



What lies beneath: microbial quality of rural groundwater explored by shotgun metagenomics

Manar-Aleslam M. Mattar¹, Walaa A. Eraqi², Mohamed Bakr Zaki^{3,4}, Akram M. Elcashlan³, Khaled A. M. Abouzid⁵, Ramy K. Aziz^{2,6}, Aymen S. Yassin^{2*} and Ali H. A. Elbeheery^{1*} 

Abstract

Background Groundwater is a vital freshwater resource, especially in rural areas; however, its microbial communities remain largely unexplored beyond traditional culture-dependent strategies. This study aimed to comprehensively analyze microbial communities in groundwater, with a focus on bacterial indicators of fecal pollution, pathogenic microorganisms, the antimicrobial resistome, virulence factors (VFs), and mobile genetic elements (MGEs). Nine groundwater samples comprising three replicates from three different hand pumps, representing different locations with different levels of contamination in Toukh, Qalyubia, Egypt, were analyzed by shotgun metagenomics.

Results Taxonomic analysis identified Pseudomonadota as the dominant phylum across all pumps. Interestingly, species richness was significantly higher in pump 3 than in either pump 1 or 2. Besides, non-metric multidimensional scaling analysis showed distinct microbial community compositions across the three pumps. Bacterial indicators of fecal pollution, including *Escherichia coli* and *Salmonella enterica*, were highly abundant, alongside pathogenic microorganisms such as *Pseudomonas aeruginosa* in all pumps. VF genes, like *hsiB1/vipA*, associated with *P. aeruginosa* pathogenicity, were highly abundant in pumps 1 and 2. Antimicrobial resistance genes (ARGs), encoding resistance to aminoglycosides, beta-lactams, tetracyclines, and sulfonamides, were detected in all pumps. Among MGEs, insertion sequences were generally more abundant than plasmids and integrons, and several ARGs were genomically colocalized with MGEs, which suggests a strong potential for ARG mobilization by horizontal gene transfer.

Conclusions These findings highlight the critical need for monitoring groundwater quality to mitigate antimicrobial resistance (AMR) and other health threats.

Keywords Groundwater, Shotgun metagenomics, Antimicrobial resistance, Virulence factors, Mobile genetic elements

*Correspondence:

Aymen S. Yassin
aymen.yassin@pharma.cu.edu.eg
Ali H. A. Elbeheery
ali.elbeheery@fop.usc.edu.eg

¹Department of Microbiology and Immunology, Faculty of Pharmacy, University of Sadat City, Sadat City 32897, Egypt

²Department of Microbiology and Immunology, Faculty of Pharmacy, Cairo University, Cairo, Egypt

³Department of Biochemistry, Faculty of Pharmacy, University of Sadat City, Sadat City, Egypt

⁴Department of Biochemistry, Faculty of Pharmacy, Menoufia National University, km Cairo-Alexandria Agricultural Road, Menoufia, Egypt

⁵Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Ain Shams University, Cairo, Egypt

⁶MARC Biosciences Research Laboratories, 6 October City, Egypt



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Groundwater is an essential resource for sustaining life on Earth, providing drinking water and driving economic activities. In addition to being an ancient source of potable water, its critical role in agriculture, livestock, and food processing has significantly boosted food production since the 1970s, particularly in semi-arid and arid regions with scarce rain and surface waters. Currently, agriculture consumes 70% of global groundwater, serving 38% of irrigated lands. By 2050, groundwater development could enhance agricultural productivity and economic growth to meet the anticipated 50% increase in food demand (UNESCO 2022).

Despite this great significance, the quality of groundwater is increasingly threatened by human activities, including industrial and agricultural activities, which discharge various pollutants into this vital resource. These pollutants include antibiotics and antimicrobial-resistant bacteria (ARB) (Liu et al. 2023). The improper use of wastewater in irrigation, such as the direct use of untreated or partially treated wastewater, can cause soil contamination by pathogens, antibiotics, and toxic chemicals, such as heavy metals, which may contaminate groundwater (Li et al. 2016; Farid et al. 2019). Additionally, 49% of antibiotics administered to humans or animals are excreted in urine, representing a major threat of increasing antimicrobial resistance (AMR) if wastewater containing this urine reaches groundwater (Polianciuc et al. 2020). This situation is a great depiction of the one health concept, which links human, animal, and environmental health in an inseparable continuum (Hernando-Amado et al. 2019).

Globally, the contamination of groundwater by antibiotics and the associated rise in AMR have triggered significant concern (Szekeres et al. 2018; Liu et al. 2023). AMR is a complex problem with a profound influence on human health. According to the World Health Organization (WHO), AMR is one of the major causes of prolonged illnesses, resulting in longer stays at hospitals and increased mortality rates. In addition, AMR increases the frequency of sepsis and sepsis-associated mortality, particularly in immunocompromised patients (Nanayakkara et al. 2021).

Given the various sources of contamination affecting groundwater, such contamination is believed to play a pivotal role in exacerbating AMR and facilitating its spread. However, most research on antimicrobial resistance genes (ARGs) focused on surface water, e.g., rivers, lakes (Nomoto et al. 2024; Díaz-Torres et al. 2024; Huang et al. 2024), with only limited studies on groundwater (Jia et al. 2019; Yan et al. 2021).

In Egypt, previous studies on groundwater quality have solely relied on traditional microbiological, culture-dependent techniques, which reveal less than 1% of

the true microbial diversity, leaving most of these communities virtually untapped (Vaz-Moreira et al. 2011). Hassanein and colleagues (2012) showed the microbiological contamination of shallow groundwater in Dakahlia, Qena, and Sohag (Hassanein et al. 2012). In Qaluybiya, Ewida and co-workers (2021) reported the pollution of shallow groundwater with iron and manganese in concentrations above the allowed limits. Microorganisms in heavy-metal-polluted water acquire mechanisms to adapt to the presence of heavy metals (Ewida et al. 2021). Moreover, heavy metals serve as co-selecting agents that can enhance resistance to both antimicrobials and heavy metals through co-resistance and cross-resistance (Nguyen et al. 2019).

To avoid the limitations of previous studies and obtain a less biased, more comprehensive profile of the actual microbial diversity in groundwater, this study used culture-independent shotgun metagenomics to investigate microbial assemblages in rural groundwater. The objectives include identifying microbial communities, with a focus on bacterial indicators of fecal pollution and potential pathogens, in addition to the profiling of ARGs, virulence factors (VFs) and mobile genetic elements (MGEs). By providing a comprehensive overview of microbial and genetic contamination, this work enhances our understanding of groundwater pollution and emphasizes the importance of a radical one health approach to address AMR and its potential impacts on public health and ecological sustainability.

Materials and methods

Sample collection

Nine freshwater samples, consisting of three replicates from each of three different hand pumps located in Toukh, Qalyubia, Egypt, were collected between March and July 2023 (Fig. 1). Because weather and rain strongly affect groundwater composition, the samples were collected around 6–7 AM, during spring and early summer months, when that area of Egypt typically experiences dry weather with minimal precipitation and stable temperatures. Pump 1 (30° 20' 27.14" N, 31° 11' 37.69"E) drawing groundwater at a depth of 35 meters is surrounded by agricultural fields and a cemetery, with the nearest residential area located approximately 500 meters away. Pump 2 (30° 20' 42.69"N, 31° 11' 46.35"E) drawing groundwater at a depth of 30 meters, is adjacent to residential buildings lacking sewage infrastructure, with inhabitants relying on swage tanks for waste disposal. Additionally, there is an agricultural field in its vicinity. Pump 3 (30° 20' 56.4036"N, 31° 12' 15.3468"E) draws groundwater at a depth of 28 meters and is situated within a residential area, where buildings are interconnected with a regular sewage system.

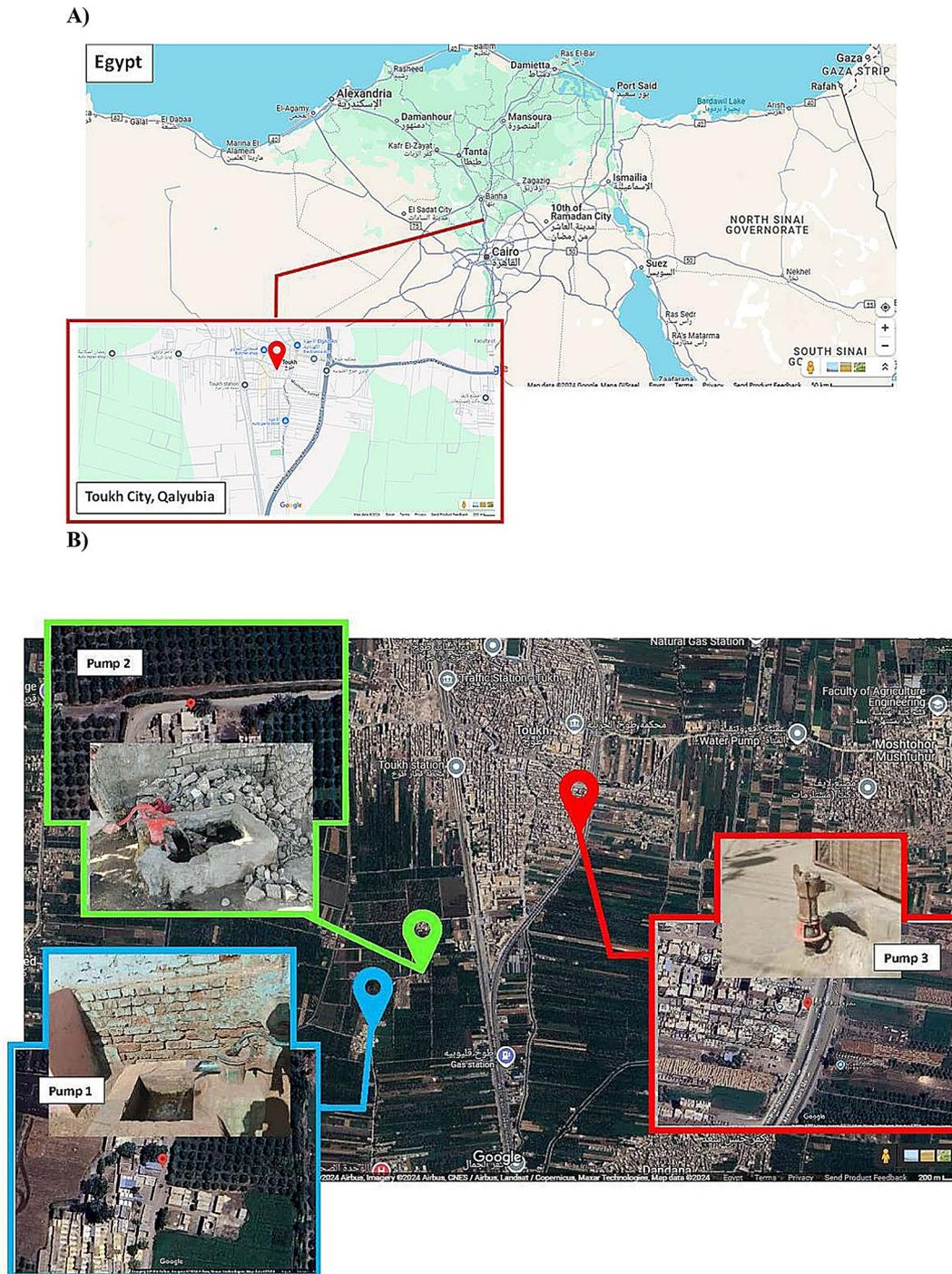


Fig. 1 Location of the hand pumps. (A) Geographic location of Toukh City, Qalyubia, Egypt (B) Satellite images showing the precise locations of the hand pumps where groundwater samples were collected, together with field photographs of the hand pumps

Each sample replicate consisted of 10 L of groundwater, collected in sterile glass bottles, which were only opened within the water flow to minimize the risk of airborne contamination. Replicates were collected on separate days. Of note, pumps were flushed before sample collection, which ensures the removal of stagnant water that may not accurately represent groundwater quality. The

samples were then transported to the laboratory for processing within four hours of sampling. Upon arrival at the laboratory, samples were filtered through mixed cellulose ester (MCE) membrane filters with a pore size of 0.2 μm (Merck-Millipore, Burlington, MA, USA) using Millipore Masterflex I/P peristaltic pump (Cole-Parmer Co.,

Vernon Hills, IL, USA). Filter membranes were stored at -20°C until DNA extraction.

DNA isolation and shotgun sequencing

For each replicate, the membrane filter was cut into small pieces and placed into a 50 mL sterile tube. Genomic DNA extraction from all samples was conducted using DNeasy PowerSoil Kit (Qiagen, Valencia, USA), following the manufacturer's protocol with a slight modification by adding 6 µL of lysozyme (Thermo Fisher Scientific, Inc., Waltham, MA, USA) and 60 µL of proteinase K (Qiagen, Hilden, Germany) to the PowerBead tube buffer in the initial step of the extraction process. Extracted metagenomic DNA was sent for shotgun sequencing using Illumina NovaSeq 6000 paired-end sequencing with a pair read length of 150 bp (Novogene, Beijing, China).

The raw reads for this study are available through the NCBI Sequence Read Archive (SRA) with BioProject accession PRJNA1196429.

Sequence quality control, taxonomic assignment, identification of fecal indicators and potential pathogenic microorganisms

FASTQ files from all samples were processed for quality control using fastp v 0.23.4 (Chen et al. 2018) with the default parameters to filter out low-quality reads and trim low-quality bases.

High quality reads were used for taxonomic classification using Kraken2 v.2.1.3 (Lu et al. 2022) with the default parameters, which mapped reads against the Kraken database PlusPF. This database includes RefSeq entries for archaea, bacteria, viruses, plasmids, human, UniVec_Core, protozoa, and fungi (<https://benlangmead.github.io/aws-indexes/k2>).

Kraken2 reports provided taxonomic information for the desired taxonomic ranks, then relative abundance estimation was conducted using Bracken v2.9 (Lu et al. 2017) with the default parameters. Visualization of the results was performed using the interactive browser application Pavian (Breitwieser and Salzberg 2020).

Alpha diversity metrics, including the Shannon index and richness, were calculated at the species level by the R package *phyloseq* v1.48.0 (McMurdie and Holmes 2012), while evenness was calculated with the *vegan* package v2.7.0 (Oksanen et al. 2024). To test the significance of the alpha diversity metrics, an ANOVA test was performed in R, followed by the Dunn-Sidak correction for adjusted *p*-value calculation (R Core Team 2024). Beta diversity was calculated using the Bray-Curtis dissimilarity metric with the *phyloseq* package v1.48.0 (McMurdie and Holmes 2012). Visualizations were generated using the *ggplot2* package v3.5.1 (Valero-Mora 2010).

The names of common bacterial indicators of fecal pollution, such as *Escherichia coli*, *Enterococcus* spp.,

Bacteroidetes spp., and *Clostridium perfringens*, were obtained from the literature (Byappanahalli et al. 2012; Dubin and Pamer 2014; Richiardi et al. 2023). Then, the relative abundance of these bacterial species was extracted from Bracken reports.

For the identification of potential pathogenic microorganisms, Kraken reports were converted into the MetaPhlAn output format using KrakenTools (Lu et al. 2022). The MetaPhlAn reports were submitted to MicroPhenoDB web application (Yao et al. 2020) to identify the potential, pathogenic microorganisms in each sample. MicroPhenoDB (Yao et al. 2020) includes the databases of disease phenotypes and pathogenic microorganisms. The identified microorganisms scores above 0.3 were found to have higher confidence of microbe-disease association, therefore a filtration step was applied, retaining only those microorganisms with scores above this threshold.

De novo assembly and assessment of VFs and ARGs

Quality-filtered sequence reads were assembled into contigs by the MEGAHIT software v1.2.9 (Li et al. 2015), with the default parameters. Redundant contigs were removed by sequence clustering at 95% identity with the cd-hit-est tool v.4.8.1 (Fu et al. 2012). The following parameters were used for clustering: -M 0 (an option to use all available memory), -T 0 (an option to use all available CPU cores), -c 0.95 (sequences must be at least 95% identical to be clustered together), and -n 10 (word length or k-mer size of 10). Open reading frames (ORFs) were predicted from assembled contigs by prodigal v2.6.3 with the "meta" mode, specific for metagenomic sequences (Hyatt et al. 2010).

ABRicate v.1.0.0 (Seemann 2020) was used for the detection of VFs and ARGs in ORFs by annotating them against the Virulence Factor Database (VFD) (Chen et al. 2016) and NCBI Bacterial Antimicrobial Resistance Reference Gene Database (Feldgarden et al. 2019) with all other parameters set to default.

Normalized relative abundances of annotated VFs and ARGs in reads per kilobase per million reads (RPKM), were estimated using the BBMap tool (Bushnell 2015) by aligning the metagenomic reads against the nucleotide ORFs generated by Prodigal.

Identification of MGEs

To annotate contigs for plasmid detection, we used the plasmidVerify tool (Antipov et al. 2019), which uses an HMMER-based hmmsearch (Finn et al. 2011) against the Pfam A database (Mistry et al. 2021) to identify plasmids through a naïve Bayesian classifier. For the detection of integrons, we used Integron-Finder v2.0.2 (Néron et al. 2022). Finally, insertion sequences (IS) were identified using isescan v1.7.2.3 (Xie and Tang 2017). Statistical

significance of the abundance of plasmids, integrons, and IS across the pumps was tested using an ANOVA test in R, followed by the Dunn-Sidak correction for adjusted *p*-value calculations (R Core Team 2024).

Results

Sequence data and statistics

Shotgun metagenomic sequencing of the nine ground-water samples generated 530,443,338 reads. On average, 43.92 to 71.15 million reads were obtained per sample (Supplementary Table 1). An average of 99.30% (range: 99.10–99.40%) of reads per sample passed QC.

De novo assembly was performed, resulting in contigs ranging in number from 812,100 to 1,377,927 per pump, with an average N50 of 800 bp per pump (precisely 894 for pump1, 876 for pump2, and 630 for pump3, Supplementary Table 2).

Taxonomic profiles and relative abundance of taxa across different pumps

Across all samples, sequences were assigned to 84 phyla, 698 families, and 2,580 genera. On average, 56.90% of the reads in each sample were unclassified (Supplementary Table 3).

Across all three pumps, Pseudomonadota was observed as the dominant phylum (Supplementary Fig. 1), and had its highest relative abundance in pump 2 (39.66% of the total reads), compared to 27.88% in pump 1 and 12.90% in pump 3.

Following Pseudomonadota, Actinomycetota was the second most abundant phylum, with its maximum relative abundance in pump 2 (13.82%). In contrast, the abundance of Actinomycetota in the remaining pumps remained below 10%. Other than these two phyla, pump 1 showed a relatively higher abundance of Bacillota, accounting for 2.38% of the total number of reads.

The dominant classes varied among the three pumps. In pump 1, Betaproteobacteria was the most abundant class (13.30%), followed by Gammaproteobacteria (7.29%). In contrast, Gammaproteobacteria dominated the microbial community in pump 2 with an abundance of 20.94%, followed by Actinomycetes (13.60%). In pump 3, Actinomycetes was the most abundant class (6.71%), followed by Betaproteobacteria (4.89%).

At the family level, pump 1 had a remarkable representation of the Comamonadaceae family (a diverse family of the Betaproteobacteria, belonging to the order Burkholderiales), constituting approximately 6.01%. In contrast, the most abundant families in pump 2 were Xanthomonadaceae and Moraxellaceae, whose relative abundances were 7.75% and 7.35%, respectively. Pump 3 had a distinct composition, with Mycobacteriaceae as the most abundant family, accounting for 2.14%.

At the genus level, *Acidovorax* was the most abundant in pump 1 (4.07%). In pump 2, *Acinetobacter* and *Stenotrophomonas* had equal relative abundance of 7.23% each. In pump 3, *Mycobacterium* (1.46%) was the most abundant genus.

In terms of diversity, which was calculated at the species level, pump 3 showed the highest alpha diversity, but no significant difference in the Shannon index was observed between pumps (Supplementary Fig. 2). Similarly, evenness was not significantly different. However, pump 3 had the highest richness, with significant pairwise differences between pump 3 and pump 1 ($p=0.0047$), and between pump 3 and pump 2 ($p=0.0021$). This pattern suggests that pump 3, the least contaminated site, supports a more diverse and richer microbial community.

Beta diversity analysis was estimated as the Bray-Curtis measure of dissimilarity. A non-metric multidimensional scaling (NMDS) plot, representing the replicates from the three pumps, confirms that the inter-pump variability across microbial communities of the three pumps was higher than the intra-pump (between replicates) variability (Supplementary Fig. 3).

Fecal indicators and common human pathogens

The microbial quality of groundwater is commonly assessed by the detection of fecal indicator bacteria, which are normally present in high numbers in the feces of humans and animals. Therefore, their presence in groundwater strongly indicates fecal (sewage) pollution and refers to a higher likelihood that fecal-related pathogens are present (Kaushik et al. 2014). The metagenomic analysis comprehensively detected a range of fecal indicators in the groundwater samples (Fig. 2). The highest abundance of fecal indicators, with diverse species, was in pump 1. *Bacteroides* spp. were identified as the most abundant fecal indicator in pump 1. Additionally, *E. coli* was among the most abundant species across all pumps with relative abundances (in RPKM) as follows: pump 1 (0.0065), pump 2 (0.0068), and pump 3 (0.0045), followed by *Salmonella enterica*, whose abundance in RPKM was 0.0034 (pump 1), 0.0033, (pump 2), and 0.0034 (pump 3).

Regarding microbial taxa with known pathogenic members, pump 1 had the highest diversity and abundance of potentially pathogenic species, followed by pump 2 then pump 3. *Pseudomonas aeruginosa*, with relative abundances—in RPKM—as follows: pump 1 (0.0317), pump 2 (0.0458), and pump 3 (0.0214), was the dominant species of potential human pathogen across all pumps. It was followed by *E. coli*, then *Salmonella*. *Mycobacterium tuberculosis*, *Bacteroides fragilis*, and *Clostridium perfringens* had their highest abundance in pump 1 (Fig. 3).

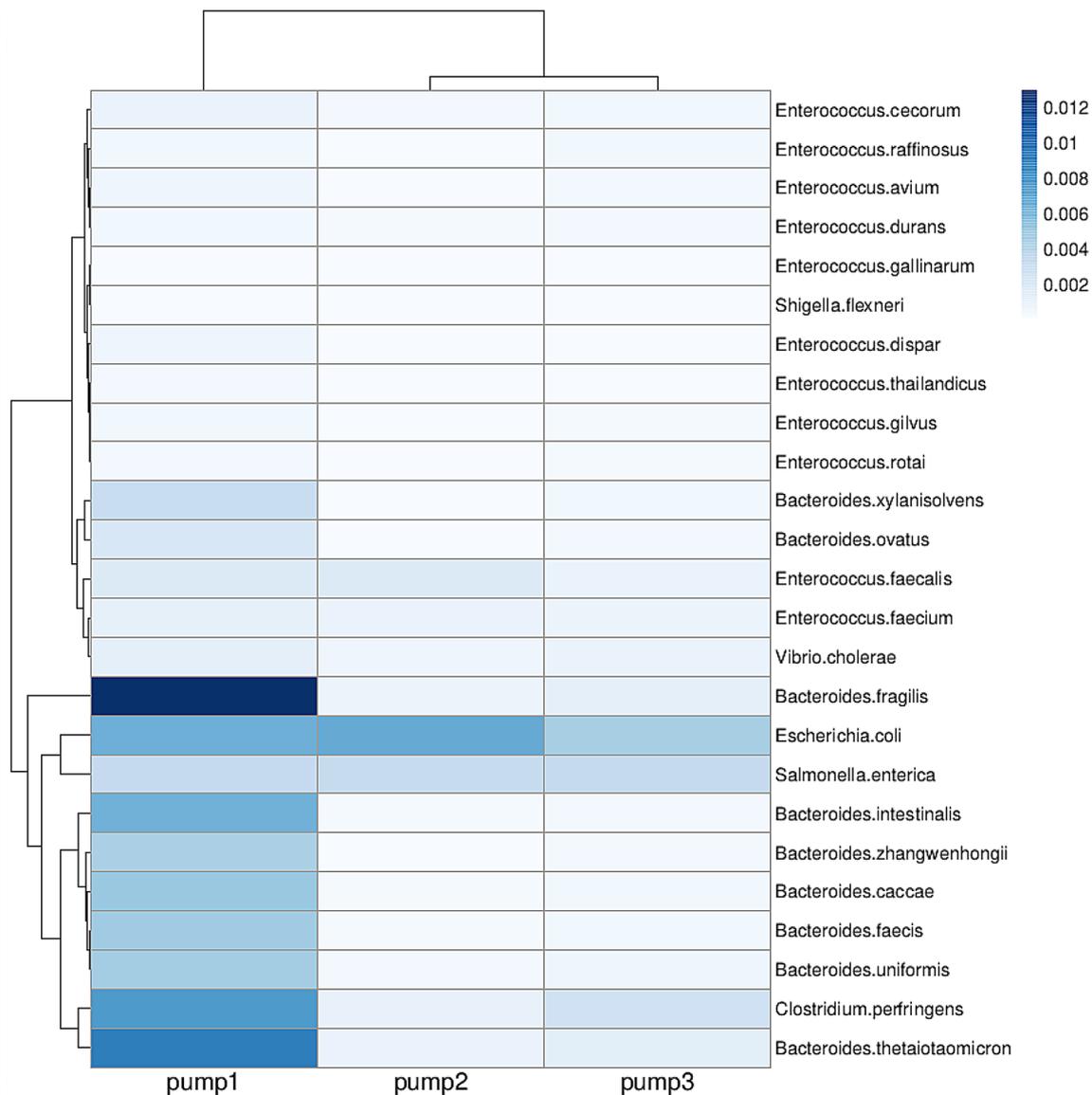


Fig. 2 Heatmap illustrating the relative abundance of fecal indicators in groundwater samples from pump 1, pump 2, and pump 3. The two dendrograms (top and left) represent hierarchical clustering of samples and fecal indicator species, respectively

Antimicrobial resistance genes

Each pump's resistome was characterized based on the antibiotic classes to which resistance was conferred. In total, 58 unique ARGs were detected, encoding resistance to 14 different antibiotic classes. Pump 1 had 27 ARGs associated with 9 drug classes, with resistance to tetracyclines being the most frequently detected (relative abundance of 21.19 RPKM), primarily represented by *tet(44)* and *tet(C)*. Resistance to aminoglycosides was the second most frequently represented by ARGs (18.73 RPKM), mostly represented by the *aph(3')-IIIa* gene. Other frequently identified resistance genes were those encoding resistance to sulfonamides (relative ARG abundance of 8.9 RPKM) and carbapenems (relative ARG abundance of 8.8 RPKM), with the main ARGs being

sul2 and *bla_{OXA-651}*, respectively. Moreover, pump 2 had 32 ARGs linked to resistance to 11 drug classes, with those conferring resistance to aminoglycosides being the most frequently detected ARGs (relative abundance of 96.75 RPKM), represented by *ant(3'')-IIa*, *aph(3')-IIc*, and *aph(6)-Id*. ARGs encoding resistance to carbapenems and other beta-lactams were the next most abundant ones (25.04 RPKM and 18.75 RPKM, respectively), represented by ARG *bla_{L1}* and *bla_{L2}*, respectively. The microbiomes of pump 3, however, only had 11 ARGs, associated with resistance to seven drug classes, predominantly aminoglycosides (10.9 RPKM), represented by ARG *aph(3')-IIIa*, *aph(6)-Id*; and beta-lactams (6.3 RPKM), e.g., ARG *bla_{RSD2-2}*. Overall, pumps 2 and 1 had higher antimicrobial resistance potential than pump 3, as reflected by

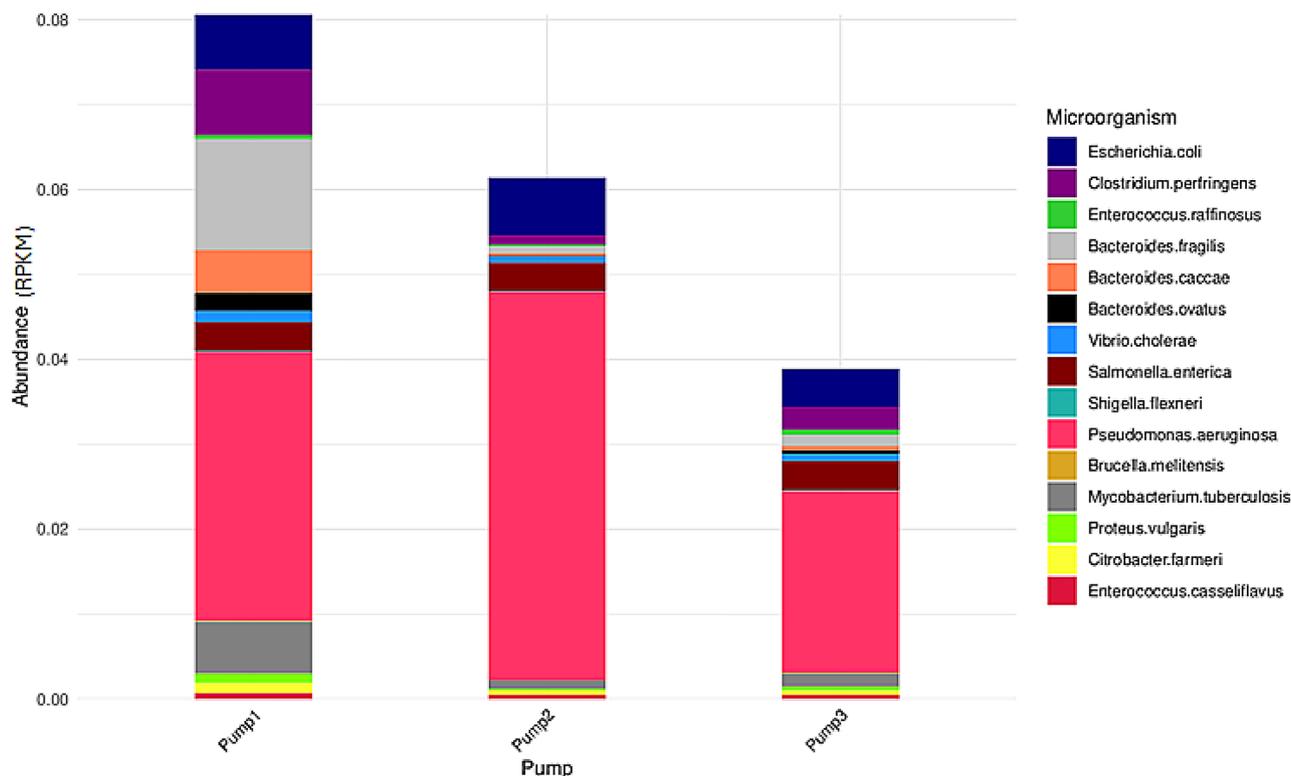


Fig. 3 Stacked bar plot showing the relative abundance of pathogenic microorganisms in groundwater samples from pump 1, pump 2, and pump 3

the highest number of distinct ARGs (32 and 27, respectively) spanning the widest range of antibiotic classes (11 and 9, respectively) (Fig. 4A, B).

Virulence factors

Forty-seven VF genes were identified across the three pumps, belonging to six VF superfamilies: adherence, invasion, motility, immune modulation, nutritional/metabolic factors, and regulation. The distribution of VFs largely varied among the three pumps. Pump 2 had the most diverse range of VF genes, with the predominant ones being (i) *hsiB1/vipA* encoding type VI secretion system tubule-forming protein VipA (6.72 RPKM), which enhances the virulence to *Pseudomonas aeruginosa*, (ii) isocitrate lyase (*icl*) (6.1 RPKM), a gene encoding probable GTP pyrophosphokinase (*relA*) (5.4 RPKM), and the transcriptional positive regulator-encoding gene *phoP* (4.9 RPKM), which are associated with the virulence of *Mycobacterium tuberculosis*. Pump 3, the second-most diverse in terms of VFs, had the highest relative abundance of *esxN*, encoding ESX-5 type VII secretion system (16.7 RPKM) and *esxM*, encoding ESX-5 type VII secretion system EsxB (7.5 RPKM), both of which are related to the pathogenicity of *Mycobacterium tuberculosis*. In contrast, pump 1 had the lowest diversity and abundance of VF genes, with only two genes identified in detectable amounts (Fig. 5): *hsiB1/vipA*, encoding a

type VI secretion system tubule-forming protein VipA (7.27 RPKM) and the acyl carrier protein-encoding gene, *acpXL* (7.4 RPKM).

This distribution of VF genes indicates that even minimally contaminated groundwater may serve as a reservoir for virulence traits.

Mobile genetic elements

The abundance of three classes of MGEs (plasmids, IS, and integrons) varied across the three pumps (Fig. 6). Overall, insertion sequences were more abundant than plasmids and integrons in all pumps. Pump 3 had the highest abundance of MGEs across all pumps.

The distribution of plasmid abundance indicated that pump 1 metagenomic replicates had the lowest levels, with an average abundance of 6,422.45 RPKM (range: 6,014.31- 7,061.44), followed by pump 2 replicates with the (mean = 7,027.64 RPKM; range: 6,476–7,886.82). The highest plasmid abundance was observed in pump 3 replicates (mean = 8,186.16 RPKM; range: 8,114.039–8,285.406). The plasmid abundance in pump 3 was significantly different from that in pump 1 (p -value = 0.022); however, no significant difference was observed between pump 1 and pump 2.

As for insertion sequences, they had comparable abundances in pump 1 and pump 2, with averages of 8,882.28, and 9,255.43 RPKM, respectively, yet lower than their

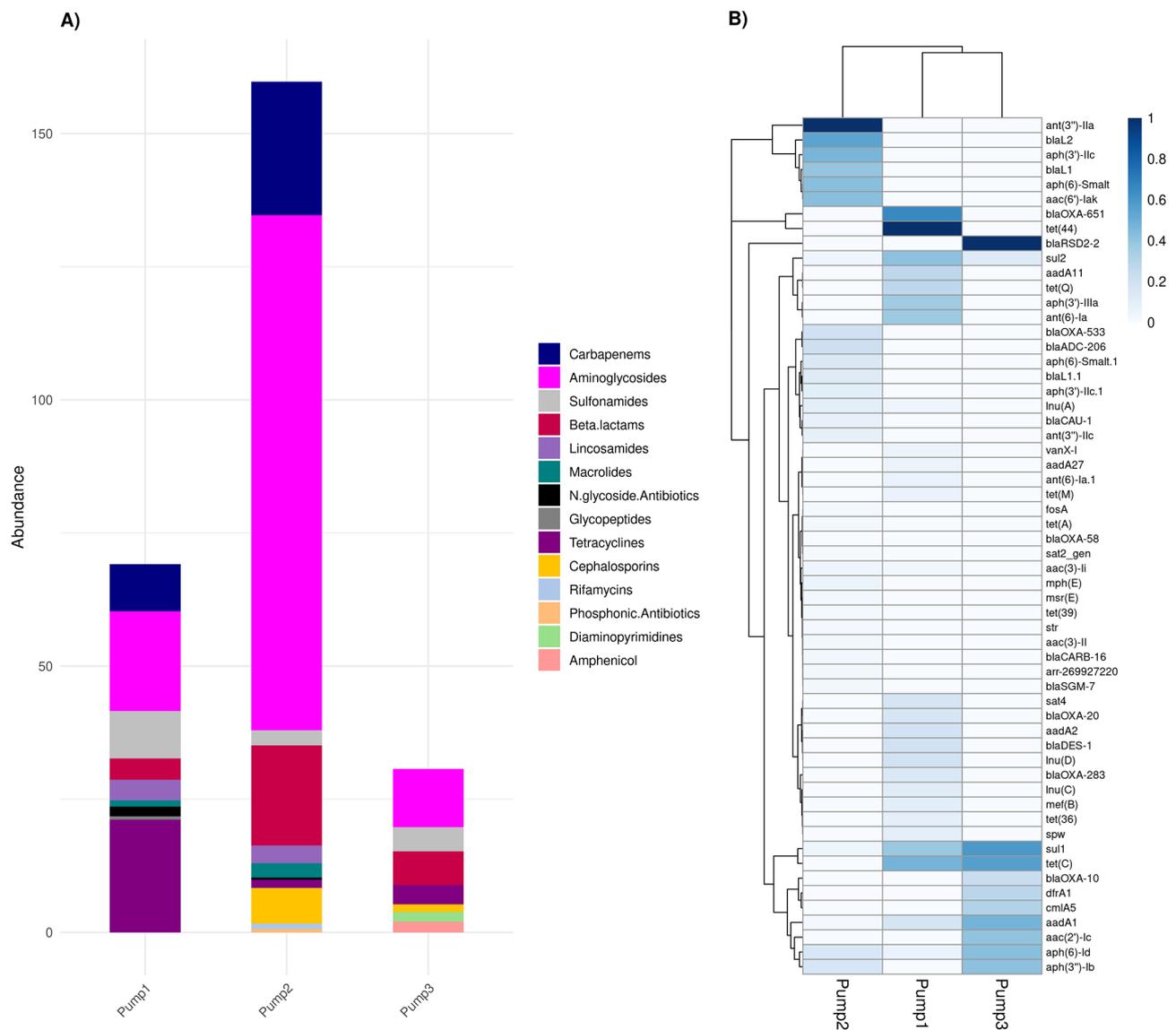


Fig. 4 Antimicrobial resistance genes and classes **(A)** stacked bar plot showing the average abundance of different antimicrobial resistance drug classes across the three groundwater replicates from pump 1, pump 2, and pump 3. **(B)** Heatmap illustrating the scaled average abundance of antimicrobial resistance genes across the three groundwater replicates from pump 1, pump 2, and pump 3

abundance in pump 3 samples (mean = 10,867.55 RPKM). It is worth noting that one of pump 2 replicate samples had an outlier (Replicate 2: relative abundance = 14,476.08 RPKM), the highest abundance among all replicates, exceeding that of the other two pumps. Because of this high variability, no statistically significant differences were detected between the pumps.

Integron abundance was generally lower than that of plasmids and IS. Pump 1 and pump 2 showed similar abundance with average 1,009.45 and 1,109.36 RPKM respectively. Pump 3 demonstrated the highest integron abundance across all pumps, with an average of 7,780.76 RPKM. This abundance was significantly different from pump 1 ($p = 0.000097$) and pump 2 ($p = 0.0001$) (Fig. 6).

The detected IS were classified into various IS families. Overall, IS3, IS5 and IS21 were the most abundant across all pumps. The mean relative abundances of IS3 in RPKM were 1,903.5 (pump 1), 2507.4 (pump 2), 1249.7 (pump 3); IS21 in RPKM (pump 1 mean = 1091.8, pump 2 mean = 997.14, pump 3 mean = 1101.56); IS5 in RPKM (pump 1 mean = 981.89, pump 2 mean = 1442.01, pump 3 mean = 760.24). Pump 3 had the highest diversity and abundance of IS families (Supplementary Fig. 4); notably, families IS200/IS605 and IS110 were highly abundant only in pump 3.

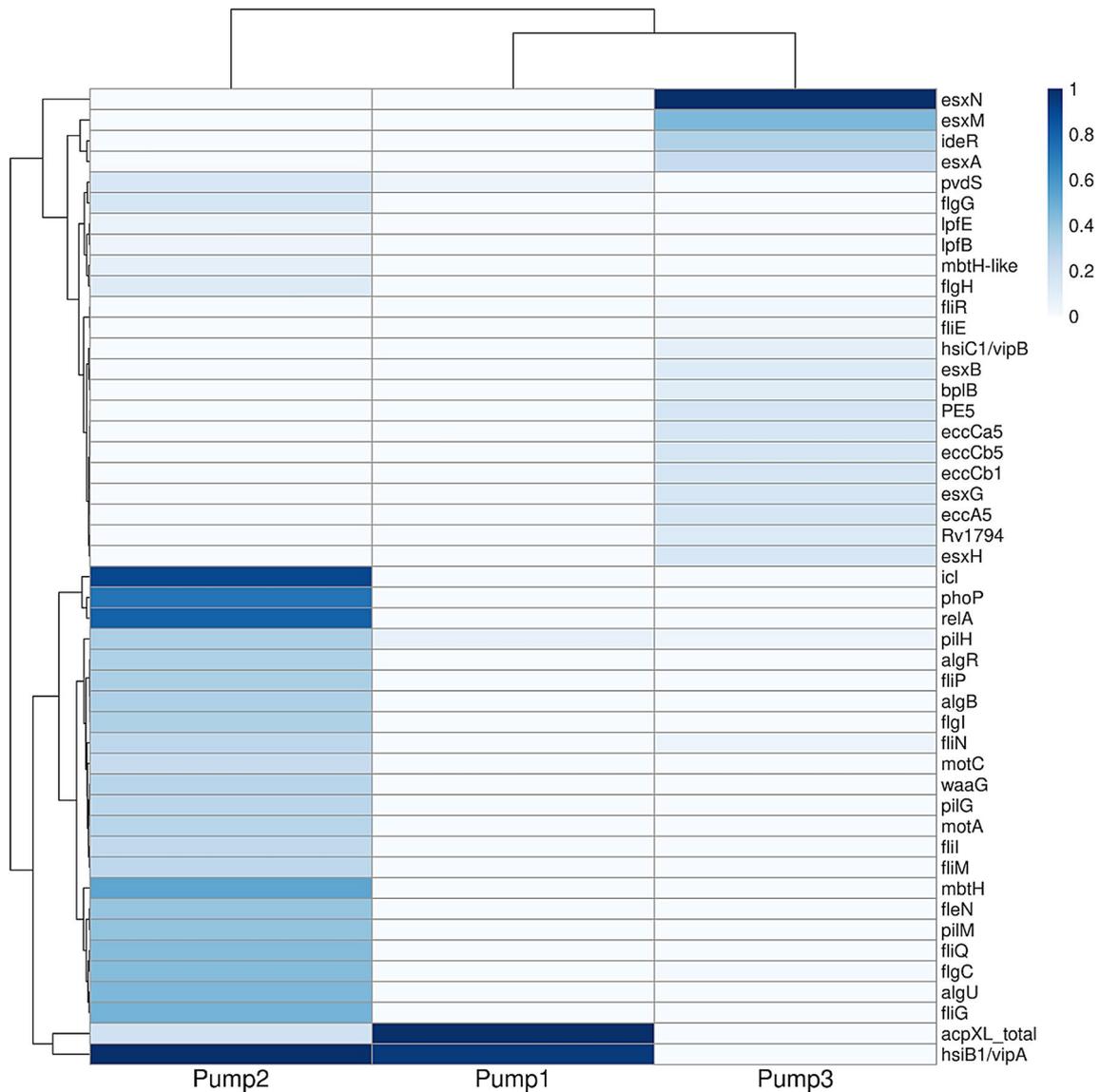


Fig. 5 Heatmap illustrating the scaled average abundance of virulence factors across the three groundwater replicates from pump 1, pump 2, and pump 3

Co-occurrence of ARGs with MGEs

Inspection of annotated contigs revealed that several ARGs were colocalized with plasmids and insertion sequences across contigs recovered from the three groundwater pumps (Table 1). In pump 1, multiple ARGs—including *aadA11*, *tet(C)*, *lnu(A)*, and *aph(6)-Id*—were found on contigs also annotated as plasmid sequences. Notably, two of these ARGs, namely *aadA11* and *tet(C)*, were also associated with insertion sequences from the IS91 and IS3 families, respectively. Other ARGs, such as *mef(B)*, *sul2*, and *bla_{OXA-20}*, were located on contigs not annotated as plasmids but carrying insertion sequences (IS21, IS91, and IS110).

In the sample from pump 2, ARGs, including *aac(3)-II*, *aph(6)-Id*, *aph(3'')-Ib*, and *bla_{OXA-58}*, were predominantly

located on plasmid-associated contigs. However, no insertion sequences were detected alongside these genes, suggesting plasmid-mediated dissemination without known transposable elements.

Like with samples from pump 1, co-occurrence of an ARG, plasmid, and insertion sequence was observed among sequences from pump 3. Specifically, contig k141_749373 harbored *bla_{RSD2-2}* along with a plasmid and IS256. In contrast, *aph(6)-Id* in contig k141_267267 was associated with a plasmid only.

These patterns suggest potential mobilization of ARGs through these potential hotspots for horizontal gene transfer.

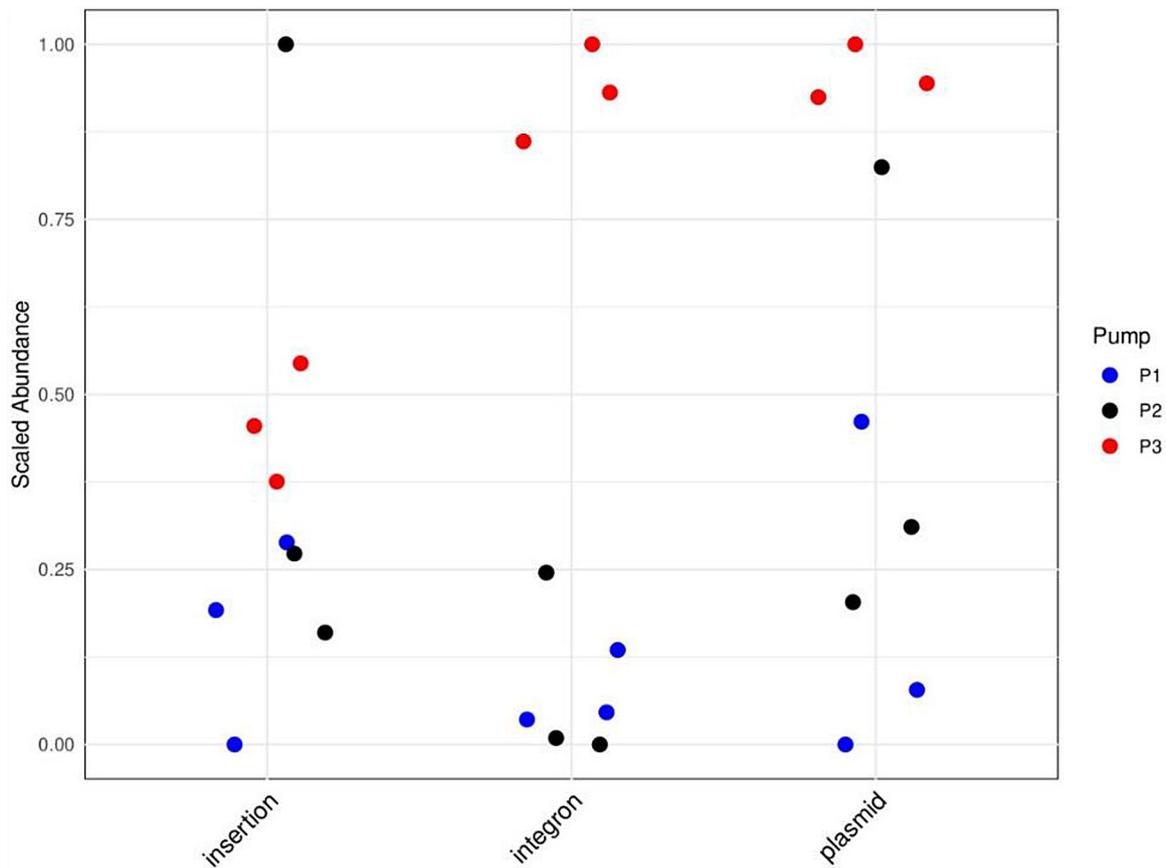


Fig. 6 A dot plot showing the scaled relative abundance of insertions, integrons and plasmids in groundwater samples from pump 1, pump 2, pump 3. Significant differences in the relative abundance of integrons between pump 3 and pump 1 ($p=0.000097$, ANOVA followed by the Dunn-Sidak *post hoc* test), pump 3 and pump 2 ($p=0.0001$, ANOVA followed by the Dunn-Sidak *post hoc* test) and in plasmid abundance between pump 3 and pump 1 ($p=0.022$, ANOVA followed by the Dunn-Sidak *post hoc* test) were observed

Table 1 Antimicrobial resistance genes, plasmids and insertion sequences co-located on the same contigs

Pump	Contig name	ARG (Antibiotic class)	Plasmid	Insertion sequence
Pump 1	k141_509352	<i>aadA11</i> (Aminoglycoside)	✓	IS91
	k141_664048	<i>tet(C)</i> (Tetracycline)	✓	IS3
	k141_453077	<i>lnu(A)</i> (Lincosamide)	✓	
	k141_601605	<i>aph(6)-Ia</i> (Aminoglycoside)	✓	
	k141_904701	<i>lnu(D)</i> (Lincosamide)	✓	
	k141_95316	<i>lnu(C)</i> (Lincosamide)	✓	
	k141_268955	<i>mef(B)</i> (Macrolide)		IS21
	k141_561264	<i>sul2</i> (Sulfonamide)		IS91
	k141_854039	<i>bla_{OXA-20}</i> (Beta-lactam)		IS110
	Pump 2	k141_146179	<i>lnu(A)</i> (Lincosamide)	✓
k141_285742		<i>aac(3)-II</i> (Aminoglycoside)	✓	
k141_395895		<i>aph(6)-Ia</i> ,	✓	
		<i>aph(3'')-Ib</i> (Aminoglycoside)		
k141_496773		<i>bla_{OXA-58}</i> (Beta-lactam)	✓	
k141_651786		<i>mph(E)</i> (Macrolide)	✓	
k141_811248		<i>str</i> (Aminoglycoside)	✓	
Pump 3	k141_749373	<i>bla_{RSD2-2}</i> (Beta-lactam)	✓	IS256
	k141_267267	<i>aph(6)-Ia</i> (Aminoglycoside)	✓	

Discussion

In this study, we comprehensively investigated the microbiological quality of groundwater from three hand pumps in a rural area through shotgun metagenomic analysis. We examined the taxonomic profiles of microbial communities to identify potential pathogens and bacterial indicators of fecal pollution. Furthermore, we studied the antimicrobial resistome, VFs, and MGEs present in these samples.

The composition of microbial communities in this study revealed that Pseudomonadota was the dominant phylum across all pumps. However, significant variability between pumps was observed at lower taxonomic ranks (such as class, order, and family). This observation aligns with the findings from the meta-analyses of global groundwater (Liu et al. 2023). These results suggest that, while groundwater microbial communities share a similar pattern at higher taxonomic levels, they have site-specific differences at finer taxonomic scales. These differences are more definitive and probably reflect localized environmental influences, such as differences in sewage disposal and proximity to agricultural fields—factors that can influence the extent of fecal contamination, which is an important determinant affecting microbial diversity (Paruch et al. 2019).

Further bioinformatics analysis showed considerable intra-pump diversity, represented by a high number of species in all pumps: pump 1: 10,362, pump 2: 10,210, pump 3: 10,357. Microbial diversity in groundwater can vary significantly based on factors such as geology, hydrology, and human activities. In some cases, groundwater has been found to harbor high microbial diversity. For example, a study comparing microbial communities in surface and groundwater (Ji et al. 2022) reported greater microbial richness and diversity in groundwater when compared with surface water. However, other research indicates that microbial diversity in groundwater is generally low, especially in pristine groundwater, reviewed in (Griebler and Lueders 2009). The observed lower microbial richness and diversity in pumps 1 and 2, which are likely the most contaminated sites, align with findings that ecosystems with higher levels of fecal pollution exhibit lower microbial diversity (Paruch et al. 2019). Paruch and colleagues demonstrated that rural waters with minimal fecal contamination had the highest microbial diversity, whereas urban waters with significant fecal pollution showed substantially reduced diversity (Paruch et al. 2019). This suggests that lower fecal contamination may help preserve the natural diversity of microbial communities in groundwater ecosystems. Such a concept also applies to several human or other host-associated microbiomes: diversity is often inversely correlated with disease (Kim et al. 2023), pollution (Sazykina et al. 2022), or external infection (Ruiz-Tagle et al. 2023).

Additionally, groundwater samples analyzed in this study showed distinct inter-pump diversity consistent with the observations of the global groundwater meta-analysis (Liu et al. 2023), indicating a site-specific composition of each microbiome. Similarly, Yan and colleagues (2020) demonstrated significant bacterial community diversity between microbial communities from near-surface groundwater samples, even when they grouped the wells by hydrochemical traits, emphasizing that microbial communities provide a higher resolution for distinguishing wells than hydrochemical characteristics alone (Yan et al. 2020).

The groundwater samples analyzed in this study revealed the presence of multiple bacterial species indicative of fecal contamination, including *Escherichia coli*, *Salmonella enterica*, *Clostridium perfringens*, *Vibrio cholerae*, and various species of *Bacteroides* (e.g., *Bacteroides fragilis* and *Bacteroides faecis*). Additionally, several *Enterococcus* species, such as *Enterococcus faecalis* and *Enterococcus faecium*, were detected across all pumps. These findings align with a study conducted on 182 wells in Italy, where 77.5% of samples exhibited fecal contamination, with *E. coli*, *Salmonella*, total coliforms, and *Enterococcus* species identified as key indicators (De Giglio et al. 2017). Similarly, a study of 50 tubewells in rural Bangladesh detected fecal bacteria, including *E. coli*, *Shigella*, and *Vibrio spp.*, indicating widespread contamination (Ferguson et al. 2012).

The findings are consistent with research conducted in Morocco, which studied 144 groundwater wells and reported contamination by fecal coliforms, total coliforms, fecal streptococci, and *Pseudomonas aeruginosa* at levels exceeding WHO and Moroccan standards. The contamination in that study was attributed to microbial pollution from agricultural manure and animal excretions due to insufficient well protection (Lotfi et al. 2020).

In this study, pumps 1 and 2 had higher levels of fecal contamination indicators compared to pump 3, likely due to their proximity to agricultural fields, consistent with the observations of the Moroccan study (Lotfi et al. 2020). This observation highlights the influence of agricultural practices on groundwater quality and the need for improved protection and regular monitoring of groundwater sources in such settings.

Regarding potentially pathogenic microorganisms, *Pseudomonas aeruginosa*, *E. coli*, *Salmonella*, and *Mycobacterium tuberculosis* were detected in the studied groundwater samples. These pathogens have been previously identified as primary contaminants in groundwater (Krauss and Griebler 2011; Pandey et al. 2014; Dong et al. 2024). As previously highlighted in literature, groundwater contamination can arise from multiple sources, including leakage from septic tanks or sewage systems, runoff from agricultural fields, and the dispersal

of manure (Pandey et al. 2014; Dong et al. 2024). In our study, pumps 1 and 2 exhibited a higher diversity of pathogenic microorganisms, likely related to their locations. Pump 1 is located next to agricultural fields, where runoff and manure application may contribute to contamination. Pump 2, on the other hand, is adjacent to residential buildings that lack proper sewage infrastructure, with waste being managed through septic tanks, and is also near an agricultural field. These findings emphasize the influence of agriculture and inadequate wastewater management on groundwater quality in these areas.

All such results, however, must always be taken with caution, since within any of these bacterial taxa, some strains are pathogenic and some are not. Thus, supporting taxonomic findings with the detection of virulence genes, discussed below, is important.

The ability of pathogenic microorganisms to infect hosts and overcome host defense mechanisms is largely mediated by virulence factors (Beceiro et al. 2013). Here we explored virulence factor genes in the metagenomes of the studied groundwater samples to assess bacterial pathogenicity. Forty-seven VF genes were identified across all three pumps, exhibiting variation among the pumps. Overall, the groundwater microbiome from pump 2 had the most diverse range of VF genes, possibly because of agricultural runoffs and sewage contamination, which potentially increase the abundance and diversity of virulence factors carried by potential pathogens (Zhu et al. 2022; Su et al. 2024).

These factors were primarily associated with three key virulence mechanisms: (1) invasion, such as type VI secretion system tubule-forming protein VipA (encoded by *hsiB1/vipA*), which play a role in the pathogenicity of *Pseudomonas aeruginosa* (Hood et al. 2010), (2) immune modulation, e.g., ESX-5 type VII secretion system and transcriptional positive regulator, PhoP, (encoded by *esxN* and *esxM*), which are crucial for the survival and ability of pathogenic *Mycobacterium tuberculosis* to cause disease (Houben et al. 2014; He et al. 2016), and (3) nutritional/metabolic factors like RelA, which is a critical regulator in *Mycobacterium tuberculosis*. In their meta-analysis of the global groundwater, Liu and colleagues reported a different set of virulence factors associated with motility and adherence, as well as the lipopolysaccharide (LPS), which is known for its endotoxin activity (Liu et al. 2023).

More specifically, we found that one of the highly abundant VF genes in pump 1 and pump 2, *hsiB1/vipA*, plays a role in the pathogenicity of *Pseudomonas aeruginosa* (Lossi et al. 2013), which is also among the most abundant pathogens in these samples. This suggests a link between VF abundance and the dominant pathogenic microbial taxa. However, a direct correlation between the most abundant virulence factors and the most

dominant pathogenic microorganisms is not necessarily to be expected. Many virulence genes can be horizontally transferred via phages and conjugative transposons, enabling their presence in multiple bacterial species, including non-pathogenic community members (Frost et al. 2005).

ARGs tend to vary between environments, and their abundance is influenced by local antibiotic consumption patterns in humans, animals and agriculture (Zhang et al. 2015; Cedeño-Muñoz et al. 2024). In our study, ARGs identified groundwater included resistance to aminoglycosides, beta-lactams, tetracyclines, and sulfonamides which are commonly used in Egyptian community pharmacies (Sabry et al. 2014) and these results are consistent with those highlighted by Szekeres and colleagues (Szekeres et al. 2018) and Yang and colleagues (Yang et al. 2024). In contrast, Liu and colleagues (Liu et al. 2023) reported rifamycin, polyketide, and quinolone resistance genes as the most abundant. This suggests regional differences in ARG profiles potentially due to distinct antibiotic usage and environmental factors.

Overall, pump 2 had the highest abundance and diversity of ARGs, along with the broadest range of antibiotic classes to which resistance is potentially conferred, followed by pump 1 and pump 3. This pattern aligns with the environmental context of the sampling sites. Pump 2 is the most susceptible to contamination due to the combined impact of inadequate sewage disposal and potential agricultural runoff from adjacent fields, both of which have been strongly linked to elevated AMR abundance and diversity (Karkman et al. 2019; Fernanda et al. 2022). Pump 1, albeit less impacted, is exposed to agricultural runoff, which likely contributed to its intermediate AMR levels. This gradient in contamination potential (pump 2 > pump 1 > pump 3) helps explain the observed variation in ARG profiles. These findings underscore the public health risks posed by untreated groundwater sources, particularly from pumps 2 and 1, and support the notion that groundwater can serve as a global reservoir for AMR (Andrade et al. 2020).

In our study, we identified various MGEs, including plasmids, insertion sequences, and integrons in all pumps; their abundance varied between pumps though IS were more abundant than plasmids and integrons across all pumps. Insertion sequences are indeed among the most abundant MGEs, both in environmental samples (Elbehery et al. 2017; Li et al. 2020) and microbial genomes (Touchon and Rocha 2007). Their prevalence is underscored by findings that transposase genes, the hallmark of insertion sequences, were the most abundant genes among over 10 million protein-coding sequences analyzed from sequenced genomes and metagenomes (Aziz et al. 2010). This high abundance of insertion sequences is probably attributable to their

small size, self-propagating nature, simplicity, minimal genetic requirements, and indispensable role in genome plasticity and environmental adaptation (Siguier et al. 2015). Despite the relatively high abundance of MGEs in pump 3, the abundance of VFs and ARGs was comparatively low, suggesting that MGEs are not always associated with increased pathogenicity and resistance. Integrons are notably abundant in both hospital-associated groundwater (Liu et al. 2024) and pristine environments, such as the South China Sea (Chen et al. 2013), and soil from Arctic regions (McCann et al. 2019). Plasmids correlated with ARG transfer in different environments. For example, they facilitate the HGT of ARGs from surface sources to groundwater in livestock farms (Lei et al. 2024), and in activated sludge (Zhang et al. 2011). Interestingly, some of the ARGs detected in the pumps were previously reported to have been mobilized by MGEs studied. For instance, sulfonamide resistance genes, such as *sul1*, were mobilized through both plasmids and integrons, whereas *sul2* was predominantly transferred with plasmids (Jiang et al. 2019). Additionally, the *aadA1* gene, which encodes resistance to aminoglycosides, was primarily mobilized by integrons (Martínez et al. 2007).

Insertion sequences potentially contribute to ARG mobility and could modulate their expression (Vandecraen et al. 2017). Specifically, IS6 family members, such as IS26 and IS257, are found on both chromosome and plasmids where they help in the rearrangement and spread of ARGs (Varani et al. 2021), interestingly, this family is detected in all pumps. Additionally, IS families IS3, IS5, and IS21 were highly abundant in all pumps, while IS200/IS605 and IS110 were highly abundant only in pump 3. Remarkably, these IS families have also been detected in other aquatic environments, such as the Red Sea brine pools, where a weak positive correlation between insertion sequences and antimicrobial resistance was observed (Elbehery et al. 2023).

The functional impact of these insertion sequences on antimicrobial resistance has been previously documented. For instance, in *Helicobacter pylori*, the insertion of IS605 into the nitroreductase gene has been linked to metronidazole resistance (Debets-Ossenkopp et al. 1999). Similarly, IS21 has been shown to correlate with multiple ARGs in drinking water treatment systems (Gu et al. 2022), while IS3 transposases have been found flanking the *floR* gene, which encodes a chloramphenicol transporter, in multidrug-resistant *Acinetobacter baumannii* isolates (Kumburu et al. 2019).

The co-localization of ARGs with MGEs, particularly plasmids and insertion sequences underscores a concerning potential for HGT in groundwater microbial communities. This association is a hallmark of mobile ARGs, which are often embedded in MGEs facilitating

their dissemination. Insertion sequences are characterized by their ability to transpose segments of DNA, including ARGs, from chromosomes to MGEs like plasmids that can disseminate these segments across taxonomic barriers (Ebmeyer et al. 2021). This concerted interaction between insertion sequences and plasmids thus plays a pivotal role in ARG mobilization, explaining why the repeated association between ARGs, plasmids and insertion sequences in pumps 1 and 3 could be alarming. The *tet(C)* gene (Table 1) was associated with both a plasmid and an IS3 family insertion sequence. Interestingly, a similar case has been reported in which *tet(C)* was carried by a plasmid-associated IS26 element in nine out of eleven *Aeromonas media* strains (Shi et al. 2021). Similarly, *aadA9*, a variant of the *aadA11* aminoglycoside adenylyltransferase gene detected in our study, was found to be flanked with exact copies of the insertion sequence IS6100 (a member of the IS6 family) on the pTET3 plasmid from *Corynebacterium glutamicum* (Tauch et al. 2002). Remarkably, the frequent detection of aminoglycoside resistance genes co-localized with MGEs across all pumps probably reflects a pattern seen in both pathogenic and nonpathogenic bacterial genomes, especially with conjugation systems (Lund et al. 2023). Overall, these findings point to the role of groundwater as a potential reservoir for mobile ARGs, emphasizing the public health risk associated with such environmental reservoirs of resistance and supporting the One Health perspective (Martak et al. 2024).

Conclusions

In conclusion, the studied groundwater samples were collected from hand pumps located in areas with varying levels of contamination: pump 1 was near agricultural land; pump 2 was situated close to residential houses lacking proper sewage disposal and also near agricultural land, while pump 3 was in an area with comparatively lower contamination than the other two. This difference in location was reflected in the quality of groundwater from pumps 1 and 2, which appeared to be inferior to that of pump 3, based on the level of contamination with bacterial indicators of fecal pollution in addition to the levels of potential pathogens, ARGs and VFs, raising an alarm concerning the municipal use of these sources of groundwater and their associated risks for public health. Additionally, the identified ARGs across all samples confer resistance to beta-lactams, tetracyclines, and sulfonamides, which are antibiotics commonly used in Egyptian community pharmacies. Finally, the detection of several potentially mobilizable ARGs, associated with plasmids and insertion sequences, highlights the groundwater's potential as a reservoir for antimicrobial resistance with a concerning mobility capacity.

Abbreviations

AMR	Antimicrobial resistance
ARB	Antimicrobial-resistant bacteria
ARGs	Antimicrobial resistance genes
IS	Insertion sequences
MCE	Mixed cellulose ester
MGEs	Mobile genetic elements
NMDS	Non-metric multidimensional scaling
ORFs	Open reading frames
RPKM	Reads per kilobase per million reads
SRA	Sequence Read Archive
VFDB	Virulence Factor Database
VFs	Virulence factors

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13213-025-01801-1>.

Supplementary Material 1

Acknowledgements

The authors extend their gratitude to the Cancer Biology, Biochemistry, and Clinical Pathology Departments at the National Cancer Institute, Egypt, for their help with DNA extraction and quantification. The authors also thank Professor Ahmed Moustafa and Mr. Amged Ouf from the American University in Cairo for their valuable help with sample filtration.

Author contributions

AHAE conceived the idea, secured fund; MMM, WAE performed sample preparation and DNA extraction; AHAE, MMM developed the data analysis pipeline; RKA, ASY, AHAE, WAE directed the research and provided scientific insights and critical review of the manuscript; KAMA, AME, MBZ contributed to the coordination and management of the project; MMM wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

This study is based on work supported by the University of Sadat City (USC), under grant No (23).

Data availability

The raw reads for this study are available through the NCBI Sequence Read Archive (SRA) with BioProject accession PRJNA1196429. All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Received: 4 January 2025 / Accepted: 27 April 2025

Published online: 15 May 2025

References

Andrade L, Kelly M, Hynds P et al (2020) Groundwater resources as a global reservoir for antimicrobial-resistant bacteria. *Water Res* 170:115360. <https://doi.org/10.1016/j.watres.2019.115360>

- Antipov D, Raiko M, Lapidus A, Pevzner PA (2019) Plasmid detection and assembly in genomic and metagenomic data sets. *Genome Res* 29:961–968. <https://doi.org/10.1101/GR.241299.118>
- Aziz RK, Breitbart M, Edwards RA (2010) Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res* 38:4207–4217. <https://doi.org/10.1093/nar/gkq140>
- Beceiro A, Tomás M, Bou G (2013) Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin Microbiol Rev* 26:185. <https://doi.org/10.1128/CMR.00059-12>
- Breitwieser FP, Salzberg SL (2020) Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* 36:1303–1304. <https://doi.org/10.1093/BIOINFORMATICS/BTZ715>
- Bushnell B (2015) BBMap. <https://sourceforge.net/projects/bbmap/>. Accessed 19 Oct 2024
- Byappanahalli MN, Nevers MB, Korajkic A et al (2012) Enterococci in the environment. *Microbiol Mol Biol Rev* 76:685–706. <https://doi.org/10.1128/MMBR.0023-12>
- Cedeño-Muñoz JS, Aransiola SA, Reddy KV et al (2024) Antibiotic resistant bacteria and antibiotic resistance genes as contaminants of emerging concern: occurrences, impacts, mitigations and future guidelines. *Sci Total Environ* 952:175906. <https://doi.org/10.1016/J.SCITOTENV.2024.175906>
- Chen B, Yang Y, Liang X et al (2013) Metagenomic profiles of antibiotic resistance genes (ARGs) between human impacted estuary and deep ocean sediments. *Environ Sci Technol* 47:12753–12760. <https://doi.org/10.1021/ES403818E>
- Chen L, Zheng D, Liu B et al (2016) VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res* 44:D694–D697. <https://doi.org/10.1093/nar/gkv1239>
- Chen S, Zhou Y, Chen Y, Gu J (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/BIOINFORMATICS/BTY560>
- R Core Team (2024) R: A language and environment for statistical computing
- De Giglio O, Caggiano G, Bagordo F et al (2017) Enteric viruses and fecal bacteria indicators to assess groundwater quality and suitability for irrigation. *Int J Environ Res Public Health* 14:558. <https://doi.org/10.3390/IJERPH14060558>
- Debets-Ossenkopp YJ, Pot RG, van Westerloo DJ et al (1999) Insertion of mini-IS605 and deletion of adjacent sequences in the nitroreductase (*rdxA*) gene cause metronidazole resistance in *Helicobacter pylori* NCTC11637. *Antimicrob Agents Chemother* 43:2657–2662. <https://doi.org/10.1128/AAC.43.11.2657>
- Díaz-Torres O, los Cobos EOV, Kreft JU et al (2024) A metagenomic study of antibiotic resistance genes in a hypereutrophic subtropical lake contaminated by anthropogenic sources. *Sci Total Environ* 927:172216. <https://doi.org/10.1016/J.SCITOTENV.2024.172216>
- Dong Y, Zhou J, Yidan H et al (2024) Pathogen contamination of groundwater systems and health risks. *Crit Rev Environ Sci Technol* 54:267–289. <https://doi.org/10.1080/10643389.2023.2236486>
- Dubin K, Pamer EG (2014) Enterococci and their interactions with the intestinal microbiome. *Microbiol Spectr*. 5: <https://doi.org/10.1128/MICROBIOLSPEC.B-AD-0014-2016>
- Ebmeyer S, Kristiansson E, Larsson DGJ (2021) A framework for identifying the recent origins of mobile antibiotic resistance genes. *Commun Biol* 4:8. <https://doi.org/10.1038/s42003-020-01545-5>
- Elbeheri AH, Aziz RK, Siam R (2017) Insertion sequences enrichment in extreme Red Sea brine pool vent. *Extremophiles* 21:271–282. <https://doi.org/10.1007/s00792-016-0900-4>
- Elbeheri AHA, Beason E, Siam R (2023) Metagenomic profiling of antibiotic resistance genes in Red Sea brine pools. *Arch Microbiol* 205:195. <https://doi.org/10.1007/s00203-023-03531-x>
- Ewida AYI, Khalil MS, Mahmoud AMA (2021) Impact of domestic wastewater treatment plants on the quality of shallow groundwater in Qalyubia, Egypt; discrimination of microbial contamination source using box-pcr. *Egypt J Bot* 61:127–139. <https://doi.org/10.21608/ejbo.2020.30986.1505>
- Farid I, Abbas M, Bassouny M et al (2019) Indirect impacts of irrigation with low quality water on the environmental safety. *Egypt J Soil Sci* 60:1–15. <https://doi.org/10.21608/ejss.2019.15434.1294>
- Feldgarden M, Brover V, Haft DH et al (2019) Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 63:e00483–e00419. <https://doi.org/10.1128/aac.00483-19>
- Ferguson AS, Layton AC, Mailloux BJ et al (2012) Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater. *Sci Total Environ* 431:314. <https://doi.org/10.1016/J.SCITOTENV.2012.05.060>

- Fernanda PA, Liu S, Yuan T et al (2022) Diversity and abundance of antibiotic resistance genes and their relationship with nutrients and land use of the inflow rivers of Taihu lake. *Front Microbiol* 13:1009297. <https://doi.org/10.3389/fmicb.2022.1009297>
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37. <https://doi.org/10.1093/NAR/GKR367>
- Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Reviews Microbiol* 3:722–732. <https://doi.org/10.1038/nrmicro1235>
- Fu L, Niu B, Zhu Z et al (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152. <https://doi.org/10.1093/BIOINFORMATICS/BTS565>
- Griebler C, Lueders T (2009) Microbial biodiversity in groundwater ecosystems. *Freshw Biol* 54:649–677. <https://doi.org/10.1111/J.1365-2427.2008.02013.X>
- Gu Q, Sun M, Lin T et al (2022) Characteristics of antibiotic resistance genes and antibiotic-resistant bacteria in full-scale drinking water treatment system using metagenomics and culturing. *Front Microbiol* 12:798442. <https://doi.org/10.3389/fmicb.2021.798442/BIBTEX>
- Hassanein NM, El-Baghady KZ, Id AKF et al (2012) Microbial communities and water quality of some groundwater systems in Egypt. *Egypt J Exp Biol (Bot)* 8:25–42
- He X, Wang L, Wang S (2016) Structural basis of DNA sequence recognition by the response regulator PhoP in *Mycobacterium tuberculosis*. *Scientific Reports* 6:1–11. <https://doi.org/10.1038/srep24442>
- Hernando-Amado S, Coque TM, Baquero F, Martínez JL (2019) Defining and combating antibiotic resistance from one health and global health perspectives. *Nat Microbiol* 2019 4:9. <https://doi.org/10.1038/s41564-019-0503-9>
- Hood RD, Singh P, Hsu FS et al (2010) A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* 7:25. <https://doi.org/10.1016/J.CHOM.2009.12.007>
- Houben ENG, Korotkov KV, Bitter W (2014) Take five — Type VII secretion systems of mycobacteria. *Biochimica et biophysica acta (BBA) - Mol Cell Res* 1843:1707–1716. <https://doi.org/10.1016/J.BBAMCR.2013.11.003>
- Huang X, Toro M, Reyes-Jara A et al (2024) Integrative genome-centric metagenomics for surface water surveillance: elucidating microbiomes, antimicrobial resistance, and their associations. *Water Res* 264:122208. <https://doi.org/10.1016/J.WATRES.2024.122208>
- Hyatt D, Chen G-L, LoCasico PF et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
- Ji L, Zhang L, Wang Z et al (2022) High biodiversity and distinct assembly patterns of microbial communities in groundwater compared with surface water. *Sci Total Environ* 834:155345. <https://doi.org/10.1016/J.SCITOTENV.2022.155345>
- Jia X, O'Connor D, Hou D et al (2019) Groundwater depletion and contamination: Spatial distribution of groundwater resources sustainability in China. *Sci Total Environ* 672:551–562. <https://doi.org/10.1016/J.SCITOTENV.2019.03.457>
- Jiang H, Cheng H, Liang Y et al (2019) Diverse mobile genetic elements and conjugal transferability of sulfonamide resistance genes (sul1, sul2, and sul3) in *Escherichia coli* isolates from *Penaeus vannamei* and pork from large markets in Zhejiang, China. *Front Microbiol* 10:1787. <https://doi.org/10.3389/fmicb.2019.01787>
- Karkman A, Pärnänen K, Larsson DGJ (2019) Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. *Nat Commun* 10:80. <https://doi.org/10.1038/s41467-018-07992-3>
- Kaushik R, Balasubramanian R, Dunstan H (2014) Microbial quality and phylogenetic diversity of fresh rainwater and tropical freshwater reservoir. *PLoS ONE* 9:e100737. <https://doi.org/10.1371/journal.pone.0100737>
- Kim G-H, Lee K, Shim JO (2023) Gut bacterial dysbiosis in irritable bowel syndrome: a case-control study and a cross-cohort analysis using publicly available data sets. *Microbiol Spectr* 11:e02125–e02122. <https://doi.org/10.1128/spectrum.02125-22>
- Krauss S, Griebler C (2011) Pathogenic microorganisms and viruses in Groundwater - acatech -. National Academy of Science and Engineering. Acatech materials
- Kumburu HH, Sonda T, Van Zwetselaar M et al (2019) Using WGS to identify antibiotic resistance genes and predict antimicrobial resistance phenotypes in MDR *Acinetobacter baumannii* in Tanzania. *J Antimicrob Chemother* 74:1484–1493. <https://doi.org/10.1093/JAC/DKZ055>
- Lei L, Chen N, Chen Z et al (2024) Dissemination of antibiotic resistance genes from aboveground sources to groundwater in livestock farms. *Water Res* 256:121584. <https://doi.org/10.1016/J.WATRES.2024.121584>
- Li D, Liu CM, Luo R et al (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li P, Wu J, Qian H et al (2016) Hydrogeochemical characterization of groundwater in and around a wastewater irrigated forest in the southeastern edge of the Tengger desert, Northwest China. *Expo Health* 8:331–348. <https://doi.org/10.1007/s12403-016-0193-y>
- Li B, Chen Z, Zhang F et al (2020) Abundance, diversity and mobility potential of antibiotic resistance genes in pristine Tibetan plateau soil as revealed by soil metagenomics. *FEMS Microbiol Ecol* 96:faa172. <https://doi.org/10.1093/femsec/faaa172>
- Liu C, Chen J, Shan X et al (2023) Meta-analysis addressing the characterization and risk identification of antibiotics and antibiotic resistance genes in global groundwater. *Sci Total Environ* 860:160513. <https://doi.org/10.1016/j.scitotenv.2022.160513>
- Liu K, Gan C, Peng Y et al (2024) Occurrence and source identification of antibiotics and antibiotic resistance genes in groundwater surrounding urban hospitals. *J Hazard Mater* 465:133368. <https://doi.org/10.1016/J.JHAZMAT.2023.133368>
- Lossi NS, Manoli E, Förster A et al (2013) The HsiB1C1 (TssB-TssC) complex of the *Pseudomonas aeruginosa* type VI secretion system forms a bacteriophage tail sheathlike structure. *J Biol Chem* 288:7536–7548. <https://doi.org/10.1074/jbc.M112.439273>
- Lotfi S, Chakit M, Najy M et al (2020) Assessment of microbiological quality of groundwater in the Saïs plain (Morocco). *Egypt J Aquat Biol Fish* 24:509–524. <https://doi.org/10.21608/EJABF.2020.73595>
- Lu J, Breitwieser FP, Thielen P, Salzberg SL (2017) Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* 3:e104. <https://doi.org/10.7717/peerj-cs.104>
- Lu J, Rincon N, Wood DE et al (2022) Metagenome analysis using the Kraken software suite. *Nat Protoc* 17:2815. <https://doi.org/10.1038/S41596-022-00738-Y>
- Lund D, Coertze RD, Parras-Moltó M et al (2023) Extensive screening reveals previously undiscovered aminoglycoside resistance genes in human pathogens. *Commun Biol* 6:812. <https://doi.org/10.1038/s42003-023-05174-6>
- Martak D, Henriot CP, Hocquet D (2024) Environment, animals, and food as reservoirs of antibiotic-resistant bacteria for humans: one health or more? *Infect Dis now* 54:104895. <https://doi.org/10.1016/j.idnow.2024.104895>
- Martínez N, Mendoza CM, Rodríguez I et al (2007) Detailed structure of integrons and transposons carried by large conjugative plasmids responsible for multidrug resistance in diverse genomic types of *Salmonella enterica* serovar Brandenburg. *J Antimicrob Chemother* 60:1227–1234. <https://doi.org/10.1093/JAC/DKM336>
- McCann CM, Christgen B, Roberts JA et al (2019) Understanding drivers of antibiotic resistance genes in high Arctic soil ecosystems. *Environ Int* 125:497–504. <https://doi.org/10.1016/J.ENVINT.2019.01.034>
- McMurdie PJ, Holmes S (2012) PhyloSeq: A bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. *Pac Symp Biocomputing* 235–246. https://doi.org/10.1142/9789814366496_0023
- Mistry J, Chuguransky S, Williams L et al (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res* 49:D412–D419. <https://doi.org/10.1093/NAR/GKAA913>
- Nanayakkara AK, Boucher HW, Fowler VG Jr et al (2021) Antibiotic resistance in the patient with cancer: escalating challenges and paths forward. *CA Cancer J Clin* 71:488–504. <https://doi.org/10.3322/caac.21697>
- Néron B, Littner E, Haudiquet M et al (2022) IntegronFinder 2.0: identification and analysis of integrons across bacteria, with a focus on antibiotic resistance in *Klebsiella*. *Microorganisms* 10:700. <https://doi.org/10.3390/microorganisms10040700>
- Nguyen CC, Hugie CN, Kile ML, Navab-Daneshmand T (2019) Association between heavy metals and antibiotic-resistant human pathogens in environmental reservoirs: A review. *Front Environ Sci Eng* 13. <https://doi.org/10.1007/s11783-019-1129-0>
- Nomoto R, Osawa K, Kinoshita S et al (2024) Metagenome and resistome analysis of beta-lactam-resistant bacteria isolated from river waters in Surabaya, Indonesia. *Microorganisms* 12:199. <https://doi.org/10.3390/MICROORGANISMS12010199/S1>
- Oksanen J, Simpson GL, Blanchet FG et al (2024) Package vegan: an R package for community ecologists. R package version 2.7-0. Accessed 5 Oct 2024 <https://github.com/vegandevs/vegan>
- Pandey PK, Kass PH, Soupir ML et al (2014) Contamination of water resources by pathogenic bacteria. *AMB Express* 4:51. <https://doi.org/10.1186/s13568-014-0051-x>

- Paruch L, Paruch AM, Eiken HG, Sørheim R (2019) Faecal pollution affects abundance and diversity of aquatic microbial community in anthropo-zoogenically influenced lotic ecosystems. *Sci Rep* 9:19469. <https://doi.org/10.1038/s41598-019-56058-x>
- Polianciuc SI, Gurzău AE, Kiss B et al (2020) Antibiotics in the environment: causes and consequences. 93:231–240. <https://doi.org/10.15386/mpr-1742>
- Richiardi L, Pignata C, Fea E et al (2023) Are indicator microorganisms predictive of pathogens in water? *Water* 2023, Vol 15, Page 2964 15:2964. <https://doi.org/10.3390/W15162964>
- Ruiz-Tagle C, Ugalde JA, Naves R et al (2023) Reduced microbial diversity of the nasopharyngeal microbiome in household contacts with latent tuberculosis infection. *Sci Rep* 13:7301. <https://doi.org/10.1038/s41598-023-34052-8>
- Sabry NA, Farid SF, Dawoud DM (2014) Antibiotic dispensing in Egyptian community pharmacies: an observational study. *Res Social Administrative Pharm* 10:168–184. <https://doi.org/10.1016/J.SAPHARM.2013.03.004>
- Sazykina MA, Minkina TM, Konstantinova EY et al (2022) Pollution impact on microbial communities composition in natural and anthropogenically modified soils of Southern Russia. *Microbiol Res* 254:126913. <https://doi.org/10.1016/j.micres.2021.126913>
- Seemann T (2020) ABRicate. <https://github.com/tseemann/abricate>. Accessed 19 Oct 2024
- Shi Y, Zhang Y, Wu X et al (2021) Potential dissemination mechanism of the *tetC* gene in *Aeromonas* media from the aerobic biofilm reactor under oxytetracycline stresses. *J Environ Sci* 105:90–99. <https://doi.org/10.1016/j.jes.2020.12.038>
- Siguier P, Goubeyre E, Varani A et al (2015) Everyman's guide to bacterial insertion sequences. <https://doi.org/10.1128/microbiolspec.MDNA3-0030-2014>. *Microbiol Spectr* 3:
- Su Z, Cui S, Wen D, Chen L (2024) Metagenomic insights into resistome, mobilome and virulome in different fecal waste. *Environ Res* 262:119861. <https://doi.org/10.1016/j.envres.2024.119861>
- Szekeres E, Chiriac CM, Baricz A et al (2018) Investigating antibiotics, antibiotic resistance genes, and microbial contaminants in groundwater in relation to the proximity of urban areas. *Environ Pollut* 236:734–744. <https://doi.org/10.1016/j.envpol.2018.01.107>
- Tauch A, Götter S, Pühler A et al (2002) The 27.8-kb R-plasmid pTET3 from *Corynebacterium glutamicum* encodes the aminoglycoside adenylyltransferase gene cassette *aadA9* and the regulated tetracycline efflux system *tet 33* flanked by active copies of the widespread insertion sequence IS6100. *Plasmid* 48:117–129. <https://linkinghub.elsevier.com/retrieve/pii/S0147619X02001208>
- Touchon M, Rocha EPC (2007) Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol* 24:969–981. <https://doi.org/10.1093/molbev/msm014>
- UNESCO (2022) World Water Development Report 2022: Groundwater— Making the Invisible Visible: Agriculture. <https://www.unesco.org/reports/wwdr/2022/en/agriculture>. Accessed 16 Nov 2024
- Valero-Mora PM (2010) ggplot2: elegant graphics for data analysis. *J Stat Softw* 35:1–3. <https://doi.org/10.18637/JSS.V035.B01>
- Vandecraen J, Chandler M, Aertsen A, Van Houdt R (2017) The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol* 43:709–730. <https://doi.org/10.1080/1040841X.2017.1303661>
- Varani A, He S, Siguier P et al (2021) The IS6 family, a clinically important group of insertion sequences including IS26. *Mob DNA* 12:1–18. <https://doi.org/10.1186/S13100-021-00239-X/FIGURES/11>
- Vaz-Moreira I, Egas C, Nunes OC, Manaia CM (2011) Culture-dependent and culture-independent diversity surveys target different bacteria: A case study in a freshwater sample. *Antonie Van Leeuwenhoek*. *Int J Gen Mol Microbiol* 100:245–257. <https://doi.org/10.1007/s10482-011-9583-0>
- Xie Z, Tang H (2017) ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33:3340–3347. <https://doi.org/10.1093/BIOINFORMATICS/BTX433>
- Yan L, Herrmann M, Kampe B et al (2020) Environmental selection shapes the formation of near-surface groundwater microbiomes. *Water Res* 170. <https://doi.org/10.1016/j.watres.2019.115341>
- Yan L, Hermans SM, Totsche KU et al (2021) Groundwater bacterial communities evolve over time in response to recharge. *Water Res* 201. <https://doi.org/10.1016/j.watres.2021.117290>
- Yang Y, Li H, Wang D et al (2024) Metagenomics of high-altitude groundwater reveal different health risks associated with antibiotic-resistant pathogens and bacterial resistome in the latitudinal gradient. *Water Res* 262:122032. <https://doi.org/10.1016/J.WATRES.2024.122032>
- Yao G, Zhang W, Yang M et al (2020) MicroPhenoDB associates metagenomic data with pathogenic microbes, microbial core genes, and human disease phenotypes. *Genomics Proteom Bioinf* 18:760–772. <https://doi.org/10.1016/J.GPB.2020.11.001>
- Zhang T, Zhang XX, Ye L (2011) Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS ONE* 6:e26041. <https://doi.org/10.1371/JOURNAL.PONE.0026041>
- Zhang Q-Q, Ying G-G, Pan C-G et al (2015) Comprehensive evaluation of antibiotics emission and fate in the river basins of China: source analysis, multimedia modeling, and linkage to bacterial resistance. *Environ Sci Technol* 49:6772–6782. <https://doi.org/10.1021/acs.est.5b00729>
- Zhu L, Lian Y, Lin D et al (2022) Insights into microbial contamination in multi-type manure-amended soils: the profile of human bacterial pathogens, virulence factor genes and antibiotic resistance genes. *J Hazard Mater* 437:129356. <https://doi.org/10.1016/j.jhazmat.2022.129356>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.