Thesis for the Degree of Master of Science

Quantitative real-time PCR-based Microbial Source Tracking in the Miho River, Cheongju, South Korea

APRAJITA BHANDARI

Department of Biotechnology

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

FEBRUARY 2024

Quantitative real-time PCR-based Microbial Source Tracking in the Miho River, Cheongju, South Korea

A Thesis submitted to the graduate school of

Jeju National University in partial fulfillment of

the requirements for the degree of Master of Science

under the supervision of Professor Kyung Hwan Boo

The thesis for the degree of Master of Science

by Aprajita Bhandari

has been approved by the dissertation committee.

2024.02.

Chair Tatsuya Unno _____

Member Hyeon Lee _____

Member Kyung Hwan Boo _____

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

BC: Byeongcheoncheon

BR: Borom Bridge

DNA Deoxyribonucleic acid

FIB: Fecal Indicating Bacteria

IPTG: Isopropyl β-D-1-thiogalactopyranoside

LEfSe: Linear discriminant analysis Effect Size

MCE: Mixed cellulose ester

MH: Miho River

MST: Microbial Source Tracking

NMDS: Non-metric dimensional scaling

PCR: Polymerase Chain Reaction

qPCR: Quantitative Polymerase Chain Reaction

SY: Song Yong-ri

WH: Wolhacheon

WWTP: Waste-water treatment plant

Quantitative real-time PCR-based Microbial Source Tracking in the Miho River, Cheongju, South Korea

APRAJITA BHANDARI

Department of Biotechnology

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

ABSTRACT

Fecal contamination of river systems has become a global concern due to its adverse impact on human health and environment. Consumption of water contaminated with fecal components can lead to diseases such as cholera, diarrhoea, and hepatitis. Fecal pollution can be an outcome of a wide range of sources such as effluents from waste-water treatment plants (WWTP), sewer tank leaks, run-offs from agricultural lands or livestock farms. Therefore, it is essential to identify and track the sources responsible for fecal dissemination. Microbial source tracking methods have evolved immensely in the past few decades. These involve discrimination of the fecal sources based on host-specific genetic markers.

In this study, we monitored Miho River which has been reported to be contaminated with fecal matter. Quantitative real-time PCR based microbial source tracking was applied to identify the sources with host-specific microbial markers. Furthermore, Illumina MiSeq sequencing was performed to analyse the taxonomical composition and distribution of microbial communities prior and after rainfall, and their association with fecal pollutants. Effect of rainfall was observed on the river samples along with the physicochemical parameters.

Our results identified pig sources to be the highest contributor in fecal pollution, followed by cow. Humans showed relatively low copies/16SrRNA gene copies. Elevated levels of *E.coli* were observed in post-rainfall samples for all sites, indicating rainfall can facilitate dissemination of coliforms. Relative and differential abundance analysis of microbial communities was performed along with qPCR-based methods, to find and study associations between genera and source-specific pollutants. These insights can be used to develop new biological markers for source tracking.

INTRODUCTION

In the past few decades, fecal pollution of river systems has been a serious concern worldwide as it is a potent risk to public health and the environment [15, 33]. The types of activities and development in a particular area can influence the sources and levels of fecal contamination, which, in turn, can affect water quality and pose a threat to human health [17, 20]. Dissemination of fecal pollutants into the river can be attributed to point sources such as WWTPs, industrial discharge, and landfill leachate or non-point sources such as agricultural runoffs, livestock farms runoff, and sewage leaks in human settlements. Although point sources can facilitate faster dissemination of fecal pollutants into the river, non-point sources are of much greater concern due to their widespread distribution and lack of identifiable sources. Water contaminated with human and animal fecal matter contains a plethora of microorganisms, and its consumption often leads to diseases such as cholera, diarrhoea, dysentery, typhoid, and polio [22, 25].According to a global report by WHO, 1.7 billion people consumed contaminated water in 2022, and approximately 505,000 deaths per year are estimated due to diarrhoea. [26, 27]. Remediation of fecal pollution is crucial due to its alarming and far-reaching consequences. The initial step towards its mitigation is to identify and monitor the fecal sources.

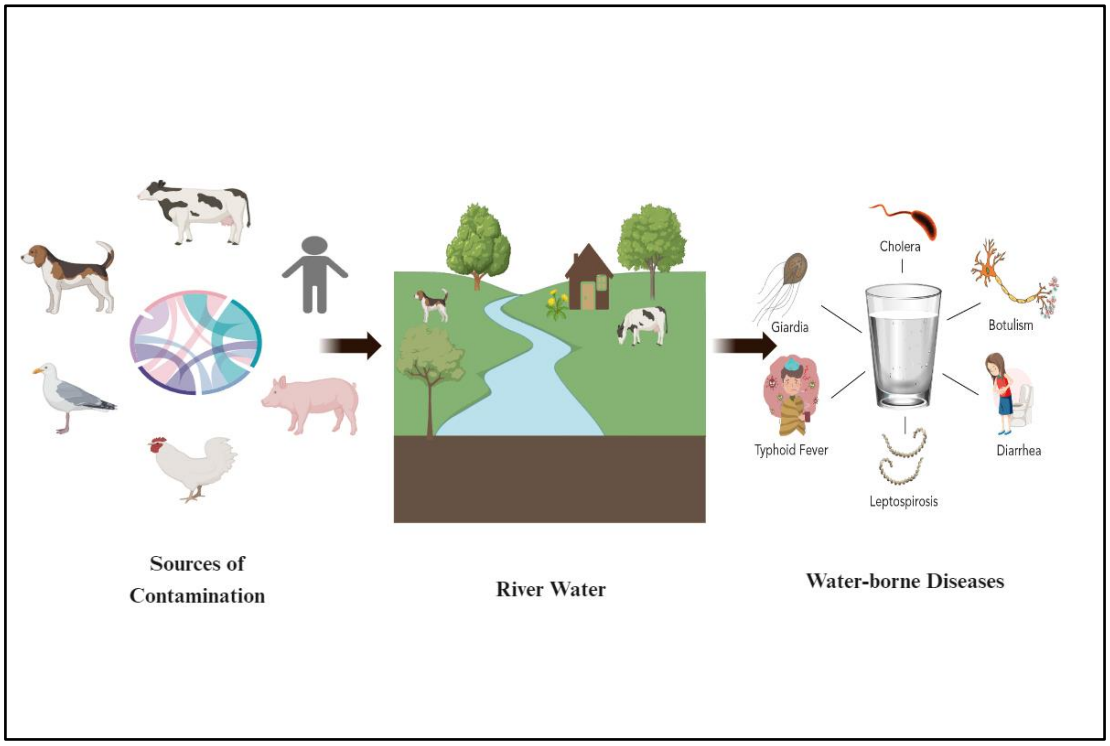
Fecal indicator bacteria (FIB) are among the widely used traditional methods for monitoring fecal pollution [3, 13, 35]. These bacteria are associated with the gut of warm-blooded animals, including humans. Therefore, the presence of total and fecal coliforms and enterococci in the river water may indicate fecal contamination. These are salient indicators as they are more persistent in the environment and accessible to detect and enumerate [30]. However, they cannot provide any information or distinguish between various sources. Microbial source tracking techniques may be used to identify the sources of fecal contamination and assess their potential risks.

Microbial source tracking is an effective technique that facilitates the identification of different sources of contamination based on specific microbial genetic markers [5, 6, 14]. It relies on the fact that certain strains of fecal bacteria are associated with specific sources, which can serve as indicative signatures of the host organism [39] These distinctive markers can be used to trace the origin of fecal contamination back to specific sources [21, 33]. To identify human-specific fecal contamination, markers such HF183 have been developed which targets 16SrRNA region of *Bacteroides* [35]. Similarly, there are markers specific to fecal sources (cow, pig, chicken, dog, gull) which are made to target 16SrRNA region of *Bacteroidales, Bacteroides-Prevotella* [44], *Brevibacterium sp., Catellicoccus marimammalium* group [4, 5, 6, 19, 21]. Quantitative real-time Polymerase Chain Reaction (qPCR) is one of the common techniques employed in MST to detect and quantify specific microbial markers indicative of the sources [36].

This study investigated seven sites in the Miho River to identify source-specific fecal pollutants through qPCR-based microbial source tracking. Many articles have reported fecal contamination of the Miho River. There are agricultural lands, human settlements, waste-water treatment plants, and livestock farms around the river. So, the contamination can be attributed to any of the sources as a result industrial discharges, agricultural runoffs, and improper waste disposal from animal farms.[18,41] Therefore, to identify which source plays a major role in the fecal pollution, microbial source tracking is essential[42]

Nowadays, many researchers believe that source tracking should be multifaceted [7]. We should not rely on a single source tracking method as their applications can be limited [31]. Many scientists have integrated library-dependent and independent methods for source tracking the pollutants [2]. We used qPCR for microbial source tracking. Additionally, we integrated physicochemical parameter analysis, effect of rainfall, and microbial community analysis through NGS based Illumina MiSeq Sequencing. Microbial community analysis

helped in evaluating taxonomic composition and shifts in their behaviour in different environments. This information can be exploited to derive associations between various sources in different environment, based on which we can develop new source tracking microbial indicators [34, 40, 43].



Fig(1): Sources of fecal contamination in a river system

METHODS

2. Sample collection and physicochemical analysis of river water

2.1 Site description

The study involved monitoring the main tributary of the Geum River, Mihocheon which originates at Eumseong and flows through Cheongju, in Chungcheongbuk-do province of South Korea. Seven sites (BC02, MH08, MH09, SY01, WH01, BR01 and MH10) spanning Miho River and its adjoining streams were selected for sampling. Table (1) shows the geographical details of the sampling sites Fig(2) shows the geographical map.

2.2 Sample collection and physicochemical analysis of the water

Water quality data of the river from recent years was retrieved from the official site of Water Environment information system. Physicochemical parameters have a significant impact on bacterial composition[10]. Physicochemical and biological parameters such as temperature, pH, DO, TOC, BOD, COD, TN, TP, electrical conductivity, and total coliforms were studied along with rainfall. Fig(3) shows correlation matrix between physicochemical parameters, rainfall, and total coliforms sites (a)BC02 and (b) MH10, it was observed that rainfall had positive correlation with total coliforms* indicating its impact on fecal pollution. Hence, sampling was conducted two times in the month of June, prior and subsequent a major rainfall event. Water samples were collected in triplicates using 500ml sterile bottles from seven sites across the river. The samples were labelled accordingly and transported to the laboratory in an ice box. Physicochemical parameters such as pH, temperature, dissolved oxygen, and electrical conductivity were also measured using a multifunction meter (Elmetron, CX-401, Zarbe, Poland). The multifunction meter was calibrated before measurement.

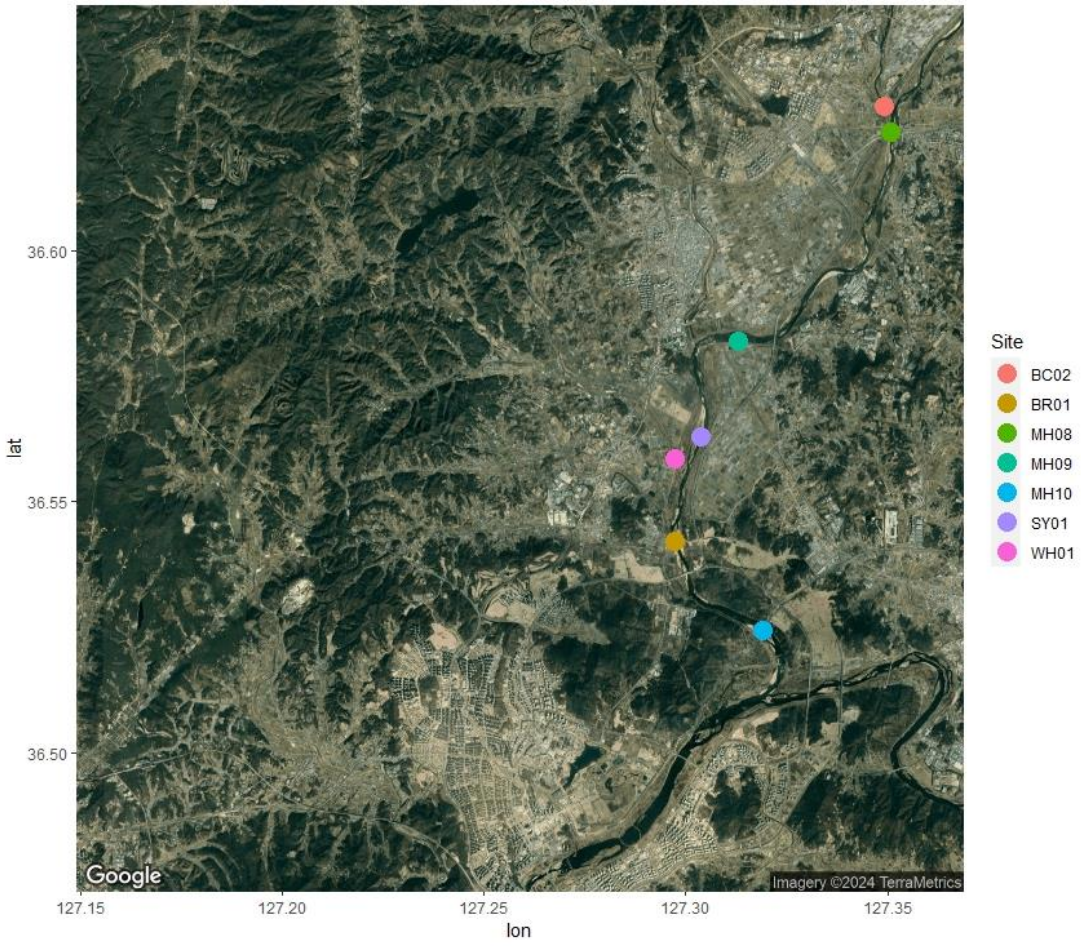
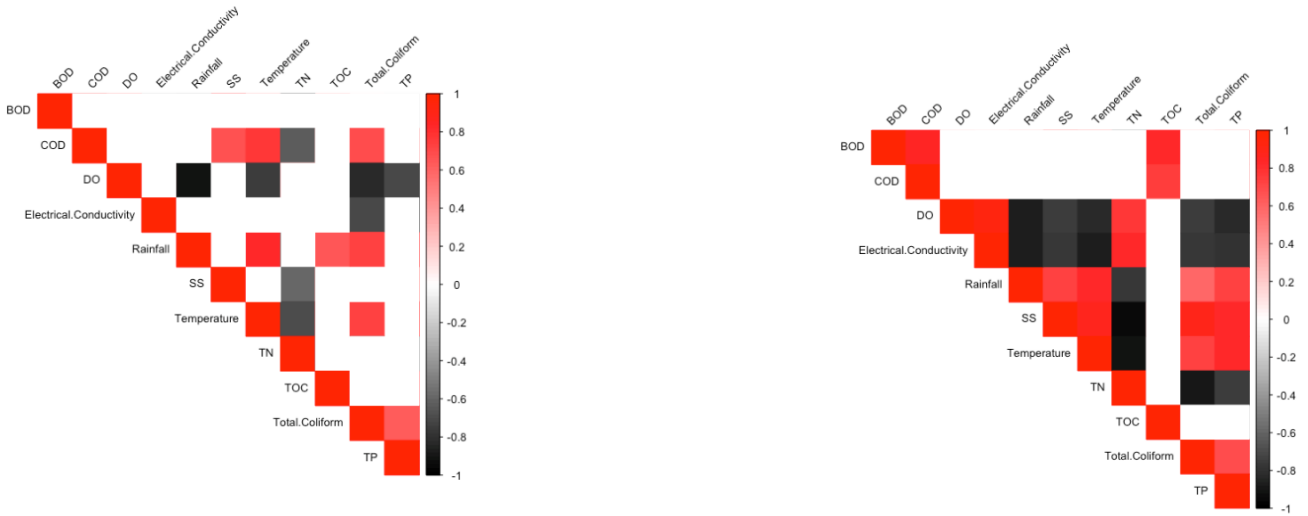


Fig 2: Geographical map showing sampling sites

Table 1: Geographical location of the sampling sites.

S.No.	Site	Site_ID	Address	Latitude	Longitude
1	Miho River	MH08	365, Tabyeon-ri, Gangnae-myeon, Heungdeok-gu	36.62	127.35
2	Byeongcheoncheon	BC02	544-1, Gungpyeong-ri, Osong-eup, Heungdeok-gu	36.63	127.35
3	Miho River	MH09	775-4, Yeyang-ri, Yeondong-myeon	36.58	127.31
4	Wolhacheon	WH01	227-1, Wolha-ri, Yeonseo-myeon	36.56	127.30
5	Borom Bridge	BR01	654-15, Sejong-dong	36.54	127.30
6	Song Yong -ri	SY01	890-27, Songyong-ri, Yeondong-myeon	36.56	127.30
7	Miho River	MH10	10-25, Hapgang-dong	36.52	127.32



(a) (b)

Fig. 3 Correlation matrix between physicochemical parameters, rainfall, and total coliforms (2022)
(a) BC02 and (b) MH10

2.3 Sample preparation and DNA extraction

River samples were prepared prior DNA extraction, in which 400 ml of each sample was vacuum filtered through a 0.22μm pore size MCE membrane filter (ADVENTEC, A020A047A) using a filtration device. The filter paper was placed in a sterile Eppendorf tube for further analysis. DNeasy Power Water Kit (Qiagen, Hilden, Germany) was used for DNA isolation and manufacturer's instructions were followed accordingly. Spectrophotometer/ Fluorometer (DeNovix, DS-11 FX+, DE, USA) was used to measure the DNA quality. Qubit fluorometer (Invitrogen, CA, USA) was used to measure the concentration of the samples**.**

2.4 Quantitative real-time PCR

Host-specific primers targeting the 16SrRNA of *Bacteroides* and *Bacteroides-Prevotella* group for human, cow and pig were selected and received from Macrogen, Seoul, South Korea. For *E.coli*, primers targeting the *uidA* gene were used. Table.2 shows details of primers used in our study.

Primers were optimized using Maxime™ PCR PreMix (i-StarTaq) and gradient PCR was performed on Bio-Rad Thermal cycler T100. For cloning, hot start PCR was performed, followed by gel electrophoresis on 1.5% Agarose Gel. The PCR product was then cloned into DH5α competent cells (BioFACT™) using Mighty TA Cloning kit (Takara Tokyo, Japan), After blue-white screening, colony PCR was performed to check the target size. Plasmids were extracted using AccuPrep® Plasmid Mini Extraction Kit (Bioneer, South Korea). Quantification of the extracted plasmid was performed using a Qubit fluorometer (Invitrogen, CA, USA). Extracted plasmids were then sent to Macrogen, Seoul, South Korea for Sanger's Sequencing.

Standard curve was set using ten-fold dilutions of the plasmid standards. All river samples were quantified using THUNDERBIRD™ SYBR™ qPCR Mix (Toyobo, Osaka, Japan),

Nuclease free water and forward and reverse primer on a Thermal Cycler Dice Real System (Takara, Tokyo, Japan). Three-step PCR was performed for human while for cow and pig two-step PCR was performed. The output data was further analysed to calculate the copy number. The final copy number were normalized with 16SrRNA gene copies, so the copy number was expressed as no.of copies/16SrRNA gene copies.

Table 2: List of primers used in this study.

Type	Target	Primer	Primer Sequence (5′-3′)	Size	References
Human-specific	*Bacteroides*	HF183F	ATCATGAGTTCACATGTCCG	82	[35, 44]
		HF183R	TACCCCGCCTACTATCTAATG		
Cow-specific	*Bacteroides cluster I*	qCS406F	GAAGGATGAAGGTTCTATGGATTGT	150	[28]
	Bacteroides–Prevotella	qBac581R	CGCTCCCTTTAAACCCAATAAA		
Pig-specific	*Prevotella cluster I*	qPS422F	CGGGTTGTAAACTGCTTTTATGAAG	150	[28]
	Bacteroides–Prevotella	qBac581R	CGCTCCCTTTAAACCCAATAAA		
E. coli	*uidA*	URL-301	TGTTACGTCCTGTAGAAAGCCC	153	[4, 24]
		URR-432	AAAACTGCCTGGCACAGCAATT		
16SrRNA	*v3-4*	338 F	ACTCCTACGGGAGGCAGCAG	55	[7]
		518 R	ATTACCGCGGCTGCTGG		

2.5 Illumina MiSeq Sequencing

Illumina MiSeq sequencing was performed to analyse the microbial communities in river samples. A 16S metagenomic library was prepared which followed a two-step PCR method. Initially, the v3-v4 hypervariable region of the 16SrRNA was amplified using a KAPA HiFi HotStart ReadyMix (PCR) kit (Roche, CA, USA) under the following conditions: 95 ℃ for 3 min, 25 cycles at 95 ℃ for 30 s. PCR products were purified using HiAccuBead. In the second PCR, unique index primers provided by Illumina were attached and amplified at 55 ℃ for 30 s, 72 ℃ for 30 s, and 72 ℃ for 5 min. Post cleanup, the concentration of all the samples was measured through a Qubit fluorometer (Invitrogen, CA, USA). Samples were pooled together and sent to Macrogen (Macrogen, Seoul, South Korea) for sequencing.

2.6 Data Analysis

Raw MiSeq data was processed with MOTHUR software (mothurv.1.48.0.) as per MiSeq SOP [32]. First, for quality check low-quality reads were trimmed, and the sequences were aligned using a SILVA DB-based file as a reference. Chimeric sequences were removed and after which OTUs were clustered based on 97% similarity. Further, sequences were binned into phylotypes according to their taxonomic classification ranging from the genus to the phylum level (1-5). For α-diversity analysis, rarefaction curves were generated, and with Chao-Shannon indices, diversity and richness were compared in before and after rainfall samples The non-metric multidimensional scaling (NMDS) was performed using braycurtis distances for investigating β-diversity [8]. Further, differential abundance in before and after rainfall samples was analysed using linear discrimination analysis effect size (LEfSe).

Statistical analysis such as ttest and Spearman's correlation analysis was performed on Excel and R(v4.2.2) was used for data visualization.

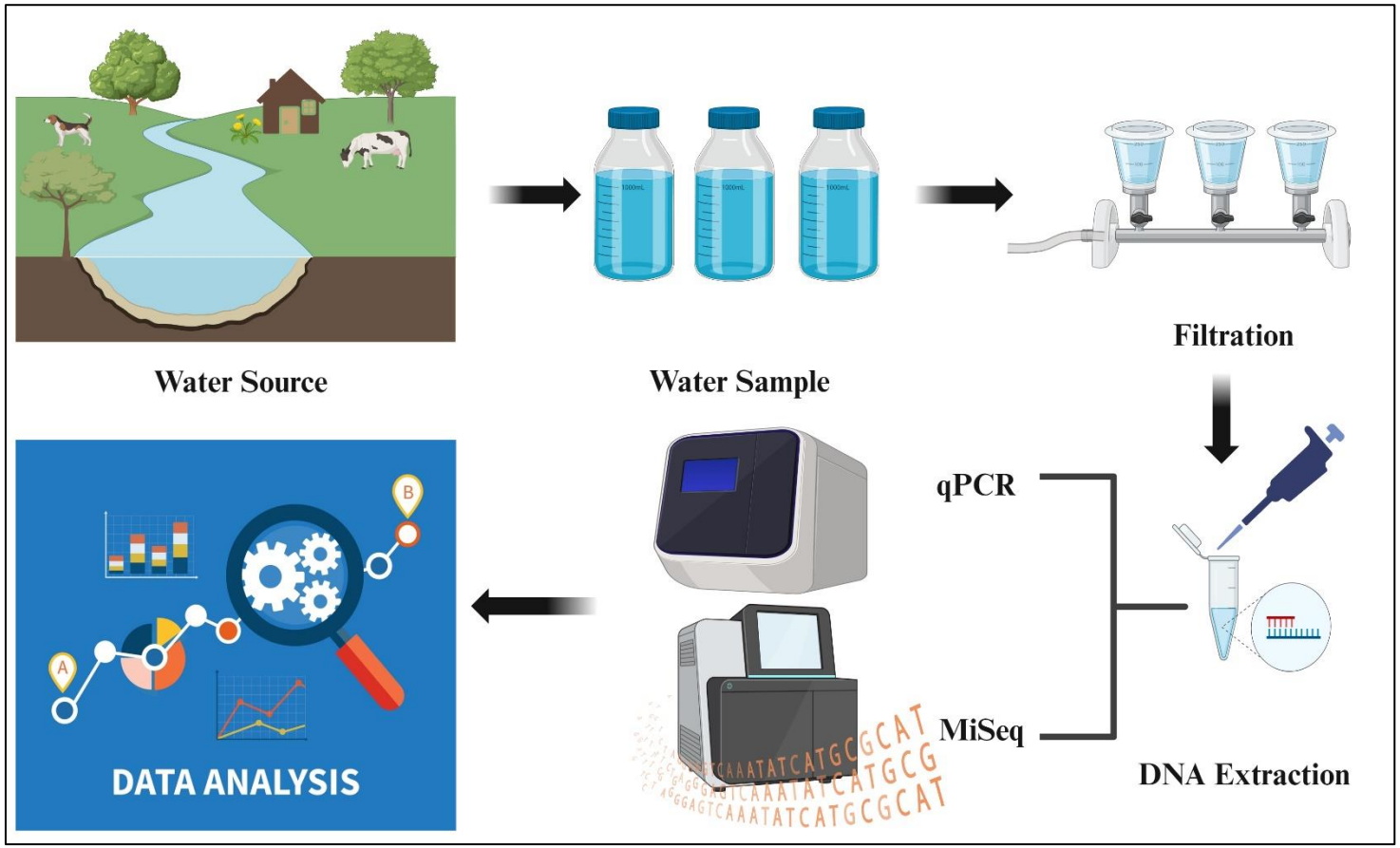


Fig 4 Overview of the experimental workflow

RESULTS AND DISCUSSION

3.1 In-stream physicochemistry and its association with fecal pollution

3.1.1 Physicochemical analysis of river samples

Non-biological properties of the environment can also provide additional information about the sources and dissemination of microbial contaminants [10, 35, 37]. Therefore, physico-chemical parameters such as pH, temperature, DO, and electrical conductivity were measured. The samples collected before rainfall ranged between (24.7-26.7°C) for temperature, pH ranged between (6.8-8.5)theO was measured between (40.1-52.5 mg/l) while electrical conductivity was recorded between (223-582.6µS/cm). Similarly, among after-rainfall samples, the temperature varied between (24.3-27.1°C), pH ranged between (6.6-7), DO was (11.3-16.6 mg/l), and electrical conductivity was measured between (220-354 µS/cm). Table Shows the physicochemical parameters measured across the sample sites.

3.1.2. Association of physicochemical analysis with fecal pollution

No significant difference was observed in temperature for before and after rainfall samples. The highest pH was observed at the upstream site BC02 in before and after rainfall samples. Before rainfall samples showed a decrease till downstream sites. However, a sharp increase in pH was observed at SY01 and BR01. This can be attributed to the runoffs from agricultural lands, as certain salts or fertilizers can increase the river's alkalinity. When the second set of samples was analyzed, they also showed a decrease in the downstream samples. Before and after rainfall samples were compared, a significant drop in the pH ($p<0.05$) was seen. This decrease can be a result of microbial activity indicating fecal contamination. Similarly, for DO($p<0.05$) and electrical conductivity($p<0.05$), the difference between the samples was significant. Low DO is often a result of high BOD and COD, which indicates the presence of microbes in the water [23].

Table 3: Physicochemical parameters measured across the sample sites.

Parameter	Temperature(°C)		pH*		DO(mg/L)*		Electrical Conductivity(µS/cm)*	
Site	Before Rainfall	After Rainfall	Before Rainfall	After Rainfall	Before Rainfall	After Rainfall	Before Rainfall	After Rainfall
BC02	26.7	27.1	8.5	7	42	12	379	278.3
MH08	25.1	26.5	7.8	6.7	49.4	12.7	548.2	354.9
MH09	26.4	26.5	7.7	6.6	42.6	13.8	491.3	303.7
SY01	24.7	27	8.2	6.7	50.1	14	554.5	267.9
WH01	24.7	24.3	6.8	6.7	40.1	14.4	223	220.6
BR01	25	25.8	8.2	6.7	48.3	11.3	547.8	317.5
MH10	26.4	25.6	7.8	6.7	52.5	16.6	582.6	250.4

(**pval***<0.05)**

3.2 Microbial source tracking through quantitative real-time PCR

Quantitative real-time PCR was performed to detect and quantify fecal coliforms and source-specific microbial markers indicative of their sources in environmental samples.

3.2.1 Quantitative analysis of fecal coliforms in river samples

For *E.coli, uidA* gene was targeted and amplified to determine its contamination level across all the sample sites [4,12]. The results were normalized with 16SrRNA and expressed as copies per 16SrRNA gene copies. In the samples collected before rainfall, a slight increase was observed between BC02 and MH08. It was followed by a decrease until MH10, except WH01, which exhibited the highest copies per 16SrRNA gene copies. (Table 3) A WWTP is located upstream of BC02, which can be a major source of this fecal contamination. Many studies have identified WWTPs as hotspots for fecal pollutants. [1, 2]. The gradual decrease in the number of copies from upstream to downstream sites can be attributed to the attenuation of coliforms due to sedimentation or dilution into another stream. A similar trend was observed in after-rainfall samples for upstream sites until MH09. Elevated contamination levels were detected at sites SY01, WH01, and BR01, which can be due to livestock or agricultural runoffs.

In a comparative analysis, it was evident that all sites showed increased copy numbers after rainfall, but the surge exhibited by BR01, SY01, and WH01 was prominent. This sudden increase can be an outcome of rainfall expediting runoffs from the ranch or livestock farms.

3.2.2 Quantitative analysis of source-specific fecal pollutants in river samples

Human, cow, and pig-specific microbial markers were utilized to source-track fecal pollutants. Site-wise analysis showed a diverse variation in the level of contamination from upstream to downstream, indicating site-specific dynamics influenced by external factors such as activity and land use around the area under study. Many studies have shown that rainfall can accelerate the process of fecal dissemination into water [16].

Human-specific pollutants had low contamination levels at SY01 and MH10, whereas sites BC02 and WH01 displayed relatively high concentrations in before-rainfall samples. The prevalence of human-specific pollutants at BC02 may correspond to the upstream WWTP of the river. BR01 and BC02 were identified as sites with the highest and lowest copy numbers, respectively, in the samples collected after rainfall. This shift in BC02 after rainfall, from the highest to lowest concentration, shows the impact of rainfall on fecal pollution, while the WWTP plant upstream BR01 may have contributed to the increase in copies after rainfall.

Cow-specific fecal contamination was higher at upstream sites BC02, MH08, MH09, and WH01 than downstream in samples collected before rainfall. MH08 had the highest number of cow-specific pollutants, followed by BC02 and WH01, while after rainfall, BC02 had the highest copies, followed by MH08 and SY01. The cow-specific pollutants may be associated with the veterinary hospital around WH01, agricultural lands, and livestock farms. Comparative analysis between before and after rainfall samples showed a substantial increase at BR01. SY01 and MH10 also showed a marginal increment, in contrast to BC02, MH08, and MH09, where a consistent decrease was observed, suggesting potential run-offs or dilution into other streams.

Pig-specific pollutants were found to be the most prevalent across the river. A trend like cow-specific pollutants was observed where upstream sites were more polluted than downstream sites. MH08 exhibited the highest concentration, indicating pigs as the source of its contamination, but post-rainfall, the level decreased. Sites BC02, MH09, MH10, and WH01 followed the same pattern. However, a substantial increase was observed at BR01, followed by SY01.

Among all the three suspected sources, pig-specific pollutants had the highest contribution to the fecal pollution of the river, followed by cows. Human-specific pollutants had relatively low concentrations in the samples. BR01 and WH01 had the most significant difference in the

concentrations in prior and after rainfall samples. T-tests performed on before and after rainfall samples showed significant differences($p > 0.05$) for human-specific pollutants. No significant differences were observed for cow and pig-specific pollutants.

Table 4 Quantitative analysis of the source-specific fecal pollutants

Site	*E.coli*		Human		Cow		Pig	
	Before Rainfall	After Rainfall	Before Rainfall	After Rainfall	Before Rainfall	After Rainfall	Before Rainfall	After Rainfall
BC02	2.63	3.4	1.24×10^{-4}	3.33×10^{-5}	1.94×10^{-4}	1.41×10^{-4}	4.76×10^{-3}	3.07×10^{-3}
MH08	2.75	4.14	7.65×10^{-6}	2.41×10^{-4}	3.37×10^{-4}	1.53×10^{-4}	4.77×10^{-3}	1.92×10^{-3}
MH09	2.29	2.9	2.67×10^{-5}	1.41×10^{-4}	1.40×10^{-4}	1.03×10^{-4}	1.68×10^{-3}	2.06×10^{-3}
SY01	1.57	3.45	1.82×10^{-5}	1.42×10^{-4}	4.35×10^{-5}	1.37×10^{-4}	6.31×10^{-4}	2.93×10^{-3}
WH01	2.95	4.71	7.20×10^{-5}	5.94×10^{-5}	1.29×10^{-4}	6.23×10^{-5}	1.79×10^{-3}	2.72×10^{-3}
BR01	1.61	6.47	2.17×10^{-5}	5.84×10^{-4}	4.85×10^{-5}	2.53×10^{-4}	8.18×10^{-4}	8.70×10^{-3}
MH10	1.17	2.92	1.11×10^{-5}	3.99×10^{-4}	2.96×10^{-5}	8.16×10^{-5}	8.78×10^{-4}	1.25×10^{-3}

*The numerical values above represent copies/16SrRNA gene copies.

3.2.3 Correlation between physicochemical parameters and source-specific fecal pollutants.

Spearman's correlation analysis was performed between physicochemical parameters and source-specific pollutants to investigate the impact of physicochemical parameters on fecal pollution. No significant correlation was observed between temperature and source-specific pollutants. Similarly, the correlation found between pH and source-specific pollutants was not significant. This shows that neither temperature nor pH have a strong influence on fecal contamination. DO and electrical conductivity showed a strong negative correlation with *E. coli* and human, cow, and pig-specific pollutants. This indicates that low levels of DO may be associated with high fecal contamination. Microbes often take up the oxygen dissolved in the water, resulting in low levels, indicating fecal pollution. [23]. Similarly, electrical conductivity showed strong negative correlations with *E. coli* and human, cow, pig specific.

Table 5. Correlation matrix between physicochemical parameters and source-specific microbial indicators.

Parameters	***E.coli***	**Human**	**Cow**	**Pig**
Temperature	-0.16	0.05	0.2	0.3
pH	-0.05	0.31	0.11	0.04
DO	-0.62*	-0.71*	-0.50*	-0.48*
Electrical Conductivity	-0.60*	-0.73*	-0.50*	-0.52*

**pval<0.0*

3.3 Identification of source-specific indicator bacteria through NGS-based methods

Due to its wide applications, next-generation sequencing has been utilized in many metagenomic studies. This study integrated NGS with qPCR-based source-tracking methods to identify and suggest potential source-specific microbial markers.

3.3.1 Microbial community analysis through Illumina MiSeq Sequencing

Illumina Miseq Sequencing was performed with Miho River samples to study the composition of microbial communities present across the sample sites and discover the shifts in their pattern under different conditions i.e. before and after rainfall. First, the relative abundance of microbial populations was calculated for all taxonomic levels. Proteobacteria were the highly abundant phyla, closely followed by Actinobacteria and Bacteroidetes Fig 5(a). Classes Betaproteobacteria, Actinobacteria, and Cyanobacteria showed high relative abundance among all. Fig 5(b) . At the order level, Burkholderiales and Micrococcales were most prevalent Fig 5(c). while *Comamnadaceae* and *GpIIa* Fig 5(d). were abundant at the family level. Fig 5(e). Bacterial populations belonging to class *Betaproteobacteria* were dominant in the river water, followed by *Actinobacteria* and *Cyanobacteria.*. Among, Genera, *Rhodoluna*, *Actinobacteria_unclassified* were among the most abundant.

For alpha diversity, Chao-Shannon indices were calculated for all river samples. In Chao plot, Fig.6 (a) all sites had significant increase post rainfall among which MH08 and MH09 were the highest. ($pval<0.01$). This also validates the fecal contamination reported at upstream sites. No significant changes were observed for BR01 which shows there was no evident Shannon plot Fig6(b). shows a significant increase in all samples. Samples from MH08 were the most diverse while WH01 had the lowest diversity.

For Beta diversity, initially a Bray-Curtis dissimilarity matrix was calculated to explore the compositional shifts in before and after rainfall samples followed by NMDS scaling for

visualization. Spearman's correlation analysis was performed and those with significant values were added to the plot. Fig 7 shows an NMDS biplot where arrows represent top 5 genus. It can be observed from the plot *Curvibacter* was associated with sites BC02, MH09 and BR01 before rainfall while in after rainfall samples, genera *Algoriphagus* and *Sandarakinorhabdus* were associated with sites MH09, SY01 and BR01.

Similarly, we performed Spearman's correlation analysis between microbial community and physicochemical parameters. The data was visualised through an NMDS biplot which showed that DO and electrical conductivity had significant impact on microbial communities within the sample. SY01, MH09 and BR01 were the most affected (Fig 8).

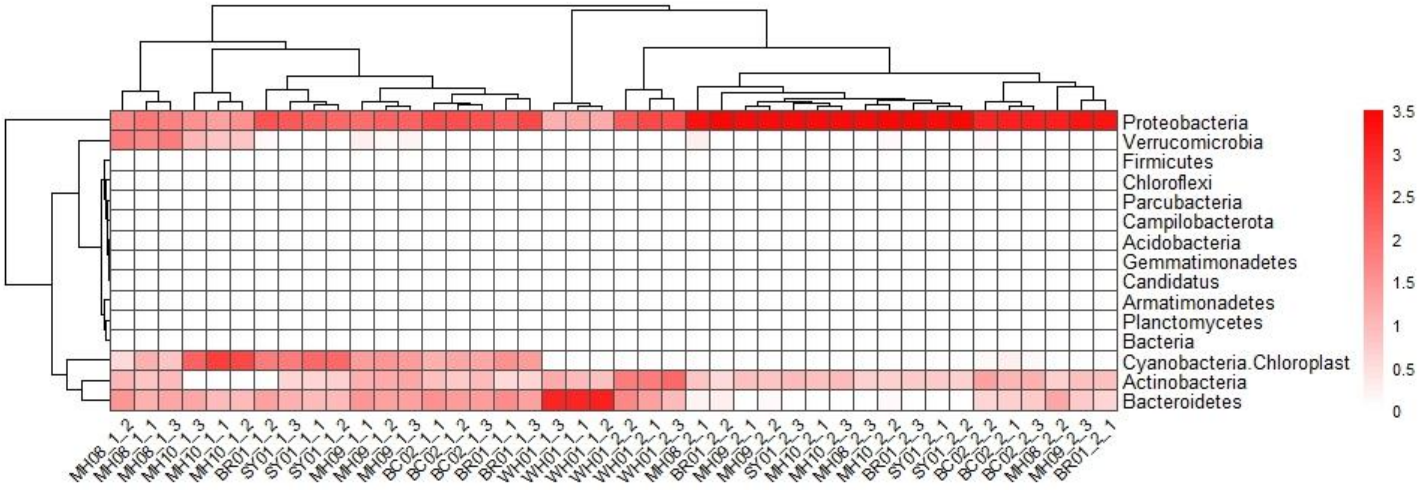


Fig 5(a) Heatmap showing the taxonomical composition at phylum level.

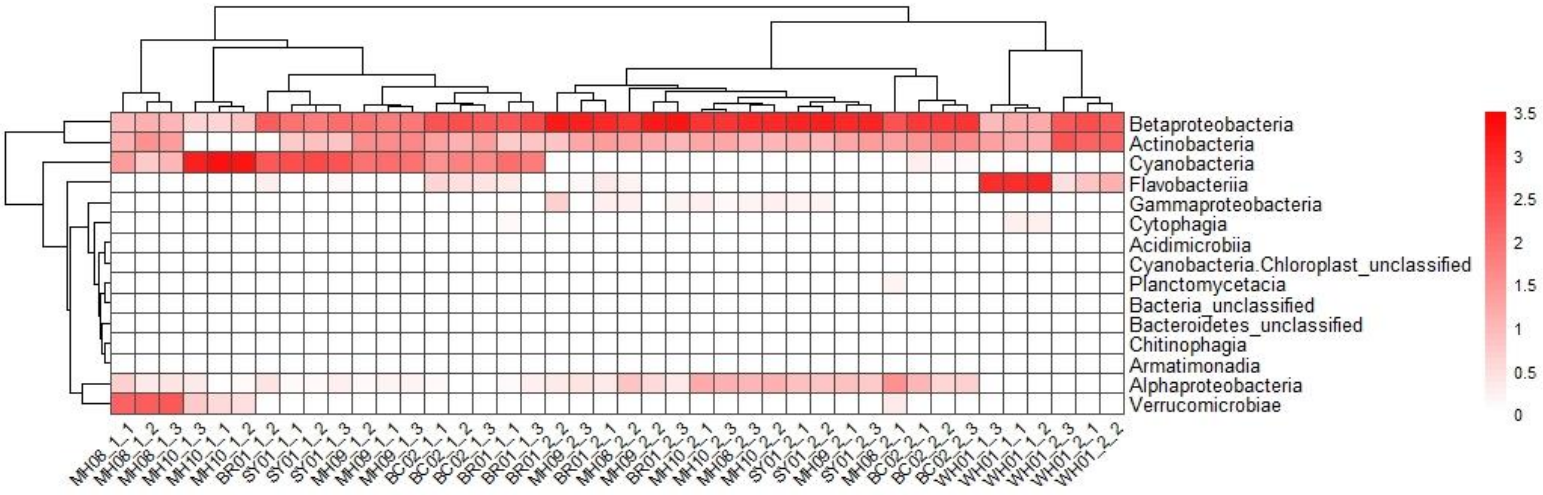


Fig 5 (b)Heatmap showing the taxonomical composition at class level

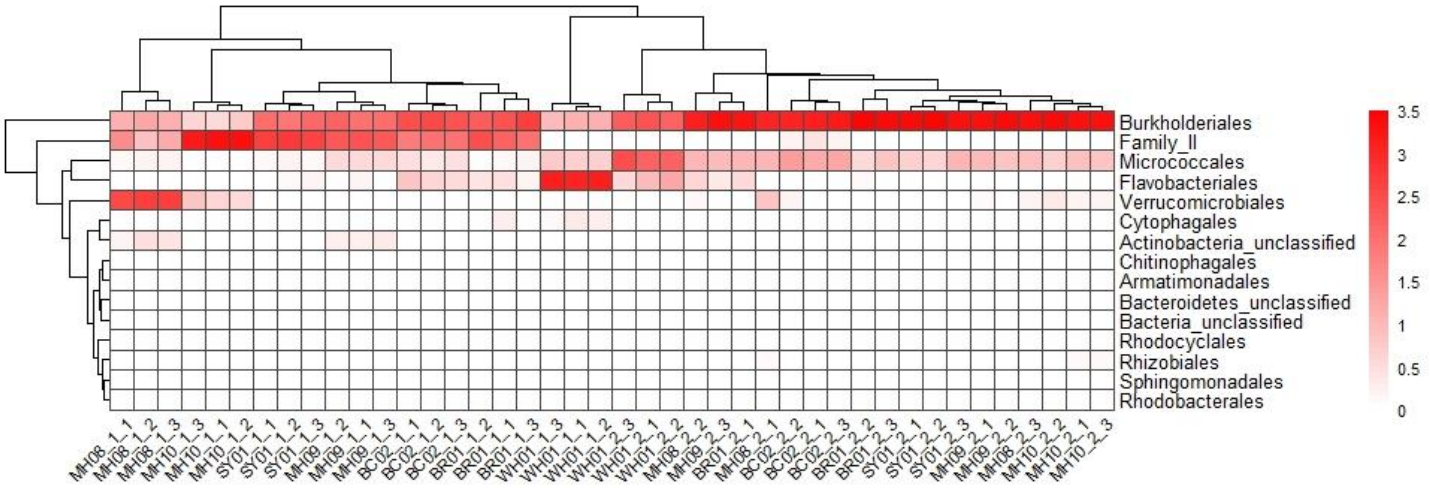


Fig 5(c)Heatmap showing the taxonomical composition at order level.

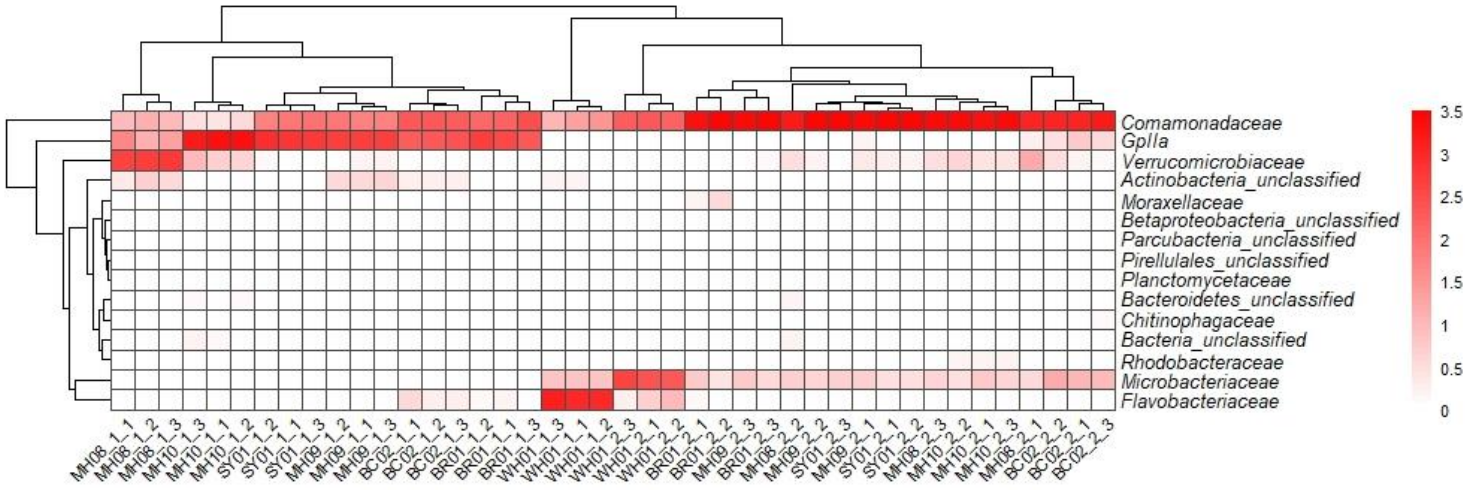


Fig 5(d) Heatmap showing the taxonomical composition at family level

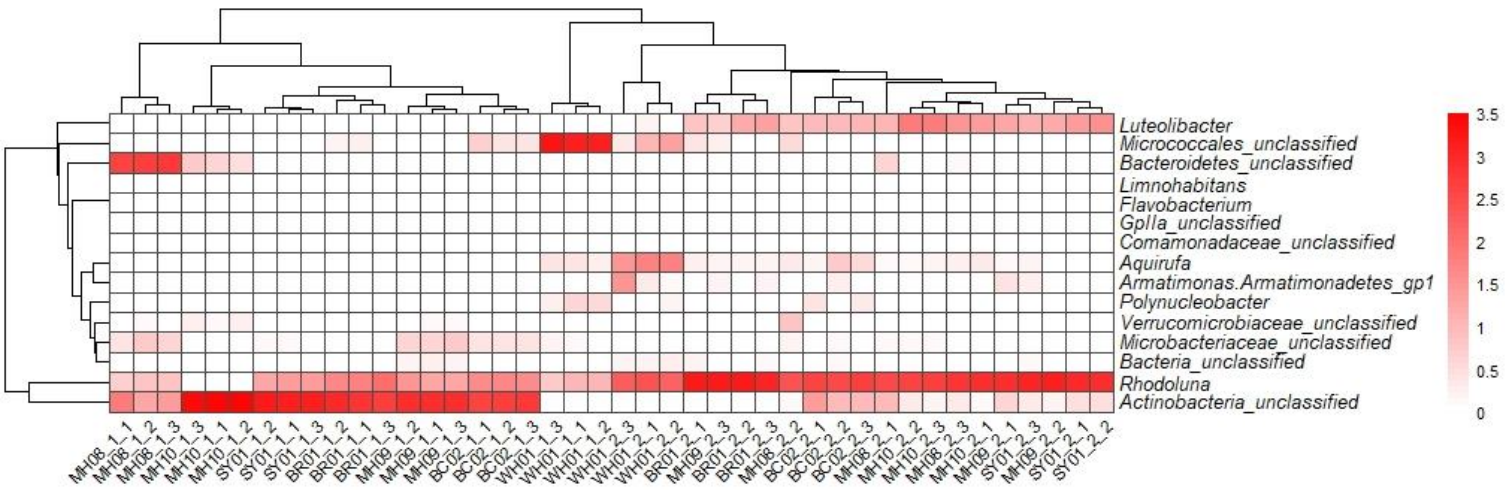
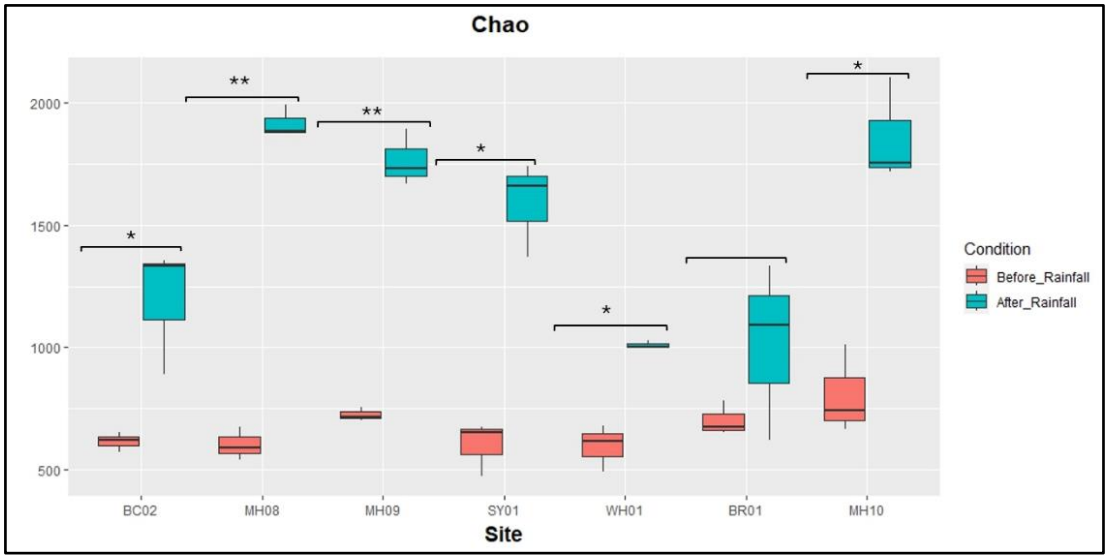
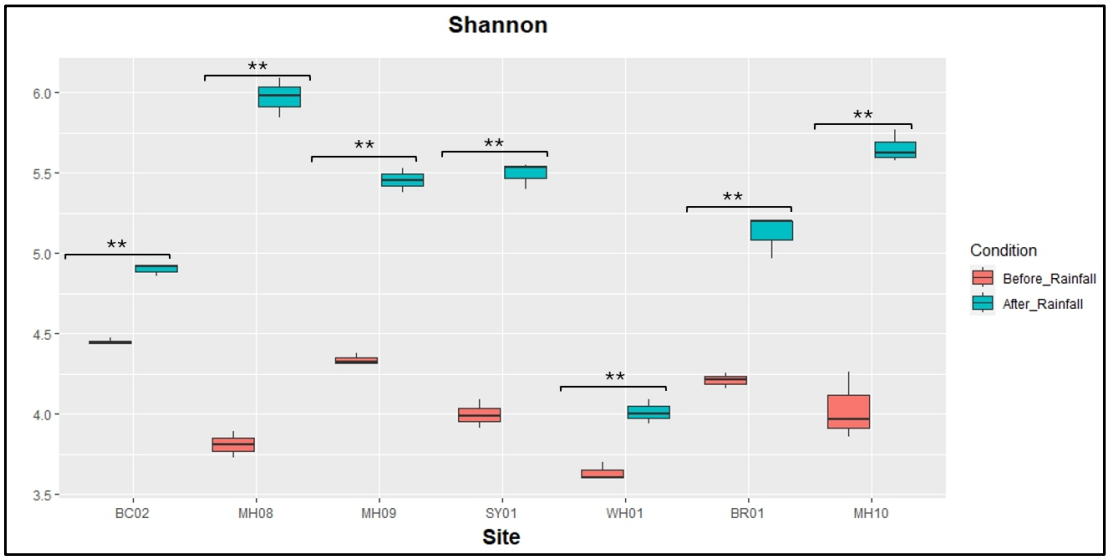


Fig 5(e): Heatmap showing the taxonomical composition at genus level



(a)



(b)

Fig 6. (a) Chao and (b) Shannon boxplot showing variations in river samples

(**pval<0.05, **pval<0.01)*

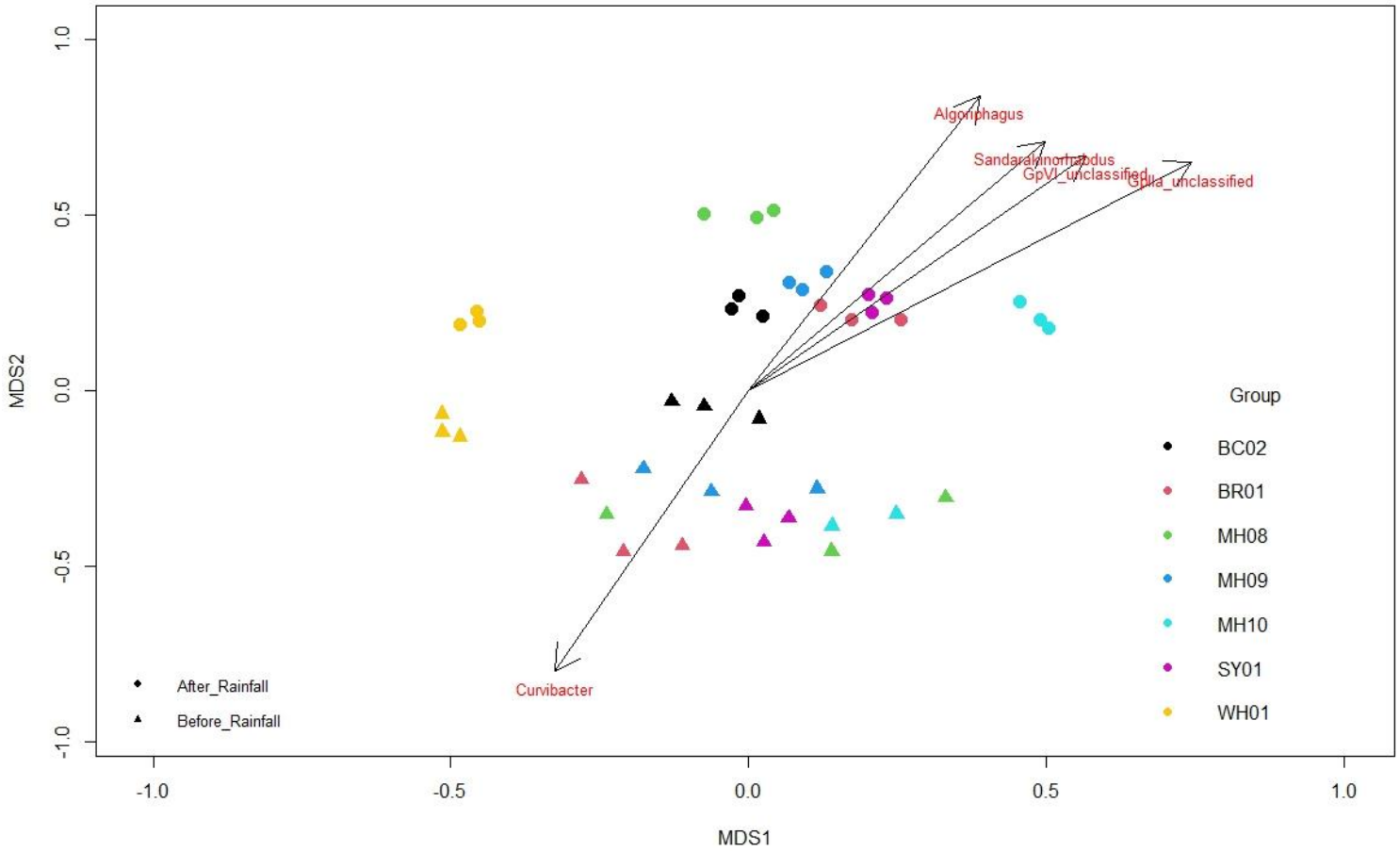


Figure 7. NMDS biplot showing correlation analysis between microbial community and abundant genera

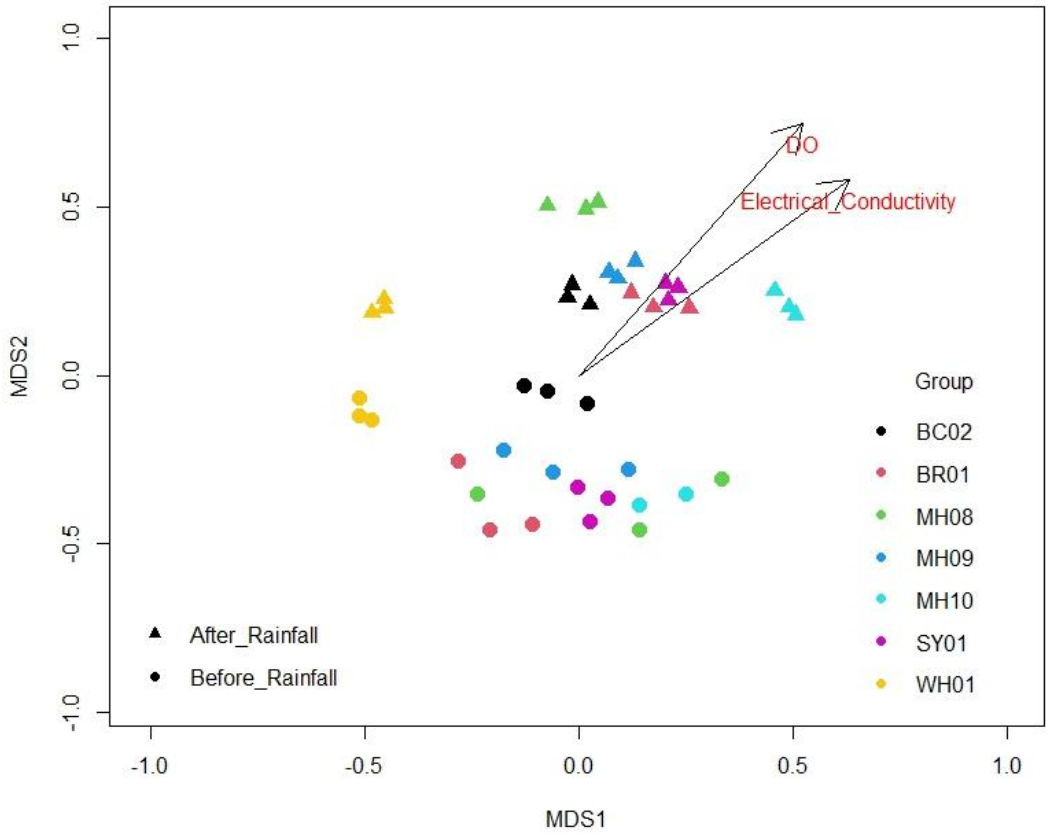
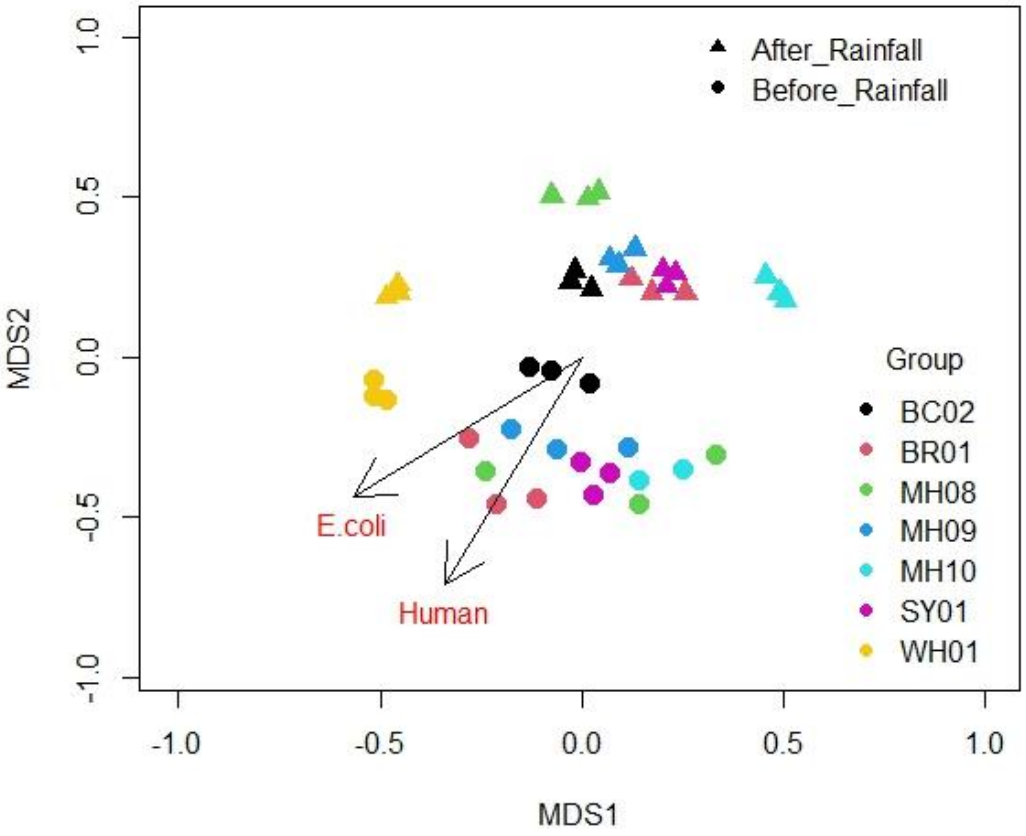


Figure 8. NMDS biplot showing correlation analysis between microbial community and physicochemical parameters.

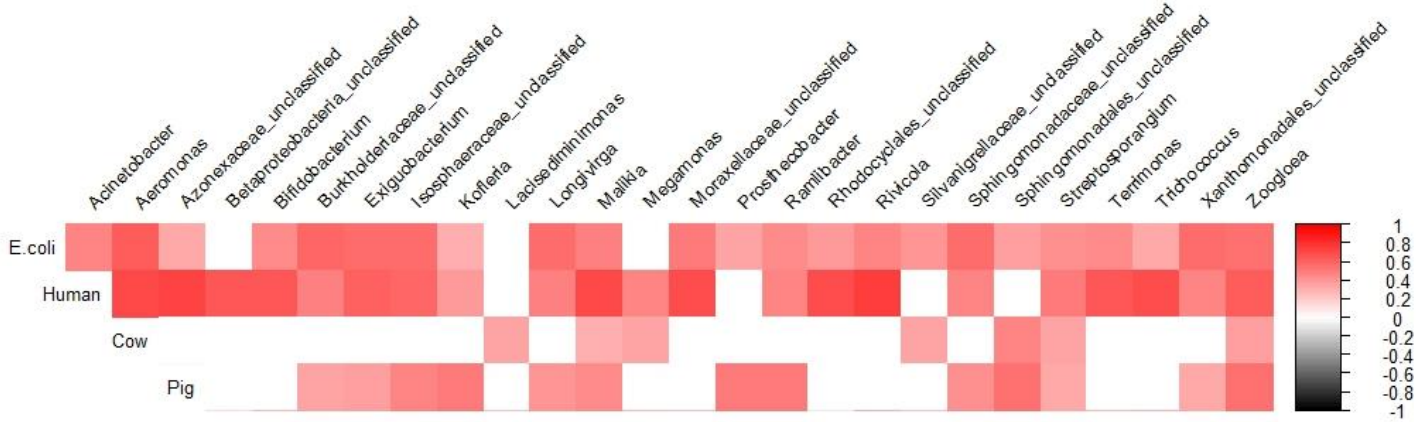
3.3.2. Association of microbial communities with source-specific fecal sources

Multivariate analysis was performed to evaluate relationship between microbial communities and source-specific pollutants i.e. human, cow, pig. Bray-Curtis distances were calculated to create a matrix based on the dissimilarity scores. This matrix was visualized through NMDS. Spearman's correlation analysis was performed between microbial community and source-specific pollutants. The proximity of the arrows to the samples before rainfall suggests that the microbial composition and abundance of *E. coli* and human copy number are more strongly associated with conditions before rainfall.



3.3.3. Investigating potential host-specific markers for microbial source tracking

Microbial source tracking, studies have suggested that a combination of various source tracking methods can be more efficient in discriminating fecal sources than a single method. [12, 14]. It is essential to discover new patterns in the distribution of microbial communities in the environment. This could be used to develop new host-specific microbial markers. For this we need evaluate which genera within microbial community could be associated with the pollutant sources. Spearman's correlation analysis was performed between genus abundance and the source-specific fecal pollutants. Most genera correlated to *E. coli* and humans which showed their potential to be developed as source-specific indicators. Genera correlating to *E. coli* were *Aeromonas, Betaproteobacteria,* and *Bifidobacterium* Similarly, human-specific also correlated with *Aeromonas, and Bifidobacterium, Malikia*, and *Rivicola. Megamonas,* which showed positive correlations with humans have previously been associated with liver disease in children and adolescents, [38, 45]. The number of genera showing correlations to cow and pig-specific pollutants was relatively less. Cow-specific showed correlation with *Malikia* and *Zoogloea* while for pig-specific *Ramlibacter, Kofleria,* and *Zoogloea* were observed. Fig(10) shows all genera which correlated with source-specific pollutants.



Fig(10) shows correlation matrix between the source-specific pollutants and abundant genera.

3.3.3. Effect of rainfall on the distribution of microbial communities

The qPCR results indicated elevated concentration of fecal pollutants in after-rainfall samples. Similarly, Chao-Shannon plot depicts significant compositional shifts in the samples collected after rainfall. This validates that rainfall impacts microbial populations significantly. For deeper investigation of prevailing microbial communities within different environment conditions, to understand the effect of rainfall on microbial composition, Linear Discriminant Analysis Effect Size (LEfSe) analysis was performed. Differential abundant genera were identified in before and after rainfall samples. This provided better insights in understanding microbial composition and their response to various environmental changes.

At BC02, (fig) *GpIIa_unclassified*, and *Flavobacterium* were found as differentially abundant genera, while in post-rainfall samples, *Comamonadaceae_unclassified* and *Rhizobiales_unclassified* were abundant. A similar pattern was observed for sites MH08, MH09, SY01 and BR01. For WH01, *Flavobacterium*, *Actinobacteria* and *Aquirufa* were abundant in the before rainfall samples and *Rhoduluna* in after rainfall samples.

Site-wise analysis showed that genera, *GpIIa_unclassified*, *Flavobacterium*, and *Actinobacteria* were the most differentially abundant within before-rainfall samples. However, post-rainfall, it shifted to *Comamonadaceae_unclassified*, *Mycobacterium*

For MH08, *Luteolibacter* , *GpII unclassified* were abundant followed by *Actinobacteria* but after rainfall *Comamonadaceae_unclassified* and *Dechloromonas* were found *to be* differentially abundant.

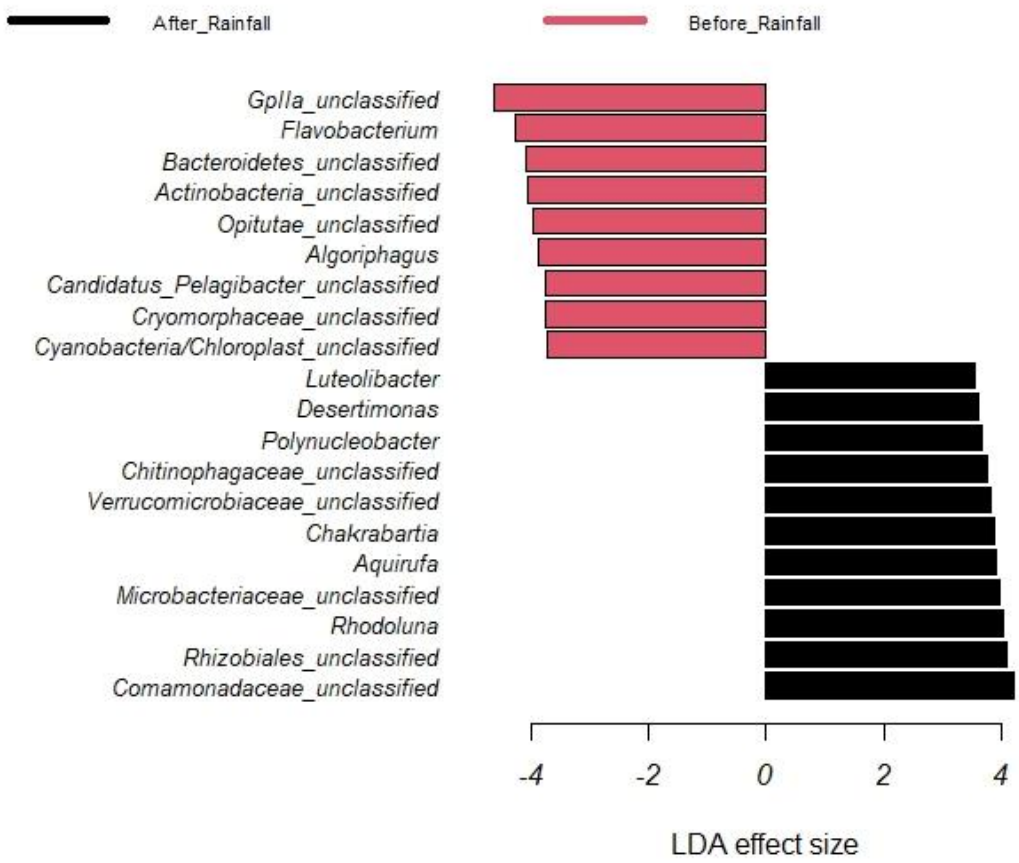


Fig 11(a) LEfSe plot showing differentially abundant genera in before and after rainfall samples obtained from BC02

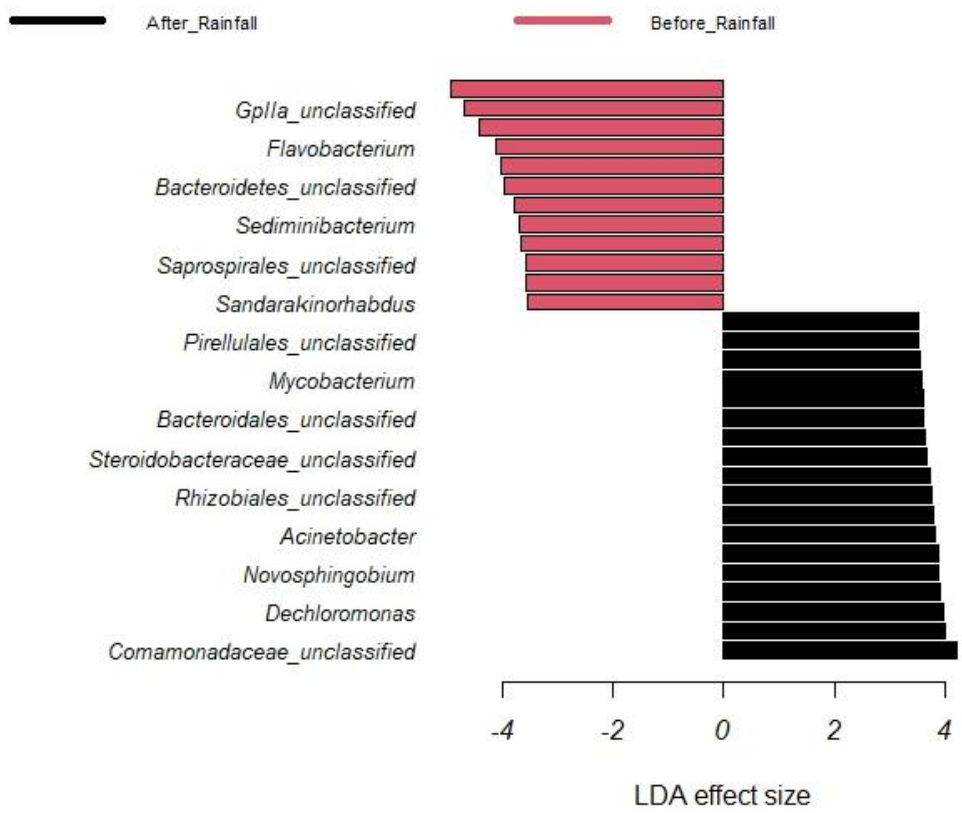


Fig 11(b) LEfSe plot showing differentially abundant genera in before and after rainfall samples obtained from MH08

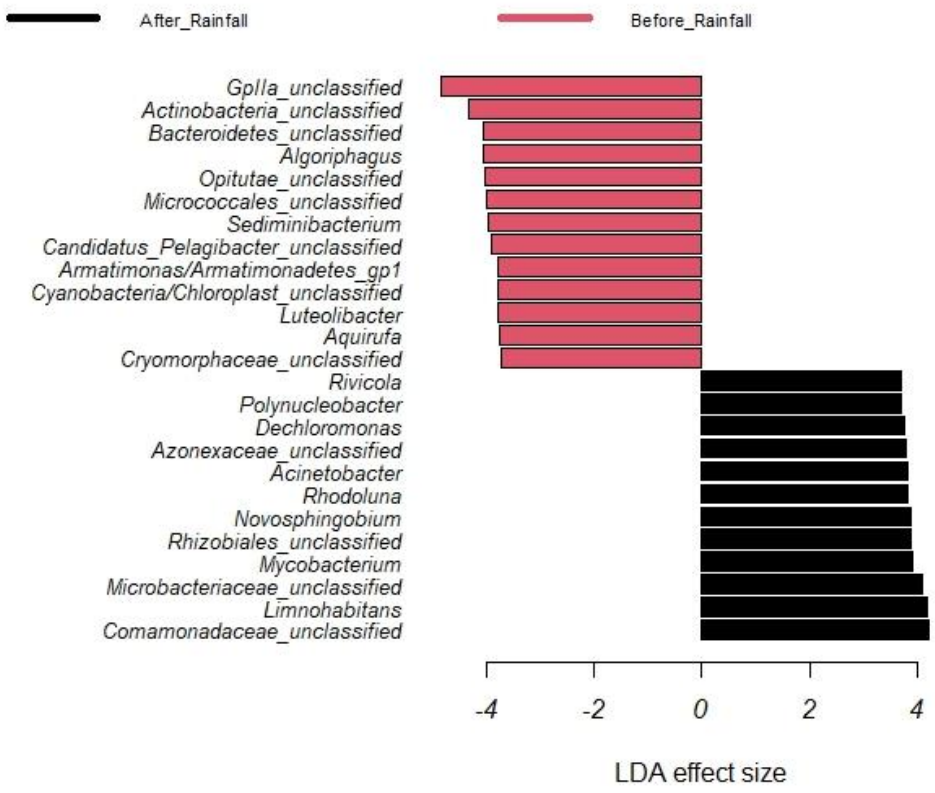


Fig11(c) LEfSe plot showing differentially abundant genera in before and after rainfall samples obtained from MH09

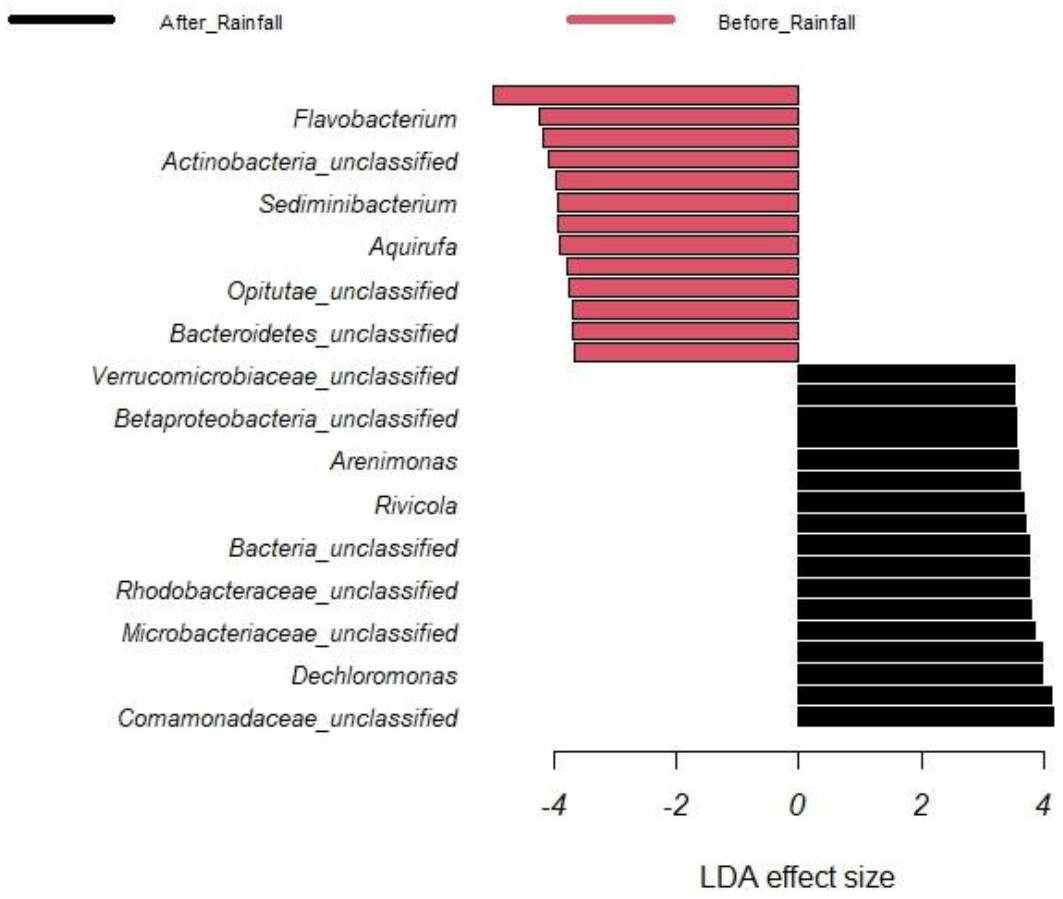


Fig 11(d) LEfSe plot showing differentially abundant genera in before and after rainfall samples obtained from SY01

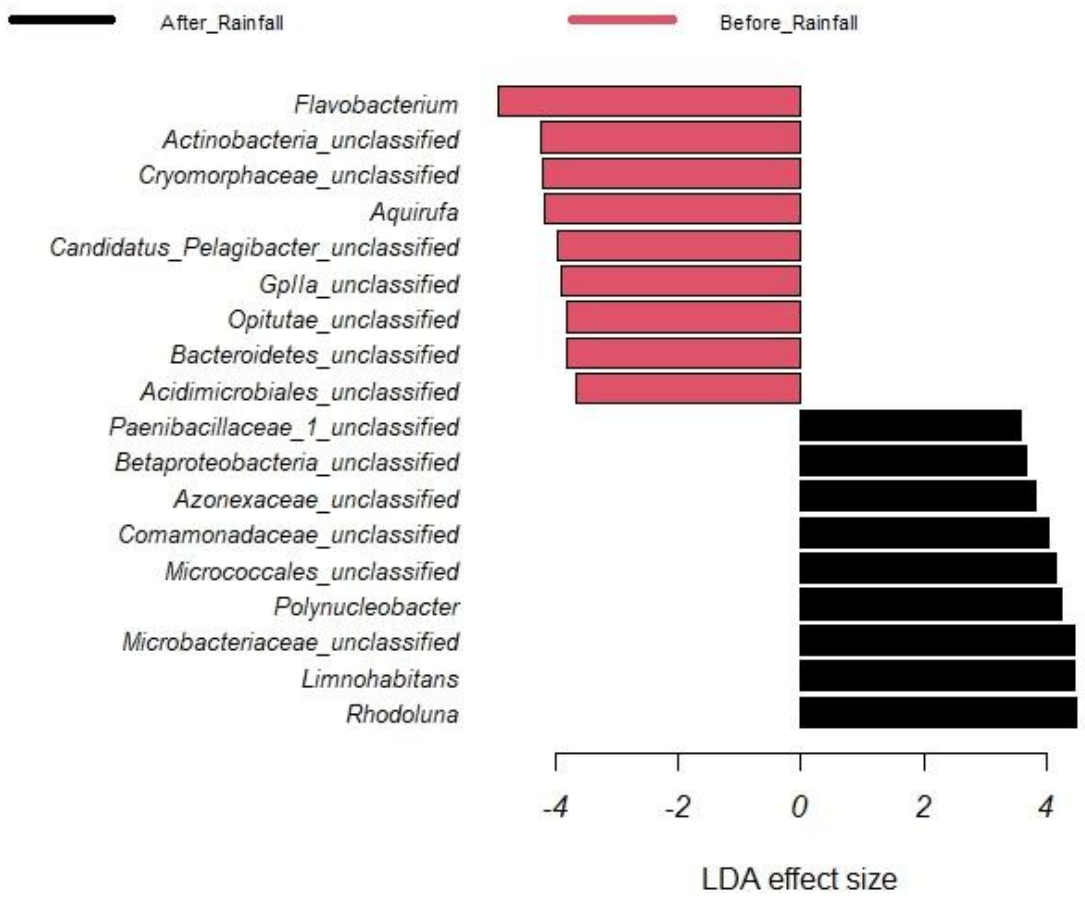


Fig 11(e) LEfSe plot showing differentially abundant genera in before and after rainfall samples obtained from WH01

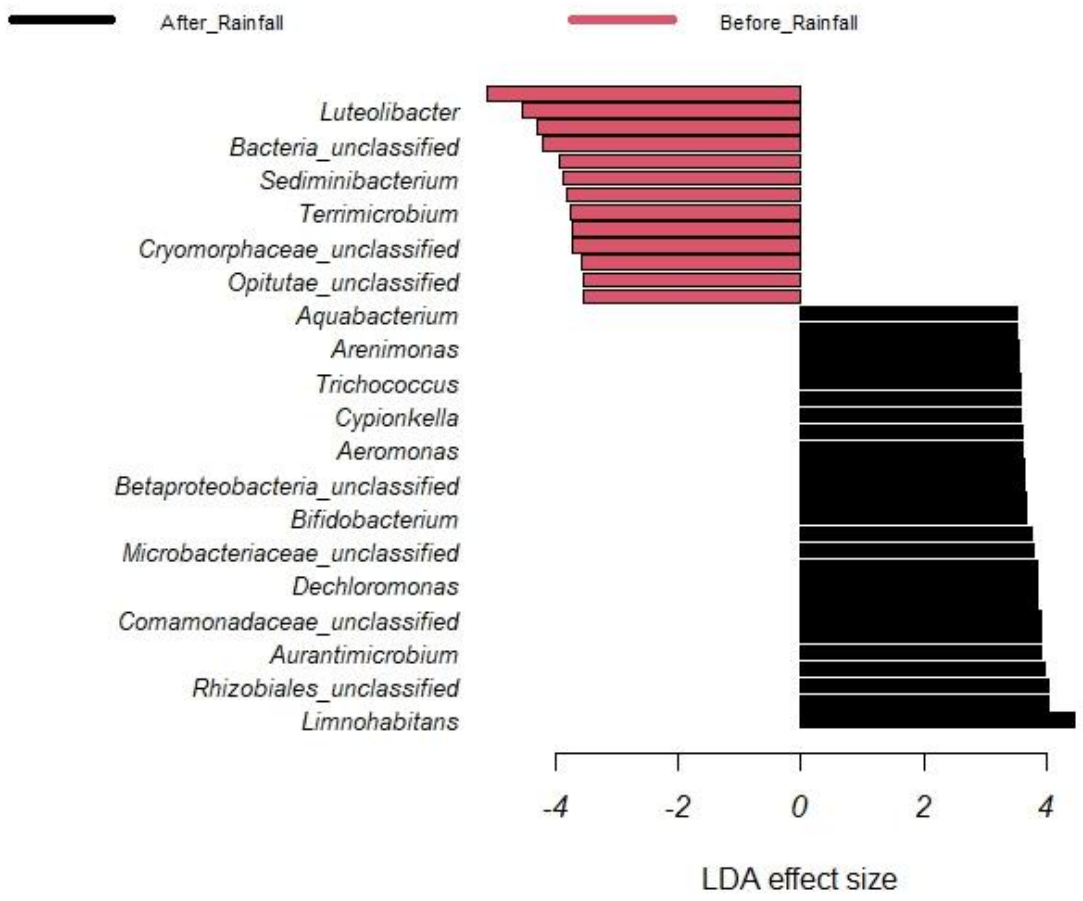


Fig 11(f) LEfSe plot showing differentially abundant genera in before and after rainfall samples obtained from MH10

CONCLUSION

In this study, we investigated seven sites on Miho River, suspected of fecal pollution. To identify the pollutant sources and for their quantification we utilised qPCR-based microbial source tracking methods. Host-specific primers were used to target human, cow, and pig-specific fecal pollutants in the river. The results showed that upstream sites were more affected than downstream as there was a decrease in the level of contamination between BC02 and MH10. However, variations between SY01 to BR01 were observed which can be associated with cattle and livestock farms located around the area. The WWTP upstream BC02 was suggested as the source of contamination and the gradual decrease was attributed to the attenuation due to sedimentation and dilution. Pig-specific contaminants were found to be as the major source, contributing to fecal pollution followed by cow-specific. Human-specific fecal pollutants were relatively low. Our results concluded that livestock farms contributed most the fecal pollution of Miho River.

Microbial source tracking through just one approach may not be effective in the identification of specific sources, especially when they are non-point sources. Therefore, we examined physicochemical parameters, effect of rainfall and microbial composition along with qPCR-based MST. In a comparative analysis, it was concluded that rainfall plays a vital role in disseminating pollutants into river as data showed elevated levels of *E.coli* in the sample collected after rainfall . DO and electrical conductivity showed strong correlations with *E.coli* and human-specific sources. Broad community analysis through NGS gave a more comprehensive understanding of microbial communities in environmental samples. It was observed that there was a pattern in the distribution of certain genera in before and after rainfall samples. Genera such as *Flavobacterium* were differentially abundant in before rainfall samples while in after-rainfall samples *Mycobacterium* were differentially abundant.

Correlation analysis between microbial communities and pollutant sources also suggested potential microbial markers for source tracking.

In conclusion, multidimensional approaches enhance the ability to identify and provide better insights for tracking fecal sources, leading to effective water quality. Our study also integrated the targeted approach of qPCR with physicochemical parameters. For further studies, a comparative analysis between qPCR and NGS- based MST methods can be suggested.

REFERENCES

1. Ahmed, W., et al. (2017). "Amplicon-based profiling of bacteria in raw and secondary treated wastewater from treatment plants across Australia." Applied Microbiology and Biotechnology **101**: 1253-1266.

2. Ahmed, W., et al. (2007). "Sourcing faecal pollution: a combination of library-dependent and library-independent methods to identify human faecal pollution in non-sewered catchments." Water research **41**(16): 3771-3779.

3. Anderson, K. L., et al. (2005). "Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments." Applied and environmental microbiology **71**(6): 3041-3048.

4. Bej, A., et al. (1991). "Polymerase chain reaction-gene probe detection of microorganisms by using filter-concentrated samples." Applied and environmental microbiology **57**(12): 3529-3534.

5. Bernhard, A. E. and K. G. Field (2000). "Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes." Applied and environmental microbiology **66**(4): 1587-1594.

6. Bernhard, A. E. and K. G. Field (2000). "A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA." Applied and environmental microbiology **66**(10): 4571-4574.

7. Blanch, A. R., et al. (2006). "Integrated analysis of established and novel microbial and chemical methods for microbial source tracking." Applied and environmental microbiology **72**(9): 5915-5926.

8. Clarke, K. R., et al. (2006). "On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages." Journal of experimental marine biology and ecology **330**(1): 55-80.

9. Devarajan, N., et al. (2016). "Occurrence of antibiotic resistance genes and bacterial markers in a tropical river receiving hospital and urban wastewaters." PloS one **11**(2): e0149211.

10. Febria, C., et al. (2010). "Characterizing seasonal changes in physicochemistry and bacterial community composition in hyporheic sediments." Hydrobiologia **647**: 113-126.

11. Field, K. G. and M. Samadpour (2007). "Fecal source tracking, the indicator paradigm, and managing water quality." Water research **41**(16): 3517-3538.

12. Gourmelon, M., et al. (2010). "Development of microbial and chemical MST tools to identify the origin of the faecal pollution in bathing and shellfish harvesting waters in France." Water research **44**(16): 4812-4824.

13. Green, H., et al. (2019). "Microbial Source-Tracking Reveals Origins of Fecal Contamination in a Recovering Watershed." Water **11**(10): 2162.

14. Griffith, J. F., et al. (2003). "Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples." Journal of Water and Health **1**(4): 141-151.

15. Valerie J. Harwood, Christopher Staley, Brian D. Badgley, Kim Borges, Asja Korajkic, Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes, *FEMS Microbiology Reviews*, Volume 38, Issue 1, January 2014, Pages 1–40, https://doi.org/10.1111/1574-6976.12031

16. Henry, R., et al. (2016). "Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters." Water research **93**: 242-253.

17. Jiang, X., et al. (2014). "A continuous-flow high-throughput microfluidic device for airborne bacteria PCR detection." Lab on a Chip **14**(4): 671-676.
    a. Rapid analysis of airborne pathogens plays a critical role in early warning of spreading infectious diseases, which is essential for public health and disease prevention. Herein we report an integrated microfluidic device that can perform airborne pathogen capture, enrichment and continuous-flow high-throughput gene analysis. The device was validated by six frequently encountered bacteria and shows great potential in environmental analysis and for public health protection.

18. Kirs, M., et al. (2017). "Multiple lines of evidence to identify sewage as the cause of water quality impairment in an urbanized tropical watershed." Water research **116**: 23-33.

19. Ko, H. Y., et al. (2018). "Host-specific Bacteroides markers-based microbial source tracking in aquaculture areas." Microbes and environments **33**(2): 151-161.

20. Kongprajug, A., et al. (2019). "Human and animal microbial source tracking in a tropical river with multiple land use activities." International Journal of Hygiene and Environmental Health 222(4): 645-654.

21. Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. 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. 2006 Jun;72(6):4214-24. doi: 10.1128/AEM.01036-05. PMID: 16751534; PMCID: PMC1489674.

22. Lee, S. H., et al. (2021). "Trends in recent waterborne and foodborne disease outbreaks in South Korea, 2015–2019." Osong public health and research perspectives 12(2): 73.

23. Middelburg, J. J. and L. A. Levin (2009). "Coastal hypoxia and sediment biogeochemistry." Biogeosciences 6(7): 1273-1293.

24. Molina IV, D. and K. L. Lowe (2012). "Use of quantitative PCR of functional genes to monitor coliforms and enteric bacteria in a constructed, urban wetland." bios 83(4): 136-144.

25. Moon, S., et al. (2013). "Epidemiological characteristics of the first water-borne outbreak of cryptosporidiosis in Seoul, Korea." Journal of Korean medical science 28(7): 983-989.

26. Nation, U. (2018). United Nations, Department of Economic and Social Affairs, Population Division (2018).

27. Nations, U. (2004). World urbanization prospects: the 2003 revision, UN.

28. Okabe, S., et al. (2007). "Quantification of host-specific Bacteroides–Prevotella 16S rRNA genetic markers for assessment of fecal pollution in freshwater." Applied Microbiology and Biotechnology **74**: 890-901.

29. Park, Y.C., Kim, P.H., Jung, Y.J. *et al.* Seasonal variation of physicochemical factor and fecal pollution in the Hansan-Geojeman area, Korea. *Fish Aquatic Sci* **19**, 17 (2016). https://doi.org/10.1186/s41240-016-0017-9.

30. Savichtcheva, O. and S. Okabe (2006). "Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives." Water research **40**(13): 2463-2476.

31. Schang, C., et al. (2016). "Evaluation of techniques for measuring microbial hazards in bathing waters: A comparative study." PloS one **11**(5): e0155848.

32. Schloss, P. D., et al. (2009). "Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities." Applied and environmental microbiology **75**(23): 7537-7541.

33. Schwarzenbach, R. P., et al. (2010). "Global water pollution and human health." Annual review of environment and resources **35**: 109-136.

34. Segata, N., et al. (2011). "Metagenomic biomarker discovery and explanation." Genome biology **12**: 1-18.

35. Seurinck, S., et al. (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." Environmental microbiology **7**(2): 249-259.

36. Shanks, O. C., et al. (2009). "Quantitative PCR for genetic markers of human fecal pollution." Applied and environmental microbiology **75**(17): 5507-5513.

37. Shehane, S., et al. (2005). "The influence of rainfall on the incidence of microbial faecal indicators and the dominant sources of faecal pollution in a Florida river." Journal of Applied Microbiology **98**(5): 1127-1136.

38. Sheng, S., et al. (2022). "Gut microbiome is associated with metabolic syndrome accompanied by elevated gamma-glutamyl transpeptidase in men." Frontiers in Cellular and Infection Microbiology **12**: 946757.

39. Simpson, J. M., et al. (2002). "Microbial source tracking: state of the science." Environmental science & technology **36**(24): 5279-5288.

40. Studer, I., et al. (2017). "Physicochemical and microbiological indicators of surface water body contamination with different sources of digestate from biogas plants." Ecological Indicators **77**: 314-322.

41. Tarek, M. H., et al. (2023). "Microbial source tracking to elucidate the impact of land-use and physiochemical water quality on fecal contamination in a mixed land-use watershed." Science of the Total Environment **872**.

42. Teaf, C. M., et al. (2018). "Toward forensic uses of microbial source tracking." Microbiology Spectrum **6**(1): 10.1128/microbiolspec. emf-0014-2017.

43. Waso, M., et al. (2018). "Microbial source tracking markers associated with domestic rainwater harvesting systems: Correlation to indicator organisms." Environmental research **161**: 446-455.

44. Wood, J., et al. (1998). "Estimation of the relative abundance of different Bacteroides and Prevotella ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequences." Applied and environmental microbiology **64**(10): 3683-3689.

45. Zhou, J., et al. (2022). "The relationship of Megamonas species with nonalcoholic fatty liver disease in children and adolescents revealed by metagenomics of gut microbiota." Scientific Reports **12**(1): 22001.

ACKNOWLEDGEMENT

First, I would like to express profound gratitude to my supervisor, Professor Tatsuya Unno, whose guidance, and support have been pivotal throughout this journey. His discipline, dedication and perseverance are the qualities I aspire to emulate. Also, I would like to thank Professor Hye Yeon Lee and Professor Kyung Hwan Boo for their evaluation and time. Their constructive feedback was helpful during this whole process.

I am grateful to all the members of Microbiome lab, including Dr. Jungman Kim, Gwang-pyo Ko, Hyejun Jo, Sunwoo Lee, Yujin Jeong, Sujin Hyeon, Minwoo Kim, Geon Choi and Youngjin Cheon, who contributed to a great research environment and gave their "꿀" tips whenever needed. Further, I would like to extend my gratitude to Dr. Ho Kyung Song, Jiwon Jeong, and Azka Tauqeer for always being a strong support system in Cheongju. They have contributed to my professional as well as personal growth.

My sincere thanks to Dr. Vineet Singh, without whom this would not have been possible. His immense faith, motivation and support helped me to navigate through the challenges. I would also like to mention my friends in Jeju (Shirjana, Roshan, Tiny, Anupama, Amreen, and Sheher Bano) who always made me feel like home with their amazing food and warm hearts.

Heartfelt gratitude to my friends in India (Kanika, Anushree, Shikha, and Sachin) for always being my greatest cheerleader and providing a reality check when needed. They have shown remarkable confidence in me precisely when I needed it the most.

Lastly, I want to express an immense gratitude to my family for their unconditional love, faith, and unwavering support, which have served as a driving force in achieving all my goals.