Dorea, (B) Macellibacteroides observed at different locations in Tiaoxi River.

**Fig. 8.** Redundancy analysis (RDA) showing a correlation of physico-chemical parameters with microbial community composition at different locations during three seasons: (A) Autumn 2014, (B) Winter 2015 and (C) Summer 2015. (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water).

**Table 1.** Range (Relative abundance percentage) of potential pathogenic bacterial genera detected in different locations of Tiaoxi River during three sampling occasions.

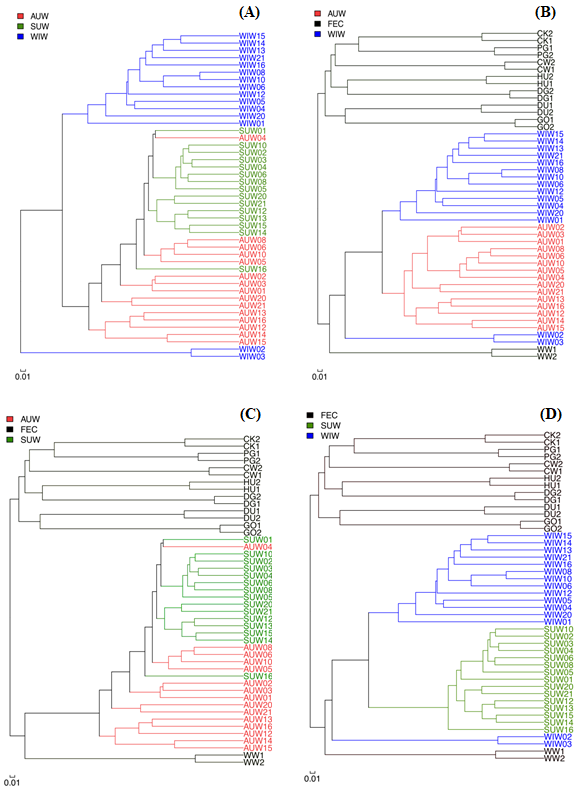
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Location** | ***Acinetobacter*** | ***Aeromonas*** | ***Arcobacter*** | ***Brevundimonas*** | ***Enterococcus*** | ***Escherichia-***  ***Shigella*** | ***Streptococcus*** | **No. of Potential pathogens (>0.1 abundance)** |
| **L-1** | 0.01-0.02 | 0.00-0.02 | 0.00-0.05 | **0.00-3.96** | 0.00-0.01 | 0.00-0.03 | 0.00-0.01 | 1 |
| **L-2** | **0.02-0.30** | 0.00-0.03 | **0.01-0.18** | **0.00-0.18** | **0.00-0.28** | **0.01-0.63** | **0.00-0.11** | **6** |
| **L-3** | **0.01-0.13** | **0.01-0.11** | **0.01-0.21** | **0.00-0.36** | **0.00-0.40** | **0.00-0.52** | **0.00-0.15** | **7** |
| **L-4** | **0.01-0.12** | **0.04-0.16** | **0.04-0.46** | **0.00-0.28** | **0.00-0.13** | **0.02-0.23** | 0.00-0.04 | **6** |
| **L-5** | 0.02-0.07 | **0.06-1.33** | **0.15-0.28** | 0.00-0.03 | 0.00-0.05 | **0.01-0.40** | 0.00-0.01 | 3 |
| **L-6** | 0.01-0.04 | 0.01-0.06 | **0.01-0.14** | 0.00-0.06 | **0.01-0.12** | **0.02-0.20** | 0.01-0.03 | 3 |
| **L-8** | 0.01-0.09 | 0.01-0.05 | 0.00-0.02 | 0.01-0.05 | 0.00-0.05 | 0.01-0.08 | 0.00-0.02 | 0 |
| **L-10** | 0.01-0.09 | 0.01-0.08 | 0.01-0.02 | 0.00-0.03 | 0.00-0.03 | 0.01-0.03 | 0.00-0.01 | 0 |
| **L-12** | 0.01-0.05 | **0.01-0.31** | **0.01-1.36** | **0.00-0.24** | 0.00-0.03 | 0.03-0.05 | 0.01-0.01 | 3 |
| **L-13** | 0.01-0.07 | 0.00-0.07 | **0.01-0.22** | **0.00-0.07** | 0.00-0.01 | 0.01-0.09 | N.D | 2 |
| **L-14** | 0.00-0.06 | **0.00-0.20** | 0.00-0.06 | **0.00-0.24** | 0.00-0.05 | 0.01-0.02 | 0.00-0.01 | 2 |
| **L-15** | **0.01-0.17** | **0.01-0.15** | 0.00-0.04 | **0.01-0.10** | **0.03-0.10** | 0.01-0.05 | 0.00-0.01 | 4 |
| **L-16** | 0.04-0.05 | **0.05-0.19** | **0.15-0.75** | **0.01-0.43** | 0.00-0.01 | **0.01-0.36** | 0.00-0.01 | 4 |
| **L-20** | 0.01-0.07 | 0.01-0.02 | 0.00-0.02 | 0.00-0.01 | **0.00-0.70** | 0.00-0.03 | N.D | 1 |
| **L-21** | 0.01-0.09 | 0.01-0.07 | **0.00-0.11** | **0.00-0.15** | **0.01-0.22** | 0.01-0.05 | N.D | 3 |



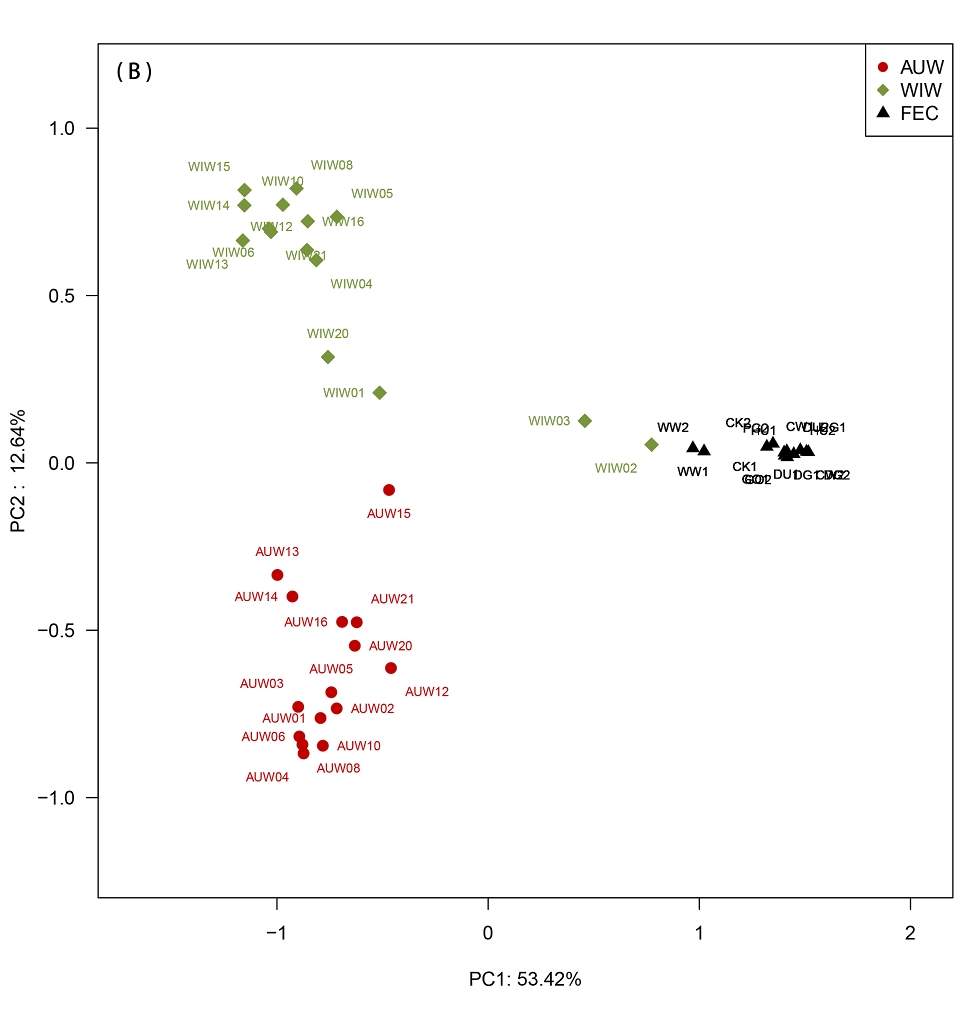
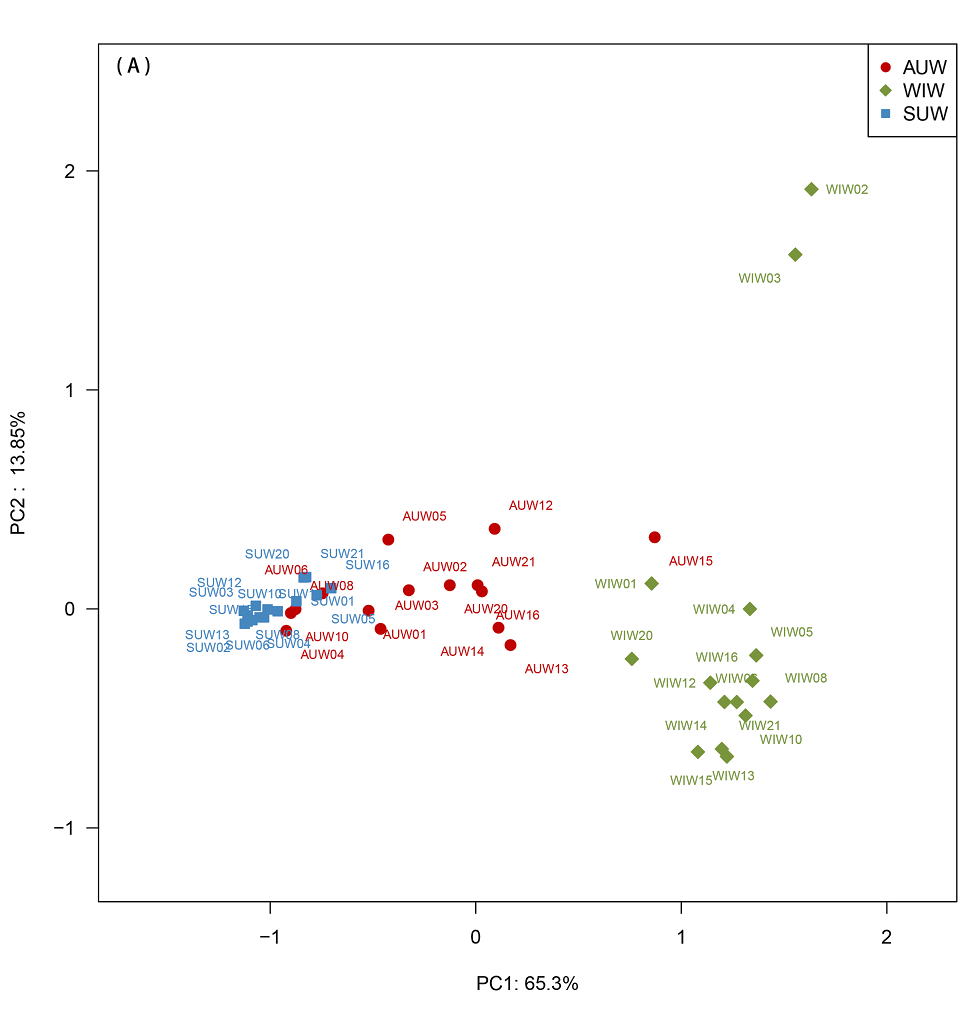
**Fig. 1.** Map of sampling locations in Tiaoxi River (Taihu watershed).



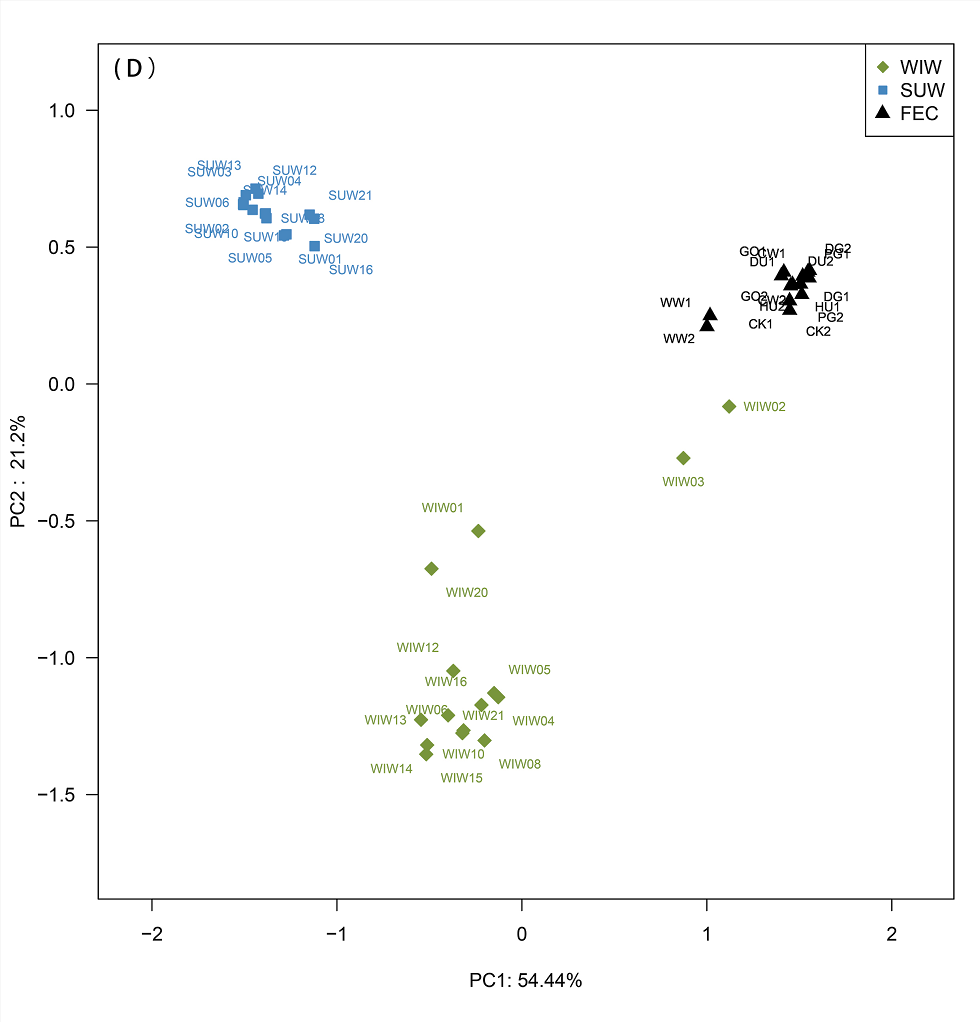
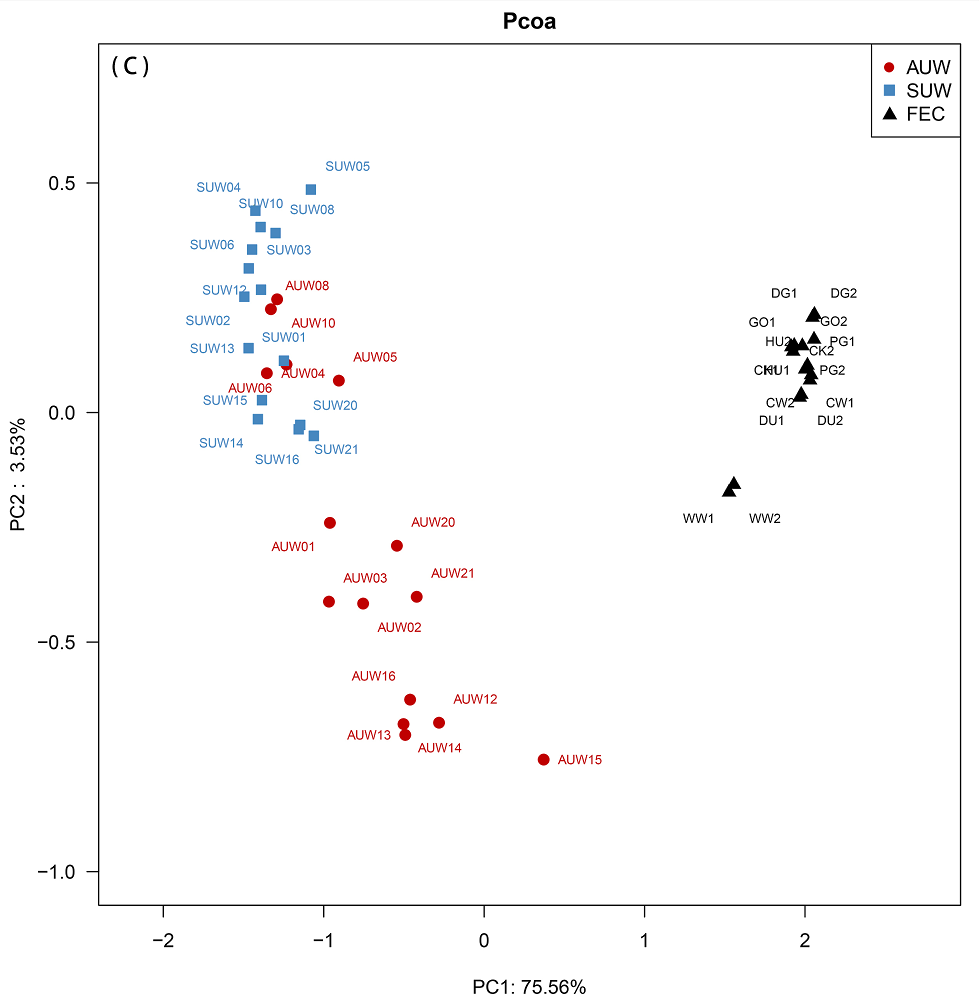
**Fig. 2.** Relative abundance of different bacterial phyla in Tiaoxi River water (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water), fecal (CK-Chicken; CW-Cow; DG-Dog; DU-Duck; GO-Goose; HU-Human; PG-Pig) and wastewater (WW) samples.



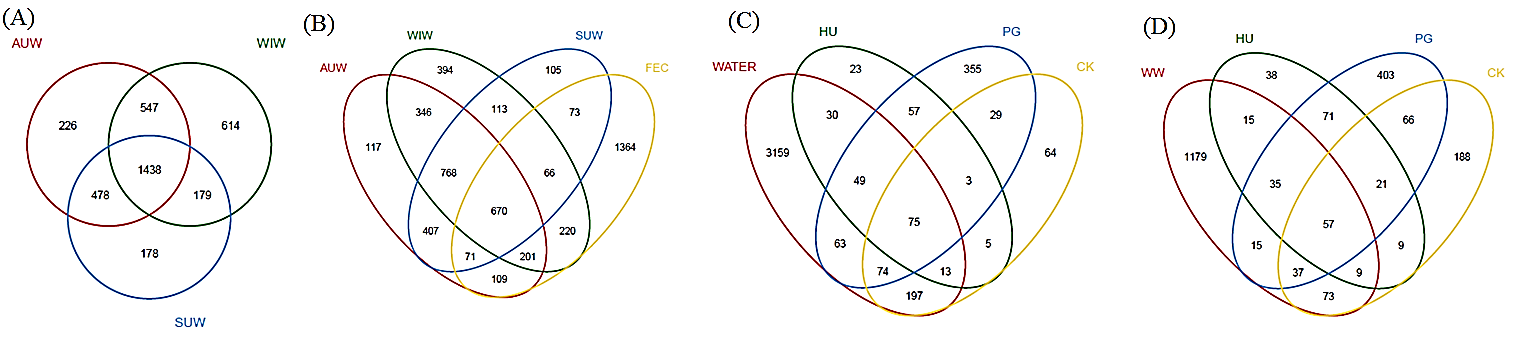
**Fig. 3.** Cluster analysis showing similarity in bacterial diversity in water samples collected in three seasons (A) and in combinations of diversity in water, fecal and wastewater samples (B, C, D). (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; CK-Chicken; CW-Cow; DG-Dog; DU-Duck; GO-Goose; HU-Human; PG-Pig; WW-Wastewater).



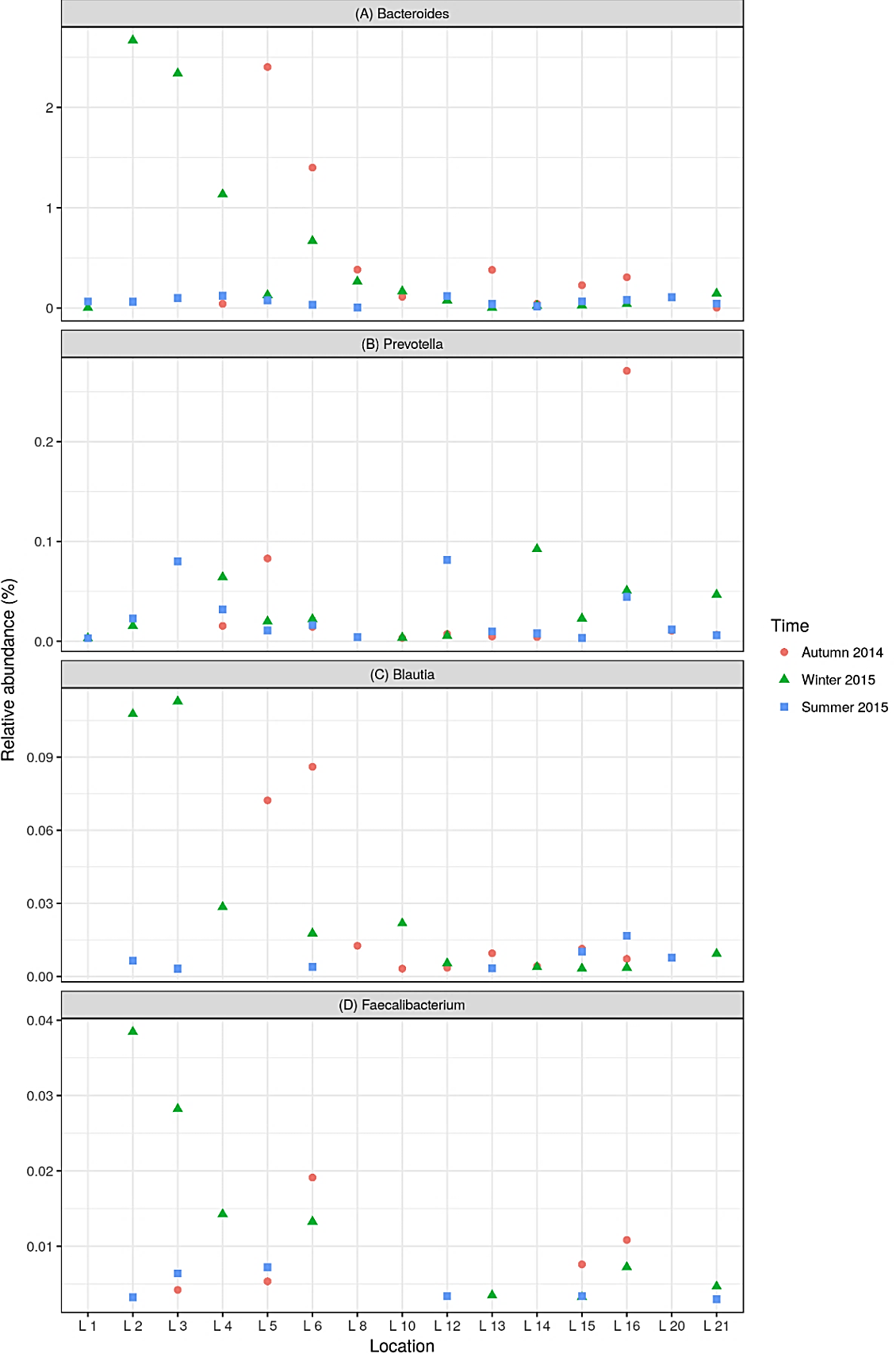
**Fig. 4.** Visualization of β-diversity among water, fecal and wastewater samples by principal coordinate analysis (PCoA). (A) Comparison between autumn, winter and summer water samples (B) Comparison of water samples (autumn and winter) with all fecal and wastewater samples (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; FEC-Fecal samples; CK-Chicken; CW-Cow; DG-Dog; DU-Duck; GO-Goose; HU-Human; PG-Pig; WW-Wastewater).

****

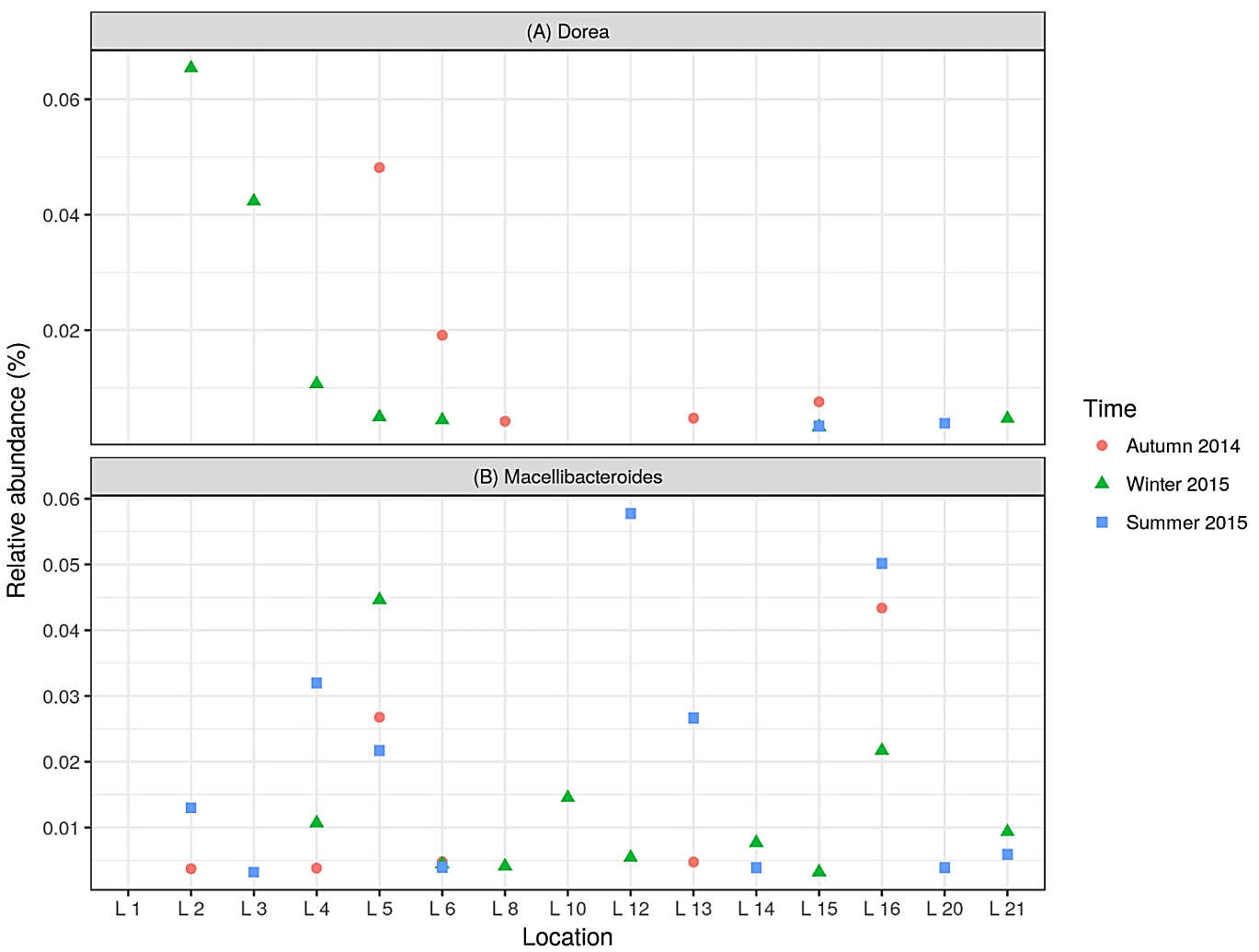
**Fig. 4.** Visualization of β-diversity among water, fecal and wastewater samples by principal coordinate analysis (PCoA). (C) Comparison of water samples (autumn and summer) with all fecal and wastewater samples, (D) Comparison of water samples (winter and summer) with all fecal and wastewater samples (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; FEC-Fecal samples; CK-Chicken; CW-Cow; DG-Dog; DU-Duck; GO-Goose; HU-Human; PG-Pig; WW-Wastewater).



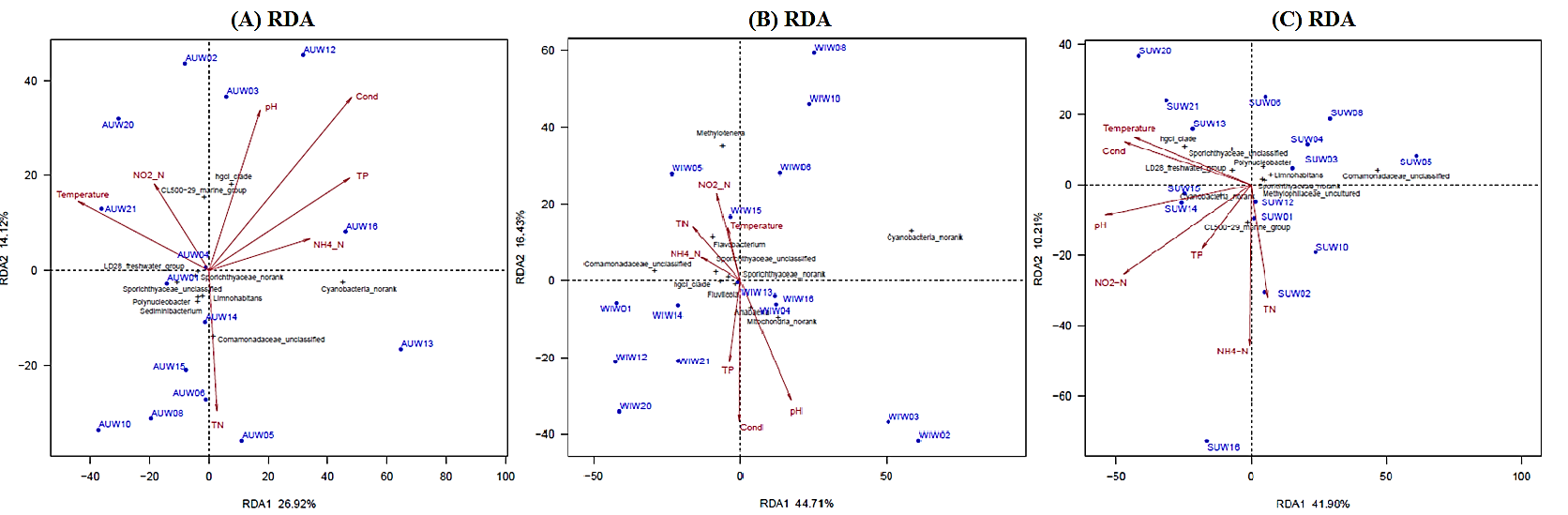
**Fig. 5.** Venn diagrams showing shared and unique OTUs identified when the sequences were compared between Autumn, Winter and Summer water samples (A), between water samples collected during three seasons and pooled fecal samples (B) and between pooled water samples and individual fecal samples (human, pig, and chicken) (C) and between fecal and wastewater samples (D) (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; WW-Wastewater; CK-Chicken; HU-Human; PG-Pig).



**Fig.6.**  Relative abundance of fecal-associated genera: (A) *Bacteroides*, (B) *Prevotella,* (C) *Blautia,* (D) *Faecalibacterium* observedat different sampling locations in Tiaoxi River.



**Fig. 7.** Relative abundance of fecal and sewage associated genera: (A) *Dorea,* (B) *Macellibacteroides* observedat different locations in Tiaoxi River*.*



**Fig. 8.** Redundancy analysis (RDA) showing a correlation of physico-chemical parameters with microbial community composition at different locations during three seasons: (A) Autumn 2014, (B) Winter 2015 and (C) Summer 2015. (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water).