nature portfolio

Corresponding author(s): Liping Ma, Xu-Xiang Zhang

Last updated by author(s): Jul 2, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No special software was used to collect data.
Data analysis	Sequencing data underwent quality control using fastp (v0.23.2) for Illumina reads and chopper (v0.8.0) for Nanopore reads. Amplicon data of 16S rRNA genes were processed and clustered using Vsearch (v2.22.1) and Usearch (v11.0) to determine microbial community composition. Metagenomic assembly was carried out with MEGAHIT (v1.1.3), and high-quality MAGs were reconstructed using the MetaWRAP pipeline (v1.3.2) integrating MetaBAT2, CONCOCT, and CheckM for binning and refinement. The recovered MAGs were subsequently dereplicated using dRep (v3.3.0) with the parameters '-sa 0.95 -nc 0.30'. Taxonomic classification and gene prediction were performed with GTDB-Tk (v2.3.2) and Prodigal (v2.6.3), while functional and virulence gene annotation used DIAMOND (v2.0.15) against the NCBI-NR and VFDB databases. Antibiotic resistance genes (ARGs) and mobile genetic elements (MGEs) were identified using UBLAST against SARG (v3.0) and BLASTP against the MobileGeneticElement database, respectively, and potential horizontal gene transfer (HGT) events were inferred with WAAFLE (v1.0.0). MetaCHIP (v1.10.6) was used to identify the primary bacterial taxa serving as donors and recipients in ARG-associated HGT events. Competitive read mapping and population genomic analyses were performed using inStrain (v1.10.0) to estimate genome wide coverage and breadth of pathogen associated MAGs across samples. Neutral community modeling (NCM) and phylogeny based null model analyses were used to evaluate ecological assembly processes shaping ARG and ARG host communities. Statistical analyses and visualizations were conducted in R (v4.3.3) using the ggplot2, vegan, and plspm packages. The custom R code for statistical analysis and data visualization for specific analyses are publicly available in the GitHub repository: https://github.com/once0728/eco-connectivity-AMR-drinking-water.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw metagenomic sequencing data generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under accession number PRJNA1288567.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No human participants were recruited for this study.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study investigates the roles and mechanisms of ecological connectivity in driving anthropogenic antimicrobial resistance (AMR) in drinking water and mitigation strategies. We conducted the systematic investigation into the migration of anthropogenic AMR pollution across interconnected urban water compartments (wastewaters, river waters, drinking water reserviors and tap waters) in a megacity, combining metagenomics and culturomics. Our study identified 1,309 antibiotic resistance genes (ARGs), and tracking their dynamics across microbial communities and fecal Enterobacteriaceae isolates (n = 3,040) demonstrated that ecological connectivity establishes a cascading contamination pathway: wastewater discharge promotes AMR accumulation in natural water bodies, facilitating its persistence and enrichment in finished drinking water. Critical human-derived ARGs, primarily conferring resistance to beta-lactams, aminoglycosides, and polymyxins, were notably enriched in clinically relevant pathogens. Further analysis clarified complex synergistic effects of biotic and abiotic drivers, including horizontal gene transfer (HGT), bacterial host proliferation, trace metals, disinfectants, and antibiotic residues, modulated by connectivity to jointly drive ARG proliferation. Novel mechanistic insights showed that integron-mediated HGT captures and rearranges exogenous ARGs, thereby assembling multi-resistant genetic determinants along connected water pathways. To enable risk assessment and targeted source control, we established a risk prioritization framework integrating dynamic patterns, mobility potential, host pathogenicity and clinical relevance to quantify health risks and identify high-risk anthropogenic ARGs. Our results elucidate the mechanistic underpinnings of AMR transmission enabled by ecological connectivity, providing a scientific foundation for targeted interventions to disrupt critical transmission links and mitigate health risks in urban water systems.

Research sample	The research samples consisted of water and sludge collected from key nodes of the urban water system in Shanghai, China, a megacity with a population of approximately 26 million. The sampling framework included: influent, activated sludge, and effluent from four full-scale wastewater treatment plants (WWTPs) employing different treatment processes; water collected upstream, downstream, and at the discharge outlet of the rivers receiving WWTP effluent; water from two main source water reservoirs; and tap water from six city districts. Samples were collected in biological triplicates over a 24 h period, and subsequently mixed as a representative composite sample. For tap water, microorganisms were collected by filtering over 2,000 L using a domestic reverse osmosis purification device operated at a constant flow rate of 1.5 $L \cdot min^{-1}$ for approximately 24 h to ensure sufficient biomass for metagenomic and genome-resolved analyses. Total DNA was extracted from each sample for metagenomic sequencing, and lactose-fermenting Enterobacteriaceae (LFE) strains were isolated from each sample for both antibiotic susceptibility testing (phenotypic analysis) and metagenomic sequencing (genotypic analysis).
Sampling strategy	The sampling strategy was designed to comprehensively cover key components of urban water system (UWS), including four wastewater treatment plants (WWTPs, including influent, activated sludge and effluent) utilizing different treatment processes with total capacity of 375,000 m^3/day and their receiving waters (upstream, discharge outlet and downstream), two primary reservoirs, and household taps from six major districts, to catalog the profile and trace the fate of ARGs across the urban water pollution pathways. Meanwhile, we isolated 3,040 lactose-fermenting Enterobacteriaceae (LFE) isolates to investigate fecal contamination-associated resistome dynamics. Biotic factors including bacterial host of ARGs and HGT mechanisms, as well as abiotic factors including water quality parameters, antibiotics, trace metals, disinfectants, and non-antibiotic pharmaceuticals, were examined to identify key drivers of resistome proliferation and dissemination in UWS. According to the observed dynamics and critical affecting factors, a comprehensive framework of AMR risk assessment within urban water pollution pathways was established to unveil potential health risks associated with the dissemination of anthropogenic ARGs.
Data collection	Data collection was performed firsthand by Ms. Yuting Qi, Dr. Xingxing Zhang, and Mr. Huafeng Liu. The data collection process comprised field sampling, in situ environmental measurements, laboratory analyses, microbial isolation, and high-throughput sequencing. Field sampling followed a predefined sampling strategy covering the major compartments of the urban water system, including wastewater treatment plants, receiving rivers, source water reservoirs, and tap water distribution networks. During sampling, physicochemical parameters, including temperature (Temp), pH, and dissolved oxygen (DO), were measured and recorded on site using portable meters in dedicated, bound field laboratory notebooks. Additional water quality parameters, including chemical oxygen demand (COD), total phosphorus (TP), ammonia nitrogen (NH3-N), and trace metals, were determined in the laboratory following standard analytical methods, with raw data logs manually cataloged and cross-verified. Microbial biomass was collected by membrane filtration, followed by DNA extraction and metagenomic and 16S rRNA amplicon sequencing using Illumina and Nanopore platforms. In addition, lactose-fermenting Enterobacteriaceae (LFE) isolates were recovered for antibiotic susceptibility testing. All primary data were systematically recorded via physical experimental logs, curated into standardized digital spreadsheets, and double-checked by the research team prior to downstream statistical and bioinformatic analyses.
Timing and spatial scale	Samples were collected in June 2023 in Shanghai, an international megacity with a population of approximately 26 million, encompassing multiple WWTPs, river systems, drinking water reservoirs, and household tap waters.
Data exclusions	1. Low-quality Sequencing Data: Raw metagenomic reads and Nanopore sequencing data were processed using fastp (v0.23.2) and chopper (v0.8.0), respectively, to remove low-quality sequences, adapters, and short reads. 2. Low-quality Genomes: Only high-quality metagenome-assembled genomes (MAGs), defined as completeness ≥75% and contamination ≤10% according to CheckM, were retained for downstream analysis.
Reproducibility	All experiments were performed using standardized and previously validated protocols with appropriate biological replicates. Reproducibility was ensured across independent sampling sites and laboratory procedures, including DNA extraction, sequencing, and bacterial isolation. All attempts to repeat the experiments were successful, and consistent results were obtained across replicates, with no failed repetitions observed.
Randomization	To minimize selection bias, a randomization strategy was employed when selecting LFE isolates for sequencing from each sample site. This approach ensured that each colony had an equal chance of being chosen, thereby enhancing the representativeness and ecological diversity of the dataset.
Blinding	Blinding was implemented during the phenotypic screening of lactose-fermenting Enterobacteriaceae (LFE) isolates. Specifically, samples were assigned randomized laboratory codes, and investigators performing colony enumeration and antibiotic susceptibility testing were blinded to the sample origin until the phenotypic data matrix was finalized and recorded.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	No plants were involved in this study.
Novel plant genotypes	Not applicable
Authentication	Not applicable