Environmental Science
Water Research & Technology

The Occurrence and Diversity of Antibiotic Resistance and Virulence Factor Genes in Wastewater from Four North American Treatment Plants

Journal:	*Environmental Science: Water Research & Technology*
Manuscript ID	EW-ART-11-2021-000820.R1
Article Type:	Paper

Water Impact Statement

The manuscript presents a comparative analysis of microbial composition and functional traits of wastewater microbiomes at different wastewater treatment plants and at treatment processes. The outcomes of the study contribute to an improved understanding of the role of wastewater treatment plants in removal and amplification of antibiotic resistance genes and pathogens.

The Occurrence and Diversity of Antibiotic Resistance and Virulence Factor Genes in Wastewater from Four North American Treatment Plants

Loan T. Le [a], Zhoujin Huang [b], Katrine Whiteson [c], Sunny Jiang [d*]

[a] School of Pharmacy and Pharmaceutical Sciences, [b] Department of Computer Sciences, [c] Department of Molecular Biology and Biochemistry, [d] Department of Civil and Environmental Engineering, University of California, Irvine, CA 92697

*Corresponding author: 844 Engineering Tower, University of California, Irvine, CA 92697. Email: sjiang@uci.edu; Tel: 949-824-5527

Abstract

The spread of antibiotic resistance in aquatic environments is an emerging environmental issue due to potential threats to public health. Wastewater treatment plants (WWTPs) could serve as a sink for removal of antibiotic-resistant bacteria (ARB), antibiotic resistance genes (ARGs), and virulence factor genes (VFGs) from wastewater before discharge to the environment, or an amplifier because the stress conditions prevailing in sewage treatment processes may favor the proliferation of ARB, ARGs, and VFGs. In this study, we performed metagenomic sequencing and analyses to examine the diversity of microbial communities and the fate of ARGs, antibiotic biosynthesis genes (ABSGs) and VFGs in sixteen (16) sewage samples collected from four large WWTPs located at two major metropolitan centers on east and west coast of U.S. Multivariate analysis revealed that the diversity and composition of microbial communities and ARGs in sewage samples were primarily associated with the wastewater treatment processes. The overall relative abundances of ARGs and VFGs declined by over 20% after treatments. However, the activated sludge process (ASP) selectively enriched ARGs encoding multidrug resistance and certain VFGs as indicated by the increase of their relative abundance in mixed liquor suspended solids. The relative abundances of sub-groups of ABSGs were also substantially amplified during ASP. These findings provide insights into the impact of conventional wastewater treatment processes on the prevalence of ARGs and VFGs to better understand the dissemination of ARGs and VFGs through human sewage. Comparison of different treatment processes in relation to ARGs removal also helps to identify strategies to reduce the spread of antibiotic resistance through sewage discharge.

Keywords: metagenome; sewage microbiome; bacterial virulence; antibiotic production; wastewater; activated sludge

1. Introduction:

Antibiotic resistance has become a serious concern to global public health because of the progressively severe ramifications for the treatment of infectious diseases. Each year in the United States, approximately 2.8 million people acquire an antibiotic-resistant infection, and more than 35,000 of these cases are lethal (CDC, 2019). The antimicrobial resistance crisis has been ascribed to the intensive use of antibiotics in medical treatment, veterinary and agriculture applications (Ventola, 2015), which intervened in the selection and spread of antibiotic-resistant bacteria (ARB) (Hughes, 2014). The exchange of antibiotic resistance genes (ARGs) between environmental microbes and clinical pathogens via the mechanism of horizontal gene transfer (HGT) further facilitates the spread of antibiotic resistance into natural environments (Lerminiaux et al., 2019). The occurrence of ARB and ARGs have increasingly been found in many environmental media, including natural rivers, soils, surface water, and even drinking water (Knapp et al., 2010; Middleton et al., 2013; Shi et al., 2013; Su et al., 2020), which increase human exposure via food chain, agricultural irrigation and recreational activities (Amarasiri et al., 2020).

The acquisition of ARGs can contribute to the enhanced virulence capacity of pathogenic bacteria to overcome antimicrobial therapies and adapt to extreme environmental conditions. Recent studies have found an increase in the expression of virulence factor genes (VFGs) in strains carrying resistant genes (Yazdanpour et al., 2020; Abd El-Baky et al., 2020). These VFGs encode virulence factors that are responsible for bacteria's ability to infect the host cells and cause disease. Virulence factors enable a microorganism to establish itself on or within a host and enhance its potential to cause disease through adherence, colonization, immune invasion, secretion systems, and toxin production.

Municipal wastewater treatment plants (WWTPs) represent an important reservoir of ARGs and VFGs. High concentrations of microbial contaminants from the households, businesses, medical facilities, and industrial wastewater discharges reach WWTPs, where bacteria are exposed to significant concentrations of antibiotic residues, disinfectants, and heavy metals (Dai et al., 2019). Diverse microbial communities in wastewater, together with the selective pressure exerted by selective agents, may promote the selection of antibiotic resistance and the HGT of resistance genes between pathogenic and environmental microorganisms. Previous studies have shown that WWTP effluents were among the main anthropogenic sources for the entry of ARGs into natural environments (Chen et al., 2013a; Wei et al., 2018a; Liu et al., 2019a). At the same time, WWTPs are designed to remove organic contaminants including antibiotic residuals and pathogenic organisms. Essentially, WWTPs are required to demonstrate the performance of controlling the concentrations of carbonaceous biochemical oxygen-demanding (CBOD) matter and the removal of nitrogen and phosphorus to meet the National Pollution Discharge Elimination System (NPDES) water quality standard before discharging to the environment (EPA, n.d.). However, there are currently no regulatory guidelines or legislations to monitor and control the levels of ARB and ARGs in sewage effluent discharge (Pazada et al., 2019). Whether WWTPs serve as amplifiers to encourage the spread of ARGs into the environment or control points to remove antibiotics and microorganisms harboring ARGs from wastewater is still an important question that requires clarification.

Extensive efforts have been made to investigate the occurrence and distribution of ARGs as well as their removal by WWTPs using metagenomic analyses (Ref). Owing to the rapidly advancing high-throughput sequencing of microorganisms' genomes and the increasing power of bioinformatic tools, metagenomics has been considered as a powerful approach to characterize the microbial community and function genes due to its ability to collect and quantify all the genomic DNA of microorganisms present in a single sample. It diminishes a quantifiable bias as it is not limited to culturable microorganisms (Thomas et al., 2012). Targeted and sequence-based metagenomics has been employed in many studies to extensively assess the phylogenetic and functional classifications of ARGs in wastewater. Recently, Li and colleagues (2015) observed a notable decrease in ARGs profile in activated and digested sludge metagenomes in comparison with the influent metagenome. Shotgun metagenomic analysis of ARGs profile in biofilm and suspended growth-based WWTPs also showed that these processes efficiently reduced ARG loads in the effluent (Petrovich et al., 2018a). Contrary to these reports, other studies have reported no significant differences in antimicrobial resistant profiles despite a substantial change in microbial communities following the treatment processes (Bengtsson-Palme et al., 2016; Rafraf et al., 2016). So far, the current knowledge on the dynamics of ARGs in wastewater treatment processes is still limited. Meanwhile, the environmental behaviors of ARGs may also vary in different WWTPs because of the distinct treatment processes. Therefore, it is necessary to understand the diversity and distribution of ARGs in WWTPs for regional management regimes to control the proliferation and dissemination of ARGs into the environment.

This study investigated the occurrence and diversity of ARGs in four municipal WWTPs using shallow shotgun metagenomic sequencing and metagenomic analyses to test the hypothesis of ARGs removal by WWTPs. Furthermore, we investigated the occurrence and prevalence of antibiotic biosynthesis genes (ABSGs) and VFGs to assess our alternative hypothesis that the prevalence of certain ARGs increase due to the evolution selection by virulent strains and the *in situ* production of antibiotics. The co-selection of ARGs and VFGs were also explored based on the co-occurrence patterns between ARG subtypes, and between ARGs and VFGs using network analysis. The results of this study provide detailed insight into the fate of ARGs, ABSGs and VFGs in WWTPs and potential human health risks imposed. Analyzing the fate of ARGs throughout treatment processes will help integrate strategies to mitigate the spread of ARGs into the environment through sewage discharge.

2. Materials and Methods

2.1. Sample collection

Sixteen sewage samples were collected from four WWTPs in the U.S. that provide service to large metropolitan regions. Each sample was taken from a different stage along the treatment trains. Plants A1 and A2 in Orange County, California treat combined approximately 185 million gallons of sewage per day (MGD) and serve a population of 2.6 million. Both plants include treatment trains of activated sludge process (ASP) and trickling filters (TF) as the secondary treatment processes. High purity oxygen is used in Plant A2 ASP. Samples were collected from Plants A1 and A2 in May 2018. The second set of samples was collected from treatment Plant B in Los Angele County, California, two weeks after collecting the first set of samples. Plant B has a design treatment capacity of 100 MGD and provides primary, secondary, and tertiary

wastewater treatment for approximately one million people northeast of Los Angeles. ASP is used as the secondary treatment process in Plant B. The third WWTP, Plant C, in the Washington DC area has a daily wastewater treatment capacity of approximately 350 MGD. The plant provides primary, secondary, and additional biological nutrient removal (BNR) as a tertiary treatment before discharging effluent to the surface water. Details of sampling locations in each treatment facility are provided in Fig S1 and Table S1.

All samples were transported on ice and processed within 24 hours of collection. In the lab, the sewage samples were pelleted by centrifugation. The cell pellets were kept frozen at -80 ºC until used for metagenomic analysis.

2.2. DNA extraction and shallow shotgun metagenomic sequencing (SSMS)

DNA extraction and shallow shotgun metagenomic sequencing (SSMS) was performed by CoreBiome (now called Diversigen; https://www.diversigen.com/). BoosterShot, a shallow shotgun sequencing approach aiming for greater than 2 million reads per sample, was used to gain a boost in data return with full species-level resolution and functional profiling not provided by amplicon sequencing. Compared to amplicon sequencing, BoosterShot provides a less biased representation of the microbial community and higher taxonomic resolution due to the inclusion of all genes in the metagenome, including antibiotic resistance genes, virulence genes, and other genes that are not conserved and therefore not included in an amplicon sequencing approach.

2.3. SSMS Processing

Initial analyses were performed by CoreBiome. Over 50 million 50-145 bp single-end reads (50,356,919) were obtained from 16 samples with an average sequencing depth of 3 million reads per sample. All raw sequences were trimmed and processed for quality control using SHI7 (Al-Ghalith et al., 2018). Sequencing data were submitted to the NCBI BioProject database under accession number PRJNA745491.

2.4. Taxonomic classification and diversity analysis

Taxonomic assignment was also carried out by CoreBiome (Diversigen). In brief, the quality-controlled sequences were aligned to a curated database containing all representative genomes in NCBI RefSeq for bacteria with additional manually curated strains. Alignments were made at 97% identity against all reference genomes. Input sequences were compared to reference sequences in the CoreBiome Venti database using fully gapped alignment with BURST. Ties were broken by minimizing the overall number of unique Operational Taxonomic Units (OTUs). For taxonomy assignment, each input sequence was assigned the lowest common ancestor that was consistent across at least 80% of all reference sequences tied for best hit. The number of counts for each OTU was normalized to the OTU's genome length. OTUs accounting for <1 millionth of all strain-level markers and those with less than 0.01% of their unique genome regions covered (and < 0.1% of the whole genome) at the species level were filtered out. Samples with fewer than 10,000 sequences were also discarded. The normalized and filtered table was used for all downstream analyses. Filtered taxonomic tables at phylum and genus levels were used for the comparison of bacterial composition in different samples.

Alpha and beta diversity analyses for taxonomy were also performed by CoreBiome bioinformatic service. Data analyses were first aggregated by WWTPs into groups: A1 - Orange County plant 1, A2- Orange County plant 2, the two plants in Orange County also combined as Plant A in data analysis, B - Los Angeles County plant, C - Washington D.C. plant; and by

treatment processes into types: PE – primary effluent, MLSS – mixed liquor suspended solids, Biofilm – biofilm retrieved from trickling filter, and SE – secondary effluent. The Chao1 index, Shannon's index, and observed OTU counts (taxonomic group) were calculated from a rarefied, filtered OTU table set to the minimum depth of 10,000 sequences using QIIME 1.9.1. Bray-Curtis beta diversity metrics were calculated from the filtered taxonomy table. The graphical presentation and downstream statistical tests were then performed using R Studio.

2.5. Identify KOs for ARGs, ABSGs, and VFGs

Gene function profiles were provided in CoreBiome Core Analysis Report. Kyoto Encyclopedia of Genes and Genomes Orthology groups (KEGG KOs) were identified directly using alignment at 97% identity against a gene database derived from the CoreBiome Venti strain database used above. Any samples that were removed during the filtering of the taxonomy tables are then also removed from the functional tables. The number of reads that mapped to each gene for each sample was then tabulated to create the absolute functional tables. The output from the KO (KEGG Orthology) table was exported into R Studio for downstream analyses.

KO groups identified in samples were compared with the KO database, the KEGG Antibiotic Resistance Signatures (https://www.genome.jp/kegg/annotation/br01600.html), the NCBI Bacterial Antimicrobial Resistance Reference Gene Database (BARRGD) (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/), the Virulence Factors Database (VFDB) (http://www.mgc.ac.cn/VFs/) and previously published literature to identify genes encoding the multidrug resistance, antibiotic resistance, and virulence factors of bacterial and viral pathogens (Chen et al., 2005; Liu et al., 2013; Katiyar et al., 2020). The antimicrobial classes included in the search criteria are presented in supplemental Table S2. Since bacteria that possess antibiotic biosynthesis pathways are also resistant to the antibiotic they produce, ABSGs (shown in Table S2) were also searched in the KO database. All virulence KOs identified in samples were categorized into virulence functions according to the VFDB. The raw counts were then normalized to the scaling factors calculated by the DESeq2 package in the R studio version 1.4.1106as described by the work of Anders et al. (2010), to account for library size variations bias. In brief, the geometric mean of raw counts for each gene was determined across all samples. The size factor (normalization factor) for each sample was calculated as the median of all ratios between gene counts and the geometric mean in a given sample using the following equation:

$$\hat{s}_j = median_i \frac{k_{ij}}{(\Pi_{v=1}^{m} k_{iv})^{1/m}}$$

where $\hat{s}_j$ is the size factor for sample j; k_{ij} is the number of reads mapping to gene i in sample j; the denominator is interpreted as a pseudo-reference sample obtained by taking the geometric mean across m samples.

The normalized count values were then computed by dividing the number of raw counts for each gene in a given sample by that sample's size factor. The normalized data were exported to R Studio (R Studio Inc., Boston, MA) for graphic representation and further statistical analyses.

2.6. Network analysis of co-occurrence patterns among ARG subtypes and VFGs

The microbial association network was constructed using the normalized ARG-KO abundance table with CoNet in Cytoscape (Faust et al., 2016) to investigate correlations between ARG-

subtypes. Pairwise Spearman's Rank Correlation Coefficient was calculated between 99 ARG subtypes that occurred in at least 6 samples (out of 16 in total). To reduce false positives in the result, the data was bootstrapped by randomizing the abundance table for 100 iterations and calculating the *p*-value using the Benjamini-Hochberg method. The robust correlations were retrieved with filtering criteria that the Spearman's Rank coefficient was > 0.8 and adjusted *p*-value was < 0.05. There were 81 statistically robust correlations and 35 unique ARG subtypes identified. The betweenness centrality was calculated in CoNet and identified 3 ARG hubs in the network. The modularity was calculated using Gephi to identify 8 distinct clusters, which densely correlated to each other. The network visualization was conducted on interactive platform Gephi version 0.92 (Bastian et al., 2009).

Additionally, network analysis of co-occurrence patterns between ARG and VFG was performed to achieve a deeper understanding of their interaction. The correlations between ARGs and VFGs were also generated using the normalized ARG-KO and VFG-KO abundance tables. Using the same method and criteria described above (pairwise Spearman's Rank Correlation Coefficient, adjusted p-value, bootstrap), the top 255 connections and 177 nodes were first identified through CoNet. After filtering out edges with 0 weight, nodes with less or equal to 2 degrees, and edges with absolute pairwise Spearman's Rank Correlation Coefficient less than 0.8 using Gephi, 87 qualified nodes and 123 qualified edges were retained in the result.

2.7. Statistical analyses

All statistical analyses and graphs, unless otherwise stated, were conducted by the R studio version 1.4.1106 (RStudio Team, 2021). Non-parametric Kruskal-Wallis test and the pairwise Wilcoxon rank-sum test were used to assess the differences in microbial diversity, ARGs, VFGs profiles among sample groups by the R statistical package "ggpubr". PAM (Partition Around Medoids) clustering algorithm was adopted to search for clustering patterns reflecting on the beta diversity Principal Coordinate Analysis (PCoA) plots. The R package "cluster" was used to perform two batches of computations with different expected numbers of clusters based on the Bray-Curtis beta-diversity metrics of both taxonomy profile and ARG profile. Differences in the composition and structure of microbial communities and ARGs profile between clusters were then evaluated using the Permutational Multivariate Analysis of Variance test (PERMANOVA) implemented in the "vegan" package in R. The partition with the highest R^2 value (coefficient of determination) was selected to represent the clustering structure. All statistical tests were considered significant at a *p*-value < 0.05.

3. Results

3.1. Microbial richness, diversities and composition across WWTPs and treatment processes

Microbial community richness, diversity and composition reflect their functions in WWTPs. Alpha diversity analyses using chao1(abundance-based richness) and Shannon (diversity) indices (Fig. S2 and Table S3) revealed that samples from Plant A2 had a greater community richness (chao1 index, p = 0.032, Wilcoxon rank-sum test) and higher community diversity (Shannon index, p = 0.016, Wilcoxon rank-sum test) compared to those samples from Plant C. Across the treatment processes, the microbial community's alpha diversity decreased from primary effluent to secondary effluent in general (Fig. S2B and S2D). The observed microbial richness and diversity indices were further confirmed by the rarefaction curve (Fig. S3). It should be noted that the secondary effluent sample from Plant B (B-2) was below the sampling depth of 10,000

sequences and therefore was excluded from the analyses. Taken together, these results suggest that the composition of microbial communities in WWTP samples was influenced by the sources of raw wastewater and treatment processes.

Although alpha diversity among samples did not differ significantly, the beta diversity across samples was distinctly different. PCoA and PAM analysis of microbial beta-diversity revealed four distinct clusters with over 60% of the variation explained by two factors – sample collection site and treatment type (Fig. S4 and Fig. S5, PERMANOVA $p=0.077$, $R^2=0.149$). The first cluster identified the dominant microbial community in Southern California raw wastewater. The second cluster represented microbial communities that perform the major function of CBOD reduction and nitrification. The third cluster included microbial communities in MLSS samples that operated under different treatment conditions. These results revealed that microbial communities in primary effluents were separated from those found in the downstream biological treatment processes. The distance between microbial communities performing wastewater treatments (clusters 2 and 3) was lesser than their distance to the corresponding primary effluent, suggesting treatment processes shaped microbial communities towards the more similar ones (therefore the similar function of biodegradation). The last group reflected the microbial community dominated in the Washington D.C. region (see supplemental Fig. S4 and S5, and legends for additional details).

The relative abundance of the microbial communities in all samples grouped by WWTPs is shown in Fig. 1 at the phylum level. The results revealed similar taxa distribution patterns with a highly uneven community predominated by four bacterial phyla in all samples. These phyla included Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. Of them, Proteobacteria was the most abundant phylum that accounted for more than 50% of the microbial composition in the samples of all plants. Meanwhile, the phylum Bacteroidetes and Firmicutes were prevalent in the primary effluents of plants A and B. However, these bacteria phyla were identified as the least dominant phyla in the primary effluent collected from Plant C, while the phylum Deinococcus-Thermus was more abundant in ASP MLSS (C-3) of Plant C. Similarly, the phylum Nitrospirae, an important group of microbes in nitrification process, exhibited higher abundances in samples retrieved from Plant C, reflecting the plant's effort on nitrogen removal. The phylum Spirochaetes, on the other hand, were more abundant in the TF biofilm of Plant A2 and ASP MLSS of Plant B than in other samples. A significant reduction in the types of microbial communities was observed following the treatment processes in three Southern California plants. Proteobacteria became increasingly dominant in later treatment processes. However, a reverse trend was observed in Plant C, with Proteobacteria dominating the primary effluent but reduced in relative abundance at later stages of treatment processes. Although the dominant phyla identified in WWTPs here have also been reported in previous studies (Wei et al., 2018; Do et al., 2019), these variations in microbial communities imply the influence of both the source of raw wastewater and treatment processes.

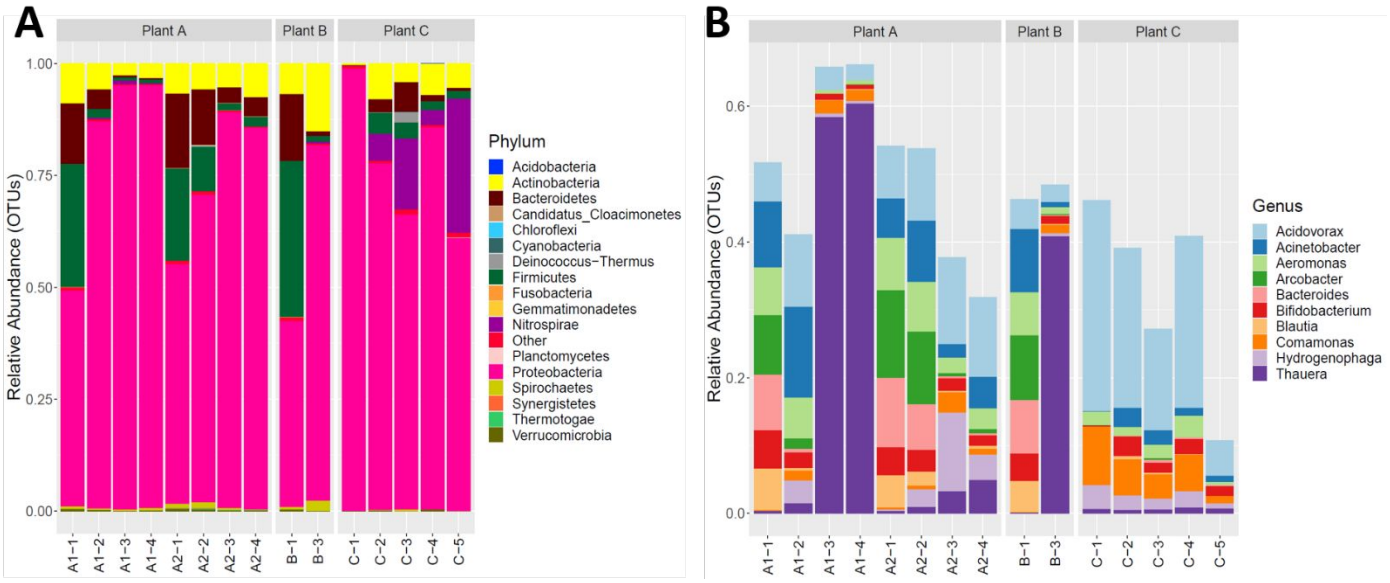

Fig. 1. Microbial community composition for all WWTP samples at the phylum level **(A)** and the genus level **(B).** The relative abundances of each phylum-level taxa and the relative abundances of the top 10 genera that are representative of OTUs. A1-1 to A1-4: Samples collected from different stages of treatment at Orange County (California) Plant 1; A2-1 to A2-4: Samples collected from different stages of treatment at Orange County (California) Plant 2; B-1 to B-3: Samples collected from different stages of treatment at Los Angeles County plant, C-1 to C-5: Samples collected from different stages of treatment at Washington D.C. plant; Detailed descriptions of each sample are listed in Table S1 and in the manuscript text.

In accordance with the dominant phylum Proteobacteria, the corresponding genera *Acidovorax* and *Acinetobacter* made up the majority of the microbial composition in all WWTP samples (Fig. 1B). The other top 10 genera included *Aeromonas*, *Arcobacter*, *Bacteroides*, *Bifidobacterium*, *Blautia*, *Comamonas*, *Hydrogenophaga*, and *Thauera*. *Thauera* was the most abundant genus in ASP MLSS samples collected from conventional ASP trains at two Southern California plants. It should be noted that plants A and B, both located in Southern California, shared a similar pattern of microbial compositions, while microbial compositions were more different in samples from Plant C. This result further confirms the observation revealed by beta diversity analysis (Fig. S4).

3.2. Diversities of ARGs are associated with WWTP processes

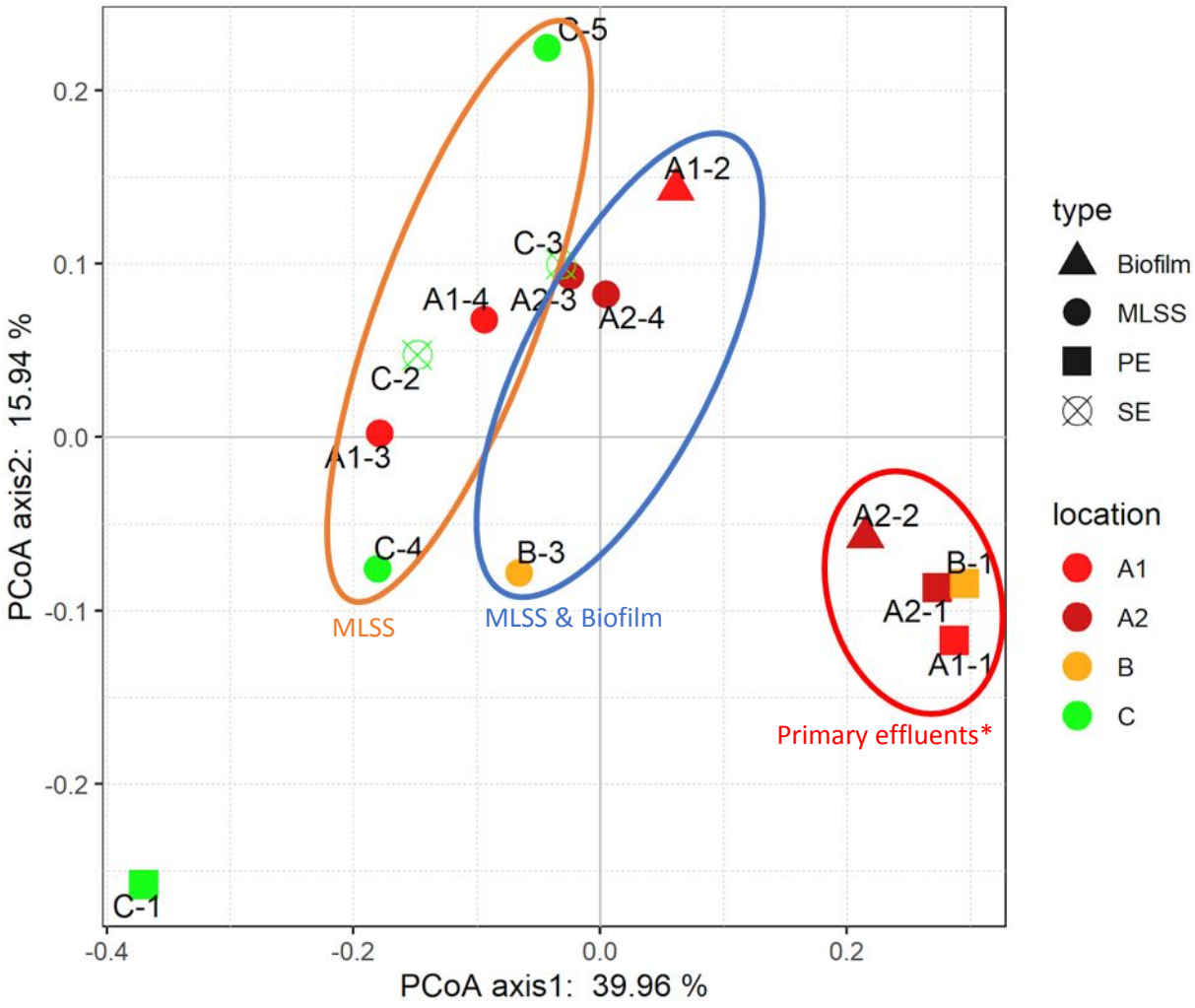

Fig. 2. Beta diversity represented by PCoA plot of the ARGs identified in WWTP samples. Two primary axes represent 55.90% of total variation (PC1=39.96%, PC2=15.94%). Each point on the graph represents the ARG profile of each sample. The shape indicates the type of the treatment process, and color indicates the location. The clusters were identified by PAM clustering algorithm and confirmed by PERMANOVA. The labels under each cluster identified the type of samples included in each cluster. * denotes that A2-2, the biofilm collected from a short contact TF, is also included as primary effluent (see text for details). Sample labels are identical as in Fig. 1.

PCoA analysis of ARGs revealed that variations in ARG abundance profiles in WWTP samples were primarily differentiated by wastewater treatment processes (Fig. 2). Indeed, three distinct clusters were identified and confirmed by PAM clustering algorithm and PERMANOVA analysis (Fig. 2 and S6, $p= 0.001$, $R^2=0.368$). These differences could be seen in significant shifts in ARG profiles between primary effluent and ARGs downstream of wastewater treatment trains (MLSS and biofilm).

All primary effluents from all three Southern California plants (A1-1, A2-1, and B-1) and the biofilm sample (A2-2) collected from the short TF at plant A2 clustered into a group (Fig 2, red

circle). The short TF (solid contact) was used as a pretreatment for ASP and had little time to change the microbial structure (Fig. S4) from the primary effluent and was grouped with primary effluent. This clustering pattern suggests that ARG compositions in Southern California raw wastewater were similar. The second cluster (Fig 2, blue circle) includes three Southern California plants MLSS samples (A2-3, A2-4, B-3) and one TF biofilm (A1-2). This group represents the ARGs among microbial communities that perform CBOD removal in Southern California plants. The third cluster (Fig 2, orange circle) includes all samples from plant C except the primary effluent (C-1), and two MLSS samples (A1-3, A1-4) from plant A1. This result suggests that this third cluster is forced by the combined factors of the treatment process and the source of wastewater. The cluster, although statistically significant, is less defined. The primary effluent from plant C (C-1) is presented as an outlier in PERMANOVA, implying the source of wastewater is the main driver for the ARG profile in this sample. Alternatively, the primary effluent (C-1) from Plant C can be clustered in the third group (Fig. S7) by PAM clustering algorithm and PERMANOVA analysis of the other three clusters (Fig. S7, $p= 0.001$, $R^2=0.359$), indicating the low stability of the cluster.

3.3. The dominance of multidrug-resistant genes in different WWTPs

Comparison of the ARGs in different WWTPs revealed similar ARG profiles across the samples (Fig. 3A). Among 111 unique KOs-associated with antibiotic resistance (identified from the 4,198 predicted KOs), nine antimicrobial classes were found, including aminoglycosides, beta-Lactams, CAMPs, FCA, MLSB, multidrug, tetracycline, trimethoprim, and vancomycin (Table S2). Of these antimicrobial classes, ARGs encoding resistance toward beta-Lactams and multi-drugs are dominant in all samples regardless of WWTPs and treatment processes. These core ARGs made up more than 60% of the total ARG abundance in the sewage samples (Fig. 3A).

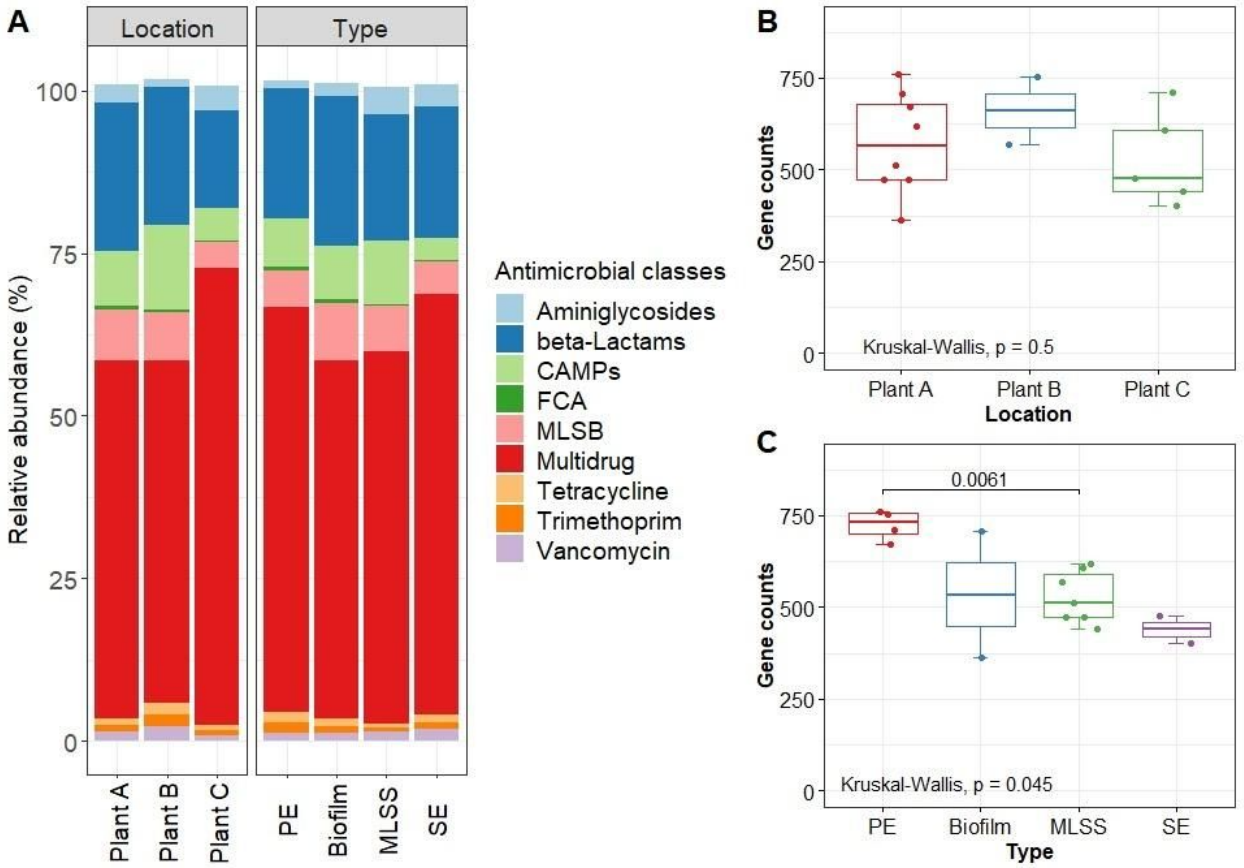

Fig. 3. The abundance of all antimicrobial classes detected in WWTP samples. Relative abundance is represented as the percentage of each antimicrobial class per total ARGs in a sample either grouped by sampling WWTP or treatment type **(A)**; and comparison of normalized total ARG counts by sampling WWTP **(B)** and by treatment type **(C).** P-values by Kruskal-Wallis with posthoc Wilcoxon rank-sum test are indicated in each boxplot. Statistical significance was considered at a $p< 0.05$. Plant A: two plants at Orange County (California); Plant B: one plant at Los Angeles County (California); Plant C: one plant at Washington D. C.; PE – primary effluent, MLSS – mixed liquor suspended solids, Biofilm – biofilm retrieved from trickling filter, and SE – secondary effluent.

There was no significant difference in normalized ARG counts comparing samples grouped by WWTP (Fig 3B, p=0.5). However, following the wastewater treatment processes, total ARG counts significantly decreased (Kruskal-Wallis, p = 0.045, Fig. 3C). Indeed, the median of total ARG counts in the ASP MLSS samples was significantly decreased by 30.2% compared to the primary effluent samples (Wilcoxon rank-sum, p = 0.0061). A similar trend was observed for the direct comparison of ARG encoding beta-Lactam resistance and multi-drug resistance, two dominant antimicrobial classes (Fig. S8 and Table S4).

We further explored the top 10 prevalent ARGs within the two dominant antimicrobial classes. Although no significant trends in the total relative abundance of each ARG were observed across four WWTPs, the compositions of ARGs in individual samples varied greatly (Fig. 4).

Particularly, gene encoding outer membrane porin of OprM efflux systems for multidrug resistance (*oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC*) was dominant in all samples, indicating that the *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC* gene is a persistent resistance gene regardless sampling WWTP. Similarly, gene encoding MexAB-OprM efflux pumps (*acrB/mexB/adeJ/smeE/mtrD/cmeB*) was also consistently found at high abundance across samples with a notable increase in relative abundance found in the BNR MLSS of Plant C (C-5). Gene *mepA* encoding MepA efflux pump was detected in high abundance in primary effluent samples of Southern California plants, while it was less dominant in sewage samples of plant C.

When comparing the variation of ARG profiles along with the treatment processes, a general decrease in relative abundance was observed as the wastewater was treated. Nevertheless, a few ARGs increased in abundance in the ASP MLSS. Notably, genes conferring multidrug resistance, MexAB-OprM efflux pumps, aminoglycosides (*aph3-II*), and CAMPs (*arnT/pmrK*) were more abundant in MLSS samples from Plant C (Fig. 4 and S9). These findings suggest that ASP may facilitate the HGT of ARGs between environmental microorganisms, and hence, be responsible for disseminating ARGs into environments.

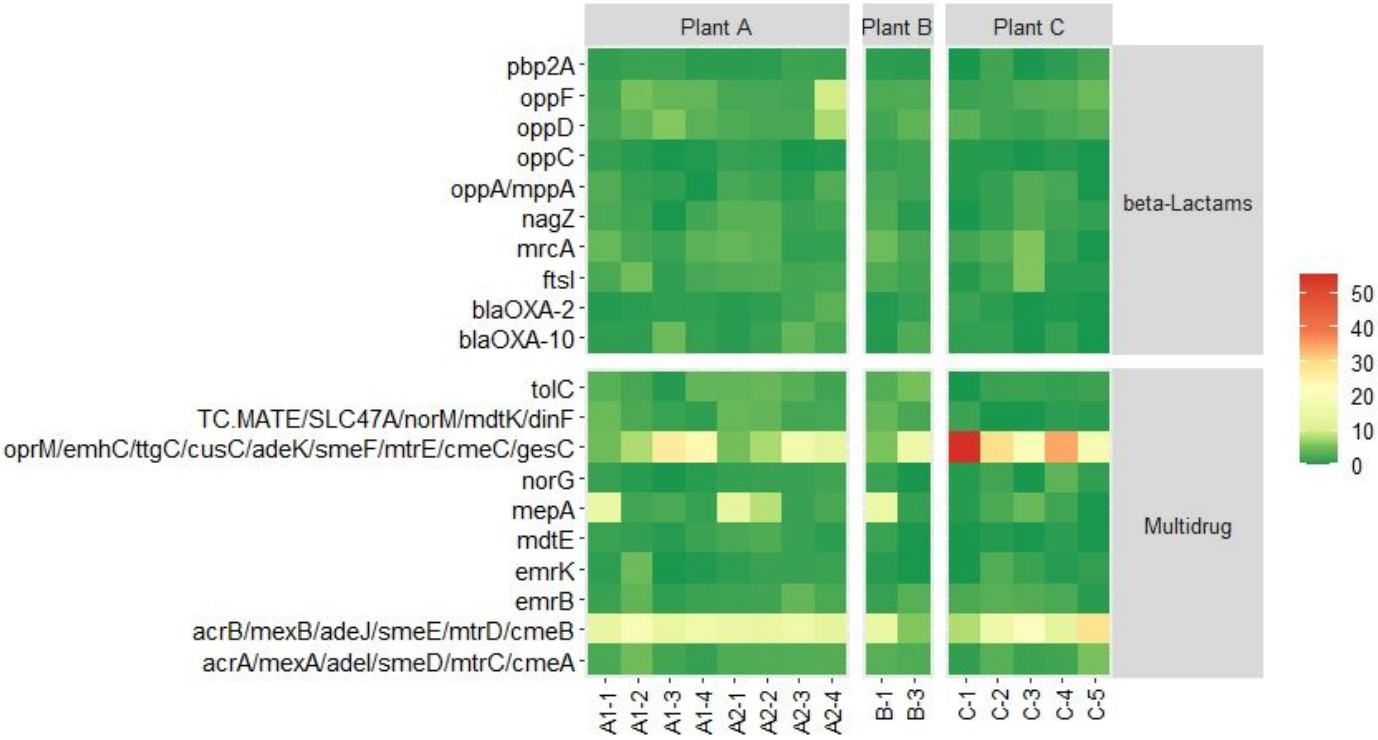

Fig. 4. Heatmap showing the relative abundances of beta-Lactams and multidrug ARGs in WWTP samples. Colors reflect relative abundances (%) of ARGs from lowest (green) to highest abundance (red). Sample labels are identical as in Fig. 1

3.4. The amplification of ABSGs following the WWTP treatment processes

Since most antibiotic compounds are produced by bacteria as their natural defense system against the other microorganisms in their vicinity, antibiotic-producing bacteria also encode genes conferring resistance to the antibiotics that they produce (Cordero et al., 2012; Peterson et al., 2018). 51 KOs related to antibiotic biosynthesis were classified into five classes: aminocoumarin, aminoglycosides, beta-Lactams, tetracyclines, and vancomycin (Table S2). Of them, genes encoding the biosynthesis of aminocoumarin and aminoglycoside classes of antibiotics were most abundant, which account for more than 70% of the total ABSGs identified in samples (Fig. S10). Similar to the ARG counts, the abundances of ABSGs did not significantly differ between WWTPs (Kruskal-Wallis, $p = 0.76$, Fig. S10). Following the wastewater treatment processes, total ABSG gene counts, however, significantly varied ($p=0.017$, Kruskal-Wallis test). Namely, the median of ABSG counts in the biofilm and MLSS samples increased by 94.3% and 164.7%, respectively, in comparison with the primary effluents ($p= 0.0061$, Wilcoxon test, Fig. S10C, Table S5).

A closer examination of individual ABSG indicated streptomycin-associated ABSGs was the most abundant subtype of ABSG-producing aminoglycosides, with higher abundances found in primary effluents (Fig. 5). Genes regulating the biosynthesis pathway of neomycin, kanamycin and gentamicin were identified as the second abundant subtype of ABSG-producing aminoglycosides in all samples. The abundance of *tobZ* declined upon treatment in plants A and B, while the gene was completely absent from samples of Plant C. Meanwhile, *kanE/ kanM2* increased in abundance from primary effluents to secondary effluents in all WWTPs. Moreover, most genes conferring the biosynthesis pathway of novobiocin (*novN*, *novO/couO*, *novR/cloR*) and streptomycin (*rfbA/rmlA/rflH*, *rfbD/rmlD*, *rfbB/rmlB/rffG*) were amplified in the ASP MLSS and present in the secondary effluents (Fig. 5). These increases in the abundance of ABSGs-producing novobiocin and streptomycin indicate that the conditions in activated sludge may favor novobiocin and streptomycin-producing microorganisms, and therefore, WWTP may be a reservoir for the proliferation of ABSGs.

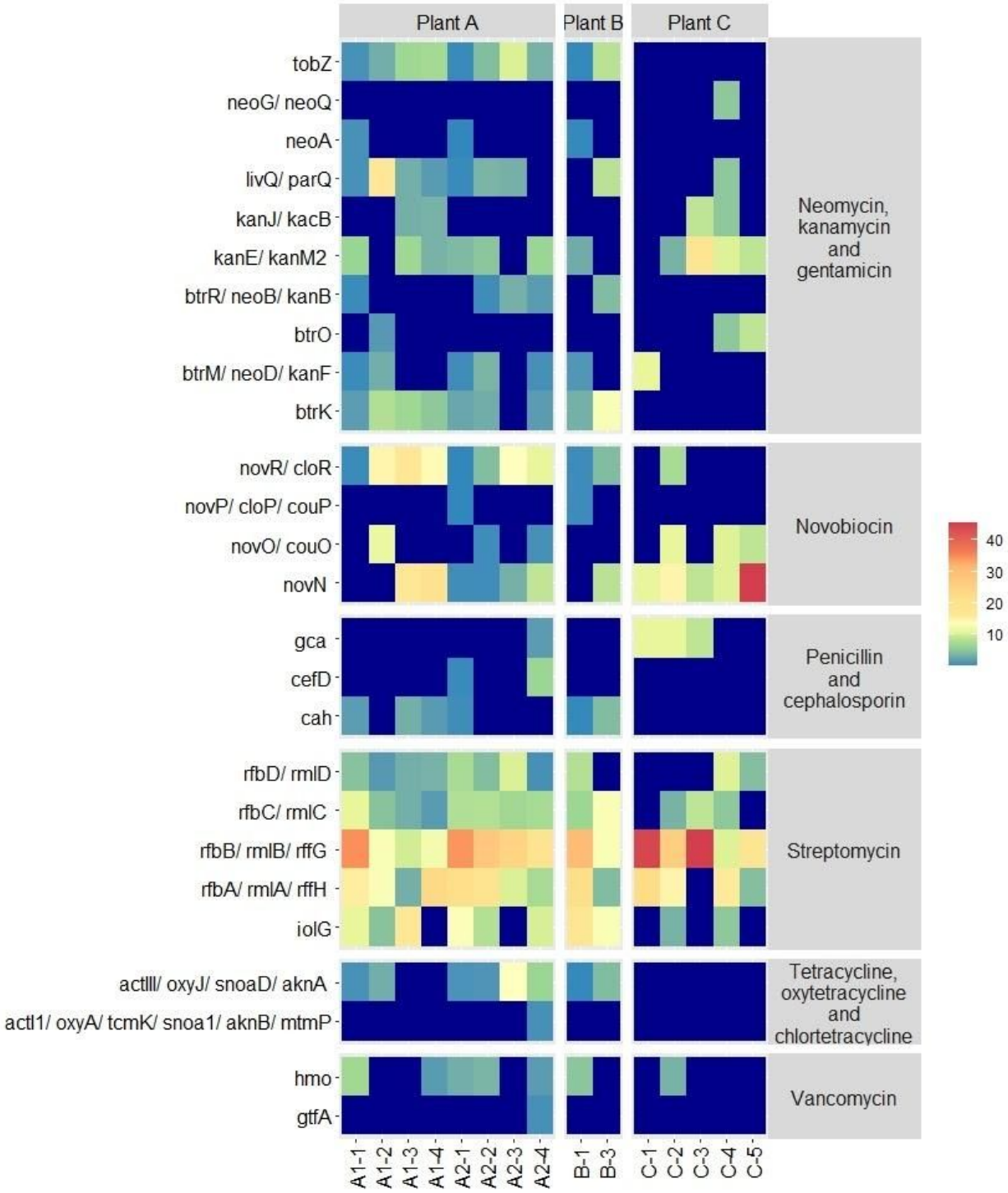

Fig. 5. Heatmap showing the relative abundance of ABSGs in WWTP samples. The relative abundances of the top ten prevalent ABSGs corresponding to each antibiotic are represented as the percentage of each gene per total ABSGs identified in samples and visualized in the heatmap. Colors reflect relative abundances of ABSGs from lowest (blue) to highest abundance (red). Dark blue indicates the absence of genes. Sample labels are identical as in Fig. 1

3.5. Occurrence and abundance of VFGs in the WWTP samples

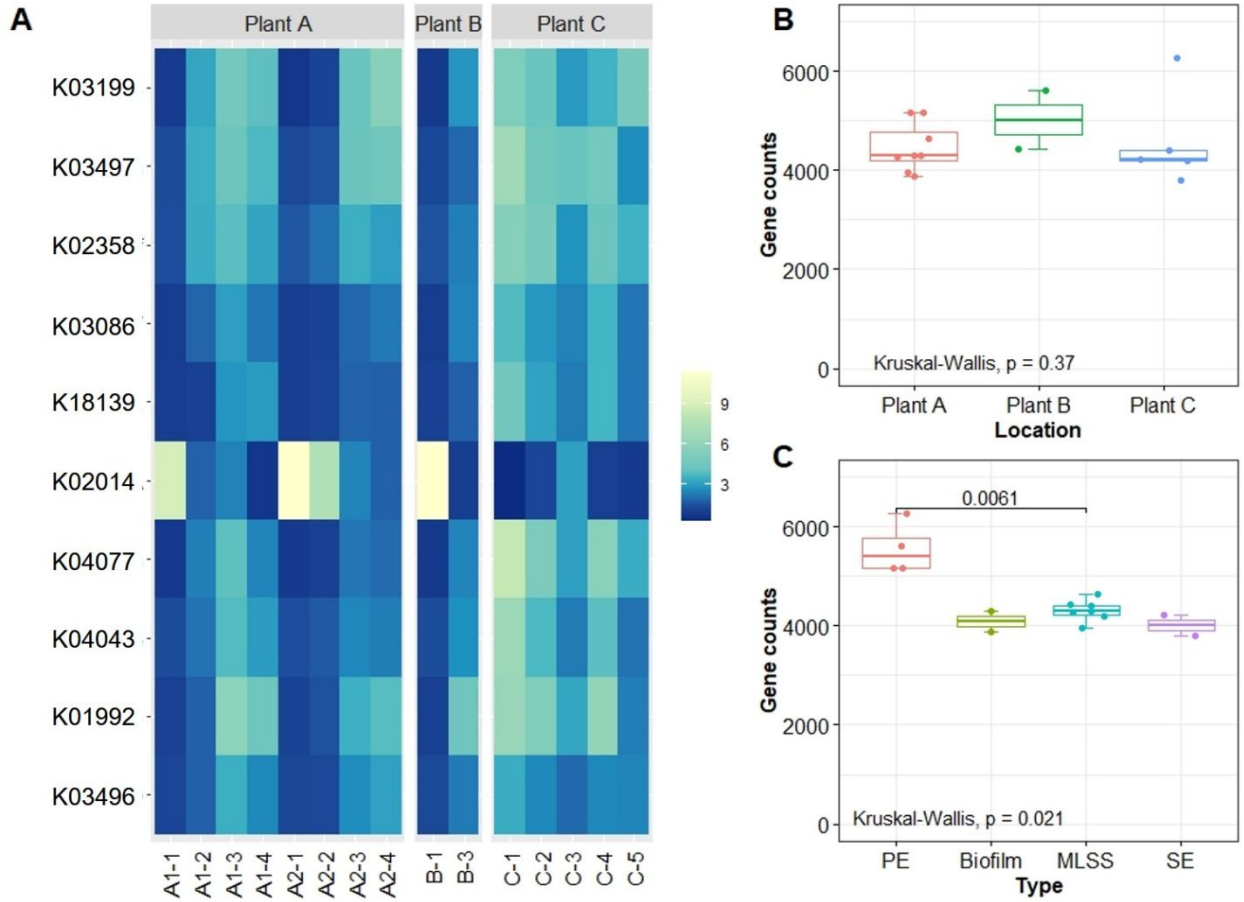

Fig. 6. Changes in abundance of VFGs in WWTP samples. **(A)** Heatmap showing the relative abundance of the top ten most abundant VFGs. Relative abundances are represented as the percentage of each virulence gene among total VFGs identified in a sample. P-values by Kruskal-Wallis with posthoc Wilcoxon rank-sum test for comparison between the normalized total VFG counts by **(B)** sampling location and by **(C)** treatment type are indicated in each boxplot with statistical significance at $p< 0.05$. Descriptions of VFG-associates KO are shown in Table S6. Sample labels are identical as in Fig. 1 and Fig. 3.

Since the acquisition of antibiotic resistance can facilitate the spread of pathogenic bacteria through the HGT of virulence plasmids with resistance determinants, we examined the occurrence and prevalence of genes encoding VFGs in our WWTP samples. Approximately 13.7 % unique virulence KOs (577 KO-based VFGs among 4,198 predicted KOs) were identified in samples (Table S6). Of them, VFGs involved in secretion systems, bacterial motility, efflux pump, iron uptake, and molecular chaperones were predominant among sewage samples. Figure 6A showed the occurrence and abundance of the top 10 most abundant VFGs detected in samples. VFG *K02014* involved in ferri-aerobactin uptake was prevalent in the primary effluents of plants A and B. Meanwhile, VFG *K03086* encoding a sigma factor of the extracytoplasmic function (ECF) family-sigma 70 was less abundant in primary effluents but increased in abundance in MLSS samples. These factors play essential roles in gene transcription of all

mycobacterial species and regulate the response to extracytoplasmic stimuli (Gomez et al., 1998). Virulence genes associated with the VirB/VirD4 type IV secretion system (*virB4* and *virB*), which mediates invasion, proinflammatory activation, and anti-apoptotic protection of endothelial cells (Schmid et al., 2004), were amplified in the MLSS and secondary effluents of all WWTPs. While the median of total VFG counts was slightly higher in sewage samples of plant B, the total gene counts of VFGs did not differ significantly between WWTPs (Kruskal-Wallis, $p = 0.37$, Fig. 6B). The median of total VFG counts, however, decreased notably following the treatment processes (Kruskal-Wallis, $p = 0.021$, Fig. 6C). Specifically, the median of total VFG counts in the MLSS samples was reduced by 20.4% in comparison with the primary effluent samples (Wilcoxon rank-sum, $p = 0.0061$, Fig. 6C, Table S5). These results highlight the efficiency of the sewage system in eliminating most virulence factors.

3.6. Co-occurrence patterns among ARG subtypes and VFGs

The calculated topological properties from the network analysis of ARG subtypes described the complicated interactions between ARGs. Figure 7A showed the co-occurrence network of ARG subtypes which consists of 46 nodes (ARG subtypes) and 114 edges. The average degree or the average number of connections was 4.957, and the average path length of the whole network was 2.964 with a network diameter of 8 edges. Modularity analysis showed a clustering coefficient of 0.487 and a modularity index of 0.397, indicating the existence of the modularity structure. Based on the modular classification, the network was partitioned into eight distinct modules, with the most densely connected nodes identified in the three largest modules: 1 (pink), 2 (light blue), and 3 (dark blue) in Figure 7A.

ARG subtype hubs were identified in three clusters (Figure 7A) as following: *lnuA_C_D_E/ lin* - Module 1 (pink); *mecR1* - Module 2 (light blue); and *marA*- Module 3 (dark blue). Each module harbored different classes of ARGs. For example, the gene *lnuA_C_D_E/ lin* was the hub of module 1 and co-occurred with *ampC* and *mecI* genes encoding resistance towards beta-Lactams, *eptB* gene conferring CAMP, and *cpxA*, *marR*, *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC* genes that are associated with multidrug resistance. Besides, *lnuA_C_D_E/ lin* also exhibited significant and positive correlations with genes from other modules such as *ompC*, *mdtO*, *blaTEM*, *catA*, *mdtL*, and *mdtF*. A possible explanation for the co-occurrence of ARG subtypes in each module is that these co-occurring genes may be shared by the same environment or harbored in the same microbial taxa.

In the co-exclusion pattern, *mdtF* was mutually exclusive with *cpxA*, *lnuA_C_D_E/ lin*, *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC*, *aph3-II* and *TC.MATE/ SLC47A/ norM/ mdtK/ dinF*. In addition, *emrE/ qac/ mmr/ smr* also revealed significant negative correlations with *macB*, *aph3-II*, and *mdtL*. These mutually exclusive genes might be shared by different environments or carried by specific bacteria species that express ARGs as their competitive advantage to inhibit the growth of other species in microbial communities.

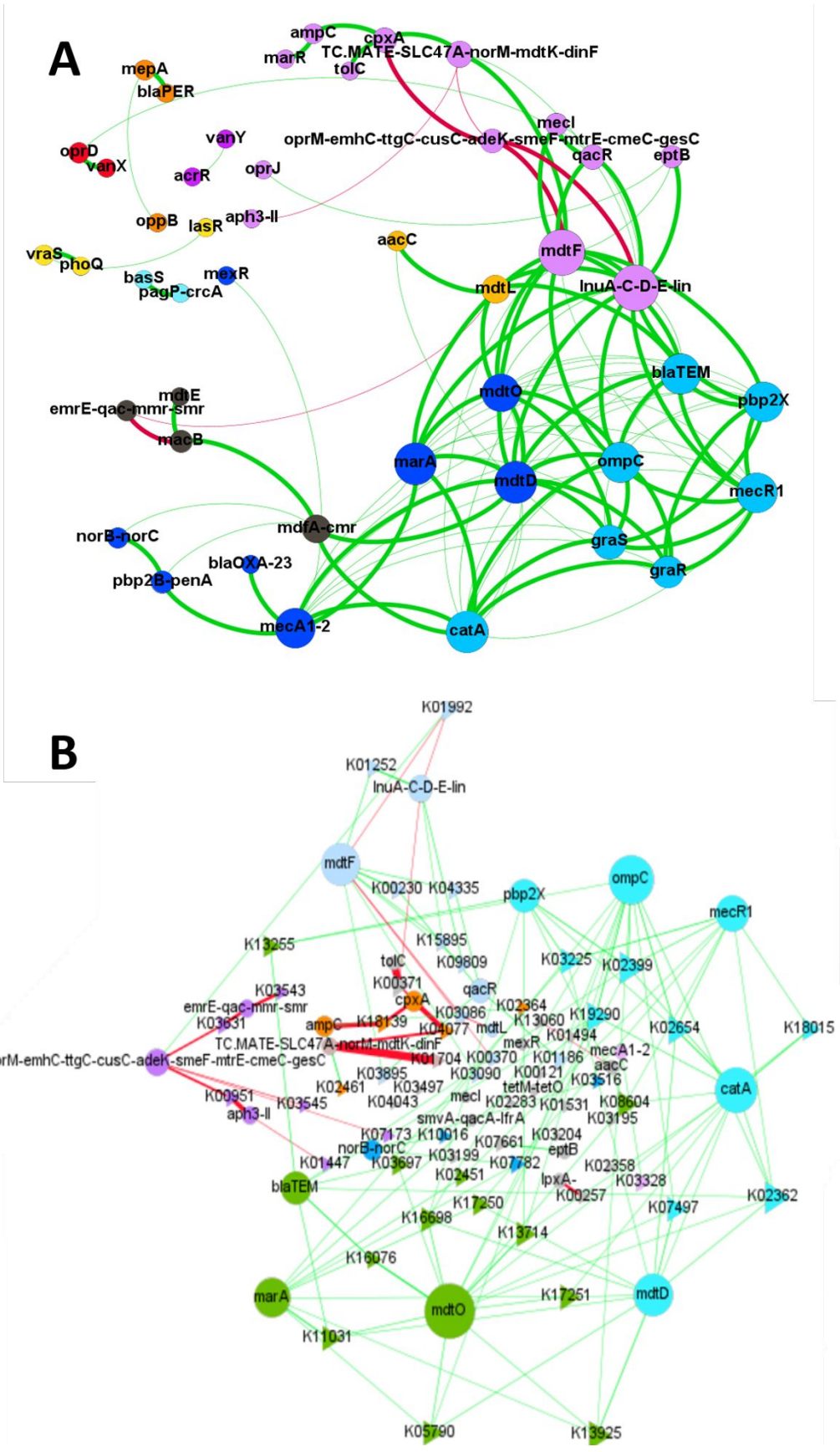
A
ampC
cpxA
marR
TC.MATE-SLC47A-norM-mdtK-dinF
tolC
mepA
blaPER
mecI
vanY
oprM-emhC-ttgC-cusC-adeK-smeF-mtrE-cmeC-gesC
oprD
vanX
qacR
eptB
acrR
oprJ
oppB
lasR
aph3-II
aacC
mdtF
vraS
phoQ
basS
mexR
pagP-crcA
mdtL
lnuA-C-D-E-lin
blaTEM
mdtO
pbp2X
mdtE
emrE-qac-mmr-smr
macB
marA
ompC
mdtD
mecR1
mdfA-cmr
norB-norC
graS
blaOXA-23
graR
pbp2B-penA
mecA1-2
catA
B
K01992
K01252
lnuA-C-D-E-lin
mdtF
K00230
K04335
pbp2X
ompC
mecR1
K13255
K15895
tolC
K09809
K03225
K02399
K00371
qacR
K03543
cpxA
emrE-qac-mmr-smr
K03086
K02364
K03631
ampC
K18139
K04077
mdtL
K13060
K19290
K02654
K18015
TC.MATE-SLC47A-norM-mdtK-dinF
mexR
K01494
oprM-emhC-ttgC-cusC-adeK-smeF-mtrE-cmeC-gesC
K01704
K00370
K01186
mecA1-2
K03895
K03090
K00121
aacC
K02461
K03497
tetM-tetO
K03516
catA
K00951
K03545
K04043
mecI
K02283
K01531
K08604
aph3-II
smvA-qacA-lfrA
K03195
K07173
K10016
K07661
K03204
norB-norC
K03199
eptB
K01447
K03697
K07782
K02358
K02451
lpxA-
K03328
blaTEM
K00257
K02362
K17250
K07497
K16698
K13714
K16076
marA
mdtO
K17251
mdtD
K11031
K05790
K13925

Fig. 7. The network analysis revealing the co-occurrence patterns among **(A)** ARG subtypes and **(B)** VFGs and ARGs. A connection between nodes represents a significant correlation between ARG subtypes, VFGs, and ARGs based on Spearman analysis (Rank coefficient $p > 0.8$ and adjusted p-value < 0.05). Colors of the connections indicate a positive correlation (green) or a negative correlation (red). The size of each node is proportional to the number of connections. The shape indicates ARGs and VFGs, and color indicates the modularity class.

The correlations between ARGs and VFGs were also demonstrated in Figure 7B. The resultant network consists of 87 nodes and 123 edges. Modularity analysis identified 26 modules in total with a modularity index of 0.776. Figure 7B shows the top seven dense modules, identified in the three largest modules: 1 (green), 2 (light blue), and 3 (purple). Every module exhibits a distinct coexistence constitution with cooccurrence of ARGs-VFGs between different classes of ARGs, indicating the co-selection of ARGs and VFGs. Amongst the most representative co-occurrence patterns between ARGs and VFGs fall into the functional category of multi-drug efflux pumps and beta-lactam resistance associated with either outer membrane proteins, secretion systems, transporters or adhesion mechanisms. For instance, the beta-lactam resistance gene *ompC* in module 1 showed strong positive associations with VFGs *K02399* and *K02654* involved in bacterial motility, *K19290* involved in serum resistance, *K03225* gene encoding type III secretion system, and *K02362* encoding siderophores. These coexistences contribute functional support for the colonization and growth of ARB in extreme niches. While most VFGs are positively correlated with ARGs, a significant co-exclusion pattern was observed between the multidrug resistance gene *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC* and six VFGs (*K00951*, *K01447*, *K03543*, *K03545*, *K03631*, *K07173*). Mutually exclusive ARGs also shared similar negative correlations with their associated VFGs. Indeed, VF *K01992* encoding the ATP-binding protein exhibited a positive correlation with *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC* but was exclusive with *lnuA_C_D_E/ lin*, *mdtF*, and *mdtL*. A significant, strong negative correlation was also identified between the multidrug resistance genes *cpxA*, *TC.MATE/ SLC47A/ norM/ mdtK/ dinF*, VFGs *K01704* and *K04077* associated with leucine biosynthesis and chaperone proteins respectively.

4. Discussion

Along with rapid population growth and economic development, the expansion of human residential areas and livestock farming has significantly contributed to pollution loads directly or indirectly into recipient bodies of water, posing profound threats to human and environmental health (Kongprajug et al., 2019). Essentially, the high antibiotic consumption in humans and in livestock resulted in the increased release of partially metabolized antibiotics into the aquatic environment through sewage systems (Fick et al., 2014). Previous studies have shown the prevalence of ARGs in influents, effluents, and activated sludge, suggesting municipal WWTPs can be a reservoir and vehicle for the transmission of pathogenic bacteria and antibiotic resistance. While the distribution of ARGs in WWTPs has been increasingly investigated, there is still a lack of comprehensive understanding of the fate of ARGs and the evolution of microbial communities along with different wastewater treatment processes. Our study demonstrates the impacts of treatment processes on the microbial community structure and the distribution of ARGs in different U.S. WWTPs.

4.1. Variations of bacterial communities in WWTPs

Shotgun metagenomic analyses revealed a distinct composition of microbial communities in primary effluents and at different stages of treatment processes (Fig. 1). The similar microbial composition observed in primary effluents is attributed to the similar source of wastewater and climate conditions in the same region that determine the microbial ecology (Wang et al., 2012). Recently, a study of global diversity and biography of bacterial communities in 269 WWTPs also reported the similarities in microbial composition between regional WWTPs (Zhang et al. 2012). Our results complement previous findings and indicate that the environmental conditions and locations of WWTPs play an influential role in shaping microbial communities in WWTPs

The diversity and composition of microbial communities are important factors influencing the fate and occurrence of ARGs. Alterations in the composition of microbial communities during the anaerobic digestion process have been shown to influence the biodegradation capacity of antibiotics and the occurrence of ARGs (Akyol et al.,2016). In the present study, WWTPs shared the major phyla including Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria regardless of geographic locations and treatment types. These results are consistent with previous reports on the predominance of these phyla in the sewage microbiome (Wei et al., 2018; Do et al., 2019). These core phyla are found to play crucial roles in organic matter degradation and nutrient cycling in wastewater treatment (Nascimento et al., 2018). While most bacteria in sewage microbiomes are important for degrading complex organic matters in wastewater, some of these bacteria could be pathogenic and hosts of ARGs. The top ten genera observed in our samples including *Arcobacter*, *Acinetobacter*, *Aeromonas*, and *Thauera* are of concern to human health. Specifically, *Aeromonas spp.* and *Arcobacter spp.* have been shown to cause a wide spectrum of human illnesses involving gastroenteritis, soft tissue, and wound infections, and have the potential to act as a reservoir of ARGs (Janda et al., 2010; Moura et al., 2012). Whereas, genera *Acinetobacter* and *Thauera* were identified as potential pathogens and host bacteria of different types of ARGs according to the UniProt database (https://www.uniprot.org/) and the Virulence Factors Database (VFDB, http://www.mgc.ac.cn/VFs/). The presence of these pathogenic bacteria and the proliferation of *Thauera* in the MLSS and effluent samples further caution that the activated sludge and effluent of WWTPs can be a vehicle for spreading pathogens and ARGs into the environment through sewage discharge.

4.2. The removal of ARGs and ABSGs in different WWTPs

We observed similar patterns of the ARG profiles from different WWTPs. In contrast with the recent report by Hendriksen et al. (2019) illustrating discrepancies in ARGs between countries across the world, the four WWTPs investigated in this study are from regions similar in sanitation and population health. The relative abundances of ARGs decreased following the wastewater treatment processes. Notably, total gene counts encoding resistance toward multi-drugs were significantly reduced, and dominant genes conferring resistance toward beta-Lactams were also slightly decreased upon the treatments (Fig. S8 and Table S5). Our results confirm previous findings that conventional WWTPs remove the majority of bacterial pathogens and their resistance genes (Yang et al., 2014; Li et al., 2015; Petrovich et al., 2018). Nevertheless, a general conclusion remains difficult because of the differences in the abundance metrics and normalization methods used.

Despite the decline of total ARGs following the treatments, genes encoding multidrug resistance such as *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC* and *acrB/mexB/adeJ/smeE/mtrD/cmeB* increased or persisted in the activated sludge. These resistant genes have been found to be involved in the MexA-MexB-OprM multidrug efflux systems that contributed to the intrinsic resistance of *Pseudomonas aeruginosa* (Li et al., 1995). Multidrug-resistant organisms are resistant to one or more classes of antimicrobial agents. The emergence and wide distribution of *Pseudomonas aeruginosa* and other multidrug-resistant gram-negative strains in nature pose a serious therapeutic challenge to the current treatment of infectious diseases as there are almost no antimicrobial agents available for the treatment of infections caused by these strains (Freire-Moran et al. 2011; Basak et al.; 2016).

Considering that activated sludge contains high concentrations of microorganisms and antibiotic residues in dense matrices, activated sludge may not only provide a suitable environment for the dissemination of ARGs through HGT between microorganisms but may contribute to the propagation of ABSGs. The antibiotic production of microorganisms has been shown to be prevalent in densely populated microbial environments such as soil (Chandra et al., 2017), where microorganisms compete and antagonize the growth of other groups by cooperating in socially cohesive units to produce biosynthetic antibiotics (Cordero et al., 2012b). Here, we observed a notable amplification of ABSGs-producing novobiocin and streptomycin in the TF biofilm and ASP MLSS (Fig. 5). Our results corroborate previous findings showing the prevalence of ABSGs-producing streptomycin in activated sludge (Petrovich et al., 2018), implying that conditions in the aerated activated sludge may have provided a selective advantage for the proliferation of microorganisms capable of producing streptomycin. Interestingly, we also noticed a higher abundance of ARGs conferring resistance to aminoglycosides (Fig. S9), suggesting the potential influence of *in situ* antibiotic production on the selection for ARGs conferring resistance to those compounds. However, future research is needed to evaluate the associations between bacterial taxa, ABSGs, and ARGs. This in turn would further clarify which bacterial taxa are associated with antibiotic production and antibiotic resistance, and how *in situ* production of antibiotics contribute to the within-process development of antibiotic resistance determinants.

4.3. The prevalence of VFGs in WWTPs and their correlation with ARGs

It has been evidenced that the acquisition of ARGs via HGT provides a specific advantage for the evolution and propagation of clinically important pathogens (Beceiro et al., 2013). Virulence factors, as elements encoded by virulence genes, are imperative for bacteria pathogens to overcome host defense systems, and the acquisition of antibiotic resistance enables pathogens to compensate for antimicrobial therapies and survive in adverse conditions. Our present study demonstrates a positive relationship between VFGs and ARGs identified in sewage samples. One of the reasonable explanations for this correlation between virulence genes and ARGs was that opportunistic pathogens likely harbor ARGs. Previous studies showed that prolonged exposure to sub-inhibitory concentrations of antibiotics in wastewater may favor the HGT of ARGs and enhance the bacterial virulence among opportunistic pathogens because of their plasticity and ability to adapt to diverse natural and nosocomial conditions through the acquisition of resistance

and persistence (Kim et al., 2014; Bruchmann et al., 2013). Contrary to this previous observation, several other studies have found an increase in antibiotic resistance accompanied by a decrease in the virulence levels of a microorganism (Evans et al., 1998; Geisinger et al., 2017). It has been reported that overexpression of the MexAB-OprM multidrug efflux system decreases the production of several extracellular virulence factors known to be regulated by quorum sensing (Evans et al., 1998). In the present study, we observed significantly negative associations in ARGs encoding multidrug efflux pump such as *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC* and VFGs involving stress response and quorum sensing. We also identified a strong negative correlation between multidrug ARGs *cpxA*, *TC.MATE/ SLC47A/ norM/ mdtK/ dinF*, and VFGs associated with leucine biosynthesis chaperone proteins. Previous studies have shown that leucine biosynthesis plays a crucial role in adaptation to nutrient starvation and increased virulence of most pathogenic bacteria and fungi (Chen et al., 2012; Orasch et al., 2019), while an increase of leucine metabolism resulted in a down-regulation of *TRI6* gene regulating the biosynthesis of mycotoxin in *Fusarium graminearum* (Subramaniam et al., 2015). These results indicate an intricate connection between ARGs and VFGs. Nevertheless, further investigation is required to understand the negative relationships between ARGs and VFGs identified in our study. These results reinforce our hypothesis that the prevalence of ARGs could be driven by the co-selection with virulence traits under selective pressure, resulting in the emergence of resistant clones.

The widespread occurrence of several classes of virulence determinants, including bacterial motility, efflux pump, secretions systems, and inflammatory mediators was identified in different WWTPs. In general, the total abundance of VFGs reduced following the treatment processes (Fig. 6C). However, ASP may selectively enrich certain VFGs as indicated by the increase of a few VFGs in MLSS samples, suggesting the need for future research to understand their potential clinical relevance and environmental impact. Preceding studies also reported increases in ARGs and VFGs in activated sludge (Zhang et al., 2016; Biswal et al., 2014), indicating activated sludge as a site for the enrichment of VFGs and ARGs. With the prevalence of VFGs identified in WWTPs, future studies that allow a clear understanding of the relationship between virulence and resistance genes are urgently encouraged.

4.4. Study limitations

This study is subject to several limitations worthy of discussion. First, an unequal and small sample size that was collected from each WWTPs could limit our ability to assess and compare microbial communities and gene profiles between different treatment processes. As an example, Plant B only consisted of three samples, but secondary effluent from Plant B was under-sampled and excluded from the analysis. As a result, the comparison of gene profiles between treatment types can be affected by random variation. In fact, it is possible that the small sample size may bias our results towards the null hypothesis, leading to a non-significant difference between the groups of comparison (Hackshaw et al.,2008). Small sample size also makes it difficult to generalize conclusions beyond the context of this study. With a larger sample, any significant differences would certainly be identified. However, the microbiome sample size is sufficient for the study as the mean sequencing depth of 3 million reads per sample was adequate to recover species-level taxonomic assignments and the diversity of different ARG subtypes present in the

sample. Indeed, Hillmann et al. (2018) recently reported that low-depth metagenomic sequencing of as few as 0.5 million reads per sample can recover broad-scale taxonomic changes and species profiles at > 0.05% relative abundance. The sample size of this study is also not much different from several impactful studies on the fate of ARGs in WWTPs that were analyzed with small samples (Yang et al., 2014; Osińska et al. 2020). Nevertheless, future research should explore a broader assessment with normally distributed samples from different treatment types of WWTPs to ascertain the influence of sewage treatments on the occurrence and mobilization of pathogenic bacteria and ARGs.

Another prominent limitation of the present study was that our functional analyses of ARGs and VFGs are mainly based on the KEGG Orthology (KO) database. The KEGG database provides detailed and well-annotated database resources for biological interpretation of genome sequences and high-level functions through the grouping of genes in KO, but it is not yet comprehensive. The functional contents in the KO database still remain incomplete, while many species and genera may have not been explicitly annotated. As a result, this underrepresentation or absence of KOs associated with species and genera in the database could limit our biological interpretation of important information from novel microbial species or genera (Dias et al., 2020). This limitation has been observed by other studies and major efforts have been made to identify and preserve functional genes of unique species globally (Kanehisa et al., 2017). In future studies, the use of additional databases and functional annotation for ARGs is necessary to expand the identification of all clinically relevant ARGs and provide greater clarity of a specific host-association of ARGs and ARB.

5. Conclusion:

Our study elucidates the influence of wastewater treatment processes on the composition of microbial communities and ARGs in four U.S. WWTPs. Particularly, the wastewater treatment process was shown to be the dominant factor in shaping the diversity and distribution of microbial communities and ARGs in WWTPs. Our findings demonstrate that conventional wastewater treatment processes are efficient at removing the majority of bacterial pathogens and ARGs in the effluent. However, the *in situ* biosynthesis of antibiotics and the evolution of virulent strains may induce the selection of certain ARGs throughout treatment processes. The relative abundances of certain ARGs encoding resistance towards multidrug and beta-Lactams, such as *oprM/emhC/ttgC/cusC/adeK/smeF/mtrE/cmeC/gesC* and *oppF*, increased in ASP MLSS. An increasing trend in prevalence was also observed for ABSGs and certain VFGs in ASP MLSS. The persistence or increase of these genes during ASP presents an environmental risk of disseminating pathogens and ARGs into the environment through sewage discharge or water reuse in urban and agriculture irrigation. Based on these results, future research should explore the relationship between microbial compositions and ARGs to identify possible host information of AGRs, especially for the potential pathogens, to control the emergence of resistant bacteria in WWTPs. In addition, continuing studies should investigate the associations between ARGs and ABSGs in order to support the observed trends. Overall, the results of this study support previous reports of possible ARGs dissemination from WWTPs to the environment, enhance our understanding of the fate of ARGs throughout treatment processes, and extend our knowledge of WWTPs regarding the prevalence and distribution of ABSGs and VFGs beyond just ARGs. In

addition, our findings provide data support for a surveillance of ARGs occurrence in WWTPs to develop antibiotic stewardship and applied engineering solutions to reduce the dissemination of ARGs through sewage discharge.

CRediT authorship contribution statement

Loan Le: Conceptualization; Data curation; Formal analysis; Investigation; Visualization; Writing - original draft, review & editing. Zhoujin Huang: Software; Data curation; Formal analysis; Investigation; Visualization; Writing- review & editing. Katrine Whiteson: Conceptualization; Supervision. Sunny Jiang: Conceptualization; Investigation; Funding acquisition; Supervision; Validation; Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financially supported by the Multidisciplinary Design Program (MDP) of the University of California, Irvine, and UCI Water-Energy Nexus Center. The authors would like to thank Dr. Zhiqian Song for his guidance and assistance of network analysis and the support from WWTPs for providing sewage samples for this analysis.

Appendix A. Supplementary data

References

CDC. *Antibiotic Resistance Threats in the United States, 2019*. Atlanta, GA: U.S. Department of Health and Human Services, CDC [Internet]. Available from: https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf [Accessed 27 March 2021].

Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. *P & T: a peer-reviewed journal for formulary management*. 2015;40(4):277–283.

Hughes D. Selection and evolution of resistance to antimicrobial drugs. *IUBMB life*. 2014;66(8):521–529.

Lerminiaux NA, Cameron A. Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian journal of microbiology*. 2019;65(1):34–44.

Knapp CW, Dolfing J, Ehlert PA, Graham DW. Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environmental science & technology*. 2010;44(2):580–587.

Middleton JH, Salierno JD. Antibiotic resistance in triclosan tolerant fecal coliforms isolated from surface waters near wastewater treatment plant outflows (Morris County, NJ, USA). *Ecotoxicology and environmental safety*. 2013;88:79–88.

Shi P, Jia S, Zhang XX, Zhang T, Cheng S, Li A. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water research*. 2013;47(1):111–120.

Su S, Li C, Yang J, Xu Q, Qiu Z, Xue B, et al. Distribution of Antibiotic Resistance Genes in Three Different Natural Water Bodies-A Lake, River and Sea. *International journal of environmental research and public health*. 2020;17(2):552.

Amarasiri M, Sano D, Suzuki S. Understanding human health risks caused by antibiotic resistant bacteria (ARB) and antibiotic resistance genes (ARG) in water environments: Current knowledge and questions to be answered. *Critical Reviews in Environmental Science and Technology*. 2020;50(19):2016–2059.

Yazdanpour Z, Tadjrobehkar O, Shahkhah M. Significant association between genes encoding virulence factors with antibiotic resistance and phylogenetic groups in community acquired uropathogenic Escherichia coli isolates. *BMC Microbiology*. 2020;20:241

Abd El-Baky RM, Ibrahim RA, Mohamed DS, Ahmed EF, Hashem ZS. Prevalence of Virulence Genes and Their Association with Antimicrobial Resistance Among Pathogenic *E. coli* Isolated from Egyptian Patients with Different Clinical Infections. *Infection and drug resistance*. 2020;13:1221–1236.

Alejandro B, María T, Germán B. Antimicrobial Resistance and Virulence: a Successful or Deleterious Association in the Bacterial World? *Clinical Microbiology Reviews*. 2013;26(2):185–230.

Dai C, Duvallet C, Zhang A, Matus M, Ghaeli N, Park S, et al. Multi-site sampling and risk prioritization reveals the public health relevance of antibiotic resistance genes found in wastewater environments. *BioRxiv*. 2019;562496.

Chen H, Zhang M. Occurrence and removal of antibiotic resistance genes in municipal wastewater and rural domestic sewage treatment systems in eastern China. *Environment International*. 2013;55:9–14.

Liu Z, Klümper U, Liu Y, Yang Y, Wei Q, Lin JG, et al. Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. *Environment International*. 2019;129:208–220.

EPA. *United States Environmental Protection Agency*. Available from https://www.epa.gov/npdes/npdes-permit-basics. [Accessed 20th August 2021].

Wei Z, Feng K, Li S, Zhang Y, Chen H, Yin H, et al. Exploring abundance, diversity, and variation of a widespread antibiotic resistance gene in wastewater treatment plants. *Environment International*. 2018;117:186–195.

Pazda M, Kumirska J, Stepnowski P, Mulkiewicz E. Antibiotic resistance genes identified in wastewater treatment plant systems – A review. *Science of The Total Environment*. 2019;697:134023.

Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*. 2012;2(1):3.

Li B, Yang Y, Ma L, Ju F, Guo F, Tiedje JM, et al. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME Journal*. 2015;9(11):2490–2502.

Petrovich M, Chu B, Wright D, Griffin J, Elfeki M, Murphy BT, et al. Antibiotic resistance genes show enhanced mobilization through suspended growth and biofilm-based wastewater treatment processes. *FEMS Microbiology Ecology*. 2018;94(5):fiy041. DOI:10.1093/femsec/fiy041.

Bengtsson-Palme J, Hammarén R, Pal C, Östman M, Björlenius B, Flach CF, et al. Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. *Science of The Total Environment*. 2016;572:697–712.

Rafraf ID, Lekunberri I, Sànchez-Melsió A, Aouni M, Borrego CM, Balcázar JL. Abundance of antibiotic resistance genes in five municipal wastewater treatment plants in the Monastir Governorate, Tunisia. *Environmental Pollution*. 2016;219:353–358.

Al-Ghalith GA, Hillmann B, Ang K, Shields-Cutler R, Knights D. SHI7 Is a Self-Learning Pipeline for Multipurpose Short-Read DNA Quality Control. *mSystems*. 2018;3(3):e00202-17.

Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. *Nucleic acids research*. 2005;33(Database issue):D325–D328.

Katiyar A, Sharma P, Dahiya S, Singh H., Kapil A., Kaur P. Genomic profiling of antimicrobial resistance genes in clinical isolates of Salmonella Typhi from patients infected with Typhoid fever in India. *Scientific Reports*. 2020;10(1):8299

Liu G, Chater KF, Chandra G, Niu G, Tan H. Molecular regulation of antibiotic biosynthesis in streptomyces. *Microbiology and molecular biology reviews: MMBR*. 2013;*77*(1):112–143.

Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biology*. 2010;11(10):R106.

Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media*. 2009;3(1):361-362.

Faust K, Raes J. CoNet app: inference of biological association networks using Cytoscape. *F1000Research*. 2016;5:1519.

RStudio Team. *RStudio: Integrated Development Environment for R, 2021*. RStudio, PBC, Boston, MA.

Do TT, Delaney S, Walsh F. 16S rRNA gene based bacterial community structure of wastewater treatment plant effluents. *FEMS Microbiology Letters*. 2019;*366*(3):fnz017. DOI:10.1093/femsec/fnz017.

Peterson E, Kaur P. Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens. *Frontiers in microbiology*. 2018;9:2928.

Gomez M, Doukhan L, Nair G, Smith I. sigA is an essential gene in Mycobacterium smegmatis. *Molecular Microbiology*. 1998;29(2):617–628.

Schmid MC, Schulein R, Dehio M, Denecker G, Carena I, Dehio C. The VirB type IV secretion system of Bartonella henselae mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. *Molecular microbiology*. 2004;52(1):81–92.

Kongprajug A, Chyerochana N, Somnark P, Leelapanang Kampaengthong P, Mongkolsuk S, Sirikanchana K. Human and animal microbial source tracking in a tropical river with multiple land use activities. *International Journal of Hygiene and Environmental Health*. 2019;222(4):645–654.

Faleye AC, Adegoke AA, Ramluckan K, Bux F, Stenström TA. Antibiotic Residue in the Aquatic Environment: Status in Africa. *Open Chemistry.* 2018;16(1):890-903.

Wang X, Hu M, Xia Y, Wen X, Ding K. Pyrosequencing analysis of bacterial diversity in 14 wastewater treatment systems in China. *Applied and environmental microbiology*. 2012;78(19):7042–7047.

Zhang T, Shao MF, Ye L. 454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *The ISME journal*. 2012;6(6):1137–1147.

Akyol Ç, Aydin S, Ince O, Ince B. A comprehensive microbial insight into single-stage and two-stage anaerobic digestion of oxytetracycline-medicated cattle manure. *Chemical Engineering Journal*. 2016;303:675–684.

Nascimento AL, Souza AJ, Andrade P, Andreote FD, Coscione AR, Oliveira FC, Regitano JB. Sewage Sludge Microbial Structures and Relations to Their Sources, Treatments, and Chemical Attributes. *Frontiers in microbiology*. 2018;9:1462.

Janda JM, Abbott SL. The genus Aeromonas: taxonomy, pathogenicity, and infection. *Clinical microbiology reviews*. 2010;23(1):35–73.

Moura A, Oliveira C, Henriques I, Smalla K, Correia A. Broad diversity of conjugative plasmids in integron-carrying bacteria from wastewater environments. *FEMS Microbiology Letters*. 2012;*330*(2):157–164.

Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, Lukjancenko O, Röder T, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat Commun*. 2019;10(1):1124.

Yang Y, Li B, Zou S, Fang HH, Zhang T. Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. *Water research*. 2014;*62*:97–106.

Li XZ, Nikaido H, Poole K. Role of mexA-mexB-oprM in antibiotic efflux in Pseudomonas aeruginosa. *Antimicrobial agents and chemotherapy*. 1995;39(9):1948–1953.

Freire-Moran L, Aronsson B, Manz C, Gyssens IC, So AD, Monnet DL, Cars O, ECDC-EMA Working Group. Critical shortage of new antibiotics in development against multidrug-resistant bacteria-Time to react is now. *Drug resistance updates: reviews and commentaries in antimicrobial and anticancer chemotherapy*. 2011;14(2):118–124.

Basak S, Singh P, Rajurkar M. Multidrug Resistant and Extensively Drug Resistant Bacteria: A Study. *Journal of pathogens*. 2016;4065603.

Chandra N, Kumar S. Antibiotics Producing Soil Microorganisms. In: Hashmi MZ, Strezov V, Varma A, editors. *Antibiotics and Antibiotics Resistance Genes in Soils: Monitoring, Toxicity, Risk Assessment and Management*. Springer International Publishing AG; 2017, vol. 51, p. 1–18.

Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F, et al. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science (New York, N.Y.)*. 2012;337(6099):1228–1231.

Zhang B, Xia Y, Wen X, Wang X, Yang Y, Zhou J, et al. The Composition and Spatial Patterns of Bacterial Virulence Factors and Antibiotic Resistance Genes in 19 Wastewater Treatment Plants. *PloS one*s. 2016;11(12):e0167422.

Biswal BK, Mazza A, Masson L, Gehr R, Frigon D. Impact of wastewater treatment processes on antimicrobial resistance genes and their co-occurrence with virulence genes in *Escherichia coli*. *Water research*. 2014;50:245–253.

Kim S, Yun Z, Ha UH, Lee S, Park H, Kwon EE, et al. Transfer of antibiotic resistance plasmids in pure and activated sludge cultures in the presence of environmentally representative micro-contaminant concentrations. *The Science of the total environment*. 2014;468-469:813–820.

Bruchmann J, Kirchen S, Schwartz T. Sub-inhibitory concentrations of antibiotics and wastewater influencing biofilm formation and gene expression of multi-resistant *Pseudomonas aeruginosa* wastewater isolates. *Environmental science and pollution research international*. 2013;20(6):3539–3549.

Geisinger E, Isberg RR. Interplay Between Antibiotic Resistance and Virulence During Disease Promoted by Multidrug-Resistant Bacteria. *The Journal of infectious diseases*. 2017;215(suppl_1):S9–S17.

Evans K, Passador L, Srikumar R, Tsang E, Nezezon J, Poole K. Influence of the MexAB-OprM multidrug efflux system on quorum sensing in Pseudomonas aeruginosa. *Journal of bacteriology*. 1998;180(20):5443–5447.

Chen JW, Scaria J, Chang YF. Phenotypic and transcriptomic response of auxotrophic Mycobacterium avium subsp. paratuberculosis leuD mutant under environmental stress. *PloS one*. 2012;7(6):e37884.

Orasch T, Dietl AM, Shadkchan Y, Binder U, Bauer I, Lass-Flörl C, et al. The leucine biosynthetic pathway is crucial for adaptation to iron starvation and virulence in *Aspergillus fumigatus*. *Virulence*. 2019;10(1):925–934.

Subramaniam R, Narayanan S, Walkowiak S, Wang L, Joshi M, Rocheleau H, Ouellet, T, Harris LJ. Leucine metabolism regulates TRI6 expression and affects deoxynivalenol production and virulence in Fusarium graminearum. *Molecular microbiology*. 2015;98(4):760–769.

Hackshaw A. Small studies: strengths and limitations. *European Respiratory Journal*. 2008;32(5):1141 LP-1143.

Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems*. 2018;3(6):e00069-18.

Osińska A, Korzeniewska E, Harnisz M, Felis E, Bajkacz S, Jachimowicz P, et al. Small-scale wastewater treatment plants as a source of the dissemination of antibiotic resistance genes in the aquatic environment. *Journal of hazardous materials*. 2020;381:121221.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*. 2017;45(D1):D353–D361.

Dias CK, Starke R, Pylro VS, Morais DK. Database limitations for studying the human gut microbiome. *PeerJ. Computer science*. 2020;6:e289.

Wittig U, De Beuckelaer A. Analysis and comparison of metabolic pathway databases. *Briefings in Bioinformatics*. 2001;2(2):126–142.