RESEARCH

Localized effect of treated wastewater effluent on the resistome of an urban watershed

Christopher N. Thornton [1],*, Windy D. Tanner[2], James A. VanDerslice[2] and William J. Brazelton[1],*

[1]School of Biological Sciences, University of Utah, 257 South 1400 East, Rm. 201, 84112, Salt Lake City, UT, USA and [2]Department of Family and Preventive Medicine, University of Utah, 257 South 1400 East, Rm. 201, 84112, Salt Lake City, UT, USA

*Correspondence address. Christopher N. Thornton, School of Biological Sciences, University of Utah, Salt Lake City, UT, USA. E-mail: christopher.thornton@utah.edu http://orcid.org/0000-0001-9907-2327 and William J. Brazelton, School of Biological Sciences, University of Utah, Salt Lake City, UT, USA. E-mail: william.brazelton@utah.edu, 257 South 1400 East, Rm. 201, 84112

Abstract

Background: Wastewater treatment is an essential tool for maintaining water quality in urban environments. While the treatment of wastewater can remove most bacterial cells, some will inevitably survive treatment to be released into natural environments. Previous studies have investigated antibiotic resistance within wastewater treatment plants, but few studies have explored how a river's complete set of antibiotic resistance genes (the "resistome") is affected by the release of treated effluent into surface waters. **Results:** Here we used high-throughput, deep metagenomic sequencing to investigate the effect of treated wastewater effluent on the resistome of an urban river and the downstream distribution of effluent-associated antibiotic resistance genes and mobile genetic elements. Treated effluent release was found to be associated with increased abundance and diversity of antibiotic resistance genes and mobile genetic elements. The impact of wastewater discharge on the river's resistome diminished with increasing distance from effluent discharge points. The resistome at river locations that were not immediately downstream from any wastewater discharge points was dominated by a single integron carrying genes associated with resistance to sulfonamides and quaternary ammonium compounds. **Conclusions:** Our study documents variations in the resistome of an urban watershed from headwaters to a major confluence in an urban center. Greater abundances and diversity of antibiotic resistance genes are associated with human fecal contamination in river surface water, but the fecal contamination effect seems to be localized, with little measurable effect in downstream waters. The diverse composition of antibiotic resistance genes throughout the watershed suggests the influence of multiple environmental and biological factors.

Keywords: antibiotic resistance; wastewater; metagenome; watershed; pollution

Background

The growing public health crisis caused by the emergence and spread of antibiotic resistance is now recognized as a global problem with a complex interplay of environmental, biological, and clinical factors. Research on antimicrobial resistance has historically been focused on human pathogens, with hospitals and other clinical settings thought to be the primary source for the dissemination and evolution of antibiotic resistance. However, in part owing to increased reporting of community-acquired, antibiotic-resistant infections [1,2], researchers have started to broaden their focus to include the role of natural environments as possible reservoirs of antimicrobial resistance and as settings for the evolution of new resistance determinants.

Antimicrobial resistance is widespread in nature. Resistance determinants can be found in nearly every environment studied to date, including deep-sea sediment [3], an isolated cave microbiome [4], and 30,000-year-old permafrost [5]. This has led to the recognition that environmental bacterial communities can serve as diverse reservoirs of antimicrobial resistance genes, termed the environmental resistome. There is substantial evidence indicating that, in the past, human pathogens have acquired resistance traits originating in non-pathogenic bacteria that inhabit natural environments (6). It is reasonable to assume that this transfer of environmental resistance factors to human pathogens is ongoing [7]. It is therefore imperative to identify source environments where resistance genes can be selected for and subsequently mobilized into human or animal pathogens.

Wastewater treatment plants (WWTPs) have been demonstrated to contain a large number of antibiotic-resistant bacteria (ARB) and antibiotic resistance genes (ARGs) associated with resistance to all known classes of antibiotic [8–11]. The treatment of wastewater typically results in substantially reduced concentrations of antibiotics and other pharmaceuticals, in addition to eliminating a significant portion of the resistant bacteria present in untreated wastewater [11]. However, despite the efficacy of modern WWTPs in removing ARB and ARGs from wastewater, some resistance determinants will inevitably persist. In some cases, ARGs have been found in treated effluent at similar or even higher rates than measured in the influent [9,12]. Antibiotic compounds and resistant bacteria that survive wastewater treatment are subsequently released into receiving bodies of water, such as lakes and rivers. Continuous discharge of these contaminants can lead to elevated background levels of resistance [13–15], enhancing the likelihood of ARGs being transferred to human commensals or pathogens in the environment.

The localization of ARGs on mobile genetic elements (MGEs), such as transposons and plasmids, enables their movement between bacterial cells of the same or different species. Transfer of resistance factors is likely to increase during exposure to selection factors such as antibiotics and other environmental pollutants [16]. The conditions in WWTPs, including the mixture of organisms from diverse environmental origins and the availability of surfaces and biofilms, can create strong selection pressures for resistance [8,17]. Even those treatments that are effective in removing bacteria from the water can promote the exchange, selection, and dispersal of genes involved in antibiotic resistance [18,19].

Recently, metagenome sequencing of wastewater effluent has been implemented as a useful tool for monitoring the spread of ARGs into natural environments (e.g., [20,21]). Few studies, however, have investigated the environmental resistome throughout a single urban watershed from its headwaters to a major drainage. In this study, we sequenced metagenomes from 72 river samples collected from an urbanized watershed with the goal of assessing the impact of point sources of human waste on the resistance profile of receiving river surface waters, focusing especially on ARGs associated with MGEs.

Data Description

Surface water samples were collected for DNA sequencing and measurements of stream chemistry and physical parameters from 24 sites along 3 of the rivers comprising the Blue River watershed: the Blue River and tributaries Indian Creek and Tomahawk Creek. The Blue River watershed was selected on the basis of its high population density, long history of waste overflow from a combined sewage system, and the presence of multiple, high-capacity WWTPs. Sampling site locations were selected on the basis of several factors, including proximity to headwaters and confluences, as well as potential sources of pollution such as wastewater treatment plants, hospitals, and drug manufacturing plants.

The surface water samples were analyzed by shotgun metagenomic sequencing, which generated a total of 8.6 billion read pairs. Additional samples were also collected for *Escherichia coli* enumeration and antibiotic susceptibility testing.

Analyses

Detection of ARGs and MGEs in river metagenomes

ARGs were detected in river water metagenome assemblies using AMRFinder v1.04 [22]. A total of 88 unique ARGs were detected in the watershed, in principle conferring resistance to 12 different classes of antibiotic and an additional 4 multidrug resistance phenotypes: MLSb, ML, MSb, and LSa. Sulfonamide resistance (26%), followed by aminoglycoside, β-lactam, and macrolide resistance (24%, 16%, and 14%, respectively), made up the largest percentage of the watershed resistome (Supplementary Table S1). The most abundant and commonly occurring ARG was *sul1*, which was detected in 22 of the 24 sampling sites (Table 1). With the exceptions of *sul1* and *blaIND*, the majority of the highest abundance genes (those composing >1% of the total watershed resistome) were found exclusively downstream from potential point sources of human pollution (Supplementary Fig. S1). A wide range of MGEs were also detected in river water metagenome assemblies. A number of these elements were located in close proximity to ≥1 or more resistance genes, and we verified many examples of ARGs encoded within complete integrons or transposable elements.

ARGs are more abundant and more diverse downstream from WWTPs

The total abundance of ARGs was found to be significantly higher in river waters sampled immediately downstream from WWTP discharge (Fig. 1; 1-way analysis of variance [ANOVA], P-adjusted = 0.00051 and P-adjusted = 0.0006728, respectively). On average, a 140-fold increase in ARG abundance was observed in samples collected from within 5 km downstream of a WWTP. A 30-fold increase in ARG diversity was also observed in samples collected from downstream surface waters. Of the 88 different ARGs detected in the watershed, 77 of them were detected at downstream sites (averaging 20 per site), while only 15 were detected at upstream sites (averaging 3 per site). ARGs associated with WWTP discharge included those associated with resistance to lincosamide, macrolide, chloramphenicol, fluoroquinolone, polypeptide, trimethoprim, tetracycline, and rifamycin antibiotics as well as with the multidrug-resistant phenotypes MLSb, ML, MSb, and LSa.

ARGs associated with resistance to β-lactam, aminoglycoside, and sulfonamide antibiotics, in contrast, were common to both upstream and downstream sites. A single sulfonamide resistance gene (*sul1*) was present in 22 of 24 sampling sites and comprised 87% of total abundance of all sulfonamide resistance genes. β-lactamases were also detected with high frequency throughout the watershed, including in 78% of the upstream samples and 75% of the downstream samples. There was no significant difference between upstream and downstream samples in the abundance of sulfonamide resistance genes (false discov-

Table 1. Top 25 resistance genes detected in the watershed

| Resistance gene	Class	Mean abundance per *rpoB* gene	Percent watershed resistome	Sample prevalence (Site prevalence))
sul1	Sulfonamide	4.48E−03	23.65	68 (22)
ant(3″)-II	Aminoglycoside	2.33E−03	12.33	12 (4)
aadA	Aminoglycoside	1.00E−03	5.28	30 (10)
msr	MSb	9.12E−04	4.82	24 (8)
mph(E)	Macrolide	8.02E−04	4.24	24 (8)
blaIND	β-Lactam	6.07E−04	3.20	24 (8)
cfxA	β-lactam	5.32E−04	2.81	17 (6)
sul2	Sulfonamide	4.51E−04	2.38	38 (13)
mph(G)	Macrolide	4.27E−04	2.25	18 (6)
aph(6)-I	Aminoglycoside	4.13E−04	2.18	20 (7)
mef(A)	Macrolide	4.00E−04	2.11	21 (7)
blaOXA-2	β-Lactam	3.55E−04	1.88	24 (8)
blaA	β-Lactam	3.51E−04	1.86	26 (9)
tet(C)	Tetracycline	3.42E−04	1.81	24 (8)
tet(Q)	Tetracycline	3.39E−04	1.79	21 (7)
tet(M-W-O-S)	Tetracycline	3.09E−04	1.63	18 (6)
erm(F)	MLS	3.03E−04	1.60	18 (6)
aph(3″)-Ib	Aminoglycoside	2.93E−04	1.55	18 (6)
aadA1	Aminoglycoside	2.80E−04	1.48	15 (5)
mef(C)	Macrolide	2.71E−04	1.43	24 (8)
blaOXA-10	β-Lactam	2.58E−04	1.36	24 (8)
tet(A-B-C-D)	Tetracycline	2.10E−04	1.11	15 (5)
mef(B)	Macrolide	1.74E−04	0.92	18 (6)
blaOXA	β-Lactam	1.65E−04	0.87	18 (6)
blaVEB	β-Lactam	1.63E−04	0.86	8 (3)
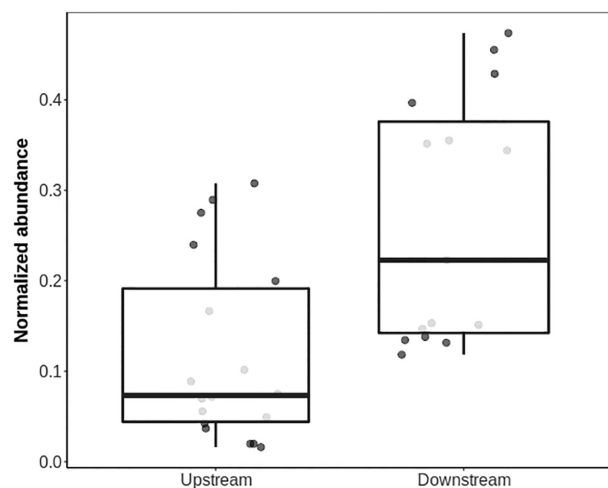
Figure 1 Total abundance of ARGs and MGEs in relation to proximity to WWTP. Normalized abundances of total ARGs (top) and MGEs (bottom) were square root–transformed for analysis. Samples are grouped on the basis of where they were collected relative to the nearest WWTP (upstream = surface waters with no impact from a WWTP; downstream = within 5 km downstream of the nearest WWTP), and all data points are shown, with the whiskers representing the maxima and minima. The solid black line is the median value, and the area between the lower and upper hinge represents the inter-quartile range, or difference between the first and third quartiles.

ery rate [FDR] = 0.647) or β-lactam resistance genes (FDR = 0.52). Aminoglycoside resistance genes, exclusively encoding agent-modifying enzymes, were detected in 64% of downstream samples and 22% of upstream samples, with a 12-fold enrichment in abundance downstream from WWTP discharge sites (FDR < 1.014E−06).

Proximity to WWTPs was an important factor influencing the abundance of ARGs in river surface waters. Total ARG abundance decreased substantially at sites located >5 km downstream from WWTPs compared with sites within 5 km (Supplementary Fig. S2; ANOVA, P-adjusted = 0.00058). The log fold-change in ARG diversity was inversely correlated with the downstream distance from WWTP discharge points (Supplementary Fig. S3; linear regression, adjusted R^2 = 0.7425, P = 0.0008314). Notably, MGE abundance, while not found to be significantly higher immediately downstream from WWTPs than upstream, also decreased with distance from the WWTPs (Supplementary Fig. S4; ANOVA, P-adjusted = 0.0007689).

ARG abundance is correlated with a marker of human fecal pollution

The relationship between total ARG abundance and the abundance of crAssphage was investigated to test whether increased abundances of ARGs could be explained by human fecal pollution. crAssphage is a highly abundant bacteriophage in human fecal metagenomes [23] and is rare in feces from non-human animals [24]. The abundances of ARGs and crAssphage were highly correlated with each other in river samples downstream from WWTPs (Fig. 2; linear regression, adjusted R^2 = 0.54, P = 5.196E−09). The highest levels of crAssphage were observed immediately downstream from WWTPs, with lower levels detected at more distant sites, following the general trend observed with total ARG abundance.

No correlation was found between crAssphage and ARG abundances in samples collected from sites without an upstream WWTP (linear regression, adjusted R^2 = −0.03, P = 0.73). Nearly all of the upstream sites were located in areas with relatively high population density, and crAssphage sequences were
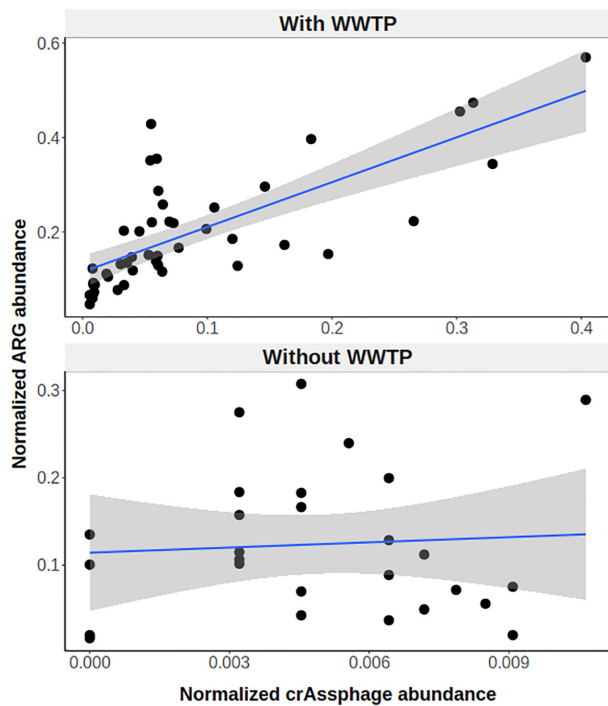
Figure 2 Relationship between ARG and crAssphage abundance. Normalized abundances of total ARG and crAssphage DNA were square root–transformed for analysis. Sites are grouped according to whether ≥1 WWTP is located upstream of the sample site (top) or not (bottom). Smoothing curves based on linear regression (blue line) are shown along with 95% confidence intervals (shaded regions). Note the smaller range of crAssphage abundances in samples without upstream WWTPs.

detected in 85% of the upstream samples. Therefore, the lack of correlation between crAssphage and ARG abundances in upstream sites is not due to a complete absence of human fecal pollution but suggests the influence of additional environmental factors that were not measured in this study.

ARGs are associated with MGEs

The potential of ARGs to be transferred between cells was investigated by identifying ARGs located on MGEs such as plasmids, transposons, and integrative and conjugative elements (ICEs). The number of unique ARGs encoded on MGEs (mARGs) was significantly higher at sites immediately downstream from WWTPs (Fig. 3; randomization, $P < 0.001$). The number of mARGs rapidly diminished with increasing distance downstream (randomization, $P < 0.001$). On average, the number of mARGs immediately downstream from WWTPs (mean, 19 per sample) was slightly higher than the number of ARGs assumed to be chromosome-encoded owing to lack of evidence to the contrary (cARG; mean, 15 per sample); however, the difference was not found to be significant (ANOVA, $P = 0.09$). Individual mARGs were more often than cARGs to be found in multiple sites downstream from WWTPs, consistent with the ability of mARGs to be shared among multiple bacterial species.

Of the 37 mARGs detected in the watershed, a majority (84%) could be found within 5 km downstream from a wastewater discharge site. Only 2 mARGs (*blaTEM* and *sul1*) were observed in sampling sites upstream from all WWTPs. The sulfonamide resistance gene *sul1* was found in 2 distinct mobile contexts throughout the watershed (Supplementary Table S2), but in up-
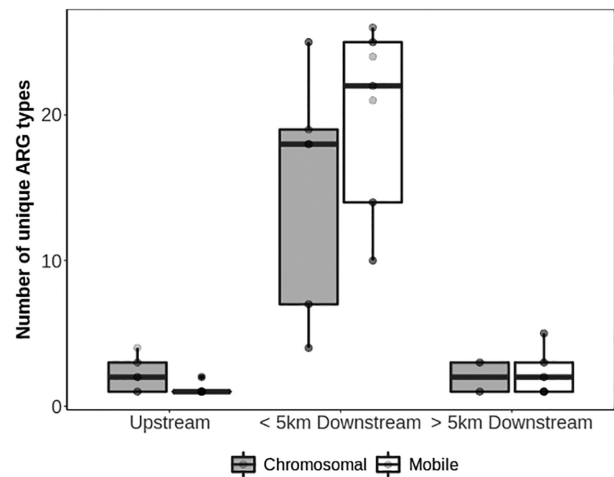

Figure 3 Diversity of ARGs and ARG-encoding plasmids/ICEs in relation to WWTP. Samples are grouped on the basis of where they were collected relative to the nearest WWTP (upstream = samples are of surface waters with no impact from a WWTP; <5 km downstream = samples were collected within 5 km downstream of the nearest WWTP; >5 km downstream = samples were collected from sites >5 km downstream of the nearest WWTP). Box plots are drawn as in Figure 1, with the exception that several data points overlap each other. Total numbers of samples for each group are 18 (upstream), 14 (<5 km downstream), and 18 (>5 km downstream).

stream sites, it appeared primarily in an integron that also contained the quaternary ammonium compound (QAC) resistance genes *qacE* and *qacG* (Fig. 4). This integron had closest sequence similarity to integron In78 (100% identity over 60% of the sequence), previously associated with *Pseudomonas aeruginosa* [25].

Downstream sites hosted mobilized genes conferring resistance to many classes of antibiotic, including tetracycline, macrolide, aminoglycoside, fluoroquinolone, and lincosamide antibiotics, as well as several likely plasmid/ICE-encoded genes responsible for the macrolide-lincosamide-streptogramin resistance phenotype (Supplementary Table S2). Unlike *sul1*, the plasmid-encoded *sul2* variant was found exclusively downstream from WWTPs and in multiple different mobile contexts, often with a rolling-circletype transposase. On 1 such contig with 100% identity to plasmid R485, *sul2* was found together with a likely insertion sequence common region (ISCR) transposase, a toxin-antitoxin system, multiple conjugation proteins, and an integrase matching those from the Tn916 family of conjugative transposons.

Resistance phenotypes detected by antibiotic susceptibility assays

E. coli colonies were cultivated from river water samples that were collected simultaneously with the samples for metagenomics sequencing. Curiously, no assembled metagenomic sequences were classified as *E. coli*, indicating that our sequencing and assembly approach was not sensitive enough to detect *E. coli* populations of this density.

E. coli colonies were tested for their susceptibility to a variety of antibiotics, including those associated with resistant Enterobacteriaceae pathogens identified in the 2013 CDC Antibiotic Resistance Threat Report (26), as well as additional antibiotics with clinical significance. Antibiotic-resistant *E. coli* were isolated at 7 of the 24 sampling sites (Supplementary Table S3). Ampicillin resistance was the most common phenotype (6% of
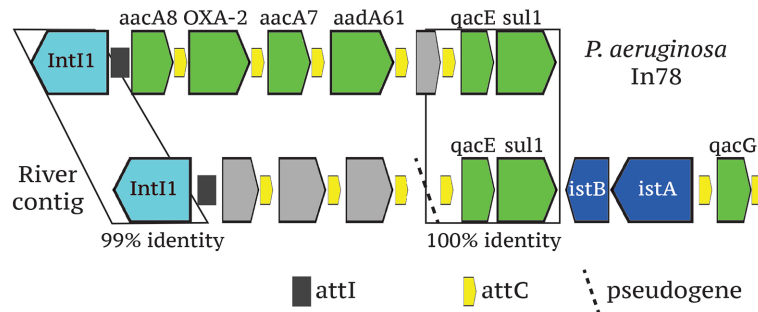
Figure 4 Novel *sul1*-bearing integron and closest match In78 from *P. aeruginosa*. In addition to the sul1-qacE cassette, the contig associated with the novel integron contained a full-length integron integrase and associated integrase and cassette attachment sites, 3 uncharacterized protein-coding genes, IS21-family transposition genes *istA* and *istB*, and a second QAC resistance gene (*qacG*). The boxed areas of the schema show the regions of alignment between integrons.

isolates), followed by amoxicillin-clavulanate, cefazolin, and cefoxitin resistance (3% of isolates each). No isolates were resistant to glycopeptide, trimethoprim, or sulfonamide antibiotics.

Colistin resistance was determined by measuring minimum inhibitory concentrations (MIC) with broth microdilution plates. The most frequently observed colistin MIC was 0.5 μg/mL. Three of the 70 isolates exhibited MICs of 2 μg/mL; colistin resistance is defined as a MIC >2 μg/mL. A single isolate exhibited a colistin MIC of 8 μg/mL. The metagenome of the site where this isolate was isolated (CPA) contains the colistin resistance gene *mcr-5*. No other genes conferring resistance to colistin were detected in the watershed.

Discussion

ARGs have been previously detected in the discharge of WWTPs [9,12, 27, 28], as well as in receiving aquatic environments such as rivers and lakes [13–15, 20, 29–31]. Here, we conducted an extensive metagenomic study to investigate the distance decay effect of 4 wastewater treatment plants on the resistance profile of an urban watershed. River samples collected immediately downstream from WWTPs had significantly higher environmental abundance and diversity of ARGs, and this effect diminished in samples collected >5 km downstream. These results provide additional support for a localized effect of WWTPs on the resistome of receiving aquatic environments.

Our results also show that ARGs associated with WWTPs are likely to be encoded on MGEs such as plasmids and transposons. Nearly half of the unique ARGs detected downstream of the WWTPs were encoded on an MGE, and all of these ARGs have been previously associated with resistance to clinically relevant drug families. For instance, the erythromycin ribosome methylase (erm) genes have been identified as part of a core WWTP resistome [32] and have consistently been found enriched in biofilms and surface waters downstream from WWTP discharge [13,14]. All of the erm genes detected in our study, including *erm(F)* and *erm(B)*, were located in assembled sequences identified as plasmids or ICEs.

A number of other resistance determinants were also associated with MGEs. These included the linked aminoglycoside resistance genes *strA* and *strB*, which have previously been detected in WWTP effluent [20,30] and other aquatic environments subjected to human pollutants [33]. The *strA-strB* genes are typically encoded on broad-host-range non-conjugative plasmids, as well as conjugative plasmids associated with Tn3-type transposons [34]. Within the Blue River watershed, *strA-strB* was en-

coded on a complete Tn3-like transposon with 100% identity to transposon Tn5393d from *Alcaligenes faecalis*. Contigs containing the complete transposon were found at 2 sites, each downstream from a potential point source of ARG pollution (LDP and UMC), while smaller contigs containing a subset of this region were found downstream from all 4 WWTPs. The *strA-strB* genes were not found in any of the upstream samples.

The spectinomycin resistance gene *aadA*, another member of the core WWTP resistome described in Munck et al. [32], was also found in samples downstream from all WWTPs and in no upstream samples. In our study, *aadA* genes frequently co-occurred with other ARGs and with multiple MGEs. For instance, *aadA*, *sul1*, *qacEdelta*, and an IS66 transposase were present on an assembled contig with >99% similarity to the conjugative tetracycline resistance plasmid pFBAOT6.

Plasmid-mediated quinolone resistance, conferred through the pentapeptide repeat protein qnr(S), was also detected immediately downstream from all WWTPs. The *qnr(S)* gene has consistently been found enriched in WWTP effluent and receiving waters [13–15].

In the 2 largest rivers investigated here, Blue River and Indian Creek, WWTPs contribute on average 15% of base flow [35] and could contribute >95% under certain conditions [36]. Despite the significant contribution of WWTPs to both streamflow and ARG abundance and diversity, the WWTP-associated ARGs did not persist in surface waters >5 km downstream from effluent discharge points. This diminishing effect of elevated ARG abundances in receiving waters with increasing distance from WWTPs has been observed elsewhere [14,20,29]. These results, combined with the strong correlation of total ARG abundance with the abundance of a human gut phage, is consistent with the interpretation that ARGs were released into surface waters with human fecal pollution [37] and then diluted in downstream waters [20,29].

The abundances of ARGs in sites upstream from all WWTPs, in contrast, were independent of human fecal pollution levels. The relatively high abundances of ARGs in these sites suggests the influence of agricultural pollution near the headwaters, but this hypothesis was not tested by the present study. The most abundant ARG in upstream samples was the *sul1* sulfonamide resistance gene, which was primarily encoded on a class 1 integron along with 2 QAC resistance genes. This same mobile element was ubiquitous throughout the entire watershed, spanning multiple streamflows, land use types, and pollution levels, suggesting that it may be maintained in natural microbial communities owing to a variety of selection pressures.

Potential Implications

Our metagenomic study of an urban watershed has shown that (i) ARGs are widespread in natural aquatic environments, (ii) WWTPs introduce a characteristic profile of mobilized ARGs into receiving rivers and streams, (iii) human fecal pollution and associated ARGs can be effectively diluted by natural microbial populations within several kilometers from discharge points, and (iv) the dispersal of specific ARG-encoding MGEs with conserved genomic structures can be traced throughout the watershed.

Characterization of the genomic context of ARGs, such as the novel, ubiquitous integron containing 3 different antimicrobial resistance genes, was enabled by assembly of the metagenomes, a computationally challenging task that is not routinely performed in such studies. The additional genomic information obtained from metagenomic assembly comes at the cost of reduced sensitivity of detection of sequences that have low abundances and those that are difficult to assemble. Nevertheless, the ability to detect ARGs in new genomic contexts and trace their dispersal among genetic elements, organisms, and environments is a powerful tool for the surveillance of antibiotic resistance in natural environments and potentially for the early detection of emerging resistance genotypes of clinical relevance. Future work should continue to develop and validate metagenomic methods for the quantitative measurement of antimicrobial resistance in natural environments.

Methods

Description of study sites

The Blue River watershed encompasses 450 km^2 and includes the southern half of the Kansas City metropolitan area below the Missouri River. The majority of water in the Kansas City metropolitan area and in many of the adjacent towns to the south and west drains into the Blue River. The Blue River basin is primarily urban and suburban, with a population density of 794.8 persons/km^2. Six WWTPs are located in the basin, 3 of which near-continuously discharge treated effluent directly into the Blue River or one of its tributaries while a fourth has occasional wet-weather discharges into the lower Blue River. WWTPs provide the dominant source of streamflow, nutrients, and pharmaceutical compounds to the middle and lower reaches of the Blue River during base flow [35]. Twelve sites on the main stem of the Blue River, 10 sites on the tributary Indian Creek, and 2 sites along Tomahawk Creek were investigated (Fig. 5). Accessible locations bracketing WWTP discharges (n = 4) were specifically targeted to capture the effect of WWTPs on the stream. Two sites along Tomahawk Creek, which does not contain any WWTPs, were sampled as additional upstream background samples for 1 WWTP site, which was located downstream from both the effluent discharge point and the confluence between Indian Creek and Tomahawk Creek. Additional samples from sites at various points throughout the watershed were collected as reference samples. Characteristics of the 4 WWTPs and a summary of the sampling sites are presented in Supplementary Tables S4 and S5, respectively.

Sample collection

A portable peristaltic pump was used to collect surface water samples by pumping water either directly from the stream/river into their respective containers or through sequential in-line filters. At each sampling site, the pump tubing inlet was submerged several centimeters beneath the water surface and held in place through the use of tube weights or anchors. An attempt was made to sample from a location in the stream/river where flow seemed to be greatest or most representative of total flow. Whenever possible pump tubing was positioned upstream to reduce unnecessary exposure to contaminants from sampling equipment and personnel, and caution was taken during the placement of the sampling equipment to avoid streambed disturbance upstream of the collection site. To avoid introducing contamination between sites the pump tubing was rinsed between sampling locations with ultrapure water and 80% ethanol and purged prior to sample collection with site water for ≥1 minute after placement of the tubing inlet. Physical and chemical parameters of the stream/river water, including dissolved oxygen, specific conductance, total dissolved solids, pH, oxidation-reduction potential, and temperature (Supplementary Table S6), were measured simultaneously with sample collection using a YSI Pro Plus portable multi-parameter water quality meter (YSI Incorporated, Yellow Springs, OH, USA), and turbidity was measured with a Mi415 turbidity meter (Martini Instruments, Adelaide, Ausralia).

For each sample collected for metagenomic sequencing, site water was filtered sequentially through a 5.0-μm mixed cellulose filter (5 μm pore size; 47 mm diameter; MF-Millipore; Merck, Darmstadt, Germany), followed by a 0.2-μm Sterivex cartridge filter (0.22 μm pore size; Sterivex; Millipore), until either clogging of the filters substantially slowed water flow or a total volume of 5 L had been filtered. This maximum volume was previously determined to be a reasonable compromise between expediency and thoroughness when sampling river water for metagenomic sequencing, despite considerable variation in sampling site water quality, and was found to be sufficient for this study as well. Depending on the level of suspended particles in the site waters, the volume of water filtered varied from 240 to 5,000 mL. Volume filtered was an important factor affecting the amount of DNA that could be extracted from the filters and, consequently, the quality of a metagenomic dataset. A correction for difference in volume filtered was factored into calculations of gene abundance to account for this. Immediately after sampling, remaining fluid was purged from the filter assembly/cartridge and the 5.0- and 0.2-μm filters placed into separate, sterile whirlpool packs and frozen on dry ice before transportation to the laboratory, where they were stored at −80°C until processing. Field replicates, in triplicate, were collected at each sampling site and processed individually.

The 5.0-μm prefilter was implemented as a screen for filtering plant debris and large eukaryotic cells to increase the coverage depth of genes associated with antibiotic resistance, which are exclusive to bacteria. In addition, the 5.0-μm prefilter removed large particles, biofilms, and aggregates, which can unpredictably swamp samples of the free-living, suspended river communities collected by the 0.2-μm filters. This smaller size fraction was the focus of the study because it was considered to be more representative of suspended cells that are dispersing through the watershed rather than being retained in sediments and debris. No sequencing was performed with the 5-μm filters; therefore, genes encoded by large cells or cells associated with aggregates and biofilms are likely to be under-represented in our study.

Samples for *E. coli* enumeration and antibiotic susceptibility testing were collected into sterile, 100-mL clear plastic bottles and then serially diluted on-site in a phosphate buffer solution. Dilutions were filtered through a 0.45-μm membrane filter and the filters aseptically removed from their assemblies and
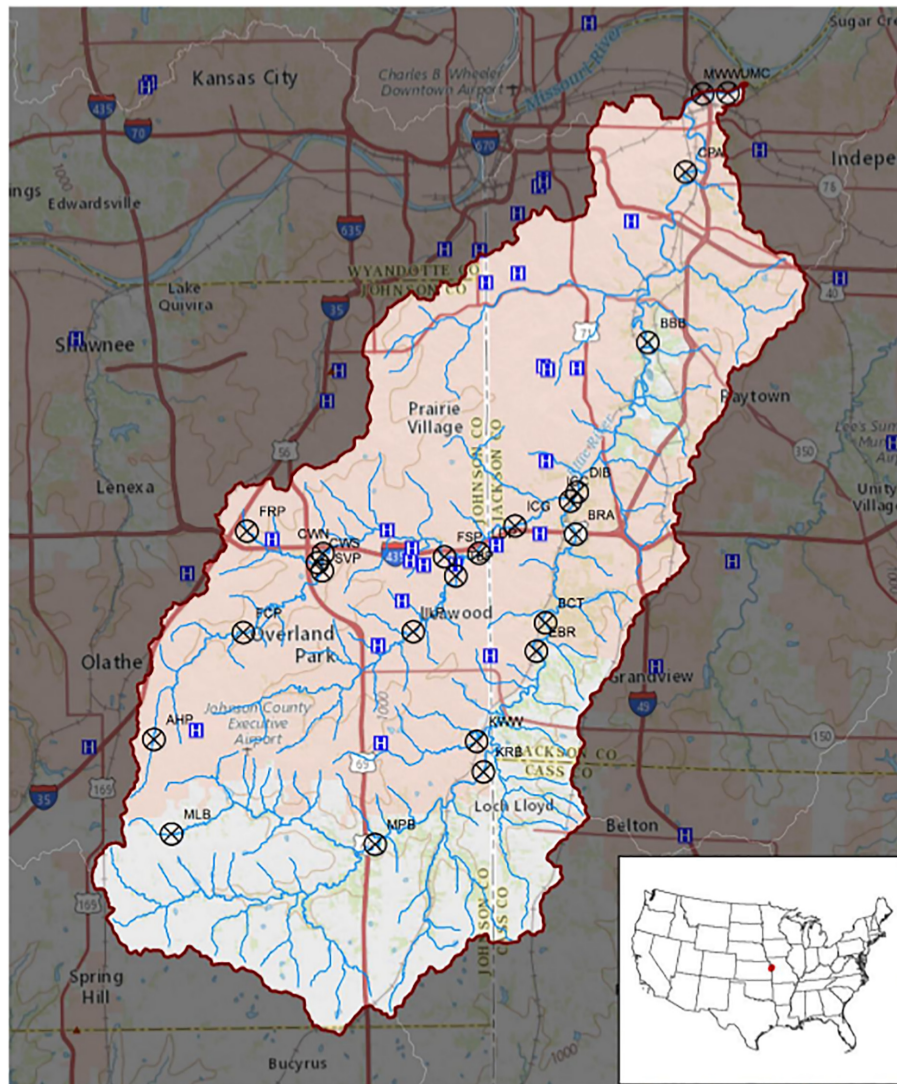
Figure 5 Location of sampling sites within the Blue River watershed. Sampling sites are marked with an X. Locations of hospitals are marked with a blue H.

placed onto U.S. Environmental Protection Agency–approved chromogenic/fluorogenic media for the detection and enumeration of *E. coli* and total coliforms in water [38]. Inoculated plates were inverted and set to incubate at 35°C for 20–24 hours. Following incubation, colonies exhibiting phenotypes typical of *E. coli* (blue-green colonies) were counted under ambient light using manufacturer's guidelines. Plates were then sealed with parafilm and stored in coolers with ice for transportation to the laboratory, where they were refrigerated for later culturing. Field blanks were collected at roughly one-quarter of the sites (chosen at random) and consisted of distilled water. Field blanks were processed in the same manner as the study samples. No growth was observed on all but 1 of the field blanks, which is suspected of having been contaminated with unsterilized forceps owing to the unique pattern of growth.

Antibiotic susceptibility testing

Once in-lab, 3 colonies per sample that had been identified as *E. coli* on the MI medium were selected for subsequent phenotypic resistance testing. The chosen colonies were plated onto

blood agar and incubated at 37°C. The purity of the cultures was reassessed on chromogenic agar (red-pink colonies) whenever blood agar colonies exhibited atypical morphology. Three of the 74 plates made for susceptibility testing were found through this additional screening measure to be contaminated and removed from testing. Following incubation for ~24 hours, sensitivity of the remaining isolates to 16 antibiotics, including ampicillin, amoxicillin, cefazolin, cefepime, ceftriaxone, ceftazidime, ceftazidime with clavulanic acid, cefoxitin, cefotaxime, cefotaxime with clavulanic acid, meropenem, ciprofloxacin, aztreonam, sulfamethoxazole-trimethoprim, gentamicin, and tetracycline, was determined through the Kirby-Bauer method with BBL Sensi-Disc Antimicrobial Susceptibility Test Discs (BD, Franklin Lakes, NJ, USA). Testing was performed according to CLSI protocols (39) with the following modification: a BBL Prompt Inoculation System (BD) was used to pick between 3 and 5 colonies from the blood agar for standardized preparation of 0.5 McFarland turbidity level bacterial inocula. This procedure has been shown to have high correlation with conventional disk diffusion techniques, which involve the suspension of colonies in saline followed by dilution to 0.5 McFarland. Within 15 minutes of prepa-

ration, the suspension was used to inoculate Mueller-Hinton (MH) agar plates using standard swabbing technique. Five or 6 antibiotic disks were applied to each of 3 plates per isolate to prevent overlapping zones. Inhibition zones were measured to the nearest millimeter for each of the 16 antibiotic disks by visual inspection of the inoculated plates following overnight incubation at 37°C, and the zones interpreted through a Zone Diameter Interpretive Chart to determine susceptibility category.

Resistance to colistin was determined through broth microdilution (BMD). Because the large size of the colistin molecule prevents sufficient diffusion through agar medium, BMD is the only procedure recommended by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) for antimicrobial susceptibility testing of colistin [40]. Minimum inhibitory concentration (MIC) for colistin was measured using MIC-strip colistin microdilution plates according to manufacturer's instruction (8 × 12 well plates; Merlin Diagnostics). Briefly, 50 μL of the standardized inoculum was added to 11 mL MH broth. The suspension was gently vortexed and a 100-μL aliquot per well transferred to microdilution strips containing MH broth supplemented with a 2-fold serial dilution of antibiotic. Final concentrations of antibiotic ranged from 0.0625 to 64 μm/mL. Inoculated plates were sealed and incubated at 37°C. Following incubation for 20–24 hours, MICs were determined through visual inspection of the plates under ambient and supplemented light. MIC was defined as the concentration of drug at which no growth was visible. Isolates were retested whenever a skipped well was observed during inspection of the BMD strips or for additional confirmation for *E. coli* isolates with MIC >2 μg/mL.

Metagenomic sequencing

Extraction of DNA from cells retained on the 0.2-μm Sterivex filters was performed according to an established laboratory protocol [41], briefly described here. For the complete protocol, see "Availability of Supporting Data and Materials." Methods for this extraction method are also openly available in protocols.io [42]. To help mitigate any variation that might be introduced during laboratory handling, DNA extraction was performed on the Sterivex filters in random order, with extraction order resolved beforehand. Filter cartridges were injected with extraction buffer, incubated at 65°C for 30 minutes, and then the fluid was extracted from the filter cartridge and distributed into 0.1-mm glass bead tubes, which were shaken for 40 seconds in a MiniBeadBeater-16607 (3,450 oscillations/min; BioSpec, Bartlesville, OK, USA). Bead tubes were centrifuged, and the supernatant was purified by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) and chloroform/isoamyl alcohol (24:1). DNA was precipitated overnight at −20°C in sodium-acetate/ethanol, and the pellets were resuspended in low-EDTA TE. DNA was not extracted from the 5-μm prefilters, which remain in frozen storage for potential future analyses.

Whenever possible, 80 ng of purified DNA was used in the construction of metagenomic libraries. Total DNA extracted from the filters was quantified on a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). The quality of the sample DNA was also assessed before library construction with a Nanodrop spectrophotometer (Thermo Fisher), and further purification conducted as needed based on Nanodrop 260/230 and 260/280 ratios using magnetic bead–based cleanup following a protocol adapted from Rohland & Reich [43]. The 260/230 curve was particularly predictive of library preparation success for this study, with a significant correlation found between Nan-

odrop 260/230 ratios and the final concentration of DNA in the prepared libraries (linear regression, adjusted R^2 = 0.283, P = 5.267E−11). Purified DNA was fragmented by sonication with a Q800R sonicator (Qsonica, Newtown, CT, USA) at 4°C, 25% amplitude, with a 10-second pulse for 60 seconds. These settings were selected to achieve a high molecular weight band on an agarose gel around the 500–700 bp range. Magnetic beads were then used to select molecules from the fragmented DNA with a target size range of 500–700 bp. The size and quality of the fragments were verified with gel electrophoresis on a random subset of samples. Libraries were prepared from the size-selected DNA for metagenomic sequencing using the NEBnext Ultra DNA library prep kit for Illumina according to manufacturer instructions. A final bead cleanup was performed on the prepared DNA to remove excess kit reagents that could interfere with sequencing, and stored at −20°C until transportation to the sequencing center. Each of the triplicate samples per site was processed separately, resulting in a total of 72 metagenomic libraries for sequencing.

Quality control and sequencing of the metagenomic libraries was conducted at the University of Utah High-Throughput Genomics Core Facility. Libraries were evaluated for quality on a Bioanalyzer DNA 1000 chip (Agilent Technologies, Santa Clara, CA, USA), and then paired-end sequencing (2 × 125 bp) was performed on an Illumina HiSeq2500 platform with HiSeq v4 chemistry. Libraries were multiplexed and pooled 4 per lane for a total of 19 lanes of Illumina sequencing, yielding >1 trillion bp of data (Supplementary Table S7). Demultiplexing and conversion of the raw sequencing base-call data were performed through the CASAVA v1.8 pipeline.

Sequence assembly

Quality control of the sequencing reads was performed with BB-Duk v37.10 and consisted of contaminant removal and quality-based trimming. First, contaminants in the form of library adapter sequence were trimmed/filtered from the reads. An adapter was considered to be present whenever an adapter reference sequence shared a 23-mer with a read, or an 8-mer if located at the extreme 3′ end. A Hamming distance of 2 was allowed between matching k-mers. Adapter sequences were also detected from completely overlapping read pairs. Additional contaminants in the form of PhiX DNA were identified and removed if 90% of a read's length was covered by the PhiX174 reference genome. PhiX DNA was used as a spike-in control during sequencing at concentrations representing ∼0.5% of reads generated per lane of flow cell. Reads were compared to the reference genome using k-mer matching with a k-mer size of 31. Matching 31-mers were allowed to differ by a Hamming distance of 1.

Reads were further trimmed on the basis of quality using the Phred algorithm in BBDuk and discarded following trimming whenever the final length of a read was <52 bp. In a study using publicly available Illumina paired-end transcriptomes, it was found that overly stringent trimming resulted in worse assemblies according to the majority of metrics measured, particularly for low-coverage datasets [44]. The adverse effects of trimming were reduced, however, as coverage was increased. Based on these findings, a gentle trimming strategy was recommended (Phred quality score between 2 and 5) except under specific scenarios when more aggressive trimming is warranted, such as in the case of exceptionally high sequencing depth. We tested the applicability of these guidelines to metagenomic assemblies with a mock microbial community [45]. In nearly all cases, aggressive trimming reduced the number of mismatches

and indels between the assembly and mapped short reads (Supplementary Fig. S5), but there was an increasingly detrimental impact on measures of assembly performance as quality score threshold was increased, which was especially evident in the lower coverage libraries. Therefore, a low quality threshold was chosen for quality-based trimming (Phred quality score of 5) because the metagenomes were determined by Nonpareil v3.3.1 [46] to not be fully saturated with reads.

Records with both reads in a pair passing quality control were co-assembled by sampling site using Megahit v1.1.1 (47) with parameters kstart = 27, kend = 127, and kstep = 20, and taxonomic bins reconstructed from the resultant assemblies using PhyloPythiaS+ [48]. The performance of the assembler on the datasets was evaluated based on read mapping rates and statistics provided by MetaQuast v4.6.1 [49]. Megahit was selected for assembly because of its reliable performance on datasets of highly complex microbial communities [50,51]. Megahit is also known to recover larger portions of strain variants than other, comparable short-read assemblers [52].

Identification and quantification of ARGs and MGEs

Contiguous sequences (or contigs) generated during assembly with length <200 bp were discarded prior to functional annotation. Putative coding DNA sequences were predicted from the remaining contigs and then translated into protein sequences using Prodigal v2.6.3 [53] in meta mode. Broad functional characterization of the gene predictions was performed through similarity searches against KEGG (release 83.2) [54] using the BLASTP subcommand of Diamond v0.9.14 [55] with the more-sensitive flag, a maximum expect value of e-5, and parameters unal = 0, max-target-seqs = 1, strand = both. Homologs of known antibiotic resistance determinants were identified from the predicted protein sequences using AMRFinder v1.04 with default settings [22]. ARGs detected through AMRFinder with family type equal to "equivalog" or "subfamily" were excluded to ensure that only the highest-confidence results were reported (see the AMRFinder documentation for a more thorough description of how ARGs are detected by AMRFinder).

The program hmmsearch (HMMER v3.1b) was used to detect remote homologs of MGE markers by searching a custom database of profile HMMs against the predicted protein sequences using manually curated gathering scores to determine which hits to report. The database was built by supplementing a set of custom profiles with select profiles taken from Pfam v32.0 [56]. Markers selected for inclusion in the database were chosen for their involvement in the transposition of DNA (i.e., integrases and transposases), and represent integrons as well as Type I and II transposable elements. To create the custom profiles, groups of related proteins belonging to the same marker family were downloaded from the NCBI RefSeq database of protein sequences. Duplicate sequences were clustered with CD-HIT, and very divergent (<20% identity) and very similar sequences (>80% identity) were removed prior to initial multiple sequence alignment (MSA) with MAFFT [57]. Subsequent steps, including filtering, final alignment, and trimming, were performed with T-Coffee [58]. First, sequence CORE (sCORE) scores were generated, and the distribution of scores analyzed to identify outliers. Sequences with poor sCORE were discarded and the 40 most informative sequences extracted. The most informative sequences were defined as the sequences diverging the most from each other and constituted the final set used in the construction of the seed MSA. Residues from the seed MSA with low transitive consistency score [59] were removed, and the resulting cu-

rated alignments used to construct the profile HMMs. Model-specific cut-offs were assigned to the profiles as in Punta et al. [60].

Coverage of the identified features was estimated by individually mapping the quality-controlled short reads to their relevant assembly using Bowtie2 v2.3.2 [61] with the very-sensitive flag and insert-size min and max parameters provided. These parameters were estimated from the insert-size mean and standard deviation calculated with BBmap v37.10. crAssphage coverage was calculated in an identical manner by mapping short reads to the crAssphage genome (GeneBank: NC_024711.1). To prevent artificial replicates from inflating coverage estimates, replicates in the mapped reads were first identified and removed using the MarkDuplicates functionality of Picard Tools v2.17.8. After duplicate filtering, mapped read best matches were sorted by name using Samtools v1.3 [62] and then used for estimating environmental abundance of the annotated sequences with count_features v1.3.0, part of the seq-annot open-source software package developed for the study.

For estimating feature abundances, read counts—the number of fragments thought to have originated from a given genomic region—were first transformed to fragment proportions (FP), a variant of transcripts per million (TPM) [63] without the application of a scaling factor. Each fragment used in the determination of the read counts was represented by a pair of aligned reads. The count of a feature was incremented whenever its coordinates fell within the interval of an alignment, defined as the region between 2 ends of a successfully mapped read pair, and an alignment interval overlapping multiple features was considered as separate evidence for the presence of each feature falling within its coordinates. Classes from the HTSeq Python library v0.9.1 [64] were incorporated into count_features for storing alignment and feature coordinates.

FP is a measure of the proportion of fragments in the underlying population produced from a given genomic region. Raw counts were replaced with nucleotide fractions by dividing read counts by effective length of the feature. FP further varies from TPM in that the effective length of a feature is equal to its actual length. In metagenomes, any given position of a feature is capable of producing fragments of any length, so consideration of fragment start position is unnecessary in the determination of the counts per bp rate. The nucleotide fraction of each feature was divided by the sum of all counts per bp rates to give the length-adjusted proportion of feature i out of n total features predicted from sample k:

$$\text{FP}_{i,k} = \frac{\text{counts}_{i,k}}{\text{length}_i(\text{bp})} \div \sum_i^{n_k} \frac{\text{counts}_{j,k}}{\text{length}_j(\text{bp})} \qquad (1)$$

ARG abundances were normalized to the abundance of the *rpoB* gene by dividing the FP of each ARG to the FP of *rpoB* in that metagenomic library (Table 1).

Detection of mobile ARGs

An assembly reconciliation program was used to merge contigs encoding 1 or more resistance genes prior to the detection of transposable ARGs. ARG-bearing contigs from across all sites were first combined and clustered at 100% identity. The dereplicated contigs were then merged using Mix v1.0 (65d) with parameters C = 100 and A = 200. Merged contigs were manually inspected for misassembly by mapping the dereplicated contigs onto the merged assembly with BLASTN (BLAST+ v2.7.1) [66].

Contigs that failed to map were then appended to the merged assembly to create the final set of unique ARG-bearing contigs. Components of MGEs were predicted from these contigs as described above with additional annotation in the form of repeat units from Repseek v6.6 [67] and attC attachment sites using cmsearch (INFERNAL v1.1.2) [68] with the covariance model from Cury et al. [69]. Transposable ARGs were defined as those encoded by a complete intracellular MGE, including autonomous transposons, composite transposons, ISCRs, and integrons. To be considered complete, an MGE was required to contain its relevant recombination module and be bounded by appropriate attachment/excision sites when applicable. For instance, integron-encoded ARGs were required to co-occur on the same contig as both an attC recombination site as well as an integron integrase.

Transferable ARGs were detected by searching the identified ARGs against a database of plasmid and ICE protein sequences using BLASTP with a reporting threshold of >99% sequence similarity. The database was composed of sequences from the ACLAME v0.4 [70] and ICEberg v2.0 [71] databases and supplemented with additional plasmid sequences downloaded from NCBI RefSeq. For each site, contigs were remapped to the set of unique ARG-bearing contigs and the results, which included ARG-encoding plasmids/ICEs as well as transposable ARGs, were matched to the ARG abundance table for each metagenome. The proportion of mobile ARG per metagenome was calculated as the abundance of transferable or transposable ARGs over the total abundance of ARGs.

Statistical analyses

All statistical analyses were performed in the R programming language v3.5.1 with the aid of reshape2 v1.4.3 [72] and several graphing and statistical libraries. Figures were generated with the data visualization library ggplot2 v3.1.0 [73], using color palettes from the dichromat package v2.0.0 [74]. Resistome percentages were calculated as the fraction of total normalized abundance for a given ARG or ARG category out of the total normalized abundance for all ARGs detected within the watershed samples. ARG diversity was calculated as the number of different ARG types detected at a site or within the watershed, where a given ARG type represents a unique entry in the AMRFinder database at the allele or exception level.

Linear regressions were performed with the modeling function lm (core stats package), and curves drawn using the geom_smooth layer of ggplot2 with linear model (lm) set as the smoothing method. Normalized ARG abundance was modeled as a linear function of the normalized abundance of crAssphage DNA. To determine whether the relationship was dependent on input from known sources of human waste, samples were grouped on the basis of whether they were collected from a site downstream from ≥ 1 WWTP or from a site with no upstream WWTP. A linear model of ARG and crAssphage abundances was compared to one that included an additional interaction term separating samples based on whether they were taken from sites located downstream from WWTPs or from sites upstream from all WWTPs. The model that included the interaction term was found to better fit the data (ANOVA, $P = 0.001927$), so the relationship between ARG and crAssphage abundances within sample groups was further tested individually. Normalized abundances were log-transformed prior to model fitting.

Geographic distance served as a proxy for river distance when assigning sites to 1 of 3 groups representing varying distances from WWTPs. Normalized coverage totals were log-transformed for between-group comparisons, and differences in total abundance assessed with ANOVA followed by the Tukey post hoc test for pairwise comparisons of group means. When abundance data were further divided by location of the resistance gene (chromosomal vs mobilized), the log-transformed data failed to meet test assumptions of normally distributed data with equal variance; in this case, significance was instead assessed through a non-parametric randomization procedure involving randomly reassigned sample labels. P-values were calculated by comparing observed F values with distributions generated from 10,000 permutations of the data. The level of statistical significance was set at 0.001 for all statistical tests performed.

Only samples collected from sites located upstream and within 5 km downstream from a WWTP were considered in the analysis of differential abundance (n = 18 and n = 15 per group, respectively). The change in abundance of ARG totals at downstream WWTP sites was calculated as the $\log_2$ of the ratio between downstream and upstream abundance, and statistical significance of differences in abundance between groups assessed using edgeR v3.24.3 [75] on unmodified read counts. Gene categories with FDR <0.05 were considered to be significantly more abundant in 1 group over the other.

Availability of Source Code and Requirements

- Project name: seq-annot
- Project home page: e.g., https://github.com/Brazelton-Lab/seq-annot
- Operating system(s): Linux
- Programming language: Python
- Other requirements: Python 3.4 or higher
- License: GPLv3
- biotoolsID: biotools:seq-annot
- RRID:SCR_018731
- Any restrictions to use by non-academics: none

Availability of Supporting Data and Materials

The raw metagenomic sequence data sets and the metagenome assemblies supporting the results reported here are available at the NCBI SRA and NCBI WGS, respectively, under BioProject accession PRJNA562643. The protocols used in the extraction of sample DNA can be found at https://baas-becking.biology.utah.edu/data/category/18-protocols. The custom software and scripts used in data processing are available from https://github.com/Brazelton-Lab. The R code used in data analysis is also available at the development GitHub page https://github.com/Brazelton-Lab/Thornton_2020. Snapshots of our code and other data further supporting this work are openly available in the *GigaScience* respository, GigaDB [76].

Additional Files

Additional File 1: Supplementary tables and figures.
Supplementary Table S1: Percent total watershed resistome by drug class.
Supplementary Table S2: ARG-bearing contigs matching known mobile genetic elements.
Supplementary Table S3: Summary of the antibiotic susceptibility testing results.

Abbreviations

ANOVA: analysis of variance; ARB: antibiotic-resistant bacteria; ARG: antibiotic resistance gene; BLAST: Basic Local Alignment Search Tool; BMD: broth microdilution; bp: base pairs; CARD: Comprehensive Antibiotic Resistance Database; cARG: chromosomal antibiotic resistance gene; CDC: Centers for Disease Control and Prevention; CLSI: Clinical & Laboratory Standards Institute; crAssphage: cross-assembly phage; EDTA: ethylenediaminetetraacetic acid; EUCAST: European Committee on Antimicrobial Susceptibility Testing; FDR: false discovery rate; FP: fragment proportions; HGT: horizontal gene transfer; HMM: hidden Markov model; ICE: integrative and conjugative element; ISCR: insertion sequence common region; KEGG: Kyoto Encyclopedia of Genes and Genomes; LSa: lincosamide-steptogramin a; MAFFT: Multiple Alignment using Fast Fourier Transform; mARG: mobile antibiotic resistance gene; MGE: mobile genetic element; MH: Mueller-Hinton; MIC: minimum inhibitory concentration; ML: macrolide-lincosamide; MLSb: macrolide-lincosamide-streptogramin B; MSb: macrolide-streptogramin B; MSA: multiple sequence alignment; QAC: quaternary ammonium compound; SRA: Sequence Read Archive; TPM: transcripts per million; WWTP: wastewater treatment plant.

Competing Interests

The authors declare that they have no competing interests.

Funding

Authors' Contributions

W.J.B. conceptualized the project, designed the experiments, and wrote the manuscript. C.N.T. assisted in the experimental design, performed the experiments, analyzed the data, developed the software, and wrote the manuscript. J.A.V. conceptualized the project and designed the experiments. W.D.T. conceptualized the project and designed and performed the experiments. All authors assisted in data collection and edited the manuscript.

Acknowledgements

References

1. Doi Y, Park YS, Rivera JI, et al. Community-associated extended-spectrum B-lactamase-producing *Escherichia coli* infection in the United States. Clin Infect Dis 2013;**56**(5):641–8.

2. Freeman J, Sexton D, Anderson D. Emergence of extended-spectrum B-lactamase-producing *Escherichia coli* in community hospitals throughout North Carolina: A harbinger of a wider problem in the United States? Clin Infect Dis 2009;**49**(2):e30–2.

3. Chen B, Yang Y, Liang X, et al. Metagenomic profiles of antibiotic resistance genes (ARGs) between human impacted estuary and deep ocean sediments. Env Sci Technol 2013;**47**(22):12753–60.

4. Bhullar K, Waglechner N, Pawlowski A, et al. Antibiotic resistance is prevalent in an isolated cave microbiome. PLoS One 2012;**7**(4):e34953.

5. D'Costa VM, King CE, Kalan L, et al. Antibiotic resistance is ancient. Nature 2011;**477**(7365):457–61.

6. Martinez, JL., Fajardo, A., Garmendia, L., Hernandez, A., Linares, JF., Martínez-Solano, L.and Sánchez, MB., A global view of antibiotic resistance, FEMS Microbiol Rev 2009;**33**: 44–65. 10.1111/j.1574-6976.2008.00142.x19054120

7. Lupo A, Coyne S, Berendonk TU. Origin and evolution of antibiotic resistance: The common mechanisms of emergence and spread in water bodies. Front Microbiol 2012;**3**, doi:10.3389/fmicb.2012.00018.

8. Guo J, Li J, Chen H, et al. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. Water Res 2017;**123**:468–78.

9. Munir M, Wong K, Xagoraraki I. Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan. Water Res 2011;**45**(2):681–93.

10. Su JQ, An XL, Li B, et al. Metagenomics of urban sewage identifies an extensively shared antibiotic resistome in China. Microbiome 2017;**5**:84.

11. Yang Y, Li B, Zou S, et al. Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. Water Res 2014;**62**:97–106.

12. Mao D, Yu S, Rysz M, et al. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater treatment plants. Water Res 2015;**85**:458–66.

13. Marti E, Jofre J, Balcazar JL. Prevalence of antibiotic resistance genes and bacterial community composition in a river influenced by a wastewater treatment plant. PLoS One 2013;**8**(10):e78906.

14. Proia L, von Schiller D, Sànchez-Melsió A, et al. Occurrence and persistence of antibiotic resistance genes in river biofilms after wastewater inputs in small rivers. Env Pollut 2016;**210**:121–8.

15. Subirats J, Triadó-Margarit X, Mandaric L, et al. Wastewater pollution differently affects the antibiotic resistance gene pool and biofilm bacterial communities across streambed compartments. Mol Ecol 2017;**26**(20):5567–81.

16. Gullberg E, Albrecht LM, Karlsson C, et al. Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. mBio 2014;**5**(5):e01918–14.

17. Rizzo L, Manaia C, Merlin C, et al. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci Total Env 2013;**447**:345–60.

18. Petrovich M, Chu B, Wright D, et al. Antibiotic resistance genes show enhanced mobilization through suspended growth and biofilm-based wastewater treatment processes. FEMS Microbiol Ecol 2018;**94**(5), doi:10.1093/femsec/fiy041.

19. Yuan QB, Guo MT, Wei WJ, et al. Reductions of bacterial antibiotic resistance through five biological treatment processes treated municipal wastewater. Env Sci Pollut Res Int 2016;**23**(19):19495–503.

20. Chu BTT, Petrovich ML, Chaudhary A, et al. Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. Appl Env Microbiol 2017;**84**(5):e02168–17.

21. , The Global Sewage Surveillance project consortium Hendriksen RS, Munk PThe Global Sewage Surveillance project consortium, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun 2019;**10**(1):1124.

22. Feldgarden M, Brover V, Haft DH, et al. Validating the AMRFinder Tool and Resistance Gene Database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob Agents Chemother 2019;**63**, doi:10.1128/AAC.00483-19.

23. Dutilh BE, Cassman N, McNair K, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun 2014;**5**(1):4498.

24. García-Aljaro C, Ballesté E, Muniesa M, et al. Determination of crAssphage in water samples and applicability for tracking human faecal pollution. Microb Biotechnol 2017;**10**(6):1775–80.

25. Aubert D, Girlich D, Naas T, et al. Functional and structural characterization of the genetic environment of an extended-spectrum β-lactamase bla$_{VEB}$ gene from a *Pseudomonas aeruginosa* isolate obtained in India. Antimicrob Agents Chemother 2004;**48**(9):3284–90.

26. CDC (Centers for Disease Control and Prevention), Antibiotic resistance threats in the United States, 2013.

27. Karkman A, Johnson TA, Lyra C, et al. High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. FEMS Microbiol Ecol 2016;**92**(3):fiw014.

28. Tang J, Bu Y, Zhang XX, et al. Metagenomic analysis of bacterial community composition and antibiotic resistance genes in a wastewater treatment plant and its receiving surface water. Ecotoxicol Env Saf 2016;**132**:260–9.

29. Czekalski N, Gascón Díez E, Bürgmann H. Wastewater as a point source of antibiotic-resistance genes in the sediment of a freshwater lake. ISME J 2014;**8**(7):1381–90.

30. Kristiansson E, Fick J, Janzon A, et al. Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. PLoS One 2011;**6**(2):e17038.

31. Rowe W, Verner-Jeffreys DW, Baker-Austin C, et al. Comparative metagenomics reveals a diverse range of antimicrobial resistance genes in effluents entering a river catchment. Water Sci Technol 2016;**73**(7):1541–9.

32. Munck C, Albertsen M, Telke A, et al. Limited dissemination of the wastewater treatment plant core resistome. Nat Commun 2015;**6**:8452.

33. Bengtsson-Palme J, Boulund F, Fick J, et al. Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. Front Microbiol 2014;**5**, doi:10.3389/fmicb.2014.00648.

34. Sundin GW, Bender CL. Dissemination of the strA-strB streptomycin-resistance genes among commensal and pathogenic bacteria from humans, animals, and plants. Mol Ecol 1996;**5**(1):133–43.

35. Wilkison DH, Armstrong DJ, Blevins DW. Effects of wastewater and combined sewer overflows on water quality in the Blue River basin, Kansas City, Missouri and Kansas, July 1998-October 2000. Water-Resources Investigations Report. US Geological Survey; 2002, doi:10.3133/wri024107.

36. Wilkison DH. Water quality in the Blue River Basin, Kansas City metropolitan area, Missouri and Kansas, July 1998 to October 2004. Scientific Investigations Report 2006-5147. US Geological Survey; 2006, doi:10.3133/sir20065147.

37. Karkman A, Pärnänen K, Larsson DGJ. Fecal pollution can explain antibiotic resistance gene abundances in anthropogenically impacted environments. Nat Commun 2019;**10**, doi:10.1038/s41467-018-07992-3.

38. United States Environmental Protection Agency (USEPA), Method 1604: total coliforms and Escherichia coli in water by membrane filtration using a simultaneous detection technique (MI Medium). Stand Methods2002;**18**:4–7.

39. Clinical and Laboratory Standards Institute, Development of in vitro Susceptibility Testing Criteria and Quality Control Parameters—4thed. M23-A4. Wayne, PA: Clinical and Laboratory Standards Institute, 2016.

40. Matuschek E, Åhman J, Webster C, et al. Antimicrobial susceptibility testing of colistin – evaluation of seven commercial MIC products against standard broth microdilution for *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Acinetobacter* spp. Clin Microbiol Infect 2018;**24**(8):865–70.

41. Brazelton WJ, Thornton CN, Hyer A, et al. Metagenomic identification of active methanogens and methanotrophs in serpentinite springs of the Voltri Massif, Italy. PeerJ 2017;**5**:e2945.

42. Thornton CN, Brazelton W. DNA Extraction from Sterivex Filters. protocols.io 2020. https://dx.doi.org/10.17504/protocols.io.bm2ek8be.

43. Rohland N, Reich D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res 2012;**22**(5):939–46.

44. MacManes MD. On the optimal trimming of high-throughput mRNA sequence data. Front Genet 2014;**5**, doi:10.3389/fgene.2014.00013.

45. Shakya M, Quince C, Campbell JH, et al. Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities:

Metagenomic and rRNA diversity characterization. Env Microbiol 2013;**15**(6):1882–99.

46. Rodriguez-R LM, Konstantinidis KT. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics 2014;**30**(5):629–35.

47. Li, D., Luo, R., Liu, CM., Leung, CM., Ting, HF., Sadakane, K., Yamashita, H.and Lam, TW., MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices, Methods 2016;**102**:3–11.10.1016/j.ymeth.2016.02.02027012178

48. Gregor I, Dröge J, Schirmer M, et al. PhyloPythiaS+: A self-training method for the rapid reconstruction of low-ranking taxonomic bins from metagenomes. PeerJ 2016;**4**:e1603.

49. Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: Evaluation of metagenome assemblies. Bioinformatics 2016;**32**(7):1088–90.

50. van der Walt AJ, van Goethem MW, Ramond JB, et al. Assembling metagenomes, one community at a time. BMC Genomics 2017;**18**, doi:10.1186/s12864-017-3918-9.

51. Vollmers J, Wiegand S, Kaster AK. Comparing and evaluating metagenome assembly tools from a microbiologist's perspective - not only size matters! PLoS One 2017;**12**(1):e0169662.

52. Awad S, Irber L, Brown CT. Evaluating metagenome assembly on a simple defined community with many strain variants. bioRxiv 2017, doi:10.1101/155358.

53. Hyatt D, Chen GL, LoCascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;**11**:119.

54. Ogata H, Goto S, Sato K, et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 1999;**27**(1):29–34.

55. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2014;**12**(1):59–60.

56. El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. Nucleic Acids Res 2019;**47**(D1):D427–32.

57. Katoh K. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;**30**(14):3059–66.

58. Notredame C, Higgins DG, Heringa J. T-coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000;**302**(1):205–17.

59. Chang JM, Di Tommaso P, Notredame C. TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. Mol Biol Evolut 2014;**31**(6):1625–37.

60. Punta M, Coggill PC, Eberhardt RY, et al. The Pfam protein families database. Nucleic Acids Res 2012;**40**(D1):D290–301.

61. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;**9**(4):357–9.

62. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;**25**(16):2078–9.

63. Li B, Ruotti V, Stewart RM, et al. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 2010;**26**(4):493–500.

64. Anders S, Pyl PT, Huber W. HTSeq–A Python framework to work with high-throughput sequencing data. Bioinformatics 2015;**31**(2):166–9.

65. Soueidan, H., Maurier, F., Groppi, A., Sirand-Pugnet, P., Tardy, F., Citti, C., Dupuy, V.and Nikolski, M., Finishing bacterial genome assemblies with Mix, BMC Bioinformatics 2013; **14 Suppl 15**:S16. 10.1186/1471-2105-14-S15-S1624564706

66. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: Architecture and applications. BMC Bioinformatics 2009;**10**(1):421.

67. Achaz G, Boyer F, Rocha EPC, et al. Repseek, a tool to retrieve approximate repeats from large DNA sequences. Bioinformatics 2007;**23**(1):119–21.

68. Nawrocki EP, Eddy SR. Infernal 1.1: 100-Fold faster RNA homology searches. Bioinformatics 2013;**29**(22):2933–5.

69. Cury J, Jové T, Touchon M, et al. Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Res 2016;**44**(10):4539–50.

70. Leplae R, Lima-Mendez G, Toussaint A. ACLAME: A CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res 2010;**38**(suppl_1):D57–61.

71. Liu M, Li X, Xie Y, et al. ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. Nucleic Acids Res 2019;**47**(D1):D660–5.

72. Wickham H. Reshaping data with the **reshape** package. J Stat Softw 2007;**21**(12), doi:10.18637/jss.v021.i12.

73. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer; 2016. https://ggplot2.tidyverse.org. Accessed: 2020-10-04.

74. Lumley T, Knoblauch K, Waichler S, et al. dichromat: Color Schemes for Dichromats. 2013. https://CRAN.R-project.org/package = dichromat. Accessed: 2020-10-04.

75. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;**26**(1): 139–40.

76. Thornton CN, Brazelton W, VanDerslice J, et al. Supporting data for "The influence of wastewater treatment on the resistome of an urban watershed." GigaScience Database 2020. http://doi.org/10.5524/100806.