Persistence of Antibiotic Resistance from Agricultural Effluents to Surface Water Revealed in Metagenome Assembled Genomes

Jin Ju Kim
Chung-Ang University
Hoon Je Seong
Chung-Ang University
Chang-Jun Cha
Chung-Ang University
Timothy Alan Johnson
Purdue University
Woo Jun Sul (sulwj@cau.ac.kr)
Chung-Ang University https://orcid.org/0000-0002-7016-1454
Jong-Chan Chae
Jeonbuk National University

Research Article

Keywords: Antibiotic resistance, Antibiotic resistance genes, Resistome, Metagenome-assembled genome (MAG), Metagenome, Freshwater, Patescibacteria, Uncultured bacteria

Posted Date: January 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1201703/v1

Abstract

Background

Despite the increasing concerns about antibiotic resistance genes (ARGs) in wastewaters released from livestock or fish farms into the natural environment, studies on unculturable bacteria related to the dissemination of antibiotic resistance are limited. Here, we assessed the impact of the microbial antibiotic resistome and mobilome in wastewaters released in Korean rivers by reconstructing metagenome-assembled genomes (MAGs) using a novel refining method.

Results

We reconstructed 1,100 MAGs using Korean river and wastewater metagenome samples and presented Additional Clustering Refiner (ACR), a method that significantly reduced contamination of genome quality. Our results indicate that ARGs harbored in the MAGs were disseminated from wastewater effluents into the downstream environment of rivers. Among the effluent-derived phyla, uncultured members of the superphylum Patescibacteria harbored ARGs co-localized with mobile genetic elements. The most abundantly detected ARGs associated with mobile genetic elements in the wastewater effluents and downstream regions of the rivers were found to originate from Patescibacteria; the detected genes represented the potential for horizontal gene transfer of beta-lactam A-, sulfonamide-, and carbapenem-resistance genes.

Conclusions

Our study tracked microorganisms that were transmitted from agricultural wastewater to receiving rivers through genome-resolved metagenomics. Based on the genomic analysis of MAGs, it was found that ARGs are more commonly co-localized with MGEs in agricultural wastewater than in river water. Furthermore, this suggests that members of Patescibacteria, which are relatively unknown, are carriers that disseminate ARGs into the environment, and a potential vector for disseminating ARGs in the environmental bacterial community. Overall, this study suggests that dissemination of ARGs by uncultured bacteria should be investigated further in multiple environments.

Background

With growing concerns about the threat to human health posed by antibiotic resistance genes (ARGs) and antibiotic-resistant bacteria (ARB), environmental contamination and the fate of ARGs and ARB are also drawing attention. Globally, animal farming accounts for more than half of all antibiotics used[1]. The intensive use of antibiotics for livestock or fish farming results in their continuous release into the environment, subsequently causing the dissemination and increased prevalence of ARB and ARGs in aquatic environments[2–5].

Most antibiotics used in livestock feed are given for non-therapeutic purposes such as growth promotion, feed conversion efficiency, and disease prevention[6,7]. This is a concern because the general classification, function, and mode of action of antibiotics used in agriculture and veterinary medicine are similar or identical to those prescribed for humans[8]. In particular, the effluents from livestock waste treatment systems and the untreated wastewater from aquaculture farms are important hotspots of bacteria that can spread antibiotic resistance into the environment[9–11]. Effluents from animal farming, which exhibit a high density of bacteria, might promote a high rate of horizontal gene transfer to environmental bacteria and pathogens. Moreover, the occurrence of horizontal gene transfer provides evidence of co-selection with virulence factors (VFs) and metal resistance genes (MRGs) as well as ARGs[12,13] and, thus, pose a potential health risk to local residents.

Previously, studies using molecular methods based on qPCR or culturable bacteria have been widely conducted to investigate the abundance of ARGs[14,15,16,17]. However, this approach has limitations with regard to the number of primers validated for targeting ARGs and identification of resistance in unculturable bacteria. Cultivation-independent metagenomics of environmental samples is widely used to overcome these limitations. Still, even this approach often presents limitations in assigning a specific bacterial host to the identified ARGs and MGEs in the metagenomic data due to fragmented metagenomic assembly. Therefore, we aimed to determine the impact of wastewater originating from animal production farms on the microbiome and resistome in river environments using metagenome-assembled genomes (MAGs). This genome-resolved metagenomic approach is critical to understand the trajectories of unknown hosts with the potential risk of transmitting ARG/VF/MRG from wastewater to river environments.

Genome-resolved metagenomics helps to determine the functional profile of microbial communities through precise taxonomic classification from the species to strain level. Many state-of-the-art binning algorithms such as CONCOCT[18], MaxBin[19], and metaBAT[20] have been developed but are still being actively studied and improved. Since the binning programs mentioned above have advantages and disadvantages due to different algorithms, bin consolidation tools try to integrate bins and compensate for weaknesses by using multiple binning programs. DAS_Tool[21] aggregates bins from different binning programs and extracts a best quality consensus bin by evaluating single-copy genes. However, an aggregation method of DAS_Tool increases completeness, while contamination occurs accordingly. In addition, because the main source of error in MAGs, which are usually derived from metagenomic samples with high microbial diversity, occurs through the binning process rather than the assembly process[22], the consequences of chimeric MAGs lead to inflated phylogenetic diversity and inadequate inferences.

In the current study, we present the binning refiner Additional Clustering Refiner (ACR), an automated program designed to be easily used with multiple binners using iterative k-means clustering of contig abundance across different samples. Through appropriate benchmarking of various methods and evaluation of genome quality, ACR clearly reduces genomic contamination and demonstrates its

superiority in removing signals from other organisms. The ACR source code is available under a GPLv3 license at https://github.com/hoonjeseong/acr.

Therefore, the objective of the current study was to determine the impact of wastewater originating from animal production farm on the microbial community and the abundance of antimicrobial resistance genes in river environments. We used MAGs to determine the likely host the the resistance genes and to understand their genetic context. Furthermore, virulence factors and metal resistance genes were also identified in the MAGs and their proximity to MGEs and if identified in close proximity they were considered as putative resistance islands (pRIs)[23,24]. We hypothesized that animal farm wastewater would be elevated in antimicrobial resistance genes and that surface waters downstream from animal farms would also be elevated in those same genes. To our knowledge, this is the first study to identify the major host of transferable ARGs at the genome-level, along with the exploration of antibiotic resistome and mobilome. Our findings could deepen the understanding of the relationship between microbiome and resistome in environments that are markedly impacted by wastewater disturbance.

Methods

Experimental design and sample collection

In this study, freshwater samples were collected from 13 regions and 6 rivers, including the Han, Geum, Mankyoung, Nakdong, Seomjin, and Dongjin Rivers in Korea for 3 years (2016–2018). A total of 57 samples divided into 18 sample sets were obtained from effluents from 14 livestock wastewater treatment plants, untreated effluents from 4 aquaculture farms, and the upstream and downstream regions of the receiving rivers (Supplementary Fig. 1a, b). Each sample set comprised 3–4 samples, including upstream, wastewater, and downstream samples (listed in Additional File 1). The distance of sampling sites from the upstream to the downstream region, including the effluent outlet, was within 300 m to avoid the influence of tributaries. All freshwater and effluent samples were filtered using 0.2-μm cellulose ester membranes to collect microorganisms after pre-filtration with 10.0-μm cellulose ester membranes. The filter membranes were stored at -80℃ until DNA extraction.

DNA extraction

Total genomic DNA was extracted from the 0.2-μm cellulose ester membrane filters using the DNeasy PowerWater Kit (Qiagen). All procedures were followed according to the protocol recommended by the manufacturer. The concentration and quality of DNA were evaluated using PicoGreen. As a control experiment and to detect any possible contamination, the steps for genomic DNA extraction from a sterile membrane filter through which no water sample was filtered, were performed. The concentration of DNA extracted from the sterile membrane filter was not sufficient for PCR and sequencing.

16S rRNA gene amplification, sequencing, and processing

For bacterial community analysis, the hypervariable V3-V4 region[25,26] of bacterial 16S rRNA genes were amplified using the following primers: forward, 5'-TCGTCGCAGCGTCAGATGTG-TATAAGAGACAGCCTACGGGNGGCWGCAG-3' and reverse, 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTAT-CTAATCC-3' . To perform PCR amplification, 2 ng of extracted DNA was amplified using Herculase II fusion DNA polymerase (Agilent Technologies, Santa Clara, CA, USA), 500 nM each of the forward and reverse primers, 1 mM of dNTP mix, and 5× reaction buffer. The cycling condition involved heat activation at 95°C for 3 min, then 25 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s followed by the final extension at 72°C for 5 min. The quantity of amplicons was normalized and pooled, and the size of libraries were verified using the TapeStation DNA screentape D1000 (Agilent). Then, the validated samples were sequenced on the Illumina MiSeq platform using paired-end (2 × 300-bp) reads at Macrogen (Seoul, South Korea). The paired-end sequences were merged using the meren/illumina-utils script “iu-merge-pairs” with default parameters (*P* value = 0.3)[27]. From the merged reads, the bacterial community and diversity were analyzed using the Quantitative Insights Into Microbial Ecology (QIIME) pipeline (v.1.9.1)[28]. A total of 5,214,161 reads were merged from the paired-end sequences of 57 samples. Using these reads, *de novo* operational taxonomic unit (OTU) clustering was performed at 97% sequence similarity using the UCLUST[29] method “pick_otus.py”. Bacterial taxonomy was assigned using the Ribosomal Database Project (RDP) classifier[30], with reference to the Greengenes database[31] (confidence threshold of 50%). The OTUs assigned to the chloroplast and mitochondria were excluded from further analysis. Alignment of the representative sequences from each OTU was conducted using the PyNAST aligner[32]. FastTree[33] was used to construct a phylogenetic tree. Parameters of alpha diversity, including Chao1, phylogenetic diversity, observed OTUs, and Simpson index, were calculated with 12,069 rarefied reads per sample using the “multiple_rarefactions.py and alpha_diversity.py” tools. Distance matrices were estimated using unweighted and weighted UniFrac distance with phylogenetic tree construction (beta_diversity.py).

Shotgun metagenomic sequencing, quality control, and assembly

To identify resistomes and mobilomes in all 57 environmental samples (Additional File 1), shotgun metagenome sequencing was conducted on the Illumina HiSeq X platform using paired-end (2 × 150-bp) reads. Following Illumina HiSeq platform sequencing, a total of 1.01 Tb and an average of 17.8 Gb of sequence data were generated from the 57 metagenomic samples. Raw sequencing reads were quality filtered to remove reads having an average quality score below five using the FaQCs (v1.36) tool[34]. Duplicate paired-end reads were removed from all the filtered reads using FastUniq[35]. As a result, a total of 981.1 Gb and an average of 17.2 Gb sequence data were used for the following analysis. For each metagenome clean read, *de novo* co-assembly was performed by running MEGAHIT[36] (v1.1.2) using multiple K-mer sizes of 21, 29, 39, 49, 59, 69, 79, 89, 99, 109, 119, 129, and 141. The samples from the same region and year, including upstream, effluent, and downstream, were co-assembled together and are listed in Additional File 1. The eighteen co-assembled contigs were obtained using the “–presets meta sensitive, –min-contig-len 500” option.

Reconstruction of metagenome-assembled genomes (MAGs)

Genome binning for MAGs was performed with MaxBin2[19] (v2.2.4) using the abundance and tetra-nucleotide frequencies of contigs (≥ 1000 bp). The contig abundance required for genome binning was estimated by read mapping using BWA-MEM[37] (v.0.7.17-r1188) and Samtools[38] (v1.8).

To obtain better quality MAGs, we employed an additional clustering refiner, a new automated refinement strategy of MAGs using additional k-means clustering. In the first step, each open reading frame of the contigs was predicted using Prodigal[39] (v2.6.3), and bacterial and archaeal single-copy proteins were identified using HMMER[40] (v3.1b2). The 120 and 122 vertically inherited bacterial and archaeal universal single-copy proteins, respectively, were used to calculate the completeness and contamination of each MAG. Then, k-means clustering with contig coverage was iterated 2–15 times, and a genome was selected when each fragmented MAG satisfied the genome score. However, some contigs clustered into the new MAG may already be included in the previously selected MAG. In that case, the new MAG should be ignored because the contig could be used as a duplicate in the iterative clustering step (Supplementary Fig. 2). To obtain the representative MAGs, all MAGs were de-replicated using dRep at the default mode of the dereplicate workflow. Briefly, the dereplicate workflow in dRep compared the Average Nucleotide Identity (ANI) value of each MAG to select the MAG with the highest genome quality among the clusters with ANI values of 99%. Then, the quality of MAGs was assessed by CheckM using lineage-specific workflow, and the taxonomy of MAGs was assigned using GTDB-Tk[41] (v0.1.3).

To normalize the sequencing depth, the quality-filtered sequences of each sample were randomly subsampled (45 million r per sample), and the relative abundance of each MAG in each sample was calculated by mapping the reads to the MAG using Bowtie2[42] (v2.2.3). Moreover, reads with alignment score less than -20 were removed to avoid misalignment of read mapping. The relative abundance of MAGs in each sample was calculated as the ratio of the number of mapped reads to the total number of subsampled 45M reads. The MAGs were considered to be present in a sample if reads were mapped to more than 40% of the length of an MAG.

Phylogenies

Phylogenies of MAGs were constructed using PhyloPhlAn2[43], based on 400 universal marker genes using the “–diversity high and –fast” option; the phylogenies were visualized using iTOL[44].

Benchmark of Additional Clustering Refiner (ACR) software

In order to validate whether ACR improved the quality of the MAGs, the performance of ACR was evaluated using high-complexity Critical Assessment of Metagenomic Interpretation (CAMI) datasets[45]. The benchmark of binning results was performed with a gold-standard assembly which includes all genomic regions by assembly from the synthetic metagenomic reads of the reference genome sequence used in CAMI dataset and the different binners were evaluated to determine how well they recover the

reference genomes. Each read from five samples of CAMI dataset was mapped to the gold-standard contigs (https://github.com/CAMI-challenge/CAMISIM/wiki/User-manual#step-4-creation-of-the-gold-standards-and-anonymization) of CAMI dataset using BWA-MEM, and initial binning results were obtained from analyses using MaxBin2, MetaBAT2, and DAS Tool. The refinement step was processed by ACR to improve the binning results (called MaxBin2_ACR and MetaBAT_ACR). DAS Tool was not an algorithm for binning but was used to compare multiple binning results and select the best binning result. Therefore, for comparison with the DAS Tool results, ACR was performed on the results obtained using MaxBin2 and MetaBAT2[20], and then, a de-replication step was performed to recover the best results from other binning programs, based on the same concept as that of the DAS Tool (referred to as dRep_ACR here). Only near complete- and medium-quality MAGs were considered in subsequent comparisons. A comparative assessment of these genome binning results was performed with a package named Assessment of Metagenome BinnERs (AMBER).

We evaluated the recovery for closely related genomes called "common strains"[45] based on which genomes were recovered within the intra-species level (ANI greater than 95% against other genomes) through binning and refinement programs. The ANI values of each gold-standard genome of CAMI dataset was compared using FastANI, and genomes in the same species were identified by single-linkage clustering between genomes with ANI values ≥ 95%. Purity and completeness were calculated in the same way as that using CAMI and AMBER previously, and binning results were obtained when the precision was higher than 95%.

Statistical analysis

To determine the beta diversity of the MAGs, we calculated the dissimilarity with unweighted UniFrac distance matrix using the R package "GUniFrac (v1.1)" with the phylogenetic tree obtained from PhyloPhlAn2. Next, PCoA ordination was visualized with the cmdscale function of the "vegan (v2.5-7)" R package. Statistical tests were conducted to compare samples on the PC1 and PC2 axes using the paired Wilcoxon rank sum test, and the results were considered significant at $P < 0.05$. The evenness and richness were calculated for each sample, and pairwise comparison was conducted among the sampling groups (upstream, effluent, and downstream). The evenness was calculated using the "alpha.shannon" package from scikit-bio tool, and the richness was calculated by counting the number of MAGs generated from each sample (in the case that 1× or more reads covered more than 40% of the genome).

Functional annotation and metabolic pathway prediction

The open reading frames of each MAG were predicted using Prodigal, and the functional annotation was performed using EggNOG mapper version 2 using the default option, based on the EggNOG 5.0 database, via the DIAMOND. For the best performance of functional annotation, subsequent analysis was processed with e-value $1e^{-10}$ cutoff according to the benchmarking results to minimize the false positives (http://github.com/jhcepas/emapperbenchmark/blob/master/-benchmark_analysis.ip-ynb). Based on the KEGG orthology, metabolic pathways of each genome were constructed. The GhostKOALA tool was used

to avoid duplicate KEGG orthology results from eggnog-mapper, and the MinPath software was used to prevent overestimation of the results obtained.

Detection of ARGs, VFs, MRGs, and MGEs

ARGs were annotated with NCBI AMRFinderPlus and additionally identified using the Resfams core database by hmmsearch with a gathering threshold option. The drug class of ARGs was identified using AMRFinderPlus. Further, the MRGs and VFs were identified through DIAMOND blastp against the BacMet2 database and Virulence Factors Database (VFDB), respectively, using the following threshold: “id ≥ 70, query cover ≥ 70 and more sensitive”.

Before annotating the MGEs, the MGE types were specifically divided into transposon, integron, CRISPR, secretion system, phage, plasmid, *attC*, and other horizontal gene transfer machinery. Additionally, based on the gene descriptions from eggnog, a keyword search was performed as follows:

- Transposon: Transposase; Transposon; Tra[0-9]; IS[0-9]; Conjugate transposon; insertion
- Integron: Integrase; Recombinase; Resov
- CRISPR: Crispr; Cas[0-9]
- Secretion system: Type [A-Z]* secretion system
- Phage: Capsid; Phage; Tail; Head; Tape measure; Antitermination
- Plasmid: Relax.*; Conjug.*; trb; Mob.*; Plasmid; Toxin; Type IV; Chromosome partitioning; Chromosome segregation
- Other horizontal gene transfer machinery: Excision.*; Exonuclease; Topoisomerase; Reverse transcrip*

In addition, transposase, integrase, and attachment site (*attC*) were further identified against the transposase database with IS finder using BLASTp (evalue < 1e-30, qcov_hsp_perc > 90) and integrase database with Integron_finder using hmmsearch (evalue < 1e-3, query coverage > 50), and against the *attC* database using cmsearch (evalue <1 and size length filtering from 40 to 200 bp).

Putative Resistance Island (pRI) analysis

In the network analysis, the number of ARG-MGE connections within each MAG were summarized at the phylum level. ARGs and MGEs were connected when they were co-localized within a distance of 4 kbp. Gephi generated the network showing the relationship between the ARGs, MGEs, and host using the Force Atlas 2 algorithm. Node size and edge thickness were determined based on the degree of each node and log value of the number of connections between two nodes, respectively.

To identify the putative Resistance Island (pRI), the 4 kbp flanking regions of ARGs, VFs, MRGs, and MGEs were searched in an MAG. When ARGs, VFs, MRGs, and MGEs were adjacent to each other (≤ 4 kbp), they were clustered (single linkage) as a pRI (containing ≥ 1 MGEs and ≥ 1 ARGs or VFs or MRGs). Moreover, the detected pRIs were classified into plasmid, phage, Class I integrons, Clusters of attC sites

lacking integron-integrase and integrative and conjugative elements. For all contigs of pRIs, plasmids were identified using plasmidVerify script from MetaSPAdes, phages were detected by VirSorter in “level 1, 2, 4 and 5” categories, and integrative and conjugative elements were defined by ConjScan and T4SSscan of MacSyFinder (mandatory ≥ 2 and accessory ≥ 2 in a contig). However, for integrative and conjugative elements, a pRI was classified as an integrative and conjugative elements only if it was within the range of integrative and conjugative elements. Class I integron designation required the presence of an integrase and an *attC* site, but clusters of attC sites lacking integron-integrase designation required multiple *attC* sites without an integrase. pRIs were considered present in a sample if sequencing reads were mapped with more than 70% breadth of 1× coverage in the pRI region of MAGs and then the prevalence of pRIs was calculated in each group (upstream, effluent, and downstream).

Quantification of ARG, VF, MRG, MGE, and pRI

Regions of ARG, VF, MRG, MGE and pRI were extracted in GFF format, and the number of reads mapped to those regions was calculated with the featureCounts[46] for quantification. Additionally, the relative abundance of each gene and region was normalized by reads per kilobase of each gene per million reads of metagenomic sample (RPKM) (Equation 1). Especially, the relative abundances of all proteins encoding ARGs, VFs, MRGs, and MGEs were corrected once more using the medium[47] of 40 universal single-copy core gene RPKM values. The ARGs, VFs, MRGs, MGEs, and pRIs were compared using the paired Wilcoxon rank sum test, and the *P*-values were corrected using the Benjamini–Hochberg method.

$$\mathrm{RPKM} = \frac{mapped\ reads\ of gene\ /gene\ length\ (kilobase)}{total\ reads\ (million)} \quad \text{(Equation 1)}$$

Results

Additional Clustering Refiner (ACR) benchmarking with Critical Assessment of Metagenome Interpretation (CAMI) dataset

We verified the ACR method by comparing the obtained results with those of other binning methods using a synthetic metagenome dataset from CAMI[45]. The gold-standard assemblies of the high-complexity datasets of CAMI were binned with MaxBin2[19], MetaBAT2[20], and DAS Tool[48]. Moreover, each original binning results were refined with ACR to compare their performances. However, since the DAS Tool uses an algorithm that combines results from multiple binners, we used ACR with dereplicated MAGs (referred as dRep_ACR) from multiple binning results, including those of the analyses using the MaxBin2 and MetaBAT2 tools, to compare the number of recovered MAGs obtained. Overall, the numbers of near-complete and medium-quality MAGs recovered by the ACR method were higher than those recovered as per the binning results of the analyses using MaxBin2, MetBAT2, and DAS Tool (Fig. 1c). In particular, the ACR analysis of MetaBAT2 and dRep recovered the largest number of medium-quality MAGs (466 and

447, respectively), while the original binning results of MetaBAT2 and DAS Tool recovered 400 and 394 medium-quality MAGs, respectively (Fig. 1c).

The CAMI data set consisted of genomes with various similarities according to ANI values. To evaluate how ACR solves the problem when closely related genomes are present, we scrutinized a set of genomes called “common strains” with an ANI greater than 95% against other genomes[45]. A total of 204 gold-standard genomes exhibited ≥ 95% ANI, when compared to other genomes. In the MAG recovery among common strains, ACR improved the binning results of MetaBAT2 and dRep (Fig. 1d, e). Particularly, refined MetBAT2 produced 108 medium-quality MAGs, recovering almost half of the gold-standard genomes. The results of the analysis using dRep_ACR showed that the number of recovered strains was lower than that obtained in case of the MetaBAT2_ACR, showing a recovery of 88 medium-quality MAGs, since the MAGs with an ANI value of 99% or more were dereplicated. The ACR method significantly reduced the contamination of MAGs and resulted in an increase in the total number of medium-quality MAGs, and implied the possibility of recovering strain-level MAGs.

Construction of >1000 metagenome-assembled genomes (MAGs) using the Additional Clustering Refiner (ACR) method

For metagenomic analysis, co-assembly and binning were performed for each sample set to recover near-complete microbial genomes. A total of 9,467 genome bins were initially generated, and then dereplicated at ANI 99%. The dereplicated 1,040 genome bins were qualified using the quality criteria of MAGs[49], and then 181 near-complete (NC) MAGs (≥ 90% completeness and <5% contamination) and 335 medium-quality MAGs (≥ 50% completeness and <10% contamination) were constructed. The initial binning results showed a broad range of contamination (0–25%) of each MAG (Fig. 1b). Thus, we applied a new method, the ACR, which split the contigs of the MAGs using k-means clustering algorithm (Fig. 1a), to obtain MAGs with better quality. This process nearly doubled the good quality MAGs to 257 near-complete and 653 medium-quality MAGs, compared to the initial binning results. ACR maintained the completeness through CheckM similar to the initial binning results but significantly reduced contamination and strain heterogeneity (Supplementary Fig. 3a). Furthermore, the chimerism was detected by GUNC[50] based on the contigs lineage homogeneity constituting each MAG. Accordingly, it was observed that MAGs (medium-quality and near-complete) that passed the GUNC also significantly improved through ACR in the initial binning results (Supplementary Fig. 3b). Consequently, we utilized ACR in our study to reconstruct 1,100 MAGs, which consisted of 257 near-complete and 683 medium-quality MAGs with an average of 89.43 ± 7.18% completeness with 5.79 ± 4.07% contamination from 57 metagenomes (Fig. 2). All MAGs were classified (bacteria (n=1,095) and archaea (n=5)) using Genome Taxonomy Database (GTDB). The MAGs were affiliated to 28 bacterial phyla, including Proteobacteria (n=371), Bacteroidota (n=295), Patescibacteria (n=136), Actinobacteriota (n=132), Verrucomicrobiota (n=35), Bdellovibrionota (n=31), Planctomycetota (n=19), Omnitrophota (n=16), Cyanobacteriota (n=14), Chloroflexota (n=9), Campylobacterota (n=8), Dependentiae (n=5), Firmicutes (n=5), and others. Despite using the GTDB reference, only 5.18% and 67.73% of MAGs were classified at the species and genus

levels, respectively. Therefore, most of the MAGs belonged to unclassified taxa in this study, indicating that they could belong to novel species and genera.

Prevalence Of Args In Korean Freshwater Mags

The ARGs in the MAGs were detected using AMRFinderPlus and Resfams, and the antibiotic class of ARGs was classified based on AMRFinderPlus classification[51]. Of the 1,100 MAGs, 1,030 possessed a total of 5,804 ARGs related to 16 antibiotic classes (Fig. 2). Except for those of the efflux class that is also related to cellular process, toxicity, pathogenicity, and transport of a wide range of substrates, the most abundant ARGs found in 738 MAGs encoded resistance against beta-lactams (1,007 genes), tetracycline (256 genes), and aminoglycosides (88 genes). A primary goal of this study was to determine the taxonomic hosts of ARGs; therefore, we focused our analysis on the genes found in near-complete and medium-quality MAGs.

Proteobacteria MAGs possessed a variety of ARGs belonging to the beta-lactam, tetracycline, macrolide, and aminoglycoside classes. Besides, ARGs for fosfomycin and sulfonamide and class C beta-lactamase (*ampC*) were found only in Proteobacteria MAGs (Additional File 2). Bacteroidota MAGs specifically harbored *aadS* (aminoglycoside resistance), *blaOXA-1* (class D beta-lactamase), *mefC*, *mphG*, *mphE*, and *msrE* (macrolide resistance), *catB* (phenicol resistance), and *tet39* and *tetX* (tetracycline resistance). Patescibacteria was the only phylum with MAGs possessing *aadA1* (aminoglycoside resistance), *blaPFM* (class B beta-lactamase), and *lnuC* (lincosamide resistance). Actinobacteriota was the only phylum with MAGs that harbored *aadA11* (aminoglycoside resistance), *ermX* (macrolide resistance), *rox* (rifamycin resistance), and *tetV* (tetracycline resistance). Other ARGs found in MAGs are described in Additional File 2.

There were 42 MAGs harboring ARGs to more than three antibiotic classes, which we refer to as multi-drug resistant (MDR) MAGs (Fig. 2). Clinically relevant ARGs targeting antibiotics underscored by World Health Organization (WHO) as important for human medicine were found in MDR MAGs[52]. Most MDR MAGs belonged to Proteobacteria, Bacteroidota, and Actinobacteriota and carried genes that confer resistance to efflux, beta-lactam, tetracycline, and aminoglycosides. In particular, the MAGs with the most diverse types of ARGs found belonged to *Moraxellaceae*, a well-known reservoir of ARGs which confer resistance to macrolides and phenicol. *Microtrichales* and *Fluviicola*, which had not been reported as encoding ARGs to date, were both found to contain MDR MAGs in our study.

Microbial dissemination from effluents into the downstream region of receiving rivers

The effect of wastewater effluents on rivers was assessed by the analysis of microbiomes using the MAG. Significantly different microbiomes among the upstream and downstream regions and effluents were detected in the multivariate analysis of MAG composition using unweighted UniFrac distance matrix [$P < 0.001$, determined by the analysis of similarities (ANOSIM)] (Fig. 3a). River water (including upstream and downstream) microbiome was clearly distinguished from the effluent microbiome. Notably, the

downstream microbiome was significantly different from upstream microbiome in both the PC 1 and PC 2 axes [$P < 0.05$, $P < 0.01$, respectively, Wilcoxon rank sum test (paired), Fig. 3a]. This indicated that the upstream microbiome was altered by the influence of the effluents, resulting in a different downstream microbiome.

We also investigated the bacterial community using 16S rRNA gene amplicon sequencing. A total of 26,422 distinct operational taxonomic units (OTUs) were obtained by removing OTUs containing 10 or fewer sequences. Principal coordinate analysis (PCoA) of OTUs showed that the effluent microbiome showed a significantly different composition compared to that of the river water microbiome [$P < 0.001$, Wilcoxon rank sum test], based on the unweighted UniFrac distance matrix (Supplementary Fig. 4). Additionally, the richness of the effluent microbiome was significantly lower than that of the river water microbiome both in the MAG and 16S rRNA gene analyses, suggesting the lower microbial diversity in effluents (Supplementary Fig. 5).

More MAGs were assigned to Patescibacteria than any other phyla in 88.9% of effluent samples, but only in 38.5% of river water samples (FDR=0.011, Fisher's exact test). Campylobacterota, Omnitrophota, Dependentiae, and Bdellovibrionota were also dominantly found in effluents, compared to the river water samples (P <0.05, Fisher's exact test). In particular, the abundance of these five taxa showed a significant increase in the downstream regions of the rivers compared to the upstream regions, suggesting that the microbial population in aquatic environments was affected by the effluents (Fig. 3b). Especially, the phyla Parcubacteria and Saccharibacteria, in the superphylum Patescibacteria, were dominant in effluents and also found in the downstream regions of the rivers (Supplementary Fig. 6).

In order to assess the dissemination at the genetic level, the ratio of the abundance of ARGs to that of the core genes was calculated. Overall, the relative abundance of ARGs in MAGs was higher in the effluents than in the river water, and higher in the downstream regions than in the upstream regions of the rivers (Fig. 4a). In particular, specific aminoglycoside (*aadA1*, *aadA5*, *aadA2*), beta-lactam (*blaOXA-10*), phenicol (*floR*), sulfonamide (*sul2*), tetracycline (*tetX*, *tetA*), and macrolide (*ereD*, *mefC*) resistance genes were abundant in the effluent MAGs, and their abundance was higher in MAGs from the downstream regions than in the upstream regions of the rivers (Fig. 4b).

In addition to ARG distribution, the abundances of metal resistance genes (MRGs), mobile genetic elements (MGEs), and virulence factors (VFs) were also significantly higher in the effluent than in the river waters (Supplementary Fig. 7). In addition, the abundance of MGEs and VFs were higher in MAGs from the downstream than in the upstream regions, indicating that MGEs and VFs, as well as ARGs, were disseminated from the effluents into the downstream regions.

For each MAG, we investigated whether it originated from effluent or river water environments. The generated MAGs were divided into 2 groups: River Group and Effluent Group. The River Group consisted of MAGs found in the upstream and downstream regions regardless of the effluent sample. MAGs in the River Group might be considered native to the river microbiome. On the other hand, Effluent Group was composed of MAGs simultaneously found in effluent and downstream, while not found in the upstream

samples, under the assumption that MAGs exist downstream due to the influence of effluent. An MAG that appeared only in effluent was also assigned to the Effluent Group (Fig. 5a). The presence of an MAG was defined as one with more than 40% genome breadth of coverage at each sampling site. A total of 542 and 460 MAGs were classified into the River Group and Effluent Group, respectively (Fig. 5a).

Although Proteobacteria, Bacteroidota and Actinobacteriota were the most prevalent phyla both in River Group and Effluent Group (Supplementary Fig. 8), the distribution of the MAGs differed between the two groups at levels lower than the order taxon (Additional File 2). In Proteobacteria MAGs classified as Betaproteobacteriales, Rhodobacterales, Sphingomonadales, and Steroidobacterales were more prevalent in the River Group, while MAGs classified as Micavibrionales and Pseudomonadales were more prevalent in the Effluent Group. MAGs assigned to the Chitinophagales, NS11-12g, and Flavobacteriales orders in Bacteroidota were more prevalent in the River Group, while MAGs classifieds as Bacteroidales were only detected in the Effluent Group. In Actinobacteria MAGs, Actinomycetales, and Microtrichales were more prevalent in the River Group.

Interestingly, the phyla Patescibacteria, Bdellovibrionota, Omnitrophota, Campylobacterota, and Dependentiae were mostly found in Effluent Group but not in River Group, with the exception of only two MAGs (Supplementary Fig. 8). This result suggested that MAGs belonging to these five phyla originated from effluents and existed in the downstream region after effluents merged into the rivers. These five phyla most commonly encoded beta-lactam resistance genes in addition to various other ARGs (Supplementary Fig. 9). MAGs from the genus *Arcobacter* (Campylobacterota phylum), considered a pathogen, harbored class B and D beta-lactamases (Additional file 2). Interestingly, *blaPFM*, which is responsible for carbapenem resistance, was detected in MAGs classified as Patescibacteria (Supplementary Fig. 9). Other carbapenem resistance genes (*blaDIM* and *blaOXA-151*) were detected in MAGs classified as Proteobacteria from Effluent Group, and *cphA* were detected in both Effluent Group and River Group. In addition, Patescibacteria, Bdellovibrionota, and Campylobacterota MAGs in the Effluent Group harbored genes that confer resistance to aminoglycosides, while Patescibacteria, Omnitrophota, and Bdellovibrionota MAGs in the Effluent Group harbored genes that confer resistance to tetracycline.

In each group, no significant difference was found with regard to the genome size and the number of ARGs, MRGs, and VFs. However, the number of MGEs was significantly higher in MAGs of the Effluent Group than in those of the River Group (Fig. 5a).

Furthermore, the detection frequency of MGEs at a given distance from ARGs in MAGs was higher in the Effluent Group than in the River Group; this difference gradually increased in proportion to the distance between the MGEs and ARGs on the genome (Fig. 5b). In particular, Patescibacteria and Campylobacterota, found mostly in Effluent Group, had a higher number of MGEs per genome size compared to the other phyla (Supplementary Fig. 10). In addition, Proteobacteria, Bacteroidota, and Actinobacteriota, which represent the common phyla in all samples, also showed a higher incidence of MGEs from ARGs in the MAGs of Effluent Group (Supplementary Fig. 11).

Co-localization Of Args And Mges In Mags

To assess the potential of horizontal transfer of ARGs from the effluent, the genetic linkage between ARGs and MGEs in the MAGs categorized at the phylum level in both groups were investigated (Fig. 6). Network analysis was performed by searching for ARGs within the 4 kbp flanking region of the MGEs. In general, this network displayed a different pattern of harboring ARGs and MGEs in hosts between Effluent Group and River Group. The number of triangles refers to the number of phyla that have a genetic link between an ARG and an MGE. The higher number of triangles occurred in Effluent Group, which means that the observed genetic link between ARG and MAG in the host is more frequent than in the River Group. In addition, in quantity and diversity of the ARGs and the MGEs harbored in phyla, Effluent Group represented a more complex network than River Group (Effluent Group: 126 nodes, 388 edges / River Group: 83 nodes, 226 edges). The ARGs were prevalent in broad ranges of taxa, with 25 and 14 phyla in Effluent Group and River Group, respectively. Proteobacteria, Bacteroidota, Verrucomicrobiota, and Actinobacteriota were the main phyla harboring ARGs with nearby MGEs in both groups.

Resistance islands persist from effluents to the downstream regions of rivers

Putative resistance islands (pRIs) containing ARGs with MGEs in 4-kbp-flanking regions of each gene were searched to ascertain the potential mobility of ARGs (Fig. 7). A total of 498 putative pRIs were found in 367 MAGs, including 3 class I integrons, 6 Clusters of *attC* sites lacking integron-integrases, 8 phages, 30 plasmids and 1 integrative and conjugative elements. Among them, only eight kinds of pRIs frequently appeared in the effluent (more than 10 sampling sites). The prevalence of these eight pRIs in the downstream regions was dependent on that in the effluent. In some cases, pRIs were detected at a higher frequency in the downstream regions than in the upstream region (Fig. 7b) due to contamination from the effluent. These pRIs carried resistance genes specific for plasmid and Clusters of *attC* sites lacking integron-integrase meditated resistance against clinically relevant antibiotics (Fig. 7).

pRI-166 containing class D beta-lactamase genes (*blaOXA*-2) and pRI-584 containing phenicol resistance genes (*floR*) were frequently found in effluents and were more abundant in the downstream regions, than in the upstream regions, suggesting a dissemination of the ARGs (Fig. 7). pRI-118 in *Pseudomonas E* had two aminoglycoside resistance genes (*aph(3")-Ib* and *aph(6)-Ib*) adjacent to Tn3 transposon. It was also remarkably widespread in more than 90% of the effluent samples and were more abundant in the downstream regions than in the upstream regions. Moreover, a short 2-kbp pRI-1238 containing genes for lincosamide resistance and IS10 family transposase was not found in the upstream regions of the rivers, but mainly found in the effluents and downstream regions of the rivers. These pRIs mainly originated from Patescibacteria (pRI-166 and pRI-1238) and Proteobacteria (pRI-584 and pRI-118). Moreover, Patescibacteria harbored a higher number of MGEs per genome size than most other phyla (Actinobacteriota, Bacteroidota, Bdellovibrionota, Chloroflexota, Omnitrophota, Planctomycetota, and Verrucomicrobiota with adjust $P < 0.05$) (Supplementary Fig. 10).

Since pRI-166, a Cluster of *attC* sites lacking integron-integrase with *blaOXA*-2 found in Patescibacteria, was composed of a short contig, a contig graph was used to extend the sequence continuity (Supplementary Fig. 12). As a result, *aadA5*, *aacA8*, and *blaGES-5*, which are responsible for the resistance to clinically relevant aminoglycosides and carbapenem, were detected in an additionally assembled contig that was not used to reconstruct MAGs. Besides, the integrase gene also found in this contig graph, pRI-166, classified as Clusters of *attC* sites lacking integron-integrase, might actually be a Class I integrons. This suggested that along with Proteobacteria, Patescibacteria could be one of the important carriers of ARGs, causing their dissemination into river environments.

Discussion

Concerns about the increase in the prevalence of ARGs and ARB in the aquatic environments of rivers receiving wastewater have been raised[53–57]. However, only little attention has been paid to investigation of uncultured bacteria carrying ARGs in the environment. Although previous studies were conducted to identify bacterial hosts of ARGs using MAG construction, there was a limitation that MAGs with high quality were not reconstructed[58,59]. Furthermore, studies on uncultured bacteria harboring ARGs have not yet been closely examined.

We compared the number of ARGs between binned and unbinned contigs to see how many ARGs were binned into MAGs. All sulfonamide and rifamycin resistance genes in 1 kbp contigs were binned into MAGs, but in most drug groups only about 30% of ARGs were binned into MAGs (Supplementary Fig. 13). Although the metagenome binning approach could only include a subset of ARGs, as has been shown previously in genome assembly[60], there are significant advantages in considering MAG-binned ARGs, including the ability to determine the host and context of ARGs. Therefore, in this study, the reconstruction of MAGs via metagenomic analysis was employed to investigate the specific bacteria and mobilomes responsible for the dissemination of ARGs from wastewater effluent of livestock or fish farms into the downstream regions of rivers.

Contamination during the reconstruction of MAG can significantly induce erroneous genomic interpretation about phylogeny and functional genes[61]. Estimation of MAG contamination can sometimes be limited in approaches based on only the single-copy marker gene proposed by CheckM[62]. This is due to "non-redundant contamination"[50], which is often replaced or expanded with non-overlapping single-copy marker genes that are not associated with a single organism, leading to a chimeric genome[50,61]. ACR substantially effectively reduces contamination (even "non-redundant contamination") by removing the signal of other organisms through the characteristics of contig abundance that vary in our environmental metagenome data (Supplementary Fig. 14). The case of a bin named ‘Y-2016_509’, which was observed as highly contaminated (completeness 99.84 and contamination 85.81), was split into two medium-quality MAGs (Y-2016-509.0 –- 80.88 and 1.72 / 83.62 and 7.93: completeness and contamination of refined two MAGs, respectively) through ACR. One showed an ANI value of 98.27 (66% alignment fraction) with *Arcobacter cryaerophilus*, and the other was taxon

assigned to the family level (*Babeliaceae*) through the RED value of GTDB. Therefore, ACR could purify and separate contaminated MAG, helping to gain more insight into unknown organisms in the environment.

Genomes recovered from the metagenome are referred to as MAGs by the phylogenetic context and are also called population genomes[63]. The diversity and strain heterogeneity of organisms determine MAGs from strain- to species-level. ACR can separate the strain-level MAGs of the “common strain” from the synthetic metagenomic benchmarking data. However, in a real metagenome, the resolution of strains from close relatives is hard to be segregated during contig assembly. Many contig fractions or only represent consensus sequences[45,64] occur in an environment with multiple strains, making it challenging to strain-level genome-resolved metagenomics. On the other hand, in an environment where bacterial diversity is relatively low according to individual samples (such as gut microbiome), organisms exhibit low strain heterogeneity, therefore, MAG recovery at strain resolution using ACR is expected.

ACR generated 1,100 microbial MAGs in the freshwater samples from 6 rivers, including river water samples and samples from effluents of livestock or fish farms; more than 93% of MAGs harbored ARGs. Despite comparing to the approximately 140,000 prokaryotic genomes of GTDB-Tk, most (>90%) MAGs in Korean freshwater samples could not be assigned at the species level (ANI < 95%). Moreover, MAGs assigned to phyla with no known cultured representatives, including Patescibacteria, Omnitrophota, and Dependentiae, were detected in Korean freshwater, especially in the effluent samples.

MDR bacteria have become a major public health threat especially in clinics[65,66]. Most MDR bacteria identified via the MAG analysis harbored resistance genes that confer resistance to aminoglycosides, beta-lactams, macrolides, and tetracyclines, which have been widely administered to humans, livestock, and fish. The corresponding genes were mostly distributed in Proteobacteria and also found in Bacteroidota, Actinobacteriota, and Patescibacteria. Although several studies reported that Proteobacteria, Bacteroidota, and Actinobacteriota are associated with multiple-ARGs in rumen[67] and cattle fecal samples[68], as well as wastewater treatment plants[58], no studies have investigated Patescibacteria harboring ARGs thus far. In this study, metagenomic analysis provided the evidence that uncultured Patescibacteria (in addition to the other taxa listed) also play a role in the dissemination of ARGs originating from the wastewater of livestock and fish farming into natural aquatic environments.

The most abundant ARGs in the effluents were also detected in the downstream regions. The resistance genes for tetracycline (*tetX*, *tetA*), beta-lactam (*blaOXA*-10), sulfonamides (*sul2*), phenicols (*florR*), macrolides (*ereD*, *mefC*), and aminoglycosides (*aadA1*, *aadA5*, *aadA2*) were prevalent in effluents as well as in the downstream region of the rivers. In particular, according to the statistics regarding the antibiotic sales in Korea provided by the Korean Ministry of Food and Drug Safety, penicillin, which belongs to the beta-lactam family, and tetracycline were the most frequently used antibiotics in the animal production industry; sulfonamides, phenicols, macrolides, and aminoglycosides were also frequently used (Additional Table 3). It has been reported that ARGs that confer resistance to tetracyclines and

sulfonamide are frequently found in sewage, receiving water, and waste sludge environments of the livestock farms[69,70].

Overall, our study is consistent with these previous studies in that tetracycline and sulfonamide resistance genes are dominant in the wastewater effluent and receiving environment. In addition, it has been reported that the tetracycline resistance genes persists in aquaculture environments even after at least 4 years without the use of antibiotics[71]. Therefore, the results of both the previous and our studies imply that the resistance genes disseminated from agricultural wastewaters to the downstream rivers could potentially remain in the river environment long-term and substantially cross environmental barriers, which may endanger human health.

The MAG analysis revealed that members of the phyla Patescibacteria, Campylobacterota, Omnitrophota, Bdellovibrionota, and Dependentiae were prevalent in effluents and disseminated into the downstream environment. Among the effluent-associated phyla, Campylobacterota had a high number of VFs and MGEs in their MAGs (Supplementary Fig. 10).

Most Campylobacterota MAGs were affiliated to *Arcobacter* and were detected in the effluents, but one Arcobacter MAG was also found in the receiving water. *Arcobacter* MAGs harbored multiple ARGs conferring resistance to aminoglycosides, beta-lactams, and macrolides and had several VFs associated with the outer membrane components (lipooligosaccharide [*waaF*, *gmhA*] and capsular polysaccharide [*ugd*]), flagella and autoinducer for quorum sensing (*luxS*). *Arcobacter* is a notorious food and waterborne pathogen for humans[73]. *Arcobacter* infection in humans causes bacteremia-associated enteritis, which is similar to the clinical symptoms of campylobacteriosis, but with a high frequency of persistent watery diarrhea[74–76]. Moreover, previous studies have reported that *Arcobacter* is positively correlated with the concentration of antibiotics[77] and can accumulate a large number of ARGs[78]. These results indicate that pathogens carrying multidrug resistance might be concentrated and released from agricultural effluents. Therefore, their environmental impact should be watched and further studied from a “One-Health” perspective.

Patescibacteria is an uncultured candidate phyla radiation (CPR), which is found in groundwater[79,80], sediments, lakes, and other aquifer environments[81–83]. Among effluent-associated phyla, a large number of Patescibacteria MAGs were reconstructed in our study (Supplementary Fig. 8) and had the most significant effect on receiving water (Fig. 3B). Although Patescibacteria MAGs have a small genome size, various ARGs have been found, and this raised questions about the potential of HGT of ARGs. To identify the exerted selective pressure on ARGs in Patescibacteria by polluted environments, we compared Patescibacteria MAGs reconstructed from ours and other studies[81,84] (Supplementary Fig. 15).

It has been found previously that individual members of the Patescibacteria phylum possess highly streamlined genomes, which may lead to metabolic dependencies for “public goods” or metabolites from the rest of the ecosystem. This relationship is called the “Black Queen Hypothesis”[85], which posits that despite dependence on “public goods”, reduced genome size acts as a selective advantage in the

environment, as it lowers the replication cost. Despite reduced genome size, the Patescibacteria obtained from this study and activated sludge possessed various types of ARGs, especially class A and D beta-lactamase and tetracycline resistance genes (Supplementary Fig. 15). Conversely, however, most ARGs were not detected in Patescibacteria MAGs originating from the pristine environment (groundwater), except for the efflux class genes. Our analysis suggests that despite the fact that Patescibacteria have reduced functional capacity and ultra-small cell size[80], they encode ARGs in agricultural wastewaters. As a result, polluted environments such as human and farm wastewater impose an ecological cost that requires, even Patescibacteria, to encode ARGs.

There are three hypothetical explanations for why Patescibacteria contain high amounts of ARGs compared to their genome size in wastewater environments: (i) a higher density of MGEs compared to those found in other phyla were found, which might drive an increased genomic plasticity in Patescibacteria (Supplementary Fig. 10), (ii) the CRISPR/Cas9 system for defense to foreign DNA (virus or plasmid) is absent in Patescibacteria[80]; thus, the probability of horizontal gene transfer via the easy uptake of genetic elements is high, and (iii) the biosynthetic gene clusters for some secondary metabolites were more abundantly detected in the Effluent Group samples, suggesting a survival advantage during natural selection (for instance, the genes for monobactam biosynthesis were more prevalent in the Effluent Group than in the River Group (FDR=0.00041; Supplementary Fig. 16). Consequently, these characteristics implied an ecological role of Patescibacteria as ARG carriers, despite their very small cell and genome size[80].

This study used genome-resolved metagenomics to identify potential mobile regions in MAGs and detected them in association with ARGs. Clusters of *attC* sites lacking integron-integrases, Class I integrons, phages, and plasmids, which are known as important mobile genetic elements in spreading ARGs,[86–89] were found co-localized with ARGs. In particular, the possible co-localization of a Class I integron with *blaOXA-2* and *sul1* in Patescibacteria, indicates the possibility that bacteria in this phylum could carry genes for resistance against beta-lactams and sulfonamides into the downstream environment. A recent study reported that Patescibacteria harbored *blaOXA* and *qacEΔ1*, which is a clinical Class I integrons marker, as detected using EpicPCR[90], supporting evidence that the Patescibacteria are carriers of beta-lactam resistance genes with Class I integrons in aquatic environments. Further, *blaOXA-2* is known to confer resistance to oxacillin and also to carbapenem depending on the host[91]; hence, the fate of these genes is of considerable interest. Thus, we suggest that Patescibacteria, a relatively under-studied bacterial phylum, may be an important carrier of ARGs.

Conclusions

We developed ACR to enable automated and accurate genome contamination refinement of MAGs to better understand microbial communities from a genome-centric perspective. This study was conducted to determine the impact of wastewater resistome on the downstream environments using 1100 MAGs. Five major phyla (Patescibacteria, Bdellovibrionota, Omnitrophota, Campylobacterota, and Dependentiae)

were consistently found in agricultural wastewater to receiving rivers, and their genomes showed common characteristics of co-localization between ARG and MGE. Among them, the relatively unknown Patescibacteria contained ARGs despite its small genome size. In particular, there was a tendency for more ARGs to be encoded with MGEs in their genomes in the polluted environments (wastewater or activated sludge) compared to the natural environment (groundwater), indicating the role of uncultured bacteria in the dissemination of ARGs. Here, we scrutinized the prevalence of ARGs and their relationship to MGEs through large-scale genome-centric metagenomics and suggested new possibilities for disseminating ARGs from wastewater/effluents into natural aquatic environments.

List Of Abbreviations

ARG
Antibiotic resistance gene
MAG
Metagenome-assembled genome
ACR
Additional Clustering Refiner
ARB
Antibiotic resistant bacteria
VF
Virulence factor
MRG
Metal resistance gene
MGE
Mobile genetic element
pRI
Putative resistance island
OTU
Operational taxonomic unit
ORF
Open reading frame
MDR
Multi-drug resistant
ANOSIM
Analysis of similarities
PCoA
Principal coordinate analysis
FDR
False discovery rate
ANI

Average nucleotide identity
CPR
Candidate phyla radiation
EpicPCR
Emulsion, Paired Isolation and Concatenation PCR
MLSB
Macrolide-Lincosamide-Streptogramin B
CAMI
Critical Assessment of Metagenomic Interpretation

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The raw sequence data from shotgun metagenome sequencing and 16S rRNA gene amplicon sequencing were submitted to NCBI SRA under BioProject accession number PRJNA658184 and PRJNA719269.

Competing interests

The authors declare that they have no competing interests

Funding

This study was funded by the Korea Ministry of Environment (MOE) as “The Environmental Health Action Program” under Project No. 2016001350005.

Authors’ contributions

These authors contributed equally: Jin Ju Kim, and Hoon Je Seong.

W.J.S., J.-C.C., and C.-J.C. conceptualized the idea, designed the experiments. J.J.K. and H.J.S. performed the experiments, wrote the manuscript and analyzed the data. J.J.K, H.J.S, T.A.J, W.J.S. and J.-C.C. reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

Correspondence to Woo Jun Sul or Jong-Chan Chae.

Acknowledgements

Not applicable

Authors' information

Department of Systems Biotechnology, Chung-Ang University, Anseong, 17546, Republic of Korea

Jin Ju Kim, Hoon Je Seong, Chang-Jun Cha and Woo Jun Sul

Department of Animal Sciences, Purdue University, West Lafayette, IN 47907, United States

Timothy A. Johnson

Division of Biotechnology, Jeonbuk National University, Iksan 54596, Republic of Korea

Jong-Chan Chae

References

1. He Y, et al. Antibiotic resistance genes from livestock waste: occurrence, dissemination, and treatment. npj Clean Water. 2020;3:4. doi:10.1038/s41545-020-0051-0.
2. Aminov RI. The role of antibiotics and antibiotic resistance in nature. Environ Microbiol. 2009;11:2970–88. doi:10.1111/j.1462-2920.2009.01972.x.
3. He LY, et al. Discharge of swine wastes risks water quality and food safety: Antibiotics and antibiotic resistance genes from swine sources to the receiving environments. Environ Int. 2016;92-93:210–9. doi:10.1016/j.envint.2016.03.023.
4. Rodriguez-Mozaz S, et al. Occurrence of antibiotics and antibiotic resistance genes in hospital and urban wastewaters and their impact on the receiving river. Water Res. 2015;69:234–42. doi:10.1016/j.watres.2014.11.021.
5. Taylor NG, Verner-Jeffreys DW, Baker-Austin C. Aquatic systems: maintaining, mixing and mobilising antimicrobial resistance? Trends Ecol Evol. 2011;26:278–84. doi:10.1016/j.tree.2011.03.004.
6. You Y, Silbergeld EK. Learning from agriculture: understanding low-dose antimicrobials as drivers of resistome expansion. Front Microbiol. 2014;5:284. doi:10.3389/fmicb.2014.00284.
7. Hong PY, Al-Jassim N, Ansari MI, Mackie RI. Environmental and Public Health Implications of Water Reuse: Antibiotics, Antibiotic Resistant Bacteria, and Antibiotic Resistance Genes. Antibiotics (Basel). 2013;2:367–99. doi:10.3390/antibiotics2030367.
8. Islam KS, Shiraj-Um-Mahmuda S, Hazzaz-Bin-Kabir MJ. J. o. P. H. i. D. C. Antibiotic usage patterns in selected broiler farms of Bangladesh and their public health implications. 2, 276–284 (2016).
9. Jia S, et al. Fate of antibiotic resistance genes and their associations with bacterial community in livestock breeding wastewater and its receiving river water. Water Res. 2017;124:259–68. doi:10.1016/j.watres.2017.07.061.

10. Vikesland PJ, et al. Toward a Comprehensive Strategy to Mitigate Dissemination of Environmental Sources of Antibiotic Resistance. Environ Sci Technol. 2017;51:13061–9. doi:10.1021/acs.est.7b03623.
11. Santos L, Ramos F. Antimicrobial resistance in aquaculture: Current knowledge and alternatives to tackle the problem. Int J Antimicrob Agents. 2018;52:135–43. doi:10.1016/j.ijantimicag.2018.03.010.
12. Pan Y, et al. Coexistence of Antibiotic Resistance Genes and Virulence Factors Deciphered by Large-Scale Complete Genome Analysis. *mSystems* **5**, doi:10.1128/mSystems.00821-19 (2020).
13. Li LG, Xia Y, Zhang T. Co-occurrence of antibiotic and metal resistance genes revealed in complete genome collection. ISME J. 2017;11:651–62. doi:10.1038/ismej.2016.155.
14. Schmidt GV, et al. Sampling and Pooling Methods for Capturing Herd Level Antibiotic Resistance in Swine Feces using qPCR and CFU Approaches. Plos One. 2015;10:e0131672. doi:10.1371/journal.pone.0131672.
15. Chen H, Zhang M. Occurrence and removal of antibiotic resistance genes in municipal wastewater and rural domestic sewage treatment systems in eastern China. Environ Int. 2013;55:9–14. doi:10.1016/j.envint.2013.01.019.
16. Karkman A, et al. High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS Microbiology Ecology* 92 (2016).
17. Zhu Y-G, et al Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proceedings of the National Academy of Sciences* **110**, 3435-3440 (2013).
18. Alneberg J, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11:1144–6. doi:10.1038/nmeth.3103.
19. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605–7. doi:10.1093/bioinformatics/btv638.
20. Kang DD, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359. doi:10.7717/peerj.7359.
21. Sieber CMK, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3:836–43. doi:10.1038/s41564-018-0171-1.
22. Mineeva O, Rojas-Carulla M, Ley RE, Scholkopf B, Youngblut ND. DeepMAsED: evaluating the quality of metagenomic assemblies. Bioinformatics. 2020;36:3011–7. doi:10.1093/bioinformatics/btaa124.
23. Hu Y, et al. The Bacterial Mobile Resistome Transfer Network Connecting the Animal and Human Microbiomes. Appl Environ Microbiol. 2016;82:6672–81. doi:10.1128/AEM.01802-16.
24. Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev. 2004;17:14–56. doi:10.1128/cmr.17.1.14-56.2004.
25. Fernandes T, Vaz-Moreira I, Manaia CM. Neighbor urban wastewater treatment plants display distinct profiles of bacterial community and antibiotic resistance genes. Environ Sci Pollut Res Int.

2019;26:11269–78. doi:10.1007/s11356-019-04546-y.

26. Helsens N, et al. Antibiotic Resistance Genes and Bacterial Communities of Farmed Rainbow Trout Fillets (Oncorhynchus mykiss). Front Microbiol. 2020;11:590902. doi:10.3389/fmicb.2020.590902.
27. Eren AM, Vineis JH, Morrison HG, Sogin M. L. A filtering method to generate high quality short reads using illumina paired-end technology. PLoS One. 2013;8:e66643. doi:10.1371/journal.pone.0066643.
28. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6. doi:10.1038/nmeth.f.303.
29. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1. doi:10.1093/bioinformatics/btq461.
30. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7. doi:10.1128/AEM.00062-07.
31. DeSantis TZ, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72. doi:10.1128/AEM.03006-05.
32. Caporaso JG, et al. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26:266–7. doi:10.1093/bioinformatics/btp636.
33. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641–50. doi:10.1093/molbev/msp077.
34. Lo CC, Chain PS. Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics. 2014;15:366. doi:10.1186/s12859-014-0366-2.
35. Xu H, et al. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One. 2012;7:e52249. doi:10.1371/journal.pone.0052249.
36. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6. doi:10.1093/bioinformatics/btv033.
37. Li HJ. a. p. a. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint* (2013).
38. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. doi:10.1093/bioinformatics/btp352.
39. Hyatt D, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119. doi:10.1186/1471-2105-11-119.
40. Eddy SR, Accelerated Profile HMM, Searches. PLoS Comput Biol. 2011;7:e1002195. doi:10.1371/journal.pcbi.1002195.
41. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927, doi:10.1093/bioinformatics/btz848%J Bioinformatics (2019).

42. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. doi:10.1038/nmeth.1923.
43. Segata N, Bornigen D, Morgan XC, Huttenhower C PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications* **4**, 1-11, doi:ARTN 230410.1038/ncomms3304 (2013).
44. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23:127–8. doi:10.1093/bioinformatics/btl529.
45. Sczyrba A, et al. Critical Assessment of Metagenome Interpretation-a benchmark of metagenomics software. Nat Methods. 2017;14:1063–71. doi:10.1038/nmeth.4458.
46. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30. doi:10.1093/bioinformatics/btt656.
47. Mende DR, Sunagawa S, Zeller G, Bork P. Accurate and universal delineation of prokaryotic species. Nat Methods. 2013;10:881–4. doi:10.1038/nmeth.2575.
48. Sieber CMK, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology. 2018;3:836–43. doi:10.1038/s41564-018-0171-1.
49. Bowers RM, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35:725–31. doi:10.1038/nbt.3893.
50. Orakov A, et al. GUNC: detection of chimerism and contamination in prokaryotic genomes. Genome Biol. 2021;22:178. doi:10.1186/s13059-021-02393-0.
51. Feldgarden M, et al. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob Agents Ch* 63, e00483-00419, doi:ARTN e00483-1910.1128/AAC.00483-19 (2019).
52. Collignon PJ, et al. World Health Organization Ranking of Antimicrobials According to Their Importance in Human Medicine: A Critical Step for Developing Risk Management Strategies to Control Antimicrobial Resistance From Food Animal Production. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2016;63:1087–93. doi:10.1093/cid/ciw475.
53. Martinez JL. Antibiotics and antibiotic resistance genes in natural environments. Science. 2008;321:365–7. doi:10.1126/science.1159483.
54. Rowe WPM, et al. Overexpression of antibiotic resistance genes in hospital effluents over time. J Antimicrob Chemother. 2017;72:1617–23. doi:10.1093/jac/dkx017.
55. Martinez JL. Environmental pollution by antibiotics and by antibiotic resistance determinants. Environ Pollut. 2009;157:2893–902. doi:10.1016/j.envpol.2009.05.051.
56. Eckert EM, Di Cesare A, Coci M, Corno G. Persistence of antibiotic resistance genes in large subalpine lakes: the role of anthropogenic pollution and ecological interactions. Hydrobiologia. 2018;824:93–108. doi:10.1007/s10750-017-3480-0.

57. Berendonk TU, et al. Tackling antibiotic resistance: the environmental framework. Nat Rev Microbiol. 2015;13:310–7. doi:10.1038/nrmicro3439.
58. Liu Z, et al. Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic resistance genes in activated sludge. Environ Int. 2019;129:208–20. doi:10.1016/j.envint.2019.05.036.
59. Liang J, et al. Identification and quantification of bacterial genomes carrying antibiotic resistance genes and virulence factor genes for aquatic microbiological risk assessment. Water Res. 2020;168:115160. doi:10.1016/j.watres.2019.115160.
60. Ricker N, Qian H, Fulthorpe RR. The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. Genomics. 2012;100:167–75. doi:10.1016/j.ygeno.2012.06.009.
61. Chen LX, Anantharaman K, Shaiber A, Eren AM, Banfield JF. Accurate and complete genomes from metagenomes. Genome Res. 2020;30:315–33. doi:10.1101/gr.258640.119.
62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. J. G. r. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. 25, 1043–1055 (2015).
63. Sangwan N, Xia F, Gilbert JA. Recovering complete and draft population genomes from metagenome datasets. Microbiome. 2016;4:8. doi:10.1186/s40168-016-0154-5.
64. Luo C, Tsementzi D, Kyrpides NC, Konstantinidis KT. Individual genome assembly from complex community short-read metagenomic datasets. ISME J. 2012;6:898–901. doi:10.1038/ismej.2011.147.
65. Kumarasamy KK, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis. 2010;10:597–602. doi:10.1016/S1473-3099(10)70143-2.
66. Pitout JD, Laupland KB. Extended-spectrum beta-lactamase-producing Enterobacteriaceae: an emerging public-health concern. Lancet Infect Dis. 2008;8:159–66. doi:10.1016/S1473-3099(08)70041-0.
67. Sabino YNV, et al. Characterization of antibiotic resistance genes in the species of the rumen microbiota. Nat Commun. 2019;10:5252. doi:10.1038/s41467-019-13118-0.
68. Liu JX, et al The fecal resistome of dairy cattle is associated with diet during nursing. *Nature Communications* **10**, 1-15, doi:ARTN 440610.1038/s41467-019-12111-x (2019).
69. Pruden A, Pei R, Storteboom H, Carlson KH. J. E. s. & technology. Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. 2006;40:7445–50.
70. Zhang Y, et al. Intracellular and extracellular antimicrobial resistance genes in the sludge of livestock waste management structures. 47, 10206–10213 (2013).
71. Tamminen M, et al. Tetracycline Resistance Genes Persist at Aquaculture Farms in the Absence of Selection Pressure. Environmental Science Technology. 2011;45:386–91. doi:10.1021/es102725n.

72. Oldfield NJ, Moran AP, Millar LA, Prendergast MM, Ketley JM. Characterization of the *Campylobacter jejuni* Heptosyltransferase II Gene, *waaF*, Provides Genetic Evidence that Extracellular Polysaccharide Is Lipid A Core Independent. 184, 2100–2107, doi:doi:10.1128/JB.184.8.2100-2107.2002 (2002).
73. Ho HTK, Lipman LJA, Gaastra W. Arcobacter, what is known and unknown about a potential foodborne zoonotic agent! Vet Microbiol. 2006;115:1–13. doi:10.1016/j.vetmic.2006.03.004.
74. Vandenberg O, et al. Arcobacter species in humans. Emerg Infect Dis. 2004;10:1863–7. doi:10.3201/eid1010.040241.
75. Lehmann D, et al. Prevalence, virulence gene distribution and genetic diversity of Arcobacter in food samples in Germany. Berl Munch Tierarztl Wochenschr. 2015;128:163–8.
76. Wybo I, Breynaert J, Lauwers S, Lindenburg F, Houf K. Isolation of Arcobacter skirrowii from a patient with chronic diarrhea. J Clin Microbiol. 2004;42:1851–2. doi:10.1128/jcm.42.4.1851-1852.2004.
77. Xiang S, et al. Response of microbial communities of karst river water to antibiotics and microbial source tracking for antibiotics. Sci Total Environ. 2020;706:135730. doi:10.1016/j.scitotenv.2019.135730.
78. Millar JA, Raghavan R. Accumulation and expression of multiple antibiotic resistance genes in Arcobacter cryaerophilus that thrives in sewage. PeerJ. 2017;5:e3269. doi:10.7717/peerj.3269.
79. Herrmann M, et al. Predominance of Cand. Patescibacteria in Groundwater Is Caused by Their Preferential Mobilization From Soils and Flourishing Under Oligotrophic Conditions. *Front Microbiol* **10**, 1407, doi:ARTN 140710.3389/fmicb.2019.01407 (2019).
80. Tian RM, et al Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* **8**, 1-15, doi:ARTN 5110.1186/s40168-020-00825-w (2020).
81. Brown CT, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature. 2015;523:208-U173. doi:10.1038/nature14486.
82. Luef B, et al. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat Commun. 2015;6:6372. doi:10.1038/ncomms7372.
83. Proctor CR, et al. Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. ISME J. 2018;12:1344–59. doi:10.1038/s41396-018-0070-8.
84. Ye L, Mei R, Liu WT, Ren H, Zhang XX. Machine learning-aided analyses of thousands of draft genomes reveal specific features of activated sludge processes. Microbiome. 2020;8:16. doi:10.1186/s40168-020-0794-3.
85. Morris JJ, Lenski RE, Zinser ER. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *Mbio* 3, e00036-00012, doi:ARTN e00036-1210.1128/mBio.00036-12 (2012).
86. Gillings MR, et al. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. Isme Journal. 2015;9:1269–79. doi:10.1038/ismej.2014.226.
87. Amos GCA, et al. The widespread dissemination of integrons throughout bacterial communities in a riverine system. ISME J. 2018;12:681–91. doi:10.1038/s41396-017-0030-8.

88. Moon K, et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome. 2020;8:75. doi:10.1186/s40168-020-00863-4.
89. Kaufman J, et al Integrative and Conjugative Elements (ICE) and Associated Cargo Genes within and across Hundreds of Bacterial Genera. *bioRxiv* (2020).
90. Sarekoski AK. The Use of epicPCR to study Antimicrobial Resistance Genes and Their Bacterial Hosts in Human Impacted Environments in West Africa. (2020).
91. Antunes NT, et al. Class D β-lactamases: are they all carbapenemases? Antimicrob Agents Ch. 2014;58:2119–25.

Figures

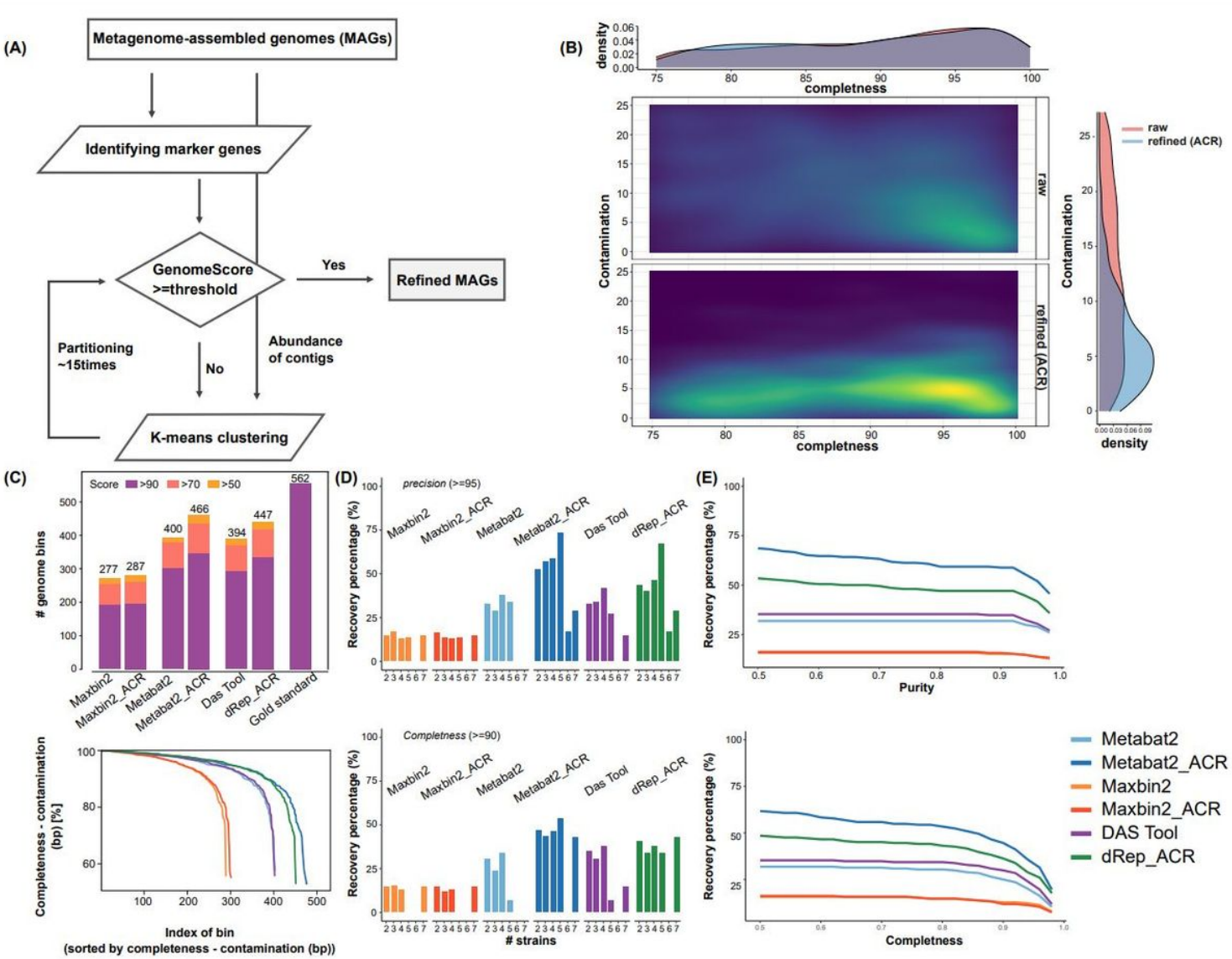

Figure 1

Novel MAG refinement approach using k-means clustering. (A) The overall workflow of ACR. (B) Completeness/Contamination density plots of ACR and MaxBin2 results of metagenomic data used in

this study. Yellow in the plot indicates high density and blue indicates low density. In the side charts at the top and right of the plot, blue is ACR-refined MAG and red is raw MAG. (C) The number of medium-quality MAGs recovered from genome binners and refiners (MaxBin2, MetaBat2, DAS Tool, and ACR) using the gold-standard assembly of CAMI's high-complexity data. "Completeness-contamination" scores were only displayed at 50 or higher. (D, E) Binning results among MAGs within the intra-species level (≥ 95% ANI).

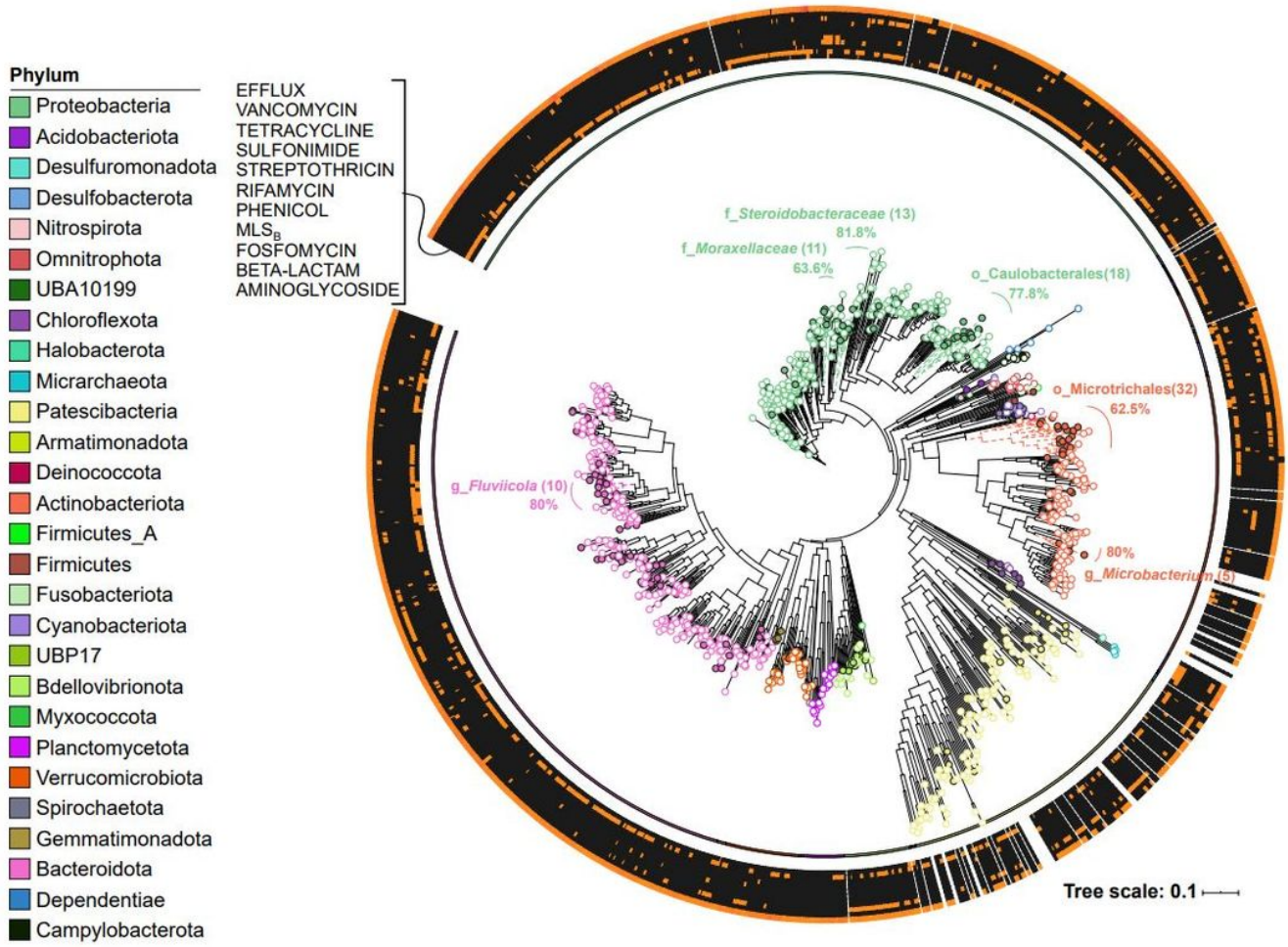

Figure 2

Metagenome-assembled genomes (MAGs) obtained in this study. Phylogenetic tree of MAGs. A total of 1,100 MAGs (1095 bacterial and 5 archaeal) were obtained. Each row of the heatmap outside the tree indicates the number of ARGs conferring resistance to each antibiotic listed in each MAG. The color filled circles indicate multidrug resistance (MDR) bacteria with genes that confer resistance to more than 3 classes of drugs. From phylum to genus, each taxa containing ≥ 50% of MDR bacteria is shown in bold (p: phylum; c: class; o: order; f: family; g: genus). The total number of MAGs in each clade is indicated in the parentheses.

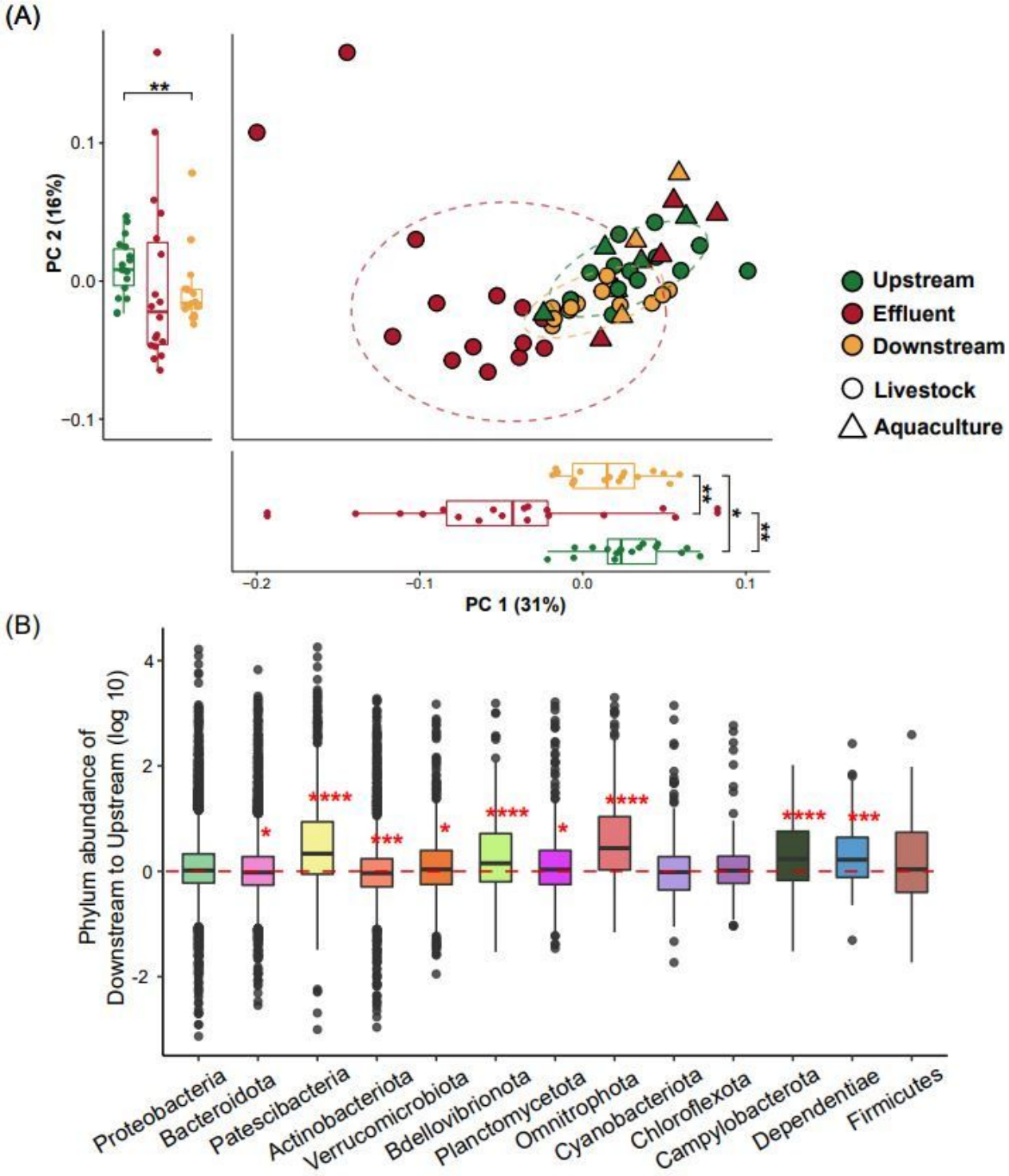

Figure 3

Impact of effluent microbiome on the downstream region of the receiving river. (A) The principal coordination analysis (PCoA) of MAG relative abundance using unweighted UniFrac distance matrix. Each circle and triangle represent livestock and aquaculture, and the colors indicate the sampling sites for the strategy. Boxplots on left and below indicate the sample scores in PC1 and PC2 (*, $P < 0.05$; **, $P < 0.01$, determined by paired Wilcoxon rank sum test). (B) For each phylum, the ratio of the abundance of

MAGs in the upstream section of the river to that in the downstream section of the river was visualized in a box plot. The ratio was in log scale, and red dashed line indicated zero (*, adjust $P < 0.05$; **, adjust $P < 0.01$; ***, adjust $P < 0.001$; ****, $P < 0.0001$, determined by paired Wilcoxon rank sum test).

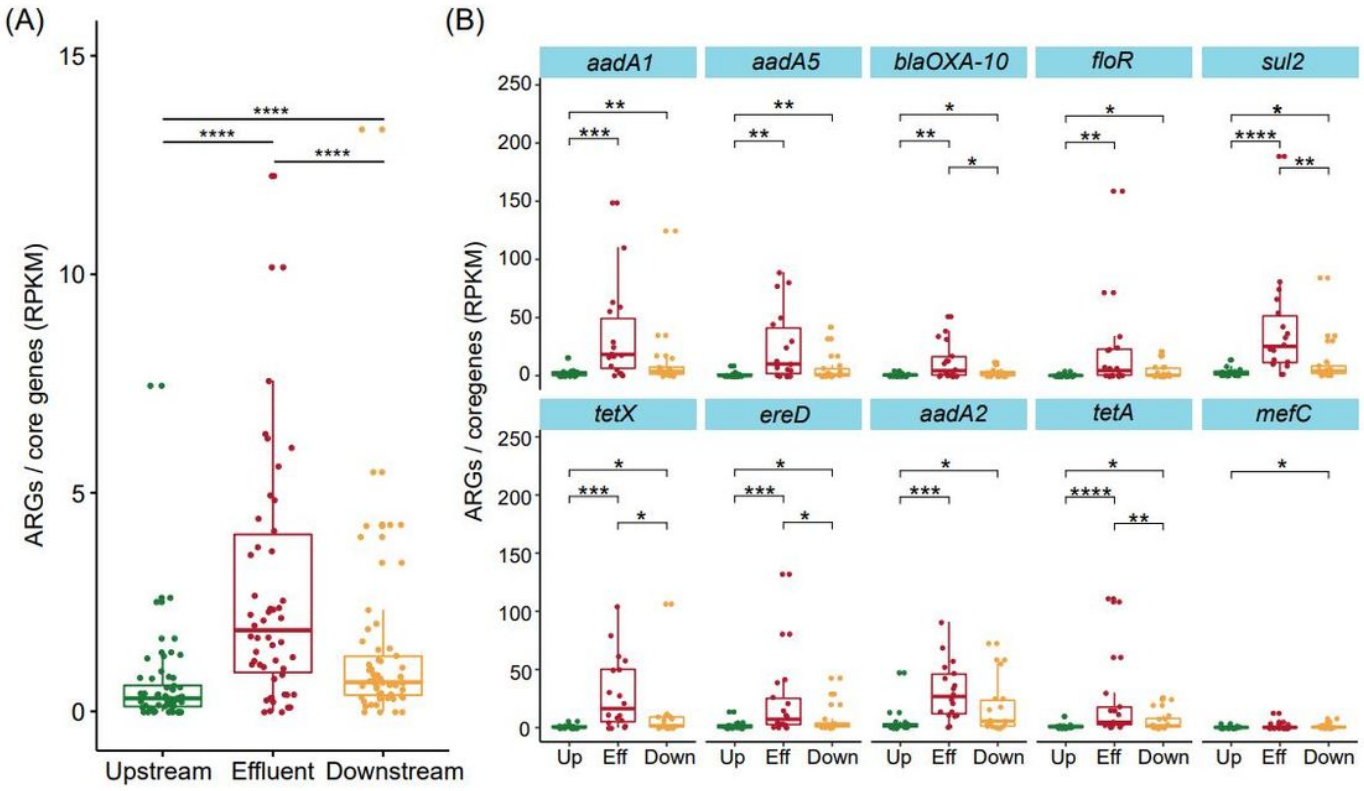

Figure 4

Abundance and distribution of ARGs in MAGs. (A) Ratio of ARGs abundance to core genes abundance. The abundance of ARGs was normalized with core genes abundance in each MAG with unit of copy per million (****, $P < 0.0001$, determined by paired Wilcoxon rank sum test). (B) Abundance of significantly dispersed ARGs. Statistical test used was paired Wilcoxon rank sum test ($P < 0.05$).

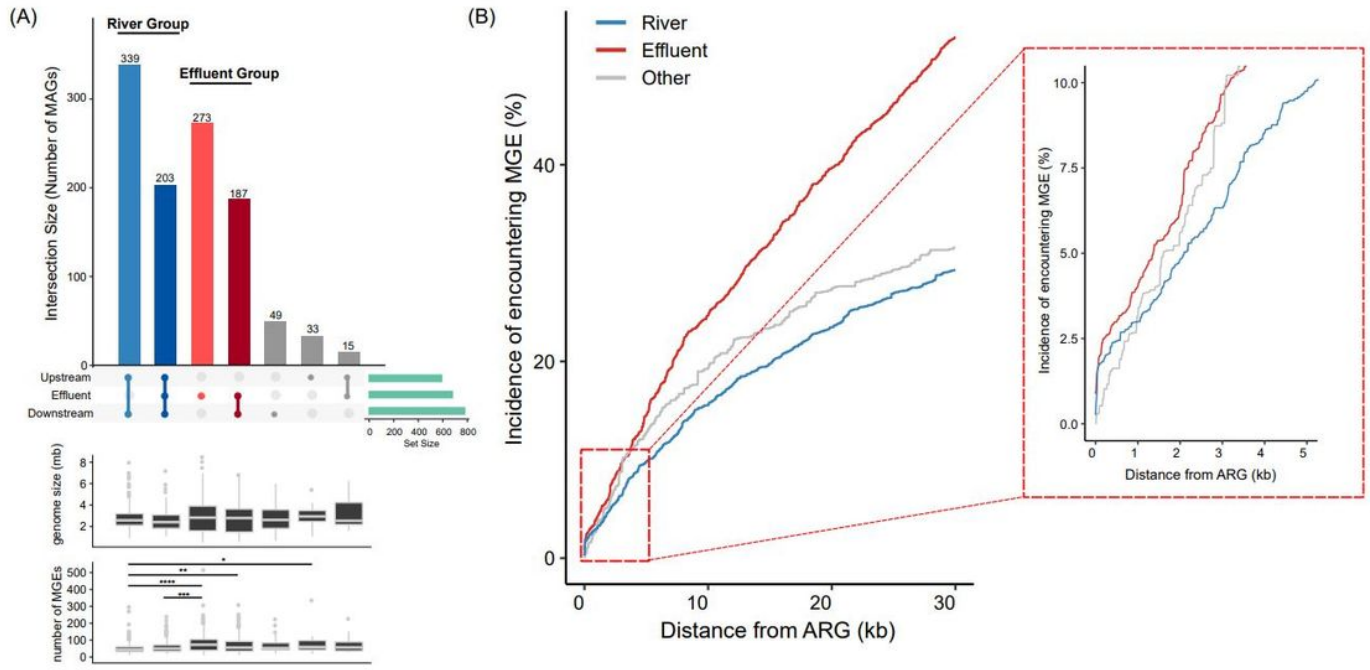

Figure 5

MAGs disseminated from effluent into the downstream region of the river. (A) Distribution of MAGs in each sampling sites. The blue, red, and grey colors indicate River Group (RG), Effluent Group (EG), and others, respectively. The labels indicate the sampling sites where MAGs were found, and the green bar graph shows the number of MAGs found in locations for each section of the river (Up, upstream; Down, downstream; Eff, effluent). Box plots below represented the genome size of MAGs and the number of MGEs, respectively (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$, paired Wilcoxon rank sum test). (B) The incidence of ARGs encountering MGEs. For all distances (10 bp intervals) of the ARG from the MGE > 1100 bp, the HGT potential of Effluent Group significantly increased than of River Group ($P \leq 0.05$, determined by Fisher's exact test).

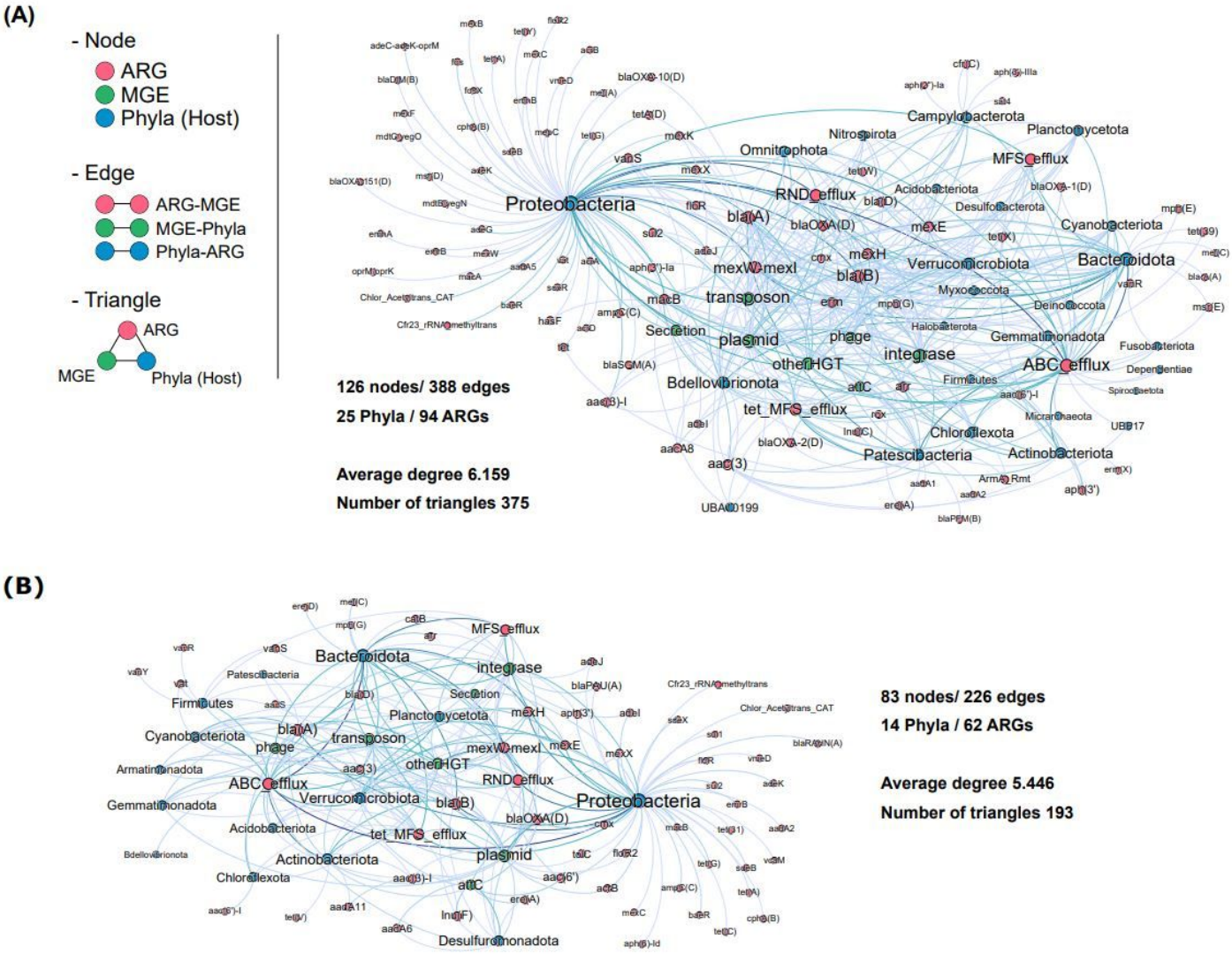

Figure 6

The host-ARG-MGE network. Networks of the relationship between ARG and MGE in each phylum-level MAGs of Effluent Group (A) and River Group (B) groups. Blue, red, and green circles indicate the phyla of host, ARG, and MGE respectively. Node size and edge thickness represent the degree of each node and the number of connections (MAGs) between nodes, respectively.

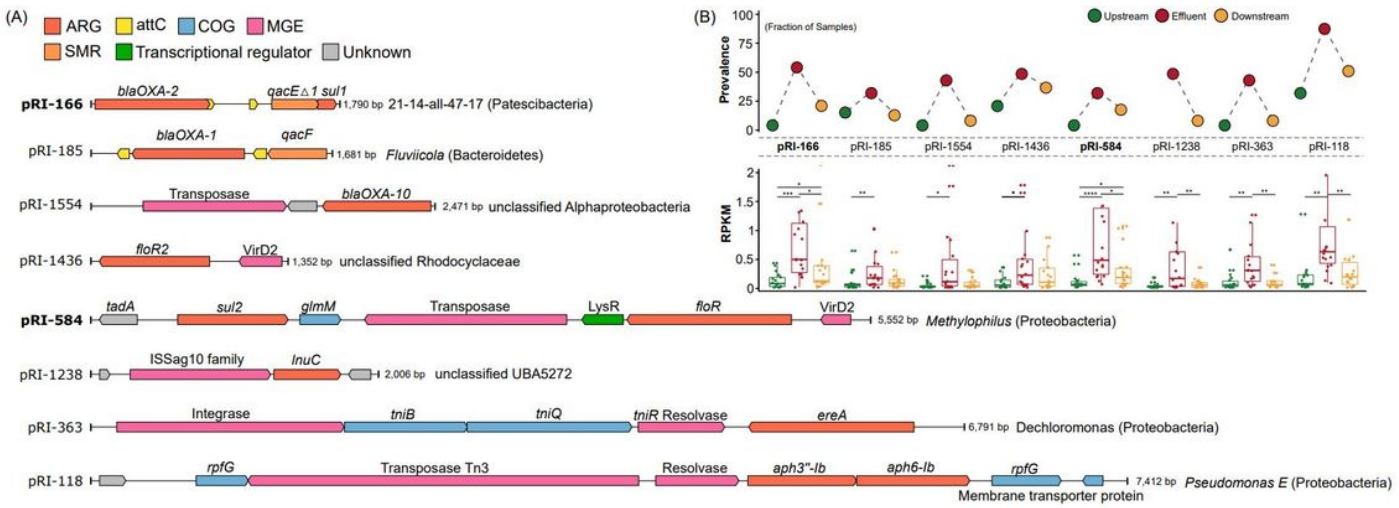

Figure 7

The prevalence of putative resistance islands (pRIs). (A) Genetic contexts of prevalent representative pRIs in effluent samples. Bacterial hosts of pRIs are given on the right. pRI-166 and pRI-185 carried Clusters of *attC* sites lacking integron-integrase mediated resistance genes, and pRI-1554, pRI-1436, pRI-584, pRI-1238, and pRI-363 are carried plasmid mediated resistance genes. COG, clusters of orthologous group; SMR, small multidrug resistance, (B) Prevalence and abundance of pRIs for 18 upstream, effluent, and downstream sites. The presence of pRIs was defined in a sample if reads mapped to at least 70% of the pRI region. pRIs in bold font and with asterisks indicate significant p-value between sample groups (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$, determined by paired Wilcoxon rank sum test, bold font adjust $P < 0.05$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf
- AdditionalFile2.xlsx
- Additionalfile3.xlsx
- SupplementaryFigure116.pdf