HT-ARGfinder: A Comprehensive Pipeline for Identifying Horizontally Transferred Antibiotic Resistance Genes and Directionality in Metagenomic Sequencing Data

Badhan Das[1], Muhit Islam Emon[1], Nazifa Ahmed Moumi[1], Justin Sein[2], Amy Pruden[3], Lenwood S. Heath[1] and Liqing Zhang[1]*

[1]Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, [2]Interdisciplinary PhD Program in Genetics, Bioinformatics and Computational Biology, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, [3]Department of Civil and Environment Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Antibiotic resistance is a continually rising threat to global health. A primary driver of the evolution of new strains of resistant pathogens is the horizontal gene transfer (HGT) of antibiotic resistance genes (ARGs). However, identifying and quantifying ARGs subject to HGT remains a significant challenge. Here, we introduce HT-ARGfinder (horizontally transferred ARG finder), a pipeline that detects and enumerates horizontally transferred ARGs in metagenomic data while also estimating the directionality of transfer. To demonstrate the pipeline, we applied it to an array of publicly-available wastewater metagenomes, including hospital sewage. We compare the horizontally transferred ARGs detected across various sample types and estimate their directionality of transfer among donors and recipients. This study introduces a comprehensive tool to track mobile ARGs in wastewater and other aquatic environments.

Keywords: antibiotic resistance gene, horizontal gene transfer, metagenomics, urban conventional activated sludge, hospital sewage, urban sewage

1 INTRODUCTION

The ability of pathogens to resist antibiotic treatments significantly increases both morbidity and mortality (Bush et al., 2011). As hosts and pathogens have coevolved over millions of years, bacterial pathogens have adjusted their virulence to adapt to host defensive mechanisms (Beceiro et al., 2013). Antimicrobial-resistant infections kill at least 700,000 people worldwide each year, and, within 30 years, resistant infections are expected to kill 10 million people annually, far outnumbering cancer mortality (O'neill, 2014). There is growing recognition that problematic pathogens encountered in the clinic originally acquired their ability to resist antibiotics from the broader pool of bacteria inhabiting natural environments. Once a resistant pathogen becomes endemic in a hospital, it can create a source of nosocomial infection and elevate the overall risk of standard medical procedures (Davies and Davies, 2010; Cheong et al., 2017).

Horizontal gene transfer (HGT) is a natural process that contributes to the evolution of resistant bacterial strains through the physical passage of DNA from one cell to another (Khan and Rao, 2019). HGT occurs through three fundamental mechanisms: transformation (uptake of naked DNA by

competent cells), transduction (uptake of foreign DNA mediated by a phage), and conjugation (bacterial mating) (Amarasiri et al., 2020). This transfer of genetic material can cause both beneficial as well as adverse consequences from a human standpoint (Kunhikannan et al., 2021). HGT is commonly harnessed for industrial biotechnological purposes and generally serves to expand genetic diversity (Le Roux and Blokesch, 2018). However, HGT is also primarily responsible for mobilizing multidrug resistance (MDR) among strains. In particular, multi-antibiotic resistant superbugs, which can be resistant to all available antibiotic therapies, have arisen mainly as a result of HGT (Mathers et al., 2015; Wang and Sun, 2015; Malhotra-Kumar et al., 2016). The transfer of plasmids carrying multiple ARGs is one primary driver of the superbug phenomenon (Sun et al., 2019).

A growing body of research has brought to light the importance of aquatic environments as a hub for the dissemination of ARGs. Aquatic environments typically receive multiple waste streams, treated and untreated, containing various mixtures of contaminants. As such, aquatic environments have been identified as ideal settings for HGT of ARGs, while human exposure to antibiotic resistance bacteria, and ARGs in aquatic environments may pose an additional health risk (Amarasiri et al., 2020). The evolution and spread of ARGs in aquatic environments have been documented across several studies conducted in various locales across the globe. Hundreds of different ARGs have been discovered in bacteria found not just in hospitals, livestock, and meatpacking wastewater but also in sewage, effluent treatment facilities, surface water, groundwater, and even drinking water (Zhang et al., 2009).

Convenient, efficient, and accurate means of identifying ARGs subject to HGT would be highly valuable to the research community and to fledgling efforts to advance environmental monitoring of antibiotic resistance. This would allow the ability to focus on ARGs that are of most concern in terms of ability to mobilize to pathogens instead of intrinsic, non-mobile counterparts. Advances in shotgun metagenomic sequencing of environmental samples holds substantial promise for this purpose. The ability to recover millions of DNA sequences from a given sample affords the opportunity to correspondingly search for ARGs and mobile genetic elements (MGEs), including plasmids, transposons, and integrons. Recently, MetaCHIP was introduced as a tool to specifically identify genes subject to HGT across a microbial community (Song et al., 2019). MetaCHIP identifies HGT in assembled metagenomic data through a combination of best hit and phylogenetic approaches, which further serves to inform estimates of the directionality of the horizontally transferred genes (i.e., the donor and recipient of the HGT). LEMON (Li et al., 2019) is a similar tool that was also recently introduced and, like MetaCHIP, works best with nearly complete metagenome-assembled genomes. However, these existing tools are not explicitly tailored to identify and track the directionality of horizontally transferred ARGs.

We propose a new pipeline, HT-ARGfinder, which detects horizontally transferred ARGs from metagenomic data derived from complex environmental microbial communities. The pipeline was applied to metagenomes obtained from a range of wastewater samples, representing presumed hotspots for ARG evolution and mobilization. Two ARG databases were applied
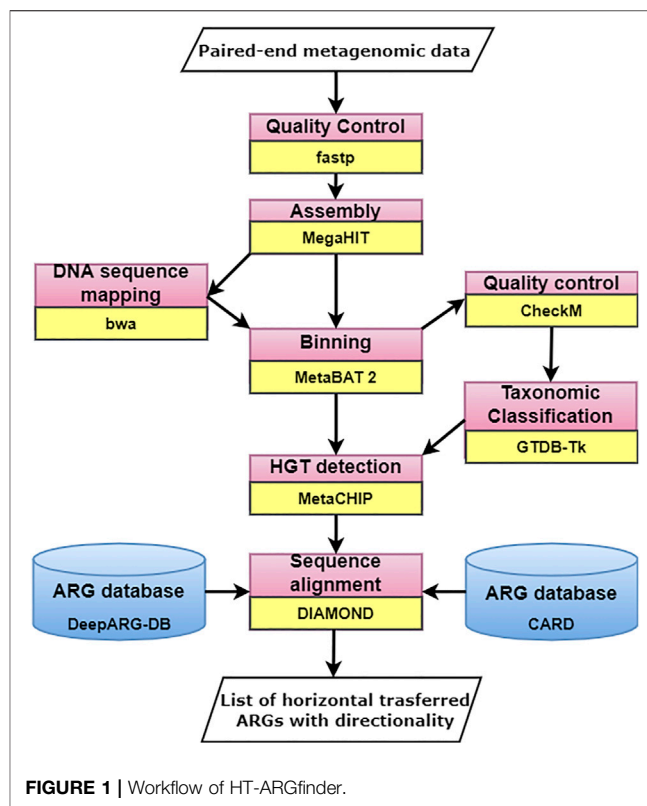

FIGURE 1 | Workflow of HT-ARGfinder.

and compared: CARD, the most well-curated and up-to-date database of known ARGs, and DeepARG-DB, which comprises 14,933 genes including the three ARG databases: CARD, ARDB, and UNIPROT, to include putative new ARGs discovered based on deep learning Arango-Argoty et al. (2018). The specific objectives of this study were to: [1]) develop HT-ARGfinder as a comprehensive and user-friendly pipeline for identifying horizontally transferred ARGs in metagenomic data sets and inferring directionality of their transfer, 2) apply the pipeline for identifying and comparing horizontally transferred ARGs detected across a range of representative wastewater environments, and 3) infer the directionality of ARG transfer among various bacterial classes in several metagenomic data sets.

2 MATERIALS AND METHODS

2.1 Workflow

HT-ARGfinder combines a series of bioinformatics tools to process metagenomic data, starting with quality control and ending with ARG detection and directionality inference. The workflow of HT-ARGfinder is shown in **Figure 1**.

---

[1]Environments were re-classified based on information reported by the authors for the consistency across this study

TABLE 1 | Summary of data sets used to demonstrate HT-ARGfinder.

Environment[1]	Accession no. (NCBI SRA)	Classification according to original study
Hospital sewage	ERX3538875, ERX3538874, ERX3538873, ERX3538872, ERX3538871, ERX3538870, ERX3538869, ERX3538868	Hospital wastewater
	SRX7901963, SRX7901962, SRX7901961, SRX7901960, SRX7901959, SRX7901958	Hospital sewage
Urban conventional activated sludge	SRX3720490, SRX3720489, SRX3720488, SRX3720487, SRX3720486, SRX3720481, SRX3720480, SRX3720479	Activated sludge
Urban sewage	ERX1795927, ERX1783568, ERX1795930, ERX1783571, ERX1795935, ERX1783576, ERX2608526, ERX2608605, ERX1783583, ERX1783584	Urban sewage

TABLE 2 | Total number of horizontally transferred genes (HGTs), ARGs according to CARD database, and ARGs according to DeepARG-DB.

Environment	Data set	HGTs	ARGs (CARD)	ARGs (DeepARG-DB)
Hospital sewage	ERX3538875	98	6	7
	ERX3538874	61	2	3
	ERX3538873	264	21	20
	ERX3538872	4	1	1
	ERX3538871	9	0	0
	ERX3538870	4	0	0
	ERX3538869	20	2	3
	ERX3538868	9	3	3
	SRX7901963	30	2	3
	SRX7901962	14	3	3
	SRX7901961	31	2	2
	SRX7901960	9	0	0
	SRX7901959	0	—	—
	SRX7901958	3	0	0
Urban conventional activated sludge	SRX3720490	20	1	1
	SRX3720489	11	1	1
	SRX3720488	19	3	3
	SRX3720487	44	2	2
	SRX3720486	8	2	2
	SRX3720481	14	2	2
	SRX3720480	11	1	2
	SRX3720479	24	0	1
Urban sewage	ERX1795927	12	3	3
	ERX1783568	8	0	0
	ERX1795930	1	1	1
	ERX1783571	4	1	1
	ERX1795935	11	1	1
	ERX1783576	18	1	1
	ERX2608526	11	0	0
	ERX2608605	11	0	0
	ERX1783583	12	1	1
	ERX1783584	9	0	0

2.1.1 Quality Control of Reads

Fastp (Chen et al., 2018), an ultra-fast FASTQ preprocessor with quality control, was incorporated for quality control and preprocessing of the paired-end reads. This tool is developed in C++ and has multi-threading support. Fastp is 2–5 times faster than other FASTQ preprocessing tools such as Trimmomatic (Bolger et al., 2014), Cutadapt (Martin, 2011), FASTQC (Brown et al., 2017), and AfterQC (Chen et al., 2017).

2.1.2 Metagenomic Assembly

MegaHIT (Li et al., 2015) was used for assembling metagenomes, which we previously found to provide the most accurate and extensive assemblies of short-read metagenomic data from complex environmental samples. MegaHIT is a *de novo* assembler that uses a succinct de Bruijn graph to assemble short reads (Bowe et al., 2012).
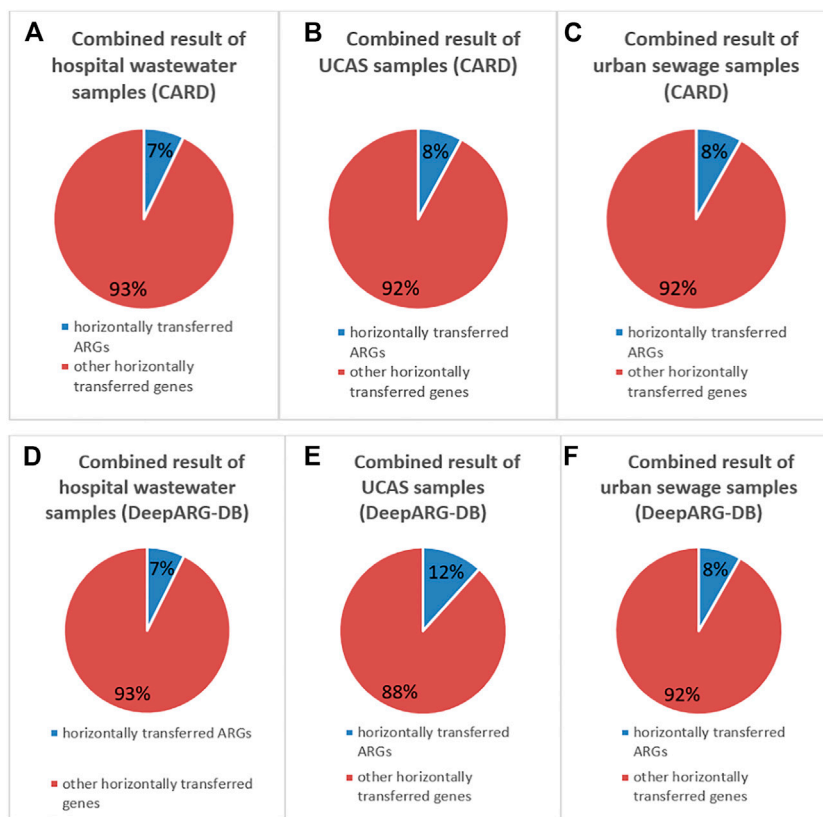
FIGURE 2 | Percentage of horizontally transferred ARGs detected [according to CARD **(A–C)**; according to DeepARG-DB **(D–F)**].

2.1.3 Binning

MetaBAT 2 (Kang et al., 2019) is incorporated for the purpose of reconstructing single genomes from microbial communities. In comparison to other binning tools such as MaxBin2 (Wu et al., 2016), CONCOCT (Alneberg et al., 2013), and MyCC (Lin and Liao, 2016), MetaBAT 2 shows better performance.

2.1.4 Quality Control of Bins

For quality control of the bins, CheckM (v1.1.2) (Parks et al., 2015) was employed. CheckM is an automated method for evaluating the quality of a genome using a broader set of marker genes specific to the position of a genome within a reference genome tree and information about the collocation of these genes. The bins were filtered based on completeness and contamination reported by CheckM. Bins having *completeness* $\geq 40\%$ and *contamination* $\leq 5\%$ were selected for the next step. According to Bowers et al. (2017), "Medium-quality drafts" SAGs and MAGs are those genomes with completeness estimates of $\geq 50\%$ and with contamination of $\leq 10\%$. All the other SAGs and MAGs ($< 50\%$ completeness or $> 10\%$ contamination) should be reported as "low quality drafts." MetaCHIP has chosen $\leq 40\%$ completless and 0% contamination for their pipeline Song et al. (2019). We have used $\leq 5\%$ contamination since its less than the recommended 10% for medium quality drafts but slightly less stringent than MetaCHIP.

2.1.5 Taxonomic Classification

After binning, the genomes are taxonomically classified using GTDB-Tk (v1.6.0), which provides objective taxonomic assignments for bacterial and archaeal genomes based on the Genome Taxonomy Database (GTDB) (Chaumeil et al., 2020).

2.1.6 HGT Detection

MetaCHIP, a pipeline for reference independent HGT identification in metagenomic data (Song et al., 2019), is used for detecting HGT at the community level. MetaCHIP implements a combination of both best match and phylogenetic approaches. First, it uses sequence similarity comparison to identify putative open reading frames (ORFs) horizontally transferred for user-defined taxonomy ranks (e.g., class, order family, or genus). Then phylogenetic trees are constructed for the putative HGT gene candidates and compared to the species trees determined by single-copy genes. Finally, reconciliation of the gene trees and species trees is performed using Ranger-DTL v2.0 (Bansal et al., 2018) to identify HGTs and donors and recipients. MetaCHIP allows users to specify the taxonomy ranks for which HGTs are identified. The default is "class," and HT-ARGfinder keeps the default one.
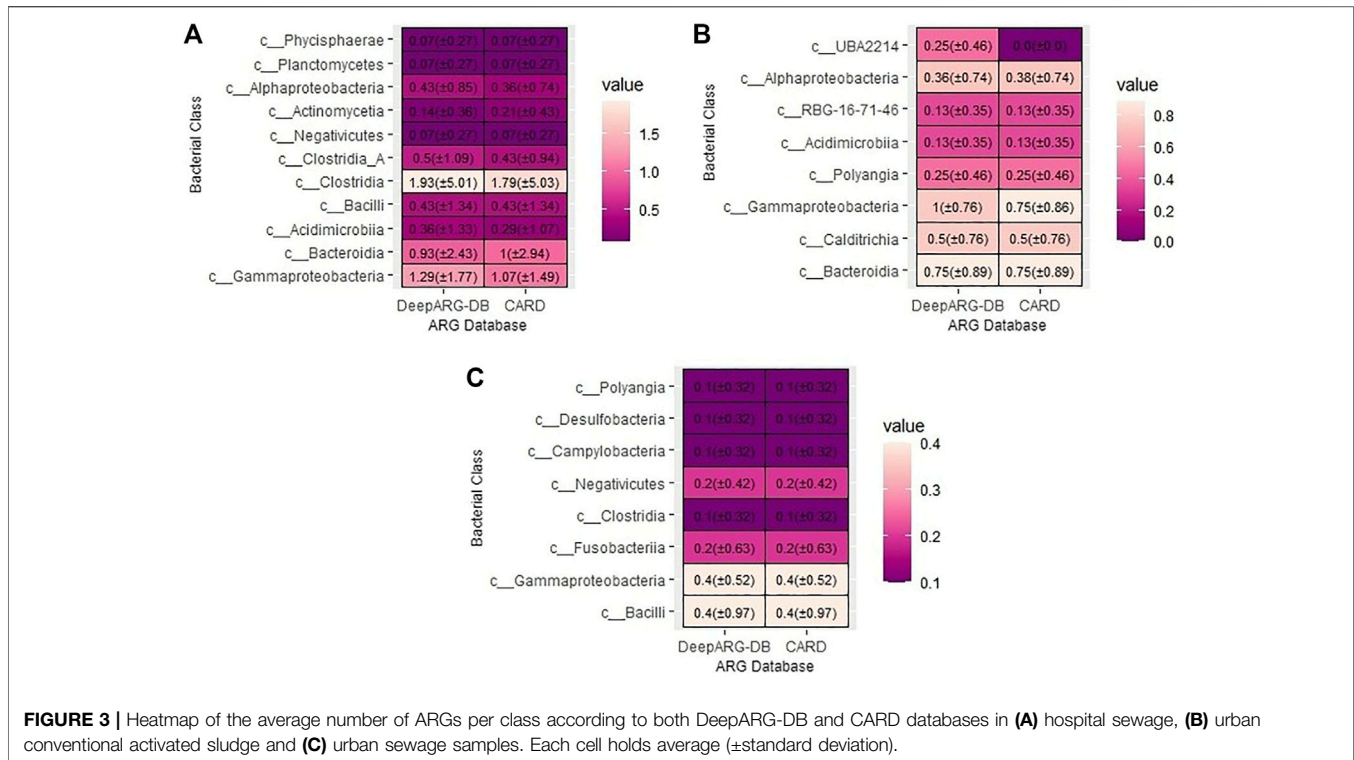
FIGURE 3 | Heatmap of the average number of ARGs per class according to both DeepARG-DB and CARD databases in **(A)** hospital sewage, **(B)** urban conventional activated sludge and **(C)** urban sewage samples. Each cell holds average (±standard deviation).

2.1.7 Databases and Sequence Alignment

The detected horizontally transferred genes are finally aligned against two reference databases, DeepARG-DB (Arango-Argoty et al., 2018) and CARD (Alcock et al., 2020) using DIAMOND (Buchfink et al., 2015) to determine ARGs. CARD represents a highly-curated database of previously described ARGs that is updated periodically. In contrast, DeepARG-DB is a larger database that encompasses the ARGs included in CARD and additional ARGs detected from a deep learning algorithm. There is an option to use new databases in our pipeline. Users can easily execute the command and include new databases to check ARGs against those databases.

2.2 Wastewater Metagenomes Included in This Study

We selected publicly-available metagenomic data from a range of wastewaters for inclusion in this study; see **Table 1**. We were particularly interested in data sets, such as hospital sewage, urban conventional activated sludge, likely to contain a diverse array of ARGs that had been subject to HGT. All data used in this study are open source and are available at NCBI Sequence Read Archive (SRA) under the Bioproject accession numbers reported in **Table 1**. These included hospital sewage, urban conventional activated sludge and urban sewage samples.

Table 1 provides a summary of these samples, including the environments analyzed as named by the original researchers and generalized categories based on common terminology used in this paper (hospital sewage, urban conventional activated sludge, and urban sewage). These datasets included Illumina HiSeq 4000 paired-end sequencing

metagenomic data, MiSeq paired-end sequencing metagenomic data, Illumina HiSeq 2500, and Illumina HiSeq 3000 paired-end metagenomic data.

2.3 Demonstrating HT-ARGfinder

We demonstrated HT-ARGfinder using the data summarized in **Table 1**. The HT-ARGfinder output consisted of two lists of horizontally-transferred ARGs according to CARD and DeepARG-DB databases and their corresponding estimated directionalities. The number of detected ARGs from these two databases was compared. We also analyzed the abundances of different bacterial classes associated with the HGT events in the samples of different environments according to CARD and DeepARG-DB. MetaCHIP was applied to identify the bacterial classes corresponding to the detected horizontally-transferred genes. We filtered out the ARGs detected using both CARD and DeepARG-DB from that list. Finally, we mapped the output lists of HT-ARGfinder to directed graphs where the nodes represent different bacterial classes involved in HGT, and the directed edges represent the direction of the HGT. This graph provides visualization of the whole HGT scenario of a metagenomic sample.

3 RESULTS

3.1 ARG Detection

HT-ARGfinder was found to be a robust tool for the detection of horizontally transferred ARGs in complex environmental metagenomic data sets. **Table 2** presents the total number of
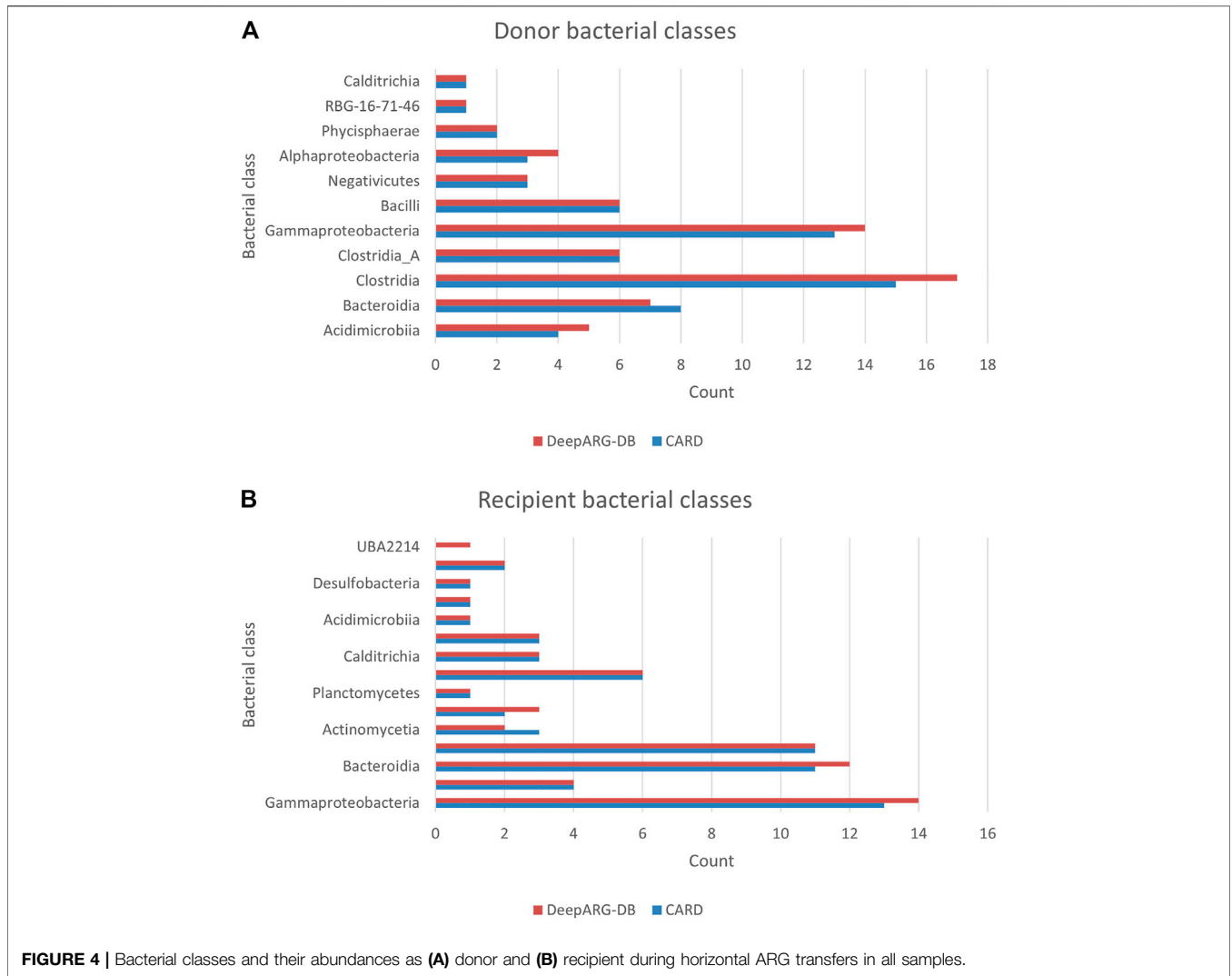
FIGURE 4 | Bacterial classes and their abundances as **(A)** donor and **(B)** recipient during horizontal ARG transfers in all samples.
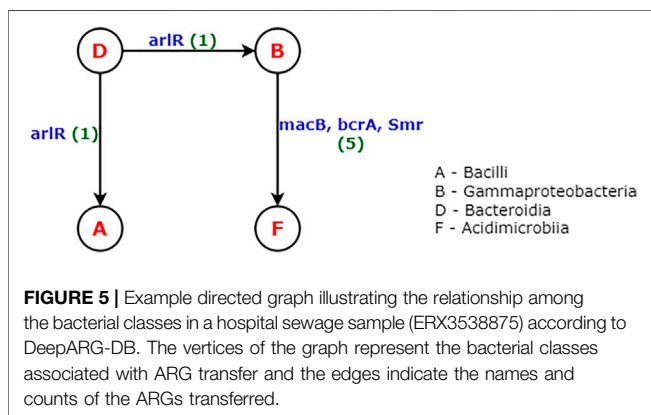

FIGURE 5 | Example directed graph illustrating the relationship among the bacterial classes in a hospital sewage sample (ERX3538875) according to DeepARG-DB. The vertices of the graph represent the bacterial classes associated with ARG transfer and the edges indicate the names and counts of the ARGs transferred.

horizontally transferred genes and the total number of ARGs according to CARD and DeepARG-DB databases. The table shows that raw hospital wastewater and urban conventional activated sludge have many HGT events. Still, fewer horizontally transferred genes were detected in the urban

wastewater treatment plant samples. Based on these trends, raw hospital wastewater, and urban conventional activated sludge samples focused on the remainder of the HT-ARGfinder demonstration.

As can be observed from **Table 2**, ARGs represent a remarkably consistent portion of the total horizontally transferred genes detected (*10% estimated for both types of environments*). As expected, DeepARG-DB tended to detect a greater number of horizontally transferred ARGs than CARD. Still, the percent of the horizontally transferred ARGs were similar (**Figure 2**). This suggests that the relative proportion of mobile ARGs does not vary substantially across the environments tested, although the overall transfer rates may vary.

3.2 Identification of Donor and Recipient Bacterial Classes

MetaCHIP provides a prediction of the donor and recipient bacteria involved in each HGT event. HT-ARGfinder

improves upon this feature by filtering events related explicitly to ARG transfers. This yielded valuable information concerning the classes of bacteria actively engaged in the horizontal transfer of ARGs. Further, the abundance of these classes and ARGs can be estimated in each environment. Heatmaps were generated to illustrate the average number of bacterial classes estimated to be engaged in ARG transfer across the samples analyzed in this study (**Figure 3**).

The hospital sewage samples were found to be enriched with Actinomycetia, Negativicutes, Clostridia_A, Clostridia, Bacilli, Acidimicrobiia, Bacteroidia, Gammaproteobacteria, Phycisphaerae, Planctomycetes, and Alphaproteobacteria. **Figure 3** shows their average abundance (**Supplementary Tables** show the number of detected ARG classes in each sample). In urban conventional activated sludge samples, we found UBA2214, Alphaproteobacteria, RBG-16-71-46, Acidimicrobiia, Polyangia, Gammaproteobacteria, Calditrichia, and Bacteroidia (**Figure 3**). In urban sewage samples, Bacilli, Gammaproteobacteria, Fusobacteriia, Clostridia, Negativicutes, Campylobacteria, Desulfobacteria, and Polyangia were found (**Figure 3**). **Figure 4** shows all the bacterial classes and their abundances as donor and recipient in these metagenomic samples according to both DeepARG-DB and CARD databases. We can see that Clostridia and Gammaproteobacteria were higher in counts as the donor, whereas Gammaproteobacteria, Bacteroidea, and Clostridia were higher in counts as the recipient.

3.3 Comparison Between CARD and DeepARG-DB

Implementation of CARD versus DeepARG-DB in the pipeline resulted in different numbers of horizontally transferred ARGs detected in some samples (**Table 1**; **Figures 2–4**). For the cases in which the discrepancy was the greatest, DeepARG-DB has consistently yielded the higher count. This is consistent with DeepARG-DB being a larger database than CARD, encompassing the ARGs included in CARD and additional ARGs identified via the deep learning module.

3.4 Relationship Among Bacterial Classes in Samples

HT-ARGfinder provides additional value in inferring the directionality of ARG transfer among bacterial classes. This can be visualized in directed graphs that illustrate the various transfer relationships among the bacterial classes represented in a sample. For example, for a metagenomic sample, let $G = (V, E)$ be a directed graph where each vertex represents a bacterial class, each directed edge between two vertices indicates the directionality of the ARG transfer between the two bacterial classes, and the edge weight indicates the number of ARGs transferred. A hospital sewage sample (ERX3538875) was used to visualize the relationship among the bacterial classes (**Figure 5**). Four bacterial classes involved in HGT of ARGs were identified in this sample: Bacilli,

Gammaproteobacteria, Bacteroidia, and Acidimicrobiia. The possible ARGs that are transferred between Gammaproteobacteria and Alphaproteobacteria encoded resistance to macrolides (macB), peptide (bcRA), and polymyxins (arnA).

4 DISCUSSION AND CONCLUSION

HT-ARGfinder provides a comprehensive and user-friendly tool specifically directed toward identifying mobile ARGs in metagenomic data sets and estimating the directionality of their transfer among bacterial classes. This pipeline can provide substantial value in targeting environmental monitoring efforts towards tracking mobile ARGs and help inform better and direct efforts to stop the spread of antibiotic resistance. Here we demonstrated the tool's utility for a spectrum of wastewater samples. We have used CARD and DeepARG-DB as the ARG databases, but this pipeline can be modified easily by replacing these databases with some recent ones, for example, HMD-ARG-DB (Li et al., 2021). It is composed of 17, 282 sequences, gathered and cleaned from 7 published ARG databases: CARD, AMRFinder (Feldgarden et al., 2019), ResFinder (Zankari et al., 2012), ARG-ANNOT (Gupta et al., 2014), DeepARG, MEGARes (Doster et al., 2020) and Resfams (Gibson et al., 2015). We have shown results using CARD and DeepARG-DB because these are well-known ones. Users can modify the pipeline by adding other databases easily.

Our tool detects all the ARGs that are transferred horizontally. These ARGs are horizontally transferred through the mobile genetic elements such as plasmids, transposons, integrons, and genomic islands. (Khezri et al., 2020) detects only the plasmid-mediated ARGs by first identifying plasmids. Since our tool detects all the ARGs transferred through HGT, it detects ARGs mediated through potential mobile genetic elements.

The analysis was consistent with the expectation of high abundances of mobile ARGs in hospital wastewater and urban conventional activated sludge samples. While the focus was on wastewater samples, the pipeline can readily be applied to various other complex metagenomes, including different aquatic and wastewater samples and gastrointestinal microbiota, broadly of interest for ARG surveillance and research. Based on this study, it was apparent that ARGs represent a substantial fraction (~ 10%) of genes actively transferred in a given wastewater microbiome. It would be interesting to examine further the remaining genes subject to HGT, which might include a novel, yet to be discovered ARGs. This is apparent in recognizing that many horizontally transferred ARGs were found when the pipeline queried DeepARG-DB. Additionally, such genes could encode other relevant functions, such as metal resistance or processes related to gene mobility. Further analysis, such as this, of the context of the ARGs, can help resolve the driver of their mobility. The directionality analysis is further of significant value in tracking the movement of ARGs across bacterial hosts. Such information could provide value in identifying which bacteria

should be targeted in mitigation efforts aimed at stemming the spread of antimicrobial resistance.

DATA AVAILABILITY STATEMENT

The source code of HTARGfinder can be found at https://github.com/Badhan023/HTARGfinder. All the data sets that have been used are publicly available in the NCBI Short Read 235 Archive (SRA).

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2022.901917/full#supplementary-material

REFERENCES

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). Card 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* 48, D517–D525. doi:10.1093/nar/gkz935

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2013). Concoct: Clustering Contigs on Coverage and Composition. arXiv preprint arXiv:1312.4038.

Amarasiri, M., Sano, D., and Suzuki, S. (2020). 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. *Crit. Rev. Environ. Sci. Technol.* 50, 2016–2059. doi:10.1080/10643389.2019.1692611

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., and Zhang, L. (2018). Deeparg: A Deep Learning Approach for Predicting Antibiotic Resistance Genes from Metagenomic Data. *Microbiome* 6, 23. doi:10.1186/s40168-018-0401-z

Bansal, M. S., Kellis, M., Kordi, M., and Kundu, S. (2018). Ranger-dtl 2.0: Rigorous Reconstruction of Gene-Family Evolution by Duplication, Transfer and Loss. *Bioinformatics* 34, 3214–3216. doi:10.1093/bioinformatics/bty314

Beceiro, A., Tomás, M., and Bou, G. (2013). Antimicrobial Resistance and Virulence: A Successful or Deleterious Association in the Bacterial World? *Clin. Microbiol. Rev.* 26, 185–230. doi:10.1128/cmr.00059-12

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Bowe, A., Onodera, T., Sadakane, K., and Shibuya, T. (2012). "Succinct de bruijn graphs," in International Workshop on Algorithms in Bioinformatics (Berlin, Heidelberg: Springer), 225–235. doi:10.1007/978-3-642-33122-0_18

Bowers, R. M., Kyrpides, N. C., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., et al. (2017). Minimum Information about a Single Amplified Genome (misag) and a Metagenome-Assembled Genome (mimag) of Bacteria and Archaea. *Nat. Biotechnol.* 35, 725–731. doi:10.1038/nbt.3893

Brown, J., Pirrung, M., and McCue, L. A. (2017). Fqc Dashboard: Integrates Fastqc Results into a Web-Based, Interactive, and Extensible Fastq Quality Control Tool. *Bioinformatics* 33, 3137–3139. doi:10.1093/bioinformatics/btx373

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and Sensitive Protein Alignment Using Diamond. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176

Bush, K., Courvalin, P., Dantas, G., Davies, J., Eisenstein, B., Huovinen, P., et al. (2011). Tackling Antibiotic Resistance. *Nat. Rev. Microbiol.* 9, 894–896. doi:10.1038/nrmicro2693

[Dataset] Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). Gtdb-tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database. *Bioinformatics* 36 (6), 1925–1927. doi:10.1093/bioinformatics/btz848

Chen, S., Huang, T., Zhou, Y., Han, Y., Xu, M., and Gu, J. (2017). Afterqc: Automatic Filtering, Trimming, Error Removing and Quality Control for Fastq Data. *BMC Bioinforma.* 18, 80. doi:10.1186/s12859-017-1469-3

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

Cheong, E., Gottlieb, T., Crook, D., Kizny Gordon, A., Kotay, S., Mathers, A., et al. (2017). The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections-A Systematic Review of the Literature. *Clin. Infect. Dis.* 64 (10), 1435–1444. doi:10.1093/cid/cix132

Davies, J., and Davies, D. (2010). Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433. doi:10.1128/mmbr.00016-10

Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., et al. (2020). Megares 2.0: A Database for Classification of Antimicrobial Drug, Biocide and Metal Resistance Determinants in Metagenomic Sequence Data. *Nucleic Acids Res.* 48, D561–D569. doi:10.1093/nar/gkz1010

Feldgarden, M., Brover, V., Haft, D. H., Prasad, A. B., Slotta, D. J., Tolstoy, I., et al. (2019). Validating the Amrfinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob. Agents Chemother.* 63, e00483–19. doi:10.1128/AAC.00483-19

Gibson, M. K., Forsberg, K. J., and Dantas, G. (2015). Improved Annotation of Antibiotic Resistance Determinants Reveals Microbial Resistomes Cluster by Ecology. *ISME J.* 9, 207–216. doi:10.1038/ismej.2014.106

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., et al. (2014). Arg-Annot, a New Bioinformatic Tool to Discover Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrob. Agents Chemother.* 58, 212–220. doi:10.1128/aac.01310-13

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). Metabat 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies. *PeerJ* 7, e7359. doi:10.7717/peerj.7359

[Dataset] Khan, A., and Rao, T. (2019). "Molecular Evolution of Xenobiotic Degrading Genes and Mobile Dna Elements in Soil Bacteria," in *Microbial Diversity in the Genomic Era* (Academic Press). doi:10.1016/B978-0-12-814849-5.00036-8

Khezri, A., Avershina, E., and Ahmad, R. (2020). Plasmid Identification and Plasmid-Mediated Antimicrobial Gene Detection in Norwegian Isolates. *Microorganisms* 9, 52. doi:10.3390/microorganisms9010052

Kunhikannan, S., Thomas, C. J., Franks, A. E., Mahadevaiah, S., Kumar, S., and Petrovski, S. (2021). Environmental Hotspots for Antibiotic Resistance Genes. *MicrobiologyOpen* 10, e1197. doi:10.1002/mbo3.1197

Le Roux, F., and Blokesch, M. (2018). Eco-Evolutionary Dynamics Linked to Horizontal Gene Transfer in Vibrios. *Annu. Rev. Microbiol.* 72, 89–110. doi:10.1146/annurev-micro-090817-062148

Li, C., Jiang, Y., and Li, S. (2019). Lemon: A Method to Construct the Local Strains at Horizontal Gene Transfer Sites in Gut Metagenomics. *BMC Bioinforma.* 20, 702–712. doi:10.1186/s12859-019-3301-8

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). Megahit: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. *Bioinformatics* 31, 1674–1676. doi:10.1093/bioinformatics/btv033

Li, Y., Xu, Z., Han, W., Cao, H., Umarov, R., Yan, A., et al. (2021). Hmd-arg: Hierarchical Multi-Task Deep Learning for Annotating Antibiotic Resistance Genes. *Microbiome* 9, 1–12. doi:10.1186/s40168-021-01002-3

Lin, H. H., and Liao, Y. C. (2016). Accurate Binning of Metagenomic Contigs via Automated Clustering Sequences Using Information of Genomic Signatures and Marker Genes. *Sci. Rep.* 6, 24175. doi:10.1038/srep24175

Malhotra-Kumar, S., Xavier, B. B., Das, A. J., Lammens, C., Butaye, P., and Goossens, H. (2016). Colistin Resistance Gene Mcr-1 Harboured on a Multidrug Resistant Plasmid. *Lancet Infect. Dis.* 16, 283–284. doi:10.1016/s1473-3099(16)00012-8

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet J.* 17, 10–12. doi:10.14806/ej.17.1.200

Mathers, A. J., Peirano, G., and Pitout, J. D. D. (2015). The Role of Epidemic Resistance Plasmids and International High-Risk Clones in the Spread of Multidrug-Resistant Enterobacteriaceae. *Clin. Microbiol. Rev.* 28, 565–591. doi:10.1128/cmr.00116-14

O'neill, J. (2014). "Antimicrobial Resistance," in *Tackling a Crisis for the Health and Wealth of Nations*.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). Checkm: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res.* 25, 1043–1055. doi:10.1101/gr.186072.114

Song, W., Wemheuer, B., Zhang, S., Steensen, K., and Thomas, T. (2019). Metachip: Community-Level Horizontal Gene Transfer Identification through the Combination of Best-Match and Phylogenetic Approaches. *Microbiome* 7, 36. doi:10.1186/s40168-019-0649-y

Sun, D., Jeannot, K., Xiao, Y., and Knapp, C. W. (2019). Horizontal Gene Transfer Mediated Bacterial Antibiotic Resistance. *Front. Microbiol.* 10, 1933. doi:10.3389/fmicb.2019.01933

Wang, B., and Sun, D. (2015). Detection of Ndm-1 Carbapenemase-Producing Acinetobacter Calcoaceticus and Acinetobacter Junii in Environmental Samples from Livestock Farms. *J. Antimicrob. Chemother.* 70, 611–613. doi:10.1093/jac/dku405

Wu, Y.-W., Simmons, B. A., and Singer, S. W. (2016). Maxbin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets. *Bioinformatics* 32, 605–607. doi:10.1093/bioinformatics/btv638

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of Acquired Antimicrobial Resistance Genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi:10.1093/jac/dks261

Zhang, X.-X., Zhang, T., and Fang, H. H. P. (2009). Antibiotic Resistance Genes in Water Environment. *Appl. Microbiol. Biotechnol.* 82, 397–414. doi:10.1007/s00253-008-1829-z