Supplementary information

For

Systematic citywide analysis reveals ecological connectivity of antimicrobial resistance genes across urban water systems

Huafeng Liu[1], Yuting Qi[1], Xingxing Zhang[1], Tong Zhang[2], Xu-Xiang Zhang[3,] *, Liping Ma[1,] *

Affiliations:

[1] School of Ecological and Environmental Sciences, East China Normal University, Shanghai, 200241, China

[2] Environmental Microbiome Engineering and Biotechnology Laboratory, The University of Hong Kong, Hong Kong SAR, 999077, China

[3] State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing, 210023, China

*Corresponding authors:

Liping Ma

School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China. Email: lpma@des.ecnu.edu.cn

Xu-Xiang Zhang

State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing, China. Email: zhangxx@nju.edu.cn

Table of Contents

levels (resistome, subtype, sequence) in the urban water system.

Supplementary Fig. 8 Profile of microbial communities across urban water system.

Supplementary Fig. 9 Identification of primary genera and potential ARG hosts across urban water system.

Supplementary Fig. 10 Reconstruction and characterization of metagenome-assembled genomes (MAGs) across urban water system.

Supplementary Fig. 11 An overview of a complete genome (100% completeness, 0% contamination).

Supplementary Fig. 12 Overview of the genetic bacterial hosts of ARGs in urban water system.

Supplementary Fig. 13 ARG–MGE co-localization and scaffold recruitment patterns of *Pseudomonas aeruginosa* (bin5) across tap water samples.

Supplementary Fig. 14 Correlation between assembly contiguity and detected HGT events.

Supplementary Fig. 15 Phylum-level summary of inferred donor-recipient relationships in ARG-associated horizontal gene transfer.

Supplementary Fig. 16 Quantitative analysis of *int*I1 based on metagenomics.

Supplementary Fig. 17 Characterization of ARG-carrying integrons in LFE strains.

Supplementary Fig. 18 (Opportunistic) pathogens detected in drinking water.

Supplementary Fig. 19 Ecological assembly mechanisms of ARGs and their host communities.

Abbreviations used in this study

UWS: urban water system

ARG: antibiotic resistance gene

ARB: antibiotic-resistant bacteria

WWTP: wastewater treatment plant

HGT: horizontal gene transfer

INF: influent

AS: activated sludge

EFF: effluent

RU: upstream of receiving water

RS: outlet site of receiving water

RD: downstream of receiving water

SW: source water

DO: dissolved oxygen

COD: chemical oxygen demand

TP: total phosphorus

NH_3-N: ammonia nitrogen

SMX: sulfamethoxazole

SMZ: sulfamethazine

ERM: erythromycin

RTM: roxithromycin

NOR: norfloxacin

OFX: ofloxacin

TET: tetracycline

OTC: oxytetracycline

BAC: benzalkonium chloride

PCMX: para-chloro-meta-xylenol

MAGs: metagenome-assembled genomes

ACCs: ARG-carrying contigs

LFE: lactose-fermenting *Enterobacteriaceae*

NMDS: non-metric multidimensional scaling

PCoA: principal coordinates analysis

PLS-SEM: partial least squares structural equation modeling

MLS: macrolides-lincosamide-streptogramin

OPA: other peptide antibiotics

ACGCs: ARG-carrying gene cassettes

Supplementary Method 1 | Quantification of antibiotics, disinfectants, antidepressants, and trace metals

Before analysis, the water samples were prepared by solid-phase extraction. Specifically, 1 L of each sample was filtered through a 0.45 μm glass fiber filter. The pH was then adjusted to 3 with 1 mol/L hydrochloric acid. To chelate metal ions and reduce interference with target analytes, 0.5 g of Na_2EDTA was added, followed by thorough vortex mixing[1,2]. An HLB cartridge (6 mL, 150 mg; Beckman, USA) was then primed for use with 6 mL of methanol and 6 mL of acidified water (pH 3). The prepared sample was percolated through the cartridge at a flow rate of 10 mL/min using a peristaltic pump. After sample loading, the cartridge was rinsed with 30-50 mL of acidified water to ensure the removal of any residual eluent. Subsequently, the sample was subjected to vacuum drying for 20 minutes. The final elution was performed using 6-mL methanol, and the eluate was collected in a 10 mL glass centrifuge tube and stored at −20°C until analysis. Quantitative analysis was performed using liquid chromatography-tandem mass spectrometry (LC-MS/MS) with the Ultimate 3000 UHPLC-Q Exactive system (Thermo Scientific, USA) equipped with an Eclipse Plus C18 column (100 mm × 4.6 mm, 5 μm).

The concentrations of metal ions (Mg, Al, Fe, Mn, Cu, and Zn) were determined using an inductively coupled plasma mass spectrometer (ICP-MS; Agilent 8900, USA). Prior to analysis, each water sample was filtered through a 0.45 μm aqueous filter membrane. Then, 5 mL of the filtered sample was accurately pipetted into a digestion tube, and 2.5 mL of concentrated nitric acid (HNO_3, 65–70% w/w, trace metal grade) was added. The mixture was subsequently digested in a microwave digestion system at 100℃ for 1 h. After digestion, the solution appeared clear and transparent with no visible precipitates. After cooling to room temperature, the sample was diluted with deionized water to a predetermined volume and analyzed by ICP-MS.

Supplementary Method 2 | 16S rRNA gene sequencing and bioinformatics analysis

The 16S rRNA gene was amplified by PCR targeting the V3-V4 regions with primers 338F (5'-ACTCCTACGGGAGGCAGCAG-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3')[3]. The thermal cycling protocol consisted of an initial denaturation step at 95°C for 3 min, followed by 35 cycles of 95°C for 20 s, 52°C

for 15 s, and 72°C for 1 min, with a final extension at 72°C for 10 min. A PE300 library preparation strategy was adopted and sequenced on the Illumina MiSeq platform (Majorbio, Shanghai, China). This ensured that each sample generated no fewer than 40,000 reads for the 16S rRNA genes. High-throughput sequencing results from this amplification product were applied for microbial diversity analysis. Specifically, bioinformatics analysis of the 16S rRNA genes was conducted using Vsearch (v2.22.1)[4]. The raw sequencing data were processed with Vsearch's fastx_filter and derep_fulllength parameters to trim primers, perform quality control, and remove redundancies. To exclude or correct erroneous sequences and select high-abundance, reliable sequences as representative, this study employed the unoise3 parameter in Usearch (v11.0)[5] for sequence denoising, followed by the construction of an ASV table using Vsearch. Alpha and beta diversity indices were calculated using the alpha_div plugin and the "vegan" package[6]. Taxonomic classification was performed using the RDP reference database (v11)[7].

Supplementary Method 3 | Identification and quantification of ARGs

All clean reads after quality control were subjected to a search for ARGs against the SARG (v3.0) database[8], employing UBLAST strategies with an e-value threshold of $\leq 10^{-7}$. A sequence was classified as an ARG-like fragment if its optimal BLASTX alignment with the ARG sequence in the referenced database exhibited a similarity of ≥ 80% and an alignment length of ≥ 75%. This study employed the unit "copy per cell" as it represents the most widely adopted and recommended standard for ARG profiling among researchers globally[9]; More importantly, it facilitates the further examination of the microscopic mechanisms (e.g., vertical transmission and horizontal transfer) associated with the proliferation and dissemination of AMR determinants in UWS[10]. The detailed calculation formula is presented as follows[9]:

$$\text{ARG Abundance (copy/cell)} = \sum_1^n \frac{N_{ARG-like\ reads} \times L_{read} / L_{ARG\ reference\ sequence}}{Cell\ number} \quad \text{(S1)}$$

where $N_{ARG\text{-}like\ reads}$ is the number of ARG-like reads annotated to one specific ARG reference sequence, L_{read} indicates the clean read length, $L_{ARG\ reference\ sequence}$ is the nucleotide sequence length of the corresponding ARG reference sequence, n is the number of mapped ARG reference sequences attributed to the ARG type or subtype,

and *Cell number* is the cell number estimated by mapping against an essential single-copy marker gene database[8].

Supplementary Method 4 | Construction of the Class 1 Integrase Database

The class 1 integrase (*int*I1) reference database was constructed based on publicly available sequences deposited in the National Center for Biotechnology Information (NCBI) database and the INTEGRALL database[11]. Briefly, annotated *int*I1-related sequences were retrieved from NCBI using the keywords "*int*I1" and "class 1 integrase", with searches restricted to the Gene and Protein categories, and complemented with curated class 1 integrase sequences obtained from INTEGRALL. After further dereplication and validation by comparing against Pfam and CDsearch database, 148 sequences were retained, composing the preliminary *int*I1 database[12]. To further expand sequence diversity, the preliminary *int*I1 database was used as a query to search against the NCBI non-redundant protein database (nr) using BLASTP, with an E-value cutoff of 1×10^{-5}. Candidate sequences were filtered based on sequence similarity (≥ 80%), and alignment coverage (≥ 50%). Duplicate or partial sequences were excluded. The retained candidates were subjected to an additional round of validation against the Pfam and CDsearch databases to exclude non-integron tyrosine recombinases. The resulting high-confidence sequences (n = 424) constituted the final class 1 integrase (*int*I1) reference database used for downstream analyses.

Supplementary Method 5 | Detection of integron gene cassettes by PCR and Nanopore sequencing

The PCR amplification targeting the gene cassettes of class 1 integron was performed using specific primer sequences and reaction conditions. The forward primer sequence was 5'-GGCATCCAAGCAGCAAG-3', and the reverse primer sequence was 5'-AAGCAGACTTGACCTGA-3'[13]. The reaction protocol included an initial denaturation step at 94°C for 5 minutes, followed by 30 cycles of denaturation at 94°C for 1 minute, annealing at 55°C for 45 seconds, and extension at 72°C for 5 minutes. A final extension was carried out at 72°C for 10 minutes to ensure complete amplification of the target region. This reaction setup was optimized to ensure high specificity and amplification efficiency under standard PCR conditions.

Following amplification, a series of quality control steps were implemented to ensure the reliability of integron gene cassette sequences prior to downstream analyses. First, PCR amplicons were examined by agarose gel electrophoresis to verify amplification specificity and confirm expected fragment sizes, and subsequently purified using a commercial silica membrane-based DNA purification kit (QIAGEN). The purified amplicons were then subjected to Nanopore sequencing following a stringent workflow for library preparation, sequencing, and quality control. Low-quality reads, chimeric sequences, and reads with excessive base-calling errors were filtered using chopper (v0.8.0)[14] prior to further analyses. Open reading frames (ORFs) were predicted using Prodigal (v2.6.3), with only complete and high confidence gene predictions retained. ORFs lacking intact start or stop codons, or those flagged as partial or truncated, were excluded to minimize artefacts arising from fragmented DNA sequences.

Supplementary Method 6 | Detection of high-risk ARGs on unbinned metagenomic contigs

To complement the MAG-based resistome risk assessment and to evaluate the potential underrepresentation of clinically relevant ARGs caused by assembly and taxonomic binning limitations, unbinned contigs (≥ 1,000 bp) were extracted from each metagenomic assembly. ORFs on these unbinned contigs were predicted using Prodigal (v2.6.3)[15]. The predicted ORFs were then searched against the SARG (v3.0) database using DIAMOND (v2.0.15), with an e-value threshold of $\leq 1 \times 10^{-10}$, a minimum amino acid identity of 80%, and a query coverage of ≥ 70%, to identify ARG-like ORFs[16]. To assess the potential association between ARGs and mobile genetic elements (MGEs), ORFs annotated as ARGs were further examined for spatial proximity to MGE-related genes on the same contig. All ORFs were screened against the MobileGeneticElement database[17] using DIAMOND with the same alignment criteria to identify MGE-related genes. An ARG was considered potentially mobilizable if an MGE-related gene was located within a 5 kb flanking region upstream or downstream on the same contig[18]. Subsequently, these unbinned contigs harboring both ARGs and nearby MGE-related genes were compared against the NCBI non-redundant (nr) database to provide additional functional annotation and taxonomic context, to identify those high-risk

ARGs associated with genetic mobility and pathogenic bacterial hosts.

Supplementary Method 7 | Absolute quantification based on mass conservation

Absolute ARG concentrations (C_{abs}, copies·mL^{-1}) were calculated from relative abundances using a mass conservation-based approach[19,20]:

$$C_{abs} = \frac{C_{rel} \times N_{cell}}{V_{sample}} \times \frac{M_{extract}}{M_{seq}} \quad \text{(S2)}$$

where C_{rel} is the relative ARG abundance (copies per cell) calculated based on metagenomic sequencing datasets, N_{cell} represents the cell number in the sequenced reads quantified by ARGs-OAP on the alignment of essential single-copy marker genes, V_{sample} is the volume of water filtered (mL), $M_{extract}$ is the total DNA mass extracted from the sample, and M_{seq} is the DNA mass represented by the sequencing reads (gram, as double strand DNA), calculated from nucleotide composition as:

$$M_{seq} = \frac{2 \times \sum(n_i \times M_i)}{N_A} \quad \text{(S3)}$$

with n_i being the total count of nucleotide i (A, T, C, G) in the sequencing reads, M_i represents its molecular weight (313.2, 304.2, 289.2, 329.2 g·mol^{-1}), and N_A is Avogadro's constant.

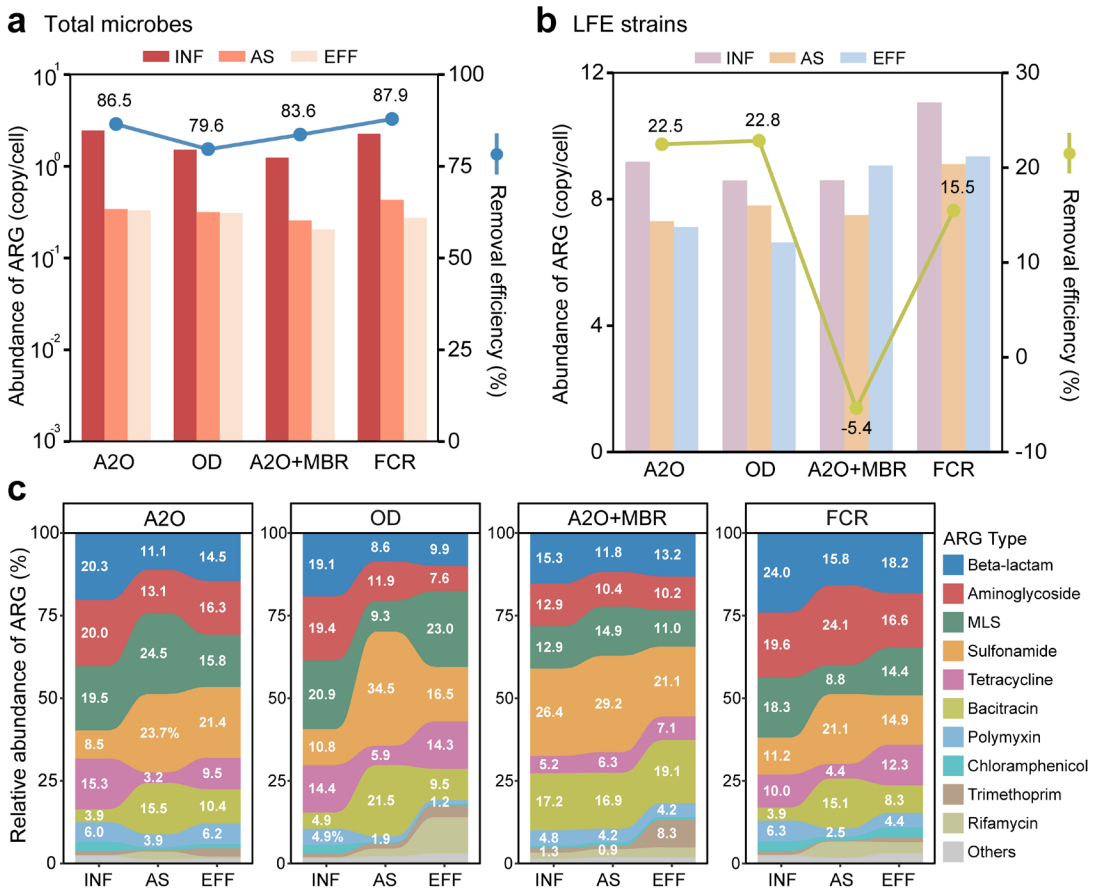
Supplementary Method 8 | Competitive genome-wide read recruitment to MAGs from tap water metagenomes

To assess the robustness of low-abundance pathogen signals in tap water samples, we performed competitive genome-wide read recruitment to MAGs using Bowtie2 alignment paired with downstream inStrain (v1.10.0)[21] profiling. All non-redundant MAGs recovered in this study were concatenated to generate a unified reference database (non-redundant MAG set), with standardized renaming and bin annotation of all contigs and scaffolds to ensure unique genome identifiers. Clean paired-end reads from drinking water samples were mapped to the combined MAGs dataset using Bowtie2[22] (v2.5.1, end-to-end mode), enabling competitive alignment of each read across all reference genomes and reducing single-reference mapping bias. SAM files were converted to sorted BAM format using SAMtools (v1.17)[23]. For key pathogen-

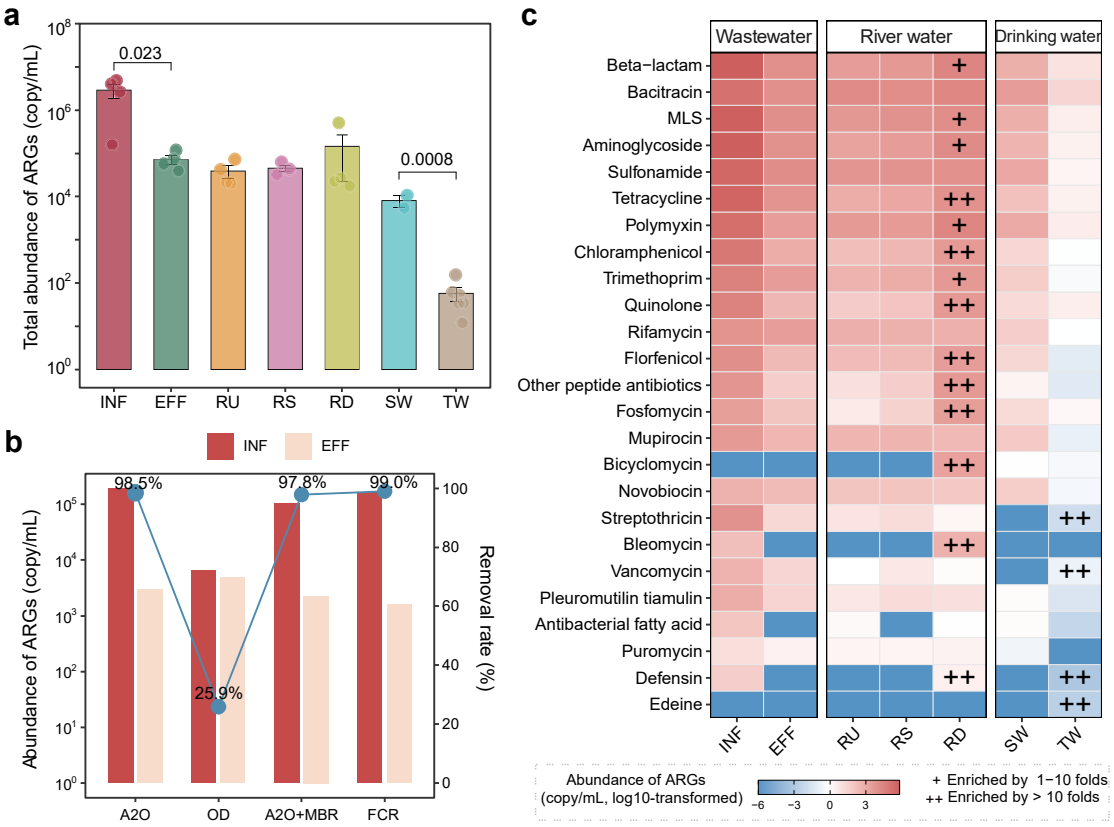
associated MAGs, including *Pseudomonas aeruginosa* (bin5), *Mycobacterium phocaicum* (bin149), *Escherichia coli* (bin140), and *Acinetobacter johnsonii* (bin1192), mapped reads were extracted and analyzed using inStrain profile to calculate genome-wide coverage, coverage breadth, and nucleotide diversity.

Supplementary Method 9 | Ecological assembly processes of ARGs and ARG host communities

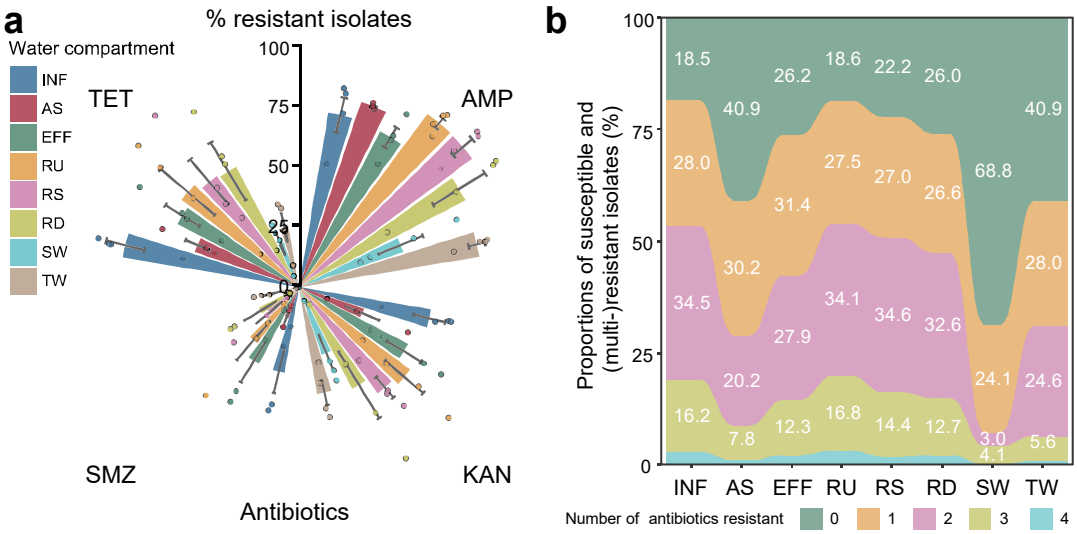
To characterize the ecological assembly mechanisms of ARG communities and their host microbial genomes, we integrated the Neutral Community Model (NCM)[24] and a phylogeny- and abundance-based null model framework[25]. First, the NCM was used to estimate the contribution of stochastic processes (e.g., dispersal and ecological drift) to ARG community assembly, with the goodness-of-fit (R^2) representing the relative importance of neutral processes. Second, for ARG host communities, the beta nearest taxon index (βNTI) was calculated to assess phylogenetic turnover and infer deterministic selection, where $|\beta NTI| > 2$ indicates heterogeneous or homogeneous selection. For comparisons with $|\beta NTI| \leq 2$, the Bray–Curtis-based Raup–Crick metric (RCbray) was further applied to partition stochastic processes, including dispersal limitation, homogenizing dispersal, and ecological drift. Together, this framework enabled quantitative estimation of the relative contributions of deterministic and stochastic processes to community assembly.



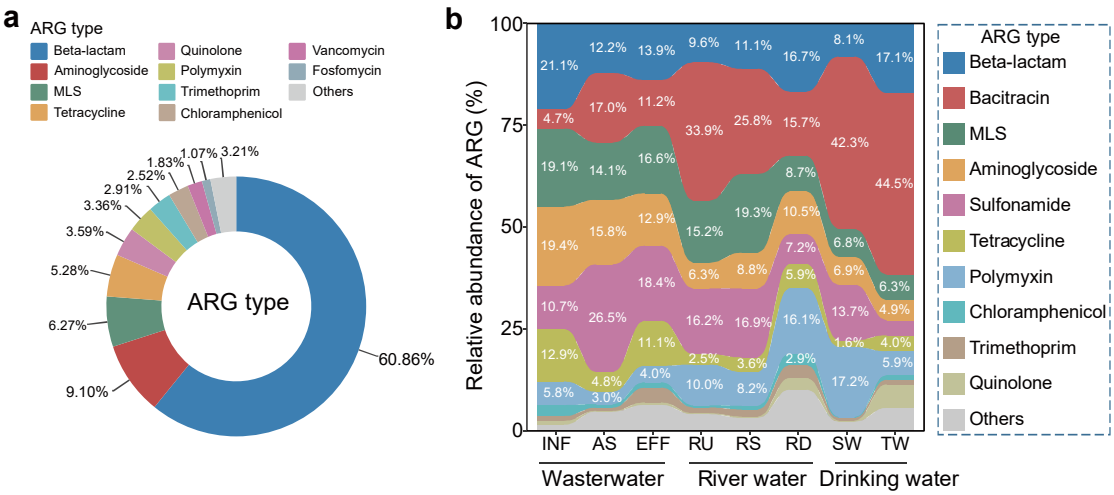
Supplementary Fig. 1 | Comparison of resistome profiles and their removal efficiency across wastewater treatment processes. a, Abundances and removal efficiencies of ARGs in total microbial communities across wastewater treatment processes. **b**, Abundances and removal efficiencies of ARGs in lactose-fermenting *Enterobacteriaceae* (LFE) strains across treatment processes. **c**, Changes in ARG types within total microbial communities across treatment processes. *A2O: anaerobic-anoxic-oxic; OD: oxidation ditch process; MBR: membrane bio-reactor; FCR: food chain reactor; INF: influent; AS: activated sludge; EFF: effluent; RU: upstream of receiving water; RS: outlet site of receiving water; RD: downstream of receiving water; SW: source water; TW: tap water*. Source data are provided as a Source Data file.



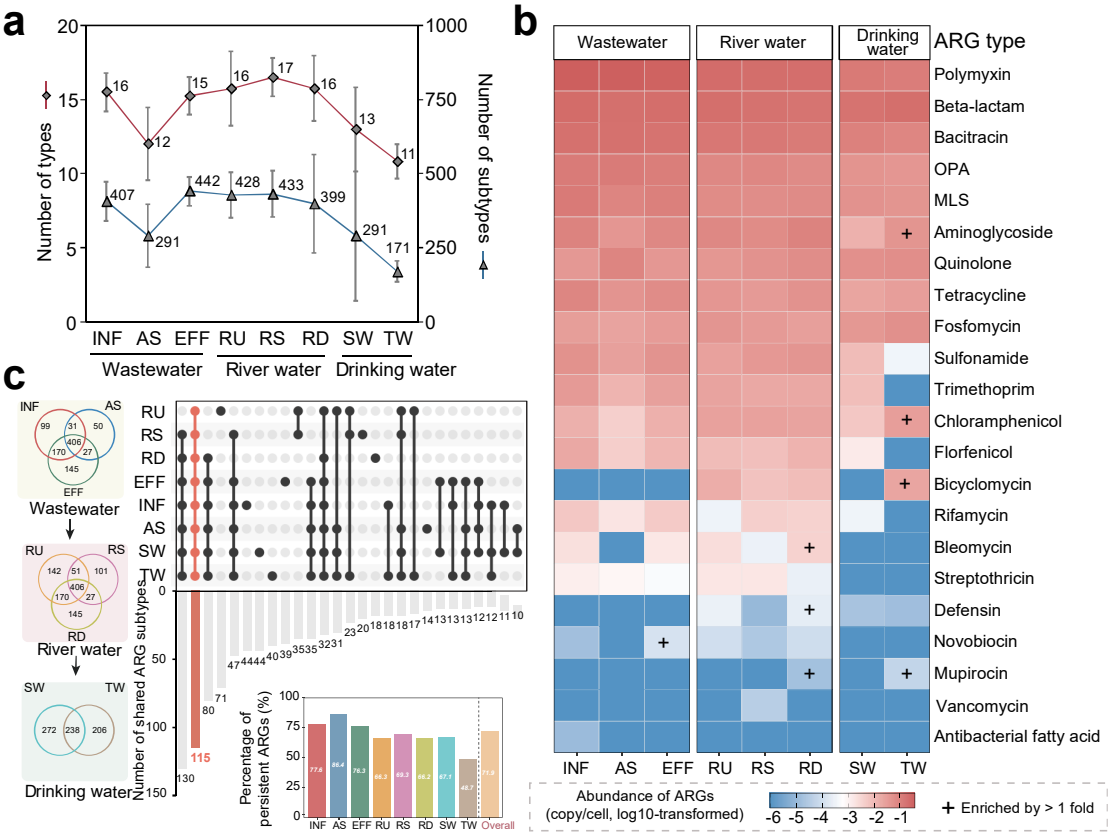
Supplementary Fig. 2 | Absolute abundance dynamics of ARGs across urban water compartments. a, Shifts in the absolute abundance of ARGs (copy/mL) within the total microbial community across different water compartments. Bars indicate mean values, error bars show standard errors of the mean (SEM), and individual dots represent different samples from each compartment (n = 4 for INF, AS, EFF, RU, RS, and RD; n = 2 for SW; n = 6 for TW). Statistical significance was assessed using unpaired, two-tailed Student's t-test, with exact p values shown above the brackets. **b**, Absolute abundances and removal efficiencies of ARGs across wastewater treatment processes. **c**, Changes in the absolute abundance of major ARG types across water compartments. Symbols '+' and '++' denote the magnitude of ARG enrichment in key comparisons: downstream *vs.* upstream receiving water (RD/RU), and tap water *vs.* source water (TW/SW). Source data are provided as a Source Data file.



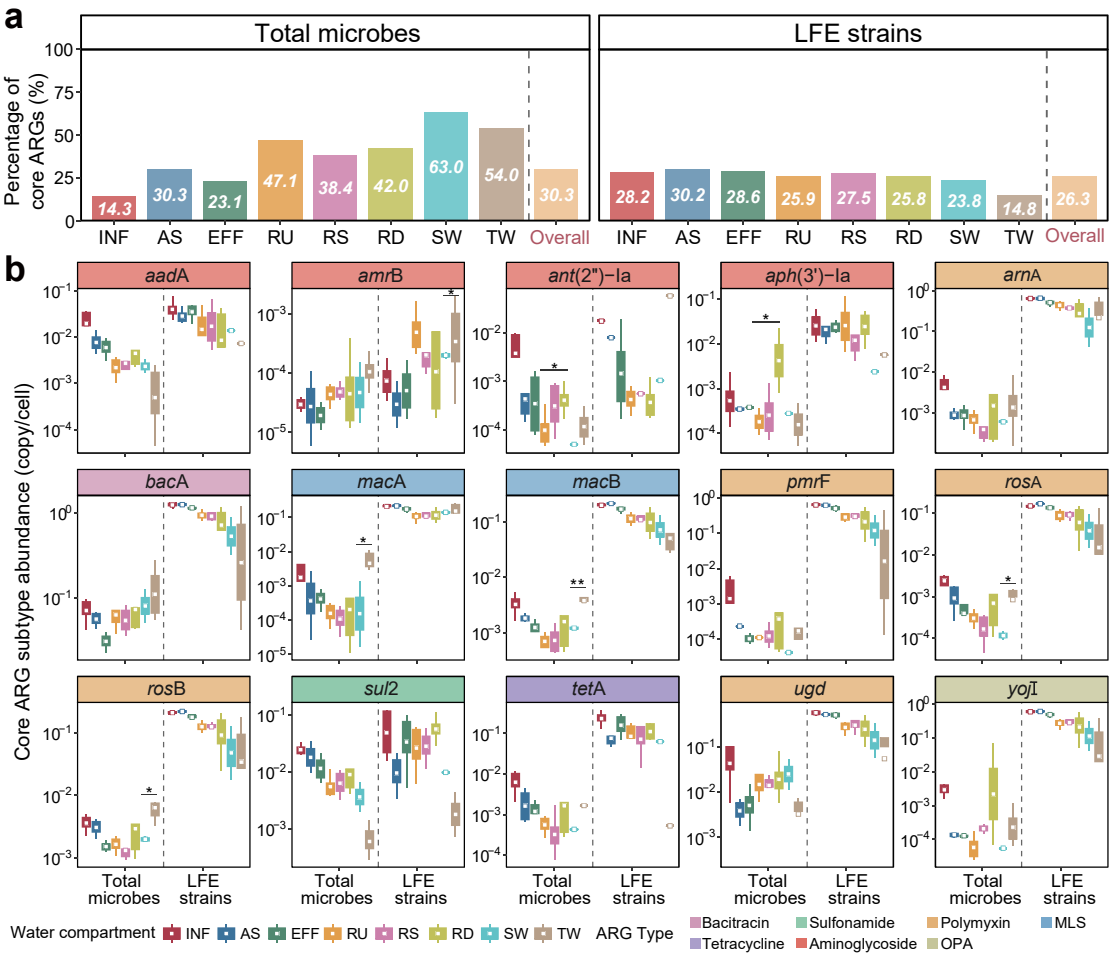
Supplementary Fig. 3 | Variation in antibiotic resistance rates among 3,040 LFE isolates from the urban water system. a, Variation in the percentage of antibiotic-resistant LFE strains across water compartments. Error bars represent the standard deviation. Bars indicate mean values, error bars show SEM, and individual dots represent different samples from each compartment (n = 4 for INF, AS, EFF, RU, RS, and RD; n = 2 for SW; n = 6 for TW). Statistical significance between the indicated pairs was determined using an unpaired, two-tailed Student's t-test. **b**, Distribution of susceptible and (multi-)resistant LFE strains. *AMP: ampicillin; TET: tetracycline; KAN: kanamycin; SMZ: sulfamethoxazole.* Source data are provided as a Source Data file.



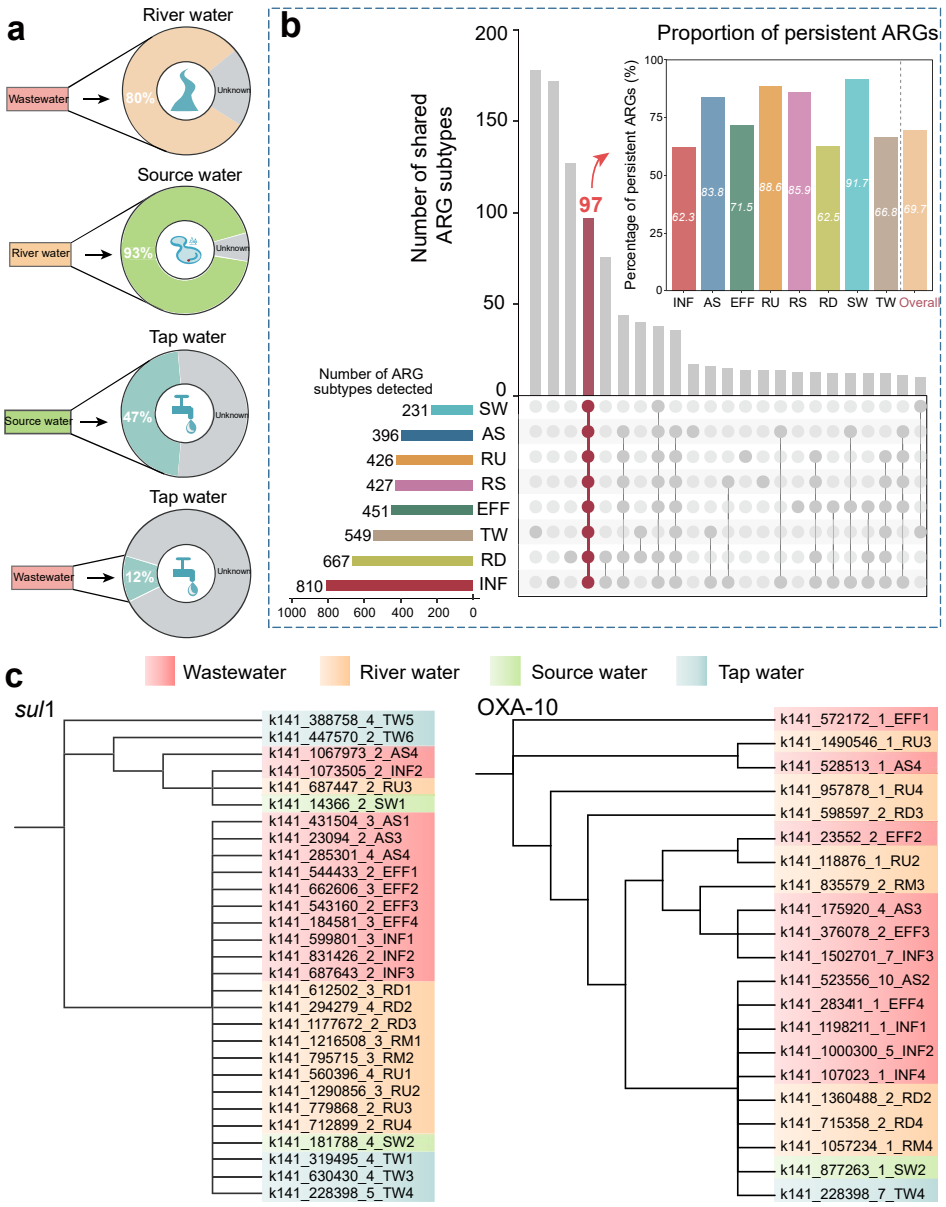
Supplementary Fig. 4 | Proportion and abundance of ARG types across the urban water system. **a**, Overall proportion of ARG types across water compartments indicated by subtype numbers. **b**, Relative abundance of top 10 ARG types, with remaining types grouped as "Others". *MLS: macrolide-lincosamide-streptogramin*. Source data are provided as a Source Data file.



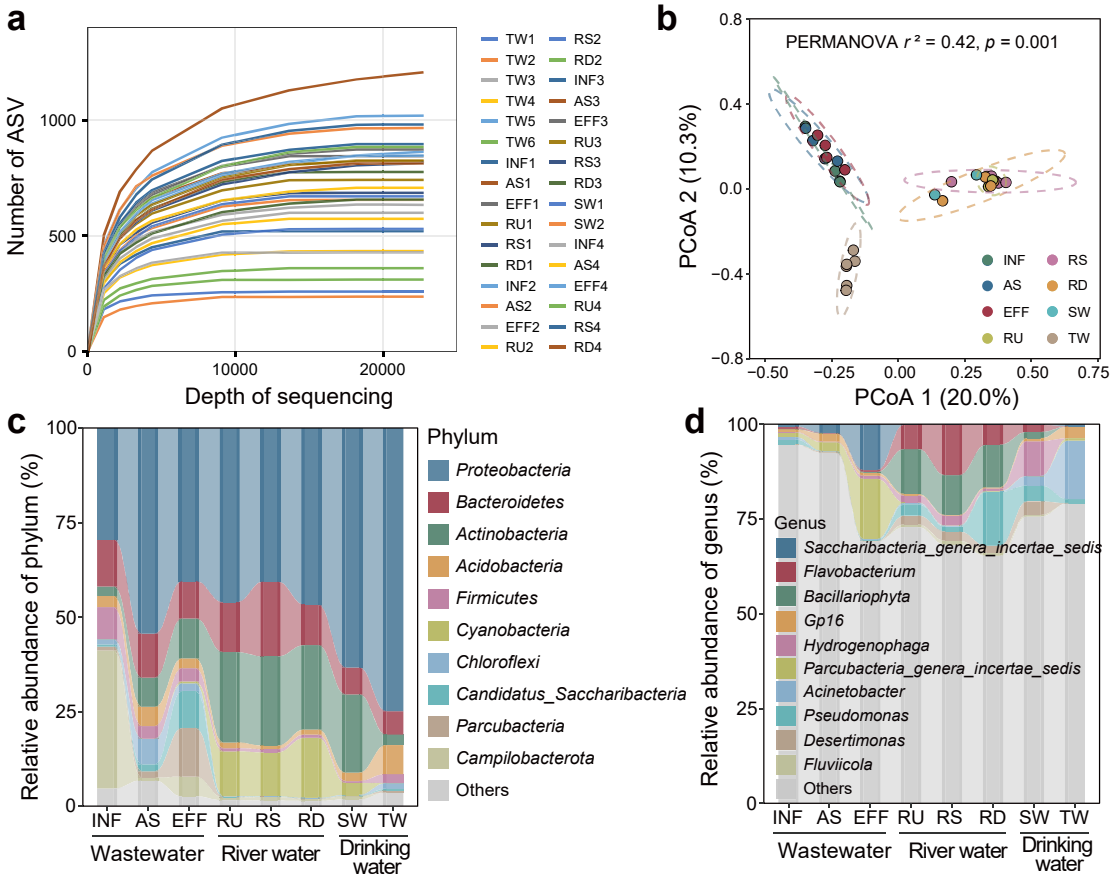
Supplementary Fig. 5 | Overview of persistent ARGs in LFE strains. a, Shifts in ARGs diversity indicated by types and subtypes detected in each water compartment. **b**, Heatmap plot showing the abundance of ARG types in the urban water system. Symbol '+' denotes ARG enrichment in the following comparisons: effluent *vs.* influent (EFF/INF), downstream *vs.* upstream receiving water (RD/RU), and tap water *vs.* source water (TW/SW). **c**, Venn and UpSet plot illustrating the number of ARG subtypes detected across different water compartments, and the count of persistent ARGs. Source data are provided as a Source Data file.



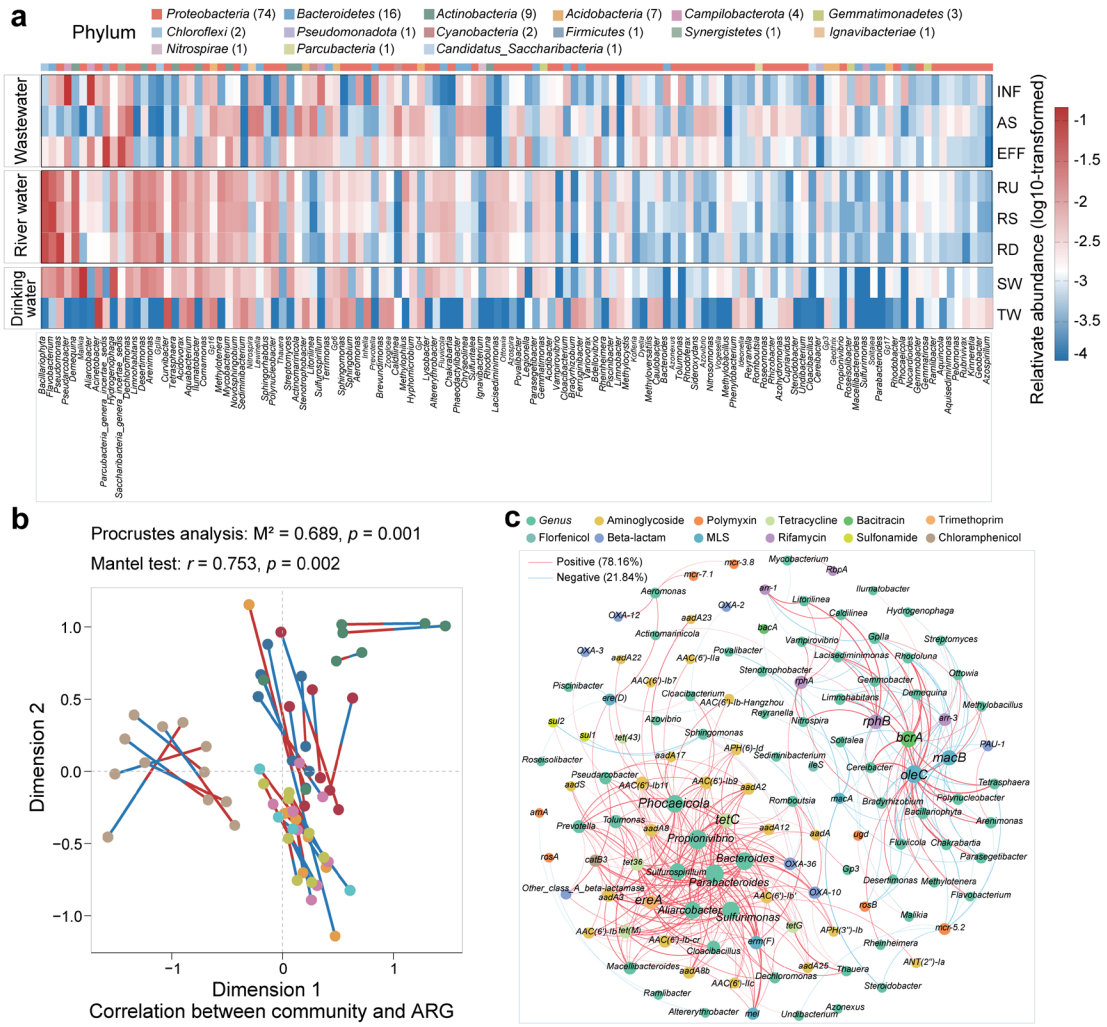
Supplementary Fig. 6 | Overview of core ARGs in total microbes and LFE strains. **a**, Proportions of core ARGs in the total microbes and LFE strains across water compartments, with "Overall" indicating total abundance. **b**, Boxplots showing the abundance shifts of core ARGs carried by total microbes and LFE strains in urban water system. Significance is denoted as follows: *$p \leq 0.05$ and **$p \leq 0.01$, determined by using Student's t-test. For the box plots, the center line represents the median; box limits indicate the 25th and 75th percentiles; and whiskers extend to the maximum and minimum values. Source data are provided as a Source Data file.



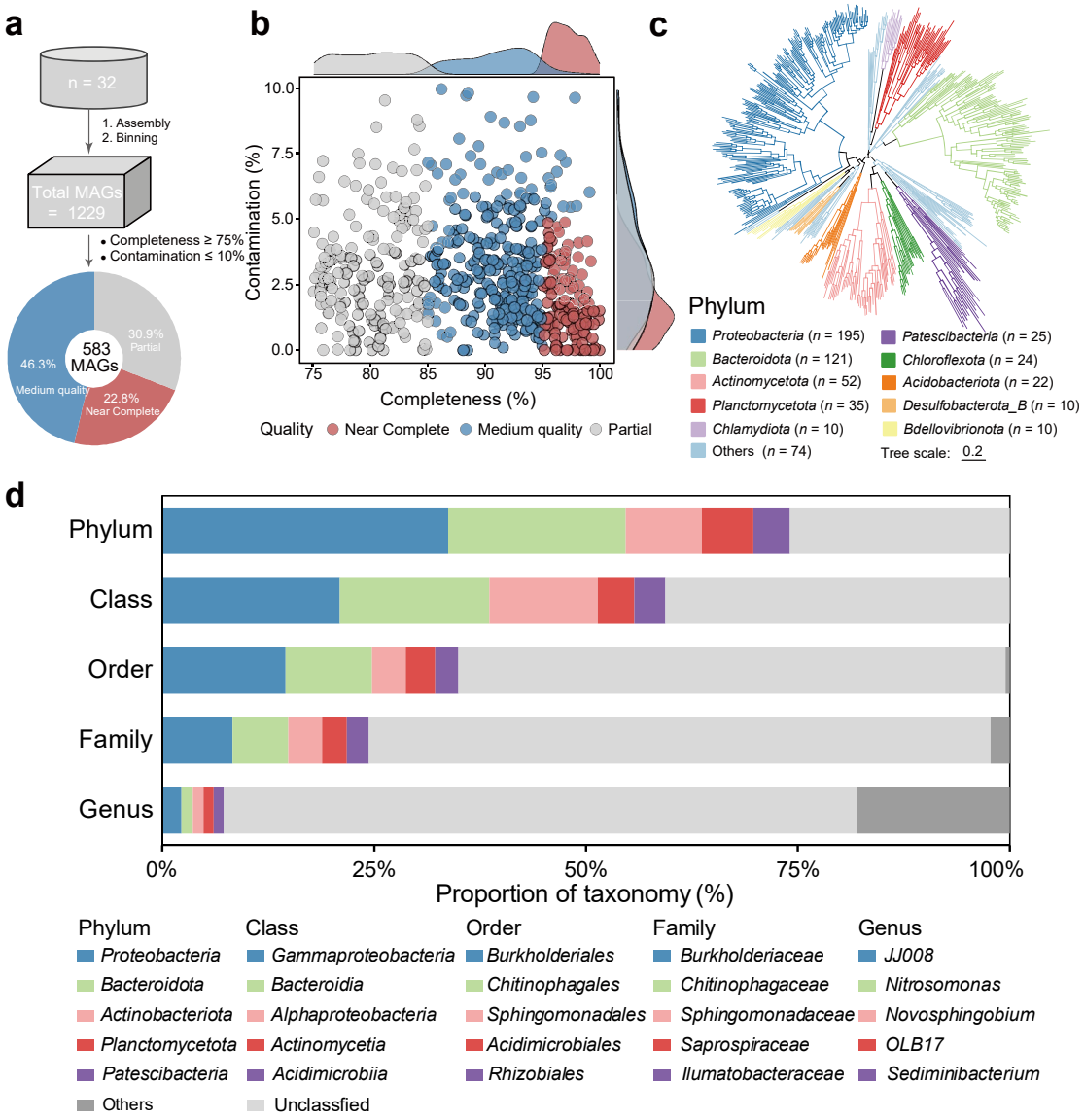
Supplementary Fig. 7 | Ecological connectivity analyses of ARGs across three hierarchical levels (resistome, subtype, sequence) in the urban water system. a, Bayesian source-tracking analysis showing the proportional contributions of sewage, receiving river, and source water to downstream ARG profiles, respectively. ARGs not assigned to any predefined source are shown as "Unknown". **b**, UpSet plot depicting the distribution of persistent ARG subtypes across water compartments. The adjacent bar chart indicates their relative proportion. **c**, Maximum-likelihood phylogenetic tree of *sul*1 and OXA-10 based on their complete ORF sequences recovered from metagenomic assemblies across urban water compartments, illustrating their sequence-level connectivity. Source data are provided as a Source Data file.



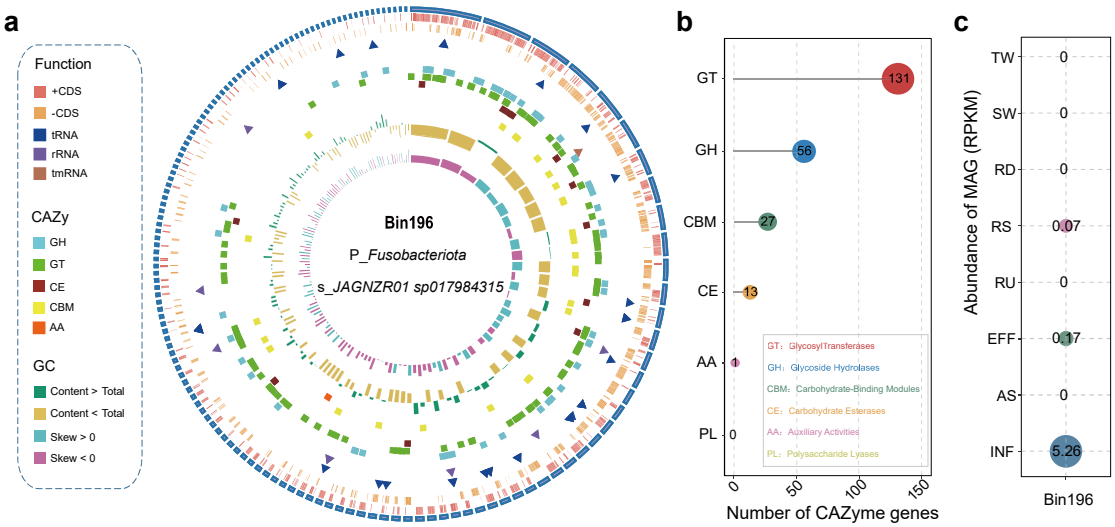
Supplementary Fig. 8 | Profile of microbial communities across urban water system. a, Rarefaction curve of amplicon sequence variants (ASVs). **b**, Principal Coordinate Analysis (PCoA) showed the significant differences across water compartments, based on Bray-Curtis dissimilarity matrix calculated from the ASV matrix. Percentages of variation explained by PCoA1 and PCoA2 is indicated in the axis labels. The microbial community compositions at the phylum (**c**) and genus (**d**) levels, with the top 10 most abundant microbial taxa presented, while all remaining phyla and genera were grouped into a single category labeled "Others". Source data are provided as a Source Data file.



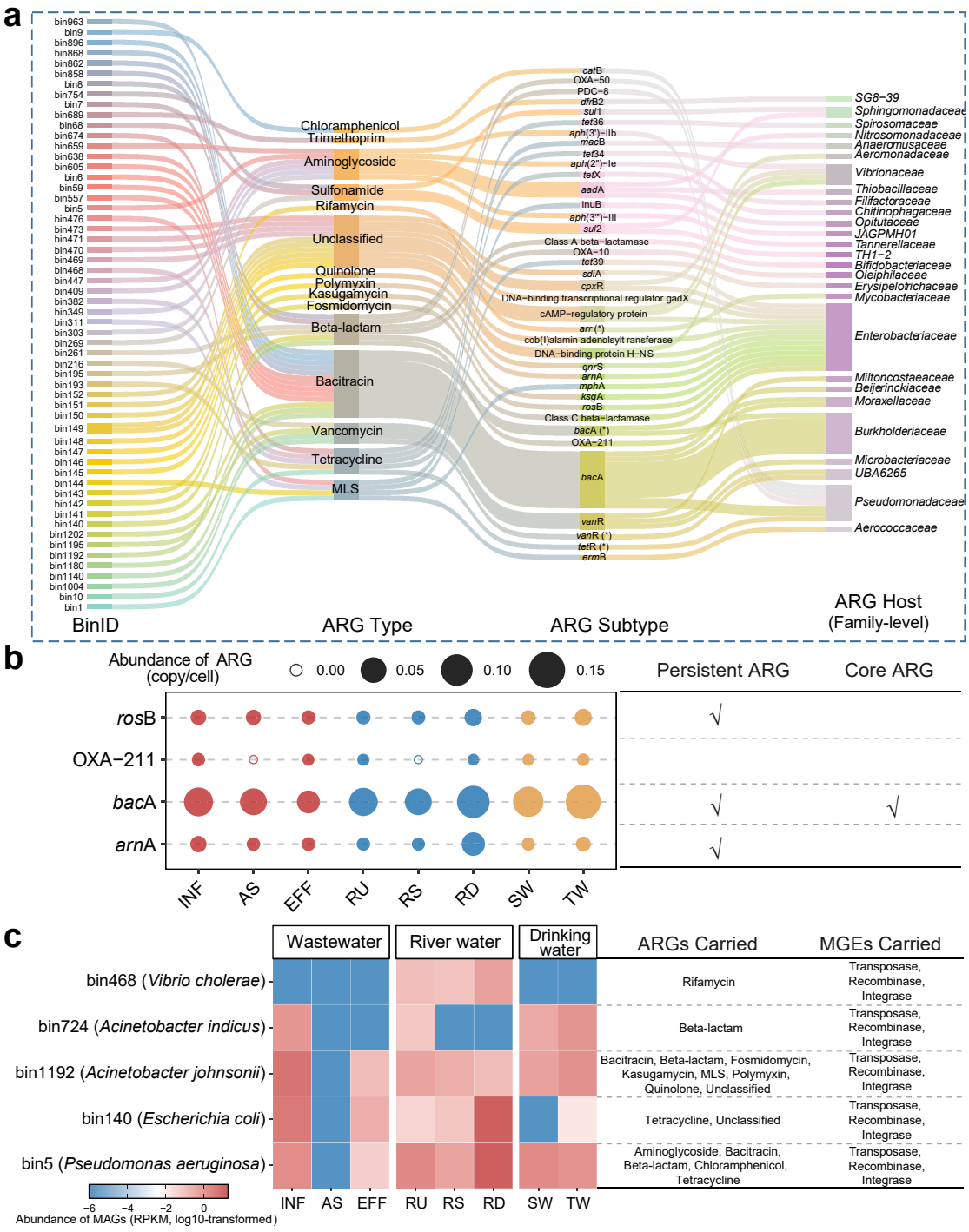
Supplementary Fig. 9 | Identification of primary genera and potential ARG hosts across urban water system. a, Heatmap presents the abundances of 124 primary genera. These genera were defined as those detected in > 50% of the samples and having an average relative abundance > 0.1% across all samples. Genera are color-coded by their respective phylum, with parenthetical numbers indicating the count of genera belonging to each phylum. **b**, Procrustes analysis illustrating the correlations between resistome and the microbial community. **c**, Co-occurrence network analysis based on ARGs abundance and bacterial communities (genus level). Nodes represent distinct ARG subtypes or bacterial genera, with ARG subtypes colored according to ARG types. Edges between nodes indicate significant correlations (Spearman's correlation coefficient $r > 0.7$, $p < 0.01$), with edge thickness proportional to the correlation strength. Node size corresponds to the number of connections. Source data are provided as a Source Data file.



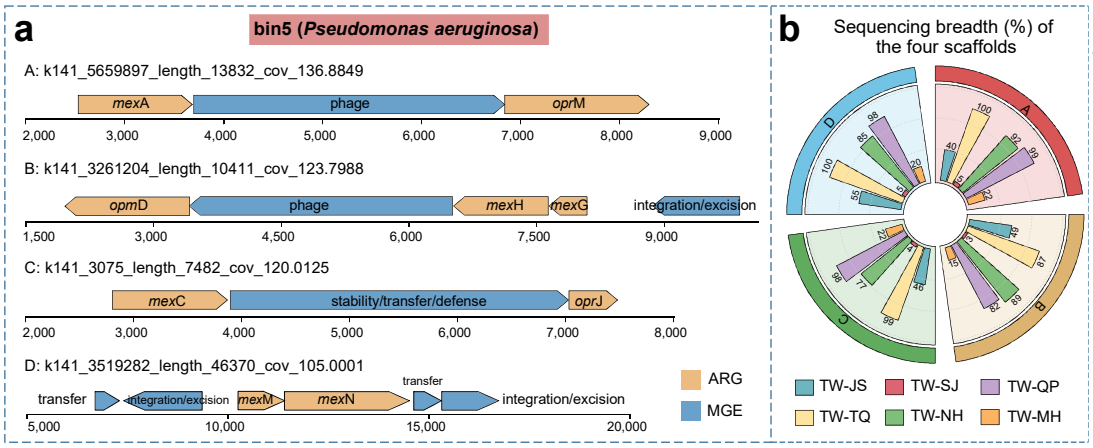
Supplementary Fig. 10 | Reconstruction and characterization of metagenome-assembled genomes (MAGs) across urban water system. a, Recovery of 583 high-quality MAGs through metagenomic binning. **b**, Quality assessment of MAGs showing completeness and contamination distributions. **c**, Phylogenetic tree of bacterial MAGs. **d**, Taxonomy composition of bacterial MAGs at different phylogenetic levels. Source data are provided as a Source Data file.



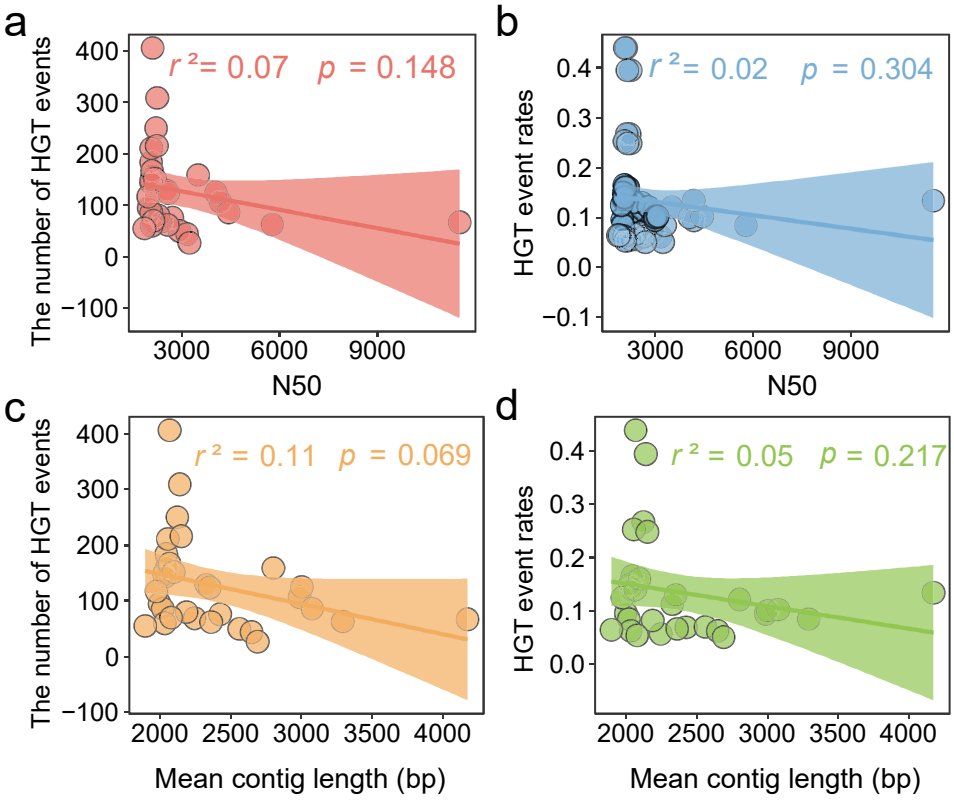
Supplementary Fig. 11 | An overview of a complete genome (100% completeness, 0% contamination). a, Circular genome map illustrating genomic features of a complete genome assigned as *JAGNZR01 sp017984315*. From the outermost to innermost rings: the first circle displays contigs labeled by serial number; the second circle highlights protein-coding sequences (CDS) on the positive strand in red and on the negative strand in yellow; the third circle indicates RNA features, with blue representing tRNA, purple representing rRNA, and brown representing tmRNA; the fourth circle depicts carbohydrate-active enzyme (CAZyme) annotations, where cyan squares represent glycoside hydrolases (GH), green squares represent glycosyltransferases (GT), brown squares represent carbohydrate esterases (CE), yellow squares represent carbohydrate-binding modules (CBM), and orange squares represent auxiliary activities (AA). **b**, Gene number of different carbohydrate-active enzyme (CAZymes) families carried by Bin196. **c**, Abundances of Bin196 across urban water compartments. Source data are provided as a Source Data file.



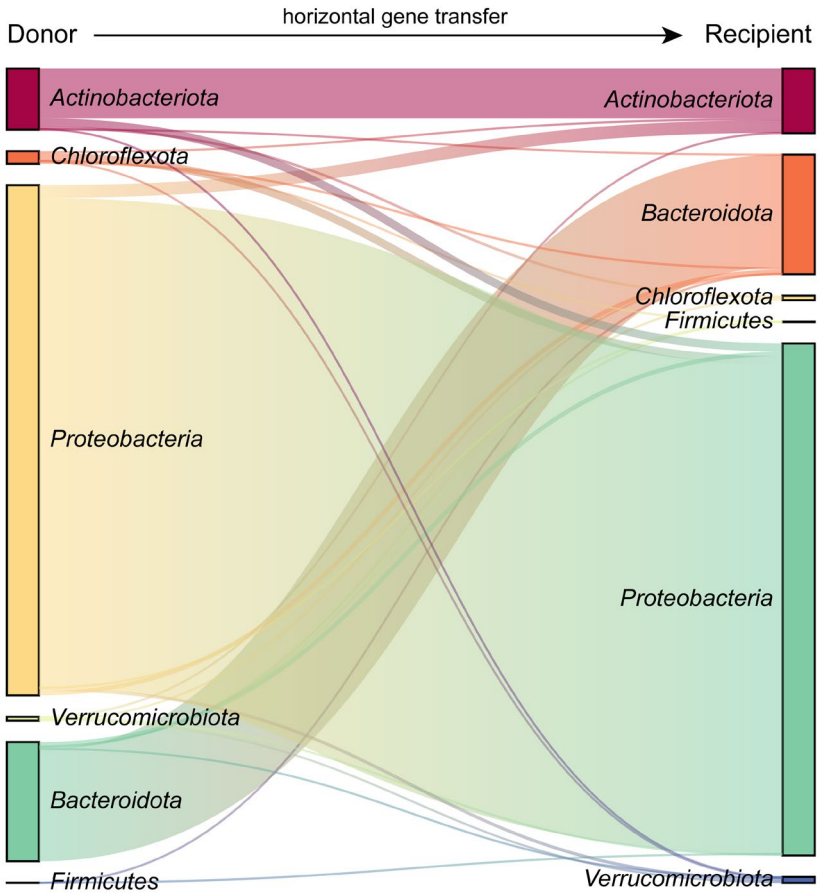
Supplementary Fig. 12 | Overview of the genetic bacterial hosts of ARGs in urban water system. a, Sankey diagram showing the categories of ARGs and their hosts. **b**, Abundances of ARG subtypes which were significantly enriched in the urban water system. **c**, Abundances of five pathogenic ARG-carrying MAGs and their carriage of ARGs and MGEs. Source data are provided as a Source Data file.



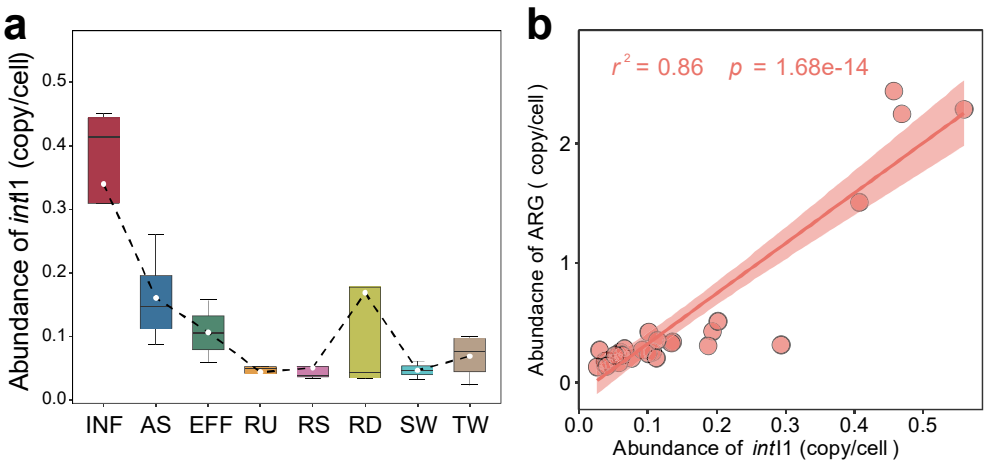
Supplementary Fig. 13 | ARG–MGE co-localization and scaffold recruitment patterns of *Pseudomonas aeruginosa* (bin5) across tap water samples. a, Genomic context and environmental distribution of ARG-MGE co-localization in *Pseudomonas aeruginosa* (bin5). Schematic representation of four representative scaffolds (A-D) showing the physical linkage between ARGs (light orange) and MGEs (blue). **b**, Polar bar plot illustrating the sequencing breadth (%) of these scaffolds across six distinct sampling sites. Source data are provided as a Source Data file.



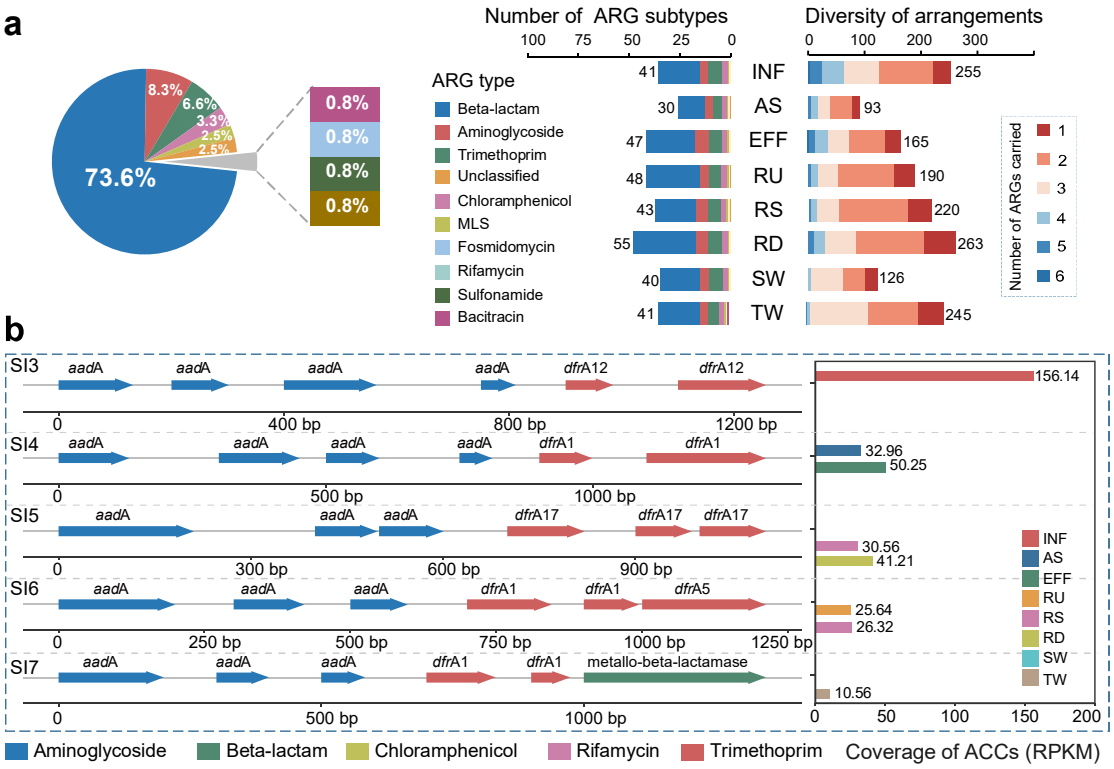
Supplementary Fig. 14 | Correlation between assembly contiguity and detected HGT events. Scatter plots show the correlations between assembly quality metrics and HGT detection outcomes across all samples. Panels depict the relationships between (**a**) N50 and the number of HGT events, (**b**) N50 and HGT event rates, (**c**) mean contig length and the number of HGT events, and (**d**) mean contig length and HGT event rates. HGT event rates were normalized by total assembly size. Each panel displays the coefficient of determination (r^2) and corresponding p-values from Spearman's rank correlation. Source data are provided as a Source Data file.



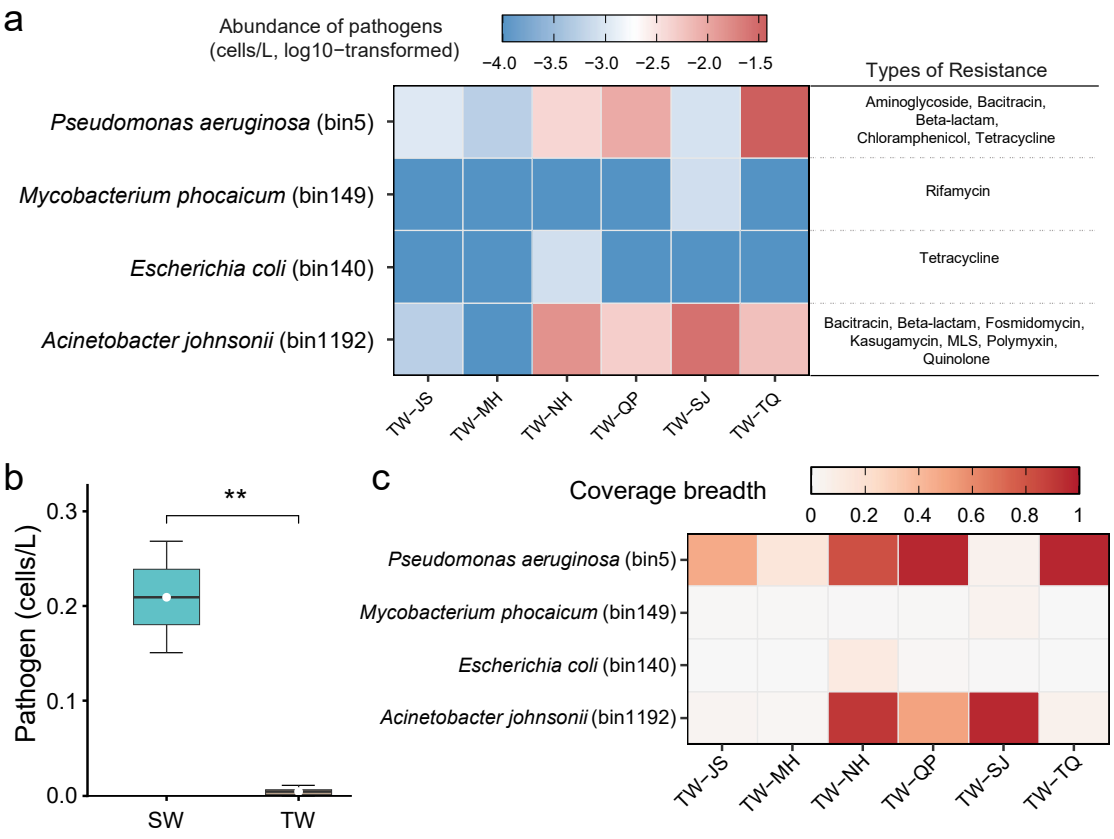
Supplementary Fig. 15 | Phylum-level summary of inferred donor-recipient relationships in ARG-associated horizontal gene transfer. Directional Sankey diagram summarizing horizontal gene transfer events inferred by MetaCHIP at the phylum level. Each flow represents inferred HGT events from donor phyla (left) to recipient phyla (right), with link widths proportional to the number of detected events. Source data are provided as a Source Data file.



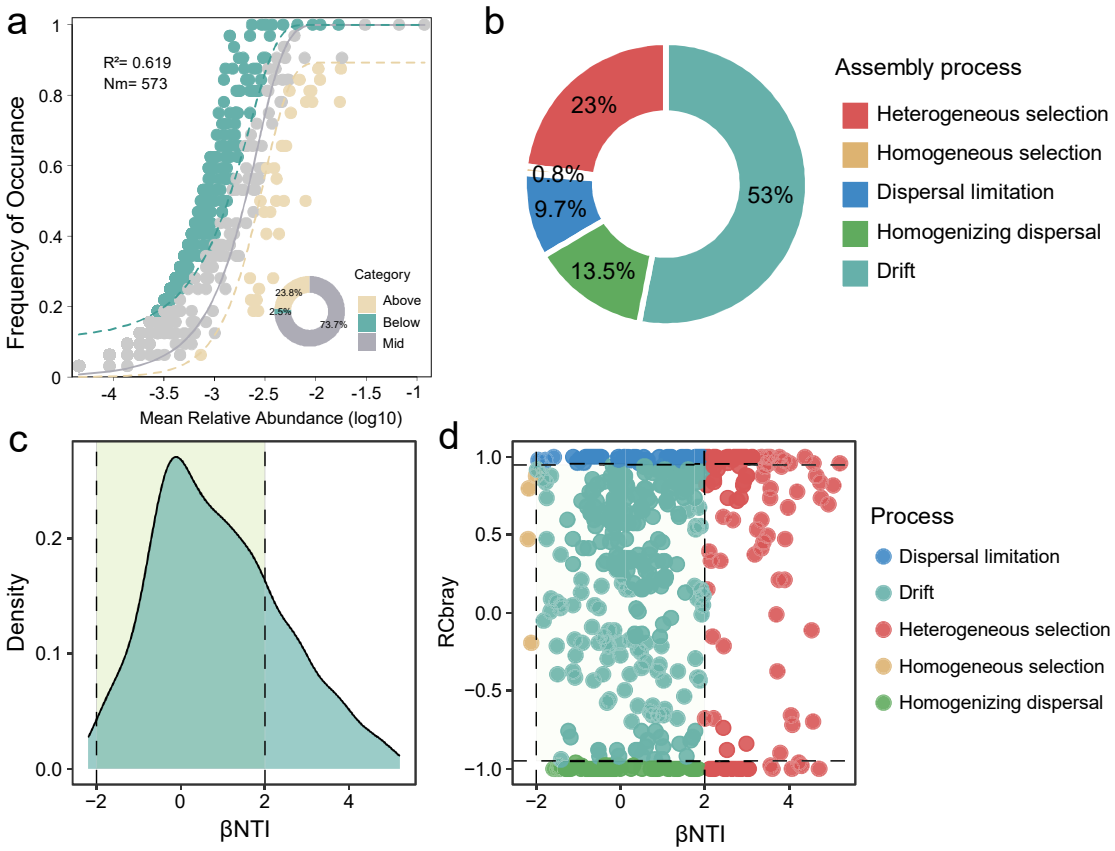
Supplementary Fig. 16 | Quantitative analysis of *int*I1 based on metagenomics. a, Abundance of *int*I1 in the urban water system. **b**, linear regression of ARG abundance versus *int*I1 abundance based on Pearson correlation analysis. Linear relationships indicate significant correlations ($p < 0.05$), with shaded areas representing 95% confidence intervals. Source data are provided as a Source Data file.



Supplementary Fig. 17 | Characterization of ARG-carrying integrons in LFE strains. a, Paired bar plots quantify ARG load and genetic arrangements within integrons, with concentric pie charts showing ARG subtype distributions. **b**, Genomic architectures of super-integrons harboring six ARGs, presented alongside their abundance profiles across water compartments. Source data are provided as a Source Data file.



Supplementary Fig. 18 | (Opportunistic) pathogens detected in drinking water. a, Heatmap illustrating the cell-equivalent concentrations of pathogen-associated metagenome-assembled genomes (MAGs) in tap water samples. Values are expressed as cells/L and $\log_{10}$-transformed. The right panel lists ARG classes identified in each MAG by genome annotation. **b**, Boxplots comparing total pathogen concentrations in source water (SW) and tap water (TW). Boxes represent the interquartile range, center lines indicate medians, whiskers denote the data range, and white dots indicate mean values. Statistical significance was determined by Student's t test (**$p \leq 0.01$). **c**, Heatmap showing the coverage breadth of pathogen-associated MAGs across six tap water samples, determined by competitive genome-wide mapping using inStrain. Source data are provided as a Source Data file.



Supplementary Fig. 19 | Ecological assembly mechanisms of ARGs and their host communities. **a**, Neutral community model (NCM) illustrating the relationship between mean relative abundance and detection frequency of ARG subtypes. Dashed lines indicate the predicted occurrence and 95% confidence intervals; points are colored based on their fit to the model (Above, Below, or Mid). The inset donut chart summarizes the proportions of these three categories. **b**, Relative contributions of deterministic and stochastic ecological processes inferred from null model analysis. The donut chart quantifies the influence of heterogeneous selection, homogeneous selection, dispersal limitation, homogenizing dispersal, and drift. **c**, Density distribution of βNTI values for ARG-hosting microbial communities. Dashed lines indicate the βNTI thresholds (−2 and 2) used to distinguish deterministic from stochastic assembly processes. **d**, Ecological assembly processes inferred from the integrated βNTI and RCbray framework. Colors denote different assembly mechanisms, including heterogeneous selection, homogeneous selection, dispersal limitation, homogenizing dispersal, and drift. Source data are provided as a Source Data file.

References

1. Hernández F, Sancho JV, Ibáñez M, Guerrero C. Antibiotic residue determination in environmental waters by LC-MS. *Trac-Trends in Analytical Chemistry* **26**, 466-485 (2007).
2. Gros M, Rodríguez-Mozaz S, Barceló D. Rapid analysis of multiclass antibiotic residues and some of their metabolites in hospital, urban wastewater and river water by ultra-high-performance liquid chromatography coupled to quadrupole-linear ion trap tandem mass spectrometry. *J Chromatogr A* **1292**, 173-188 (2013).
3. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**, e1 (2013).
4. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: A versatile open source tool for metagenomics. *Peerj* **4**, e2584 (2016).
5. Zhou Y, Liu YX, Li X. USEARCH 12: Open-source software for sequencing analysis in bioinformatics and microbiome. *Imeta* **3**, e236 (2024).
6. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**, 927-930 (2003).
7. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y*, et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**, D633-642 (2014).
8. Yin XL, Zheng XW, Li LG, Zhang AN, Jiang XT, Zhang T. ARGs-OAP v3.0: Antibiotic-resistance gene database curation and analysis pipeline optimization. *Engineering-Prc* **27**, 234-241 (2023).
9. Yin XL, Chen X, Jiang XT, Yang Y, Li B, Shum MHH*, et al.* Toward a universal unit for quantification of antibiotic resistance genes in environmental samples. *Environ Sci Technol* **57**, 9713-9721 (2023).
10. Berendonk TU, Manaia CM, Merlin C, Fatta-Kassinos D, Cytryn E, Walsh F*, et al.* Tackling antibiotic resistance: The environmental framework. *Nat Rev Microbiol* **13**, 310-317 (2015).
11. Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics* **25**, 1096-1098 (2009).
12. Ma LP, Li AD, Yin XL, Zhang T. The prevalence of integrons as the carrier of antibiotic resistance genes in natural and man-made environments. *Environ Sci Technol* **51**, 5721-5728 (2017).

13. Lévesque C, Piché L, Larose C, Roy PH. PCR mapping of integrons reveals several novel combinations of resistance genes. *Antimicrob Agents Ch* **39**, 185-191 (1995).
14. De Coster W, Rademakers R. NanoPack2: Population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**, btad311 (2023).
15. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics* **11**, 119 (2010).
16. Ma LP, Xia Y, Li B, Yang Y, Li LG, Tiedje JM, Zhang T. Metagenomic assembly reveals hosts of antibiotic resistance genes and the shared resistome in pig, chicken, and human feces. *Environ Sci Technol* **50**, 420-427 (2016).
17. Parnanen K, Karkman A, Hultman J, Lyra C, Bengtsson-Palme J, Larsson DGJ*, et al.* Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nat Commun* **9**, 3891 (2018).
18. Zhang Z, Zhang Q, Wang T, Xu N, Lu T, Hong W*, et al.* Assessment of global health risk of antibiotic resistance genes. *Nat Commun* **13**, 1553 (2022).
19. Jiang X-T, Yu K, Li L-G, Yin X-L, Li A-D, Zhang T. AQMM: Enabling absolute quantification of metagenome and metatranscriptome. *bioRxiv*, 218347 (2017).
20. Yin X, Yang Y, Deng Y, Huang Y, Li L, Chan LYL, Zhang T. An assessment of resistome and mobilome in wastewater treatment plants through temporal and spatial metagenomic analysis. *Water Res* **209**, 117885 (2022).
21. Olm MR, Crits-Christoph A, Bouma-Gregson K, Firek BA, Morowitz MJ, Banfield JF. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat Biotechnol* **39**, 727-736 (2021).
22. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
23. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO*, et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
24. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**, 732-740 (2006).
25. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka A. Quantifying community assembly processes and identifying features that impose them. *The ISME Journal* **7**, 2069-2079 (2013).