Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-022-22197-4

RESEARCH ARTICLE

# Exploring antibiotic resistance genes, mobile gene elements, and virulence gene factors in an urban freshwater samples using metagenomic analysis

**Sohail M. Noman[1] · Muhammad Shafiq[1] · Shabana Bibi[2,3] · Bharti Mittal[4] · Yumeng Yuan[1] · Mi Zeng[1] · Xin Li[1] · Oluwaseyi Abraham Olawale[5] · Xiaoyang Jiao[1] · Muhammad Irshad[6]**

Received: 8 February 2022 / Accepted: 20 July 2022


## Abstract

Antibiotic resistance genes (ARGs) and antimicrobial resistance elements (AMR) are novel environmental contaminants that pose a significant risk to human health globally. Freshwater contains a variety of microorganisms that might affect human health; its quality must be assessed before use. However, the dynamics of mobile genetic elements (MGEs) and ARG propagation in freshwater have rarely been studied in Singapore. Therefore, this study used metagenomics to compare diversity, virulence factor composition, and ARG and MGE co-occurrence with bacterial communities in paired ($n=8$) environmental freshwater samples. KneadData, FMAP, and Kraken2 were used for bioinformatics analysis and R (v4.1.1) for statistical analysis. Sequence reads with a total of 9043 species were taxonomically classified into 66 phyla, 130 classes, 261 orders, 584 families, and 2477 genera. *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, and *Firmicutes* were found the Phyla in all samples. Analysis of QIIME output by PICRUSt and ß-diversity showed unique clusters and functional microbial community structures. A total of 2961 ARGs were found that conferred resistance to *multidrug*, *aminoglycosides*, *tetracyclines*, *elfamycins*, and more. The classified ARG mechanism revealed significant distribution of virulence factors in bacterial cells. Transposes and transposon were highly correlated to ARG gene transfer. Co-occurrence network analysis showed several MGEs appear to use the same ARGs (*intI* and *rho*) and were dominant in all samples. Furthermore, ARGs are also highly correlated with bacteria like *Campylobacter* and *Escherichia*. This study enhances the understanding of antibiotic risk assessment and provides a new perspective on bacterial assembly contamination and the functional prevalence of ARGs and MGEs with antibiotic resistance bacteria. Moreover, it raises public awareness because these contaminants put people's lives at risk of acquiring bacterial infections. In addition, it can also help propose hybrid water treatment approaches.

**Keywords** Antibiotic resistance genes · Bacterial diversity · Environmental infections · Freshwater · Metagenomic · Mobile genetic elements · Virulence gene factors

## Introduction

Antibiotic resistance genes (ARGs) and antibiotic-resistant bacteria (ARBs) are responsible for contaminating water resources and are significant threats to the environment and human health globally, which are discharged into the environment and present in water treatment plants (WTPs) in urban areas (Yoo et al. 2019). Despite the prevalence of ARGs, the dynamics of mobile gene elements (MGEs) in WTPs have rarely been studied in Singapore (Sahani et al. 2022). According to the environmental assessment report from the United Nations (https://www.unwater.org/) (UNEP 2021), antibiotic resistance is the most pressing worldwide public health issue. Horizontal gene transfer (HGT) is an important mechanism by which genes are transferred from bacterial species to different recipients like plants, animals, and fungal species and becomes the major cause of pathogenic evolution and promotes ARGs and ARBs. Various reports identify WTP reservoirs as a possible source of ARGs and ARBs released into the environment (Barancheshme and Munir 2018; Che et al. 2019; Alexander et al. 2020) because sewage from hospitals and

Responsible Editor: Gerald Thouand

✉ Xiaoyang Jiao
xyjiao@stu.edu.cn

Extended author information available on the last page of the article

Published online: 08 August 2022

homes contains a huge amount of antibiotics and human disease-causing bacteria that are likely to act as discerning pressure on ARBs and resistance genes (RGs). In addition, the number of ARGs and ARBs may also be increased due to certain circumstances in the WTPs during the treatment process (Dávalos et al. 2021).

When it comes to addressing ARGs as a public health issue, it is still not obvious whether WTPs pose the highest microbiological hazard or risk (Fang et al. 2014; Mohiuddin and Schellhorn 2015; Zhao et al. 2018). Currently, no comparison studies have been conducted in Singapore on the overall incidence, including a variety of distinct ARGs implicated in urban and ambient WTP processes (Mitchell et al. 2018). In spite of the fact that a lot of research has been conducted globally to determine the differences in abundance and diversity of ARGs and MGEs, studies are still lacking to show antibiotic contaminants are the result of urban and ambient WTP activities (White et al. 2016; Fang et al. 2019; Palermo et al. 2019; Chopyk et al. 2020). Indeed, there are still questions about which ARGs are significant for bacterial populations to acquire and exchange via MGEs (Saparbaevna Alexyuk et al. 2017). Hence, when it comes to understanding the diversity and composition of microbial communities in environmental samples, metagenomic approaches are considered the most reliable and cost-effective methodologies (Grant 2022; Shilpa et al. 2022). Therefore, metagenomics can be applied to better understand the variation in ARGs and MGEs profiles (Meneghine et al. 2017).

The main aim of this study is to investigate the ARGs, MGEs, and virulence factors in urban freshwater samples using metagenomic analysis and further identify the resistance mechanisms and bacterial contamination. In addition, the comparative analysis of functional observed bacterial operational taxonomic units (OTUs) competencies is anticipated in the co-occurrence network. The fundamental significance of this study is to raise awareness in the population who drinks freshwater about the potential host and bacterial interactions because these contaminants are putting their lives at risk of acquiring antibiotic resistant bacterial infections and other viral diseases.

## Method

### Data samples

This study utilizes eight paired (forward and reverse reads) environmental metagenome data samples associated with the antimicrobial resistance bio project of urban freshwater samples. All raw data samples were collected and registered by the National University of Singapore and are freely available in the online repository (https://www.ncbi.nlm.nih.gov/) with the accession number (PRJNA400857). An initial study about the parameters of the water samples can be found online in the same repository with the accession number (PRJEB30238). In addition, this study provides further bioinformatics pipeline analysis of environmental bacterial interactions with human health on raw data.

### Bioinformatics analysis preprocessing

For subsequent bioinformatics analysis, Knead Data software was used for quality control on the Fastq raw samples based on Trimmomatic and Bowtie2 de-hosting (Kumar Awasthi et al. 2020). The results of the Knead Data were used to identify the cumulative effects and data quality control coherence as well as to cluster the tags into OTUs based on similarity indexes (Brown et al. 2017; Sohail et al. 2019a, b). Kraken2 software was utilized further to assign the OTUs taxonomic ranks (Yang et al. 2021), and diversity calculations were done via R software (v4.1.1) (Liu et al. 2021; Mammola et al. 2021). In addition, Phyloseq was used to perform the alpha and beta diversity based on the taxonomic OTU ranks (Arenas et al. 2021; Vieira and Pecchia 2021). FMAP software was used further to compare and annotate the reads with an ARG database called CARD to determine the resistance of ARGs to antibiotics like aminoglycosides, tetracyclines, elfamycins, cycloserine, quinolones, cephalosporins, isoniazid, lincosamildes, bicyclomycins, fosfomycin, fosmidomycin, multidrug, and peptide drugs (Danko et al. 2021; Yadav and Kapley 2021).

### Metagenomic assembly and plasmid/chromosomal sequence

In total, eight clean WGS Metagenomics samples were submitted to the PATRIC genome assembly pipeline (v3.6.9) on the Illumina platform for assembly (Davis et al. 2020; Parrello et al. 2021). Rectifying bases, correcting misassemblies, and filling gaps were all handled by the SPAdes (v3.12.0) assembler and Pilon (v1.23) (Antipov et al. 2020). Furthermore, the QUAST (v5.0.2) (Taş et al. 2021) was used to rate the quality of assemblies. Plasflow (Galaxy v1.0) was used to predict plasmid and chromosomal sequences from all metagenomic assemble contigs with the default parameters using trained neural network models with 96% accuracy on genomes and plasmid sequences (Bibi et al. 2021; Taş et al. 2021).

### Functional annotation of MAGs

Genes of interest and functional categorization were predicted using the PATRIC (v3.6.9) annotation pipeline with bacteria as the taxonomic target domain, employing the RAST toolkit (RASTtk) (Yousafi et al. 2021). The assembly contigs of all samples were submitted to a pipeline with unique genome identifiers assigned to them, and the genetic codon for translation of most bacteria and bacteriophages was used to call functional features. The KEGG database was mapped for each annotated feature to anticipate the dataset's functional assignments (Bibi et al. 2019; Biswas et al. 2021).

The BLASTP and k-mer-based detection methods were used to identify and quantify the resistome profiles of metagenomic data after annotating them (Saparbaevna Alexyuk et al. 2017; Yadav and Kapley 2021; Danko et al. 2021). This was done to facilitate the detection and quantification of resistome profiles of metagenomic data. To provide antibiotic resistance, virulence factors, known drug targets, and classification of the detected ARGs in all samples into different mechanism categories, PATRIC BLASTs were utilized for all genes, including genes known in the genome against specialty gene databases like ARDB, NDARO, CARD, and PATRIC AMR-related curation (Yousafi et al. 2021). Following the analysis and measurements of ARGs' frequency and relative distribution, the abundances of different resistance mechanisms were tallied by merging the abundances of ARG subtypes associated with a specific resistance mechanism category. All metagenomic assembled contigs were combined to determine the percentage and proportion of different ARG types on the plasmid, chromosomal, and unclassified sequences (Antipov et al. 2020; Davis et al. 2020; Danko et al. 2021; Parrello et al. 2021; Taş et al. 2021).

### Statistical analyses

R (v4.1.1) and SPSS (v28.0) software were used to analyze the data, and graphics were further created with GraphPad Prism (v8.0.2), and *p*-values represent a two-sided statistical test (Sohail et al. 2019a, b; Liu et al. 2021; Taş et al. 2021). To investigate the association between ARG subtypes and bacterial species, Spearman's correlation analysis was conducted using the R software package called Hmisc (v4.6) (Frank 2021) and Psych (v2.1.9) (Revelle 2021). A co-occurrence network was generated using Gephi network-building software (v0.9.2), for which we chose pairs where the correlation coefficient was ($\geq 0.4$) and the *p*-value was ($\leq 0.05$) (Yuan et al. 2021). Prior to statistical analysis, the Plasflow result's bar chart was used to calculate the ratios of various ARG types found on the plasmid, chromosomal, and unclassified sequences, including the frequency and relative distribution of ARG types across all datasets. In addition, heatmaps and circos plots were visualized by using the Galaxy tool (Rasche and Hiltemann 2020).

## Results

### Data quality control and assembly assessment

This study analyzed a total of eight paired environmental metagenomic urban freshwater data samples with a total number of 15,203,491 sequence reads. After filtering the quality control of raw data with the Fastq criteria, we were able to obtain a dataset consisting of a total (10,817,393; 71.15%) high-quality sequence reads, including 10.75% in S1, 8.57% in S2, 12.66% in S3, 17.45% in S4, 21.31% in S5, 10.83% in S6, 14.81% in S7, and 3.63% in S8. The quality control comparison information of raw and filtered reads is presented in Table 1 with the total assembly size in kilobytes (kb) and genome length in base pairs (bp) of each sample. Furthermore, the graphical presentation of raw/filtered data quality done by MultiQC is present in the supplementary file (section-I).

All eight metagenomic assembly genomes (MAGs) contained a total of 8827 contigs, each with a minimum number of 300 bp per contig. The average short-read coverage was

**Table 1** Samples information with their average length, genome length, and assembly size

| Samples | Samples ID | Total reads | Classified reads | Avg length | Assembly size (in kb) | Genome length (in bp) |
|---|---|---|---|---|---|---|
| SRR5997540 | S1 | 1,284,460 | 1,162,657 | 183.90 | 12,464 | 12,382,321 |
| SRR5997543 | S2 | 2,095,861 | 926,931 | 186.80 | 6806 | 6,968,361 |
| SRR5997544 | S3 | 2,194,503 | 1,369,213 | 186.30 | 2572 | 2,633,599 |
| SRR5997546 | S4 | 2,377,924 | 1,887,852 | 191.00 | 5056 | 5,176,671 |
| SRR5997548 | S5 | 2,815,510 | 2,304,889 | 187.80 | 3953 | 4,047,347 |
| SRR5997549 | S6 | 1,896,447 | 1,171,043 | 185.00 | 666 | 681,446 |
| SRR5997550 | S7 | 2,078,551 | 1,602,187 | 185.20 | 4209 | 4,309,853 |
| SRR5997552 | S8 | 460,235 | 392,621 | 189.40 | 666 | 681,446 |

employed for all datasets, including the long-read coverage, scaffolding reads, and average GC content. The assembly showed high quality in terms of contig length of scaffold L50, coarse, and fine consistency as percentages. Quality assessments demonstrated that each genome has the best quality and a high probability of containing more annotated features and ARGs when BLAST against the curated databases.

## Bacterial abundance and diversity assessment

A total of 9043 bacterial species OTUs were found in the dataset. Table 2 shows the percentage of bacterial reads categorized at each level of taxonomic classification, including Kingdom, Phylum, Class, Order, Family, Genus, and Species. For the co-occurrence network analysis, we focused on the environmental homogeneous bacterial categorization of classes, families, genera, and species, whose taxonomic abundance bar chart is provided in the supplementary file (section-II). Figure 1 creates by the Krona chart library by merging the assembly of all samples, which shows *Proteobacteria* has a higher percentage in all samples than *Terrabacteria*. It is useful to look over the various taxonomic classifications and compare the abundance of OTUs according to the Kraken confidence scores (Mreyoud et al. 2022). Furthermore, all the remaining classification levels of taxonomic bar charts are included in the supplementary file (section-III) with their percentage counts for the top 10 bacterial distributions in each sample.

The complexity of an organism's community is intuitively perceived as diversity in the ecological sense (Yousafi et al. 2021). The Shannon and Simpson indices were used in this study to assess the alpha diversity within a single sample, and they were based on the number and relative abundance of taxa at various ranks (e.g., species). The lower the bacterial diversity, the higher the Simpson index will be (Yuan et al. 2021). We found that the Shannon and Simpson diversity percentages for each bacterial species in each data sample were statistically different by utilizing a confidence interval ($P \leq 0.05$), as shown in Table 2. In addition, QIIME software (v1.8) assessed beta diversity to determine the degree to which two communities differ in terms of beta diversity. It was shown that species complexity differs between communities based on Bray–Curtis's metrics that calculate the sum of smaller numbers for species in each community divided by the sum of all counts in each community. Samples were subjected to beta diversity to explain the distribution patterns of two communities in terms of diversity, which was presented in the supplementary file (section-IV).

## Evaluation of functional genome annotations

All MAGs were included for the annotated features in the RASTtk system, which included proteins with functional assignments and Enzyme Commission (EC) numbers as well as Gene Ontology (GO) and KEGG pathway mappings (Gabashvili et al. 2022; Wu et al. 2022). Two types of protein families, namely, PLFams and PGFams, were annotated from the PATRIC database, which was also included in the annotated features (Baniya and Digennaro 2021).

### ARG abundance and diversity assessment

The ARG abundances in each sample are shown in Fig. 2. Per sample distribution, the average abundance of ARGs was calculated as follows: ($6.28 \times 10^{-2}$; for S1), ($1.21 \times 10^{-2}$; for S2), ($5.24 \times 10^{-2}$; for S3), ($1.0 \times 10^{-1}$; for S4), ($3.21 \times 10^{-2}$; for S5), ($0.79 \times 10^{-3}$; for S6), ($1.36 \times 10^{-2}$; for S7), and ($1.0 \times 10^{-1}$; for S8), respectively. It was found in all samples that mostly ARGs accounted for *Aminoglycoside*, *Elfamycins*, *Quinolone*, and *Cycloserine*, with the remainder being made up of *Peptide antibiotics*, *Tetracycline*, *Isoniazid*, *Multidrug*, and *unclassified* ARGs. In addition, *aminoglycoside* was shown to be in high abundance in S1 with a ratio of ($5.17 \times 10^{-2}$) compared to *Cycloserine* with a difference

**Table 2** Bacterial taxonomic classification and diversity percentage index including Shannon and Simpson

| Sample ID | Classification percentage | | | | | | | Diversity percentage at the species level | |
|---|---|---|---|---|---|---|---|---|---|
| | Kingdom | Phylum | Class | Order | Family | Genus | Species | Shannon index (H)/(H/LN (N)) | Simpson index (1-D) |
| S1 | 90.52 | 89.73 | 88.58 | 87.74 | 79.25 | 76.91 | 65.05 | 5.18/0.59 | 0.97 |
| S2 | 44.22 | 42.50 | 40.09 | 39.69 | 34.34 | 37.53 | 34.03 | 6.42/0.72 | 0.97 |
| S3 | 62.39 | 60.70 | 58.27 | 57.30 | 53.85 | 51.77 | 45.10 | 6.80/0.76 | 0.99 |
| S4 | 79.39 | 78.45 | 77.29 | 76.52 | 70.96 | 72.94 | 66.00 | 5.25/0.59 | 0.96 |
| S5 | 81.86 | 80.68 | 77.31 | 71.43 | 67.31 | 65.46 | 50.77 | 5.88/0.66 | 0.98 |
| S6 | 61.75 | 60.13 | 57.91 | 56.77 | 53.19 | 50.92 | 44.20 | 6.86/0.77 | 0.99 |
| S7 | 77.08 | 75.98 | 73.63 | 72.92 | 57.75 | 69.19 | 61.57 | 5.61/0.64 | 0.95 |
| S8 | 85.31 | 84.46 | 83.24 | 82.72 | 71.86 | 78.07 | 70.82 | 5.05/0.59 | 0.97 |

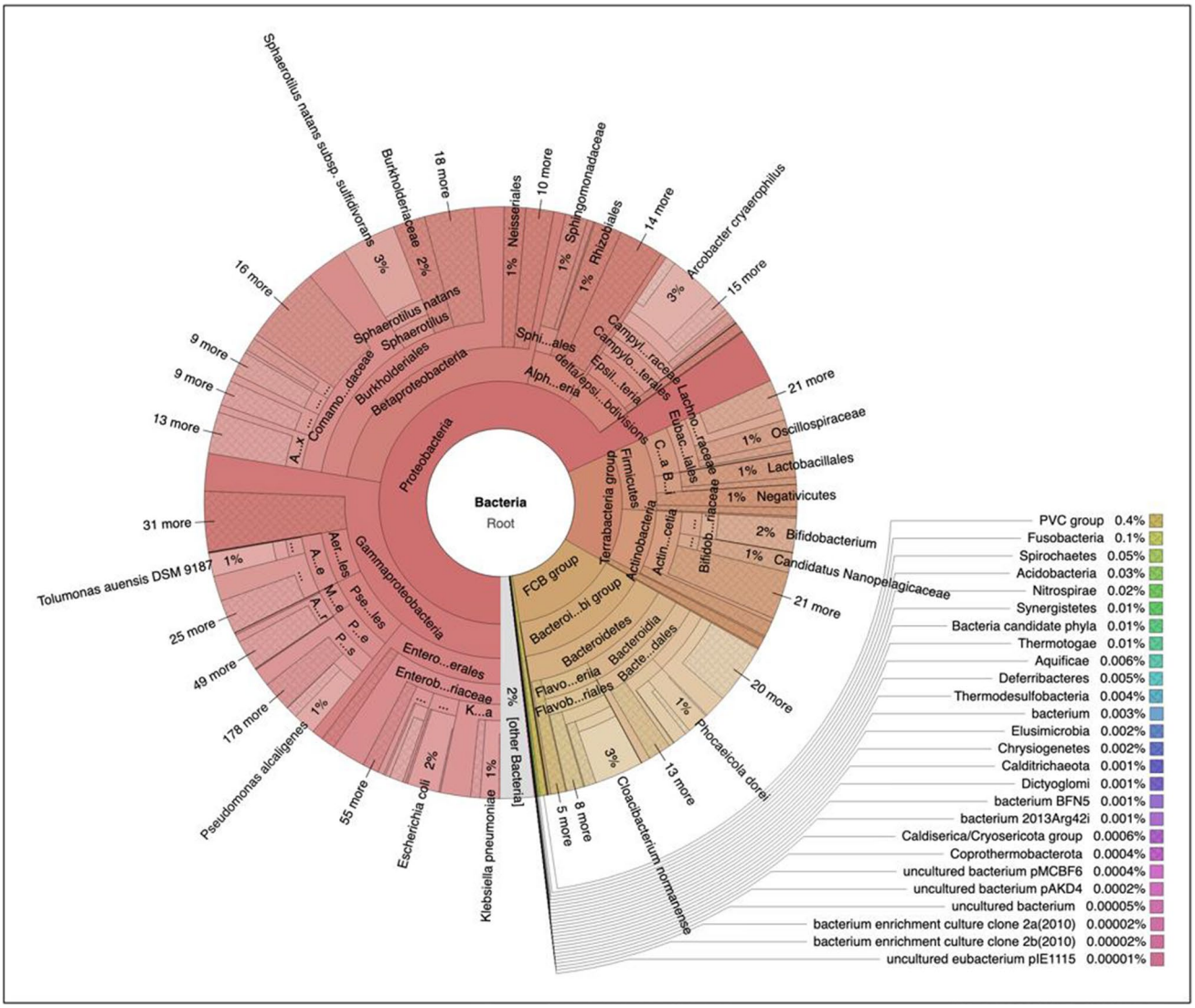


**Fig. 1** Krona bacterial distribution pie chart of all samples together

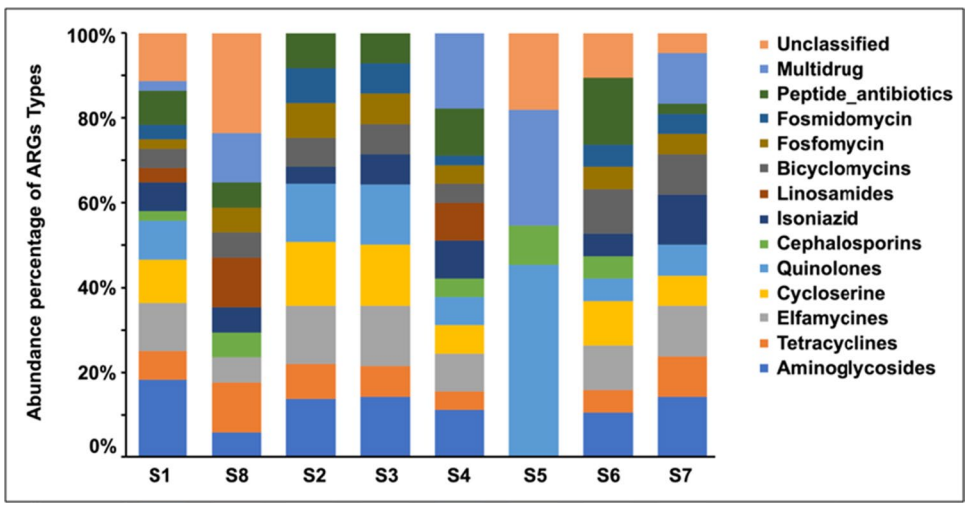


**Fig. 2** Abundance distribution of ARG types

of a bit lower in abundance ratio of ($3.56 \times 10^{-2}$) in S2. Further, in sample 3, ARG types occurred with a similar abundance ratio of ($3.24 \times 10^{-2}$), namely, *Aminoglycoside*, *Elfamycins*, and *Quinolones*. The lowest in the abundance of ARG types was *Lincosamides*, and *Cephalosporins* were a bit lower among all with a ratio of ($2.27 \times 10^{-2}$). The abundance clustered heatmap of ARGs with each sample was presented in Fig. 3, which shows the abundance of 14 ARG types detected was calculated based on summing the coverage of ARG subtypes belonging to the same ARG type. Abundance values were transformed using log ($x+1$), and clustering was based on Euclidean distances, which illustrate three distinct groups of antibiotic resistomes in the urban freshwater samples. This study found that the rise and reduction in the abundance of ARGs occurred between comparative urban freshwater samples, with *aminoglycoside* being the most dominant.

### ARG proportion assessment in plasmid and chromosomal sequence

Prior to the contigs process from the assembly, ARG types were classified for the chromosomes, plasmid, and unclassified sequences. Our results showed that chromosomal contigs had a substantially higher percentage of ARG types than plasmid and unclassified contigs with ARGs, according to ARG detection. In addition, unclassified encoded in plasmid and multidrug were proven to be the most dominant encoded in chromosomal contigs in terms of abundance of ARG subtypes belonging to each ARG. *Bicyclomycins* have ARGs encoded only on chromosomes and unclassified contigs, with a higher proportion on chromosome, while ARG counts on *Tetracyclines* were much dominant in plasmid contigs. Figure 4 shows the relative abundance of ARGs, where the genes conferring resistance to unclassified (25%) are the most prevalent in plasmids. Further, *multidrug* (17%), *Aminoglycoside* (10%), *Fosmidomycin* (7%), *Elfamycins* (6%), *Cycloserine* (6%), *Quinolones* (6%), *Isoniazid* (5%), *Bicyclomycins* (4%), and *Fosfomycin* (3%) being more frequently encoded in chromosomes. Moreover, our results showed that ARGs found in plasmids and chromosomes were in different proportions, and some ARG types were found on unclassified chromosomes.

### Categories of ARG resistance mechanism

In this study, the blast results were done on three major databases, namely, CARD, NDARO, and PATRIC AMR-related curation databases. In our results, the antibiotic resistance mechanism was classified into five major categories, namely, efflux pump, antibiotic inactivation, antibiotic target protection, antibiotic target replacement, and target modifying enzyme. In addition, Fig. 5 shows the unclassified categories of mechanisms as others. From our results, an antibiotic efflux pump is detected to be the dominant resistance mechanism in all samples, followed by antibiotic inactivation and the unclassified.

The percentage abundance of ARG antibiotic subtypes for all categories of the resistance mechanism was counted as (28% for S1, 1% for S2 and S3, 19% for S4, 33% for

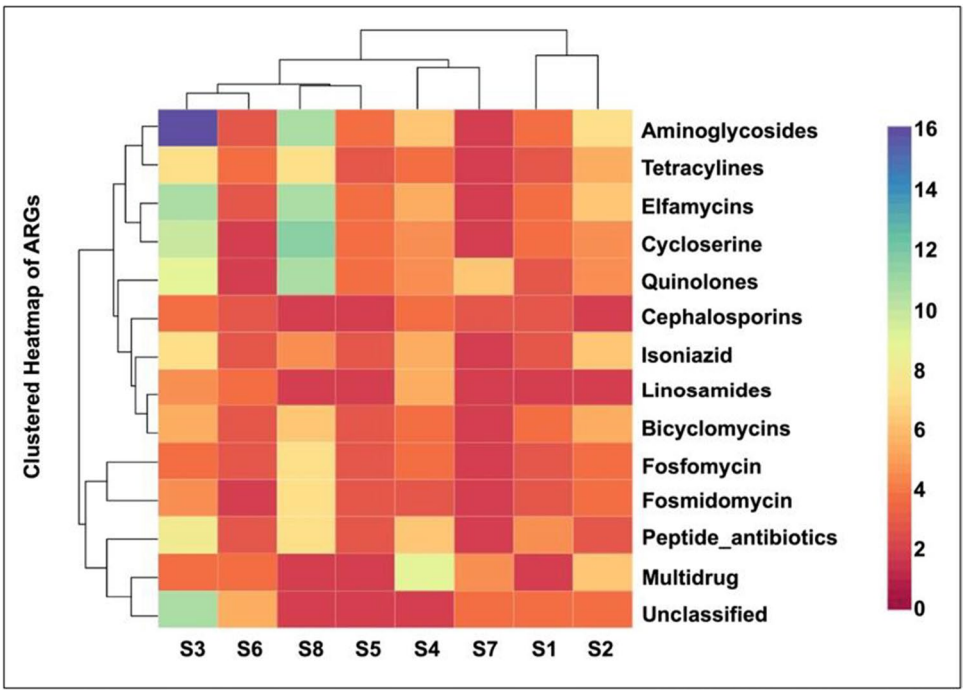


**Fig. 3** Clustered heatmap of ARG types

S5, 2% for S6, 3% for S7, and 14% for S8), respectively. It shows that S5 has more efflux pumps, antibiotic inactivation, and unidentified bacteria than any other. Genes associated with multiple drug resistance (MDR) are more prevalent in samples S1, S4, and S5 and are also included in the antibiotic efflux pump subclass. Further in antibiotic inactivation, the *beta-lactam* resistance gene was more prevalent in S4, S5, and S8 and *aminoglycoside* RGs in S1. Efflux pumps transport proteins in microorganisms that allow them to manage their internal environment by eliminating toxins and are frequently related to multiple drug resistance.

### Virulence gene assessment related to different factors

A total of 197 virulence genes were examined in this study after being compared to three major databases, namely, VFDB, vectors, and PATRIC_VF based on a high sequence similarity technique using BLAST. The demonstrated results in Fig. 6 show that sample S1 has the highest abundance of virulence genes compared to other samples, with genes in virulence and adherence being most dominant aside from the unclassified. Further, analysis showed that S8 had no virulence genes. *Escherichia coli* and *Salmonella enterica* have the most virulence genes in samples S1, S5, S6, and S7;

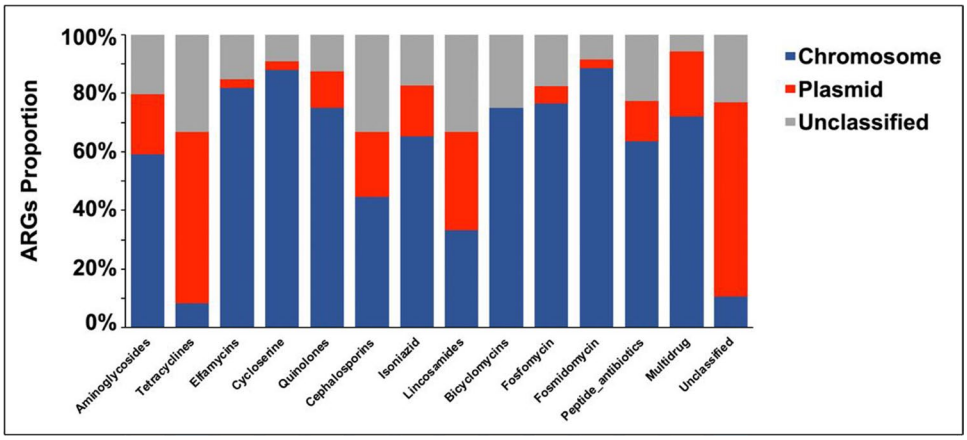


**Fig. 4** The proportion of ARG types located on plasmid, chromosomal, and unclassified sequences

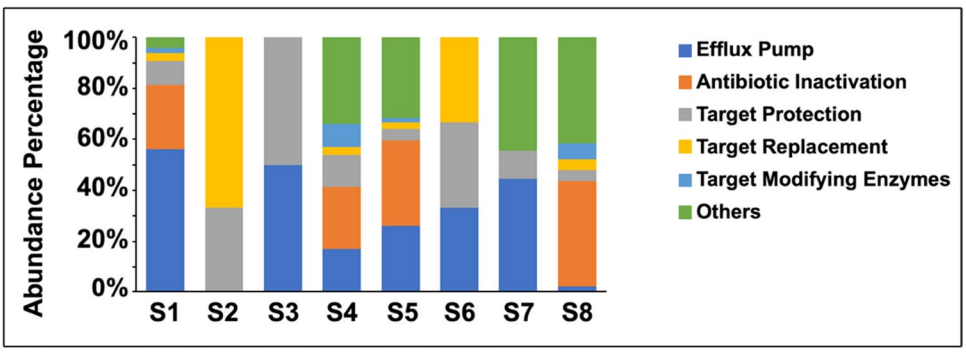


**Fig. 5** Percentages of ARG resistance mechanism with subtypes

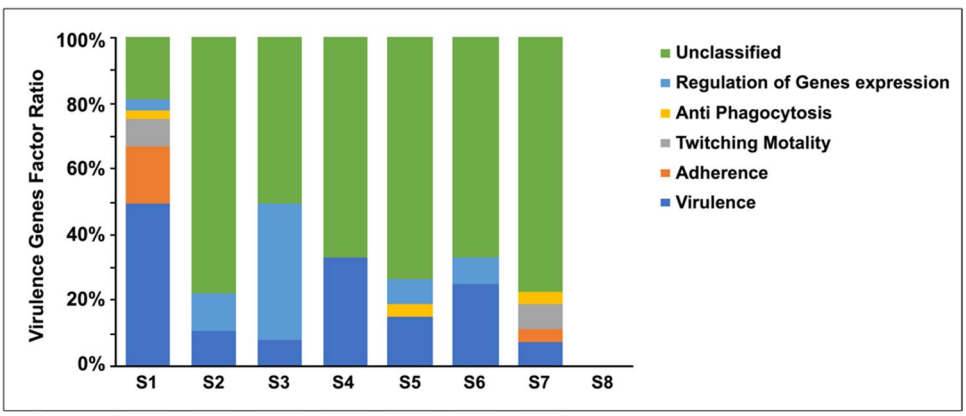


**Fig. 6** Number of virulence genes relative to different virulence factors

*Mycobacterium tuberculosis* in S2; and *Salmonella enterica* in S4. The common genes that were shared between S1, S4, S6, and S7 have the same virulence factors as *rfba* (*transketolase*) in *Salmonella enterica* (*Gram-negative bacteria*) and *tktA* (*glucose-1-phosphate thymidylyltransferase*) in *Escherichia coli* (*Gram-negative bacteria*). For adherence, S1 exclusively contained virulence genes from *Escherichia coli*, while sample S7 only contained virulence genes from *Pseudomonas aeruginosa* (*an encapsulated Gram-negative bacteria*) and no further shared genes between both samples. Moreover, in the regulation of gene expression virulence factor, the distributed genes in S1, S2, S3, S5, and S6 were found in *Escherichia coli*, *Shigella flexneri*, *Salmonella enterica*, and *Mycobacterium tuberculosis* with the dominant gene called *hfg* (*RNA-binding protein*).

**Table 3** Information about microbial subsystems and gene classification

| Classification | Subsystems | Genes |
|---|---|---|
| Metabolism | 156 | 8256 |
| Stress response, defense, and virulence | 61 | 1770 |
| Protein processing | 49 | 3690 |
| Energy | 43 | 3211 |
| Membrane transport | 42 | 1343 |
| Cellular processes | 26 | 1359 |
| DNA processing | 22 | 1234 |
| Cell envelop | 19 | 841 |
| RNA processing | 17 | 898 |
| Miscellaneous | 16 | 557 |
| Regulation and cell signaling | 7 | 192 |

### Expression level of MGEs and MRGs

In this study, four major types of MGEs and two metal resistance genes (MRGs) were discovered from the entire WGS assembly samples and the expression rate of each sample was then calculated as shown in Fig. 7. Transposase was the most highly expressed in S1, S4, S7, S5, S8, and S6, followed by transposon in S1, S4, S8, and S5 as compared to other MGEs. In each sample, MRGs were found to be present, with zinc being the most prevalent. S1, S4, S5, S7, and S8 are the only ones that contain ions, according to the results. In addition, S1 and S5 have higher concentrations of Ion.

### Microbial gene classifications and co-occurrence analysis

A subsystem is called a set of proteins that implements a structural biological complexity and subsystems specific to each genome which was included in the PATRIC annotation process (Davis et al. 2016). Table 3 depicts an overview of subsystem classification for all samples including genes. The classification assessment discovered a total of genes (8256 in metabolism), (1770 in virulence stress), (3690 in protein process), (3211 in energies), (1343 in membrane transport), (1359 in the cellular process), (1234 in DNA process), (841 in RNA process), (841 in Cell envelope), (898 in RNA processing), and (192 in cell signaling).

We investigated the co-occurrence patterns using a network analysis approach among four categories, namely, antibiotics and their RGs; mobile elements and their RGs, ARGs, and MGEs; and ARGs and microbial taxa as shown in Fig. 8a–d. Spearman pairs of ($p \leq 0.05$) and ($r \geq 0.4$) were utilized for co-occurrence analysis patterns. In addition, the bacterial genus was considered for possible ARG-type hosts based on co-occurrence results. We selectively studied the top 100 genes for the antibiotics and mobile elements based on their abundance and matched profiles respectively. As the fundamental focus of this study is human and environmental health, we have chosen the top homogenous bacteria and their highly associated species in all samples namely *Campylobacter*, *Escherichia*, *Staphylococcus*, *Streptomyces*, *Enterococcus*, *Bacillus*, *Leptospira*, *Listeria*, *Pseudomonas*, *Clostridium*, *Klebsiella*, and *Salmonella*. The nodes were sized and colored on the frequency ratios bases, and edges were colorized by weights

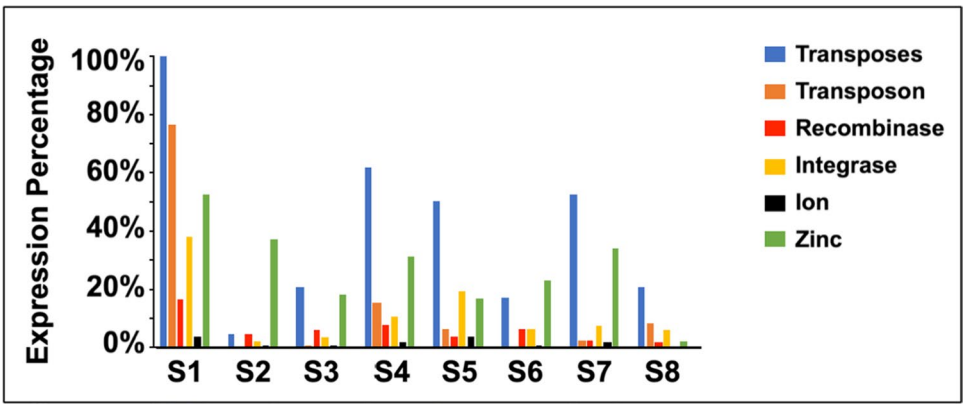


**Fig. 7** Expression level of MGEs and MRGs per assembly sample

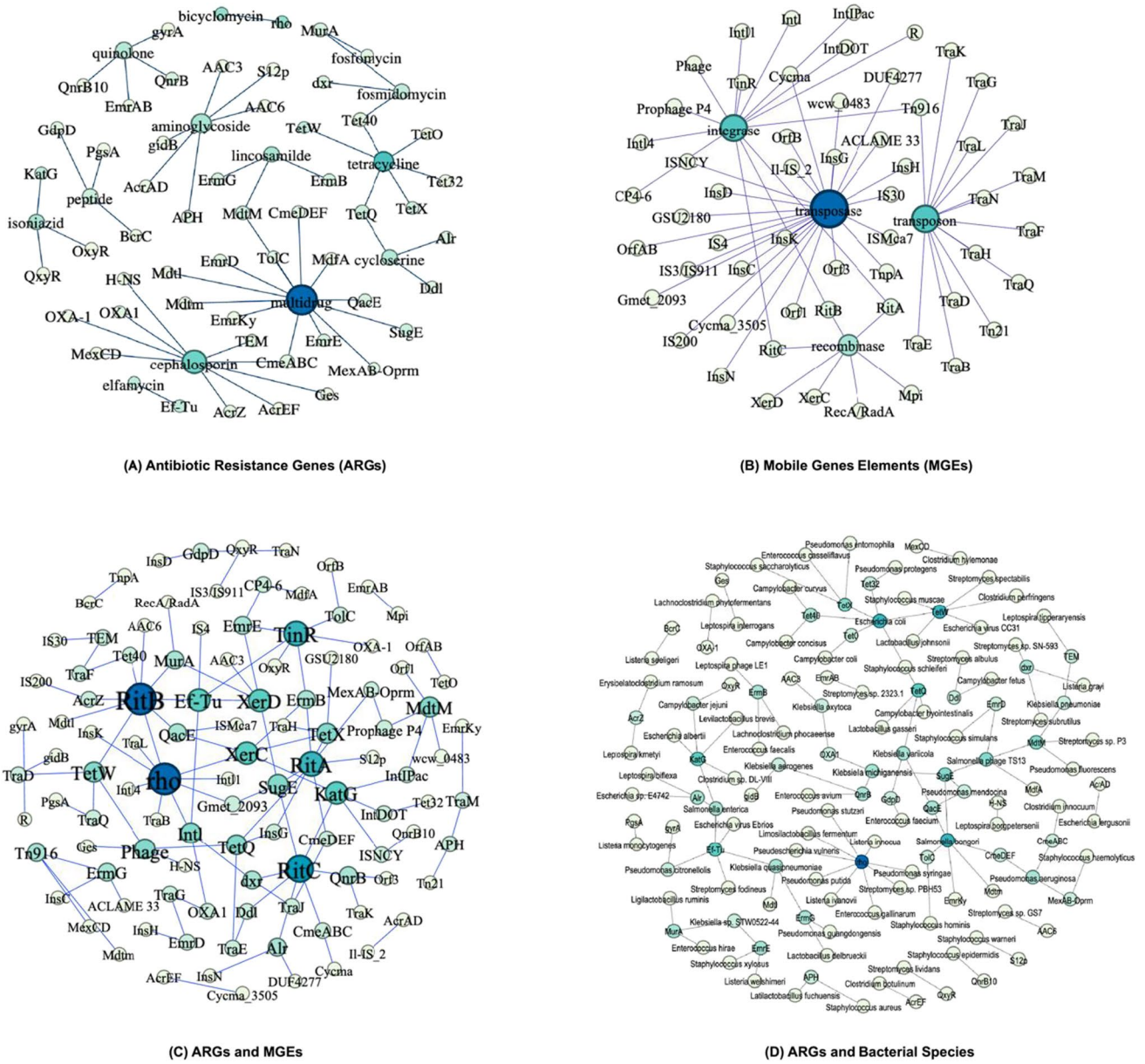


**Fig. 8** Co-occurrence network analysis pattern

that occurred in correlation analysis. Figure 8a clearly shows that *multidrug* is highly influenced and correlated with ARGs in environmental samples followed by *tetracycline*, *cephalosporin*, *aminoglycoside*, *fosmidomycin*, *lincosamildes*, *quinolone*, *isoniazid*, *fosfomycin*, *bicyclomycins*, *elfamycins*, and *peptide drugs*. In addition, four mobile elements were analyzed in Fig. 8b, namely, integrase, transposes, recombinase, and transposon. The figure clearly shows that transposes have a huge correlation frequency with MGEs in environmental samples followed by transposon, integrase, and recombinase. Further, ARGs and MGEs were analyzed in Fig. 8c and show *Rit* (*A*, *B*, *C*), *Rho*, *Tin*, *Xer* (*C*, *D*), *Kat*, *Phage*, *Tet* (*M*, *Q*, *W*, *X*), *EF-TU*, *SugE*, *QnrB*, *MdtM*, and *Int* are highly correlated genes followed by *TEM*, *ACR*, *Tn*, *Gmet*, *Tra*, *Emr* (*D*, *G*), *OXA*, *dxr*, *Ddl*, *Alr*, *Erm*, *Emr*, *Mur*, *Qac*, and *Gdp*. Moreover, Fig. 8d shows that homogenous microbial bacteria, namely, *Campylobacter*, *Escherichia*, *Staphylococcus*, *Streptomyces*, *Enterococcus*, *Bacillus*, *Leptospira*, *Listeria*, *Pseudomonas*, *Clostridium*, *Klebsiella*, and *Salmonella* are correlated with ARGs (*rho*, *Ef-Tu*, *Mur*, *Emr*, *APH*, *Erm*, *Alr*, *Kat*, *Acr*, *Tet*, *Bcr*, *Oxy*, *TEM*, *dxr*, *Qac*, *Qnr*, *Tol*, *AAC*, *gid*, *Mdt*, and *Cme*) in environmental freshwater samples. The raw formation of co-occurrence network analysis data is present in the supplementary file (section-V) with their weights and ratios.

## Discussion

WTPs are known as reservoirs of ARGs worldwide, and antibiotic virulence is a hot topic in infectious disease research (Sohail et al. 2018; Uba Muhammad et al. 2018). Literature is available to support in silico studies for the identification of virulence factors caused by multidrug-resistant pathogenic agents (Farman et al. 2019). Moreover, it demonstrated that ARGs and MGEs are commonly found in the water resources of remote areas as well, so it is time to go for the extensive metagenomics analysis of different samples from different regions around the world to reduce the significant issue of antibiotic resistance (Szekeres et al. 2018). This study was conducted to investigate and identify bacterial contamination, antibiotic resistance, MGEs, virulence factors, and the resistance mechanism in urban freshwater samples. It is noted that WTPs are now hotspot research topics for the identification of ARGs and ARBs.

The percentage of bacterial reads categorized at each level of taxonomic classification was studied, and the co-occurrence network explained the environmental homogeneous bacterial categorization of three important levels: Family, Genus, and Species. To explain the taxonomic abundance, a bar chart was generated in the supplementary file (section-II, III). Samples were subjected to beta diversity to explain the distribution patterns of two communities in terms of diversity as shown in the supplementary file (section-IV). The ARG abundances in each sample are demonstrated in Fig. 2. Results depicted that the profiles of major ARGs that confer *aminoglycosides*, *tetracyclines*, *elfamycins*, *cycloserine*, *quinolones*, *cephalosporins*, *isoniazid*, *lincosamildes*, *bicyclomycins*, *fosfomycin*, *fosmidomycin*, *multidrug*, and *peptide drugs* resistance were abundant in Singapore urban freshwater. These antibiotics are among the most commonly prescribed drugs in Singapore and are among the major pharmaceutical products found in the WTPs (Goh et al. 2020). In addition, ARGs to *tetracyclines*, *cycloserine*, *aminoglycosides*, and *multidrug* are frequently detected in WTPs worldwide, along with those conferring resistance to *fosmidomycin* and *tetracyclines*. Because of the strong sorption of the *tetracycline* and *fosmidomycin* antibiotics, their mobility in the environment may be facilitated by transport with freshwater (Anh et al. 2021).

The relative abundance of ARGs for each sample was calculated and shows the significant genes conferring resistance to unclassified as (25%) which found as the most prevalent in plasmids. Further, *multidrug* (17%), *Aminoglycoside* (10%), *Fosmidomycin* (7%), *Elfamycins* (6%), *Cycloserine* (6%), *Quinolones* (6%), *Isoniazid* (5%), *Bicyclomycins* (4%), and *Fosfomycin* (3%) being more frequently encoded in chromosomes. Moreover, our results show that ARGs found in plasmids and chromosomes were in different proportions, and some ARG types were found on unclassified chromosomes as in Fig. 4. In bacteria, the plasmid and chromosome differentiate by the circular representation of double-stranded extra-chromosomal DNA structure. Relatively high multidrug resistance in aerobic processes has also been reported and potentially explained by the presence of many microstressors in freshwater, which select for bacteria with multiple defense mechanisms and dissemination of their resistance through HGT. *Multidrug* with high solubility and chemical stability can persist in the environment for a long period of time, resulting in a high abundance of ARGs in the WTPs (Akhil et al. 2021). Therefore, *tetracycline* and *multidrug* resistances are frequently detected in effluent and surface water, indicating a general resilience to the WTP which has a major impact on freshwater after treatments.

In our results, the antibiotic resistance mechanism was classified into five major categories, namely, efflux pump, antibiotic inactivation, antibiotic target protection, antibiotic target replacement, and target modifying enzyme (Fig. 5). This mechanism is associated with the disruption of a significant process of bacterial cells that are actually playing a role in bacterial division and growth of bacteria. The bacteria are actually associated with the virulence factors; therefore, the sample S1 has the highest abundance of virulence genes compared to other samples, with genes in virulence and adherence being most dominant aside from the unclassified samples (Fig. 6). Transposons are associated with antibiotic resistance mechanisms and are helpful in the transfer mechanism of bacterial genes. Figure 7 demonstrates transposase was the most highly expressed in S1, S4, S7, S5, S8, and S6, followed by transposon in S1, S4, S8, and S5 as compared to other MGEs. In each sample, MRGs were found to be present, with zinc being the most prevalent. S1, S4, S5, S7, and S8 are the only ones that contain Ions, according to the results. In addition, S1 and S5 have higher concentrations of Ion. The zoonotic bacteria and the pathogens have been widely studied in the selected sample and it is noticed that multiple bacteria such as *Campylobacter*, *Escherichia*, *Staphylococcus*, *Streptomyces*, *Enterococcus*, *Bacillus*, *Leptospira*, *Listeria*, *Pseudomonas*, *Clostridium*, *Klebsiella*, and *Salmonella* are highly correlated with ARGs (*rho*, *Ef-Tu*, *Mur*, *Emr*, *APH*, *Erm*, *Alr*, *Kat*, *Acr*, *Tet*, *Bcr*, *Oxy*, *TEM*, *dxr*, *Qac*, *Qnr*, *Tol*, *AAC*, *gid*, *Mdt*, and *Cme*) in the selected environmental freshwater samples (Fig. 8).

According to previous studies (Kori et al. 2019; Yoo et al. 2020; Zhang et al. 2020; R. Zhao et al. 2020; Ahmad et al. 2021; Mukherjee et al. 2021), there is growing concern about ARGs and MGEs in the WTP due to the prospect that anaerobic digestors could serve as a new ARB source. This point is worth mentioning because this study used samples

from Singapore's freshwater treatment plant. In addition, previous studies have highlighted that fresh drinking water contain bacteria that pose a threat to human health. Furthermore, this study focused on the homogenous bacteria like *campylobacter*, *Escherichia*, *Staphylococcus*, *Streptomyces*, *Enterococcus*, *Bacillus*, *Leptospira*, *Listeria*, *Pseudomonas*, *Clostridium*, *Klebsiella*, and *Salmonella*. The top homogenous bacteria list was published by the World Health Organization (WHO), and can be found at (https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed). From our perspective, freshwater is a significant reservoir of ARBs, and water treatment methods may be employed in the future to minimize the amount of this reservoir of resistance. The primary focus of this study was to understand the bacterial community structure and functional characteristics of ARGs and MGEs associated with Singapore's freshwater treatment plant. However, the biotechnological applications of novel functional traits were also discovered as a result of this metagenomic study. This study demonstrated a comprehensive in silico analysis of ARGs and their relative abundance with respect to diversity involving bacterial communities in urban freshwater. The spread and awareness of these antibiotic-related diseases are very important. Hence, our study will add knowledge to antibiotic research and could be helpful for future understanding of antibacterial resistance and its management in freshwater sediments.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s11356-022-22197-4.

**Acknowledgements** We would like to thank our colleagues who contributed in the study.

**Author contribution** Conceptualizing, methodology, empirical estimations, writing, and drafting of the original draft by Sohail M. Noman; supervision by Xiaoyang Jiao; data collection by Sohail M. Noman, Muhammad Shafiq, Yumeng Yuan, Mi Zeng, and Xin Li; review and editing by Sohail M. Noman, Shabana Bibi, Bharti Mittal, Oluwaseyi Abraham Olawale, and Muhammad Irshad. All authors have read and approved the final manuscript.

**Funding** This research work has been supported by Natural Science Foundation of China (NSFC) for international young scientist under grant No. 42150410383.

**Data availability** All raw data used in this study can be found in online repositories. The names of the repository and accession number(s) can be found as: (https://www.ncbi.nlm.nih.gov/), PRJNA400857.

## Declarations

**Ethical approval** This article does not contain any studies concerned with the experiment on human or animals.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

## References

Ahmad T, Gupta G, Sharma A, Kaur B, El-Sheikh MA, Alyemeni MN (2021) Metagenomic analysis exploring taxonomic and functional diversity of bacterial communities of a Himalayan urban fresh water lake. PLoS ONE 16(3 March), 1–19. https://doi.org/10.1371/journal.pone.0248116

Akhil D, Lakshmi D, Senthil Kumar P, Vo DVN, Kartik A (2021) Occurrence and removal of antibiotics from industrial wastewater. Environ Chem Lett 19(2):1477–1507. https://doi.org/10.1007/S10311-020-01152-0

Alexander J, Hembach N, Schwartz T (2020) Evaluation of antibiotic resistance dissemination by wastewater treatment plant effluents with different catchment areas in Germany. Scientific Reports 2020 10:1, 10(1), 1–9. https://doi.org/10.1038/s41598-020-65635-4

Anh HQ, Le TPQ, Da Le N, Lu XX, Duong TT, Garnier J, Rochelle-Newall E, Zhang S, Oh NH, Oeurng C, Ekkawatpanit C, Nguyen TD, Nguyen QT, Nguyen TD, Nguyen TN, Tran TL, Kunisue T, Tanoue R, Takahashi S, … Nguyen TAH (2021) Antibiotics in surface water of East and Southeast Asian countries: a focused review on contamination status, pollution sources, potential risks, and future perspectives Sci Total Environ 764(1):142873. https://doi.org/10.1016/J.SCITOTENV.2020.142865

Antipov D, Raiko M, Lapidus A, Pevzner PA (2020) MetaviralSPAdes: assembly of viruses from metagenomic data. Bioinformatics 36(14):4126–4129. https://doi.org/10.1093/BIOINFORMATICS/BTAA490

Arenas F, Navarro-Ródenas A, Marqués-Gálvez JE, Ghignone S, Mello A, Morte A (2021) Different patterns in root and soil fungal diversity drive plant productivity of the desert truffle Terfezia claveryi in plantation. Environ Microbiol 23(10):5917–5933. https://doi.org/10.1111/1462-2920.15688

Baniya A, Digennaro P (2021) Genome announcement of Steinernema khuongi and its associated symbiont from Florida. *G3 Genes|Genomes|Genetics*, *11*(4), 1–16. https://doi.org/10.1093/G3JOURNAL/JKAB053

Barancheshme F, Munir M (2018) Strategies to combat antibiotic resistance in the wastewater treatment plants. Front Microbiol 8(JAN):2615. https://doi.org/10.3389/FMICB.2017.02603/BIBTEX

Bibi S, Hasan MM, Wang Y-B, Papadakos SP, Yu H (2021) Cordycepin as a promising inhibitor of SARS-CoV-2 RNA dependent RNA polymerase (RdRp). Curr Med Chem 28(1):33–41. https://doi.org/10.2174/0929867328666210820114025

Bibi S, Wang Y-B, Tang D-X, Kamal MA, Yu H (2019) Prospects for discovering the secondary metabolites of Cordyceps sensu lato by the integrated strategy. Med Chem 17(2):97–120. https://doi.org/10.2174/1573406416666191227120425

Biswas P, Hasan MM, Dey D, dos Santos Costa AC, Polash SA, Bibi S, Ferdous N, Kaium MA, Rahman MH, Jeet FK, Papadakos S, Islam K, Uddin MS (2021) Candidate antiviral drugs for COVID-19 and their environmental implications: a comprehensive analysis. Environ Sci Pollut Res 2021 28:42, 28(42), 59570–59593. https://doi.org/10.1007/S11356-021-16096-3

Brown J, Pirrung M, Mccue LA (2017) FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 33(19):3137–3139. https://doi.org/10.1093/BIOINFORMATICS/BTX373

Che Y, Xia Y, Liu L, Li AD, Yang Y, Zhang T (2019) Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome 7(1):1–13. https://doi.org/10.1186/S40168-019-0663-0/FIGURES/5

Chopyk J, Nasko DJ, Allard S, Callahan MT, Bui A, Ferelli AMC, Chattopadhyay S, Mongodin EF, Pop M, Micallef SA, Sapkota AR (2020) Metagenomic analysis of bacterial and viral assemblages from a freshwater creek and irrigated field reveals temporal and spatial dynamics. Sci Total Environ 706(1):135409. https://doi.org/10.1016/J.SCITOTENV.2019.135395

Danko D, Bezdan D, Afshin EE, Ahsanuddin S, Bhattacharya C, Butler DJ, Chng KR, Donnellan D, Hecht J, Jackson K, Kuchin K, Karasikov M, Lyons A, Mak L, Meleshko D, Mustafa H, Mutai B, Neches RY, Ng A, … Zubenko S (2021) A global metagenomic map of urban microbiomes and antimicrobial resistance. Cell 184(13):3376-3393.e17https://doi.org/10.1016/J.CELL.2021.05.002

Dávalos AF, Garcia PK, Montoya-Pachongo C, Rengifo A, Guerrero D, Díaz-Ordoñez L, Díaz G, Ferro BE (2021) Identification of nontuberculous mycobacteria in drinking water in Cali, Colombia. Int J Environ Res Public Health *2021, Vol. 18, Page 8451*, *18*(16), 8462. https://doi.org/10.3390/IJERPH18168451

Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, Overbeek R, Santerre J, Shukla M, Wattam AR, Will R, Xia F, Stevens R (2016) Antimicrobial resistance prediction in PATRIC and RAST. Sci Rep 6(1):1–12. https://doi.org/10.1038/srep27930

Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, MacHi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, … Stevens R (2020) The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res 48(D1):D606-D612https://doi.org/10.1093/NAR/GKZ943

Fang H, Huang K, Yu J, Ding C, Wang Z, Zhao C, Yuan H, Wang Z, Wang S, Hu J, Cui Y (2019) Metagenomic analysis of bacterial communities and antibiotic resistance genes in the Eriocheir sinensis freshwater aquaculture environment. Chemosphere 224(1):202–211. https://doi.org/10.1016/J.CHEMOSPHERE.2019.02.068

Fang H, Cai L, Yang Y, Ju F, Li X, Yu Y, Zhang T (2014) Metagenomic analysis reveals potential biodegradation pathways of persistent pesticides in freshwater and marine sediments. Sci Total Environ 470–471:983–992. https://doi.org/10.1016/J.SCITOTENV.2013.10.076

Farman M, Yasir M, Al-Hindi RR, Farraj SA, Jiman-Fatani AA, Alawi M, Azhar EI (2019) Genomic analysis of multidrug-resistant clinical Enterococcus faecalis isolates for antimicrobial resistance genes and virulence factors from the western region of Saudi Arabia. Antimicrob Resist Infect Control 8(1):1–11. https://doi.org/10.1186/S13756-019-0508-4/FIGURES/5

Frank H (2021) Harrell Miscellaneous Package 'Hmisc.' https://hbiostat.org/R/Hmisc/

Gabashvili E, Kobakhidze S, Chkhikvishvili T, Tabatadze L, Tsiklauri R, Dadiani K, Koulouris S, Kotetishvili M (2022) Metagenomic and recombination analyses of antimicrobial resistance genes from recreational waters of Black Sea coastal areas and other marine environments unveil extensive evidence for their both intrageneric and intergeneric transmission across Genetical. Mar Genomics 61(1):100931. https://doi.org/10.1016/J.MARGEN.2021.100916

Goh JL, Jaffar JLY, Xu H, Lum VWM (2020) A retrospective study on end-of-life care in the emergency department of a tertiary hospital in Singapore. Proceedings of Singapore Healthcare 30(2):138–144. https://doi.org/10.1177/2010105820960175

Grant KR (2022) Next-Generation Amplicon Sequencing: A cost-effective method for exploring microbial biodiversity. In *Molecular genetics and genomics tools in biodiversity conservation* (1st ed., Vol. 1, pp. 203–236). Springer, Singapore. https://doi.org/10.1007/978-981-16-6005-4_10

Kori JA, Mahar RB, Vistro MR, Tariq H, Khan IA, Goel R (2019) Metagenomic analysis of drinking water samples collected from treatment plants of Hyderabad City and Mehran University Employees Cooperative Housing Society. Environ Sci Pollut Res 26(28):29052–29064. https://doi.org/10.1007/s11356-019-05859-8

Kumar Awasthi, M., Ravindran, B., Sarsaiya, S., Chen, H., Wainaina, S., Singh, E., Liu, T., Kumar, S., Pandey, A., Singh, L., & Zhang, Z. (2020). Metagenomics for taxonomy profiling: tools and approaches. 11(1):356–374. https://doi.org/10.1080/21655979.2020.1736238

Liu C, Cui Y, Li X, Yao M (2021) microeco: an R package for data mining in microbial community ecology. FEMS Microbiol Ecol 97(2):1–10. https://doi.org/10.1093/FEMSEC/FIAA255

Mammola S, Carmona CP, Guillerme T, Cardoso P (2021) Concepts and applications in functional diversity. Funct Ecol 35(9):1869–1885. https://doi.org/10.1111/1365-2435.13882/SUPPINFO

Meneghine AK, Nielsen S, Varani AM, Thomas T, Alves LMC (2017) Metagenomic analysis of soil and freshwater from zoo agricultural area with organic fertilization. PLoS ONE 12(12):e0190189. https://doi.org/10.1371/JOURNAL.PONE.0190178

Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, Salazar GA, Pesseat S, Boland MA, Hunter FMI, Ten Hoopen P, Alako B, Amid C, Wilkinson DJ, Curtis TP, Cochrane G, Finn RD (2018) EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. Nucleic Acids Res 46(1):D726–D735. https://doi.org/10.1093/nar/gkx967

Mohiuddin M, Schellhorn HE (2015) Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. Front Microbiol 6(SEP):971. https://doi.org/10.3389/FMICB.2015.00960/BIBTEX

Mreyoud Y, Song M, Lim J, Ahn TH (2022) MegaD: deep learning for rapid and accurate disease status prediction of metagenomic samples. Life 2022, Vol. 12, Page 669, 12(5):680. https://doi.org/10.3390/LIFE12050669

Mukherjee M, Laird E, Gentry TJ, Brooks JP, Karthikeyan R (2021) Increased antimicrobial and multidrug resistance downstream of wastewater treatment plants in an urban watershed. Front Microbiol 12(1):1278. https://doi.org/10.3389/FMICB.2021.657353/BIBTEX

Palermo CN, Fulthorpe RR, Saati R, Short SM (2019) Metagenomic analysis of virus diversity and relative abundance in a eutrophic freshwater harbour. Viruses 2019, Vol. 11, Page 792, 11(9):802. https://doi.org/10.3390/V11090792

Parrello B, Butler R, Chlenski P, Pusch GD, Overbeek R (2021) Supervised extraction of near-complete genomes from metagenomic samples: A new service in PATRIC. PLoS ONE 16(4):250092–250102. https://doi.org/10.1371/JOURNAL.PONE.0250092

Rasche H, Hiltemann S (2020) Galactic Circos: user-friendly Circos plots within the Galaxy platform. GigaScience 9(6):1–6. https://doi.org/10.1093/GIGASCIENCE/GIAA065

Revelle, W. (2021, September 22). *PSYCH: Procedures for Psychological, Psychometric, and Personality Research*. Cran R Projects; Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=psych

Sahani S, Hansa, Sharma YC, Kim TY (2022) Emerging contaminants in wastewater and surface water. New Trends in Emerging Environment Contaminants 9–30https://doi.org/10.1007/978-981-16-8367-1_2/COVER/

SaparbaevnaAlexyuk M, SabirzhanovnaTurmagambetova A, GennadievichAlexyuk P, PavlinovichBogoyavlenskiy A, EleazarovichBerezin V (2017) Comparative study of viromes from freshwater samples of the Ile-Balkhash region of Kazakhstan captured

through metagenomic analysis. VirusDisease 28(1):18–25. https://doi.org/10.1007/s13337-016-0353-5

Shilpa B, N., & Meena, S. S. (2022) Exploring the plastic degrading ability of microbial communities through metagenomic approach. Materials Today: Proceedings 57:1924–1932. https://doi.org/10.1016/J.MATPR.2022.02.308

Sohail MN, Jiadong R, Uba MM, Irshad M, Iqbal W, Arshad J, John AV (2019a) A hybrid Forecast Cost Benefit Classification of diabetes mellitus prevalence based on epidemiological study on Real-life patient's data. Sci Rep 9(1):1–10. https://doi.org/10.1038/s41598-019-46631-9

Sohail MN, Ren J, Muhammad MU (2019b) A euclidean group assessment on semi-supervised clustering for healthcare clinical implications based on real-life data. Int J Environ Res Public Health 16(9):1581–1593. https://doi.org/10.3390/ijerph16091581

Sohail N, Jiadong R, Musa Uba M, Irshad M, Khan A (2018) Classification and cost benefit analysis of diabetes mellitus dominance. Int J Comput Sci Netw Secur 18(10):29–35. http://paper.ijcsns.org/07_book/201810/20181006.pdf

Szekeres E, Chiriac CM, Baricz A, Szőke-Nagy T, Lung I, Soran ML, Rudi K, Dragos N, Coman C (2018) Investigating antibiotics, antibiotic resistance genes, and microbial contaminants in groundwater in relation to the proximity of urban areas. Environ Pollut 236(1):734–744. https://doi.org/10.1016/J.ENVPOL.2018.01.107

Taş N, de Jong AE, Li Y, Trubl G, Xue Y, Dove NC (2021) Metagenomic tools in microbial ecology research. Curr Opin Biotechnol 67(1):184–191. https://doi.org/10.1016/J.COPBIO.2021.01.019

Uba Muhammad M, Jiadong R, Sohail MN, Irshad M, Bilal M, Osi AA (2018) A logistic regression modeling on the prevalence of diabetes mellitus in the North Western Part of Nigeria. Benin J St 1(1):1–10. https://srg-uniben.org/pdf/1_bjs.pdf

UNEP (2021) Progress on SDG 6 Indicators | UNEP - UN Environment Programme. UN Environment Programme. https://www.unep.org/resources/report/progress-sdg-6-indicators

Vieira FR, Pecchia JA (2021) Bacterial community patterns in the Agaricus bisporus cultivation system, from compost raw materials to mushroom caps. Microb Ecol 1(1):1–13. https://doi.org/10.1007/S00248-021-01833-5/FIGURES/5

White RA, Chan AM, Gavelis GS, Leander BS, Brady AL, Slater GF, Lim DSS, Suttle CA (2016) Metagenomic analysis suggests modern freshwater microbialites harbor a distinct core microbial community. Front Microbiol 6(JAN):1543. https://doi.org/10.3389/FMICB.2015.01531/BIBTEX

Wu F, Speth DR, Philosof A, Crémière A, Narayanan A, Barco RA, Connon SA, Amend JP, Antoshechkin IA, Orphan VJ (2022) Unique mobile elements and scalable gene flow at the prokaryote–eukaryote boundary revealed by circularized Asgard archaea genomes. Nat Microbiol 2022:1–13. https://doi.org/10.1038/s41564-021-01039-y

Yadav S, Kapley A (2021) Antibiotic resistance: global health crisis and metagenomics. Biotechnology Reports 29(1):604–612. https://doi.org/10.1016/J.BTRE.2021.E00604

Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, Bian ZX, Zhang L (2021) A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. Genomics 1(1):1–8. https://arxiv.org/abs/2109.00809v1

Yoo K, Yoo H, Lee J, Choi EJ, Park J (2019) Exploring the antibiotic resistome in activated sludge and anaerobic digestion sludge in an urban wastewater treatment plant via metagenomic analysis. J Microbiol 2020 58(2):123–130. https://doi.org/10.1007/S12275-020-9309-Y

Yoo K, Yoo H, Lee J, Choi EJ, Park J (2020) Exploring the antibiotic resistome in activated sludge and anaerobic digestion sludge in an urban wastewater treatment plant via metagenomic analysis. J Microbiol 58(2):123–130. https://doi.org/10.1007/s12275-020-9309-y

Yousafi Q, Amin H, Bibi S, Rafi R, Khan MS, Ali H, Masroor A (2021) Subtractive proteomics and immuno-informatics approaches for multi-peptide vaccine prediction against Klebsiella oxytoca and validation through in silico expression. Int J Pept Res Ther 27(4):2685–2701. https://doi.org/10.1007/S10989-021-10283-Z/FIGURES/8

Yuan Y, Chen Y, Yao F, Zeng M, Xie Q, Shafiq M, Noman SM, Jiao X (2021) Microbiomes and resistomes in biopsy tissue and intestinal lavage fluid of colorectal cancer. Frontiers in Cell and Developmental Biology 9(1):737008. https://doi.org/10.3389/FCELL.2021.736994

Zhang L, Zhong M, Li X, Lu W, Li J (2020) River bacterial community structure and co-occurrence patterns under the influence of different domestic sewage types. J Environ Manage 266(February):110590. https://doi.org/10.1016/j.jenvman.2020.110590

Zhao R, Yu K, Zhang J, Zhang G, Huang J, Ma L, Deng C, Li X, Li B (2020) Deciphering the mobility and bacterial hosts of antibiotic resistance genes under antibiotic selection pressure by metagenomic assembly and binning approaches. Water Res 186(1):33–41. https://doi.org/10.1016/j.watres.2020.116318

Zhao Y, Zhang XX, Zhao Z, Duan C, Chen H, Wang M, Ren H, Yin Y, Ye L (2018) Metagenomic analysis revealed the prevalence of antibiotic resistance genes in the gut and living environment of freshwater shrimp. J Hazard Mater 350:10-18. https://doi.org/10.1016/J.JHAZMAT.2018.02.004

## Authors and Affiliations

**Sohail M. Noman[1] · Muhammad Shafiq[1] · Shabana Bibi[2,3] · Bharti Mittal[4] · Yumeng Yuan[1] · Mi Zeng[1] · Xin Li[1] · Oluwaseyi Abraham Olawale[5] · Xiaoyang Jiao[1] · Muhammad Irshad[6]**

Sohail M. Noman
mn.sohail@stumail.ysu.edu.cn

Muhammad Shafiq
drshafiqnjau@yahoo.com

Shabana Bibi
shabana_bibi@ynu.edu.cn

Bharti Mittal
counseler4genetics@gmail.com

Yumeng Yuan
19ymyuan@stu.edu.cn

Mi Zeng
20mzeng1@stu.edu.cn

Xin Li
21xli1@stu.edu.cn

Oluwaseyi Abraham Olawale
oluseyi88@gmail.com

Muhammad Irshad
mirsha@polyu.edu.hk

[1] Department of Cell Biology and Genetics, Shantou University Medical College, Shantou 515041, Guangdong, China

[2] Department of Biosciences, Shifa Tameer-E-Milat University, Islamabad 43600, Pakistan

[3] Department of Herbal Laboratory, College of Ecology and Environmental Sciences, Yunnan University, Yunnan, Kunming 650091, China

[4] Department of Science Education and Research, Nitte University, Deralakatte, Mangalore 575018, India

[5] Genomac Hub, Ogbomoso 210271, Oyo State, Nigeria

[6] Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom 0000, Hong Kong