Application of Microbial Source Tracking Techniques to Characterize Fecal Pollution entering Taihu Lake (China)

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Kiran Kumar Vadde

June 2018

UNIVERSITY OF LIVERPOOL

Abstract

Contamination of water bodies with human and animal fecal sources could significantly affect human health by disseminating pathogenic microorganisms. Therefore, accurate identification of sources that contribute to fecal pollution is vital for executing effective management strategies. Microbial source tracking (MST) is a promising approach to identify the sources of fecal contamination, which could be valuable for best management practices. The waterbody of interest, Taihu Lake, is one of the largest freshwater Lakes in China and serves as an important source for drinking water in addition to many other roles. Though this Lake is connected to several Rivers, Tiaoxi River provides most of the inflow (>60%). Previous reports indicated that Tiaoxi river is facing serious issues with fecal contamination and suggested implementing MST study. In this regard, 25 sampling locations were selected across the Tiaoxi River to monitor fecal contamination, and samples (water and sediment) were collected in three seasons (autumn 2014, winter and summer 2015). Physico-chemical and culture-based microbiological analysis of water samples were carried out for preliminary assessment of fecal contamination at these locations. The results showed that TN, TP, NO_2-N, and NH_4-N were the major nutrients that contributed to pollution in this River, and fecal coliform counts were high (>250 CFU/100 mL) in 15 locations indicating that a MST study was needed to ascertain sources of fecal contamination. Before applying MST, microbial community analysis was carried out in 45 water samples (collected from 15 locations in three seasons) to identify the diversity and composition of bacteria, including fecal and pathogenic bacteria, using Illumina high throughput sequencing. The Operational Taxonomic Units (OTUs) data comparison between total water samples with individual fecal sources indicated that chicken (9.8%), pig (7.1%), and human fecal samples (4.5%) have shared OTUs with total water samples, indicating the presence of avian, pig and human fecal contamination in this River. The genus level bacterial community data revealed that members of five fecal associated genera (*Bacteroides*, *Prevotella*, *Blautia*, *Faecalibacterium,* and *Dorea*) were present at several locations, pointing to human or animal fecal contamination in those locations. Furthermore, seven potential pathogenic bacterial genera namely *Acinetobacter, Aeromonas*, *Arcobacter*, *Brevundimonas*, *Enterococcus*, *Escherichia-Shigella*, and *Streptococcus* were also detected with high relative abundance (>0.1%), specific PCR assays are needed for accurate identification of their pathogenicity. As MST validation is required prior to its application in any new geographical area, a comprehensive evaluation of ten MST assays including two universal/general *Bacteroidales* (BacUni and GenBac3), four human-associated (HF183 SYBR, HF183 Taqman, BacHum and Hum2), one swine associated (Pig-2-Bac), one livestock/domestic animal associated (BacCow) and two avian associated MST qPCR assays (GFD and AV4143) targeting sewage, human and animal fecal DNA was carried out to determine the suitable MST assays for identifying fecal pollution sources at Taihu watershed. The

results showed that BacUni, HF183 Taqman, Pig-2-Bac and GFD markers were the best performers and are recommended for tracking total and host-associated fecal contamination in this region. The above evaluated MST markers were quantified in 15 locations (in water and sediment samples) of Tiaoxi River. Total *Bacteroidales* marker was detected in all the water and sediment samples, confirming the presence of fecal contamination. The human-associated marker was frequently detected at four locations at high concentrations (4.83-5.62 $\log_{10}$ copies/100ml) indicating that those locations were heavily contaminated with fecal pollution. Swine associated marker was frequently detected in samples from two locations and the avian associated marker was detected with high concentrations at 4 locations, correlating with the land use patterns and pointing to the entry of pig and avian fecal sources into Tiaoxi River. Among five bacterial pathogens monitored, *Campylobacter jejuni* was detected exceeding levels of lowest infection dose in 2 locations that are highly polluted with avian fecal source. Similarly, *Shigella* spp. were detected at two locations that are highly contaminated with human fecal sources, and Shiga toxin producing *E.coli* (STEC) at 2 locations that are contaminated with either human or pig fecal sources. The bacterial pathogen quantification results correlate with the findings of host associated fecal markers, demonstrating the potential of MST in predicting the presence of pathogenic organisms and the concomitant risk to human health.

// Acknowledgments

I have to thank many people for helping and supporting me in this four-year PhD journey.

Firstly, I am thankful to my Primary Supervisor, **Dr. Sekar Raju**, Associate Professor, Department of Biological Sciences, Xi'an Jiaotong-Liverpool University (XJTLU) for giving me this opportunity and his guidance and support during the course of the PhD.

I thank my Co-supervisor, **Professor Alan J. McCarthy**, Institute of Integrative Biology, University of Liverpool (UoL) for his support and guidance during this PhD. I greatly admire him for his humanitarian and charity works.

I would like to thank **Dr. Rong Rong** (Co-supervisor, XJTLU) for her support in my PhD study. I'm grateful to my assessors **Dr. David Ruiz Carrillo** (XJTLU), **Dr. Meriel Jones** (UoL), and **Dr. Peter Miller** for the suggestions and research insights during my PhD process.

Thanks to **Dr. Jianjun Wang** (Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences) for his support with physico-chemical analysis, **Dr. Wenwei Li** (University of Science and Technology of China, Suzhou Campus) for providing lab space and facilities to perform some of the experiments and **Dr. Heather Allison** (UoL) for providing bacterial cultures required for my PhD work.

I am fortunate to have befriended with **Dr. Tatsuhiko Kadowaki**, who helped me, motivated me during tough times, supported me financially when I needed. Thank you **Dr. Paul Loughnane** (UoL), one of the kindest person I have seen, for his warm welcome and support during my stay at UoL.

I am grateful to **Long, Shen, Madanqi, Eleanora and Lisa** for their help with field sampling and other lab work. A special thanks to **George, Abhipsa, Jingting, Hao, Xiaofeng, Qiushi, Sally** and all other PhD students of Biological sciences for all the fun chats we had and all the

support. A massive thanks to Senior **Kiran** for his moral support during early days of my PhD. I'm also grateful to **Adharsh, Shiva, Naveen, Mohammed, Bhooma** and other PhD students of Civil Engineering for their support during my PhD.

Thanks to all the people in the Department of Biological Sciences, XJTLU for helping me throughout my PhD and Lab H of UoL during my Research Visit.

Financial support for this project was provided through XJTLU research consumables budget and Natural Science Foundation of Jiangsu Province (BK20141211; Principal Investigator: Dr Sekar Raju). Financial support was provided to me through Postgraduate Research Scholarship (PGRS-12-01-08). Thanks to XJTLU Research Ethics Committee for giving permision to handle human and animal fecal samples in this study (Ethics approval number: EXT 16-07-01).

Thanks is not enough to the two people who brought me into this world, **my parents**, for their everlasting love and support which they provided me. I really miss **my granddad** who left us but his memories and thoughts will never leave me until my breath. I'm extremely thankful to **my mother-in-law and father-in-law** (who passed away leaving us in a great pain), the way they supported me and for taking care of my kids during my absence. Thanks to **my little sister, my grandma, aunts, uncles, and cousins** for all the support over the last four years.

Finally, thank you very much **Ramya, my better half**. It's been a hard time for me but you kept me smiling and sane. I am lucky to have two little stars, **my kids,** who kept me lively with their adorable conversations and funny deeds. Thank you dears!

Table of Contents

Abstract..........ii

Acknowledgments..........iv

Table of Contents..........vi

List of Tables..........ix

List of Figures..........xi

List of Abbreviations..........xiv

List of IUPAC Degenerate Base Symbols..........xvii

List of Publications and Conference Presentations..........xviii

Chapter-1: General Introduction..........1

1.1 Water Pollution..........2

1.2 Fecal Pollution..........3

1.3 Traditional indicators of fecal pollution..........4

1.4 Microbial Source Tracking..........4

1.5 Water quality issues in China..........12

1.6 Taihu lake water quality..........13

1.7 Research Objectives..........16

Chapter-2: Physico-Chemical and Microbiological Analysis of Tiaoxi River Water (Taihu Watershed)..........17

Abstract..........18

2.1 Introduction..........19

2.2 Materials and Method..........21

2.3 Results..........27

2.4 Discussion..........27

2.5 Conclusions..44

Chapter-3: Next Generation Sequencing approach to characterize Fecal and Potential Pathogenic Bacterial Diversity in Tiaoxi river water...45

Abstract..46

3.1 Introduction..48

3.2 Materials and Method..51

3.3 Results..54

3.4 Discussion..71

3.5 Conclusions..77

Chapter-4: Evaluation of Human and Animal-associated Microbial Source Tracking quantitative PCR markers for monitoring Fecal Pollution in Taihu Watershed……...79

Abstract..80

4.1 Introduction..81

4.2 Materials and Method..83

4.3 Results..88

4.4 Discussion..95

4.5 Conclusions..100

Chapter-5: Quantification of microbial source tracking and pathogenic bacterial markers in water and sediment of Taihu watershed...101

Abstract..102

5.1 Introduction..103

5.2 Materials and Method..105

5.3 Results..110

5.4 Discussion..125

5.5 Conclusions..131

Chapter-6: Summary and Future Directions..132

References...139

Supplementary files……..157

List of Tables

Table 1.1: Summary of commonly used MST methodologies with advantages and disadvantages....10

Table 2.1: Description of sampling locations along with coordinates and corresponding land use types....24

Table 2.2: Summary of the results of physico-chemical and microbiological characteristics of Tiaoxi river water with statistical analyses....29

Table 2.3: Spearman's correlation coefficient (r) values observed between different water quality parameters....38

Table 2.4: Range of physico-chemical and microbiological characteristics for locations in which pollution is classified as relatively low, moderate or high....40

Table 2.5: Loadings of thirteen variables on Varimax rotated principal components……43

Table 3.1: Bacterial diversity index values comparison among water and fecal samples with normalized sequence reads…....58

Table 3.2: Range (Relative abundance percentage) of Potential pathogenic bacterial genera detected in different locations of Tiaoxi River during three sampling occasions....69

Table 4.1: List of primers and probes used for validation of MST qPCR assays in samples collected from the Taihu watershed region....85

Table 4.2: Performance characteristics of MST qPCR assays tested using fecal and sewage samples.... 90

Table 4.3: Performance of universal/general Bacteriodales MST assays on fecal and sewage samples.... 91

Table 4.4: Performance of human-associated Bacteroidales MST assays on fecal and sewage samples.... 92

Table 4.5: Performance of animal associated MST assays on fecal samples.........................94

Table 5.1: List of primers and probes used for the quantification of MST markers and genes of pathogenic bacteria... 108

Table 5.2: Detection Frequencies of MST markers in water and sediment samples of Tiaoxi River, Taihu watershed (2014-2015)...113

Table 5.3: Detection Frequencies of bacterial pathogen marker genes in water and sediment samples of Tiaoxi River, Taihu watershed (2014-2015)...119

Table 5.4: Concentration of Pathogenic Bacterial marker genes in water and sediment samples of Tiaoxi River..120

Table 5.5: Correlation between FC, MST markers and pathogenic bacterial genes...........126

List of Figures

Figure 1.1: Map of Taihu lake watershed (A) and Tiaoxi river watershed (B).....................15

Figure 1.2: Schematic representation of the research objectives of this study.....................16

Figure 2.1: Map of the study area with locations selected for surface water sampling in Tiaoxi River Basin...23

Figure 2.2: Spatial and temporal variations in physico-chemical parameters: Total Nitrogen (A), Total phosphorus (B), Nitrate-N (C), Nitrite-N (D)...32

Figure 2.3: Spatial and temporal variations in physico-chemical parameters: Phosphate-P, (A), Ammonium-N (B), Total Organic Carbon (C), Chlorophyll *a* (D)................................33

Figure 2.4: Spatial and temporal variations in microbiological parameters: Total viable count (A), Total coliforms (B), Fecal coliforms (C)..36

Figure 2.5: Dendrogram showing clustering of sampling locations based on surface water quality characteristics of the Tiaoxi river..39

Figure 2.6: Sampling locations representing high (HP), medium (MP) and low (LP) pollution risk locations..41

Figure 3.1: Rarefaction curves generated with normalized sequence reads for water, wastewater and fecal samples 57

Figure 3.2: Relative abundance of bacterial phyla in Tiaoxi river water, fecal and wastewater samples…... 59

Figure 3.3: Cluster analysis showing similarity in bacterial diversity in water samples collected in three seasons (A) and in combinations of diversity in water and fecal samples (B, C, and D)..61

Figure 3.4: Visualization of β-diversity among water and fecal samples by Principal coordinate analysis. (A) Comparison of bacterial diversity between water samples collected

during autumn, winter and summer (B) Comparison of bacterial diversity in autumn and winter water samples with fecal and wastewater samples..62

Figure 3.4: Visualization of β-diversity among water and fecal samples by Principal coordinate analysis. (C) Comparison between autumn and summer water samples with all fecal samples (D) Comparison between winter and summer water samples with all fecal samples.. 63

Figure 3.5: 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 season water samples with all fecal samples (B) and between pooled water samples and individual fecal samples (human, pig, and chicken) (C) and between fecal and wastewater samples (D) 64

Figure 3.6: Relative abundance of fecal-associated genera (A) *Bacteroides*, (B) *Prevotella*, (C) *Blautia* (D) *Faecalibacterium* observed at different sampling locations in Tiaoxi river.67

Figure 3.7: Relative abundance of fecal and sewage associated genera at Tiaoxi river sampling locations (A) *Dorea*, (B) *Macellibacteroides*...68

Figure 3.8 Redundancy analysis (RDA) showing a correlation of physico-chemical parameters with microbial community composition at a different location during three seasons (A) Autumn 2014, (B) Winter 2015, (C) Summer 2015...72

Figure 5.1: Concentration of MST fecal markers quantified in water samples at different sampling locations of Tiaoxi River. A) Total *Bacteroidales*; B) Human associated *Bacteroidales*; C) Swine associated *Bacteroidales* and D) Avian associated MST marker……………………………………………………………………………….....114

Figure 5.2: Concentration of MST markers quantified in sediment samples in different sampling locations in Tiaoxi River. A) Total *Bacteroidales*; B) Human associated

Bacteroidales; C) Swine associated *Bacteroidales* and D) Avian associated MST marker..……………………………………………………………………………...115

Figure 5.3: Concentrations (mean and standard deviations) of marker genes of bacterial pathogens detected in water (A) and sediment samples (B) of Tiaoxi River.......................118

List of Abbreviations

ACE	Abundance based Coverage Estimator
APHA	American Public Health Association
ARA	Antibiotic Resistance Analysis
ARGs	Antibiotic Resistance Genes
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CA	Cluster Analysis
CFU	Colony Forming Units
Chl *a*	Chlorophyll *a*
COD	Chemical Oxygen Demand
CSO	Combined Sewer Overflow
Ct	Threshold Cycle
CUP	Carbon Utilization Profiling
DNA	Deoxyribonucleic acid
DNQ	Detected but not quantifiable
dNTP	Deoxynucleotide triphosphate
EC	Electrical Conductivity
FC	Fecal Coliforms
FIB	Fecal Indicator Bacteria

LD	Library-Dependent
LD-MST	Library-Dependent Microbial source tracking
LI	Library-Independent
LI-MST	Library-Independent Microbial source tracking
LLOQ	Lower Limit Of Quantification
LOD	Limit of Detection
MEP	Ministry of Environmental Protection
MST	Microbial Source Tracking
NBSC	National Bureau of Statistics of China
NCBI	National Centre for Biotechnology Information
NGS	Next Generation Sequencing
NH_4-N	Ammonia nitrogen
NO_2-N	Nitrite-N
NO_3-N	Nitrate-N
NPS	Non-point sources
OTU	Operational Taxonomic Unit
PCA	Principal Component Analysis
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PFGE	Pulsed-Field Gel Electrophoresis

PO_4-P	Phosphate
PS	Point Source
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative PCR
rep-PCR	Repetitive palindromic Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
TC	Total Coliforms
Tm	Melting Temperature
TN	Total Nitrogen
TOC	Total Organic Carbon
TP	Total Phosphorous
TVC	Total Viable Count
UNICEF	United Nations Children's Fund
USEPA	U.S. Environmental Protection Agency
WHO	World Health Organization
WT	Water temperature
WWTPs	Wastewater Treatment Plants

IUPAC Degenerate Base Symbols

A	Adenosine
C	Cytidine
G	Guanosine
T	Thymidine
U	Uracil
W	Weak (A or T)
S	Strong (G or C)
M	Amino (A or C)
K	Keto (G or T)
R	Purine (A or G)
Y	Pyrimidine (C or T)
B	Not A
D	Not C
H	Not G
V	Not T
N	Any base (or unspecified base)

List of Publications:

1. Vadde, K.K., Wang, J., Cao, L., Yuan, T., McCarthy, A. J. & Sekar, R. 2018. Assessment of water quality and identification of pollution risk locations in Tiaoxi River (Taihu Watershed), China. **Water** (MDPI, Switzerland), 10, 183. (https://doi.org/10.3390/w10020183).

2. Vadde, K.K., McCarthy, A. J. & Sekar, R. Evaluation of human and animal associated microbial source tracking quantitative PCR markers for monitoring fecal pollution in Taihu Watershed, China. Journal of Applied Microbiology (In review).

3. Vadde, K.K., McCarthy, A. J., Rong. R. & Sekar, R. Quantification of microbial source tracking and pathogenic bacterial markers in water and sediment of Taihu watershed. Manuscript prepared for publication in Frontier in Microbiology.

4. **Vadde, K.K**., Feng, Q, Wang J, McCarthy, A. J., & Sekar, R. Application of next-generation sequencing approach to characterize fecal and potential pathogenic bacterial diversity in Tiaoxi river water. (In preparation).

List of Conference Presentations:

1. Sekar, R., **Vadde, K.K**., Wang, J., Cao, L., Yuan, T. & McCarthy, A. J. Microbiological and physico-chemical characteristics of surface water collected from Tiaoxi river (Taihu watershed), China. Poster presentation at Water Microbiology Conference (WMC) held at University of North Carolina, USA during 17-19 May 2016.

1. Vadde, K.K., Wang, J., McCarthy, A. J. & Sekar, R. Application of next-generation sequencing to study the microbial community composition and fecal pollution in Taihu watershed. Oral presentation at the Cold Spring Harbor Asia Conference on 'Microbiology and Environment', held in Suzhou, China during 26-30 September 2016.

CHAPTER-1

General Introduction

1.1 Water Pollution

Water is a vital resource for the human and natural environment (Oki and Kanae, 2006). Pollution of water occurs when contaminants such as chemical, physical, pathogenic microbes or radioactive substances (pollutants) enter the natural water bodies (Rivers, Lakes, groundwater, and oceans) leading to changes in water quality and harming human health and the environment (Briggs, 2003). Insufficient remediation or management of wastewater from urban, agricultural and industrial sources could lead to accumulation of chemical and biological pollutants at hazardous levels in water resources (Stevens et al., 2009). In terms of human health, it has been reported that unsafe drinking water and inadequate sanitation facilities are primary concerns affecting one-third of the world population. In addition, other threats include pathogen or chemical toxicant exposure through food (via edible plants irrigated with contaminated water or toxic chemical bioaccumulation in fish and other seafood) or recreational activities like swimming and surfing in contaminated water (Schwarzenbach et al., 2010).

Globally, particularly in low-income countries, drinking water contamination is a major concern as it causes infection or illness (Bylund et al., 2017). According to the World Health Organization, diarrheal disease is the second major cause of child death (<5 years) in the world though it is preventable (WHO, 2017). Globally, around 480,000 children (<5 years) die due to diarrheal diseases which accounts for nearly 8% of total child deaths (UNICEF, 2018). In China, 4% of total child deaths under the age of 5 years are due to diarrheal diseases (WHO, 2015). In most of the developing countries, diarrheal diseases are primarily transmitted through contaminated drinking water or inadequate sanitation facilitates (Ashbolt, 2004). Primarily, diarrheal diseases are due to contamination of water with human or animal feces transmitting pathogenic microorganisms (WHO, 2017).

1.2 Fecal pollution

Fecal contamination of the aquatic environment by human and animal sources is one of the critical concerns for public health and the environment. Fecal pollution occurs when feces from septic leakages, agricultural runoff, inadequately treated sewage from treatment plants, and wildlife or livestock enters water bodies (Harwood et al., 2000, Bernhard and Field, 2000b). Fecal pollution of waters can transmit a wide range of pathogenic microorganisms (Ashbolt, 2015). Therefore, proper water quality monitoring and remediation procedures are required to prevent waterborne diseases. Globally, waterborne outbreaks due to fecal contamination occur in all countries, although, the occurrence and severity levels are higher in developing countries (Shuval, 2003, Tornqvist et al., 2011). Most of the pathogenic microorganisms associated with waterborne diseases originate from feces of mammals (Leclerc et al., 2002). Therefore, in terms of human health, it is crucial to prevent contamination of water bodies with mammalian fecal sources (Stewart et al., 2007). Similarly, avian fecal pollution originating from domestic and wildlife are also very important as it can transmit pathogenic microorganisms such as *Campylobacter* species, *Giardia lamblia,* and *Cryptosporidium parvum* (Kuhn et al., 2002, Ahmed et al., 2016b). Both human and animal fecal pollution enter water bodies in different ways through point sources or non-point sources. Point sources refer to contamination of water bodies through single, discrete, identifiable sources like ditches and pipes. Non-point sources, which are commonly termed “diffuse sources”, refer to contamination that cannot be attributed to a single, discrete source; some examples are storm runoff, illegal dumping, wildlife and septic leakages. Mostly, it is the non-point sources of pollution that are considered the primary cause for the declining of water quality in a watershed (Santo Domingo et al., 2007). Therefore, appropriate monitoring is required to determine the influence of different fecal sources on the water bodies and to enable effective management of water resources.

1.3 Traditional indicators of fecal pollution

Culture-based methods to enumerate fecal indicator bacteria (FIB) have been routinely used as indicators of fecal contamination in the aquatic environment traditionally (Fremaux et al., 2009). FIB, which includes fecal coliforms (FC), fecal *Escherichia coli*, and fecal *Enterococcus* spp., are normally found in the intestinal tract of warm-blooded animals and are excreted in feces. As direct monitoring of pathogens is cost-effective and technically complex due to their high diversity and low concentration, FIB cultivation and enumeration have been commonly used to determine public health risk associated with human and animal fecal pollution (Schriewer et al., 2010, Shahryari et al., 2014). However, several studies have indicated that the correlation between traditional fecal indicator bacteria (FIB) and pathogen presence is not perfect (Walters et al., 2009, Fremaux et al., 2009, Shahryari et al., 2014). Additionally, FIB occurrence and proliferation in the natural aquatic environment has been reported (Byappanahalli et al., 2003). Furthermore, FIB detection in water cannot determine the origin of the fecal source (Field and Samadpour, 2007). Determining the sources of fecal pollution is essential to assess the potential public health risks associated and also to implement effective remediation measures by taking legitimate actions (Santo Domingo et al., 2007). In this regard, various microbial source tracking (MST) techniques have emerged to ascertain the source or origin of fecal contamination in environmental samples (Bernhard and Field, 2000b, Harwood et al., 2000, Scott et al., 2005, Layton et al., 2006, Kildare et al., 2007, Ahmed and Katouli, 2008).

1.4 Microbial Source Tracking

Microbial source tracking (MST) is a group of methods that examine similarities/genetic matches between fecal indicator bacteria of different hosts and contaminated environmental samples to distinguish the fecal source of origin (Bernhard and Field, 2000b). Typically, MST methods are categorized into library-dependent (LD) and library independent (LI) methods

based on the requirement of reference source "library" construction. Some of the commonly used LD and LI methods with their advantages and disadvantages are summarized in Table 1.1.

1.4.1 Library-dependent MST method

LD-MST methods require the construction of libraries or databases of bacterial strains (typically based on phenotypic or genotypic characteristics) isolated from reference fecal sources (known), and these libraries are used for comparison against isolates of water samples (unknown) to ascertain the sources of fecal pollution (Harwood et al., 2000, Wiggins et al., 2003). LD-MST approach relies on the assumption that the target organism of a particular host source has specific phenotypic and genotypic characteristics and these remain constant for a period in the environment to which a fingerprint library can be constructed to predict the source of isolates present in unknown samples (Ritter et al., 2003, Robinson et al., 2007). Many of the LD-MST techniques require culturing of target organisms from reference fecal sources and unknown water samples collected from a watershed of interest; though recently developed Next Generation Sequencing (NGS) based LD-MST methods do not require culturing (Cao et al., 2011, Unno et al., 2012). The culture based LD-MST methods rely on either genotypic or phenotypic analysis of FIB. Ribotyping of *E. coli* has been the most frequently used culture-based genotypic LD-MST method (Parveen et al., 1999, Scott et al., 2003, Carson et al., 2001) and in the culture based phenotypic LD-MST method, antibiotic resistance profiling of fecal coliforms (Harwood et al., 2000, Whitlock et al., 2002), fecal enterococci (Hagedorn et al., 1999, Harwood et al., 2003) were commonly used. Several evaluation studies have revealed that culture based LD-MST methods performed poorly making them less useful for source tracking of environmental samples, as well as the disadvantage of needing to culture large site-specific libraries or databases of phenotypic or genotypically characterized isolates (Moore et al., 2005, Harwood et al., 2014). However, culture-independent LD-MST methods such as NGS based methods have gained more attention in recent years, though these methods are site-

specific (specific to a geographical location) and requires optimizations and validations before applying to environmental samples (Ahmed et al., 2015b, Staley et al., 2018, Unno et al., 2018). The culture based (phenotypic and genotypic) and culture-independent (NGS) LD-MST methods are briefly given below.

1.4.1.1 Culture based LD-MST: phenotypic methods

Phenotypic methods depend on comparison of physical or biochemical characteristics such as antibiotic resistance profiles or carbon sources utilized by the target organism. The commonly used phenotypic methods are antibiotic resistance analysis (ARA) and carbon utilization profiling (CUP), in which ARA has mostly used method due to its simple and inexpensive nature requiring less technical skills. However, these methods require complex statistical analysis to predict the source of contamination, which is a major disadvantage. The details of these two phenotypic approaches along with their applicability to field studies for source tracking have been reviewed in several papers (Stoeckel and Harwood, 2007, Ahmed and Katouli, 2008, Field and Samadpour, 2007).

1.4.1.2 Culture based LD-MST: Genotypic methods

Genotyping utilizes comparison of genetic material of target organisms to differentiate between the fecal sources (Griffith et al., 2003). Ribotyping, Repetitive palindromic Polymerase Chain Reaction (rep-PCR), and pulsed-field gel electrophoresis (PFGE) are the most commonly used genotypic methods. These methods are highly reproducible when performed by a skilled technician. However, these methods are expensive; require high technical skills and complex statistical analysis to predict the source (Scott et al., 2002, McLellan et al., 2003). The complete details of genotypic methods can be found in previous review papers (Yan and Sadowsky, 2007, Meays et al., 2004, Stoeckel and Harwood, 2007).

1.4.1.3 Culture independent LD-MST: NGS based methods

Advances in sequencing technologies with the introduction of NGS has improved the ability to characterize bacterial communities in different sources such as environmental and fecal samples, which further facilitated the development of culture-independent LD-MST in recent years (Knights et al., 2011, Unno et al., 2012). This method relies on construction or creation of large sequence libraries in the form of operational taxonomic units (OTUs) for different fecal sources at a geographical location and their comparison with OTUs of environmental samples (Unno et al., 2010, Knights et al., 2011). This can be discrete examination of shared OTUs from fecal and environmental samples (called community based LD-MST using shared OTU method) (Unno et al., 2010) or by using a software program named SourceTracker which uses a Bayesian algorithm to indicate the contribution of OTUs from fecal source in environmental samples (Knights et al., 2011).

Community-based LD-MST using the shared OTU method can facilitate qualitative assessment of fecal pollution (Boehm et al., 2013). Although there is a criticism in assigning these OTUs or short sequence reads generated by NGS to species level (Nguyen et al., 2016), they can be used to characterize bacterial species present in different fecal sources and group them as host-specific or shared species (OTUs) between several fecal sources. Host-specific OTUs have been compared with OTUs of environmental samples to enable the potential fecal source of contamination and magnitude of the contamination based to be ascertained, based on total abundance (Unno et al., 2010, Unno et al., 2012).

As discussed above, the SourceTracker method utilizes a Bayesian algorithm to track the fecal pollution by determining the shared OTUs of the fecal origin with environmental samples (Knights et al., 2011). This method has gained much attention in recent years, as it was developed to predict the quantitative presence of fecal contamination. However, some validation studies have reported that this method has high variability in quantification results

among technical replicates and showed lower confidence (Henry et al., 2016, Brown et al., 2017). Furthermore, a validation study conducted very recently in Florida, USA indicated the geographical variability (spatial) of this method and recommended that further work is mandatory to assess the temporal variability in fecal communities (Staley et al., 2018).

Overall, both culture independent LD-MST methods have spatio-temporal limitations, and in-depth evaluation and optimizations are needed before their application in a new geographical setting (Staley et al., 2018, Unno et al., 2018).

1.4.2 Library independent MST method

LI-MST methods do not involve preparation of libraries for fecal sources. LI-MST methods need either culturing of host-specific microbes from environmental samples or molecular analysis of host-specific viral or bacterial genetic material isolated from environmental samples in order to identify human and animal fecal sources (Griffith et al., 2003, Hagedorn and Liang, 2011). Some of the commonly used LI-MST techniques are given below.

1.4.2.1 Bacteriophage typing

The viruses that infect specific bacteria (in some cases species-specific) are bacteriophages. Coliphages are a broad group of viruses that infect *E.coli*. There are further categories based on the genome (RNA or DNA), nature of attachment to target cell before infecting, such as some adsorb to the cell wall called somatic coliphages and some attach to sex pilus of *E.coli* called F+ RNA or DNA coliphages (Vinje et al., 2004). In LI-MST, F+RNA coliphages are largely used in discrimination of human fecal sources from non-human sources in environmental samples (Vergara et al., 2015, Shahrampour et al., 2015). The brief procedure includes initial F+RNA coliphages isolation in the presence of DNAse to separate them from F+DNA types and then followed by serotyping (using antisera that inhibit its infection to *E.coli*) or genotyping (using labeled probes) (Havelaar et al., 1990, Wolf et al., 2008). Though this technique is fast and relatively easy to perform, the main disadvantage is it can only

discriminate human fecal sources from animal sources, but not amongst animal sources. Furthermore, F+ RNA coliphages have unpredictable survival kinetics in the environment (McQuaig et al., 2012).

1.4.2.2 Viral PCR and qPCR:

Generally, viruses are very specific to the host. With limited host range, viruses have been used as targets to differentiate human and animal fecal sources in environmental samples. The viral LI-MST method is relatively simple which includes isolation of virus from samples, followed by viral genome extraction and amplification of targets by PCR or qPCR. Enterovirus and adenovirus that are specific to human, bovine and porcine have been used as targets to indicate human, bovine, and livestock-associated fecal contamination in environmental samples (Noble et al., 2003, Wolf et al., 2010). However, an evaluation study has reported that bovine enteroviruses lack specificity and had cross reactivity with other animals like horse and geese (Field and Samadpour, 2007). Another study reported that though human viral fecal markers showed good sensitivity to sewage samples, they failed to detect viral DNA in human fecal samples, as their abundance is low in the human population (Griffith et al., 2003).

1.4.2.3 Bacterial PCR and qPCR:

Several host specific bacteria have been suggested for their use as a fecal indicator in MST studies and most of these methods rely on extraction of DNA from environmental samples, followed by PCR or qPCR amplification of marker gene of bacteria specific to a host. Some of the recommended host specific bacteria are *Bacteroidales* (Bernhard and Field, 2000b), *Bifidobacterium* spp. (Bonjoch et al., 2004), *Methanobrevibacter* spp. (Ufnar et al., 2007, Harwood et al., 2013), *Rhodococcus coprophilus* (Wicki et al., 2012), *Escherichia coli* (Kim et al., 2010) and *Enterococcus* spp. (Scott et al., 2003). Among these, *Bacteroidales* based MST makers targeting 16S rRNA genes have shown most promising results.

Table 1.1 Summary of commonly used MST methodologies with advantages and disadvantages.

Method	Library-dependent	Culture-dependent	Method Description	Advantages	Disadvantages
Rep-PCR	Yes	Yes	Requires PCR amplification of specific repetitive sequences and DNA fragments separation by agarose gel, producing unique fingerprints of bacterial strains.	Differentiates isolates among multiple hosts, Easy and rapid, higher accuracy (based on the size of library).	Reference libraries are required, concerns over reproducibility, Geographical and temporal-specific.
Ribotyping	Yes	Yes	Involves digestion of genomic DNA with restriction enzymes, followed by Southern blotting of digested genomic DNA.	Highly reproducible differentiates isolates among multiple hosts, higher accuracy (based on the size of library).	Reference libraries are required, Geographical and temporal-specific, specialized training is required.
Antibiotic resistance profiling	Yes	Yes	A biochemical method that discriminates bacteria based on response and resistance pattern to antibiotics.	Easy, rapid and inexpensive.	Geographical and temporal-specific, higher false-positives, complex analysis.
Next-Generation Sequencing	Yes	No	DNA extraction, sequencing using NGS, computational analysis.	Comprehensive bacterial communities identification, multiple fecal sources can be identified if libraries are available.	Geographical and temporal-specific requires the construction of large libraries for different hosts.
Bacteriophage typing	No	Yes	Involves source specific bacteriophage isolation and enumeration, followed by serotyping or genotyping of the phages.	Highly specific, relatively inexpensive	Only distinguishes between human and animal only, less sensitive due to low abundance.
Bacterial PCR and qPCR	No	No	Involves amplification of host-associated bacterial markers, several markers were developed to discriminate between human and animal fecal sources.	Specificity to host, accuracy ranged from medium to high, fast and easy to perform.	Markers available for relatively few hosts, marker unknown persistence of markers, expensive, errors in amplification.
Viral PCR and qPCR	No	No	Involves amplification of host-associated viral markers.	Host specificity, fast and easy to perform	Less abundance in the environment, markers availability for very few hosts.

Currently, PCR/qPCR assays targeting members of the order *Bacteroidales* are the widely used in MST, as they have shown higher potential in tracking fecal sources (Wuertz et al., 2011). *Bacteroidales* are obligate anaerobic bacteria living in the gastrointestinal tract of mammals (Wexler, 2007). They are present at a higher concentration than conventional indicator bacteria in mammalian feces, though this ratio varies in different animal species (Meays et al., 2004). The *Bacteroidales* abundance has been reported as low in avian species, such as chicken, goose, and gull (Jeter et al., 2009, Lu et al., 2008, Ohad et al., 2016). *Bacteroidales* are strictly anaerobic bacteria, which means they cannot survive for a long time in environmental samples and their presence indicates recent fecal contamination, an important feature for microbial source tracking. It was reported that *Bacteroidales* could persist for about one week in the environment (Bae and Wuertz, 2009, Walters and Field, 2009, Bae and Wuertz, 2012). However, it has been reported that *Bacteroides thetaiotaomicron* could survive for a longer period in the environment (Xu et al., 2003).

Although most of the *Bacteroidales* LI-MST assays are based on traditional PCR amplification to detect presence or absence of target genes initially, in recent years they have been replaced with quantitative PCR (qPCR) assays to quantify the markers in environmental samples. Marker gene detection and quantification by qPCR can be accomplished either by non-specific fluorescent reporters like SYBR Green (Seurinck et al., 2005, Okabe et al., 2007) or using specific labeling probes, based on Taqman assays (Layton et al., 2006, Kildare et al., 2007, Mieszkin et al., 2009, Green et al., 2014). However, TaqMan based assays are recommended for MST studies on environmental sample due to probe and primer requirement in the assay to provide more specificity (Kildare et al., 2007). Currently, total or universal *Bacteroidales* can be detected by four assays: BacUni (Kildare et al., 2007), GenBac (Dick and Field, 2004, Siefring et al., 2008), AllBac (Layton et al., 2006) and BacPre1 (Okabe et al., 2007). Several human-associated *Bacteroidales* MST assays have been developed and field-tested for

quantification of human fecal markers in environmental samples (Layton et al., 2006, Kildare et al., 2007, Seurinck et al., 2005, Reischer et al., 2007, Shanks et al., 2009, Haugland et al., 2010, Green et al., 2014). Similarly, several host-associated assays were developed for pig, ruminant, dog, gull, Canada geese, horse and elk (Converse et al., 2009, Dick et al., 2005, Dorai-Raj et al., 2009, Fremaux et al., 2010, Mieszkin et al., 2009, Silkie and Nelson, 2009, Stricker et al., 2008). Though numerous MST studies focused on the evaluation of source specificity and cross reactivity of fecal sources, few studies have validated the geographical applicability of these host-specific markers and it has been recommended that local validation is needed prior to their use at any new study area (Ahmed et al., 2007, Reischer et al., 2013, Jenkins et al., 2009, Boehm et al., 2016, Odagiri et al., 2015).

1.5 Water quality issues in China

Rapid urbanization, industrialization, and socio-economic development have led to a severe ground and surface water pollution in China, which has been acknowledged as a heavily degraded natural resource of China (Han et al., 2016). Several lakes and rivers are contaminated by a high degree of pollution and the limitations of water resources is a major concern to China's development (Jiang, 2009). Though China has higher total renewable freshwater resources of 2796 billion m^3 per year, it still has low water reserves of 2039 m^3 per person which are only 34% of the world average per capita per year (6016 m^3) (NBSC, 2015, WHO, 2015). The total water consumption in China had gradually increased with an annual average rise of 0.97% from the year 2000 (550 billion m^3) to 2013 (618 billion m^3), followed by a slight decrease in 2014 and 2015. Agriculture sector continued to be the major water consumer in China accounting for 61-69% of total water consumption annually and it had a 385 billion-m^3 water consumption in 2015. The remaining water consumption came from industry (21-24%), domestic usage (10-13%) and ecological protection (1-2%) purposes (NBSC, 2015, Jiang, 2015). The wastewater discharge from these consumers could further pollute the existing

freshwater sources, which is evident from increase in water pollution. By 2015, the total amount of wastewater discharge was almost 73 billion m^3 and about 68% of the discharge came from domestic sources, which is also increasing annually (NBSC, 2015). Although wastewater discharge is increasing, treatment plant provision is inadequate and infrastructure developments are lagging behind the actual requirement (Hagedorn and Liang, 2011). As of 2014, it was reported that 82% of urban areas are equipped with wastewater treatment plants but only 1% of rural areas have these facilities, and discharge from these areas could degrade the aquatic environment of receiving sources (Jiang, 2015). According to the WHO, more than 36% of the population in rural China still lack access to improved sanitation and drinking water (WHO, 2015). Though drinking water quality is improving in recent years, nearly 200 million people are still consuming unsafe drinking water in China (Tao and Xin, 2014, Han et al., 2016). Approximately 60,000 people are dying in China annually due to water pollution-related diseases (Qiu, 2011).

Depending on pollution levels, the Ministry of Environmental Protection (MEP), China has classified water quality into classes I-III (excellent, good, satisfactory), IV-V (moderately and heavily polluted), and >V (seriously polluted). Based on this grading system, the overall surface water quality of China for 2015 was graded as Class I-III (64.5%), Class IV-V (26.7%), and Class >V (8.8%) and one-third of lakes, rivers were reported as seriously polluted, and human contact was not recommended in these waters (MEP, 2016). For the Taihu Lake basin, the study of interest here, water quality was graded as Class I-III (41.1%), Class IV-V (52.9%), and Class >V (5.9%) (Rose et al., 2015, Sha Long et al., 2018) and it was reported that this lake has eutrophication and fecal pollution issues (Hagedorn and Liang, 2011, He et al., 2016a).

1.6 Taihu Lake water quality

Taihu Lake is one of the top five larger freshwater lakes in China, located in southeast Jiangsu province and the Yangtze River basin (Chen et al., 2016), one of the most industrialized regions,

where this lake serves multiple functions as drinking water source, flood protection, fisheries, tourism, and transportation (Fig. 1.1A) (Qin et al., 2007). Pollution of water bodies is a serious concern in areas experiencing rapid urbanization. Rapid industrial development coupled with an increase in the population of the Taihu watershed has resulted in declining water quality and ecosystem health (Chen et al., 2016). Increased anthropogenic activities combined with existing land use practices can increase pollutant loadings, such as nutrients and microbes into water bodies that can affect public health (Carroll et al., 2006). On the other hand, rainfall events can further accelerate the pollutant loadings due to the entry of storm water runoff from urban areas, as well as from agricultural areas due to common practices such as manure use as fertilizers and livestock grazing near the water bodies (Ackerman and Weisberg, 2003). Although entry of pathogenic organisms is a major concern, nutrient loading can cause enrichment of water resources resulting in eutrophication (Carroll et al., 2006). Previous reports showed that pollution from inflow rivers in a watershed could contribute to severe ecological and sanitary problems (Singh et al., 2005, Wang et al., 2007b). Therefore, it is important to assess the water quality of connecting rivers to prevent the entry of polluted waters and provide effective management of Taihu Lake water quality. Taihu Lake watershed is a complex network and currently, it is connected to 220 rivers and tributaries approximately (Wang et al., 2007b). Water from tributaries enters the Lake from the west side and goes out to Rivers in the eastern side, primarily through Taihu Bay (Xu et al., 2013). Among the 13 main inflow Rivers, Tiaoxi River is the major water source contributor to Taihu Lake (Fig. 1.1B) (Zhang et al., 2016b). Therefore, Tiaoxi River could significantly influence the water quality of this Lake and it is crucial to monitor the water quality of Tiaoxi River for active regulation of Taihu Lake watershed. However, recent reports showed that Tiaoxi River is heavily contaminated with pollution from multiple sources such as farmland, domestic sewage, and industry, which consequently influences Taihu Lake water quality (Hagedorn and Liang, 2011, Lv et al., 2015).

Figure 1.1 Map of Taihu Lake watershed (A) and Tiaoxi River watershed (B).

Further literature about the Tiaoxi River water quality is given in the next chapters based on the type of pollution the chapter is focused on (for example; physico-chemical analysis or fecal pollution monitoring chapters).

Overall, to the best of our knowledge, no MST study has been carried out at Tiaoxi River thus far and therefore, this current study was aimed at application of library independent-MST Techniques to characterize fecal pollution entering Taihu Lake.

1.7 Research Objectives

- Preliminary identification of locations (hot spots) with possible fecal pollution based on physicochemical and culture-based microbiological analysis.
- Investigation of the bacterial community diversity in water samples of Tiaoxi River and their temporal and spatial variations.
- Evaluation of human and animal associated microbial source tracking quantitative PCR markers for monitoring fecal pollution in Taihu Watershed, China
- Investigation of the MST markers and genes of specific bacterial pathogens (by qPCR of environmental DNA) in water and sediment samples of Tiaoxi River and their correlation with conventional FIB.

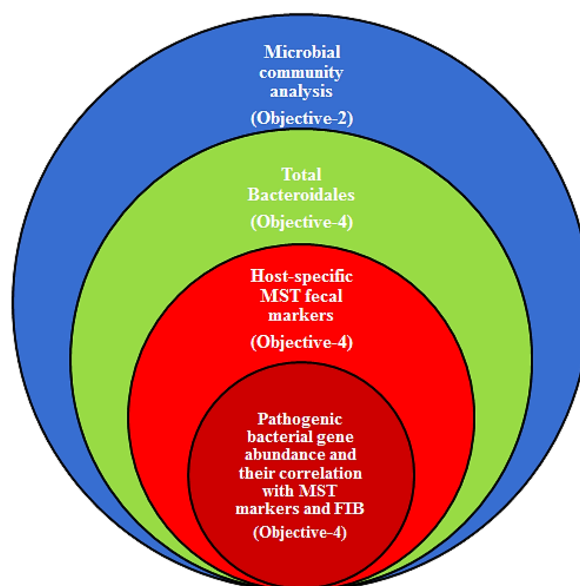


Figure 1.2 Schematic representation of research objectives of this study

CHAPTER-2

Physico-Chemical and Microbiological Analysis of Tiaoxi River Water (Taihu Watershed)

Abstract

This chapter was aimed to evaluate the physico-chemical and microbiological characteristics of Tiaoxi River and to determine the spatial and seasonal variations in the water quality. Water samples were collected from 25 locations across the Tiaoxi River in three seasons in 2014–2015. Fourteen water quality parameters including multiple nutrients and indicator bacteria were assessed, and the data analyzed by multivariate statistical analyses. The physico-chemical analysis showed high levels (>1 mg/L) of total nitrogen (TN) in all locations for all seasons. Total phosphorus (TP), nitrite-N (NO_2-N), and ammonium-N (NH_4-N) exceeded the acceptable limits in some locations and fecal coliform counts were high (>250 CFU/100 mL) in 15 locations. Hierarchical cluster analysis showed that the sampling sites could be grouped into three clusters based on water quality, which were categorized as low, moderate, and high pollution areas. Principal component analysis (PCA) applied to the entire dataset identified four principal components, which explained 83% of the variation; pH, conductivity, TP, and NO_3-N were found to be the key parameters responsible for variations in water quality. The overall results indicated that some of the sampling locations in the Tiaoxi River are heavily contaminated with pollutants from various sources, which can be correlated with land use patterns and anthropogenic activities.

2.1 Introduction

Surface water pollution remains a major problem worldwide, caused by both natural processes and anthropogenic activities (Noori et al., 2010). Evaluation of surface water quality in drinking water sources is important, as it can be one of the main pathways for the dissemination of toxic chemicals and pathogenic microorganisms (Lodder et al., 2010, Ouyang, 2005). Identifying the source(s) of contamination and developing appropriate management strategies is essential to minimize potential public health risks (Carroll et al., 2006). The surface water quality in a region can be affected by both point and non-point sources of pollution (Nnane et al., 2011). Point source (PS) pollution occurs through a single identifiable source such as effluents from industries and wastewater treatment plants (WWTPs) (Hill, 2010) whereas, non-point sources (NPS) include run-off associated with a particular land use pattern such as urban (e.g. stormwater, sewage overflows), agriculture (e.g. fertilizers, pesticides, animal manure) or forestry land uses (Bu et al., 2014). Entry of these sources into the water can represent improper discharge of toxic chemicals and pathogenic microorganisms, therefore, water quality monitoring and sanitary risk identification is essential to protect the population from waterborne diseases and to develop appropriate preventive measures. Since environmental systems like rivers and lakes are affected by multiple sources, it is important to understand the spatial and temporal variations in physico-chemical and microbiological parameters for assessment and management (Razmkhah et al., 2010). However, assessing multiple water quality parameters generates large and complex datasets and multivariate statistical techniques are required to interpret the results (Li et al., 2011, Mei et al., 2014, Wang et al., 2012). Multivariate techniques like cluster analysis (CA) and principal component analysis (PCA), have been successfully applied to better understand the water quality and ecological status of studied systems (Shrestha and Kazama, 2007, Singh et al., 2005). In addition to assessing the water quality, multivariate techniques have several applications such as identification of

possible factors/ sources that influence water systems (Ogwueleka, 2015, Sheikhy Narany et al., 2014), data reduction (Mustapha et al., 2013, Sharma et al., 2013), spatial and temporal variations (Ogwueleka, 2015) and grouping of sampling sites (Gatica et al., 2012).

The current waterbody of interest, Taihu Lake, is the third largest freshwater lake in China, serving multiple functions such as drinking water source, flood protection, fisheries, tourism, and transportation (Qin et al., 2007). As mentioned earlier, this lake is facing serious water quality issues and previous reports showed that input rivers significantly contribute to the eutrophication, algal blooming and fecal pollution to this lake (Wang et al., 2011a, Du et al., 2017, Hagedorn and Liang, 2011). Therefore, it is essential to assess water quality of inflow rivers to prevent the entry of polluted waters and provide effective management of Taihu Lake water quality. At present, Taihu Lake is connected to more than 200 Rivers (Qin et al., 2007), and Tiaoxi River is the main inflow river contributing approximately 60% of the total source water for this lake (Liu et al., 2011, Zhang et al., 2016b). Although Tiaoxi River significantly influences Taihu Lake by providing source water, studies focusing on the detailed water quality analysis of Tiaoxi River or identification of pollution hotspots are limited (Liu et al., 2011, Tang et al., 2012, Zheng et al., 2017). Liu et al. (2011) studied the size distribution and composition of phosphate in the East Tiaoxi River and Tang et al. (2012) investigated the estrogen pollution in the Tiaoxi River by chemical and bioassays. Recently, Zheng et al. (2017) studied the profiling of antibiotic resistance genes (ARGs) in water samples collected from different catchment areas in East Tiaoxi River and analyzed some water quality parameters, namely chemical oxygen demand (COD), TN, TP and NH_3-N and correlated with ARGs and pathogenic bacteria. Therefore, the main objective of the study reported here was to provide a more comprehensive assessment of Tiaoxi River (east, west and combined River) water quality and identify pollution hotspots based on physico-chemical and microbiological data through the application of multivariate statistical approaches.

2.2. Materials and Method

2.2.1 Study Area

This study was carried out in the Tiaoxi River, which is one of the major Rivers connected to Taihu Lake. The mainstream of the Tiaoxi River is 158 km in length, comprising east and west Tiaoxi Rivers. The annual inflow of Tiaoxi River to Taihu Lake is 2.7 billion m^3 which are approximately 60% of the total source water input of the lake (Liu et al., 2011). Furthermore, it serves as a drinking water source and also supplies water for agricultural and industrial purposes for many cities located in northern Zhejiang Province, China (Tang et al., 2012). Upstream, the river flows through agricultural areas, while the downstream part flows within the urban cities of northern Zhejiang Province and is subject to industrial inputs. The River collects waters from a population of approximately one million inhabitants, primarily located in moderately sized cities such as Huzhou along the River stream (Zhang et al., 2016b).

2.2.2 Sampling Locations and Sample Collection

Twenty-five sampling locations were selected for this study covering domestic, agricultural and industrial areas. The land use patterns of areas close to sampling locations are shown in Table 2.1 Surface water samples were collected along 100 km in the main streams of east and west Tiaoxi Rivers and from Tiaoxi River junctions with other tributaries (Fig. 2.1) in autumn 2014, winter and summer 2015. Water samples were collected in sterile 5L polypropylene containers and kept at ambient temperature until they were brought to the laboratory. The water samples were processed within 24h of the collection. Due to unforeseen circumstances, it was not possible to collect samples from a few locations in one or more seasons.

2.2.3 Physico-chemical Parameters

In China, the surface water quality is regulated by the Ministry of Environmental Protection (MEP) of People's Republic of China (PRC) and quality standards have been set for surface

water (MEP, 2016). Based on MEP standards, surface water bodies in China are classified into five categories based on their utilization and protection objectives. The values of physico-chemical and microbiological parameters of Tiaoxi River water were compared to the class III water quality category standards (applicable to second class of protected area for centralized sources of drinking water, protected area for common fishes and swimming areas) as cited in "Environmental Quality Standards for Surface Water (GB3838–2002, GHZB1–1999)", MEP, PRC guidelines for surface water. The following physico-chemical parameters were analyzed in this study: air and water temperature, pH, conductivity (EC), total nitrogen (TN), total phosphorous (TP), nitrate (NO_3-N), nitrite (NO_2-N), phosphate (PO_4-P), ammonia nitrogen (NH_4-N), total organic carbon (TOC) and chlorophyll *a* (Chl *a*).

Temperature and conductivity were measured on site with COM-100 handheld meter instruments (HM Digital Inc., USA) and pH was measured using Eutech pH 700 instrument (Thermo Fisher Scientific Inc., USA). TN and TP were determined by peroxodisulphate oxidation and spectrophotometric method. NO_3-N, NO_2-N, NH_4-N, and PO_4-P were measured using a continuous flow analyser (Skalar SA 1000, Breda, Netherlands) as described in Wang et al. (Wang et al., 2011b)). The TOC was measured by high-temperature oxidation with a Shimadzu analyser (model 5000; Tokyo, Japan). Chl *a* was measured following by the procedures recommended by American Public Health Association (APHA, 2005, Paerl et al., 2011).

2.2.4 Microbiological Parameters

Total viable count (TVC) was determined by using plate count agar (LabM, UK). Briefly, 100 µl of each serially diluted water sample was aseptically plated onto agar plates in duplicates and incubated at 37 °C for a maximum duration of 48 h. The mean colony counts were expressed as CFU/mL (Reasoner and Geldreich, 1985).

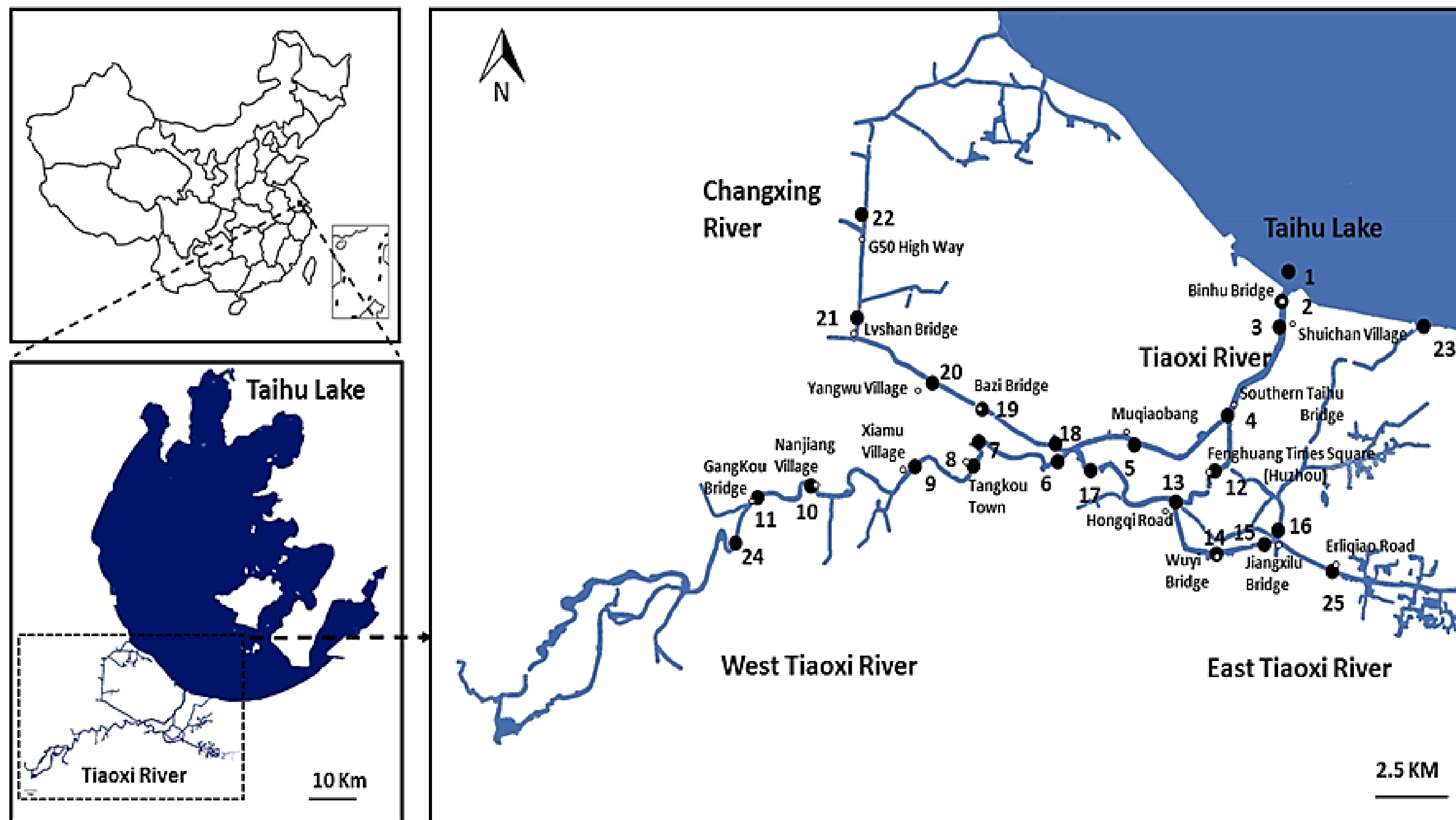


Figure 2.1 Map of the study area with locations selected for surface water sampling in Tiaoxi River Basin.

Table 2.1 Description of sampling locations along with coordinates and corresponding land use types.

Sampling location	Description of location and land use types	Coordinates	
		Latitude	**Longitude**
1	Taihu Lake and Tiaoxi River junction; 1 km inside the Taihu Lake; Aquaculture/fishing area.	N30°57′3.15″	E120°07′42.64″
2	Suburban area with aquaculture and fish handling/processing area.	N30°56′25.30″	E120°07′35.72″
3	Fishermen village; People live on boats stationed at this location.	N30°55′57.65″	E120°07′37.27″
4	Sub-urban area with residential apartments, businesses, and parks; East and West Tiaoxi River junction near south Taihu bridge.	N30°53′50.96″	E120°06′0.95″
5	Urban area with construction sites and various factories; Heavy ferry transportation were noticed in this area. Close to WWTP.	N30°53′19.40″	E120°03′18.16″
6	Suburban and industrial area with various factories; West Tiaoxi River and Changxin River junction. Active in Pig and Poultry farming.	N30°52′55.15″	E120°0′58.87″
7	Residential, farming and small industrial area close to a village; Various farm animals in small scale were noticed at the Riverbank.	N30°53′14.16″	E119°58′38.58″
8	Close to a town with businesses and residences, Ferry/boat docking area. Active in Pig and Poultry farming.	N30°53′1.82″	E119°58′48.08″
9	Rural agricultural area with sparse residential apartments.	N30°52′43.41″	E119°56′43.37″
10	Rural agriculture area with few industries (e.g. shipping industries and oil station); Heavy ferry transportation was noticed in this area.	N30°52′21.55″	E119°53′55.85″
11	Rural with a high number of residential apartments; Heavy ferry transportation was noticed in this area.	N30°52′8.11″	E119°52′15.52″

12	Urban area with businesses (e.g. many shopping malls) and residential apartments; Tourist boats docked close to this location.	N30°52′54.56″	E120°06′1.47″
13	Urban area with residential apartments and construction sites; Second junction between west and east Tiaoxi River. Pig and Poultry farms are present within 1km.	N30°51′56.74″	E120°04′25.11″
14	Sub-urban area with construction sites, residential apartments, and businesses; Ferry docking (large-scale) area.	N30°50′53.74″	E120°05′38.57″
15	Sub-urban area with residential apartments and businesses; Junction between east Tiaoxi and a small River that connects to Taihu; Sampled close to ferry docking (large-scale) area.	N30°50′59.27″	E120°06′21.50″
16	Sub-urban and residential/business area; Junction between the main River and a canal, which connects to Taihu Lake.	N30°51′27.75″	E120°07′32.13″
17	Sub-urban and sparse residential area; Sampled at the third junction between west and east Tiaoxi River.	N30°52′40.51″	E120°01′58.88″
18	Sub-urban and industrial area; Sampled in the junction of ChangXing and Tiaoxi River; Sampled near ferry docking station.	N30°53′11.17″	E120°0′52.95″
19	Rural/village, sparse residential and industrial area.	N30°54′2.88″	E119°58′42.16″
20	Rural/village and sparse residential/industrial area. Pig and Poultry farms are present near to this location.	N30°54′33.91″	E119°57′31.34″
21	Rural/village, residential and sparse industrial area. Pig and Poultry farms are present near to this location.	N30°55′52.05″	E119°55′9.61″
22	Rural/village and industrial area; Heavy ferry transportation; Sampled close to a factory and ferry docking station.	N30°57′45.22″	E119°55′19.98″
23	Rural/village area; Sampled in a small canal, which connects to Taihu Lake.	N30°55′53.87″	E120°11′35.48″
24	Rural/village and sparse residential /industrial area.	N30°51′0.12″	E119°51′28.68″
25	Sub-urban area with businesses and industries; Many small Rivers branch off from East Tiaoxi River.	N3050045.36”	E120°08′21.54″

Total coliform (TC) bacteria were determined using Harlequin™ E. coli/Coliform medium (LabM, UK). Briefly, 500 µl of 1:10 diluted sample was spread over the entire surface of the plate and incubated at 37 °C for 24 hrs. The average colony counts were expressed as CFU/1mL (Baylis and Patrick, 1999).

Fecal coliform (FC) counts were carried out according to the membrane filtration method suggested by APHA (APHA, 2005). Water samples were filtered through 0.45 µm nitrocellulose filters (Millipore, UK) and the filters were placed on mFC agar (Difco, Germany), and incubated at 44.5 °C for 24 h. Colonies that exhibiting any shades of blue were counted and expressed as CFU/100 ml surface water (APHA, 2005).

2.2.5 Statistical Analyses

The variations in physico-chemical and microbiological parameters (excluding fecal coliforms) across the sampling locations (spatial) and seasons were analyzed by one-way analysis of variance (ANOVA). Only the data, which was available for all three seasons, were used for statistical analysis. In total, physico-chemical and microbiological data (excluding fecal coliforms) collected from 19 locations in three seasons were used for further statistical analyses. The correlations between the bacterial counts (TVC, TC) and physicochemical parameters of water samples were determined by Spearman's non-parametric rank correlation test. The data transformation and the statistical analyses were performed using SPSS 22 version software (SPSS Inc., Chicago, IL, USA).

2.2.5.1 Cluster Analysis

Cluster analysis (CA) is a common technique for statistical data analysis and exploratory data mining applied in many fields of research as well as for water quality assessment (Kazi et al., 2009, Shrestha and Kazama, 2007). With the aim of studying the spatial variability of water quality in Tiaoxi River basin and grouping similar sampling locations, hierarchical

agglomerative clustering was performed using the normalized dataset. Ward's method of linkage with squared Euclidean distances was used as a measure of similarity (Shrestha and Kazama, 2007). Previous studies indicated that CA reliably classifies surface water quality and the results can be used as a guide for developing sampling strategies for the future (Alberto et al., 2001, Singh et al., 2004).

2.2.5.2 Principal Component Analysis /Factor Analysis

Principal component analysis (PCA)/Factor analysis (FA) provides information on the most meaningful parameters which describe whole dataset rendering data reduction with minimum loss of original information (Helena et al., 2000). It is a powerful technique for pattern recognition that attempts to explain the variance of a large set of inter-correlated variables and transforming into a smaller set of independent (uncorrelated) variables called principal components (Ouyang, 2005). In order to classify the variations of water quality indicators (thirteen in total) namely water temperature, pH, EC, TN, TP, NO_3-N, NO_2-N, PO_4-P, NH_4-N, TOC, Chl *a*, TVC and TC were used. PCA was executed using normalized variables to extract significant principal components (PCs) and these PCs were subjected to the varimax rotation (raw) generating factors to further reduce the contribution of variables with minor significance (Abdul-Wahab et al., 2005, Shrestha and Kazama, 2007).

2.3 Results and Discussion

Since this chapter has a lot of statistical analysis, the results and discussion sections of this chapter were combined together in order to give a better explanation.

2.3.1 Physico-chemical and Microbiological Parameters

The results (range values) of physico-chemical and microbiological analyses from 25 sampling locations within Tiaoxi River across three seasons are summarized in Table 2.2. The complete

results of physico-chemical and microbiological analyses from 25 sampling locations within Tiaoxi River for three seasons were presented in Supplementary Table S2.1.

2.3.1.1 Physico-chemical Parameters

The surface water temperature was in the range of 22.3-26.6 °C in autumn, 6–8.8 °C in winter and 29-31.2 °C in summer. Water samples in all locations were within the pH range set by Ministry of Environmental Protection (MEP), People Republic of China (PRC) for surface water and also natural waters pH limits (6.5–8.5) set for aquatic life and irrigation purposes (Chapman, 1996). Most natural water systems require a pH range of 6.5–8 to support a diverse aquatic population (Pearce et al., 1999). Significant statistical differences in pH were observed seasonally ($P<0.05$) however, no spatial variation was observed. EC values were within the range of 124-400 µS/cm in all the locations and the values varied significantly with seasons ($p<0.05$) (Table 2.2). For most freshwaters, the EC ranges from 10 to 1000 µS/cm and elevated levels of above 1000 µS/cm can be seen in polluted water or water bodies that receive large quantities of land runoff (Chapman, 1996). In streams and Rivers, the conductivity is affected by various factors such as the type of soils, bedrocks, and presence of inorganic dissolved solids. The sewage or wastewater could raise the conductivity due to the presence of chloride, phosphate, and nitrate (USEPA, 2012).The seasonal and spatial variations in TN, TP, NO_3-N, NO_2-N, PO_4-P, NH_4-N, TOC, and Chl $_a$ are shown in Figures 2.2 (A-D) and 2.3 (A-D). Among all these parameters tested only TP, NH_4-N and Chl *a* showed significant ($P<0.001$) seasonal and spatial variations (Table 2.2). The parameters such as TN, NO_3-N, NO_2-N, and TOC showed significant ($p<0.05$) spatial variations. Based on MEP guidelines, TN values were outside the acceptable limits (≥ 1mg/L) in all locations for all seasons. The highest TN values (4.13mg/L) were reported in samples collected from location 18 in Autumn 2014 (Fig. 2.2A), and this location is near to suburban/industrial area (junction of ChangXing and Tiaoxi River) with ferry docking facilities. The sources for elevated levels of TN in water bodies include

Table 2.2 Summary of the results of physico-chemical and microbiological characteristics of Tiaoxi River water with statistical analyses.

Parameters	Acceptable Range	Range (Minimum – Maximum)			P value	
	(by MEP)	Autumn 2014	Winter 2015	Summer 2015	Season	Spatial
WT (°C)	-	22.8-26.6	6-8.8	27.2-30.8	0.0001*	0.712
pH	6.5-8.5	7.2 – 8	7.4 – 7.9	7.3 – 7.9	0.112*	0.4546
EC (µS/cm)	-	153 – 400	164 – 356	124 – 234	0.0102*	0.6564
TN (mg/l)	≤1	1.78 - 4.13	1.3 - 4.03	1.88 - 3.11	0.5209	0.0001***
TP (mg/l)	≤0.2 (≤0.05[a])	0.07 - 0.18	0.07 - 0.19	0.08 - 0.14	0.0001***	0.0001***
NO_3-N (mg/l)	≤ 10	0.84 - 3.43	0.376 - 3.39	1.07 - 2.02	0.2464	0.0220*
NO_2-N (mg/l)	≤0.15	0.02 - 0.16	0.002 - 0.05	0.04 - 0.18	0.9987	0.0011**
PO_4-P (µg/l)	-	2.4 - 38.2	3.2 - 35.24	6.8 - 51.9	0.1324	0.0001***
NH_4-N (mg/l)	≤1	0.013 – 1	0.05 - 1.025	0.02 - 0.81	0.0001***	0.0001***
TOC (mg/l)	-	2.38 - 8.46	14.9 - 268.9	1.9 - 13.7	0.2929	0.0083**
Chl *a* (µg/L)	-	36.3 - 103.4	29.8 - 89.3	49.1 - 132.6	0.0001***	0.0001***
TVC (Log10 CFU/mL)	-	3.57 – 4.28	3.06 – 4.34	3.60 – 4.19	0.3078	0.2454
TC (Log10 CFU/mL)	1	1.60 – 3.30	2.0 – 3.31	2.22– 3.61	0.328	0.0025**
FC (Log10 CFU/100mL)	-		2.0 – 3.45	1.69 –3.62	-	0.0055**

[a] for Lakes, * Statistically significant difference at $p < 0.05$; ** Statistically significant difference at $p < 0.01$; *** Statistically significant difference at $p < 0.001$; MEP: Ministry of Environmental Protection, PR China.

runoff from agricultural croplands and animal manure, discharge from wastewater treatment plants and leakage from septic tanks (USEPA, 2013b). Presence of elevated levels of TN and ammonia in water are considered indicative of freshly polluted water by environmental management engineers (Farhadinejad et al., 2012). As per MEP guidelines, the acceptable TP levels for class III water bodies is <0.2mg/L but for Lakes and reservoirs, the TP levels should be <0.05mg/L. Here, TP levels were outside the acceptable range in Location 1 (Taihu Lake, ~1km inside from Taihu Lake/Tiaoxi River junction) and the levels were closer to acceptable limits in location 16 for autumn (175.58 µg/L) and winter (187.42 µg/L) seasons (Fig.2.2B). Location 16 is a suburban mixed residential and business area and the samples were collected at a junction between the main River and a canal that connects to Taihu Lake (Table 2.1). Zheng et al. (2017) reported higher levels of TP and TN in catchment 8 (urban land) of East Tiaoxi River which has two WWTPs out of which one is located near sampling locations 15 and 16 of the current study. The higher levels of TP observed at location 16 could be due to effluents from WWTP. As stated above, the concentrations of TP showed significant seasonal and spatial variation ($p<0.001$) and TP levels were comparatively high for most of the locations in winter 2015.

Wang et al. (2007) reported similar TP levels for Rivers surrounding Taihu Lake (Wang et al., 2007b). Possible runoff from fertilized lawns and cropland, animal manure and also domestic sewage entry into the water are likely causes (Yan et al., 2015). NO_3-N levels were within acceptable limits (<10mg/L) as suggested by MEP but elevated levels of NO_2-N (>0.15 mg/L) were observed in location 3 of autumn season and location 15 and 16 during the summer season (Fig. 2.2 C-D). Sources of NO_2-N include human sewage, livestock manure, fertilizers and erosion of natural deposits (Lucassen et al., 2004). Location 3 is a fishing village where people live on boats stationed at this location. Locations 15 and 16 are in a suburban area and the sampling was conducted in a junction between Tiaoxi River and a canal that connects to Taihu

Lake where boats/ferries were docked. The higher levels of NO_2-N observed in these three locations may be due to the entry of human sewage into the water. Previous study showed that presence of higher concentrations of NO_2-N in water is a potential problem due to its toxicity to human (more potential health effects are seen in infants) and livestock when consumed (Fewtrell, 2004).

NH_4-N levels were outside standard limits (>1 mg/L) in location 16 in autumn and at location 24 in winter (Fig. 2.3 B). NH_4-N enters into water mostly from anthropogenic sources such as human sewage, municipal effluent discharges, livestock manure, and agricultural runoff. Elevated levels of NH_4-N in surface water primarily exerts toxic effects on the higher aquatic organisms such as fish and shrimps (USEPA, 2013a). Xu et al. (2009) reported similar results for surface water quality in the Taihu watershed (Xu et al., 2009). There are no specific standards for PO_4-P as per MEP, PR China but a concentration of <20µg/L is commonly present in streams and Rivers. Elevated levels of >20µg/L indicates pollution and can lead to excessive algal growth (Shock and Pratt, 2003). In the present study, PO_4-P levels were high in all locations (Fig. 2.3A) on one or more occasions, however, no significant correlation between PO_4-P and Chl *a* was observed (Table 2.2). Similarly, for TOC there are no specific standards set by MEP, PRC. Both PO_4-P and TOC levels showed statistically significant P<0.05) spatial variation (Table 2.2, Fig. 2.3 A &C). Most of the surface waters with low nutrients levels have Chl *a* levels of <2.5µg/L but higher levels can be seen if there is high nutrient availability (Chapman, 1996). In the current study, all locations showed higher Chl *a* levels in all seasons indicating high algal growth in Tiaoxi River water (Fig. 2.3 D). Chlorophyll *a* levels were high in the summer season followed by autumn and winter, which can be correlated with warm temperature in summer and autumn seasons and availability of nutrients. In general, the concentrations of Chl *a* was high if the location has high TP

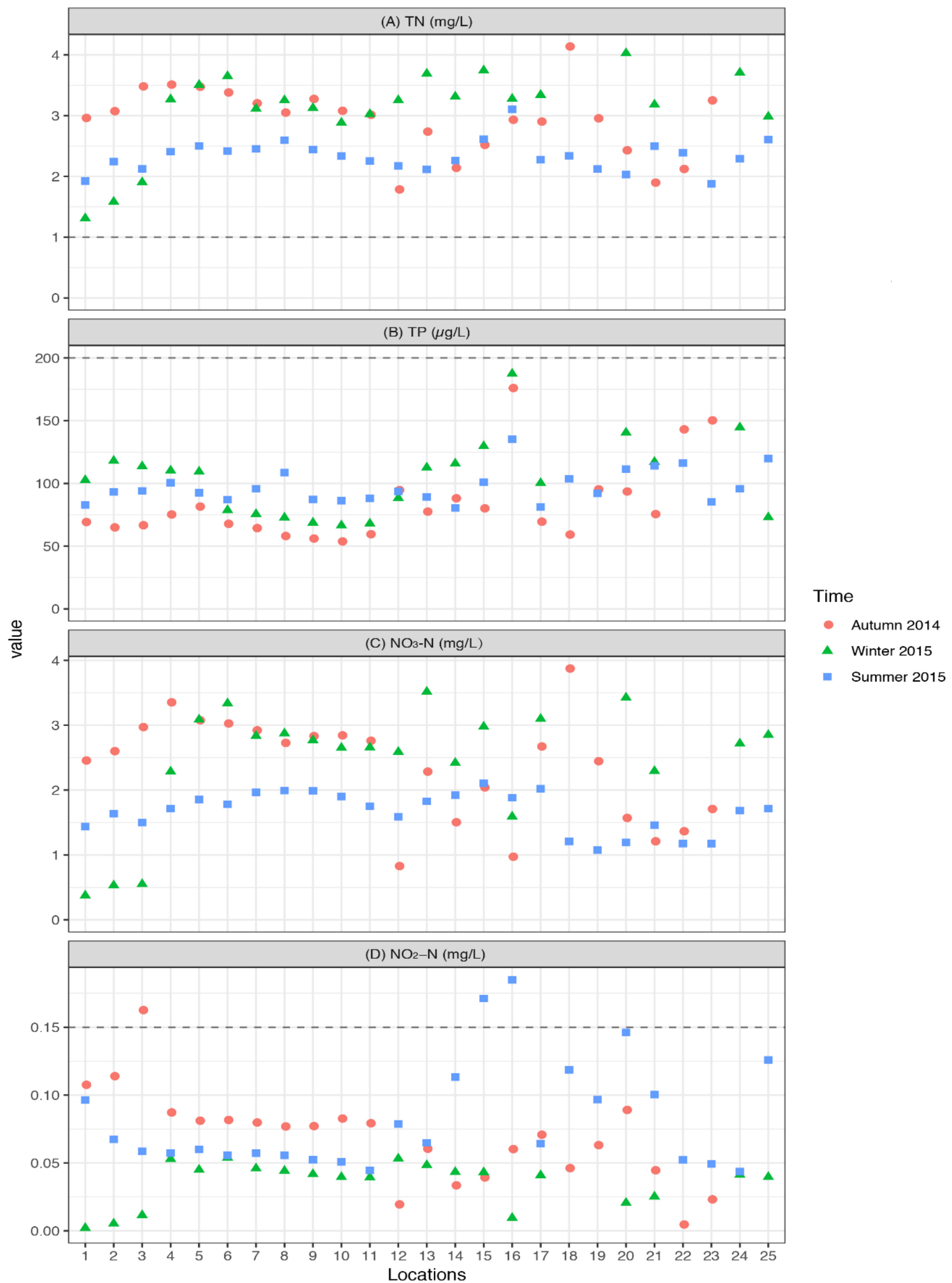


Figure 2.2 Spatial and temporal variations in physico-chemical parameters: Total Nitrogen (A), Total phosphorus (B), Nitrate-N (C), Nitrite-N (D), "----- line" indicates acceptable range for parameter as suggested by MEP, China .

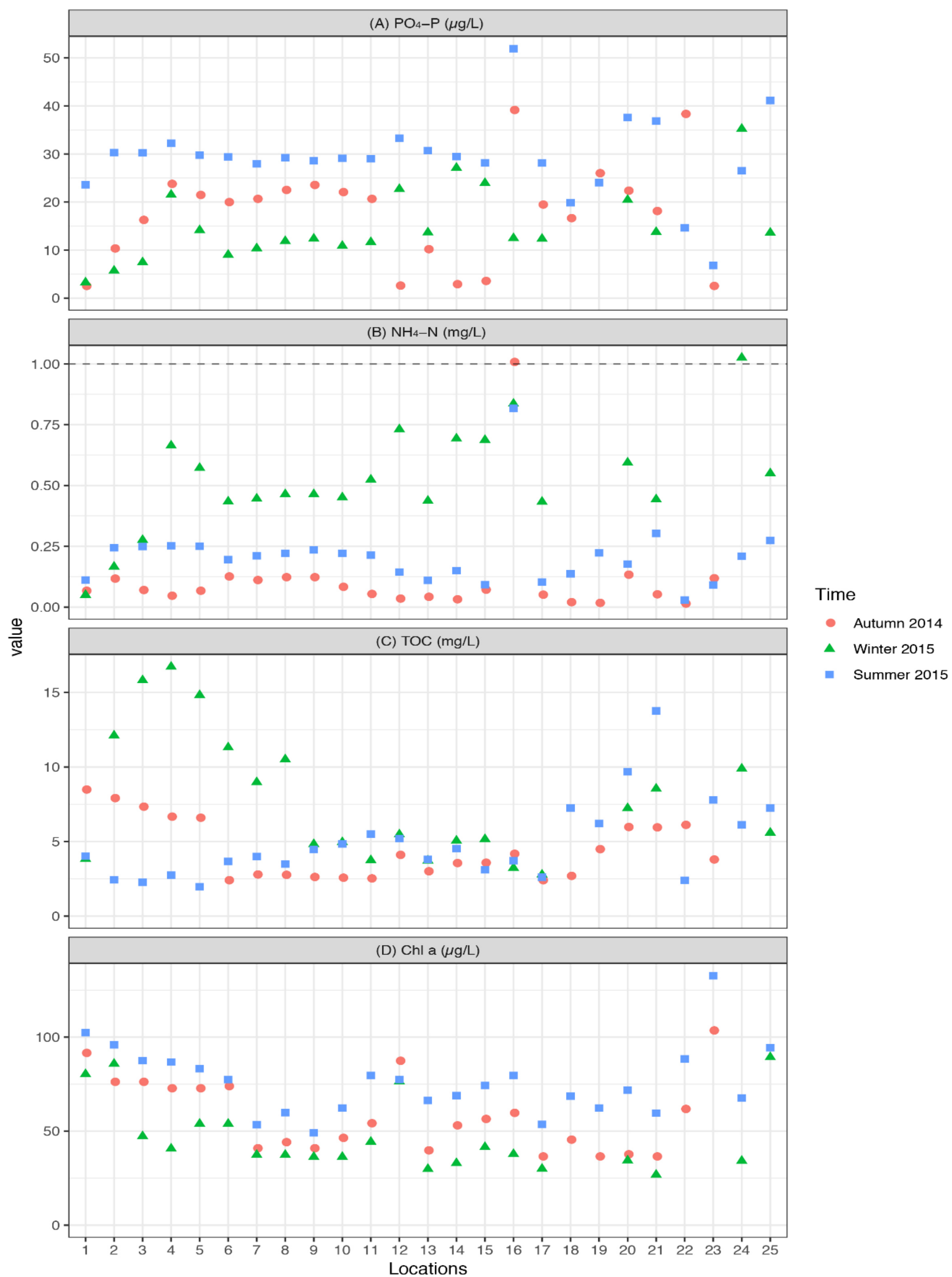


Figure 2.3 Spatial and temporal variations in physico-chemical parameters: Phosphate-P, (A), Ammonium-N (B), Total Organic Carbon (C), Chlorophyll *a* (D). "----- line" indicates acceptable range for parameter as suggested by MEP, China.

concentration. Xu et al. (2010) reported similar results for Chl *a* levels in Taihu Lake water (Xu et al., 2010). The prevalence of cyanobacteria and higher concentrations of Chl *a* during the summer season in Taihu Lake has been reported previously (Du et al., 2017, Xu et al., 2010). Highest Chl *a* levels were observed at locations 1, 12 and 23 during the summer season, and Chl *a* concentrations were statistically significant both spatially and seasonally ($p<0.05$) (Table 2.2).

2.3.1.2 Microbiological Parameters

TVC was carried out to enumerate aerobic/facultative anaerobic mesophiles in the surface water, primarily to determine whether these counts showed any relationship with physico-chemical parameters and coliform counts. TVC values neither showed any seasonal or spatial significance statistically nor followed the similar trend as total coliform and fecal coliform numbers (Table 2.2, Fig. 2.4A). As per MEP standards, the suggested standard limit for total/fecal coliforms for level III water bodies is <10000/L (or <10/ml) but elevated levels of total coliforms were observed in all locations for all seasons, and much higher levels were observed in seven locations (locations 2, 3, 5, 12, 15, 16 and 17) on one or more occasions with highest at location 16 (3.61 Log10 CFU/mL) during the summer season (Fig. 2.4B). The results correlate well with the land use pattern or possible mixing of waste into the above locations where boats/ferries were docked or leakage of waste into the river through human activities was observed. Hagedorn and Liang (2011) also indicated a serious fecal contamination of Tiaoxi River and reported higher levels (2.54 log10 CFU/ml) of *E.coli* for the water samples collected near FENGKOU drinking water station (Hagedorn and Liang, 2011). TC showed statistically significant ($p<0.005$) differences between locations (Table 2.2, Fig. 2.4C). Previously, total coliforms were considered as bacterial water quality indicators to assess fecal contamination in recreational waters in the USA, as required by the Beaches Environmental Assessment and Coastal Health Act (BEACHAct, 2000) to reduce health risks. However, it

was reported that some members of the coliform group live in the environment (i.e. outside of the gastrointestinal tract), which may show a false indication for fecal contamination in water (Pisciotta et al., 2002). Therefore, TC counts are no longer used as an indicator for recreational waters as they are widespread in nature, but are still used to assess drinking water quality (USEPA, 2012). Fecal coliform (FC) counts are used as guidelines for microbial water quality to assess fecal contamination. In the present study, most of the locations showed higher levels of FC (>400CFU/100ml) in winter and summer seasons (Fig. 2.4C) as compared to USEPA standards, however, no guidelines were suggested by MEP for FC in surface water in China. High FC count was observed in five locations (Locations 2, 3, 5, 12, 16) in one or more occasions with the highest at location 16 (3.62 log10 CFU/100ml) during the summer season. As indicated previously, these are the locations near either the residential areas where people live on boats without adequate sanitation facilities or urban residential areas with multiple waste inputs into the Rivers such as leakage of waste from unknown sources. The higher levels of FC observed in these locations could be correlated with the discharge of effluent from WWTP located near locations 15 and 16. Only some fecal coliforms are pathogenic and a previous study showed that FC presence does not always correlate with pathogen presence (Schriewer et al., 2010). However, high FC count implies impaired water quality and increased risk associated with the presence of pathogen (Haller et al., 2009). FC levels were comparatively higher in summer compared to winter and this may be due to the runoff and heavy rainfall that occurred before summer sampling in 2015. The increased concentrations of fecal coliforms after rainfall events were widely acknowledged in the scientific literature (Ackerman and Weisberg, 2003, Muirhead et al., 2004, Carroll et al., 2006). The higher levels of FC observed could also be due to warm temperatures, which can facilitate FC bacteria accustomed to such conditions (Heaney et al., 2015). FC numbers showed significant spatial ($p<0.05$) variation.

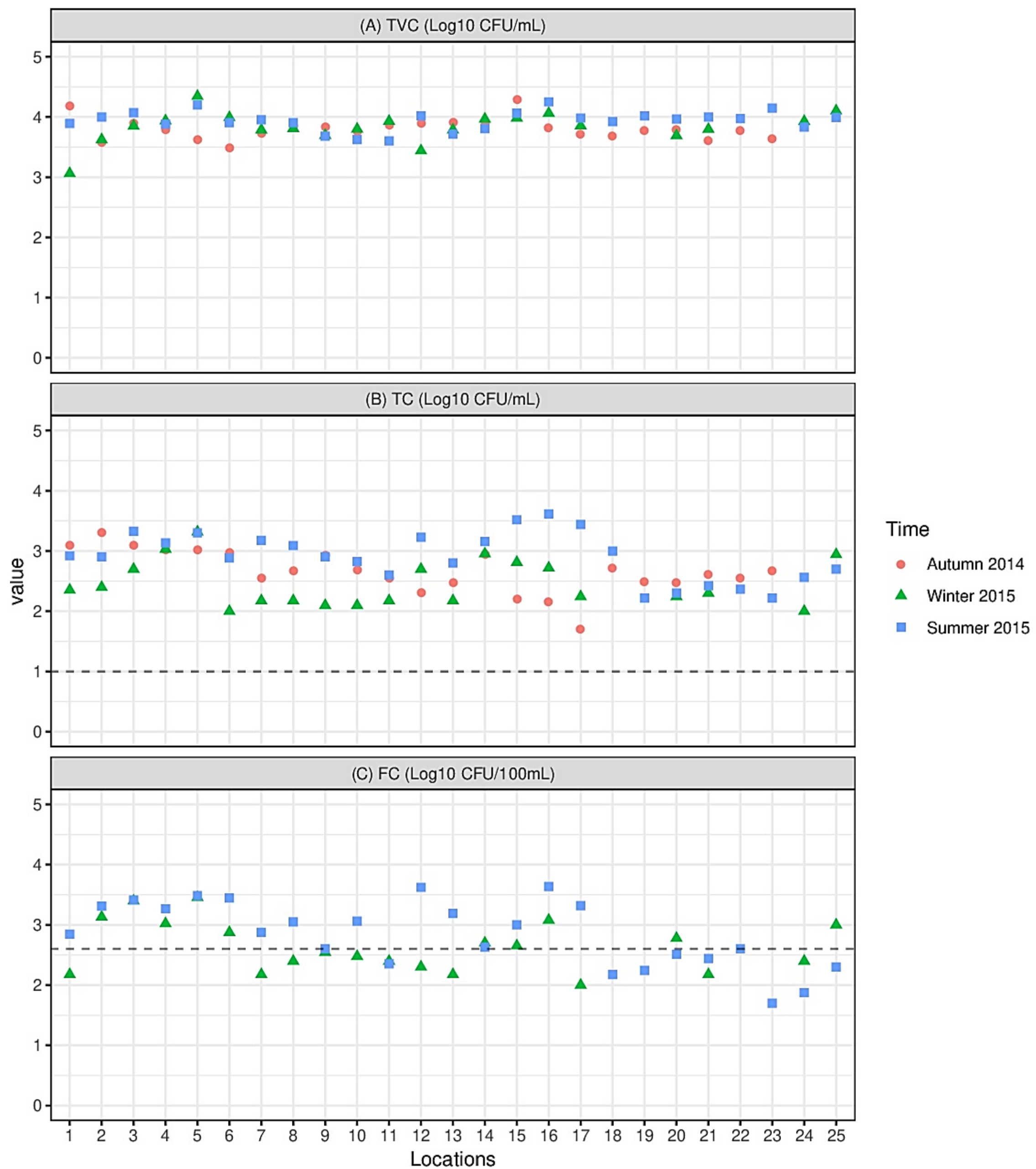


Figure 2.4 Spatial and temporal variations in microbiological parameters: Total viable count (A), Total coliforms (B), Fecal coliforms (C). "----- line" indicates acceptable range for parameter as suggested by MEP, China or USEPA.

2.3.2 Correlation between Variables

Results of correlation analysis (Table 2.3) showed that TN has a moderate positive correlation with NO_3-N (r=0.651) and a weak correlation with PO_4-P (r=0.486) and NH_4-N (r=0.448), which suggests that NO_3-N could be the major contributor of TN (Mei et al., 2014). TP has a

moderate positive correlation with NO_2-N and NH_4-N, indicating probably a common source of organic contamination such as untreated sewage (Vieira et al., 2012). TP also has a strong positive correlation with EC and moderate positive correlation with the microbiological parameters (TVC). Microbiological parameters (TVC and TC) showed a positive correlation with each other (r=0.796) and also with some of the nutrients (TP and NO_2-N) indicating possible microbial contaminations from sewage (Nnane et al., 2011). pH had a strong positive correlation with EC, and NO_3-N had a strong negative correlation with EC. No obvious correlation ($r<0.4$) was observed among the other parameters.

2.3.3 Cluster Analysis for Spatial Grouping

Cluster analysis (CA) was applied to group sampling locations with similar water quality characteristics. Dendrogram was generated by CA grouped 19 locations into three clusters at (Dlink/Dmax) <60 (Fig. 2.5). CA results are convincing, as the generated clusters share similar characteristic features and land use pattern. Based on the physico-chemical and microbiological results, each cluster was classified into respective pollution categories (Table 2.4). Cluster 1 includes 8 locations (locations 6-11, 13 and 17) and consists of mixed land use, either rural or urban/suburban residential areas with little industrial activity, corresponding to a relatively low level of pollution. Cluster 2 comprises 4 locations (4, 5, 15 and 16) which are mostly the junctions of east and west Tiaoxi River or other streams. These locations are predominantly close to urban and semi-urban residential areas with a large-scale business, ferry transportation, and ferry docking activities, and in some of these locations entry of wastewater in the River was noticed during sampling (Fig. 2.6). This cluster was classified as highly polluted based on the physico-chemical and microbiological results. Cluster 3 comprises 7 locations (1, 2, 3, 12, 14, 20 and 21), and includes mixed land use and can be categorized as moderately polluted locations based on the physico-chemical and microbiological analysis. Sampling location 20 and 21 were close to sparse residential/industrial areas and locations 2 and 3 are residential

Table 2.3 Spearman's correlation coefficient (r) values observed between different water quality parameters.

	WT	pH	EC	TN	TP	NO_3-N	NO_2-N	PO_4-P	NH_4-N	TOC	Chl *a*	TVC	TC
WT	1												
pH	0.03	1											
EC	0.22	0.80^{**}	1										
TN	-0.15	-0.35	-0.35	1									
TP	0.17	0.74^{**}	0.78^{**}	0.08	1								
NO_3-N	-0.16	-0.61^{**}	-0.73^{**}	0.65^{**}	0.45^{*}	1							
NO_2-N	0.34	0.16	0.46^{*}	0.23	0.54^{*}	-0.24	1						
PO_4-P	0.27	0.06	-0.02	0.49^{*}	0.28	0.10	-0.01	1					
NH_4-N	-0.17	0.25	0.23	0.45^{*}	0.53^{*}	-0.08	0.16	0.73^{**}	1				
TOC	0.57^{*}	-0.15	0.24	-0.08	0.21	-0.16	0.23	0.15	0.12	1			
Chl *a*	-0.02	0.06	0.33	-0.23	0.12	-0.41	0.34	-0.37	0.01	0.33	1		
TVC	0.09	0.29	0.30	0.27	0.50^{*}	-0.13	0.59^{**}	0.01	0.32	0.125	0.37	1	
TC	-0.03	0.18	0.19	0.24	0.36	-0.12	0.50^{*}	-0.11	0.25	0.037	0.47^{*}	0.80^{**}	1

* The correlation was significant at $p < 0.05$; ** The correlation was significant at $p < 0.01$.

areas where a few people are living on boats (Fig. 2.6). The CA enabled us to categorize sampling locations based on water quality so that in future studies, the number of sampling locations can be minimized for cost-effective monitoring of water quality in Tiaoxi River by choosing a few locations from each cluster based on the distance distribution and pollution levels in those locations. Previous studies have reported that a similar strategy has been successfully applied in water quality monitoring programs elsewhere (Pejman et al., 2009, Phung et al., 2015, Shrestha and Kazama, 2007, Singh et al., 2005), and the Tiaoxi River Taihu catchment is therefore similarly amenable to this rational approach.

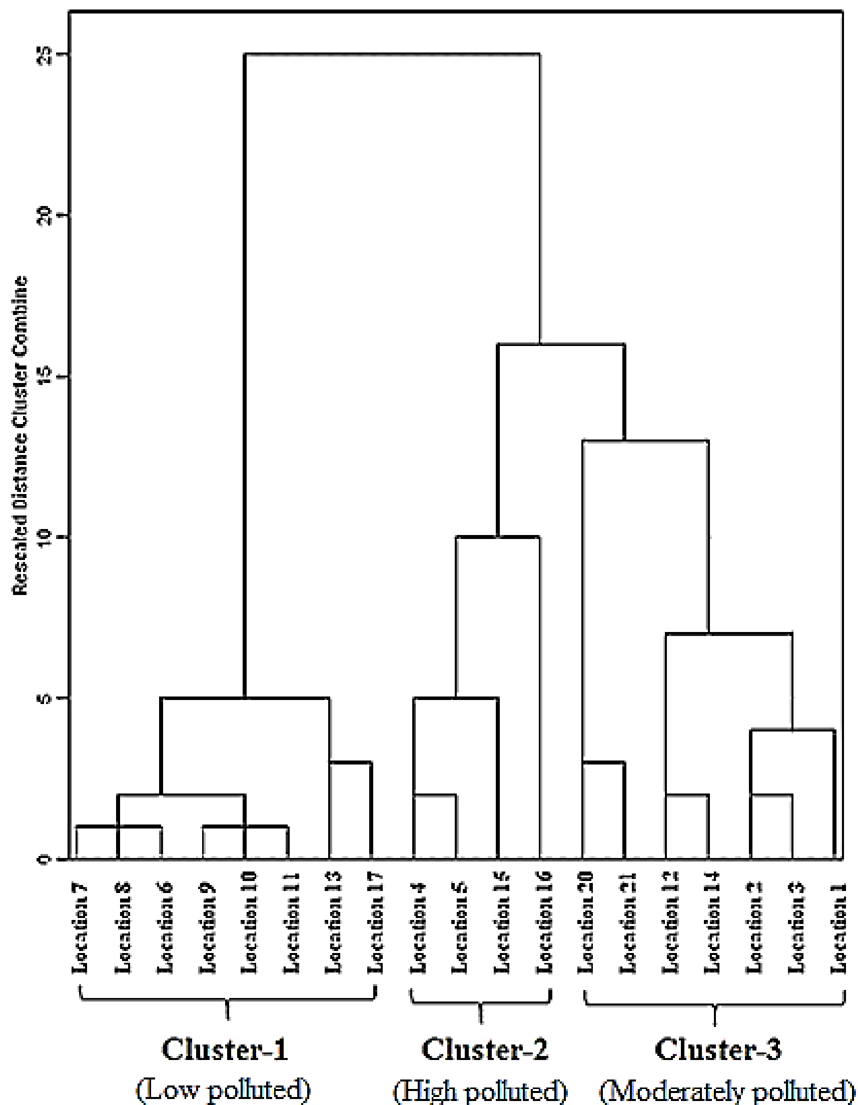


Figure 2.5 Dendrogram showing clustering of 19 sampling locations based on surface water quality characteristics of the Tiaoxi River.

Table 2.4 Range of physico-chemical and microbiological characteristics for locations in which pollution is classified as relatively low, moderate or high.

Parameter	Cluster 1- Relatively low pollution (Locations: 6,7,8,9,10,11,13,17)	Cluster 2- Highly Polluted (Locations: 4,5,15,16)	Cluster 3- Moderately polluted (Locations: 2,3,12,14,20,21)
pH	7.25-7.67	7.25-7.88	7.3-7.86
Conductivity	125-257	143-400	136-356
TN (mg/L)	2.25-3.69	2.40-3.74	1.31-3.47
TP (µg/L)	53.4-112.6	74.89-187.4	64.6-140.4
NO_3-N (mg/L)	1.75-3.51	0.96-3.34	0.37-3.42
NO_2-N (mg/L)	0.03-0.08	0.01-0.18	0.01-0.16
PO_4-P (µg/L)	9.01-30.7	3.51-50.9	2.46-37.6
NH_4-N (mg/L)	0.04-0.52	0.04-1.01	0.03-0.73
TOC (mg/L)	2.39-11.31	1.97-16.71	2.27-15.81
Chl *a* (µg/L)	29.88-79.6	37.82-86.7	26.8-102.3
TVC (CFU/mL)	3040-9800	4190-22050	1150-15500
TC (CFU/mL)	50-2766	145-4133	175-2133
FC (CFU/100mL)	100-2800	450-4325	150-2600

Figure 2.6 Sampling locations representing high (HP), medium (MP) and low (LP) pollution risk locations.

2.3.4 Principal Component Analysis /Factor Analysis for Source Identification

Principal component analysis (PCA)/Factor analysis (FA) was performed using log-transformed data to identify the factor(s) that influence the water quality during the entire sampling period and in seasons (autumn, winter, and summer). Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were carried out to verify the suitability of data for PCA/FA. KMO value of 0.5 or more is required to perform PCA and lower KMO value indicates that the dataset is not suitable for the PCA (Ogwueleka, 2015).

In this study, the KMO value for the entire dataset was 0.53, however, Bartlett's test gave a P value of <0.001 indicating the suitability of the data for PCA. The significance of the factor is evaluated by eigenvalue in PCA; the higher the eigenvalues, the higher the significance of factors, with 1.0 or greater eigenvalues considered significant (Shrestha and Kazama, 2007). The PCA for the entire dataset yielded four PCs (with eigenvalues ≥ 1), which explained over 83% of the total variance in the dataset. The variable loadings on varimax rotated PCs for the entire data set is provided in Table 2.5. Variable loading is classified as 'strong', 'moderate' and 'weak', corresponding to their absolute loading values of >0.75, 0.75-0.50 and 0.50-0.30, respectively (Liu et al., 2003).

The first component (VF1) accounted for 32.2% of total variance and has a strong positive loading for pH, EC and TP, and strong negative loading for NO_3-N indicating variability in physico-chemical sources (Table 2.5). Normally EC is used to indicate natural pollution and can be due to soil erosion or weathering effects on water quality during seasonal changes (Ogwueleka, 2015). This component also suggests that most of the variation is due to pH and EC changes. The second component (VF2) is responsible for 25.4% of the total variance and showed strong positive loading for TN, PO_4-P, and NH_4-N. This component also gave moderate negative loading to Chl *a* indicating nutrient pollution and this could be interpreted as influences from agricultural and domestic waste (Ruzdjak and Ruzdjak, 2015). The third

component (VF3) explained 14.2% of the total variance and has a strong positive loading for TVC and TC. This component also has a moderate positive loading for Chl *a* and represents influences of mainly microbial origin. The microbial factor TC can be associated with sewage pollution in the River. The fourth component (VF4), accounting for 12% of the total variance has a strong positive loading for WT and TOC. This component represents physicochemical sources and could be interpreted as influences from organic pollution caused by domestic and industrial discharges. Similar results have been reported by other authors for water quality assessment by PCA/FA (Mei et al., 2014, Ruzdjak and Ruzdjak, 2015).

Table 2.5 Loadings of thirteen variables on Varimax rotated principal components.

	Components			
	VF1	**VF2**	**VF3**	**VF4**
WT	0.134	0.121	-0.173	**0.857**
pH	**0.911**	0.131	-0.003	-0.160
EC	**0.926**	-0.086	0.182	0.207
TN	-0.474	**0.837**	0.181	-0.056
TP	**0.755**	0.443	0.393	0.172
NO_3-N	**-0.808**	0.388	-0.123	-0.195
NO_2-N	0.385	0.279	0.495	0.392
PO_4-P	0.088	**0.938**	-0.024	0.127
NH_4-N	0.175	**0.866**	0.195	-0.076
TOC	0.013	-0.170	0.114	**0.861**
Chl *a*	0.145	**-0.602**	**0.589**	0.168
TVC	0.183	0.183	**0.848**	0.010
TC	0.030	-0.030	**0.885**	-0.140
Eigenvalue	4.186	3.297	1.843	1.567
% Total variance	32.197	25.365	14.174	12.054
Cumulative % variance	32.197	57.562	71.735	83.786

Values in bold indicate strong or moderate loadings.

2.5 Conclusions

- Physico-chemical analysis results indicated that TN, TP, NO_2-N, and NH_4-N were the major nutrient pollutants of the Tiaoxi River.
- The microbiological results indicated that TC and FC counts were high in 15 locations, suggesting fecal contamination of those locations.
- Cluster analysis grouped sampling locations into three clusters and these three clusters can be classified as relatively low, moderate and high pollution areas based on the land use patterns, physico-chemical and microbiological data.
- Principal component analysis (PCA) indicated that pH, conductivity, TP, and NO_3-N were the key parameters responsible for variations in Tiaoxi water quality.
- Based on physico-chemical and microbiological results, 15 locations were assessed as possible fecal polluted locations, and used for microbial source tracking study to identify sources for fecal contaminations.

CHAPTER-3

Next Generation Sequencing approach to characterize Fecal and Potential Pathogenic Bacterial Diversity in Tiaoxi River water (Taihu Watershed)

Abstract

This study was aimed at characterizing bacterial diversity including fecal and potential pathogenic bacteria in Tiaoxi River water by a next-generation sequencing (NGS) approach and to examine the relationship between physico-chemical parameters and bacterial community composition. Forty-five water samples (collected from 15 locations in three seasons) were used for bacterial community analysis along with fecal samples collected from 7 different hosts (human and animal fecal samples) and sewage samples from a wastewater treatment plant. DNA was extracted from water, fecal and sewage samples and bacterial diversity and community composition was studied using the Illumina MiseqPE250 platform to sequence the V3-V4 region of 16S rRNA genes. The bacterial diversity results showed that Proteobacteria, Actinobacteria, Bacteroidetes, and Cyanobacteria were dominant phyla in most of the Tiaoxi River water samples, although their percentage distribution varied among the samples tested. In fecal samples, phyla Bacteroidetes, Firmicutes, and Proteobacteria were abundant, whereas wastewater samples were dominated by Proteobacteria, Bacteroidetes, and Chloroflexi. The hierarchical cluster analysis and principal coordinate analysis (PCoA) indicated that the community composition was significantly different between water, fecal and sewage samples. At the genus level, ~180 different genera were detected in the River water samples among which five bacterial genera of fecal origin and seven potential pathogens were detected in many locations. Members of *Bacteroides* and *Prevotella* were frequently observed in many locations which indicate possible human and animal fecal pollution in river water. Genus *Dorea* was observed in some of the locations and their presence in the water indicates potential human fecal pollution. Potentially pathogenic bacterial genera such as *Arcobacter, Aeromonas, Enterococcus,* and *Shigella* were found with higher abundance (>0.10%) in some of the locations. The redundancy analysis (RDA) results showed that pH, conductivity and temperature were the main environmental factors that contributed to shaping the bacterial

composition of water samples although a few other nutrients parameters were also found to have a seasonal influence. Overall, the results obtained by NGS indicate possible human and animal fecal pollution in several locations in the Tiaoxi River.

3.1 Introduction

Waterborne diseases are one of the major concerns of the world and cause about 2.2 million deaths annually, with the majority of these deaths 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, which are regarded as major challenging environmental issues in China (Jiang, 2009, Han et al., 2016). Although drinking water quality is improving rapidly in recent years in China, it has been reported that 200 million people still consume unsafe drinking water and 60,000 people die every year due to water related issues such as diarrheal diseases in China (Qiu, 2011, Han et al., 2016). As diarrheal diseases are primarily caused due to contamination of water source with human or animal feces transmitting pathogenic microorganisms (WHO, 2017), monitoring of fecal contamination and pathogenic microorganisms in waters used for human consumption and recreational activities have become mandatory to reduce human health risk.

Most of the conventional fecal monitoring studies rely on enumeration of fecal indicator bacteria (FIB), which are considered as the "golden standard" to assess the microbiological water quality and pathogen presence in environmental waters (Savichtcheva and Okabe, 2006, Figueras and Borrego, 2010). However, FIB can originate from multiple hosts and it has been reported that human health risks varies on the exposure to the nature of host fecal source. For example, exposure to human fecal contamination is considered as a serious concern as human feces contains many potentially pathogenic microbes. Conventional FIB enumeration methods do not determine the origin of the fecal source, and previous reports showed poor or no correlation between FIB and pathogenic organisms (Fremaux et al., 2009, Shahryari et al., 2014). Therefore, several qualitative and quantitative "Microbial Source Tracking (MST)" methods were developed in recent decade to overcome this limitation and MST methods focused on determining the origin of the fecal sources (Seurinck et al., 2005, Kildare et al.,

2007, Mieszkin et al., 2009, Green et al., 2014). In addition, several quantitative polymerase chain reaction (qPCR) methods have been widely used to determine the presence and abundance of pathogens (Ahmed et al., 2009b, Sidhu et al., 2012, Oster et al., 2014), and these methods can identify the targeted pathogens that are specifically chosen for monitoring, but not all the pathogens. As a result, it is challenging to monitor the wide-range of pathogens in a watershed with culture-based or culture independent qPCR based methods. Therefore, a comprehensive monitoring method is required to characterize the diversity of microorganisms at a watershed and to ascertain the presence of dominant pathogens in order to assess the public health risks. In this regard, next-generation sequencing (NGS) method targeting the 16S rRNA gene has gained attention to explore the diversity of bacterial communities and their influence on microbial water quality (Staley et al., 2013, Ibekwe et al., 2013). Although community based NGS methods (OTU comparison between fecal source and environmental samples) were proposed for MST study (as culture independent library dependent MST method) (Unno et al., 2010, Jeong et al., 2011), they are considered as a qualitative method for assessing fecal pollution since the OTU comparison results shows discrepancies with 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 spatio-temporal limitations of SourceTracker method pointed to the necessity of optimization and validation prior to their application in a new geographical area (Ahmed et al., 2015a, Henry et al., 2016, Brown et al., 2017, Staley et al., 2018). However, the NGS based microbial community analysis approach is still considered as a valuable tool for premilinary investigation of water samples to assess public health risk associated with fecal contamination or pathogens (Tan et al., 2015). Although quantitative (qPCR) methods are appropriate for assessment of fecal contamination or pathogens, the NGS method can evaluate bacterial

diversity (including fecal and pathogenic bacteria) of water or environmental samples and their relative abundance, which could provide useful information to prioritize more specific exposure assessment of suitable targets using quantitative (qPCR) methods (Tan et al., 2015). Most of the NGS studies monitoring microbial communities in water samples have relied on targeting hypervariable regions (V1, V3, V4, V6) of the 16S rRNA gene (Guo et al., 2013). The 454-pyrosequencing platform was preferred earlier to assess microbial communities in water samples however in recent years, the Illumina platform has gained much attention as several studies revealed it is superior over pyrosequencing and also 454 sequencing has been discontinued (Sinclair et al., 2015, Newton et al., 2015, Loman et al., 2012). Furthermore, the commonly used open source software packages for sequence analysis such as QIIME and MOTHUR have been upgraded to analyze pair end sequence data produced by Illumina sequencers, improving the performance (Kozich et al., 2013, Caporaso et al., 2010).

As indicated in the previous chapter(s), Taihu Lake is suffering from water quality issues such as eutrophication, cyanobacterial blooms and fecal pollution due to various discharges into the watershed and Tiaoxi River was reported as one of the main contributors. Fecal pollution is one of the major concerns to water bodies, particularly if water is used for drinking, recreation and aquaculture due to possible human exposure to pathogens. Although previous studies focused on assessing the microbial community in water and sediment samples of Taihu Lake (Wilhelm et al., 2011, Paerl et al., 2011, Cai et al., 2013, Zhao et al., 2017), the microbial water quality assessment and studies aiming at characterizing the microbial communities with respect to fecal and pathogen diversity at Tiaoxi River are very limited. Zheng et al. (2017) studied the antibiotic resistance genes (ARGs) pattern and analyzed conventional water quality factors in samples collected from East Tiaoxi River and indicated that ARGs variation at different catchment areas could be due to changes in bacterial communities . Therefore, the current study was aimed at assessing the diversity and relative abundance of bacteria including fecal and

potential bacterial pathogens in Tiaoxi River water by 16S rRNA genes targeted NGS. The specific objectives were to i) study the bacterial diversity in Tiaoxi River water including the seasonal variations, and ii) determine the relative abundance of bacteria of fecal origin and potential pathogens and iii) to assess the influence of environmental factors on bacterial diversity by statistical analysis.

3.2 Materials and Method

3.2.1 Collection of water and fecal samples

As stated in the previous chapter 2, 15 out of 25 sampling locations in the Tiaoxi River were identified as hotspots of fecal contamination, and samples collected from those locations were used further studies. The details of land use type around the selected sampling locations are given in Table 2.1. Sampling was carried out at these locations on three occasions: autumn 2014, winter 2015 and summer 2015 and 5L of surface water samples were collected in sterile polypropylene bottles and brought to the laboratory on ice. The processing of samples was carried out within 8 hrs of sampling. 250mL of water sample was filtered using 0.22-µm polycarbonate membrane filters (Millipore, UK) and filters were stored at -20ºC until DNA extraction.

A total of 120 fresh individual fecal samples from several hosts such as chicken (CK), cow (CW), dog (DG), duck (DU), goose (GO), human (HU), and pigs (PI) and primary effluents from wastewater treatment plants (WWTP) were collected from Huzhou (Zhejiang province) and Suzhou (Jiangsu province) areas. 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. Safety guidelines were given to them in handling the samples and their consent for use of samples in this study was obtained. Ethical approval for handling fecal samples in this study was acquired from Xian Jiaotong Liverpool University Research Ethics Committee. 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 Tiaoxi River water sampling locations in Zhejiang Province. All the samples were carried on ice to the laboratory and the samples were stored at -20ºC within 6 hours of sample collection. Approximately 0.5gms of different individual fecal samples of a host were pooled together to form a composite sample of a respective host. The composite samples were used for further DNA extractions. 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 (at 4000×g for 10min at 4ºC) and the DNA was extracted immediately.

3.2.2 Physico-chemical assessment

The physico-chemical assessment of the Tiaoxi River was reported in our earlier chapter and the physico-chemical data were used to evaluate their influence on bacterial diversity of Tiaoxi River. The parameters that were included for the analysis include conductivity (EC), pH, water temperature (WT), Total Nitrogen (TN), and Total phosphorous (TP), ammonia nitrogen (NH_4-N) and nitrite (NO_2-N).

3.2.3 Extraction of genomic DNA from water and fecal samples

Genomic DNA was extracted from membrane filters (water samples) using the PowerSoil DNA Isolation Kit (MoBio Inc., Carlsbad, CA). Initially, the membrane filters were cut into pieces, placed into the PowerBead tubes aseptically to extract the genomic DNA as instructed by the manufacturer.

DNA was extracted in duplicate for each composite fecal/sewage sample (respective host) using PowerFecal® DNA isolation kit (MoBio, Carlsbad, CA USA), according to the instructions provided by the manufacturer. Approximately 0.25 grams of composite feces or biomass from primary effluent samples were used for DNA extraction. To avoid any cross contamination with other hosts, DNA extraction was conducted for each type of fecal source

separately. The DNA concentrations and quality were assessed by using NanoDrop ND 2000UV spectrophotometer (Thermo Fisher Scientific., Vienna, Austria) and the DNA extracts were stored at -20°C until further analysis.

3.2.4 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 Illumina sequencing, sequence data processing and statistical analyses were carried out by outsourcing at the Shanghai Majorbio Pharmaceutical Technology Limited, China. Further data analysis, interpretation of results for NGS and preparation of thesis chapter were carried out by myself (PhD candidate), with the direction of my supervisors. The hypervariable region (V4 region) of bacterial 16S rRNA genes present in water and fecal DNA samples was amplified by PCR using a universal primer set as described elsewhere (Kozich et al., 2013). The barcoded 515F (5’-GTGCCAGCMGCCGCGGTAA-3’) and 806R (5’-GGACTACHVGGGTWTCTAAT-3’) was used fasor amplification. Triplicate PCR reactions were carried out for each sample with each reaction mixture contains 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.

3.2.5 Sequence data processing and analysis

The quality check for raw data sequence files from Illumina was carried out using FastQC software (Andrews, 2010). The raw sequences were demultiplexed, 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 (around 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 sequences obtained in this study were submitted to the National Center for Biotechnological Information (NCBI) Short Read Archive (SRA) database under the accession numbers SAMN09469451 to SAMN09469511.

3.2.6 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 compared the bacterial diversity between different samples (beta diversity) by using QIIME (Caporaso et al., 2010). The relationship between the environmental parameters and bacterial community in the samples was performed by redundancy analysis (RDA) by using R language vegan package (http://www.R-project.org/ 2013).

3.3 Results

3.3.1 Assessment of physico-chemical parameters

The detailed results of physio-chemical parameters are provided in Table S2.1 and results were discussed in detail in chapter 2. Among the tested parameters, conductivity and pH were in

acceptable limits as suggested by Ministry of Environmental Protection (MEP), China (MEP, 2016) but TN was higher than acceptable levels in all the locations during three occasions. NH_4-N, NO_2-N, and TP had surpassed acceptable levels in some locations. The relationship between these parameters and bacterial diversity are discussed in this chapter.

3.3.2 Quality check for NGS study:

A three-step quality control (QC) measures were applied to ensure the quality of the sequence data obtained in this study. In the first step, high-quality genomic DNA was used for NGS after confirming with spectrophotometric (Nanodrop) and agarose gel electrophoresis methods. In the second step, the amplicons generated after PCR amplification were assessed for appropriate sizes (~ 250bp) by 2% agarose gel and quantified to ensure the products were in appropriate concentration (Supplementary fig. S3.1 and Table S3.1). Finally, the real-time QC monitoring of sequencing run generated by the Illumina Miseq sequencing systems was done by using Sequencing Analysis Viewer (SAV, v2.4 version) and a probability of $Q20 \geqslant 90$ and $Q30 \geqslant 80$ was set as a selection criteria. The quality check for raw sequence data performed using FastQC software produced mean quality Phred score (error rate), GC%, the presence of ambiguous bases (N), adapter sequences and other statistics for each set of paired reads (Andrews, 2010). The mean Phred score was calculated to be above 37 and the percentage of ambiguous bases and remaining adapter sequences was found to be below 1.0 % cut off, indicating sequencing accuracy of 99.9% (Cliften, 2015).

3.3.3 Bacterial diversity and community composition in water and fecal samples

In total, 61 samples (45 water samples collected in three seasons and 16 host-specific fecal and wastewater samples) were used for sequencing and in-depth monitoring of bacterial diversity and composition by NGS. A total of 1,694,935 bacterial 16S rRNA reads were generated for water, wastewater and fecal samples with sequence libraries of size ranging from 18640 reads (goose fecal) to 37367 reads (water sample). Although different sequence size libraries were

generated for the water, wastewater and fecal samples, the sequence data were normalized by considering library with lowest sequence reads as cutoff and reanalyzing the data to show normal distribution of variance and for further comparison between samples (Ahmed et al., 2015a). The average number of OTUs, species diversity (Simpson and Shannon) indices and richness (ACE and Chao1) values for bacterial diversity observed in water, wastewater and fecal samples were given in Table 3.1 and a detailed diversity index along with number of OTUs for each sample was provided in supplementary files (Table S3.2). The rarefaction curves based on total number of OTUs with normalized sequence reads were presented in Fig. 3.1. For water samples (n=45), a total of 39192 OTUs were generated ranging from 564 to 1292) OTUs. The highest diversity and species richness was observed at location 20 (WIW20) and 21 (WIW21) during winter and lowest at location 2 (WIW2) and 3 (WIW3) as indicated by OTU richness, Chao and Shannon indices (Table. S3.1). Similarly, for fecal samples (n=14), 5254 OTUs were generated ranging from 138 to 640 OTUs (Table S3.2) 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 3.1). A total of 2360 OTUs were generated for wastewater samples (n=2) with an average of 1180 and these samples had the highest diversity when compared to fecal and water samples (Table 3.1).

The RDP classifier categorized all the OTUs of water into 20 bacterial phyla, however their relative abundance varied with the type of sample (Fig.3.1). Proteobacteria, Actinobacteria, Bacteroidetes, and Cyanobacteria are the most dominant phyla accounting for about 90% of total abundance in all location except in location 2 (WIW2) and 3 (WIW3) samples collected in the winter season. These two samples had Proteobacteria, Firmicutes, Bacteroidetes, and Cyanobacteria as abundant phyla. About 180 different genera were identified in Tiaoxi River water, although genera related to Cyanobacteria and Actinobacteria (hgcI_clade) showed higher relative abundance in all the locations of Tiaoxi River water, except location 2 (WIW2)

and 3 (WIW3) samples collected in the winter season (Table S3.3). Although different genera such as *Microcystis, Flavobacterium, Sediminibacterium* and *Fluviicola* were also abundant in Tiaoxi River water samples, this study was mainly focused on bacteria that are fecal associated or potentially pathogenic in nature. In the current study, several fecal-associated and potential pathogenic bacterial genera were detected in Tiaoxi River water samples, which are discussed in detail in the following sections.

In fecal samples, only 16 phyla were observed and the most abundant phyla were Bacteroidetes, Firmicutes, and Proteobacteria constituting 90% of relative abundance (Fig.3.1). At genus level, 143 genera were observed but *Prevotella, Bacteroides and Lactobacillus* were the most abundant genera in fecal samples (Table S3.3). In wastewater samples, 20 different phyla were observed but the phylum Proteobacteria was the most abundant followed by *Bacteroidetes* and *Chloroflexi.* Genera related to Proteobacteria such as *Dechloromonas* and *Arcobacter* were abundant in wastewater samples (Table S3.3).

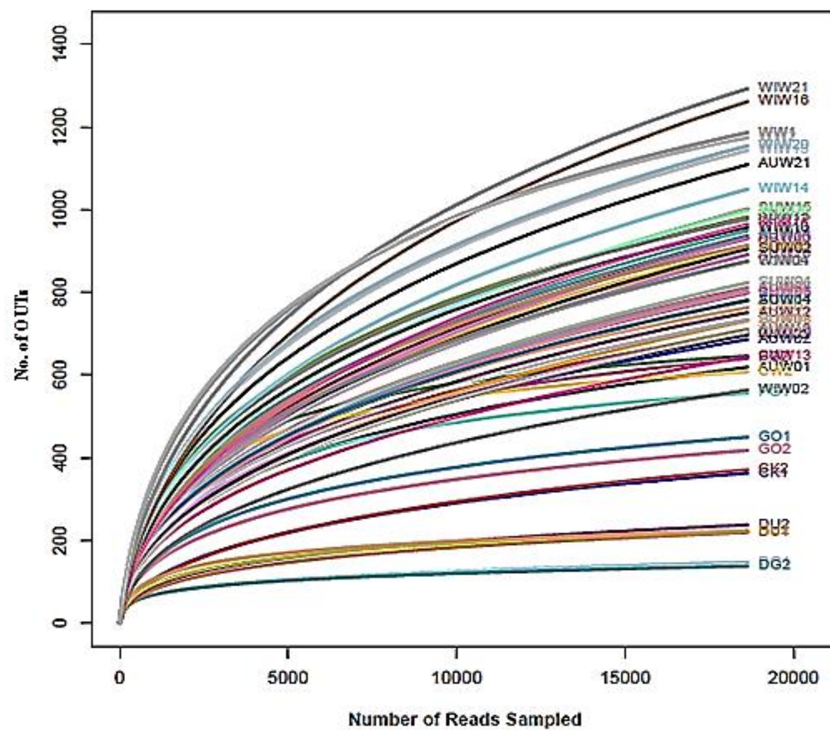


Figure 3.1 Rarefaction curves generated with normalized sequence reads (n=18640) for water, wastewater and fecal sample

Table 3.1 Bacterial diversity values based on the 16S rRNA gene sequences by NGS for water and fecal samples (normalized to n=18640 per sample).

Sample group	OTUs (±SD)	Shannon (±SD)	Simpson (±SD)	ACE index (±SD)	Chao1 (±SD)	Coverage (±SD)
Total water samples [a]	870±163	4.64±0.33	0.033±0.014	1510±235	1322±231	0.98±0.01
Chicken feces[b]	367±5.6	3.43±0.01	0.062±0.001	504±25	475±20	0.99±0.02
Cow feces[b]	627±26	5.13±0.01	0.015±0.006	685±31	687±29	0.99±0.02
Dog feces[b]	142±6	3.32±0.03	0.060±0.006	193±35	187±3	0.99±0.02
Duck feces[b]	229±58	3.15±0.13	0.09±0.010	328±59	310±19	0.99±0.02
Goose feces[b]	434±22	4.13±0.05	0.033±0.009	544±23	563±24	0.99±0.01
Human feces[b]	223±1	3.25±0.39	0.115±0.037	268±7	289±7	0.99±0.02
Pig feces[b]	598±59	4.39±0.01	0.036±0.001	685±80	698±101	0.99±0.01
Wastewater[b]	1180±9	5.59±0.04	0.010±0.001	1444±21	1461±40	0.98±0.01

[a] where n=45 samples, [b] where n=2 composite samples.

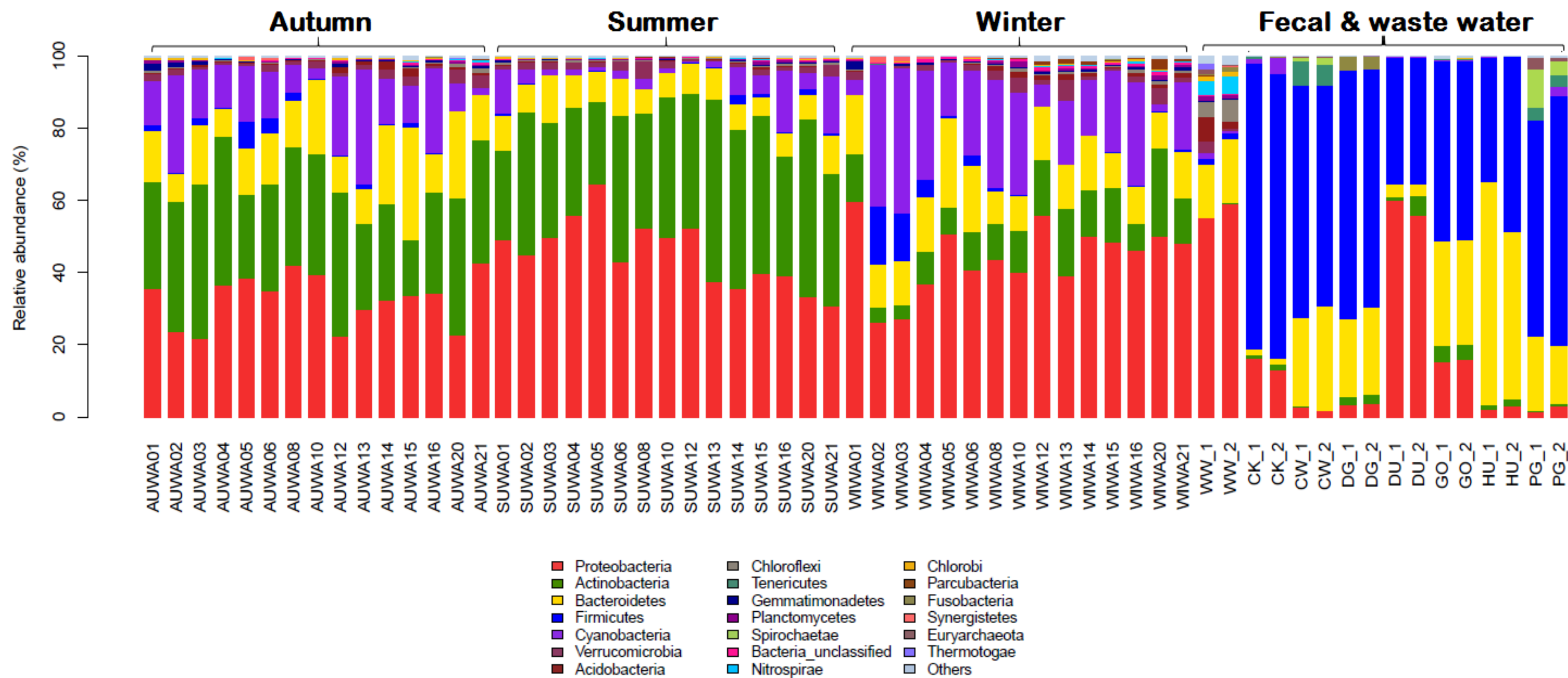


Figure 3.2 Relative abundance of bacterial phyla in Tiaoxi River water, fecal and wastewater samples.

3.3.4 Comparison of bacterial community structure between water and fecal samples

Hierarchical cluster analysis was performed by the UPGMA method to find out the similarities between different water and fecal samples (β-diversity). When three season water samples were compared, water samples from location 2 (WIW2) and 3 (WIW3) during winter clustered separately from other water samples indicating that the bacterial composition was different in these samples (Fig. 3.3A). As stated previously, the phylum and genus level bacterial compositions at these locations were different from those of other water samples (Fig.3.3 and Table S3.3). The β-diversity analyzed by PCoA also showed similar results and revealed that these two samples clustered separately from other water samples (Fig. 3.4A). The results of β-diversity analyzed by PCoA and cluster analysis also indicated that significant seasonal variation was observed for water samples that formed distinct seasonal clusters, though few autumn and summer water samples were closely related (Fig. 3.3A and 3.4A). 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.3.3B-D & Fig.3.4B-D). However, WIW2 and WIW3 water samples clustered close to WW and fecal samples indicating the presence of fecal related microbiota in these samples.

The Venn diagram analysis was performed with unambiguous OTUs for better analysis of shared and specific OTUs present in water, wastewater and fecal samples (Fig. 3.5A, B, C and 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. 3.5A). The analysis was also performed between three different season water samples with total fecal samples, the results indicated that 39.1%, 41.6%, and 38.7% of

OTUs from autumn, winter, and summer season water samples were shared with the OTUs of fecal samples (Fig. 3.5B). These findings point out at the possible fecal contamination in the

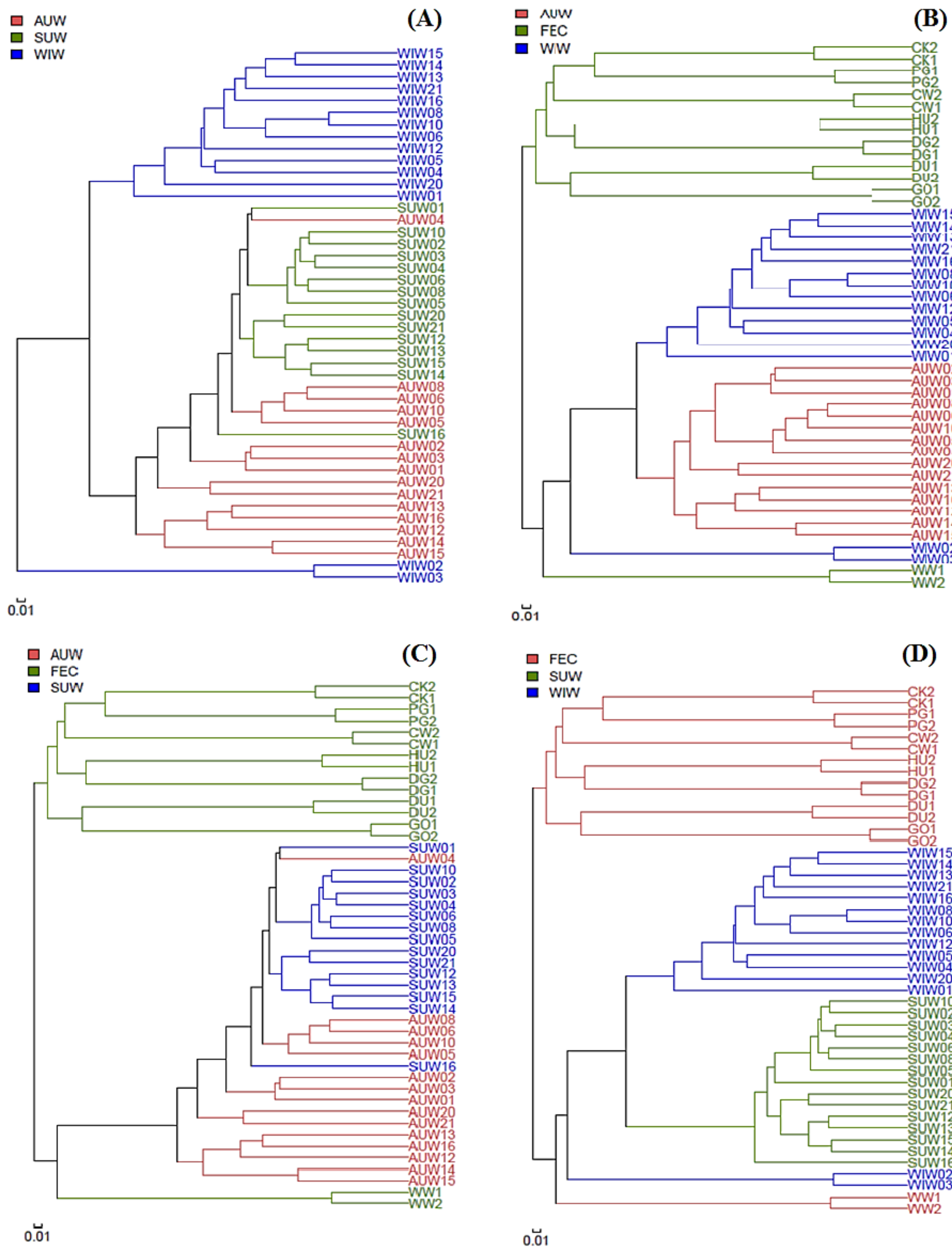


Figure 3.3 Cluster analysis showing similarity in bacterial diversity in water samples collected in three seasons (A) and in combinations of diversity in water and fecal samples (B, C, D). (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; FEC-Fecal samples).

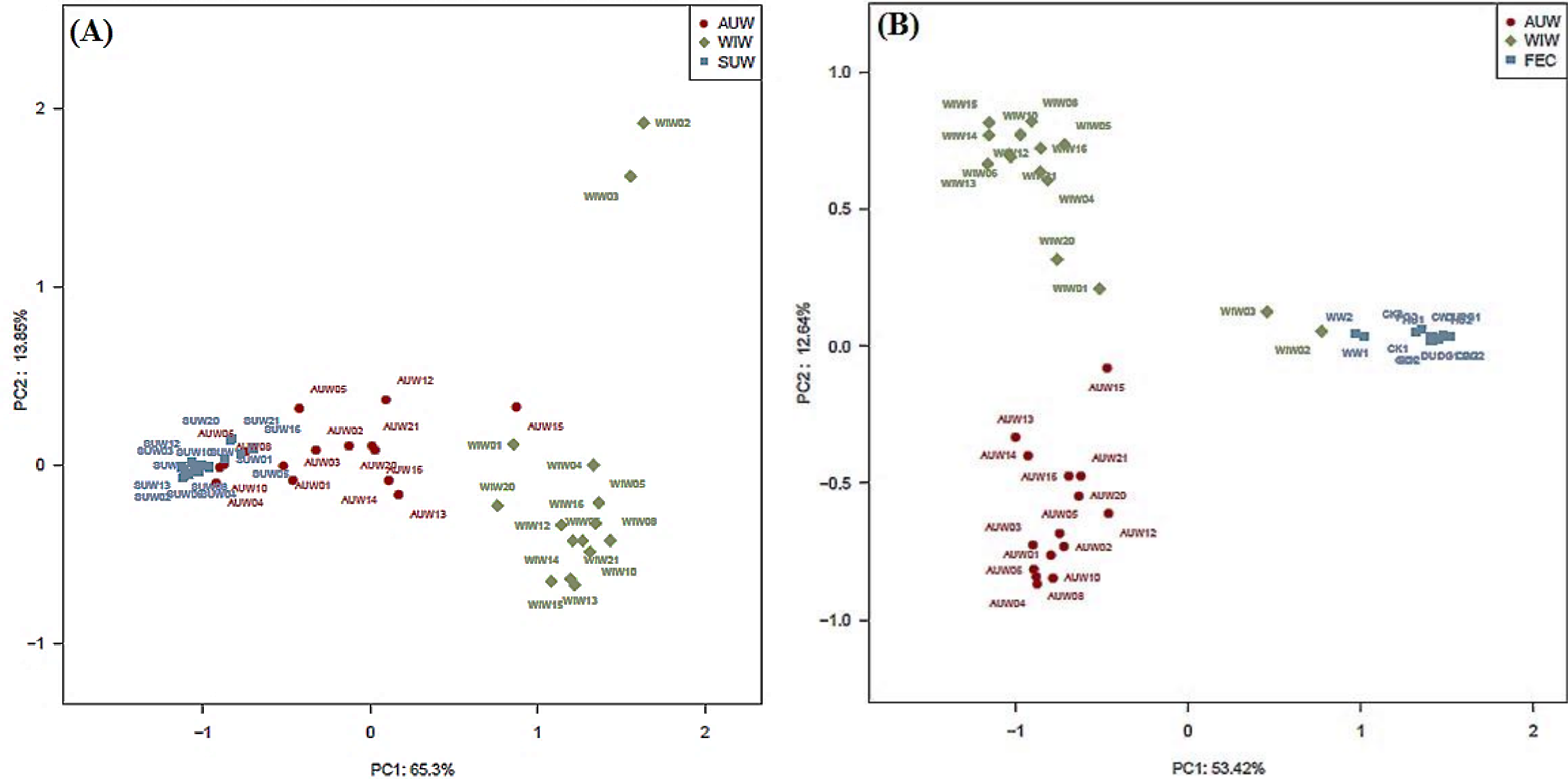


Figure 3.4 Visualization of β-diversity among water and fecal samples by Principal coordinate analysis. (A) Comparison of bacterial diversity between water samples collected during autumn, winter and summer (B) Comparison of bacterial diversity in autumn and winter water samples with fecal and wastewater samples. (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; FEC-Fecal samples).

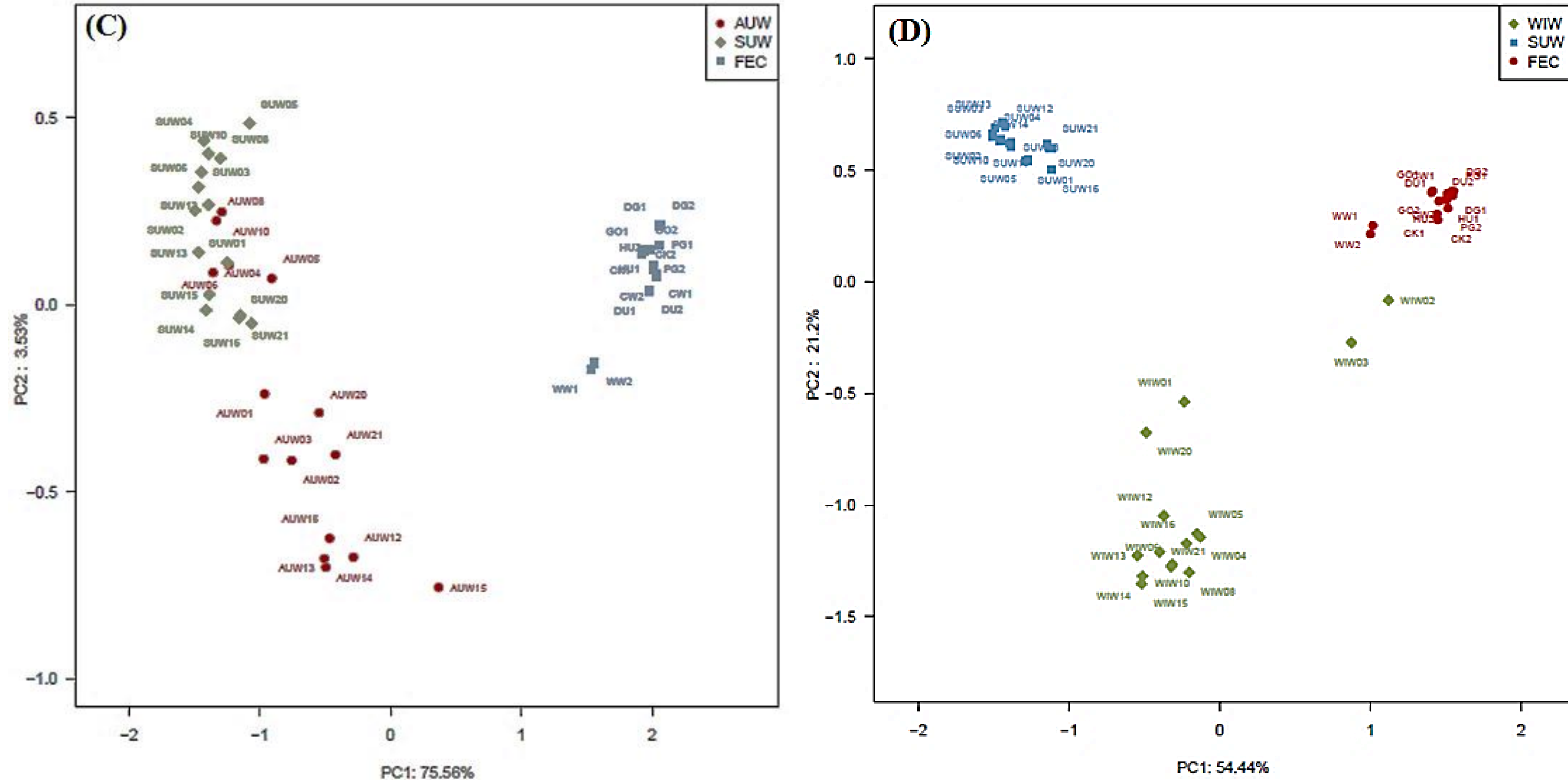


Figure 3.4 Visualization of β-diversity among water and fecal samples by Principal coordinate analysis. (C) Comparison between autumn and summer water samples with all fecal samples (D) Comparison between winter and summer water samples with all fecal samples. (AUW-Autumn Water; WIW=Winter Water; SUW-Summer Water; FEC-Fecal samples).

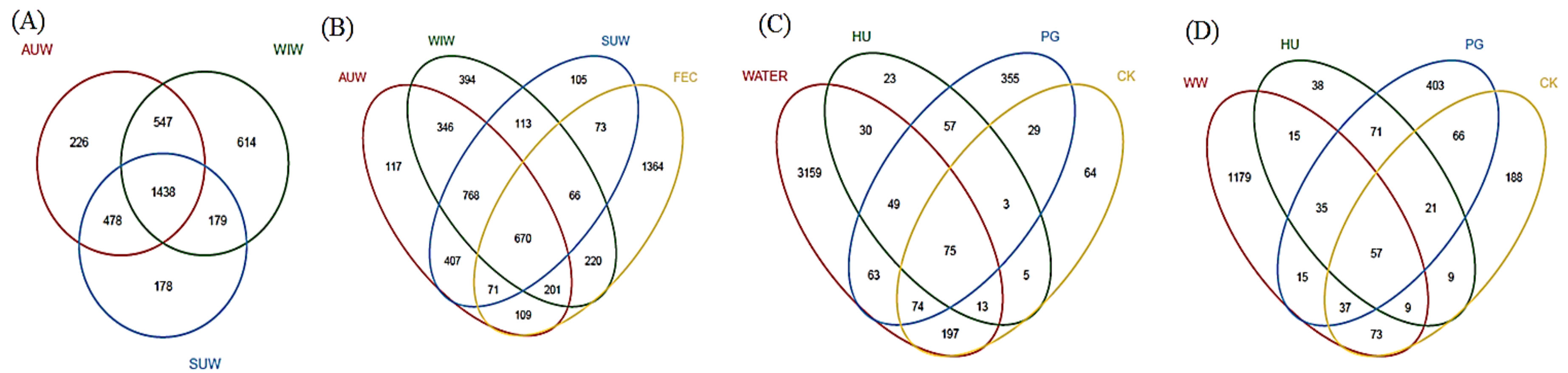


Figure 3.5 Venn diagrams showing shared and unique OTUs identified when the sequences were compared between Autumn, Winter and Summer water samples (A), between water samples collected during three season water samples with all fecal samples (B) and between pooled water samples and individual fecal samples (human, pig, and chicken) (C) and between fecal and wastewater samples (D).

water samples. 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.3.5C) indicating fecal contamination of water samples from these sources. 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. 3.5D). 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, which could have led to high diversity in these samples (Table 3.1).

3.3.5 Abundance of fecal or sewage-associated genera in Tiaoxi River water samples

As discussed previously, 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 potential pathogens in nature. Five fecal and one sewage-associated genus that include *Bacteroides*, *Prevotella*, *Blautia*, *Faecalibacterium*, *Dorea* and *Macellibacteroides* were detected in Tiaoxi River water samples. *Bacteroides* was detected in most of the locations (39 out of 45 samples tested) of Tiaoxi River (Fig. 3.6A). Genus *Bacteroides*, a member of Order *Bacteroidales*, represents Gram-negative obligately anaerobic bacteria that are abundantly present in the human and animal gut. *Bacteroides* contributed to 10-30 % of total fecal bacteria in mammalian feces (Layton et al., 2006) and their presence in environmental samples indicates potential mammalian fecal pollution. *Bacteroides* concentrations were comparatively high in the winter season and highest abundance was observed at WIW2 (2.6%) and WIW3 samples (2.4%) (Fig. 3.6A). Genus *Prevotella* was also observed at several locations (35 out of 45 samples tested) of Tiaoxi River with the highest concentration at location 16 during autumn season (Fig. 3.6B). *Prevotella*, a member of *Bacteroidales*, are Gram-negative obligately anaerobic bacteria that are present abundantly in human and animal gut as well as other body

parts (Lee et al., 2011). Genus *Blautia* was detected in several samples (26 out of 45 samples) of Tiaoxi River water (Fig. 3.6C). *Blautia*, which belongs to Order *Clostridiales*, are Gram-positive obligately anaerobic bacteria present in high abundance in the mammalian gut microbiome (Garcia-Mazcorro et al., 2012). The highest *Blautia* abundance was detected in WIW2 (0.10%) and WIW3 (0.11%) samples of Tiaoxi River water. Genus *Faecalibacterium,* which are obligately anaerobic bacteria and consist of only one known species, was detected in 19 out of 45 water samples tested of Tiaoxi River (Fig.3.6D). *Faecalibacterium* are abundantly present in the healthy human intestinal microbiota representing nearly 5% of the total bacterial population (Miquel et al., 2013). They were also reported to be present in avian fecal samples (Sun et al., 2016, Green et al., 2012). Therefore, their presence in water indicates possible human or avian fecal pollution. Water samples collected from location 2 (WIW2) and 3 (WIW3) during winter season showed highest *Faecalibacterium* abundance (0.038 and 0.028%). Genus *Dorea* was detected in 14 out of 45 water samples of Tiaoxi River and highest abundance was detected at WIW2 (0.06%) and WIW3 (0.04%) water samples. (Fig.3.7A). *Dorea* spp. are Gram-positive obligately anaerobic bacteria present mainly in human faeces (Taras et al., 2002) and presence of this genus indicates possible human fecal contamination. Sewage associated genus *Macellibacteroides* was also detected in some locations (27 out of 45 samples) of Tiaoxi River (Fig. 3.7B). *Macellibacteroides* belongs to *Bacteroidales* and are strict anaerobic Gram-negative bacteria mostly isolated from WWTPs (Jabari et al., 2012); though these bacteria were detected in goose fecal samples here (Table S3.2). These bacteria were frequently detected at locations 4, 5, 6 and 16 in all seasons. The results match with the land use pattern as the upstream of Location 6 is active in poultry farming and location 4, 5 and 16 were reported to have WWTPs (Zheng et al., 2017). A few more human-associated fecal bacteria such as *Parabacteroides, Bifidobacterium* were also detected in low concentrations at

different locations (Table S3.3 and S3.4) indicating potential human contaminations in these locations.

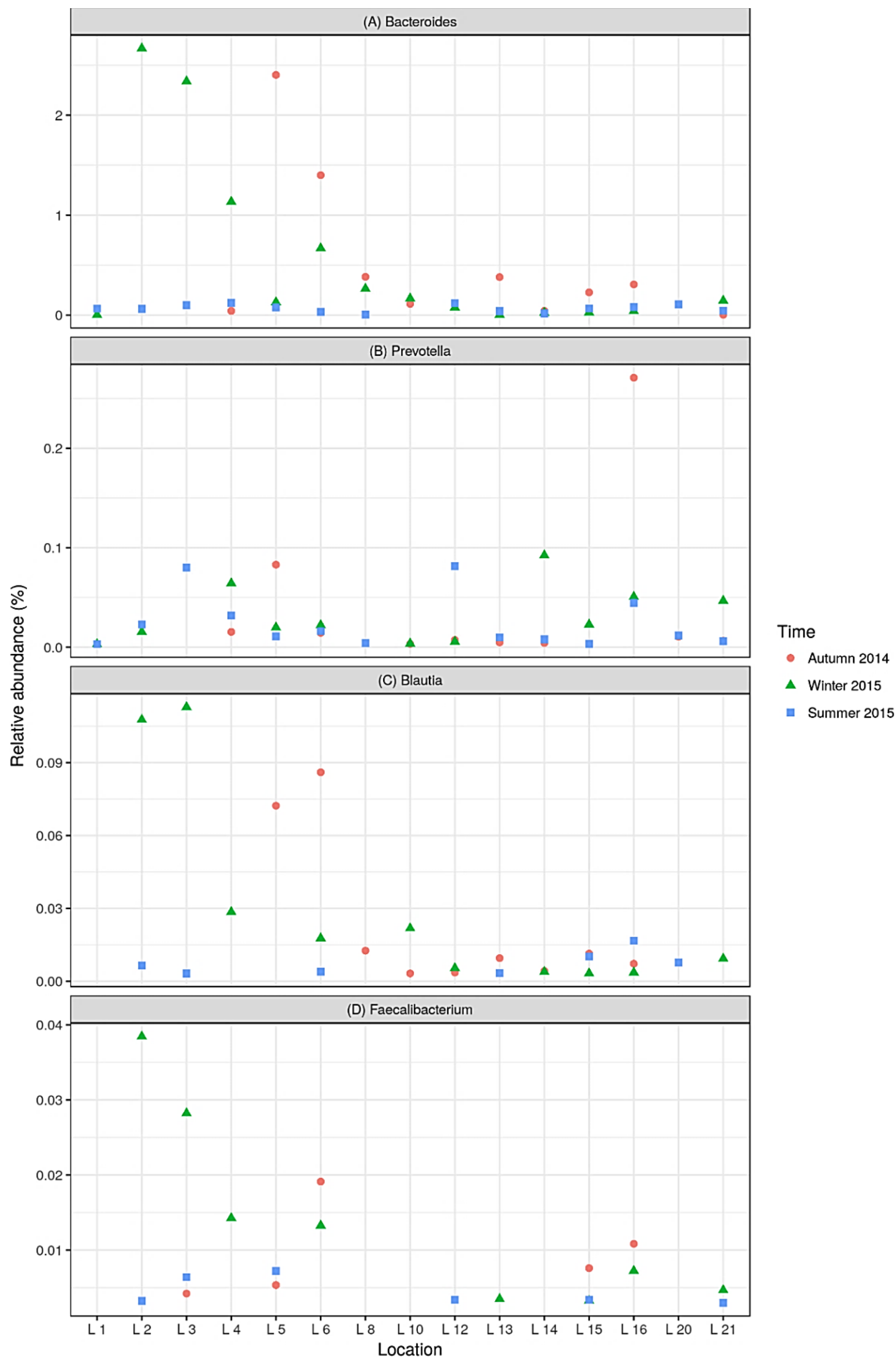


Figure 3.6 Relative abundance of fecal-associated genera (A) *Bacteroides*, (B) *Prevotella,* (C) *Blautia* ((D) *Faecalibacterium* observed at different sampling locations in Tiaoxi River.

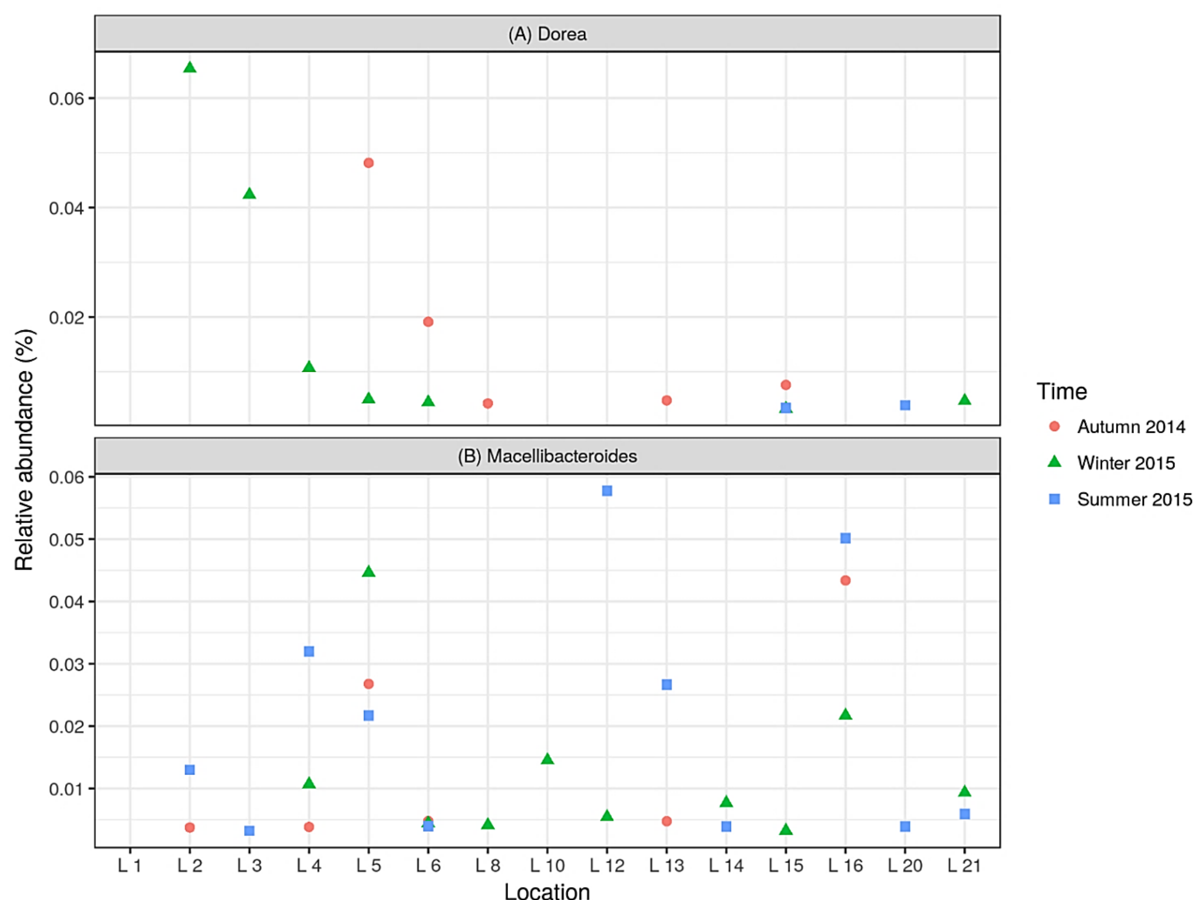


Figure 3.7 Relative abundance of fecal and sewage associated genera locations (A) *Dorea,* (B) *Macellibacteroides* observed at different locations in Tiaoxi River.

3.3.6 Abundance of potentially pathogenic genera in Tiaoxi River water samples

The sequencing data were 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 potential pathogenic bacterial genera were detected in Tiaoxi River water samples (Table S3.4). However, only seven groups (*Acinetobacter, Aeromonas, Arcobacter, Brevundimonas, Enterococcus, Escherichia-Shigella,* and *Streptococcus*) showed relative abundance greater than 0.1 at different locations of Tiaoxi River water and these data are presented in Table 3.2.

Table 3.2 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

Several species of *Acinetobacter* and *Aeromonas* are considered as opportunistic pathogens, which cause nosocomial infections and gastroenteritis (Khosravi et al., 2015, Laukova et al., 2018). The highest *Acinetobacter* concentration was observed at location 2 and *Aeromonas* at location 5, though these genera were not classified up to species level. As stated earlier, these locations have higher levels of human and animal fecal contamination. The genus *Arcobacter* detected in this study showed 97% identity to *A. cryaerophilus,* which is an emerging enteropathogen to human and animals (Fernandez et al., 2015). *Arcobacter* was one of the most frequently (9 out of 15 locations with above 0.1% abundance) detected pathogens in Tiaoxi River and the highest abundance (1.3%) was observed at location 12. *Brevundimonas* was the most commonly detected potential pathogen in Tiaoxi River, which was found with higher abundance (>0.10) at 10 monitoring locations and the highest abundance (3.9%) was detected at location-1 (Taihu Lake) during winter season. The sequences of this genus were identified as *Brevundimonas alba, B. bullata* and *B. vesicularis*, in which the latter is an opportunistic pathogen that causes bacteremia (Zhang et al., 2012). Although *Enterococcus and Escherichia-Shigella* were detected with high 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 pathogens with relative abundance >0.1 were detected at location 3 and six of these potential pathogens with higher abundance was observed at location 2 and 4 emphasizing that these locations could be of potential human health risk.

3.3.7 Relationship between environmental parameters and bacterial community composition

RDA analysis was carried out to find out the relationship between environment parameters (pH, temperature, conductivity, TN, TP, NO_2-N, NH_4-N) on the bacterial community composition observed at different locations and seasons. RDA1 and RDA2 explained 26.92% and 14.72% of the total variations observed in autumn 2014, respectively (Fig. 3.8A). The RDA biplot

indicated that the parameters temperature, conductivity, pH, TP and TN variations altered the bacterial community composition. In particular, the pH, conductivity and TP influenced the community composition in the locations 3 and 12. In winter 2015, the pattern was different. The RDA1 and RDA2 contributed 44.7% and 16.43% to the total variations, respectively (Fig. 3.8B). The pH and conductivity were found to be major factors influencing the bacterial community composition. The RDA biplot pattern was entirely different in the samples collected during summer 2015 (Fig. 3.8C). The RDA1 and RDA2 contributed 41.9% and 10.2%, respectively to the total variations, respectively. The parameters such as temperature, conductivity, pH, NH_4-N and NO_2-N were found to be influenced the community composition. The overall analysis indicated that pH, conductivity, temperature and some of the nutrients were the main environmental factors that showed strong influence on the bacterial community composition.

3.4 Discussion

Recent advances in NGS technologies coupled with reduced cost has increased the application of microbial community analysis in different fields of research such as monitoring microbial quality and diversity in aquatic environment and studying the microflora associated with fecal samples (Ley et al., 2008, Thomas et al., 2012). Vierheilig et al. (2015) used NGS based 16S rRNA gene sequencing for water quality monitoring by studying the bacterial community composition in water, feces, soil and sediment samples, and the results revealed that this approach can be used for monitoring bacterial contamination and re-growth in the environmental waters. Newton et al. (2011) assessed human fecal pollution in urban harbor by 16S rRNA targeted pyrosequencing. Marti et al. (2017) assessed the microbial contamination in benthic and hyporheic sediments in a peri-urban River by MST and 16S rRNA sequencing and the study concluded that the NGS enabled detection of bacterial contamination in water and that the NGS based method was sensitive to track community changes as compared to MST

Figure 3.8 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.

markers. In this study, bacterial communities of water, fecal and wastewater samples were studied by Illumina sequencing by targeting V3-V4 hypervariable region of the 16S rRNA gene. 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 community associated with chicken, dog, duck and human fecal samples were less compared to other fecal, wastewater and Tiaoxi River samples (Table 3.1) and this could be due to dominance/abundance of relatively small number of taxa (Jeong et al., 2011). In chicken and duck fecal samples, *Proteobacteria* and *Firmicutes* were dominant which 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 and indicated that diet influences the gut flora leading to less diversity or the presence of a small number of dominant taxa. They also reported that carnivores and omnivores (human, dog, duck and chicken) have less diverse gut microbiomes compared to herbivores. High diversity values were observed in wastewater samples here are consistent with previous studies (Newton et al., 2013, Shanks et al., 2013). Higher diversity in wastewater samples could be due to presence of mixed bacterial communities such as bacteria from human feces (collected from large number of people) releasing into wastewater and from environmental samples releasing into WWTP streams (Newton et al., 2013). Moreover, the wastewater contains high amount of nutrients, which support the growth of diverse microbial populations (McLellan et al., 2013, Arfken et al., 2015). The bacterial diversity observed in this study can be different from other studies and several factors could contribute to the difference in intestinal microflora caused by dietary habits of livestock and humans (Ley et al., 2008, Dowd et al., 2008a, Dowd et al., 2008b). Other factors include the nature of gene marker and hypervariable region selected for sequencing, number of reads analyzed, variance in similarity cutoff values for determining OTUs, and difference in DNA isolation procedure or

kits employed (Gihring et al., 2012). Although the hierarchal cluster analysis results indicated that human fecal samples are closely related to dog fecal samples (Fig. 3.3B, C and D), the Venn diagram of human fecal samples with pig and chicken fecal samples (as they are common livestock's of study area) revealed that human fecal samples harbor microbes that are closely related to pig fecal samples (Fig. 3.5D). This could be due to more shared OTUs from genera *Bacteroides* and *Prevotella* present in these samples (Dick et al., 2005).

Many species of genera *Bacteroides, Prevotella* and *Blautia* are abundant in the mammalian gut and were advocated as fecal indicators (Savichtcheva and Okabe, 2006, Newton et al., 2011). 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 human feces associated *Blautia* sequences in the samples. Species of *Bacteroides* have host specificity and limited survival in environment, making them ideal indicators of recent fecal contamination and are often used for microbial source tracking (MST) (Bernhard and Field, 2000b, 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* were mainly isolated from human feces (Wei et al., 2001, Kitahara et al., 2005) and presence of these bacterial species in the environment indicates human fecal contamination. These species were frequently detected with high concentration (OTUs) at location 2 and 3 of Tiaoxi River (Table S3.3) and highest concentrations were observed in the samples collected during winter (WIW2 and WIW3 samples). During winter sampling (Jan/Feb 2015), the rainfall was relatively less than autumn winter sampling and considered as a dry period (only 66mm precipitation, NBSC (2016) and fecal contamination at these locations could be due to direct discharge of sewer and septic waste but not merely through runoff (Ohad et al., 2015). The samples collected at location 3 were close to a fishing village where people live on the

boats without proper sanitation facilities. The high *Bacteroides* concentration could be associated with entry of sewage at this location and based on land use pattern, the presence of higher *Bacteroides* concentration at location 2 could be associated with the transport of these bacteria from location 3 (Marti et al., 2013). The presence of *Parabacteroides merdae,* which was mainly identified from human feces, at locations 2 and 3 also confirms the potential human fecal contamination at these locations (McLuskey et al., 2016). In the present study, *B. propionicifaciens* was detected only in goose and duck fecal samples however; it was detected frequently at location 6 of the Tiaoxi River water samples indicating avian fecal contamination at this location. The presence of *B. massiliensis*, *B. nordii* and *B. stercoris* with high concentrations at location 2 and 3 point to potential human risks, as these species are associated with anaerobic bacteremia and abdominal infections (Song et al., 2004, Fenner et al., 2005, Otte et al., 2017). *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). The results obtained in this study are in agreement with previous studies, which reported that the above *Bacteroides* species are isolated from effluents of WWTPs (Nishiyama et al., 2009, Ueki et al., 2011). However, they were also detected in goose fecal samples in the current study. Overall, the species level identification of *Bacteroides* are matching with the land use pattern in the current study (Table S3.3). The remaining fecal indicator bacterial genera such as *Prevotella, Blautia, Faecalibacterium* and *Dorea* detected in this study were not reliably classified to species level indicating the limitations of using short reads of 16S rRNA generated by NGS (Nguyen et al., 2016). However, *Prevotella* was detected at locations 4, 5, 6, 12, 14, 16 and 21 in three seasons (Fig. 3.5B) indicating 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 sewage entry or transport of bacteria from upstream locations such as location 13 where poultry and pigs farms are observed (Table 2.1; S5.2). While studying the fecal contaminations in an urban River, Newton et al. (2011) noticed an increased in fecal indicators and human-associated sequences after the heavy rainfall and combined sewer overflows (CSOs). The fecal indicator bacteria *Blautia, Faecalibacterium* and *Dorea* are significantly higher at locations 2 and 3 during winter season, highlighting the presence of fecal contamination at these locations again. Overall, the presence of fecal indicator bacteria at different locations of Tiaoxi River water stresses the application of MST techniques to determine the source of fecal pollution at Tiaoxi River.

As fecal indicator bacteria were observed at several locations, further analysis was performed to detect potential pathogenic bacteria, which revealed that seven potential pathogenic bacterial genera (*Acinetobacter, Aeromonas, Arcobacter, Brevundimonas, Enterococcus, Escherichia-Shigella,* and *Streptococcus*) were present in Tiaoxi River water (Table 3.2). Most of these potential pathogenic bacterial genera were detected with higher concentration (>0.1 abundance) at locations 2, 3 and 4, which were highly contaminated with feces. However, most of these potential pathogenic bacteria are opportunistic pathogens as an example, *Acinetobacter,* an emerging opportunistic pathogen comprises more than 20 species and only three species are considered as pathogens (Shamsizadeh et al., 2017). Similarly, *Aeromonas* has 21 species but only three species are mainly considered as pathogens (Janda and Abbott, 2010). Therefore, the sequencing data was carefully examined again to check the pathogenic species of these genera. Only two species level classified pathogenic bacterial genera were identified: *A. cryaerophilus* and *B. vesicularis*. *A. cryaerophilus* that causes acute to chronic diarrhoea in human was frequently detected at locations 12 and 16, demonstrating potential health risk associated with water samples at these locations. However, *Arcobacter* genus, which was included in *Campylobacter* genus earlier, was created or reclassified recently and often

confused with *Campylobacter* species due to their phylogenetic similarity (Collado and Figueras, 2011). 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 (Table S3.3) and the remaining potential pathogenic genera such as *Enterococcus, Escherichia-Shigella*, and *Streptococcus* were not classified to species level with 97% similarity and this could be due to limitation of short sequence reads or the low resolution of hyperviable region V4 in classifying pathogens using NGS method (Janda and Abbott, 2007, Nguyen et al., 2016). However, NGS could be useful as a screening method to assess a wide-range of fecal and potential pathogenic bacteria, which could be further verified by sensitive methods such as qPCR to determine the potential human health risk.

3.5 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 indicating the bacterial compositions were different in these samples.
- 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.
- Five bacterial genera associated with fecal sources (*Bacteroides, Prevotella, Blautia, Faecalibacterium*, and *Dorea*) were present at several locations indicating possible human and animal fecal contamination in these locations.

- Seven potential pathogenic bacteria (*Acinetobacter, Aeromonas, Arcobacter, Brevundimonas, Enterococcus, Escherichia-Shigella,* and *Streptococcus*) were observed in several locations at abundances >0.1 and more specific PCR assays are needed for accurate detection and quantification of these genera.
- Overall, the results indicate that NGS could be a valuable tool to screen for a wide variety of bacteria including fecal and pathogenic in nature as an initial step to identify human health risk and to prioritize sites for further assessment using more specific methods such as quantitative PCR.

CHAPTER-4

Evaluation of Human and Animal Associated Microbial Source Tracking Quantitative PCR Markers for Monitoring Fecal Pollution in Taihu Watershed, China

Abstract

The aim of this chapter was to evaluate the performance of existing universal, human and animal associated microbial source tracking (MST) qPCR assays for their applicability to ascertain host-associated fecal pollution in Taihu watershed, China. Ten MST qPCR assays were evaluated using DNA extracts from sewage and fecal DNA extracts from human, chicken, cow, duck, goose, dog and pig. The BacUni and GenBac3 (universal *Bacteroidales* markers) amplified all DNA samples from fecal and sewage sources but higher abundance was obtained with BacUni. The four human-associated *Bacteroidales* assays (HF183 Taqman, BacHum, HF183 SYBR, and Hum2) exhibited a sensitivity of 53-80% and had cross-reactivity with chicken (40-70%) and dog (10-20%) fecal DNA samples. However, the HF183 Taqman assay did not show any cross-reactivity with either pig or cow fecal DNA and it quantified the target in all DNA samples from sewage. Pig-2-Bac assay showed high sensitivity (90%) with pig fecal DNA, low cross-reactivity (20%) with cow fecal DNA, and no amplification of human fecal DNA. BacCow, which was tested as a livestock/domestic animals associated assay, amplified only cattle (100%), pig (20%) and chicken (40%) fecal DNA but not all livestock/domestic animal fecal DNA samples. Among the avian associated markers tested, GFD showed less sensitivity (70%) than AV4143 (90%) but it did not cross-react with human fecal DNA, suggesting that this assay could be used to differentiate between chicken and human fecal contaminations. This validation study demonstrates that BacUni, HF183 Taqman, Pig-2-2Bac and GFD assays are the most suitable for differentially identifying and monitoring human and animal fecal contamination in Taihu watershed.

4.1 Introduction

Fecal pollution of watercourses leads to eutrophication and is a serious threat to human health by disseminating pathogenic microorganisms (Wade et al., 2003, Lapointe et al., 2015). Although pathogen detection methods and wastewater treatment technologies have improved significantly, waterborne disease outbreaks have been reported frequently in several countries (Brookes et al., 2004, Ahmed et al., 2015b, Unno et al., 2012). The World Health Organization (WHO) reported that diarrheal disease leads to the death of around 525,000 children (<5 years old) each year in the world and a significant portion of this diarrheal disease was due to the consumption of poorly sanitized water (WHO, 2017). Although human health risk is the main concern, pathogen contamination can also lead to substantial economic loss due to the closure of water bodies used for fisheries or recreational activities (Mitch et al., 2010). In China, rapid industrialization and urbanization combined with inadequate water treatment regimes has led to severe water pollution; one report stated that more than half of the populations in rural China consume water polluted with animal and human feces (Hagedorn and Liang, 2011). In China, as elsewhere, microbial quality of surface waters is evaluated using fecal indicator bacteria (FIB), as suggested in the Surface Water Criteria (GB3838-2002) by the Ministry of Ecology and Environmental Protection (MEP), People's Republic of China. FIB in water can originate directly from human or animal feces inputs, but also from sewage; simple detection and enumeration of FIB by traditional culture-based methods provides no evidence on the specific source of the contamination. Furthermore, previous studies indicated the occurrence of non-fecal FIB in environmental waters (Byappanahalli et al., 2003, Yamahara et al., 2007).

As human health risks vary with the type of host fecal source exposure, understanding the origin of fecal sources and their impact on water quality is very important for effective management of a watershed and to enable appropriate remedial action (Soller et al., 2010). In this regard, microbial source tracking (MST) techniques emerged as promising approaches to

distinguish human from other animal sources of fecal contamination. Initially, several library dependent MST methods were developed that rely on comparison of genetic or phenotypic traits between reference fecal sources (known) and environmental samples (unknown) to ascertain the origin of fecal pollution (Parveen et al., 1999, Harwood et al., 2000); however, these methods are time-consuming and not very cost effective. For this reason, library independent MST (LI-MST) methods targeting host-associated genetic markers were developed as alternative indicators to distinguish human fecal sources from other fecal sources (Scott et al., 2002, Kildare et al., 2007, Green et al., 2014, Shanks et al., 2009, Seurinck et al., 2005).

LI-MST methods focused on *Bacteroidales* 16S rRNA gene markers associated with a specific host have given promising results, although some studies reported cross-reactivity of these markers (Roslev and Bukh, 2011, Layton et al., 2013, Reischer et al., 2013, Boehm et al., 2016, Harris et al., 2016). *Bacteroidales* is an order of anaerobic bacteria universally found in the mammalian intestinal tract at high concentrations, but at lower abundances in avian species, such as chicken, goose, and gull (Wexler, 2007, Lu et al., 2008, Jeter et al., 2009). Several methods targeting *Bacteroidales* 16S rRNA gene makers have been developed worldwide to differentiate human fecal sources from other feces such as a ruminant, swine, dog, and birds (Kildare et al., 2007, Green et al., 2014, Mieszkin et al., 2010, Mieszkin et al., 2009, Gourmelon et al., 2010). However, it was reported recently that avian fecal markers could be better distinguished from other fecal sources by targeting bacterial taxonomic groups such as *Helicobacter* spp. rather than *Bacteroidales* 16S rRNA gene (Green et al., 2012).

Previous reports specified that geographical differences could significantly affect the performance of these MST methods and recommended proper assessment prior to field application at new study areas (Reischer et al., 2013, Odagiri et al., 2015, Boehm et al., 2016).

Most of these MST methods were developed in North America, Europe, Australia, or New Zealand and their performance and applicability in Asia, particularly in China, has been little studied (Jia et al., 2014, He et al., 2016a, Fan et al., 2017). In China, studies on qPCR based MST assays to monitor fecal pollution are very limited. He et al. (2016a) validated five MST and four mitochondrial DNA fecal source tracking (FST) markers for their applicability to study fecal pollution in Taihu Lake watershed. They reported that mitochondrial DNA based human FST methods were superior (but slight cross-reactivity was observed with pig fecal samples which are the major livestock at Taihu watershed) over those *Bacteroidales* based MST (BacH, HF183 SYBR) assays tested; the most widely used MST makers such as HF183 Taqman and BacHum qPCR assays were not included. They also reported that Pig-2-Bac assay (*Bacteroidales* based) showed higher accuracy than other mitochondrial DNA markers. Jia et al. (2014) used swine specific *Bacteroidales* assay (Pig-2-Bac) to assess the levels of swine fecal pollution in Hongqi River and Fan et al. (2017) developed two new assays based on a genome fragment enrichment (GFE) method targeting *Bacteroidales*-like sequence present in pig fecal samples (Fan et al., 2017).

This study report the results of a detailed and comprehensive study to determine the sensitivity and specificity of a range of existing general and host-specific MST qPCR markers, using fecal samples collected from the Taihu watershed region, China in order to identify the most suitable MST qPCR assays for detecting host associated fecal pollution across this large and important water catchment.

4.2 Materials and Method

4.2.1 Overview of study area and selection of MST qPCR assay for evaluation

The current evaluation study was carried out at the Taihu watershed region located in Jiangsu and Zhejiang province, People's Republic of China (PRC). The samples were primarily collected from rural and urban areas of Tiaoxi River region. Tiaoxi River is the main inlet River

connected to Taihu Lake and it serves as water source for drinking, industrial and agricultural activities for many cities located in Zhejiang Province (Tang et al., 2012). The preliminary selection of MST assays for this study was based on the identity of the livestock population in Zhejiang province and as per the data released by the National Bureau of Statistics, PRC in 2015 (NBSC, 2015). Pigs predominate in this province accounting for more than 90% of the livestock population in this region excluding poultry (NBSC, 2015). Cattle, goats, sheep, and buffalos are rare and together constitute less than 10% of the total livestock population. China is the major consumer and producer of poultry in the world (Sun et al., 2016) and poultry production is the common livestock in Zhejiang province, which often leads to issues concerning the disposal of poultry waste (Zheijiang, 2016). MST assays developed and designed elsewhere were selected to validate their applicability; the details of the MST assays are provided in Table 4.1.

4.2.2 Collection and processing of fecal samples

To determine the specificity and sensitivity of host-associated MST qPCR markers, individual and composite fecal and wastewater samples were used as target sources (Ahmed et al., 2016b). In total, 61 fresh individual and composite fecal samples from various hosts were collected in Huzhou (Zhejiang province) and Suzhou (Jiangsu province) areas and tested. Further details of fecal and wastewater sample collection along with composite fecal samples preparation is provided in Supplementary note S4.1. Fresh fecal samples collected in sterile containers were provided by healthy human volunteers (n=10) aged between 16 and 40 years. Safety guidelines were provided and consent for use of the samples in this study was obtained. Ethical approval was acquired from XJTLU Research Ethics Committee for handling fecal and sewage samples in this study. Individual fecal samples from animal hosts representing pig, chicken, dog, and cow (n=10 for each) and composite fecal source from duck and goose (n=3 pooled samples) were collected from farms, pet stores and backyards of households located near the Taihu Lake

Table 4.1 List of primers and probes used for validation of MST qPCR assays in samples collected from the Taihu watershed region.

Assay	Primer/probe	Concentration	Oligonucleotide sequence (5'–3')	Annealing temperature (oC)	Reference
BacUni	BacUni- 520F	400nM	CGTTATCCGGATTTATTGGGTTTA	60	Kildare et al. 2007
(Taqman)	BacUni-690R1	400nM	CAATCGGAGTTCTTCGTGATATCTA		
	BacUni-656P	80nM	FAM-TGGTGTAGCGGTGAAA-MGB		
GenBac3	GenBacF3	1000nM	GGGGTTCTGAGAGGAAGGT	60	Siefring et al. 2008
(Taqman)	GenBac4R	1000nM	CCGTCATCCTTCACGCTACT		
	GenBact2P	80nM	FAM-CAATATTCCTCACTGCTGCCTCCCGTA-TAMRA		
HF183	HF183F	1000nM	ATCATGAGTTCACATGTCCG	60	Green et al. 2014
(TaqMan)	BacR287R	1000nM	CTTCCTCTCAGAACCCCTATCC		
	BacP234P	80nM	FAM-CTAATGGAACGCATCCC-MGB		
BacHum	BacHum-160F	400nM	TGAGTTCACATGTCCGCATGA	60	Kildare et al. 2007
(Taqman)	BacHum-241R	400nM	CGTTACCCCGCCTACTATCTAATG		
	BacHum-193P	80nM	6-FAM-TCCGGTAGACGATGGGGATGCGTT-TAMRA		
HF183	HF183F	100nM	ATCATGAGTTCACATGTCCG	53	Seurinck et al. 2005
(SYBR Green)	Bac242R	100nM	TACCCCGCCTACTATCTAATG		
HumM2	Hum2F	400nM	CGTCAGGTTTGTTTCGGTATTG	60	Shanks et al. 2009
(Taqman)	Hum2R	400nM	TCATCACGTAACTTATTTATATGCATTAGC		
	HumM2P	80nM	FAM-TATCGAAAATCTCACGGATTAACTCTTGTGTACGC-TAMRA		
Pig-2-Bac	Pig-2-Bac41F	300nM	GCATGAATTTAGCTTGCTAAATTTGAT	60	Mieszkin et al. 2009
(Taqman)	Pig-2-Bac163R	300nM	ACCTCATACGGTATTAATCCGC		
	Pig-2-Bac113P	200nM	VIC-TCCACGGGATAGCC-MGB		
BacCow	BacCow-128F	400nM	CCAACYTTCCCGWTACTC	60	Kildare et al. 2007
(Taqman)	BacCow-305R	400nM	GGACCGTGTCTCAGTTCCAGTG		
	BacCow-257P	80nM	6-FAM-TAGGGGTTCTGAGAGGAAGGTCCCCC- TAMRA		
AV4143	Av4143F	500nM	TGCAAGTCGAACGAGGATTTCT	60	Ohad et al. 2016
(Taqman)	Av4143R	500nM	TCACCTTGGTAGGCCGTTACC		
	Av4143P	250nM	FAM-AGGTGGTTTTGCTATCGCTTT-BHQplus		
GFD	GFD-F	100nM	TCGGCTGAGCACTCTAGGG	57	Green et al. 2012
(SYBR Green)	GFD-R	100nM	GCGTCTCTTTGTACATCCCA		

watershed region in the Huzhou area. Fresh fecal dropping from animals were collected in sterile containers. All fecal samples were brought to the laboratory on ice and were stored at -20 ºC within 6 hours of collection. Primary effluents (500 mL; n=6) were collected from a wastewater treatment plant (WWTP) situated in Suzhou and brought to the laboratory on ice. Biomass from primary influents was collected by centrifugation at 4000×g for 10 min at 4ºC and DNA was extracted immediately.

4.2.3 Extraction of Genomic DNA

DNA extraction from the fecal/sewage samples was performed using the PowerFecal® DNA isolation kit that uses Inhibitor Removal Technology® (IRT) (MoBio, Carlsbad, CA USA), following manufacturer's instructions. About 250 mg of individual fecal/raw sewage samples were used for DNA extraction from all the host fecal sources except avian fecal samples for which only 0.10g was used. To avoid any cross contamination with other hosts, DNA extraction was conducted from each type of fecal source separately. Blanks with no fecal samples were also performed simultaneously in each batch. The quality and quantity of extracted DNA were confirmed by spectrophotometry (NanoDrop ND 2000C, Thermo Fisher Scientific., Wilmington, USA) and extracts were stored at -20ºC until further analysis.

4.2.4 Preparation of DNA standards for qPCR assays

Plasmid DNA standards were used for all qPCR assays and prepared by amplifying the target genes for each assay with the respective primer set using the fecal DNA extracts. The amplified products were purified using a PCR purification kit (Axygen Biosciences, CA, USA), ligated to pMD 19 vector (Takara, Bio Inc., Shiga, Japan) (Supplementary Figure S4.1) and transformed into competent *E. coli* cells (Tiangen Biotech, China). Plasmid DNA was extracted from the positive clones using QIAprep Spin Miniprep Kit (Qiagen, Mississauga, ON), and the extracted plasmids were sequenced using respective primers (Sangon Biotech, China). Plasmid

DNA standards were quantified using NanoDrop ND 2000C spectrophotometer and these concentrations were used for calculation of gene copy numbers. Standard curves for each assay were developed using tenfold serial diluted plasmid standards containing the respective target gene (10^8-10^1 copies/µl for each reaction).

4.2.5 Quality assurance for DNA samples and qPCR assay conditions

The DNA extracts were evaluated for the absence of PCR inhibitors and for the presence of amplifiable fecal DNA using the Bac-Uni qPCR assay (Odagiri et al., 2015) which detects universal *Bacteroidales* 16S rRNA genes. Two dilutions of DNA extracts (1:10 and 1:100) were assayed and the DNA extract was judged as free from PCR inhibitors if the two sample dilutions gave matching concentrations of Bac-Uni amplification products (Odagiri et al., 2015, Reischer et al., 2013).

All the qPCR reactions were performed in triplicates with 20µl final reaction volume. The reaction mixture for all Taqman chemistry based qPCR assays includes 10 µl of TaqMan Environmental Master Mix 2.0 (Applied Biosystems, California, USA), 2 µl of the probe/primer set with a final concentration as given in Table.1 and 8µl of 10-fold diluted target DNA template. For the two SYBR Green chemistry based qPCR assays (HF183 SYBR and GFD), the reaction mixture contained 10µl of SYBR Green Master Mix 2.0 (Thermo Scientific, USA), 2 µl of primer mixture (Table 4.1) and 8µl of 10-fold diluted target DNA templates as stated earlier. The correct amplification products for these SYBR Green assays were chosen based on the melting curve analysis as described by the publisher (Green et al., 2012, Seurinck et al., 2005). The annealing temperature of all assays is provided in Table 4.1 and all the assays were performed using the Applied Biosystems amplification system (ABI 7500fast, CA) with StepOnePlus software.

4.2.6 qPCR performance characteristics and data analysis

For each MST assay, the lower limit of quantification (LLOQ) was determined from the standard curve and the lowest concentration of standard gene copies that can be confidently detected in all triplicates was considered as the LLOQ (Schriewer et al., 2013, Ahmed et al., 2016a). All the qPCR results were normalized to gene copies/ng of DNA and the samples considered positive if the concentrations were above LLOQ. The qPCR assay results interpretation was as stated in previous studies (Nshimyimana et al., 2017, Ahmed et al., 2016b). Sensitivity and specificity of all qPCR assays were determined by the formulae below (Ahmed et al., 2013). To determine human-associated assay sensitivity, human fecal and sewage samples were considered as target samples.

Sensitivity= a/ (a+b) where, a and b represent true positives and false negatives, respectively.

Specificity =c/ (c+d) where c and d represent true negatives and false positives respectively.

The statistical and qPCR data analyses were carried out using either Microsoft Excel or SPSS version 22.0. The linear regression analysis was performed using Microsoft Excel; the statistical significance in the abundance of Bac-Uni and GenBac3 markers of fecal and sewage samples was done using SPSS 22.0 (IBM Inc., Chicago, IL, USA).

4.3 Results

4.3.1 Assessment of qPCR inhibitors and quality assurance of fecal DNA samples

Reliable quantification and specific detection of genetic markers by qPCR poses many challenges. The qPCR assays should be carefully designed and optimized to obtain maximum achievable specificity and sensitivity (Bustin et al., 2009). In the current study, BacUni qPCR assay was carried out to test the presence of amplifiable DNA in feces/sewage samples and to ensure that the DNA extracts were free of PCR inhibitors. The Ct mean values for the BacUni

marker in 1:10 and 1:100 diluted human fecal samples are given in Supplementary Table S4.1. All the fecal/sewage DNA samples (n=61) showed matching concentrations of BacUni markers in the two different dilutions tested (1:10 and 1:100) indicating the absence or negligible amounts of PCR inhibitors, and the presence of amplifiable DNA. All further assays were performed with 1:10 diluted DNA extracts.

4.3.2 Performance characteristics and LLOQ of MST qPCR assays

The amplification efficiencies of all MST qPCR assays tested were in the range 86% to 102% and the correlation coefficient (r^2) values were ≥0.98. The detailed performance characteristics of all the qPCR assays are provided in Supplementary Table S4.2; all of the values were within the limits recommended in the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin et al., 2009). All of the qPCR standards were reanalyzed to determine the master standard curve with standardized slope, amplification efficiencies, and correlation coefficient (r^2) values. The details of master standard curves and LLOQ for each tested MST assays are presented in Table 4.2 (Fig. S4.1A-D).

4.3.3 Performance of universal *Bacteroidales* assays

Both BacUni and GenBac3 assays, targeting universal/ general *Bacteroidales*, exhibited 100% sensitivity to fecal and sewage samples as they amplified DNA from all the samples (n=61). The mean concentration of these markers in fecal and sewage samples is given in Table 4.3. Comparatively, BacUni showed slightly higher total *Bacteroidales* concentrations (copies per nanogram (ng) of DNA) than GenBac3 assay in all of the tested samples. The normalized qPCR results for BacUni and GenBac3 assays were given in Supplementary Table S4.3A & B.

Table 4.2 Performance characteristics of MST qPCR assays tested using fecal and sewage samples

Assay	Slope	y-intercept	R^2	Efficiency	LLOQ* (cp/rxn)
BacUni	-3.32	43.3	0.99	100.0	24.5
GenBac3	-3.27	40.3	0.99	101.8	14.1
BacHum	-3.26	38.4	0.99	102	36.8
HF183 Taqman	-3.37	39.6	0.99	99.7	11.5
HumM2	-3.34	42.2	0.99	98.9	75
HF183 SYBR	-3.3	36.2	0.99	100.9	10
Pig-2-Bac	-3.27	41.1	0.99	102	30
BacCow	-3.31	41.6	0.99	100.3	100
AV4143	-3.5	43	0.99	93	10
GFD	-3.41	36.9	0.99	96	11.3

* Lower limit of Quantification

4.3.4 Performance of human-associated *Bacteroidales* assays

The specificity and sensitivity of the four human-associated assays on human and animal fecal and sewage samples were compared in order to select the most suitable assay for the Taihu watershed. The HF183 SYBR marker was the most sensitive marker among human-associated assays, though it showed high cross-reactivity. It was found in 80% of human origin samples (7/10 human feces and 5/5 sewage samples) at an average concentration of 2.72 log_{10} gene copies per ng of DNA and showed cross reactivity with chicken (7/10), cow (3/10), duck (1/3), dog (2/10) and goose (1/3) fecal DNA samples making it least specific (69.5%) (Table 4.4).

Table 4.3 Performance of universal/general *Bacteriodales* MST assays on fecal and sewage samples

Source	No. of samples tested	Bac-Uni		GenBac3	
		No. of positive samples	Mean (±SD) Concentration Log10 gene copies per ng	No. of positive samples	Mean (±SD) Concentration Log10 gene copies per ng
Human	10	10	6.95 (0.41)	10	5.91 (0.35)
Sewage	5	5	5.36 (0.82)	5	4.19 (1.30)
Pig	10	10	5.39 (0.67)	10	4.44 (0.79)
Chicken	10	10	3.86 (0.12)	10	2.85 (0.27)
Cow	10	10	5.31 (0.90)	10	4.80 (0.94)
Dog	10	10	4.25 (0.32)	10	3.82 (0.49)
Duck*	3	3	4.12 (0.31)	-	-
Goose*	3	3	5.81 (0.12)	-	-
Average			5.19 (1.52)		4.37 (1.56)
Sensitivity		100%		100%	

***composite fecal samples.

Table 4.4 Performance of human-associated *Bacteroidales* MST assays on fecal and sewage samples

Source	**No. of samples tested**	**BacHum**		**HF183 Taqman**		**HF183 SYBR**		**Hum2**	
		No. of positive samples	Mean Concentration (Log10 gene copies per ng)	No. of positive samples	Mean Concentration (Log10 gene copies per ng)	No. of positive samples	Mean Concentration (Log10 gene copies per ng)	No. of positive samples	Mean Concentration (Log10 gene copies per ng)
Human	10	6	3.98	6	4.51	7	3.99	7	3.28
Sewage	5	5	2.41	5	2.73	5	1.46	1	1.76
Pig	10	0	0	0	0	0	0	2	2.46
Chicken	10	6	2.53	7	3.22	7	2.08	4	2.03
Cow	10	1	1	0	0	3	1.39	2	2.12
Dog	10	2	2.1	2	1.28	2	1.39	1	2.24
Duck*	3	0	0	0	0	1	1.41	0	0
Goose*	3	0	0	0	0	1	1.68	1	1.97
Target			3.19		3.63		2.72		2.52
Non-Target			2.22		1.5		1.62		2.16
Sensitivity		73.3%		73.3%		80%		53.3%	
Specificity		80.4%		80.4%		69.5%		78.2%	

*Composite fecal samples

The BacHum maker was the highly specific marker (80.4%) among tested human-associated markers, along with HF183 Taqman. It had a sensitivity of 73.3% (6/10 human feces and 5/5 sewage samples) at an average concentration of 3.19 log_{10} copies per ng of DNA and was detected (above LLOQ) in three different host fecal DNA samples; chicken (6/10), cow (1/10) and dog (2/10). Similar to BacHum, the HF183 Taqman marker was highly specific (80%) among human-associated assays and was found in 73% of target source (6/10 human feces and 5/5 sewage samples) at an average concentration of 3.63 log_{10} copies per ng of DNA. However, HF183 Taqman marker was found in only two different host fecal DNA samples, chicken (7/10) and dog (2/10). The final human-associated assay, Hum2 was the least sensitive marker and was detected in only 53% samples of human origin (7/10 human feces and 1/5 sewage samples) at an average concentration of 2.52 log_{10} copies per ng of DNA. It was also detected in chicken (4/10), cow (2/10), pig (2/10), dog (1/10) and goose (1/3) fecal DNA samples, making it less specific (78%). In general, chicken (40 to 70%) and dog fecal samples (10 to 20%) had cross-reactivity with all of the human-associated assays. Overall, HF183 Taqman was the only assay that did not exhibit any cross-reactivity with cattle or swine fecal DNA samples, while HF183 SYBR, Hum2, and BacHum showed cross-reactivity with either cattle or swine fecal DNA samples. The normalized qPCR results for HF183 Taqman and BacHum assays were given in Supplementary Table S4.3A & B.

4.3.5 Performance of swine and ruminant associated *Bacteroidales* assays

The performance of the swine associated assay (Pig-2-Bac) was evaluated with 10 pig fecal samples and the target was found in 90% of pig fecal samples (9/10) at an average concentration of 2.81 log10 copies per ng of DNA (Table 4.5). The Pig-2-Bac marker was highly specific (95%) and it had a low level of cross-reactivity with cow fecal samples (2/10). The normalized qPCR results for Pig-2-Bac assay are given in Supplementary Table S4.5.

Table 4.5 Performance of animal associated MST assays on fecal samples

Source	**No. of samples tested**	**Pig-2-Bac**		**BacCow**		**AV4143**		**GFD**	
		No. of positive samples	Mean Concentration (Log10 gene copies per ng)	No. of positive samples	Mean Concentration (Log10 gene copies per ng)	No. of positive samples	Mean Concentration (Log10 gene copies per ng)	No. of positive samples	Mean Concentration (Log10 gene copies per ng)
Human	10	0	0	0	0	1	1.7	0	0
Pig	10	9	3.03	2	2.52	0	0	0	0
Chicken	10	0	0	2	2.32	10	4.13	7	2.16
Cow	10	2	1.75	10	4.3	2	1.3	1	1.2
Dog	10	0			0	0	0	3	1.4
Target average			2.81		4.3		4.13		2.16
Non-Target average			0.87		1.39		1.5		1.3
Sensitivity		90%		100%		100%		70%	
Specificity		95%		77.5%		95%		92.50%	

BacCow marker was found in 100% of cattle fecal samples (10/10) at an average concentration of 2.81 log10 copies per ng of DNA (Table 4.5). However, the BacCow marker showed some cross-reactivity with pig (2/10) and chicken (2/10) fecal DNA samples but was not found in any of the human fecal DNA samples at above LLOQ levels, resulting in 90% specificity.

4.3.6 Performance of avian associated MST assays

The avian associated MST assays performed very distinctively on the tested fecal DNA samples (Table 4.5). The AV4143 marker was found in 100% of chicken fecal samples but was also detected in human (1/10) and cow (2/10) fecal DNA. The mean concentration of AV4143 marker was 4.13 log10 copies per ng of DNA in chicken feces. Those human and cow samples for which there was cross reactivity comprised ca. 3 log-fold fewer gene copies than chicken faeces.

The GFD markers were only detected in 7 of the 10 chicken faeces samples and cross-reacted with dog fecal DNA samples (3/10). The mean concentration of GFD markers was 2.16 log10 copies per ng of fecal DNA sample, so ca. 100-fold less than AV4143. The normalized qPCR results for GFD assay was given in Supplementary Table S4.6.

4.4 Discussion

4.4.1 Assessment of universal *Bacteroidales* assays for their applicability at Taihu watershed

The high sensitivity of the BacUni (Universal *Bacteroidales*) marker for human and animal feces, excepting poultry, has been frequently reported in Asian countries such as India (Odagiri et al., 2015) and Singapore (Nshimyimana et al., 2017), and also in the USA (Kildare et al., 2007), and Kenya (Jenkins et al., 2009). The GenBac3 general *Bacteroidales* assay is the most widely used assay in the USA to quantify *Bacteroidales* in environmental samples (Shanks et al., 2009, Ervin et al., 2013). In our study, both BacUni and GenBac3 showed 100% sensitivity

to fecal samples but the BacUni assay amplified higher copies per ng of DNA than GenBac3 assay in all of the tested samples. We conclude that the BacUni assay is more suitable for quantification of total *Bacteroidales* in the Taihu watershed, China; Odagiri et al. (2015) also reported that the BacUni marker quantified higher copy numbers than GenBac3 in fecal DNA samples tested in Odisha, India.

4.4.2 Assessment of human-associated *Bacteroidales* assays for their applicability at Taihu watershed

The ability of human-associated markers to identify human fecal sources as distinct from other sources in an aquatic environment is vital to the MST approach. Variation, due for example to the impact of dietary pattern on gut microbiomes (Wu et al., 2011) or geographical population differences (Yatsunenko et al., 2012), could significantly affect the performance of MST assays and this has been supported by several validation studies across different countries in recent years (Jenkins et al., 2009, Ahmed et al., 2009a, Reischer et al., 2013, Odagiri et al., 2015, Boehm et al., 2016, He et al., 2016b, Nshimyimana et al., 2017, Malla et al., 2018). Therefore, it is important that the performance of these different human faecal markers be validated in context. Here, with the exception of the HF183 SYBR assay, all of the human-associated assays showed lower sensitivity (53 to 73%) than expected (81% to 100%). This reduced sensitivity range could well be due to the geographical variability in human gut microbiomes (Yatsunenko et al., 2012). The HF183 SYBR assay which performed well in Bangladesh (Ahmed et al., 2010) and in Singapore as a human sewage indicator (Nshimyimana et al., 2014), showed less specificity in our study. Our results on the specificity of this assay are more in line with previous studies conducted in India (Odagiri et al., 2015) and Nepal (Malla et al., 2018). Although the remaining three assays are less sensitive compared to HF183 SYBR, they are more specific. The BacHum and HF183 Taqman assays showed relatively more specificity (80.4%), followed by Hum2 with 72.5% specificity. All of the human target assays gave some

level of cross-reactivity with other faecal samples, with the highest levels recorded in chickens (Table 4.4). The high cross-reactivity with chicken fecal samples by human-associated markers (HumM2, HF183 Taqman, and HF183 SYBR) has been described in earlier studies directed in South Asia (Odagiri et al., 2015, Nshimyimana et al., 2017). The occurrence of the HF183 marker in dogs and chicken has also been reported previously, with the HF183 forward primers with Bac708 reverse primer picking up a target sequence in chicken fecal samples (Ahmed et al., 2012, Balleste et al., 2010, Gourmelon et al., 2007). The BacHum assay's cross-reactivity with chicken fecal DNA has also been indicated previously in several studies (Reischer et al., 2013, Odagiri et al., 2015, Nshimyimana et al., 2017). With the exception of HumM2, none of the markers exhibited cross-reactivity with pig fecal DNA, which is encouraging as pigs are major livestock animals in Zhejiang province (Taihu watershed region). An assay that showed zero cross-reactivity to pig fecal DNA and highly sensitive to sewage samples would be considered as very suitable for source tracking in the Taihu watershed region. In this study, HF183 Taqman and BacHum assays showed the same accuracy, but HF183 Taqman was found to be the more suitable human-associated fecal maker than BacHum as it amplified DNA from all of the sewage samples with higher abundance and did not show any cross-reactivity with pig and cattle fecal samples (Table 4.4). The HF183 Taqman's cross-reactivity to chicken fecal samples can be negated by employing avian associated assays, such as GFD assay, in tandem to verify the existence of true human fecal pollution. An evaluation study carried out in Singapore (Nshimyimana et al., 2017) indicated similar sensitivity and specificity for HF183 Taqman to that reported here, though that study recommended another human-associated assay, B. theta (Taqman), for source tracking of human fecal and sewage in Singapore. The inconsistencies in the performance of *Bacteroidales* assays in regions other than the original has been reported elsewhere (Reischer et al., 2013, Layton et al., 2013, Boehm et al., 2016, Harris et al., 2016). As indicated previously, this could be due to geographical differences,

therefore location validation is highly recommended prior to application in a new region (Wuertz et al., 2011); here, that validation is recognized for HF183 Taqman.

4.4.3 Assessment of swine and ruminant-associated Bacteroidales assays for their applicability at Taihu watershed

The Swine associated marker Pig-2-Bac has been broadly evaluated and used for MST studies across different countries, including China and Nepal (Marti et al., 2011, Pisanic et al., 2015, Arfken et al., 2015, Jia et al., 2014, Malla et al., 2018, He et al., 2016a). He et al. (2016a) validated the Pig-2-Bac assay on target and non-target fecal samples of Taihu watershed region and reported that it was more sensitive and specific than mitochondrial DNA based swine markers. They have also applied the Pig-2-Bac marker for identifying swine fecal pollution in the Taige River of Taihu watershed. Jia et al. (2014) used Pig-2-Bac assay for quantification of swine fecal markers in Hongqi River, Yongan River, and Taige River. The results of the swine associated assay in our study are in complete agreement with the findings of He et al. (2016a), and the suitability of the Pig-2-Bac assay for identifying contamination by pig fecal sources at Taihu watershed is confirmed.

The BacCow marker, which was originally developed to detect fecal sources of cow or cattle origin (Kildare et al., 2007), has shown cross-reactivity with fecal samples collected from other ruminants (e.g. deer) and non-ruminants such as horse, pig, dog, and chicken in a California based validation study leading to its reclassification as ruminant-associated marker (Raith et al., 2013). Similar findings were reported in evaluation studies conducted in Australia and Europe, indicating that this assay had cross-reactivity with non-targets such as chicken, goose, dog, pig and duck (Ahmed et al., 2013, Reischer et al., 2013). In a validation study conducted in India, BacCow markers were reported to be found in all types of composite livestock/domestic animal feces (cow, buffalo, goat, sheep, dog, and chicken) but not in tested human samples and it was recommended that the BacCow assay could be used to detect fecal

pollution by livestock/domestic animals (Odagiri et al., 2015). More recently, BacCow was detected in all fecal samples tested including human origin (composite sewage) in a validation study conducted in Nepal (Malla et al., 2018). In order to check the true specificity and sensitivity of BacCow, our study was conducted with individual fecal samples instead of composite fecal samples. Here, we detected BacCow in all of the cattle fecal samples tested but found some cross-reactivity with pig (2/10) and chicken (2/10) samples in agreement with the findings of Ahmed et al (2013) and Reischer et al (2013). Since the BacCow marker was not identified in all of the livestock/domestic fecal samples tested, it is not applicable in the Taihu watershed as a livestock/domestic animal fecal source-tracking marker.

4.4.4 Assessment of avian associated assays for their applicability at Taihu watershed

The ability to draw conclusions on the microbial flora in avian fecal samples is still ambiguous worldwide, making it difficult to develop reliable and specific qPCR assays for specific detection of avian fecal contaminations (Ohad et al., 2016). This could be due to variation in food intake by the avian sources regionally and seasonally (Ahmed et al., 2016a). Though *Bacteroides* and its closely related organisms are commonly used for identifying the source of fecal contamination for human and animals (Kildare et al., 2007, Bernhard and Field, 2000b, Layton et al., 2006), previous studies reveal that *Bacteroides* are rarely present in avian sources (Lu et al., 2009). The phylum level mapping of avian fecal samples showed that Firmicutes, Proteobacteria, and Fusobacteria are the main phyla (Lu et al., 2009, Dick et al., 2005). *Bacteroidales* members were not frequently reported in avian gut or excreta and they were found to be nearly absent in a few studies (Zhu et al., 2002) or they were identified in varying frequencies in other studies (Scupham et al., 2008). Therefore, MST markers targeting avian fecal markers are still limited (Fremaux et al., 2010, Green et al., 2012, Ohad et al., 2016). In the current study, AV4143 assay targeting *Lactobacillus* showed higher sensitivity and specificity than the GFD assay targeting *Helicobacter* spp. However, GFD markers were not

detected in human fecal samples making this assay very suitable for application in the Taihu watershed as it can reliably differentiate chicken and human fecal samples. The GFD assay was similarly validated in Australia, New Zealand, and North America with similar results for sensitivity and specificity (Ahmed et al., 2016b).

4.5 Conclusions

- Both BacUni and GenBac3 markers were highly sensitive to fecal and sewage samples and BacUni markers were comparatively more abundant than the other markers.
- HF183 Taqman was found to be a more suitable assay among human associated assays as it is highly specific and did not show any cross reactivity with pig and cow fecal samples.
- The swine associated assay, Pig-2-Bac was highly specific and sensitive which could be a potential marker for tracking swine fecal pollution.
- BacCow marker was not detected in all animal samples tested; therefore, it is not recommended for detecting total livestock fecal contamination.
- GFD marker was highly specific and did not cross-react with human fecal samples, indicating its efficiency to discriminate avian from human fecal sources.

CHAPTER-5

Quantification of Microbial Source Tracking Markers and Pathogenic Bacterial Genes in Water and Sediment of Taihu Watershed

Abstract:

This chapter was aimed at investigating the fecal pollution and the presence of pathogenic bacteria in water and sediments of Tiaoxi River by qPCR assays to assess the human health risks. Samples collected from 15 locations in the Tiaoxi River during three seasons in 2014-2015 were used for quantification of MST markers comprising total, human and swine associated *Bacteroidales* genes together with avian associated markers, and specific gene markers for several pathogenic bacterial species. The total *Bacteroidales* markers were detected in all water and sediment samples at an average concentration of 6.22-log10 gene copies/100ml and 6.11-log10 gene copies/gram, respectively however, host-associated MST marker detection varied. The human (97%) and avian (89%) markers were the most frequently detected host-associated markers in water samples with a mean concentration of 6.22 and 2.70 log10 gene copies/100ml, respectively. However, in sediment samples, human-associated markers were detected more often (86%) than swine and avian markers with a mean concentration of 3.91 log10 gene copies/gram. Swine associated markers were frequently detected in water and sediment samples collected from two locations that correlated with the presence of pig farms nearby. Among the five bacterial pathogens tested, *Shigella* spp. and *Campylobacter jejuni* were the most frequently detected pathogens in water (60% and 62%, respectively) and sediment samples (91% and 53%, respectively). Shiga toxin-producing E.coli (STEC) and pathogenic *Leptospira* spp. were less frequently detected in water (55% and 33%, respectively) and sediment samples (51% and 13%, respectively) whereas *E.coli* O157: H7 was detected in sediment samples only (11%). Overall, the higher prevalence and concentrations of *Campylobacter*, *Shigella* spp. and STEC along with MST markers at several locations of Tiaoxi River indicates poor water quality and significant human health risk at those locations. The data generated here would be valuable to water quality monitoring authorities to minimize health risk associated with pathogens

5.1 Introduction

Fecal contamination of drinking water sources, shellfish harvest, and recreational waters is a major concern to public health as it promotes human exposure to pathogenic microorganisms (Napier et al., 2017). Therefore, continuous monitoring and proper protection of these waters are required. Traditionally, fecal indicator bacteria (FIB) are used to monitor pollution in environmental waters and to assess the associated public health risks (Griffith et al., 2009). However, there are several limitations of using FIB for microbial water quality monitoring, as these bacteria can persist and replicate outside of the host making it difficult to predict identify recent water contamination (Byappanahalli et al., 2003, Jamieson et al., 2005) and poor correlation of FIB and pathogen presence (Ahmed et al., 2013, McQuaig et al., 2012). The major limiting factor is that FIB detection does not indicate the source or origin of fecal contamination (Field and Samadpour, 2007), which is very important in order to characterize the public health risk potential (Santo Domingo et al., 2007), implement remediation measures and prevent further contamination. Therefore, microbial source tracking (MST) techniques have been developed over the last decade to unequivocally identify the sources and origins of fecal pollution.

Both library-dependent (LD-MST) and library-independent MST (LI-MST) methods were developed for identifying the sources of fecal pollution, though LD-MST methods have several limitations in correctly assigning fecal contamination of host-specific sources (Field and Samadpour, 2007, Harwood et al., 2003). However, quantitative PCR (qPCR) based LI-MST techniques have proven to be widely applicable to study fecal contamination in environmental waters as they can accurately quantify the host-specific MST target sequences (Layton et al., 2006, Reischer et al., 2007, Kildare et al., 2007). Among the LI-MST methods, *Bacteroidales* are often used as target organism as they are obligate anaerobic bacteria found in the human and animal gut at higher levels than facultatively anaerobic *E. coli* (Bernhard and Field, 2000b);

host-associated *Bacteroidales* 16S rRNA gene markers have been developed for different hosts to discriminate human feces from other animal feces detected in the environment (Bernhard and Field, 2000a). Specifically, several host-associated *Bacteroidales* markers are available to identify human (Kildare et al., 2007, Green et al., 2014), ruminant (Raith et al., 2013), cow (Shanks et al., 2008), and pig fecal sources (Mieszkin et al., 2009). Avian fecal markers that target bacterial taxonomic groups such as *Helicobacter* spp. (Green et al., 2012) have also been developed.

Although MST methods provide information on the sources of fecal pollution, these methods do not provide evidence or confirm the presence of bacterial pathogens and the associated public health risk. Determining the correlation between MST data and direct pathogen detection has been addressed in a relatively few cases (Ridley et al., 2014, Bradshaw et al., 2016). The correlation of MST markers with pathogen presence in environmental samples has given mixed results in the MST (Tambalo et al., 2012, Marti et al., 2013). Field studies to evaluate the correlation of MST marker and pathogen presence along with the probability of pathogens to cause illness at a watershed are more important than normal fecal monitoring by FC enumeration or MST methods in terms of public health risk. A study of this nature has not previously been carried out on the Taihu watershed region and the results generated here will also inform water quality monitoring management at other watersheds in Asia, and beyond.

Therefore, this study was aimed at assessing the presence and abundance of MST markers and genes of bacterial pathogens at Tiaoxi River water and sediments. The specific objectives were 1) to assess the abundance of universal and host-associated MST markers and genes of bacterial pathogens at Tiaoxi River, and 2) to determine whether any correlation exists between the MST markers and host-associated bacterial pathogens tested, and also to assess the potential human health risk associated with quantified pathogen based on lowest infectious dose.

5.2 Materials and Method

5.2.1 Overview of the study area

As described earlier, Tiaoxi River flows in northern Zhejiang province covering upstream agricultural areas and downstream urban cities of northern Zhejiang Province. It has been estimated that the River collects water from one million inhabitants residing in the moderately sized cities of Zhejiang Province (Zhang et al. 2016). Poultry is a common livestock resource in Zhejiang province, and waste disposal is an issue (Zheijiang, 2016). Excluding poultry, pigs comprise more than 90% of the remaining livestock population in this region (NBSC, 2015).

5.2.2 Sample collection and processing

As mentioned in previous Chapter(s), 15 out of 25 sampling locations in the Tiaoxi River were identified as hotspots of fecal contamination, and samples collected from those locations were used further studies. The details of land use type around the selected sampling locations are given in Table 2.1 (Chapter-2). Sampling was carried out at these locations on three occasions: autumn 2014, winter 2015 and summer 2015. Water samples were collected in sterile 5L polypropylene containers and sediment was collected using a sediment sampler; the samples in triplicate were transferred to sterile 50mL tubes. The water and sediment samples were transported to the laboratory on ice and were processed immediately. Sediment samples were frozen at -20°C and the water samples (250 ml) were filtered through 0.22 µm polycarbonate membrane filters (Millipore, UK) and stored at -20°C prior to DNA extraction.

5.2.3 Enumeration of fecal coliforms

The culture based microbiological assessment data of the Tiaoxi River were reported earlier in chapter-2 and the fecal coliforms data were used to evaluate the correlation between FIB, MST markers and pathogenic bacterial genes. The enumeration of fecal coliforms (FC) was carried out by standard membrane filtration technique (APHA, 2005) using mFC agar (Difco, Germany) according to manufacturer’s instructions. Water samples were filtered through 0.45

μm nitrocellulose filters (Millipore, UK) in duplicates and the filters were placed on mFC agar and incubated at 44.5 °C for 24 h. Colonies exhibiting any shade of blue color were counted and expressed as CFU/100 ml (APHA, 2005).

5.2.4 Molecular analysis

5.2.4.1 DNA Extraction

DNA was extracted from membrane filters (water samples) and sediment samples using PowerSoil DNA Isolation Kit (MoBio Inc., Carlsbad, CA). The membrane filters were cut into pieces, and placed into the PowerBead tubes aseptically for DNA extraction. Sediment samples were centrifuged initially for 20 min at 3000×*g* to remove excess water and 0.25g of sediment was transferred to a PowerBead tube and the DNA was extracted from sediments and membrane filters as per the manufacturer's instructions. The quality and quantity of DNA were analyzed using NanoDrop ND 1000UV spectrophotometer (Thermo Fisher Scientific., Vienna, Austria). The DNA extracts were stored at -20°C until further analysis.

5.2.4.2 Plasmid DNA standards for qPCR assays

Plasmid DNA standards were prepared for both MST markers and genes of pathogenic bacteria and used for qPCR assays. The plasmid DNA standards for MST markers were prepared as mentioned in Chapter-4 (section 4.2.4). For pathogenic bacterial genes, the target gene of pathogenic bacteria were PCR-amplified from respective genomic DNA of target organisms (*Campylobacter jejuni* sub sp. *jejuni* ATCC 29428, *Shigella sonnei* ATCC 9290, *Leptospira borgpetersenii* serovar Hardjo ATCC 14028 and *Escherichia coli* ATCC 35150) using the primers designed elsewhere (Table 5.1). The amplified PCR products were purified using a QIAquick PCR purification kit (Qiagen, GmbH, Germany) following manufacturer's instructions. The purified PCR products were cloned separately into a pGEM-T Vector (Promega, Madison, WI), followed by transformation of the recombinant plasmid into *E. coli* JM109 competent cells. The plasmid DNA was extracted from positive clones using QIAprep

Spin Miniprep Kit (Qiagen, GmbH, Germany) and target gene presence in the recombinant plasmid was confirmed by PCR and sequencing (GATC, Germany). Plasmid DNA was quantified with a NanoDrop ND-1000 UV spectrophotometer to analyze the purity and concentration for calculating the gene copy number using the formula mentioned in Oster et al. (2014). The sequencing results for pathogenic bacterial genes was given in Supplementary Note S1 and schematic representation of pGEM®-T Easy cloning vector with restriction sites was given in Supplementary Figure S5.1.

5.2.4.3 Quantitative PCR

A seven-point 10-fold serial diluted recombinant plasmid DNA with target sequence was used to generate standard curve (with a range 10^1 to 10^7 copies/reaction) in each qPCR assay. All qPCR reactions were run in triplicates with a final reaction volume of 20 µL. The sequences of the primers and probes along with concentrations used in the current study are presented in Table 5.1. Prior to the quantification, the absence of PCR inhibitors was analyzed in at least 10% of DNA samples by applying BacUni qPCR assay that amplifies the 16S rRNA gene of *Bacteroidales* (Odagiri et al., 2015). PCR inhibition was evaluated by determining the BacUni marker concentrations in the undiluted and 1:4 diluted DNA samples. The undiluted DNA samples were judged as free from PCR inhibitors if the BacUni markers concentrations were matching in the two dilutions (undiluted and 1:4 diluted) (Odagiri et al., 2015, Reischer et al., 2013). The accuracy and efficiency of the standard curve were determined by including a positive control of 10^3 copies of plasmid standard as unknown in each assay (Oster et al., 2014).

Four previously designed qPCR assays targeting total, human, swine, and avian associated fecal source were selected for MST study at Taihu watershed region (Kildare et al., 2007, Green et al., 2014, Mieszkin et al., 2009, Green et al., 2012). Three TaqMan assays (BacUni, HF183 Taqman and Pig-2-Bac) were selected for detection of total, human and swine associated

Table 5.1 List of primers and probes used for quantification of MST markers and genes of pathogenic bacteria.

Target source/ organism	Assay	Primer/probe	Concentration	Oligonucleotide sequence (5'–3')	Annealing temperature (oC)	Reference
Total *Bacteroidales*	BacUni	BacUni- 520F	400nM	CGTTATCCGGATTTATTGGGTTTA	60	Kildare et al. 2007
		BacUni-690R1	400nM	CAATCGGAGTTCTTCGTGATATCTA		
		BacUni656P	80nM	FAM-TGGTGTAGCGGTGAAA-MGB		
Human associated *Bacteroidales*	HF183	HF183F	1000nM	ATCATGAGTTCACATGTCCG	60	Green et al. 2014
		BacR287R	1000nM	CTTCCTCTCAGAACCCCTATCC		
		BacP234P	80nM	FAM-CTAATGGAACGCATCCC-MGB		
Swine associated *Bacteroidales*	Pig-2-Bac	Pig-2-Bac41F	300nM	GCATGAATTTAGCTTGCTAAATTTGAT	60	Mieszkin et al. 2009
		Pig-2-Bac163R	300nM	ACCTCATACGGTATTAATCCGC		
		Pig-2-Bac113P	200nM	VIC-TCCACGGGATAGCC-MGB		
Avian associated Marker	GFD	GFD-F	100nM	TCGGCTGAGCACTCTAGGG	57	Green et al. 2012
		GFD-R	100nM	GCGTCTCTTTGTACATCCCA		
C. jejuni	mapA	mapA F	400nM	CTGGTGGTTTTGAAGCAAAGATT	60	Best et al., 2003
		mapA R	400nM	CAATACCAGTGTCTAAAGTGCGTTTAT		
		mapA P	80nM	FAM-TTGAATTCCAACATCGCTAATGTATAAAAGCCCTTT-TAMRA		
Pathogenic *Leptospira* spp.	LipL32	LipL32F	300nM	AAG CAT TAC CGC TTG TGG TG	60	Stoddard et al., 2009
		LipL32R	300nM	GAA CTC CCA TTT CAG CGA TT		
		LipL32P	200nM	FAM-AAAGCCAGGACAAGCGCCG-BHQ1		
Shigella spp.	ipaH	ipaH F	400nM	CTTGACCGCCTTTCCGATA	64	Oster et al., 2014
		ipaH R	400nM	AGCGAAAGACTGCTGTCGAAG-		
		ipaH P	80nM	FAM-AAC AGG TCG CTG CAT GGC TGG AA-TAMRA		
E.coli (STEC)	Stx2	Stx2F	200nM	CAGGCAGATACAGAGAGAATTTCG	61	Beutin et al., 2008
		Stx2R	200nM	CCGGCGTCATCGTATACACA		
E.coli O157:H7	eae	eae-F	200nM	GTAAGTTACACTATAAAAGCACCGTCG	56	Ibekwe et al. 2002
		eae-R	200nM	TCTGTGTGGATGGTAATAAATTTTTG		

*Bacteroidale*s and one SYBR green- GFD assay was selected for detection of avian associated fecal markers, as validated in our previous study for Taihu watershed region (Unpublished data; the details of the assays and limit of detection (LOD) are given in the supplementary Table S5.1). Five qPCR assays targeting *stx2, eae, LipL32, ipaH*, and *mapA* genes of pathogenic bacteria were selected for this study. Two qPCR assays targeting *eae* gene specific for *E. coli* O157: H7 (Ibekwe et al., 2002) and stx2 gene specific for Shiga toxin producing *E. coli* (Beutin et al., 2008) were applied using SYBR Green chemistry. Three assays targeting LipL32 gene specific for pathogenic *Leptospira sp.* (Stoddard et al., 2009), ipaH gene specific for *Shigella spp*. (Wang et al., 2007a) and mapA gene specific for *Campylobacter jejuni* (Best et al., 2003) used TaqMan chemistry.

The Taqman qPCR assays (20 µl of master mix), contained 10µl of TaqMan Environmental Master Mix 2.0 (Applied Biosystems, UK), 2µL of template DNA, 6µl nuclease-free water and 2µL of primers and probes with the final concentrations as shown in Table 1. SYBR Green assays (20 µl of master mix), contained 10µl of SYBR Green PCR Master Mix (Thermofisher Technologies, Foster City, CA), 7.0 µL nuclease-free water, 2 µL of template DNA and 1µL of primer mixture with a final concentration as shown in Table 1.

5.2.5 Data processing and statistical analyses

All the assays with R^2 values of above 0.95 and efficiencies between 80 and 110% were considered as acceptable for detection and quantification of target markers in environmental samples and if these criteria were not met by any assay, the samples were tested again. The details of the limit of detection (LOD), limit of quantification (LOQ) and final assessment of qPCR results for each MST and pathogen quantification assays are provided as supplementary materials (Table S5.1). The MST qPCR results for each assay of this study was processed based on LOD as mentioned in supplementary Table S5.1 and for genes of pathogenic bacteria as

described by Oster et al 2014 (Oster et al., 2014) (Table S5.1). For statistical analysis, the concentrations of FC, MST markers, and genes of pathogenic bacteria were log transformed and non-detects were assigned as 0 (Bradshaw et al., 2016). The data were not normally distributed (based on Kolmogorov- Smirnov test) after transformation, therefore, Kruskal-Wallis non-parametric ANOVA with Dunn's post-test was used for determining statistical significance among different sampling locations. The correlation among FC, MST markers and genes of pathogenic bacteria in water samples was analyzed by Spearman's coefficient correlation. The correlation analysis was not carried out for sediment samples, as most of the genes of pathogenic bacteria and some of the MST markers were not quantifiable in the tested samples.

5.3 Results

In total, 138 samples (69 water and 69 sediment) were collected from 25 sampling locations in three different seasons (Autumn 2014, Winter and Summer 2015) were investigated for the presence and abundance of fecal coliforms, MST markers and bacterial pathogens in the Tiaoxi River.

5.3.1 Enumeration of fecal coliforms in water samples

Fecal coliform (FC) counts data was used for initial assessment of fecal contamination in sampling locations and the detail results were presented in Table S2.1. Fecal indicator bacteria guidelines suggested by Ministry of Environmental Protection (MEP), China for surface water includes both Total coliforms and FC, therefore FC count observed in the present study was compared to USEPA standards (USEPA, 2012). Elevated levels of FC (>250CFU/100ml) were observed in fifteen locations (1-6, 8, 10, 12-16, 20 and 21) on one or more occasions (Chapter-2, Fig. 2.4C). In general, the FC counts were high during summer, which could be linked to the warm (optimal) weather that could support FC bacteria to acclimatize to those conditions and a rainfall event that occurred the day before the sampling might increase the transport of fecal

matter and proliferation of FC bacteria (Muirhead et al., 2004, Heaney et al., 2015). Based on the elevated levels of FC count, only 15 locations were considered as preliminary hotspots of fecal contamination and were selected for quantification of MST fecal markers and genes of pathogenic bacteria in DNA extracted from water and sediment samples.

5.3.2 Quality assurance of extracted DNA and performance of qPCR assays

DNA samples tested for PCR inhibition did not show any major difference between the mean CT values obtained for undiluted and 1:4 diluted DNA extracts, indicating that the samples were free from potential PCR inhibitors. Therefore, all the qPCR assays were conducted using undiluted DNA samples. The amplification efficiency and linear range of quantification (R^2) for all the qPCR assays were determined using standard curves generated by serial dilutions of known copy numbers. The qPCR amplification efficiencies for MST fecal markers and pathogenic bacterial genes ranged from 83 to 105.8%, with R^2 values between 0.97 and 0.99. The compiled amplification efficiencies and linear range of quantification for all qPCR assays carried out are provided in the Supplementary Table S5.1. The average Ct values for all negative controls (NTC) was > 38 and samples were tested again if average Ct value of NTCs were <38.

5.3.3 Detection frequency and concentration of MST markers in water and sediment samples

The presence and distribution of MST markers in water and sediment samples at 15 sampling locations determined by qPCR is shown in Table 5.2. Since the monitoring locations selected for MST study are presumed hot spots of fecal contamination based on FC count, the total *Bacteroidales* marker was detected in all water and sediment samples at the 15 locations. In water samples, the mean concentration of total *Bacteroidales* marker was 6.22-$\log_{10}$ gene copies/100ml with concentrations ranging from 4.62 to 7.63-$\log_{10}$ gene copies/100ml

(Fig.5.1A). For sediment, the mean concentration was 6.11-log_{10} gene copies/gram and concentrations were ranged from 4.37 to 7.82 log_{10} gene copies/ gram (Fig.5.2A). Significant statistical variation in total *Bacteroidales* concentration among different location was observed in both water ($P<0.015$) and sediment samples ($P<0.003$). In relative terms, the total *Bacteroidales* concentration was high during winter for both water and sediment samples. Based on the quantification of total *Bacteroidales*, the water samples from location 16 and sediment samples from location 15 were found to be most fecally polluted, regardless of the fecal source (human versus animal species).

For the host associated MST marker analysis in water samples, human-associated markers (97%) are the most frequently detected, followed by avian (89%) and swine (84%). In sediment samples, human-associated MST marker was detected more (86%), followed by swine (64%). The avian marker was positive for several sediment samples but they were below the limit of quantification. The human-associated marker, HF183 Taqman was detectable at most of the locations tested in water samples (44/45) in all three seasons with a mean concentration of 3.75-log_{10} gene copies/100ml and the concentrations were ranged from 2.91 to 5.6 log_{10} gene copies/100ml (Fig.5.1B). For sediment samples, the HF183 Taqman marker was detected with high frequency (39/45) at concentrations ranging from 3.8 to 5.6-log_{10} gene copies/gram with an average of 3.91 log_{10} gene copies/gram (Fig.5.2B). During the summer, human-associated markers were detected at a higher prevalence in water and sediment samples, probably due to the runoff water received from heavy rainfall (349 and 268 mm in July and August 2015) that occurred before summer sampling (NBSC, 2016). There was a significant statistical difference in prevalence of human markers in sediment samples ($P<0.004$) but not for water samples. Based on HF183 Taqman assay, water samples collected from locations 12 and 16 and sediment samples collected from location 15 had the highest concentration of human-associated marker during three seasons. Location 12 is very close to Huzhou city and as stated

Table 5.2 Detection frequencies of MST markers in water and sediment samples of Tiaoxi River, Taihu watershed (2014-2015).

Sample type	No. of samples tested (n)	No. of positive samples (%) [a]			
		Total *Bacteroidales*	Human associated markers	Swine associated markers	Avian associated markers
Water					
Autumn	15	15 (100%)	15 (100%)	10 (66%)	13 (86%)
Winter	15	15 (100%)	15 (100%)	13 (86%)	15 (100%)
Summer	15	15 (100%)	14 (93%)	15 (100%)	12 (86%)
Total	**45**	**45 (100%)**	**44 (97%)**	**38 (84%)**	**40 (89%)**
Sediment					
Autumn	15	15 (100%)	13 (86%)	6 (40%)	DNQ
Winter	15	15 (100%)	11 (73%)	8 (53%)	DNQ
Summer	15	15 (100%)	15 (100%)	15 (100%)	4
Total	**45**	**45 (100%)**	**39 (86%)**	**29 (64%)**	**4 (8.8%)**

**DNQ=Detected but not quantifiable. *[a] DNQs as negative.*

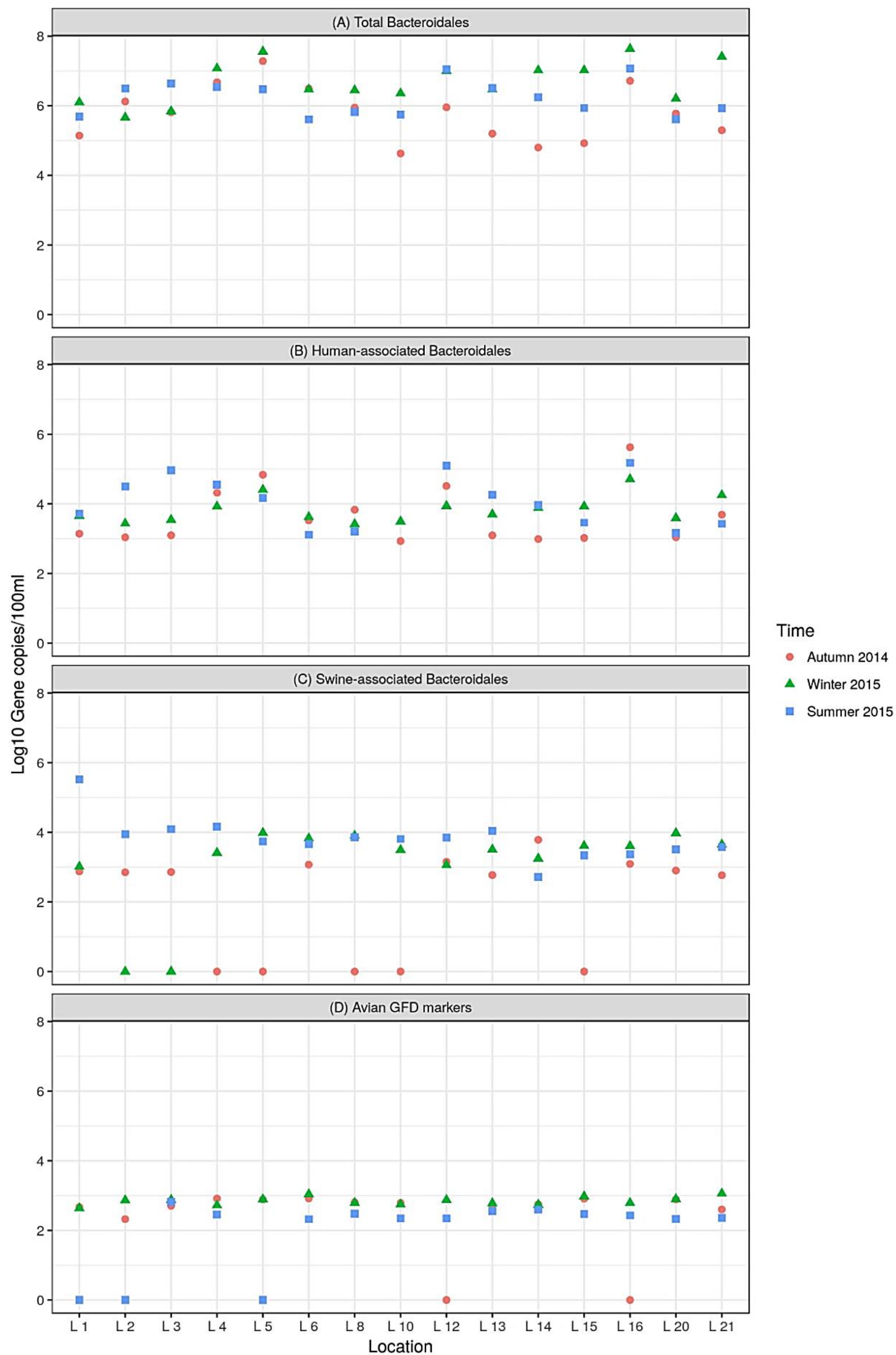


Figure 5.1 Concentration of MST fecal markers quantified in **water** samples at different sampling locations of Tiaoxi River. A) Total *Bacteroidales*; B) Human associated *Bacteroidales*; C) Swine associated *Bacteroidales* and D) Avian associated MST marker.

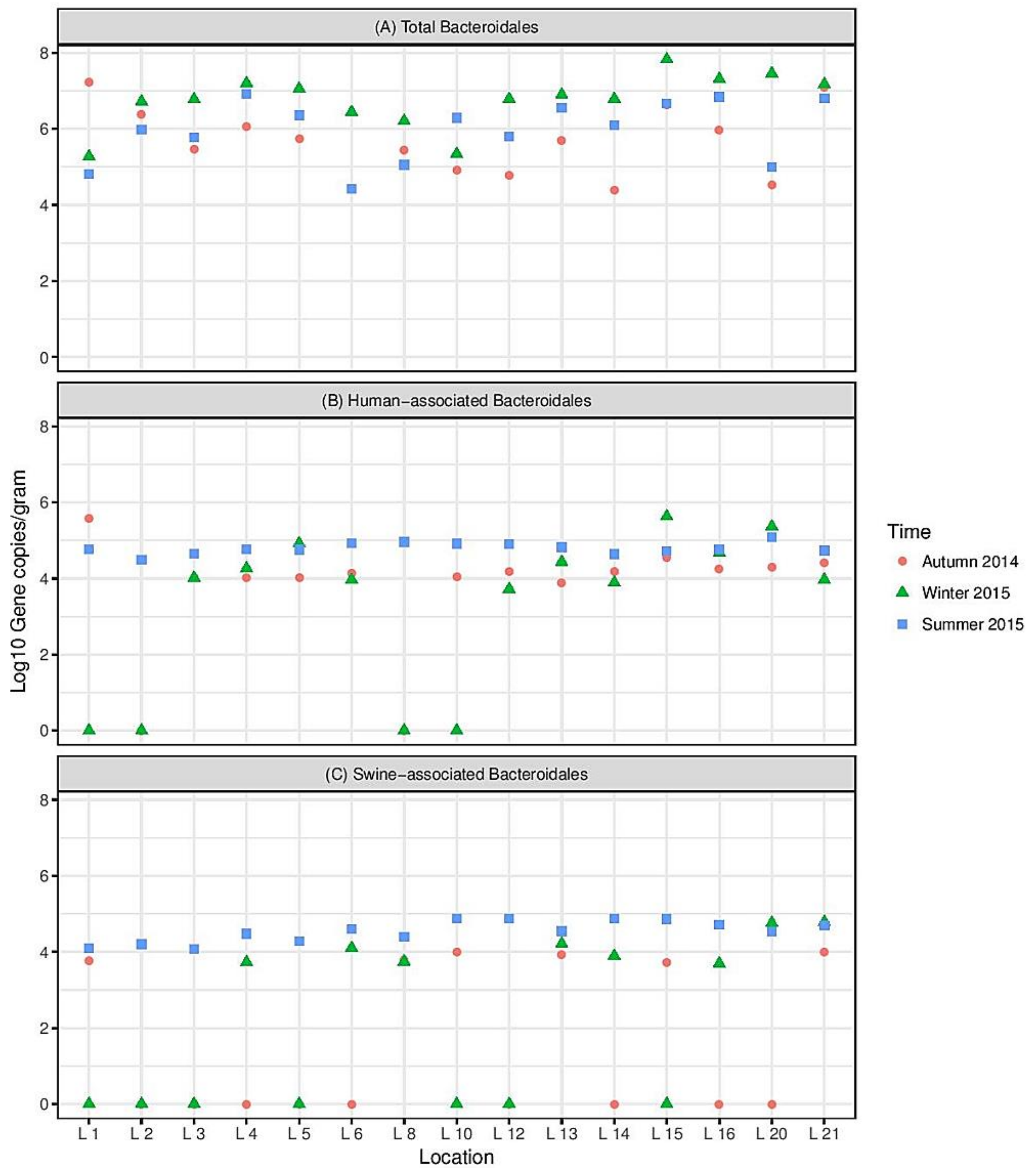


Figure 5.2 Concentration of MST markers quantified in **sediment** samples in different sampling locations in Tiaoxi River. A) Total *Bacteroidales*; B) Human associated *Bacteroidales*; C) Swine associated *Bacteroidales* and D) Avian associated MST marker.

earlier, the locations 15 and 16 are suburban areas close to a WWTP and the sampling was carried out at the junction of Tiaoxi River and a canal that enters Taihu Lake (Table. 2.1 of Chapter-2).

With respect to swine fecal contamination, swine associated marker was detected less often in water (38/45) and sediment samples (29/45) than human markers. The mean concentration of swine marker in water samples was 2.96-log_{10} gene copies/100ml and concentrations were ranged from 2.77 to 5.56-log_{10} gene copies/100ml of the water sample (Fig. 5.1C). The mean concentration of swine markers in sediment samples was 2.75-log_{10} gene copies/gram and concentrations were ranged from 3.6 to 4.7 log_{10} gene copies/ gram (Fig.5.2C). The variation in the concentration of swine markers detected in water ($P> 0.001$) and sediment samples ($P>0.001$) was statistically significant. Higher concentration of swine associated marker was observed in water samples at location 1 during summer and in sediment samples at location 10 during summer. The samples at Location 1 were collected 1km inside the Taihu Lake from the junction between Tiaoxi River and Taihu Lake and runoff during the summer was probably a contributory factor. Location 10 was close to a rural area where agricultural input to the River upstream was likely to enhance the swine marker content. Location 10 was close to a rural area where agricultural input to the River upstream was likely to enhance the swine marker content.

The avian associated fecal pollution was found to be the second dominant host associated fecal pollution in Tiaoxi River water samples. The avian markers were detected in 89% of water samples (40/45) at a concentrations ranging from 2.30 to 5.56-log_{10} gene copies/100ml at an average of 2.70 log_{10} gene copies/100ml (Fig.5.1D). In case of sediment samples, the avian markers were detected only in 4 of the 15 samples collected in the summer season. During autumn and winter seasons, though some samples were positive to the avian MST assay, they were below the quantification limits. The positive samples were tested again to confirm their presence but the results were not reproducible. The highest avian marker concentration was

observed at location 6 for water and sediment samples. Location 6 is in suburban area near Qijia village, where several swine and poultry farms are present (Supplementary Figure S5.2).

5.3.4 Concentration of pathogenic bacterial genes in water and sediment samples

The presence and distribution of gene markers for five bacterial pathogens is presented in Table 5.3. Considering detected not quantifiable (DNQs) as positive samples, the most commonly detected pathogens in water and sediment samples respectively were *Campylobacter jejuni* (62% and 53%) and *Shigella* spp. (60% and 88%), followed by STEC (55% and 51%) and pathogenic *Leptospira* spp. (33% and 13%). The *E.coli* O157: H7 was detected only in sediment samples (11%). The concentrations of genes of bacterial pathogens detected in water and sediment samples at each location are given in Table 5.4. Using the limit of quantification (LOQ) as selection criteria, *Campylobacter jejuni* (mapA) was present in quantifiable range in 20 out of 45 water samples (2.31 to 2.88-$\log_{10}$ gene copies/100ml) and 12 of 45 sediment samples (3.30 to 3.86-$\log_{10}$ gene copies/gram (Fig.5.3). The highest *mapA* gene concentration was observed in water samples collected at Location 16 and in sediment samples collected at location 20 (Table 4). In case of *Shigella* spp. although *ipaH* gene was detected in several water samples, it was quantified only in 2 out of 45 samples (locations 3 and 12) with concentrations of 2.32 and 2.35 $\log_{10}$ gene copies/100ml respectively. In sediments, the *ipaH* gene was quantified in 14 samples with a concentrations ranging 3.32-to 3.47 $\log_{10}$ gene copies/gram (Fig.5.3) and highest concentration was observed at location 12. Similarly, the *stx2 (*Shiga toxin-producing *E.coli)* was quantified only in 2 out of 45 water samples (locations 5 and 2) with a concentration of 2.31 and 2.42 $\log_{10}$ gene copies/100ml respectively. In sediment, it was quantified in 13 out of 45 samples with a concentration range of 3.32 to 3.65-$\log_{10}$ gene copies/gram (Fig.5.3) and highest concentration of *stx2* gene was observed at location-14. The *LipL32* gene (Pathogenic *Leptospira* spp.) was quantified in 15 out of 45 water samples with the concentration ranged from 2.43 to 3.13 $\log_{10}$ gene copies/100ml with highest levels at location-21. *LipL32* gene was

quantified in 6 (out of 45) sediment samples with a concentration of 3.32 to 3.65-log10 gene copies/gram (Fig.5.3) and highest concentration was observed at location 20. The *eae* gene of *Escherichia coli* O157: H7 was quantified only in 3 sediment samples collected in autumn 2014 with a concentration of 3.32-to 4.03 log_{10} gene copies/gram and highest concentration was observed at location 20.

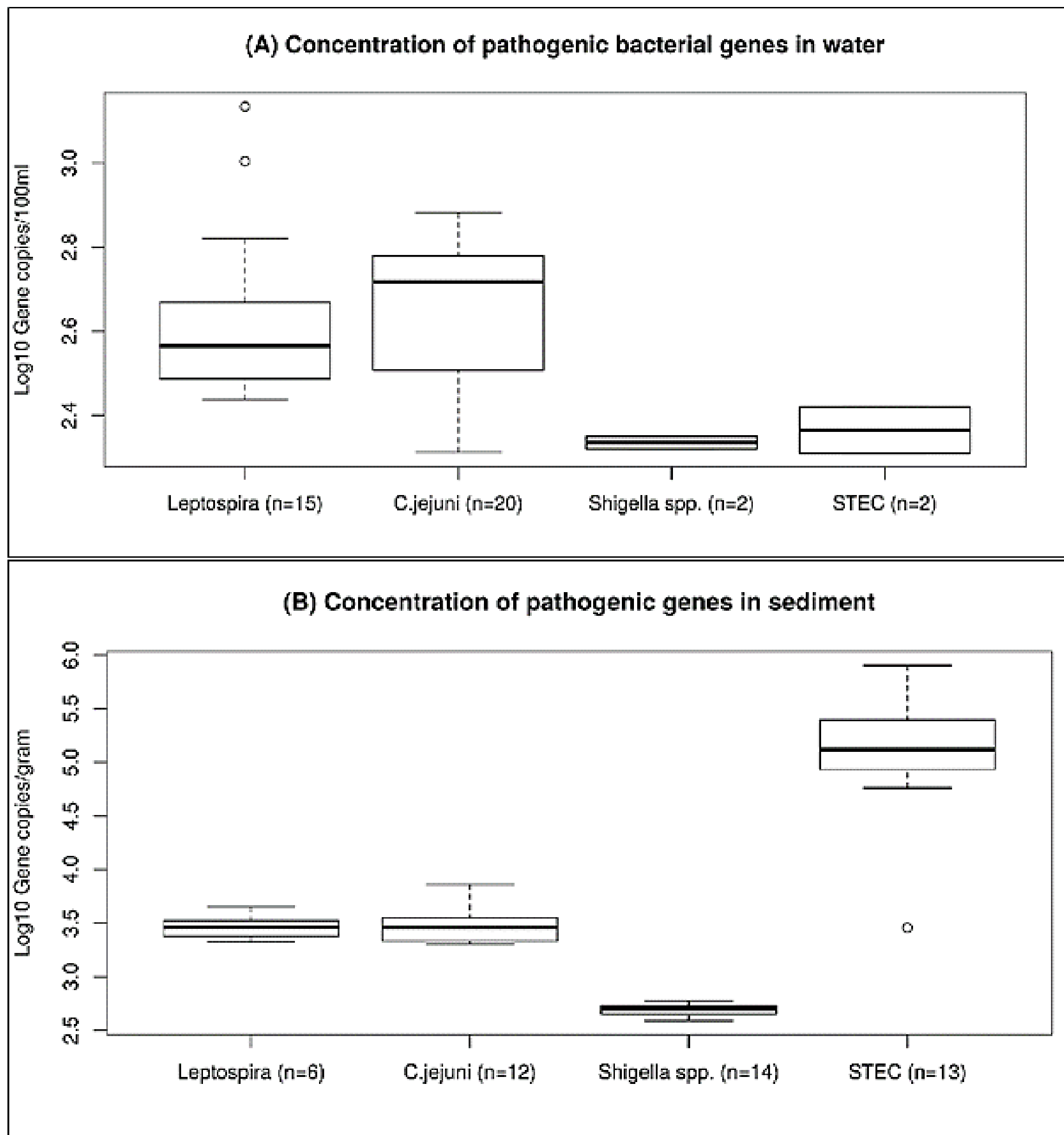


Figure 5.3 Concentrations (mean and standard deviations) of marker genes for bacterial pathogens detected in water (A) and sediment samples (B) of Tiaoxi River.

Table 5.3 Detection frequencies of pathogenic bacterial marker genes in water and sediment samples of Tiaoxi River, Taihu watershed (2014-2015).

Sample type	No. of samples tested (n)	No. of positive samples [a]				
		***Leptospira* (LipL32)**	***Campylobacter* (mapA)**	***Shigella* (ipaH)**	***STEC* (stx2)**	***EHEC O157:H7* (eae)**
Water						
Autumn	15	5 (33%)	4 (26.6)	15 (100%)	0	0
Winter	15	10 (66%)	9 (60%)	8 (53%)	12 (80%)	0
Summer	15	0	15 (100%)	4 (26.6)	12 (80%)	0
Total	**45**	**15 (33%)**	**28 (62%)**	**27 (60%)**	**25 (55%)**	**0**
Sediment						
Autumn	15	0	6 (40%)	15 (100%)	11 (73%)	5 (33%)
Winter	15	6	10 (66%)	11 (73%)	3 (20%)	0
Summer	15	0	8 (53%)	14 (97%)	9 (60%)	0
Total	**45**	**6 (13%)**	**24 (53%)**	**40 (88%)**	**23 (51%)**	**5 (11%)**

*[a] Considering DNQ's (Detected Not Quantifiable) as positive samples.

Table 5.4 Concentration of pathogenic bacterial marker genes in Tiaoxi River water and sediment samples

	Campylobacter jejuni					
	Water (log10 gene copies/100ml)			**Sediment (log10 gene copies/gram)**		
Location	**Autumn 2014**	**Winter 2015**	**Summer 2015**	**Autumn 2014**	**Winter 2015**	**Summer 2015**
L-1	DNQ	N.D	**2.76**	N.D	**3.75**	DNQ
L-2	N.D	N.D	**2.66**	N.D	DNQ	DNQ
L-3	N.D	DNQ	**2.77**	N.D	N.D	**3.31**
L-4	N.D	**2.49**	**2.52**	**3.37**	**3.71**	DNQ
L-5	DNQ	N.D	**2.63**	N.D	**3.52**	**3.31**
L-6	**2.32**	N.D	**2.72**	DNQ	DNQ	DNQ
L-8	N.D	**2.38**	**2.60**	N.D	**3.58**	**3.32**
L-10	N.D	DNQ	**2.76**	DNQ	N.D	N.D
L-12	N.D	**2.38**	**2.81**	N.D	N.D	DNQ
L-13	N.D	**2.31**	**2.72**	N.D	**3.46**	DNQ
L-14	N.D	DNQ	**2.87**	DNQ	N.D	**3.34**
L-15	N.D	N.D	**2.79**	DNQ	**3.46**	N.D
L-16	DNQ	DNQ	**2.88**	N.D	DNQ	N.D
L-20	N.D	DNQ	**2.81**	**3.86**	N.D	N.D
L-21	N.D	N.D	**2.75**	N.D	**3.47**	DNQ

DNQ: Detected not Quantifiable, ND: Not Detected.

	Leptospira spp.					
	Water (log10 gene copies/100ml)			Sediment (log10 gene copies/gram)		
Location	**Autumn 2014**	**Winter 2015**	**Summer 2015**	**Autumn 2014**	**Winter 2015**	**Summer 2015**
L-1	**2.46**	**2.54**	N.D	N.D	N.D	N.D
L-2	**2.59**	**2.67**	N.D	N.D	**3.33**	N.D
L-3	N.D	N.D	N.D	N.D	N.D	N.D
L-4	N.D	**2.52**	N.D	N.D	N.D	N.D
L-5	N.D	**2.61**	N.D	N.D	**3.52**	N.D
L-6	N.D	**3.00**	N.D	N.D	N.D	N.D
L-8	N.D	N.D	N.D	N.D	N.D	N.D
L-10	**2.49**	**2.44**	N.D	N.D	N.D	N.D
L-12	N.D	N.D	N.D	N.D	N.D	N.D
L-13	N.D	**2.44**	N.D	N.D	N.D	N.D
L-14	N.D	N.D	N.D	N.D	**3.47**	N.D
L-15	N.D	N.D	N.D	N.D	**3.38**	N.D
L-16	N.D	**2.54**	N.D	N.D	**3.45**	N.D
L-20	**2.55**	**2.82**	N.D	N.D	**3.65**	N.D
L-21	**2.44**	**3.13**	N.D	N.D	N.D	N.D

Shigella spp.						
	Water (log10 gene copies/100ml)			**Sediment (log10 gene copies/gram)**		
Location	**Autumn 2014**	**Winter 2015**	**Summer 2015**	**Autumn 2014**	**Winter 2015**	**Summer 2015**
L-1	DNQ	N.D	N.D	3.33	DNQ	DNQ
L-2	DNQ	N.D	N.D	3.39	DNQ	DNQ
L-3	**2.33**	N.D	N.D	3.34	DNQ	DNQ
L-4	DNQ	N.D	N.D	3.40	DNQ	DNQ
L-5	N.D	N.D	N.D	DNQ	DNQ	N.D
L-6	DNQ	N.D	N.D	3.33	DNQ	DNQ
L-8	DNQ	N.D	DNQ	3.40	N.D	DNQ
L-10	DNQ	DNQ	N.D	3.36	DNQ	DNQ
L-12	**2.35**	DNQ	N.D	3.48	N.D	DNQ
L-13	DNQ	DNQ	N.D	3.41	DNQ	DNQ
L-14	DNQ	DNQ	N.D	3.42	N.D	DNQ
L-15	DNQ	DNQ	DNQ	3.43	N.D	DNQ
L-16	DNQ	DNQ	N.D	3.44	DNQ	DNQ
L-20	DNQ	DNQ	N.D	3.43	DNQ	DNQ
L-21	DNQ	DNQ	DNQ	3.40	DNQ	DNQ

Shiga toxin-producing E.coli						
	Water (log10 gene copies/100ml)			**Sediment (log10 gene copies/gram)**		
Location	**Autumn 2014**	**Winter 2015**	**Summer 2015**	**Autumn 2014**	**Winter 2015**	**Summer 2015**
L-1	N.D	DNQ	N.D	N.D	DNQ	DNQ
L-2	N.D	DNQ	DNQ	DNQ	N.D	N.D
L-3	N.D	DNQ	DNQ	**3.46**	N.D	N.D
L-4	N.D	DNQ	DNQ	**5.35**	N.D	DNQ
L-5	N.D	**2.31**	DNQ	**5.12**	N.D	N.D
L-6	N.D	N.D	DNQ	**5.32**	DNQ	**3.36**
L-8	N.D	N.D	DNQ	**4.90**	DNQ	DNQ
L-10	N.D	N.D	DNQ	**5.02**	N.D	N.D
L-12	N.D	DNQ	DNQ	**4.97**	N.D	N.D
L-13	N.D	DNQ	N.D	**4.76**	N.D	N.D
L-14	N.D	DNQ	N.D	**5.90**	DNQ	DNQ
L-15	N.D	DNQ	DNQ	**5.65**	DNQ	**3.41**
L-16	N.D	DNQ	DNQ	N.D	DNQ	DNQ
L-20	N.D	DNQ	DNQ	**5.43**	DNQ	**3.33**
L-21	N.D	**2.42**	DNQ	N.D	DNQ	DNQ

	Escherichia coli O157: H7					
	Water (log10 gene copies/100ml)			**Sediment (log10 gene copies/gram)**		
Location	**Autumn 2014**	**Winter 2015**	**Summer 2015**	**Autumn 2014**	**Winter 2015**	**Summer 2015**
L-1	N.D	N.D	N.D	N.D	N.D	N.D
L-2	N.D	N.D	N.D	N.D	N.D	N.D
L-3	N.D	N.D	N.D	N.D	N.D	N.D
L-4	N.D	N.D	N.D	N.D	N.D	N.D
L-5	N.D	N.D	N.D	N.D	N.D	N.D
L-6	N.D	N.D	N.D	N.D	N.D	N.D
L-8	N.D	N.D	N.D	N.D	N.D	N.D
L-10	N.D	N.D	N.D	N.D	N.D	N.D
L-12	N.D	N.D	N.D	DNQ	N.D	N.D
L-13	N.D	N.D	N.D	**3.32**	N.D	N.D
L-14	N.D	N.D	N.D	**3.45**	N.D	N.D
L-15	N.D	N.D	N.D	DNQ	N.D	N.D
L-16	N.D	N.D	N.D	N.D	N.D	N.D
L-20	N.D	N.D	N.D	**4.03**	N.D	N.D
L-21	N.D	N.D	N.D	N.D	N.D	N.D

5.3.5 Correlation between FC, MST markers and genes of bacterial pathogens

All the correlations in this study were considered as significant only when Rho(r) and *p* values were > 0.5 and < 0.05, respectively. The concentrations of FC were positively correlated with BacUni ($r = 0.667$ and $p < 0.01$) and HF183 Taqman ($r = 0.572$ and $p < 0.05$) but not with Pig-2-Bac and GFD markers (Table 5.5), indicating the common source of fecal pollution such as effluents from WWTP or human waste entry into the River. The FC did not show correlation with any of the genes of bacterial pathogens tested in this study, in agreement with the previous studies (Bradshaw et al., 2016, Zhang et al., 2016a); FC points to the potential risk of exposure to pathogens but does not demonstrate their specific presence. The BacUni marker showed strong positive correlation with HF183Taqman ($r = 0.832$ and $p < 0.01$), suggesting human source could be a major contributor to total *Bacteroidales* content. BacUni also showed positive correlation with *stx2* gene ($r = 0.6$ and $p < 0.05$). HF183 Taqman showed negative correlation with GFD marker ($r = -0.582$ and $p < 0.05$) and positive correlation with stx2 gene ($r = 0.593$ and $p < 0.05$). Since there are multiple fecal sources (such as pig, cow, and poultry) of stx2 gene presence in the environment, it is difficult to draw conclusions on stx2 gene correlation with BacUni and HF183 Taqman. No significant correlation was observed for Pig-2-Bac and GFD markers with the marker genes of bacterial pathogens addressed in this study.

5.4 Discussion

Fecal pollution of surface waters is a serious concern to the aquatic ecosystem and human health. Fecal bacteria can enter into the environment through several sources such as effluents from wastewater treatment plants, septic leaks, urban and storm runoff water (Marsalek and Rochfort, 2004, Kapoor et al., 2015). In this study, fecal coliforms present in surface water collected were enumerated from 25 sampling locations to enable identification of locations for further investigation. Fecal coliform counts were higher than suggested limits at the fifteen

Table 5.5 Correlation between FC, MST markers and pathogenic bacterial marker genes.

	Correlation coefficient								
	FC	BacUni	HF183Taq	Pig-2-Bac	GFD	*LipL32*	*mapA*	*ipaH*	*stx2*
FC	1	.667**	.572*	-0.234	-0.27	-0.248	0.095	-0.361	0.397
BacUni		1	.832**	-0.036	-0.396	-0.04	-0.018	-0.168	.600*
HF183Taq			1	-0.061	-.582*	-0.265	0.1	-0.121	.593*
Pig-2-Bac				1	-0.268	0.346	0.5	0.068	-0.421
GFD					1	-0.095	0	0.057	-0.275
LipL32						1	-0.251	-0.225	0.207
mapA							1	0.379	-0.286
ipaH								1	-0.018
stx2									1

**. Correlation is significant at 0.01 level (2-tailed).

*. Correlation is significant at 0.05 level (2-tailed).

locations and increased concentrations were observed during summer. Acclimatization of existing FC bacteria to warm temperatures or entry of fresh feces from different sources such as human, animal or sewage into the surface water due to runoff from rainfall event occurred before sampling could have elevated the levels of FC during summer (Sidhu et al., 2012). Therefore, MST was conducted to assess the presence of fresh fecal pollution in the Tiaoxi River and to determine the fecal sources at these locations.

Overall, the distribution of MST markers among different locations correlated well with the land use pattern and results indicated that there was a mixed input of fecal pollution at several locations. Presence of Total *Bacteroidales* markers with high frequency in water and sediment samples at all locations (Fig. 5.1A & 5.2A) shows possible fresh fecal source entry and transportation to other locations within the study area (Marti et al., 2013). With respect to host-associated markers, human-associated markers were consistently identified (in both water and sediment) in most of the sampling locations during three occasions (Fig. 5.1B & 5.2B). Except for summer season (July/August 2015) where rainfall event (~349/268mm) took place, autumn (Oct 2014) and winter (Jan/Feb 2015) were dry seasons (with only 32 & 66mm precipitation) indicating that fecal contamination at these locations might not be merely through runoff but could be due to direct discharge of sewer and septic waste (Ohad et al., 2015). The detection of *Shigella* spp. that solely originates from human fecal sources at several locations also points out to the human fecal contamination in the studied area. High levels of human-associated markers were frequently observed in water samples at location-3, 5, 12 and 16 and in sediment samples at location- 15 on one or more occasions. The higher levels of human-associated markers observed in water samples at location 5 and 16 and sediment samples at 15 could be associated with effluents from WWTP located near these locations (Zheng et al., 2017). The higher levels of human-associated markers in water samples at Location 3 (rural area) and 12 (urban area) (Table 2.1 of chapter-2) that do not have agricultural or effluent influx from

WWTP, indicates that the major source of human-associated markers at these location could be due to sewage (Kapoor et al., 2015). Swine farming is the dominant livestock-based agricultural activity in Zhejiang province. The swine associated marker was frequently detected in water samples collected at locations 6, 12, 13, 20 and 21 and sediment samples collected in location 13 and 21. The results obtained are consistent with land use pattern as these locations have active pig-farming operations (Fig. 5.1C & 5.2C). Location 6 is near Qijia village (Table 2.1 of chapter-2 and Fig.S5.1) where commercial and household backyard pig and poultry farms were observed during the sampling. On the upstream side of Location 13, WWTP and active pig farming (Table S5.2) were found near to this location (Zheng et al., 2017). Location 20 and 21 were close to changxing port, which has several farms for pigs and poultry (Table. S5.2). The only possible explanation for consistent detection of swine associated markers at location 12 is transport of fecal bacteria from location 13, as the distance between two locations is ~1km. In China, both backyard and commercial based poultry farming are common and existence of such farms nearby leads to release of poultry feces into the watershed (Zhuang et al., 2017). The provincial government of Zhejiang had concerns over the illegal discharge of poultry wastes into the watershed and has recently initiated monitoring control measures recently (Zheijiang, 2016). The avian associated marker quantification results prove that poultry fecal pollution was high in the study area as markers were detected in 89 % of water samples (40/45) (Fig. 5.1D). Based on GFD assay, water samples collected at location 6, 15, 20 and 21 (Fig. 5.1D) and sediments (summer season) collected at locations 6 and 20 (Table S5.3) had high levels of avian markers. Active commercial and backyard poultry farming near locations 6, 20 and 21 was observed (Table S5.2 and Fig.S5.1) and results correlate well with the land use pattern. Overall, MST results indicate potential occurrence of pathogens at these locations, which was followed up with in-depth monitoring of host-associated pathogens.

Although Quantitative Microbial Risk Assessment (QMRA) tools can identify the probability of infection or illness more appropriately by utilizing pathogen concentration data, hypothetical water consumption or contact volume and hydrological data of watershed (Schmidt and Emelko, 2011), it should be noted that it is impossible to determine the lowest infection dose of pathogen for an individual or population since immunity of a person plays an import role (Fullerton et al., 2012). However, a comparison of cell count or gene copies concentration data of bacterial pathogens for water or environmental samples with lowest infection dose could provide initial assessment of human health risk (Ahmed et al., 2009b). Currently, *Campylobacter jejuni* was found to be most frequently detected pathogen in the study area (Table 5.3). *C. jejuni* originates primarily from chicken and other avian feces (Lund et al., 2004). High frequency of *C.jejuni* detected could be due to high avian fecal contamination, although this organism can survive up to 4 months in the environment (Murphy et al., 2006). As *mapA* gene marker is a single copy gene for *C. jejuni,* the gene copy number can be converted to cell counts to assess human health risk (Ahmed et al., 2009b). Although *C. jejuni* was frequently detected in water and sediment samples at location-4, 6, 8 and 13, the concentrations exceeded the lowest infectious dose of 500 organisms (Ahmed et al., 2009b) only at location 6 and 13 in one or more occasions indicating these locations could pose significant health risks. Globally, *Shigella* spp. is one among the major bacterial cause of diarrhea. It is considered as one of the top four pathogens that causes moderate to severe diarrhea to the children of Africa and South Asia (Huynh et al., 2015). *Shigella sonnei* is the most commonly recovered species in infected patients from United States, while in Asia and developing countries, *S. dysenteriae* and *S. flexneri* were the major causative species (Wu et al., 2009). Humans are considered as the common and natural reservoirs for *Shigella* (Ishii et al., 2013) and presence of these bacteria indicates human fecal contamination (Oster et al., 2014). The lowest infectious dose of *Shigella* spp. to cause diarrhea is 10-200 organisms,

though it differs from person to person and species to species of this genus (Aragon et al., 2007). For instance, *S. flexneri* requires <140 organisms or cells while *S. dysenteriae* it is <100 cells and for some virulent strains it is <10 cells (Aragon et al., 2007). In the present study, *ipaH* gene was detected in 68 out of 90 water and sediment samples tested (Table 5.3). As *ipaH* gene occurs in 5 to 10 copies in plasmid and genomic DNA of *Shigella* spp. (Greenberg et al., 2010), the *Shigella* spp. detected in water and sediment samples at location-3 and 12 could pose a human health risk. The monitoring of *Shigella* and *Campylobacter* is a useful tool for watershed managers/monitoring agencies as these organisms are associated with specific sources (host associated). Shiga toxin-producing *E. coli* (STEC) can cause gastrointestinal disease leading to mild or severe bloody diarrhea (Haack et al., 2015). The infectious dose for some STEC strains to cause infection is <100 organisms (Thorpe, 2004). Here, although STEC was detected in nearly 50% of water samples tested in the study area (Table 3), only samples collected from locations 5 and 21 could pose significant health risk. STEC may originate from multiple reservoirs or sources (Beutin et al., 2008). Pathogenic *Leptospira* species cause leptospirosis by colonization of the renal tubules of infected reservoir hosts such as dogs, rats, and cattle and enters into the environment via urine (Monahan et al., 2008). Pathogenic *Leptospira* spp. detection frequency was comparatively lower than STEC, *Campylobacter* and *Shigella* spp. Pathogenic *Leptospira* spp. requires high bacterial count to cause infection due to the acid sensitivity of the bacteria. Although the exact low infection dose of pathogenic *Leptospira* spp. for human is unknown, it is considered that >500 cells are required for infection (Ganoza et al., 2010). Therefore, the pathogenic *Leptospira* quantified in the present study were not considered as significant human health risk. *E. coli* O157: H7 (*eae* gene) was not detected in any of the water samples, but detected at low frequency (33%) in sediment samples (autumn season) (Fig. 5). It has been reported that cattle are the main reservoir for *E. coli O157* (Ahmed et al., 2009b) and the results presented here are consistent with the paucity of cattle farming in

the watershed area (NBSC, 2015). Overall, the quantification of the genes of bacterial pathogens in water and sediment samples indicate that *C. jejuni* and STEC are the major concerns for human health risk in a few locations of the study area and direct contact with water or sediment samples at these locations could cause illness to the human population.

5.5 Conclusions

- The human-associated marker was frequently detected in high concentrations at locations 3, 5, 12 and 16 indicating high human fecal pollution in those locations.
- Swine associated marker was frequently detected in samples from locations 13 and 21 and the avian associated marker was detected in high concentrations at locations 6, 15, 20 and 21, matching with the land use pattern and pointing to the entry of pig and avian fecal sources into Tiaoxi River.
- *Campylobacter jejuni* was detected in exceeding levels of lowest infection dose at location 6 and 13. Similarly, *Shigella* spp. were detected exceeding levels of lowest infection dose at location 3 and 12, and Shiga toxin producing *E.coli* (STEC) at location 5 and 21.
- The bacterial pathogen quantification results correlate with the findings of host associated fecal markers and demonstrating the potential of MST in predicting the presence of host associated pathogenic organisms.
- Overall, the results of MST and bacterial pathogen quantification indicate severe fecal contamination in the Tiaoxi River.

CHAPTER-6

Summary and Future Directions

6.1 Summary

The overall goal of this PhD thesis was to apply LI-MST techniques to characterize fecal pollution and to identify the abundance of specific bacterial pathogens in the Tiaoxi River (Taihu watershed) to determine human health risk. This thesis comprised of: analysis of the physico-chemical and microbiological parameters of Tiaoxi River water to assess the preliminary hotspots of fecal contamination; determination of the microbial community in water samples collected from these preliminary hotspots by Illumina sequencing (NGS) to characterize the diversity of bacteria and to demonstrate the presence of fecal and potentially pathogenic bacteria; validation of the performance of existing universal, human and animal associated MST qPCR assays for their applicability to detect host-associated fecal pollution at Taihu watershed; and finally, investigation of the presence and abundance of MST markers and genes of bacterial pathogens at Tiaoxi River.

In Chapter 2, a detailed physico-chemical and microbiological analysis was carried out to assess the quality of Tiaoxi River water. Fourteen water quality parameters including multiple nutrients and indicator bacteria were analysed in the water samples collected from 25 locations across the Tiaoxi River in three seasons (autumn 2014, winter and summer 2015), and the data were analyzed by multivariate statistical analyses. The results indicated that TN, TP, NO_2-N, and NH_4-N were the major nutrients that contributed to pollution in this River. The concentration of multiple nutrients and bacterial counts varied with the type of land-use and the locations where non-point sources possibly entered the Tiaoxi River. The entry of wastes into this River from non-point sources was observed in several locations during the sampling events and these wastes negatively affect the Tiaoxi River and water quality. The elevated levels of nutrients contribute to eutrophication and this is evident from the presence of high concentrations of Chl *a,* particularly in location 1 (Taihu Lake) due to algal blooms in the Lake. Chl *a* levels were comparatively high in the samples collected during the summer season, as

expected. The total coliform and fecal coliform counts were higher in urban/semi-urban locations where wastes from non-point sources entered into the River. The microbiological results also indicated that fecal pollution was high in 15 locations, suggesting the entry of human or animal fecal matter. Cluster analysis showed that the sampling locations can be grouped into three clusters based on the land use patterns and physico-chemical and microbiological data, and these three clusters can be classified as relatively low, moderate and high pollution areas. The PCA of the entire dataset reduced the parameters to four principal components that explain 83% of the total variance. The four components that contribute significant variation in water can be classified as natural, nutrient, microbial and organic type of pollutants. The overall physico-chemical and microbiological results indicate the possibility of fecal source entry into the River through various sources, and a microbial source tracking study is needed to identify the fecal sources.

In Chapter 3, a comprehensive study to identify the diversity and composition of bacterial communities in water and fecal samples was conducted using the Illumina MiseqPE250 platform. Forty five water samples (from 15 locations in three seasons) were analyzed with duplicates of 7 fecal hosts and sewage samples as reference. About 20 different phyla were observed in Tiaoxi River water, fecal and wastewater samples, though their relative abundance varied with the type of samples. Proteobacteria, Actinobacteria, Bacteroidetes, and Cyanobacteria were the dominant phyla accounting for 90% of bacterial abundance in most of the water samples. The phyla Bacteroidetes, Firmicutes, and Proteobacteria were abundant in fecal samples, whereas wastewater samples had Proteobacteria, Bacteroidetes, and Chloroflexi as the dominant phyla. Hierarchical cluster analysis and PCoA showed that fecal and wastewater samples clustered separately from water samples. Comparison of OTU data from individual fecal sources (Venn diagram analysis) with total water samples specified 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. When sequence data were further analyzed at the genus level, the results indicated that five fecal related genera (*Bacteroides*, *Prevotella*, *Blautia*, *Faecalibacterium and Dorea)* were present at several locations, with *Bacteroides* detected in most of the samples tested (39/45) indicating the presence of fecal contamination. The presence of *Dorea* (14 out of 45 samples) and *Bacteroides* species (*B. caccae* and *B. plebeius*) points to potential human fecal contamination at these locations. Examination of the sequence data of water samples also revealed the presence of seven potential pathogenic bacterial genera (*Acinetobacter, Aeromonas, Arcobacter, Brevundimonas, Enterococcus, Escherichia-Shigella,* and *Streptococcus*) with high abundance (>0.1); only sequences of two genera could be classified to species level with high sequence similarity (97%), *A. cryaerophilus* (formerly *Campylobacter cryaerophilus*) and *Brevundimonas vesicularis.* However, as *Arcobacter* and *Campylobacter* species are genotypically similar, more specific PCR assays are needed for differentiation and identification. Overall, Illumina sequencing results indicated the presence of fecal contamination and potentially pathogenic bacteria at Tiaoxi River, demanding a MST study to determine the major fecal source, and application of more specific qPCR assays to detect the presence of pathogens.

In Chapter 4, validation of two universal/general *Bacteroidales* (BacUni and GenBac3), four human-associated (HF183 SYBR, HF183 Taqman, BacHum and Hum2), one swine associated (Pig-2-Bac), one livestock/domestic animal associated (BacCow) and two avian associated MST qPCR assays (GFD and AV4143) using sewage, human and animal fecal samples was carried out to determine the suitable MST assays for identifying fecal pollution sources at Taihu watershed. Results indicated that BacUni and GenBac3 markers (universal/general *Bacteroidales* assay) were highly sensitive (100%) to fecal and sewage samples, with the BacUni marker yielding comparatively higher abundance data. All of the four human-

associated assays showed cross reactivity with chicken and dog-fecal samples; HF183 Taqman and BacHum were the more specific of these. However, HF183 Taqman is a more suitable assay at this geographical region as it did not show any cross reactivity with pig and cow fecal samples. The swine associated assay, Pig-2-Bac, was highly specific and sensitive demonstrating its potential in tracking swine fecal pollution. BacCow marker (Livestock/domestic animal associated assay) was not detected in all animal samples tested indicating that it cannot be applicable for detecting total livestock fecal contamination at this geographical location. Among the avian associated assays, although the GFD marker was relatively less sensitive compared to AV4143 it was highly specific and did not cross-react with human fecal samples, indicating its ability to discriminate avian and human fecal sources. Therefore, BacUni, HF183 Taqman, Pig-2-Bac and GFD markers are recommended as a suite of marker assays for tracking total and host-associated fecal contamination at Taihu watershed region. The data also reinforce the importance of conducting regional MST validation studies prior to their application in any new geographical region.

In Chapter 5, the above MST markers were quantified in water and sediment samples collected from 15 locations in Tiaoxi River, along with five pathogenic bacterial genes to assess the human health risks. The total *Bacteroidales* marker was detected in all the water and sediment samples of 15 monitoring locations, suggesting the presence of fecal contamination. Although human-associated markers were detected frequently at several locations, location-3, 12 and 16 had high concentrations on one or more occasions indicating that they are major human fecal contaminated sites of the Tiaoxi River region. Swine associated marker was detected in all water and sediment samples from location 13 and 21, correlating with the land use pattern and pointing to the entry of pig fecal source into Tiaoxi River. The avian associated marker was detected at high concentrations in all water samples at locations- 6, 15, 20 and 21, which also correlates well with the land use pattern at these locations. Among bacterial pathogens

monitored, *Campylobacter jejuni* was frequently detected in water and sediment samples at locations 4, 6, 8 and 13, with concentrations exceeding the lower limit of infection dose at location 6 and 13 (on one or more occasions) indicating that these waters are unsafe for human contact or consumption. *Shigella* spp. detected in water and sediment samples at locations 3 and 12 and STEC detected at locations 5 and 21 water and sediment samples also show that these sites are unsafe for human contact or consumption.

In summary, this thesis has demonstrated the geographical and regional applicability of a group of MST qPCR assays and pointed out the presence of multiple fecal sources in water and sediment of Tiaoxi River. The MST data generated here could be of significant importance for water quality managers/engineers of Tiaoxi River watershed to take appropriate actions. Furthermore, the pathogen quantification data are useful for the Tiaoxi River water quality managers to assess additional human health threats using microbial risk assessment tools in combination with their access to complete information on the Tiaoxi River watershed.

6.2 Future directions

The findings in this thesis have opened up several potential opportunities for future research. The sequence data from fecal sources in this study revealed that the genus *Faecalibacterium* is abundantly present in human samples, as opposed to *Bacteroides* that are abundant in both humans and animals, suggesting that *Faecalibacterium* would be a useful target for the design of human-specific MST markers. A recent study (Sun et al., 2016) also provided initial evidence for this and supported the use of *Faecalibacterium* as a human-specific MST marker. Therefore, development of a Taqman based MST qPCR assay could provide a more specific human-associated MST assay.

The microbial community analysis study revealed the presence of several pathogenic genera such as *Acinetobacter*, *Aeromonas*, *Arcobacter* and *Enterococcus* with high concentrations at several locations of Tiaoxi River. Therefore, further research is required to identify their

pathogenicity by applying more specific PCR assays and assess human health risk associated with these bacteria.

Recent developments in metagenomic studies such as NGS-based MST methods coupled with SourceTracker software have gained much attention due to the ability to characterize rare members of the fecal biome that can act as alternative indicators of fecal contamination in environmental samples, though this method requires optimization and validation before applying to environmental samples (Unno et al., 2018, Staley et al., 2018). The fecal libraries constructed in this PhD project could be valuable for future studies that use NGS-based MST methods using SourceTracker software at Taihu watershed, as this method requires site-specific larger libraries for accurate identification of fecal sources. If sufficient funds are available, application of SourceTracker (NGS) based MST methods along with qPCR based MST as a toolbox may provide more data that are robust.

References

Abdul-Wahab, S. A., Bakheit, C. S. & Al-Alawi, S. M. 2005. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environ Model Softw,* 20**,** 1263-1271.

Ackerman, D. & Weisberg, S. B. 2003. Relationship between rainfall and beach bacterial concentrations on Santa Monica bay beaches. *J Water Health,* 1**,** 85-9.

Ahmed, W., Goonetilleke, A., Powell, D. & Gardner, T. 2009a. Evaluation of multiple sewage-associated Bacteroides PCR markers for sewage pollution tracking. *Water Res,* 43**,** 4872-7.

Ahmed, W., Hamilton, K. A., Gyawali, P., Toze, S. & Haas, C. N. 2016a. Evidence of Avian and Possum Fecal Contamination in Rainwater Tanks as Determined by Microbial Source Tracking Approaches. *Appl Environ Microbiol,* 82**,** 4379-86.

Ahmed, W., Harwood, V. J., Nguyen, K., Young, S., Hamilton, K. & Toze, S. 2016b. Utility of Helicobacter spp. associated GFD markers for detecting avian fecal pollution in natural waters of two continents. *Water Research,* 88**,** 613-622.

Ahmed, W. & Katouli, M. 2008. Phenotypic variations of enterococci in surface waters: analysis of biochemical fingerprinting data from multi-catchments. *J Appl Microbiol,* 105**,** 452-8.

Ahmed, W., Masters, N. & Toze, S. 2012. Consistency in the host specificity and host sensitivity of the Bacteroides HF183 marker for sewage pollution tracking. *Letters in Applied Microbiology,* 55**,** 283-289.

Ahmed, W., Sawant, S., Huygens, F., Goonetilleke, A. & Gardner, T. 2009b. Prevalence and occurrence of zoonotic bacterial pathogens in surface waters determined by quantitative PCR. *Water Res,* 43**,** 4918-28.

Ahmed, W., Sritharan, T., Palmer, A., Sidhu, J. P. & Toze, S. 2013. Evaluation of bovine feces-associated microbial source tracking markers and their correlations with fecal indicators and zoonotic pathogens in a Brisbane, Australia, reservoir. *Appl Environ Microbiol,* 79**,** 2682-91.

Ahmed, W., Staley, C., Sadowsky, M. J., Gyawali, P., Sidhu, J. P., Palmer, A., Beale, D. J. & Toze, S. 2015a. Toolbox Approaches Using Molecular Markers and 16S rRNA Gene Amplicon Data Sets for Identification of Fecal Pollution in Surface Water. *Appl Environ Microbiol,* 81**,** 7067-77.

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

Ahmed, W., Stewart, J., Gardner, T., Powell, D., Brooks, P., Sullivan, D. & Tindale, N. 2007. Sourcing faecal pollution: a combination of library-dependent and library-independent methods to identify human faecal pollution in non-sewered catchments. *Water Res,* 41**,** 3771-9.

Ahmed, W., Yusuf, R., Hasan, I., Goonetilleke, A. & Gardner, T. 2010. Quantitative PCR assay of sewage-associated Bacteroides markers to assess sewage pollution in an urban lake in Dhaka, Bangladesh. *Can J Microbiol,* 56**,** 838-45.

Alberto, W. D., Pilar, D. M. D., Valeria, A., Fabiana, P. S., Cecilia, H. A. & Angeles, B. M. D. L. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. a case study: Suquia river basin (Cordoba–Argentina). *Wat Res,* 35**,** 2881–2894.

Andrews, S. 2010. *A quality control tool for high throughput sequence data* [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed 11th September 2018].

Apha 2005. Standard methods for the examination of water and wastewater, 21st ed. American Public Health Association, Washington, DC.

Aragon, T. J., Vugia, D. J., Shallow, S., Samuel, M. C., Reingold, A., Angulo, F. J. & Bradford, W. Z. 2007. Case-control study of shigellosis in San Francisco: the role of sexual transmission and HIV infection. *Clin Infect Dis,* 44**,** 327-34.

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.

Ashbolt, N. J. 2004. Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology,* 198**,** 229-38.

Ashbolt, N. J. 2015. Microbial Contamination of Drinking Water and Human Health from Community Water Systems. *Curr Environ Health Rep,* 2**,** 95-106.

Bae, S. & Wuertz, S. 2009. Rapid decay of host-specific fecal Bacteroidales cells in seawater as measured by quantitative PCR with propidium monoazide. *Water Res,* 43**,** 4850-9.

Bae, S. & Wuertz, S. 2012. Survival of host-associated bacteroidales cells and their relationship with Enterococcus spp., Campylobacter jejuni, Salmonella enterica serovar Typhimurium, and adenovirus in freshwater microcosms as measured by propidium monoazide-quantitative PCR. *Appl Environ Microbiol,* 78**,** 922-32.

Balleste, E., Bonjoch, X., Belanche, L. A. & Blanch, A. R. 2010. Molecular indicators used in the development of predictive models for microbial source tracking. *Appl Environ Microbiol,* 76**,** 1789-95.

Baylis, C. L. & Patrick, M. 1999. Comparison of a range of chromogenic media for enumeration of total coliforms and E. coli in foods. *Leatherhead International Technical Notes.*

Beachact 2000. Beaches Environmental Assessment and Coastal Health Act. Washington, DC.: 106th U.S. Congress.

Bernhard, A. E. & Field, K. G. 2000a. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Appl Environ Microbiol,* 66**,** 1587-94.

Bernhard, A. E. & Field, K. G. 2000b. 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**,** 4571-4.

Best, E. L., Powell, E. J., Swift, C., Grant, K. A. & Frost, J. A. 2003. Applicability of a rapid duplex real-time PCR assay for speciation ofCampylobacter jejuniandCampylobacter colidirectly from culture plates. *FEMS Microbiology Letters,* 229**,** 237-241.

Beutin, L., Kruger, U., Krause, G., Miko, A., Martin, A. & Strauch, E. 2008. Evaluation of major types of Shiga toxin 2E-producing Escherichia coli bacteria present in food, pigs, and the environment as potential pathogens for humans. *Appl Environ Microbiol,* 74**,** 4806-16.

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**,** 6812-28.

Boehm, A. B., Wang, D., Ercumen, A., Shea, M., Harris, A. R., Shanks, O. C., Kelty, C., Ahmed, A., Mahmud, Z. H., Arnold, B. F., Chase, C., Kullmann, C., Colford, J. M., Luby, S. P. & Pickering, A. J. 2016. Occurrence of Host-Associated Fecal Markers on Child Hands, Household Soil, and Drinking Water in Rural Bangladeshi Households. *Environmental Science & Technology Letters,* 3**,** 393-398.

Bonjoch, X., Balleste, E. & Blanch, A. R. 2004. Multiplex PCR with 16S rRNA gene-targeted primers of bifidobacterium spp. to identify sources of fecal pollution. *Appl Environ Microbiol,* 70**,** 3171-5.

Bradshaw, J. K., Snyder, B. J., Oladeinde, A., Spidle, D., Berrang, M. E., Meinersmann, R. J., Oakley, B., Sidle, R. C., Sullivan, K. & Molina, M. 2016. Characterizing relationships among fecal

indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. *Water Res,* 101**,** 498-509.

Briggs, D. 2003. Environmental pollution and the global burden of disease. *Br Med Bull,* 68**,** 1-24.

Brookes, J. D., Antenucci, J., Hipsey, M., Burch, M. D., Ashbolt, N. J. & Ferguson, C. 2004. Fate and transport of pathogens in lakes and reservoirs. *Environ Int,* 30**,** 741-59.

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.

Bu, H., Meng, W. & Zhang, Y. 2014. Spatial and seasonal characteristics of river water chemistry in the Taizi River in Northeast China. *Environ Monit Assess,* 186**,** 3619-32.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J. & Wittwer, C. T. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem,* 55**,** 611-22.

Byappanahalli, M. N., Shively, D. A., Nevers, M. B., Sadowsky, M. J. & Whitman, R. L. 2003. Growth and survival of Escherichia coli and enterococci populations in the macro-alga Cladophora (Chlorophyta). *FEMS Microbiol Ecol,* 46**,** 203-11.

Bylund, J., Toljander, J., Lysen, M., Rasti, N., Engqvist, J. & Simonsson, M. 2017. Measuring sporadic gastrointestinal illness associated with drinking water - an overview of methodologies. *J Water Health,* 15**,** 321-340.

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. *Microb Ecol,* 66**,** 73-83.

Cao, Y., Van De Werfhorst, L. C., Sercu, B., Murray, J. L. & Holden, P. A. 2011. Application of an integrated community analysis approach for microbial source tracking in a coastal creek. *Environ Sci Technol,* 45**,** 7195-201.

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**,** 335-6.

Carroll, S. P., Dawes, L., Hargreaves, M. & Goonetilleke, A. 2006. Water quality profile of an urbanising catchment – Ningi Creek Catchment. *Technical Report, Caboolture Shire Council.* Caboolture Shire Council.: School of Urban Development, Queensland University of Technology.

Carson, C. A., Shear, B. L., Ellersieck, M. R. & Asfaw, A. 2001. Identification of fecal Escherichia coli from humans and animals by ribotyping. *Appl Environ Microbiol,* 67**,** 1503-7.

Chapman, D. 1996. *Water quality assessments - A guide to use of biota, sediments and water in Environmental monitoring.*, University Press, Cambridge.

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.

Converse, R. R., Blackwood, A. D., Kirs, M., Griffith, J. F. & Noble, R. T. 2009. Rapid QPCR-based assay for fecal Bacteroides spp. as a tool for assessing fecal contamination in recreational waters. *Water Res,* 43**,** 4828-37.

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**,** 3184-91.

Dick, L. K. & Field, K. G. 2004. Rapid estimation of numbers of fecal Bacteroidetes by use of a quantitative PCR assay for 16S rRNA genes. *Appl Environ Microbiol,* 70**,** 5695-7.

Dorai-Raj, S., J, O. G. & Colleran, E. 2009. Specificity and sensitivity evaluation of novel and existing Bacteroidales and Bifidobacteria-specific PCR assays on feces and sewage samples and their application for microbial source tracking in Ireland. *Water Res,* 43**,** 4980-8.

Dowd, S. E., Callaway, T. R., Wolcott, R. D., Sun, Y., Mckeehan, T., Hagevoort, R. G. & Edrington, T. S. 2008a. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol,* 8**,** 125.

Dowd, S. E., Sun, Y., Wolcott, R. D., Domingo, A. & Carroll, J. A. 2008b. Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) for microbiome studies: bacterial diversity in the ileum of newly weaned Salmonella-infected pigs. *Foodborne Pathog Dis,* 5**,** 459-72.

Du, C., Li, Y., Wang, Q., Liu, G., Zheng, Z., Mu, M. & Li, Y. 2017. Tempo-spatial dynamics of water quality and its response to river flow in estuary of Taihu Lake based on GOCI imagery. *Environ Sci Pollut Res Int*.

Ervin, J. S., Russell, T. L., Layton, B. A., Yamahara, K. M., Wang, D., Sassoubre, L. M., Cao, Y. P., Kelty, C. A., Sivaganesan, M., Boehm, A. B., Holden, P. A., Weisberg, S. B. & Shanks, O. C. 2013. Characterization of fecal concentrations in human and other animal sources by physical, culture-based, and quantitative real-time PCR methods. *Water Research,* 47**,** 6873-6882.

Fan, L., Shuai, J., Zeng, R., Mo, H., Wang, S., Zhang, X. & He, Y. 2017. Validation and application of quantitative PCR assays using host-specific Bacteroidales genetic markers for swine fecal pollution tracking. *Environ Pollut,* 231**,** 1569-1577.

Farhadinejad, T., Khakzad, A., Jafari, M., Shoaee, Z., Khosrotehrani, K., Nobari, R. & Shahrokhi, V. 2012. The study of environmental effects of chemical fertilizers and domestic sewage on water quality of Taft region, Central Iran. *Arab J Geosci,* 7**,** 221-229.

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**,** 1335-7.

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**,** 145-7.

Fewtrell, L. 2004. Drinking-water nitrate, methemoglobinemia, and global burden of disease: A discussion. *Environ Health Perspect,* 112**,** 1371-1374.

Field, K. G. & Samadpour, M. 2007. Fecal source tracking, the indicator paradigm, and managing water quality. *Water Res,* 41**,** 3517-38.

Figueras, M. J. & Borrego, J. J. 2010. New perspectives in monitoring drinking water microbial quality. *Int J Environ Res Public Health,* 7**,** 4179-202.

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**,** 31-7.

Fremaux, B., Boa, T. & Yost, C. K. 2010. Quantitative real-time PCR assays for sensitive detection of Canada goose-specific fecal pollution in water sources. *Appl Environ Microbiol,* 76**,** 4886-9.

Fremaux, B., Gritzfeld, J., Boa, T. & Yost, C. K. 2009. Evaluation of host-specific Bacteroidales 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. *Water Res,* 43**,** 4838-49.

Fullerton, K. E., Scallan, E., Kirk, M. D., Mahon, B. E., Angulo, F. J., De Valk, H., Van Pelt, W., Gauci, C., Hauri, A. M., Majowicz, S., O'brien, S. J. & International Collaboration Working, G. 2012. Case-control studies of sporadic enteric infections: a review and discussion of studies conducted internationally from 1990 to 2009. *Foodborne Pathog Dis,* 9**,** 281-92.

Ganoza, C. A., Matthias, M. A., Saito, M., Cespedes, M., Gotuzzo, E. & Vinetz, J. M. 2010. Asymptomatic renal colonization of humans in the peruvian Amazon by Leptospira. *PLoS Negl Trop Dis,* 4**,** e612.

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**,** 340-7.

Gatica, E. A., Almeida, C. A., Mallea, M. A., Del Corigliano, M. C. & Gonzalez, P. 2012. Water quality assessment, by statistical analysis, on rural and urban areas of Chocancharava River (Río Cuarto), Córdoba, Argentina. *Environ Monit Assess,* 184**,** 7257-74.

Gihring, T. M., Green, S. J. & Schadt, C. W. 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ Microbiol,* 14**,** 285-90.

Gourmelon, M., Caprais, M. P., Le Mennec, C., Mieszkin, S., Ponthoreau, C. & Gendronneau, M. 2010. Application of library-independent microbial source tracking methods for identifying the sources of faecal contamination in coastal areas. *Water Sci Technol,* 61**,** 1401-9.

Gourmelon, M., Caprais, M. P., Segura, R., Le Mennec, C., Lozach, S., Piriou, J. Y. & Rince, A. 2007. Evaluation of two library-independent microbial source tracking methods to identify sources of fecal contamination in French estuaries. *Appl Environ Microbiol,* 73**,** 4857-66.

Green, H. C., Dick, L. K., Gilpin, B., Samadpour, M. & Field, K. G. 2012. Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. *Appl Environ Microbiol,* 78**,** 503-10.

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.

Greenberg, J., Price, B. & Ware, A. 2010. Alternative estimate of source distribution in microbial source tracking using posterior probabilities. *Water Res,* 44**,** 2629-37.

Griffith, J. F., Cao, Y., Mcgee, C. D. & Weisberg, S. B. 2009. Evaluation of rapid methods and novel indicators for assessing microbiological beach water quality. *Water Res,* 43**,** 4900-7.

Griffith, J. F., Weisberg, S. B. & McGee, C. D. 2003. Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. *J Water Health,* 1**,** 141-51.

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.

Haack, S. K., Duris, J. W., Kolpin, D. W., Fogarty, L. R., Johnson, H. E., Gibson, K. E., Focazio, M., Schwab, K. J., Hubbard, L. E. & Foreman, W. T. 2015. Genes Indicative of Zoonotic and Swine Pathogens Are Persistent in Stream Water and Sediment following a Swine Manure Spill. *Appl Environ Microbiol,* 81**,** 3430-41.

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.

Hagedorn, C., Robinson, S. L., Filtz, J. R., Grubbs, S. M., Angier, T. A. & Reneau, R. B., Jr. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Appl Environ Microbiol,* 65**,** 5522-31.

Haller, L., Amedegnato, E., Poté, J. & Wildi, W. 2009. Influence of freshwater sediment characteristics on persistence of fecal indicator bacteria. *Water Air Soil Pollut,* 203**,** 217-227.

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

Harris, A. R., Pickering, A. J., Harris, M., Doza, S., Islam, M. S., Unicomb, L., Luby, S., Davis, J. & Boehm, A. B. 2016. Ruminants Contribute Fecal Contamination to the Urban Household Environment in Dhaka, Bangladesh. *Environmental Science & Technology,* 50**,** 4642-4649.

Harwood, V. J., Boehm, A. B., Sassoubre, L. M., Vijayavel, K., Stewart, J. R., Fong, T. T., Caprais, M. P., Converse, R. R., Diston, D., Ebdon, J., Fuhrman, J. A., Gourmelon, M., Gentry-Shields, J., Griffith, J. F., Kashian, D. R., Noble, R. T., Taylor, H. & Wicki, M. 2013. Performance of viruses and bacteriophages for fecal source determination in a multi-laboratory, comparative study. *Water Res,* 47**,** 6929-43.

Harwood, V. J., Staley, C., Badgley, B. D., Borges, K. & Korajkic, A. 2014. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiol Rev,* 38**,** 1-40.

Harwood, V. J., Whitlock, J. & Withington, V. 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. *Appl Environ Microbiol,* 66**,** 3698-704.

Harwood, V. J., Wiggins, B., Hagedorn, C., Ellender, R. D., Gooch, J., Kern, J., Samadpour, M., Chapman, A. C., Robinson, B. J. & Thompson, B. C. 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. *J Water Health,* 1**,** 153-66.

Haugland, R. A., Varma, M., Sivaganesan, M., Kelty, C., Peed, L. & Shanks, O. C. 2010. Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. *Syst Appl Microbiol,* 33**,** 348-57.

Havelaar, A. H., Pot-Hogeboom, W. M., Furuse, K., Pot, R. & Hormann, M. P. 1990. F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. *J Appl Bacteriol,* 69**,** 30-7.

He, X., Liu, P., Zheng, G., Chen, H., Shi, W., Cui, Y., Ren, H. & Zhang, X. X. 2016a. Evaluation of five microbial and four mitochondrial DNA markers for tracking human and pig fecal pollution in freshwater. *Sci Rep,* 6**,** 35311.

He, X. W., Liu, P., Zheng, G. L., Chen, H. M., Shi, W., Cui, Y. B., Ren, H. Q. & Zhang, X. X. 2016b. Evaluation of five microbial and four mitochondrial DNA markers for tracking human and pig fecal pollution in freshwater. *Scientific Reports,* 6.

Heaney, C. D., Myers, K., Wing, S., Hall, D., Baron, D. & Stewart, J. R. 2015. Source tracking swine fecal waste in surface water proximal to swine concentrated animal feeding operations. *Science of the Total Environment,* 511**,** 676-683.

Helena, B., Pardom, R., Vega, M., Barrado, E., Fernandez, J. M. & Fernandez, L. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Wat. Res.,* 34**,** 807-816.

Henry, R., Schang, C., Coutts, S., Kolotelo, P., Prosser, T., Crosbie, N., Grant, T., Cottam, D., O'brien, P., Deletic, A. & Mccarthy, D. 2016. Into the deep: Evaluation of SourceTracker for assessment of faecal contamination of coastal waters. *Water Research,* 93**,** 242-253.

Hill, M. K. 2010. *Understanding Environmental Pollution.,* Cambridge, Cambridge University Press.

Huynh, A., Mcgrath, C., Johnson, D. & Burrell, L. M. 2015. Shigella sonnei bacteraemia occurring in a young man with shigellosis. *BMJ Case Rep,* 2015.

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.

Ibekwe, A. M., Watt, P. M., Grieve, C. M., Sharma, V. K. & Lyons, S. R. 2002. Multiplex Fluorogenic Real-Time PCR for Detection and Quantification of Escherichia coli O157:H7 in Dairy Wastewater Wetlands. *Applied and Environmental Microbiology,* 68**,** 4853-4862.

Ishii, S., Segawa, T. & Okabe, S. 2013. Simultaneous quantification of multiple food- and waterborne pathogens by use of microfluidic quantitative PCR. *Appl Environ Microbiol,* 79**,** 2891-8.

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**,** 2522-7.

Jamieson, R., Joy, D. M., Lee, H., Kostaschuk, R. & Gordon, R. 2005. Transport and deposition of sediment-associated Escherichia coli in natural streams. *Water Res,* 39**,** 2665-75.

Janda, J. M. & Abbott, S. L. 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol,* 45**,** 2761-4.

Janda, J. M. & Abbott, S. L. 2010. The genus Aeromonas: taxonomy, pathogenicity, and infection. *Clin Microbiol Rev,* 23**,** 35-73.

Jenkins, M. W., Tiwari, S., Lorente, M., Gichaba, C. M. & Wuertz, S. 2009. Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays. *Water Res,* 43**,** 4956-66.

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**,** 585-94.

Jeter, S. N., Mcdermott, C. M., Bower, P. A., Kinzelman, J. L., Bootsma, M. J., Goetz, G. W. & Mclellan, S. L. 2009. Bacteroidales diversity in ring-billed gulls (Laurus delawarensis) residing at Lake Michigan beaches. *Appl Environ Microbiol,* 75**,** 1525-33.

Jia, S., He, X., Bu, Y., Shi, P., Miao, Y., Zhou, H., Shan, Z. & Zhang, X. X. 2014. Environmental fate of tetracycline resistance genes originating from swine feedlots in river water. *J Environ Sci Health B,* 49**,** 624-31.

Jiang, Y. 2009. China's water scarcity. *J Environ Manage,* 90**,** 3185-96.

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

Kapoor, V., Pitkanen, T., Ryu, H., Elk, M., Wendell, D. & Domingo, J. W. S. 2015. Distribution of Human-Specific Bacteroidales and Fecal Indicator Bacteria in an Urban Watershed Impacted by Sewage Pollution, Determined Using RNA-and DNA-Based Quantitative PCR Assays. *Applied and Environmental Microbiology,* 81**,** 91-99.

Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Baig, J. A. & Shah, A. Q. 2009. Assessment of water quality of polluted lake using multivariate statistical techniques: a case study. *Ecotoxicol Environ Saf,* 72**,** 301-9.

Khosravi, A. D., Sadeghi, P., Shahraki, A. H., Heidarieh, P. & Sheikhi, N. 2015. Molecular Methods for Identification of Acinetobacter Species by Partial Sequencing of the rpoB and 16S rRNA Genes. *J Clin Diagn Res,* 9**,** DC09-13.

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**,** 3701-15.

Kim, S. Y., Lee, J. E., Lee, S., Lee, H. T., Hur, H. G. & Ko, G. 2010. Characterization of Enterococcus spp. from human and animal feces using 16S rRNA sequences, the esp gene, and PFGE for microbial source tracking in Korea. *Environ Sci Technol,* 44**,** 3423-8.

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**,** 2143-7.

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**,** 761-3.

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**,** 5112-20.

Kuhn, R. C., Rock, C. M. & Oshima, K. H. 2002. Occurrence of Cryptosporidium and Giardia in wild ducks along the Rio Grande River valley in southern New Mexico. *Appl Environ Microbiol,* 68**,** 161-5.

Lapointe, B. E., Herren, L. W., Debortoli, D. D. & Vogel, M. A. 2015. Evidence of sewage-driven eutrophication and harmful algal blooms in Florida's Indian River Lagoon. *Harmful Algae,* 43**,** 82-102.

Laukova, A., Kubasova, I., Kandricakova, A., Strompfova, V., Zitnan, R. & Simonova, M. P. 2018. Relation to enterocins of variable Aeromonas species isolated from trouts of Slovakian aquatic sources and detected by MALDI-TOF mass spectrometry. *Folia Microbiol (Praha).*

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**,** 4214-24.

Layton, B. A., Cao, Y. P., Ebentier, D. L., Hanley, K., Balleste, E., Brandao, J., Byappanahalli, M., Converse, R., Farnleitner, A. H., Gentry-Shields, J., Gidley, M. L., Gourmelon, M., Lee, C. S., Lee, J., Lozach, S., Madi, T., Meijer, W. G., Noble, R., Peed, L., Reischer, G. H., Rodrigues, R., Rose, J. B., Schriewer, A., Sinigalliano, C., Srinivasan, S., Stewart, J., Van De Werfhorst, L. C., Wang, D., Whitman, R., Wuertz, S., Jay, J., Holden, P. A., Boehm, A. B., Shanks, O. & Griffith, J. F. 2013. Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. *Water Research,* 47**,** 6897-6908.

Leclerc, H., Schwartzbrod, L. & Dei-Cas, E. 2002. Microbial agents associated with waterborne diseases. *Crit Rev Microbiol,* 28**,** 371-409.

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

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**,** 1647-51.

Li, S., Li, J. & Zhang, Q. 2011. Water quality assessment in the rivers along the water conveyance system of the middle route of the south to north water transfer project (China) using multivariate statistical techniques and receptor modeling. *J Hazard Mater,* 195**,** 306-17.

Liu, C.-W., Lin, K.-H. & Kuo, Y.-M. 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci Total Environ,* 313**,** 77-89.

Liu, J., Liang, X., Yang, J., Ye, Y., Su, M., Nie, Z. & Chen, Y. 2011. Size distribution and composition of phosphorus in the East Tiao river, China: the significant role of colloids. *J Environ Monit,* 13**,** 2844-50.

Lodder, W. J., Van Den Berg, H. H., Rutjes, S. A. & De Roda Husman, A. M. 2010. Presence of enteric viruses in source waters for drinking water production in The Netherlands. *Appl Environ Microbiol,* 76**,** 5965-71.

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**,** 434-9.

Lu, J., Santo Domingo, J. W., Hill, S. & Edge, T. A. 2009. Microbial diversity and host-specific sequences of Canada goose feces. *Appl Environ Microbiol,* 75**,** 5919-26.

Lu, J., Santo Domingo, J. W., Lamendella, R., Edge, T. & Hill, S. 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. *Appl Environ Microbiol,* 74**,** 3969-76.

Lucassen, E. C. H. E. T., Smolders, A. J. P., Vander Salm, A. L. & Roelofs, J. G. M. 2004. High groundwater nitrate concentrations inhibit eutrophication of sulphate-rich freshwater wetlands. *Biogeochemistry,* 00**,** 249–267.

Lund, M., Nordentoft, S., Pedersen, K. & Madsen, M. 2004. Detection of Campylobacter spp. in chicken fecal samples by real-time PCR. *J Clin Microbiol,* 42**,** 5125-32.

Lv, H., Xu, Y., Han, L. & Zhou, F. 2015. Scale-dependence effects of landscape on seasonal water quality in Xitiaoxi catchment of Taihu Basin, China. *Water Sci Technol,* 71**,** 59-66.

Malla, B., Ghaju Shrestha, R., Tandukar, S., Bhandari, D., Inoue, D., Sei, K., Tanaka, Y., Sherchand, J. B. & Haramoto, E. 2018. Validation of host-specific Bacteroidales quantitative PCR assays and their application to microbial source tracking of drinking water sources in the Kathmandu Valley, Nepal. *J Appl Microbiol*.

Marsalek, J. & Rochfort, Q. 2004. Urban wet-weather flows: sources of fecal contamination impacting on recreational waters and threatening drinking-water sources. *J Toxicol Environ Health A,* 67**,** 1765-77.

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. *Front Microbiol,* 8**,** 19.

Marti, R., Zhang, Y., Lapen, D. R. & Topp, E. 2011. Development and validation of a microbial source tracking marker for the detection of fecal pollution by muskrats. *J Microbiol Methods,* 87**,** 82-8.

Mclellan, S. L., Daniels, A. D. & Salmore, A. K. 2003. Genetic characterization of Escherichia coli populations from host sources of fecal pollution by using DNA fingerprinting. *Appl Environ Microbiol,* 69**,** 2587-94.

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. *Environ Microbiol,* 15**,** 2213-27.

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**,** 9482-91.

Mcquaig, S., Griffith, J. & Harwood, V. J. 2012. Association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution. *Appl Environ Microbiol,* 78**,** 6423-32.

Meays, C. L., Broersma, K., Nordin, R. & Mazumder, A. 2004. Source tracking fecal bacteria in water: a critical review of current methods. *J Environ Manage,* 73**,** 71-9.

Mei, K., Liao, L., Zhu, Y., Lu, P., Wang, Z., Dahlgren, R. A. & Zhang, M. 2014. Evaluation of spatial-temporal variations and trends in surface water quality across a rural-suburban-urban interface. *Environ Sci Pollut Res,* 21**,** 8036–8051.

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**,** 3045-54.

Mieszkin, S., Yala, J. F., Joubrel, R. & Gourmelon, M. 2010. Phylogenetic analysis of Bacteroidales 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. *J Appl Microbiol,* 108**,** 974-84.

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**,** 255-61.

Mitch, A. A., Gasner, K. C. & Mitch, W. A. 2010. Fecal coliform accumulation within a river subject to seasonally-disinfected wastewater discharges. *Water Res,* 44**,** 4776-82.

Monahan, A. M., Callanan, J. J. & Nally, J. E. 2008. Proteomic analysis of Leptospira interrogans shed in urine of chronically infected hosts. *Infect Immun,* 76**,** 4952-8.

Moore, D. F., Harwood, V. J., Ferguson, D. M., Lukasik, J., Hannah, P., Getrich, M. & Brownell, M. 2005. Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed. *J Appl Microbiol,* 99**,** 618-28.

Muirhead, R. W., Davies-Colley, R. J., Donnison, A. M. & Nagels, J. W. 2004. Faecal bacteria yields in artificial flood events: quantifying in-stream stores. *Water Res,* 38**,** 1215-24.

Murphy, C., Carroll, C. & Jordan, K. N. 2006. Environmental survival mechanisms of the foodborne pathogen Campylobacter jejuni. *J Appl Microbiol,* 100**,** 623-32.

Mustapha, A., Aris, A. Z., Juahir, H., Ramli, M. F. & Kura, N. U. 2013. River water quality assessment using environmentric techniques: case study of Jakara River Basin. *Environ Sci Pollut Res Int,* 20**,** 5630-44.

Napier, M. D., Haugland, R., Poole, C., Dufour, A. P., Stewart, J. R., Weber, D. J., Varma, M., Lavender, J. S. & Wade, T. J. 2017. Exposure to human-associated fecal indicators and self-reported illness among swimmers at recreational beaches: a cohort study. *Environ Health,* 16**,** 103.

Nbsc 2015. National Bureau of Statistics of China.

Nbsc. 2016. *National Bureau of Statistics of China* [Online]. Available: http://www.stats.gov.cn/tjsj/ndsj/2016/indexeh.htm [Accessed 6 June 2018].

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. *Microb Ecol,* 65**,** 1011-23.

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**,** 6972-81.

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**,** 1901-7.

Nnane, D. E., Ebdon, J. E. & Taylor, H. D. 2011. Integrated analysis of water quality parameters for cost-effective faecal pollution management in river catchments. *Water Res,* 45**,** 2235-46.

Noble, R. T., Allen, S. M., Blackwood, A. D., Chu, W., Jiang, S. C., Lovelace, G. L., Sobsey, M. D., Stewart, J. R. & Wait, D. A. 2003. Use of viral pathogens and indicators to differentiate between human and non-human fecal contamination in a microbial source tracking comparison study. *J Water Health,* 1**,** 195-207.

Noori, R., Sabahi, M. S., Karbassi, A. R., Baghvand, A. & Taati Zadeh, H. 2010. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination,* 260**,** 129-136.

Nshimyimana, J. P., Cruz, M. C., Thompson, R. J. & Wuertz, S. 2017. Bacteroidales markers for microbial source tracking in Southeast Asia. *Water Res,* 118**,** 239-248.

Nshimyimana, J. P., Ekklesia, E., Shanahan, P., Chua, L. H. & Thompson, J. R. 2014. Distribution and abundance of human-specific Bacteroides and relation to traditional indicators in an urban tropical catchment. *J Appl Microbiol,* 116**,** 1369-83.

Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P. R., Panigrahi, P. & Jenkins, M. W. 2015. Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Sci Total Environ,* 502**,** 462-70.

Ogwueleka, T. C. 2015. Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. *Environ Monit Assess,* 187**,** 137.

Ohad, S., Ben-Dor, S., Prilusky, J., Kravitz, V., Dassa, B., Chalifa-Caspi, T., Kashi, Y. & Rorman, E. 2016. The Development of a Novel qPCR Assay-Set for Identifying Fecal Contamination Originating from Domestic Fowls and Waterfowl in Israel. *Frontiers in Microbiology,* 7.

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**,** 5037-47.

Okabe, S., Okayama, N., Savichtcheva, O. & Ito, T. 2007. Quantification of host-specific Bacteroides-Prevotella 16S rRNA genetic markers for assessment of fecal pollution in freshwater. *Appl Microbiol Biotechnol,* 74**,** 890-901.

Oki, T. & Kanae, S. 2006. Global hydrological cycles and world water resources. *Science,* 313**,** 1068-72.

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**,** 14148-57.

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.

Ouyang, Y. 2005. Evaluation of river water quality monitoring stations by principal component analysis. *Water Res,* 39**,** 2621-35.

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**,** 1973-83.

Parveen, S., Portier, K. M., Robinson, K., Edmiston, L. & Tamplin, M. L. 1999. Discriminant analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. *Appl Environ Microbiol,* 65**,** 3142-7.

Pearce, G. R., M Ramzan Chaudhry & Ghulam., S. 1999. A simple methodology for water quality monitoring. HR Wallingford.: Department for International Development.

Pejman, A. H., Nabi Bidhendi, G. R., Karbassi, A. R., Mehrdadi, N. & Bidhendi, M. E. 2009. Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. *Int. J. Environ. Sci. Tech.,,* 6**,** 467-476.

Phung, D., Huang, C., Rutherford, S., Dwirahmadi, F., Chu, C., Wang, X., Nguyen, M., Nguyen, N. H., Do, C. M., Nguyen, T. H. & Dinh, T. A. 2015. Temporal and spatial assessment of river surface water quality using multivariate statistical techniques: a study in Can Tho City, a Mekong Delta area, Vietnam. *Environ Monit Assess,* 187**,** 229.

Pisanic, N., Nadimpalli, M., Rinsky, J. L., Stewart, J., Wing, S., Love, D. C., Hall, D. & Heaney, C. D. 2015. Pig-2-Bac as a biomarker of occupational exposure to pigs and livestock-associated Staphylococcus aureus among industrial hog operation workers. *Environ Res,* 143**,** 93-7.

Pisciotta, J. M., Rath, D. F., Stanek, P. A., Flanery, D. M. & Harwood, V. J. 2002. Marine bacteria cause false-positive results in the Colilert-18 rapid identification test for Escherichia coli in Florida waters. *Appl Environ Microbiol,* 68**,** 539-544.

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

Qiu, J. 2011. China faces up to 'terrible' state of its ecosystems. *Nature,* 471**,** 19.

Raith, M. R., Kelty, C. A., Griffith, J. F., Schriewer, A., Wuertz, S., Mieszkin, S., Gourmelon, M., Reischer, G. H., Farnleitner, A. H., Ervin, J. S., Holden, P. A., Ebentier, D. L., Jay, J. A., Wang, D., Boehm, A. B., Aw, T. G., Rose, J. B., Balleste, E., Meijer, W. G., Sivaganesan, M. & Shanks, O. C. 2013. Comparison of PCR and quantitative real-time PCR methods for the characterization of ruminant and cattle fecal pollution sources. *Water Research,* 47**,** 6921-6928.

Razmkhah, H., Abrishamchi, A. & Torkian, A. 2010. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran). *J Environ Manage,* 91**,** 852-60.

Reasoner, D. J. & Geldreich, E. E. 1985. A new medium for the enumeration and subculture of bacteria from potable water. *Appl Environ Microbiol,* 49**,** 1-7.

Reischer, G. H., Ebdon, J. E., Bauer, J. M., Schuster, N., Ahmed, W., Astrom, J., Blanch, A. R., Bloschl, G., Byamukama, D., Coakley, T., Ferguson, C., Goshu, G., Ko, G., De Roda Husman, A. M., Mushi, D., Poma, R., Pradhan, B., Rajal, V., Schade, M. A., Sommer, R., Taylor, H., Toth, E. M., Vrajmasu, V., Wuertz, S., Mach, R. L. & Farnleitner, A. H. 2013. Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ Sci Technol,* 47**,** 8548-56.

Reischer, G. H., Kasper, D. C., Steinborn, R., Farnleitner, A. H. & Mach, R. L. 2007. A quantitative real-time PCR assay for the highly sensitive and specific detection of human faecal influence in spring water from a large alpine catchment area. *Lett Appl Microbiol,* 44**,** 351-6.

Ridley, C. M., Jamieson, R. C., Hansen, L. T., Yost, C. K. & Bezanson, G. S. 2014. Baseline and storm event monitoring of Bacteroidales marker concentrations and enteric pathogen presence in a rural Canadian watershed. *Water Research,* 60**,** 278-288.

Ritter, K. J., Carruthers, E., Carson, C. A., Ellender, R. D., Harwood, V. J., Kingsley, K., Nakatsu, C., Sadowsky, M., Shear, B., West, B., Whitlock, J. E., Wiggins, B. A. & Wilbur, J. D. 2003. Assessment of statistical methods used in library-based approaches to microbial source tracking. *J Water Health,* 1**,** 209-23.

Robinson, B. J., Ritter, K. J. & Ellender, R. D. 2007. A statistical appraisal of disproportional versus proportional microbial source tracking libraries. *J Water Health,* 5**,** 503-9.

Rose, G., Wooldridge, D. J., Anscombe, C., Mee, E. T., Misra, R. V. & Gharbia, S. 2015. Challenges of the Unknown: Clinical Application of Microbial Metagenomics. *International Journal of Genomics*.

Roslev, P. & Bukh, A. S. 2011. State of the art molecular markers for fecal pollution source tracking in water. *Appl Microbiol Biotechnol,* 89**,** 1341-55.

Ruzdjak, A. M. & Ruzdjak, D. 2015. Evaluation of river water quality variations using multivariate statistical techniques: Sava River (Croatia): a case study. *Environ Monit Assess,* 187**,** 215.

Santo Domingo, J. W., Bambic, D. G., Edge, T. A. & Wuertz, S. 2007. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Water Res,* 41**,** 3539-52.

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**,** 2463-76.

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

Schmidt, P. J. & Emelko, M. B. 2011. QMRA and decision-making: are we handling measurement errors associated with pathogen concentration data correctly? *Water Res,* 45**,** 427-38.

Schriewer, A., Goodwin, K. D., Sinigalliano, C. D., Cox, A. M., Wanless, D., Bartkowiak, J., Ebentier, D. L., Hanley, K. T., Ervin, J., Deering, L. A., Shanks, O. C., Peed, L. A., Meijer, W. G., Griffith, J. F., Santodomingo, J., Jay, J. A., Holden, P. A. & Wuertz, S. 2013. Performance evaluation of canine-associated Bacteroidales assays in a multi-laboratory comparison study. *Water Research,* 47**,** 6909-6920.

Schriewer, A., Miller, W. A., Byrne, B. A., Miller, M. A., Oates, S., Conrad, P. A., Hardin, D., Yang, H. H., Chouicha, N., Melli, A., Jessup, D., Dominik, C. & Wuertz, S. 2010. Presence of Bacteroidales as a predictor of pathogens in surface waters of the central California coast. *Appl Environ Microbiol,* 76**,** 5802-14.

Schwarzenbach, R. P., Egli, T., Hofstetter, T. B., Von Gunten, U. & Wehrli, B. 2010. Global Water Pollution and Human Health. *Annual Review of Environment and Resources,* 35**,** 109-136.

Scott, T. M., Jenkins, T. M., Lukasik, J. & Rose, J. B. 2005. Potential use of a host associated molecular marker in Enterococcus faecium as an index of human fecal pollution. *Environ Sci Technol,* 39**,** 283-7.

Scott, T. M., Parveen, S., Portier, K. M., Rose, J. B., Tamplin, M. L., Farrah, S. R., Koo, A. & Lukasik, J. 2003. Geographical variation in ribotype profiles of Escherichia coli isolates from humans, swine, poultry, beef, and dairy cattle in Florida. *Appl Environ Microbiol,* 69**,** 1089-92.

Scott, T. M., Rose, J. B., Jenkins, T. M., Farrah, S. R. & Lukasik, J. 2002. Microbial source tracking: current methodology and future directions. *Appl Environ Microbiol,* 68**,** 5796-803.

Scupham, A. J., Patton, T. G., Bent, E. & Bayles, D. O. 2008. Comparison of the cecal microbiota of domestic and wild turkeys. *Microb Ecol,* 56**,** 322-31.

Seurinck, S., Defoirdt, T., Verstraete, W. & Siciliano, S. D. 2005. Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. *Environ Microbiol,* 7**,** 249-59.

Sha Long, Lin Zhao, Tongtong Shi, Jingchen Li, Jiangye Yang, Hongbo Liu, Guozhu Mao, Zhi Qiao & Yang., Y. 2018. Pollution control and cost analysis of wastewater treatment at industrial parks in Taihu and Haihe water basins, China. *Journal of Cleaner Production,* 172 (2435-2442.

Shahrampour, D., Yavarmanesh, M., Najafi, M. B. H. & Mohebbi, M. 2015. Application of F(+)RNA Coliphages as Source Tracking Enteric Viruses on Parsley and Leek Using RT-PCR. *Food and Environmental Virology,* 7**,** 381-389.

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. *Environ Monit Assess,* 186**,** 7055-62.

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., Atikovic, E., Blackwood, A. D., Lu, J., Noble, R. T., Domingo, J. S., Seifring, S., Sivaganesan, M. & Haugland, R. A. 2008. Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. *Appl Environ Microbiol,* 74**,** 745-52.

Shanks, O. C., Kelty, C. A., Sivaganesan, M., Varma, M. & Haugland, R. A. 2009. Quantitative PCR for genetic markers of human fecal pollution. *Appl Environ Microbiol,* 75**,** 5507-13.

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**,** 2906-13.

Sharma, A., Bora, C. R. & Shukla, V. 2013. Evaluation of seasonal changes in physico-chemical and bacteriological characteristics of water from the Narmada river (India) using multivariate analysis. *Nat Resour Res.,* 22**,** 283-296.

Sheikhy Narany, T., Ramli, M. F., Aris, A. Z., Sulaiman, W. N. & Fakharian, K. 2014. Spatiotemporal variation of groundwater quality using integrated multivariate statistical and geostatistical approaches in Amol-Babol Plain, Iran. *Environ Monit Assess,* 186**,** 5797-815.

Shock, C. C. & Pratt, K. 2003. Phosphorus effects on surface water quality and phosphorus TMDL development. Western Nutrient Management Conference, March 6-7, 2003 Salt Lake City, UT. Salt Lake City, UT, 211-220.

Shrestha, S. & Kazama, F. 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environ Model Softw,* 22**,** 464-475.

Shuval, H. 2003. Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. *J Water Health,* 1**,** 53-64.

Sidhu, J. P. S., Hodgers, L., Ahmed, W., Chong, M. N. & Toze, S. 2012. Prevalence of human pathogens and indicators in stormwater runoff in Brisbane, Australia. *Water Research,* 46**,** 6652-6660.

Siefring, S., Varma, M., Atikovic, E., Wymer, L. & Haugland, R. A. 2008. Improved real-time PCR assays for the detection of fecal indicator bacteria in surface waters with different instrument and reagent systems. *J Water Health,* 6**,** 225-37.

Silkie, S. S. & Nelson, K. L. 2009. Concentrations of host-specific and generic fecal markers measured by quantitative PCR in raw sewage and fresh animal feces. *Water Res,* 43**,** 4860-71.

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.

Singh, K. P., Malik, A., Mohan, D. & Sinha, S. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)--a case study. *Water Res,* 38**,** 3980-92.

Singh, K. P., Malik, A. & Sinha, S. 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. *Anal Chim Acta,* 538**,** 355-374.

Soller, J. A., Schoen, M. E., Bartrand, T., Ravenscroft, J. E. & Ashbolt, N. J. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Res,* 44**,** 4674-91.

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**,** 5565-70.

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.

Stevens, G., Mascarenhas, M. & Mathers, C. 2009. Global health risks: progress and challenges. *Bull World Health Organ,* 87**,** 646.

Stewart, J. R., Santo Domingo, J. W. & Wade, T. J. 2007. *Microbial Source Tracking,* Washington, DC, USA.

Stoddard, R. A., Gee, J. E., Wilkins, P. P., Mccaustland, K. & Hoffmaster, A. R. 2009. Detection of pathogenic Leptospira spp. through TaqMan polymerase chain reaction targeting the LipL32 gene. *Diagn Microbiol Infect Dis,* 64**,** 247-55.

Stoeckel, D. M. & Harwood, V. J. 2007. Performance, design, and analysis in microbial source tracking studies. *Appl Environ Microbiol,* 73**,** 2405-15.

Stricker, A. R., Wilhartitz, I., Farnleitner, A. H. & Mach, R. L. 2008. Development of a Scorpion probe-based real-time PCR for the sensitive quantification of Bacteroides sp. ribosomal DNA from human and cattle origin and evaluation in spring water matrices. *Microbiol Res,* 163**,** 140-7.

Sun, D., Duan, C. R., Shang, Y. N., Ma, Y. X., Tan, L. L., Zhai, J., Gao, X., Guo, J. S. & Wang, G. X. 2016. Application of Faecalibacterium 16S rDNA genetic marker for accurate identification of duck faeces. *Environmental Science and Pollution Research,* 23**,** 7639-7647.

Tambalo, D. D., Fremaux, B., Boa, T. & Yost, C. K. 2012. Persistence of host-associated Bacteroidales gene markers and their quantitative detection in an urban and agricultural mixed prairie watershed. *Water Research,* 46**,** 2891-2904.

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. *Front Microbiol,* 6**,** 1027.

Tang, X., Ullah, N., Yu, C., Qin, Z., Huang, R., Zhu, X., Shen, C. & Chen, Y. 2012. Chemical and bioassay analysis of estrogen pollution in the surface water of the Tiaoxi River, the source river for Taihu Lake. *Bull Environ Contam Toxicol,* 89**,** 816-9.

Tao, T. & Xin, K. 2014. Public health: A sustainable plan for China's drinking water. *Nature,* 511**,** 527-8.

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**,** 423-8.

Thomas, T., Gilbert, J. & Meyer, F. 2012. Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp,* 2**,** 3.

Thorpe, C. M. 2004. Shiga toxin-producing Escherichia coli infection. *Clin Infect Dis,* 38**,** 1298-303.

Tornqvist, R., Jarsjo, J. & Karimov, B. 2011. Health risks from large-scale water pollution: trends in Central Asia. *Environ Int,* 37**,** 435-42.

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**,** 448-53.

Ufnar, J. A., Wang, S. Y., Ufnar, D. F. & Ellender, R. D. 2007. Methanobrevibacter ruminantium as an indicator of domesticated-ruminant fecal pollution in surface waters. *Appl Environ Microbiol,* 73**,** 7118-21.

Unicef. 2018. *Diarrhoeal Disease: Current Status and Progress.* [Online]. Available: https://data.unicef.org/topic/child-health/diarrhoeal-disease/ [Accessed 30 April 2018].

Unno, T., Di, D. Y., Jang, J., Suh, Y. S., Sadowsky, M. J. & Hur, H. G. 2012. Integrated online system for a pyrosequencing-based microbial source tracking method that targets Bacteroidetes 16S rDNA. *Environ Sci Technol,* 46**,** 93-8.

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**,** 7777-82.

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. *Environ Microbiol*.

Usepa. 2012. *United States Environmental Protection Agency. Water: Monitoring & Analysis.* [Online]. [Accessed 17th Jan 2017].

Usepa 2013a. Aquatic Life Ambient Water Quality Criteria for Ammonia – Freshwater. *In:* WATER, O. O. (ed.). Washington, D.C.: United States Environmental Protection Agency.

Usepa 2013b. United States Environmental Protection Agency. Total Nitrogen. *In:* WATER, O. O. (ed.). Washington D.C.

Vergara, G. G. R. V., Goh, S. G., Rezaeinejad, S., Chang, S. Y., Sobsey, M. D. & Gin, K. Y. H. 2015. Evaluation of FRNA coliphages as indicators of human enteric viruses in a tropical urban freshwater catchment. *Water Research,* 79**,** 39-47.

Vieira, J. S., Pires, J. C. M., Martins, F. G., Vilar, V. J. P., Boaventura, R. A. R. & Botelho, C. M. S. 2012. Surface water quality assessment of Lis river using multivariate statistical methods. *Water Air Soil Pollut,* 223**,** 5549-5561.

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 Sci Technol,* 72**,** 1962-72.

Vinje, J., Oudejans, S. J., Stewart, J. R., Sobsey, M. D. & Long, S. C. 2004. Molecular detection and genotyping of male-specific coliphages by reverse transcription-PCR and reverse line blot hybridization. *Appl Environ Microbiol,* 70**,** 5996-6004.

Wade, T. J., Pai, N., Eisenberg, J. N. S. & Colford, J. M. 2003. Do U.S. Environmental Protection Agency Water Quality Guidelines for Recreational Waters Prevent Gastrointestinal Illness? A Systematic Review and Meta-analysis. *Environmental Health Perspectives,* 111**,** 1102-1109.

Walters, S. P. & Field, K. G. 2009. Survival and persistence of human and ruminant-specific faecal Bacteroidales in freshwater microcosms. *Environ Microbiol,* 11**,** 1410-21.

Walters, S. P., Yamahara, K. M. & Boehm, A. B. 2009. Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: implications for their use in assessing risk in recreational waters. *Water Res,* 43**,** 4929-39.

Wang, F., Tian, P., Yu, J., Lao, G.-m. & Shi, T.-c. 2011a. Variations in pollutant fluxes of rivers surrounding Taihu Lake in Zhejiang Province in 2008. *Physics and Chemistry of the Earth, Parts A/B/C,* 36**,** 366-371.

Wang, J., Soininen, J., Zhang, Y., Wang, B., Yang, X. & Shen, J. 2011b. Contrasting patterns in elevational diversity between microorganisms and macroorganisms. *J. Biogeogr.,* 38**,** 595–603.

Wang, L., Li, Y. & Mustaphai, A. 2007a. Rapid and simultaneous quantitation of Escherichia coli 0157:H7, Salmonella, and Shigella in ground beef by multiplex real-time PCR and immunomagnetic separation. *J Food Prot,* 70**,** 1366-72.

Wang, X.-l., Lu, Y.-l., Han, J.-y., He, G.-z. & Wang, T.-y. 2007b. Identification of anthropogenic influences on water quality of rivers in Taihu watershed. *J Environ Sci,* 19**,** 475–481.

Wang, X., Cai, Q., Ye, L. & Qu, X. 2012. Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: A case study of the Xiangxi River basin, China. *Quat Int,* 282**,** 137-144.

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**,** D581-91.

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**,** 6044-54.

Wexler, H. M. 2007. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev,* 20**,** 593-621.

Whitlock, J. E., Jones, D. T. & Harwood, V. J. 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. *Water Res,* 36**,** 4273-82.

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

Who. 2017. *World Health Organization: Diarrhoeal disease fact sheets.* [Online]. Available: http://www.who.int/en/news-room/fact-sheets/detail/diarrhoeal-disease [Accessed 22 April 2018].

Wicki, M., Auckenthaler, A., Felleisen, R., Liniger, M., Loutre, C., Niederhauser, I., Tanner, M. & Baumgartner, A. 2012. Improved detection of Rhodococcus coprophilus with a new quantitative PCR assay. *Appl Microbiol Biotechnol,* 93**,** 2161-9.

Wiggins, B. A., Cash, P. W., Creamer, W. S., Dart, S. E., Garcia, P. P., Gerecke, T. M., Han, J., Henry, B. L., Hoover, K. B., Johnson, E. L., Jones, K. C., Mccarthy, J. G., Mcdonough, J. A., Mercer, S. A., Noto, M. J., Park, H., Phillips, M. S., Purner, S. M., Smith, B. M., Stevens, E. N. & Varner, A. K. 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. *Appl Environ Microbiol,* 69**,** 3399-405.

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.

Wolf, S., Hewitt, J. & Greening, G. E. 2010. Viral multiplex quantitative PCR assays for tracking sources of fecal contamination. *Appl Environ Microbiol,* 76**,** 1388-94.

Wolf, S., Hewitt, J., Rivera-Aban, M. & Greening, G. E. 2008. Detection and characterization of F+ RNA bacteriophages in water and shellfish: application of a multiplex real-time reverse transcription PCR. *J Virol Methods,* 149**,** 123-8.

Wu, C. H., Huang, L. T., Huang, I. F., Liu, J. W., Chen, J. B., Liang, C. D., Hwang, K. P. & Tiao, M. M. 2009. Acute non-outbreak shigellosis: ten years experience in southern Taiwan. *Chang Gung Med J,* 32**,** 59-65.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F. D. & Lewis, J. D. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science,* 334**,** 105-8.

Wuertz, S., Wang, D., Reischer, G. H. & Farnleitner, A. H. 2011. *Library-Independent Bacterial Source Tracking Methods.,* Springer, New York, NY, Springer Science+Business Media.

Xu, H., W. Paerl, H., Qin, B., Zhu, G. & Gaoa, G. 2010. Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. *Limnol. Oceanogr,* 55**,** 420–432.

Xu, H., Yang, L.-Z., Zhao, G.-M., Jiao, J.-G., Yin, S.-X. & Liu, Z.-P. 2009. Anthropogenic impact on surface water quality in Taihu Lake Region, China. *Pedosphere,* 19**,** 765–778.

Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V. & Gordon, J. I. 2003. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science,* 299**,** 2074-6.

Xu, S., Huang, B., Wei, Z. B., Luo, J., Miao, A. J. & Yang, L. Y. 2013. Seasonal variation of phytoplankton nutrient limitation in Lake Taihu, China: a monthly study from year 2011 to 2012. *Ecotoxicol Environ Saf,* 94**,** 190-6.

Yamahara, K. M., Layton, B. A., Santoro, A. E. & Boehm, A. B. 2007. Beach sands along the California coast are diffuse sources of fecal bacteria to coastal waters. *Environ Sci Technol,* 41**,** 4515-21.

Yan, C. A., Zhang, W., Zhang, Z., Liu, Y., Deng, C. & Nie, N. 2015. Assessment of water quality and identification of polluted risky regions based on field observations & GIS in the Honghe River watershed, China. *PLoS One,* 10**,** e0119130.

Yan, T. & Sadowsky, M. J. 2007. Determining sources of fecal bacteria in waterways. *Environ Monit Assess,* 129**,** 97-106.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., Knight, R. & Gordon, J. I. 2012. Human gut microbiome viewed across age and geography. *Nature,* 486**,** 222-7.

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**,** 448-52.

Zhang, Q., Eichmiller, J. J., Staley, C., Sadowsky, M. J. & Ishii, S. 2016a. Correlations between pathogen concentration and fecal indicator marker genes in beach environments. *Sci Total Environ,* 573**,** 826-830.

Zhang, Y., Shi, K., Zhou, Y., Liu, X. & Qin, B. 2016b. Monitoring the river plume induced by heavy rainfall events in large, shallow, Lake Taihu using MODIS 250m imagery. *Remote Sens Environ,* 173**,** 109-121.

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.

Zheijiang, G. o. 2016. *Zhejiang Scrutinizes Livestock, Poultry Pollution* [Online]. Government of Zheijiang. Available: http://english.zj.gov.cn/art/2016/4/27/art_5798_2098514.html [Accessed].

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.

Zhu, X. Y., Zhong, T., Pandya, Y. & Joerger, R. D. 2002. 16S rRNA-based analysis of microbiota from the cecum of broiler chickens. *Appl Environ Microbiol,* 68**,** 124-37.

Zhuang, F. F., Li, H., Zhou, X. Y., Zhu, Y. G. & Su, J. Q. 2017. Quantitative detection of fecal contamination with domestic poultry feces in environments in China. *AMB Express,* 7**,** 80.

Supplementary Files

Table S2.1 Range, mean and standard deviation of different water quality parameters measured at different locations in Taihu watershed during 2014-2015.

Locations	TN (mg/L)			TP (µg/L)			NO_3-N (mg/L)			NO_2-N (mg/L)			PO_4-P (µg/L)		
	Range	Mean	S.D	Range	Mean	S.D	Range	Mean	S.D	Range	Mean	S.D	Range	Mean	S.D
Location 1	1.31-2.95	2.06	0.83	68.88-102.61	84.76	16.95	0.37-2.44	1.42	1.04	0.01-0.10	0.07	0.06	2.46-23.58	9.77	11.97
Location 2	1.58-3.06	2.30	0.74	64.6-118.04	91.96	26.74	0.53-2.59	1.59	1.03	0.01-0.11	0.06	0.05	5.70-30/28	15.40	13.08
Location 3	1.90-3.47	2.50	0.85	66.29-113.6	91.31	23.77	0.55-2.96	1.67	1.22	0.01-0.16	0.08	0.08	7.45-30.24	17.97	11.50
Location 4	2.40-3.50	3.06	0.58	74.89-110.24	95.23	18.27	1.71-3.34	2.45	0.83	0.05-0.08	0.07	0.02	21.52-23.67	25.81	5.67
Location 5	2.50-3.50	3.16	0.57	81.27-109.4	94.40	14.16	1.85-3.08	2.67	0.71	0.04-0.08	0.06	0.02	14.13-29.75	21.77	7.82
Location 6	2.41-3.64	3.15	0.65	67.43-87.09	77.73	9.86	1.78-3.33	2.71	0.82	0.05-0.08	0.06	0.02	9.01-29.39	19.43	10.20
Location 7	2.45-3.20	2.92	0.41	63.94-95.78	78.42	16.11	1.96-2.91	2.57	0.53	0.04-0.07	0.06	0.02	10.36-27.94	19.61	8.83
Location 8	2.59-3.25	2.97	0.34	57.63-108.64	79.69	26.20	1.99-2.87	2.53	0.47	0.04-0.07	0.06	0.02	11.86-29.23	21.18	8.76
Location 9	2.44-3.27	2.95	0.44	55.76-87.27	70.58	15.84	1.99-2.82	2.53	0.47	0.04-0.07	0.06	0.02	12.37-28.60	21.48	8.29
Location 10	2.33-3.07	2.76	0.38	53.42-86.25	68.74	16.52	1.90-2.83	2.46	0.50	0.03-0.08	0.06	0.02	10.87-29.11	20.66	9.19
Location 11	2.25-3.02	2.76	0.44	59.31-88.12	71.80	14.78	1.75-2.75	2.39	0.55	0.03-0.07	0.05	0.02	11.64-29.11	20.41	8.69
Location 12	1.77-3.25	2.40	0.77	88.33-94.32	92.09	3.27	0.82-2.5	1.67	0.89	0.01-0.07	0.05	0.03	2.52-33.28	19.50	15.63
Location 13	2.73-3.69	2.85	0.79	77.06-112.64	92.97	18.09	1.82-3.51	2.54	0.87	0.04-0.06	0.06	0.01	10.08-30.71	18.16	11.02
Location 14	2.13-3.31	2.57	0.65	80.45-115.82	94.68	18.67	1.49-2.41	1.95	0.46	0.03-0.11	0.06	0.04	2.79-29.45	19.79	14.76
Location 15	2.51-3.74	2.96	0.68	79.76-129.68	103.47	25.05	2.03-2.97	2.37	0.53	0.03-0.17	0.08	0.08	3.51-28.15	18.54	13.18
Location 16	2.92-3.27	3.10	0.18	135.26-187.42	166.09	27.35	0.96-1.88	1.48	0.47	0.01-0.18	0.08	0.09	12.47-51.91	34.48	20.11
Location 17	2.27-3.34	2.84	0.54	69.24-100.27	83.58	15.65	2.02-3.09	2.59	0.54	0.04-0.07	0.06	0.02	12.47-28.14	19.96	7.91
Location 18	2.33-4.13	3.23	1.27	58.83-103.63	81.23	31.68	1.20-3.86	2.54	1.88	0.04-0.11	0.08	0.05	16.54-19.85	18.20	2.34
Location 19	2.12-2.95	2.54	0.59	92.1-94.86	93.48	1.95	1.07-2.44	1.76	0.97	0.06-0.09	0.08	0.02	24.02-25.88	24.95	1.31
Location 20	2.03-4.02	2.83	1.06	93.3-140.48	115.07	23.80	1.19-3.42	2.06	1.20	0.02-0.14	0.09	0.06	20.44-37.60	26.79	9.41
Location 21	1.88-3.18	2.52	0.65	75.31-117.14	102.15	23.30	1.45-2.29	1.65	0.57	0.02-0.04	0.06	0.04	13.75-36.85	22.89	12.28
Location 22	2.11-2.38	2.25	0.19	116.25-142.74	129.50	18.73	1.17-1.36	1.27	0.13	0.01-0.05	0.03	0.03	14.63-38.26	26.45	16.71
Location 23	1.87-3.24	2.56	0.97	85.34-149.96	117.65	45.69	1.17-1.70	1.44	0.37	0.02-0.04	0.04	0.02	2.40-6.80	4.60	3.11
Location 24	2.29-3.70	3.00	1.00	95.78-144.62	120.20	34.54	1.68-2.71	2.20	0.73	0.04-0.04	0.04	0.00	26.51-35.24	30.88	6.17
Location 25	2.60-2.98	2.79	0.27	72.97-119.81	96.39	33.12	1.71-2.85	2.28	0.80	0.03-0.12	0.08	0.06	13.60-41.13	27.37	19.47

Locations	NH_4-N (mg/L)			TOC (mg/L)			Chl *a* (µg/L)			TVC (x 10^3 CFU/mL)			TC (x 10^2 CFU/mL)			FC (x 10^2 CFU/100mL)		
	Range	**Mean**	**S.D**	**Range**	**Mean**	**S.D**	**Range**	**Mean**	**S.D**	**Range**	**Mean**	**S.D**	**Range**	**Mean**	**S.D**	**Range**	**Mean**	**S.D**
Location 1	0.05-0.06	0.08	0.03	3.83-8.45	5.44	2.62	80.3-102.3	91.33	11.1	1.15-15.50	8.15	7.18	2.25-12.40	7.66	5.11	1.50-7.00	4.25	3.89
Location 2	0.11-0.24	0.18	0.06	2.44-12.11	7.48	4.85	75.9-95.9	85.87	10.0	3.74-10.00	5.98	3.49	2.50-20.40	10.30	9.17	13.50-20.50	17.00	4.95
Location 3	0.06-0.27	0.20	0.11	2.27-15.81	8.47	6.84	47.3-87.5	70.23	20.7	7.05-11.80	8.88	2.56	5.00-21.33	12.91	8.18	25.00-26.00	25.50	0.71
Location 4	0.04-0.66	0.32	0.32	2.75-16.71	8.70	7.21	40.7-86.7	66.67	23.6	6.12-8.65	7.46	1.27	10.55-13.67	11.66	1.75	10.50-18.50	14.50	5.66
Location 5	0.06-0.57	0.30	0.26	1.97-14.81	7.79	6.51	53.9-83.2	69.90	14.8	4.19-22.05	14.01	9.06	10.50-20.75	17.08	5.71	28.50-30.50	29.50	1.41
Location 6	0.12-0.43	0.25	0.16	3.67-11.31	5.79	4.83	53.9-77.4	68.33	12.6	3.04-9.80	6.95	3.50	1.00-9.50	6.05	4.47	7.50-28.00	17.75	14.5
Location 7	0.11-0.44	0.26	0.17	4.00-11.31	5.25	3.29	37.4-53.4	43.83	8.5	5.43-9.00	6.84	1.90	1.50-15.00	6.67	7.29	1.50-7.50	4.50	4.24
Location 8	0.12-0.46	0.27	0.18	2.75-10.51	5.58	4.28	37.4-59.8	47.07	11.5	6.45-8.00	7.21	0.78	1.50-12.33	6.19	5.56	2.50-11.25	6.88	6.19
Location 9	0.12-0.46	0.27	0.17	2.59-4.82	3.96	1.20	36.3-49.1	42.03	6.5	4.80-6.82	5.52	1.13	1.25-8.50	5.92	4.05	3.50-4.00	3.75	0.35
Location 10	0.08-0.45	0.25	0.19	2.54-4.96	4.11	1.36	36.3-62.3	48.27	13.1	4.20-6.30	5.39	1.08	1.25-6.66	4.25	2.75	3.00-11.50	7.25	6.01
Location 11	0.05-0.52	0.26	0.24	2.50-5.50	3.92	1.51	44.3-79.6	59.27	18.3	4.00-8.50	6.64	2.35	1.50-4.00	3.02	1.33	2.25-2.59	2.38	0.18
Location 12	0.14-0.73	0.30	0.38	4.08-5..47	4.92	0.74	76.4-77.4	80.33	6.0	2.75-10.40	6.98	3.89	2.05-17.00	8.02	7.92	2.00-42.00	22.00	28.2
Location 13	0.04-0.43	0.20	0.21	2.98-3.81	3.50	0.46	29.88-66.3	45.26	18.9	5.20-8.18	6.49	1.53	1.50-6.33	3.59	2.48	1.50-15.50	8.50	9.90
Location 14	0.03-0.69	0.29	0.35	3.54-5.06	4.38	0.77	32.98-68.9	51.56	18.0	6.40-9.25	7.85	1.43	8.90-14.33	10.74	3.11	4.25-5.00	4.63	0.53
Location 15	0.07-0.68	0.28	0.35	3.55-5.16	3.94	1.08	41.52-74.3	57.34	16.4	9.60-19.15	13.45	5.04	1.60-33.00	13.70	16.8	4.50-10.00	7.25	3.89
Location 16	0.81-1.01	0.89	0.10	3.23-4.16	3.70	0.47	37.82-79.6	58.94	20.9	6.50-17.80	11.95	5.66	1.45-41.33	16.01	22.0	12.00-43.25	27.63	22.1
Location 17	0.04-0.43	0.20	0.21	2.39-2.78	2.59	0.20	29.98-53.6	39.96	12.2	5.22-9.60	7.32	2.20	0.50-27.66	9.97	15.3	1.00-20.75	10.88	13.9
Location 18	0.01-0.13	0.08	0.08	2.67-7.25	4.96	3.24	45.2-68.6	56.90	16.6	4.75-8.40	6.58	2.58	5.15-10.00	7.58	3.43	1.5	1.50	0.00
Location 19	0.01-0.22	0.12	0.15	4.46-6.21	5.34	1.24	36.3-62.3	49.30	18.4	5.91-10.40	8.16	3.17	1.66-3.10	2.38	1.02	1.75	1.75	0.00
Location 20	0.13-0.59	0.30	0.26	5.94-9.68	7.62	1.90	34.4-71.8	47.87	20.8	4.90-9.20	6.73	2.22	1.75-2.95	2.23	0.63	3.25-6.00	4.63	1.94
Location 21	0.05-0.44	0.27	0.20	5.92-13.76	9.41	3.99	26.8-59.6	40.90	16.9	4.10-10.00	6.78	2.99	2.00-4.05	2.90	1.05	1.50-2.75	2.13	0.88
Location 22	0.01-0.02	0.02	0.01	2.40-6.08	4.24	2.60	61.6-88.4	75.00	19.0	5.93-9.40	7.67	2.45	2.33-3.55	2.94	0.86	4	4.00	0.00
Location 23	0.09-0.11	0.10	0.02	3.77-7.79	5.78	2.84	103.-132.6	118.0	20.7	4.35-14.00	9.18	6.82	1.66-4.65	3.16	2.11	0.5	0.50	0.00
Location 24	0.20-1.02	0.62	0.58	6.12-9.89	8.00	2.67	34.2-67.6	50.90	23.6	6.80-8.45	7.63	1.17	1.00-3.66	2.33	1.88	0.75-2.50	1.63	1.24
Location 25	0.27-0.55	0.41	0.20	7.25-5.59	6.42	1.17	89.3-94.3	91.80	3.5	9.80-12.65	11.23	2.02	5.00-8.75	6.88	2.65	2.00-10.00	6.00	5.66

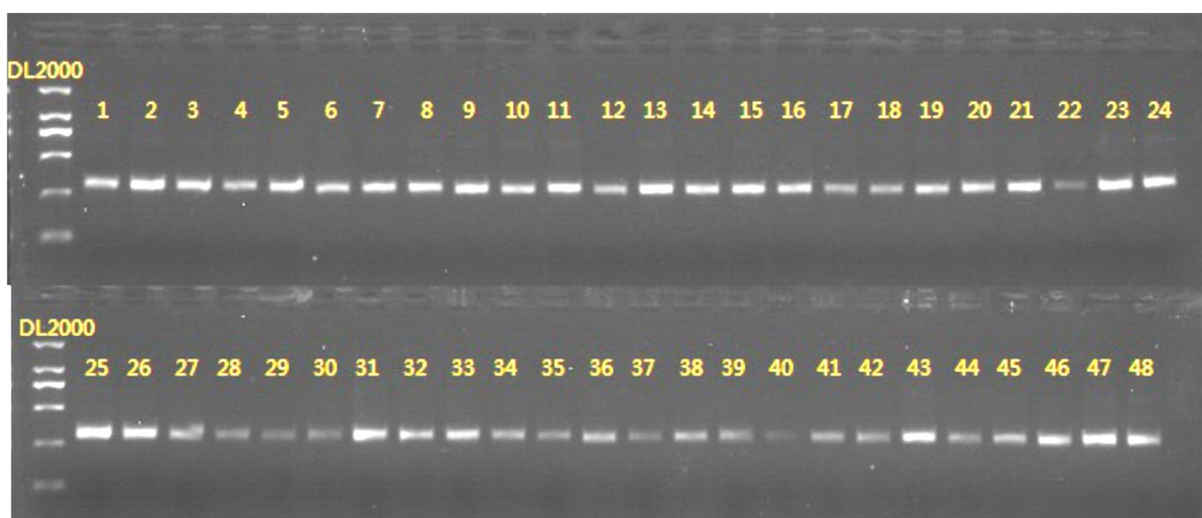


Figure S3.1A Agarose gel pictures of PCR products generated for DNA samples using 515F and 807R primers. DL2000: 100bp markers, Serial number 1-48: water or fecal DNA sample tested (Sample ID for each serial number indicated in gel is given below).

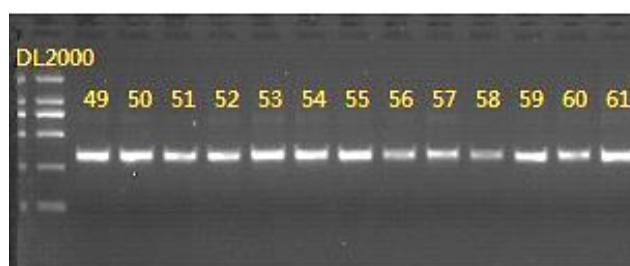


Figure S3.1B Agarose gel pictures of PCR products generated for DNA samples using 515F and 807R primers. DL2000: 100bp markers, Serial number 49-61: fecal or wastewater DNA sample tested (Sample ID for each serial number indicated in gel is given below).

Table S3.1 Description and result of quality check for water, fecal and wastewater samples.

Serial number	Sample ID	Description	Result
1	AUW01	Autumn water sample, Location-1	Grade A
2	AUW02	Autumn water sample, Location-2	Grade A
3	AUW03	Autumn water sample, Location-3	Grade A
4	AUW04	Autumn water sample, Location-4	Grade A
5	AUW05	Autumn water sample, Location-5	Grade A
6	AUW06	Autumn water sample, Location-6	Grade A
7	AUW08	Autumn water sample, Location-8	Grade A
8	AUW10	Autumn water sample, Location-10	Grade A
9	AUW12	Autumn water sample, Location-12	Grade A
10	AUW13	Autumn water sample, Location-13	Grade A
11	AUW14	Autumn water sample, Location-14	Grade A
12	AUW15	Autumn water sample, Location-15	Grade A
13	AUW16	Autumn water sample, Location-16	Grade A
14	AUW20	Autumn water sample, Location-20	Grade A
15	AUW21	Autumn water sample, Location-21	Grade A
16	WIW01	Winter water sample, Location-1	Grade A
17	WIW02	Winter water sample, Location-2	Grade A
18	WIW03	Winter water sample, Location-3	Grade A
19	WIW04	Winter water sample, Location-4	Grade A

Serial number	Sample ID	Description	
20	WIW05	Winter water sample, Location-5	Grade A
21	WIW06	Winter water sample, Location-6	Grade A
22	WIW08	Winter water sample, Location-8	Grade A
23	WIW10	Winter water sample, Location-10	Grade A
24	WIW12	Winter water sample, Location-12	Grade A
25	WIW13	Winter water sample, Location-13	Grade A
26	WIW14	Winter water sample, Location-14	Grade A
27	WIW15	Winter water sample, Location-15	Grade A
28	WIW16	Winter water sample, Location-16	Grade A
29	WIW20	Winter water sample, Location-20	Grade A
30	WIW21	Winter water sample, Location-21	Grade A
31	SUW01	Summer water sample, Location-1	Grade A
32	SUW02	Summer water sample, Location-2	Grade A
33	SUW03	Summer water sample, Location-3	Grade A
34	SUW04	Summer water sample, Location-4	Grade A
35	SUW05	Summer water sample, Location-5	Grade A
36	SUW06	Summer water sample, Location-6	Grade A
37	SUW08	Summer water sample, Location-8	Grade A
38	SUW10	Summer water sample, Location-10	Grade A
39	SUW12	Summer water sample, Location-12	Grade A
40	SUW13	Summer water sample, Location-13	Grade A
41	SUW14	Summer water sample, Location-14	Grade A
42	SUW15	Summer water sample, Location-15	Grade A
43	SUW16	Summer water sample, Location-16	Grade A
44	SUW20	Summer water sample, Location-20	Grade A
45	SUW21	Summer water sample, Location-21	Grade A
46	HU-1	Human fecal sample	Grade A
47	HU-2	Human fecal sample	Grade A
48	CW-1	Cow fecal sample	Grade A
49	CW-2	Cow fecal sample	Grade A
50	DG-1	Dog fecal sample	Grade A
51	DG-2	Dog fecal sample	Grade A
52	PG-1	Pig fecal sample	Grade A
53	PG-2	Pig fecal sample	Grade A
54	DU-1	Duck fecal sample	Grade A
55	DU-2	Duck fecal sample	Grade A
56	GO-1	Goose fecal sample	Grade A
57	GO-2	Goose fecal sample	Grade A
58	CK-1	Chicken fecal sample	Grade A
59	CK-2	Chicken fecal sample	Grade A
60	WW-1	Wastewater sample	Grade A
61	WW-2	Wastewater sample	Grade A

Grade A: The PCR product size is correct and the concentration is suitable for subsequent experiments.

Grade B: The PCR product size is correct and the concentration is low, and subsequent experiments can be tried.

Grade C: The PCR product is too weak or not detected, and subsequent experiments cannot be performed.

Table S3.2 Detailed diversity index values for autumn and winter season water samples.

Sample ID	Reads	OTU	ace	chao1	Shannon	Simpson	coverage
AUW01	18640	619	838	837	4.73	0.019	0.989646
AUW02	18640	686	1194	1034	4.37	0.0419	0.986481
AUW03	18640	700	1128	981	4.55	0.0354	0.987393
AUW04	18640	781	1487	1218	4.64	0.0232	0.983423
AUW05	18640	908	1697	1445	4.84	0.0225	0.980472
AUW06	18640	931	1855	1521	4.85	0.019	0.979292
AUW08	18640	811	1462	1233	4.73	0.0192	0.983047
AUW10	18640	710	1534	1232	4.66	0.0195	0.984174
AUW12	18640	751	1416	1169	4.32	0.0384	0.984227
AUW13	18640	949	1628	1383	4.54	0.039	0.980418
AUW14	18640	937	1738	1529	5.04	0.0146	0.980258
AUW15	18640	982	1370	1401	5.23	0.0118	0.981921
AUW16	18640	892	1569	1349	4.45	0.0483	0.981545
AUW20	18640	891	1826	1475	4.65	0.0269	0.979399
AUW21	18640	1109	1617	1609	5.11	0.017	0.977843
WIW01	18640	734	1327	1127	4.54	0.0318	0.984925
WIW02	18640	564	1095	875	3.46	0.1275	0.987929
WIW03	18640	696	1195	1017	3.97	0.0869	0.986105
WIW04	18640	875	1590	1323	4.75	0.0249	0.982082
WIW05	18640	705	1389	1146	4.15	0.0475	0.984442
WIW06	18640	999	1788	1487	4.67	0.0394	0.978916
WIW08	18640	913	1280	1247	4.15	0.083	0.983047
WIW10	18640	957	1352	1386	4.42	0.0627	0.981652
WIW12	18640	977	1348	1285	4.94	0.0224	0.982725
WIW13	18640	1143	1588	1566	5.03	0.0279	0.978916
WIW14	18640	1049	1543	1605	4.84	0.029	0.978433
WIW15	18640	965	1451	1444	4.39	0.0508	0.979238
WIW16	18640	1261	1851	1832	4.58	0.0678	0.973712
WIW20	18640	1154	1621	1607	5.25	0.0149	0.97838
WIW21	18640	1292	1839	1807	5.22	0.0194	0.974893

Table S3.2 Detailed diversity index values for fecal, wastewater and summer season water samples.

Sample ID	Reads	OTU	ace	chao1	Shannon	Simpson	coverage
SUW01	18640	763	1269	1118	4.87	0.017	0.985461
SUW02	18640	905	1843	1598	4.75	0.0215	0.979024
SUW03	18640	803	1490	1229	4.67	0.0227	0.983101
SUW04	18640	823	1618	1351	4.64	0.0218	0.981599
SUW05	18640	798	1448	1199	4.51	0.0288	0.983047
SUW06	18640	800	1485	1190	4.49	0.0306	0.982833
SUW08	18640	732	1510	1295	4.42	0.0283	0.982994
SUW10	18640	700	1355	1104	4.49	0.025	0.984871
SUW12	18640	703	1355	1063	4.72	0.0178	0.985354
SUW13	18640	644	1349	1036	4.38	0.0325	0.985891
SUW14	18640	780	1450	1170	4.72	0.0217	0.983906
SUW15	18640	1001	1986	1695	4.94	0.0228	0.977629
SUW16	18640	994	1731	1447	5	0.0216	0.979936
SUW20	18640	930	1885	1558	4.6	0.033	0.97838
SUW21	18640	875	1573	1285	4.66	0.0254	0.981545
CK1	18640	363	486	461	3.44	0.0608	0.993884
CK2	18640	371	522	490	3.42	0.0647	0.99324
CW1	18640	646	707	708	5.13	0.0156	0.994957
CW2	18640	608	663	666	5.12	0.0144	0.995279
DG1	18640	147	218	189	3.3	0.0614	0.998069
DG2	18640	138	168	185	3.34	0.0601	0.998337
DU1	18640	220	287	296	3.05	0.1028	0.996674
DU2	18640	238	370	324	3.24	0.0816	0.996191
GO1	18640	450	561	580	4.16	0.0324	0.993509
GO2	18640	418	528	546	4.09	0.0342	0.993777
HU1	18640	223	274	284	2.97	0.1533	0.997264
HU2	18640	224	263	295	3.52	0.0781	0.997586
PG1	18640	556	628	626	4.38	0.0349	0.994528
PG2	18640	640	742	770	4.39	0.0388	0.992436
WW1	18640	1186	1459	1490	5.56	0.011	0.982725
WW2	18640	1174	1429	1433	5.61	0.0108	0.983906

Table S3.3 Relative abundance of important bacterial genera (most abundant, fecal or potential pathogenic) observed in Tiaoxi River water samples.

Taxon	AUW01	AUW02	AUW03	AUW04	AUW05	AUW06	AUW08	AUW10	AUW12	AUW13	AUW14	AUW15	AUW16	AUW20	AUW21
Acinetobacter	0.003	0.022	0.004	0.057	0.032	0.019	0.034	0.003	0.014	0.005	0.000	0.015	0.047	0.074	0.051
Aeromonas	0.003	0.022	0.034	0.042	0.062	0.014	0.013	0.003	0.014	0.000	0.000	0.011	0.065	0.014	0.077
Arcobacter	0.000	0.011	0.008	0.042	0.158	0.139	0.008	0.006	0.018	0.005	0.000	0.000	0.759	0.014	0.000
Bacteroides	0.000	0.000	0.000	0.042	2.403	1.401	0.382	0.112	0.121	0.381	0.042	0.228	0.307	0.000	0.003
Bifidobacterium	0.000	0.000	0.000	0.004	0.037	0.053	0.013	0.000	0.000	0.014	0.000	0.004	0.000	0.000	0.000
Blautia	0.000	0.000	0.000	0.000	0.072	0.086	0.013	0.003	0.004	0.010	0.004	0.011	0.007	0.000	0.000
Brevundimonas	0.000	0.000	0.000	0.000	0.003	0.024	0.017	0.003	0.000	0.014	0.084	0.106	0.007	0.004	0.003
CL500-29_marine_group	5.405	14.087	7.488	11.115	3.586	4.971	7.162	5.438	20.898	3.782	4.312	2.418	4.105	5.179	8.159
Clostridium_sensu_stricto_1	0.003	0.000	0.000	0.008	0.011	0.024	0.004	0.010	0.007	0.010	0.013	0.015	0.018	0.021	0.029
Comamonadaceae_unclassified	5.278	4.072	3.672	9.513	8.168	8.102	10.081	9.990	3.645	4.096	5.367	3.804	8.889	3.290	3.133
Corynebacterium_1	0.000	0.000	0.000	0.000	0.003	0.000	0.004	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000
Cyanobacteria_norank	8.442	10.999	12.241	11.414	14.839	11.974	7.275	2.820	19.037	29.487	10.974	9.305	21.528	3.870	1.825
Dechloromonas	0.003	0.026	0.017	0.379	0.335	0.139	0.021	0.038	0.068	0.033	0.017	0.049	0.723	0.035	0.170
Dorea	0.000	0.000	0.000	0.000	0.048	0.019	0.004	0.000	0.000	0.005	0.000	0.008	0.000	0.000	0.000
Enterococcus	0.000	0.000	0.000	0.000	0.051	0.053	0.038	0.013	0.007	0.000	0.000	0.038	0.000	0.081	0.064
Escherichia-Shigella	0.000	0.007	0.000	0.015	0.407	0.196	0.088	0.032	0.029	0.090	0.004	0.046	0.361	0.000	0.013
Faecalibacterium	0.000	0.000	0.004	0.000	0.005	0.019	0.000	0.000	0.000	0.000	0.000	0.008	0.011	0.000	0.000
Flavobacterium	0.499	0.453	0.959	0.103	0.171	0.330	0.470	0.757	0.089	0.504	3.251	5.668	0.614	9.417	3.095
Fluviicola	0.488	0.281	1.649	2.615	0.731	0.784	1.168	3.388	1.300	0.452	0.897	0.873	1.098	3.884	2.245
hgcI_clade	12.330	14.397	23.254	16.411	10.268	12.868	11.615	11.635	15.121	12.677	11.393	5.797	21.365	18.449	14.436
Lactobacillus	0.000	0.000	0.000	0.004	0.016	0.010	0.000	0.000	0.011	0.005	0.000	0.011	0.007	0.000	0.013
Macellibacteroides	0.000	0.004	0.000	0.004	0.027	0.005	0.000	0.000	0.000	0.005	0.000	0.000	0.043	0.000	0.000
Microcystis	0.904	0.004	0.013	0.004	0.000	0.000	0.000	0.003	0.257	1.475	1.500	0.847	0.842	0.046	0.000
Prevotella_9	0.000	0.000	0.000	0.015	0.083	0.014	0.000	0.003	0.007	0.005	0.004	0.000	0.271	0.011	0.006
Pseudobutyrivibrio	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Roseburia	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sediminibacterium	1.733	1.250	1.632	1.063	1.504	2.643	3.550	5.450	0.121	1.346	2.187	2.802	0.177	0.794	1.539
Sporichthyaceae_unclassified	7.100	3.960	6.773	6.833	4.608	5.989	6.876	8.200	1.428	3.953	7.106	5.030	1.257	8.343	7.806
Streptococcus	0.000	0.000	0.000	0.004	0.013	0.033	0.025	0.000	0.004	0.000	0.004	0.004	0.004	0.000	0.000
Turicibacter	0.003	0.000	0.000	0.000	0.005	0.000	0.004	0.006	0.004	0.005	0.000	0.004	0.000	0.000	0.003
Weissella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000

Table S3.3 Relative abundance of important bacterial genera (most abundant, fecal or potential pathogenic) observed in Tiaoxi River water samples.

Taxon	WIW01	WIW02	WIW03	WIW04	WIW05	WIW06	WIW08	WIW10	WIW12	WIW13	WIW14	WIW15	WIW16	WIW20	WIW21
Acinetobacter	0.021	0.096	0.134	0.125	0.074	0.049	0.017	0.036	0.025	0.028	0.062	0.172	0.051	0.010	0.093
Aeromonas	0.009	0.008	0.018	0.164	1.334	0.066	0.004	0.004	0.057	0.049	0.204	0.159	0.051	0.020	0.075
Arcobacter	0.027	0.019	0.014	0.210	0.283	0.027	0.004	0.011	0.035	0.017	0.065	0.045	0.159	0.027	0.117
Bacteroides	0.003	2.669	2.339	1.134	0.129	0.668	0.265	0.167	0.076	0.003	0.023	0.026	0.040	0.000	0.145
Bifidobacterium	0.000	0.077	0.049	0.036	0.005	0.013	0.004	0.004	0.000	0.000	0.004	0.000	0.004	0.000	0.005
Blautia	0.000	0.108	0.113	0.029	0.000	0.018	0.000	0.022	0.005	0.000	0.004	0.003	0.004	0.000	0.009
Brevundimonas	3.968	0.181	0.367	0.285	0.035	0.062	0.058	0.040	0.248	0.073	0.246	0.019	0.438	0.010	0.154
CL500-29_marine_group	1.754	0.050	0.092	0.699	0.729	0.708	0.682	1.196	0.997	1.749	0.892	1.494	0.152	0.372	0.462
Clostridium_sensu_stricto_1	0.000	0.054	0.049	0.029	0.020	0.053	0.000	0.022	0.003	0.035	0.012	0.006	0.022	0.060	0.028
Comamonadaceae_unclassified	13.345	0.861	2.844	12.648	18.488	6.527	6.531	5.629	21.956	6.724	17.684	14.633	15.790	10.780	11.656
Corynebacterium_1	0.003	0.015	0.007	0.000	0.000	0.004	0.000	0.000	0.000	0.003	0.004	0.000	0.000	0.000	0.000
Cyanobacteria_norank	2.095	34.286	31.842	23.913	14.487	23.509	29.727	28.183	5.419	17.512	13.503	20.621	26.393	1.456	9.667
Dechloromonas	0.012	0.000	0.007	0.107	0.228	0.093	0.062	0.051	0.076	0.073	0.150	0.208	0.300	0.402	0.173
Dorea	0.000	0.065	0.042	0.011	0.005	0.004	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.005
Enterococcus	0.012	0.285	0.402	0.128	0.045	0.124	0.058	0.033	0.033	0.003	0.046	0.097	0.004	0.000	0.019
Escherichia-Shigella	0.009	0.635	0.529	0.235	0.040	0.111	0.054	0.025	0.027	0.003	0.027	0.003	0.011	0.007	0.051
Faecalibacterium	0.000	0.038	0.028	0.014	0.000	0.013	0.000	0.000	0.000	0.003	0.000	0.003	0.007	0.000	0.005
Flavobacterium	1.883	0.104	0.000	4.865	20.337	3.796	2.126	2.239	4.051	2.524	4.209	3.184	2.290	2.233	3.091
Fluviicola	5.749	0.761	2.212	1.320	0.550	1.872	0.782	1.221	1.008	1.257	1.246	0.939	0.785	0.670	1.401
hgcI_clade	5.173	0.400	0.466	2.665	1.919	2.341	2.755	3.180	2.863	6.379	4.824	4.461	2.749	7.978	3.806
Lactobacillus	0.003	0.027	0.028	0.021	0.005	0.009	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.014
Macellibacteroides	0.000	0.000	0.000	0.011	0.045	0.004	0.004	0.015	0.005	0.000	0.008	0.003	0.022	0.000	0.009
Microcystis	1.022	0.462	2.812	2.172	0.015	0.018	0.004	0.004	0.188	0.031	0.919	1.046	0.832	0.285	2.601
Prevotella_9	0.003	0.015	0.000	0.064	0.020	0.022	0.000	0.004	0.005	0.000	0.092	0.023	0.051	0.000	0.047
Pseudobutyrivibrio	0.000	0.004	0.004	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.004	0.000	0.005
Roseburia	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sediminibacterium	0.066	0.000	0.007	0.121	0.213	0.642	0.418	0.472	0.264	0.698	0.246	0.305	0.130	0.686	0.280
Sporichthyaceae_unclassified	3.413	0.065	0.123	2.037	1.780	3.704	3.739	4.223	5.779	5.610	3.993	5.305	3.031	8.443	4.236
Streptococcus	0.000	0.115	0.155	0.039	0.005	0.018	0.000	0.000	0.005	0.000	0.000	0.000	0.007	0.000	0.000
Turicibacter	0.003	0.027	0.021	0.021	0.005	0.004	0.008	0.004	0.005	0.000	0.000	0.003	0.000	0.007	0.000
Weissella	0.000	0.004	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table S3.3 Relative abundance of important bacterial genera (most abundant, fecal or potential pathogenic) observed in Tiaoxi River water samples.

Taxon	SUW01	SUW02	SUW03	SUW04	SUW05	SUW06	SUW08	SUW10	SUW12	SUW13	SUW14	SUW15	SUW16	SUW20	SUW21
Acinetobacter	0.006	0.302	0.013	0.005	0.025	0.040	0.099	0.099	0.051	0.077	0.027	0.081	0.056	0.019	0.009
Aeromonas	0.021	0.033	0.118	0.073	0.094	0.020	0.052	0.084	0.319	0.077	0.023	0.044	0.192	0.023	0.018
Arcobacter	0.058	0.185	0.217	0.462	0.253	0.012	0.016	0.023	1.362	0.230	0.019	0.031	0.424	0.000	0.021
Bacteroides	0.064	0.062	0.099	0.123	0.076	0.032	0.004	0.000	0.119	0.040	0.016	0.068	0.081	0.109	0.045
Bifidobacterium	0.000	0.003	0.006	0.000	0.000	0.000	0.000	0.000	0.014	0.003	0.000	0.000	0.006	0.027	0.006
Blautia	0.000	0.007	0.003	0.000	0.000	0.004	0.000	0.000	0.000	0.003	0.000	0.010	0.017	0.008	0.000
Brevundimonas	0.006	0.007	0.006	0.005	0.004	0.004	0.008	0.005	0.007	0.003	0.000	0.010	0.008	0.000	0.003
CL500-29_marine_group	3.428	13.904	4.695	3.346	2.328	5.146	4.079	12.823	8.531	10.194	8.568	8.808	8.559	4.040	2.205
Clostridium_sensu_stricto_1	0.006	0.010	0.000	0.041	0.014	0.008	0.012	0.000	0.003	0.000	0.012	0.020	0.006	0.016	0.021
Comamonadaceae_unclassified	13.482	13.351	16.794	17.515	30.012	17.424	22.663	18.233	11.741	9.215	7.203	7.499	7.930	4.714	5.864
Corynebacterium_1	0.000	0.003	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.003	0.004	0.000
Cyanobacteria_norank	9.793	3.173	0.927	1.056	0.851	1.814	1.483	1.056	0.414	1.255	6.153	3.946	5.191	3.532	7.641
Dechloromonas	1.123	1.151	0.809	1.481	2.154	1.005	1.495	0.915	1.389	0.576	0.202	0.448	1.513	0.066	0.107
Dorea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.004	0.000
Enterococcus	0.006	0.003	0.003	0.005	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.088	0.011	0.709	0.223
Escherichia-Shigella	0.030	0.010	0.032	0.018	0.014	0.024	0.004	0.005	0.054	0.010	0.008	0.054	0.042	0.031	0.006
Faecalibacterium	0.000	0.003	0.006	0.000	0.007	0.000	0.000	0.000	0.003	0.000	0.000	0.003	0.000	0.000	0.003
Flavobacterium	0.107	0.163	0.230	0.206	0.228	0.388	0.123	0.141	0.282	0.077	0.000	0.044	0.170	0.101	0.369
Fluviicola	2.709	0.254	0.432	0.379	0.463	0.553	0.227	0.225	1.399	0.999	0.953	0.410	1.683	0.733	1.094
hgcI_clade	13.741	14.225	15.626	13.474	8.506	22.510	16.806	11.843	14.169	24.513	20.909	21.439	15.018	27.283	20.209
Lactobacillus	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.019	0.003
Macellibacteroides	0.000	0.013	0.003	0.032	0.022	0.004	0.000	0.000	0.058	0.027	0.004	0.000	0.050	0.004	0.006
Microcystis	0.356	0.026	0.032	0.023	0.004	0.004	0.000	0.000	0.000	0.003	0.058	0.024	0.011	0.000	0.042
Prevotella_9	0.003	0.023	0.080	0.032	0.011	0.016	0.004	0.000	0.082	0.010	0.008	0.003	0.045	0.012	0.006
Pseudobutyrivibrio	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.006	0.004	0.000
Roseburia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sediminibacterium	1.017	1.876	2.757	2.646	2.559	2.559	1.964	1.839	1.841	2.377	1.159	0.590	0.396	1.904	2.178
Sporichthyaceae_unclassified	4.125	5.718	5.987	7.322	6.921	7.781	6.318	7.263	9.139	10.048	8.564	7.713	4.369	11.491	9.186
Streptococcus	0.003	0.000	0.000	0.000	0.000	0.004	0.000	0.005	0.010	0.000	0.004	0.007	0.014	0.000	0.000
Turicibacter	0.000	0.003	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.004	0.000
Weissella	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table S3.3 Relative abundance of important bacterial genera (most abundant, fecal or potential pathogenic) observed in fecal and wastewater samples.

Taxon	CK_1	CK_2	CW_1	CW_2	DG_1	DG_2	DU_1	DU_2	GO_1	GO_2	HU_1	HU_2	PG_1	PG_2	WW_1	WW_2
Acinetobacter	0.094	0.275	0.038	0.000	0.029	0.009	4.294	18.764	3.357	2.516	0.003	0.000	0.000	0.006	0.444	0.713
Aeromonas	0.000	0.000	0.000	0.000	0.006	0.003	0.000	0.000	0.025	0.016	0.007	0.007	0.000	0.003	1.057	1.418
Arcobacter	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	2.060	2.302	0.000	0.000	0.000	0.000	3.636	4.962
Bacteroides	0.114	0.415	2.375	3.259	14.261	11.038	0.034	0.023	8.681	7.602	2.105	2.699	0.160	0.231	1.271	1.604
Bifidobacterium	0.038	0.033	0.015	0.000	0.586	0.664	0.038	0.051	0.003	0.005	1.222	1.713	0.020	0.044	0.012	0.004
Blautia	0.080	0.033	0.030	0.024	11.913	12.101	0.004	0.009	0.003	0.000	2.938	5.251	0.210	0.272	0.003	0.023
Brevundimonas	0.264	0.173	0.000	0.000	0.000	0.000	0.008	0.005	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CL500-29_marine_group	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.008
Clostridium_sensu_stricto_1	5.703	3.383	1.684	1.402	2.206	2.613	0.011	0.023	0.092	0.102	0.281	0.425	13.557	14.928	0.015	0.011
Comamonadaceae_unclassified	1.100	0.748	0.004	0.003	0.000	0.000	0.241	0.138	0.633	0.612	0.000	0.000	0.005	0.003	2.763	2.691
Corynebacterium_1	0.062	0.132	0.169	0.106	0.006	0.000	0.124	0.777	3.523	3.036	0.000	0.000	0.150	0.219	0.003	0.000
Cyanobacteria_norank	1.270	4.172	0.068	0.000	0.006	0.000	0.320	0.391	0.184	0.177	0.036	0.018	0.005	2.610	1.385	0.568
Dechloromonas	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.016	0.000	0.000	0.000	0.000	5.802	5.660
Dorea	0.003	0.008	0.143	0.112	0.006	0.000	0.000	0.000	0.000	0.005	0.598	1.198	0.050	0.066	0.009	0.008
Enterococcus	4.905	2.610	0.474	0.009	2.354	2.355	5.224	4.016	3.923	4.458	0.271	0.079	0.010	0.031	0.012	0.000
Escherichia-Shigella	9.165	6.486	0.045	0.024	0.386	0.572	1.404	1.021	0.135	0.113	0.775	1.101	0.145	0.353	0.129	0.137
Faecalibacterium	0.073	0.045	0.008	0.006	0.229	0.327	0.004	0.005	0.015	0.016	2.958	3.444	0.085	0.053	0.055	0.034
Flavobacterium	0.146	0.107	0.000	0.000	0.000	0.003	0.478	0.446	1.144	1.030	0.000	0.000	0.000	0.003	0.144	0.133
Fluviicola	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.019
hgcI_clade	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000
Lactobacillus	42.559	56.131	0.026	0.006	22.158	15.667	0.930	4.182	15.005	12.275	0.007	0.004	10.366	14.179	0.031	0.027
Macellibacteroides	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	1.153	1.271	0.000	0.000	0.000	0.000	1.161	1.433
Microcystis	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
Prevotella_9	0.000	0.012	0.000	0.003	7.002	12.667	0.000	0.000	0.031	0.011	49.289	32.620	1.111	0.481	0.279	0.316
Pseudobutyrivibrio	0.000	0.000	0.669	0.745	0.000	0.000	0.000	0.000	0.000	0.005	4.128	6.341	0.135	0.231	0.015	0.004
Roseburia	0.000	0.000	0.154	0.142	0.000	0.000	0.000	0.000	0.000	0.000	1.817	2.991	0.010	0.037	0.003	0.004
Sporichthyaceae_unclassified	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008
Streptococcus	0.354	0.156	0.004	0.003	2.841	3.437	0.222	0.253	2.469	2.854	0.206	0.320	9.955	10.797	0.012	0.008
Turicibacter	8.159	5.306	0.395	0.225	0.158	0.151	0.290	0.230	0.098	0.134	0.072	0.122	0.760	0.977	0.003	0.011
Weissella	0.090	0.115	0.000	0.000	0.000	0.000	0.060	0.051	4.593	4.405	0.029	0.014	0.005	0.000	0.000	0.000

Table S3.4 Number of OTUs assigned to fecal associated or potential pathogenic bacteria detected in autumn season water samples using RDP Classifier.

OTU ID	AUW01	AUW02	AUW03	AUW04	AUW05	AUW06	AUW08	AUW10	AUW12	AUW13	AUW14	AUW15	AUW16	AUW20	AUW21
Arcobacter cryaerophilus	0	1	0	0	12	4	0	0	2	0	0	0	68	1	0
Bacteroides caccae	0	0	0	0	51	33	7	1	3	12	3	9	2	0	0
Bacteroides graminisolvens	0	0	0	1	3	0	0	1	1	0	0	0	12	0	1
Bacteroides massiliensis	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
Bacteroides nordii	0	0	0	0	17	12	8	1	1	1	1	2	0	0	0
Bacteroides paurosaccharolyticus	0	0	0	1	2	0	0	0	0	0	0	0	1	0	0
Bacteroides plebeius	0	0	0	0	6	5	1	0	1	1	0	2	7	0	0
Bacteroides propionicifaciens	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0
Bacteroides stercoris	0	0	0	2	7	4	0	0	1	1	0	1	2	0	0
Brevundimonas vesicularis	0	0	0	0	1	3	2	0	0	0	0	1	0	0	0
Parabacteroides merdae	0	0	0	0	17	6	3	2	1	3	1	2	0	0	0

Table S3.4 Number of OTUs assigned to fecal associated or potential pathogenic bacteria detected in winter season water samples using RDP Classifier.

OTU ID	WIW01	WIW02	WIW03	WIW04	WIW05	WIW06	WIW08	WIW10	WIW12	WIW13	WIW14	WIW15	WIW16	WIW20	WIW21
Arcobacter cryaerophilus	0	1	0	5	9	0	0	0	1	1	4	2	10	1	3
Bacteroides caccae	0	57	51	18	0	14	5	6	0	0	0	1	0	0	4
Bacteroides graminisolvens	0	1	1	2	2	1	1	0	0	0	1	1	2	0	2
Bacteroides massiliensis	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
Bacteroides nordii	0	26	20	7	0	2	0	4	0	0	0	0	0	0	2
Bacteroides paurosaccharolyticus	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Bacteroides plebeius	0	11	5	6	3	0	2	0	0	0	0	0	1	0	2
Bacteroides propionicifaciens	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Bacteroides stercoris	0	11	4	5	2	2	1	0	1	0	1	2	2	0	5
Brevundimonas vesicularis	3	3	4	2	0	2	8	6	4	6	2	1	0	0	0
Parabacteroides merdae	0	21	25	6	1	7	2	1	0	0	0	0	0	0	1

Table S3.4 Number of OTUs assigned to fecal associated or potential pathogenic bacteria detected in summer season water samples using RDP Classifier.

OTU ID	SUW01	SUW02	SUW03	SUW04	SUW05	SUW06	SUW08	SUW10	SUW12	SUW13	SUW14	SUW15	SUW16	SUW20	SUW21
Arcobacter cryaerophilus	2	9	12	17	19	0	0	1	50	21	2	1	27	0	0
Bacteroides caccae	0	0	2	1	0	1	0	0	0	0	0	1	0	1	0
Bacteroides graminisolvens	0	1	0	5	5	0	0	0	5	3	1	1	2	0	0
Bacteroides massiliensis	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
Bacteroides nordii	0	0	0	0	1	0	1	0	0	0	0	0	0	4	0
Bacteroides paurosaccharolyticus	0	0	3	1	1	0	0	0	0	1	0	0	1	0	0
Bacteroides plebeius	1	3	1	2	2	1	0	0	0	1	1	0	1	0	0
Bacteroides propionicifaciens	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bacteroides stercoris	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Brevundimonas vesicularis	0	1	0	1	1	1	0	0	1	1	0	0	1	0	1
Parabacteroides merdae	1	1	1	1	1	1	0	0	0	0	0	1	0	1	0

Table S3.4 Number of OTUs assigned to fecal associated or potential pathogenic bacteria detected in fecal or wastewater samples using RDP Classifier.

OTU ID	CK1	CK2	CW1	CW2	DG1	DG2	DU1	DU2	GO1	GO2	HU1	HU2	PG1	PG2	WW1	WW2
Arcobacter cryaerophilus	0	0	0	0	0	0	0	0	402	429	0	0	0	0	412	552
Bacteroides caccae	1	7	0	0	0	0	1	0	5	1	11	14	0	5	0	1
Bacteroides graminisolvens	0	0	0	0	0	0	0	0	1052	926	0	0	0	0	129	179
Bacteroides massiliensis	0	0	0	0	3	1	0	0	0	0	3	6	0	0	2	3
Bacteroides nordii	0	3	0	0	0	0	0	2	4	0	1	0	0	3	1	0
Bacteroides paurosaccharolyticus	0	0	0	0	0	0	0	0	66	39	0	0	0	0	4	12
Bacteroides plebeius	0	1	0	0	8	7	0	0	1	8	111	175	0	1	9	7
Bacteroides propionicifaciens	0	0	0	0	0	0	1	1	153	159	0	0	0	0	0	0
Bacteroides stercoris	1	0	0	0	185	193	0	0	0	0	22	16	0	1	15	6
Brevundimonas vesicularis	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Parabacteroides merdae	7	4	0	1	1	0	0	0	2	0	10	21	0	2	4	2

Supplementary Note S4.1 Fecal sampling and composite sample preparation:

Individual pig fecal samples were collected from pig farms located at Changxing Zhicheng (N30°59′26.53″; E119°53′24.67″) and Wuxing Daishan (N30°53′24.92″; E120°11′59.15″) areas of Zhejiang Province. Individual Dog fecal samples were collected from different pet stores located in Hong Qiao (N31°33′45.68″; E120°20′20.19″) of Zhejiang Province. Individual chicken, duck and goose samples were collected from poultry farms located at Cao Jian Duan Village (N30°54′9.94″; E120°18′40.81″), Baishui (N30°54′19.78″; E119°49′17.86″) and Lijiagang (N31°02′18.59″; E119°51′39.48″) areas of Zhejiang province. Individual cow fecal samples were collected from slaughterhouses at Changxing Zhicheng (N30°59′26.53″; E119°53′24.67″) of Zhejiang Province. Individual human fecal samples were 10 different individuals of age 16-40 years. Primary effluents were collected on six different occasions from a WWTP located in Suzhou (N31°17′37.54″; E120°34′10.00″), Jiangsu province. For preparation of composites for duck and goose samples, approximately 0.5gms of individual fecal samples (>11 samples) of a host (Goose or duck) were pooled together to form a composite sample of respective host. The composite samples were used for further DNA extractions.

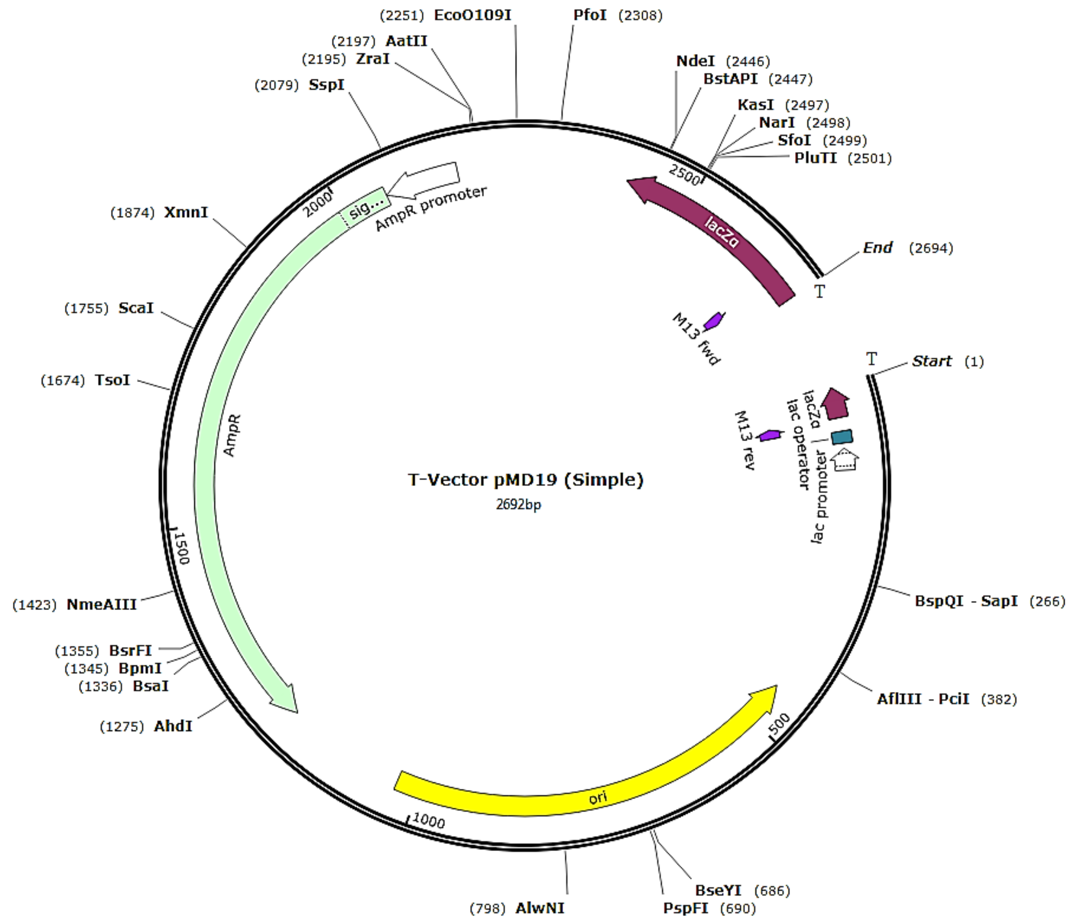


Figure S4.1 Schematic representation of pMD19 TA cloning vector with restriction sites. Ori: origin of replication, AmpR: Ampicillin resistance gene, lacZ: lacZ gene interrupted by T-overhang cloning site.

Table S4.1 Confirmation for absence of PCR inhibitors in Human fecal samples with BacUni assay.

Samples	Ct value for 1:10 dilution	Ct value for 1:100 dilution	Coefficient of variation (CV) %
H-1	18.32	22.30	0.41
H-2	18.53	21.99	0.16
H-3	17.84	22.04	0.52
H-4	23.31	27.69	0.59
H-5	16.04	20.02	0.41
H-6	10.44	14.20	0.31
H-7	13.41	17.27	0.36
H-8	10.62	14.38	0.31
H-9	12.26	16.29	0.44
H-10	20.56	23.98	0.14

Table S4.2. The detailed performance characteristics (range) of all the qPCR assays.

Assay	Slope	R^2 value	Efficiency (%)
BacUni	-3.2 to -3.32	0.996 to 0.997	100.0 to 105.8
GenBac	-3.26 to - 3.27	0.994 to 0.996	101.8 to 102.4
HF183 Taqman	-3.32 to -3.35	0.996 to 0.999	98.16 to 99.7
BacHum	-3.25 to -3.29	0.994 to 0.998	101 to 103.9
HF183 SYBR	-3.26 to -3.30	0.997 to 0.999	100.7 to 102.3
Hum2	-3.22 to -3.36	0.994 to 0.996	98.4 to 105
Pig-2-Bac	-3.27 to -3.35	0.997 to 0.998	98.5 to 102
BacCow	-3.31 to -3.42	0.995 to 0.997	95.9 to 100.3
AV4143	-3.5 to -3.56	0.99 to 0.998	91.1 to 93.3
GFD	-3.27 to 3.44	0.997 to 0.999	95.2 to 102

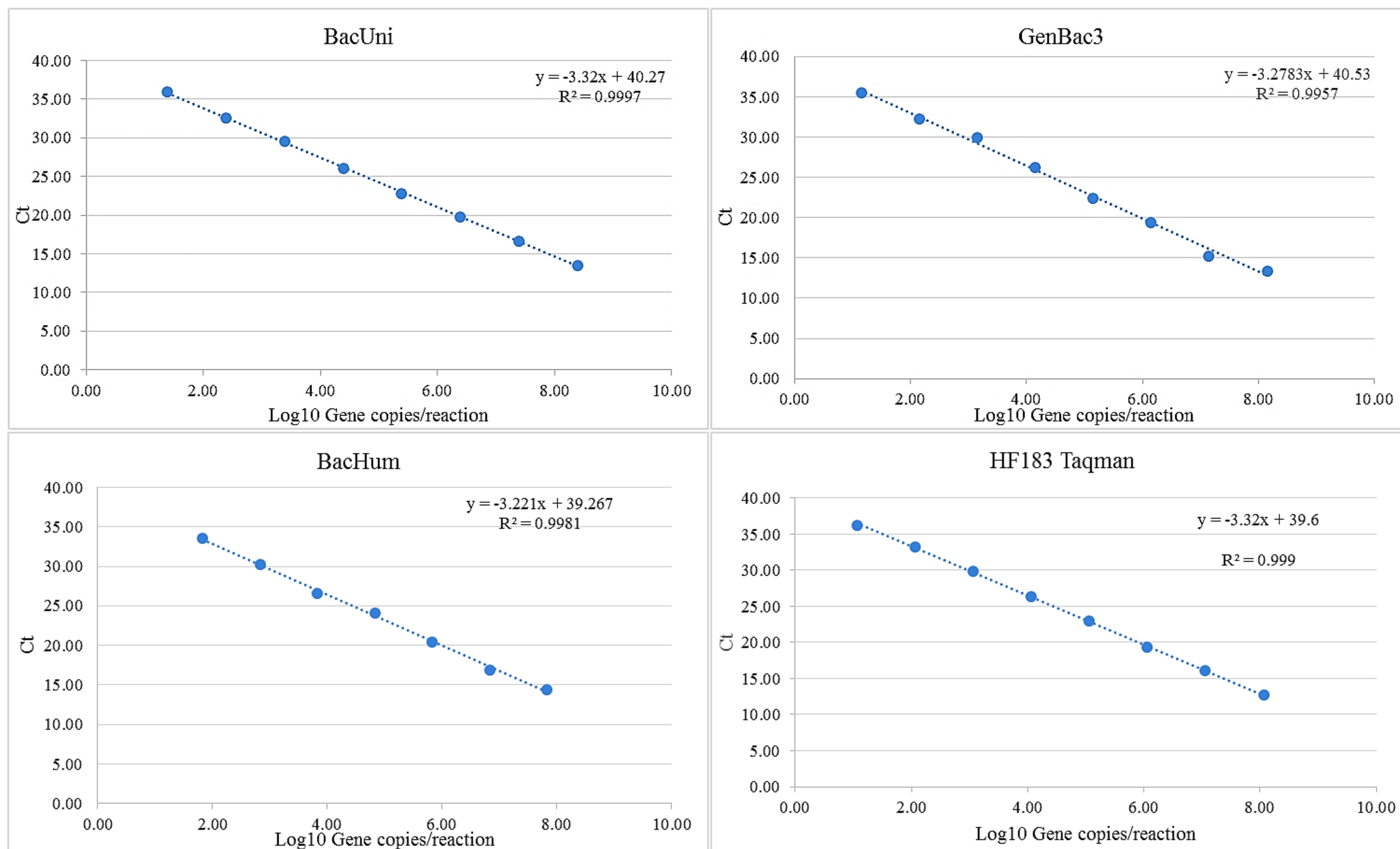
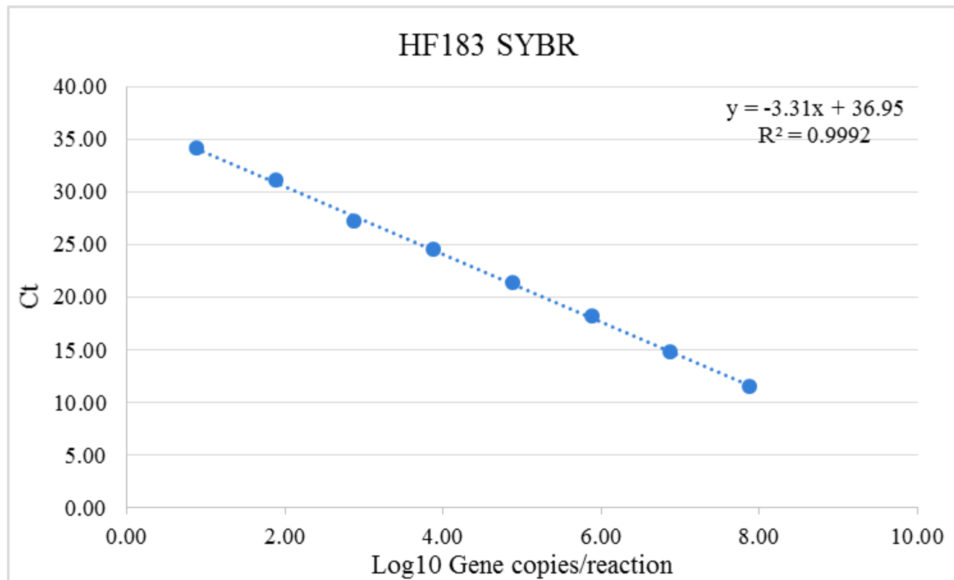
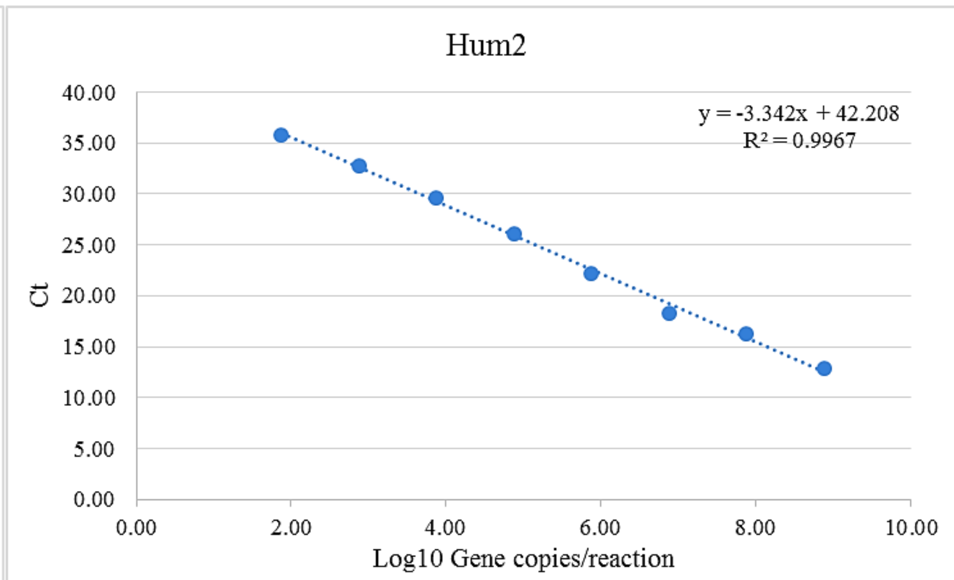
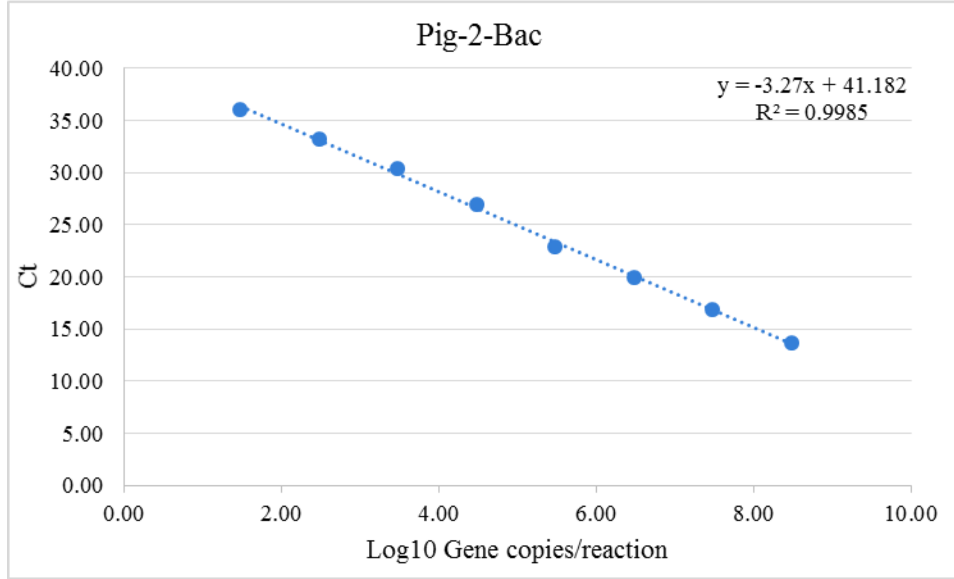


Fig.S4.1A Standardized master standard curve for MST assays: Universal Bacteroidales (BacUni, GenBac3), Human associated (BacHum, HF183 Taq) assays







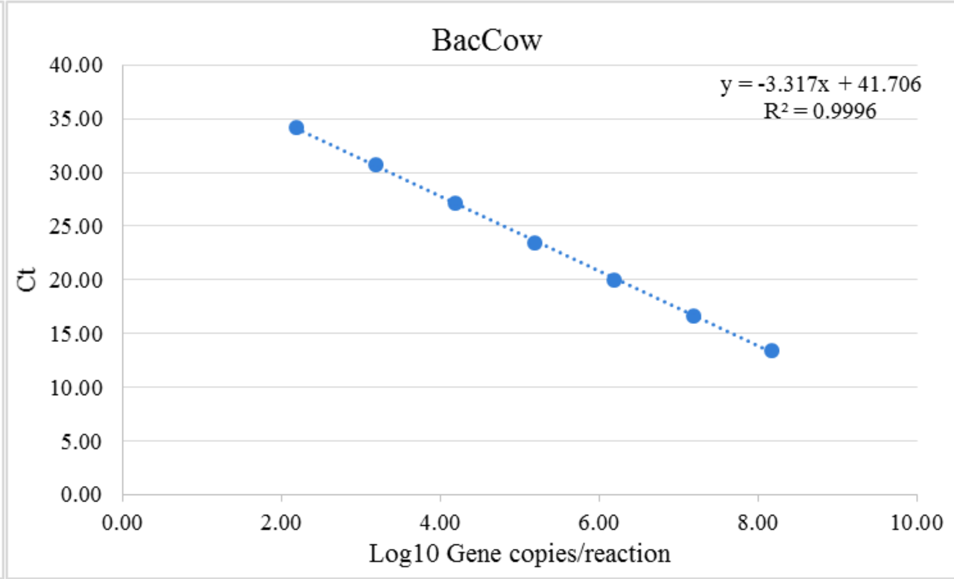


Fig.S4.1B Standardized master standard curve for MST assays: Human (HF183 SYBR, Hum2), swine (Pig-2-Bac) and livestock associated (BacCow) assays

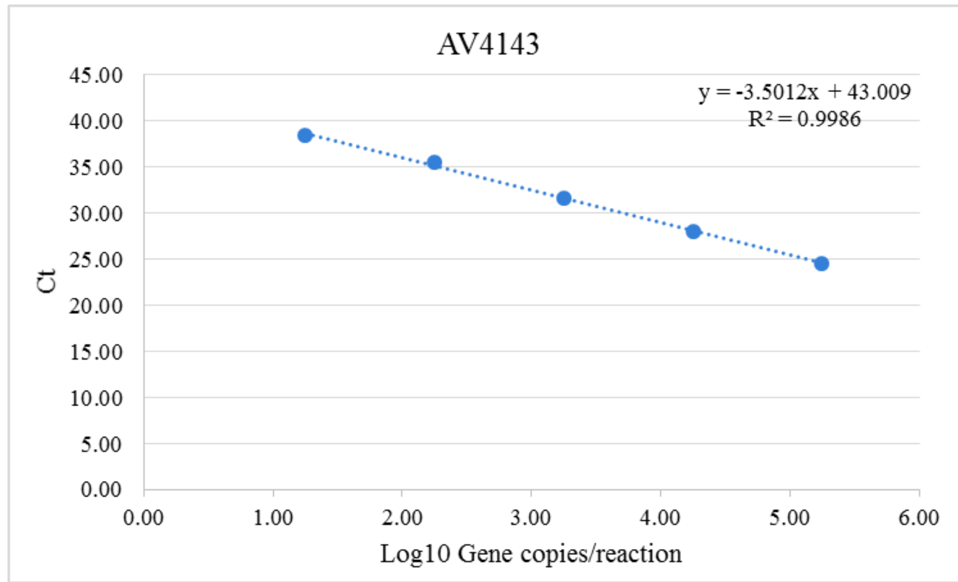


Fig.S4.1C Standardized master standard curve for Avian AV4143 MST assays.

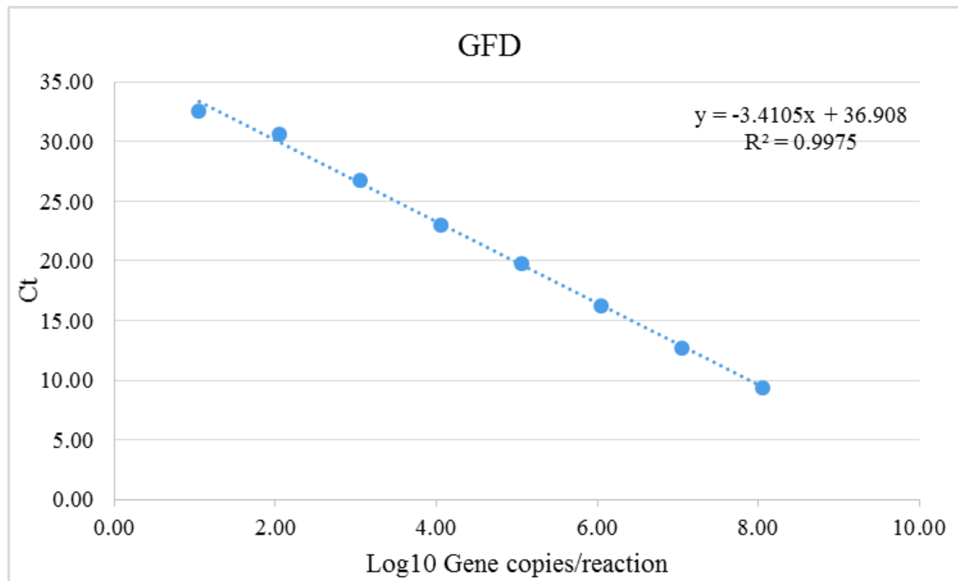


Fig.S4.1D Standardized master standard curve for Avian GFD MST assays.

Table S4.3A Normalized qPCR results for BacUni assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
HF-1	17.87	9969989	10.32	966084.2
HF-2	18.53	6123175	11.68	524244.4
HF-3	17.62	11965760	6.4	1869650
HF-4	23.19	202054.6	2.44	82809.25
HF-5	16.04	37966246	8.2	4630030
HF-6	10.44	2.30E+09	61.6	37323397
HF-7	13.41	2.60E+08	10.92	23844304
HF-8	10.62	2.01E+09	59.16	33949307
HF-9	12.26	6.06E+08	53.8	11263160
HF-10	20.56	1381838	4.24	325905.2
			Mean	9.11E+06
			Mean Log10	6.959518
Sew-1	16.48	27592901	97.2	283877.6
Sew-2	16.33	30658296	103.64	295815.3
Sew-3	16.27	32097229	125.52	255714.1
Sew-4	15.88	42806989	130.2	328778.7
Sew-5	17.51	12923909	103.88	124411.9
			Mean	230696
			Mean Log10	5.363
Dg-1	19.88	2279801	5.8	393069.2
Dg-2	25.6	34524.9	4.4	7846.567
Dg-3	28.1	5559.593	1.28	4343.432
Dg-4	19.54	2915760	8.96	325419.7
Dg-5	23.81	127945.7	5.12	24989.4
Dg-6	23.14	209987.2	3.6	58329.78
Dg-7	24.53	76039.6	2.36	32220.17
Dg-8	24.54	75469.66	7.56	9982.759
Dg-9	24.1	103469	5.8	17839.48
Dg-10	23.02	228438	2.68	85238.04
			Mean	17782.79
			Mean Log10	4.25

Table S4.3A Normalized qPCR results for BacUni assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Pg-1	23.53209	157324.1	6.56	23982.34
Pg-2	20.36221	1602571	7.64	209760.6
Pg-3	19.43734	3154466	13.44	234707.3
Pg-4	18.91039	4639765	16.92	274217.8
Pg-5	20.12624	1904828	8.84	215478.2
Pg-6	19.35824	3342564	10.68	312974.1
Pg-7	19.25645	3601214	13.92	258707.9
Pg-8	17.50967	12939722	16.48	785177.3
Pg-9	18.70834	5379574	10.92	492635
Pg-10	20.187	1821940	8.48	214851.4
			Mean	245924
			Mean Log10	5.3908
Cw-1	18.50	6248728	8.56	729991.6
Cw-2	17.72	11093013	12.96	855942.4
Cw-3	17.99	9104505	11.12	818750.4
Cw-4	18.91	4648436	10.84	428822.5
Cw-5	18.86	4825872	6.76	713886.3
Cw-6	18.71	5368124	10.88	493393.7
Cw-7	18.07	8599842	7.72	1113969
Cw-8	17.50	13001359	12.32	1055305
Cw-9	19.75	2515858	5.92	424976
Cw-10	19.46	3109837	8.88	350206.8
			Mean	302249.2
			Mean Log10	5.393651
Ck-1	30.93	697.761	3.68	189.609
Ck-2	26.74	15052.39	2.2	6841.994
Ck-3	26.58	16904.49	4	4226.122
Ck-4	31.29	536.1745	2.8	191.4909
Ck-5	29.90	1489.379	1.68	886.5351
Ck-6	29.39	2154.077	1.6	1346.298
Ck-7	30.76	790.3176	3.88	203.6901
Ck-8	23.37	177101.5	2.6	68115.98
Ck-9	28.12	5478.404	9.28	590.3453
Ck-10	29.44	2077.827	2.04	1018.543
			Mean	7274.3
			Mean Log10	3.8618

Table S4.3A Normalized qPCR results for BacUni assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Go-1	17.85	10106566	13.04	775043.4
Go-2	18.25	7500891	9.12	822466.1
Go-3	18.3	7279173	17.28	421248.4
			Mean	672919.3
			Mean Log10	5.827963
Du-1	24.38	84725.53	2.96	28623.49
Du-2	23.98	113569.4	25.48	4457.199
Du-3	23.97	114232.5	6.4	17848.82
			Mean	16976.5
			Mean Log10	4.229848

Table S4.3B Normalized qPCR results for GenBac3 assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
HF-1	19.91672	993845.6	10.32	91962.77
HF-2	20.42835	729737.9	11.68	56815.46
HF-3	20.28812	830992.7	6.4	114371.8
HF-4	25.6144	18787.86	2.44	7236.282
HF-5	19.17645	1748803	8.2	194221.3
HF-6	12.56791	1.69E+08	61.6	2627521
HF-7	16.1425	17707214	10.92	1217028
HF-8	12.80404	1.54E+08	59.16	2319451
HF-9	15.15588	29066762	53.8	492465.5
HF-10	22.70599	130151.3	4.24	31829.11
			Mean	715290.2
			Mean Log10	5.91
Sew-1	19.26651	1738526	97.2	15384.86
Sew-2	19.28086	1943105	103.64	14284.89
Sew-3	19.22315	1653461	125.52	12280.49
Sew-4	19.22266	1618692	130.2	11843.12
Sew-5	19.25615	1647356	103.88	14500.27
			Mean	13658.73
			Mean Log10	4.19

Table S4.3B Normalized qPCR results for GenBac3 assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
PI-1	25.25697	23449.9	6.56	3455.847
PI-2	22.33795	184647.1	7.64	22849.19
PI-3	21.46402	387859.2	13.44	23931.97
PI-4	20.95633	546898.1	16.92	27112.02
PI-5	22.4407	173693.9	8.84	18378.35
PI-6	21.27174	351325.8	10.68	34451.05
PI-7	21.22514	470261.6	13.92	27307.78
PI-8	19.97712	1040931	16.48	55206.83
PI-9	20.98568	518555.9	10.92	41155.26
PI-10	22.25741	190745.1	8.48	21778.57
			Mean	27562.69
			Mean Log10	4.440322
CK-1	30.87891	620.6648	3.68	120.8449
CK-2	28.64624	2449.321	2.2	963.1974
CK-3	28.96644	1784.459	4	423.4798
CK-4	30.83603	483.5309	2.8	163.659
CK-5	30.40692	718.2353	1.68	368.2226
CK-6	30.26552	738.4081	1.6	426.8157
CK-7	30.41188	598.7581	3.88	158.8842
CK-8	26.35772	10165.37	2.6	4038.189
CK-9	29.80449	1069.376	9.28	101.5844
CK-10	30.21149	841.5911	2.04	347.6488
			Mean	711.2526
			Mean Log10	2.852024
Cw-1	19.88041	3287283	8.56	113722.5
Cw-2	19.80479	3486195	12.96	79191.78
Cw-3	20.37262	2269026	11.12	62048.41
Cw-4	21.10569	1304987	10.84	38121.79
Cw-5	20.84742	1584392	6.76	73230.29
Cw-6	20.54687	1993179	10.88	56141.87
Cw-7	20.79168	1667789	7.72	66672.86
Cw-8	20.39838	2228350	12.32	55004.73
Cw-9	21.55309	932113	5.92	51051.24
Cw-10	21.46426	999421.4	8.88	36215.28
			Mean	63140.08
			Mean Log10	4.800305

Table S4.3B Normalized qPCR results for GenBac3 assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Dg-1	22.80296	362309.6	5.8	21742.72
Dg-2	28.16595	6331.176	4.4	673.8274
Dg-3	28.71439	4188.899	1.28	1578.442
Dg-4	22.7716	372569.6	8.96	14386.59
Dg-5	25.82025	37204.78	5.12	2986.217
Dg-6	25.82508	37040.29	3.6	4232.734
Dg-7	26.20794	27743.21	2.36	4940.113
Dg-8	24.61798	92351.22	7.56	4688.106
Dg-9	26.15587	28885.99	5.8	2084.658
Dg-10	25.10013	63994.72	2.68	9439.658
			Mean	6675.307
			Mean Log10	3.824471

Table S4.4A Normalized qPCR results for HF183Taqman assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
H-1	27.70961	4457.197	10.32	431.8989
H-2	36.67342	10.25684	11.68	0.878154
H-3	Undetermined	0	6.4	0
H-4	27.4856	5187.858	2.44	2126.171
H-5	32.06174	233.4698	8.2	28.47193
H-6	34.91648	33.73492	61.6	
H-7	21.88154	231347	10.92	21185.62
H-8	35.6557	20.44224	59.16	
H-9	15.60139	16312183	53.8	303200.4
H-10	25.9325	14861.6	4.24	3505.094
			Mean	33047.86
			Mean Log10	4.519143
R-1	23.98239	55716.37	97.2	573.2136
R-2	22.94991	112161.1	103.64	1082.218
R-3	24.51417	38858.05	125.52	309.5765
R-4	23.76085	64741.58	130.2	497.2471
R-5	24.71755	33855.17	103.88	325.9065
			Mean	557.6324
			Mean Log10	2.746348
PI-1	Undetermined	0	16.4	0
PI-2	Undetermined	0	19.1	0
PI-3	Undetermined	0	33.6	0
PI-4	36.13846	14.73841	42.3	0.348426
PI-5	36.46095	11.84518	22.1	0.535981
PI-6	Undetermined	0	26.7	0
PI-7	Undetermined	0	34.8	0
PI-8	Undetermined	0	41.2	0
PI-9	Undetermined	0	27.3	0
PI-10	Undetermined	0	21.2	0

Table S4.4A Normalized qPCR results for HF183Taqman assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
CK-1	37.98047	4.230075	3.68	1.14
CK-2	30.94072	499.0518	2.2	226.84
CK-3	32.26527	203.3906	4	50.84
CK-4	38.1917	3.665918	2.8	1.30
CK-5	25.66446	17821.7	1.68	10608
CK-6	32.90628	131.7291	1.6	82.33
CK-7			3.88	
CK-8	31.85766	268.0968	2.6	103.11
CK-9	29.77588	1098.9	9.28	118.41
CK-10	32.62491	159.4004	2.04	78.13
			Mean	1402.43
			Mean Log10	3.22
Dg-1	Undetermined	22.19253	5.8	0
Dg-2	35.35321	63.77336	4.4	5.70
Dg-3	34.82946	31.26203	1.28	27.95
Dg-4	35.96917	23.4196	8.96	1.84
Dg-6	34.76284	14.37812	5.12	7.31
Dg-7	Undetermined	24.63728	3.6	0
Dg-8	34.56477	32.45933	2.36	18.14
Dg-8	Undetermined	26.54006	7.56	0
Dg-9	Undetermined	41.589	5.8	0
Dg-10	Undetermined	17.62012	2.68	0
			Mean	19.4587
			Mean Log10	1.289114
Cw-1	Undetermined	0	8.56	0
Cw-2	Undetermined	0	12.96	0
Cw-3	Undetermined	0	11.12	0
Cw-4	34.98201	32.26967	10.84	2.97
Cw-5	Undetermined	0	6.76	0
Cw-6	Undetermined	0	10.88	0
Cw-7	35.694	19.91851	7.72	2.58
Cw-8	35.39976	24.31366	12.32	1.97
Cw-9	36.37638	12.54382	5.92	2.11
Cw-10	35.32625	25.55555	8.88	2.87

Table S4.4A Normalized qPCR results for HF183Taqman assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Go-1	35.18	28.21799	13.04	2.16
Go-2	36.38	12.51313	9.12	1.37
Go-3	35.88	17.55966	17.28	1.01
Du-1	35.99	16.29833	2.96	5.50
Du-2	35.26	26.72896	25.48	1.04
Du-3	35.88	17.55966	6.4	2.74

Table S4.4B Normalized qPCR results for BacHum assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
H-1	27.56228	6519.678	10.32	631.751
H-2	36.1857	58.71541	11.68	5.027
H-3	36.50873	49.21855	6.4	7.690
H-4	27.14693	8179.909	2.44	3352.42
H-5	32.00788	575.0819	8.2	70.131
H-6	31.85858	623.9417	61.6	10.12
H-7	22.83475	86217.91	10.92	7895.41
H-8	29.91566	1803.043	59.16	30.47
H-9	15.67388	4306813	53.8	80052.28
H-10	25.9734	15527.77	4.24	3662.20
			Mean	9571.75
			Mean Log10	3.980
R-1	24.70478	31047.46	97.2	319.4183
R-2	24.6314	32317.05	103.64	311.8202
R-3	24.78232	29760.08	125.52	237.0944
R-4	25.09661	25065.96	130.2	192.5189
R-5	25.04409	25795.4	103.88	248.3192
			Mean	261.8342
			Mean Log10	2.418026
P-1	37.49517	28.71749	16.4	1.751066
P-2	36.68573	44.68326	19.1	2.339438
P-3	37.2166	33.43666	33.6	0.995139
P-4	38.0034	21.75678	42.3	0.514345
P-5	Undetermined		22.1	0
P-6	Undetermined	0	26.7	0
P-7	Undetermined	0	34.8	0
P-8	Undetermined	0	41.2	0
P-9	Undetermined	0	27.3	0
P-10	Undetermined	0	21.2	0

Table S4.4B Normalized qPCR results for BacHum assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Ck-1	36.1951	58.41487	3.68	15.87361
Ck-2	31.93817	597.4026	2.2	271.5466
Ck-3	31.58739	723.5546	4	180.8887
Ck-4	35.2	100.5912	2.8	35.92543
Ck-5	33.98239	195.6036	1.68	116.4307
Ck-6	32.02557	569.5527	1.6	355.9704
Ck-7			3.88	0
Ck-8	28.78822	3337.594	2.6	1283.69
Ck-9	39.53803	9.409763	9.28	1.013983
Ck-10	30.2127	1533.021	2.04	751.4808
			Mean	332.994
			Mean Log10	2.522436
Cw-1	36.10709	61.29122	8.56	7.160189
Cw-2	34.76732	127.4064	12.96	9.830737
Cw-3	33.34525	277.0167	11.12	24.91158
Cw-4	Undetermined	0	10.84	0
Cw-5	34.64015	136.5705	6.76	20.20273
Cw-6	33.42395	265.3616	10.88	24.38985
Cw-7	34.83159	123.0119	7.72	15.93418
Cw-8	32.74911	383.6293	12.32	31.13874
Cw-9	32.80243	372.618	5.92	62.94222
Cw-10	35.01717	111.1543	8.88	12.51738
			Mean	62.94222
			Mean Log10	1.798942
Dg-1	35.59	81.29278	5.8	14.016
Dg-2	34.98	113.4342	4.4	25.7805
Dg-3	36.35	53.67598	1.28	41.93436
Dg-4	32.31	487.6048	8.96	54.42018
Dg-5	35.71	76.13564	5.12	14.87024
Dg-6	35.85	70.53099	3.6	19.59194
Dg-7			2.36	0
Dg-8	34.61	138.8379	7.56	18.3648
Dg-9	35.69	76.97186	5.8	13.27101
Dg-10	34.34	160.8985	2.68	60.03675
			Mean	101.9711
			Mean Log10	2.084772

Table S4.4B Normalized qPCR results for BacHum assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Go-1	36.78	42.44084	13.04	3.254666
Go-2	36.79	42.20967	9.12	4.628254
Go-3	35.25	97.88136	17.28	5.664431
Du-1	36.19	58.57766	2.96	19.78975
Du-2	34.36	159.1505	25.48	6.246095
Du-3	35.94	67.14783	6.4	10.49185

Table S4.5 Normalized qPCR results for Pig-2-Bac assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
P-1	29.69927	3383.407	16.4	206.3053
P-2	28.44808	8195.862	19.1	429.1027
P-3	25.62132	60491.26	33.6	1800.335
P-4	24.82464	106255.9	42.3	2511.96
P-5	32.117	612.1661	22.1	27.69983
P-6	27.62523	14665.33	26.7	549.2633
P-7	26.14039	41906.97	34.8	1204.223
P-8	24.97767	95358.09	41.2	2314.517
P-9	26.46454	33322.67	27.3	1220.611
P-10	27.62027	14716.86	21.2	694.1914
			Mean	1095.821
			Mean Log10	3.03974
H-1	34.74931	95.1674	10.32	9.221647
H-2	37.41711	14.42798	11.68	1.235272
H-3	37.07684	18.35276	6.4	2.867619
H-4	37.09207	18.15624	2.44	7.441083
H-5	38.43076	7.045551	8.2	0.859214
H-6	37.74312	11.45746	61.6	0.185998
H-7	Undetermined	0	10.92	0
H-8	38.27981	7.839186	59.16	0.132508
H-9	38.00023	9.552753	53.8	0.17756
H-10	38.09177	8.954006	4.24	2.111794

Table S4.5 Normalized qPCR results for Pig-2-Bac assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Cw-1	Undetermined	0	8.56	0
Cw-2	Undetermined	0	12.96	0
Cw-3	32.3634	514.2803	11.12	46.24823
Cw-4	32.2792	545.8303	10.84	56.35335
Cw-5	36.18755	34.41941	6.76	5.091628
Cw-6	35.83752	44.08569	10.88	4.051993
Cw-7	36.12146	36.06614	7.72	4.671779
Cw-8	35.32595	63.29957	12.32	5.137952
Cw-9	Undetermined	0	5.92	0
Cw-10	Undetermined	0	8.88	0
			Mean	51.30079
			Mean Log10	1.710124
Dg-1	39.49083	3.329401	5.8	0.574035
Dg-2	Undetermined	0	4.4	0
Dg-3	39.80171	2.672359	1.28	2.087781
Dg-4	Undetermined	0	8.96	0
Dg-5	36.0503	37.92738	5.12	7.407692
Dg-6	38.36295	7.391593	3.6	2.05322
Dg-7	Undetermined	0	2.36	0
Dg-8	Undetermined	0	7.56	0
Dg-9	Undetermined	0	5.8	0
Dg-10	36.60196	25.67663	2.68	9.580833
Ck-1	Undetermined	0	3.68	0
Ck-2	Undetermined	0	2.2	0
Ck-3	Undetermined	0	4	0
Ck-4	Undetermined	0	2.8	0
Ck-5	Undetermined	0	1.68	0
Ck-6	Undetermined	0	1.6	0
Ck-7	Undetermined	0	3.88	0
Ck-8	Undetermined	0	2.6	0
Ck-9	Undetermined	0	9.28	0
Ck-10	Undetermined	0	2.04	0

Table S4.6 Normalized qPCR results for GFD assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
CK-1	28.34796	143.4526	3.68	38.98167
CK-2	29.13884	82.33183	2.2	37.42356
CK-3	23.96371	3115.038	4	778.7596
CK-4	29.44949	66.19879	2.8	23.64242
CK-5	33.8944	2.921448	1.68	1.738957
CK-6	31.99987	11.04688	1.6	6.904299
CK-7	30.56651	30.21847	3.88	7.788266
CK-8	28.52854	126.3712	2.6	48.60432
CK-9	27.11831	340.115	9.28	36.65033
CK-10	27.08161	348.9934	2.04	171.0752
			Mean	162.1624
			Mean Log10	2.20995
H-1	29.39362	68.84706	10.32	6.671227
H-2	33.83386	3.048284	11.68	0.260983
H-3	34.24408	2.285488	6.4	0.357108
H-4	32.48035	7.8839	2.44	3.231107
H-5	Undetermined	0	8.2	0
H-6	35.55049	0.91339	61.6	0.014828
H-7	32.97448	5.572884	10.92	0.510337
H-8	35.22396	1.14872	59.16	0.019417
H-9	Undetermined	0	53.8	0
H-10	Undetermined	0	4.24	0
P-1	33.74834	3.236921	16.4	0.197373
P-2	36.8093	0.377438	19.1	0.019761
P-3	33.06635	5.224792	33.6	0.1555
P-4	33.18999	4.790387	42.3	0.113248
P-5	32.44208	8.098589	22.1	0.366452
P-6	33.87931	2.952564	26.7	0.110583
P-7	31.67419	13.88483	34.8	0.398989
P-8	33.06524	5.228866	41.2	0.126914
P-9	31.44464	16.31282	27.3	0.597539
P-10	Undetermined	0	21.2	0

Table S4.6 Normalized qPCR results for GFD assay:

Sample-ID	Ct value	Quantity	Dilution	Quantity (GC/ng)
		(Raw data)		(normalized)
Cw-1	32.55093	7.502767	8.56	
Cw-2	30.28869	36.72644	12.96	
Cw-3	30.30537	36.29892	11.12	
Cw-4	30.79982	25.65277	10.84	
Cw-5	31.4771	15.9453	6.76	
Cw-6	30.57813	29.97284	10.88	
Cw-7	31.16508	19.85023	7.72	
Cw-8	31.3578	17.33834	12.32	
Cw-9	30.66953	28.11006	5.92	
Cw-10	28.43083	135.3441	8.88	15.24145
			Mean	15.24145
			Mean Log10	1.183026
Dg-1	30.9276	23.45175	5.8	
Dg-2	30.40531	33.83919	4.4	
Dg-3	30.28034	36.94243	1.28	28.86127
Dg-4	31.11202	20.60374	8.96	
Dg-5	32.46751	7.955313	5.12	
Dg-6	29.6832	66.18141	3.6	18.38373
Dg-7	30.51287	31.37807	2.36	13.29579
Dg-8	30.87511	24.33212	7.56	
Dg-9	30.92163	23.55022	5.8	
Dg-10	32.63765	7.059609	2.68	
			Mean	20.76301
			Mean Log10	1.36165

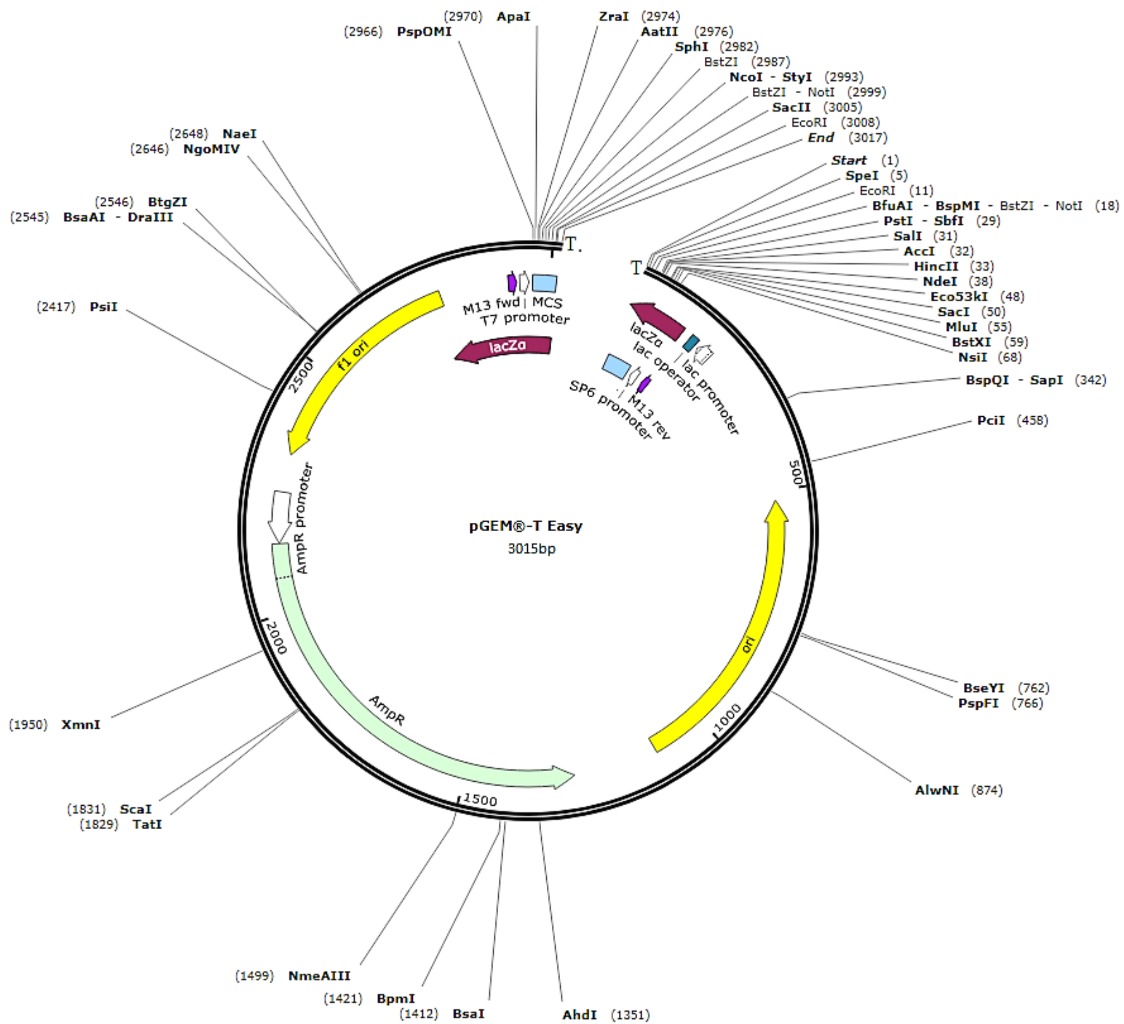


Figure S5.1 Schematic representation of pGEM®-T Easy cloning vector with restriction sites. Ori: origin of replication, AmpR: Ampicillin resistance gene, lacZ: lacZ gene interrupted by T-overhang cloning site, MCS: Multiple cloning site, T7 promoter: T7 RNA polymerase promoter, SP6 promoter: SP6 RNA polymerase promoter.

Supplementary Note S5.1:

Pathogenic bacterial gene sequences used for cloning into pGEM easy vector were given below (primer/probe sequences were marked in red).

1. *Campylobacter jejuni* mapA gene sequence (96bp):

CAATACCAGTGTCTAAAGTGCGTTTATTGGCACAACATTGAATTCCAACATCGCTAATGT
ATAAAAGCCCTTTAATCTTTGCTTCAAAACCACCAG

2. Pathogenic *Leptospira* LipL32 gene sequence (242bp):

GAACTCCCATTTCAGCGATTACAGCTGGGATCCAAACATAGAGATAGTATGCTTTTTTGT
TTCCGTCGACTAAACCGTCCGGCGCTTGTCCTGGCTTTACATATCCGTAATAGTTGATCAC
AGATCCATAGGGAAGGAACGTTTTTACGGTTTCGTTTGTTCCTGGAACTGTGCTCTCGCTC
AGAACAAAAGAGCTTTTAAGGCTTGGCAAACCACCGAACGCACCACAAGCGGTAATGCT
T

3. *Shigella* sp. ipaH gene sequence (117bp):

AGCGAAAGACTGCTGTCGAAGCTCCGCAGAGGCACTGAGTTTTTCCAGCCATGCAGCGA
CCTGTTCACGGAATCCGGAGGTATTGCGTGCAGAGACGGTATCGGAAAGGCGGTCAAG

4. *E.coli* stx2 gene sequence (68bp):

CAGGCAGATACAGAGAGAATTTCGTCTGGCACTGTCTGAAACTGCTCCTGTTTATACGAT
GACGCCGG

5. *E.coli* O157: H7 eae gene sequence (106bp):

GTAAGTTACACTATAAAAGCACCGTCGTATATGATAAAAGTGGATAAGCAAGCCTATTA
TGCTGATGCTATGTCCATTTGCAAAAATTTATTACCATCCACACAGA

Table S5.1 Standardized qPCR statistics for pathogen quantifications:

Assay	Compiled Slope	Compiled Y-intercept	Compiled R^2 value	Compiled Efficiency	LOQ (cp/µl)	LOD (cp/µl)
BacUni	-3.32	40.3	0.99	100.0	100[a]	-
HF183 Taqman	-3.37	38.7	0.99	99.7	10 [a]	-
Pig-2-Bac	-3.27	41.1	0.99	102	30 [a]	-
GFD	-3.27	36.9	0.99	96	10 [a]	-
Leptospira (LipL32)	-3.33	40.06	0.99	99.38	10 [b]	3 [b]
Campylobacter (mapA)	-3.22	39.3	0.99	104	10 [b]	3 [b]
Shigella spp. (ipaH)	-3.16	36.9	0.99	100.08	10 [b]	3 [b]
STEC (stx2)	-3.37	33.6	0.99	102.	10 [b]	3 [b]
E.coli O157: H7 eae	-3.34	34.7	0.99	98.9	10 [b]	3 [b]

LOD: Limit of Detection, LOQ: Limit of Quantification.

[a] Based on MST validation study by our group (Unpublished).

[b] Based on Oster et al. 2014.

Data Processing:

For MST assays, The Data were classified as quantifiable if two or more replicates were above LOD.

For Pathogens, The Data were classified as quantifiable (Q), detectable but not quantifiable (DNQ), and nondetectable (ND). A sample was classified as Q if two or more replicates were above the LOQ, DNQ if two or more replicates were between the LOD and LOQ, and ND if two or more replicates were below the LOD.

Table S5.2 GPS coordinates of some of the Pig and Poultry farms at location 13 and 21.

Sampling Location	potential source/farm type	Distance from sampling location	Latitude	Longitude
13	Poultry farms	<500mts	N30°51′45.67″	E120°04′10.66″
13	Pig farms	<500mts	N30°51′44.25″	E120°04′42.14″
21	Poultry farms	<2kms	N30°55′39.67″	E119°54′45.23″
21	Pig farms	<2kms	N30°55'21.0"	E119°54'51.8"

Table S5.3 Normalized qPCR results for summer season GFD assay:

Sample-ID	Ct value	Quantity (Log10 copies/ gram)
L-1	31.898	3.375
L-2	-	0
L-3	34.504	DNQ
L-4	34.273	DNQ
L-5	34.568	DNQ
L-6	31.766	3.415
L-8	34.19	DNQ
L-10	-	0
L-12	-	0
L-13	34.012	DNQ
L-14	34.136	DNQ
L-15	32.059	3.326
L-16		0
L-20	31.127	3.610
L-21	-	0

Figure S5.2 Household backyard poultry farming near Location **6 and 8.**