communications earth & environment

Article

A Nature Portfolio journal

https://doi.org/10.1038/s43247-025-02139-x

Profiles of antibiotic resistome risk in diverse water environments

Yiwen Yang [1,6] ✉, Shuang Cai[2,6], Chunhao Mo[3], Junjie Dong [4], Sheng Chen[5] & Zhiguo Wen [1] ✉

The water environment is considered to be a reservoir for antibiotic resistant bacteria and antibiotic resistance genes. However, profiles of antibiotic resistome risk in different water environments remain largely underexplored. Here we found that the number, abundance and risk of antibiotic resistance genes in wastewater, especially slaughterhouse wastewater, were higher than those in natural water. Then, 6167 high-quality metagenome-assembled genomes were obtained. The main hosts were *Escherichia*, *Desulfobacter*, *Citrobacter* and *Pseudomonas_E*, respectively. Moreover, distinct of patterns of horizontal gene transfer were observed in different microbes. Overall, microbial composition and resistance risk varied in different water environments and there was a correlation between microbial composition and resistome risk. Therefore, models based on microbial composition were constructed with an accuracy of 86.87 ± 1.18% for predicting the risk of resistance in unknown water environments, providing an essential reference for dealing with antibiotic-resistant pollution and for water management.

Water is the most abundant substance on Earth and is essential for sustaining life. All living beings, whether plants, animals, or humans, rely on water for survival. We are surrounded by different water environments, such as freshwater sources, wastewater treatment plants, and water ecosystems. The water quality of these environments is directly related to environmental safety and public health. Understanding the relationships between these different water environments and public health is critical for ensuring a safe water supply and preventing waterborne diseases.

The presence of antibiotic-resistant microorganisms in the water environment has become a growing concern. The emergence and spread of antibiotic resistance pose a major threat to public health, limit the effectiveness of antibiotic therapy, and increase the risk of infection. Studies have shown that different levels of antimicrobial resistance exist in different water environments. Freshwater sources are important water sources for humans; however, many antibiotic residues and antibiotic resistance genes (ARGs) have been identified in recent years[1–3]. Wastewater treatment plants are a key part of the treatment of wastewater, but studies have shown that antibiotic and antibiotic-resistant microorganisms present in wastewater treatment plants can enter the environment[4,5], posing a potential risk to water ecosystems and public health. In addition, water ecosystems, such as rivers, lakes, and marine, are important reservoirs of antibiotic-resistant microorganisms[2,6–8], which may be transmitted through the food chain and water cycle. For example, ARGs can travel through different water environments, such as via surface runoff and groundwater flow, and can ultimately enter the human body, including through drinking water.

The spread of antimicrobial resistance is a complex process involving multiple factors, including microbial communities, residual antibiotics, heavy metals, pesticides, mobile genetic elements (MGEs), and physicochemical factors[9,10]. ARGs can be transmitted in the water environment through a number of pathways, including direct contact in the water, movement, and migration of microorganisms in the water, and interactions between microorganisms in the water and their hosts[11–13]. In addition, human activities, such as the discharge of hospital wastewater, municipal wastewater, and livestock wastewater, play an important role in the spread of antibiotic resistance[14–17]. Together, these factors accelerate the spread of ARGs and resistant microorganisms and pose a potential threat to public health.

Despite the growing interest in the issue of antibiotic resistance in the water environment, there is still a lack of studies and research in this area. In particular, there is a dearth of comparative studies on antibiotic resistome risk in different water environments. In this study, the profiles of antibiotic resistome risk in different water environments were investigated.

[1]Key Laboratory of Feed Biotechnology of Ministry of Agriculture and Rural Affairs, Institute of Feed Research, Chinese Academy of Agricultural Sciences, Beijing, China. [2]State Key Laboratory of Animal Nutrition and Feeding, Ministry of Agriculture and Rural Affairs Feed Industry Center, China Agricultural University, Beijing, China. [3]Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, State Key Laboratory of Swine and Poultry Breeding Industry, College of Animal Science, South China Agricultural University, Guangzhou, China. [4]School of Software, Dalian University of Technology, Dalian, China. [5]State Key Lab of Chemical Biology and Drug Discovery and the Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hong Kong, China. [6]These authors contributed equally: Yiwen Yang, Shuang Cai. ✉e-mail: yangyiwen@caas.cn; wenzhiguo@caas.cn

Understanding the variation in the antibiotic resistome in different water environments is essential for developing effective strategies to address resistome risk in water environments and to protect public health.

Materials and methods

Sample collection

A total of 283 water metagenomic samples, including 9 drinking water samples, 58 groundwater samples, 42 marine samples, 46 municipal wastewater samples, 36 slaughterhouse wastewater samples, 32 swine wastewater samples, 20 duck wastewater samples, and 40 hospital wastewater samples, were analyzed (Supplementary Table 1). Among them, the natural water included drinking water, groundwater, and marine water. Wastewater included municipal wastewater, slaughterhouse wastewater, swine wastewater, duck wastewater, and hospital wastewater.

Among them, samples of swine wastewater and duck wastewater were collected in this study. Water samples were collected, pre-treated, and high throughput sequenced with reference to our previous study[10]. These wastewater samples were preserved in sterile plastic bottles, kept at low temperatures, and transported to the laboratory for pretreatment within 12 h. The wastewater was passed through a 0.22 filter membrane, which was used for DNA extraction and metagenomic sequencing.

Metagenomic assembly and bacterial composition analysis

The quality of the metagenome sequence was controlled using Trimmomatic v_0.33 (SLIDINGWINDOW:4:20, MINLEN:50)[18]. The clean sequencing data were assembled using MEGAHIT (v1.2.9) software with the default parameters (k-lists 21, 29, 39, 59, 79, 99, 119 and 141)[19]. Next, Prodigal v2.63 was used to predict open reading frames (ORFs)[20]. CD-HIT (v4.8.1) was used to cluster ORFs (-c: 0.95, -aS: 0.90)[21]. Salmon software was used to calculate gene abundance with default parameters. Salmon combines a new dual-phase parallel inference algorithm and feature-rich bias models with an ultra-fast read mapping procedure. The bacterial composition of the different samples was analyzed with Kraken2 software[22]. Plasmid sequences were identified using the PLASME tool.

Gene annotation and antibiotic resistome risk analysis

DIAMOND (v2.0.15.153) was used to identify antibiotic resistance genes (ARGs), mobile genetic elements (MGEs) and virulence factors (VFs) based on the Comprehensive Antibiotic Resistance Database (CARD v3.2.5), mobile orthologous groups database (mobileOG-db 1.6 v1) and virulence factor (VF) database (VFDB_setA, Oct 7, 2022) with the following parameters: -e 1e-6, --query-cover 70, and --id 60[23–26]. ARG-OAP v3.2 was used to calculate the abundance of ARGs, MGEs, and VFs (normalized to copies per prokaryotic cell number)[27]. The MetaCompare pipeline was used to identify the number of contigs carrying ARGs, MGEs, and pathogen sequences[28]. According to the number and proportion of these contigs, the resistome risk of each sample was calculated using the MetaCompare pipeline.

Metagenome binning

Contigs longer than 2000 kb were selected for metagenomic binning. Metagenome-assembled genomes (MAGs) were generated using MetaBAT2, MaxBin, and CONCOCT in the MetaWRAP pipeline[29]. CheckM software was used to assess the completeness and contamination of these MAGs[30]. The MAGs with completeness greater than 50% and contamination less than 10% were used for subsequent analysis. The taxonomy of these MAGs was classified based on the Genome Taxonomy Database (GTDB, R214, April 2023) using GTDB-Tk (v 2.3.2)[31]. The number of ARGs in these MAGs was identified using DIAMOND software. MetaCHIP pipeline was used to analyze horizontal gene transfer.

Machine learning for predicting resistome risk

The resistome risk score (MetaCompare) and microbial composition (Kraken2) of the 283 water samples were used for machine learning. First, the missing value of microbial abundance was input to 0. Then the microbial abundance was normalized. The outlier of resistome risk scores was removed. Five models were used for training, including Decision Tree, KNN, Naive Bayes, Random Forest, SVM, and XGBoost. Performance measures were obtained through five-fold cross-validation of the training data.

Data analysis and presentation

The data were prepared in the WPS office (v12.1). SPASS (v22.0) was used for significance and correlation analyses. R software (v4.2.2) was used for plotting the PCoA and heatmap. A phylogenetic tree of MAGs was constructed with Interactive Tree Of Life (iTOL, v6.74)[32]. GraphPad Prism 8 software was used for violin, stacked bar, and column bar plots. Adobe Illustrator 22.1 was used for the graphic layout. Figures 6A and 7 were created in BioRender (Yang, Y.,2025, https://BioRender.com/o45u108).

Results

Profile of antibiotic resistance genes in different water environments

A total of 283 metagenomic samples from different water environments, including natural water and wastewater, were collected. The natural water used was drinking water, groundwater, and marine. Wastewater included municipal wastewater, slaughterhouse wastewater, swine wastewater, duck wastewater, and hospital wastewater. Significant differences were found in the ARG composition among the different water environments (Fig. 1A) ($R = 0.6674$, $P = 0.001$). The number of subclasses of ARGs detected in wastewater was significantly higher than the number of subclasses of ARGs detected in natural water (82–534 and 2–145, respectively) (Fig. 1B) ($P < 0.01$). Differences in the number of ARGs detected between different wastewaters and between different natural waters were not significant ($P > 0.05$). The highest abundance of ARGs was detected in slaughterhouse wastewater (9.98 ± 1.04 copies/cell number), and the lowest abundance of ARGs was detected in marine (0.157 ± 0.013 copies/cell number) (Fig. 1C). The total abundance of ARG in municipal wastewater was not significantly different from that in natural water. The ARG abundance in the other wastewater was significantly higher than that in the natural water ($P < 0.01$). The ARG abundances in the municipal wastewater, drinking water, and groundwater were 0.767 ± 0.113, 0.705 ± 0.210, and 0.443 ± 0.171 copies/cell number, respectively. The ARG abundances in duck wastewater, swine wastewater, and hospital wastewater were 3.762 ± 0.677, 2.104 ± 0.257, and 2.521 ± 0.208 copies/cell number, respectively. The abundance units of ARGs were then transformed to Transcripts Per Million (TPMs) and the same results were found, with the highest abundance of ARGs in the slaughterhouse wastewater (Supplementary Data 1). The dominant ARG types varied in different water environments (Fig. 1D), with multidrug resistance genes dominating in groundwater, municipal wastewater, slaughterhouse wastewater, and hospital wastewater. The dominating The dominant ARG types in marine and duck wastewater environments were aminoglycoside and multidrug resistance genes. The dominant ARG types in swine wastewater were aminoglycoside, MLS, and tetracycline resistance genes. The dominant ARG types in drinking water were rifamycin and phosphonic acid resistance genes. These ARGs were mainly attributed to antibiotic efflux, antibiotic inactivation, and antibiotic target alteration classes of resistance genes (Supplementary Fig. 1). These results indicate that the composition of ARGs in different water environments varies significantly and that the abundance and number of ARGs in wastewater (especially slaughterhouse wastewater) are significantly higher than those in natural water (ANOSIM).

Antibiotic resistome risk in different water environments

Then, we assessed the ARG risk in different water environments in terms of their mobility and correlation with pathogens. The abundance of MGEs was found to be highest in slaughterhouse wastewater (96.90 ± 8.57 copies/cell number), which was significantly higher than that in other water environments ($P < 0.01$) (Fig. 2A). This was followed by hospital wastewater, duck wastewater, drinking water, swine wastewater, municipal wastewater and

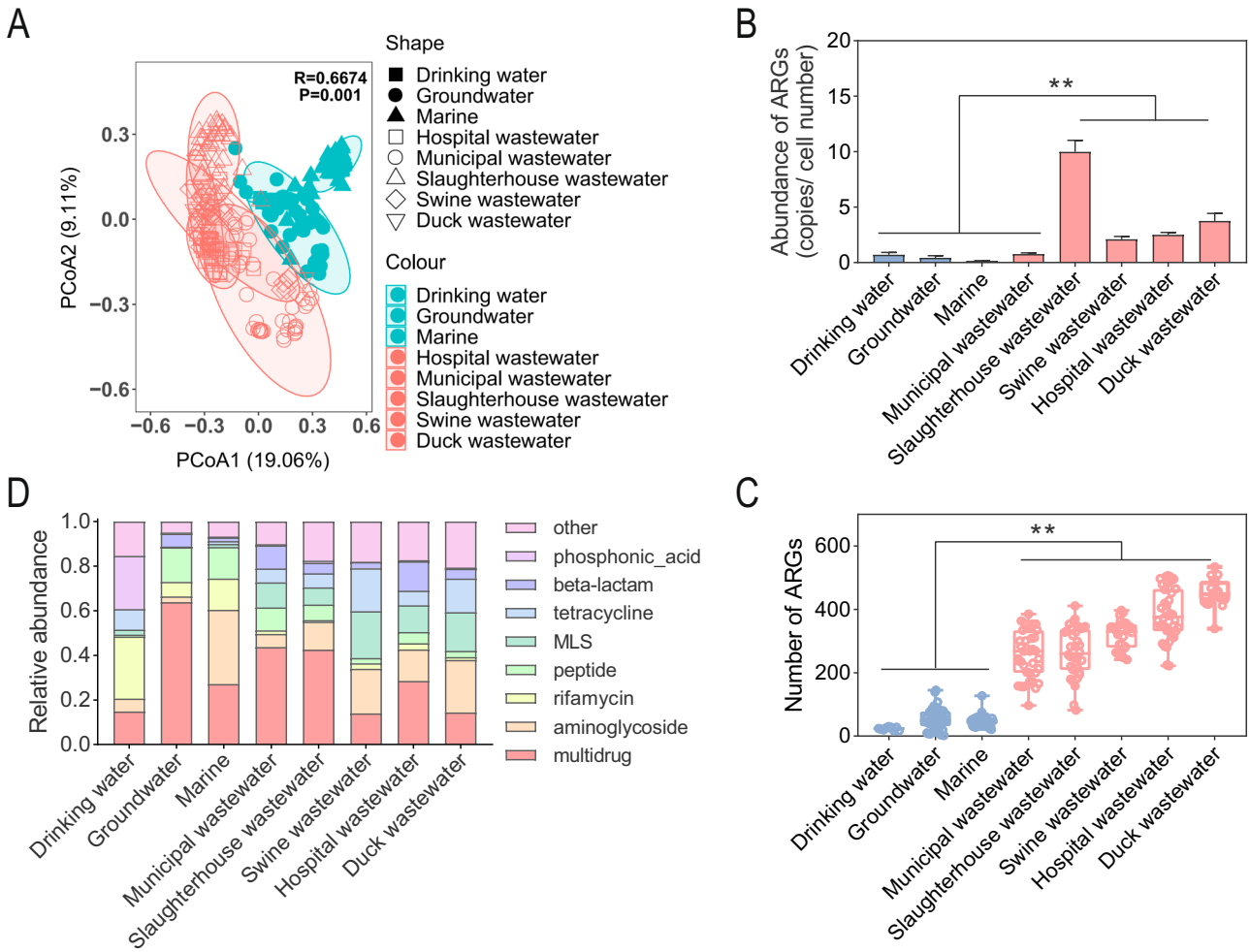

Fig. 1 | Comparison of ARG profiles in different water environments. A PCoA of ARGs in water samples. **B** Abundance, **C** number, and **D** composition of ARGs. * and ** indicate *P* values less than 0.05 and 0.01, respectively. The upper, middle, and lower lines of the box plot indicate the 25% value, the mean value, and the 75% value, respectively. All bar graphs present the mean ± standard deviation.

groundwater, which had MGE abundances of 65.62 ± 6.55, 56.90 ± 7.01, 39.63 ± 11.56, 32.39 ± 4.62, 24.04 ± 3.73 and 17.67 ± 1.96 copies/cell number, respectively. The lowest MGE abundance was found in the marine environment with an abundance of 9.74 ± 0.69 copies/cell number. These MGEs were mainly distributed in the Integration/excision, Transfer, and Phage classes of MGEs (Supplementary Fig. 2). And in these water environments, we also detected different types of ARGs in the plasmid sequences, further validating the risk of ARGs transfer in the water environments (Supplementary Data 2). The results for the abundance of VFs were similar to those of MGEs (Fig. 2B). Again, the abundance of VFs was highest in slaughterhouse wastewater, followed by hospital wastewater, duck wastewater, drinking water, swine wastewater, municipal wastewater, and groundwater. The lowest abundance of VFs was found in marine environments. These VFs mainly consisted of adherence, immune modulation, motility, effector delivery system, stress survival, and nutritional/metabolic factor classes of VFs. Correlation analysis of these MGEs and VFs with ARGs revealed that tetracycline, phosphonic acid, MLS, and aminoglycoside resistance genes were positively correlated with the majority of MGEs and VFs (Fig. 2C). Peptide and Rifamycin resistance genes were negatively correlated with most of the MGEs and VFs. Then, we used the MetaCompare pipeline to assess the risk to the resistome based on the abundance, mobility, and pathogenicity of ARGs and found that slaughterhouse wastewater had the highest resistome risk score (59.56 ± 5.91), followed by hospital wastewater, municipal wastewater, swine wastewater, groundwater, duck wastewater, and drinking water (32.54 ± 1.36, 30.61 ± 3.65, 26.87 ± 0.49, 25.73 ± 1.11, 24.66 ± 0.86 and 23.22 ± 1.59, respectively) (Fig. 2D). The lowest resistome risk score (20.27 ± 0.11) was found in marine environments.

Bacterial communities in different water environments

Bacteria are important ARG hosts and the main carriers of ARGs. We analyzed the bacterial communities in different water environments and found that groundwater had the lowest microbial richness index (615–4773), which was significantly lower than that in marine (6050–7081), pig wastewater (6007–7368), duck wastewater (5978–7316) and hospital wastewater (4582–6116) (Fig. 3A). The richness indices in the drinking water and slaughterhouse wastewater were 3590–5945 and 1899–6136, respectively. PCoA of the bacterial communities revealed that the microbial composition in marine differed significantly from that in other water environments (Fig. 3B). The differences in microbial composition in other water environments were not significant. The major phylum in these water environments was Proteobacteria (0.445–0.688), followed by Firmicutes (0.02–0.30), Actinobacteria (0.05–0.29), and Bacteroidetes (0.03–0.22) (Fig. 3C). The relative abundance of Bacteroidetes in marine environments was significantly higher than that in other water environments ($P < 0.05$). The relative abundance of Firmicutes in slaughterhouse wastewater, duck wastewater, and swine wastewater was significantly higher ($P < 0.05$) than that in other water environments. Correlation analysis of bacterial phyla with the

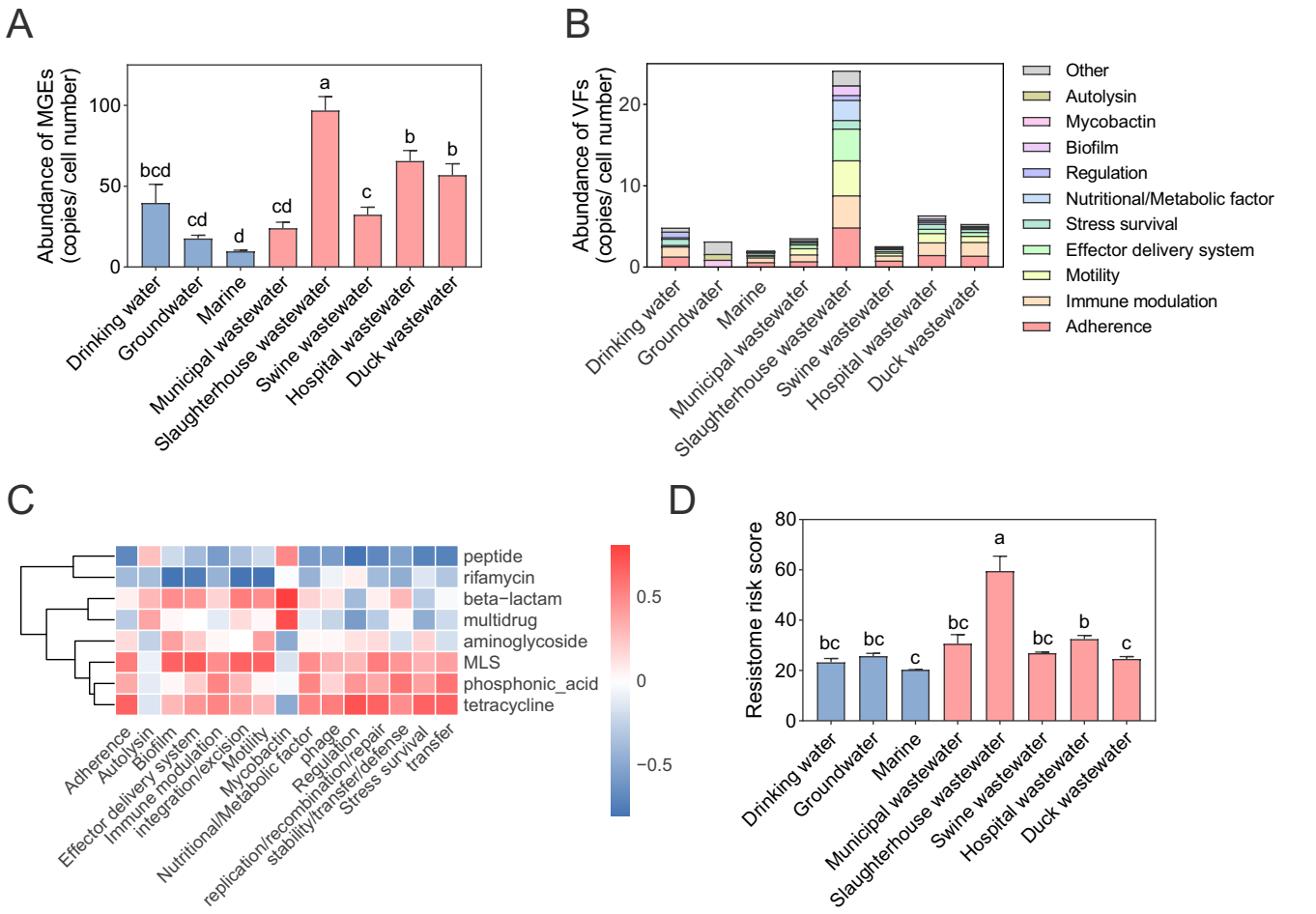

Fig. 2 | Comparison of antibiotic resistome risk in different water environments. The abundance of **A** MGEs and **B** VFs. **C** The correlation between ARGs and MGEs and VFs. **D** Antibiotic-resistance risk scores in different water environments. Different letters indicate significant differences ($P < 0.05$). All bar graphs present the mean ± standard deviation.

abundance of different types of ARGs revealed that Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes were positively correlated with each of the four types of ARGs (Fig. 3D). Proteobacteria were associated with MLS, multidrug, beta-lactam and peptide resistance genes. Firmicutes were positively correlated with tetracycline, MLS, beta-lactam, and aminoglycoside resistance genes. The abundance of Actinobacteria was positively correlated with the ARGs related to tetracycline, MLS, phosphonic acid, and rifamycin-like resistance.

Hosts of antibiotic-resistance genes in different water environments

To further explore the relationship between microbial communities and ARGs, a total of 6167 high-quality MAGs were assembled using metagenomic binning from different water environment samples with greater than 50% genome integrity and less than 10% contamination (Supplementary Data 3). The GC content, N50, and size of these MAGs were 0.217–0.749, 1727–625,897 bp, and 319,964–15,240,531 bp, respectively. A total of 94.61% of them were bacterial genomes ($n = 5861$), and 5.39% were archaeal genomes ($n = 306$) (Fig. 4). The major bacterial phyla were Pseudomonadota, Bacteroidota, Patescibacteria, Bacillota_A, and Actinomycetota. The number of ARGs carried by these MAGs was analyzed, and it was found that 2719 MAGs carried ARGs and that they were considered the main ARG hosts. The profiles of the dominant ARG hosts differed among the different water environments. The most important ARG hosts (those with the 15 most abundant ARGs) in slaughterhouse wastewater all carried more than 20 ARGs. Duck wastewater had 5 major hosts with more than 20 ARGs. Municipal wastewater and hospital wastewater had 2 major hosts with more than 20 ARGs. There was only one host with more than 20 ARGs in the swine wastewater and groundwater. All the marine and drinking water samples had fewer than 20 ARGs. The major ARG hosts in drinking water, groundwater, marine, municipal wastewater, slaughterhouse wastewater, swine wastewater, duck wastewater, and hospital wastewater were *Mycobacterium*, *Escherichia*, *Stutzerimonas*, *Mangrovibacter*, *Citrobacter*, *Escherichia*, *Desulfobacter*, and *Pseudomonas_E*, respectively, which carried 8, 61, 19, 45, 52, 66, 62 and 27 ARGs, respectively. Among these major ARG hosts, *Escherichia*, *Desulfobacter,* and *Citrobacter* carried the highest number of ARGs. In addition, *Pseudomonas_E* was the most widely distributed ARG host in different water environments, and it was the major host in groundwater, municipal wastewater, slaughterhouse wastewater, swine wastewater, and municipal wastewater, carrying 11-38 ARGs. Then, a phylogenetic tree was produced for the MAGs carrying the 10 most abundant ARGs in different water environments (Fig. 5). These major ARG hosts were mainly distributed in Pseudomonadota, followed by Actinomycetota. The ARG hosts of both natural water (drinking water, groundwater, and marine water) and wastewater (municipal wastewater, slaughterhouse wastewater, swine wastewater, and hospital wastewater) were distributed among these phyla. These results suggest that the composition of ARG hosts in different water environments varies, with *Escherichia*, *Desulfobacter*, *Citrobacter,* and *Pseudomonas_E* being the main ARG hosts in water environments.

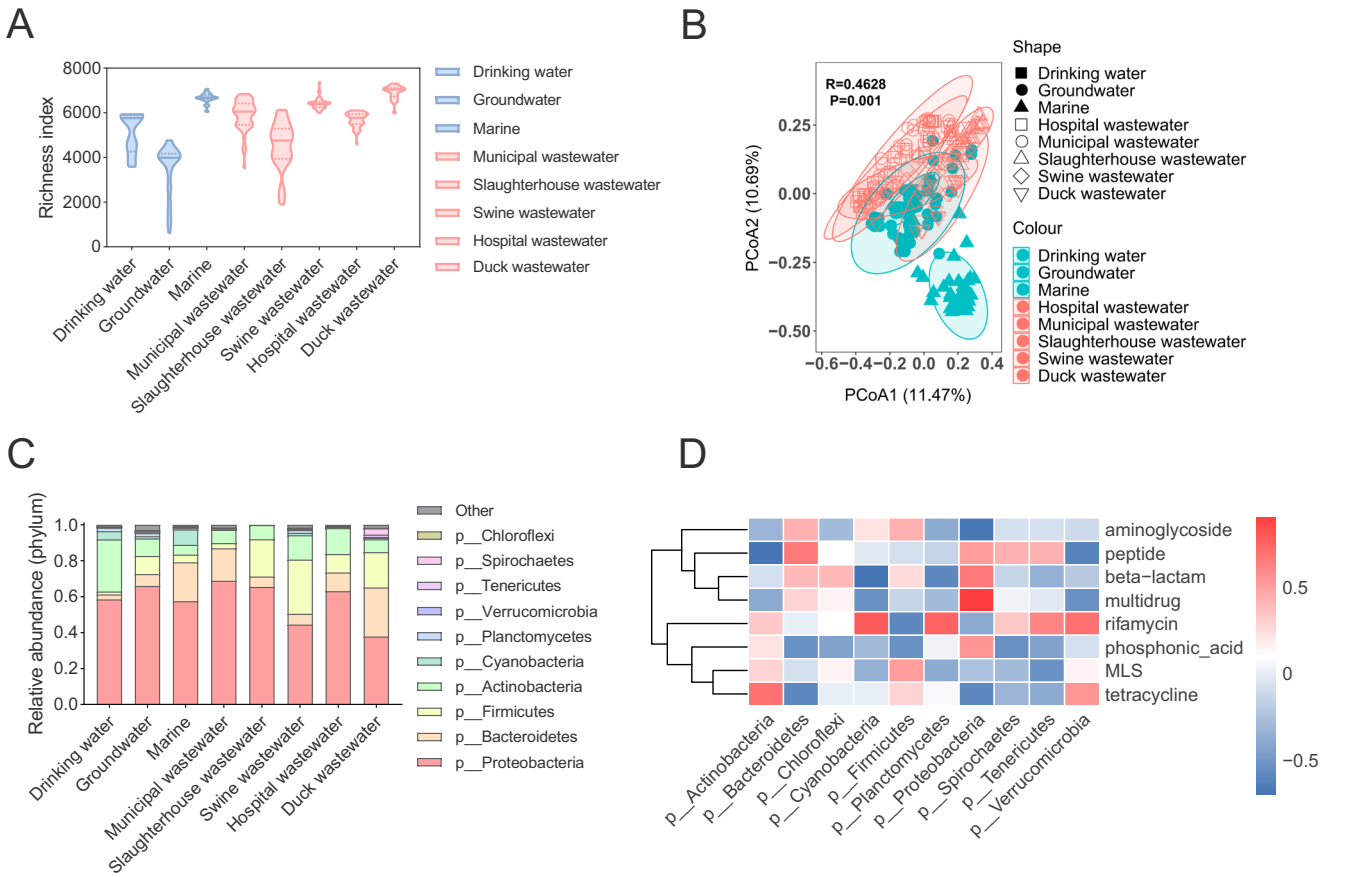

Fig. 3 | Bacterial composition in different water environments. A Richness index, **B** PCoA, and **C** relative abundance of bacterial communities in the water samples. **D** Correlations between bacterial communities and ARGs. The upper, middle, and lower lines of the violin plot indicate the 25% value, the mean value, and the 75% value, respectively.

Horizontal gene transfer analysis of these MAGs revealed different transfer patterns in these different water environments (Supplementary Fig. 3). For example, horizontal gene transfer in slaughterhouse wastewater occurred mainly in Pseudomonadota, in duck wastewater in Bacillota_A, and in the marine in Bacteroidota. This also increases the risk of ARG transfer between different hosts.

Machine learning for predicting resistome risk of water environment

The above results show that there are different resistome risks and ARG hosts present in various water environments. And there is a correlation between microbial communities and resistome risk. It indicates that the microbial composition can represent the resistome risk. Then we used the machine learning method to predict resistome risk based on the microbial composition in the water environment (Fig. 6A). We set the median of resistome risk scores as the standard threshold (Fig. 6B). When the scores are greater than 25, it is high risk; when the score is below 25, it is low risk. Six machine learning methods were used (Fig. 6C and Supplementary Table 2). And XGBoost was found to be the best method with an accuracy was 86.87 ± 1.18%, followed by Random Forest (85.65 ± 1.23%), Decision Tree (82.13 ± 1.33%), Naive Bayes (81.41 ± 1.45%), SVM (70.67 ± 0.48%) and KNN (60.95 ± 0.48%). Their AUCn (Area Under the Curve) were 0.95,0.95,0.8,0.87,0.74 and 0.63, respectively. In summary, the XGBoost model has high accuracy and can be used to predict resistance risk in unknown aquatic environments.

Discussion

In the present study, the resistance profiles of different water environments were investigated. The results showed that the antibiotic resistome risk in wastewater, especially slaughterhouse wastewater, was higher than that in natural water (Fig. 1). The lowest risk was found in marine. Studies found that the abundance of ARGs in wastewater, such as livestock wastewater and municipal wastewater, was generally higher than 10^3–10^8 copies/mL[5,33,34]. The abundance of ARGs in natural water, such as marine, rivers, and groundwater, is generally 10^{-1}–10^4 copies/mL[35–37]. These results are similar to our findings that the wastewater resistome risk was greater than the natural water resistome risk. This may be due to the presence of large amounts of antibiotic residues and bacteria in wastewater, which provides an ideal environment for the development of antibiotic resistance[38,39].

The resistance risk in natural water is lower than in wastewater, which is an expected result. However, there is a need to pay attention to the characteristics of resistome risk in different water environments. Whether it is wastewater or natural water, it is all around us and affects our health. Antibiotic-resistant microorganisms and ARGs in these wastewater may not be completely removed by treatment processes[40], allowing them to enter neighboring natural water environments. In addition, when wastewater is discharged into natural water environments such as rivers, lakes, or marine environments, the resistant microorganisms and ARGs could interact with microorganisms in natural water, leading to the spread of resistance (Fig. 7)[41]. Resistant microorganisms and ARGs may spread to different locations through the water cycle. In addition, ARGs undergo horizontal transfer in different water environments (Supplementary Fig. 3)[42,43], furthering the complexity and risk of drug resistance. Therefore, we need to take measures to reduce the transmission risk of antibiotic resistance in wastewater and to strengthen wastewater

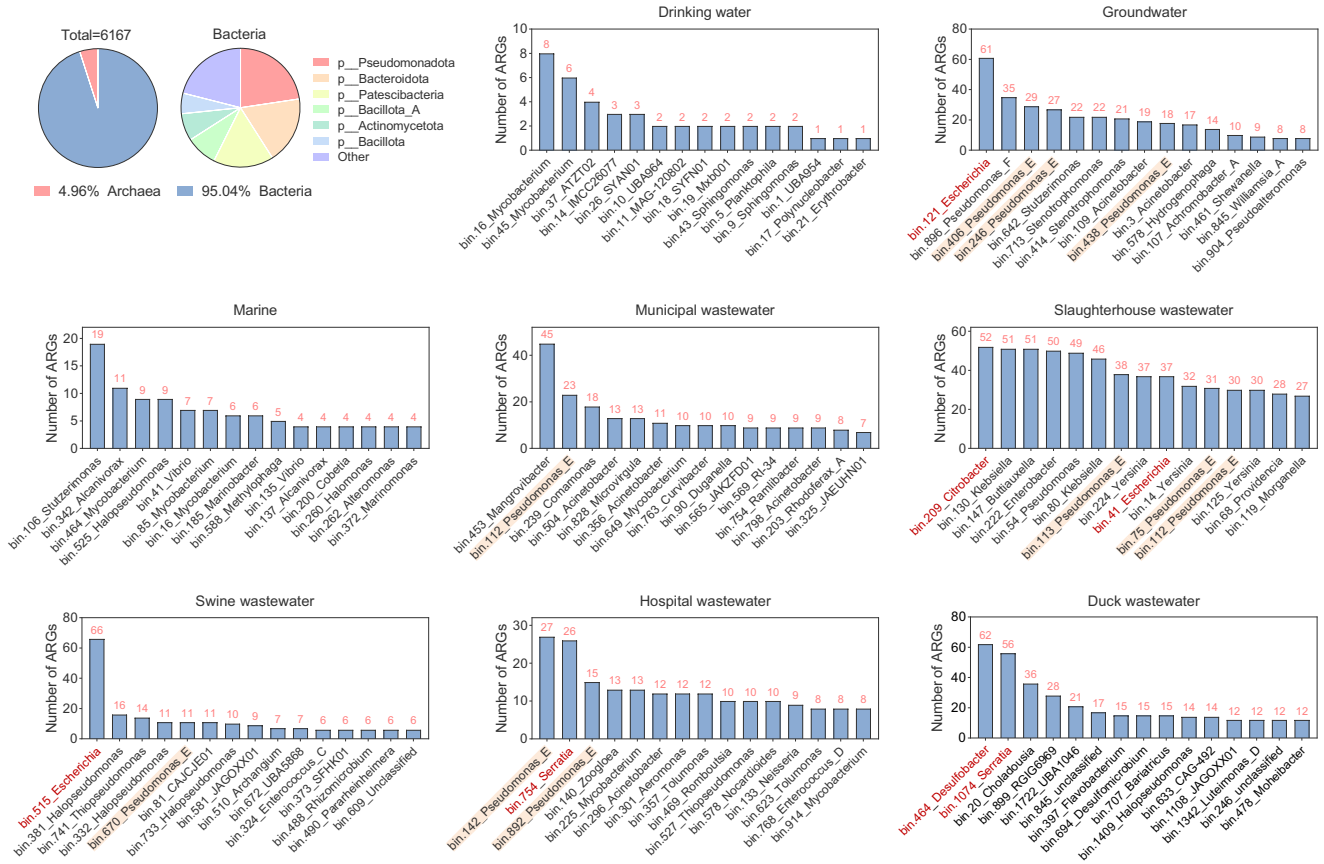

Fig. 4 | Metagenome-assembled genomes from different water samples. The pie chart shows the composition of the metagenome-assembled genomes. The bar chart shows the number of ARGs in the metagenome-assembled genomes.

treatment and monitoring of resistance. Of course, we should also strengthen the monitoring of antibiotic resistance in natural water to avoid the impact of external sources of pollution such as wastewater.

The current methods for monitoring the resistance risk in the environment are mainly traditional methods such as high-throughput sequencing analysis and isolation culture, of which high-throughput analysis requires higher sequencing costs and longer analysis time, and has a high technical threshold. The isolation culture method has a lower cost and shorter time, but it can only analyze very few culturable antibiotic-resistant microorganisms. In recent years, machine learning has been used to predict environmental resistance risk, which is faster and more accurate[44,45]. Studies have shown significant variations in ARG hosts across different environments. For instance, *Flavobacteriales*, *Acinetobacter*, *Pseudomonas,* and *Burkholderiaceae* are the main ARG hosts in river-lake systems[46], while *Enterobacteriaceae* dominate in soil[47], and *Streptococcus* and *Clostridium* in pig farm air[48]. This study revealed different ARG hosts in various water environments, highlighting a correlation between microbial composition and ARGs, suggesting that microbial composition can serve as an indicator of resistome risk. Therefore, we constructed a prediction model using a machine learning method (Fig. 6). This model only needs to input the results of microbial composition in the unknown water sample to predict its antibiotic resistance risk, which can greatly reduce the labor cost and time cost, and its accuracy is up to 86.87 ± 1.18% with good repeatability. This is a highly rewarding result. Of course, in the future, we need to add more data to improve its accuracy for better application in production.

In addition, to address the risk of spreading antibiotic resistance in water, effective preventive and control measures are essential. For example, the construction and management of wastewater treatment facilities should be strengthened to ensure that antibiotic residues and resistant microorganisms in wastewater can be effectively removed. The use of advanced treatment technologies, such as biological treatment and membrane filtration, can effectively reduce the release of resistance-related substances in wastewater. Management and monitoring in agricultural production should be strengthened to rationalize the use of antibiotics and pesticides and ensure that they are not used in excess. Promote sustainable agricultural practices, such as organic farming and eco-agriculture, to reduce the use of chemical substances and radically reduce the development and spread of antibiotic resistance. The public awareness and awareness of the problem of antibiotic resistance should be increased, and the rational use of antibiotics and other drugs should be promoted. Personal hygiene and infection control measures in medical institutions should be strengthened to reduce the abuse and misuse of antibiotics and prevent the further spread of drug resistance. A sound antibiotic resistance monitoring system was established to regularly monitor and assess resistance risk in the water environment. Relevant research should be carried out to gain an in-depth understanding of the transmission mechanisms and influencing factors of antibiotic resistance to provide a scientific basis for the formulation of targeted prevention and control strategies.

Conclusion

In this study, the profiles of antibiotic resistome risk in different water environments were investigated. The number, abundance, and risk score of ARGs were found to be higher in wastewater, especially slaughterhouse wastewater, than those in natural water. The dominant ARG hosts in different water environments were varied. And different patterns of horizontal

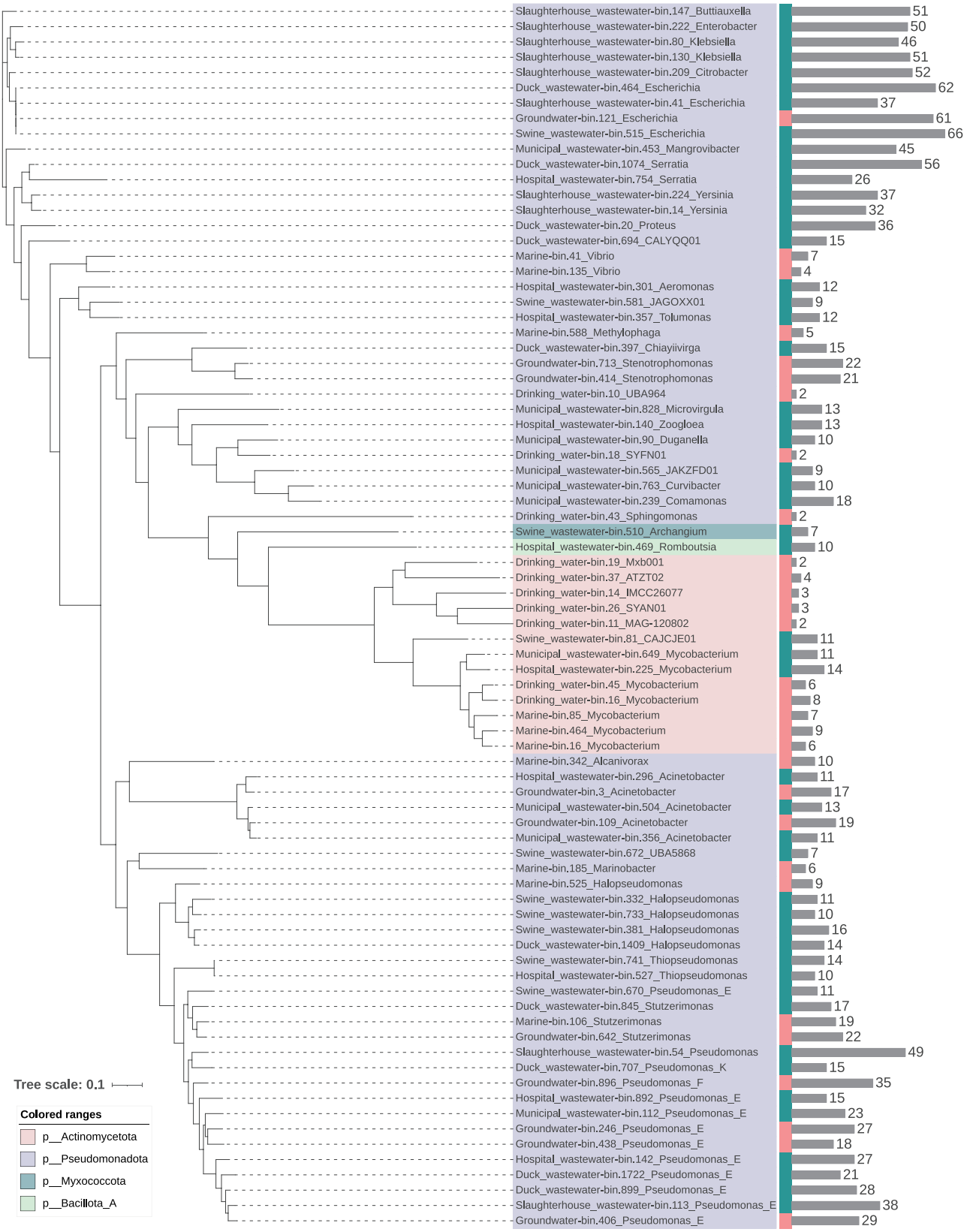

Fig. 5 | The phylogenetic tree of metagenome-assembled genomes. The values are the numbers of ARGs.

gene transfer were observed between different ARG hosts. Overall there is a correlation between microbial composition and resistome risk. Therefore, based on microbial composition, models were constructed with an accuracy of 86.87 ± 1.18% for predicting the risk of resistance in unknown water environments. These findings have important implications for the control of antibiotic resistance and for water management.

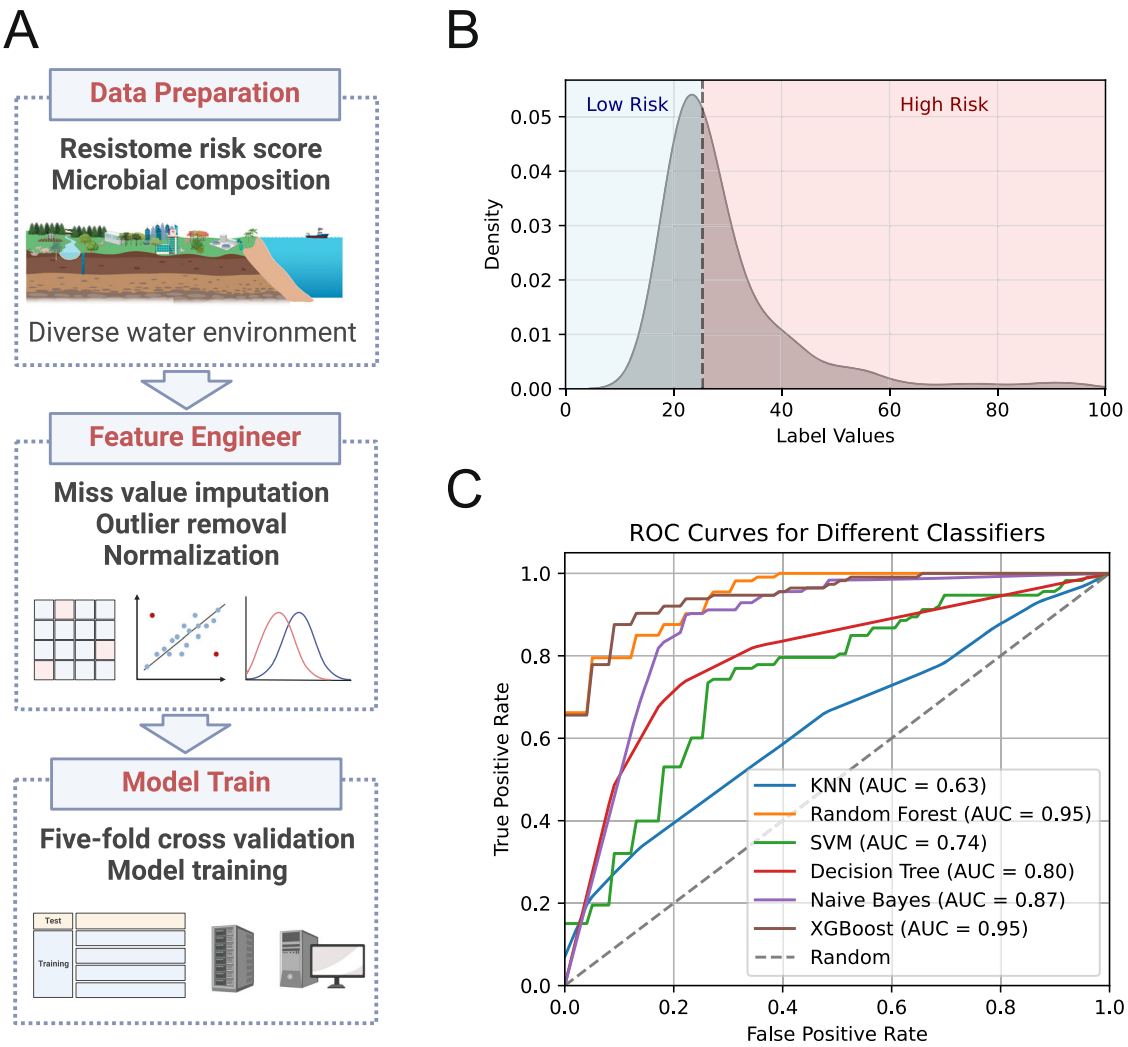

Fig. 6 | Machine learning for predicting resistome risk of water environment. A Workflow of model training. **B** The distribution of resistome risk score in water samples. Label values were the reistome risk scores. **C** Receiver Operating Characteristic (ROC) Curve for different classifiers.

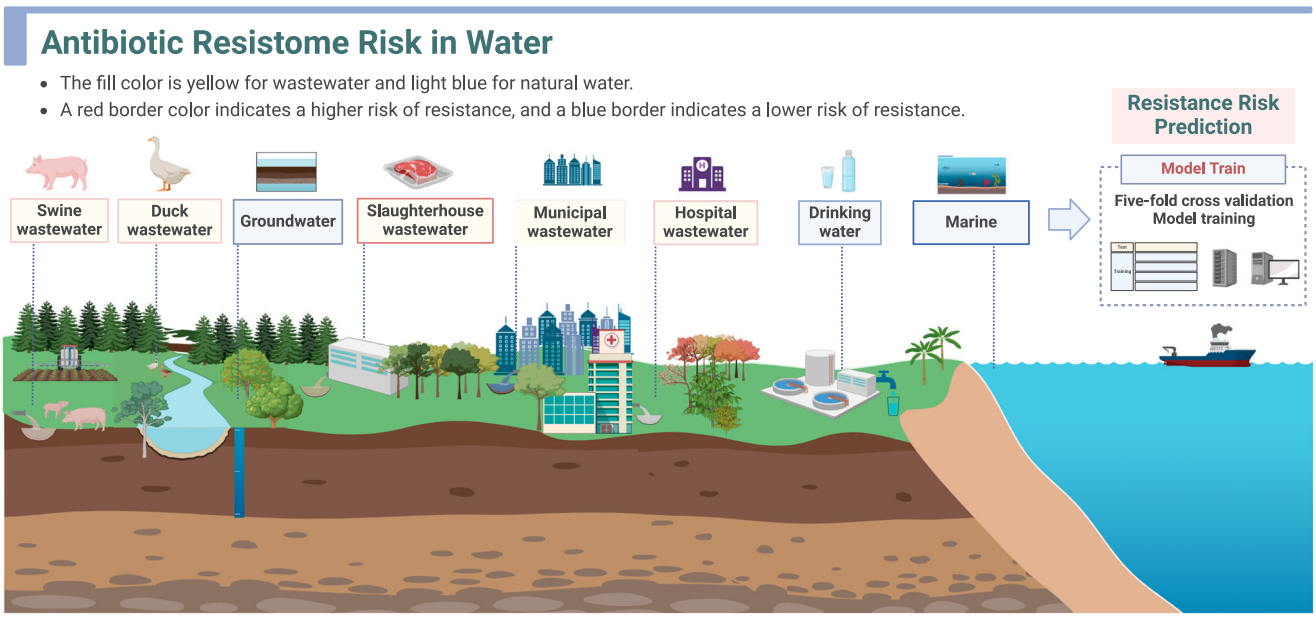

Fig. 7 | Overview diagram of antibiotic resistome risk in different water environments.

Data availability

All the sequencing data have been deposited in The National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) with accession number: PRJNA1050755. And the data of machine learning for resistome risk were available on https://github.com/yyiw/Machine-learning-for-resistome-of-water-environment. Source data used to generate figures for this study are available from https://figshare.com/articles/dataset/dataset_of_water_resistome/28358663?file=52172720.

Received: 4 May 2024; Accepted: 14 February 2025;
Published online: 27 February 2025

References

1. Ke, Y. et al. Antibiotic resistome alteration along a full-scale drinking water supply system deciphered by metagenome assembly: Regulated by seasonality, mobile gene elements and antibiotic resistant gene hosts. *Sci. Total Environ.* **862**, 160887 (2023).
2. Huang, F. et al. Antibiotic resistome in groundwater and its association with mountain springs and river. *Ecotoxicol. Environ. Saf.* **252**, 114603 (2023).
3. Alm, E. W., Zimbler, D., Callahan, E. & Plomaritis, E. Patterns and persistence of antibiotic resistance in faecal indicator bacteria from freshwater recreational beaches. *J. Appl. Microbiol.* **117**, 273–285 (2014).
4. Di Cesare, A. et al. Contribution of plasmidome, metal resistome and integrases to the persistence of the antibiotic resistome in aquatic environments. *Environ. Pollut.* **297**, 118774 (2022).
5. Yang, Y., Liu, Z., Xing, S. & Liao, X. The correlation between antibiotic resistance gene abundance and microbial community resistance in pig farm wastewater and surrounding rivers. *Ecotoxicol. Environ. Saf.* **182**, 109452 (2019).
6. Zhang, Y. et al. Metagenomic insights into resistome coalescence in an urban sewage treatment plant-river system. *Water Res.* **224**, 119061 (2022).
7. Xu, N. et al. A global atlas of marine antibiotic resistance genes and their expression. *Water Res.* **244**, 120488 (2023).
8. Liu, C., Chen, J., Yang, Y., Teng, Y. & Chen, H. Biogeography and diversity patterns of antibiotic resistome in the sediments of global lakes. *J. Environ. Sci.***127**, 421–430 (2023).
9. Yang, Y. et al. Profiles of bacteria/phage-comediated ARGs in pig farm wastewater treatment plants in China: Association with mobile genetic elements, bacterial communities and environmental factors. *J. Hazard Mater.* **404**, 124149 (2021).
10. Yang, Y. et al. The fate of antibiotic resistance genes and their association with bacterial and archaeal communities during advanced treatment of pig farm wastewater. *Sci. Total Environ.* **851**, 158364 (2022).
11. Carlson, J. C. et al. Mechanisms of antimicrobial resistant Salmonella enterica transmission associated with starling–livestock interactions. *Vet. Microbiol.* **179**, 60–68 (2015).
12. Laroche, E., Petit, F., Fournier, M. & Pawlak, B. Transport of antibiotic-resistant Escherichia coli in a public rural karst water supply. J. Hydrol 392, 12–21 (2010).
13. Gothwal, R. & Thatikonda, S. Modeling transport of antibiotic resistant bacteria in aquatic environment using stochastic differential equations. *Sci. Rep.* **10**, 15081 (2020).
14. Aubertheau, E. et al. Impact of wastewater treatment plant discharge on the contamination of river biofilms by pharmaceuticals and antibiotic resistance. *Sci. Total Environ.* **579**, 1387–1398 (2017).
15. Liu, X. et al. Longitudinal monitoring reveals the emergence and spread of blaGES-5-harboring carbapenem-resistant Klebsiella quasipneumoniae in a Hong Kong hospital wastewater discharge line. *Sci. Total Environ.* **903**, 166255 (2023).
16. Wang, B. et al. Distribution and migration of antibiotic resistance genes, as well as their correlation with microbial communities in swine farm and its surrounding environments. *Environ. Pollut.* **316**, 120618 (2023).
17. Gaviria-Figueroa, A., Preisner, E. C., Hoque, S., Feigley, C. E. & Norman, R. S. Emission and dispersal of antibiotic resistance genes through bioaerosols generated during the treatment of municipal sewage. *Sci. Total Environ.* **686**, 402–412 (2019).
18. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
19. Li, D., Liu, C., Luo, R., Sadakane, K. & Lam, T. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
20. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
21. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
22. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. **15**, R46 (2014).
23. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
24. Yang, Y. et al. Pet cats may shape the antibiotic resistome of their owner's gut and living environment. *Microbiome* **11**, 235 (2023).
25. Brown, C. L. et al. mobileOG-db: a manually curated database of protein families mediating the life cycle of bacterial mobile genetic elements. *Appl. Environ. Microbiol*. **88**, e00991-22 (2022).
26. Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–D699 (2023).
27. Yin, X. et al. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics* **34**, 2263–2270 (2018).
28. Oh, M. et al. MetaCompare: a computational pipeline for prioritizing environmental resistome risk. *FEMS Microbiol. Ecol*. **94**, fiy079 (2018).
29. Uritskiy, G. V., Diruggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
30. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
31. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
32. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–W478 (2011).
33. Liang, C. et al. Fates of intracellular and extracellular antibiotic resistance genes during the cattle farm wastewater treatment process. *Bioresour. Technol.* **344**, 126272 (2022).
34. Lee, J. et al. Quantitative and qualitative changes in antibiotic resistance genes after passing through treatment processes in municipal wastewater treatment plants. *Sci. Total Environ.* **605–606**, 906–914 (2017).
35. Gros, M. et al. Antibiotics, antibiotic resistance and associated risk in natural springs from an agroecosystem environment. *Sci. Total Environ.* **857**, 159202 (2023).
36. Lu, J. et al. Fate of land-based antibiotic resistance genes in marginal-sea sediment: territorial differentiation and corresponding drivers. *Chemosphere* **288**, 132540 (2022).

37. Gao, H. et al. Occurrence and distribution of antibiotics and antibiotic resistance genes in water of Liaohe River Basin, China. *J. Environ. Chem. Eng.* **10**, 108297 (2022).
38. Leinert, J. L., Weichert, S., Jordan, A. J. & Adam, R. Non-typhoidal salmonella infection in children: influence of antibiotic therapy on postconvalescent excretion and clinical course—a systematic review. *Antibiotics* **10**, 1187 (2021).
39. Ishikawa, N. K. et al. Determination of tylosin excretion from sheep to assess tylosin spread to agricultural fields by manure application. *Sci. Total Environ.* **633**, 399–404 (2018).
40. Wang, J. L. & Chen, X. Y. Removal of antibiotic resistance genes (ARGs) in various wastewater treatment processes: an overview. *Crit. Rev. Environ. Sci. Technol.* **52**, 571–630 (2022).
41. Huang, F. et al. Removal of antibiotic resistance genes during livestock wastewater treatment processes: review and prospects. *Front. Vet. Sci.* **9**, 1054316 (2022).
42. Yang, Y. et al. Changes in the carbon metabolism of Escherichia coli during the evolution of doxycycline resistance. *Front. Microbiol.* **10**, 2506 (2019).
43. Shao, S. C., Hu, Y. Y., Cheng, J. H. & Chen, Y. C. Research progress on distribution, migration, transformation of antibiotics and antibiotic resistance genes (ARGs) in aquatic environment. *Crit. Rev. Biotechnol.* **38**, 1195–1208 (2018).
44. Sun, Y., Clarke, B., Clarke, J. & Li, X. Predicting antibiotic resistance gene abundance in activated sludge using shotgun metagenomics and machine learning. *Water Res.* **202**, 117384 (2021).
45. Wu, D. et al. Metagenomic and machine learning meta-analyses characterize airborne resistome features and their hosts in china megacities. *Environ. Sci. Technol.* **57**, 16414–16423 (2023).
46. Ding, J. et al. Hydraulic conditions control the abundance of antibiotic resistance genes and their potential host microorganisms in a frequently regulated river–lake system. *Sci. Total Environ.* **946**, 174143 (2024).
47. Zheng, D. et al. Global biogeography and projection of soil antibiotic resistance genes. *Sci. Adv.***8**, 12 (2022).
48. Gao, F. et al. Airborne bacterial community and antibiotic resistome in the swine farming environment: metagenomic insights into livestock relevance, pathogen hosts and public risks. *Environ. Int.***172**, 11 (2023).

Acknowledgements

The study was supported by the National Natural Science Foundation of China (32202734) and Project funded by the China Postdoctoral Science Foundation (2021M700047, 2022T150226). We acknowledge that no specific sampling permissions were required for the sites used in this study, as they are publicly accessible.

Author contributions

Yiwen Yang: Conceptualization, Investigation, Methodology, Writing - original draft, Project administration, Funding acquisition. Shuang Cai: Conceptualization, Investigation, Methodology, Review & editing. Chunhao Mo: Formal analysis, Investigation. Junjie Dong: Methodology, Review & editing. Sheng Chen: Conceptualization, Review & editing. Zhiguo Wen: Conceptualization, Investigation, Review & editing, Project administration.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-025-02139-x.

Correspondence and requests for materials should be addressed to Yiwen Yang or Zhiguo Wen.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Alice Drinkwater. A peer review file is available.