Environmental Science
Water Research & Technology

Factors associated with elevated levels of antibiotic resistance genes in sewer sediments and wastewater

Journal:	*Environmental Science: Water Research & Technology*
Manuscript ID	EW-ART-03-2020-000230
Article Type:	Paper

Water Impact Statement

Characterizing the sewer microbiome is of interest for protecting public health and interpreting wastewater-based epidemiology data. Our sewer investigation indicated microbial community and antibiotic resistance varied by matrix and season. NDM-1, a resistance gene of high health importance, was frequently observed in sewer sediments and rarely in sewage, indicating the potential relevance of in-sewer processes for select microbial agents.

Factors associated with elevated levels of antibiotic resistance genes in sewer sediments and wastewater

Eramo, Alessia,[a] Morales Medina, William R.,[b] Fahrenfeld, N.L.[a*]

[a]Civil & Environmental Engineering, Rutgers, The State University of New Jersey, 500 Bartholomew Rd., Piscataway, NJ 08854

[b]Microbiology and Molecular Genetics, Rutgers, The State University of New Jersey

*nfahrenf@rutgers.edu, 1-848-445-8416

Abstract

The sewer environment is a potential hotspot for the proliferation of antibiotic resistance genes (ARGs) and other hazardous microbial agents. Understanding the potential for ARG proliferation and retardation and/or accumulation in sewer sediments is of interest for protecting the health of sewage workers and the broader community in the event of sewer overflows as well as for interpreting sewage epidemiology data. To better understand this understudied environment for antibiotic resistance, a field survey was conducted to identify the factors that may control ARGs in sewer sediments and sewage. qPCR was performed for select ARGs and amplicon sequencing was performed for paired samples from combined and separate sanitary sewer systems. Metagenomic sequencing was performed on combined sewer sediments. The relative abundances of *sul*1, *tet*(O), *tet*(W), *erm*F, and *van*A were higher in wastewater compared to sewer sediments, while NDM-1 was greater in sewer sediment and *erm*F was similar between the two matrices. NDM-1 was observed in sewer sediment but rarely above detection in wastewater in this study. This may indicate that larger/more frequent wastewater samples are needed for detection and/or that retardation and/or accumulation in sewage sediment may need to be considered when interpreting wastewater-based epidemiology data for ARGs. Random forest analyses indicated that season and conductivity were important variables and to a lesser extent so were pH, TSS, heavy metals, and sewer type for explaining the variance of the ARGs. These

variables explained the 19-61% of the variance of *sul*1, *tet*(O), *tet*(G), and *tet*(W) quantified in wastewater. These variables performed less well for explaining the variance in sewer sediments (0.2-24%). Sewer sediment and wastewater had distinct microbial community structures and biomarkers for each are described. Metagenomics indicated that a high diversity of ARGs, including several of medical importance, were observed in the combined sewer sediment. This work provides insight into the complex sewer microbiome and the potential hazard posed by different sewer matrices.

Keywords: ARG, combined sewer, heavy metals, amplicon sequencing, metagenomics, sewage

1. Introduction

Antibiotic resistance is a public health threat and a comprehensive risk assessment requires an understanding of the fate of antibiotic resistance genes (ARGs) from environmental hotspots.[1] The potential role of sewage collection systems as one such hotspot is of interest, particularly given the risk posed by separate sanitary and combined sewer overflows (CSOs). In cities with combined sewer infrastructure, overflow events contribute to waterborne-disease outbreaks[2] and present a risk to public health by serving as a source of pathogens[3] and antibiotic resistant genes and bacteria.[4, 5] Understanding the potential for growth, retardation of transport, decay, horizontal gene transfer, and selection for antibiotic resistant microbes in sewers is of interest also for protecting the health of sewage workers, mitigating the impacts of sewer overflows, and interpreting sewage epidemiology data.[6]

More remains to be understood about the biological processes that occur in sewer deposits and their potential effects on the fate of antibiotic resistant bacteria. Microbes are present in wastewater, sewer biofilms,[7] and wastewater solids that settle during conveyance and collect at joints or other discontinuities in the sewer system.[8] The activity of microbes in sewers is apparent from studies of microbially induced corrosion,[9] dichlorination of polychlorinated biphenyls,[10, 11] and growth of fecal indicators in storm sewer sediments.[12] Retardation of pathogen transport during conveyance has also been indicated by a controlled release of inactivated polio virus.[13] However, limited efforts have been made to understand the factors that affect antibiotic resistant bacteria and their genes in sewer systems, particularly in the sewer sediment matrix. Observations of antibiotic concentrations above the predicted no-effect levels in sewers indicate the potential for selection.[6] However, a study simulating a hospital sewer line carrying fluoroquinolone antibiotics indicated that despite accumulation of these drugs, there

was no evidence of selection for fluoroquinolone resistance.[14] This observation was potentially due to sorption resulting in lower bioavailability and selection pressure. In contrast, correlations between some antimicrobial residues and heavy metals have been reported for select antibiotic resistances in wastewater influent [e.g., ciprofloxacin resistance with ciprofloxacin and arsenic concentrations[15]].

The aims of this study are to (1) quantify the loading and describe the diversity of ARGs in sewer sediments, (2) compare the loading of ARGs in sewer sediments to the wastewater being conveyed by the sewers, and (3) understand what factors are associated with elevated ARG loads in both matrices. To achieve these aims, a field survey was conducted collecting sewer sediments and wastewater from combined and separate sanitary sewer systems for different seasons (fall/winter versus summer). qPCR for select ARGs was performed on both matrices. Metagenomic sequencing (whole genome shotgun sequencing) was performed on the combined sewer sediment samples to better understand the diversity of ARGs present in this understudied matrix. Wastewater and sewer sediment quality data were collected and analyses including machine learning (i.e., random forest) were performed to determine the factors explaining the variance in ARG data and related to elevated ARGs relative abundances. The results of this study provide insight into the hazard posed by the sewer environment. Results can also provide insight into the impact of solids settling during conveyance on interpretation of sewage epidemiology data and the potential hazard imposed by the release of different matrices during overflow events.

2. Materials and methods

2.1 Sewer Sediment and Wastewater Sampling

A total of sixty samples were collected across five different sewers systems during two sampling campaigns. Sewer sediments and post-screen wastewater influent (i.e., two matrices) were each collected in triplicate. The collection systems all include municipal wastewater and have health care facilities in the catchment. The treatment facilities were all > 100 MGD design flow. To compare seasons, sampling was performed for Fall/Winter between September 2016 and January 2016 and for Summer between June 2017 and September 2017. Sewer sediment sampling locations were selected based on the presence of solids deposition sufficient to collect ~1 L, which varied by system, location and sampling day. Samples were collected from a variety of accumulation points in the sewer systems as described in Table 1. When possible, different locations within a given sewer system were selected for different sampling events. Samples from each system were collected during baseflow conditions at least one week apart. Sewer systems were labeled C1-C3 for the three combined sewer systems and S1-S2 for the two separate sanitary sewer systems sampled. On the same day of sediment sampling, a 24-hr time paced composite wastewater influent sample (2 L, collected via autosampler) was also collected from the corresponding downstream wastewater treatment plant. (In one exception, combined sewer system C3 wastewater was collected one day after sediment samples due to precipitation that may have influenced the planned 24-hr composite sample.) Field blanks consisting of autoclaved deionized water left open for the duration of sediment sampling were collected during each sampling season then preserved and analyzed using the same biomolecular techniques as the field samples. Samples were preserved in sterile sample containers on ice during transport to the lab where they were immediately processed.

2.1 Chemical Characterization of Field Samples

Sewer sediment samples were sieved < 2 mm and subsamples were analyzed for moisture content, pH, conductivity, particle size analysis, and select heavy metals. Moisture content was measured by drying aliquots to constant weight. Sediment pH and conductivity were measured according to standard methods.[16, 17] All chemical analyses were conducted in duplicate for 20% of samples for QA/QC. Particle size analysis was conducted by a sieve method. For each sample with sufficient volume, approximately 200-650 g of sediment were dried at ~100 ºC overnight to achieve constant mass. Samples were homogenized with a mortar and pestle and sieved through a series of stacked stainless-steel U.S. Standard sieves numbered 35, 60, 120, and 230 (ASTM E-11 Certified), which correspond to aperture sizes 500, 250, 125 and 63 µm. The stack was placed on a mechanical shaker for approximately 10 min and the dry mass passing through each sieve was measured. The fraction > 63 µm would be classified as sand and the fraction < 63 µm would be classified silt/clay if these samples were from soil. Biomass is expected to associate with the latter fraction.

Conductivity, pH, oxidation-reduction potential, total suspended solids (TSS) and volatile suspended solids (VSS) in wastewater samples were measured with a multimeter (Orion Star A329, Thermo Scientific) according to standard methods.[18, 19] Chemical oxygen demand (COD) was analyzed according to Hach Method 8000 with Hach COD vials (20-1500 mg/L range) and a DR2700 spectrophotometer (Hach, Loveland, CO). Sediment and wastewater samples were submitted to an outside lab (TestAmerica, Edison, NJ) for analysis of total arsenic, cadmium, copper, and nickel according to EPA Method SW846 6010C.[20] These metals were selected given that they have previously been associated with selection for antibiotic resistance in environmental matrices and bacterial cultures.[21-25] Metal concentrations are reported on a dry weight basis.

2.3 Biomolecular analyses

DNA was extracted from the field blanks, wastewater, and sewer sediment samples for qPCR, amplicon sequencing and (for select sewer sediment samples) metagenomic sequencing. Wastewater samples (~150 mL) and field blanks were concentrated on 0.22 μm nitrocellulose filters (Millipore Corporation, Billerica, MA). Filters or sieved sediment aliquots (~0.5 g wet weight), were added to DNA lysing tubes and stored at -20°C prior to DNA extraction. DNA extractions were conducted using a commercial kit (FastDNA Spin Kit for Soil, MP Biomedicals, Solon, OH) following the manufacturer's directions. qPCR was performed for select ARG [*sul*1,[26] *tet*(G),[27] *tet*(W),[28] *tet*(O),[28] *erm*F,[29] NDM-1,[30] *van*A[31]] and 16S rRNA gene copies for total bacterial population.[32] These sulfonamide and tetracycline resistance genes were selected because they are commonly observed in environmental matrices. NDM-1 was investigated because Carbapenem-resistant Enterobacteriaceae are classified as an "urgent threat" by the US CDC[33] and NDM-1 has emerged in multidrug resistant clinical infections, raising alarm.[34] *van*A is also a medically important gene because it encodes for resistance to vancomycin, considered a drug of last resort for antibiotic resistant infections. For *sul*1, *tet*(G), *tet*(W), *tet*(O), ermF, *van*A, and 16S rRNA, a standard SybrGreen (5 μL SsoFast EvaGreen, BioRad, Hercules, CA) chemistry with 0.4 μM forward and reverse primers, and 1 μL diluted (1:100) DNA extract in a 10 μL reaction was used. For NDM-1, a standard TaqMan protocol (5 μL SsoFast Probes Mix, BioRad, Hercules, CA) with 0.72 μM forward and reverse primers, 0.22 μM probe, and 1 μL diluted DNA extract in a 10 μL reaction was used. QA/QC on the qPCR included a no-template control on each plate, a seven-point calibration curve, and melt curve and/or gel electrophoresis to verify the specificity of qPCR products. qPCR calibration curve R^2 and efficiency values were 0.99 ± 0.01 and 87 ± 11 %, respectively. The limits of quantification

(LOQs) based on the lowest standard on the curve and factoring in dilutions were approximately 2.0×10^6 copies/g and 6.7×10^4 copies/mL for sediment and wastewater, respectively. All sample results were within the LOQs or not detected except for two *van*A results which were below the LOQ and therefore unquantifiable.

Amplicon sequencing was conducted to understand if differences were observed between sewer types/matrices that could be linked to ARG abundance and to define prominent community members in the different matrices. Amplicon sequencing (Illumina MiSeq, 300 bp, paired end) was performed on samples from both matrices targeting the V3-4 region of the 16S rRNA gene at a commercial lab (MrDNA, Shallowater, TX). A total of 40 samples were submitted for amplicon sequencing: two samples from each wastewater and sewer sediment per season (from the second and third fall/winter sampling events and first and second summer sampling events) for each of the five sewer systems. Sequences were analyzed using Quantitative Insights Into Microbial Ecology (QIIME) version 1.9 run through Oracle Virtual Box VM and the Rutgers School of Engineering High Performance Computing Cluster. Sequences were trimmed using Trimmomatic[35] and stitched using PandaSeq.[36] Sequences were otherwise analyzed following the tutorial for next generation sequencing.[37] Briefly, after extracting barcodes (barcode_extract.py), samples were demultiplexed and quality filtered (split_libraries_fastq.py), chimeras were removed (identify_chimeric_seqs.py), followed by assigned operational taxonomic units (OTUs, pick_de_novo_otus.py). Samples were rarefied at 55,029, the minimum number of sequences observed per sample. Rarefaction curves were generated using the mothur (v1.35.1) rarefaction.single function for order level OTU tables generated in QIIME (Fig. S1). Sequences are available in the NCBI database under Accession Numbers (SAMN10356326-SAMN10356393).

Metagenomic sequencing was performed on six sewer sediment samples to provide a deeper understanding of the diversity of ARGs observed in sewer sediments. For each combined sewer system, sediment DNA from two fall/winter sampling events were pooled and two summer sampling events were pooled to create one pooled sample representing each season. Pooled samples were submitted for Illumina HiSeq sequencing (MRDNA, Shallowater, TX). Libraries were prepared using Nextera DNA sample preparation kits (Illumina, San Diego, CA) following the manufacturer's user guide, pooled, diluted, and sequenced paired-end (150 bp) for 300 cycles. Sequences were analyzed using the MG-RAST pipeline.[38] Pipeline options included removal of artificial replicates produced by sequencing artifacts (dereplication), screening and removal of H. sapiens sequences (H.sapiens), and dynamic trimming for sequences with 5 bp below a 15 phred score. To investigate the presence of antibiotic resistance mechanisms, genes called as proteins in MG-RAST were queried against the Comprehensive Antibiotic Resistance Database [CARD[39] version 3.0.2] using BlastX with an E-value cutoff of 10^{-5}.[40] The threshold for amino acid identity was ≥90% and sequence alignment set to ≥25 amino acids.[41, 42] Resulting sequences were normalized to total clean reads (sequences passing quality control which included dereplication and trimming described above) per sample, reported as parts per million [(ARG reads / total clean reads) x 10^6]. Sequences are available under accession numbers [Reviewer Token: https://www.mg-rast.org/mgmain.html?mgpage=token&token=CbYyXYsUBH07ly69t_mpBoq09dFvoLsWO2FCu9cYuMJs6t4b2l Accession Numbers upon public release] (paired forward and reverse runs) in the MG-RAST database.

2.4 Statistical analyses

Statistical tests were performed in R. qPCR data were log-normalized before analysis. A random forest regression model (randomForest package) was used to determine the factors effecting the observed ARG concentrations in wastewater or sediment and the relative importance of the factors for each matrix. When a factor resulted in a negative mean square error (MSE) increase, the analysis was repeated excluding that factor as suggested by Mendez (2011).[43] Next, PERMANOVA (vegan package) was performed to test for differences in ARG relative abundances (ARG copies / 16S rRNA gene copies, allowing for cross-matrix comparisons) due to matrix, season, and/or sewer type. PERMANOVA was also used to test for differences in chemical parameters in a given matrix between season and sewer type (comparisons were not made between matrices for the chemical parameters because the units were not necessarily consistent). Arsenic and cadmium sediment concentrations were Box Cox transformed prior to analysis because concentrations were below detection in 20% of the samples. Next, correlations between ARGs (gene copies per g dry weight) and metals (concentration, dry weight), conductivity, or the < 63 μm sediment fraction were tested using Spearman rank tests.

To investigate whether shifts in community structure could be attributed to various sample characteristics, a Bray-Curtis similarity matrix was calculated on log-normalized subsampled (N= 55,029 sequences) OTU data at the class level followed by cluster analysis with a SIMPROF test and non-metric multidimensional scaling (nMDS) in PRIMER 7. ANOSIM was used to test for significance of community shifts ($\alpha<0.05$) between and across season, sewer type, and matrix. Richness of each sample, described as the number of OTUs observed for rarefaction at 55,029 reads was compared across season, matrix, and sewer type using a Wilcoxon rank sum test. The same test was used to compare Shannon Diversity indices between

samples. Additionally, the linear discriminant analysis effect size (LEfSe) tool[44] was used to identify biomarkers for the different matrices using the default settings.

Network analyses were performed as previously described[45] on the metagenome data to explore connections between ARG concentrations and the microbial community. Briefly, 16S rRNA gene taxonomy was obtained from the metagenomic data in MG-RAST using contigLCA and filtered for OTUs with abundance >0.5% in at least one sample. Next a matrix of pairwise t-tests was performed (psych package corr.test) with a BH correction for multiple comparisons. Results were plotted (igraph) using only data that resulted in an adjusted p-value <0.01 and rho>0.8. Diversity indices were calculated for the annotated ARGs including Shannon Diversity, Richness, Evenness, and inverse Simpson. The indices were compared across season and sewer system by a Friedman test.

3. Results

ARGs in Sewer Sediment and Wastewater and Explanatory Factors

qPCR was performed for select ARGs in paired wastewater and sewer sediment samples collected from combined and separate sanitary sewer systems across two seasons (Fig. 1). To describe the ARG relative abundance, PERMANOVA was performed for the descriptive factors (matrix, season, and sewer type) across wastewater and sewer sediment. Matrix (wastewater versus sewer sediment) resulted in significantly different relative abundances of *sul*1, *tet*(O), *tet*(W), *erm*F, and *van*A (all $p<0.002$) with wastewater having the higher relative abundances of these ARGs and sewer sediment having a higher relative abundance of NDM-1 ($p=0.001$). Season was associated with differences in select ARG relative abundances. Higher *sul*1 and

237 *tet*(G) relative abundances were observed in winter/fall compared to the summer (both $p<0.001$)
238 and higher *tet*(O) and *van*A relative abundances were observed in the summer (both $p<0.026$).
239 Sewer type resulted in different relative abundances of *sul*1, NDM-1, and *tet*(G), and *van*A (all
240 $p<0.043$), with some interactions by season. For example, *sul*1 and *tet*(G) were in greater
241 relative abundance for the combined sewer sediment and wastewater in the winter/fall. In the
242 summer the relative abundance of these genes varied by sewer type but not consistently: higher
243 relative abundances in separate sanitary sewer sediment than combined sewer sediment for *tet*(G)
244 and the opposite for *sul*1 in sewer sediment. NDM-1, which was only consistently above
245 detection in the sewer sediments, where it was observed at greater relative abundances in the
246 combined sewer sediment than separate sewer sediment for both seasons. When ARGs in the
247 sewer sediment were compared on a dry weight basis (i.e., gene copies per g sediment) rather
248 than 16S rRNA normalized, the separate sanitary sewer had higher concentrations of *sul*1, *erm*F,
249 *tet*(G), and *tet*(W) (all $p<0.006$, Fig. S2). (Seasonal patterns for ARGs in sewer sediment on a
250 dry weight basis were similar to those described above for 16S rRNA gene copy normalized
251 ARGs.)

252 Several potentially explanatory chemical parameters were measured in wastewater and sewer
253 sediment (Table 2 and Fig. S3). In wastewater, differences between combined and separate
254 sanitary sewers were observed for conductivity and ORP (both $p=0.023$) but not the other water
255 quality parameters. Conductivity and ORP were lower in the separate sanitary sewer
256 wastewater. For sewer sediment, significant differences were observed between combined and
257 separate sanitary sewers for pH ($p=0.029$), conductivity ($p=0.027$), copper ($p=0.027$), and arsenic
258 ($p=0.006$). Copper and conductivity were higher in the separate sewer sediments while pH and
259 arsenic were higher in the combined sewer sediments. Arsenic was also higher in sewer

sediment samples collected in the summer (p=0.02). Nickel was rarely observed in either matrix. No other differences were observed by sewer type or by season.

Correlations were tested between the heavy metal content of sewer sediments and wastewaters to provide insight into their potential to associate with ARGs. Further, antibiotic resistance has been correlated with heavy metals in other environments.[22, 23, 46] For wastewater, positive correlations were observed between *tet*(W), *tet*(O), and *erm*F concentrations (gene copies per volume) and arsenic (Spearman's r>0.39; p<0.03), and between *erm*F (gene copies per volume) and nickel (Spearman's r=0.37; p=0.04). In sewer sediment, a moderate positive correlation was observed between *tet*(G) (gene copies per g dry mass) and copper (Spearman's r = 0.45; p=0.01) and total metals measured (Spearman's r = 0.40; p=0.03), driven by copper's higher concentrations compared to the other metals. A weak positive correlation was observed between NDM-1 in sewer sediment and copper on a 16S rRNA-normalized basis (Spearman's r = 0.38; p = 0.04). Weak positive correlations were also observed between *van*A copies/g and copper and *van*A copies/g and total metals (Spearman's r = 0.42 and 0.39; p < 0.03). Negative correlations were observed between *tet*(O) (gene copies per g dry mass) and arsenic, cadmium, and nickel (Spearman's r < -0.46; p < 0.01).

Given that sewer sediments can be released to receiving surface water bodies during sewer overflows, the association of heavy metals with different sewer sediment particle size fractions (Fig. S4) was explored. Correlations between heavy metals and the <63 μm size fractions in sewer sediment were tested, given this fraction may be expected to sorb metals and allow for release to surface water during sewer overflow events. Positive correlations were observed between the fraction of <63 μM particles in the different sewer systems and the four metals analyzed (Spearman's r = 0.46-0.76; all p<0.04). Regression analysis indicated a strong linear

relationship between arsenic and the < 63µm size fraction (R^2=0.84) and less linearity for the other metals (R^2=0.17-0.22).

Relative Importance of Sewer Factors

The random forest regression was used to determine predictive factors for ARG relative abundance in sewer sediment and wastewater samples (Fig. 2). Random forest analysis can help identify important variables related to the response variable, provide insight into the discriminative ability of individual predictor variables, and identify a small number of variables sufficient for good prediction of the response variable.[47-49] Performing the random forest analysis on an individual matrix and the including all the factors measured for that matrix showed that at least three times the variance was explained for *sul*1, *tet*(G), and *tet*(W) in wastewater compared to sediment (Fig. 2). (The analysis was not performed for NDM-1 in wastewater given that the gene was observed in only four samples for that matrix.) After removing variables that resulted in negative MSE, the remaining wastewater variables explained a small (*tet*(W) 19.01% and *tet*(O) 20.0%) to moderate (*sul*1 53.1% and *tet*(W) 60.91%) amount of the variance in the relative abundance of these genes. Among the remaining variables, season was the most important factor (indicated by the greatest increase in MSE) followed by conductivity and sewer type (results are shown in Table 2) for *sul*1. Season and conductivity were the most important factors for *tet*(G), and *tet*(W), while copper was the most important factor for *tet*(O). Other contributing variables were pH, COD, TSS, VSS, arsenic, nickel, and/or sewer type.

The variables for the sediment random forests explained a small amount of the variance for *sul*1 (13.9%), *tet*(G) (18.4%), *tet*(O) (24.2%), and *tet*(W) (0.23%). Conductivity followed by season were the most important factors for *sul*1 and *tet*(G), while conductivity was most important for

306 *tet*(O) and pH was most important for *tet*(W). Other contributing variables for explaining the
307 variance of ARGs in sewer sediment were metals, moisture, and/or sewer type. Random forest
308 performed poorly for explaining the variance in *erm*F and *van*A relative abundances in both
309 matrices and poorly for NDM-1 in sediments. Overall, this analysis suggested gene-to-gene
310 differences and matrix effects in the estimative power of the parameters tested.

311 **Combined Sewer Sediment ARGs via Metagenomics**

312 To provide an understanding of the diversity of ARGs in combined sewer sediments, samples
313 from the three combined sewers were analyzed with metagenomic (whole genome shotgun)
314 sequencing. Sequences were annotated for ARGs and the associated antibiotic drug classes and
315 mechanisms using CARD. There were 659-882 ARGs annotated per sample and a Shannon
316 diversity index of 4.94 to 5.10, richness ranged from 68.1 to 89.1, evenness from 0.74 to 0.76,
317 and an inverse Simpson of 0.98-0.99, with no significant differences by sewer or season (all
318 $p>0.08$). Cluster analysis on the ARG profiles indicated clustering by sewer system for C1 at
319 91.2% similarity and C3 with 88.8% similarity (Fig. 3a). Significant differences in the clustering
320 by ARG profile were observed between the sewer systems and for C2 between the seasons (all
321 $p=0.01$, SIMPROF). The most prevalent drug classes observed were multidrug (35±3% of
322 annotations), macrolide (13±1%), and tetracycline (10±1%). Antibiotic efflux was the most
323 commonly annotated resistance mechanism (60-65% of ARG annotations) followed by antibiotic
324 target alteration (16-21%) and antibiotic target protection (10-11%) (Fig. S5). All of the ARGs
325 that were detected with qPCR were detected in the metagenomes. Of interest given the relevance
326 for public health, mcr-1 (encoding for colistin resistance) was detected in five out of six
327 combined sewer sediment samples at 0.5 to 2.3ppm of total reads. *mec*A (encoding for

methicillin resistant *S. aureus* at 6-8 ppm of total reads) and *van*A (encoding for vancomycin resistance at 9-33 ppm of total reads) were observed in all the combined sewer sediment samples.

Microbial Community Analysis via Amplicon Sequencing and Metagenomics

Microbial community analysis was performed on the amplicon sequencing data to determine if there were differences in community structure between matrix, season, and/or sewer type, given that these could help explain any observed differences in ARG relative abundances. As expected, the microbial community structure differed by matrix: wastewater was significantly different from sewer sediment ($p=0.001$, ANOSIM). Richness, determined by number of observed OTUs, was lower in wastewater compared to sewer sediments (138 ± 19 vs 183 ± 61 OTUs per sample; $p=3.9\times10^{-4}$). Shannon diversity index was similar in the two matrices: 2.6 ± 0.3 for wastewater and 2.9 ± 0.4 for sewer sediment. The wastewater samples clustered with greater than 70.4% similarity and eight of the out ten sample pairings collected from the same system in the same season (seasonal replicates), clustered without significant differences (all $p>0.063$). The sewer sediment samples clustered with 47.2% similarity and only four of the ten seasonal replicate pairs clustered without significant difference (all $p=1.0$). Sewer sediments collected from S2 clustered more closely with wastewater (70.4% similarity) than the other sewer sediments. Sewer sediments from C3 formed a separate cluster from the other sewer sediments, the C3 cluster had 60.5% similarity. Neither microbial community structure (both $p>0.20$; ANOSIM) nor richness (both $p>0.25$) were significantly different across season or sewer type.

Dominant microbial classes in the amplicon sequencing data for wastewater and sewer sediment were also evaluated. Classes that were detected at abundances >0.01 are summarized in Fig. 4. Actinobacteria, Bacteroidia, Flavobacteriia, Bacilli, Clostridia, and classes of Proteobacteria

were detected most frequently across samples. Archaea and various classes from the Chloroflexi phylum were detected in sewer sediment but not in wastewater samples. LEfSE biomarker analysis of wastewater and sewer sediment revealed different biomarkers for the matrices. Wastewater was characterized primarily by Proteobacteria, Gammaproteobacteria, and Epsilonproteobacteria. Sewer sediment was characterized primarily by Archaea, Euryarchaeota, and Bacteroidia. The three matrices did not have any biomarkers in common above a linear discriminant analysis (LDA) score of 3 (Fig. S6).

Network analysis was performed on the combined sewer metagenome data. Linkages were determined between the relative abundance of ARG annotations based on the CARD database and genus level 16S rRNA gene annotations resulting from the MG-RAST analysis (Fig. S7). Among the remaining genera were those that contain pathogens and/or opportunistic pathogens (along with commensal organisms) including Aeromonas, Bacillus, Bacteroides, among others. The majority of the antibiotic resistance mechanism remaining were for efflux followed by target alternation and target protection. Most genera had linkage to a single ARG therefore ARGs expected to be seen together, e.g., those located in a single operon, were not observed as linked to the same genera.

4. Discussion

Loading and Diversity of ARGs in Sewer Sediments and Compared to Wastewater

ARGs (*sul*1, *tet*(G), *tet*(O), *tet*(W), *erm*F, and vanA) were observed in all sewer sediment samples at similar or lower relative abundances compared to wastewater, while NDM-1 was observed in sewer sediment samples and rarely above detection in wastewater. Sewer sediment and wastewater ARGs are related given the exchange of solids between the two matrices: solids from wastewater can settle at junctions and locations of low flow and settled solids may erode and be re-suspended during high flow and CSO events.[50, 51] These results suggest that *sul*1, *tet*(G), *tet*(O), *tet*(W), and *erm*F, and *van*A ARGs quantified here did not necessarily accumulate in sewer sediment compared to the wastewater reaching the treatment plant. NDM-1 was more commonly observed in the sewer sediment, potentially due to preferential partitioning, selection within the sediment, decay within the wastewater during conveyance, or temporal differences between the sediment measured (representing accumulation over a period of days or longer) and the 24-hr composite influent samples. For NDM-1 this may indicate a potential hazard for this gene and matrix combination. NDM-1 is a gene of high risk given that multiple hosts can use the gene to confer resistance and it has been found on broad range plasmids.[52-55] NDM-1 has previously been observed in wastewater, hospital wastewater, and surface waters receiving wastewater effluent and feces.[56-59] Interestingly, samples collected from C3, the sediment stockpile in the CSO detention tank that is sometimes dry, exhibited consistent ARG abundances with the other sewer sediments that are consistently in contact with wastewater. Thus, there is evidence that ARGs persist in sewer sediments even without constant exchanges with the mobile bed load and wastewater matrix.

Of particular interest is the abundance and diversity of ARGs in combined sewer sediments given the likelihood of release to surface water environment without treatment. Combined sewer sediments contained a diverse range of ARGs, with more than double the Shannon diversity indices for metagenomes from polluted surface water sediments [2.07-2.89[60]] were analyzed using a similar pipeline and an older version of CARD. Across three combined sewer systems, multidrug resistance and efflux pumps were the most prevalent antibiotic resistance mechanisms, similar to unimpacted estuarine sediments.[61] The most prevalent drug classes observed in sewer sediments were multidrug, tetracycline, and macrolide. This observation is similar to the dominant drug classes reported for wastewater metagenomes. Multidrug resistance was a dominant drug class annotated (20.2%) in WWTP influent in Hong Kong along with tetracycline (23.1%) and aminoglycoside (14.8%).[62] In municipal wastewater in China, the most prevalent ARGs in order of abundance were to tetracycline, β-lactams, macrolides, aminoglycosides, and multidrug.[63] In contrast to sewer sediments, anthropogenically impacted river sediments ARG annotations were dominated by the aminoglycoside and multidrug,[60] amphenicol, macrolide, tetracycline [compared to unimpacted sediment[64]], and sulfonamide, fluoroquinolone, and aminoglycoside resistance genes.[61]

Factors Associated with Elevated ARGs in Sewer Sediment and Wastewater

To better understand factors associated with elevated ARGs in the sewer environment, a random forest regression was performed to determine variables that helped explain the variance in the relative abundance of the genes measured by qPCR. The analysis indicated a small to moderate amount of the variance in wastewater *sul*1, *tet*(G), *tet*(O), and *tet*(W) was explained primarily by season and conductivity with contributions as well from pH, TSS, metals, and sewer type. In contrast, the variables included in the random forest explained little to a small amount of the

variance of these ARGs in sewer sediment with the most important factors again being conductivity and season with contributions as well from pH, metals, moisture, and sewer type. Wastewater is a better studied matrix than sewer sediments and it is possible other factors not measured here may be important for driving the sewer sediment ARG loading. Examples may include sewer sediment age, concentrations of other selecting agents (e.g., sorbed antibiotics), and/or exchange rates with nutrients or selecting factors in the mobile bed load.

Seasonal differences between ARGs were observed. Overall higher concentrations were seen in the fall and winter for *sul*1, *tet*(G) and *tet*(W). The seasonal variations observed are not thought to be due to changes in sewer water temperature, which was not previously shown to impact the sewage microbiome.[65] Antibiotic use is up to three times higher in winter[66] and may be an ultimate cause of observed differences given that antibiotics use has been correlated to clinical antibiotic resistant infections.[67]

It was hypothesized that ARGs would be similar in the wastewaters from different sewer systems given that they were collected during periods without rain. We also hypothesized that ARG concentrations would be different in the sewer sediments from combined versus separate sewer systems given that combined sewers would be expected to convey more stormwater (although, separate systems are subject to varying amounts of infiltration and inflow). Based on the random forest analysis of 16S rRNA gene copy normalized ARGs there was some contribution of the sewer system to both the wastewater and sewer sediment, although there were interactions with other factors and inconsistent patterns of which system type contained higher levels. While differences were not necessarily consistent between sewer systems for ARGs normalized to 16S rRNA gene copies, concentrations (on a dry weight basis) of all ARGs genes in separate sewer

system sediments were higher than in combined sewer system sediments. This can be explained by the higher moisture content of the separate sewer system sediments collected.

Metals were of interest because antibiotic resistance has been associated with heavy metals in wastewater[68] and other environments.[22, 23, 46] The only significant positive correlations between heavy metals and ARGs were for *tet*(G) and NDM-1 versus copper in sewer sediments. The positive correlation for *tet*(G) could be due to coselection, as suggested by previous reports of plasmids carrying both copper and tetracycline resistance.[69] Otherwise, the metal concentrations detected in this study may not have been high enough to trigger coselection for antibiotic resistance.[21, 24] Correlations between metals and silt/clay indicate that sorption may be an important mechanism for metals retained in sewer sediment. Metals were detected less frequently in wastewater than sediment samples. Only copper was detected consistently in the wastewater samples and correlations were not observed between it and the ARGs measured with qPCR.

Microbial Community Analysis via Amplicon Sequencing and Metagenomics

Microbial communities were studied in parallel with analysis of ARGs because differences in ARGs could be attributable to different microbial community members present in different matrixes particularly if these are associated with different characteristic or intrinsic resistances.[70]

Significantly different microbial community structures were observed in sewer sediments compared to wastewater. The sewer sediment microbial community was most similar to the wastewater community in S2, which was collected from a wet well. The sewer sediment least similar to wastewater and other samples came from system C3 which were collected from a CSO retention basin stock pile. Therefore, the fact that these samples were not in consistent contact with wastewater and the mobile bed load may have resulted in a shift in the microbial

community structure given that the paired wastewater samples from this system had much greater similarity to wastewater samples from other systems. The sewer sediment samples were less similar to one another than the wastewater samples, which was potentially a function of differences in sampling (wastewater was collected as 24-hr composites, sewer sediments were collected as composites from a single point of accumulation within the sewer system and likely represent a greater range of time, although aging the sewer sediment was beyond the scope of this project). Interestingly, the sewer sediment microbial community from combined sewer systems did not necessarily cluster separately from that of the separate sewer sediment. This is despite the fact that the combined sewer sediments potentially had greater inputs from the storm water microbial community in addition to wastewater compared to separate sewer sediment. This may be due to resuspension of sewer sediments during storms (thus not allowing for sufficient deposition of storm water sediments), infiltration and inflow in separate sewers resulting in a similar contribution of stormwater microbes, and/or the greater microbial loading and diversity in wastewater compared to stormwater thereby obscuring any stormwater sediment signal in the combined sewers.

Implications for Sewer Operations, CSOs, and Interpreting Sewage Epidemiology Data

The data collected here is of interest for understanding the potential hazard presented by the different sewer matrices during overflows and maintenance as well as for interpreting sewage epidemiology data. With respect to sewage overflows, sewer sediments contained ARGs and heavy metals that can be released to the environment if mobilized and not treated during wet weather flows. The association of metals with solids indicates that end-of-pipe treatment methods that remove settleable solids have the potential to remove these contaminants from the effluent during CSO events.[71] The detection of ARGs in sewer sediments at abundances similar

480 to or higher than wastewater highlights that sewer sediments can be a source of microbial
481 contaminants released during CSO events. The relative importance of these matrices may be a
482 function of the volume of each matrix and potential differences in fate and transport upon release
483 to surface water. Based on comparison of data from different seasons, the ARG hazard from
484 release of wastewater and sewer sediments is similar or higher during fall and winter compared
485 to summer. However, in the summer, downstream factors such as warmer water temperature in
486 summer can still promote outbreaks of water-borne bacteria[72] and the higher likelihood of
487 contact during recreation in the summer should be taken into consideration to assess the overall
488 hazard posed by sewer overflows.

489 The presence of ARGs in the sewer sediments indicates that a portion of ARGs in wastewater are
490 retarded during transport, which was expected based on previous research monitoring viruses.[13]
491 A diverse range of ARGs were present in combined sewer sediments and several ARGs of
492 medical importance were observed including NDM-1, which was quantified using qPCR.
493 Although it was assumed that sediment deposition would be widespread, accumulation of sewer
494 sediment was not always observed. Numerous manhole locations (systems C1 and C2) were
495 rejected because sediment was not found at the bottom of the pipe, while on the same day in the
496 vicinity, accumulation was appreciable. Systems with sewer solids accumulation that perform
497 maintenance (e.g., jetting) will mobilize these ARGs and the associated heavy metals. More
498 information about the genetic context and host[70] as well as any potential differences in exposure
499 rates to the two matrices during maintenance would be necessary to compare the relative risk
500 posed by the different matrices. Further study would also be needed to understand if and how
501 retardation of ARGs in sewer sediments may impact the interpretation of sewage epidemiology
502 data for different systems. Based on this study, sewer sediment was more likely to have NDM-1

503 at detectable levels. This indicates that either large sample sizes or more frequent sampling
504 would be needed for wastewater, otherwise the burden of this gene could be underestimated in
505 wastewater-based epidemiological studies focuses on wastewater without considering retardation
506 or preferential accumulation in sewer sediments.

507 **5. Conclusions**

508 This research provides insight towards understanding an understudied potential hotspot for
509 ARGs: sewers. Sewer sediments were found to contain NDM-1 more frequently than
510 wastewater but not the other ARGs quantified in this study: wastewater contained a higher
511 relative abundance of sul1, *tet*(G), and *tet*(W). The chemical parameters measured and factors
512 considered for this study explained the variance of some of these genes moderately well in
513 wastewater but at most a small portion of the variance in sewer sediments. Important variables
514 for ARGs in wastewater included season and conductivity, followed by pH, TSS, metals, and
515 sewer type. The microbial community structures were different between the two matrices, which
516 may explain some differences in these ARGs relative abundances. Metagenomic results
517 indicated that the sewer system was more important than season for the ARG profiles observed
518 in combined sewer sediments. A high diversity of ARGs were observed and several ARGs of
519 medical importance were observed, highlighting a potential hazard. This work can help inform
520 mitigation strategies for sewer overflows and preventative sewer maintenance. Observations of
521 ARGs and heavy metals in sewer sediments indicate that there is retardation during transport and
522 their potential for release during sewer overflows if these sediments are eroded. The fact that
523 NDM-1 was above detection in sewer sediments and few wastewater samples may indicate
524 potential retardation/preferential accumulation in sewer sediments or temporal variation in the
525 wastewater that was captured in settled sediment not apparent in the 24-hr wastewater

composites that should be considered when interpreting wastewater-based epidemiology data for ARGs. A better understanding of system hydraulics and other factors such as exposure rates to the different matrices for sewage workers and the public as well as genetic context and host of the ARGs would help inform the potential risk posed by sewer sediments and the need to account for the impact of settling/resuspension on interpretation of sewage epidemiology data. Future studies may wish to include wastewater sampling within the sewer pipe, seek better information on the accumulation rate and age of sewer sediments (not available here), and include monitoring for a broader range of ARGs.

Acknowledgements

Thanks to our utility partners for providing wastewater samples and access to sewer sampling locations. Funding for this project was provided by a grant from the National Science Foundation (grant number 1510461). Additional support was provided by the Eagleton Fellowship Program to AE and NIH Bridges to the Doctorate Program (grant number R25GM058389).

References

1. N. J. Ashbolt, A. Amézquita, T. Backhaus, P. Borriello, K. K. Brandt, P. Collignon, A. Coors, R. Finley, W. H. Gaze and T. Heberer, Human health risk assessment (HHRA) for environmental development and transfer of antibiotic resistance, *Environ Health Perspect*, 2013, **121**, 993.
2. J. A. E. ten Veldhuis, F. H. L. R. Clemens, G. Sterk and B. R. Berends, Microbial risks associated with exposure to pathogens in contaminated urban flood water, *Water Res*, 2010, **44**, 2910-2918.
3. E. Donovan, K. Unice, J. D. Roberts, M. Harris and B. Finley, Risk of gastrointestinal disease associated with exposure to pathogens in the water of the Lower Passaic River, *Appl Environ Microbiol*, 2008, **74**, 994-1003.
4. S. Young, A. Juhl and G. D. O'Mullan, Antibiotic-resistant bacteria in the Hudson River Estuary linked to wet weather sewage contamination, *J Water Health*, 2013, **11**, 297-310.
5. A. Eramo, H. Delos Reyes and N. L. Fahrenfeld, Partitioning of antibiotic resistance genes and fecal indicators varies intra and inter-storm during combined sewer overflows, *Front Microbiol*, 2017a, **8**, 1-13.
6. N. Fahrenfeld and K. J. Bisceglia, Emerging investigators series: sewer surveillance for monitoring antibiotic use and prevalence of antibiotic resistance: urban sewer epidemiology, *Environ. Sci.: Water Res. Technol.*, 2016, **2**, 788-799.
7. R. Crabtree, Sediments in Sewers, *Water and Environment Journal*, 1989, **3**, 569-578.
8. R. Ashley and R. Crabtree, Sediment origins, deposition and build-up in combined sewer systems, *Water Sci Technol*, 1992, **34**, 1-12.
9. R. L. Islander, J. S. Devinny, F. Mansfeld, A. Postyn and H. Shih, Microbial ecology of crown corrosion in sewers, *J Environ Eng*, 1991, **117**, 751-770.
10. L. A. Rodenburg, S. Du, D. E. Fennell and G. J. Cavallo, Evidence for widespread dechlorination of polychlorinated biphenyls in groundwater, landfills, and wastewater collection systems, *Environ Sci Technol*, 2010, **44**, 7534-7540.
11. V. Krumins, W. Sun, J. Guo, S. Capozzi, D. E. Fennell and L. A. Rodenburg, Sewer sediment bacterial communities suggest potential to transform persistent organic pollutants, *Water Environ. Res.*, 2018, **90**, 2022-2029.
12. R. P. Marino and J. J. Gannon, Survival of fecal coliforms and fecal streptococci in storm drain sediment, *Water Res.*, 1991, **25**, 1089-1098.
13. T. Hovi, M. Stenvik, H. Partanen and A. Kangas, Poliovirus surveillance by examining sewage specimens. Quantitative recovery of virus after introduction into sewerage at remote upstream location, *Epidemiol. Infect.*, 2001, **127**, 101-106.
14. P.-Å. Jarnheimer, J. Ottoson, R. Lindberg, T.-A. Stenström, M. Johansson, M. Tysklind, M.-M. Winner and B. Olsen, Fluoroquinolone antibiotics in a hospital sewage line; Occurrence, distribution and impact on bacterial resistance, *Scand. J. Infect. Dis.*, 2004, **36**, 752-755.
15. A. R. Varela, S. André, O. C. Nunes and C. M. Manaia, Insights into the relationship between antimicrobial residues and bacterial populations in a hospital-urban wastewater treatment plant system, *Water Res.*, 2014, **54**, 327-336.
16. United States Department of Agriculture, Diagnosis and improvement of saline and alkali soils. 1954, **Agricultural Handbook No. 60**.

17. US Environmental Protection Agency, EPA Test Method 9056D: Soil and waste pH. 2004.

18. US Environmental Protection Agency, Field Measurement of Oxidation Reduction Potential (ORP). 2013.

19. American Water Works Association, Standard Methods for the Examination of Water and Wastewater. 2012.

20. US Environmental Protection Agency, Method 6010C: Inductively coupled plasma-atomic emission spectrometry. 2007.

21. S. Chen, X. Li, G. Sun, Y. Zhang, J. Su and J. Ye, Heavy metal induced antibiotic resistance in bacterium LSJC7, *Int J Molec Sci*, 2015, **16**, 23390-23404.

22. C. Seiler and T. U. Berendonk, Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture, *Front Microbiol*, 2012, **3**, 399.

23. X. Ji, Q. Shen, F. Liu, J. Ma, G. Xu, Y. Wang and M. Wu, Antibiotic resistance gene abundances associated with antibiotics and heavy metals in animal manures and agricultural soils adjacent to feedlots in Shanghai; China, *J Hazard Mat*, 2012, **235**, 178-185.

24. E. Gullberg, L. M. Albrecht, C. Karlsson, L. Sandegren and D. I. Andersson, Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals, *MBio*, 2014, **5**, e01918-01914.

25. C. Baker-Austin, M. S. Wright, R. Stepanauskas and J. McArthur, Co-selection of antibiotic and metal resistance, *Trends Microbiol*, 2006, **14**, 176-182.

26. R. Pei, S.-C. Kim, K. H. Carlson and A. Pruden, Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG), *Water Res.*, 2006, **40**, 2427-2435.

27. R. I. Aminov, J. C. Chee-Sanford, N. Garrigues, B. Teferedegne, I. J. Krapac, B. A. White and R. I. Mackie, development, validation, and application of PCR primers for detection of tetracycline efflux genes of Gram-negative bacteria, *Appl Environ Microbiol*, 2002, **68**, 1786-1793.

28. R. I. Aminov, N. Garrigues-Jeanjean and R. I. Mackie, Molecular ecology of tetracycline resistance: development and validation of primers for detection of tetracycline resistance genes encoding ribosomal protection proteins, *Appl Environ Microbiol*, 2001, **67**, 22-32.

29. J. Chen, Z. Yu, F. C. Michel, T. Wittum and M. Morrison, Development and application of real-time PCR assays for quantification of erm genes conferring resistance to macrolides-lincosamides-streptogramin B in livestock manure and manure management systems, *Appl. Environ. Microbiol.*, 2007, **73**, 4407-4416.

30. S. M. Diene, N. Bruder, D. Raoult and J.-M. Rolain, Real-time PCR assay allows detection of the New Delhi metallo-β-lactamase (NDM-1)-encoding gene in France, *Int. J. Antimicrob. Agents*, 2011, **37**, 544-546.

31. S. Dutka-Malen, S. Evers and P. Courvalin, Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR, *J Clinical Microbiol* 1995, **33**, 24-27.

32. G. Muyzer, E. C. De Waal and A. G. Uitterlinden, Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis or polymerase chain reaction-amplified genes coding for 16S rRNA, *Appl Environ Microbiol* 1993, **59**, 695-700.

33. Center for Disease Control U.S. (CDC), Antibiotic Resistance Threats in the United States, 2019. Atlanta, GA: U.S. Department of Health and Human Services., 2019.

34. K. K. Kumarasamy, M. A. Toleman, T. R. Walsh, J. Bagaria, F. Butt, R. Balakrishnan, U. Chaudhary, M. Doumith, C. G. Giske, S. Irfan, P. Krishnan, A. V. Kumar, S. Maharjan, S. Mushtaq, T. Noorie, D. L. Paterson, A. Pearson, C. Perry, R. Pike, B. Rao, U. Ray, J. B. Sarma, M. Sharma, E. Sheridan, M. A. Thirunarayan, J. Turton, S. Upadhyay, M. Warner, W. Welfare, D. M. Livermore and N. Woodford, Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study, *Lancet Infect Diseases*, 2010, **10**, 597-602.

35. A. M. Bolger, M. Lohse and B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, 2014, **30**, 2114-2120.

36. A. P. Masella, A. K. Bartram, J. M. Truszkowski, D. G. Brown and J. D. Neufeld, PANDAseq: paired-end assembler for illumina sequences, *BMC Bioinformatics*, 2012, **13**, 31.

37. QIIME, 454 Overview Tutorial: de novo OTU picking and diversity analyses using 454 data, http://qiime.org/tutorials/tutorial.html, (accessed January 11, 2019).

38. F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens and A. Wilke, The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes, *BMC bioinformatics*, 2008, **9**, 386.

39. B. Jia, A. R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K. K. Tsang, B. A. Lago, B. M. Dave, S. Pereira and A. N. Sharma, CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database, *Nucleic Acids Res*, 2016, gkw1004.

40. Y. Yang, B. Li, F. Ju and T. Zhang, Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach, *Environ Sci Technol*, 2013, **47**, 10197-10205.

41. E. Kristiansson, J. Fick, A. Janzon, R. Grabic, C. Rutgersson, B. Weijdegård, H. Söderström and D. J. Larsson, Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements, *PloS one*, 2011, **6**, e17038.

42. T. Zhang, X.-X. Zhang and L. Ye, Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge, *PloS one*, 2011, **6**, e26041.

43. G. Mendez and S. Lohr, Estimating residual variance in random forest regression, *Comput Stat Data Anal*, 2011, **11**, 2937-2950.

44. N. Segata, J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W. S. Garrett and C. Huttenhower, Metagenomic biomarker discovery and explanation, *Genome Biol*, 2011, **12**, R60.

45. D. Baral, B. I. Dvorak, D. Admiraal, S. Jia, C. Zhang and X. Li, Tracking the sources of antibiotic resistance genes in an urban stream during wet weather using shotgun metagenomic analyses, *Environ Sci Technol*, 2018, **52**, 9033-9044.

46. A.-D. Li, L.-G. Li and T. Zhang, Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants, *Front Microbiol*, 2015, **6**.

47. R. Genuer, J.-M. Poggi and C. Tuleau-Malot, Variable selection using random forests, *Pattern Recognition Lett*, 2010, **31**, 2225-2236.

48. K. J. Archer and R. V. Kimes, Empirical characterization of random forest variable importance measures, *Comput Stat Data Anal*, 2008, **52**, 2249-2260.

49. U. Grömping, Variable importance assessment in regression: linear regression versus random forest, *American Stat*, 2009, **63**, 308-319.

50. R. Ashley and M. Verbanck, Mechanics of sewer sediment erosion and transport, *J Hydraul Res,* 1996, **34**, 753-770.

51. J. Passerat, N. K. Ouattara, J. M. Mouchel, V. Rocher and P. Servais, Impact of an intense combined sewer overflow event on the microbiological water quality of the Seine River, *Water Res*, 2011, **45**, 893-903.

52. P. L. Ho, W. U. Lo, M. K. Yeung, C. H. Lin, K. H. Chow, I. Ang, A. H. Y. Tong, J. Y.-J. Bao, S. Lok and J. Y. C. Lo, Complete sequencing of pNDM-HK encoding NDM-1 Carbapenemase from a multidrug-resistant *Escherichia coli* strain isolated in Hong Kong, *PLOS ONE*, 2011, **6**, e17989.

53. T. R. Walsh, J. Weeks, D. M. Livermore and M. A. Toleman, Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study, *Lancet Infectious Dis*, 2011, **11**, 355-362.

54. A. Potron, L. Poirel and P. Nordmann, Plasmid-mediated transfer of the blaNDM-1 gene in Gram-negative rods, *FEMS Microbiol. Lett.*, 2011, **324**, 111-116.

55. B. Jovcic, Z. Lepsanovic, V. Suljagic, G. Rackov, J. Begovic, L. Topisirovic and M. Kojic, Emergence of NDM-1 metallo-β-lactamase in *Pseudomonas aeruginosa* clinical isolates from Serbia, *Antimicrob. Agents Chemother.*, 2011, **55**, 3929-3931.

56. Z. S. Ahammad, T. R. Sreekrishnan, C. L. Hands, C. W. Knapp and D. W. Graham, Increased waterborne blaNDM-1 resistance gene abundances associated with seasonal human pilgrimages to the Upper Ganges River, *Environ. Sci. Technol.*, 2014, DOI: 10.1021/es405348h.

57. F. Yang, D. Mao, H. Zhou, X. Wang and Y. Luo, Propagation of New Delhi Metallo-β-lactamase genes (blaNDM-1) from a wastewater treatment plant to its receiving river, *Environ. Sci. Technol. Lett.*, 2016, **3**, 138-143.

58. C. Zhang, S. Qiu, Y. Wang, L. Qi, R. Hao, X. Liu, Y. Shi, X. Hu, D. An, Z. Li, P. Li, L. Wang, J. Cui, P. Wang, L. Huang, J. D. Klena and H. Song, Higher Isolation of NDM-1 Producing Acinetobacter baumannii from the Sewage of the Hospitals in Beijing, *PLOS ONE*, 2013, **8**, e64857.

59. Y. Luo, F. Yang, J. Mathieu, D. Mao, Q. Wang and P. J. J. Alvarez, Proliferation of Multidrug-Resistant New Delhi Metallo-β-lactamase genes in municipal wastewater treatment plants in Northern China, *Environ. Sci. Technol. Lett.*, 2014, **1**, 26-30.

60. N. Fahrenfeld, H. Delos Reyes, A. Eramo, D. M. Akob, A. Mumford and I. M. Cozzarelli, Shifts in microbial community structure and function in unconventional oil and gas waste impacted surface waters revealed by metagenomics., *Sci. Total Environ.*, 2017, **580**, 1205-1213.

61. H. Chen, X. Bai, L. Jing, R. Chen and Y. Teng, Characterization of antibiotic resistance genes in the sediments of an urban river revealed by comparative metagenomics analysis, *Sci Total Environ*, 2018.

62. Y. Yang, B. Li, S. Zou, H. H. Fang and T. Zhang, Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach, *Water Res*, 2014, **62**, 97-106.

63. J. Tang, Y. Bu, X.-X. Zhang, K. Huang, X. He, L. Ye, Z. Shan and H. Ren, Metagenomic analysis of bacterial community composition and antibiotic resistance genes in a wastewater treatment plant and its receiving surface water, *Ecotox Environ Safety*, 2016, **132**, 260-269.

64. N. P. Marathe, C. Pal, S. S. Gaikwad, V. Jonsson, E. Kristiansson and D. J. Larsson, Untreated urban waste contaminates Indian river sediments with resistance genes to last resort antibiotics, *Water Res*, 2017, **124**, 388-397.

709 65. J. L. Vandewalle, G. W. Goetz, S. M. Huse, H. G. Morrison, M. L. Sogin, R. G. Hoffmann, K.
710 Yan and S. L. McLellan, Acinetobacter, Aeromonas and Trichococcus populations dominate the
711 microbial community within urban sewer infrastructure, *Environ Microbiol*, 2012, **14**, 2538-2552.
712 66. S. Coutu, L. Rossi, D. A. Barry, S. Rudaz and N. Vernaz, Temporal variability of antibiotics
713 fluxes in wastewater and contribution from hospitals, *PLoS One*, 2013, **8**, e53592.
714 67. L. Sun, E. Y. Klein and R. Laxminarayan, Seasonality and temporal correlation between
715 community antibiotic use and resistance in the United States, *Clin Infect Dis*, 2012, **55**, 687-694.
716 68. P. Gao, M. Munir and I. Xagoraraki, Correlation of tetracycline and sulfonamide antibiotics with
717 corresponding resistance genes and resistant bacteria in a conventional municipal wastewater
718 treatment plant, *Sci Total Environ*, 2012, **421-422**, 173-183.
719 69. E. Gullberg, L. M. Albrecht, C. Karlsson, L. Sandegren and D. I. Andersson, Selection of a
720 multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals, *MBio*, 2014, **5**,
721 e01918-01914.
722 70. J. L. Martínez, T. M. Coque and F. Baquero, What is a resistance gene? Ranking risk in
723 resistomes, *Nat Rev Microbiol*, 2015, **13**, 116-123.
724 71. M. A. Wilson, O. Mohseni, J. S. Gulliver, R. M. Hozalski and H. G. Stefan, Assessment of
725 hydrodynamic separators for storm-water treatment, *J Hydraul Eng*, 2009, **135**, 383-392.
726 72. A. Sterk, F. M. Schets, R. Husman, A. Maria, T. Nijs and J. F. Schijven, Effect of climate change
727 on the concentration and associated risks of *Vibrio spp*. in Dutch recreational waters, *Risk*
728 *Analysis*, 2015, **35**, 1717-1729.
729

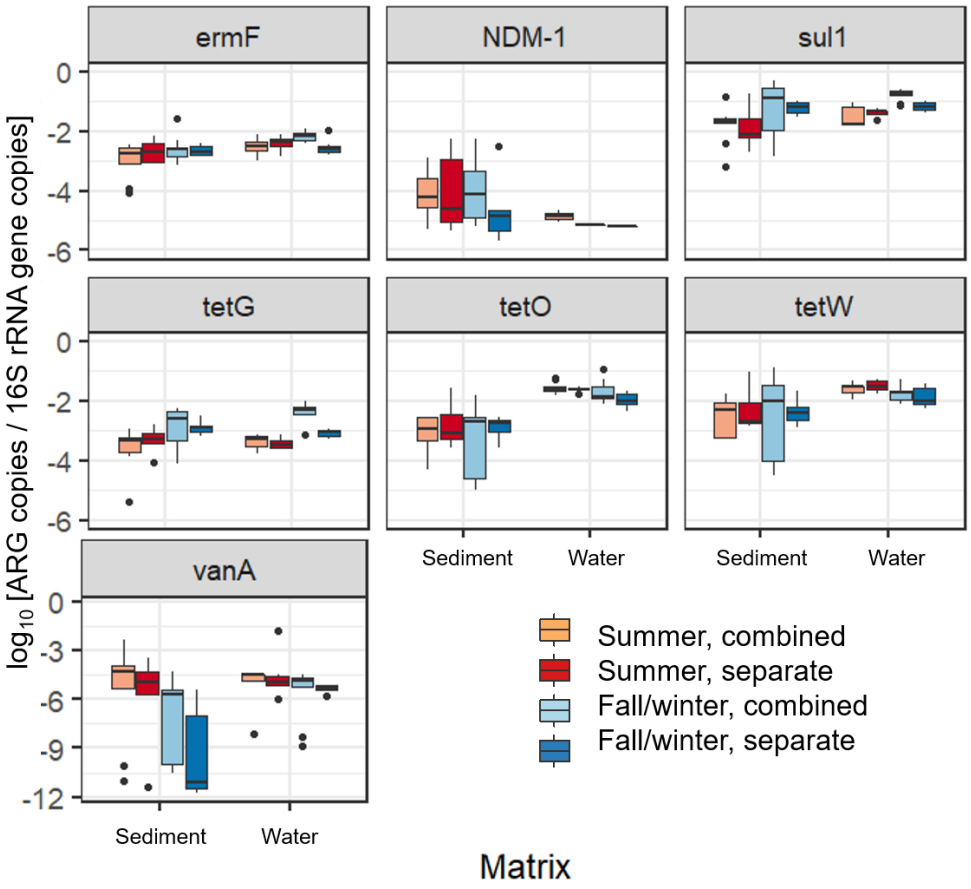


Figure 1. ARG copies normalized to 16S rRNA gene copies for sewer sediment ("Sediment") and wastewater ("Water"). Samples were collected from either combined or separate sanitary sewer systems in triplicate during each season (each box represents N=9 for combined and N=6 for separate sanitary sewers). Note the scale is different for vanA.

195x188mm (150 x 150 DPI)

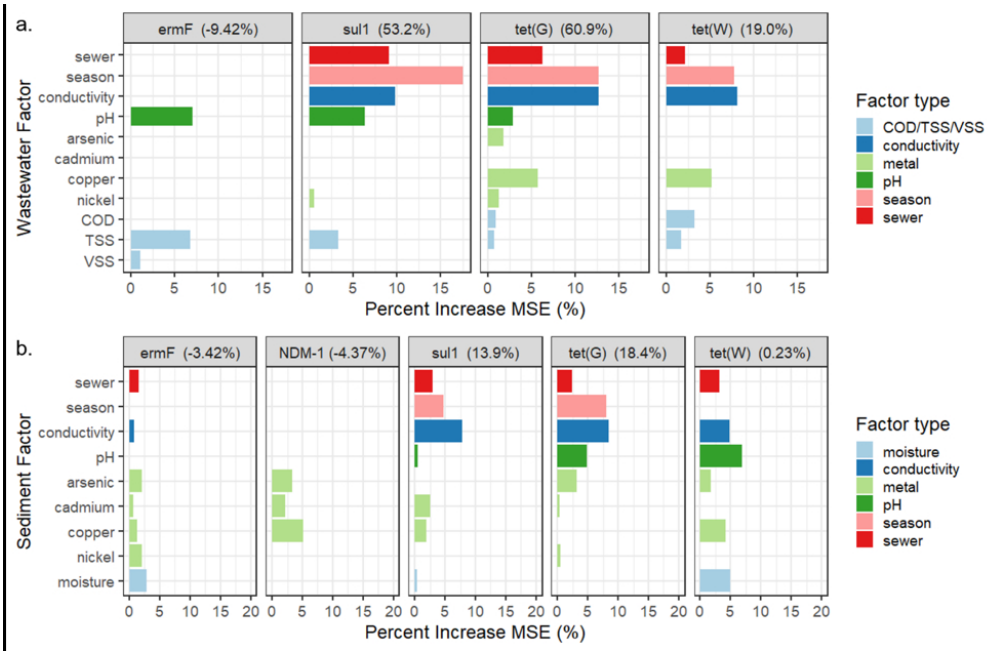


Figure 2. Percent increase in mean square error (MSE) for the different factors included in the random forest regression for (a) wastewater or (b) sediment factors for ARGs. The percent of variance explained for each regression is listed following the ARG name. (No data is shown for NDM-1 in wastewater given that it was only observed in four out of thirty samples.)

165x106mm (150 x 150 DPI)

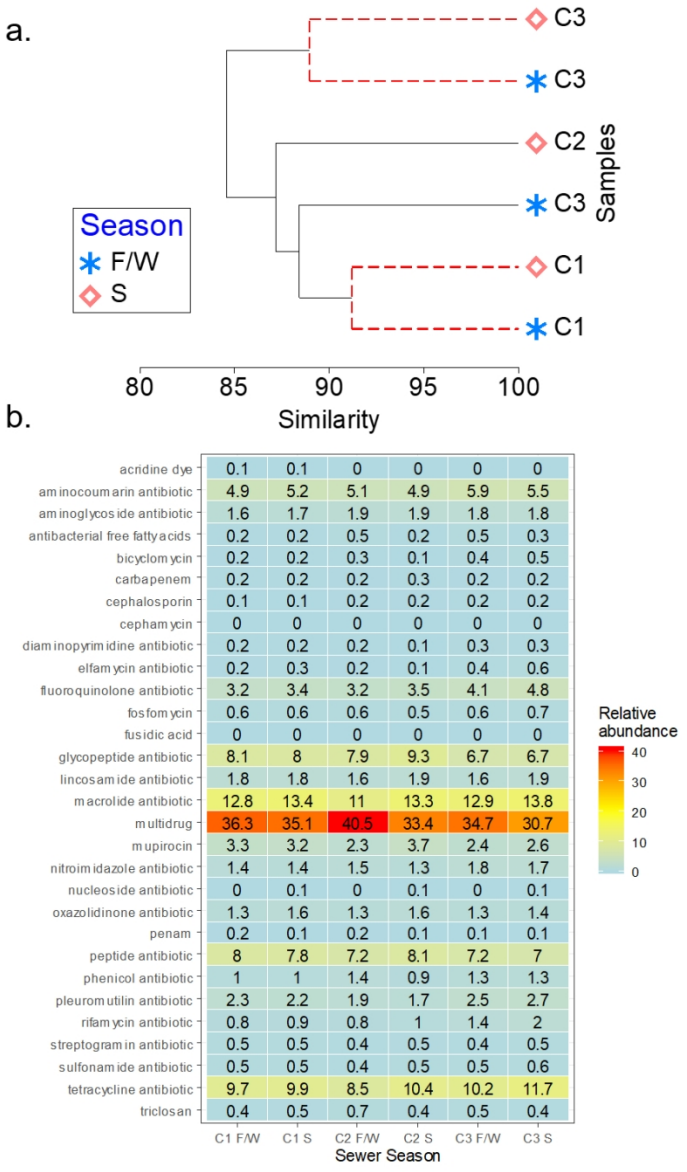


Figure 3. (a) Cluster analysis for the ARG profiles observed in combined sewer sediment systems (C1-C3) collected in two different seasons (fall/winter "F/W" versus summer "S"). Samples connected by black bars are significantly different (SIMPROF p>0.05). (b) Heat map of antibiotic classes relative abundance observed in the sewer sediment metagenomes.

160x262mm (150 x 150 DPI)

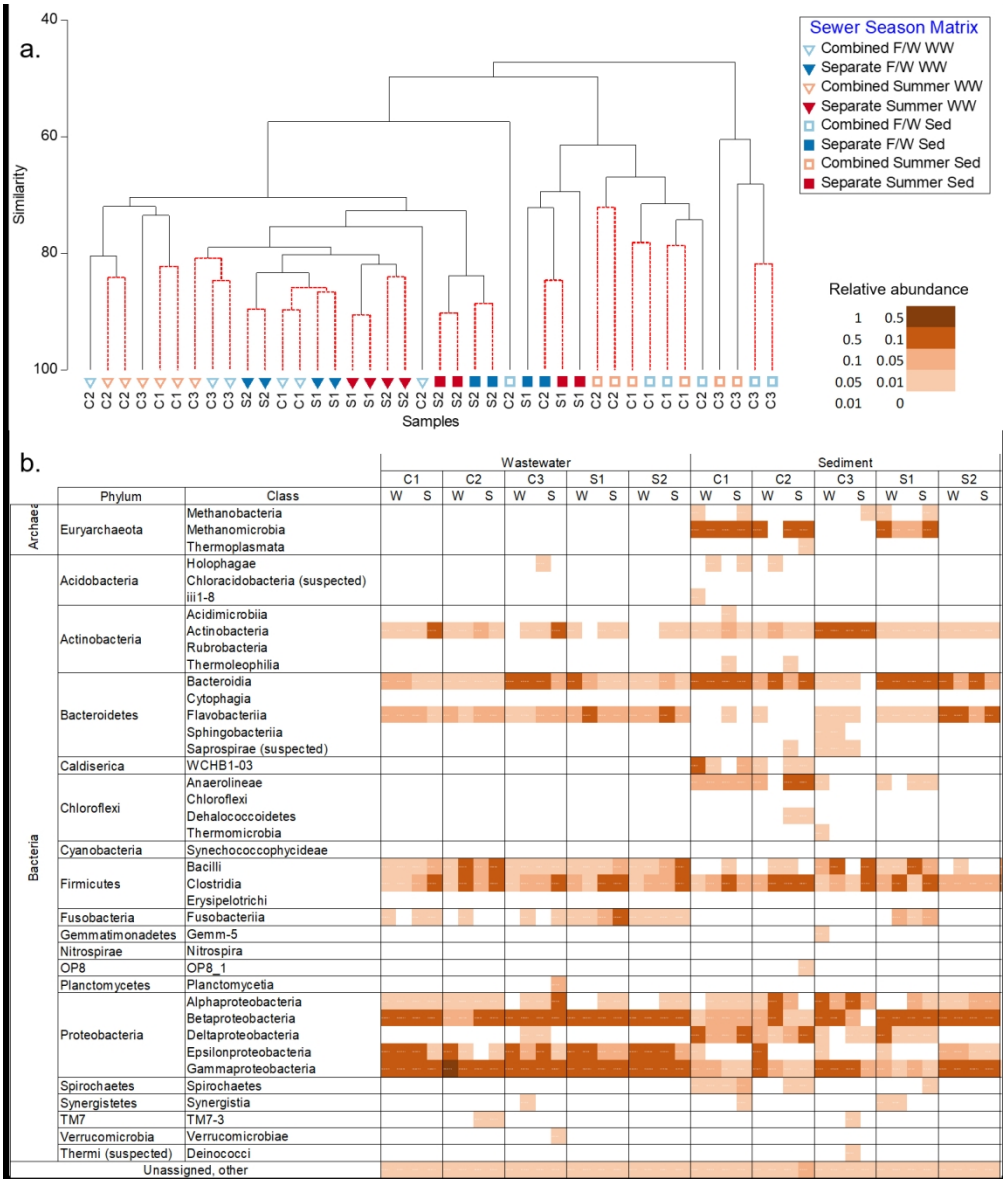


Figure 4. (a) Cluster analysis for the microbial community profiles observed in combined sewer (C1-C3) and separate sanitary sewer (S1 and S2) systems. Paired wastewater ("WW") and sewer sediment ("S") samples were collected triplicate during two different seasons (fall/winter "F/W" versus summer "S") and replicate samples were sequenced. Samples connected by black bars are significantly different (SIMPROF p>0.05). (b) Heat map of bacterial and archaeal phyla and classes with relative abundances greater than 1% in at least one sample from wastewater or sewer sediment. Results for replicate samples from fall/winter ("W") and summer ("S") are shown.

266x312mm (150 x 150 DPI)

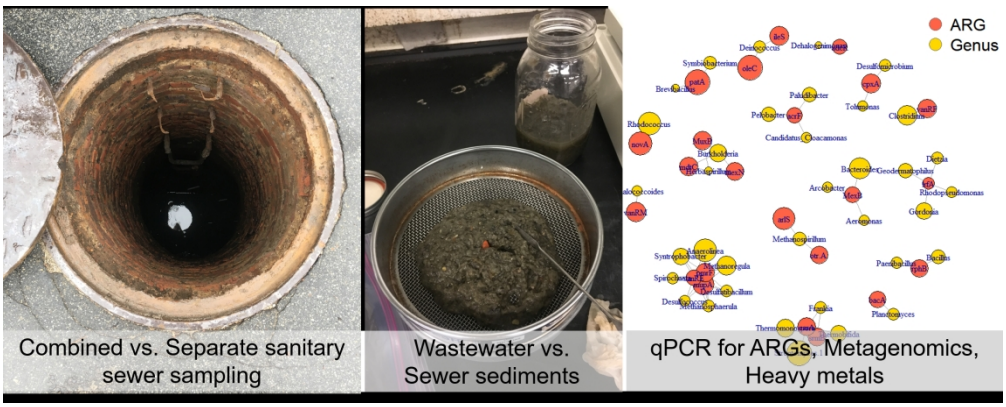
Table. 1 Sewer type, sampling period, and description of where sediment samples* were collected.

System ID	Sewer Type	Fall/Winter Sampling Months	Summer Sampling Months	Sediment Sampling Location/Type
C1	combined	October-November	June-July	Sediment deposits from bottom of sewer pipe collected via manhole
C2	combined	October-November	July-August	Sediment deposits from bottom of sewer pipe collected via manhole
C3	combined	October-November	June-July	Sewer sediment discharged during CSO events and stockpiled in CSO detention tank
S1	separate	September	June-July	Sediment deposits from pump or metering stations
S2	separate	December-January	August-September	Wet well

*N=3 samples per season for each sewer, exact locations varied by availability of access and sediment deposits

Table 2. Wastewater and sewer sediment chemical and quality data for combined ("C") and separate sanitary ("S") sewer systems. Average values are reported ± standard deviation (N=6).

	Wastewater Influent						Sewer Sediment		
System ID	**pH**	**Conductivity (µS/cm)**	**COD (mg/L)**	**TSS (mg/L)**	**VSS (mg/L)**	**ORP (mV)**	**pH**	**Conductivity (µS/cm)**	**Moisture Content (%)**
C1	7.3 ± 0.1	891 ± 93.0	697 ± 160	274 ± 107	220 ± 126	278 ± 90.0	7.0 ± 0.6	119 ± 45.0	25 ± 8.0
C2	7.4 ± 0.2	2020 ± 522	603 ± 330	179 ± 56.0	150 ± 35	253 ± 127	7.1 ± 0.5	575 ± 702	35 ± 17
C3	7.0 ± 0.3	863 ± 257	759 ± 220	420 ± 121	344 ± 108	228 ± 71	6.9 ± 0.8	659 ± 466	30 ± 9.0
S1	7.3 ± 0.3	895 ± 158	516 ± 61	277 ± 70	223 ± 52	296 ± 159	6.6 ± 0.7	466 ± 423	26 ± 16
S2	7.1 ± 0.1	784 ± 120	737 ± 139	443 ± 163	326 ± 115	350 ± 104	6.3 ± 0.5	224 ± 79.0	65 ± 7.0



491x194mm (150 x 150 DPI)