Global Diversity and Distribution of Antibiotic Resistance Genes in Human Wastewater Treatment Systems

Congmin Zhu[1,2,3#], Linwei Wu[3,4#]†, Daliang Ning[3#], Renmao Tian[3,5], Shuhong Gao[3], Bing Zhang[6,7], Jianshu Zhao[8], Ya Zhang[3], Naijia Xiao[3], Yajiao Wang[3], Mathew R. Brown[9], Qichao Tu[10], Global Water Microbiome Consortium[*], Feng Ju[11,12], George F. Wells[13], Jianhua Guo[14], Zhili He[15,16], Per H. Nielsen[17], Aijie Wang[18], Yu Zhang[19], Ting Chen[20], Qiang He[21,22], Craig S. Criddle[23], Michael Wagner[24], James M. Tiedje[25], Thomas P. Curtis[9], Xianghua Wen[6], Yunfeng Yang[6], Lisa Alvarez-Cohen[26,27], David A. Stahl[28], Pedro J. J. Alvarez[29], Bruce E. Rittmann[30], and Jizhong Zhou[3,27,31,32,33]†

[1]School of Biomedical Engineering, Capital Medical University, Beijing, China
[2]Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Beijing, China
[3]Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA
[4]Institute of Ecology, Key Laboratory for Earth Surface Processes of the Ministry of Education, College of Urban and Environmental Sciences, Peking University, Beijing, China
[5]Institute for Food Safety and Health, Illinois Institute of Technology, Bedford Park, IL, USA
[6]State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China
[7]College of Life and Environmental Sciences, Minzu University of China, Beijing, China
[8]Center for Bioinformatics and Computational Biology, School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[9]School of Engineering, Newcastle University, Newcastle upon Tyne, UK
[10]Institute for Marine Science and Technology, Shandong University, Qingdao, China
[11]Key Laboratory of Coastal Environment and Resources of Zhejiang Province, School of Engineering, Westlake University, Hangzhou, Zhejiang, China
[12]Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang, China
[13]Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL, USA

[14]Australian Centre for Water and Environmental Biotechnology, The University of Queensland, Brisbane, Queensland, Australia
[15]Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, China
[16] College of Agronomy, Hunan Agricultural University, Changsha, China
[17]Department of Chemistry and Bioscience, Center for Microbial Communities, Aalborg University, Aalborg, Denmark
[18]Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China
[19]State Key Laboratory of Environmental Aquatic Chemistry, Research Center for Eco-Environmental Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China
[20]Institute for Artificial Intelligence and Department of Computer Science and Technology, Tsinghua University, Beijing, China
[21]Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN, USA
[22]Institute for a Secure and Sustainable Environment, The University of Tennessee, Knoxville, TN, USA
[23]Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA
[24]Department of Microbiology and Ecosystem Science, Division of Microbial Ecology, Research Network 'Chemistry meets Microbiology', University of Vienna, Vienna, Austria
[25]Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA
[26]Department of Civil and Environmental Engineering, College of Engineering, University of California, Berkeley, CA, USA
[27]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[28]Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA
[29]Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA
[30]Biodesign Swette Center for Environmental Biotechnology, Arizona State University, Tempe, AZ, USA
[31]School of Biological Sciences, University of Oklahoma, Norman, OK, USA

[32]School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA
[33]School of Computer Sciences, University of Oklahoma, Norman, OK, USA
*A list of authors and their affiliations appears at the end of the paper.

[#]These authors contributed equally: Congmin Zhu, Linwei Wu, Daliang Ning.
†**Corresponding author:** Linwei Wu, e-mail: linwei.wu@pku.edu.cn; Jizhong Zhou, e-mail: jzhou@ou.edu

Classification: Biological Sciences, Microbiology

Keywords: antibiotic resistance, microbiome, metagenomics, activated sludge, wastewater treatment plants

Abstract

Antibiotic resistance poses a significant threat to human health, and wastewater treatment plants (WWTPs) are important reservoirs of antibiotic resistance genes (ARGs). Here, we analyse the antibiotic resistomes of 226 activated sludge samples from 142 WWTPs across six continents, using a consistent pipeline for sample collection, DNA sequencing and analysis. We find that ARGs are diverse and similarly abundant, with a core set of 20 ARGs present in all WWTPs. ARG composition differs across continents and is distinct from that of the human gut and the oceans. ARG composition strongly correlates with bacterial taxonomic composition, with *Chloroflexi*, *Acidobacteria* and *Deltaproteobacteria* being the major carriers. ARG abundance positively correlates with presence of mobile genetic elements, and 57% of the 1,112 recovered high-quality genomes possess putatively mobile ARGs. Resistome variations appear to be driven by a complex combination of stochastic processes and deterministic abiotic factors.

Introduction

Antibiotic resistance (i.e., the ability of bacteria to survive and replicate in the presence of an antibiotic[1]) poses an increasingly urgent global public health challenge[2]. Many bacterial pathogens have developed resistance to major antibiotics, with some resisting multiple drugs and causing untreatable infections[3, 4]. Owing to the global broad use of antibiotics, antibiotic resistant bacteria (ARB) and their antibiotic resistance genes (ARGs) are emerging and spreading globally among people, food, animals, plants, and environmental compartments; i.e., soil, water, and air[5, 6]. The environment provides an immense gene pool from which numerous ARGs could be acquired by pathogens to resist antibiotics[7]. Since many ARGs are found on mobile genetic elements (MGEs) and are therefore often horizontally transmitted, antibiotic use also imposes a selective pressure on the whole microbiome, not just pathogens.

In addition to studying the acquisition of antimicrobial resistance in pathogens, it is important to examine how antibiotic use and other environmental variables (such as temperature[8], pH[9], gross domestic product (GDP)[10], population density[11]) affect the aggregate collection of resistance genes of commensal microbiomes; i.e., the resistome. Reliable information on the global occurrence and biotic/abiotic drivers of ARGs is urgently needed to inform public health actions and antibiotic-use decisions. Previous studies have reported global maps of resistomes for soil[12], inland water[13], urban mass-transit systems[14], sewage[15], and the human gut[16], providing baseline information for understanding ARG diversity and health risks in the environment.

The sewage of ~52% of the global population is delivered to wastewater treatment plants (WWTPs)[17, 18], an essential infrastructure for the protection of human and ecosystem health[19, 20]. However, WWTPs are among the most important reservoirs of ARGs and ARB because they receive wastewater from homes, hospitals, and pharmaceutical manufacturing facilities. Most WWTPs employ the activated sludge (AS) process, an open aerobic enrichment-culture system of microbial flocs or granules. Different anoxic/aerobic AS variants remove organic carbon, nitrogen, and phosphorus and can function within treatment trains to remove pathogens, micropollutants, and ARB[21-23]. The activated sludge could also be a spawning ground for resistance evolution, making it an important platform to study the rules governing the development of ARGs in the environment.

Recent studies have investigated resistome dynamics over time[24, 25] or across treatment compartments in one specific WWTP[24, 25], and the resistome diversity and distribution in several local WWTPs[9, 26, 27]. However, their findings exhibit limited concordance, possibly due to small sample sizes or non-unified protocols. For instance, co-occurrence network analysis suggested the bacterial phyla of Actinobacteria, and Bacteriodetes as main hosts of ARGs in WWTPs[26], but metagenome-assembled genome (MAG)-based methods revealed the most frequent hosts to be Proteobacteria[28]. Moreover, few studies have assessed the environmental factors driving resistomes in WWTPs[9]. Hence, our understanding of global ARG diversity in WWTPs and the underlying mechanisms affecting ARGs in WWTPs remains incomplete. Meta-analysis based on localized experiments is problematic due to differences in experimental systems, sampling methods, and analytical approaches[29, 30]. To discern the global picture of ARGs in WWTPs, a survey is needed that is systematic, methodologically consistent, and globally representative.

To meet this need, a Global Water Microbiome Consortium (GWMC) was established (http://gwmc.ou.edu/) to oversee and coordinate a systematic global campaign for collection, sequencing, and analysis of ~1,200 AS samples using identical protocols[31]. Among these samples, 226 metagenomes (i.e., a collection of genomes and genes from all microorganisms[32]) were identified by shotgun sequencing. The resistomes (i.e., collections of ARGs)[33] were analyzed to address fundamental questions: (i) What are the diversity and distributions of global AS resistomes? (ii) What are the associations among the resistomes and microbiomes? and (iii) What biotic and abiotic mechanisms control the diversity, structure, and distributions of global AS resistomes?

Results and Discussion

Diversity of global AS resistomes

To determine the resistomes of AS, the community DNA of 226 samples from 114 representative WWTPs across six continents (Fig. 1a) was sequenced. A total of 2.8 terabases (Tb) with an average of 12.3 ± 3.9 Gb per sample (Supplementary Data 1) were obtained. Rarefaction analysis of the sequencing reads mapping to bacterial 16S rRNA genes (Supplementary Fig. 1a, b) and ARGs (Supplementary Fig. 1c, d) showed that the sequencing depth was sufficient to represent the diversity of AS microbiomes and resistomes.

Overall, 36,147,212 contigs longer than 1 kb were assembled from all filtered metagenomic reads, and 34,860,381 non-redundant open reading frames (ORFs) were predicted. 37,029 (0.11%) of the ORFs were annotated as ARG sequences. A total of 179 different ARGs, relevant to 15 drug classes, were identified (Supplementary Table 1). To assess geographical distribution, ARG abundance was normalized to the ARG copy number per bacterial cell[34]. The core ARGs in activated sludge, meaning those present in all AS samples analyzed, encompassed 20 genes that accounted for 83.8% of the total ARG abundance (Supplementary Data 2). The three most abundant ARGs were Tetracycline_Resistance_MFS_Efflux_Pump (15.2%), ClassB (13.5%), and vanT gene in vanG cluster (11.4%), which respectively confer Tetracycline, Beta-lactam, and Glycopeptide resistance (Supplementary Data 2).

Since different ARGs might be associated with the same resistance mechanism or drug class, the relative abundances of ARGs were aggregated based on their resistance mechanisms and drug classes (Fig. 1b and Supplementary Fig. 2a). ARGs encoding antibiotic inactivation were the most abundant, accounting for about 55.7% of the total ARG abundance. The next most prevalent were ARGs for antibiotic-target alteration (25.9%) and efflux pumps (15.8%). When ARGs were aggregated by drug class, ARGs conferring resistance to Beta-lactam (46.5%), Glycopeptide (24.5%), and Tetracycline (16.2%) were the most abundant. The relative abundances of ARGs encoding major resistance mechanisms or drug classes were relatively consistent across samples.

Global distribution of AS resistomes

Global variation in ARG abundance. The total ARG abundance showed no significant difference across the six continents (Supplementary Fig. 2b; $p = 0.78$, Kruskal-Wallis test). However, the mean ARG richness (Fig. 1c) and Shannon's H index (Supplementary Fig. 2c) were significantly higher in Asia than in other continents except Africa. ARG abundance varied across samples from different countries ($p = 0.034$, Kruskal-Wallis test): Samples from Chile (2.87 ± 0.40) and Canada (3.10 ± 0.35) were the lowest in mean ARG abundance, while samples from Switzerland (4.30 ± 0.20) and Colombia (4.26 ± 0.86) were the highest (Supplementary Fig. 3a). However, *post hoc* analysis indicated that total ARG abundance was not significantly different between any country pairs (*p.adj > 0.05,* Dunn *post hoc* tests).

Global variations in ARG compositions. To identify structural differences of resistomes across continents, PERMANOVA (Permutational multivariate analysis of variance) was performed at the individual gene level (Table 1). The resistomes were all significantly different ($p < 0.05$) when comparing pairwise continents. Principal coordinate analysis (PCoA) and clustering analysis at the gene level showed a strong regional separation (Fig. 1d, Supplementary Fig. 4a, and Supplementary Note 1). A weaker regional separation was observed at the drug-class level, versus the gene level (Supplementary Fig. 4b and Supplementary Note 1).

ARG differences across different habitats. To determine whether the structure of AS resistomes resembled those from other habitats, we conducted a comparative analysis of resistomes across different environments (AS, human gut[35], soil[36], ocean[37], and sewage[15]) according to the read-based annotations. Comparison of the results obtained from contig- and read-based approaches on our AS samples demonstrated that the major conclusions remain consistent regardless of the approach used (Supplementary Fig. 5 and Supplementary Note 2). PCoA revealed that the resistomes were distinctly different across habitats (Fig. 1e). AS resistomes were much more similar to sewage and soil resistomes than to ocean or human gut resistomes (Fig. 1e), even when aggregated by resistance mechanisms or drug classes (Supplementary Fig. 3b, c). The similar ARG compositions among AS, sewage, and soil could be due to the interconnection of these environments, as sewage is the influent of WWTPs, and soils could also be an important source of the influent's compositions, especially in combined sewer systems that collect both domestic sewage and stormwater.

Relationships between the resistomes and microbiomes

Associations of the resistomes to bacterial community structure. To understand the relationships between resistomes and bacterial community structure, we performed Procrustes analyses. The bacterial community structure was represented either by 16S rRNA genes extracted from metagenomes (Fig. 2a) or amplified 16S rRNA genes (Fig. 2b). Procrustes analysis yielded a matrix-matrix correlation coefficient of 0.74 for metagenome 16S-based bacterial community structure, and a matrix-matrix correlation coefficient of 0.70 for 16S amplicon-based bacterial community structure (protest, $p < 0.001$), suggesting a strong association between WWTP bacterial community structure and the resistomes. These results are consistent with previous

studies on local WWTPs[9, 27] and soil[38], demonstrating that bacterial community composition plays a pivotal role in shaping the resistomes.

To further determine whether the relationships between the resistomes and microbiomes depend on phylogenetic lineages, we determined the linkages of the total ARG abundance and the top four major ARG groups to the relative abundances of major phyla (Fig. 2c and Supplementary Note 3). Bacteroidetes, the most abundant phylum, were positively correlated with the ARG abundance based on amplicon 16S rRNA gene data (rho = 0.28, adjusted p = 0.0001). Based on metagenome-derived 16S rRNA genes, the ARG abundance was also positively correlated with Chloroflexi (rho = 0.48, adjusted $p < 2.7\text{x}10^{-13}$), Acidobacteria (rho = 0.28, adjusted $p = 9.4\text{x}10^{-5}$), Gemmatimonadetes (rho = 0.24, adjusted p = 0.001), Nitrospirae (rho = 0.20, adjusted p = 0.009), and Deltaproteobacteria (rho = 0.20, adjusted p = 0.008), suggesting that these taxa may be major carriers of ARGs. Strong correlations between the ARG abundance and taxonomic groups were also observed in other environments but with different patterns (Supplementary Fig. 6 and Supplementary Note 3). These results suggest that the resistomes in AS could be strongly tied to microbial physiology.

ARG-associated metagenome-assembled genomes. To further understand the association between ARGs and their bacterial hosts, the shotgun sequences of these global AS samples were assembled into contigs and binned into genomes (See Methods for details). A total of 1,112 dereplicated high-qualityand MAGs were recovered with 536 Bacteroidota, 272 Proteobacteria and 43 Actinobacteria. We detected that 1,054 of them contain at least one ARG and 28 were identified as potential human pathogens based on the taxonomic information and presence of virulence factors[39-41] (Supplementary Note 4). As shown in the MAGs-based phylogenetic tree in Fig. 2d, the total ARG abundance and major ARG classes varied greatly among different phylogenetic groups. Chloroflexi (7.2±3.0 ARG counts), Acidobacteria (6.6±3.0), Deltaproteobacteria (4.5±2.8), Gemmatimonadota (3.5±2.1), and Bacteroidetes (3.3±1.7) were the top five carriers of ARGs (Fig. 2e), which was consistent with their positive correlations with the ARG abundance. Bacteroidetes and Proteobacteria were reported to be the main hosts of ARGs in local WWTPs[26, 28], consistent with our synthetic analyses using both correlation- and MAG-based methods. This is likely due to their ability to disseminate resistance genes via horizontal gene transfer (HGT)[42] and their adaptability to antibiotic-rich environments[43].

Collectively, all the above analyses indicate that the identified taxa may play significant roles in ARG persistence and dissemination in activated sludge systems.

Mobility of resistomes and MAGs

MGEs facilitate the horizontal transfer of ARGs, contributing to antibiotic resistance dissemination and evolution in microbial communities[44]. For determining the diversity of MGEs, a total of 2,200 non-redundant ORFs were identified as 56 MGE genes (Supplementary Data 3). The three most abundant MGEs were *tnpA*, *IS91*and *tniA*, and the corresponding MGEs classes were transposase, insertion_element_IS91 and plasmid in AS (Fig. 3a and Supplementary Note 5). The total MGE abundance showed significant difference across the six continents (Fig. 3b; $p = 1.2\text{x}10^{-6}$, Kruskal-Wallis test) and different countries ($p = 5.8\text{x}10^{-7}$, Kruskal-Wallis test). Linear regressions showed that the MGE richness was positively correlated with the ARG richness (R = 0.38, adjusted $p = 2.8\text{x}10^{-9}$). Furthermore, the total ARG abundance was positively correlated with the abundance of their nearby MGEs (R = 0.20, adjusted $p = 0.003$; Supplementary Note 5).

We further quantified mobility based on the ARGs sharing between distinct hosts. Following the method applied to human microbiomes[16, 45], mobile ARGs were identified as identical or near identical sequences present in different bacterial hosts. From these 1,112 dereplicated MAGs, 3,646 ORFs were annotated as ARG sequences, which were further clustered into 2,368 ARG clusters at 99% nucleotide identity. Subsequently, 29% of the ARG clusters (682/2,368) covering 54% of all ARG sequences (1,959/3,646) were assigned to multiple species, suggesting possible recent horizontal gene transfer across distantly related organisms. In comparison, 10% of the ARG clusters from the human microbiome MAGs were multi-species ARGs[16]. Remarkably, the proportion of potentially mobile ARGs in AS was surprisingly higher than that in the human microbiome. This may be due to the high density of bacterial cells and well-mixed nature of AS, which enhances the probability of bacterial physical contact and subsequently increases the likelihood of horizontal gene transfer. Note that the non-mobile/intrinsic ARGs still contribute to the gene pool in the environment, as they might be captured by mobile genetic elements in a certain stage of evolution and become mobile ARGs[46].

The potential ARG mobility for MAGs varied across phylogenetic lineages (Fig. 2d). Of the 1,112 MAGs, 57.6% (641/1,112) were identified as carrying multi-species mobile ARGs.

Among MAGs harboring multi-species mobile ARGs, the proportion of Bacteroidetes phylum was higher than those with immobile ARGs (Fig. 3c), suggesting that the Bacteroidetes phylum could be more prone to horizontal gene transfer to survive in AS with antibiotics. In terms of resistance mechanisms and drug classes, the relative abundances of glycopeptide and macrolide-lincosamide-streptogramin resistance genes also were higher in mobile than immobile ARGs, suggesting that these classes could potentially be more mobile in AS (Fig. 3c). Most mobile ARG clusters can transfer across multi-species, while only 4% (26/682) ARG clusters exhibit the ability to move across multi-phyla (Fig. 3d). Notably, 65% (17/26) multi-phyla mobile ARG clusters are associated with antibiotic inactivation. Horizontal transfer of antibiotic inactivation resistance genes plays a crucial role in microbial survival by enhancing adaptability, accelerating the dissemination of resistance, and conferring evolutionary advantages in antibiotic-rich environments[47]. Horizontal transfer poses considerable challenges to public health.

Drivers of global AS resistomes

We quantitatively assessed the relative contribution of stochastic vs. deterministic processes to the global AS resistome variations with the metric of normalized stochasticity ratio (NST)[48]. The NST estimated for resistomes was generally above 0.5 for all continents except Europe (Fig. 4a and Supplementary Note 6), suggesting that stochastic processes may play a role in the AS resistome variations. Multiple regression on matrices (MRM)-based variance partition analysis (VPA) also revealed that substantial variations (67.4%) of the resistomes remained unexplained by the measured environmental variables and geographical distance (Fig. 4b and Supplementary Note 6). While these results align with previous findings that stochastic processes are important in shaping bacterial community assembly in AS[31], it is critical to note that apparent stochasticity could mask unmeasured deterministic pressures, such as environmental stresses from antibiotics[43], heavy metals[49], or microplastics[50]. Additionally, methodological limitations, including sequencing depth and database biases, might constrain our ability to resolve deterministic signals. Thus, while stochastic processes likely contribute to AS resistome variations, deterministic factors should not be overlooked.

To further discern the roles of individual deterministic factors, we examined the environmental variables having significant correlations ($p < 0.05$) with changes in ARG abundance by using univariate models (Supplementary Table 2). The mixed liquor suspended

298 solid (MLSS), temperature, and city population showed positive correlations with the ARG
299 abundance (Supplementary Fig. 7a-c and Supplementary Note 7). Conversely, the ARG
300 abundance was negatively correlated with pH, solids retention time and influent biochemical
301 oxygen demand (BOD) (Supplementary Fig. 7d-f and Supplementary Note 7), which have been
302 reported to play important roles in regulating the structure of the AS bacterial community[31, 36].
303 Unlike previous observations indicating that the abundance of sewage ARGs is strongly
304 correlated with socio-economic factors[15], we found no significant correlation between ARG
305 abundance and per capita GDP or country-level antibiotics use[51] for where the WWTP is located
306 (Supplementary Table 2). The non-correlation may suggest that the antibiotic concentrations in
307 AS might be insufficient to pose a significant selective pressure for ARGs maintenance and
308 propagation[52]. However, the resolution of antibiotic use data (only from 15 country-level
309 observations) may be too low to reveal its impact on the ARG abundance in AS.

310 A more in-depth analysis using partial least squares (PLS) further revealed potential
311 direct and indirect effects of biotic and abiotic drivers (Fig. 4c). PLS analysis indicated that the
312 bacterial community structure, MGEs, temperature, and city population could affect the AS
313 resistome, which further influenced the AS ecosystem functioning for pollutant removal.
314 Temperature had a direct influence on ARG abundance (Pearson $r = 0.39$, partial $R^2 = 0.08$) and
315 indirectly affected ARG abundance through the bacterial community structure (Pearson $r = 0.54$,
316 partial $R^2 = 0.14$ of the first principal component score (PC1) representing the community
317 structure). Because temperature is a primary driver of biological processes[53], temperature likely
318 has important effects on ARG abundance and distributions[8]. Although the potential mechanisms
319 underlying the relationships between ARGs and temperature are not clear, temperature could
320 facilitate horizontal gene transfer, population growth, biotic interactions, and community
321 turnovers[54-56]. ARG abundance was also directly influenced by the abundance of proximal MGEs
322 (Pearson $r = 0.30$ partial $R^2 = 0.09$). Several studies have shown that MGEs can carry multiple
323 ARGs and contributed to their spread within bacterial populations, thereby increasing the ARG
324 abundance[57, 58]. Another factor that had a direct positive effect on ARG abundance was the city
325 population (Pearson $r = 0.30$, partial $R^2 = 0.05$). A higher population may be associated with an
326 increased use and sewage discharge of antibiotics, exacerbating the emergence and spread of
327 ARGs in bacteria[10]. Overall, although the abiotic environmental variables had significant effects
328 on the resistome, their impact was relatively small (partial $R^2 < 0.1$, Fig. 4c), which are consistent

with the null model-based stochasticity ratio (Fig. 4a) and MRM-VPA analysis (Fig. 4b) showing that stochastic processes may play a more important role.

Concluding remarks

Understanding the global ARG abundance, diversity, and distributions, along with their controlling mechanisms is critical to risk assessment and mitigation of antibiotic resistance. By analyzing the AS resistomes via well-coordinated international efforts, this study showed that ARGs are highly abundant, diverse, and widely distributed across global WWTPs, this corroborates that WWTPs are an important reservoir of environmental ARGs[5, 59-61]. By offering a global-scale characterization of ARGs, this study provides inter-continental and inter-country comparisons of the resistomes in WWTPs. Our results revealed that the structures of activated sludge resistomes differed among continents and were far distant from those of the human gut and oceans, but they exhibited close similarity to those of sewage and soils. We also recovered thousands of dereplicated high-quality MAGS, which could enable more in-depth analyses of ARG hosts and the quantification of ARG mobility. In addition, our analyses indicate that resistome variations in activated sludge may be driven by stochastic processes, such as random gene exchanges and drift[62]. However, deterministic factors such as temperature and city populations still played important roles in the evolution and proliferation of ARGs in global WWTPs.

Methods

Global sampling and DNA sequencing

A total of 1,186 AS samples were collected by the GWMC from 269 WWTPs across 23 countries with varying geographic locations, latitudes, and climate zones[31]. There was a unified protocol (http://gwmc.ou.edu/files/Sampling_Shipping_Protocol_General_20141103.pdf) developed at GWMC for sampling, preserving samples, collecting metadata, collecting DNA, and sequencing so that potential effects of the variations on experimentation would be minimized. A total of 226 representative samples out of 1,186 AS samples had sufficient metadata to be used for metagenomic sequencing.

Detailed information about the procedure of DNA extraction is described in Wu et al[31]. In brief, MoBio PowerSoil DNA isolation kit was used to isolate community DNA from mixed

liquor samples (3 mL). We vortexed 12 bead tubes at maximum speed for 10 minutes, following the manufacturing protocol, to minimize variations in cell lysis efficiency between samples. Then, we constructed genomic DNA libraries by following the manufacturer's instructions with an average insert size of 300 bp using KAPA Hyper Prep Kit (KR0961). DNA LabChip 1000 kit from Agilent was used to assess the quality of all libraries, and all qualified libraries were sequenced at the Oklahoma Medical Research Foundation (OMRF) with paired-end sequencing on Illumina HiSeq3000. The sequenced reads were deposited in the Sequence Read Archive (BioProject accession number PRJNA509305).

Metagenomic sequences processing

An internal metagenomic pipeline (ARMAP, http://zhoulab5.rccc.ou.edu/pipelines/ARMAP_web/job_submission.php) was used to process the metagenomic data. First, all sequenced reads were subjected to FastQC for quality evaluation with quality profile, duplication rates, and contamination rates. Using CD-HIT (v4.6.8)[63], a 100% identity cutoff was used to remove duplicates. Quality trimming and filtering were performed using NGS QC Toolkit (v2.3.3)[64]. The pair-end adaptor library was used to detect reads with residual adaptors. Raw reads were filtered with the following constraints: (i) reads with more than one ambiguous N base were removed; (ii) 3′-ends of reads were trimmed to the first high-quality base with quality score ≥ 20; and (iii) trimmed reads with the length > 120 bp (80% of the sequence read length) were further filtered with an average quality score cutoff of 20. The paired-end reads (fasta) of each sample after quality trimming and filtering were assembled by Megahit (v1.0.5)[65] into contigs in a time- and cost-efficient way, using the following parameters: –min-contig-len = 1000, --k-min = 31, --k-max = 131, --k-step = 20 and –min-count = 1. All assembled contigs were imported into the NGS QC Toolkit for calculation of the contig length profiles (N50Stat.pl).

ARGs annotation for open reading frames

Open reading frames (ORFs) of protein-coding genes were predicted from the assembled contigs of each metagenome by Prodigal (v2.6.3)[66] with '*-p* meta' option. A non-redundant ORFs catalog was constructed by protein clustering using MMseqs2[67], with a minimum identity threshold of 95% and a minimum sequence coverage of 90% (--min-seq-id 0.95 -c 0.9 --cluster-mode 2 --cov-mode 1). The coverages of the non-redundant genes in each sample were

determined by CoverM (v0.6.1) (https://github.com/wwood/CoverM) using default settings. Then, non-redundant ORFs were functionally annotated to the Comprehensive Antibiotic Resistance Database (CARD)[68] and the ResFams database[69]. Genes were first assigned as ARGs by annotating with the CARD using their recommended tool Resistance Gene Identifier (RGI) (v6.0.0), requiring a hit scoring above the family-specific threshold under the CARD homologue model, with the top hit taken if several are achieved. The remaining unannotated genes were filtered and subsequently annotated with Resfams protein families, requiring the score to a ResFams hidden Markov model exceeded the gathering threshold for that model. The ORF annotated to Resfam were represented as gene families. The following criteria were used to remove potential false positive ARGs: (i) genes that confer resistance via overexpressing of resistant target alleles (e.g. resistance to antifolate drugs via mutated DHPS and DHFR); (ii) global gene regulators, two-component system proteins, and signaling mediators; (iii) efflux pumps that confer resistance to multiple antibiotics; (iv) genes modifying cell wall charge (e.g. those conferring resistance to polymixins and defensins). Raw unnormalized abundance value was calculated for each ARG in a sample as the summed coverage depths of all ORFs that were annotated to that ARG in the given sample.

To assess the ARG distributions in AS samples, the raw abundance of ARGs was normalized and expressed as "copy of ARG per cell" using the equation (1).

$$Abundance = \sum_{i=1}^{n} \frac{Coverage_{i(\text{ARG-like gene})}}{Coverage_{16S\ \text{sequence}}} \times N_{16S\ \text{copy number}}$$

$$= \sum_{i=1}^{n} \frac{N_{i(\text{ARG-like sequence})} \times L_{reads}/L_{i(\text{ARG ORF})}}{N_{16S\ \text{sequence}} \times L_{reads}/L_{16S\ \text{sequence}}} \times N_{16S\ \text{copy number}} \quad (1)$$

Where $Coverage_{i(\text{ARG-like gene})}$ is the coverage of a specific ARG ORF, which is calculated from the number of reads annotated to this ORF ($N_{i(\text{ARG-like sequence})}$), the sequence length (bp) of the reads (L_{reads}), and the length (bp) of the corresponding ARG ORF ($L_{i(\text{ARG ORF})}$). For the coverage of 16S rRNA gene ($Coverage_{16S\ \text{sequence}}$) calculation, $N_{16S\ \text{sequence}}$ is the number of the 16S rRNA gene sequences identified for the metagenomic data by Metaxa2 (v2.248)[70], L_{reads} represents the sequence length of the reads, $L_{16S\ \text{sequence}}$ is the average length of 16S rRNA genes (1,432 bp) in Greengenes database[71]. $N_{16S\ \text{copy number}}$ is the average copy number of 16S rRNA genes per cell in the community, and *n* is the number of annotated ARGs for a

specific category. The average copy number in the community was calculated as the abundance-weighted mean 16S rRNA gene copy number, where the 16S rRNA gene copy number of each genus was estimated through the rrnDB database based on its closest relatives with known rRNA gene copy number[72, 73]. It is noted that the normalized ARG abundance (gene copies per cell) depends on the algorithms for identifying ARGs and 16S rRNA genes. There could be false positives and false negatives; thus, the resultant ARG abundance may not reflect the real values in the community. However, we can still conduct relative comparison across different samples, under the assumption that the estimations across samples are subjected to the same degree of bias. In this way, we can compare the abundance of ARGs of samples and explore the underlying mechanisms shaping the resistomes.

Mobile genetic elements (MGEs) annotation

To determine the diversity of MGEs in the AS, we annotated MGEs for the non-redundant ORFs by BLASTN (-perc_identity 0.5 -evalue 1e-10 -max_target_seqs 1) against the previously published database of MGEs[74]. This database consists of MGEs with 278 different genes and more than 2,000 unique sequences. The raw abundance of each MGE in a sample was calculated as the summed coverage depths of all ORFs annotated to that MGE and normalized as "copy of MGE per cell" in the same manner as for the ARGs.

To quantify the mobility potential of ARGs, we performed co-localization analysis between ARGs and MGEs on all assembled contigs. We first annotated the ARGs and MGEs on all contigs and then identified the contigs carrying both ARGs and MGEs for calculating the minimum distance between them. ARGs with potential mobility were defined as sharing a nearby area (<10 kb)[75] with an MGE. We calculated the proportions of mobile ARGs in each sample. We also calculated the raw abundance of MGEs co-located (<10 kb) with ARGs using the coverage of the corresponding contigs in the given sample, which was determined by CoverM (v0.6.1) using default settings. The raw abundances of MGEs were then normalized as "copy per cell" with the above method.

Taxonomic profiling of the metagenomic sequences

Bacterial-community profiling at the genus level was done using Metaxa2 (v2.248)[70], based on the bacterial 16S rRNA reads extracted from the high-quality metagenomic reads. The bacterial profile was also represented by the OTU table based on 16S rRNA amplicon sequencing data,

which was published by Wu et al[31]. The relative abundance of a taxonomic category was calculated as the sum of reads annotated to that category normalized by the total number of taxonomic reads in each sample.

MAG recovery, taxonomic annotation, and phylogenetic tree construction

All assembled contigs longer than 1 kbp were binned with Matabat2[76], Maxbin2[77], and CONCOCT (v0.4.1)[78] based on contig composition and coverages. Before binning, Bowtie2[79] was used to align short-read sequences to contigs (options: -very-fast), and SAMtools[80] was used to sort and convert SAM files to BAM format. Then, DAStools[81] was used to refine binned contigs with default parameters where Usearch[82] was used as the search engine. We performed CheckM (v1.0.6)[83] to estimate the completeness and contamination of each bin. To get the nonredundant consolidation, the dRep[84] dereplication workflow was used with options 'dereplicate_wf -p 16 -pa 0.9 -sa 0.95 -nc 0.3 -comp 70 -con 10 -str 100 -strW 0'. Bin scores were given as completeness-5×contamination+0.5×log(N50), and only the highest-scoring MAGs from each cluster (> 95% average nucleotide identity) were retained in the dereplicated set. The bins with high completeness (>90%) and few contaminants (<5%) were retained as high-quality MAGs and were used for downstream analyses.

The taxonomy of the represented MAGs was assigned using GTDB-tk v2.1.0[85] based on the Genome Taxonomy Database[86]. Besides, to identify the pathogenic genomes, we first selected the potential ones by referring to two published reference pathogen lists that consisted of 140 potentially human pathogenic genera[40] and 538 potentially human pathogenic species[41]. Then, we searched the ORFs of taxonomically predicted potentially pathogenic genomes against the experimentally verified bacterial virulence factor database VFDB (last update: Dec.11, 2020)[39] with BLASTN. The genomes with virulence factors with a global identity > 70% were considered pathogens. The phylogenetic relationships of all MAGs were inferred by a maximum likelihood alignment-based approach with PhyloPhlAn3[87] (--diversity high, --fast, with configurations --db_aa diamond, --map_dna diamond, --map_aa diamond, --msa mafft, --trim trimal, --tree1 iqtree). Visualization and annotation of the tree were done using GraPhlAn[88]. It should be noted that it has proven difficult to assemble genomes for populations below 1% relative abundance owing to insufficient sequencing depth or difficulty in binning and assembly of individual genomes from complex metagenomes[89].

476 **ARG host and mobility annotation for MAGs**

477 For the near-complete MAGs, ARGs of MAGs' contigs were also identified based on CARD[68]
478 and the ResFams database[69] as above. The mobile ARGs were defined as identical or near
479 identical sequences present in different species[16, 45]. Since our recovered MAGs were
480 dereplicated at an average nucleotide identity of 95%, they represented species-level genome
481 bins[90, 91]. Thus, we searched for mobile ARGs as those present in two or more MAGs. To
482 achieve this, we first clustered the nucleotide sequences of all detected ARG ORFs into ARG
483 clusters with 99% identity, using the 'cluster' command of MMseqs2[67] with '–min-seq-id 0.99 -c
484 0.9 –cov-mode 0'. We then labeled any ARG cluster that was found in multiple MAGs as 'multi-
485 species', which was considered as the evidence of recent horizontal gene transfer. This strategy
486 of searching for ARG clusters across species to detect recent horizontal gene transfer is
487 equivalent to that used in some other studies on human microbiomes[16, 45].

488 **Analyzing metagenomic samples from other environments**

489 To compare AS resistomes with other environments, we selected the public global metagenomic
490 projects in human gut[35], sewage[15], soil[36], and oceans[37] and collected samples from these public
491 databases. The raw metagenomic sequences were downloaded from the European Bioinformatics
492 Institute Sequence Read Archive database (sewage: PRJEB13831, soil: ERP020652, gut:
493 ERP004605, ocean: ERP001736). To avoid bias caused by data processing, we re-processed the
494 raw sequences with the same quality trimming and filtering parameters with our pipeline to
495 obtain high-quality sequences. Rather than using the contig-based approach to annotate ARGs
496 which requires significant time and vast computational resources for the assembly step, here we
497 profiled the abundance of ARGs through a read-based mapping strategy. The read-based
498 approach enabled an efficient comparison of resistomes between environments. We annotated
499 ARGs from the high-quality metagenomic sequences by DeepARG (v2)[92] using the default
500 options (--id 50, -e 1e-10, -k 1000 of short reads mode), which can infer ARGs from short reads.
501 The abundances of ARGs were normalized to the unit of "copy per cell"[34] in a similar manner as
502 described above, although the calculation of ARG coverage was slightly different with the
503 equation (2).

504 $$Abundance = \sum_{i=1}^{n} \frac{Coverage_{i(\mathrm{ARG})}}{Coverage_{16\mathrm{S\ sequence}}} \times N_{16\mathrm{S\ copy\ number}}$$

$$= \sum_{i=1}^{n} \frac{N_{i(\text{ARG-like sequence})} \times L_{reads} / L_{i(\text{ARG reference sequence})}}{N_{16S\ \text{sequence}} \times L_{reads} / L_{16S\ \text{sequence}}} \times N_{16S\ \text{copy number}} \quad (2)$$

Where $N_{i(\text{ARG-like sequence})}$ is the number of ARG-like reads annotated as one specific ARG reference sequence, $L_{i(\text{ARG reference sequence})}$ is the sequence length (bp) of the corresponding ARG reference sequence.

To compare the results of the two ARG detection methods (contig-based and read-based approaches), we performed Procrustes analyses between the resultant AS ARG abundance matrices from the two methods using the function 'procrustes' of *vegan* R package[93]. We also examined the correlation between the total abundance from two methods using the function 'lm' of R.

Statistical analyses

The global map was created using the function 'tm_shape' of *spData* R package (10.32614/CRAN.package.spData). Richness and Shannon's H index were computed using the *vegan* R package[93] to measure the diversity of ARGs or MGEs based on a rarified count matrix, which was obtained by rounding the coverages and sub-sampling to the lowest sample's level. The richness and Shannon's H diversity rarefaction curves for bacteria and ARGs were respectively based on the reads mapping to the bacterial 16S rRNA genes and ARGs. The curves were computed using the function 'rarefaction.individual' of *rareNMtests*[94] and plotted using the *ggplot2*[95] R packages. Kruskal–Wallis and the Dunn *post hoc* test were used to compare the means of ARG abundance or diversity between continents or countries, using R function 'kruskal.test' and function 'dunnTest' of *FSA* R package[96]. To visualize the variation of resistomes across samples, the principal coordinate analysis (PCoA) was performed on the resistome Bray-Curtis dissimilarity matrix based on gene relative abundances, using the function 'pcoa' of *ape* R package[97]. The heat map of genes was generated using the function 'aheatmap' of *NMF* R package[98]. PERMANOVA was applied to assess the resistome dissimilarities among continents using the function 'adonis2' of *vegan* R package. Procrustes analysis was performed to test the association between bacterial taxonomic composition and the resistome using the function 'procrustes' of *vegan* R package in which the ordinations of the bacterial taxonomic composition and the resistome were generated from PCoA. To disentangle the relative contributions of stochastic and deterministic processes to AS resistome, null model-based NST

approach[48] was applied to community ARG data. Normalized stochasticity ratio (NST) was used to quantify ecological stochasticity in communities within continents and was analyzed in R using the *NST* package[48].

To estimate the relative contributions of the environmental effects versus the distance effects on the resistome dissimilarities, we performed a variation partition analysis (VPA) based on multiple regression on matrices (MRM). Briefly, we first selected a non-redundant set of environmental variables that contained missing data in less than 20% of all samples. The final set included mixed liquid temperature, air temperature, precipitation, design capacity, volume of aeration tanks, plant age, mixed liquor suspended solids (MLSS), solids retention time (SRT), dissolved oxygen (DO), pH, and influent biochemical oxygen demand (BOD), effluent BOD, food to microorganism (F/M) ratio and city GDP. The variance inflation factors (VIF) were less than 10, indicating a low level of collinearity among these variables. MRM was performed using the function 'MRM' of *ecodist* R package [99]. Geographic distance was log-transformed. A Euclidean distance matrix was calculated for each environmental variable. In VPA, the R^2 of the selected environmental variables as independent matrices (R^2_E), geographical distance as an independent matrix (R^2_G), and all matrices (R^2_T) were used to compute the three components of variations: (i) pure environmental variation = $R^2_T - R^2_G$; (ii) pure geographical distance = $R^2_T - R^2_E$; and (iii) spatially structured environmental variation = $R^2_G + R^2_E - R^2_T$. Univariate models predicting the total ARG abundance (ARG copies per cell) as a function of various environmental and site variables were performed using R function 'lm' and 'summary'. For each variable, we fitted a linear and a quadratic model and results are shown for the model with lower Akaike information criteria (AIC) value.

The partial least squares (PLS) model with a partial R^2 index based on PLS[100] was used to explore the relationships among the microbiome (PC1 of bacterial community structure), resistome (the total ARG abundance, PC1 of ARG composition, abundances of the top three resistance mechanisms), the abundance of MGEs located near (< 10kb) ARGs, six environmental variables which significantly correlated ($p < 0.05$) with the total ARG abundance based on the univariate models, and ecosystem functions (the removal rate of BOD, COD, total nitrogen and total phosphorus). Based on predictive performance counting in the explained variation (R^2_Y) and model significance (P for R^2_Y and $Q^2_Y < 0.05$, where significant Q^2_Y helps to avoid overfitting), Each optimum PLS model was forward selected from all factors that might

affect the dependent variable. To visualize relevant associations, we only included the most relevant variable(s) with Variable Influence on Projection (VIP) values larger than 1. When used as independent variables in PLS, the ARG composition was represented by the PC1 from PCoA of Bray-Curtis distance. We used a partial R^2 index[101] on the basis of PLS to represent the proportion of variance explained by each independent variable (equation 3). We also calculated the pairwise correlation coefficient (as well as the R^2) among the factors and the significance was based on Pearson correlation as reference. The PLS-related analysis was performed using the *ropls* package[102] and the Mantel test using the *vegan* package[93] in R.

$$R^2_{\mathrm{PLS}j} = R^2_Y \times \frac{\sum_f (W^2_{jf} \times SSY_f)}{SSY_{cum}} = \frac{\sum_f (W^2_{jf} \times SSY_f)}{SSY} \qquad (3)$$

Where $R^2_{\mathrm{PLS}j}$ is the partial R^2 of variable j based on PLS, W_{jf} is the PLS weight of variable j on component f, SSY_f is the sum of squares of Yexplained by component f, SSY_{cum} is the cumulative sum of squares of Yexplained by all components, R^2_Y is the percentage of Y dispersion (i.e., sum of squares) explained by the PLS model, and SSY is the Y dispersion, that is, sum of squares of Y.

Data availability

The DNA sequences of the 16S rRNA gene and metagenomes generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) database under the project accession number PRJNA509305 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA509305]. The assembled MAG sequences have been deposited in Zenodo under the DOI 10.5281/zenodo.14916172. [https://doi.org/10.5281/zenodo.14916172]. The source data underlying figures and supplementary figures are provided as a Source Data file with this paper. The raw metagenomic sequences of metagenomic samples from other environments used in this study are available in the European Bioinformatics Institute Sequence Read Archive database (sewage: ERP015409 [https://www.ebi.ac.uk/ena/browser/view/PRJEB13831], soil: ERP020652 [https://www.ebi.ac.uk/ena/browser/view/PRJEB18701], gut: ERP004605

590 [https://www.ebi.ac.uk/ena/browser/view/PRJEB5224], ocean: [ERP001736
591 https://www.ebi.ac.uk/ena/browser/view/PRJEB1787]).

592 **Code availability**

593 No custom algorithms or software were used to generate and analyze data. The R script for
594 partial least squares is publicly available on GitHub at https://github.com/congminz/GWMC.

595 **Global Water Microbiome Consortium members contributing to this study**

596 Dany Acevedo[34], Miriam Agullo-Barcelo[35], Pedro J. J. Alvarez[29], Lisa Alvarez-Cohen[26,27],
597 Juliana Calabria de Araujo[36], Érika Ferreira de Abreu Mac Conell[36], Kevin Boehnke[37], Philip
598 Bond[35], Charles B. Bott[38,39,40], Patricia Bovio-Winkler[41], Rebecca K. Brewster[42], Faizal Bux[43],
599 Angela Cabezas[44], Léa Cabrol[45,46], Si Chen[23], Craig S. Criddle[23], Thomas P. Curtis[9], Claudia
600 Etchebehere[41], Amanda Ford[38], Janeth Sanabria Gómez[34], James S. Griffin[47], April Z. Gu[48],
601 Moshe Habagil[49], Lauren Hale[4], Marc Harmon[50], Qiang He[21,22], Zhili He[15,16], Harald Horn[51],
602 Zhiqiang Hu[52], David R. Johnson[53], Deborah Garbani Marcantini[53], Alexander Keucken[49,54],
603 Sheena Kumari[43], Cintia Dutra Leal[36], Zarraz MP Lee[55], Yong Li[56], Zhenxin Li[57], Mengyan Li[29],
604 Yu Liu[55,58], Leda C. Mendonça-Hagler[59], Francisca Gleire Rodriguez de Menezes[60], Arthur J.
605 Meyers[61], Per H. Nielsen[17], Daliang Ning[3], Andrew Palmer[62], Prathap Parameswaran[30], Valeria
606 Reginatto[63], Francis L. de los Reyes III[64], Bruce E. Rittmann[30], Simona Rossetti[65],Jatinder
607 Sidhu[62], Kylie Smith[62], Oscarina Viana de Sousa[60], David A. Stahl[28], Kyle Stephens[66],
608 Chenxiang Sun[6], Renmao Tian[3,5], James M. Tiedje[25], Nicholas B. Tooker[67], Qichao Tu[10], Joy D.
609 Van Nostrand[4], Michael Wagner[24], Steve Wakelin[68], Aijie Wang[18], Bei Wang[69], Joseph E.
610 Weaver[64], George F. Wells[13], Xianghua Wen[6], Stephanie West[51], Sung-Geun Woo[23], Linwei
611 Wu[3,4], Jer-Horng Wu[70], Liyou Wu[3], Chuanwu Xi[42], Naijia Xiao[3], Tao Yan[71], Min Yang[19],
612 Yunfeng Yang[6], Michelle Young[33], Bing Zhang[6,7], Chiqian Zhang[52], Qian Zhang[72,73], Wen
613 Zhang[74], Ya Zhang[3], Yu Zhang[19], Wanlin Zheng[6], Hongde Zhou[69], Jizhong Zhou[3,27,31,32,33],
614 Weiqin Zhuang[71]
615
616 [34]Environmental Microbiology and Biotechnology Laboratory, Engineering School of
617 Environmental & Natural Resources, Engineering Faculty, Universidad del Valle–Sede
618 Meléndez, Cali, Colombia

[35]Advanced Water Management Centre, The University of Queensland, Brisbane, QLD, Australia
[36]Universidade Federal de Minas Gerais, Departamento de Engenharia Sanitária e Ambiental, Belo Horizonte, Brazil
[37]Anesthesiology Department, The University of Michigan, Ann Arbor, MI, USA
[38]Hampton Roads Sanitation District (HRSD), Virginia Beach, VA, USA
[39]Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, USA
[40]Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA, USA
[41]Microbial Ecology Laboratory, Microbial Biochemistry and Genomics Department, Biological Research Institute "Clemente Estable", Montevideo, Uruguay
[42]Department of Environmental Health Sciences, The University of Michigan, Ann Arbor, MI, USA
[43]Institute of Water and Wastewater Technology, Durban University of Technology, Durban, South Africa
[44]Instituto Tecnológico Regional Centro Sur, Universidad Tecnológica, Durazno, Uruguay
[45]Aix-Marseille University CNRS IRD, MIO UM110 Mediterranean Institute of Oceanography, Marseille, France
[46]Escuela de Ingenieria Bioquimica, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
[47]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA、
[48]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[49]Vatten & Miljö i Väst AB (VIVAB), Falkenberg, Sweden
[50]Golden Heart Utilities, Fairbanks, AK, USA
[51]Karlsruhe Institute of Technology, Engler-Bunte-Institut, Water Chemistry and Water Technology, Karlsruhe, Germany
[52]Department of Civil and Environmental Engineering, University of Missouri, Columbia, MO, USA
[53]Department of Environmental Microbiology, Eawag, Dübendorf, Switzerland

650 [54]Water Resources Engineering, Faculty of Engineering, Lund University, Lund, Sweden
651 [55]Advanced Environmental Biotechnology Centre, Nanyang Environment & Water Research
652 Institute, Nanyang Technological University, Singapore
653 [56]College of Resource & Environment Southwest University, Chongqing, China
654 [57]School of Environment, Northeastern Normal University, Changchun, China
655 [58]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore
656 [59]Plant Biotechnology Program, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
657 [60]Federal University of Ceará, UFC, Fortaleza, Brazil
658 [61]University of Tennessee, Center for Environmental Biotechnology, Knoxville, TN, USA
659 [62]CSIRO Land and Water, Ecosciences Precinct, QLD, Australia
660 [63]Departamento de Química, Universidade de São Paulo, Faculdade de Filosofia Ciências e
661 Letras de Ribeirão Preto – FFCLRP, Ribeirão Preto, SP, Brazil
662 [64]Department of Civil, Construction, and Environmental Engineering, North Carolina State
663 University, Raleigh, NC, USA
664 [65]CNR-IRSA, National Research Council, Water Research Institute, Rome, Italy
665 [66]Tryon Creek & Columbia Blvd. Wastewater Treatment Plants, Bureau of Environmental
666 Services, City of Portland, OR, USA
667 [67]Department of Civil and Environmental Engineering, Northeastern University, Boston, MA,
668 USA
669 [68]Scion Research, Christchurch, New Zealand
670 [69]School of Engineering, University of Guelph, Guelph, Ontario, Canada
671 [70]Department of Environmental Engineering, National Cheng Kung University, Tainan City,
672 Taiwan, China
673 [71]Department of Civil and Environmental Engineering, University of Hawaii at Manoa,
674 Honolulu, HI, USA
675 [72]Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand
676 [73]Biotechnology Institute, University of Minnesota, St. Paul, MN, USA
677 [74]Department of Civil Engineering, University of Arkansas, Fayetteville, AR, USA

678 # References

1. Balaban, N.Q. et al. Definitions and guidelines for research on antibiotic persistence. *Nature reviews. Microbiology* **17**, 441-448 (2019).
2. Organization, W.H. Antimicrobial resistance: global report on surveillance. (World Health Organization, 2014).
3. Holmes, A.H. et al. Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet (London, England)* **387**, 176-187 (2016).
4. Nathan, C. & Cars, O. Antibiotic resistance--problems, progress, and prospects. *The New England journal of medicine* **371**, 1761-1763 (2014).
5. Berendonk, T.U. et al. Tackling antibiotic resistance: the environmental framework. *Nat Rev Microbiol* **13**, 310-317 (2015).
6. Larsson, D.G.J. et al. Critical knowledge gaps and research needs related to the environmental dimensions of antibiotic resistance. *Environ Int* **117**, 132-138 (2018).
7. Larsson, D. & Flach, C.-F. Antibiotic resistance in the environment. *Nature Reviews Microbiology*, 1-13 (2021).
8. Diehl, D.L. & LaPara, T.M. Effect of temperature on the fate of genes encoding tetracycline resistance and the integrase of class 1 integrons within anaerobic and aerobic digesters treating municipal wastewater solids. *Environmental science & technology* **44**, 9128-9133 (2010).
9. Ju, F. et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *The ISME journal* **13**, 346-360 (2019).
10. Zhu, Y.G. et al. Continental-scale pollution of estuaries with antibiotic resistance genes. *Nature microbiology* **2**, 16270 (2017).
11. Zhang, Z. et al. Assessment of global health risk of antibiotic resistance genes. *Nature communications* **13**, 1553 (2022).
12. Zheng, D. et al. Global biogeography and projection of soil antibiotic resistance genes. *Science advances* **8**, eabq8015 (2022).
13. Wang, B. et al. Global diversity, coexistence and consequences of resistome in inland waters. *Water research* **253**, 121253 (2024).
14. Danko, D. et al. A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell* **184**, 3376-3393.e3317 (2021).
15. Hendriksen, R.S. et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature communications* **10**, 1-12 (2019).
16. Lee, K. et al. Population-level impacts of antibiotic usage on the human gut microbiome. *Nature communications* **14**, 1191 (2023).
17. Tiseo, L. Global share of population with wastewater collection systems by region 2018. https://www.statista.com/statistics/1016760/population-connected-to-wastewater-collection-system-globally-by-region (2018).
18. World Population. https://www.worldometers.info/world-population (2018).
19. Wang, X. et al. Probabilistic evaluation of integrating resource recovery into wastewater treatment to improve environmental sustainability. *Proceedings of the National Academy of Sciences* **112**, 1630-1635 (2015).
20. Grant, S.B. et al. Taking the "waste" out of "wastewater" for human water security and ecosystem sustainability. *Science* **337**, 681-686 (2012).
21. Kartal, B., Kuenen, J.v. & Van Loosdrecht, M. Sewage treatment with anammox. *Science* **328**, 702-703 (2010).

22. van Loosdrecht, M.C. & Brdjanovic, D. Anticipating the next century of wastewater treatment. *Science* **344**, 1452-1453 (2014).
23. Yang, C. et al. Phylogenetic Diversity and Metabolic Potential of Activated Sludge Microbial Communities in Full-Scale Wastewater Treatment Plants. *Environmental science & technology* **45**, 7408-7415 (2011).
24. Sun, C., Li, W., Chen, Z., Qin, W. & Wen, X. Responses of antibiotics, antibiotic resistance genes, and mobile genetic elements in sewage sludge to thermal hydrolysis pre-treatment and various anaerobic digestion conditions. *Environment international* **133**, 105156 (2019).
25. Zheng, W., Huyan, J., Tian, Z., Zhang, Y. & Wen, X. Clinical class 1 integron-integrase gene–A promising indicator to monitor the abundance and elimination of antibiotic resistance genes in an urban wastewater treatment plant. *Environment international* **135**, 105372 (2020).
26. An, X.L. et al. Tracking antibiotic resistome during wastewater treatment using high throughput quantitative PCR. *Environment international* **117**, 146-153 (2018).
27. Munck, C. et al. Limited dissemination of the wastewater treatment plant core resistome. *Nature communications* **6**, 8452 (2015).
28. Yuan, L. et al. Pathogenic and Indigenous Denitrifying Bacteria are Transcriptionally Active and Key Multi-Antibiotic-Resistant Players in Wastewater Treatment Plants. *Environmental science & technology* **55**, 10862-10874 (2021).
29. Thompson, L.R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457-463 (2017).
30. Zhou, J. et al. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. *mBio* **6**, e02288-02214 (2015).
31. Wu, L. et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nature microbiology* **4**, 1183–1195 (2019).
32. Marchesi, J.R. & Ravel, J. The vocabulary of microbiome research: a proposal. *Microbiome* **3**, 31 (2015).
33. Wright, G.D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature reviews. Microbiology* **5**, 175-186 (2007).
34. Ma, L. et al. Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. *Microbiome* **5**, 154 (2017).
35. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nature biotechnology* **32**, 834-841 (2014).
36. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233-237 (2018).
37. Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
38. Forsberg, K.J. et al. Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612-616 (2014).
39. Liu, B., Zheng, D., Jin, Q., Chen, L. & Yang, J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* **47**, D687-D692 (2019).
40. Cai, L., Ju, F. & Zhang, T. Tracking human sewage microbiome in a municipal wastewater treatment plant. *Appl Microbiol Biotechnol* **98**, 3317-3326 (2014).

41. Zhang, S., He, Z. & Meng, F. Floc-size effects of the pathogenic bacteria in a membrane bioreactor plant. *Environment international* **127**, 645-652 (2019).
42. Brito, I.L. Examining horizontal gene transfer in microbial communities. *Nature Reviews Microbiology* **19**, 442-453 (2021).
43. Brown, C.L. et al. Selection and horizontal gene transfer underlie microdiversity-level heterogeneity in resistance gene fate during wastewater treatment. *Nature communications* **15**, 5412 (2024).
44. Brito, I.L. Examining horizontal gene transfer in microbial communities. *Nature reviews. Microbiology* **19**, 442-453 (2021).
45. Brito, I.L. et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435-439 (2016).
46. Hu, Y., Gao, G.F. & Zhu, B. The antibiotic resistome: gene flow in environments, animals and human beings. *Frontiers of medicine* **11**, 161-168 (2017).
47. Darby, E.M. et al. Molecular mechanisms of antibiotic resistance revisited. *Nature Reviews Microbiology* **21**, 280-295 (2023).
48. Ning, D., Deng, Y., Tiedje, J.M. & Zhou, J. A general framework for quantitatively assessing ecological stochasticity. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 16892-16898 (2019).
49. Liu, Z.-T. et al. Organic fertilization co-selects genetically linked antibiotic and metal(loid) resistance genes in global soil microbiome. *Nature communications* **15**, 5168 (2024).
50. Wang, Y.-F. et al. Microplastic diversity increases the abundance of antibiotic resistance genes in soil. *Nature communications* **15**, 9788 (2024).
51. The Center for Disease, D.E.P. (2021).
52. Zheng, W., Huyan, J., Tian, Z., Zhang, Y. & Wen, X. Clinical class 1 integron-integrase gene - A promising indicator to monitor the abundance and elimination of antibiotic resistance genes in an urban wastewater treatment plant. *Environment international* **135**, 105372 (2020).
53. Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).
54. MacFadden, D.R., McGough, S.F., Fisman, D., Santillana, M. & Brownstein, J.S. Antibiotic resistance increases with local temperature. *Nature Climate Change* **8**, 510-514 (2018).
55. Guo X et al. Climate warming leads to divergent succession of grassland microbial communities. *Nature Climate Change* **8**, 813-818 (2018).
56. Yuan, M.M. et al. Climate warming enhances microbial network complexity and stability. *Nature Climate Change* **11**, 343-348 (2021).
57. Kim, M., Park, J., Kang, M., Yang, J. & Park, W. Gain and loss of antibiotic resistant genes in multidrug resistant bacteria: One Health perspective. *Journal of microbiology (Seoul, Korea)* **59**, 535-545 (2021).
58. Jiang, Y. et al. Pooled Plasmid Sequencing Reveals the Relationship Between Mobile Genetic Elements and Antimicrobial Resistance Genes in Clinically Isolated Klebsiella pneumoniae. *Genomics, proteomics & bioinformatics* **18**, 539-548 (2020).
59. Aarestrup, F.M. & Woolhouse, M.E.J. Using sewage for surveillance of antimicrobial resistance. *Science* **367**, 630-632 (2020).

60. Pärnänen, K.M.M. et al. Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. *Science advances* **5**, eaau9124 (2019).
61. Mao, D. et al. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. *Water research* **85**, 458-466 (2015).
62. Hou, L. et al. Fecal pollution mediates the dominance of stochastic assembly of antibiotic resistome in an urban lagoon (Yundang lagoon), China. *Journal of hazardous materials* **417**, 126083 (2021).
63. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)* **28**, 3150-3152 (2012).
64. Patel, R.K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* **7**, e30619 (2012).
65. Li, D., Liu, C.M., Luo, R., Sadakane, K. & Lam, T.W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)* **31**, 1674-1676 (2015).
66. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**, 119 (2010).
67. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology* **35**, 1026-1028 (2017).
68. Alcock, B.P. et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research* **48**, D517-D525 (2020).
69. Gibson, M.K., Forsberg, K.J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *The ISME journal* **9**, 207-216 (2015).
70. Bengtsson-Palme, J. et al. METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour* **15**, 1403-1414 (2015).
71. DeSantis, T.Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* **72**, 5069-5072 (2006).
72. Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R. & Schmidt, T.M. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res* **43**, D593-598 (2015).
73. Wu, L. et al. Microbial functional trait of rRNA operon copy numbers increases with organic levels in anaerobic digesters. *The ISME journal* **11**, 2874-2878 (2017).
74. Pärnänen, K. et al. Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nature communications* **9**, 3891 (2018).
75. Sun, J. et al. Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. *Nature communications* **11**, 1427 (2020).
76. Kang, D.D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
77. Wu, Y.W., Simmons, B.A. & Singer, S.W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics (Oxford, England)* **32**, 605-607 (2016).
78. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nature methods* **11**, 1144-1146 (2014).

79. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
80. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
81. Sieber, C.M.K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature microbiology* **3**, 836-843 (2018).
82. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)* **26**, 2460-2461 (2010).
83. 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 research* **25**, 1043-1055 (2015).
84. Rahman, S.F., Olm, a.M.R., Morowitz, a.M.J. & Msystems, b.J.F.B.J. Machine Learning Leveraging Genomes from Metagenomes Identifies Influential Antibiotic Resistance Genes in the Infant Gut Microbiome. **3**, e00123-00117 (2018).
85. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P. & Parks, D.H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics (Oxford, England)* **38**, 5315-5316 (2022).
86. Parks, D.H. et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature biotechnology* **38**, 1079-1086 (2020).
87. Asnicar, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature communications* **11**, 2500 (2020).
88. Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029 (2015).
89. Albertsen, M. et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**, 533-538 (2013).
90. Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications* **9**, 5114 (2018).
91. Saheb Kashaf, S., Almeida, A., Segre, J.A. & Finn, R.D. Recovering prokaryotic genomes from host-associated, short-read shotgun metagenomic sequencing data. *Nature protocols* **16**, 2520-2541 (2021).
92. Arango-Argoty, G. et al. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 23 (2018).
93. Oksanen, J. et al. vegan community ecology package version 2.6-2 April 2022. (2022).
94. Cayuela, L. & Gotelli, N.J. rareNMtests: Ecological and biogeographical null model tests for comparing rarefaction curves. (2014).
95. Wickham, H. Ggplot2: Elegant Graphics for Data Analysis. (ggplot2: Elegant Graphics for Data Analysis, 2009).
96. Derek H. Ogle, J.C.D., A. Powell Wheeler and Alexis Dinno FSA: Simple Fisheries Stock Assessment Methods. . (2023).
97. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics (Oxford, England)* **35**, 526-528 (2019).
98. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC bioinformatics* **11**, 367 (2010).
99. Goslee, S.C. & Urban, D.L.J.J.o.S.S. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. **22**, 1-19 (2007).

100. Zhang, Y. et al. Experimental warming leads to convergent succession of grassland archaeal community. *Nature Climate Change* **13**, 561-569 (2023).
101. Guo, X. et al. Climate warming accelerates temporal scaling of grassland soil microbial biodiversity. *Nature Ecology & Evolution* **3**, 612-619 (2019).
102. Thévenot, E.A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *Journal of proteome research* **14**, 3322-3335 (2015).

Acknowledgements

The authors thank A. Al-Omari, R. Bart, D. Crowley, S. D. Hardeman, T. Hensley, G. Harwood, J. Keller, M. Taylor, and, B. Pathak for helping with sampling and metadata collection. This work was supported by the Office of the Vice President for Research at the University of Oklahoma. The data synthesis performed by Linwei Wu and Ya Zhang was also partially supported by the U.S. National Science Foundation (NSF) (EF-2025558). The contribution by Linwei Wu was also partially supported by the National Natural Science Foundation of China (grant 32371724). The data synthesis performed by Congmin Zhu and Ting Chen was supported by the National Natural Science Foundation of China (grants 82202299, 61872218 and 61721003), National Key R&D Program of China (2019YFB1404804), Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University Initiative Scientific Research Program, and Guoqiang Institute, Tsinghua University.

Author Contributions Statement

All authors contributed experimental assistance and intellectual input to this study. The original concept was conceived by Jiz.Z. Experimental strategies and sampling design were developed by Jiz.Z., X.W., T.P.C., Q.H., Z.H. and D.N. Sample collections were coordinated by Q.H., Ya.Z., D.N., X.W., T.P.C., B.Z., M.R.B., G.F.W., Jiz.Z. and other GWMC members. Q.T. and D.N. managed shipping. N.X., B.Z., Yu.Z., Y.W, D.N. and some GWMC members did DNA extraction. B.Z., and S.G. performed Shotgun sequencing with the help from Lin.W. Data analyses were performed by C.Z., Lin.W., Jian.Z., R.T., D.N., and Jiz.Z. The manuscript was written by C.Z., Lin.W., Jiz.Z. and D.N. with the help from B.E.R., L.A.-C., M.W., C.S.C., J.M.T., P.J.J.A., A.W., P.H.N., F.J., J.G., D.A.S., T.C. and Y.Y.

Competing Interests Statement

The Authors declare no competing interests.

Tables

Table 1. Differences of the AS resistomes between continents. PERMANOVA (Permutational multivariate analysis of variance using distance matrices) was performed based on Bray-Curtis dissimilarity matrix at the level of individual ARGs. The upper triangle (shaded grey) shows the F values of PERMANOVA, while the lower triangle shows the two-sided *p* values. Sample size: n = 6, 59, 14, 20, 106, and 21 biologically independent samples for Africa, Asia, Australasia, Europe, North America, and South America, respectively.

	Africa	Asia	Australasia	Europe	North America	South America
Africa		5.04	2.18	5.78	2.66	2.03
Asia	0.001		13.60	37.64	39.14	10.93
Australasia	0.011	0.001		7.17	4.46	3.20
Europe	0.002	0.001	0.001		14.12	13.11
North America	0.003	0.001	0.001	0.001		7.94
South America	0.028	0.001	0.003	0.001	0.001	

Figure Legends/Captions

Figure 1. The abundance and global distribution of the AS resistomes. a) Map of the sampling locations. **b)** Average relative ARG abundance (copy of ARGs per cell) across different continents based on resistance mechanism, drug class, and the nine most abundant ARGs. MLS: macrolide-lincosamide-streptogramin. **c)** Richness of the ARGs. Richness index was calculated based on a rarified matrix of resistance gene coverage, which was rounded and subsampled to the lowest sample's level. In the boxplots, hinges show the 25, 50, and 75 percentiles. The upper whisker extends to the largest value no further than 1.5 * IQR from the upper hinge, where IQR is the inter-quartile range between the 25% and 75% quartiles; The lower whisker extends to the smallest value at most 1.5 * IQR from the lower hinge. Sample size: n = 6, 59, 14, 20, 106, and 21 samples for Africa, Asia, Australasia, Europe, North America, and South America, respectively. Significant differences (Dunn's test with two-sided *p*-values adjusted by the Bonferroni method < 0.05) between continent pairs are indicated in the plot. **d)** Principal coordinate analysis (PCoA) reveals distinct ARG composition diversity in six continents. **e)** PCoA reveals distinct ARG diversity in different environments. Source data are provided as a Source Data file.

Figure 2. The linkage of the AS resistomes to microbiomes. a) Relationships detected by Procrustes analysis between the resistomes and bacterial community structure as measured by 16S genes extracted from the metagenomes. Metagenomic shotgun sequencing was performed for all activated sludge samples, and the 16S sequences were extracted and grouped at the genus level using Metaxa2. **b)** Relationships detected by Procrustes analysis between the resistomes and bacterial community structure as measured by the 16S amplicon sequencing data. The dotted ends of lines represent the resistome position, while the undotted ends represent the bacteriome position. Vegan Procrustes test 'protest' with 999 permutations yielded a matrix-matrix correlation coefficient of 0.74 (protest, $p = 0.001$) for metagenome 16S-based bacterial community structure, and a matrix-matrix correlation coefficient of 0.70 (protest, $p = 0.001$) for 16S amplicon-based bacterial community structure. **c)** The association between the ARG abundance (total ARG abundance and the top four major ARG groups) and the relative abundance of top 15 major bacterial phyla from 16S rRNA gene amplicon data (16S) or metagenomes (Shotgun). The circle-filled color corresponds to Spearman's correlation coefficient. The asterisks '*' denote significant correlations (two-sided $p < 0.05$ after adjustment for multiple testing). **d)** The phylogenetic tree of metagenome-assembled genomes (MAGs) from global AS samples. The leaf colors indicate phylum groups. The bar heights outside the circle are respectively proportional to the ARG count annotated in MAGs, and those red bars represent the MAGs carrying multi-species mobile ARGs. Inner rings show the resistance gene abundances of the five major drug classes, with darker colors indicating higher abundances. **e)** The mean count and relative abundances of ARGs encoding major resistance mechanisms or drug classes across phylogenetic groups. Error bars indicate standard deviations. Numbers on the top indicate the number of MAGs belonging to the phylogenetic groups. Source data are provided as a Source Data file.

Figure 3. Mobility of ARGs from assembly and MAG-based analyses. a) Relative MGE abundance identified from the non-redundant ORFs on gene level and group level. **b)** Boxplots of the MGE Shannon's H index across six continents. Hinges show the 25, 50, and 75 percentiles. The upper whisker extends to the largest value no further than 1.5 * IQR from the upper hinge, where IQR is the inter-quartile range between the 25% and 75% quartiles; The lower whisker extends to the smallest value at most 1.5 * IQR from the lower hinge, and dots indicate values of individual samples. Sample size: n = 6, 59, 14, 20, 106, and 21 samples for Africa, Asia, Australasia, Europe, North America, and South America, respectively. Significant differences (Dunn's test with two-sided *p*-values adjusted by the Bonferroni method < 0.05) between continent pairs are indicated in the plot. **c)** The relative abundance of mobile or immobile ARGs based on taxonomic composition, resistance mechanisms and drug class. **d)** Multi-phyla mobile ARGs based on gene sharing between MAGs. Nodes represent ARG sequences with labels indicating the gene/gene family name. Node colors indicate the phylogenetic groups of MAGs in which the ARG is present. Node shapes indicate different resistance mechanisms. Source data are provided as a Source Data file.

Figure 4. Drivers for the AS resistomes. a) Normalized stochasticity ratio (NST) quantifies the relative importance of stochasticity in governing resistomes. Sample size: n = 6, 59, 14, 20, 106, and 21 samples for Africa, Asia, Australasia, Europe, North America, and South America, respectively. **b)** The Variance partition analysis (VPA) results indicated that the relative contributions of geographic distance (Geo), environmental variables (ENV), and their interactions to the variation of the AS resistomes all reached a significant level (two-sided $p < 0.05$). **c)** PLS models of the relationships among microbiome (PC1 of bacterial community structure), resistome (the total ARG abundance, PC1 of ARG composition, abundances of the top three resistance mechanisms), the abundance of MGEs located near (<10kb) ARGs, ARG-correlated environmental variables, and ecosystem functions (the removal rate of BOD, COD, total nitrogen, total phosphorus). Directions for all arrows are from independent variable to a dependent variable in the forward selected PLS models ($p < 0.05$); only the variables with variable influence on projection > 1 are presented. The numbers near the pathway arrow indicate the proportion of variance explained for every dependent variable, with the top row representing the partial R^2 index based on PLS and the bottom row representing Pearson correlation R^2. The asterisks denote the significance levels with *** $p < 0.01$, ** $p < 0.05$ and * $p < 0.10$ (two-sided). The colors of pathways are related to the positive (blue) or negative (red) relationships. The widths of pathways are related to the partial R^2 index. Source data are provided as a Source Data file.