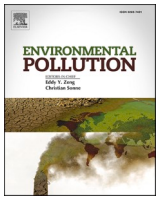
Environmental Pollution 357 (2024) 124424

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol



Towards the definition of an antibiotic resistome signature in wastewater and downstream environments☆

Diogo Cachetas [a], Ivone Vaz-Moreira [a,*], Vítor Pereira [b,c], Célia M. Manaia [a]

[a] Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Rua de Diogo Botelho 1327, 4169-005, Porto, Portugal
[b] Centre of Biological Engineering University of Minho, Braga, Portugal
[c] LABBELS –Associate Laboratory, Braga/Guimarães, Portugal

ARTICLE INFO

Keywords:
Metagenomics
Resistome
Environmental impact
One-health
Monitoring

ABSTRACT

Domestic wastewater is a significant reservoir of antibiotic resistance genes, which pose environmental and public health risks. We aimed to define an antibiotic resistome signature, represented by core genes, i.e., shared by ≥ 90% of the metagenomes of each of three conceptual environmental compartments – wastewater (influent, sludge, effluent), freshwater, and agricultural soil. The definition of resistome signatures would support the proposal of a framework for monitoring treatment efficacy and assessing the impact of treated wastewater discharge into the environment, such as freshwater and agricultural soil.

Metagenomic data from 163 samples originating from wastewater (n = 81), freshwater (n = 58), and agricultural soils (n = 24) across different regions (29 countries, 5 continents), were analysed regarding antibiotic resistance diversity, based on annotation against a database that merged CARD and ResFinder databases. The relative abundance of the total antibiotic resistance genes (corresponding to the ratio between the antibiotic resistance genes and total reads number) was not statistically different between raw and treated wastewater, being significantly higher than in freshwater or agricultural soils. The latter had the significantly lowest relative abundance of antibiotic resistance genes. Genes conferring resistance to aminoglycosides, beta-lactams, and tetracyclines were among the most abundant in wastewater environments, while multidrug resistance was equally distributed across all environments. The wastewater resistome signature included 27 antibiotic resistance genes that were detected in at least 90% of the wastewater resistomes, and that were not frequent in freshwater or agricultural soil resistomes. Among these were genes responsible for resistance to tetracyclines (n = 8), macrolide-lincosamide-streptogramin B (n = 7), aminoglycosides (n = 4), beta-lactams (n = 3), multidrug (n = 2), sulphonamides (n = 2), and polypeptides (n = 1). This comprehensive assessment provides valuable insights into the dynamics of antibiotic resistance in urban wastewater systems and their potential ecological implications in diverse environmental settings. Furthermore, provides guidance for the implementation of One Health monitoring approaches.

1. Introduction

Sewage, a major component of urban wastewater, has been identified as an important source of antibiotic resistance genes (ARGs) of human origin (Hendriksen et al., 2019; Munk et al., 2022). The wastewater treatment, unfortunately not globally available (WHO and UN-Habitat, 2022), can only partially remove those biological contaminants (Manaia, 2023; Pärnänen et al., 2019). The presence of ARGs in treated wastewater has a significant impact on the downstream environment, with frequent reports of subsequent contamination of receiving rivers, lakes, soils, or wastewater-irrigated vegetables with antibiotic resistant bacteria (ARB) and ARGs (Cerqueira et al., 2019; Christou et al., 2017; Larsson and Flach, 2022; Manaia, 2023). Under a One Health perspective, in which human health cannot be dissociated from animal well-being and environmental quality (Hernando-Amado et al., 2019), the integrated monitoring of antibiotic resistant bacteria and ARGs across distinct environments is fundamental. However, integrated monitoring may be doubly blurred by the ubiquitous occurrence

☆ This paper has been recommended for acceptance by Sarah Harmon.

* Corresponding author. CBQF, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Rua de Diogo Botelho, 1327, 4169-005, Porto, Portugal.
E-mail address: ivmoreira@ucp.pt (I. Vaz-Moreira).

https://doi.org/10.1016/j.envpol.2024.124424
Received 15 March 2024; Received in revised form 18 May 2024; Accepted 20 June 2024
Available online 21 June 2024

of non-acquired (i.e., not spread by horizontal gene transfer) ARGs and by the extremely low abundance of some ARGs in clean environments (e.g., disinfected wastewater, soils, vegetables) (Fortunato et al., 2018; Manaia, 2023, 2017; Marano et al., 2021).

Besides the culture-based assessment of antibiotic resistant bacteria, metagenomics and quantitative PCR (qPCR) have been proposed for the monitoring of ARGs in non-clinical contexts (Keenum et al., 2022; Manaia, 2023; Teixeira et al., 2023; The Water Research Foundation, 2023; Tiwari et al., 2022). While qPCR is a targeted method that allows the screening of specific genetic determinants, metagenomics, as a non-targeted method, offers a comprehensive overview. When supported by robust annotation databases, such as those existing for ARGs (e.g., CARD (Alcock et al., 2020) and ResFinder (Bortolaia et al., 2020), metagenomics permits the detection of hundreds of ARGs and thousands of allelic variants (Davis et al., 2023; Ferreira et al., 2023; Munk et al., 2022). The ever-increasing number of metagenomes deposited in public databases is a crucial resource to inform about ARGs biogeography and distribution across humans, animals, and the environment (Eckert et al., 2020; Munk et al., 2022; Pillay et al., 2022). However, unfortunately, often, these resources are not supported by reliable metadata (e.g., water quality parameters, temperature), which could inform on the abiotic conditions that may influence ARGs ecology. An important application benefiting from metagenomics data is the possibility of customizing ARGs monitoring protocols towards its operation and interpretation simplification, with the analysis of a selected group of genes (Ferreira et al., 2023; Teixeira et al., 2023).

In this study, we explored metagenomes available in public databases, therefore covering a wide geographic distribution, aiming the definition of a signature resistome, defined based on ARGs that were shared by $>$ 90% of each conceptual compartment represented by domestic wastewater (influent, sludge, effluent), freshwater or agricultural soil. Our vision is that identifying core and exclusive ARGs in different types of environments will facilitate the assessment of wastewater treatment efficacy, or the impacts of discharge of treated wastewater in the receptor environments (e.g., freshwater, agricultural soil).

2. Material and methods

In silico analyses were performed using a full Ubuntu terminal environment with Windows Subsystem for Linux (WSL) and Ubuntu 22.04 LTS from Canonical Group Limited, on a 16 GB RAM computer with 8 logical processors.

2.1. Metagenomes dataset

The metagenomes search was done using the terms "freshwater", "wastewater", and "agricultural soil" in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (accessed on July 25, 2022). Illumina paired-end sequencing data, assigned to a Bioproject and supported by metadata (e.g., sample type, geography), were selected and categorized as raw wastewater, treated wastewater, activated sludge, freshwater, or agricultural soil (Table S1). Due to technical limitation, a 50 GB threshold for metagenome sequence read archive (SRA) files was established. Redundant samples were discarded by eliminating metagenomes corresponding to the same location and assigned to the same Bioproject, selecting the replicate with the highest number of reads.

2.2. Metagenome's processing and annotation

The raw metagenomic reads were imported into a KBase narrative (Arkin et al., 2018), with the "Import SRA File as Reads From Web" (v1.0.7) app and the paired-end reads were merged, which were imported to the local machine in FASTQ format. The reads were aligned using KMA (k-mer alignment), a k-mers aligner that allows the trimming of the reads and matching of the k-mers between the query and the database (Clausen et al., 2018). After, the reads were aligned against two ARGs databases, the Comprehensive Antibiotic Resistance Database (CARD, accession date June 14, 2022, with 4634 sequences) (Alcock et al., 2020) and the ResFinder database (accession date June 14, 2022, with 3153 sequences) (Bortolaia et al., 2020). To avoid redundant annotations, in which distinct names were attributed to the same nucleotide sequence, a merged database that combined CARD and ResFinder databases was prepared for this study. To merge both databases, the respective sequences were organized into clusters using the CD-HIT-est in default mode, with a nucleotide sequence identity threshold of 90% (Huang et al., 2010). By clustering the ARGs sequences using CD-HIT-est, most of the allelic variants of the same ARG were grouped in the same cluster, being the longest selected as cluster reference (Table S3). This approach contributed to overcoming the existence of distinct ARG annotations with homotypic synonyms in both databases and, therefore, simplified the definition of signature resistomes.

The databases, individually or merged, were indexed to KMA for the alignment of the raw reads using the command line "*kma -ipe *.fastq.1 *.fastq.2 -o *_kma -t_db -t 8 -1t1 -mem_mode -ef*", with the parameters: paired-end files (*-ipe*), because the metagenomes were sequenced with paired-end Illumina; force each query sequence to match to only one template (*-1t1*), to avoid overestimation; ConClave algorithm carried out based on the mapping scores rather than alignment scores (*-mem_mode*), to use less memory; generation of a mapstat file with extended features (*-ef*), needed to calculate the ARGs relative abundance further. The k-mer length was set to default (https://gensoft.pasteur.fr/docs/kma/1.2.22/KMAspecification.pdf). To guarantee an accurate annotation, avoiding fragile alignments based on short hits, only sequences that aligned with the template by $>$ 200 bp were considered.

For resistome analysis, both databases CARD and ResFinder queried independently, or merged (CARD + ResFinder) were considered. The ARGs were classified according to the respective annotation in the database, referring to resistance phenotypes and ARGs classes: aminoglycosides, amphenicols, beta-lactams, disinfectants, fluoroquinolones, MLSB (macrolides-lincosamide-streptogramin B), multidrug, polypeptides, sulphonamides, tetracyclines, and other (including aminocoumarins, ansamycins, bicyclomycin, diaminopyrimidines, elfamycins, fosfomycins, fusidanes, glycopeptides, macrolides, mupirocin, nitroimidazoles, nucleosides, rifamycins, triterpenes). The relative abundance of ARGs, or ARGs classes, was calculated based on the number of annotated reads *versus* the total number of reads, as indicated in each figure or output. As a complement, the number of 16S rRNA gene reads and the taxonomic annotation were determined for each metagenome. The SILVA rRNA database version 111 (Quast et al., 2013) was processed with SortMeRNA to obtain a representative 16S rRNA database (Kopylova et al., 2012), and the metagenome raw reads were annotated based on the alignment against the SILVA rRNA reference database, using KMA (Clausen et al., 2018), and with Kaiju app (Menzel et al., 2016) against the NCBI BLAST nr (no Euks) database.

2.3. Resistome analysis and resistome signature

The first step to define a resistome signature consisted of searching for ARGs that were present in at least 90% of the metagenomes of a given type of sample (core resistome of wastewater, freshwater, or agricultural soil). To minimize the chance of false negative results were also identified the ARGs present in at least 75% of the metagenomes. To analyse the ARGs annotation distribution a Python pipeline was developed, using scripts available on GitHub at https://github.com/pg42866/Metagenomic-Analysis (Fig. S1). The core resistomes were determined based on the merged database (CARD + ResFinder), and also based on each database individually.

A wastewater ARGs signature included ARGs present in the wastewater core resistome but that were absent from the freshwater or agricultural soil core resistomes. This selection of wastewater ARGs included clusters of sequences ($>$90% sequence identity). When the cluster was

composed by multiple sequences a representative *consensus* sequence was created by multiple sequence alignments using the EMBL-EBI Clustal Omega tool (McWilliam et al., 2013), and the *consensus* sequence designed with the EMBL-EBI Emboss Cons tool (McWilliam et al., 2013) using default settings. For verification of the ARGs signatures, the ARGs *consensus* sequences were aligned against the metagenomes examined in this study using the KMA aligning method (Clausen et al., 2018), and against the freshwater, soil, and human-associated metagenomes available at the MGnify database (https://www.ebi.ac.uk/metagenomics) using the MGnify sequence search tool. Following the criteria used by CARD-RGI, a Bit-score cut-off (blastP) of 500 was defined. An ARG was confirmed as a suitable biomarker candidate when the *consensus* sequence was not identified in more than 40% of the freshwater and agricultural soil metagenomes under study (Fig. S5) or gave a Bit-score (blastP) above the cut-off value against the MGnify freshwater and soil databases. ARGs that gave poor matches (blastP Bit-score <500) against human associated metagenomes during the validation against the MGnify database were also excluded.

3. Results

3.1. ARG and taxonomic profiling of wastewater, freshwater, and agricultural soils

The dataset generated for the study included a total of 163 metagenomes from the following environments: wastewater (n = 81; 30 influent, 21 effluent, and 30 sludge); freshwater (n = 58); and agricultural soil (n = 24) (Table S2), with origin in 29 countries and five continents (Fig. 1). The number of reads ranged from 0.7 to 370 million reads in wastewater (1.3–60 million reads for influent; 1.4–46 million reads for effluent; 0.7–370 million reads for sludge), 1.2–284 million reads in freshwater and 3.5–321 million reads in agricultural soil (Table S2). The 16S rRNA gene was annotated in these metagenomes as a simplified method to assess bacteria diversity. The relative abundance (per total number of reads) of this gene was similar for all the metagenomes, in the order of magnitude of −5 log-units. The taxonomic annotation was dominated by the phylum *Pseudomonadota*, mainly the classes *Alpha-*, *Beta-*, and *Gammaproteobacteria*, in all types of samples (Fig. 2a and b), although with significant differences between samples (Table S4). *Bacteroidota* was the second most prevalent phylum ($p < 0.01$) in sludge, freshwater, and agricultural soil, while *Actinomycetota* was the second most prevalent in wastewater influent, and effluents.

The ARGs annotation for the 163 metagenomes against the two databases, including allelic variants, yielded 1670 genes against CARD database and 1413 against ResFinder database. Resistance to aminoglycosides, beta-lactams, multidrug, tetracyclines, and macrolide-lincosamide-streptogramin B (MLSB) were among the most prevalent classes in both cases (Fig. 2c and d and Fig. S2). The abundance of ARG reads relative to the total number of reads was not significantly different (p-value >0.01) in influent and effluent wastewater, for both ResFinder and CARD annotations (Fig. 3a and c, Table S6). Accordingly, the ratio of ARG per 16S rRNA gene reads (referring to total bacteria) was also not significantly different in effluent and in influent wastewater metagenomes (Fig. 3b and d). As expected, freshwater environments presented the significantly lowest relative abundance of ARGs (in terms of total reads), although not significantly lower than the wastewater effluent in terms of ARGs relative abundance per 16S rRNA gene abundance. Agricultural soils presented significantly lower relative abundance (per total reads and 16S rRNA gene) of putatively acquired ARGs than wastewater influent, according to the ResFinder annotation. However, the same was not observed for the total resistome (CARD annotation), in which agricultural soils were only significantly different from freshwater (Fig. 2). Because CARD includes both intrinsic and acquired ARGs, it was observed a dominance of genes associated with multidrug efflux pumps, some of them intrinsic in certain taxonomic groups, in all samples, leading to a limited discrimination of the resistomes. In contrast, ResFinder mainly includes ARGs associated with

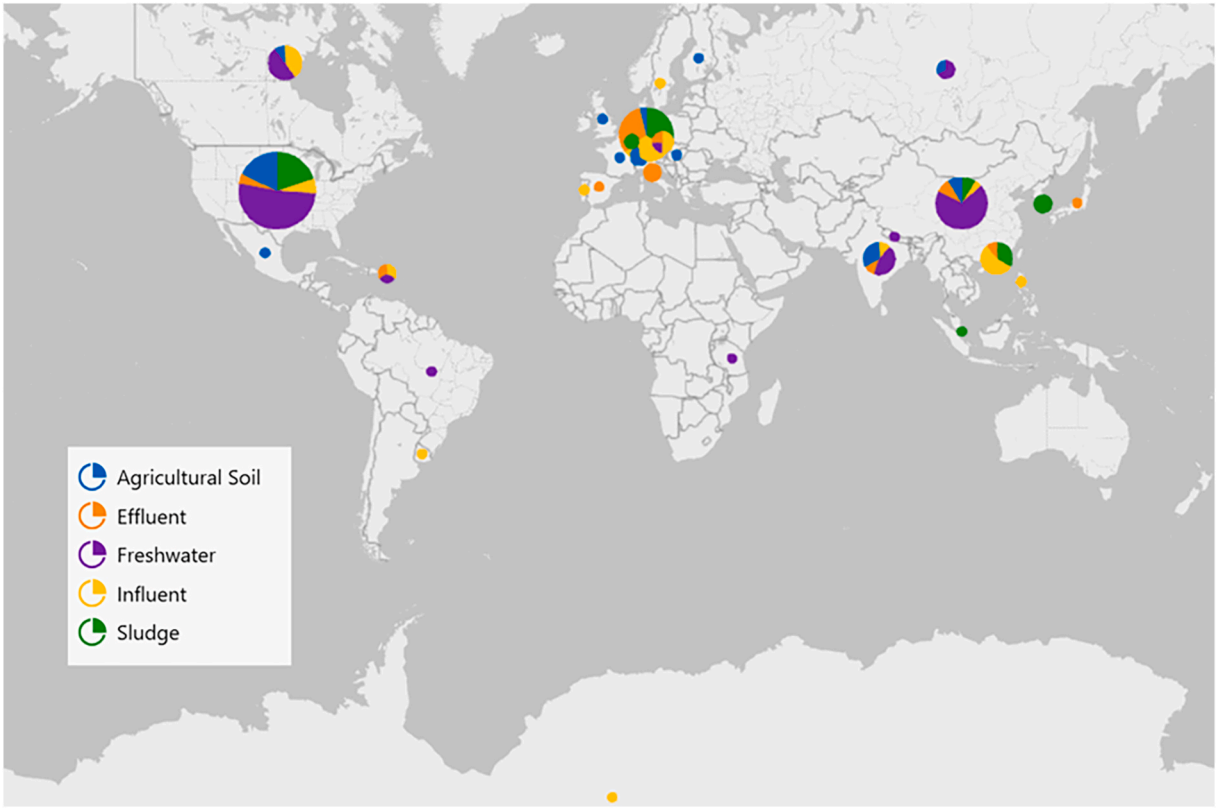


Fig. 1. Global geographic distribution of the selected metagenomes according to the origin (influent, effluent, sludge, freshwater, or agricultural soil).

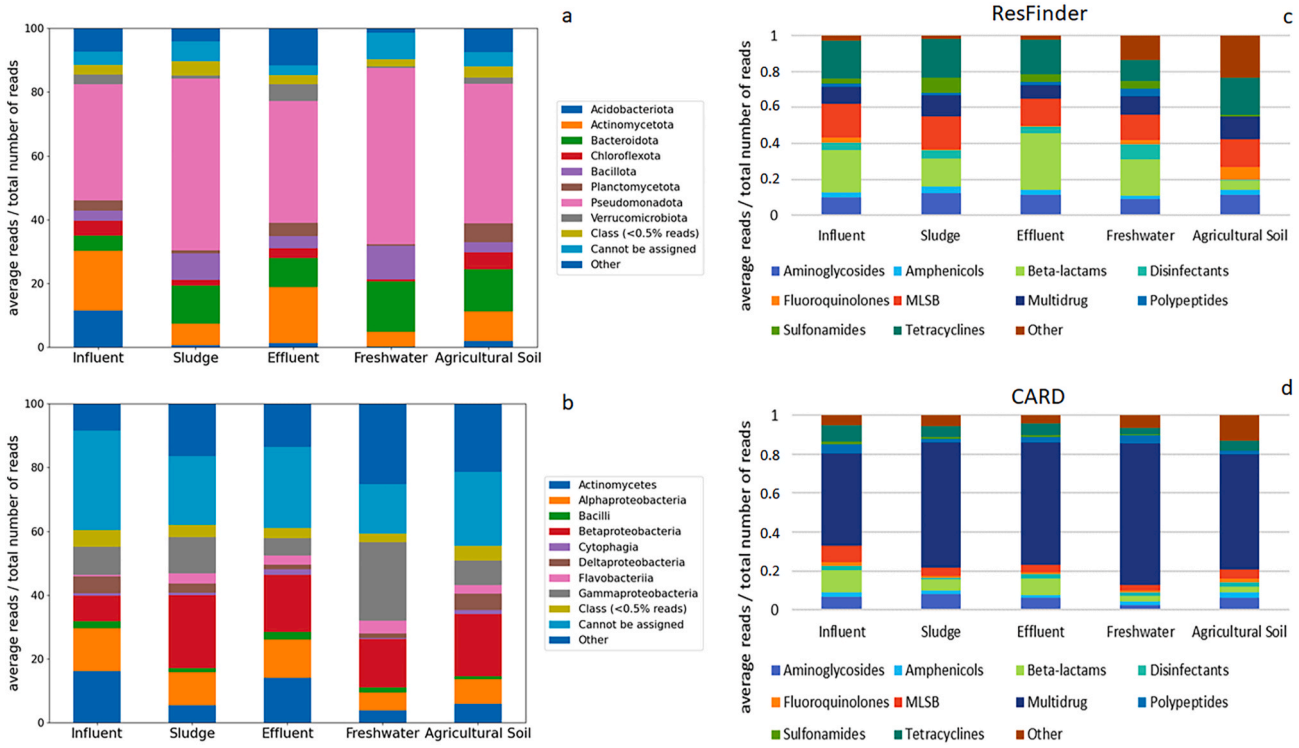


Fig. 2. Taxonomic and ARGs profiles for influent, sludge, effluent, freshwater, and agricultural soil. Average relative abundance for the taxonomic annotations of the raw metagenomes using Kaiju at the Phylum (a) and class (b) level. Average relative abundance (reads per total number of reads) for the top 10 most abundant ARGs classes, based on the ResFinder (c) and CARD (d) annotations. (the statistical analysis results are detailed in Tables S4 and S5).

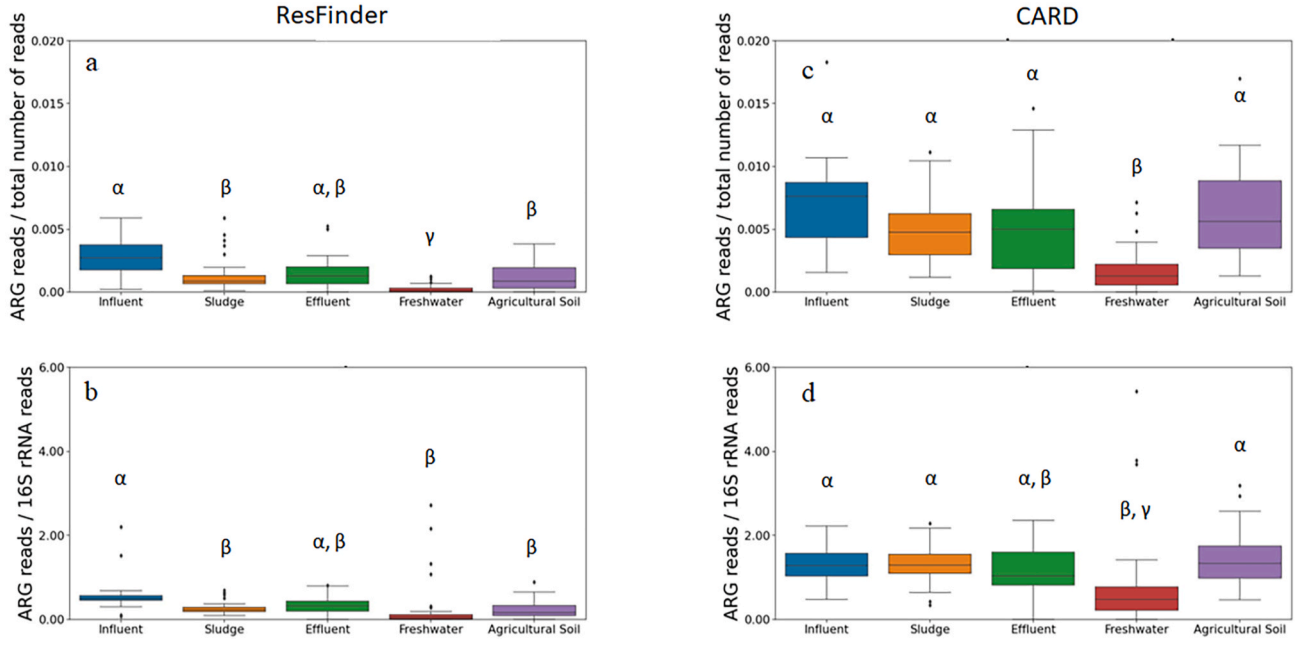


Fig. 3. Boxplots of the ARGs profile in terms of ARGs reads relative abundance per total number of reads considering the ResFinder annotation (**a**) and the CARD annotation (**c**), and ARGs reads per 16S rRNA gene reads for the ResFinder annotation (**b**) and CARD annotation (**d**). α,β,γ indicate the statistical differences ($p < 0.01$) between types of samples (detailed p-values in Table S6).

acquired resistance, being observed a higher diversity of antibiotic resistance classes with a predominance of beta-lactam, MLSB and tetracycline resistance genes (Fig. 2c). The significantly lower relative abundance of beta-lactam resistance genes clearly differentiated soils from (waste)water, while a lower relative abundance of tetracycline and higher relative abundance of disinfectant genes distinguished freshwater from all other sample types (Fig. 2c).

The relative abundance per total read patterns of ARGs organized by drug family (Fig. S2), shows that across wastewater samples a pattern/signature was identified after organizing/grouping the ARGs by drug family (characterized by a higher relative abundance of aminoglycosides, beta-lactams, MLSB, and tetracyclines resistance genes), while the

receiving environments do not show a defined profile. This can be a consequence of fewer ARG classes represented in the receiving environments. While the wastewater metagenomes have a median number of 11 ARG classes (ranging from 6 to 11 in the influent and sludge, and 1–11 in the effluent), the agricultural soil metagenomes have 7 (ranging from 1 to 10), and the freshwater metagenomes a median of 4.5 (ranging from 1 to 11) ARG classes, considering the ResFinder annotation (Fig. S2a). For 19% of the freshwater metagenomes (11 out of 58) only one ARG class was detected, which could be due to the lower ARG load of these samples. This lower ARGs richness and diversity is also reflected in a lower dispersion of the freshwater and agricultural soil samples in a Principal Component Analysis (Figs. S3a and S3b). The agricultural soil and freshwater metagenomes cluster together and apart from most of the wastewater metagenomes. This same trend was not observed for the PCA based on the taxonomic annotation (Figs. S3c and S3d).

According to the CARD annotation, in treated wastewater the relative abundance (per total reads number) of ARGs associated with polypeptides and tetracyclines resistance was significantly lower ($p < 0.01$) than for the raw influent (Fig. 2d–Table S5). In opposition, the relative abundance of multidrug resistance genes was significantly higher in treated wastewater. However, it must be noted that in the CARD annotation, >40% of the ARG reads corresponded to that class. Part of these genes are intrinsic efflux systems, possibly involved in housekeeping processes (Zhang and Feng, 2016). In comparison to freshwater, effluents presented higher ($p < 0.01$) relative abundance (per total number of reads) of genes associated with aminoglycosides, beta-lactams, and tetracyclines resistance than the freshwater, according to CARD annotation (Fig. 2d; E vs F in Table S5). Agricultural soils presented lower relative abundance (per total reads number) of ARGs for beta-lactams and sulfonamides, according to both CARD and ResFinder annotations (Fig. 2c and d; E vs A in Table S5). Interestingly, according to the ResFinder annotation other significant differences were observed, especially with soils presenting a higher relative abundance of genes related to resistance to fluoroquinolones (Fig. 2, Table S5).

3.2. Identifying core ARGs and proposing resistome signatures

The ARGs that were present in at least 90% of metagenomes of each type of sample, were further investigated aiming at defining sample-specific signatures. The number of signature ARGs varied according to the database used, being 25 with ResFinder (all core in wastewater samples - 23 in influent, 8 in sludge, 7 in effluent) and 50 with CARD (40 in influent, 20 in sludge, 12 in effluent, and 12 in agricultural soil) (Fig. S4). To avoid this difference resulting from the different content of the two databases and, in some cases, also different annotations, we merged the two databases (CARD + ResFinder), which led to the identification of 1278 clusters of ARGs sequences with >90% similarity (Table S3). Of those, 579 clusters contained a single sequence, from one of the databases. Considering the annotation with the merged database (CARD + ResFinder), 52 ARGs clusters were observed to be resistome signatures in at least one type of sample (42 in influent, 21 in sludge, 13 in effluent, and 12 in agricultural soil) (Fig. 4). CARD being the most comprehensive database, the merged result is close to the one obtained with CARD. However, two of the signatures were only detected via ResFinder – cluster 379 (*msrD*) and cluster 273 (*mefA*) (Fig. S4 and Table S3).

The core wastewater resistome was observed to be richer than that of freshwater or agricultural soil. While 49 ARGs clusters were identified in >90% of the wastewater metagenomes, only 12 ARGs clusters were common in 90% of agricultural soil metagenomes and none in freshwater metagenomes (Fig. 4). Considering the wide diversity of metagenomes examined, produced from samples collected in different geographies, in distinct occasions and with different equipment, the variability observed may be influenced also by technical and methodological factors. To avoid misinterpretations or underestimations, we also highlighted the ARGs clusters that were identified in at least 75% of the metagenomes (highlighted in orange in Fig. 4 and Fig. S4). For sludge and effluent samples, most of the ARGs signatures identified for the raw influent were also identified in those samples, if not in 90% in 75% of the metagenomes. In the freshwater samples, that presented the highest variability (Fig. S2), with the cut-off of 75% it was possible to identify 5 possible ARGs signatures – cluster 701 (*rpoB*), 702 (*rpoB2*), 706 (*mexF*), 715 (*axyY*), and 718 (*muxB*) (Fig. 4). Most of the signatures found for agricultural soil samples were also identified for the other types of sample, mainly wastewater – clusters 701 (*rpoB*), 702 (*rpoB2*), 706 (*mexF*), 714 (*mexB*), 715 (*axyY*), 718 (*muxB*), 725 (*muxC*), 729 (*ceoB*), and 731 (*mexK*). Most of them are associated with intrinsic resistance. In addition, three gene clusters were identified as exclusive signatures for agricultural soil: clusters 379 (*rphA*), 959 (*vanS*), and 1151 (*vanR*) (Fig. 4).

3.3. Proposal of ARGs wastewater biomarkers

The objective of this work was to identify ARGs signatures and based on these, identify possible biomarkers, i.e. genes that can be used as measurable parameter to indicate the antibiotic resistance load, for purposes such as monitoring wastewater treatment efficacy or assessing the impact of treated wastewater discharge into the receiving environments. For that, suitable biomarkers need to be genes that are common in wastewater but not in the receptor environments. To identify the genes candidates to biomarkers, ARGs observed as core in at least one of the wastewater metagenomes (influent, sludge, and/or effluent) but not in more than 75% of the freshwater and/or agricultural soil metagenomes were selected. This procedure reduced the list to 37 gene clusters that were considered as possible wastewater biomarkers (Table 1). The list encompasses genes associated with resistance to different classes of antibiotics, such as aminoglycosides (e.g., *ant(3″), aadA, aac(6′), aph, acrD*), aminocoumarins (*mdtB*), beta-lactams (bla_{TEM} *and* bla_{OXA}), disinfectants (*qacH*, *qacL*, *qacE*), fluoroquinolones (*aac(6')-Ib-cr, emrB*), MLSB (*ermB, ermF, mefA, mphA, mphE, mrsD, msrE*), multidrug (*mel, acrB, smeB, mexI, mdtP*), polypeptides (*eptB*), sulphonamides (*sul1 and sul2*), and tetracyclines (e.g., *tetA, tetC, tetG, tetQ, tetM, tetO, tetX*) (Table 1).

Because we worked with clusters of highly similar sequences, for each gene cluster of the candidate biomarker genes was defined a representative *consensus* sequence (Table S7). The definition of these representative sequences facilitates the definition of protocols for monitoring, for example based on quantitative methods for which it is necessary to design primers. We used the *consensus* sequences, or the sequence of single-sequence clusters, to do a cross-validation aligning those sequences against the metagenomes under study (Fig. S5) and a largest metagenomes database (EMBL-EBI MGnify) (Table S8), to validate that the selected genes are common in wastewater or human-associated samples but not in freshwater or agricultural soil. From the cross-validation 27 out of the 37 genes were validated as possible biomarker candidates. The remaining 10 were excluded because: i) they were well identified in freshwater and/or soil metagenomes, using a larger database or presented a high frequency (>40%) in the agricultural soil metagenomes used in this study; or ii) they were not well identified in human-associated metagenomes in a largest metagenomes database (Table 1).

4. Discussion

In this study, we aimed to use metagenomics data available in public databases, covering a broad geographic diversity, to investigate if we could identify antibiotic resistome signatures, i.e., ARGs sets that can be considered characteristic of wastewater, not detected in freshwater or agricultural soil (Manaia, 2023; Teixeira et al., 2023). We had a double interest in providing further insight into antibiotic resistance ecology, including regarding risks of transmission to humans (Zhang et al., 2021) and in recommending possible monitoring approaches (Tarek and

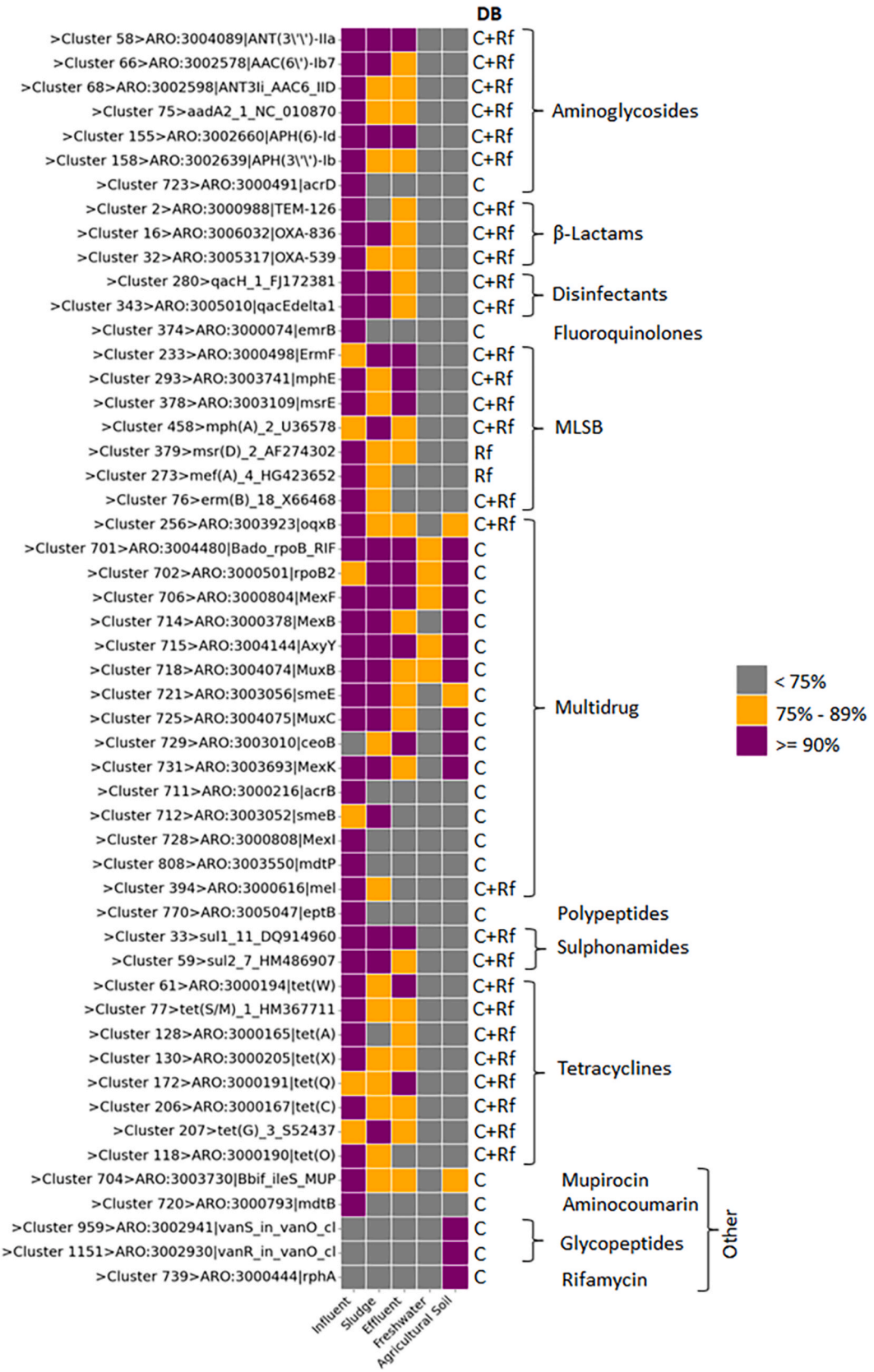


Fig. 4. Core resistomes based on the annotation of the influent (n = 30), sludge (n = 30), effluent (n = 21), freshwater (n = 58), and agricultural soil (n = 24) metagenomes using the merged database (CARD + ResFinder). In purple are indicated the genes of the core resistome (identified in ≥90% of the metagenomes) and in orange the ones identified in 75–89% of the metagenomes. On the right side are indicated the ARGs classes and the databases (DB) from where the sequences present in the cluster were collected, CARD (C) or ResFinder (Rf). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
List of ARGs selected as possible wastewater biomarkers, organized by ARG class, with the validation result. A validated biomarker consists of a gene that was observed to be common in wastewater or other human-associated metagenomes, but not in freshwater or agricultural soil metagenomes. (detailed information about the sequences in each cluster available in Table S3).

ARG Class	Cluster no.	No. sequences in the cluster[a]	Reference Gene (acc. number)	Genes/genes variants in the cluster	Validated as a biomarker? (if No, why)
Aminoglycosides	58	20 (7 C, 13 Rf)	*ant(3″)*-IIa (X02340.1)	*ant(3″)-IIa*, *ant(3″)-Ia*, *aadA* variants	*Yes*
	75	16 (7 C, 9 Rf)	*aadA2_1* (NC_010870)	*aadA* variants	*Yes*
	68	17 (9 C, 8 Rf)	*ant(3″)-Ii-aac(6′)-IId* (AF453998.2)	*ant(3″)-Ii-aac(6′)-IId*, *aac(6′)-Ib* variants, *aadA* variants	*No (poorly detected in human associated metagenomes)*
	155	6 (1 C, 5 Rf)	*aph(6)-Id* (AF024602.1)	*aph(6)-Id* variants	*Yes*
	158	6 (1 C, 5 Rf)	*aph(3″)-Ib* (AF313472.2)	*aph(3″)-Ib* variants	*Yes*
	723	1 (C)	*acrD* (AP009048.1)	*acrD*	*No (detected in freshwater or soil metagenomes)*
Aminoglycosides/ Fluoroquinolones	66	18 (12 C, 6 Rf)	*aac(6′)-Ib7* (KR091911.1)	*aac(6′)-Ib* variants, *aac(6′)-Ib-cr* variants	*No (poorly detected in human associated metagenomes)*
Aminocoumarin	720	1 (C)	*mdtB* (U00096.1)	*mdtB*	*No (detected in freshwater or soil metagenomes)*
Beta-lactams	2	381 (198 C, 183 Rf)	$bla_{TEM\text{-}126}$ (AY628199.1)	bla_{TEM} variants	*Yes*
	16	65 (39 C, 26 Rf)	$bla_{OXA\text{-}836}$ (NG_065881.1)	bla_{OXA} variants	*Yes*
	32	36 (20 C, 16 Rf)	$bla_{OXA\text{-}539}$ (NG_052064.1)	bla_{OXA} variants	*Yes*
Disinfectants	280	3 (1C, 2 Rf)	*qacH_1* (FJ172381)	*qacH*, *qacL*	*No (poorly detected in human associated metagenomes)*
	343	3 (2C, 1 Rf)	*qacEdelta1* (U49101.1)	*qacE* variants	*No (poorly detected in human associated metagenomes)*
Fluoroquinolones, Multidrug	374	2 (C)	*emrB* (U00096.1)	*emrB*, *Kpne_KpnH*	*No (detected in freshwater or soil metagenomes)*
MLSB	76	16 (1 C, 15 Rf)	*ermB_18* (X66468)	*ermB* variants	*Yes*
	233	4 (1 C, 3 Rf)	*ermF* (M17124.1)	*ermF* variants	*Yes*
	273	3 (Rf)	*mefA* (HG423652)	*mefA* variants	*Yes*
	458	2 (1 C, 1 Rf)	*mphA_2* (U36578)	*mphA* variants	*Yes*
	293	3 (1C, 2 Rf)	*mphE* (DQ839391.1)	*mphE* variants	*Yes*
	379	2 (Rf)	*msrD_2* (AF274302)	*msrD* variants	*Yes*
	378	2 (1 C, 1 Rf)	*msrE* (EU294228.1)	*msrE*	*Yes*
Multidrug, MLSB	394	2 (1 C, 1 Rf)	*mel* (AF227521.1)	*mel*, *mefA*	*Yes*
Multidrug	711	1 (C)	*acrB* (U00096.3)	*acrB*	*No (detected in freshwater or soil metagenomes)*
	712	1 (C)	*smeB* (AF173226.1)	*smeB*	*No (detected in freshwater or soil metagenomes)*
	728	1 (C)	*mexI* (AE004091.2)	*mexI*	*No (detected in freshwater or soil metagenomes)*
	808	1 (C)	*mdtP* (AP009048.1)	*mdtP*	*Yes*
Polypeptides	770	1 (C)	*eptB* (FO203501.1)	*eptB*	*Yes*
Sulphonamides	33	34 (1 C, 33 Rf)	*sul1* (DQ914960)	*sul1* variants	*Yes*
	59	20 (1 C, 19 Rf)	*sul2* (HM486907)	*sul2* variants	*Yes*
Tetracyclines	128	7 (1 C, 6 Rf)	*tet(A)* (AF534183.1)	*tet(A)* variants	*Yes*
	206	4 (1 C, 3 Rf)	*tet(C)* (AY043299.1)	*tet(C)* variants	*Yes*
	207	4 (1 C, 3 Rf)	*tet(G)_3* (S52437)	*tet(G)* variants	*Yes*
	172	5 (1 C, 4 Rf)	*tet(Q)* (Z21523.1)	*tet(Q)* variants	*Yes*
	77	15 (1 C, 14 Rf)	*tet(S/M)_1* (HM367711)	*tet(S/M)*, *tet(M)* variants	*Yes*
	118	8 (2 C, 6 Rf)	*tet(O)* (M18896.2)	*tet(O)* variants, *tet(O/M/O)*, *tet(O/32/O)* variants	*Yes*
	61	19 (5 C, 14 Rf)	*tet(W)* (AJ222769.3)	*tet(W) variants*, tet(W/N/W), *tet(O/W) variants, tet(O/W/O) variants,* tet(W/32/O),	*Yes*
	130	7 (3 C, 4 Rf)	*tet(X)* (M37699.1)	*tet(X)* variants	*Yes*

[a] Number of sequences from CARD (C) and from ResFinder (Rf).

Garner, 2023; Teixeira et al., 2023; Zhang et al., 2020).

The wastewater core resistome has been investigated. Raza et al. (2022) used shotgun metagenomics to study the core resistome of wastewater samples (influent and effluent samples of twelve urban WWTPs from major cities in South Korea), and identified 16 ARGs as the major core genes. de Nies et al. (2022) also used shotgun metagenomics to analyse the resistome of a municipal biological WWTP, sampled weekly over 1.5 years, having observed that a core group of 15 antibiotic resistance categories was found across all time points. Recently, Teixeira et al. (2023) tested candidate biomarkers, based on qPCR, aiming to monitor ARGs that may suggest the degree of anthropogenic contamination of wastewater and downstream environments. Tarek and Garner (2023) identified indicator ARGs, based on published metagenomic sequencing data, and developed a framework to identify indicator ARGs that incorporated their clinical and environmental relevance, abundance in wastewater, geographic ubiquity, ARG mobility and association with mobile genetic elements, and the availability of quantitative analytical methods for monitoring (Tarek and Garner, 2023). In our study, we aimed to overcome some of the constraints observed in these studies, such as: i) the reduced number of samples or absence of comparison with other environments, by reusing metagenomics data available in public databases for the different environments (wastewater, freshwater, and agricultural soil) and from different geographies; and ii) the use of a single ARGs annotation database, by querying two of the most comprehensive ARGs annotation databases (CARD and ResFinder) that were merged to uniformize genes nomenclature.

The ARGs richness was directly related to the expected anthropogenic pollution gradient (influent > sludge > effluent > agricultural soil > freshwater), as also described by other authors (Di Cesare et al., 2023). This gradient was also reflected in the number of ARGs belonging

to the core resistome of the different environments. In this study, we found core resistomes with distinct profiles for the different sample types (wastewater and agricultural soil), although with some co-incidences (Fig. 4). For freshwater were not identified ARGs for >90% of the metagenomes analysed. Among the 12 agricultural soil core ARGs, nine were also core in wastewater. The exceptions were the genes *rphA*, *vanS*, and *vanR*, exclusive of the agricultural soil metagenomes. Vancomycin resistance genes were previously described as part of the core resistome of cropland soil (Du et al., 2020). Regarding the wastewater resistome, more genes were identified as belonging to the core resistome for the influent samples than for the sludge or effluent samples. This is certainly a consequence of the capacity of wastewater treatment to remove bacteria. Indeed, wastewater treatment has been associated with an increase in the ARGs hosts diversity, reflected in a broad diversity of plasmid recipient bacteria (Jacquiod et al., 2017). Also, previous studies have already described a higher similarity, in terms of bacterial communities and ARGs profiles, for the wastewater before treatment (influent) than after treatment (effluent), even when considering the same WWTP (Bengtsson-Palme et al., 2016; Lira et al., 2020).

There is a wide range of shared ARGs across the WWTP continuum (Quintela-Baluja et al., 2019). In our study, considering the ARGs present in the core resistome (>90% of the analysed metagenomes for each type of sample), it was observed that 38% of the ARGs in influent were detected also in activated sludge, and that 52% of the ARGs found in activated sludge were present in the final effluent core resistome. In a recent study, it was concluded that around 50% of the ARGs in the influent remained in activated sludge, and about 85% of the remaining ARGs in the sludge were carried over to the effluent (Honda et al., 2023). In a few cases, we identified core genes in sludge or effluent metagenomes that were not observed in the influent core. Mostly, they were in at least 75% of the influent metagenomes, but did not reach 90%, except for the gene *ceoB*, a cytoplasmic membrane component of the CeoAB-OpcM efflux pump. (Fig. 4). This may be related to a lower relative abundance in the influent samples and an increase after the treatment due to selective pressures, making the detection easier. This was previously described for other ARGs (Manaia et al., 2018; Wang et al., 2022).

Most of the core ARGs in wastewater were annotated as being associated with resistance to aminoglycosides, MLSB, tetracyclines, and multidrug. de Nies et al. (de Nies et al., 2022), based on an extended sampling (1.5 years) of a biological WWTP, described the categories aminoglycosides, beta-lactam, and multidrug, followed by tetracycline and MLSB as being persistent over time. Li et al. (2018), although not using the definition of core resistome, also defined those categories as the most represented in activated sludge from urban wastewater treatment plant metagenomes, considering 59 metagenomes. Surprisingly, we did not identify beta-lactam resistance genes among the most represented ARGs in the wastewater core resistome. We believe that this is a consequence of the high genetic diversity within each beta-lactam resistance gene family (>5480 sequences, in 360 clusters, considering the merged CARD + ResFinder database), reducing the probability of coincidence among metagenomes, together with the methodological approach we used of create gene clusters based on the sequence similarity instead of gene family. For example, the $bla_{CTX\text{-}M}$ gene variants were divided among eight gene clusters (clusters number 5, 8, 27, 44, 116, 151, 218, and 459) (Table S3). However, the definition of gene clusters based on sequence similarity is critical when defining putative biomarkers. As putative wastewater contamination biomarkers we only considered signature resistome ARGs, which are present in >90% wastewater samples (influent, sewage, or effluent) but that are not part of the core resistome of freshwater and agricultural soil, possible receptor environments. Among the latter environments, freshwater did not yield a core resistome and in agricultural soil, most of these were associated with multidrug resistance (Fig. 4). Interestingly, all those genes were annotated with the CARD database, highlighting the importance of using multiple databases for ARGs annotation.

CARD and ResFinder are two of the most comprehensive curated ARGs databases. However, the inconsistency of ARGs nomenclature between both databases can cause some hurdles in the results interpretation. In this work, we surpassed this problem by merging CARD and ResFinder databases into a common nomenclature, based on sequences similarity clusters to enable the definition of resistomes signatures. From the ARGs profiles (Fig. 2c and d, and Fig. S2) we can conclude that ResFinder was better annotating the categories beta-lactams and MLSB, while CARD mostly identified multidrug resistance. This observation was previously described (Lal Gupta et al., 2020), and mostly results from the fact that beta-lactams and MLSB resistance genes are mainly acquired genes, while multidrug resistance genes are frequently chromosomal (Che et al., 2019).

The wastewater ARGs signature was established based on the core resistome, comprising 49 ARGs for wastewater and 12 for agricultural soil (Fig. 4). The wastewater core ARGs that were not part of the agricultural soil core, constituted the 37 ARGs signature for wastewater from which it was possible to select monitoring biomarker candidates (Fig. S5, Table 1). After a cross-validation against a larger metagenomes database, 10 out of the 37 ARGs were considered as inadequate candidates because they were not frequently detected in human-associated metagenomes or were frequently detected in freshwater or soil metagenomes (Table 1, Table S8). This information can be used to design monitoring approaches, for example targeted quantitative PCR (qPCR), for a reduced number of ARGs. Our work proposes 27 biomarkers specific for wastewater that can be used for risk assessment to monitor wastewater and wastewater impacted environments (e.g. agricultural soil and pristine water bodies) (Table 1). From those 27 possible biomarkers, at least 14 (*ant(3″)-Iia*; *aadA*; *aph(3″)-Ib*; *ermB*; *ermF*; *mel*; *sul1*; *sul2*; *tetA*; *tetC*; *tetG*; *tetO*; *tetW*; *tetX*) were previously described as good wastewater biomarkers (Manaia, 2023; Tarek and Garner, 2023; Teixeira et al., 2023; Zhang et al., 2020). Some of those genes (e.g. *sul1*) may also be reported in pristine environments, but are markedly more abundant in human impacted environments (Pruden et al., 2012). Our results mostly (5 out of the 7 ARGs proposed) validate the biomarkers recently proposed by Teixeira et al. (2023). The authors, however, used a different approach based on the validation of the absence in clean environments of a group of 60 pre-selected antibiotic resistance and housekeeping genes and mobile genetic elements (MGEs) that have been frequently associated with anthropogenic impact, have been classified as being clinically relevant, and/or observed to occur in wastewater environments. In this way, our study contributes to enlarge the number of wastewater biomarkers that can be used to monitor the impact that the discharge of wastewater may have in the receptor environments. Also, this study confirms that wastewater contains ARGs that are seldomly observed in pristine sources and that a large fraction (26/49) of these can be spread to the environment via biosolids (sludge) or effluents.

5. Conclusions

Based on two ARGs databases (CARD and ResFinder), our study successfully defined the profiles of the core resistome of wastewater (influent, sludge, and effluent) and agricultural soil. Wastewater core resistome was dominated by ARGs conferring resistance to aminoglycosides, MLSB, tetracyclines, and multidrug, while multidrug resistance ARGs dominated agricultural soil core resistome. The freshwater resistome is the poorest and least conserved, hindering the definition of a core resistome (ARGs present in >90% of the freshwater samples).

After solving the problem of non-coincident nomenclatures, by analyzing the merged (CARD + ResFinder) databases' gene clusters, using the two databases proved to be the best approach to increasing the coverage of the ARGs nomenclature. Our work again highlights the importance of unifying ARGs nomenclature across databases for a more accessible data comparison and integration.

Overall, metagenomics data analysis proved to be an excellent

approach to define the environmental resistomes and allowed to define 27 ARGs as wastewater biomarkers (*ant(3″)-Iia*, *aadA2*, *aph(6)-Id*, *aph (3″)-Ib*, $bla_{TEM-126}$, $bla_{OXA-836}$, $bla_{OXA-539}$, *ermF*, *mphE*, *msrE*, *mphA*, *msrD*, *mefA*, *ermB*, *mdtP*, *mel*, *eptB*, *sul1*, *sul2*, *tetW*, *tet(S/M)*, *tetA*, *tetX*, *tetQ*, *tetC*, *tetG*, and *tetO*). Those biomarkers can be an essential tool for tracking the emergence and dissemination of antibiotic resistance in the receptor environments due to wastewater discharge, a critical step in safeguarding both environmental health and human well-being.

Funding

This work was developed within the scope of the project HSoil4Food - Healthy soils for healthy foods (NORTE-01-0145-FEDER-000066) co-financed by the European Regional Development Fund (FEDER) through the Northern Regional Operational Program. This work was also supported by funding from the National Funds from the Fundação para a Ciência e a Tecnologia (FCT) through project UIDB/50016/2020.

CRediT authorship contribution statement

Diogo Cachetas: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ivone Vaz-Moreira:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Vítor Pereira:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Célia M. Manaia:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2024.124424.

References

Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.L.V., Cheng, A.A., Liu, S., Min, S.Y., Miroshnichenko, A., Tran, H.K., Werfalli, R.E., Nasir, J.A., Oloni, M., Speicher, D.J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A.N., Bordeleau, E., Pawlowski, A.C., Zubyk, H.L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G.L., Beiko, R.G., Brinkman, F.S.L., Hsiao, W.W.L., Domselaar, G.V., McArthur, A.G., 2020. Card 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48, D517–D525. https://doi.org/10.1093/NAR/GKZ935.

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M.W., Henderson, M.L., Riehl, W. J., Murphy-Olson, D., Chan, S.Y., Kamimura, R.T., Kumari, S., Drake, M.M., Brettin, T.S., Glass, E.M., Chivian, D., Gunter, D., Weston, D.J., Allen, B.H., Baumohl, J., Best, A.A., Bowen, B., Brenner, S.E., Bun, C.C., Chandonia, J.M., Chia, J. M., Colasanti, R., Conrad, N., Davis, J.J., Davison, B.H., Dejongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J.N., Fang, G., Faria, J.P., Frybarger, P.M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H.L., He, F., Jain, R., Joachimiak, M.P., Keegan, K.P., Kondo, S., Kumar, V., Land, M.L., Meyer, F., Mills, M., Novichkov, P.S., Oh, T., Olsen, G.J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S.S., Price, G.A., Ramakrishnan, S., Ranjan, P., Ronald, P.C., Schatz, M.C., Seaver, S.M.D., Shukla, M., Sutormin, R.A., Syed, M.H., Thomason, J., Tintle, N.L., Wang, D., Xia, F., Yoo, H., Yoo, S., Yu, D., 2018. KBase: the United States department of energy systems biology knowledgebase. Nat. Biotechnol. https://doi.org/10.1038/nbt.4163.

Bengtsson-Palme, J., Hammarén, R., Pal, C., Östman, M., Björlenius, B., Flach, C.F., Fick, J., Kristiansson, E., Tysklind, M., Larsson, D.G.J., 2016. Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. Sci. Total Environ. 572, 697–712. https://doi.org/10.1016/J.SCITOTENV.2016.06.228.

Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L., Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., Malhotra-Kumar, S., Westh, H., Pinholt, M., Anjum, M.F., Duggett, N.A., Kempf, I., Nykäsenoja, S., Olkkola, S., Wieczorek, K., Amaro, A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O., Aarestrup, F.M., 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J. Antimicrob. Chemother. 75, 3491–3500. https://doi.org/10.1093/JAC/DKAA345.

Cerqueira, F., Matamoros, V., Bayona, J., Elsinga, G., Hornstra, L.M., Piña, B., 2019. Distribution of antibiotic resistance genes in soils and crops. A field study in legume plants (Vicia faba L.) grown under different watering regimes. Environ. Res. 170, 16–25. https://doi.org/10.1016/j.envres.2018.12.007.

Che, Y., Xia, Y., Liu, L., Li, A.-D., Yang, Y., Zhang, T., 2019. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore metagenomic sequencing. Microbiome 7, 44. https://doi.org/10.1186/s40168-019-0663-0.

Christou, A., Agüera, A., Bayona, J.M., Cytryn, E., Fotopoulos, V., Lambropoulou, D., Manaia, C.M., Michael, C., Revitt, M., Schröder, P., Fatta-Kassinos, D., 2017. The potential implications of reclaimed wastewater reuse for irrigation on the agricultural environment: the knowns and unknowns of the fate of antibiotics and antibiotic resistant bacteria and resistance genes – a review. Water Res. 123, 448–467. https://doi.org/10.1016/j.watres.2017.07.004.

Clausen, P.T.L.C., Aarestrup, F.M., Lund, O., 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinf. 19, 1–8. https://doi.org/10.1186/S12859-018-2336-6/TABLES/2.

Davis, B.C., Brown, C., Gupta, S., Calarco, J., Liguori, K., Milligan, E., Harwood, V.J., Pruden, A., Keenum, I., 2023. Recommendations for the use of metagenomics for routine monitoring of antibiotic resistance in wastewater and impacted aquatic environments. Crit. Rev. Environ. Sci. Technol. 53, 1731–1756. https://doi.org/10.1080/10643389.2023.2181620.

de Nies, L., Busi, S.B., Kunath, B.J., May, P., Wilmes, P., 2022. Mobilome-driven segregation of the resistome in biological wastewater treatment. Elife 11. https://doi.org/10.7554/eLife.81196.

Di Cesare, A., Sabatino, R., Sbaffi, T., Fontaneto, D., Brambilla, D., Beghi, A., Pandolfi, F., Borlandelli, C., Fortino, D., Biccai, G., Genoni, P., Corno, G., 2023. Anthropogenic pollution drives the bacterial resistome in a complex freshwater ecosystem. Chemosphere 331, 138800. https://doi.org/10.1016/j.chemosphere.2023.138800.

Du, S., Shen, J.-P., Hu, H.-W., Wang, J.-T., Han, L.-L., Sheng, R., Wei, W.-X., Fang, Y.-T., Zhu, Y.-G., Zhang, L.-M., He, J.-Z., 2020. Large-scale patterns of soil antibiotic resistome in Chinese croplands. Sci. Total Environ. 712, 136418 https://doi.org/10.1016/j.scitotenv.2019.136418.

Eckert, E.M., Di Cesare, A., Fontaneto, D., Berendonk, T.U., Bürgmann, H., Cytryn, E., Fatta-Kassinos, D., Franzetti, A., Larsson, D.G.J., Manaia, C.M., Pruden, A., Singer, A. C., Udikovic-Kolic, N., Corno, G., 2020. Every fifth published metagenome is not available to science. PLoS Biol. 18, e3000698 https://doi.org/10.1371/journal.pbio.3000698.

Ferreira, C., Otani, S., Aarestrup, F.M., Manaia, C.M., 2023. Quantitative PCR versus metagenomics for monitoring antibiotic resistance genes: balancing high sensitivity and broad coverage. FEMS Microbes 4. https://doi.org/10.1093/femsmc/xtad008.

Fortunato, G., Vaz-Moreira, I., Becerra-Castro, C., Nunes, O.C., Manaia, C.M., 2018. A rationale for the high limits of quantification of antibiotic resistance genes in soil. Environ. Pollut. 243, 1696–1703. https://doi.org/10.1016/j.envpol.2018.09.128.

Hendriksen, R.S., Munk, P., Njage, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., Kaas, R.S., Clausen, P.T. L.C., Vogt, J.K., Leekitcharoenphon, P., van de Schans, M.G.M., Zuidema, T., de Roda Husman, A.M., Rasmussen, S., Petersen, B., Bego, A., Rees, C., Cassar, S., Coventry, K., Collignon, P., Allerberger, F., Rahube, T.O., Oliveira, G., Ivanov, I., Vuthy, Y., Sopheak, T., Yost, C.K., Ke, C., Zheng, H., Baisheng, L., Jiao, X., Donado-Godoy, P., Coulibaly, K.J., Jergović, M., Hrenovic, J., Karpíšková, R., Villacis, J.E., Legesse, M., Eguale, T., Heikinheimo, A., Malania, L., Nitsche, A., Brinkmann, A., Saba, C.K.S., Kocsis, B., Solymosi, N., Thorsteinsdottir, T.R., Hatha, A.M., Alebouyeh, M., Morris, D., Cormican, M., O'Connor, L., Moran-Gilad, J., Alba, P., Battisti, A., Shakenova, Z., Kiiyukia, C., Ng'eno, E., Raka, L., Avsejenko, J., Bērziņš, A., Bartkevics, V., Penny, C., Rajandas, H., Parimannan, S., Haber, M.V., Pal, P., Jeunen, G.-J., Gemmell, N., Fashae, K., Holmstad, R., Hasan, R., Shakoor, S., Rojas, M.L.Z., Wasyl, D., Bosevska, G., Kochubovski, M., Radu, C., Gassama, A., Radosavljevic, V., Wuertz, S., Zuniga-Montanez, R., Tay, M.Y.F., Gavačová, D., Pastuchova, K., Truska, P., Trkov, M., Esterhuyse, K., Keddy, K., Cerdà-Cuéllar, M., Pathirage, S., Norrgren, L., Örn, S., Larsson, D.G.J., Heijden, T. Van der, Kumburu, H. H., Sanneh, B., Bidjada, P., Njanpop-Lafourcade, B.-M., Nikiema-Pessinaba, S.C., Levent, B., Meschke, J.S., Beck, N.K., Van, C.D., Phuc, N. Do, Tran, D.M.N., Kwenda, G., Tabo, D., Wester, A.L., Cuadros-Orellana, S., Amid, C., Cochrane, G., Sicheritz-Ponten, T., Schmitt, H., Alvarez, J.R.M., Aidara-Kane, A., Pamp, S.J., Lund, O., Hald, T., Woolhouse, M., Koopmans, M.P., Vigre, H., Petersen, T.N., Aarestrup, F.M., 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat. Commun. 10, 1124. https://doi.org/10.1038/s41467-019-08853-3.

Hernando-Amado, S., Coque, T.M., Baquero, F., Martínez, J.L., 2019. Defining and combating antibiotic resistance from one health and global health perspectives. Nat. Microbiol. 49 (4), 1432–1442. https://doi.org/10.1038/s41564-019-0503-9, 2019.

Honda, R., Matsuura, N., Sorn, S., Asakura, S., Morinaga, Y., Van Huy, T., Sabar, M.A., Masakke, Y., Hara-Yamamura, H., Watanabe, T., 2023. Transition of antimicrobial resistome in wastewater treatment plants: impact of process configuration,

geographical location and season. npj Clean Water 6, 46. https://doi.org/10.1038/s41545-023-00261-x.
Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682. https://doi.org/10.1093/BIOINFORMATICS/BTQ003.
Jacquiod, S., Brejnrod, A., Morberg, S.M., Abu Al-Soud, W., Sørensen, S.J., Riber, L., 2017. Deciphering conjugative plasmid permissiveness in wastewater microbiomes. Mol. Ecol. 26, 3556–3571. https://doi.org/10.1111/mec.14138.
Keenum, I., Liguori, K., Calarco, J., Davis, B.C., Milligan, E., Harwood, V.J., Pruden, A., 2022. A framework for standardized qPCR-targets and protocols for quantifying antibiotic resistance in surface water, recycled water and wastewater. Crit. Rev. Environ. Sci. Technol. 52, 4395–4419. https://doi.org/10.1080/10643389.2021.2024739.
Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217. https://doi.org/10.1093/BIOINFORMATICS/BTS611.
Lal Gupta, C., Kumar Tiwari, R., Cytryn, E., 2020. Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. Environ. Int. 138, 105667 https://doi.org/10.1016/j.envint.2020.105667.
Larsson, D.G.J., Flach, C.F., 2022. Antibiotic resistance in the environment. Nat. Rev. Microbiol. https://doi.org/10.1038/s41579-021-00649-x.
Li, L.G., Yin, X., Zhang, T., 2018. Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. Microbiome 6, 93. https://doi.org/10.1186/S40168-018-0480-X/FIGURES/5.
Lira, F., Vaz-Moreira, I., Tamames, J., Manaia, C.M., Martínez, J.L., 2020. Metagenomic analysis of an urban resistome before and after wastewater treatment. Sci. Rep. 10, 1–9. https://doi.org/10.1038/s41598-020-65031-y.
Manaia, C.M., 2023. Framework for establishing regulatory guidelines to control antibiotic resistance in treated effluents. Crit. Rev. Environ. Sci. Technol. 53, 754–779. https://doi.org/10.1080/10643389.2022.2085956.
Manaia, C.M., 2017. Assessing the risk of antibiotic resistance transmission from the environment to humans: non-direct proportionality between abundance and risk. Trends Microbiol. 25, 173–181. https://doi.org/10.1016/j.tim.2016.11.014.
Manaia, C.M., Rocha, J., Scaccia, N., Marano, R., Radu, E., Biancullo, F., Cerqueira, F., Fortunato, G., Iakovides, I.C., Zammit, I., Kampouris, I., Vaz-Moreira, I., Nunes, O.C., 2018. Antibiotic resistance in wastewater treatment plants: tackling the black box. Environ. Int. 115, 312–324. https://doi.org/10.1016/j.envint.2018.03.044.
Marano, R.B.M., Gupta, C.L., Cozer, T., Jurkevitch, E., Cytryn, E., 2021. Hidden resistome: enrichment reveals the presence of clinically relevant antibiotic resistance determinants in treated wastewater-irrigated soils. Environ. Sci. Technol. 55, 6814–6827. https://doi.org/10.1021/acs.est.1c00612.
McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., Lopez, R., 2013. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 41, W597–W600. https://doi.org/10.1093/NAR/GKT376.
Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 71 (7), 1–9. https://doi.org/10.1038/ncomms11257, 2016.
Munk, P., Brinch, C., Møller, F.D., Petersen, T.N., Hendriksen, R.S., Seyfarth, A.M., et al., 2022. Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. Nat. Commun. 13, 7251. https://doi.org/10.1038/s41467-022-34312-7.
Pärnänen, K.M.M., Narciso-Da-Rocha, C., Kneis, D., Berendonk, T.U., Cacace, D., Do, T. T., Elpers, C., Fatta-Kassinos, D., Henriques, I., Jaeger, T., Karkman, A., Martinez, J. L., Michael, S.G., Michael-Kordatou, I., O'Sullivan, K., Rodriguez-Mozaz, S., Schwartz, T., Sheng, H., Sørum, H., Stedtfeld, R.D., Tiedje, J.M., Giustina, S.V. Della, Walsh, F., Vaz-Moreira, I., Virta, M., Manaia, C.M., 2019. Antibiotic resistance in European wastewater treatment plants mirrors the pattern of clinical antibiotic resistance prevalence. Sci. Adv. 5 https://doi.org/10.1126/sciadv.aau9124.
Pillay, S., Calderón-Franco, D., Urhan, A., Abeel, T., 2022. Metagenomic-based surveillance systems for antibiotic resistance in non-clinical settings. Front. Microbiol. 13 https://doi.org/10.3389/fmicb.2022.1066995.
Pruden, A., Arabi, M., Storteboom, H.N., 2012. Correlation between upstream human activities and riverine antibiotic resistance genes. Environ. Sci. Technol. 46, 11541–11549.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/NAR/GKS1219.
Quintela-Baluja, M., Abouelnaga, M., Romalde, J., Su, J.-Q., Yu, Y., Gomez-Lopez, M., Smets, B., Zhu, Y.-G., Graham, D.W., 2019. Spatial ecology of a wastewater network defines the antibiotic resistance genes in downstream receiving waters. Water Res. 162, 347–357. https://doi.org/10.1016/j.watres.2019.06.075.
Raza, S., Shin, H., Hur, H.-G., Unno, T., 2022. Higher abundance of core antimicrobial resistant genes in effluent from wastewater treatment plants. Water Res. 208, 117882 https://doi.org/10.1016/j.watres.2021.117882.
Tarek, M.H., Garner, E., 2023. A proposed framework for the identification of indicator genes for monitoring antibiotic resistance in wastewater: insights from metagenomic sequencing. Sci. Total Environ. 854, 158698 https://doi.org/10.1016/j.scitotenv.2022.158698.
Teixeira, A.M., Vaz-Moreira, I., Calderón-Franco, D., Weissbrodt, D., Purkrtova, S., Gajdos, S., Dottorini, G., Nielsen, P.H., Khalifa, L., Cytryn, E., Bartacek, J., Manaia, C.M., 2023. Candidate biomarkers of antibiotic resistance for the monitoring of wastewater and the downstream environment. Water Res. 247, 120761 https://doi.org/10.1016/j.watres.2023.120761.
The Water Research Foundation, 2023. Standardizing methods with QA/QC standards for investigating the occurrence and removal of antibiotic resistant bacteria/antibiotic resistance genes (ARB/ARGs) in surface water. Wastewater, and Recycled Water.
Tiwari, A., Kurittu, P., Al-Mustapha, A.I., Heljanko, V., Johansson, V., Thakali, O., Mishra, S.K., Lehto, K.-M., Lipponen, A., Oikarinen, S., Pitkänen, T., Heikinheimo, A., 2022. Wastewater surveillance of antibiotic-resistant bacterial pathogens: a systematic review. Front. Microbiol. 13 https://doi.org/10.3389/fmicb.2022.977106.
Wang, Y., Han, Y., Li, L., Liu, J., Yan, X., 2022. Distribution, sources, and potential risks of antibiotic resistance genes in wastewater treatment plant: a review. Environ. Pollut. 310, 119870 https://doi.org/10.1016/j.envpol.2022.119870.
WHO and UN-Habitat, 2022. Progress on wastewater treatment (SDG target 6.3) [WWW Document]. URL. https://sdg6data.org/en/indicator/6.3.1. accessed 5.24.23.
Zhang, A.-N., Gaston, J.M., Dai, C.L., Zhao, S., Poyet, M., Groussin, M., Yin, X., Li, L.-G., van Loosdrecht, M.C.M., Topp, E., Gillings, M.R., Hanage, W.P., Tiedje, J.M., Moniz, K., Alm, E.J., Zhang, T., 2021. An omics-based framework for assessing the health risk of antimicrobial resistance genes. Nat. Commun. 12, 4765. https://doi.org/10.1038/s41467-021-25096-3.
Zhang, G., Feng, J., 2016. The intrinsic resistance of bacteria. Yi chuan = Hered 38, 872–880. https://doi.org/10.16288/j.yczz.16-159.
Zhang, Y., Wang, J., Lu, J., Wu, J., 2020. Antibiotic resistance genes might serve as new indicators for wastewater contamination of coastal waters: spatial distribution and source apportionment of antibiotic resistance genes in a coastal bay. Ecol. Indic. 114, 106299 https://doi.org/10.1016/j.ecolind.2020.106299.