Journal of Water & Health

IWA PUBLISHING

Journal of Water and Health Vol 23 No 1, 43 doi: 10.2166/wh.2024.399

Influence of sewage effluent discharge on putative pathogen community in drinking water sources: insights from full-length 16S rRNA gene amplicon sequencing

Jie Zeng, Tomohiro Nakanishi *, Ayato Hara and Sadahiko Itoh

Department of Environmental Engineering, Graduate School of Engineering, Kyoto University, Kyoto University Katsura, Nishikyo, Kyoto 615-8540, Japan
*Corresponding author. E-mail: nakanishi.tomohiro.8r@kyoto-u.ac.jp

TN, 0009-0003-3549-5779

ABSTRACT

The discharge of sewage effluent is a major source of microbial contamination in drinking water sources, necessitating a comprehensive investigation of its impact on pathogenic bacterial communities. This study utilized full-length 16S rRNA gene amplicon sequencing to identify putative pathogenic bacteria and analyze their community structures in drinking water sources subjected to different levels of fecal pollution: urban rivers with low, moderate, and high sewage effluent mixing ratios, and mountain streams with minimal human impact. The sewage effluent itself was also analyzed. Mountain streams primarily harbored environmental pathogens, whereas urban rivers exhibited significantly higher concentrations of fecal indicator bacteria (FIB) (i.e., *Escherichia coli* and *Clostridium perfringens*) along with markedly more diverse enteric pathogens with a higher relative abundance. Furthermore, within urban rivers, the putative pathogen communities displayed significant variation, closely aligning with the sewage effluent mixing ratios. The effectiveness of FIBs as indicators of enteric pathogens was found to be largely dependent on the levels of fecal pollution. This study offers novel insights into the impact of sewage effluent discharge on putative pathogenic bacterial communities with enhanced species-level resolution.

Key words: drinking water sources, fecal indicator bacteria, mountain streams, pathogenic bacterial community, urban rivers

HIGHLIGHTS

- Characterization of pathogenic bacterial communities in water sources using full-length 16S sequencing.
- Distinct pathogen communities are identified between mountain streams and urban rivers.
- Pathogen communities in urban rivers correlated with sewage effluent mixing ratios.
- The effectiveness of fecal indicator bacteria varied across water sources.

INTRODUCTION

Waterborne outbreaks related to drinking water remain a significant public health concern. Urban rivers, essential sources of water supply, are particularly vulnerable to microbial contamination due to extensive human activities. Municipal wastewater treatment plants (WWTPs) are a major source of pollution in these rivers, discharging treated sewage that can still harbor a wide range of microbes and pathogens (Yang *et al.* 2022). For example, Li *et al.* (2015) detected 37 bacterial pathogen species in sewage effluents. Similarly, Chen *et al.* (2024) analyzed 113 metagenomes from WWTP effluents worldwide and found prevalent bacterial pathogens such as *Acinetobacter baumannii* and *Bacteroides uniformi*. Therefore, it is crucial to assess the contributions of WWTP effluents to the microbial risks, especially in water sources situated near the effluent outflows.

Microbial risk management in drinking water sources usually relies on fecal indicator bacteria (FIB), such as *Escherichia coli* and *Clostridium perfringens* (WHO 2017). However, these indicators do not provide direct information on the presence of pathogens and have several limitations such as prevalence in non-intestinal environments (Ishii *et al.* 2006; Devane *et al.* 2020; Li *et al.* 2021), varied suitability across water types and pollution sources (Goh *et al.* 2019; Richiardi *et al.* 2023) and oversight of environmental pathogens that could also pose health risks (Falkinham 2020). Relying solely on FIB for predicting health risks may not be sufficient, which emphasizes the demand for direct pathogen monitoring in drinking water sources.

Additionally, profiling the entire pathogen community using high-throughput sequencing (HTS), compared to specific pathogens, is particularly beneficial.

Several studies using HTS have found that sewage effluent discharge can alter the pathogen communities in receiving waters (Ibekwe *et al.* 2016; Liu *et al.* 2018; Pascual-Benito *et al.* 2020). For instance, Liu *et al.* (2018) observed that the pathogenic bacterial compositions at the site receiving WWTP effluent were significantly different from those in natural and suburban areas. Pascual-Benito *et al.* (2020) reported higher abundances of genera containing pathogenic species, such as *Escherichia*, *Shigella*, *Klebsiella*, and *Enterobacter*, in an intermittent mediterranean stream after receiving the WWTP effluent. However, these studies did not specify the mixing ratios of effluent to river flow (pollution levels). Understanding the relationship between pollution levels and the degree of impact on pathogenic bacterial communities can provide valuable insights for microbial risk management in drinking water sources, which requires further investigation.

Furthermore, these studies conducted amplicon sequencing targeting partial 16S rRNA gene (henceforth called 16S), which inherently provides limited species-level resolution. Since bacterial pathogenicity is often considered at the species level, conventional partial 16S analysis is insufficient for accurately capturing the pathogenic bacterial community. In contrast, recent advancements in long-read sequencing technologies now enable full-length 16S (FL-16S) amplicon sequencing, which covers all nine variable (V1–V9) regions of 16S, significantly improving species-level resolution in many cases. For example, Johnson *et al.* (2019) conducted an *in silico* comparison of 16S sequences across variable regions, demonstrating that nearly 100% of FL-16S sequences could be confidently assigned to the correct species, whereas up to 56% of V4 amplicons failed to achieve such annotation. Similarly, Jeong *et al.* (2021) compared FL-16S and V3–V4 sequencing using 24 human gut microbiota samples, reporting that FL-16S provided superior species identification accuracy and more effectively captured the microbial community alpha-diversity and compositional profiles. Particularly, there has been growing interest in nanopore sequencing, which offers unique capabilities for onsite and real-time analysis, leading to promising approaches in microbial water quality monitoring (Acharya *et al.* 2019; Urban *et al.* 2021). Utilizing a nanopore sequencer for FL-16S sequencing holds great promise for studying the impact of sewage effluent on the pathogenic bacterial community at the species level. However, to the best of our knowledge, its application to drinking water sources has been limited.

Therefore, this study aims (1) to characterize pathogenic bacterial communities across drinking water sources with varying fecal pollution levels, (2) to explore the effects of sewage effluent discharge at different levels on pathogenic bacterial communities in receiving river water, and (3) to assess the effectiveness of FIBs in predicting the occurrence of pathogenic bacterial species. We sampled two types of drinking water sources in Japan: mountain streams minimally impacted by human activities and urban rivers with sewage effluent mixing ratios ranging from low to high. Additionally, sewage effluent from one WWTP was also examined. Pathogenic bacteria at the species level were comprehensively identified from these water bodies using FL-16S nanopore sequencing. This study particularly provides new insights into the pathogenic bacterial communities in drinking water sources with improved species-level resolution and the impact of sewage effluent discharge on these communities.

MATERIALS AND METHODS

Water sampling

Mountain streams minimally affected by human activities were collected from four small mountain streams (MS_A–D) in Shiga and Kyoto Prefecture, Japan, which served as the water sources for small water supply systems (Zeng *et al.* 2024). These streams originated from spring water and were sheltered by forests. For urban rivers impacted by sewage effluent, samples were collected in the Katsura River basin, Japan (Figure 1). The mean mixing ratio of sewage effluent to river flow at each sampling site was roughly estimated (see Text S1 in Supplementary Information). UR_A was located 7 km upstream of the nearest WWTP, with a low mixing ratio of ~3%. UR_B was positioned 5 km downstream of a large WWTP (treatment capacity: 954,000 m^3/day), presenting a high mixing ratio of ~40%. UR_C was situated in the Yodogawa River, characterized by high flow due to the convergence of three rivers, and located 12 km downstream of the nearest WWTP, with a moderate mixing ratio of ~10%. UR_C is also close to the water intake for drinking water treatment plants.

Water samples were collected at MS_A–C once a month from March 2021 to January 2022, and collected at MS_D in March, May, August, and December 2021. The total sample numbers for MS_A–D were 11, 9, 11 and 4, respectively. Four samples from UR_A, and 6 samples from UR_B–C, were taken from August to December, 2021. Sewage effluent samples were collected from the WWTP nearest to UR_B, located upstream, during the same period.

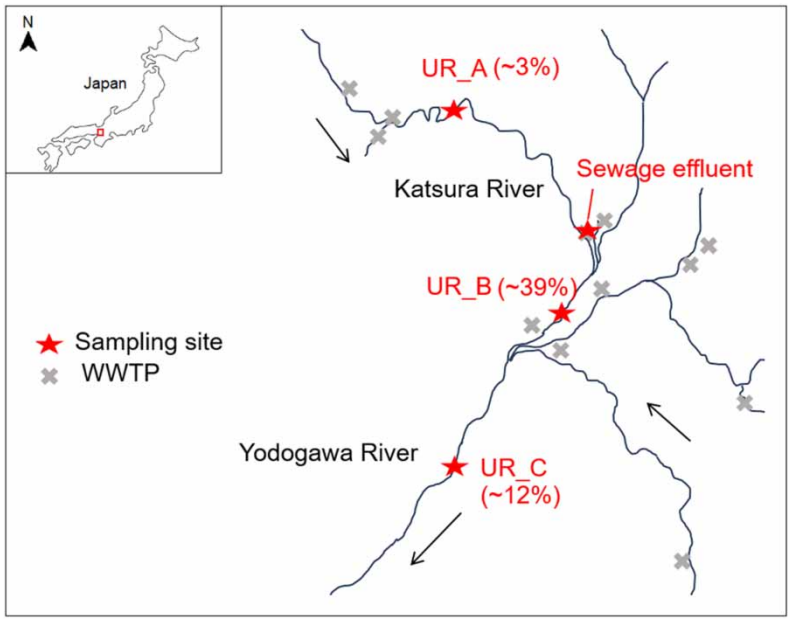

Figure 1 | Sampling sites for urban rivers and sewage effluent. The number in parentheses indicates the mixing ratio of sewage effluent to river flow, which was roughly estimated in this study (see Text S1).

Measurement of water quality

Electrical conductivity (EC) and turbidity were measured by a field-type multi-digital water quality meter (LAQUA WQ-320PC-S, Horiba, Japan), and a turbidity meter (TR-55, Kasahara Chemical Industry, Japan), respectively. The standard plate count (SPC) and heterotrophic plate count (HPC) were determined by the pour plate method using Standard Methods Agar 'DAIGO' and R2A Agar (both from Nihon Pharmaceutical, Japan), respectively. For SPC, plates were counted after incubation at 36 °C for 24 h. For HPC, bacteria were cultured at 20 °C for 7 days.

The *E. coli* concentration was determined by the most probable number (MPN) method using ONPG-MUG broth supplemented with IPTG (ES coli-catch, Eiken Chemical, Japan) based on the manufacturer's protocol. The *C. perfringens* spores were measured by a triple-layered method with modified Handford agar (Eiken Chemical, Japan) (Handford 1974). It should be noted that all black colonies with a diameter of 1–3 mm in the second layer were regarded as *C. perfringens*, even though other species within the *Clostridium* genus could be potentially detected.

DNA extraction

Water samples were transported to the laboratory immediately after sample collection and stored at 4 °C. About 1–2 L of water samples were filtered by 0.22 μm polyethersulfone membrane (Millipore Express PLUS, diameter 47 mm, USA) and the membranes were stored at −80 °C until DNA extraction. DNA was extracted using DNeasy PowerSoil Pro Kit (Qiagen, Germany) as described in our previous study (Zeng *et al.* 2024). All DNA samples were stored at −20 or −80 °C for downstream usage.

Library preparation and nanopore sequencing

16S Barcoding Kit 1–24 (SQK-16S024) provided by Oxford Nanopore Technologies (ONT) was used for PCR amplification. To prepare the reaction solution, 25 μL LongAmp Hot Start Taq 2X Master Mix (New England Biolabs, Japan), 10 μL barcoded primers in an ONT kit, 25 ng of extracted DNA (as measured by Nanodrop, ThermoFisher Scientific), and nuclease-free water were mixed to a total volume of 50 μL. The PCR reactions were conducted using a Takara Thermal Cycler Dice (TaKaRa Bio, Japan) with the following cycles: 1 cycle of 95 °C for 1 min; 25 cycles of 95 °C for 20 s, 50 °C for 30 s and 65 °C for 2 min; 1 cycle of 65 °C for 5 min. It is noted that PCR protocols were evaluated to minimize the PCR bias using mock communities (see Text S2-1, S2-2), and the above protocol was employed (Figure S1).

The quality of PCR products was checked using agarose gel electrophoresis with 2% agarose. If the target band was clearly observed and no unspecific bands occurred, the PCR products were cleaned up using Agencourt AMPure XP beads

(Beckman Coulter) with a beads/sample ratio of 0.6 and 70% fresh ethanol. The cleaned-up amplicons were quantified using a Qubit 3.0 fluorometer (Invitrogen) with a Qubit dsDNA HS Assay Kit. The final sequencing libraries, consisting of 10 μL with 50–100 fmole amplicons, were prepared by pooling barcoded samples at equivalent molar ratios. Further details can be referred to the protocol of '16S Barcoding Kit 1-24 (SQK-16S024)' (Version: 16S_9086_v1_revR_14Aug2019, ONT). Libraries were sequenced with MinION MK1B and R9.4.1 flow cell (ONT) for about 48 h.

Processing of sequence data

The raw sequencing data in fast5 format were basecalled using the super high accuracy mode with a Qscore threshold of 10, and then demultiplexed by Guppy v.5.0.16 (ONT). The demultiplexed reads of 1,300–1,700 bp were retained by seqkit seq command (Shen *et al.* 2016), and aligned to the curated 'Microbial Identification using rRNA Operon Region' (MIrROR) database (Seol *et al.* 2022) with minimap2 (-cx map-ont -N 19 –secondary = yes, v2.24) (Li 2018). Before use, the non-redundant MIrROR database was prepared as demonstrated in Text S3. Reads with low coverage (hit length/reference sequence length <20%) and low percent of identity (<90%) were considered poor quality (e.g., PCR artifacts), and subsequently removed. Top hits of the remaining reads were utilized for taxa annotation to the species level. Species with low abundance were removed using a 0.1% cut-off threshold. These filtering steps were employed to minimize the impact of PCR artifacts and high error rates in nanopore reads. The selection of minimap2 software and the subsequent cut-off thresholds were determined using mock communities, as shown in Text S2 and Table S4. Rarefaction curves were then plotted using a vegan package in R (Oksanen *et al.*, 2022; R Core Team 2022). Finally, the read counts of each species were transformed into relative abundance using Phyloseq package (McMurdie & Holmes 2013).

Pathogenic bacterial community analysis

A newly published pathogen list by Bartlett *et al.* (2022), containing 1,513 human-infective pathogenic bacteria species, was employed to screen the pathogenic bacteria species from all detected species. Pathogenic species within genera *Klebsiella* (Graf *et al.* 2021), *Streptococcus* (Ma *et al.* 2023), *Clostridium* (Li *et al.* 2021), *Escherichia* (Graf *et al.* 2021), *Enterobacter* (Graf *et al.* 2021), *Bacteroides* (Li *et al.* 2021), *Clostridioides* (Hernández *et al.* 2019), *Yersinia* (Numberger *et al.* 2019), and *Parabacteroides* (Cui *et al.* 2022) were classified as enteric species in this study. The pathogen richness was evaluated by the species or genus number of pathogenic bacteria in a sample. The pathogen's relative abundance was evaluated by the total relative abundance of pathogenic bacteria in a sample.

Non-metric multidimensional scaling (NMDS) analysis and analysis of similarities (ANOSIM) tests were conducted to identify the differences in the pathogenic community among three types of water, three sampling sites for urban rivers, as well as between seasons. For seasonal differences, water samples were classified into summer (November to April) and winter (May to October) seasons, with the summer months typically experiencing higher precipitation in Japan. NMDS analysis and ANOSIM test were conducted based on Bray–Curtis distance using the vegan package. The metaMDS function was used to create NMDS plots. The envfit function was used to test for vectors (genera and water quality variables) that significantly correlated with NMDS ordinations ($R^2 \geq 0.2$) with a permutation of 999 times. For the water quality variables, bacterial concentrations were logarithmically transformed with a base of 10. The ANOSIM test was conducted with the anosim function with a permutation of 999 times. All significant tests were conducted with a confidence level of 95%.

Statistical analysis

For analysis in pathogen richness and relative abundance, Welch's *t*-test was used for the differences between seasons using Origin (OriginLab). One-way Analysis of Variance (ANOVA) was conducted to investigate differences among water types in R. If the ANOVA results were significant, pairwise *t*-tests with Holm's correction were performed to identify differences between every two groups. The relationship between FIB concentrations and pathogen relative abundances was examined by Spearman's ranks correlation coefficient (r_s) analysis using Origin. All these analyses were conducted with a confidence level of 95%.

RESULTS

Basic water quality

Basic water quality data are summarized in Table S5. Urban rivers showed higher levels of EC, SPC and HPC than mountain streams. Compared to sewage effluent, urban rivers presented higher turbidity and lower EC, with comparable SPC and HPC.

The measurement results of FIBs are shown in Table 1. Geometric mean concentrations of *E. coli* and *C. perfringens* in urban rivers ranged 161–591 MPN/100 mL and 30–201 CFU/100 mL, respectively, which were around 1 order of magnitude higher than those in mountain streams. Among UR_A–C, the FIB levels followed the sequence of UR_B > UR_C > UR_A, which mirrored the pattern of sewage effluent mixing ratios. Notably, UR_B exhibited comparable *E. coli* concentration to sewage effluent, which is reasonable given the mixing ratio of ~40% at this location.

Summary of FL-16S nanopore sequencing

A total of 57 water samples were analyzed, and 65,254–749,742 reads were obtained for each sample after demultiplexing, as shown in Table S6. Among demultiplexed reads, 87–97% passed the length filter and showed an average length ranging from 1,430 to 1,450 bp. After removing the reads with low coverage and sequence similarity with reference reads, 38,903–614,912 sequences were assigned to the species level. Following the removal of low-abundance species (<0.1%), the rarefaction curves indicated the datasets were sufficiently large to capture community richness (Figure S2).

Putative pathogenic species detected by FL-16S nanopore sequencing

The detected putative pathogens are demonstrated at both species (Figure 2) and genus level (Figure S3). A total of 79 pathogenic species spanning 45 genera were detected across 57 water samples. In mountain streams, a total of 27, 20, 12, and 14 pathogenic species were detected from sites MS_A–D, respectively. Species within the *Aeromonas*, *Legionella*, *Pseudomonas*, and *Serratia* genera were frequently detected, with a relative abundance ranging from 0.1 to 2.4% (Figure 2(a)). The highest relative abundance was observed for *Aeromonas hydrophila* (2.4%) in MS_B sample collected on October 2021. Enteric species *Yersinia ruckeri*, *Streptococcus pasteurianus* and *Parabacteroides goldsteinii* were sporadically detected in three samples of mountain streams with a relative abundance of 0.6, 0.3 and 0.1%, respectively. The overall positive rate for enteric species was 8.6% across all 35 mountain streams samples (data not shown).

Figure 2(b) demonstrates the putative pathogens detected in urban rivers and sewage effluent. A total of 35 pathogenic species were detected from sewage effluent samples, including genera such as *Streptococcus*, *Pseudomonas*, *Clostridioides*, *Clostridium*, *Escherichia*, *Enterobacter*, *Klebsiella*, *Moraxella*, *Raoultella*, *Bacillus*, *Aeromonas*, and *Acinetobacter*. Among these, 10 enteric species spanning 6 genera were observed. In UR_A, characterized by a low sewage mixing ratio, a total of 11 pathogenic species were detected, including only three enteric species within *Klebsiella* and *Clostridioides*. *Pseudomonas* species were prevalent in UR_A, accounting for 0.1–0.3% of the total bacterial community. In UR_C, with a moderate sewage mixing ratio, 16 pathogenic species were detected, including 4 enteric species within *Klebsiella*, *Clostridioides*, and *Streptococcus*. *Roseomonas* was frequently detected in UR_C. UR_B, with the highest sewage mixing ratio, exhibited a total of 29 pathogenic species, including 11 enteric species across 6 genera (i.e., *Streptococcus*, *Clostridioides*, *Escherichia*, *Enterobacter*, *Klebsiella*, *Bacteroides*). *Acinetobacter* (0.2–2.7%), *Moraxella* (1.9%), *Raoultella* (1.4%), and *Streptococcus* (1.0%) showed high relative abundance in certain samples from UR_B, while these species were also found in sewage effluent samples. *Acinetobacter* and *Aeromonas* were detected in all three river sites.

Table 1 | FIB concentrations in mountain streams, urban rivers and sewage effluent

	E. coli **(MPN/100 mL)**			***C. perfringens*** **(CFU/100 mL)**		
Water	**Positive samples (/total samples)**	**Min.–Max.**	**Geometric mean**	**Positive samples (/total samples)**	**Min.–Max.**	**Geometric mean**
MS_A	9/11	9–110	24	11/11	0.1–7	1
MS_B	8/9	2–110	10	8/9	0.4–7	1
MS_C	9/11	1–46	12	10/11	0.3–10	2
MS_D	4/4	1.5–24	4	4/4	0.1–2	1
UR_A	3/3	120–230	161	3/3	12–167	30
UR_B	5/5	90–2,400	591	5/5	50–1,030	201
UR_C	5/6	46–2,100	198	5/5	7–167	38
Sewage effluent	4/5	120–1,500	504	5/5	260–1,100	629

Note: The detection limit for *E. coli* was 0.3 MPN/100 mL for mountain streams and 30 MPN/100 mL for urban river and sewage effluent. The detection limit for *C. perfringens* was 0.07 CFU/100 mL for mountain streams.

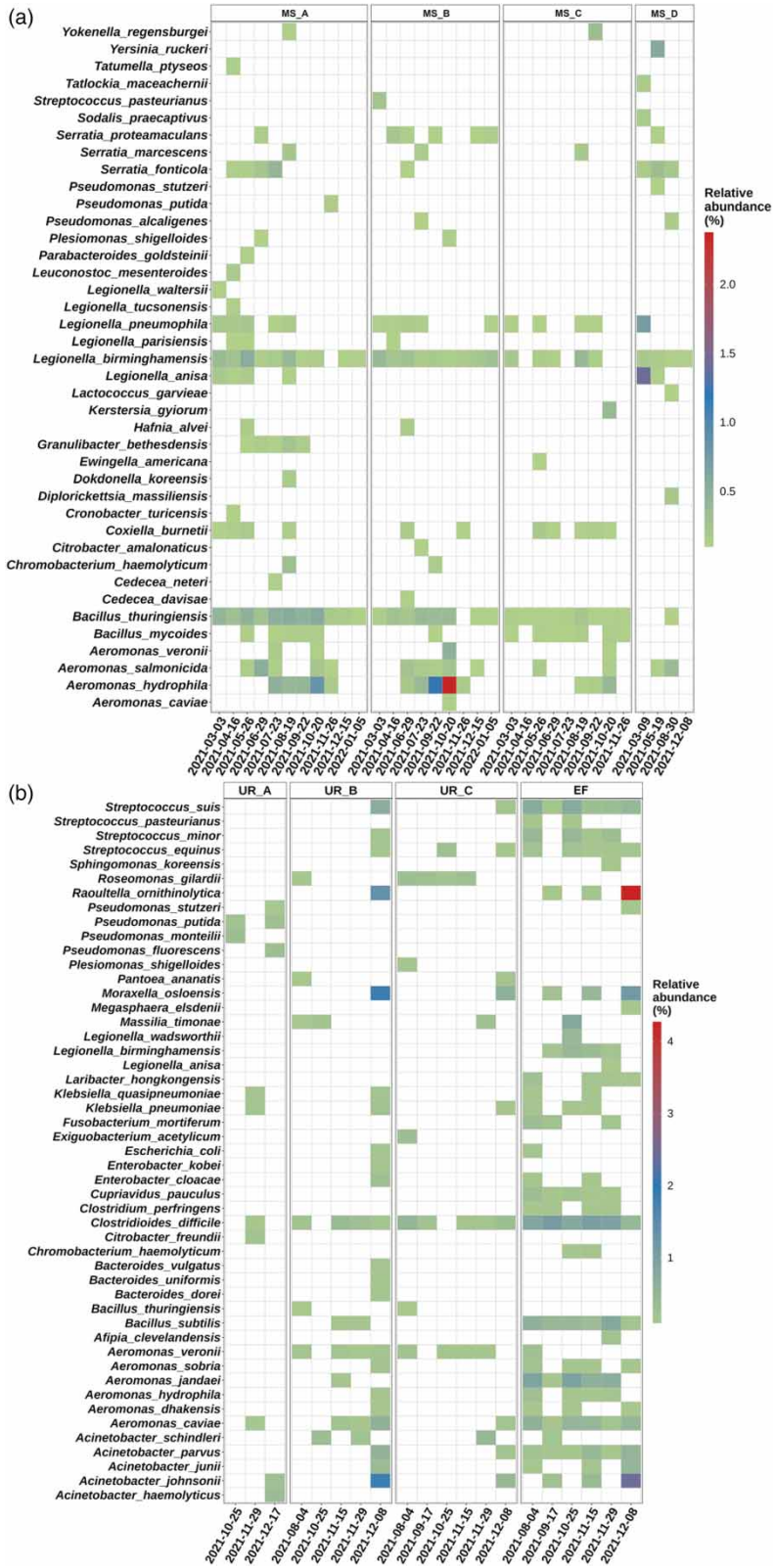

Figure 2 | Putative pathogenic species in mountain streams (a), urban rivers and sewage effluent (b), EF indicates sewage effluent. The white color indicates not detected.

Pathogen richness and relative abundance in different water bodies

The richness and relative abundance of total and enteric pathogenic bacteria are demonstrated in Figure 3. For total pathogenic bacteria, the average richness and relative abundance in mountain streams and urban rivers ranged from 3 to 7 species and 0.6–2.2%, respectively (Figure 3(a) and 3(b)). These values were significantly lower than sewage effluent, which had an average richness of 18 species and a relative abundance of 6.7% (Holm-adjusted $p < 0.001$). No significant differences in richness and relative abundance were observed between mountain streams and urban rivers (Holm-adjusted $p > 0.05$).

For enteric pathogenic species, few were found at each site for mountain streams, with less than one species and <0.2% in relative abundance each on average (Figure 3(c) and 3(d)). Importantly, urban rivers exhibited a greater number of enteric species with higher relative abundance than mountain streams (Holm-adjusted $p < 0.01$). Among UR_A–C, the average values followed the sequence UR_B (2.5 species, 0.58%) >UR_C (1.5 species, 0.35%) >UR_A (<1 species, 0.11%), although these differences were not statistically tested due to the small datasets. Sewage effluent samples harbored more diverse and abundant species than urban rivers (Holm-adjusted $p < 0.001$), which presented six enteric species with a relative abundance of 2.0% on average.

The seasonal variations in pathogen richness and relative abundance for mountain streams and urban rivers were investigated, as shown in Figure 4. Mountain streams (Figure 4(a) and 4(b)) exhibited significantly higher pathogen richness and relative abundance during the summer months ($p < 0.001$). Conversely, for urban rivers (Figure 4(c) and (d)), all average values for winter were higher than those for summer; however, the differences were not statistically significant ($p > 0.05$).

The relationships between cultivable FIB concentrations and total/enteric pathogen relative abundance were also explored (Figure 5). For mountain streams, no statistically significant rank correlations were observed between FIBs and total or enteric pathogens (Figure 5(a) and 5(b), $p > 0.05$). However, for urban rivers (Figure 5(c) and 5(d)), strong correlations

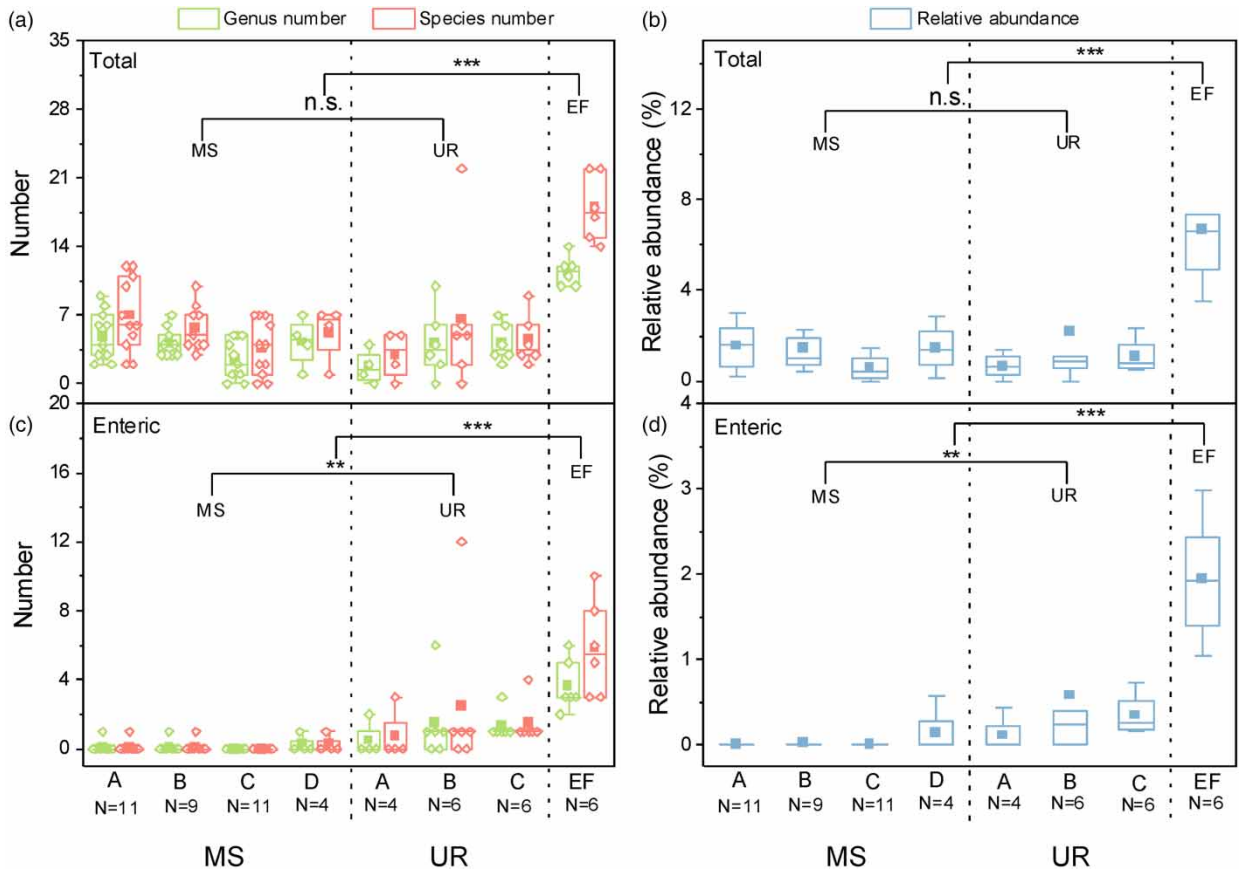

Figure 3 | Richness and relative abundance of total (a, b) and enteric pathogenic bacteria (c, d). Negative samples were included for calculations. MS, UR, and EF indicate mountain stream, urban river and sewage effluent, respectively. The shape of square indicates average value. The differences among water types are shown: 'n.s.' means not significant; '**' and '***' mean $p < 0.01$ and 0.001, respectively.

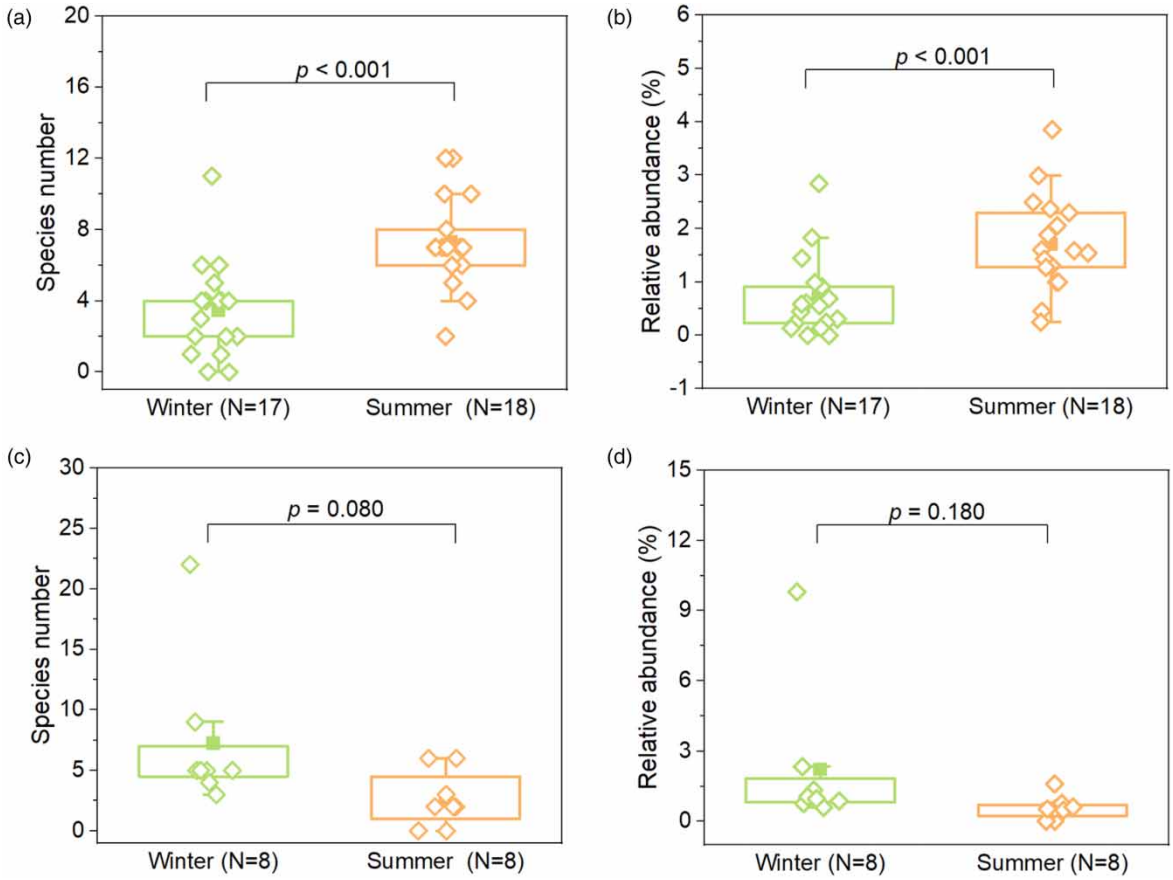

Figure 4 | The seasonal variations in pathogen richness and relative abundance for mountain streams (a, b) and urban river (c, d). The shape of square indicates average value.

were observed between *C. perfringens* and pathogen relative abundance (both total and enteric), with coefficients of 0.64 and 0.76, respectively ($p < 0.05$). Although correlations were also observed with *E. coli* concentrations in urban rivers, the coefficients were lower and *p*-values were higher than those with *C. perfringens*.

Differences in the structure of pathogenic bacterial community

The differences in the pathogenic bacterial composition collapsed at the genus level were investigated using NMDS and ANOSIM methods (Figure 6). As shown in Figure 6(a), urban river ellipse greatly overlapped with that of sewage effluent, with an ANOSIM-R value of merely 0.22 ($p = 0.039$). Among sites UR_A–C, slight but significant separation was observed (ANOSIM: $r = 0.29$, $p = 0.028$), and the distance to the sewage effluent followed the trend UR_A > UR_C > UR_B. UR_B, with the highest sewage mixing ratio, overlapped with sewage effluent. Particularly, the samples collected on 8 December 2021 showed a high similarity to the effluent sample collected on the same day (Figure S4).

Pathogenic communities in urban rivers were significantly different from those in non-human-impacted mountain streams (Figure 6(b)), as urban river ellipse was significantly separated from that of mountain streams (ANOSIM: $r = 0.678$, $p = 0.001$). These differences were primarily captured by the NMDS1 dimension. Water quality parameters, including EC, SPC, HPC, *E. coli* and *C. perfringens* concentrations, were positively correlated with the NMDS1 dimension. Moreover, the clustering of urban rivers was driven by many bacteria in the NMDS1 dimension, including *Acinetobacter*, *Pseudomonas*, *Clostridioides*, *Roseomonas*, and *Massilia*, while the clustering of mountain streams was only significantly influenced by *Bacillus*, *Legionella* and *Serratia*.

Seasonal variation trends in pathogenic bacterial compositions were different between urban rivers and mountain streams. Summer and winter clusters almost completely overlapped for river water (Figure 6(c); ANOSIM: $r = 0.051$, $p = 0.256$).

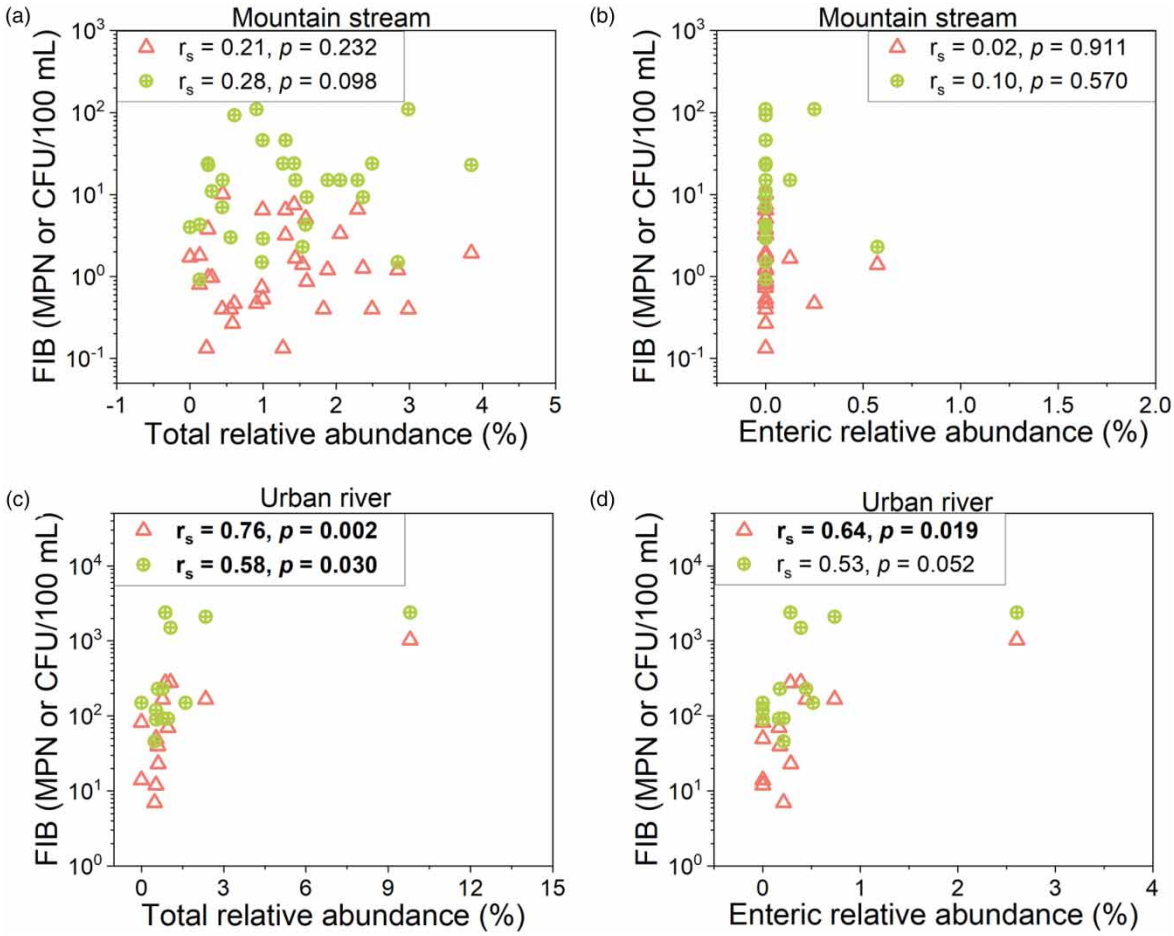

Figure 5 | Relationships between total (a, c) or enteric (b, d) pathogen relative abundance and FIB concentration for mountain streams (a, b) and urban rivers (c, d). FIB-negative samples were included for analysis but not shown in the figures. Green symbols indicate *E. coli* and red symbols indicate *C. perfringens*. r_s indicates Spearman's rank correlation coefficient. r_s in bold indicates significant relationship.

However, they were slightly separated for mountain streams (ANOSIM: $r = 0.286$, $p = 0.001$), with *Aeromonas*, *Bacillus*, *Serratia* and SPC contributing to the summer cluster.

DISCUSSION

Nanopore sequencing as a screening tool for bacterial pathogens

Modern practices of water quality management rely on measurement of FIB. Despite their usefulness, FIB measurements cannot answer 'what pathogens are present'. Comprehensive detection of pathogenic bacteria through FL-16S nanopore sequencing provides a valuable complement to routine FIB measurements. Nanopore sequencers offer several advantages, including relatively low initial cost, portability, and real-time analysis capabilities. These features enhance their potential to be incorporated into routine analysis (Acharya *et al.* 2019; Urban *et al.* 2021).

One drawback of nanopore sequencers is their lower read accuracy compared to other sequencing platforms (Akacin *et al.* 2022). To reduce misclassification caused by sequencing errors, we selected an appropriate data analysis method by using standard mock community samples, which consisted of known bacterial species at predetermined ratios (Tables S1 and S2). Specifically, we compared the performance of two software tools: Minimap2 (Li 2018), a read mapping tool optimized for erroneous nanopore reads but without error-correction, and NanoCLUST (Rodriguez-Perez *et al.* 2021), which corrects errors through read clustering. Additionally, we optimized relative abundance cutoffs, ranging from 0.001 to 1%, to reduce false positive detection. As detailed in Text S2-3 and Table S4, using Minimap2 with an abundance cut-off of 0.1% achieved

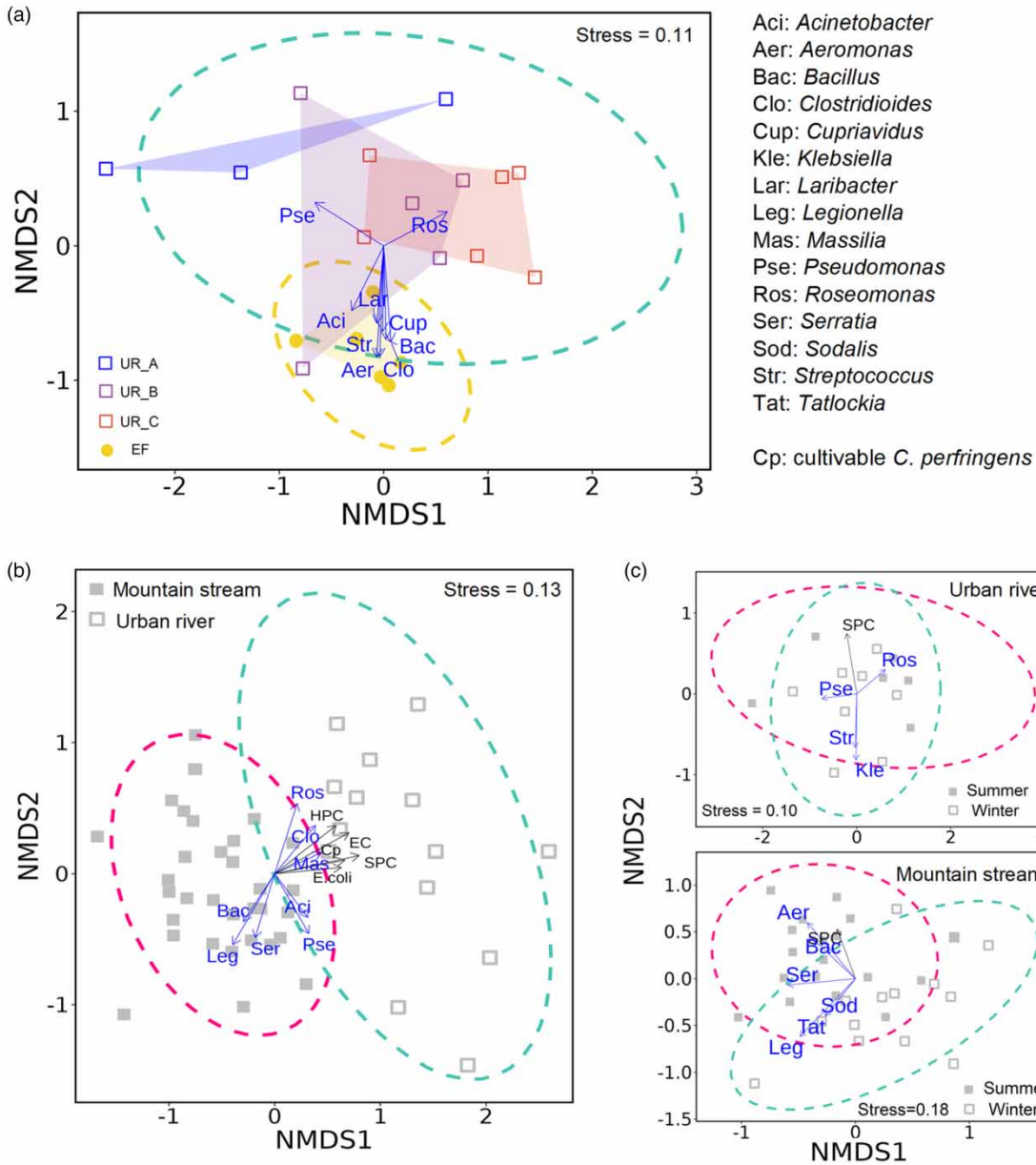

Figure 6 | NMDS plots of pathogenic bacterial communities collapsed at the genus level. (a) Urban rivers and WWTP effluent; (b) urban rivers and mountain streams. Genus (blue) and water quality (black) vectors indicate the direction and strength of significant correlations within the NMDS ordinations. Ellipses were plotted with a confidence interval of 95%.

acceptable accuracy and sensitivity in species detection. Therefore, although not perfect (Table S4, Figure S5), our analysis approach could significantly address the issues of sequencing errors.

It is important to acknowledge that differentiating closely related species based on 16S sequence is inherently challenging due to the limited interspecies variability within this gene region. For example, pathogenic *E. coli* and *Shigella* can share over 99% sequence similarity in FL-16S sequences (Devanga Ragupathi *et al.* 2018), complicating conclusive identification. Additionally, the pathogenicity inferred in our approach remains inconclusive, as 16S sequences do not provide direct information about pathogenic traits. Therefore, our FL-16S sequencing approach should be viewed as a preliminary screening tool for the broad detection of bacterial pathogens. For more definitive results, the detection of other gene regions is highly recommended.

Mountain streams and urban rivers had distinct pathogenic bacterial communities

First, we compared the composition of putative pathogens between mountain streams (minimally impacted by human activities) and urban rivers (differentially affected by sewage effluent). Although urban rivers and mountain streams presented similar richness (3–7 species) and relative abundance (0.6–2.2%) of total pathogenic bacteria, their compositions were significantly different (Figure 6(b)).

In mountain stream samples, environmental pathogens constituted the majority of the pathogenic communities. Notably, *Bacillus*, *Legionella* and *Serratia* were exclusively prevalent in mountain streams. Conversely, enteric pathogenic species occurred sporadically with very low frequency and abundance (Figure 3), presumably originating from wildlife.

In contrast, urban rivers exhibited a much more diverse array of putative enteric pathogens with higher relative abundance, including species within *Klebsiella*, *Streptococcus*, *Clostridium*, *Escherichia*, *Bacteroides*, *Enterobacter*, and *Clostridioides*. Especially, *Clostridioides difficile* was frequently detected, significantly driving the compositional differences between mountain streams and urban rivers (Figure 6(b)), likely due to its resistance to environmental stress through spore formation. While environmental pathogens were also observed in urban rivers, their compositions differed from those in mountain streams. For example, *Acinetobacter*, *Pseudomonas*, *Roseomonas*, and *Massilia* were exclusively found in urban rivers.

Sewage effluent discharge influenced the pathogenic bacterial community in urban rivers

This study clearly demonstrated that sewage effluent significantly impacted the pathogenic bacterial communities in urban rivers (Liu *et al.* 2018; Pascual-Benito *et al.* 2020). FIB levels in urban rivers were much higher than in non-human-impacted mountain streams (Table 1), indicating fecal pollution. The NMDS analysis (Figure 6(a)) showed a significant overlap in pathogenic community composition between urban rivers and sewage effluents, with UR_B (the site with the highest sewage mixing ratio) placed nearest to sewage effluent outfall. Moreover, the occurrence of putative enteric pathogens was more frequent in urban rivers than in mountain streams (Figure 3(c) and 3(d)). Although we categorized pathogenic species from nine genera as enteric pathogens to distinguish them from environmental pathogens, their specific sources, human or animal, were not verified in this study. Nonetheless, these enteric species detected in the urban rivers, especially in UR_B, are most likely of human origin, as no major animal sources, such as livestock farms, are present within the service area of the nearby WWTPs. These findings suggest that the pathogenic bacterial community in urban rivers was most likely impacted by sewage effluent discharge.

Sewage effluent usually contains higher pathogen levels, as demonstrated in this study and previous studies (Li *et al.* 2015; Numberger *et al.* 2019), along with nutrients (Xie *et al.* 2022; Lu *et al.* 2023). The discharge of sewage effluent could introduce 'new' pathogens into receiving waters, and some of these pathogens may survive and even proliferate (Xie *et al.* 2022), thus contributing to shaping the pathogenic bacterial community. These findings were also consistent with previous researches indicating that the discharge of sewage effluent could influence and even shift the overall bacterial community in receiving waters (García-Armisen *et al.* 2014; Mansfeldt *et al.* 2020; Wang *et al.* 2022). Our results reconfirmed these observations with improved species-level resolution using the FL-16S amplicon sequencing approach.

Mixing ratios of sewage effluent caused variable impacts on pathogenic bacterial communities in urban rivers

This study also observed that the pathogenic bacterial communities in urban rivers varied depending on the mixing ratio of the sewage effluent. Some putative pathogens observed in sewage effluent, e.g., *Streptococcus* and *Moraxella*, exhibited relative abundance gradients corresponding to the mixing ratios (i.e., UR_B > UR_C > UR_A) in urban rivers (Figure 2). Additionally, the mean richness and proportion of putative pathogens followed the sequence of mixing ratios (Figure 3). Most importantly, NMDS analysis demonstrated clear separation among pathogenic communities in UR_A – C, with their similarity to the sewage effluent mirroring the trend of mixing ratios (Figure 6(a)). These results indicated that the impact of sewage effluent discharge on pathogenic community increased with the mixing ratio. A previous study at the genus-level resolution also reported the response of pathogen community to the gradients of pollution levels in a typical wastewater-receiving river (Yang *et al.* 2020).

It is of great importance to investigate the thresholds of effluent mixing ratio for significantly increasing the microbial risk in urban rivers. Ruprecht *et al.* (2021) explored the thresholds of WWTP effluent concentration required to induce shifts in the overall bacterial community, reporting that concentrations above 10% caused significant shifts, while 1% had no significant effect. Consistent with these findings, this study observed that the pathogenic community structures of UR_B (~40% mixing ratio) and UR_C (~10%) were much closer to those in sewage effluent than UR_A (3%) (Figure 6(a)). Although

unable to establish a quantitative relationship, this study demonstrated the potential response of the pathogenic community to gradients of sewage effluent mixing ratios: a ratio around 1% may not significantly influence the pathogenic bacterial community, while ratios above 10% may have a significant impact.

However, several limitations should be noted: (i) the mixing ratios estimated in this study did not account for the differential decay or proliferation of microbes after discharge, which could be significant (Wakelin *et al.* 2008; Price *et al.* 2018; Xie *et al.* 2022); (ii) this study did not assess the microbial risk posed by sewage effluent discharge, and linking risk levels with effluent mixing ratios is recommended in further research.

Seasonal variations in pathogenic bacterial community

We also explored the seasonal variation patterns of pathogenic bacterial community, as recognizing the seasonal dynamics of pathogens is crucial for predicting and managing microbial risks. Different patterns were observed across different water types. During the summer months (May to October), mountain streams exhibited a higher diversity and relative abundance of putative pathogens compared to the winter months (Figure 4). This variation is likely attributed to increased precipitation and elevated water temperatures in the summer; increased precipitation enhances runoff into these streams, introducing a wide array of microorganisms, while higher water temperatures facilitate the growth of certain bacterial species (Hirotani *et al.* 1999; Coffey *et al.* 2018). These factors collectively contribute to a higher SPC and the relative abundance of pathogenic species within the genera *Aeromonas*, *Bacillus*, and *Serratia* (Figure 6(c)).

Conversely, urban rivers appeared to harbor more pathogens during the winter months, although this trend was not statistically significant. Several combined factors may contribute to this potential seasonal variation. Firstly, in the investigated urban rivers, river flow levels significantly decrease due to reduced precipitation during the winter months. Since the flow rate of sewage effluent discharge remains relatively constant, this can result in temporarily higher effluent mixing ratios, leading to an increase in pathogen levels during winter. Additionally, lower water temperatures and reduced sunlight in winter may enhance the persistence of pathogens (Coffey *et al.* 2018). These findings also suggest that the impacts of sewage effluent discharge on pathogen communities in urban rivers may exhibit seasonality. Indeed, previous studies have demonstrated that the effect of WWTP effluents on the overall bacterial community can be seasonal or weather-related (García-Armisen *et al.* 2014; Price *et al.* 2018).

E. coli and *C. perfringens* as indicators for fecal pathogenic bacteria

FIB may not reliably predict enteric pathogen presence across diverse pollution scenarios and water types (Goh *et al.* 2019; Richiardi *et al.* 2023). We therefore assessed two common FIBs (*E. coli* and *C. perfringens*) for their performance in predicting the occurrence of enteric bacterial pathogens.

In mountain streams, no significant correlations were observed between the levels of FIBs (i.e., *E. coli* and *C. perfringens*) and enteric pathogens, with FIBs typically occurring in the absence of enteric pathogens (Figure 5(b)). This suggests that in less fecally-polluted environments, detection of FIBs, even at concentrations higher than 1 MPN or CFU/100 mL, does not necessarily indicate the presence of enteric pathogens. This limited predictive ability may arise from the fact that FIBs can also originate from natural environments (Ishii *et al.* 2006; Devane *et al.* 2020; Zeng *et al.* 2024). Conversely, in urban rivers polluted with sewage effluents, FIB levels significantly correlated with the abundance of enteric pathogens (Figure 5(d)). This could be attributed to the predominance of FIBs originating from sewage effluents over those from natural sources, thereby improving their indicative performance. Consequently, our findings emphasize that the efficacy of FIBs in indicating the presence of enteric pathogens may be highly dependent on the levels of fecal pollution and may be particularly unreliable in environments with minimal levels of fecal contamination.

The rank correlation of *C. perfringens* with putative pathogens was comparable to, or slightly higher than, that of *E. coli*, suggesting *C. perfringens* may serve as a more useful indicator of pathogenic bacteria for sewage effluent-impacted waters. However, studies on its reliability as a general fecal pathogen indicator have produced mixed results. For example, *C. perfringens* showed stronger predictive value for pathogens than *E. coli* in tropical environments (Viau *et al.* 2011), whereas in rivers contaminated with raw sewage, its indicative performance was inferior to that of *E. coli* or F-RNA phage (Devane *et al.* 2014). The specific reasons underlying the better performance of *C. perfringens* in the rivers studied remain unclear. Nevertheless, the difference in environmental persistence between *E. coli* and *C. perfringens* may play a role. While *C. perfringens* may reflect historical contamination due to higher persistence (via spore-forming capability), *E. coli* is generally more responsive to recent fecal contamination but may exhibit faster die-off under certain environmental

conditions, potentially explaining the weaker correlation observed in this study. These results suggest that the effectiveness of FIB may vary depending on the specific characteristics of the water body, the nature and distribution of contamination sources, the types of target pathogens, and various environmental conditions. Further research is warranted to examine these factors in depth.

CONCLUSION

This study characterized the putative pathogenic bacterial communities in drinking water sources with varying fecal pollution levels using FL-16S amplicon sequencing. A comparison of putative pathogens in mountain streams and urban rivers revealed similar richness and relative abundance but different compositions: mountain streams predominantly harbored environmental pathogens, while urban rivers, impacted by sewage effluent discharge, contained diverse enteric pathogens. The discharge of sewage effluent was found to alter the composition of pathogenic communities in urban rivers, particularly increasing the diversity and proportion of enteric pathogens. Furthermore, within urban rivers, putative pathogenic communities displayed significant variation, closely aligning with the sewage effluent mixing ratios. Moreover, the efficacy of FIBs (i.e., *E. coli* and *C. perfringens*) in indicating the presence of enteric pathogens was largely dependent on the levels of fecal pollution. Overall, this study provides new insights into the influences of sewage effluent discharge on the pathogenic bacterial community.

ACKNOWLEDGEMENTS

The authors profoundly appreciate the individuals involved in the small water supply systems for their continuous support in facilitating the water sampling process. This work was partially supported by JSPS KAKENHI Grant Number 22H01626, and a research grant from the Takahashi Industrial and Economic Research Foundation of Japan.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Acharya, K., Khanal, S., Pantha, K., Amatya, N., Davenport, R. J. & Werner, D. (2019) A comparative assessment of conventional and molecular methods, including MinION nanopore sequencing, for surveying water quality, *Sci. Rep.*, **9** (1), 15726. https://doi.org/10.1038/s41598-019-51997-x.

Akacin, I., Ersoy, S., Doluca, O. & Gungormusler, M. (2022) Comparing the significance of the utilization of next generation and third generation sequencing technologies in microbial metagenomics, *Microbiol. Res.*, **264**, 127154. https://doi.org/10.1016/j.micres.2022.127154.

Bartlett, A., Padfield, D., Lear, L., Bendall, R. & Vos, M. (2022) A comprehensive list of bacterial pathogens infecting humans, *Microbiology*, **168** (12), 001269. https://doi.org/10.1099/mic.0.001269.

Chen, K., Liang, J., Wang, Y., Tao, Y., Lu, Y. & Wang, A. (2024) A global perspective on microbial risk factors in effluents of wastewater treatment plants, *J. Environ. Sci. (China)*, **138**, 227–235. https://doi.org/10.1016/j.jes.2023.04.012.

Coffey, R., Paul, M. J., Stamp, J., Hamilton, A. & Johnson, T. (2018) A review of water quality responses to Air temperature and precipitation changes 2: nutrients, algal blooms, sediment, pathogens, *J. Am. Water Resour. Assoc.*, **55** (4), 844–868. https://doi.org/10.1111/1752-1688.12711.

Cui, Y., Zhang, L., Wang, X., Yi, Y., Shan, Y., Liu, B., Zhou, Y. & Lü, X. (2022) Roles of intestinal parabacteroides in human health and diseases, *FEMS Microbiol. Lett.*, **369** (1). fnac072. https://doi.org/10.1093/femsle/fnac072.

Devane, M. L., Moriarty, E. M., Wood, D., Webster-Brown, J. & Gilpin, B. J. (2014) The impact of major earthquakes and subsequent sewage discharges on the microbial quality of water and sediments in an urban river, *Sci. Total Environ*, **485–486**, 666–680. https://doi.org/10.1016/j.scitotenv.2014.03.027.

Devane, M. L., Moriarty, E., Weaver, L., Cookson, A. & Gilpin, B. (2020) Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring, *Water Res.*, **185**, 116204. https://doi.org/10.1016/j.watres.2020.116204.

Devanga Ragupathi, N. K., Muthuirulandi Sethuvel, D. P., Inbanathan, F. Y. & Veeraraghavan, B. (2018) Accurate differentiation of *Escherichia coli* and *Shigella* serogroups: challenges and strategies, *New Microbes and New Infections*, **21**, 58–62. https://doi.org/10.1016/j.nmni.2017.09.003.

Falkinham III, J. O. (2020) Living with *Legionella* and other waterborne pathogens, *Microorganisms*, **8** (12), 2026. https://doi.org/10.3390/microorganisms8122026.

García-Armisen, T., Inceoglu, O., Ouattara, N. K., Anzil, A., Verbanck, M. A., Brion, N. & Servais, P. (2014) Seasonal variations and resilience of bacterial communities in a sewage polluted urban river, *PLoS One*, **9** (3), e92579. https://doi.org/10.1371/journal.pone.0092579.

Goh, S. G., Saeidi, N., Gu, X., Vergara, G. G. R., Liang, L., Fang, H., Kitajima, M., Kushmaro, A. & Gin, K. Y. (2019) Occurrence of microbial indicators, pathogenic bacteria and viruses in tropical surface waters subject to contrasting land use, *Water Res.*, **150**, 200–215. https://doi.org/10.1016/j.watres.2018.11.058.

Graf, J., Ledala, N., Caimano, M. J., Jackson, E., Gratalo, D., Fasulo, D., Driscoll, M. D., Coleman, S. & Matson, A. P. (2021) High-resolution differentiation of enteric bacteria in premature infant fecal microbiomes using a novel rRNA amplicon, *mBio*, **12** (1), e03656-20. https://doi.org/10.1128/mBio.03656-20.

Handford, P. M. (1974) A new medium for the detection and enumeration of clostridium perfringens in foods, *J. Appl. Bacteriol.*, **37** (4), 559–570. https://doi.org/10.1111/j.1365-2672.1974.tb00481.x.

Hernández, M., de Frutos, M., Rodríguez-Lázaro, D., López-Urrutia, L., Quijada, N. M. & Eiros, J. M. (2019) Fecal microbiota of toxigenic clostridioides difficile-associated diarrhea, *Front. Microbiol.*, **9**, 3331. https://doi.org/10.3389/fmicb.2018.03331.

Hirotani, H., Sese, C. & Kagawa, H. (1999) Correlations of *Aeromonas hydrophila* with indicator bacteria of water quality and environmental factors in a mountain stream, *Water Environ. Res.*, **71** (2), 132–138. https://doi.org/10.2175/106143099x121544.

Ibekwe, A. M., Ma, J. & Murinda, S. E. (2016) Bacterial community composition and structure in an urban river impacted by different pollutant sources, *Sci. Total Environ*, **566–567**, 1176–1185. https://doi.org/10.1016/j.scitotenv.2016.05.168.

Ishii, S., Ksoll, W. B., Hicks, R. E. & Sadowsky, M. J. (2006) Presence and growth of naturalized *Escherichia coli* in temperate soils from lake superior watersheds, *Appl. Environ. Microbiol.*, **72** (1), 612–621. https://doi.org/10.1128/AEM.72.1.612-621.2006.

Jeong, J., Yun, K., Mun, S., Chung, W. H., Choi, S. Y., Nam, Y. D., Lim, M. Y., Hong, C. P., Park, C., Ahn, Y. J. & Han, K. (2021) The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology, *Sci. Rep.*, **11** (1), 1727. https://doi.org/10.1038/s41598-020-80826-9.

Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E. & Weinstock, G. M. (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis, *Nat. Commun.*, **10** (1), 5029. https://doi.org/10.1038/s41467-019-13036-1.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*, **34** (18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, B., Ju, F., Cai, L. & Zhang, T. (2015) Profile and fate of bacterial pathogens in sewage treatment plants revealed by high-throughput metagenomic approach, *Environ. Sci. Technol.*, **49** (17), 10492–10502. https://doi.org/10.1021/acs.est.5b02345.

Li, E., Saleem, F., Edge, T. A. & Schellhorn, H. E. (2021) Biological indicators for fecal pollution detection and source tracking: a review, *Processes*, **9** (11), 2058. https://doi.org/10.3390/pr9112058.

Liu, S., Wang, C., Wang, P., Chen, J., Wang, X. & Yuan, Q. (2018) Variation of bacterioplankton community along an urban river impacted by touristic city: with a focus on pathogen, *Ecotoxicol. Environ. Saf.*, **165**, 573–581. https://doi.org/10.1016/j.ecoenv.2018.09.006.

Lu, Q., Zhao, R., Li, Q., Ma, Y., Chen, J., Yu, Q., Zhao, D. & An, S. (2023) Elemental composition and microbial community differences between wastewater treatment plant effluent and local natural surface water: a Zhengzhou city study, *J Environ Manage*, **325** (Pt A), 116398. https://doi.org/10.1016/j.jenvman.2022.116398.

Ma, M., Wang, S., Zhu, X., Li, X., Bao, Y., Chen, X. & Wu, Z. (2023) The identification of streptococcus pasteurianus obtained from six regions in China by multiplex PCR assay and the characteristics of pathogenicity and antimicrobial resistance of this zoonotic pathogen, *Pathogens*, **12** (4), 615. https://doi.org/10.3390/pathogens12040615.

Mansfeldt, C., Deiner, K., Mächler, E., Fenner, K., Eggen, R. I. L., Stamm, C., Schönenberger, U., Walser, J.-C. & Altermatt, F. (2020) Microbial community shifts in streams receiving treated wastewater effluent, *Sci. Total Environ.*, **709**, 135727. https://doi.org/10.1016/j.scitotenv.2019.135727.

McMurdie, P. J. & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, *PLoS One*, **8** (4), e61217. https://doi.org/10.1371/journal.pone.0061217.

Numberger, D., Ganzert, L., Zoccarato, L., Muhldorfer, K., Sauer, S., Grossart, H. P. & Greenwood, A. D. (2019) Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing, *Sci. Rep.*, **9** (1), 9673. https://doi.org/10.1038/s41598-019-46015-z.

Oksanen, J.,, Simpson, G. L., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M. O., Lahti, L., McGlinn, D., Ouellette, M., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C. & Weedon, J. (2022)*Vegan: Community Ecology Package*. Available at: https://CRAN.R-project.org/package=vegan.

Pascual-Benito, M., Ballesté, E., Monleón-Getino, T., Urmeneta, J., Blanch, A. R., García-Aljaro, C. & Lucena, F. (2020) Impact of treated sewage effluent on the bacterial community composition in an intermittent Mediterranean stream, *Environ. Pollut.*, **266**, 115254. https://doi.org/10.1016/j.envpol.2020.115254.

Price, J. R., Ledford, S. H., Ryan, M. O., Toran, L. & Sales, C. M. (2018) Wastewater treatment plant effluent introduces recoverable shifts in microbial community composition in receiving streams, *Sci. Total Environ.*, **613–614**, 1104–1116. https://doi.org/10.1016/j.scitotenv.2017.09.162.

R Core Team (2022) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Richiardi, L., Pignata, C., Fea, E., Bonetta, S. & Carraro, E. (2023) Are indicator microorganisms predictive of pathogens in water?, *Water*, **15** (16), 2964. https://doi.org/10.3390/w15162964.

Rodriguez-Perez, H., Ciuffreda, L. & Flores, C. (2021) NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data, *Bioinformatics*, **37**, 1600–1601. https://doi.org/10.1093/bioinformatics/btaa900.

Ruprecht, J. E., Birrer, S. C., Dafforn, K. A., Mitrovic, S. M., Crane, S. L., Johnston, E. L., Wemheuer, F., Navarro, A., Harrison, A. J., Turner, I. L. & Glamore, W. C. (2021) Wastewater effluents cause microbial community shifts and change trophic status, *Water Res.*, **200**, 117206. https://doi.org/10.1016/j.watres.2021.117206.

Seol, D., Lim, J. S., Sung, S., Lee, Y. H., Jeong, M., Cho, S., Kwak, W. & Kim, H. (2022) Microbial identification using rRNA operon region: database and tool for metataxonomics with long-read sequence. *Microbiol. Spectrum*, **10** (2), e02017–e02021. https://doi.org/10.1128/spectrum.02017-21.

Shen, W., Le, S., Li, Y. & Hu, F. (2016) Seqkit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, *PLoS One*, **11** (10), e0163962. https://doi.org/10.1371/journal.pone.0163962.

Urban, L., Holzer, A., Baronas, J. J., Hall, M. B., Braeuninger-Weimer, P., Scherm, M. J., Kunz, D. J., Perera, S. N., Martin-Herranz, D. E., Tipper, E. T., Salter, S. J. & Stammnitz, M. R. (2021) Freshwater monitoring by nanopore sequencing, *Elife*, **10**, e61504. https://doi.org/10.7554/eLife.61504.

Viau, E. J., Goodwin, K. D., Yamahara, K. M., Layton, B. A., Sassoubre, L. M., Burns, S. L., Tong, H.-I., Wong, S. H., Lu, Y. & Boehm, A. B. (2011) Bacterial pathogens in Hawaiian coastal streams-associations with fecal indicators, land cover, and water quality, *Water Res.*, **45** (11), 3279–3290. https://doi.org/10.1016/j.watres.2011.03.033.

Wakelin, S. A., Colloff, M. J. & Kookana, R. S. (2008) Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow, *Appl. Environ. Microbiol.*, **74** (9), 2659–2668. https://doi.org/10.1128/AEM.02348-07.

Wang, J., Chen, Y., Cai, P., Gao, Q., Zhong, H., Sun, W. & Chen, Q. (2022) Impacts of municipal wastewater treatment plant discharge on microbial community structure and function of the receiving river in Northwest Tibetan Plateau, *J. Hazard. Mater.*, **423** (Pt B), 127170. https://doi.org/10.1016/j.jhazmat.2021.127170.

WHO. (2017) *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum*. Geneva: World Health Organization.

Xie, Y., Liu, X., Wei, H., Chen, X., Gong, N., Ahmad, S., Lee, T., Ismail, S. & Ni, S. Q. (2022) Insight into impact of sewage discharge on microbial dynamics and pathogenicity in river ecosystem, *Sci. Rep.*, **12** (1), 6894. https://doi.org/10.1038/s41598-022-09579-x.

Yang, Y., Hou, Y., Ma, M. & Zhan, A. (2020) Potential pathogen communities in highly polluted river ecosystems: geographical distribution and environmental influence, *Ambio*, **49** (1), 197–207. https://doi.org/10.1007/s13280-019-01184-z.

Yang, Y., Che, Y., Liu, L., Wang, C., Yin, X., Deng, Y., Yang, C. & Zhang, T. (2022) Rapid absolute quantification of pathogens and ARGs by nanopore sequencing, *Sci. Total Environ.*, **809**, 152190. https://doi.org/10.1016/j.scitotenv.2021.152190.

Zeng, J., Nakanishi, T. & Itoh, S. (2024) Two-year monitoring of microbiological water quality in small water supply systems: implications for microbial risk management, *Environmental Management*, **74**, 256–267. https://doi.org/10.1007/s00267-024-01988-0.

First received 8 August 2024; accepted in revised form 2 December 2024. Available online 11 December 2024