Abundance, diversity, and host assignment of total, intracellular, and extracellular antibiotic resistance genes in riverbed sediments

AUTHORS: A.S. Deshpande,[a] N.L. Fahrenfeld[b*]

[a] Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ 08901

[b] Civil and Environmental Engineering, Rutgers University, 500 Bartholomew Rd., Piscataway, NJ 08854; nfahrenf@rutgers.edu

Abstract

Human health risk assessment for environmental antibiotic resistant microbes requires not only quantifying the abundance of antibiotic resistance genes (ARGs) in environmental matrices, but also understanding their hosts and genetic context. Further, differentiating ARGs in intracellular and extracellular DNA (iDNA and eDNA) fractions may help refine our understanding of ARG transferability. The objectives of this study were to understand the (O1) abundance and diversity of extracellular, intracellular, and total ARGs along a land use gradient and (O2) impact of bioinformatics pipeline on the assignment of putative hosts for the ARGs observed in the different DNA fractions. Sediment samples were collected along a land use gradient in the Raritan River, New Jersey, USA. DNA was extracted to separate eDNA and iDNA and qPCR was performed for select ARGs and the 16S rRNA gene. Shotgun metagenomic sequencing was performed on DNA extracts for the different DNA fractions. ARG hosts were assigned via two different bioinformatic pipelines: network analysis of raw reads versus assembly. Results of the two pipelines were compared to evaluate their performance in terms of number and diversity of linkages and accuracy of *in silico* matrix spike host assignments. No differences were observed

in the 16S rRNA gene normalized *sul*1 concentration between the DNA fractions. The overall microbial community structure was more similar for iDNA and total DNA compared to eDNA and generally clustered by sampling site. ARGs associated with mobile genetic elements increased in iDNA for the downstream sites. Regarding host assignment, the raw reads pipeline via network analysis identified 247 ARG hosts as compared to 53 hosts identified by assembly pipeline. Other comparisons between the pipelines were made including ARG assignment to taxa with pathogens and practical considerations regarding processing time.

Keywords: ARG, *sul*1, metagenomic sequencing, assembly, network analysis, eDNA, iDNA

1. Introduction

Increasing rates of antibiotic resistant infections, including some community-acquired (CDC, 2019), have raised concerns about environmental sources of antibiotic resistant microbes. The abundance of environmental antibiotic resistance genes (ARGs) is well-documented [e.g., (Hong et al., 2013)] in comparison to our understanding of their hosts and genetic contexts. Understanding the genetic context of an ARG can provide insight into (1) the hazard posed by the host organism and (2) the potential mechanisms and rates of gene transfer (Ashbolt et al., 2013; Martinez et al., 2015; Vikesland et al., 2017). For example, a chromosomally encoded ARG is expected to transfer less readily than a plasmid encoded ARG (Mazel and Davies, 1999; Rowe-Magnus and Mazel, 2002). Further, extracellular ARGs propagate by transformation whereas intracellular ARGs can propagate by transduction or conjugation, indicating a fundamental difference in the fate and transmissivity of these two fractions of DNA (Mao et al., 2014; Zhang et al., 2013). However, there is no consensus for how to perform host assignment

for ARGs in environmental metagenomic studies and it is not clear how extracellular DNA (eDNA) in environmental samples can impact the results.

A growing body of literature reports ARG relative abundances and diversity through metagenomic sequencing of riverbed sediments (Chen et al., 2019; Chen et al., 2020; Jiang et al., 2018). Given that DNA can persist in sediment for months to millennia due to adsorption to sediment particles (Deere et al., 1996; Lorenz and Wackernagel, 1987; Turner et al., 2015; Willerslev et al., 2014), there is motivation for understanding the extracellular ARG loads in this matrix (Alawi et al., 2014; Chen et al., 2018; Corinaldesi et al., 2005; Guo et al., 2018; Mao et al., 2014; Zhang et al., 2018). A potentially important source of extracellular ARGs to the water environment is wastewater effluents from facilities using disinfectants that disrupt cell membranes but not destroy the released DNA (Dodd, 2012). While there are reports of eDNA metagenomes from sewage sludge (Calderon-Franco et al., 2021; Zhou et al., 2019) and aquaculture farm sediment (Chen et al., 2018), there is a paucity of data differentiating DNA fractions in riverbed sediments towards identifying ARGs and assigning their hosts. Of particular interest is how the potentially spatially variable concentrations of eDNA may affect ARG profiles and host assignments in environmental metagenomes.

Assigning a gene to a putative host in environmental metagenomic studies can been done by read- or assembly-based pipelines (Scholz et al., 2012). Read-based pipelines often apply network analysis, relying upon correlation between ARG and host 16S rRNA gene abundances. Assembly of raw reads increases the confidence in gene prediction (Loman et al., 2013) and allows for the construction of novel genomes and genomic elements (Howe et al., 2014). However, non-uniform sequencing depths for different organisms in a sample or high community complexity can cause the assembly to be highly fragmented (Breitwieser et al., 2019;

66 Xie et al., 2010) or result in low mapping percentage of reads to the assembly (Vollmers et al.,
67 2017).

68 The objectives of this study were to understand the (O1) spatial variability in abundance and
69 diversity of extracellular, intracellular, and total ARGs along a land use gradient and (O2) impact
70 of bioinformatics pipeline on the assignment of putative hosts for the ARGs observed in different
71 DNA fractions. For (O1), qPCR for select ARGs and shotgun metagenomic sequencing was
72 performed on eDNA, intracellular DNA (iDNA), and total DNA from riverbed sediments. For
73 (O2), host assignments based on network analysis of raw reads and assembly were compared to
74 understand their performance in terms of number and diversity of ARG-host linkages. *In silico*
75 matrix spikes were performed to provide a measure of the accuracy of host assignments. Overall,
76 these results can help inform the choice of methods applied in future studies of antibiotic
77 resistance in riverbed sediments towards better defining the potential hazard posed by this matrix
78 for environmental antibiotic resistance.

79 **2. Material and Methods**

80 **2.1 Sampling**

81 Riverbed sediment samples were collected from five public access sites along the Raritan River
82 in New Jersey, USA (Fig. 1, Table S1), coinciding with long term civic science monitoring for
83 macroinvertebrates and/or pathogens. The sites have varying influence by wastewater effluent,
84 urban activities, agricultural activities, and tides (Table S2). Composite sediment core samples (5
85 cores/site) were collected using a soil probe (JMC, Clements Associates Inc., Newton, IA) from
86 the upper 8-9 cm of riverbed sediment. Field blanks, which consisted of autoclaved deionized
87 water, were left open for the duration of the sampling. All the samples were stored and

transported to lab in coolers on ice and then immediately stored at 4°C until processing. All processing occurred within one week of sample collection. Water and sediment quality parameters were measured for each site and details are provided in *SI 1*.

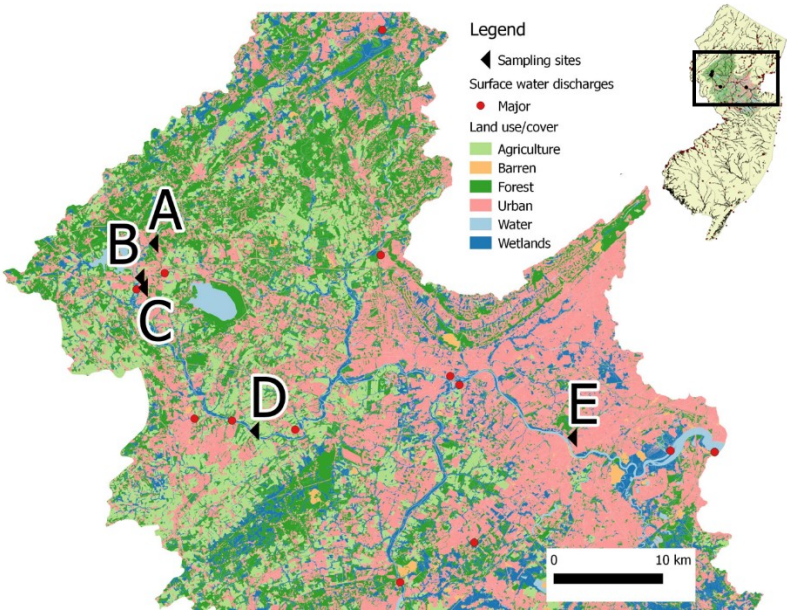


Fig. 1: *Land use map of study catchment with sampling sites (black triangles) and major (red) surface water discharges. Insert map shows location of the study area within the state of New Jersey, USA.*

2.2 Total, eDNA, and iDNA extraction

Field samples were prepared for DNA extraction by combining composite samples in the laboratory. Composite sediment samples were sieved (2000 µm) and homogenized for each site. Total DNA extraction was performed on the homogenized sediment (0.5 g) using a FastDNA® Spin kit for soil (MP Biomedicals, Solon, OH, USA) following the manufacturer's instructions. DNA extracts were resuspended in 100 µL of DES buffer and stored at -20°C until analysis.

eDNA was separated from iDNA according to previously developed protocols (Corinaldesi et al., 2005; Mao et al., 2014) with few minor modifications. Based on previously developed protocols, sodium phosphate buffer and polyvinyl polypyrrolidone (PVPP) was added to 1.5 g of sediment (wet weight). The mixture was mixed on a shaker table (250 rpm, 5 min) followed by centrifugation at 7100 × g for 25 minutes. The supernatant containing eDNA was removed from the tubes with care to not disturb the pellet containing iDNA. Sodium phosphate buffer was again added to the pellet, and the process was repeated, as above. Following separation, the pellets were stored at -20°C until iDNA extraction. The supernatants were combined and filtered (Cyclopore, Whatman). eDNA concentration was performed on the filtrate using cetyl trimethyl ammonium bromide (CTAB, 1% w/v) buffer. A phenol chloroform isoamyl alcohol extraction was performed followed by an alcohol precipitation during which GlycoBlue™ coprecipitant (Invitrogen™, Thermo Fischer Scientific) was added during the alcohol precipitation steps to enhance precipitation and make the pellet visible (details in *SI 2*). iDNA extraction was performed on preserved pellets using FastDNA® Spin kit for soil (MP Biomedicals, Solon, OH, USA) following the manufacturer's instructions.

Matrix spikes were performed using positive controls for eDNA and iDNA on sample splits to estimate the extraction recovery for each DNA fraction. For the eDNA positive control, pUC19 plasmids containing the *van*Z gene were used. *Escherichia coli* (*E. coli*) DH5a cells were used as positive control for the iDNA extraction. The pre-quantified plasmids carrying the *van*Z gene and *E. coli* cells were spiked into the samples prior to eDNA extraction for each sampling site (details in *SI 3*).

2.3 qPCR and shotgun metagenomic sequencing

qPCR was performed to quantify ARGs encoding for sulfonamide resistance (*sul*1, *sul*2) (Pei et al., 2006), tetracycline resistance (*tet*(G)) (Aminov et al., 2002), and the 16S rRNA gene (Muyzer et al., 1993) in all samples, matrix spikes, and field blanks. These genes were selected because tetracycline and sulfonamide are the most widely used antibiotics (Luo et al., 2010). Also, *sul*1 is detected frequently in the environment (Deng et al., 2022; Jiang et al., 2021; Miao et al., 2021; Munir et al., 2011) and it is included in the list of genetic determinants that act as indicators of antibiotic resistance in the environment (Berendonk et al., 2015). Additionally, qPCR was performed for *van*Z (Jensen et al., 1998), which served as the eDNA matrix spike. Reaction recipes, thermocycling conditions, controls, and calibration curves are described in *SI 4* and *Table S5*.

To understand the diversity of ARGs and their hosts in different DNA fractions, DNA extracts of sediment samples from each site (eDNA, iDNA and total DNA) in addition to one replicate each of iDNA and total DNA from Site E were submitted for shotgun metagenomic sequencing for QA/QC at a commercial laboratory (DNA link USA, Inc., Los Angeles, CA) using NovaSeq6000 platform (150 bp, paired end). Prior to sequencing, DNA quality and purity was confirmed via Nanodrop by measuring the A260:A280 ratio and by analysis on 1% agarose gel and quantity was measured with Qubit by the sequencing lab. TruSeq DNA library preparation kit (Illumina, San Diego, CA) was used for DNA library construction. Sequences are available in the National Center for Biotechnology Information Sequence Read Archive under accession number PRJNA802588.

2.4 Processing of raw reads and network analysis, assembly, and binning

To understand the impact of different bioinformatics approaches on host assignments, two types of pipelines were tested: (P1) raw reads processing by network analysis and (P2) assembly. Details of all bioinformatics processing can be found in *SI 5- SI 8*. Briefly, (P1) raw reads were trimmed and ARGs identified from predicted proteins then Spearman correlations were generated between ARGs and hosts at the family level. Next, (P2) trimmed reads were assembled, contigs were used for predicting open reading frames (ORF), BLASTP was performed against the CARD database to identify ARGs and mobile genetic elements (MGEs) were identified in the ARG containing contigs (ACCs) by BLASTP against the ACLAME database. Bacterial families which could contain NIAID's priority pathogens (https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens) were identified in both the pipelines. For binning, contigs obtained after assembly were binned and classified.

For quality control, reads were generated from an assembled genome of a methicillin resistant *Staphylococcus aureus* (strain MRSA252, NCBI Accession BX571856) (Holden et al., 2004) at genome coverage of 50X, 20X, and 1X. These reads were added to the environmental metagenomes as an *in silico* matrix spike and the samples re-analyzed to assess the performance of the pipelines. The detection and host assignment of three methicillin resistance ARGs, *mec*A, *mec*I and *mec*R1, to the host-family for *S. auerus* (*Staphylococcaceae*) were tabulated for each pipeline. These three genes were present in the chromosomal genome of MRSA but not observed in the metagenomes of the riverbed sediments.

2.5 Statistical tests

All statistical tests were performed in Rstudio version 3.6.2 (www.r-project.org). Data visualization, filtering, sorting and grouping was done in Excel or R. Random Forest was

performed to understand the factors potentially impacting iDNA and eDNA recovery (randomForest package) (Liaw and Wiener, 2002). Censored statistical analyses were performed for qPCR data using regression on order statistics in NADA package (Lee, 2020). ANOVA was performed to identify any significant differences for the qPCR results. A binomial test was performed for presence/absence of the other ARGs analyzed by qPCR. Linear discriminant analysis effect size (LEfSe) test (Segata et al., 2011) was performed for total microbial community and hosts of ARGs (obtained by both pipelines) to identify the biomarkers of each DNA fraction and pipeline. Cluster analysis was performed on total microbial community of all samples using the SIMPROF test (PrimerE, Auckland, NZ). Finally, Jaccard indices were calculated for the adjacency matrices of networks to understand the degree of similarity of the networks (Tantardini et al., 2019).

For the bioinformatics outputs, Bray-Curtis dissimilarity matrices were created for the family level bacterial community and ARGs obtained through the two pipelines. Non-metric multidimensional scaling (nMDS) was performed for ordination to visualize the bacterial, ARG, or ARG-host community structures by spatial and/or DNA fraction factors. Shannon, Simpson and Inverse Simpson diversity indices were calculated for total microbial community and ARGs and their values were compared across different sites and DNA fractions using 2-way ANOVA with a posthoc TukeyHSD test. For the ANOVA, normality of data was confirmed using a Shapiro-Wilk normality test, equal variances confirmed with Bartlett test. Heatmaps were made using the ggplot2 (Wickham, 2016) and reshape (Wickham, 2007) packages to represent the ARG relative abundances in the metagenomes. Chord diagrams were drawn to represent ARGs and their host assignments in the metagenomes using the circlize package (Gu et al., 2014).

3. Results

3.1 Recovery and abundance of ARGs in different DNA fractions across sites

To compare extracellular and intracellular DNA using qPCR and metagenomics, first an assessment of recovery from a series of matrix spikes was evaluated. High recoveries were observed for sediment samples (Table S6). The average recoveries for sediment samples were 72.6±14.6 % for eDNA and 81.8±21.4 % for iDNA. Random forest analysis indicated that pH (7.9% increase in mean square error) followed by clay content (5.5%), silt content (5.0%), sand content (4.5%), moisture (4.0%), site (3.5%), and conductivity (2.4%) explained 31.3% of the variance in the recovery. DNA fraction was not an important factor as indicated by the negative increase in mean square error prior to that variable's removal from the Random Forest model.

Gene copy balances between the different DNA fractions were tested with the expectation that in cases of similar recovery across DNA fractions that the sum of iDNA plus eDNA should be similar to the total DNA extracted. The sum of eDNA and iDNA was similar to the total DNA (paired t-test, all $p > 0.18$) with respect to DNA concentration as measured via nanodrop (Table S7) and *sul*1 and 16S rRNA gene copies analyzed by qPCR.

The sulfonamide resistance gene *sul*1 was the only ARG of the ARGs tested that was quantifiable via qPCR in all DNA fractions (Fig. 2). *sul*1 absolute gene copy concentrations in iDNA, total DNA and eDNA were similar (ANOVA, $p=0.11$). No significant differences were observed between the 16S rRNA gene normalized *sul*1 copies across river matrices for total DNA, iDNA and eDNA (ANOVA, $p =0.5$). *tet*(G) was observed in four of the total DNA, two of the iDNA, and one of the eDNA samples from the five sites sampled ($p=0.15$, prop.test). *sul*2 was detected in all total and iDNA samples but only one eDNA sample ($p=0.004$, prop.test) (Table S8).

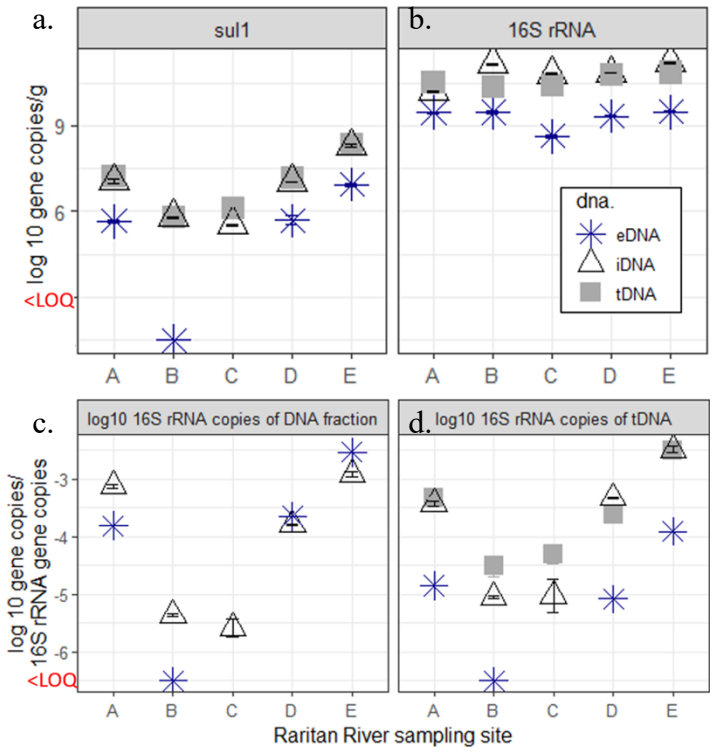


Fig. 2 *(a, b) Abundance of sul1 and 16S rRNA gene copies per gram (wet weight) in sediment samples (of eDNA, iDNA and total DNA ("tDNA")) and (c, d) normalized to 16S rRNA gene copies from the respective DNA fraction and tDNA. Error bars represent standard deviation of technical replicates (N=3). Data shown on the lowest line marked <LOQ was detected but not quantifiable in samples.*

3.2 Classification of metagenomics reads reveals differences in total microbial community

Gel electrophoresis of samples showed that a DNA band was present for each sample. The A260/ 280 ratios of all DNA extracts ranged from 1.6-1.9 and A260/230 < 2.2. Metagenomic data were generated for each sampling site and DNA fraction for the riverbed sediment samples

to study the diversity of ARGs and their hosts (Fig. S1). The microbial communities generated from trimmed reads obtained from iDNA and total DNA clustered more closely (up to 99% similarity) than eDNA (at most 89% similarity, Fig. 3a) for a given sampling site. Replicates for the tidally influenced sites were seen close to each other in nDMS (> 97% similarity). Significant differences were observed between DNA fractions (eDNA-iDNA and eDNA-total DNA) for Shannon, Simpson and Inverse Simpson diversity indices (TukeyHSD, all $p<0.05$) (Table S9). Generally, eDNA had lower diversity than iDNA or total DNA (Table S9).

LEfSe analysis on the data revealed that there were 48 OTUs that served as biomarkers for the eDNA microbial community, while there was only one iDNA and two total DNA biomarkers (Fig. S2). The phylum Proteobacteria was the biomarker in iDNA fraction while the families *Methylocystaceae* and *Kaistiaceae* were biomarkers in total DNA. For eDNA, the biomarkers included the phyla Planctomycetes, Lentisphaerae and Nanoarchaeota in addition to other orders, classes and families (Fig. S2).

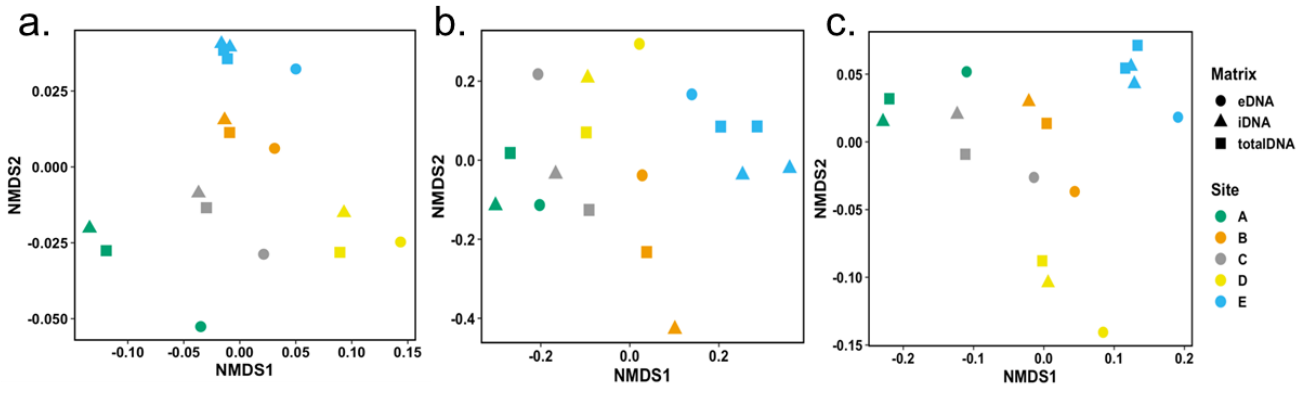


Fig. 3: *nMDS of (a) total microbial community at all sites identified by Kaiju software (stress= 0.05) (b) ARGs at all sites identified by the Assembly pipeline (stress= 0.16), (c) ARGs in raw reads pipeline (stress= 0.09).*

3.3 eDNA, iDNA and total DNA ARG diversity and MGEs

After assembly, 7871 ARG carrying contigs (ACCs) were identified in the 17 samples which contained 392 ARGs in total. Across all sampling sites, 221, 267 and 313 different ARGs were observed in eDNA, iDNA and total DNA fractions, respectively. These ARGs were grouped into 18 major drug classes (Fig. 4) and eight different resistance mechanisms (Fig. S3). The rifamycin drug class constituted 60.8± 5.2 % of total abundance, followed by multidrug resistance (29.1± 3.7% of total abundance). The ARG types detected in all samples and at all the sites were rifamycin, multidrug, glycopeptide, peptide, beta-lactam, aminoglycoside, and aminocoumarin.

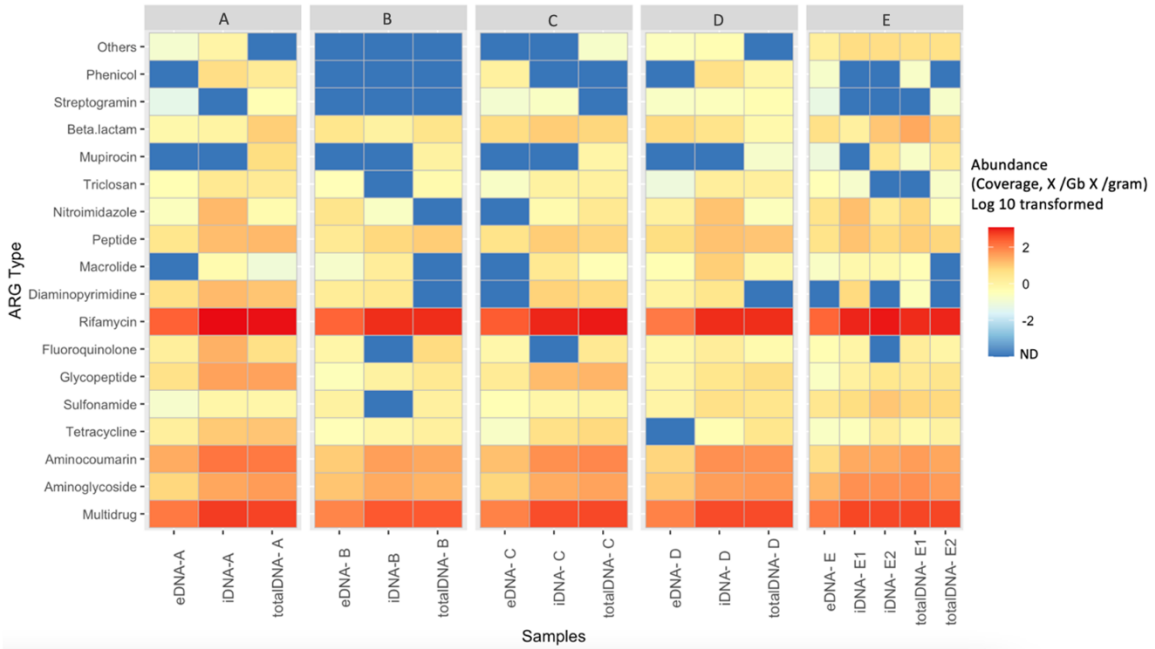
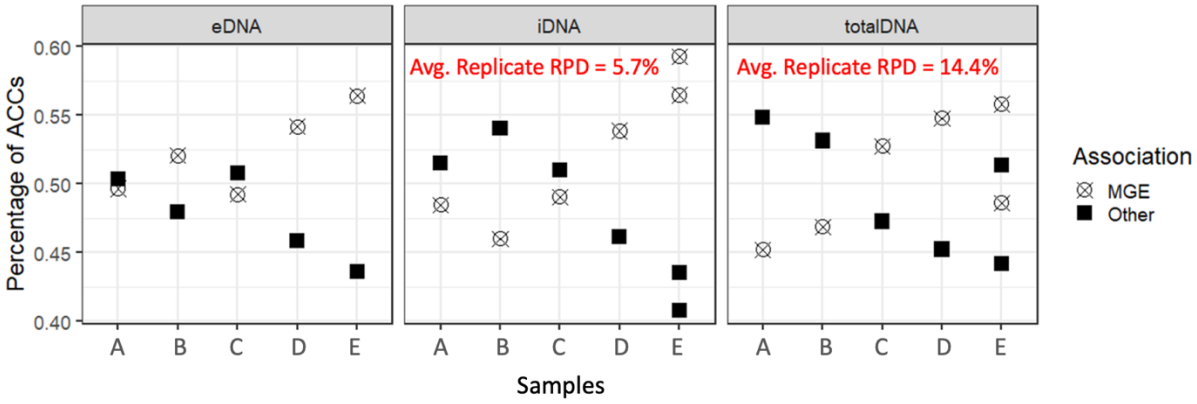


Fig. 4 *Heatmap of ARG relative abundance according to drug class by sampling site and DNA fraction in the bed sediment metagenomes. Field replicates are shown for iDNA and total DNA for Site E. "ND" stands for not detected*

Comparing ARGs in the different DNA fractions, the total normalized abundance (Eq.1) of ARGs in eDNA, iDNA, and total DNA was significantly different (Kruskall-Wallis test, p= 0.0067) (Fig. S4). The average ARG abundance was 349±66 for eDNA which was significantly less than the 1542±315 for iDNA and 1509±246 for total DNA (p= 0.0065, pairwise.t.test, Fig. S4). There were no differences in ARG diversity indices for the ARGs between different DNA fractions (2-way ANOVA, all p>0.05, Table S10).

A total of 2879 ACCs were identified to be associated with MGEs through the assembly pipeline. The percentage of ARGs associated with MGEs from the different DNA fractions and sites was similar (two-way ANOVA, p > 0.2). However, moving downstream, the percent of ARGs associated with MGEs increased with an increase in the percent of flow constituted by wastewater discharge for the iDNA (Spearman rho 0.91, p=0.01) but not for other DNA fractions (both p>0.13) (Fig. 5).

ARGs obtained through the raw-reads pipeline were also grouped into 18 drug classes (Fig. S5). Rifamycin and multidrug drug classes constituted 38.3± 2.9 % and 43.6± 2.8 % of total abundance respectively. Unlike the assembly pipeline, all ARG drug classes were present in all the samples.



271

272 **Fig. 5** *Percentage of ACCs identified with the assembly pipeline associated with MGEs or*
273 *chromosomal/unclassified ("other") by DNA fraction and site. The average relative percent*
274 *difference (RPD) is listed for Site E iDNA and total DNA replicate samples.*

275 **3.4 ARGs hosts and diversity identified through two pipelines**

276 In total, 1443 ARGs (11.5% of total ARGs) were assigned to hosts at least at the phylum level
277 after assembly and classification by CAT. At the family level, 23, 30 and 33 ARG hosts were
278 identified for eDNA, iDNA and tDNA respectively. The most commonly observed hosts were
279 *Aeromonadaceae* representing 24% of tDNA, 46.3% of eDNA, and 17.5% of iDNA hosts,
280 *Mycbacteriaceae* representing 26% of tDNA and 17.5% of iDNA hosts and *Pseudomonadaceae*
281 representing 15.1% of eDNA hosts (Fig.S6). In all the three fractions of DNA, most of the
282 tetracycline resistance genes were assigned to *Mycobacteriaceae* family while *Aeromonadaceae*
283 were the hosts of most of beta-lactam group of ARGs. Host-ARG associations are shown site by
284 site in Fig. S7. Significant differences were observed in the diversity of ARG host assignments
285 with generally the greatest Shannon diversity observed for the iDNA or total DNA followed by
286 eDNA (all $p < 0.025$, TukeyHSD) (Table S11).

Bins (624) were identified across the 17 samples, of which, 40 were high quality bins. Of these, three bins carrying ARGs were recovered from eDNA and iDNA. The bin from eDNA fraction contained the ARG *rsm*A and was assigned to phylum Nitrospirae. One of bins from the iDNA fraction carrying multiple ARGs (*sul*1, *rsm*A, *qac*E and *qac*Edelta1) belonged to phylum Nitrospirae and the other belonged to Class Gammaproteobacteria and carried a single ARG *rsm*A.

Network analysis revealed a wide diversity of potential ARG hosts for eDNA, iDNA and total DNA (Shannon diversity 4.91±0.01). For eDNA, 207 bacterial families correlated strongly with 182 ARGs (Fig. S8). Five families showed strong correlations with more than 15 ARGs including *Enterobacteriaceae*, *Neisseriaceae*, and *Shewanellaceae*. ARGs linked with more than 10 families included genes harboring resistance to multiple drugs (multidrug), beta-lactam, rifamycin, fluoroquinolone, streptogramin and other antibiotics. For the iDNA, strong correlations were observed between 220 bacterial families and 189 ARGs (Fig. S9). Fourteen families including *Sphingomonadaceae*, showed strong correlations with more than 15 ARGs. At least 47 ARGs were assigned to more than 10 bacterial hosts. For total DNA, 221 bacterial families showed high correlation with 182 ARGs (Fig. S10). Ten families including *Legionellaceae*, and *Mycobacteriaceae* were identified as the potential hosts of more than 15 ARGs. The majority (65%) of the ARGs showing strong correlations with more than 10 bacterial families had mechanisms of antibiotic efflux.

The Jaccard index for the network pairs of eDNA-iDNA was 0.455, iDNA-total DNA was 0.457, total DNA-eDNA was 0.47. Jaccard index of 0 indicates that there is no overlap while index of one indicates that there is perfect overlap between networks (Yang et al., 2016). The Jaccard

309 indices indicate ~50% overlap in the network pairs, (statistical comparison of these indices is
310 not possible with the present experiment design).

a) b)

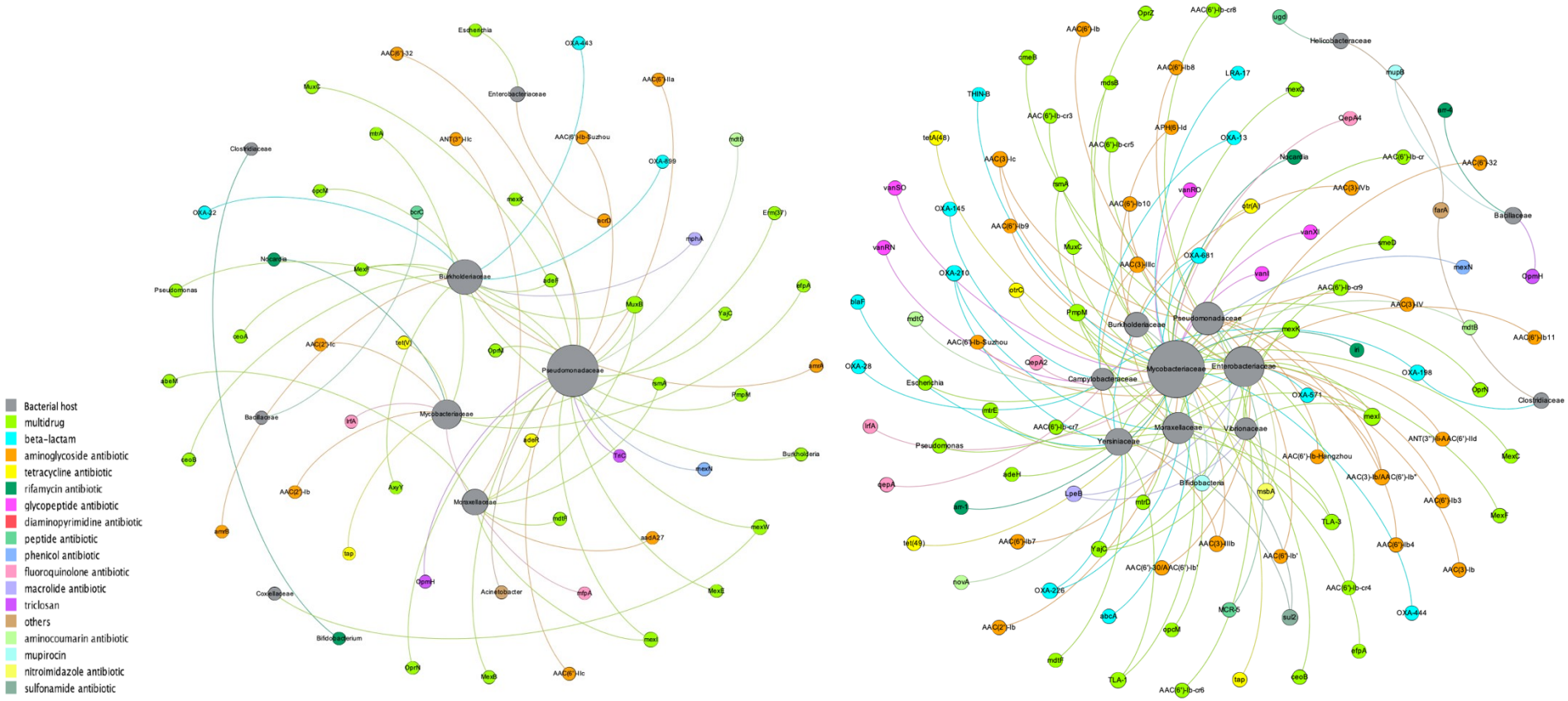


Fig 6: *Linkage of hosts (bacterial families) which contain pathogens in NIAID's list of priority pathogens to ARGs obtained from the a) Assembly and b) Raw reads pipeline*

3.5 Comparison of the Network and Assembly pipeline

The ARG-host linkages identified by network and assembly pipelines were compared first by identifying common hosts (Table S13- S15). For all DNA fractions, most of the hosts identified by the assembly pipeline were also identified by the network analysis pipeline (average 85.3±5.5%) with more hosts identified by the network pipeline (247 hosts) compared to assembly (53 total hosts). nMDS plots of ARGs generally showed clustering by sampling site and less similarity for iDNA and total DNA than seen for the total community nMDS (Fig.3 b,c). LEfSe was performed to determine the biomarkers for each pipeline and compare the host-ARG assignments (i.e., input was hosts as a function of number of ARGs). A total of eight families were identified as biomarkers of the ARG hosts for the assembly pipeline compared to 132 families for the network pipeline. Of the ARG hosts, seven bacterial families for the assembly pipeline and 12 for the network analysis were identified as putative hosts of ARGs that could contain NIAID's priority pathogens (Fig. 6).

3.6 Quality assurance *in silico* matrix spikes

In silico matrix spikes of a control genome to the field sample metagenomes (Table S12) demonstrated the expected host-ARG correlation was observed for the network analysis when the environmental metagenome was spiked with 50X and 20X MRSA genome coverage. The assembly pipeline and binning for samples spiked with 50X of the *S. aureus* genome resulted in the correct host assignment of the selected methicillin resistant ARGs (*mec*A, *mec*I, *mec*R1) at least at family level. At 20X depth, two out of the three tested samples had successful association of host to ARGs at family level for the assembly pipeline, while binning was successful in linking the correct host to these ARGs at genus level.

Adding 1X of the MRSA genome, the raw-reads pipeline failed to assign the correct host to the ARGs. Potential mis-associations of the ARGs *mec*A, *mec*I, *mec*R1 to hosts (*Silvanigrellaceae*, *Gomontiellaceae*) with abundance less than 0.02 (less than 1X depth) were observed for the network analyses. These are considered potential mis-linkages given that these *mec* genes were below detection in the field metagenomes and therefore were only expected to link with *S. aureus*. Multiple ARG hits for *mec*A, *mec*I and *mec*R1(up to 681 hits) were observed for the raw-reads pipeline for 1X, 20X and 50X coverages.

As expected, a single ARG hit for *mec*A, *mec*I, and *mec*R1 was observed in the samples that had 20X and 50X *S. aureus* genome addition for the assembly pipeline. With 1X genome coverage of *S. aureus* the assembly pipeline resulted in no hits for *mec*I and multiple hits for *mec*A and *mec*R1. *mec*A was assigned to the host *Staphylococcaceae* (family level) in two out of the three samples at 1X depth. At a depth of 1X, binning also failed to associate the host to the expected ARGs.

4 Discussion

4.1 Quantitative comparisons of ARGs in different DNA fractions

Raritan River bed sediment had lower concentrations of 16S rRNA gene copies in eDNA compared to iDNA and total DNA fractions across the sampling sites. This result is in accordance with previous studies on terrestrial sludge and sediment samples (Dong et al., 2019; Zhang et al., 2013). The ARGs were either less frequently detected (*sul*2, *tet*(G)) or had similar concentration (*sul*1) in eDNA than the iDNA fraction. This is in contrast to *sul*1, *sul*2, and other ARG observations in aquatic sediment samples (Mao et al., 2014; Zhang et al., 2018).

Interestingly, the eDNA concentrations for the targeted genes were not necessarily highest at Sites C through E that were influenced by WWTP effluent. Extracellular ARGs have been detected in oxidant disinfected WWTP effluent (Oliveira et al., 2020) and in river water downstream of WWTP effluent discharge (Liu et al., 2020). Here the accumulation of eDNA ARGs in sediments may also be a function of settling/deposition rates, eDNA loss rates in the sediments and the water column prior to settling, sediment type and sorption, etc. For example, while environmental eDNA is subject to enzymatic degradation, the eDNA that adsorbs to soil particles is protected from such degradation with clay particles providing more protection to bound eDNA (Barnes et al., 2014; Lopatkin et al., 2017; Nielsen et al., 2007; Ogram et al., 1994). This may explain why the highest recovery for eDNA matrix spikes was observed at Site D which also had the lowest sand content and the highest fractions of both clay and silt. Further, a previous study on riverbed sediments with a higher clay percentage (7% -22%) reported a higher concentration of eDNA than iDNA (Mao et al., 2014).

The matrix spike recoveries of eDNA and iDNA were similar to or greater than those previously reported for sediment samples using the same eDNA extraction method as here. eDNA recovery was 37-68% for cloned CESA9 gene and $80.2 \pm 9.4\%$ for *E. coli* DNA as internal standard while recovery was 57-94% for plant pathogen bacterial cells and $102.2 \pm 4.1\%$ for *E. coli* cells as iDNA internal standards (Chen et al., 2018; Mao et al., 2014; Zhang et al., 2018).

4.2 Microbial community structure and ARG diversity

The eDNA microbial community generally had lower diversity and a different structure compared to the iDNA and total DNA. This observation was somewhat different from a previously published study on wastewater where community structures were conserved between

eDNA and iDNA at the family level using a cutoff criterion of 0.5% abundance (Calderon-Franco et al., 2021). The loading of eDNA appears environment specific as eDNA was >40% of the total DNA (as calculated by comparing DNAse treated samples representing intracellular DNA only to non-DNAase treated samples representing total DNA) reduced the alpha diversity in drinking water samples disinfected by monochloramine (Sakcham et al., 2019). Here, eDNA measured by nanodrop was ~13% of total DNA. As expected, the greatest relative abundance and diversity of ARGs was observed in the iDNA and total DNA rather than eDNA.

Through the assembly pipeline, rifamycin resistance genes were ubiquitous and abundant (60.8± 5.2 %) in the sediment samples, similar to a previous metagenomic study (Ma et al., 2014). Rifamycin is naturally produced by soil bacteria (Li et al., 2020) and was isolated from bacteria in the family *Pseudonocardiaceae* (Sensi, 1983). This family was detected in all the samples through the raw reads pipeline. Multidrug resistance was the second most abundant category (29.1± 3.7%), which has been found to be abundant in river sediments (Chen et al., 2019; Jiang et al., 2018; Li et al., 2021). Notably, vancomycin resistance genes were observed in all samples. Vancomycin is a last resort drug used for treating antibiotic resistant infections (Moellering, 1998) with vancomycin resistant *Enterococci* considered a "serious threat" (CDC, 2019).

The bed sediment microbial community structure and ARG diversity were similar between iDNA and total DNA for a given site. Thus, measuring only total DNA for these riverbed samples would not overestimate the ARGs or drastically shift the microbial community in iDNA. Likewise, separation of eDNA did not cause a significant shift in microbial community structure of iDNA. A recent review highlighted that significant differences in microbial community were reported after the removal of eDNA in environments such as drinking water and soil with the effect more apparent at lower taxonomic levels such as genus or amplicon sequence variant

(ASV) level (Bairoliya et al., 2021). However, some studies have also reported no change in richness and evenness in community after eDNA removal in sediment samples (Bairoliya et al., 2021).

4.3 ARG host assignment in bed sediment by assembly

Assembly of the bed sediment metagenomes resulted in low percentage of ACCs that were classified by CAT. This is consistent with a recent study on pig farms also using CAT for taxonomic classification of contigs (~19% contigs classified at phylum level) (Zhang et al., 2021). Likewise, there was a low number of high-quality bins carrying ARGs. Other studies that reported a higher number of ARG carrying bins (up to 635 bins) used activated sludge samples (Liu et al., 2019; Zhao et al., 2020). Microbial community complexity or other differences in studies (e.g., geography) could explain these differences.

In this study, most of tetracycline and beta- lactam resistance genes were linked to *Mycobacteriaceae* and *Aeromonadaceae* respectively through the assembly pipeline. Tetracycline resistance has been linked to *Mycobacterium* genus frequently through network analysis in metagenomic sequencing studies on sludge samples (Fan et al., 2018; Xia et al., 2019). *Aeromonadaceae* has also been identified as the host of beta-lactam genes through assembly of metagenomics sequences from pig farms (Zhang et al., 2021).

4.4 Comparison of raw reads and assembly

The host-ARG associations obtained through both pipelines were compared with the raw reads analysis pipeline identifying more ARGs than the assembly pipeline, likely due to the low mapping percentage of assembly (39-55%), and/or due to repeat resolution by assembler (Nurk

et al., 2017). Comparison of this observation to the present literature is complicated by the application of different cutoff criteria by different researchers as explained in *SI 5*.

Taxonomic families that could contain NIAID's priority pathogens were identified as ARG hosts through both pipelines. ARGs in pathogenic bacteria pose a greater concern because they can directly infect humans and resist the antibiotic treatment for treating serious infections (WHO, 2014). The number of ARGs linked to the families containing putative pathogenic hosts varied by pipeline, with more such hosts identified by the raw-reads pipeline than the assembly pipeline (Fig. 6). The ability to link ARGs to hosts at deeper taxonomic levels will be needed to best understand the hazard, as these families will also contain many commensal organisms. Although, if gene transfer is most likely from hosts that are most closely related, this may be a reasonable first approximation. More targeted methods such as qPCR would be desirable to confirm the presence of the pathogens.

A benefit of the assembly pipeline is that it provides genetic context, allowing for the observation that the percentage of ARGs associated with MGEs increased moving downstream, potentially due to the increase in wastewater influence (Table S2). Wastewater treatment plants discharge extracellular ARGs due to lysing during disinfection (e.g., (Eramo et al., 2019; Yuan et al., 2019)). Other sources of eDNA in the environment include active DNA release from live cells, DNA released due to cell lysis and viruses (Ibáñez de Aldecoa et al., 2017).

The assembly pipeline used more memory and took ~4 times longer to run, making the raw-reads pipeline more practical to perform. A drawback of the network analysis, which is based on Spearman's correlations and required multiple samples for predicting host-ARG linkages, is that with the present study design only linkages for ARGs and hosts across the sampling sites was

447 possible, rather than for a given sampling site. In contrast, geospatial comparisons were feasible
448 with assembly. The diversity of the ARG hosts was lower for the assembly pipeline, which may
449 indicate that the assembly pipeline was more conservative than the network analysis, assuming
450 both provided assignments with similar accuracy.

451 The *in silico* spike-in study confirmed that poor genome coverage of some organisms in
452 metagenome can lead to failure in host-assignments, either by not capturing the assignments or
453 by providing potentially incorrect assignments to hosts. Removal of low abundance taxa may
454 help in resolving this issue. It also confirmed that accurate host-ARG assignments for the MRSA
455 spike were observed in the reads-based pipeline as well as assembly pipeline at 20X, 50X depth.

456 **5. Conclusion**

457 ARGs in riverbed sediments were either more commonly observed via qPCR in iDNA and total
458 DNA compared to eDNA or similar between DNA fractions. In contrast, metagenomics clearly
459 demonstrated the eDNA fraction had less total ARG abundance compared to iDNA and total
460 DNA. Removing eDNA resulted in shifts in total microbial community structure but iDNA and
461 total DNA communities were generally closely clustered, indicating that total DNA was
462 reasonable representation of the iDNA for this system. Comparing pipelines for ARG host
463 assignment, most of the ARG hosts identified by the assembly pipeline were also identified by
464 the raw reads pipeline (which provided many more ARG-host linkages) and potential pathogenic
465 hosts were identified in both the pipelines. A benefit of assembly pipeline was demonstrated
466 through the observation that ARGs assigned to mobile genetic units increased for iDNA moving
467 downstream, useful information for understanding the potential for ARG propagation. The
468 pipeline comparisons provided here can inform how to compare studies that applied different

methods for ARG-host assignment. Further work is needed to confirm that pathogenic hosts were present in our samples, for example via qPCR or longer read sequencing or emerging techniques for high throughput ARG-host identification (Gallego et al., 2020).

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

The research was supported by a grant from the National Science Foundation (# 1846815). Thanks to Will Boni for assistance with field sampling and Genevieve Ehasz for help with qPCR.

Contributors

Conceptualization, Methodology, Formal analysis: ASD, NLF; Investigation, Writing-original draft: ASD; Writing – review & editing, project administration, funding acquisition: NLF.

References

Alawi, M., Schneider, B. and Kallmeyer, J. 2014. A procedure for separate recovery of extra- and intracellular DNA from a single marine sediment sample. J Microbiol Methods 104, 36-42.

Aminov, R.I., Chee-Sanford, J.C., Garrigues, N., Teferedegne, B., Krapac, I.J., White, B.A. and Mackie, R.I. 2002. Development, validation, and application of PCR primers for

489 detection of tetracycline efflux genes of gram-negative bacteria. Appl Environ Microbiol
490 68(4), 1786-1793.
491 Ashbolt, N.J., Heberer, T., van den Eede, C., Schönfeld, J., Krone, S., Backhaus, T., Li, X.-Y.,
492 Borriello, P., Amezquita, A., Larsson, D.G.J., Brandt, K., Smalla, K., Silley, P.,
493 Lawrence, J. and Ryan, J. 2013. Human Health Risk Assessment (HHRA) for
494 environmental development and transfer of antibiotic resistance. Environ. Health
495 Perspect. 121(9), 993-1001.
496 Bairoliya, S., Koh Zhi Xiang, J. and Cao, B. 2021. Extracellular DNA in environmental
497 samples: Occurrence, extraction, quantification, and impact on microbial biodiversity
498 assessment. Appl Environ Microbiol, AEM0184521.
499 Barnes, M.A., Turner, C.R., Jerde, C.L., Renshaw, M.A., Chadderton, W.L. and Lodge, D.M.
500 2014. Environmental conditions influence eDNA persistence in aquatic systems. Environ
501 Sci Technol 48(3), 1819-1827.
502 Berendonk, T.U., Manaia, C.M., Merlin, C., Fatta-Kassinos, D., Cytryn, E., Walsh, F.,
503 Burgmann, H., Sorum, H., Norstrom, M., Pons, M.N., Kreuzinger, N., Huovinen, P.,
504 Stefani, S., Schwartz, T., Kisand, V., Baquero, F. and Martinez, J.L. 2015. Tackling
505 antibiotic resistance: the environmental framework. Nat Rev Microbiol 13(5), 310-317.
506 Breitwieser, F.P., Lu, J. and Salzberg, S.L. 2019. A review of methods and databases for
507 metagenomic classification and assembly. Brief Bioinform 20(4), 1125-1136.
508 Calderon-Franco, D., van Loosdrecht, M.C.M., Abeel, T. and Weissbrodt, D.G. 2021. Free-
509 floating extracellular DNA: Systematic profiling of mobile genetic elements and
510 antibiotic resistance from wastewater. Water Res 189, 116592.
511 US Centers for Disease Control (CDC). 2019 Antibiotic Resistance Threats in the United States.

512 Chen, B., Lin, L., Fang, L., Yang, Y., Chen, E., Yuan, K., Zou, S., Wang, X. and Luan, T. 2018.
513 Complex pollution of antibiotic resistance genes due to beta-lactam and aminoglycoside
514 use in aquaculture farming. Water Res 134, 200-208.
515 Chen, H., Chen, R., Jing, L., Bai, X. and Teng, Y. 2019. A metagenomic analysis framework
516 for characterization of antibiotic resistomes in river environment: Application to an urban
517 river in Beijing. Environ Pollut 245, 398-407.
518 Chen, H., Li, Y., Sun, W., Song, L., Zuo, R. and Teng, Y. 2020. Characterization and source
519 identification of antibiotic resistance genes in the sediments of an interconnected river-
520 lake system. Environ Int 137, 105538.
521 Corinaldesi, C., Danovaro, R. and Dell'Anno, A. 2005. Simultaneous recovery of extracellular
522 and intracellular DNA suitable for molecular studies from marine sediments. Appl
523 Environ Microbiol 71(1), 46-50.
524 Deere, D., Porter, J., Pickup, R.W. and Edwards, C. 1996. Survival of cells and DNA of
525 Aeromonas salmonicida released into aquatic microcosms. J Appl Bacteriol 81(3), 309-
526 318.
527 Deng, Y., Mao, C., Lin, Z., Su, W., Cheng, C., Li, Y., Gu, Q., Gao, R., Su, Y. and Feng, J. 2022.
528 Nutrients, temperature, and oxygen mediate microbial antibiotic resistance in sea bass
529 (Lateolabrax maculatus) ponds. Sci Total Environ 819, 153120.
530 Dodd, M.C. 2012. Potential impacts of disinfection processes on elimination and deactivation
531 of antibiotic resistance genes during water and wastewater treatment. Journal of
532 Environmental Monitoring 14(7), 1754-1771.

533 Dong, P., Wang, H., Fang, T., Wang, Y. and Ye, Q. 2019. Assessment of extracellular
534 antibiotic resistance genes (eARGs) in typical environmental samples and the
535 transforming ability of eARG. Environ Int 125, 90-96.
536 Eramo, A., Morales Medina, W.R. and Fahrenfeld, N.L. 2019. Viability-based quantification of
537 antibiotic resistance genes and human fecal markers in wastewater effluent and receiving
538 waters. Sci Total Environ 656, 495-502.
539 Fan, X.Y., Gao, J.F., Pan, K.L., Li, D.C., Dai, H.H. and Li, X. 2018. Functional genera,
540 potential pathogens and predicted antibiotic resistance genes in 16 full-scale wastewater
541 treatment plants treating different types of wastewater. Bioresour Technol 268, 97-106.
542 Gallego, S., Barkay, T. and Fahrenfeld, N.L. 2020. Tagging the *van*A gene in wastewater
543 microbial communities for cell sorting and taxonomy of *van*A carrying cells. Sci. Total
544 Environ. 732, 138865.
545 Gu, Z.G., Gu, L., Eils, R., Schlesner, M. and Brors, B. 2014. circlize implements and enhances
546 circular visualization in R. Bioinformatics 30(19), 2811-2812.
547 Guo, X.P., Yang, Y., Lu, D.P., Niu, Z.S., Feng, J.N., Chen, Y.R., Tou, F.Y., Garner, E., Xu, J.,
548 Liu, M. and Hochella, M.F., Jr. 2018. Biofilms as a sink for antibiotic resistance genes
549 (ARGs) in the Yangtze Estuary. Water Res 129, 277-286.
550 Holden, M.T., Feil, E.J., Lindsay, J.A., Peacock, S.J., Day, N.P., Enright, M.C., Foster, T.J.,
551 Moore, C.E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S.D., Chillingworth,
552 C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd,
553 L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K.D.,
554 Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M.A.,
555 Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K.,

Whitehead, S., Barrell, B.G., Spratt, B.G. and Parkhill, J. 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci U S A 101(26), 9786-9791.

Hong, P.-Y., Al-Jassim, N., Ansari, M. and Mackie, R. 2013. Environmental and public health implications of water reuse: antibiotics, antibiotic resistant bacteria, and antibiotic resistance genes. Antibiotics 2(3), 367-399.

Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M. and Brown, C.T. 2014. Tackling soil diversity with the assembly of large, complex metagenomes. Proc Natl Acad Sci U S A 111(13), 4904-4909.

Ibáñez de Aldecoa, A.L., Zafra, O. and González-Pastor, J.E. 2017. Mechanisms and regulation of extracellular DNA release and its biological roles in microbial communities. Front Microbiol 8, 1390.

Jensen, L.B., Ahrens, P., Dons, L., Jones, R.N., Hammerum, A.M. and Aarestrup, F.M. 1998. Molecular analysis of Tn1546 in *Enterococcus faecium* isolated from animals and humans. J Clin Microbiol 36(2), 437-442.

Jiang, H., Zhou, R., Zhang, M., Cheng, Z., Li, J., Zhang, G., Chen, B., Zou, S. and Yang, Y. 2018. Exploring the differences of antibiotic resistance genes profiles between river surface water and sediments using metagenomic approach. Ecotoxicol Environ Saf 161, 64-69.

Jiang, X., Liu, L., Chen, J., Fan, X., Xie, S., Huang, J. and Yu, G. 2021. Antibiotic resistance genes and mobile genetic elements in a rural river in Southeast China: Occurrence, seasonal variation and association with the antibiotics. Sci Total Environ 778, 146131.

Lee, L. 2020. Nondetects and data analysis for environmental data. R package version 1.6-1.1.

579 Li, B., Chen, Z., Zhang, F., Liu, Y. and Yan, T. 2020. Abundance, diversity and mobility
580 potential of antibiotic resistance genes in pristine Tibetan Plateau soil as revealed by soil
581 metagenomics. FEMS Microbiol Ecol 96(10).
582 Li, Y., Chen, H., Song, L., Wu, J., Sun, W. and Teng, Y. 2021. Effects on microbiomes and
583 resistomes and the source-specific ecological risks of heavy metals in the sediments of an
584 urban river. J Hazard Mater 409, 124472.
585 Liaw, A. and Wiener, M. 2002. Classification and regression by randomForest. R news 2(3),
586 18-22.
587 Liu, M., Hata, A., Katayama, H. and Kasuga, I. 2020. Consecutive ultrafiltration and silica
588 adsorption for recovery of extracellular antibiotic resistance genes from an urban river.
589 Environ Pollut 260, 114062.
590 Liu, Z., Klumper, U., Liu, Y., Yang, Y., Wei, Q., Lin, J.G., Gu, J.D. and Li, M. 2019.
591 Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic
592 resistance genes in activated sludge. Environ Int 129, 208-220.
593 Loman, N.J., Constantinidou, C., Christner, M., Rohde, H., Chan, J.Z., Quick, J., Weir, J.C.,
594 Quince, C., Smith, G.P., Betley, J.R., Aepfelbacher, M. and Pallen, M.J. 2013. A
595 culture-independent sequence-based metagenomics approach to the investigation of an
596 outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. JAMA 309(14), 1502-1510.
597 Lopatkin, A.J., Meredith, H.R., Srimani, J.K., Pfeiffer, C., Durrett, R. and You, L. 2017.
598 Persistence and reversal of plasmid-mediated antibiotic resistance. Nat Commun 8(1),
599 1689.
600 Lorenz, M.G. and Wackernagel, W. 1987. Adsorption of DNA to sand and variable degradation
601 rates of adsorbed DNA. Appl Environ Microbiol 53(12), 2948-2952.

Luo, Y., Mao, D., Rysz, M., Zhou, Q., Zhang, H., Xu, L. and P, J.J.A. 2010. Trends in antibiotic resistance genes occurrence in the Haihe River, China. Environ Sci Technol 44(19), 7220-7225.

Ma, L., Li, B. and Zhang, T. 2014. Abundant rifampin resistance genes and significant correlations of antibiotic resistance genes and plasmids in various environments revealed by metagenomic analysis. Appl Microbiol Biotechnol 98(11), 5195-5204.

Mao, D., Luo, Y., Mathieu, J., Wang, Q., Feng, L., Mu, Q., Feng, C. and Alvarez, P.J. 2014. Persistence of extracellular DNA in river sediment facilitates antibiotic resistance gene propagation. Environ Sci Technol 48(1), 71-78.

Martinez, J.L., Coque, T.M. and Baquero, F. 2015. What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol 13(2), 116-123.

Mazel, D. and Davies, J. 1999. Antibiotic resistance in microbes. Cell Mol Life Sci 56(9-10), 742-754.

Miao, J., Yin, Z., Yang, Y., Liang, Y., Xu, X. and Shi, H. 2021. Abundance and Dynamic Distribution of Antibiotic Resistance Genes in the Environment Surrounding a Veterinary Antibiotic Manufacturing Site. Antibiotics (Basel) 10(11).

Moellering, R.C., Jr. 1998. The specter of glycopeptide resistance: current trends and future considerations. Am J Med 104(5A), 3S-6S.

Munir, M., Wong, K. and Xagoraraki, I. 2011. Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. Water Res 45(2), 681-693.

623 Muyzer, G., de Waal, E.C. and Uitterlinden, A.G. 1993. Profiling of complex microbial
624 populations by denaturing gradient gel electrophoresis analysis of polymerase chain
625 reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59(3), 695-700.
626 Nielsen, K.M., Johnsen, P.J., Bensasson, D. and Daffonchio, D. 2007. Release and persistence
627 of extracellular DNA in the environment. Environ Biosafety Res 6(1-2), 37-53.
628 Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P.A. 2017. metaSPAdes: a new
629 versatile metagenomic assembler. Genome Res 27(5), 824-834.
630 Ogram, A.V., Mathot, M.L., Harsh, J.B., Boyle, J. and Pettigrew, C.A. 1994. Effects of DNA
631 polymer length on its adsorption to soils. Appl Environ Microbiol 60(2), 393-396.
632 Oliveira, M., Nunes, M., Barreto Crespo, M.T. and Silva, A.F. 2020. The environmental
633 contribution to the dissemination of carbapenem and (fluoro)quinolone resistance genes
634 by discharged and reused wastewater effluents: The role of cellular and extracellular
635 DNA. Water Res 182, 116011.
636 Pei, R., Kim, S.C., Carlson, K.H. and Pruden, A. 2006. Effect of river landscape on the
637 sediment concentrations of antibiotics and corresponding antibiotic resistance genes
638 (ARG). Water Res 40(12), 2427-2435.
639 Rowe-Magnus, D.A. and Mazel, D. 2002. The role of integrons in antibiotic resistance gene
640 capture. Int J Med Microbiol 292(2), 115-125.
641 Sakcham, B., Kumar, A. and Cao, B. 2019. Extracellular DNA in monochloraminated drinking
642 water and its influence on DNA-based profiling of a microbial community. Environ Sci
643 Technol Lett 6(5), 306-312.

644 Scholz, M.B., Lo, C.C. and Chain, P.S. 2012. Next generation sequencing and bioinformatic
645 bottlenecks: the current state of metagenomic data analysis. Curr Opin Biotechnol 23(1),
646 9-15.
647 Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. and Huttenhower, C.
648 2011. Metagenomic biomarker discovery and explanation. Genome Biol 12(6), R60.
649 Sensi, P. 1983. History of the development of rifampin. Rev Infect Dis 5 Suppl 3, S402-406.
650 Tantardini, M., Ieva, F., Tajoli, L. and Piccardi, C. 2019. Comparing methods for comparing
651 networks. Sci Rep 9(1), 17557.
652 Turner, C.R., Uy, K.L. and Everhart, R.C. 2015. Fish environmental DNA is more concentrated
653 in aquatic sediments than surface water. Biological Conservation 183, 93-102.
654 Vikesland, P.J., Pruden, A., Alvarez, P.J.J., Aga, D., Bürgmann, H., Li, X.-d., Manaia, C.M.,
655 Nambi, I., Wigginton, K., Zhang, T. and Zhu, Y.-G. 2017. Toward a comprehensive
656 strategy to mitigate dissemination of environmental sources of antibiotic resistance.
657 Environ. Sci. Technol. 51(22), 13061-13069.
658 Vollmers, J., Wiegand, S. and Kaster, A.K. 2017. Comparing and evaluating metagenome
659 assembly tools from a microbiologist's perspective - Not only size matters! PLoS One
660 12(1), e0169662.
661 World Health Organization (W.H.O.) (2014) Antimicrobial resistance: global report on
662 surveillance, World Health Organization.
663 Wickham, H. 2007. Reshaping data with the reshape package. J Stat Software 21(12), 1-20.
664 Wickham, H. 2016. ggplot2-Elegant Graphics for Data Analysis. Springer International
665 Publishing. Cham, Switzerland.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergard, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L.S., Pearman, P.B., Cheddadi, R., Murray, D., Brathen, K.A., Yoccoz, N., Binney, H., Cruaud, C., Wincker, P., Goslar, T., Alsos, I.G., Bellemain, E., Brysting, A.K., Elven, R., Sonstebo, J.H., Murton, J., Sher, A., Rasmussen, M., Ronn, R., Mourier, T., Cooper, A., Austin, J., Moller, P., Froese, D., Zazula, G., Pompanon, F., Rioux, D., Niderkorn, V., Tikhonov, A., Savvinov, G., Roberts, R.G., MacPhee, R.D., Gilbert, M.T., Kjaer, K.H., Orlando, L., Brochmann, C. and Taberlet, P. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. Nature 506(7486), 47-51.

Xia, H., Chen, J., Chen, X., Huang, K. and Wu, Y. 2019. Effects of tetracycline residuals on humification, microbial profile and antibiotic resistance genes during vermicomposting of dewatered sludge. Environ Pollut 252(Pt B), 1068-1077.

Xie, G., Chain, P.S., Lo, C.C., Liu, K.L., Gans, J., Merritt, J. and Qi, F. 2010. Community and gene composition of a human dental plaque microbiota obtained by metagenomic sequencing. Mol Oral Microbiol 25(6), 391-405.

Yang, J.J., Kwon, H. and Lee, J.M. 2016. Complementary characteristics of correlation patterns in morphometric correlation networks of cortical thickness, surface area, and gray matter volume. Sci Rep 6, 26682.

Yuan, Q.B., Huang, Y.M., Wu, W.B., Zuo, P., Hu, N., Zhou, Y.Z. and Alvarez, P.J.J. 2019. Redistribution of intracellular and extracellular free & adsorbed antibiotic resistance genes through a wastewater treatment plant by an enhanced extracellular DNA extraction method with magnetic beads. Environ Int 131, 104986.

688 Zhang, R.M., Liu, X., Wang, S.L., Fang, L.X., Sun, J., Liu, Y.H. and Liao, X.P. 2021.
689 Distribution patterns of antibiotic resistance genes and their bacterial hosts in pig farm
690 wastewater treatment systems and soil fertilized with pig manure. Sci Total Environ 758,
691 143654.
692 Zhang, Y., Niu, Z., Zhang, Y. and Zhang, K. 2018. Occurrence of intracellular and extracellular
693 antibiotic resistance genes in coastal areas of Bohai Bay (China) and the factors affecting
694 them. Environ Pollut 236, 126-136.
695 Zhang, Y., Snow, D.D., Parker, D., Zhou, Z. and Li, X. 2013. Intracellular and extracellular
696 antimicrobial resistance genes in the sludge of livestock waste management structures.
697 Environ Sci Technol 47(18), 10206-10213.
698 Zhao, R., Yu, K., Zhang, J., Zhang, G., Huang, J., Ma, L., Deng, C., Li, X. and Li, B. 2020.
699 Deciphering the mobility and bacterial hosts of antibiotic resistance genes under
700 antibiotic selection pressure by metagenomic assembly and binning approaches. Water
701 Res 186, 116318.
702 Zhou, S., Zhu, Y.J., Yan, Y., Wang, W.G. and Wang, Y.Y. 2019. Deciphering extracellular
703 antibiotic resistance genes (eARGs) in activated sludge by metagenome. Water Res 161,
704 610-620.
705
706