Sewage-Borne Pathogens and Indigenous Denitrifiers are Active and Key Multi-Antibiotic Resistant Players in Wastewater Treatment Plants

Ling Yuan [1, 2], Yubo Wang [1, 2], Lu Zhang [1, 2], Alejandro Palomo [3], Jizhong Zhou [4], Barth F. Smets [3], Helmut Bürgmann [5], Feng Ju [1, 2] *

[1] Key Laboratory of Coastal Environment and Resources Research of Zhejiang Province, School of Engineering, Westlake University, Hangzhou, China

[2] Institute of Advanced Technology, Westlake Institute for Advanced Study, 18 Shilongshan Road, Hangzhou 310024, China

[3] Department of Environmental Engineering, Technical University of Denmark, Denmark

[4] Department of Microbiology and Plant Biology, Institute for Environmental Genomics, University of Oklahoma, Norman, OK, 73019, USA

[5] Department of Surface Waters - Research and Management, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Switzerland

*** Corresponding to:**

Dr. Feng Ju (Assistant Professor)

Address: Westlake University, 18 Shilongshan Road, Hangzhou 310024, China

Tel.: 571-87963205 (lab), 571-87380995 (office)

Fax: 0571-85271986

E-mail: jufeng@westlake.edu.cn

Abstract

The global rise and environmental spread of antibiotic resistance greatly challenge the treatment of deadly bacterial infections. Wastewater treatment plants (WWTPs) harbor and discharge antibiotic resistance genes (ARGs) as emerging environmental contaminants. However, the knowledge gap on the host identity and functionality limits rational assessment of transmission and health risks of ARGs emitted from WWTPs to the environment. Here, we developed an innovative genome-centric quantitative metatranscriptomic approach that breaks existing methodological limitations by integrating genome-level taxonomy and activity-based analyses to realize high-resolution qualitative and quantitative analyses of bacterial hosts of ARGs (i.e., multi-resistance, pathogenicity, activity and niches) throughout 12 urban WWTPs. We found that ~45% of 248 population genomes recovered actively expressed resistance against multiple classes of antibiotics, among which bacitracin and aminoglycoside resistance in *Proteobacteria* was the most prevalent. Both sewage-borne pathogens and indigenous denitrifying bacteria were transcriptionally active ARG hosts, contributing ~60% of detected resistance activities. Remarkably, eighteen antibiotic-resistant pathogens from the influent survived wastewater treatment, indicating their potential roles as persistent and clinically relevant resistance disseminators. The prevalence of ARGs in most transcriptionally active denitrifying populations (~90%) may suggest their adaptation to local metabolic niches and antimicrobial stressors. While wastewater treatment dramatically reduced the absolute expression activities of ARGs (> 99%), their relative expression levels remained almost unchanged in the majority of resistant populations. The prevalence of active ARG hosts including globally emerging pathogens (e.g., *A. cryaerophilus*) throughout and across WWTPs prioritizes future examination on the health risks related with resistance propagation and human exposure in the receiving environment.

Key words: Antibiotic resistance; Wastewater treatment plants; Denitrifying and pathogenic bacteria; Genome-centric metatranscriptomics; Metagenome-assembled genomes

Introduction

The extensive use of antibiotics and the resulting accelerated bacterial resistance dissemination have largely promoted the rise of antibiotic resistance as one of the greatest global public health threats [1,2]. Most of the antibiotic wastes together with antibiotic resistant bacteria and antibiotic resistance genes (ARGs) emitted from anthropogenic sources in urban areas eventually enter wastewater treatment plants (WWTPs) which are considered as hotspots for the release of ARGs and their hosts into the environment [3-5]. The prevalence and high diversity of ARGs in WWTPs have been widely noted [4,6-8] through metagenomic approaches [9,10]. However, the fragmented nature of reported metagenomic assemblies and the lack of activity-based resistome monitoring make it impossible to solidly predict either the identity of active ARG hosts or their important functional traits (e.g., decontamination, multi-resistance and niche breadth), restraining objective evaluation of environmental transmission and health risks of antibiotic resistance.

Theoretically, genome-centric metatranscriptomics can overcome the above technical bottlenecks by providing both high-resolution genome-level taxonomy and functional characterization and global gene expression activities of environmental microorganisms. So far, few studies have exploited this methodology to explore the host identity and activity of ARGs in WWTPs [11] where microorganisms are the functional agents of water purification and environmental protection [12-14]. Although functional bacteria [12,14,15] and ARGs [6,16,17] in WWTPs were extensively studied independently through culture-independent approaches [13,18,19], the extent to which indigenous microbes and especially the key functional bacteria in different compartments of WWTPs may represent hitherto-unrecognized recipients or even disseminators of ARGs remains unknown. This is of particular interest as the locally-adapted microbes dedicated to organic and nutrients removal in the activated sludge are under continuous and long-term exposure to subinhibitory levels of antimicrobial contaminants (e.g., antibiotics, heavy metals and biocides). Considering

the fact that enteric microbes including pathogens are being continuously introduced into WWTPs with sewage inflow, their regular close contact with indigenous microbes that are potentially under stress from exposure to antimicrobials may create conditions where resistance exchange involving pathogens followed by multi-resistance selection and potential local niche adaptation is favored (Fig. 1). This may represent expectable but not yet evaluated ecological and health risks[20]. Efforts are needed to fill all these knowledge gaps on the antibiotic resistance in WWTPs with improved methodology.

In this study, metagenome-assembled genomes (MAGs) analysis integrated with quantitative metatranscriptomics was exploited to answer the following questions about bacterial populations hosting ARGs in the 12 urban WWTPs. First, who are the ARG hosts and what are their functional roles in the WWTPs? Second, what are the potential ecological risks associated with antibiotic resistance in the WWTPs? For example, could some ARGs likely be hosted by pathogens? Which ARGs may transfer between indigenous and/or pathogenic bacteria? And who are the important ARG hosts that actively express ARGs throughout and across WWTPs especially in the treated effluent? To address these questions, we first resolved the genome phylogenies of active ARG hosts in the WWTPs and found *Proteobacteria* and *Actinobacteriota* as the two most common bacterial hosts. We then checked the multi-resistance, pathogenicity, distribution, activity, survival and other functional traits (e.g., biological nitrogen removal) of all the identified ARG hosts from the WWTPs, leading to the key finding that sewage-borne pathogens and indigenous denitrifiers are active and key players of wastewater (multi-)antibiotic resistance (Fig. 1).

Methods

Genome-centric reanalysis of WWTPs microbiome data

Between March and April 2016, a total of 47 microbial biomass samples were taken

from the primarily clarified influent, the denitrifying bioreactors, the nitrifying bioreactors and the secondarily clarified effluent of 12 urban WWTPs across Switzerland. Total DNA and RNA extractions, processing of the mRNA internal standards , data pretreatment, and metagenome assembly were performed as previously described in our earlier publication[6]. The metagenomes and metatranscriptomes generated by our preliminary study have been used to gain a comprehensive overview on the fate and expression patterns of known antibiotic, biocide and metal resistance genes throughout the varying compartments of the WWTPs[6]. However, the identity, multi-resistance, pathogenicity, distribution, activity and other functional traits of ARG hosts remained unknown, due to the fragmented nature of metagenome assemblies obtained. In this study, we filled this knowledge gap by re-analysis of the datasets using a genome-centric metatranscriptomic strategy as described below. The generated sequence datasets are deposited in China National GeneBank (CNGB) with an accession number CNP0001328.

Genome binning, annotation and phylogenetic analysis

Metagenome-assembled genomes (MAGs) were recovered using MetaWRAP (v1.2.2)[21] pipeline. Briefly, with metaBAT2 in the binning module, MAGs were reconstructed from the 47 single-sample assemblies. Contamination and completeness of the recovered MAGs were evaluated by CheckM (v1.0.12)[22], and only those genomes with quality score (defined as completeness – 5× contamination) ≥ 50 were included in the succeeding analysis. The draft genomes were dereplicated using dRep (v1.4.3)[23] with default parameters, which resulted in a total of 248 unique MAGs from all the 47 metagenome-datasets. The accession numbers of 248 recovered MAGs are listed in Dataset S2.

Taxonomy affiliation of each MAG was determined by GTDB-Tk (v0.3.2)[24] classify_wf. Open reading frames (ORFs) were predicted from MAGs using Prodigal (v2.6.3)[25]. Phylogenetic analysis of MAGs was conducted with FastTree (v2.1.10)[26] based on a set of 120 bacterial domain-specific marker genes from GTDB, and the

phylogenetic tree was visualized in iTOL[27].

ARG annotation and mobility assessment

The annotation of ARGs from the recovered MAGs was accomplished using DeepARG (v2)[28] with options '--align --genes --prob 80 --iden 50'. Predicted ARGs of antibiotic classes with less than 10 reference sequences in the database were removed to avoid mis-annotation due to possible bias. In total, 496 ORFs annotated in 162 MAGs were identified as ARGs with resistance functions to 14 specific antibiotic classes, while 313 ORFs annotated in 117 MAGs were identified as ARGs of multidrug class and were listed in Dataset S3 but not included in the downstream analysis. The 248 high-quality MAGs were then categorized as "multi-resistant" (113), "single-resistant" (49) and "non-resistant" (86), according to whether > 1, = 1, or = 0 ARG classes were annotated in the genome, respectively.

Considering the importance of the plasmid for spreading ARGs, the presence of plasmid sequences in the metagenomic contigs was checked by PlasFlow (v1.1)[29] which utilizes neural network models trained on full genome and plasmid sequences to predict plasmid sequences from metagenome-assembled contigs. A strict parameter '--threshold 0.95' was employed to robustly compare the occurrence frequencies of plasmid contigs in the binned (i.e., MAGs) and un-binned contigs. Moreover, mobile genetic elements (MGEs) were additionally identified by hmmscan[30] against Pfam[31], with options '--cut-ga'. ARGs with potential mobility were defined as either located on the plasmid contig or shared a nearby area (< 10 kb) with an MGE[32].

Identification of pathogenic genomes

The potentially pathogenic genomes were firstly identified referring to two published reference pathogen lists that contained 140 potentially human pathogenic genera[33] and 538 potentially human pathogenic species[34]. 3642 experimentally verified virulence factors downloaded from pathogenic bacteria virulence factor database (VFDB, last update: Jun 27 2020)[35] were then used to construct a blast database. ORFs from

taxonomically predicted potentially pathogenic genomes were searched against the constructed virulence factor database by BLASTN, and those genomes with an ORF with global nucleic acid identity > 70% to any virulence factor were considered as potential pathogens.

Nitrification-denitrification genes annotation

To explore certain functional traits (i.e., biological nitrogen removal driven by nitrification and denitrification in WWTPs) of ARG hosts in the WWTPs, nitrification-denitrification genes (NDGs) were annotated. Briefly, all MAG-predicted ORFs were searched against a nitrogen cycle database (NCycDB)[36] using diamond. Those ORFs annotated as nitrification or denitrification genes with global nucleic acid identity > 85% to the reference sequences in the NCycDB database were directly interpreted as functional genes related to nitrogen removal in the WWTPs. Other ORFs were further checked by BLASTN against the NCBI nt database, ORFs with global nucleic acid identity > 70% to the reference sequences were also identified as annotatable functional genes. Together, 283 ORFs from 88 MAGs were annotated as NDGs. With the intention to display the distribution patterns of NDGs in the MAGs, a network was constructed and visualized in Gephi (v0.9.2) [37]. The network was divided into seven parts according to nitrification (3) and denitrification (4) pathway steps.

Quantitative analyses of genome-centric metatranscriptomics

Quantification at the genome level. In order to calculate the relative abundance and the expression level of each MAG, 47 metagenomic datasets of clean DNA reads and 47 metatranscriptomic datasets of clean mRNA reads were mapped across the 47 individual assemblies and the 47 ORF libraries using bowtie2 (v2.3.4.1)[38], respectively. The resulting .sam files contained mapping information of both MAGs and un-binned contigs, and subsequent filtering extracted mapping results of each MAG. Then, the relative abundance and expression level of each MAG was

calculated and normalized to RPKM (reads per kilobase per million) values as the total number of bases (bp) that mapped to the genome, divided by the MAG size (bp) and the sequencing depth (Gb).

Quantification at the gene level. To overcome the limitation of relative abundance in the metatranscriptomic analysis[39], absolute expression values (AEV) were calculated for the 496 ARGs annotated in the 248 high-quality MAGs based on spiked mRNA internal standards [6] and mapping results. Briefly, AEV was calculated using the following equation:

Absolute expression value (AEV, copies/g) =

$$\frac{N_{added\,standard\,copies}}{m_{biomass}} \times \frac{N_{gene-mapped\,reads}/L_{gene\,sequence}}{N_{recovered\,standard\,reads}/L_{standard\,sequence}} \qquad (1)$$

where $N_{\text{added standard copies}}$ is the copy numbers of added mRNA internal standards, $m_{biomass}$ is the mass of collected volatile suspended solids, $N_{gene-mapped\,reads}$ is the number of reads mapped to the gene in the metatranscriptomic dataset, $L_{gene\,sequence}$ is the length of the gene, $N_{recovered\,standard\,reads}$ is the number of reads mapped to the mRNA internal standard in the metatranscriptomic dataset, $L_{standard\,sequence}$ is the length of the mRNA internal standard. This calculation is optimized by weighing different lengths of reported genes, and only genes with >50% of their lengths covered by mapped reads were considered. In this study, if the sample range is not otherwise specified, AEV of a gene refers to the average AEV across all 47 samples.

While AEV is the absolute expression activity of a given gene, relative expression ratio (RER) is a comparison between the given gene and the single-copy marker genes in the genome, which calculated by relativizing the AEV of the given gene by the median AEV of the single-copy marker genes in the genome as shown below:

$$\text{Relative expression ratio (RER)} = \frac{AEV_{gene}}{median(AEV_{single-copy\,marker\,genes})} \qquad (2)$$

The single-copy marker genes in the recovered genomes were determined by

GTDB-tk[24] which searched 120 ubiquitous single-copy marker genes of bacteria[40] in the genome, and those unique marker genes in the genome were used to calculate basic expression level of the genome. Ideally, if RER > 1, this gene would be regarded as over expressed compared with the house-keeping marker genes, and if RER = 1, it indicates that this gene expresses at a same level as the marker genes. Similarly, if RER < 1, it indicates that this gene is under expressed compared with the marker genes. Our proposal of these two metrics (i.e., AEV and RER) offer complementary insights into a given gene of interest: AEV quantifies its absolute expression activity in a sample, thus proportionally corresponds to the changing concentration of its host cells within a given microbial community, while RER measures its relative expression compared with basic expression level of its host genome. Thus, RER is a more sensitive parameter to monitor microbial response to environmental changes. Finally, the aggregate AEV and average RER of ARGs in the genome were used to represent the absolute and relative expression activity of the antibiotic resistance function in this genome, respectively.

Statistical analysis

All statistical analyses were considered significant at $p < 0.05$. The similarity of microbial community structure between the nitrification and denitrification bioreactors was examined by mantel test in R using the function 'mantel' in the vegan package[41]. The difference of relative expression ratio of individual ARGs and ARGs in the recovered MAGs between the influent and effluent wastewater was determined by Mann-Whitney U test using function 'wilcox.test' with option 'paired=FALSE'. The test of difference in relative expression ratio of ARGs between the four compartments was performed with Kruskal-Wallis test in python using function 'kruskalwallis' in scipy package. The average RER of ARGs and denitrification genes in the MAGs was calculated after removing outliers (based on the 3σ principle).

Results and discussion

Metagenome-assembled genomes recovered from the WWTPs microbiome

The key functions of urban WWTPs such as removal of organic carbon and nutrients are largely driven by uncultured microorganisms [13,14]. To explore the key microbial functional groups including uncultured representatives, 1844 metagenome-assembled genomes (MAGs) were reconstructed from 47 samples taken from varying compartments in the 12 Swiss WWTPs. A total of 248 unique and high-quality MAGs were retained for further analysis after dereplication and quality filtration. These genomes accounted for 14-62% and 7-75% of paired metagenomic and metatranscriptomic reads, respectively, and therefore represented a significant fraction of the microbial community in the WWTPs (Dataset S1). Basic information on the MAGs recovered was listed in Dataset S2. Phylogenetic analysis based on 120 single-copy marker genes of the 248 MAGs showed their grouping and taxonomic classification into 15 phyla (Fig. 2). The phylum with the largest number of MAGs recovered was *Proteobacteria* (88), followed by *Patescibacteria* (68), *Bacteroidota* (39), *Actinobacteriota* (22), *Firmicutes* (11) and *Myxococcota* (4). The phylum-level microbial community composition in the 12 WWTPs were overall similar to a recent study that recovered thousands of MAGs from activated sludge of global WWTPs that were also mostly assigned to *Proteobacteria*, *Bacteroidota* and *Patescibacteria*[42].

Further comparisons of abundance percentage and expression percentage of the 248 MAGs across all samples clearly showed distinct DNA- and mRNA-level compositional profiles across phyla and genomes. Overall, 3, 22 and 88 MAGs assigned to *Campylobacterota*, *Actinobacteriota* and *Proteobacteria* exhibited a high average abundance percentage of 1.9%, 0.8% and 0.6%, corresponding to an average expression percentage of 4.1%, 0.7%, and 0.7%, respectively. In contrast, *Patescibacteria* showed low average abundance percentage (0.12%) and expression percentage (0.01%). This newly defined superphylum, belonging to a recently

discovered candidate phylum radiation [43,44], was found to be the second most frequent populations in the 12 WWTPs of this study. These *Patescibacteria* populations, however, might have been overlooked by previous large-scale 16S rRNA-based surveys [13,15,45] due to their special 16S rRNA gene features, i.e., they appear to encode proteins and have self-splicing introns rarely found in the 16S rRNA genes of bacteria [46]. Our first discovery of their survival at extremely low gene expression level (Fig. 2) calls for further investigation of the original sources and potential functional niches of these ultra-small populations (cells < 0.2 μm) in WWTPs [47].

Host identity, expression activities and mobility of ARGs

To understand taxonomic distribution and activity of ARGs in the MAGs recovered from the WWTPs, a genome-centric metatranscriptomic approach was exploited to examine ARGs in genomic and transcriptomic contexts of all 248 genomes. Together, 496 ORFs carried by 162 (65.3%) MAGs were identified as ARGs encoding resistance functions of 14 antibiotic classes (Dataset S3). The predicted 162 ARG hosts were further categorized as "multi-resistant" (113 MAGs, 45.6%) and "single-resistant" (49 MAGs, 19.8%) (Fig. 3a, Dataset S2). Among those multi-resistant MAGs, W60_bin3 and W72_bin28 affiliated with *Aeromonas media* and *Streptococcus suis*, respectively, were found to harbor the largest numbers of ARGs, i.e., they both carried 11 ARGs conferring resistance to 9 and 4 antibiotic classes, respectively, followed by 3 MAGs from *Aeromonas media* (2) and *Acinetobacter johnsonii* (1) that carried 10 ARGs (Fig. 3a and Dataset S2).

Taxonomically, ARG hosts were found in 11 out of 15 phyla (except for *Verrucomicrobiota* A, *Bdellovibrionota*, *Nitrospirota and Gemmatimonadota*, each containing no more than 2 MAGs) (Fig. 3b). MAGs assigned to the phylum of *Proteobacteria* were the most frequent hosts of ARGs encoding resistance of 13 antibiotic classes. Eighty-four out of the 88 *Proteobacteria*-affiliated MAGs were ARG hosts and nearly all of them (83 out of 84 MAGs) were transcriptionally active for resistance to at least one antibiotic class (Fig. 3b). *Actinobacteriota* were also

active hosts of ARGs of 10 antibiotic classes, especially for glycopeptide and tetracycline (Fig. 3b). In contrast, *Patescibacteria* were transcriptionally inactive hosts of ARGs, i.e., 10 out of 68 MAGs encoded ARGs with only one population (W73_bin6) displaying transcription of beta-lactam and aminoglycoside resistance (Fig. 3b, Dataset S4). *Patescibacteria* were recently revealed to harbor small but mighty populations with reduced genomes (~1 Mbp) and usually truncated metabolic pathways[48], and an under representation of ARGs in their genomes may be a strategic outcome from their process of reducing redundant and nonessential functions.

Among the 14 resistance types of ARGs identified (Fig. 3c), bacitracin (78, 31.5%) and aminoglycoside (68, 27.4%), being most prevalent in *Proteobacteria*, were found to be the two most frequent resistance types, followed by beta-lactam (47, 19.0%) and fosmidomycin (45, 18.1%). In contrast, sulfonamide- (2, 0.8%) and chloramphenicol-resistance (1, 0.4%) were both hosted by few MAGs, all belonging to *Proteobacteria* (Fig. 3b). Absolute quantification revealed that the sulfonamide resistance genes showed the highest expression level with an average AEV of 2.53×10^{11} copies/g, followed by those of tetracycline (1.51×10^{11} copies/g) and peptide (1.46×10^{11} copies/g). In contrast, the fluoroquinolone resistance genes displayed the lowest average AEV (1.42×10^{9} copies/g, Dataset S4). Among all 496 ARGs, 460 ARGs were confirmed to have transcriptional activity in at least one sample (Dataset S4). This indicated that most ARGs are expressed under the environmental condition of the WWTPs. ARG expression could be induced by specific antibiotics or their co-selective or -expressive antimicrobial agents (e.g., other antibiotics and heavy metals) in wastewater, but may also be constitutively expressed or only under global control. These results reveal that multiple antibiotic resistance was widely distributed and expressed in the WWTPs microbiome.

Plasmid is an evolutionarily important reservoir and transfer media for ARGs. From our study, 11 ARGs were found to locate on the plasmid contigs (Dataset S5), three of which were carried by potential pathogens (see Fig. 4) i.e., *tet*39 and ANT(3")-IIc

carried by *Acinetobacter johnsonii* and *lnu*A carried by *Streptococcus suis* as later discussed. It is notable that plasmid sequences, especially when present in multi-copies or shared across bacteria, are largely excluded from and poorly represented in the reconstructed genomes which are supposed to consist of single-copy genomic regions with nearly the same coverage [49]. For example, our firsthand data from one WWTP showed that only 2.3% contigs from MAGs were predicted by PlasFlow as plasmid sequences, while 6.4%, 7.1%, 7.7% and 9.9% contigs from un-binned contigs assembled from influent, denitrifying sludge, nitrifying sludge and effluent metagenomes were predicted as plasmid-originated. Besides, 35 ARGs identified from the MAGs were located near to an MGE (<10kb) including six cases that ARG and MGE are directly adjacent on the same contigs (Dataset S5). These results together reveal considerable mobility and thus dissemination potentials of wastewater ARGs mediated by plasmids or other MGEs.

Pathogenicity, distribution and activities of ARG hosts across WWTP compartments

Whether environmental ARGs are hosted by clinically relevant pathogens is central to assessing their health risks. Compared with reported metagenomic contigs or gene fragments [50,51], MAGs provide a more complete genome context allowing for more robust host identification at higher resolution, down to the species or even strain level. In total, 20 potentially pathogenic MAGs were identified based on the published reference pathogen lists [34] and verified the presence of virulence factors. Seventeen out of the 20 pathogenic MAGs were found to encode multi-antibiotic resistance, and the aforementioned five MAGs that encode the largest number of ARGs (10 or 11) all belonged to the pathogenic group. These potentially pathogenic organisms were transcriptionally active and dominant in the influent (Fig. 4, Fig S1), suggesting the possibility that they originated from the human intestinal tract. It is noteworthy that members of pathogenic group were almost absent in the downstream denitrifying and nitrifying bioreactors but were observed again in the effluent where they were not

completely eliminated (Fig. 4). We suspected that these influent-abundant pathogens were mainly planktonic cells that generally failed to invade or inhabit activated sludge flocs, but passively drifted into the final effluent with wastewater flow. Among the 20 pathogenic MAGs, 3 were assigned to *Aliarcobacter cryaerophilus*, a globally emerging foodborne and zoonotic pathogen[52]. Although these 3 pathogens were classified as either non-resistant or single-resistant, their considerable transcriptional activities in the effluent (average RPKM in effluent > 1, Dataset S6) deserve further attention. The 9 MAGs classified as *Aeromonas media*, a well-known gram negative, rod-shaped and facultative anaerobic opportunistic human pathogen[53], were all identified as being resistant to more than three classes of antibiotics and transcriptionally active in the effluent (RPKM in effluent: 0.58~0.67, Dataset S6). In addition, other potential pathogens survived wastewater treatment included *Acinetobacter johnsonii* (4 MAGs), *Streptococcus* (3 MAGs) and *Pseudomonas fluvialis* (1 MAG) (Fig. 4 and Dataset S6). Together, 18 (multi-)antibiotic resistant pathogens from the wastewater influent may have roles as persistent pathogenic agents and ARGs disseminators in the WWTP effluents, as they could successfully enter into the receiving rivers where health risks associated with their local propagation, resistance transfer and human exposure call for research attention.

The comparative profiles in relative abundance and expression level of the 162 ARG hosts as well as the 86 non-resistant MAGs across 47 samples showed that both the population distribution and the expression profiles dramatically shifted across influent, denitrification, nitrification, and effluent compartments (Fig. S1), probably driven by environmental heterogeneity and habitat filtering. Interestingly, although the denitrification and nitrification compartments differed significantly (paired t-test $p < 0.001$) in dissolved oxygen (0.02 vs. 2.04 mg/L), organic carbon (14.3 vs. 11.2 mg/L), ammonia nitrogen (8.1 vs. 1.9 mg/L), nitrate nitrogen (3.9 vs. 9.3 mg/L) and hydrolytic retention time (3.9 vs. 8.3 days) [6], the two compartments shared almost the same genomic and transcriptomic composition (mantel statistic r = 0.900 and 0.957, *p*

< 0.001; Fig. 4), suggesting that a set of core species survive and thrive in the classic anoxic-aerobic cycles of activated sludge process. Unlike the tightly clustered profiles in the influent, the effluent had highly dispersive population distribution and expression patterns that partially resembled those of activated sludge and influent, revealing prominent impacts from wastewater treatment and diverse emission of live resistant bacterial populations.

Multi-antibiotic resistance associated with biological nitrogen removal

Biological nitrogen removal is one of the key goals of wastewater treatment processes driven by nitrifiers and denitrifiers which were found to be closely associated with antibiotic resistance in this study. Together, 88 MAGs were found to be potentially involved in wastewater nitrogen removal (Dataset S2). Compared with nitrification, a much higher diversity of microbes (7 phyla vs. 3 phyla, 87 vs. 5 unique MAGs) showed genetic potential for denitrification/truncated denitrification, and it was noteworthy that 4 MAGs simultaneously expressed denitrification and nitrification genes (Dataset S2). This finding from WWTP systems echoes the widely accepted ecological concepts that nitrification is often carried out by specialist taxa while denitrification can involve a wide range of taxa [54]. Detailed description of nitrification-denitrification genes (NDGs) distribution in the 88 MAGs is available in the Supplementary Information S1 which suggests the presence of these functional bacteria and genes as the basis for biological nitrogen removal from wastewater.

Among these MAGs, a portion of nitrifying populations (3/5 MAGs) and most of denitrifying (without nitrifying) populations (75/83 MAGs) were multi-resistant (71/88 MAGs) or single-resistant (7/88 MAGs), while the majority of non-resistant populations (76/86 MAGs) were not involved in either nitrification or denitrification (Fig. 5b), revealing antibiotic resistance may be an important trait for successful survival and routine functioning of nitrogen-removing bacteria under WWTP conditions, i.e., in the presence of wastewater-borne antimicrobial stressors. The two ammonia-oxidizing MAGs classified as *Nitrosomonas* (W68_bin8 and W79_bin32),

both expressed resistance to bacitracin, and W68_bin8 additionally expressed resistance to fosmidomycin and tetracycline. The two nitrite-oxidizing MAGs classified as *Nitrospirota* (W81_bin21 and W77_bin34) did not encode detectable ARGs. Besides, 306 out of 496 ARGs were in the MAGs of potential denitrifiers, revealing that denitrifying bacteria are important previously unrecognized hosts of diverse ARGs in the WWTPs (Dataset S2). When both taxonomic affiliation and nitrogen removal function of the 248 MAGs were considered, we found that multi-antibiotic resistant *Proteobacteria* (58/88 MAGs, 65.9%) played a predominant role in the nitrification and denitrification, while *Patescibacteria* (66/68 MAGs, 97.1%) and *Bacteroidota* (26/39 MAGs, 66.7%) were dominated by non-resistant or single-resistant populations without a detectable NDG (Dataset S2). Combined, the above results strongly indicate the high prevalence of ARGs in nitrogen-removing functional organisms, especially denitrifying *Proteobacteria*, a hitherto unrecognized hotspot of multi-antibiotic resistance in WWTP systems, and suggest that antibiotic resistance is a trait of microbes performing a central function of the WWTP process, and can thus likely not be easily removed from these systems.

Differential antibiotic-resistant activities across WWTP compartments

The absolute expression and relative expression levels of ARGs were examined both in the functional groups involved in nitrogen removal (Fig. 6a) and other resistant members (Fig. 6b) across WWTP compartments. Notably, 14 out of 18 resistant pathogens were also identified as denitrifiers, thus may assist biological nitrogen removal from wastewater (Fig. 6a). The 18 pathogenic populations (e.g., MAGs from *Acinetobacter johnsonii* and *Aeromonas media*) were found actively expressing ARGs in the WWTPs, and they overall contributed to ~38% of ARGs expression in the recovered MAGs (Fig. 6a and Dataset S7). Although nitrifiers were overall not active in the expression of ARGs (e.g., W68_bin8 from *Nitrosomonas*: 2.04×10^{8} copies/g, W68_bin12 from *Caldilineales*: 4.03×10^{8} copies/g), some denitrifiers, especially those indigenous denitrifiers (shared > 95% total activities of denitrification genes in

the nitrifying and denitrifying sludge, ≤ 5% total activities in the influent and effluent) highly expressed ARGs in the WWTPs (e.g., 3 MAGs from *Phycicoccus* and 2 MAGs from *Tetrasphaera* > 6×10^{11} copies/g). This contrasting pattern between nitrifying and denitrifying bacteria suggests considerable differences in their resistance response and survival strategy to tackle antibiotic stresses in wastewater. Together, the resistant members from sewage-borne pathogenic group (marked in red, Fig. 6) and indigenous denitrifying group (marked in green, Fig. 6a) were both found to be key hosts of ARGs actively expressing resistance in the WWTPs: these two groups contributed ~60% of ARGs expression in the recovered MAGs (Dataset S7).

Of the 64 resistant MAGs without an identifiable NDG but expressed ARGs in the WWTPs, 35 MAGs primarily expressed antibiotic resistance in the nitrifying and denitrifying bioreactors (>95% total activities) rather than in the influent and effluent (≤ 5% total activities, Fig. 6b, Dataset S7). These indigenous resistant bacteria of activated sludge were dominated by populations of phylum *Bacteroidota* (15 MAGs), *Proteobacteria* (11 MAGs) and *Actinobacteriota* (8 MAGs, Fig. 6b). For instance, chemoorganotrophic *Microthrix* (3 MAGs) are associated with activated sludge flocs formation and filamentous bulking[55], while chemolithoautotrophic *Gallionellaceae* (4 MAG assigned to UBA7399), a poorly characterized family in WWTPs microbiome, are known to harbor aerobic nitrite-oxidizing bacteria (e.g., *Nitrotoga* [56]) and ferrous iron-oxidizing bacteria [57]. Unsurprisingly, the absolute expression of ARGs decreased dramatically (>99%) in most effluent populations, due to the efficient removal of bacterial cells in the WWTPs (e.g., 88%-99%[6]). However, the effluent had maintained high absolute expression (AEV > 1×10^{10} copies/g) of multi-antibiotic resistance in 6 populations, among which, two denitrifying *Malikia spinosa* strains (Fig. 6a) and one *Beggiatoaceae* spp. (Fig. 6b) were identified as the three most pronounced contributors of multi-antibiotic resistant activities in the effluent microbiota (8.88×10^{10}, 2.56×10^{10} and 1.67×10^{10} copies/g, respectively).

While the comparative profiles of absolute expression (i.e., AEV dynamics)

enable us to sort out host bacteria actively expressing ARGs, RER provides an additional insight into the relative expression and regulation of ARGs under varying wastewater stresses and environmental changes throughout WWTPs. Overall, relative expression of ARGs were only ~0.4-fold of the average expression level of the single-copy genes in the host genomes, implying that antibiotic resistance was a generally inactive function with below-average expression level in the WWTP microbiome. Moreover, most ARGs exhibited relatively stable RER dynamics across compartments (Fig. 6). Of 130 MAGs that expressed ARGs in the influent and/or effluent, only 16 (e.g. 2 MAGs from *Zoogloea*) showed significant decrease (Mann-Whitney FDR-$p < 0.05$) in the RER of ARGs from influent to effluent, and 27 (e.g. 5, 3, 3, 2 MAGs from *Aeromonas media*, *Acinetobacter johnsonii*, *Phycicoccus* and *Nitrosomonas*) showed significant increase (Mann-Whitney FDR-$p < 0.05$) in the RER of ARGs. In contrast, no significant change was observed for the remaining majority MAGs (87/130, 66.9%) (Dataset S8). This result was consistent with the observation at the level of ARGs (Supplementary information S2), indicating that antibiotic resistance activities were overall weakly affected by changing environmental conditions within WWTPs. However, the expression pattern of NDGs was quite different from that of ARGs. The relative expression of denitrification genes and nitrification genes were 4.6-fold and 80.1-fold of average level in the host genomes, respectively, indicating that biological nitrogen removal is a functionally important and metabolically active bioprocess in the WWTPs. The significant upregulation (Mann-Whitney FDR-$p < 0.05$) of denitrification genes from influent to the downstream activated sludge bioreactors was noted in ~53% of the denitrifiers (46/87 MAGs) (Dataset S8). Notably, two multi-resistant denitrifying populations assigned to *Rhodocyclaceae* and *Flavobacterium* (Fig. 6a), together with four functionally unassigned populations associated with *Streptococcus*, *GCA-2746885*, *49-20* and *UBA9655* (Fig. 6b), actively expressed ARGs (RER >1) across the four treatment compartments. Therefore, these persistently active resistant populations were important reservoirs of wastewater-borne antibiotic resistance.

Research significance and methodological remarks

To the best of our knowledge, this is the first study to gain so far the most complete insights into the key functional traits of ARG hosts in WWTPs based on both absolute expression activity of ARGs and their relative expression activity in the host genomes. Our findings demonstrated that pathogens carried by sewage influent and indigenous activated sludge denitrifiers in the WWTPs were important living hosts and hotspots of ARGs in which multi-antibiotic resistance genes were not only present but also expressed even in the treated effluent. Further, the almost unchanged relative expression of ARGs in the majority of resistant populations and those resistant bacteria surviving wastewater treatment indicate that these populations are robust under environmental conditions and leave the WWTPs alive, raising environmental concerns regarding their role forin dissemination of multi-antibiotic resistance into downstream aquatic ecosystems. Future studies are thus needed to examine the propagation and health risks of wastewater-derived multi-antibiotic resistance determinants with regards to their ability to successfully colonize the receiving environment of and/or regarding human exposure to their pathogenic hosts via such environmental reservoirs.

Our study also demonstrates a new methodological framework that integrates metagenome-centric genomic and quantitative metatranscriptomic analyses to overcome the limitations of existing mainstream metagenomic approaches widely employed for host tracking and risk assessment of environmental ARGs: (i) poor taxonomic resolution, (ii) lack of resistance activity monitoring, and (iii) lack of absolute quantification of ARGs. This new meta-omics framework is not only directly applicable for host tracking of ARGs in other environmental samples or of functional genes other than ARGs, but also sets a foundation for developing related bioinformatics pipelines and tools. Despite the demonstrated power of the framework in resolving key host traits of ARGs, its metagenome-assembled genome analysis necessarily focused on chromosomal ARGs while underestimated plasmid ARGs,

although we also recovered resistance contigs of plasmid origin from the MAGs recovered (Dataset S5). Notably, it is hard to link (mobile) multi-resistance plasmids with their host phylogeny with the same confidence as for chromosomal MAGs, nor can it be completely excluded that the bacteria from the studied MAGs do not harbor additional ARG on plasmids, whether or not ARG can be identified on their host chromosomes. As the importance of plasmids for spreading antibiotic resistance is well known, the current approach cannot capture the full picture of ARG-host relationships. On the other hand, the high representativeness of the transcribed resistome covered by the studied MAGs (65%) indicates that an important part of the resistance load is represented. This limitation of our study would, at least in theory, be circumventable by a massive application of single-cell genomics although at present this approach would still be limited in practice by cost and labor considerations.

Conclusions

Using genome-centric metatranscriptomic approaches, this study resolves the key functional traits (i.e., identity, multi-resistance, pathogenicity and activity) and metabolic niches of ARG hosts throughout urban WWTPs. The distribution and expression activities of the resistant populations significantly changed across WWTP compartments, driven by habitat filtering and sharp environmental gradients. Both long-lasting indigenous activated sludge denitrifiers and sewage-borne human pathogens were identified as major contributors (~60%) to ARGs activities in the recovered genomes. Although wastewater treatment dramatically reduces the total transcriptional activities of ARGs (> 99% activities), the relative expression of ARGs in the majority (66.9%) of the host MAGs remained almost unchanged from the influent to final effluent except for upregulated expression noticed in some populations (20.7%). This study provides new genome-centric metatranscriptomic insights into the identity and functions of ARG hosts throughout WWTPs, which were previously poorly understood dimensions essential for an improved understanding on

the environmental prevalence and spread of antibiotic resistance.

Acknowledgements

This research was supported by Natural Science Foundation of China via Project 51908467. FJ would like to also thank The National Key Research and Development Program of China (grant no. 2018YFE0110500) for financial support. We thank Dr. WeiZhi Song at the UNSW Sydney and Mr. Guoqing Zhang at the Westlake University for helpful discussion on the part of the bioinformatics procedures.

Conflict of interest

The authors declare no conflict of interest.

References

1 Pruden, A., Pei, R., Storteboom, H., Carlson, K. H. J. E. s. & technology. Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. **40**, 7445-7450 (2006).

2 Berendonk, T. U. *et al.* Tackling antibiotic resistance: the environmental framework. *Nat. Rev. Microbiol.* (2015).

3 Michael, I. *et al.* Urban wastewater treatment plants as hotspots for the release of antibiotics in the environment: a review. *Water Res* **47**, 957-995, doi:10.1016/j.watres.2012.11.027 (2013).

4 Guo, J., Li, J., Chen, H., Bond, P. L. & Yuan, Z. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Res* **123**, 468-478, doi:10.1016/j.watres.2017.07.002 (2017).

5 Karkman, A., Do, T. T., Walsh, F. & Virta, M. P. J. Antibiotic-Resistance Genes in Waste Water. *Trends Microbiol* **26**, 220-228, doi:10.1016/j.tim.2017.09.005 (2018).

6 Ju, F. *et al.* Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J* **13**, 346-360, doi:10.1038/s41396-018-0277-8 (2019).

7 An, X. L. *et al.* Tracking antibiotic resistome during wastewater treatment using high throughput quantitative PCR. *Environ Int* **117**, 146-153, doi:10.1016/j.envint.2018.05.011 (2018).

8 Mao, D. *et al.* Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. *Water Res* **85**, 458-466, doi:10.1016/j.watres.2015.09.010

575 (2015).

576 9 Arango-Argoty, G. *et al.* DeepARG: a deep learning approach for predicting antibiotic
577 resistance genes from metagenomic data. *Microbiome* **6**, 1-15 (2018).

578 10 Yang, Y., Li, B., Ju, F. & Zhang, T. Exploring variation of antibiotic resistance genes in activated
579 sludge over a four-year period through a metagenomic approach. *Environ. Sci. Technol.* **47**,
580 10197-10205 (2013).

581 11 Liu, Z. *et al.* Metagenomic and metatranscriptomic analyses reveal activity and hosts of
582 antibiotic resistance genes in activated sludge. *Environ Int* **129**, 208-220,
583 doi:10.1016/j.envint.2019.05.036 (2019).

584 12 Ju, F., Guo, F., Ye, L., Xia, Y. & Zhang, T. Metagenomic analysis on seasonal microbial variations
585 of activated sludge from a full-scale wastewater treatment plant over 4 years. *Environmental*
586 *microbiology reports* **6**, 80-89 (2014).

587 13 Wu, L. *et al.* Global diversity and biogeography of bacterial communities in wastewater
588 treatment plants. *Nature Microbiology*, doi:10.1038/s41564-019-0426-5 (2019).

589 14 Ju, F. & Zhang, T. Bacterial assembly and temporal dynamics in activated sludge of a full-scale
590 municipal wastewater treatment plant. *The ISME journal* **9**, 683 (2015).

591 15 Ju, F., Xia, Y., Guo, F., Wang, Z. & Zhang, T. Taxonomic relatedness shapes bacterial assembly
592 in activated sludge of globally distributed wastewater treatment plants. *Environ. Microbiol.*
593 **16**, 2421-2432 (2014).

594 16 Yang, Y., Li, B., Ju, F. & Zhang, T. Exploring variation of antibiotic resistance genes in activated
595 sludge over a four-year period through a metagenomic approach. *Environ Sci Technol* **47**,
596 10197-10205, doi:10.1021/es4017365 (2013).

597 17 Ju, F. *et al.* Antibiotic resistance genes and human bacterial pathogens: Co-occurrence,
598 removal, and enrichment in municipal sewage sludge digesters. *Water Res* **91**, 1-10,
599 doi:10.1016/j.watres.2015.11.071 (2016).

600 18 Boolchandani, M., D'Souza, A. W. & Dantas, G. Sequencing-based methods and resources to
601 study antimicrobial resistance. *Nat Rev Genet*, doi:10.1038/s41576-019-0108-4 (2019).

602 19 Xia, Y., Wen, X., Zhang, B. & Yang, Y. Diversity and assembly patterns of activated sludge
603 microbial communities: A review. *Biotechnol. Adv.* **36**, 1038-1047 (2018).

604 20 Bouki, C., Venieri, D. & Diamadopoulos, E. Detection and fate of antibiotic resistant bacteria
605 in wastewater treatment plants: a review. *Ecotoxicol Environ Saf* **91**, 1-9,
606 doi:10.1016/j.ecoenv.2013.01.016 (2013).

607 21 Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP-a flexible pipeline for genome-resolved
608 metagenomic data analysis. *Microbiome* **6**, 158, doi:10.1186/s40168-018-0541-1 (2018).

609 22 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing
610 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
611 *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

612 23 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic
613 comparisons that enables improved genome recovery from metagenomes through
614 de-replication. *ISME J* **11**, 2864-2868, doi:10.1038/ismej.2017.126 (2017).

615 24 Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify
616 genomes with the Genome Taxonomy Database. *Bioinformatics*,
617 doi:10.1093/bioinformatics/btz848 (2019).

618 25 <Prodigal prokaryotic gene recognition and translation initiation site identification.pdf>.

619 26 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees
620 for large alignments. *PLoS One* **5**, e9490, doi:10.1371/journal.pone.0009490 (2010).

621 27 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display
622 and annotation. *Bioinformatics* **23**, 127-128, doi:10.1093/bioinformatics/btl529 (2007).

623 28 Arango-Argoty, G. *et al.* DeepARG: a deep learning approach for predicting antibiotic
624 resistance genes from metagenomic data. *Microbiome* **6**, 23, doi:10.1186/s40168-018-0401-z
625 (2018).

626 29 Krawczyk, P. S., Lipinski, L. & Dziembowski, A. PlasFlow: predicting plasmid sequences in
627 metagenomic data using genome signatures. *Nucleic Acids Res* **46**, e35,
628 doi:10.1093/nar/gkx1321 (2018).

629 30 Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity
630 searching. *Nucleic Acids Res* **39**, W29-37, doi:10.1093/nar/gkr367 (2011).

631 31 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future.
632 *Nucleic Acids Res* **44**, D279-285, doi:10.1093/nar/gkv1344 (2016).

633 32 Sun, J. *et al.* Environmental remodeling of human gut microbiota and antibiotic resistome in
634 livestock farms. *Nat Commun* **11**, 1427, doi:10.1038/s41467-020-15222-y (2020).

635 33 Cai, L., Ju, F. & Zhang, T. Tracking human sewage microbiome in a municipal wastewater
636 treatment plant. *Appl Microbiol Biotechnol* **98**, 3317-3326, doi:10.1007/s00253-013-5402-z
637 (2014).

638 34 Zhang, S., He, Z. & Meng, F. Floc-size effects of the pathogenic bacteria in a membrane
639 bioreactor plant. *Environ Int* **127**, 645-652, doi:10.1016/j.envint.2019.04.002 (2019).

640 35 Liu, B., Zheng, D., Jin, Q., Chen, L. & Yang, J. VFDB 2019: a comparative pathogenomic
641 platform with an interactive web interface. *Nucleic Acids Res* **47**, D687-D692,
642 doi:10.1093/nar/gky1080 (2019).

643 36 Tu, Q., Lin, L., Cheng, L., Deng, Y. & He, Z. NCycDB: a curated integrative database for fast and
644 accurate metagenomic profiling of nitrogen cycling genes. *Bioinformatics* **35**, 1040-1048,
645 doi:10.1093/bioinformatics/bty741 (2019).

646 37 Bastian, M., Heymann, S. & Jacomy, M. in *International AAAI conference on weblogs and*
647 *social media.* (AAAI Press Menlo Park, CA).

648 38 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,

649 357-359, doi:10.1038/nmeth.1923 (2012).

650 39 Satinsky, B. M., Gifford, S. M., Crump, B. C. & Moran, M. A. Use of internal standards for
651 quantitative metatranscriptome and metagenome analysis. *Methods Enzymol* **531**, 237-250,
652 doi:10.1016/B978-0-12-407863-5.00012-5 (2013).

653 40 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny
654 substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004, doi:10.1038/nbt.4229
655 (2018).

656 41 Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927-930
657 (2003).

658 42 Ye, L., Mei, R., Liu, W. T., Ren, H. & Zhang, X. X. Machine learning-aided analyses of thousands
659 of draft genomes reveal specific features of activated sludge processes. *Microbiome* **8**, 16,
660 doi:10.1186/s40168-020-0794-3 (2020).

661 43 Castelle, C. J. & Banfield, J. F. Major new microbial groups expand diversity and alter our
662 understanding of the tree of life. *Cell* **172**, 1181-1197 (2018).

663 44 Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain
664 Bacteria. *Nature* **523**, 208-211 (2015).

665 45 Zhang, T., Shao, M.-F. & Ye, L. 454 Pyrosequencing reveals bacterial diversity of activated
666 sludge from 14 sewage treatment plants. *The ISME journal* **6**, 1137-1147 (2012).

667 46 Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain
668 Bacteria. *Nature* **523**, 208-211, doi:10.1038/nature14486 (2015).

669 47 Proctor, C. R. *et al.* Phylogenetic clustering of small low nucleic acid-content bacteria across
670 diverse freshwater ecosystems. *ISME J* **12**, 1344-1359, doi:10.1038/s41396-018-0070-8
671 (2018).

672 48 Tian, R. *et al.* Small and mighty: adaptation of superphylum Patescibacteria to groundwater
673 environment drives their genome simplicity. *Microbiome* **8**, 51,
674 doi:10.1186/s40168-020-00825-w (2020).

675 49 New, F. N. & Brito, I. L. What Is Metagenomics Teaching Us, and What Is Missed? *Annu Rev*
676 *Microbiol* **74**, 117-135, doi:10.1146/annurev-micro-012520-072314 (2020).

677 50 Ma, L. *et al.* Catalogue of antibiotic resistome and host-tracking in drinking water deciphered
678 by a large scale survey. *Microbiome* **5**, 1-12 (2017).

679 51 Forsberg, K. J. *et al.* Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**,
680 612-616 (2014).

681 52 Barboza, K. *et al.* First isolation report of Arcobacter cryaerophilus from a human diarrhea
682 sample in Costa Rica. *Rev Inst Med Trop Sao Paulo* **59**, e72,
683 doi:10.1590/S1678-9946201759072 (2017).

684 53 Batra, P., Mathur, P. & Misra, M. C. Aeromonas spp.: An Emerging Nosocomial Pathogen. *J Lab*
685 *Physicians* **8**, 1-4, doi:10.4103/0974-2727.176234 (2016).

686 54 Kuypers, M. M. M., Marchant, H. K. & Kartal, B. The microbial nitrogen-cycling network. *Nat*
687 *Rev Microbiol* **16**, 263-276, doi:10.1038/nrmicro.2018.9 (2018).

688 55 Rossetti, S., Tomei, M. C., Nielsen, P. H. & Tandoi, V. "Microthrix parvicella", a filamentous
689 bacterium causing bulking and foaming in activated sludge systems: a review of current
690 knowledge. *FEMS Microbiol Rev* **29**, 49-64, doi:10.1016/j.femsre.2004.09.005 (2005).

691 56 Kinnunen, M., Gülay, A., Albrechtsen, H. J., Dechesne, A. & Smets, B. F. Nitrotoga is selected
692 over Nitrospira in newly assembled biofilm communities from a tap water source community
693 at increased nitrite loading. *Environ. Microbiol.* **19**, 2785-2793 (2017).

694 57 Hallbeck, L. & Pedersen, K. The family gallionellaceae. *The prokaryotes*, 853-858 (2014).

695

Figures

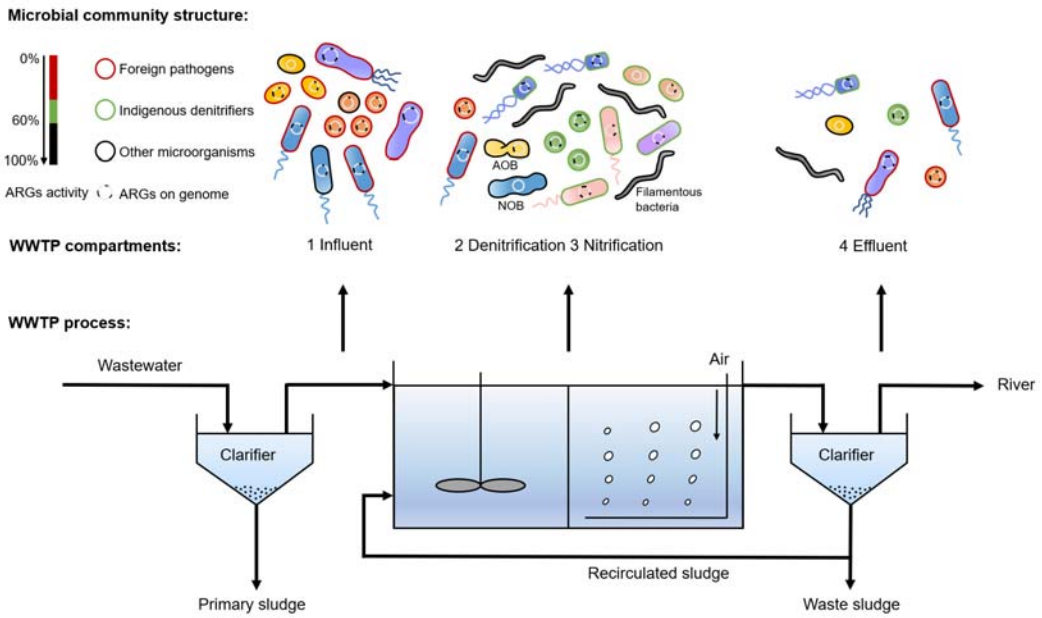

Fig. 1 Sewage-borne pathogens and indigenous denitrifiers as active and key players of multi-antibiotic resistance in the urban wastewater treatment plants (WWTPs). Microbial samples were taken from the influent, denitrification and nitrification bioreactors, and the effluent of 12 urban WWTP systems. Metagenomic sequencing, assembly and binning together with metatranscriptomic analysis enables a genome-level high-resolution and systematic view on the host identity, multi-resistance, pathogenicity, activity of diverse antibiotic resistance genes (ARGs) throughout the WWTPs. Sewage-borne pathogens (marked by red border, defined as MAGs that taxonomically predicted as human pathogens and harbored at least one experimentally verified virulence factor) derived from human intestinal tracts were abundant in the influent. Diverse microorganisms lived in the denitrifying and nitrifying sludge, including the indigenous denitrifiers (marked by green border, defined as MAGs that shared > 95% total expression activities of denitrification genes in the nitrifying and denitrifying sludge while ≤ 5% total expression activities in the influent and effluent). Most members of sewage-borne pathogens and indigenous denitrifiers were identified to host multi-antibiotic resistance genes and were not completely eliminated from the final effluent, thus they represented hitherto-unraveled disseminators of WWTP-released ARGs. Overall, sewage-borne pathogens and indigenous denitrifiers contributed ~60% of all antibiotic resistance activities detected in the recovered genomes and were considered as active and key players of antibiotic resistance in the WWTPs.

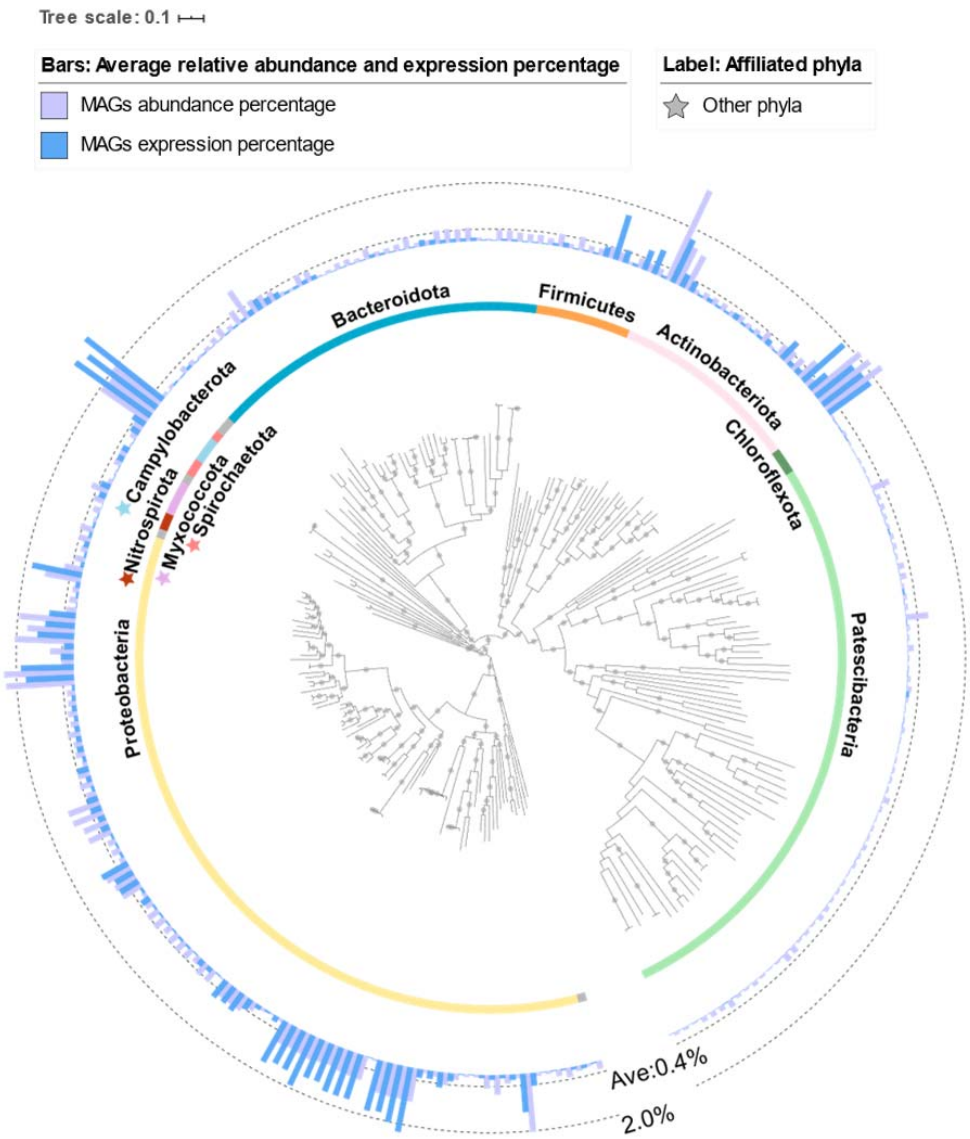

Fig. 2 Phylogenetic tree of 248 high-quality MAGs recovered from 12 urban WWTPs. The tree was produced from 120 bacterial domain-specific marker genes from GTDB using FastTree and subsequently visualized in iTOL. Labels indicate phyla names and, to facilitate an easier differentiation, the color of the front stars beside the phyla label is the same as the color of the corresponding phyla; phyla in which only one MAG were recovered were taken as others. The relative abundance and expression level of each MAG were calculated based on RPKM values across all samples. Abundance percentage and expression percentage were proportions of relative abundance and expression level, respectively, and were shown by external bars (purple: abundance percentage; blue: expression percentage). The dashed circles represent the scale for abundance and expression percentage (inside: average 0.4%, outside: 2.0%). Bootstraps >75% are indicated by the grey dots.

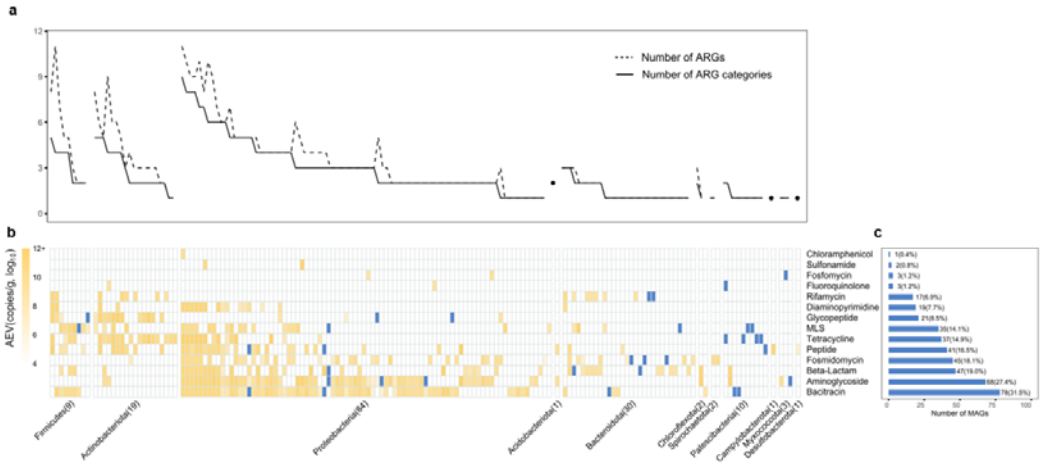

Fig. 3 The distribution and activity of ARGs in the recovered genomes. a. richness of ARGs and ARG classes detected in 162 resistant MAGs. The MAGs were categorized into multi-resistant (113) and single-resistant (49) populations based on the number of antibiotic classes in each MAG. **b.** taxonomic distribution and absolute expression value (AEV, copies/g-VSS) of ARG classes across MAGs. Yellow color intensity represents average AEV of ARGs from each ARG class in the genome. Blue color represents the corresponding MAG harbored but not expressed the corresponding ARG. **c.** number of MAGs assigned to each class of ARGs.

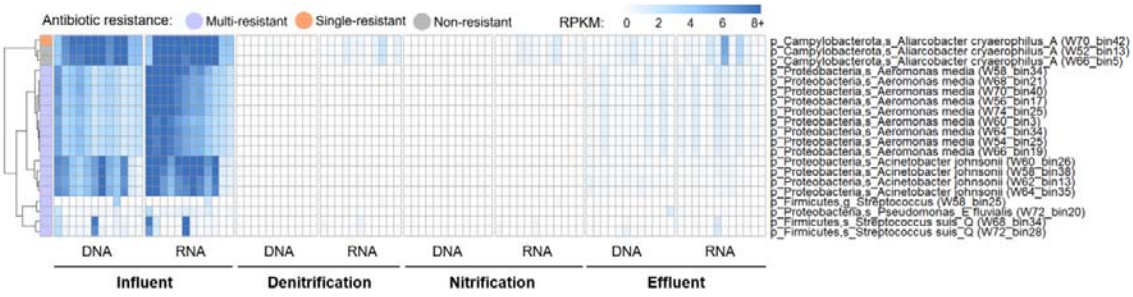

Fig. 4 The cross-compartment distribution and expression pattern of pathogenic populations in the WWTPs. Heatmap for relative abundance and expression level of 20 potentially pathogenic populations MAGs in the influent, dentification, nitrification and effluent compartments. Blue color intensity represents genome relative abundance and expression level normalized by RPKM values. Left annotation column shows antibiotic resistant patterns of potential pathogens.

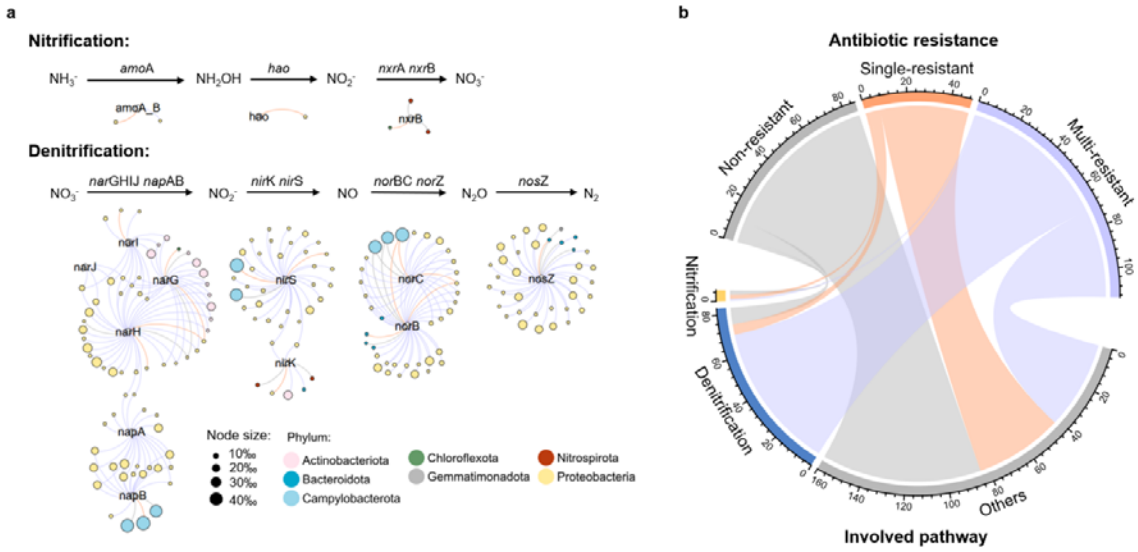

Fig. 5 The distribution of MAGs annotated with NDGs and their relationship with antibiotic resistance. a. Network reveals distribution of NDGs in nitrifiers and denitrifiers. Each node represented a NDG or MAG (colored by taxonomy and size scaled by expression percentage), and each edge connected a MAG to a NDG which represented the MAG expressed the NDG in at least one sample. Color of edge represents antibiotic resistant pattern of the linked MAG (purple: multi-resistant, orange: single-resistant, grey: non-resistant) **b.** Relationship between antibiotic resistance and nitrogen-removing metabolism in the related MAGs. The width of the string represents the number of MAGs.

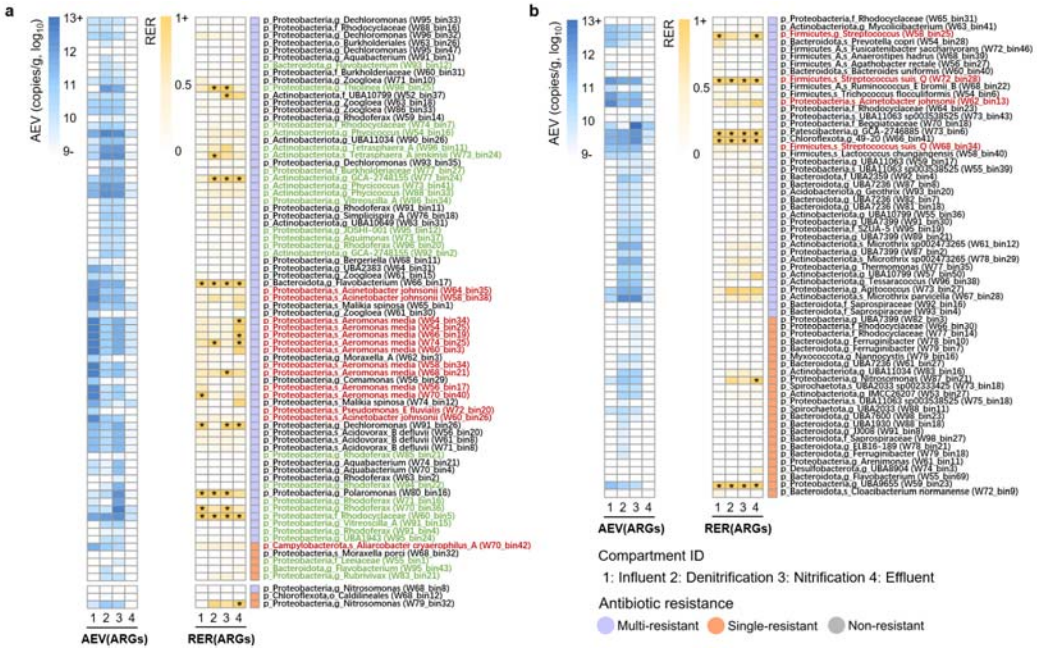

Fig. 6 The absolute expression value (AEV) and relative expression ratio (RER) for ARGs in the WWTP bacterial populations. a. the AEV and RER of ARGs in MAGs putatively involved in nitrogen removal. **b.** the AEV and RER of ARGs in other MAGs. The case of RER>1 is marked with an asterisk. Right annotation column illustrated antibiotic resistant pattern of MAGs. Column names of heatmap represent compartment ID in the WWTPs. MAGs marked in red were sewage-borne pathogenic group and MAGs marked in green were indigenous denitrifying group.