Science of the Total Environment 749 (2020) 141451

Contents lists available at ScienceDirect

Science of the Total Environment

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

Differential prevalence and host-association of antimicrobial resistance traits in disinfected and non-disinfected drinking water systems

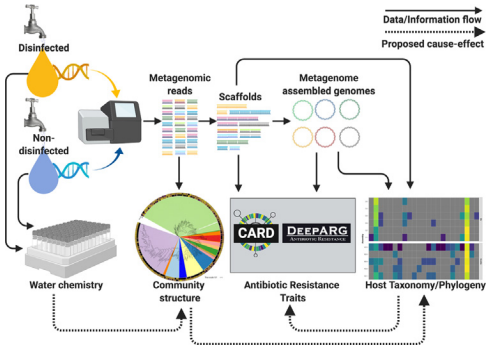
Maria Sevillano [a], Zihan Dai [b], Szymon Calus [b], Quyen M. Bautista-de los Santos [b], A. Murat Eren [c,d], Paul W.J.J. van der Wielen [e,f], Umer Z. Ijaz [b], Ameet J. Pinto [a,*]

[a] Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA
[b] Infrastructure and Environmental Research Division, School of Engineering, University of Glasgow, G12 8LT Glasgow, UK
[c] Department of Medicine, University of Chicago, Chicago, IL, USA
[d] Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA, USA
[e] KWR Watercycle Research Institute, Nieuwegein, Netherlands
[f] Laboratory of Microbiology, Wageningen University, Wageningen, Netherlands

HIGHLIGHTS

- The drinking water resistome is distinct in systems with and without disinfection residuals.
- Presence/absence of disinfectant residuals is associated with antimicrobial resistance traits.
- Resistant trait distributions are strongly associated with microbial community.
- Mycobacterial populations with intrinsic resistance were only recovered from systems with disinfection residual.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 May 2020
Received in revised form 10 July 2020
Accepted 1 August 2020
Available online 5 August 2020

Editor: Damia Barcelo

Keywords:
Antibiotic resistance genes
Drinking water
Metagenomics
Disinfection

ABSTRACT

Antimicrobial resistance (AMR) in drinking water has received less attention than its counterparts in the urban water cycle. While culture-based techniques or gene-centric PCR have been used to probe the impact of treatment approaches (e.g., disinfection) on AMR in drinking water, to our knowledge there is no systematic comparison of AMR trait distribution and prevalence between disinfected and disinfectant residual-free drinking water systems. We used metagenomics to assess the associations between disinfectant residuals and AMR prevalence and its host association in full-scale drinking water distribution systems (DWDSs) with and without disinfectant residuals. While the differences in AMR profiles between DWDSs were associated with the presence or absence of disinfectant, they were also associated with overall water chemistry and more importantly with microbial community structure. AMR genes and mechanisms differentially abundant in disinfected systems were primarily associated with nontuberculous mycobacteria (NTM). Finally, evaluation of metagenome assembled genomes (MAGs) also suggests that NTM possessing AMR genes conferring intrinsic resistance to key antibiotics were prevalent in disinfected systems, whereas such NTM genomes were not detected in disinfectant residual free DWDSs. Altogether, our findings provide insights into the drinking water resistome and its association with potential opportunistic pathogens, particularly in systems with disinfectant residual.

* Corresponding author.
E-mail address: a.pinto@northeastern.edu (A.J. Pinto).

https://doi.org/10.1016/j.scitotenv.2020.141451
0048-9697/

1. Introduction

Regulation compliant drinking water can contain low concentrations of chemicals and microorganisms which stem from source water, treatment process, distribution systems, and premises plumbing (Bautista-De los Santos et al., 2016b; Nescerecka et al., 2014). While chemicals (e.g., disinfection by-products (DBP) (Richardson et al., 2007)) and microorganisms (Pinto et al., 2014; Prest et al., 2016; Vaz-Moreira et al., 2013) have been extensively studied in drinking water, our understanding of emerging biological contaminants (Pruden et al., 2006) like antibiotic resistance bacteria (ARB) and genes (ARG) is limited. In fact, drinking water can constitute a potential exposure route to microorganisms with intrinsic and acquired antibiotic resistance genes (i.e., resistome (Wright, 2007)) and it can be particularly concerning if these traits are associated with waterborne pathogens (Ashbolt et al., 2013). This is of particular concern for treatment processes, which are known to shape the drinking water microbiome (Zhang and Liu, 2019) and play a role in selecting antimicrobial resistant opportunistic pathogens (Sanganyado and Gwenzi, 2019).

Drinking water disinfection, via chlorination, remains the most widely used treatment strategy for pathogen control (Cutler and Miller, 2005). However, its unintended impacts (e.g., DBP formation) have prompted the evaluation of alternative strategies to mitigate associated risks. These range from the use of alternate disinfectants such as chloramines, to advanced oxidation processes that avoid the use of disinfectants and sometimes eliminate the need for maintaining a disinfectant residual. A disinfectant residual free approach towards production and distribution of safe drinking water is practiced in some European countries (e.g., Denmark, Netherlands, Switzerland) (Rosario-Ortiz et al., 2016). Among the range of treatment strategies to manage microbial growth in the drinking water distribution systems (DWDS), disinfection is likely to have a significant effect on AMR prevalence; either via direct (Karumathil et al., 2014; Shrivastava et al., 2004; Zhang et al., 2017) or indirect (i.e., DBP mediated) selection (Lv et al., 2014) of resistant microbial populations. For instance, sub-lethal chlorine concentrations are associated with upregulation of ARGs in pathogens (Karumathil et al., 2014; Shrivastava et al., 2004) and can promote conjugative transfer of ARGs within strains and across genera (Zhang et al., 2017). Culture dependent work has shown that the presence of chlorine (Armstrong et al., 1982) and monochloramine (Xi et al., 2009) can result in increased ARB and ARG (Shi et al., 2012) prevalence. For instance, Xu et al. showed that a terminal chlorination step in advanced ozone-biological activated carbon treatment enhanced ARG concentrations (Xu et al., 2016). Others have argued that the change in resistance profiles are likely driven by changes in bacterial community composition during drinking water treatment rather than selection for the resistance traits themselves (Jia et al., 2015).

A key challenge for contextualizing the importance of ARGs in DWS has been the inability to resolve their host-association. The presence of ARGs in pathogenic microorganisms or mobile genetic elements can present higher exposure risk, compared to their presence in non-pathogenic and/or innocuous microbes (Bengtsson-Palme and Larsson, 2015; Martínez et al., 2014). qPCR is the gold standard for ARG detection and quantitation, but requires a priori knowledge of the ARGs that need to be quantified and it lacks host association information. In contrast, high throughput sequencing (HTS) (e.g., metagenomics) facilitates the identification of a broader range of ARGs and enables insights about the functionality of detected ARGs and their host association. Multiple studies have shown high concordance between these two techniques, and show that metagenomic methods can indeed be useful for relative quantitation (Munk et al., 2017; Plaire et al., 2017; Stedtfeld et al., 2017; Willmann et al., 2015; Zhou et al., 2016). However, there are caveats associated with the use of metagenomic data. Cost of sequencing can be significantly higher than qPCR with the trade-off being the ability to capture larger number of genes. Additionally, the detection power of metagenomics is limited to medium-to-high abundance genes and rare traits require extremely deep sequencing effort (Fitzpatrick and Walsh, 2016). In contrast, qPCR can detect target genes at extremely low levels. Further, metagenomic based ARG identification is impacted by the use of varying thresholds for functional annotation (Arango-Argoty et al., 2018; Clooney et al., 2016; Li et al., 2015; Xavier et al., 2016), requiring additional curation efforts to ensure the trait assignment is robust.

In this study, we used a metagenomic approach to compare the prevalence of ARGs and their hosts in DWS with different disinfection strategies. To do this, we analyzed tap water samples from several locations in the UK and the Netherlands that either maintain (Dis) or lack disinfectant residual (NonDis) in the DWDS, respectively. Specifically, our goals were to (1) survey ARGs across a range of Dis and NonDis DWDS, (2) assess their abundance and diversity in relation to presence/absence of disinfectant residual, (3) employ *de novo* assembly and genome binning to determine the host-association of high prevalence ARGs, and (4) evaluate the relationship between the prevalence of ARG harboring microbes and/or mobile genetic elements to the presence/absence and concentration of a disinfectant residual in the DWDS.

2. Methods

2.1. Drinking water sampling and water quality analyses

A total of 39 samples were collected from 11 drinking water systems (DWS) including systems with a disinfectant residual (Dis), located in the UK ($n = 21$), and without a disinfectant residual (NonDis), located in the Netherlands ($n = 18$) in 2013 and 2015, respectively. Bulk water was collected from different countries to capture the differences in disinfection strategy. In the UK, it is common practice to maintain and regularly monitor disinfectant residual while a residual disinfectant is not maintained in the DWDS in the Netherlands. A detailed sampling protocol was described previously (Bautista-de los Santos et al., 2016a). Briefly, the drinking water faucet was flushed for 20 min, to minimize impact of stagnant water from premises plumbing. A grab sample was either collected from the tap in sterile (by autoclaving) Nalgene containers, transported to the laboratory at 4 °C and immediately filtered or filtered onsite using sterile equipment. Drinking water was filtered in triplicate through 0.2 μm Sterivex filters (SVGP01050 EMD Millipore, USA) using a peristaltic pump (Watson-Marlow 323S/D, UK) until the filter clogged or up to a 15 l volume for each filter. Water quality parameters measurements were performed as described previously (Bautista-de los Santos et al., 2016a). This included on-site measurements of temperature, pH, conductivity, and dissolved oxygen using an Orion 5 Star Meter (Thermo Fisher Scientific, USA). Total chlorine was measured on site with US EPA compliant HACH kit on a DR 2800 VIS Spectrophotometer (Hach Lange, UK). Nitrogen species (ammonia, nitrite, and nitrate) were measured in the laboratory using standard methods 4500-NH_3-F, 4500-NO_2-B, and 4500-NO_3-B respectively. Total Organic Carbon (TOC) was measured with a Shimadzu TOC-LCPH Analyzer (Shimadzu, Japan). Detailed measurements and descriptions of sampling sites can be found in supplementary materials (Table S1).

2.2. DNA extraction and shotgun sequencing

Filter membranes were aseptically removed from the Sterivex cartridge and transferred to 2 ml Lysing Matrix E tubes (SKU 116914100, MP Biomedicals, USA) and DNA extraction and purification were performed in a Maxwell® 16 DNA extraction system (Promega) using the LEV DNA kit (AS1290, Promega, USA). Briefly, 300 μl of lysing buffer and 30 μl of Proteinase K were added to the Lysing Matrix E tubes containing the filter membrane, followed by incubation at 56 °C for 20 min. Subsequently, 500 μl of chloroform:isoamyl alcohol (24:1, pH 8.0) was added to the tube and the tube was vortexed, followed by bead beating for 40 s at 6 m/s using a FastPrep 24 instrument (MP Biomedicals, USA) and centrifugation at 14,000*g* for 10 min. The aqueous phase of the supernatant was transferred to a 2 ml centrifuge tube and two more bead

beating steps were performed by replacing the aqueous phase with fresh lysing buffer (75 μl and 50 μl, respectively) prior to each bead beating step and followed by centrifugation at 14,000*g* for 10 min. Subsequent DNA purification from the aqueous phase was carried out by Maxwell LEV DNA kit. The extracted DNA was quantified using a Qubit HS dsDNA kit (Q32854, Life Technologies, UK) with a Qubit 2.0 Fluorometer (Life Technologies, UK). Negative controls consisting of reagent blanks (no input material) and filter blanks (filter membranes from unused Sterivex filters) were processed identically as the samples for DNA extraction ($n = 8$).

Libraries were prepared using the Nextera XT DNA Sample Preparation Kit (FC-131-1096, Illumina Inc.) according to the manufacturer's protocol. DNA extracts from the reagent and filter blanks were spiked with genomic DNA from an even and an uneven mock community consisting of genomic DNA from 10 organisms (Table S2) and included in the library preparation and sequencing run. All samples were cleaned up with HighPrep PCR magnetic beads (AC-60050, MagBio Inc.) according to manufacturer's instructions to remove very short fragments, evaluated for fragment size using High Sensitivity DNA Kit on Agilent Bioanalyzer (5067-4626, Agilent Inc.) and then quantified with qPCR according to Illumina guidelines. Libraries from all samples and spiked negative controls were normalized based on qPCR results and pooled in equimolar concentration. Finally, the pooled samples were quantified with Qubit HS dsDNA assay and concentrated using HighPrep PCR magnetic beads. The sequencing library was then subject to metagenomic sequencing on four lanes of Illumina HiSeq 2500 flow cell (paired end, dual indexing, 2 × 250 bp read length, Rapid Run Mode) at University of Liverpool Centre for Genomic Research (CGR, Liverpool, UK). UK and the Netherlands samples were processed identically and sequenced together.

2.3. Metagenomic data processing

Quality filtering was performed at the CGR on the raw FASTQ files by trimming reads to remove Illumina adapter sequences using Cutadapt (Compeau et al., 2011) v1.2.1. The option -O 3 was used, so the 3′ end of any reads which match the adapter sequence for 3 bp or more were trimmed. The reads were further trimmed using Trimmomatic (Bolger et al., 2014) v0.35 with a minimum Phred score of Q20. Filtered paired end reads were interleaved and co-assembled for each drinking water system using MetaSpades (Nurk et al., 2017) v3.10.1, for a total of 11 co-assemblies (6 Dis and 5 NonDis), to minimize heterogeneities between sampling sites. The resulting scaffolds were filtered by size selection and only scaffolds 500 bp or longer were used for downstream analyses. Coverage information was obtained by mapping trimmed reads from each sample against scaffolds with bwa-mem (Li and Durbin, 2009) v0.7.12 and followed by using genomecov from bedtools (Quinlan and Hall, 2010). Furthermore, contaminant analyses were incorporated into the processing workflow as described in detail by Dai et al. (2020). A scaffold was considered present in a sample if it was not detected in the negative controls or if its relative abundance in sample was greater than the negative control and coverage across the length of scaffold was more uniform in the sample as compared to the negative control. Using this approach, each scaffold was labelled as a "true" scaffold (i.e., present in the sample) or a "contaminant" scaffold. Only true scaffolds were used for downstream analyses. Open reading frames (ORFs) were identified using prodigal (Hyatt et al., 2010) v2.6.3. The predicted ORFs were then searched against *rpoB* gene specific hmm profile (pf04563) from the Pfam-A database using hmmsearch (hmmer.org). The *rpoB* normalized coverage of each true scaffold was determined by dividing its coverage in each sample by the cumulative coverage of all scaffolds containing *rpoB* genes in that sample. Scaffold level coverage was used for normalization purposes to account for unequal mapping density of reads across scaffolds.

2.3.1. Annotation of antibiotic resistance genes

To obtain a cross-validated list of ARGs, Predicted ORFs were queried using two approaches. First, ORFs were mapped against the Comprehensive Antibiotic Resistance Database (McArthur et al., 2013) (CARD, homologous model, v3.0.3) using DIAMOND (Buchfink et al., 2014) v0.8.2 to identify antibiotic resistance ontologies (ARO) with an e-value threshold of 10^{-5}. Second, the same ORFs were further interrogated using the DeepARG (Arango-Argoty et al., 2018) v2 pipeline with default parameters. The alignment results from both approaches were merged and further filtered. Specifically, only alignments with alignment lengths of 25 aa or longer, query coverage greater or equal to 70%, and percent identity greater or equal to 50% (Pearson, 2013) were retained. We considered alignments with percent identities to the CARD reference genes ranging from 50% to 80% and greater than 80% as loose and strict as recommended previously (Arango-Argoty et al., 2018). Further, we only retained loose alignments (i.e., 50–80% sequence similarity) to the CARD references if they were also annotated by DeepARG to the same antibiotic resistance genes. Finally, any ARO identified as a "mutant" was discarded. We use a combined approach to produce robust annotations that balance strict homology-based annotations, known to drastically under predict (i.e., false negatives) AMR prevalence due to heavy reliance on reference database, and loose ARG annotations, which risk the inclusion of false positives by annotating homologous genes as associated with antibiotic resistance (false positives). Therefore, a cross-validated list of ARGs was obtained by examining the predicted gene sequences from open-reading frames (ORFs) in assembled scaffolds against CARD and the DeepARG pipeline and combining results that were consistent using both approaches.

2.3.2. Microbial community membership and structure

Paired-end libraries of all samples were analyzed using phyloFlash (Gruber-Vodicka et al., 2019) v3.3b1 software with default settings as well as their reformatted version of the SILVA database (Quast et al., 2013) (Release 132), downloaded using phyloFlash_makedb.pl –remote. The FASTA output files containing SSU rRNA gene sequences assembled by phyloFlash using SPAdes and EMIRGE were combined and filtered to remove sequences less than 500 bp length. The sequences within this FASTA file were further clustered with vsearch (Rognes et al., 2016) v2.13.6 using 'cluster_fast' and with an identity threshold of 99% to generate operational taxonomic units (OTUs). Cluster centroids were then used as representative SSU rRNA gene sequences for their respective OTUs. The quality filtered paired-end reads were mapped to these representative SSU rRNA gene sequences using BBMap (Bushnell, 2015) v38.63 with the following flags: 'ambiguous = best', 'mappedonly = t', 'pairedonly = t', and 'minid = 0.97'. An OTU table of relative abundances was generated by dividing the number of mapped reads to representative OTU sequences by the total number of reads from the sample mapped to all representative OTU sequences. All representative OTUs sequences were aligned using SINA (Pruesse et al., 2012) v1.2.11. This alignment was subsequently used to construct a Maximum-likelihood phylogenetic tree with IQ-TREE (Nguyen et al., 2015) v1.6.12 using '-fast' flag. Phylogenetic tree visualization and annotation was performed on iTOL (Letunic and Bork, 2019).

2.3.3. Identifying host and mobile genetic element association of ARGs

All scaffolds containing ORFs identified as ARGs were classified with Kaiju (Menzel et al., 2016) v1.7.2 using the NCBI non-redundant database as reference (index and taxonomy downloaded from http://kaiju.binf.ku.dk/server, file: nr_euk 2019-06-25). Finally, we also analyzed the ARO containing scaffolds to determine if they were associated with mobile genetic elements by determining if they were likely plasmid or viral scaffolds. To do this, the scaffolds were aligned against a local database consisting of plasmid sequences or viral sequences obtained from NCBI RefSeq (Release 86) with an identity and query coverage threshold of 97% and 70%, respectively. Further, the scaffolds containing AROs were analyzed for viral origin by (1) aligning against

the IMG/VR database (Paez-Espino et al., 2017a) using BLAST with an identity and query coverage threshold of 97% and 70%, respectively, (2) using the k-mer based approach of VirFinder (Ren et al., 2017) with default settings, and (3) using the IMG/VR protocol for viral detection using the virus discovery pipeline as described previously (Paez-Espino et al., 2017b).

2.3.4. Genome binning, annotation, and phylogenomic analyses

Metagenome assembled genomes (MAGs) were recovered by binning using the CONCOCT (Alneberg et al., 2014) implementation in anvi'o (Eren et al., 2015). Resulting bins were refined using RefineM (Parks et al., 2017) and then manually curated in anvi'o and deduplicated by dRep (Olm et al., 2017) v2.3.2. The quality of all refined bins was assessed by CheckM (Parks et al., 2015). Detailed binning and curation procedure to obtain MAGs is described by Dai et al. (2020). Taxonomic assignment was performed based on GTDB-Tk (Parks et al., 2018) v0.1.3. The web-based Resistance Gene Identifier (RGI) tool of the CARD was used to assess the presence of ARGs in all MAGs. DNA sequences were queried to recover perfect or strict hits only (according to the RGI nomenclature) and the high quality/coverage option was selected. The coding density of a subset of MAGs attributable to AROs was calculated as the cumulative length of ORFs annotated as ARGs within MAG divided by the length of the MAG to estimate proportional length of ORFs annotated as AROs within a MAG. Reads from all samples were mapped to the MAGs using BBMap (Bushnell, 2015) v38.24 with a 90% identity threshold and the following flags: 'ambiguous = best' and 'pairedonly = t'. A MAG was regarded as detected in a sample if at least 25% of its bases were covered by at least one read in a sample. The abundance of a MAG in a sample was calculated as sample reads mapped per total number of reads in sample (in million) per genome length in kbp (RPKM).

A total of 356 reference genomes within the genera *Mycobacterium*, *Mycolicibacterium*, *Mycolicibacter*, *Mycoliciballus*, and *Mycobacteroides* were downloaded using ncbi-genome-download (https://github.com/kblin/ncbi-genome-download) with flags: '–section refseq' '–format fasta', and '–assembly-level complete'. These genera were chosen based on Gupta et al. proposed division of the *Mycobacterium genus* into five distinct monophyletic groups (Gupta et al., 2018). Next, dRep v2.3.2 was used to obtain representative reference genomes with a clustering criterion of 99% identity, which resulted in 63 representative genomes. Phylogenomic trees were generated in anvi'o by first using the program 'anvi-get-sequences-for-hmm-hits' to recover individually-aligned and concatenated 38 single-copy ribosomal protein genes from 'Bacteria_71' hmm profiles(Lee and Ponty, 2019) using five mycobacterial MAGs recovered from this study and 63 complete representative reference genomes (Table S3), and then using the program 'anvi-gen-phylogenomic-tree' that infers evolutionary associations between genomes using FastTree (Price et al., 2010) v2.1.10. Phylogenomic tree visualization and annotation was performed on iTOL (Letunic and Bork, 2019).

2.4. Data analyses and statistics

Statistical analyses were conducted in R software (R Development Core Team, 2018) and visualizations generated with ggplot2 (Wickham, 2011) package. Non parametric testing was performed with R base statistic packages using function wilcox.test() using presence/absence of disinfectant as the grouping factor. Alpha diversity metrics were obtained with sepecnumber() and diversity() functions from vegan (Oksanen et al., 2015) package 2.4-0. Beta diversity was evaluated with pairwise comparisons of *rpoB* normalized ARO abundances and SSU rRNA gene based OTU relative abundances using Bray Curtis distances in metaMDS() function from vegan. Permutational ANOVA was performed using the perm.oneway.anova() function from wPerm package (https://cran.r-project.org/web/packages/wPerm/index.html). Bray Curtis dissimilarities in resistome were correlated with the Euclidean distances of environmental parameters using Spearman's correlation method implemented in bioenv() function from vegan. As a complement, a parsimonious dbRDA model of resistome abundance was developed by stepwise regression constrained by environmental variables based on Akaike Information Criteria (AIC) using capscale() and ordistep() from the vegan package. Environmental parameters were standardized using deconstand() from vegan. Mantel testing using Spearman's correlation between Bray Curtis distances of relative abundances of taxa and *rpoB* normalized ARO were performed with mantel() function within vegan.

3. Results and discussion

To investigate the prevalence of antibiotic resistance genes (ARGs) as a function of disinfectant residual strategy (i.e., disinfectant residual present (Dis) vs disinfectant residual free (NonDis)), we generated 39 shotgun metagenomes containing an average of 7.4×10^6 paired-end reads per sample after quality control (Table S4) and an average metagenome assembly size of 434.4 Mbp. Our metagenomes correspond to different full-scale drinking water systems (DWS) in the UK and the Netherlands, each with their unique source water, treatment system configuration, and distribution system. Thus, these DWS inherently harbor different drinking water microbiomes, irrespective of the sampling time-frame. Our goal was to determine whether the presence/absence of disinfectant residuals was associated with antimicrobial resistance and its host association and whether this association is discernible in light of these underlying inevitable confounding variables - that can be addressed through statistical analyses and a balanced study design. We achieved a balanced study design by incorporating a similar number of samples between the two groups (Dis or NonDis) and a DWS-specific co-assembly approach, which should capture DWS specific communities and their resistomes. We relied on consensus ARG annotation approach that combined homology to the CARD database with DeepARG based annotations to identify loose (50–80% sequence similarity) and strictly (>80% sequence similarity) annotated AMR traits. Further, we examine the results through the group wise comparisons of the relative abundance of ARGs between Dis and NonDis systems. Therefore, our results don't necessarily reflect quantitative exposures, considering the fact that the typical absolute abundance (i.e., concentration) of microbes in Dis systems are significantly lower than that in NonDis systems. Rather, the focus of this study was to understand the ecology of AMR in Dis and NonDis systems and assess the role of water chemistry and microbial community on ARG distribution and their host associations in systems that maintain or do not maintain a disinfectant residual.

3.1. Antimicrobial resistance traits are prevalent in drinking water systems irrespective of presence/absence of disinfectant residual

We identified a total of 807 significant hits comprising 134 AROs using a combination of CARD-based homology search and DeepARG in Dis and NonDis groups (Fig. S1A, Table S5), with a median query coverage centered at 100% (Fig. S1B). For both the Dis and NonDis categories, 80% of the annotations were within 50–80% sequence similarity. The relative abundance of an ARO was determined as that of ARO containing scaffold coverage normalized by *rpoB* coverage in the sample. The *rpoB* gene, unlike the 16S rRNA gene, is a single copy gene that can be used as a molecular marker to estimate the bacteriological diversity of a sample and normalize the abundance of functional genes of interest (Case et al., 2007; Dahllof et al., 2000; Vos et al., 2012).

A greater number of scaffolds from the Dis samples were annotated with AROs as compared to the NonDis samples, with the mean *rpoB* normalized coverage of these AROs being significantly higher ($p < 0.001$) in the Dis ($2.7 \pm 10 \times 10^{-2}$) compared to the NonDis ($2.5 \pm 11 \times 10^{-3}$) samples. Of the 134 AROs detected in this study, a total of 83 and 94 AROs were present in Dis and NonDis samples, respectively with 40

and 51 AROs exclusive to Dis and NonDis, respectively (Fig. 1A, Table S6). This is likely due to higher microbial diversity in NonDis systems (Bautista-De los Santos et al., 2016b; Dai et al., 2020) which may result in higher diversity of ARGs. Of the shared AROs ($n = 43$), a higher proportion of AROs exhibited greater relative abundance in Dis systems ($n = 28$) as compared to NonDis ($n = 15$) (Fig. 1B). Of the shared AROs, only three were frequently (i.e., detected in ≥50% of samples in each strategy) detected in samples within each strategy (Fig. 1C) and showed varying levels of similarity with their corresponding reference AROs. The frequent and proportionally abundant AROs in Dis systems consisted of *MexF*, *adeF*, and *golS* all of these are associated antibiotic efflux. *MexF* is a multidrug transporter of the RND efflux family associated with bacteria within the genera *Pseudomonas* (Fetar et al., 2011). It has been shown that *P. fluorescens* in biofilms responded to sublethal sodium hypochlorite concentrations by increased transcription of MexEF pumps, among other genes (Lipus et al., 2019). Further, *adeF* is a membrane protein for multidrug efflux in the RND family, associated with *Acinetobacter baumannii* (Coyne et al., 2010) and *golS* is a regulator for the expression of a multidrug efflux pump in response to silver, also within the RND family (Checa et al., 2007).

Among the AROs categorized as exclusively present in either strategy type, only three and one ARO were frequently detected (i.e., ≥ 50% detection rate) in Dis and NonDis systems, respectively (Fig. 1D). These consisted of *efpA*, *RbpA*, and *mfpA* in Dis systems (Table S6). *efpA* is a transporter associated with *Mycobacterium tuberculosis* that mediates resistance to rifamycin and isoniazid, which is upregulated in exposure to oxidative stress (Wilson et al., 1999) and may be elicited by disinfectant residual in Dis systems. *RbpA* is a RNA polymerase binding protein which confers resistance to rifampin and is also associated with *Mycobacterium tuberculosis*. *mfpA* is a *qnr* homolog and a pentapeptide repeat protein that confers resistance to fluoroquinolones via target protection in *Mycolicibacterium smegmatis*. Di Cesare et al. reported that the *qnrS* harboring microbes demonstrated enhanced survival and post-disinfection growth, whereas this was not the case for alternate disinfectants (e.g., peracetic acid, UV) (Di Cesare et al., 2016). Tetracycline efflux pump, *tetA*, was the only exclusive and frequently detected ARO in NonDis systems. Among these genes, *RbpA* consistently showed homology to corresponding genes in the reference database, while the other genes were highly divergent from their corresponding CARD reference sequences (Fig. 1D).

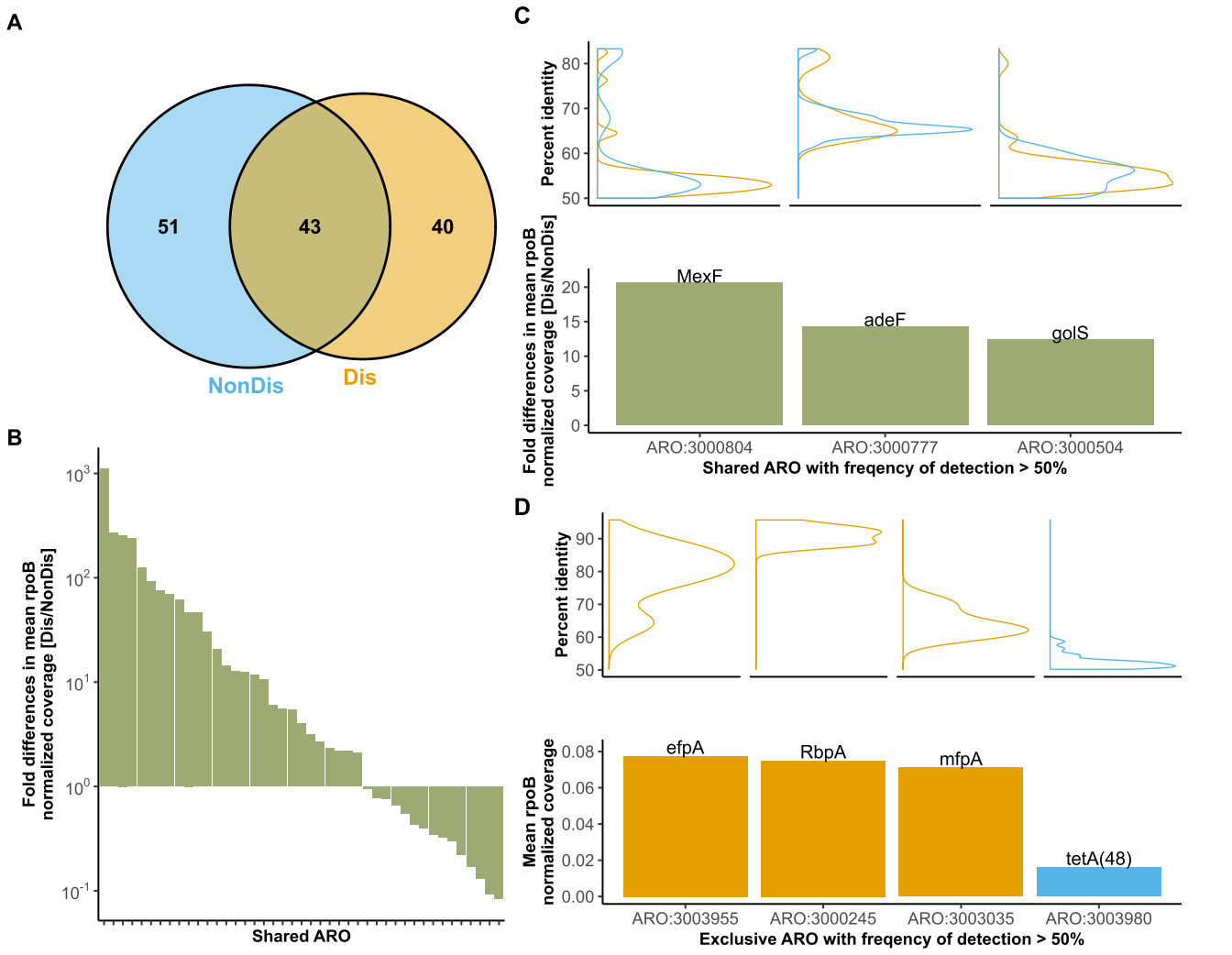


Fig. 1. (A) Venn diagram of annotated antibiotic resistance ontologies (AROs) for NonDis samples in blue and Dis samples in orange. (B) Fold difference in mean *rpoB* normalized AROs shared between Dis and NonDis systems indicated that shared AROs were enriched in Dis systems. (C) Fold difference in mean *rpoB* coverage of shared AROs with frequency of detection >50% also indicated enrichment in Dis samples and the distribution of percent identities to reference sequences in the CARD database are shown on the top. (D) Mean *rpoB* coverage of exclusive AROs with frequency of detection >50% for Dis systems in orange and NonDis systems in blue shows a more abundant and frequent resistome in Dis systems compared to NonDis systems and the distribution of percent identities to reference sequences in the CARD database are shown on the top.

3.2. Efflux and target alteration are dominant antibiotic resistance mechanisms in disinfected and non-disinfected systems

The AROs detected in this study were associated with six different resistance mechanisms. Specifically (1) efflux, (2) target alteration, (3) enzymatic inactivation, (4) target protection, and (5) target replacement (Fig. S2A). The efflux resistance mechanism was the most abundant group in both Dis and NonDis samples, significantly higher ($p < 0.001$) for Dis compared to NonDis systems (Fig. 2). The detected efflux genes include those associated with ATP binding cassette (ABC), resistance nodulation cell division (RND), small multidrug resistance (SMR), and major facilitator super family antibiotic efflux (MFS) gene families for both Dis and NonDis systems. Efflux pumps are widespread and present in all organisms and exhibit multiple ecological roles (e.g., cell-cell communication, homeostasis, extrusion of abiotic factors (Blanco et al., 2016)), therefore their prevalence and abundance in databases is likely high, as well as the potential for them to be annotated. Some genes in this category will only mediate resistance if overexpressed, while others have basal intrinsic resistance (Lupo et al., 2012). Genes involved in enzymatic inactivation of antibiotics, (i.e., acetyltransferases, beta-lactamases, phosphotransferases) accounted for 32% and 26% of the AROs found in Dis and NonDis samples, respectively with diverse *AAC*, Bla_{OXA}, and *APH* associated genes detected. AROs involved in enzymatic inactivation and target protection were also significantly different between strategies ($p < 0.001$) and exhibited higher relative abundance in Dis systems. The significance for target protection category was largely driven by the absence of target protection genes in NonDis systems. In contrast, genes involved in target replacement mechanism were significantly more abundant ($p < 0.001$) in NonDis samples. The genes associated with this mechanism were exclusively *dfr* and *sul* related genes which are both related to folic metabolism and are widespread in the environment and are predominantly located on plasmids or transposons (Frye and Jackson, 2013). The inactivation of ARGs such as *sul1* from municipal wastewater is most effective with chlorination, followed by UV and ozonation (Zhuang et al., 2015) which could explain the infrequency of detection of these genes in Dis systems.

Mechanisms of resistance are important to ascertain target pathways and in fact disinfectants often share mechanisms of action with antibiotics (Chapman, 2003). The aforementioned mechanisms of resistance were associated with 16 drug classes (Figs. S2B, S3). The order of abundance of drug classes was relatively similar between strategies. For Dis systems, multidrug resistance was the most abundant category (49.75% of AROs), followed by aminoglycoside (16.5%), beta-lactam (6.4%), rifamycin (5.4%), with the remaining drug classes accounting for less than 5% of the AROs individually. For NonDis systems, the order of drug classes was as follows: multidrug (39.16%), aminoglycoside (12.82%), peptide (9.80%), diaminopyrimidine (7.46%), beta-lactam (5.83%), bacitracin (5.36%), with the remaining drug classes accounting for less than 5% of the AROs individually. Consistent with the observation of prevalence of efflux mechanism of resistance regardless of strategy, multidrug determinants constituted majority of the AROs.

Group-wise comparisons indicated statistically significant differences between Dis and NonDis groups for resistance traits associated with aminoglycoside ($p < 0.01$), diaminopyrimidine ($p < 0.001$), fosfomycin ($p < 0.001$), lincosamide ($p < 0.001$), multidrug (<0.001), sulfonamide ($p < 0.01$) (Fig. 2). The cumulative *rpoB* normalized coverage of aminoglycoside resistance was higher in Dis compared to NonDis samples. This drug class consisted of acetyltransferases and phosphotransferases that elicit antibiotic inactivation that were detected in both Dis and NonDis systems, and included a transcriptional activator of a two-component system that mediates resistance through efflux in NonDis samples (*kdpE*). Previous studies have documented aminoglycoside as a dominant determinant in aquatic environments including drinking water metagenomes (Ma et al., 2017) and *E. coli* isolates (Zhang et al., 2009). Diaminopyrimidine resistance was significantly higher in NonDis samples as opposed to Dis samples. This determinant exclusively consisted of dihydrofolate reductases which mediate resistance through target replacement, and was more diverse in NonDis systems. Diaminopyrimidine (trimethoprim) is widely distributed in natural waters (Danner et al., 2019), it has been shown to readily react with chlorine (Dodd and Huang, 2007), but in contrast it has less potential to react with UV (Kim et al., 2015). This could suggest that the antibiotic selective pressure is reduced in Dis systems while

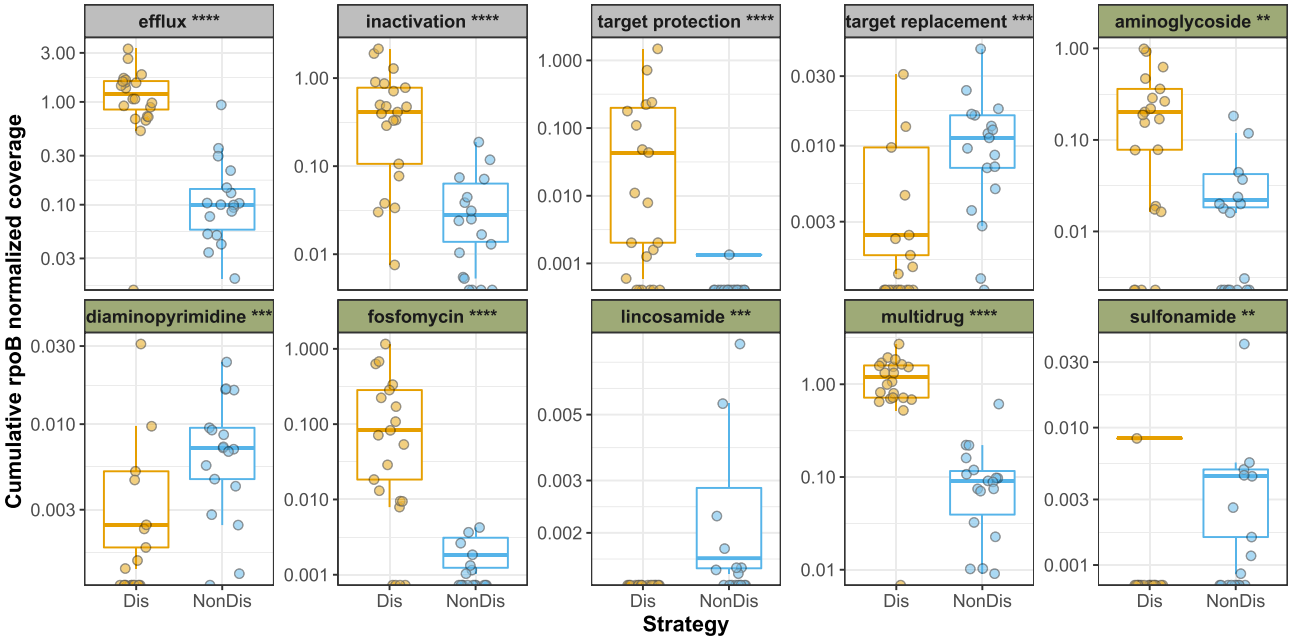


Fig. 2. Comparison of cumulative *rpoB* normalized coverage between Dis and NonDis treatment strategies (orange and blue, respectively) for statistically significant mechanisms of resistance in grey panels and statistically significant drug classes in green panels.

remaining potentially unchanged in NonDis systems at treatments steps prior to distribution. Consequently, the microbial communities -and their associated resistance traits- which thrive in these conditions, have the potential to persist downstream (e.g., in the distribution system). Genes associated with fosfomycin drug class were more abundant and frequently detected in Dis strategy. Fosfomycin is a broad-spectrum bactericidal drug that inhibits bacterial cell wall biogenesis by inactivating the enzyme UDP-N acetylglucosamine-3-enolpyruvyltransferase, also known as MurA (Collignon et al., 2016; Thompson et al., 2015). The observed determinants within this class consisted exclusively of Fos genes and primarily FosX.

3.3. Antimicrobial resistance traits while not conserved across drinking water systems, indicated association with presence/absence of disinfectant residual

While the relative abundance of AMR genes and associated mechanisms drug classes demonstrated significant differences between Dis and NonDis systems, there were very few universally conserved AMR genes within each disinfection strategy. Of the 83 AROs detected in Dis systems, only 11 AROs were present in 50% or more Dis samples, whereas only 11 out of 94 AROs that were detected in NonDis systems had a frequency of detection greater or equal to 50%. Further, the traits that were widely distributed across DWS within each strategy were largely limited to efflux related genes (Dis: 6/11, NonDis: 8/11). This indicates that while AMR traits are found in DWSs from both strategies, the core resistome across all sampled DWSs or even within each disinfection strategy group was highly limited in diversity. This could either be a result of the variable impact of factors influencing the prevalence of AMR traits between individual DWSs or AMR traits may not be under direct selection due to the presence or absence of disinfection. Thus, we evaluated the potential association between other water chemistry parameters on the resistome structure and membership and compared that with the effect exhibited by presence/absence of disinfectant.

Bray-Curtis distance clustering of samples based on *rpoB* normalized coverage of AROs in each sample showed that the structure of the resistance traits clustered based on the presence or absence of a disinfectant residual (Fig. 3A). This suggests that despite the limited nature of the core resistome, the abundance and membership of a resistome was associated with the presence or absence of disinfectant residual. Permutational ANOVA analyses indicated a significant difference in ARO composition based on grouping of samples with respect to presence/absence of disinfectant residual ($p < 0.001$). We further evaluated the contribution of other measured chemical parameters on the distribution of AMR traits through a combination of (1) correlations of the Bray-Curtis dissimilarities between samples estimated using resistome data and corresponding Euclidean distances estimated using measured environmental (i.e., temperature) and water chemistry (i.e., pH, conductivity, dissolved oxygen (DO), TOC, nitrate, and chlorine concentration) parameters and (2) distance based redundancy analysis (dbRDA) followed by (3) variance partitioning analyses. The best subset of environmental and water chemistry parameters found were chlorine and temperature, which moderately and significantly correlated with resistome distribution (Spearman's $\rho = 0.54$, $p < 0.001$) Additionally, a statistically significant model ($p < 0.001$) was achieved through a stepwise regression procedure based on AIC applied to constrained and null dbRDA models. The final model included temperature, pH, conductivity, total chlorine, and TOC, and excluded other water quality parameters to achieve a parsimonious model from backward and forward selection. The first two principal components explained 71% of the variation (Fig. 3B, Table S7) with an adjusted R^2 of 0.28. Since conductivity exhibited a high variance inflation factor (2.05). Conductivity measurements could be a proxy for source water type or treatment process. Usually high conductivity waters are recovered from ground water sources and changes in conductivity can be observed after water softening, for example. Conductivity, although generally higher in NonDis systems, reflects high and low measurements for samples in both strategies. This observation coupled with the multicollinearity exhibited with other variables, suggests that this factor is neither the only nor the best to explain variation in resistance traits prevalence. Therefore, variance partitioning analysis was performed after removing this variable from the model to allocate the variation of *rpoB* normalized abundances of resistance traits among four explanatory variables, i.e., temperature, pH, total chlorine and TOC, as well as obtaining an overall variation effect. The environmental variable explaining the highest individual fraction was pH (adjusted $R^2 = 0.07$), followed by temperature (adjusted $R^2 = 0.04$), chlorine (adjusted $R^2 = 0.04$), and TOC (adjusted $R^2 = 0.03$). Chlorine concentration (i.e., presence/absence of disinfectant, residual disinfection strategy), although not the main explanatory variable is important in combination with the other explanatory variables (adjusted $R^2 = 0.10$). Thus, while the presence/absence and concentration of chlorine was not the sole explanatory variable, it was consistently identified as significant in all considered models. Further, approximately 76% of the variation in AMR trait distribution could not

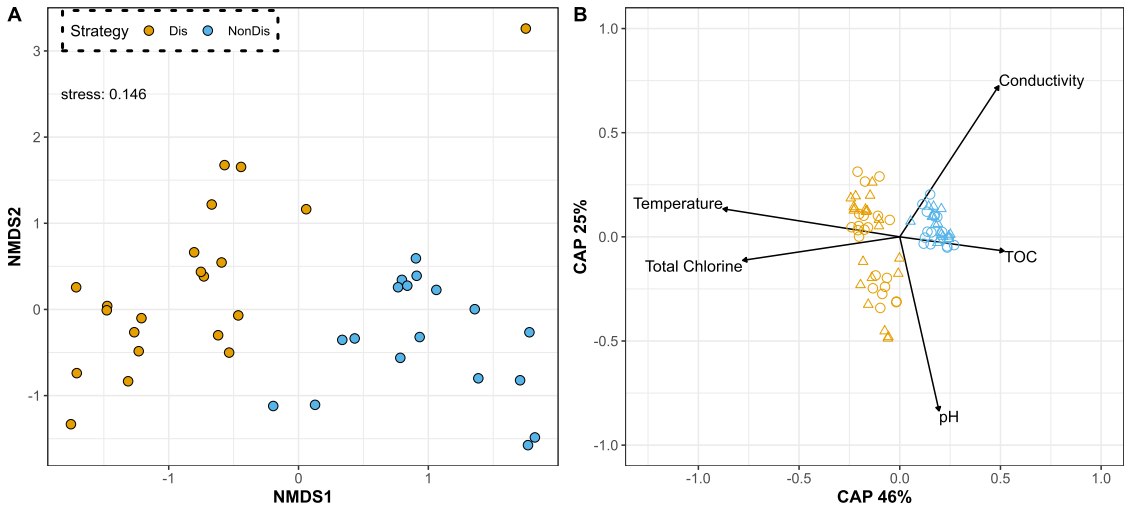


Fig. 3. (A) Ordination plot using Bray-Curtis distance describing AROs' *rpoB* normalized coverage dissimilarity colored by disinfection strategy (Orange: Dis and Blue: NonDis). The clustering suggests that the resistome composition of Dis and NonDis systems is different. (B) Distance based redundancy analyses (dbRDA) plot showing the Bray-Curtis distance of resistome data, with Dis samples in orange and NonDis samples in blue, constrained by the Euclidean distances of its respective environmental parameters as vectors.

be allocated to any of the measured environmental and chemical parameters (Table S8). This clearly indicates that AMR trait distribution in DWSs is impacted by several other factors that were not measured by this study. These unaccounted factors could range from source waters, treatment process types, DWDS characteristics, etc.

3.4. Microbial community structure is a strong determinant of AMR trait distributions

In order to address additional factors that might impact AMR trait distributions, we (1) estimated the bacterial community membership and structure in each sample to evaluate its association with AMR trait distribution and (2) also evaluated the host association of detected AROs through a combination of taxonomic annotation of ARO contigs and binning of contigs into metagenome assembled genomes (MAGs). PhyloFlash analyses resulted in the reconstruction of 481 bacterial 16S rRNA gene sequences. As expected, NonDis samples were more diverse as compared to samples from Dis systems. Taxa within the phylum Proteobacteria exhibited the highest relative abundance (RA, considering the bacterial fraction only) across in Dis and NonDis systems and their abundance ranged from 4.7–99.9% in Dis samples, with mean RA of 70.6% and 10.9–62.2% in NonDis samples with mean RA of 43.5%. Actinobacteria and Cyanobacteria, were more abundant in Dis samples as compared to NonDis samples (mean RA in Dis = 12.8% vs in NonDis = 8.6%, and mean RA in Dis = 7.8%, vs in NonDis = 1.02%, respectively). Further, Patescibacteria and Omnitrophicaeota were more abundant to NonDis than Dis systems (mean RA: Dis = 8.3%, NonDis = 29%, and Dis = 1.9%, NonDis = 5.3%, respectively) (Fig. 4A, Table S9). To determine the association between bacterial community structure and AMR trait prevalence, we performed Mantel test with distance matrices constructed using pairwise Bray Curtis distances between samples that were estimated using relative abundance of assembled 16S rRNA genes and *rpoB* normalized ARO abundance. The results showed that the AMR distributions were highly correlated with bacterial community structure across all drinking water samples (Mantel $R = 0.75$, <0.001) (Fig. 4B), with stronger correlations for Dis (Mantel $R = 0.84$, <0.001) (Fig. S4A) as compared to NonDis systems (Mantel $R = 0.76$, <0.001) (Fig. S4B). This suggests that community membership is strongly associated with AMR trait distribution and has an important role in shaping the resistome, which is consistent with a previous report (Jia et al., 2015).

We further evaluated the taxonomic affiliation of all scaffolds to assess whether ARO containing contigs exhibiting significantly higher relative abundance in Dis compared to NonDis systems exhibited a conserved taxonomic signal (Table S10). Only two ARO containing low abundance scaffolds were annotated as being of viral origin using either

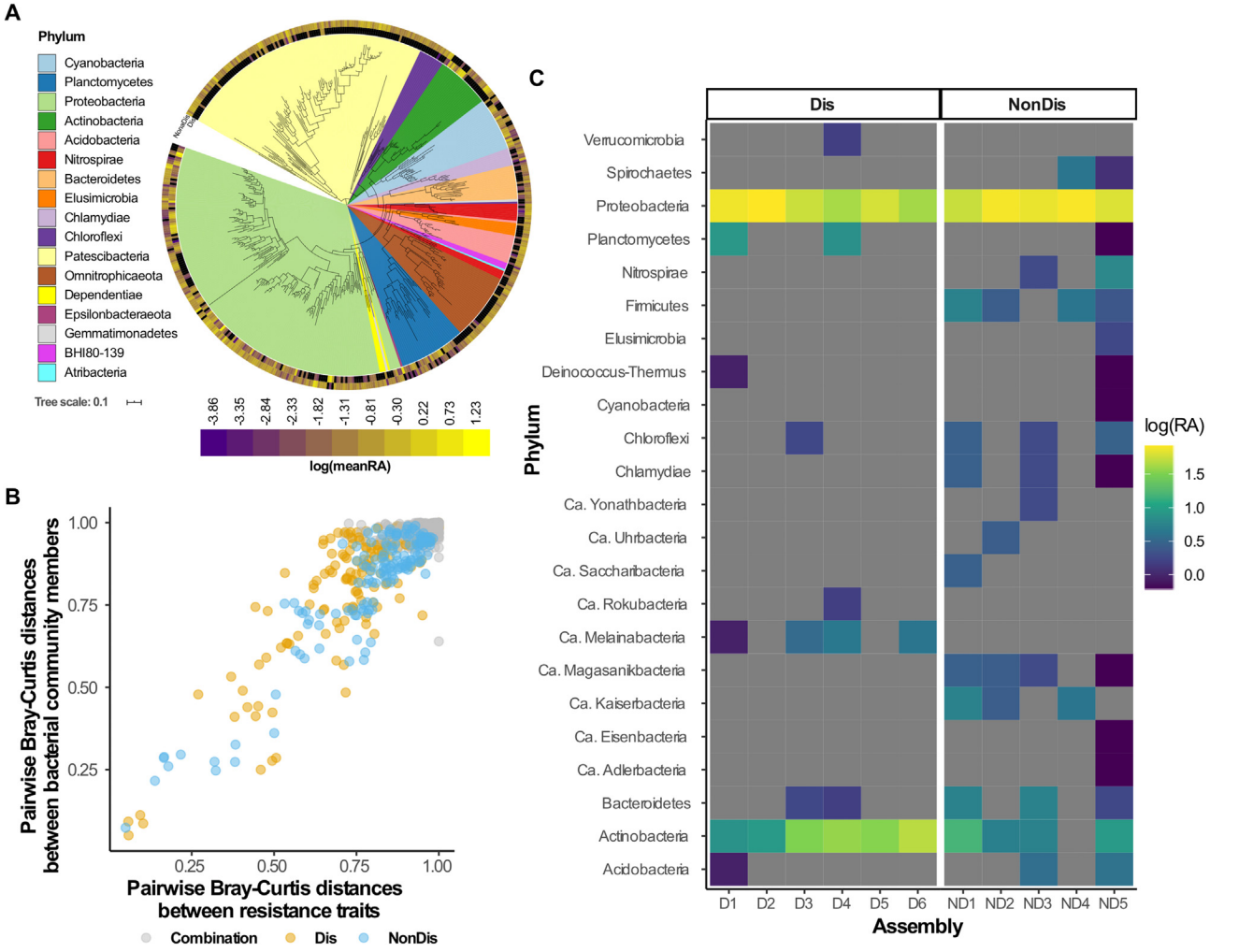


Fig. 4. (A) The phylogenetic tree of 481 reconstructed SSU sequences with mean relative abundance in Dis and NonDis samples with the outer circle indicating relative abundance in Dis (inner ring) and NonDis (outer ring) systems. (B) Global correlation between Bray-Curtis pairwise distances of community RA and Bray-Curtis pairwise distances of *rpoB* normalized coverage of AROs. Pairwise comparison between Dis samples, NonDis and Dis vs NonDis are shown in orange, blue, and grey, respectively. (C) ARO carrying scaffolds classified at phylum level, Dis samples on the left and NonDis samples on the right. Log10 transformations of scaffold relative abundance are used for visualization purposes.

of the approaches that rely on the IMG/VR database, while no ARO containing scaffolds were annotated as viral scaffolds using NCBI or VirFinder. While our sampling protocol was not designed for viral recovery, our results suggest that particle (size ≥0.2 μm) and host-associated mobile genetic elements (i.e., viruses and plasmids) do not contribute significantly to AMR traits in drinking water systems. A majority of the ARO containing contigs were classified as Proteobacteria at the phylum level in Dis and NonDis systems (Fig. 4C, Table S11). The taxonomic classification at phylum level of ARO containing scaffolds indicated high abundance of Proteobacteria, regardless of the disinfection strategy. The proportion of ARO containing scaffolds classified as Proteobacteria in the systems ranged from 42 to 91% in Dis, and 58–88% in NonDis. The ARO containing contigs from NonDis systems were associated with more candidate phyla (e.g., Patescibacteria) and in general were more taxonomically diverse than Dis systems; this was consistent with the observation of higher diversity of the microbial communities as a whole. While ARO containing scaffolds within NonDis systems exhibited high taxonomic diversity, those from Dis systems were largely confined to Proteobacteria and Actinobacteria (7–54%). Furthermore, we observed that the approximately 50–100% of the Actinobacterial ARO associated scaffolds were classified to the genus *Mycobacterium* depending on sample, whereas no ARO containing scaffolds in the NonDis systems were classified as *Mycobacterium*.

3.5. Genome-resolved metagenomics indicates that ARO harboring mycobacteria are responsive to chlorine concentrations

The 787 scaffolds containing 807 AROs were binned into 65 of the 112 MAGs assembled from Dis (n = 62) and NonDis systems (n = 50) (Dai et al., 2020), with majority of the ARO containing MAGs assembled from Dis systems (n = 48/65). MAGs from the Dis system primarily belonged to Proteobacteria (n = 33/48), Actinobacteriota (n = 5), Planctomyceota (n = 5), and Cyanobacteriota (n = 4), while those from the NonDis systems belonged to Proteobacteria (n = 14/17), Acidobacteriota (n = 2), and Actinobacteriota (n = 1) (Fig. 5A), according to the GTDB-Tk taxonomy. The primary resistance mechanism associated with these MAGs were efflux which was detected in present in 57/65 MAGs, followed by inactivation present in 23/65 MAGs (Fig. 5A). All Actinobacteriotal MAGs from Dis systems were associated with mycobacteria and harbored multiple AROs. In contrast, no mycobacterial MAGs were assembled from NonDis systems. Specifically, mycobacteria MAGs were recovered from D3, D4, and D5 metagenome assemblies. The complete statistics associated with each mycobacterial MAG are presented in Table 1. Genome taxonomy database annotation indicated that Bins 4 and 13 were closely related to *Mycobacterium llatzerense* with 99.4 and 97.0% sequence average nucleotide identity to the same draft genome (GCF_000746215) which was assembled from plant associated samples. Bin 3 exhibited 99.4% sequence similarity to a mycobacterial isolate sampled from tap water sample in Germany (GCF_002013415), while Bins 9 and 10 did not have any closely related mycobacterial genomes at the species level. Phylogenetic placement of the bins based on multigene alignment of ribosomal proteins indicated that all assembled bins belonged to environmental and rapidly growing mycobacteria (Fig. 5B). While Bin 3 placed in the rapid growing *Mycobacterium chelonae-absceccus* complex (i.e., *Mycobacteroides*), Bins 4 and 13 were closely phylogenetically clustered with the *fortuitum-vaccae* clades. Although Bins 9 and 10 were distantly related from most mycobacterial reference genomes, they were closely associated with mycobacterial isolates from hydrocarbon contaminated and river estuary sediment, respectively. All five assembled MAGs contained at least one ARO, with Bins 10 and 13 containing three and two, respectively (Table 1). Specifically, the identified AROs were *RbpA*, *mtrA*, and *murA*. In two of these bins, *RbpA* (rifamycin resistance) and *mtrA* (macrolide and penam resistance) co-occur. *M. llatzarense* has been isolated from DWDS and is described as a NTM that resists amoebal phagocytosis

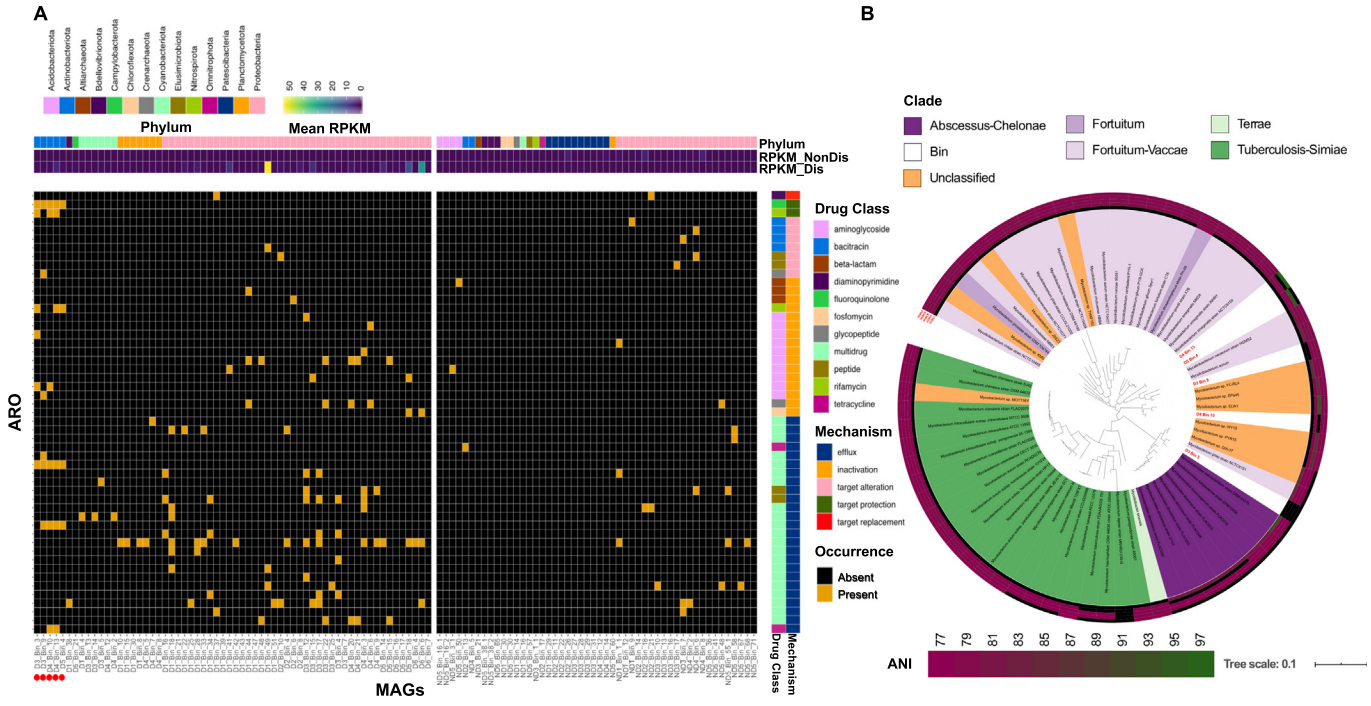


Fig. 5. (A) Heatmap depicting the presence or absence of resistance traits (ARO) in metagenome assembled genomes (MAGs), from Dis and NonDis samples in right and left panel, respectively. The top bars of the heatmap denote Phylum, RPKM in NonDis systems, and RPKM in Dis systems from top to bottom. The bars to the right of the heatmap show drug classes and mechanisms from left to right. The red circles correspond to selected MAGs shown in panel B. (B) Phylogenomic tree contextualizing selected MAGs (in bold red font) and representative Mycobacterium reference genomes. The labels are colored by clades. The heatmap corresponds to ANI values between selected MAGs (D5_Bin_D4, D4_Bin_13, D4_Bin_10, D3_Bin_9, and D3_Bin_3, respective to outer most ring) and representative reference genomes.

Table 1
Statistics associated with recovered bins along with taxonomy and resistance profile information.

MAG	Completeness (%)	Redundancy (%)	Size (Mbp)	Closest reference genome	ANI	ARG	ARG family	Mechanism	Drug class
Bin3	69.57	3.56	4.07	Mycobacterium spp. (GCF_002013415.1)	99.4	RbpA	RNA polymerase binding protein	Target protection	Rifamycin
Bin4	63.07	3.99	5.58	*Mycobacterium llatzerense* (GCF_000746215.1)	96.3	mtrA	Resistance-nodulation-cell division	Efflux	Macrolide, penam
Bin9	71.45	1.16	4.18	*Mycobacterium spp.*(N/A)	N/A	mtrA	Resistance-nodulation-cell division	Efflux	Macrolide, penam
Bin10	98.49	0.00	5.76	*Mycobacterium spp.* (N/A)	N/A	RbpA	RNA polymerase binding protein	Target protection	Rifamycin
						murA	murA transferase	Target alteration	Fosfomycin
						mtrA	Resistance-nodulation-cell division	Efflux	Macrolide, penam
Bin13	90.02	0.97	4.86	*Mycobacterium llatzerense* (GCF_000746215.1)	97	mtrA	Resistance-nodulation-cell division	Efflux	Macrolide, penam
						RbpA	RNA polymerase binding protein	Target protection	Rifamycin

and survives disinfection (Delafont et al., 2017; Lalucat et al., 2019). Similarly, Bin3 associated with the *M. chelonae-absceccus* complex also demonstrated the presence of gene conferring rifamycin resistance. The enrichment of mycobacterial MAGs with AMR traits in Dis systems can be attributed to several factors. For instance, mycobacteria are known to harbor intrinsic antibiotic resistance (Fedrizzi et al., 2017) and their colonization, survival, and persistence in DWDS and their resistance to disinfectants at low nutrient concentrations has been well documented (Dantec et al., 2002; Falkinham III et al., 2015; Loret and Dumoutier, 2019). Further, intracellular colonization of mycobacteria in protozoa may allow for survival and proliferation of these genera (Thomas et al., 2008) and sheltering within biofilms can also enhance their survival (Liu et al., 2016; Loret and Dumoutier, 2019; Revetta et al., 2016; Zhang et al., 2018).

4. Conclusions

In conclusion, in this study we identified and characterized the AMR traits of tap water from DWS with and without disinfectant residual, Dis and NonDis, respectively. We observed that presence/absence of disinfectant plays a significant role in AMR trait prevalence, composition, and abundance. Although both systems have a diverse and uneven AMR trait distribution, they exhibited higher relative abundance in Dis systems compared to NonDis systems. Further, while the ARO distribution was significantly associated with presence/absence of chlorine, a large proportion of the variance in ARO trait distribution was unexplained by water chemistry alone. Further, the precise ARO traits detected in each system varied significantly suggesting that AMR traits themselves may not be under selection in drinking water systems. Rather, consistent with previous reports, bacterial community structure was the strongest determinant of ARO presence/absence and distribution. In particular, the ARO host association indicated significant and consistent differences across Dis and NonDis system suggesting that disinfection mediated selection pressures likely play a role at the community level with indirect implications for AMR prevalence. For instance, we recovered ARO containing nontuberculous mycobacterial MAGs only from Dis systems. An important direction for future research is coupling quantitative measurements with functional metagenomics to determine absolute concentrations of ARGs and assess if the associated hosts are active and AMR traits are functional. This information will be vital to translate our characterization of the AMR traits in drinking water systems to quantitative microbiological risk assessment.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.141451.

Data availability

DNA-sequencing data are available from SRA under BioProject ID PRJNA533545.

CRediT authorship contribution statement

Maria Sevillano: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Zihan Dai:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - review & editing. **Szymon Calus:** Investigation, Methodology, Writing - review & editing. **QM Bautista-de los Santos:** Investigation, Methodology, Writing - review & editing. **A. Murat Eren:** Resources, Software, Writing - review & editing. **Paul W.J.J. van der Wielen:** Investigation, Methodology, Resources, Writing - review & editing. **Umer Z. Ijaz:** Formal analysis, Funding acquisition, Resources, Software, Supervision, Writing - review & editing. **Ameet J. Pinto:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

Declaration of competing interest

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

Acknowledgements

MS was supported by the College of Engineering at Northeastern University. This work was funded by Engineering and Physical Science Research Council (EP/M016811/1) and the National Science Foundation (NSF-CBET 1749530). UZI is supported by NERC Independent Research Fellowship (NERC NE/L011956/1).

References

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. Nat. Methods 11 (11), 1144–1146. https://doi.org/10.1038/nmeth.3103.

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L.S., Vikesland, P., Zhang, L., 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome 6 (1), 1–15. https://doi.org/10.1186/s40168-018-0401-z.

Armstrong, J.L., Calomiris, J.O.N.J., Seidler, R.J., 1982. Selection of antibiotic-resistant standard plate count bacteria during water treatment. Appl. Environ. Microbiol. 44 (2), 308–316.

Ashbolt, N.J., Amézquita, A., Backhaus, T., Borriello, P., Brandt, K.K., 2013. Human health risk assessment (HHRA) for environmental development and transfer of antibiotic resistance. Environ. Health Perspect. 121 (9), 993–1002. https://doi.org/10.1289/ehp.1206316.

Bautista-de los Santos, Q.M., Schroeder, J.L., Blakemore, O., Moses, J., Haffey, M., Sloan, W., Pinto, A.J., 2016a. The impact of sampling, PCR, and sequencing replication on discerning changes in drinking water bacterial community over diurnal time-scales. Water Res. 90, 216–224. https://doi.org/10.1016/j.watres.2015.12.010.

Bautista-De los Santos, Q.M., Schroeder, J.L., Sevillano-Rivera, M.C., Sungthong, R., Ijaz, U.Z., Sloan, W.T., Pinto, A.J., 2016b. Emerging investigators series: microbial communities in full-scale drinking water distribution systems-a meta-analysis. Environmental

Science: Water Research and Technology 2, 631–644. https://doi.org/10.1039/c6ew00030d.
Bengtsson-Palme, J., Larsson, D.G.J., 2015. Antibiotic resistance genes in the environment: prioritizing risks. Nat. Rev. Microbiol. 13 (6), 396. https://doi.org/10.1038/nrmicro3399-c1.
Blanco, P., Hernando-Amado, S., Reales-Calderon, J., Corona, F., Lira, F., Alcalde-Rico, M., Bernardini, A., Sanchez, M., Martinez, J., 2016. Bacterial multidrug efflux pumps: much more than antibiotic resistance determinants. Microorganisms 4 (1), 14. https://doi.org/10.3390/microorganisms4010014.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12 (1), 59–60. https://doi.org/10.1038/nmeth.3176.
Bushnell, B., 2015. BBMap short-read aligner, and other bioinformatics tools. http://sourceforge.net/projects/bbmap/.
Case, R.J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W.F., Kjelleberg, S., 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Applied Environmental Microbiology 73 (1), 278–288. https://doi.org/10.1128/AEM.01177-06.
Chapman, J.S., 2003. Disinfectant resistance mechanisms, cross-resistance, and co-resistance. Int. Biodeterior. Biodegradation 51 (4), 271–276. https://doi.org/10.1016/S0964-8305(03)00044-1.
Checa, S.K., Espariz, M., Audero, M.E.P., Botta, P.E., Spinelli, S.V., Soncini, F.C., 2007. Bacterial sensing of and resistance to gold salts. Mol. Microbiol. 63 (5), 1307–1318. https://doi.org/10.1111/j.1365-2958.2007.05590.x.
Clooney, A.G., Fouhy, F., Sleator, R.D., O'Driscoll, A., Stanton, C., Cotter, P.D., Claesson, M.J., 2016. Comparing apples and oranges?: next generation sequencing and its impact on microbiome analysis. PLoS One 11 (2), 1–16. https://doi.org/10.1371/journal.pone.0148028.
Collignon, P.C., Conly, J.M., Andremont, A., McEwen, S.A., Aidara-Kane, A., Griffin, P.M., Agerso, Y., Dang Ninh, T., Donado-Godoy, P., Fedorka-Cray, P., Fernandez, H., Galas, M., Irwin, R., Karp, B., Matar, G., McDermott, P., Mitema, E., Reid-Smith, R., Scott, H.M., Singh, R., Dewaal, C.S., Stelling, J., Toleman, M., Watanabe, H., Woo, G.J., 2016. World Health Organization ranking of antimicrobials according to their importance in human medicine: a critical step for developing risk management strategies to control antimicrobial resistance from food animal production. Clin. Infect. Dis. 63 (8), 1087–1093. https://doi.org/10.1093/cid/ciw475.
Compeau, P.E.C., Pevzner, P.A., Tesler, G., Papoutsoglou, G., Roscito, J.G., Dahl, A., Myers, G., Winkler, S., Pippel, M., Sameith, K., Hiller, M., Francoijs, K.-J., Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., Nagarajan, N., Pop, M., Abbas, M.M., Malluhi, Q.M., Balakrishnan, P., Lantz, H., Notredame, C., Soler, L., Hjerde, E., Klopp, C., Dominguez Del Angel, V., Bocs, S., Binzer-Panchal, M., Gibrat, J.-F., Leskosek, B.L., Sterck, L., Vinnere Pettersson, O., Bouri, L., Capella-Gutierrez, S., Amselem, J., Vlasova, A., Buermans, H.P.J., den Dunnen, J.T., Davis, J.M.G., Bolton, R.E., Garrett, J., Ekblom, R., Wolf, J.B.W., Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17 (1), 10–12. https://doi.org/10.14806/ej.17.1.200.
Coyne, S., Rosenfeld, N., Lambert, T., Courvalin, P., Périchon, B., 2010. Overexpression of resistance-nodulation-cell division pump AdeFGH confers multidrug resistance in Acinetobacter baumannii. Antimicrob. Agents Chemother. 54 (10), 4389–4393. https://doi.org/10.1128/AAC.00155-10.
Cutler, D., Miller, G., 2005. The role of public health improvements in health advances: the twentieth-century United States. Demography 42 (1), 1–22. https://doi.org/10.1353/dem.2005.0002.
Dahllof, I., Baillie, H., Kjelleberg, S., 2000. rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. Applied Environmental Microbiology 66 (8), 3376–3380. https://doi.org/10.1128/AEM.66.8.3376-3380.2000.
Dai, Z., Sevillano-Rivera, M., Calus, S., Bautista-de los Santos, Q.M., Eren, A.M., van der Wielen, P.W.J.J., Ijaz, U.Z., Pinto, A.J., 2020. Disinfection exhibits systematic impacts on the drinking water microbiome. Microbiome 8, 42. https://doi.org/10.1186/s40168-020-00813-0.
Danner, M.C., Robertson, A., Behrends, V., Reiss, J., 2019. Antibiotic pollution in surface fresh waters: occurrence and effects. Sci. Total Environ. 664, 793–804. https://doi.org/10.1016/j.scitotenv.2019.01.406.
Dantec, C. Le, Duguet, J., Montiel, A., Dumoutier, N., Dubrou, S., Vincent, V., 2002. Chlorine disinfection of atypical mycobacteria isolated from a water distribution system. Applied Environmental Microbiology 68 (3), 1025–1032. https://doi.org/10.1128/AEM.68.3.1025.
Delafont, V., Samba-louaka, A., Cambau, E., Bouchon, D., 2017. Mycobacterium llatzerense, a waterborne mycobacterium, that resists phagocytosis by Acanthamoeba castellanii. Sci. Rep. 7, 46270. https://doi.org/10.1038/srep46270.
Di Cesare, A., Fontaneto, D., Doppelbauer, J., Corno, G., 2016. Fitness and recovery of bacterial communities and antibiotic resistance genes in urban wastewaters exposed to classical disinfection treatments. Environ. Sci. Technol. 50 (18), 10153–10161. https://doi.org/10.1021/acs.est.6b02268.
Dodd, M.C., Huang, C.H., 2007. Aqueous chlorination of the antibacterial agent trimethoprim: reaction kinetics and pathways. Water Res. 41 (3), 647–655. https://doi.org/10.1016/j.watres.2006.10.029.
Eren, A.M., Esen, C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T.O., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3, e1319. https://doi.org/10.7717/peerj.1319.
Falkinham III, J.O., Hilborn, E.D., Arduino, M.J., Pruden, A., Edwards, M.A., 2015. Epidemiology and ecology of opportunistic premise plumbing pathogens. Environ. Health Perspect. 123 (8), 749–758. https://doi.org/10.1289/ehp.1408692.
Fedrizzi, T., Meehan, C.J., Grottola, A., Giacobazzi, E., Serpini, G.F., Tagliazucchi, S., Fabio, A., Bettua, C., Bertorelli, R., Sanctis, V. De, Rumpianesi, F., Pecorari, M., Jousson, O., Tortoli, E., Segata, N., 2017. Genomic characterization of nontuberculous mycobacteria. Sci. Rep. 7, 45258. https://doi.org/10.1038/srep45258.
Fetar, H., Gilmour, C., Klinoski, R., Daigle, D.M., Dean, C.R., Poole, K., 2011. mexEF-oprN multidrug efflux operon of Pseudomonas aeruginosa: regulation by the MexT activator in response to nitrosative stress and chloramphenicol. Antimicrob. Agents Chemother. 55 (2), 508–514. https://doi.org/10.1128/AAC.00830-10.
Fitzpatrick, D., Walsh, F., 2016. Antibiotic resistance genes across a wide variety of metagenomes. FEMS Microbiol. Ecol. 92 (2), fiv168. https://doi.org/10.1093/femsec/fiv168.
Frye, J.G., Jackson, C.R., 2013. Genetic mechanisms of antimicrobial resistance identified in Salmonella enterica, Escherichia coli, and Enteroccocus spp. isolated from U.S. food animals. Front. Microbiol. 4. https://doi.org/10.3389/fmicb.2013.00135.
Gruber-Vodicka, H.R., Seah, B.K.B., Pruesse, E., 2019. phyloFlash – rapid SSU rRNA profiling and targeted assembly from metagenomes. bioRxiv https://doi.org/10.1101/521922.
Gupta, R.S., Lo, B., Son, J., 2018. Phylogenomics and comparative genomic studies robustly support division of the genus mycobacterium into an emended genus mycobacterium and four novel genera. Front. Microbiol. 9, 67. https://doi.org/10.3389/fmicb.2018.00067.
Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11 (1). https://doi.org/10.1186/1471-2105-11-119.
Jia, S., Shi, P., Hu, Q., Li, B., Zhang, T., Zhang, X.X., 2015. Bacterial community shift drives antibiotic resistance promotion during drinking water chlorination. Environ. Sci. Technol. 49 (20), 12271–12279. https://doi.org/10.1021/acs.est.5b03521.
Karumathil, D.P., Yin, H., Kollanoor-johny, A., 2014. Effect of chlorine exposure on the survival and antibiotic gene expression of multidrug resistant Acinetobacter baumannii in water. Int. J. Environ. Res. Public Health 11 (2), 1844–1854. https://doi.org/10.3390/ijerph110201844.
Kim, H.Y., Kim, T.H., Yu, S., 2015. Photolytic degradation of sulfamethoxazole and trimethoprim using UV-A, UV-C and vacuum-UV (VUV). J. Environ. Sci. Health A 50 (3), 292–300. https://doi.org/10.1080/10934529.2015.981118.
Lalucat, J., Gomila, M., Ramirez, A., Gasco, J., 2019. Mycobacterium llatzerense sp. nov., a facultatively autotrophic, hydrogen-oxidizing bacterium isolated from haemodialysis water. Int. J. Syst. Evol. Microbiol. 58 (12), 2769–2773. https://doi.org/10.1099/ijs.0.65857-0.
Lee, M.D., Ponty, Y., 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35 (20), 4162–4164. https://doi.org/10.1093/bioinformatics/btz188.
Letunic, I., Bork, P., 2019. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47 (W1), W256–W259. https://doi.org/10.1093/nar/gkz239.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25 (14), 1754 Understanding, Monitoring, and Controlling Biofilm Growth in Drinking Water Distribution Systems.1760. https://doi.org/10.1093/bioinformatics/btp324.
Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J.M., Zhang, T., 2015. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME J. 9 (11), 2490–2502. https://doi.org/10.1038/ismej.2015.59.
Lipus, D., Vikram, A., Gulliver, D., Bibby, K., 2019. Upregulation of peroxide scavenging enzymes and multidrug efflux proteins highlight an active sodium hypochlorite response in Pseudomonas fluorescens biofilms. Biofouling 35 (3), 329–339. https://doi.org/10.1080/08927014.2019.1605357.
Liu, S., Gunawan, C., Barraud, N., Rice, S.A., Harry, E.J., Amal, R., 2016. Understanding, monitoring, and controlling biofilm growth in drinking water distribution systems. Environ. Sci. Technol. 50 (17), 8954–8976. https://doi.org/10.1021/acs.est.6b00835.
Loret, J., Dumoutier, N., 2019. Non-tuberculous mycobacteria in drinking water systems: a review of prevalence data and control means. International Journal of Hygene and Environmental Health 222 (4), 628–634. https://doi.org/10.1016/j.ijheh.2019.01.002.
Lupo, A., Coyne, S., Berendonk, T.U., 2012. Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. Front. Microbiol. 3. https://doi.org/10.3389/fmicb.2012.00018.
Lv, L., Jiang, T., Zhang, S., Yu, X., 2014. Exposure to mutagenic disinfection byproducts leads to increase of antibiotic resistance in Pseudomonas aeruginosa. Environ. Sci. Technol. 48 (14), 8188–8195. https://doi.org/10.1021/es501646n.
Ma, L., Li, B., Jiang, X.T., Wang, Y.L., Xia, Y., Li, A.D., Zhang, T., 2017. Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. Microbiome 5 (1), 154. https://doi.org/10.1186/s40168-017-0369-0.
Martínez, J.L., Coque, T.M., Baquero, F., 2014. What is a resistance gene ? Ranking risk in resistomes. Nature Reviews in Microbiology 13 (3), 116–123. https://doi.org/10.1038/nrmicro3399.
McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M.R., O'Brien, J.S., Pawlowski, A.C., Piddock, L.J.V., Spanogiannopoulos, P., Sutherland, A.D., Tang, I., Taylor, P.L., Thaker, M., Wang, W., Yan, M., Yu, T., Wright, G.D., 2013. The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. 57 (7), 3348–3357. https://doi.org/10.1128/aac.00419-13.
Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7 (1), 1–9. https://doi.org/10.1038/ncomms11257.
Munk, P., Andersen, V.D., Knegt, L. De, Jensen, M.S., Knudsen, B.E., Lukjancenko, O., Mordhorst, H., Clasen, J., Agersø, Y., Folkesson, A., Pamp, J., 2017. A sampling and metagenomic sequencing-based methodology for monitoring antimicrobial resistance in swine herds. Journal of Antimicromial Chemotherapy 72 (2), 385–392. https://doi.org/10.1093/jac/dkw415.
Nescerecka, A., Rubulis, J., Vital, M., Juhna, T., Hammes, F., 2014. Biological instability in a chlorinated drinking water distribution network. PLoS One 9 (5), 1–11. https://doi.org/10.1371/journal.pone.0096354.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274. https://doi.org/10.1093/molbev/msu300.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: a new versatile metagenomic assembler. Genome Res. 27 (5), 824–834. https://doi.org/10.1101/gr.213959.116.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. (2015) Vegan: community ecology package. http://CRAN. R-project. org/package= vegan.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 11 (12), 2864–2868. https://doi.org/10.1038/ismej.2017.126.

Paez-Espino, D., Chen, I.M.A., Palaniappan, K., Ratner, A., Chu, K., Szeto, E., Pillay, M., Huang, J., Markowitz, V.M., Nielsen, T., Huntemann, M., Reddy, T.B.K., Pavlopoulos, G.A., Sullivan, M.B., Campbell, B.J., Chen, F., McMahon, K., Hallam, S.J., Denef, V., Cavicchioli, R., Caffrey, S.M., Streit, W.R., Webster, J., Handley, K.M., Salekdeh, G.H., Tsesmetzis, N., Setubal, J.C., Pope, P.B., Liu, W.T., Rivers, A.R., Ivanova, N.N., Kyrpides, N.C., 2017a. IMG/VR: a database of cultured and uncultured DNA viruses and retroviruses. Nucleic Acids Res. 45, D457–D465. https://doi.org/10.1093/nar/gkw1030.

Paez-Espino, D., Pavlopoulos, G.A., Ivanova, N.N., Kyrpides, N.C., 2017b. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. Nat. Protoc. 12 (8), 1673–1682. https://doi.org/10.1038/nprot.2017.063.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25 (7), 1043–1055. https://doi.org/10.1101/gr.186072.114.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., Tyson, G.W., 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2 (11), 1533–1542. https://doi.org/10.1038/s41564-017-0012-7.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P., Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36 (10), 996–1004. https://doi.org/10.1038/nbt.4229.

Pearson, W.R., 2013. An introduction to sequence similarity ("homology") searching. Curr. Protoc. Bioinformatics 42 (1), 3.1.3–3.1.8. https://doi.org/10.1002/0471250953.bi0301s42.

Pinto, A.J., Schroeder, J., Lunn, M., Sloan, W.R.L., 2014. Spatial-temporal survey and occupancy-abundance modeling to predict bacterial community dynamics in the drinking water microbiome. Mbio 5 (3), e01135–14. https://doi.org/10.1128/mBio.01135-14.

Plaire, D., Puaud, S., Elalouf, M., 2017. Comparative analysis of the sensitivity of metagenomic sequencing and PCR to detect a biowarfare simulant (Bacillus atrophaeus) in soil samples. PLoS One 12 (5). https://doi.org/10.1371/journal.pone.0177112.

Prest, E.I., Weissbrodt, D.G., Hammes, F., Van Loosdrecht, M.C.M., Vrouwenvelder, J.S., 2016. Long-term bacterial dynamics in a full-scale drinking water distribution system. PLoS One 11 (10). https://doi.org/10.1371/journal.pone.0164445.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5 (3). https://doi.org/10.1371/journal.pone.0009490.

Pruden, A.M.Y., Pei, R., Storteboom, H., Carlson, K.H., 2006. Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. Environ. Sci. Technol. 40 (23), 7445–7450. https://doi.org/10.1021/es060413l.

Pruesse, E., Peplies, J., Glöckner, F.O., 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28 (14), 1823–1829. https://doi.org/10.1093/bioinformatics/bts252.

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 (D1), D590–D596. https://doi.org/10.1093/nar/gks1219.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26 (6), 841–842. https://doi.org/10.1093/bioinformatics/btq033.

R Development Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing https://www.R-project.org/.

Ren, J., Ahlgren, N.A., Lu, Y.Y., Fuhrman, J.A., Sun, F., 2017. VirFinder: a novel k -mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome 5 (1), 69. https://doi.org/10.1186/s40168-017-0283-5.

Revetta, R.P., Gerke, T.L., Domingo, J.W.S., Ashbolt, N.J., 2016. Changes in bacterial composition of biofilm in a metropolitan drinking water distribution system. J. Appl. Microbiol. 121 (1), 294–305. https://doi.org/10.1111/jam.13150.

Richardson, S.D., Plewa, M.J., Wagner, E.D., Schoeny, R., DeMarini, D.M., 2007. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. Mutation Research/Reviews in Mutation Research 636 (1–3), 178–242. https://doi.org/10.1016/j.mrrev.2007.09.001.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584.

Rosario-Ortiz, F., Rose, J., Speight, V., Gunten, U.v., Schnoor, J., 2016. How do you like your tap water? Science 351 (6276), 912–914. https://doi.org/10.1126/science.aaf0953.

Sanganyado, E., Gwenzi, W., 2019. Antibiotic resistance in drinking water systems: occurrence, removal, and human health risks. Sci. Total Environ. 669, 785–797. https://doi.org/10.1016/j.scitotenv.2019.03.162.

Shi, P., Jia, S., Zhang, X., Zhang, T., Cheng, S., Li, A., 2012. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. Water Res. 47 (1), 111–120. https://doi.org/10.1016/j.watres.2012.09.046.

Shrivastava, R., Upreti, R.K., Jain, S.R., Prasad, K.N., Seth, P.K., Chaturvedi, U.C., 2004. Suboptimal chlorine treatment of drinking water leads to selection of multidrug-resistant Pseudomonas aeruginosa. Ecotoxicol. Environ. Saf. 58 (2), 277–283. https://doi.org/10.1016/S0147-6513(03)00107-6.

Stedtfeld, R.D., Stedtfeld, T.M., Fader, K.A., Williams, M.R., Bhaduri, P., Quensen, J., Zacharewski, T.R., Tiedje, J.M., Hashsham, S.A., 2017. TCDD influences reservoir of antibiotic resistance genes in murine gut microbiome. FEMS Microbiol. Ecol. 93 (5), fix058. https://doi.org/10.1093/femsec/fix058.

Thomas, V., Loret, J., Jousset, M., Greub, G., 2008. Biodiversity of amoebae and amoebae-resisting bacteria in a drinking water treatment plant. Environ. Microbiol. 10 (10), 2728–2745. https://doi.org/10.1111/j.1462-2920.2008.01693.x.

Thompson, M.K., Keithly, M.E., Sulikowski, G.A., Armstrong, R.N., 2015. Diversity in fosfomycin resistance proteins. Perspectives in Science 4, 17–23. https://doi.org/10.1016/j.pisc.2014.12.004.

Vaz-Moreira, I., Egas, C., Nunes, O.C., Manaia, C.M., 2013. Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. FEMS Microbiol. Ecol. 83 (2), 361–374. https://doi.org/10.1111/1574-6941.12002.

Vos, M., Quince, C., Pijl, A.S., de Hollander, M., Kowalchuk, G.A., 2012. A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. PLoS One 7 (2). https://doi.org/10.1371/journal.pone.0030600.

Wickham, H., 2011. Ggplot2. WIREs Computational Statistics 3, 180–185. https://doi.org/10.1002/wics.147.

Willmann, M., El-hadidi, M., Huson, D.H., Autenrieth, I.B., Peter, S., Weidenmaier, C., 2015. Antibiotic selection pressure determination through sequence-based metagenomics. Antimicrob. Agents Chemother. 59 (12), 7335–7345. https://doi.org/10.1128/AAC.01504-15.Address.

Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P., Rane, S., Brown, P.O., Schoolnik, G.K., 1999. Exploring drug-induced alterations in gene expression in mycobacterium tuberculosis by microarray hybridization. Proc. Natl. Acad. Sci. 96 (22), 12833–12838. https://doi.org/10.1073/pnas.96.22.12833.

Wright, G.D., 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat. Rev. Microbiol. 5 (3), 175–186. https://doi.org/10.1038/nrmicro1614.

Xavier, B.B., Das, A.J., Cochrane, G., De Ganck, S., Kumar-Singh, S., Aarestrup, F.M., Goossens, H., Malhotra-Kumar, S., 2016. Consolidating and exploring antibiotic resistance gene data resources. J. Clin. Microbiol. 54 (4), 851–859. https://doi.org/10.1128/JCM.02717-15.

Xi, C., Zhang, Y., Marrs, C.F., Ye, W., Simon, C., Foxman, B., Nriagu, J., 2009. Prevalence of antibiotic resistance in drinking water treatment and distribution systems. Applied Environmental Microbiology 75 (17), 5714–5718. https://doi.org/10.1128/AEM.00382-09.

Xu, L., Ouyang, W., Qian, Y., Su, C., Su, J., Chen, H., 2016. High-throughput pro fi ling of antibiotic resistance genes in drinking water treatment plants and distribution systems. Environ. Pollut. 213, 119–126. https://doi.org/10.1016/j.envpol.2016.02.013.

Zhang, Y., Liu, W.T., 2019. The application of molecular tools to study the drinking water microbiome–current understanding and future needs. Crit. Rev. Environ. Sci. Technol. 49 (13), 1188–1235. https://doi.org/10.1080/10643389.2019.1571351.

Zhang, X.X., Zhang, T., Fang, H.H.P., 2009. Antibiotic resistance genes in water environment. Appl. Microbiol. Biotechnol. 82 (3), 397–414. https://doi.org/10.1007/s00253-008-1829-z.

Zhang, Y., Gu, A.Z., He, M., Li, D., Chen, J., 2017. Subinhibitory concentrations of disinfectants promote the horizontal transfer of multidrug resistance genes within and across genera. Environ. Sci. Technol. 51 (1), 570–580. https://doi.org/10.1021/acs.est.6b03132.

Zhang, J., Li, W., Chen, J., Qi, W., Wang, F., Zhou, Y., 2018. Impact of biofilm formation and detachment on the transmission of bacterial antibiotic resistance in drinking water distribution systems. Chemosphere 203, 368–380. https://doi.org/10.1016/j.chemosphere.2018.03.143.

Zhou, Y., Wylie, K.M., Feghaly, E. El, Mihindukulasuriya, K.A., Elward, A., Haslam, D.B., Storch, G.A., Weinstock, M., 2016. Metagenomic approach for identification of the pathogens associated with diarrhea in stool specimens. J. Clin. Microbiol. 54 (2), 368–375. https://doi.org/10.1128/JCM.01965-15.

Zhuang, Y., Ren, H., Geng, J., Zhang, Yingying, Zhang, Yan, Ding, L., Xu, K., 2015. Inactivation of antibiotic resistance genes in municipal wastewater by chlorination, ultraviolet, and ozonation disinfection. Environ. Sci. Pollut. Res. 22 (9), 7037–7044. https://doi.org/10.1007/s11356-014-3919-z.