

# **Sensitivity and consistency of long- and short-read metagenomics and epicPCR for the detection of antibiotic resistance genes and their bacterial hosts in wastewater**

Esther G. Lou<sup>1, #</sup>, Yilei Fu<sup>2, #</sup>, Qi Wang<sup>2</sup>, Todd J. Treangen<sup>2</sup> and Lauren B. Stadler<sup>1, \*</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, Rice University, 6100 Main Street, Houston, TX 77005, USA

<sup>2</sup>Department of Computer Science, Rice University, 6100 Main Street, Houston, TX 77005, USA

<sup>#</sup>These two authors contributed equally to this work.

\*Corresponding author: [lauren.stadler@rice.edu](mailto:lauren.stadler@rice.edu)

## **Abstract**

Wastewater surveillance is a powerful tool to assess the risks associated with antibiotic resistance in communities. One challenge is selecting which analytical tool to deploy to measure risk indicators, such as antibiotic resistance genes (ARGs) and their respective bacterial hosts. Although metagenomics is frequently used for analyzing ARGs, few studies have compared the performance of long-read and short-read metagenomics in identifying which bacteria harbor ARGs in wastewater. Furthermore, for ARG host detection, untargeted metagenomics has not been compared to targeted methods such as epicPCR. Here, we 1) evaluated long-read and short-read metagenomics as well as epicPCR for detecting ARG hosts in wastewater, and 2) investigated the host range of ARGs across the WWTP to evaluate host proliferation. Results highlighted long-read revealed a wider range of ARG hosts compared to short-read metagenomics. Nonetheless, the ARG host range detected by long-read metagenomics only represented a subset of the hosts detected by epicPCR. The ARG-host linkages across the influent and effluent of the WWTP were characterized. Results showed the ARG-host phylum linkages were relatively consistent across the WWTP, whereas new ARG-host species linkages appeared in the WWTP effluent. The ARG-host linkages of several clinically relevant species found in the effluent were identified.

**Keywords:** antibiotic resistance genes (ARGs), antibiotic resistant bacteria (ARB), wastewater treatment plant (WWTP), metagenomics, epicPCR

## 1 **1. Introduction**

2           The worldwide propagation and dissemination of antibiotic resistance have raised serious  
3 public health concerns. An estimated 1.27 million deaths were attributed to bacterial antibiotic  
4 resistant infections in 2019<sup>1</sup>. To mitigate this risk, a comprehensive understanding of antibiotic  
5 resistance in humans, animals, and the environment is needed<sup>2</sup>. Wastewater treatment plants  
6 (WWTPs) are regarded as hotspots of antibiotic resistance in the environment<sup>3,4</sup>, and their role in  
7 the dissemination of antibiotic resistance is complex. Wastewater treatment processes generally  
8 remove antibiotic resistance genes (ARGs) and antibiotic resistant bacteria (ARB) from sewage  
9 through different mechanisms such as, anaerobic<sup>5,6</sup> and aerobic processes<sup>7,8</sup>, coagulation and  
10 sedimentation<sup>9</sup>, membrane filtration<sup>10,11</sup>, and disinfection<sup>12,13</sup>. In spite of the significant removal  
11 of the overall abundance of ARGs and ARB by WWTPs<sup>14–16</sup>, certain ARGs and ARB can be  
12 persistent or even enriched across the treatment units<sup>17–19</sup>, representing a major source of ARGs  
13 and ARBs in receiving waterbodies. For example, ARGs and microorganisms discharged by  
14 WWTPs can persist in receiving-river biofilms<sup>20</sup> and sediments<sup>21,22</sup>. Furthermore, ARGs and  
15 ARB after establishing in the effluent-receiving environment can propagate to distant areas away  
16 from the discharge point<sup>21,23</sup>.

17           Previous studies have proposed metrics for evaluating risks associated with ARGs in the  
18 environment to public health<sup>24,25</sup>. One key component is identification of the bacterial host of the  
19 ARG, as a pathogenic bacteria harboring an ARG is a far greater public health risk as compared  
20 to a non-pathogenic environmental bacterial host of the same ARG<sup>26,27</sup>. A second component is  
21 whether an ARG is associated with mobile genetic elements (MGEs), which can directly and  
22 indirectly mediate horizontal gene transfer (HGT) of ARGs among microorganisms<sup>28–30</sup>. Thus,  
23 methods are needed that not only identify ARGs in environmental samples, but also provide

24 information that link those ARGs to their microbial hosts and contextual information about  
25 whether the ARGs are associated with MGEs.

26         One of the most widely used methods to analyze antibiotic resistance in wastewater is  
27 metagenomic sequencing<sup>31–33</sup>. Next-generation sequencing (i.e., short-read) coupled with *de*  
28 *novo* assembly recovers ARG-host and ARG-MGE linkages by screening taxonomical markers  
29 and MGEs on the assembled ARG-carrying contigs<sup>29,34,35</sup>. However, this method suffers from  
30 limited detection sensitivity due to the low percentage of raw reads mapping back to the  
31 assembled contigs<sup>36</sup> and intergenomic assembly errors<sup>37</sup>. Importantly, a large portion of reads  
32 associated with MGEs fail to assemble because of the extended homologous and mosaic  
33 sequences found in those regions<sup>38–40</sup>. Third-generation sequencing technologies (i.e., long-read)  
34 are excellent at tracking ARG hosts in environmental samples<sup>28,41</sup> because long-read sequencing  
35 can directly reveal the genetic context of ARGs thanks to the extended read length. For example,  
36 two wastewater metagenomic studies reported greater numbers of long reads generated via  
37 Oxford Nanopore Technology (ONT) than numbers of short-read assembled contigs, and an  
38 average long read length of 2-10 kbp, which was significantly longer than the average length of  
39 short-read assembled contigs<sup>28,42</sup>. However, because both long- and short-read metagenomic  
40 sequencing are untargeted methods, their ability to detect low abundance and rare ARGs is  
41 limited<sup>43,44</sup>. On the other hand, emerging targeted methods such as single cell fusion PCR  
42 methods, called epicPCR (Emulsion, Paired Isolation, and Concatenation PCR), can overcome  
43 sensitivity limitations. In epicPCR single cells are isolated and encapsulated in a polyacrylamide  
44 bead within which PCR takes place to fuse a target ARG with the 16S rRNA gene<sup>45</sup>. As a result,  
45 PCR amplifies the signal of the ARG and its associated host 16S DNA from the background

46 environmental metagenome, improving detection sensitivity. No studies to date have directly  
47 compared ARG hosts detected using untargeted sequencing and targeted fusion PCR methods.

48 In this study, we compared the sensitivity and consistency of ARG hosts detection via  
49 three different methods: short-read sequencing, long-read sequencing, and epicPCR. In addition,  
50 for the untargeted metagenomic sequencing methods, we also analyzed the genetic context of the  
51 detected ARGs, including their associations with MGEs. We then applied these methods to  
52 samples collected across a WWTP to characterize ARG hosts shifts across the wastewater  
53 treatment process, and to identify high-risk ARGs associated with putative pathogens and MGEs  
54 in the final effluent. The results of this work reveal breadth vs. sensitivity tradeoffs associated  
55 with method selection for identifying ARG hosts in wastewater monitoring programs.

56

## 57 **2. Materials and Methods**

### 58 **2.1. Sample collection, DNA extraction, and pretreatment for epicPCR**

59 Wastewater samples were collected from a conventional WWTP (City of West  
60 University Place WWTP, Houston, Texas, USA) that treats an average of 2 million gallons of  
61 municipal sewage per day. This WWTP employs a conventional aerobic activated sludge process  
62 as secondary treatment, followed by chlorination disinfection (gaseous Cl<sub>2</sub>, 2-4 mg/L effective  
63 chlorine concentration, 20 mins contact time). Nine grab samples were collected from three  
64 sampling locations, WWTP influent, secondary effluent, and final effluent on three consecutive  
65 dry days (n=3 for each sampling location). All samples were collected at the same time of the  
66 day to avoid diurnal variations. After collection, samples were kept on ice, immediately  
67 transported to the lab, and processed within 45 minutes of collection.

68 DNA was extracted from all samples prior to conducting long- and short-read  
69 metagenomic sequencing. A 50 mL influent sample, 250 mL secondary effluent sample, and 500  
70 mL final effluent sample were filtered through a cellulose nitrate membrane filter (pore size 0.22  
71  $\mu\text{m}$ , diameter 47 mm; Millipore Sigma) to concentrate biomass. Next, filters were cut into small  
72 pieces using sterilized forceps and transferred to a 2 mL tube containing 0.1 mL glass beads for  
73 bead-beating, followed by DNA extraction. A Maxwell RSC Instrument (Cat. Num. AS4500,  
74 Promega) using Maxwell RSC PureFood GMO and Authentication kits (Cat. Num. AS1600,  
75 Promega) were used to extract DNA. For epicPCR, all influent (n=3) and final effluent samples  
76 (n=3) were centrifuged to concentrate cells for cell counting, polymerization, and cell lysis as  
77 previously described<sup>85</sup>. Only samples with good cell separation and partitioning in  
78 polyacrylamide beads (i.e., one single cell per 35-50 polyacrylamide beads) were used in the  
79 downstream experiments to avoid false positive detections. Details of DNA extraction and  
80 sample pretreatment for epicPCR are provided in Supplementary Information Section 1.2.

81

## 82 **2.2. Sequencing epicPCR product using MinION (ONT)**

83 We selected three ARG targets for epicPCR analysis: *sull*, *ermB*, and *tetO*. They were  
84 chosen because of their wide host range as previously reported<sup>24,86,87</sup>. For example, *ermB*, the  
85 macrolide-lincosamide-streptogramin B (MLS<sub>B</sub>) resistance gene, is of clinical relevance because  
86 it is enriched in human-related environments, harbored by human pathogens, and often carried  
87 on MGEs<sup>24</sup>. Primer sequences for the three targets used in this study are listed in Supplementary  
88 Information Table 1. Details of the epicPCR experiments consisting of fusion PCR and nested  
89 PCR are provided in Supplementary Information 1.2. After attaining nested PCR products,  
90 library preparation and sequencing were performed following the protocol “Native barcoding

91 amplicons (with EXP-NBD104, EXP-NBD114, and SQK-LSK 109)” (ONT). The pooled library  
92 was loaded on an R9.4 flow cell (MIN-FLO106, ONT) in a MinION device. The sequencing run  
93 was monitored via the software MinKNOW (v.20.10), targeting a >1000X depth per sample.

94

### 95 **2.3. Metagenomics sequencing (long- and short-read)**

96 DNA extracts of all samples were measured using a Qubit Broad Range dsDNA assay kit  
97 and Qubit 2.0 fluorometer. DNA quality was then evaluated using electrophoresis to ensure a  
98 DNA size greater than 3 kbps. For short-read sequencing, DNA extracts were shipped on dry ice  
99 to BGI Tech Solutions (Hong Kong) Co., Ltd for DNBseq general DNA library construction and  
100 DNBseq platform sequencing. For long-read sequencing, 500 ng of DNA from each of the three  
101 sample replicates were combined for library preparation following the protocol “Genomic DNA  
102 by ligation (SQK-LSK 109)” (ONT). Each of the three libraries (influent, secondary effluent, and  
103 final effluent) was loaded onto a Flow Cell R9.4 (MIN-FLO106, ONT) and sequenced with a  
104 MinION device. The sequencing run was controlled via MinKNOW (v.20.10). Long- and short-  
105 read sequencing statistics are provided in Supplementary Information Table 2.

106

### 107 **2.4. Analysis of epicPCR reads for ARG host range profiling**

108 Raw reads were basecalled via guppy\_basecaller (Version 4.4.1+1c81d62). Basecalled  
109 reads were trimmed by Porechop (<https://github.com/rrwick/Porechop>) and filtered using  
110 Nanofilt with a minimum quality score of 7<sup>88</sup>. Next, all reads were searched against the  
111 corresponding linker primer sequence (RL-*sull*-519F', RL-*ermB*-519F', and RL-*tetO*-519F')  
112 using BLAST. The output reads were filtered using the perfect match criteria (100% identity and  
113 100% length coverage) to exclude partially fused fragments, and only complete ARG-16S rRNA

114 fusion structures were included for the downstream analysis. Then, we used a customized script  
115 to split the fusion structures into the ARG and the 16S rRNA gene portions based on the reverse  
116 linker position. We then conducted taxonomic classification on the 16S rRNA gene portion using  
117 Emu<sup>89</sup> and the SILVA ribosomal RNA gene database (release 138, 2019). To avoid false  
118 positives, two actions were taken to further filter the reads: 1) The ARG portion of all split reads  
119 was aligned against the SARG database<sup>90</sup> using BLAST, 2) for each sample, only hosts that were  
120 identified consistently from at least two sample replicates were counted. The results of epicPCR  
121 sequencing statistics can be found in Supplementary Information Section 2.1.

122

## 123 **2.5. Analysis of metagenomic sequencing reads generated by long-read and short-read** 124 **sequencing**

125 We processed long- and short-read sequencing data in an integrated pipeline as shown in  
126 Supplementary Information Fig. 1. Our metagenomic analysis included: (1) identification of  
127 ARGs on long reads (via long-read sequencing) or short-read-assembled contigs (via short-read  
128 sequencing); (2) filter ARG-carrying long reads and contigs to include only those that were  
129 chromosome-associated for the host classification step; (3) identification of MGEs that were  
130 located on the same read or contig as the ARGs, and (4) identification of ARG host by  
131 taxonomic classification of the chromosomal reads or contigs that were associated with ARGs.  
132 Detailed methods describing the pipeline used to detect ARG-carrying long reads via BLAST,  
133 ARG-carrying contigs via CARD's Resistance Gene Identifier (RGI)<sup>91</sup>, and ARG-MGE linkages  
134 are provided in Supplementary Information Table 7. ARG relative abundance was calculated by  
135 normalizing the copy number of ARGs detected on long reads or assembled contigs to the total  
136 giga base pairs (Gbp) of the sample<sup>41</sup>.



137 ARG-carrying reads or contigs were categorized as “chromosome,” “plasmid,” or  
138 “unclassified” via PlasFlow (V1.1)<sup>92</sup>. The “unclassified” reads and contigs were re-classified via  
139 megaBLAST against the NCBI nt database with a minimum bit score of 50, an E value threshold  
140 of  $e^{-10}$ , and a 70% sequence similarity cutoff, followed by keyword match (“chromosome”) to  
141 retrieve chromosome-associated long reads and contigs. All ARG-carrying reads and contigs  
142 classified as “chromosome” were subject to taxonomic classification using Centrifuge (V1.0.4)<sup>93</sup>.  
143 ARG-host linkages were identified by summarizing the associations between each ARG and the  
144 taxonomic classification result of the corresponding ARG-carrying read or contig. Putative  
145 pathogens were scanned according to the WHO resistant pathogen list<sup>94</sup>. In addition, three  
146 publicly available datasets from NCBI SRA were downloaded, each consisting of long-read (via  
147 Nanopore) and short-read (via Illumina) data based on sequencing the same wastewater  
148 sample<sup>28,31</sup>. These datasets were run through identical pipelines for analyzing long- and short-  
149 read sequencing data as used in this study to identify ARG-host linkages. Details regarding the  
150 three datasets are provided in Supplementary Information Table 3. Plasmid-associated ARG-  
151 carrying reads and contigs were subject to plasmid mobility prediction using MOB-suite  
152 (v3.0.3)<sup>95</sup> and MOBscan (<https://castillo.dicom.unican.es/mobscan/>). Furthermore, to compare  
153 long-read sequencing with epicPCR for ARG host profiling, we processed Centrifuge  
154 specifically for those long reads that were found to carry *ermB*, *sull*, and *tetO*.

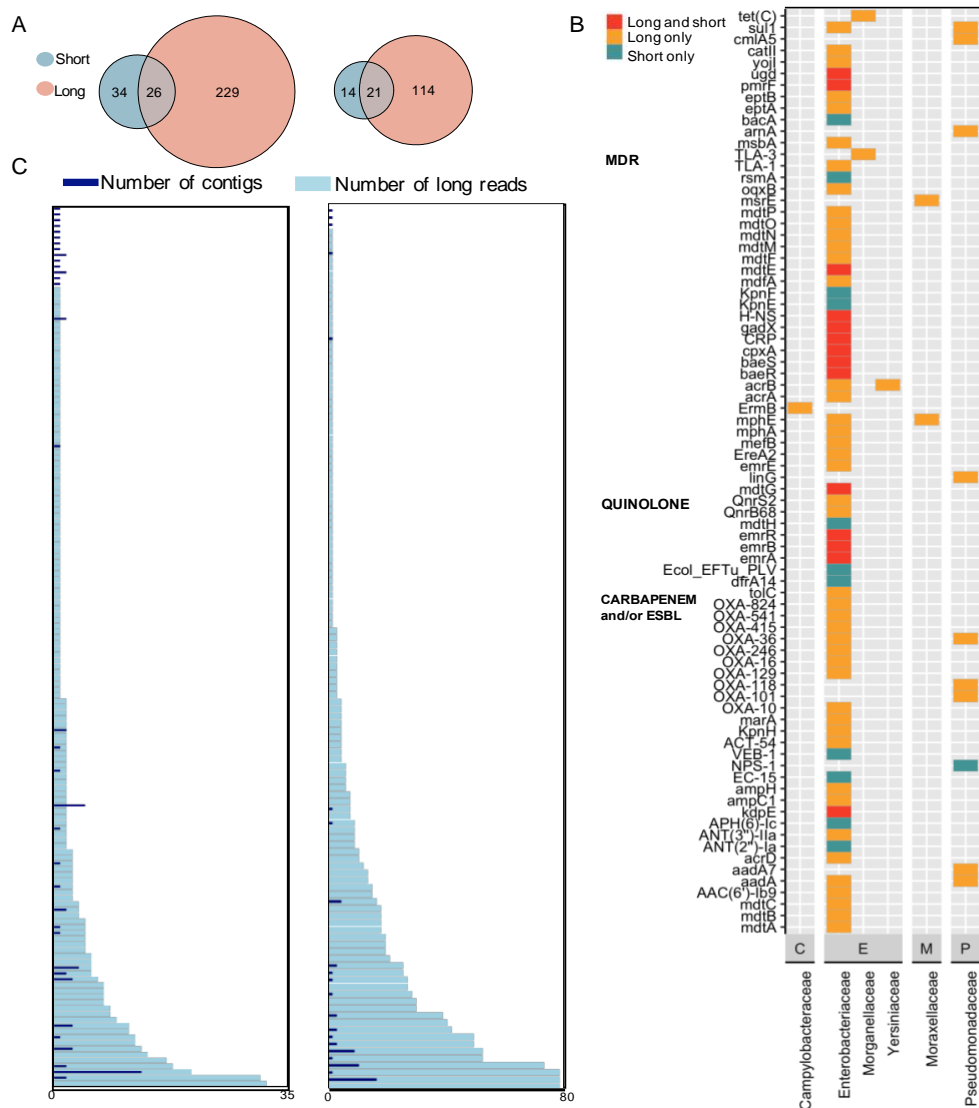
155

### 156 **3. Results and Discussion**

157

#### 158 **3.1. Long-read sequencing demonstrated superior performance for ARG host identification** 159 **as compared to short-read sequencing**

160 Long-read sequencing identified a greater number of linkages of ARGs and their hosts  
161 than short-read sequencing (Fig. 1a). This result highlights that long-read sequencing produced a  
162 more diverse host profile than short-read sequencing, even though both methods produced  
163 similar total community bacterial composition profiles (Supplementary Information Table 4) and  
164 consistent ARG subtype profiles (Supplementary Information Fig. 2a). However, these two  
165 methods showed inconsistency in ARG host identifications. In total, 26 ARG-host family  
166 linkages, or 21 ARG subtype-host family linkages, were consistently detected by both methods,  
167 which accounts for only a small fraction of the corresponding total linkages detected by each  
168 method (Fig. 1a). Although several studies have focused on the consistency of long-read and  
169 short-read sequencing in resistome analysis<sup>28,41,46</sup> and sample-wise taxonomic abundance  
170 estimation<sup>47-50</sup>, ours is the first to explicitly compare their ability to characterize ARG host range  
171 in wastewater and reveals inconsistencies across the methods (discussed below).  
172



**Fig. 1.** Comparison of long- and short-read sequencing for wastewater ARG host identification. **A.** Venn diagrams illustrating the detections of unique linkages of a specific ARG and its host family (left), and of unique linkages of a specific ARG drug class subtype and its host family (right). **B.** Profile of ARG bacterial hosts on the WHO priority list. Highlighted ARGs are those conferring multidrug resistance (MDR), fluoroquinolone resistance, and those encoding ESBL-production and/or carbapenemase-production. The colors denote a detection of an ARG-host (orange: detected only by long-read sequencing, turquoise: detected only by short-read sequencing, red: detected by both sequencing technologies). Family-level hosts are grouped by Order on x-axis (C: Campylobacteriales, E: Enterobacterales, M: Moraxellales, P: Pseudomonadales). **C.** The number of reads (X-axis) via long-read sequencing (light blue bars) and the number of contigs via short-read sequencing (dark blue bars) supporting each unique linkage of ARG subtype and host family (Y-axis; specific linkages are not annotated on the graph). The left panel consists of data generated in this study, and the right panel is using publicly available data of a sample collected from the influent of a WWTP in Boston, MA (sample ID: B\_ww\_1, Supplementary Information Table 3).

174 We further compared the ARG-host linkages identified by each method, focusing  
175 specifically on those putative hosts included on the WHO's resistant pathogen list. Long- and  
176 short-read sequencing altogether recovered 117 ARG-host linkages, covering 80 ARGs  
177 (corresponding to 17 subtypes) and 31 putative pathogenic species (Fig. 1b). Both methods  
178 identified *Escherichia coli* as the putative pathogenic host that carried the highest abundance of  
179 ARGs. In fact, according to global surveillance of clinical cases, among all bacterial pathogens  
180 associated with or attributable to antibiotic resistance, *E. coli* ranks first as the cause of direct or  
181 indirect deaths<sup>1</sup>. Not surprisingly, most putative hosts identified were within the family  
182 *Enterobacteriaceae*, which includes the vast majority of commensal and enteric bacteria that live  
183 in the gastrointestinal tract of humans<sup>51-53</sup>. Consistent with previous studies, *Enterobacteriaceae*  
184 was found to harbor multiple classes of clinically relevant ARGs, especially those encoding  
185 ESBL (extended-spectrum beta-lactamase)-production and/or carbapenemase-production<sup>51,54-56</sup>.  
186 Long-read sequencing detected ARG-host connections across six host families and 68 ARGs,  
187 whereas short-read sequencing only detected *Enterobacteriaceae* and *Pseudomonadaceae* as the  
188 host families for 24 ARGs (Fig. 1b). Hence, despite having a significantly shallower sequencing  
189 depth (long-read method sequenced only 10.12% of the total bases sequenced by short-read  
190 method for the same wastewater sample; Supplementary Information Table 2), long-read  
191 sequencing detected a more comprehensive profile of putative pathogenic hosts of ARGs than  
192 short-read sequencing (Supplementary Information Table 5).

193 To investigate the inconsistency between the two sequencing methods, we compared the  
194 number of reads supporting the ARG-host linkages detected by long-read sequencing and the  
195 number of contigs supporting the ARG-host linkages detected by short-read sequencing (Fig.  
196 1c). In addition, we also compared our results with an existing publicly available dataset

197 containing short- and long-read metagenomic sequencing data of the same wastewater microbial  
198 community (ID: B\_ww\_1<sup>31</sup>; Fig. 1c, right; Supplementary Information Fig. 3). As expected,  
199 long-read sequencing demonstrated more ARG subtype-host family linkages as compared to  
200 short-read sequencing. Quantitatively, the numbers of long reads were generally greater than the  
201 numbers of contigs across the vast majority of the ARG subtype-host family linkages (Fig. 1c).  
202 For those reads and contigs that supported the same linkages, their numbers were moderately  
203 correlated (n=21, Spearman's Rho=0.43, p<0.05 for this study; n=18, Spearman's Rho=0.47,  
204 p<0.05 for B\_ww\_1), which indicates some degree of consistency between these two methods in  
205 quantifying the linkages of ARG subtypes and host families.

206 Thus, our results show that long-read sequencing demonstrated superior performance  
207 over short-read sequencing in detecting ARG hosts in two respects: 1) it captured a wider host  
208 range for different ARGs (Fig. 1b) and ARG subtypes (Fig. 1c, Supplementary Information Fig.  
209 1a, 4); and 2) quantitatively, it detected ARG-host linkages by generating greater numbers of  
210 reads supporting the linkages as compared to short-read derived contigs (Fig. 1c, Supplementary  
211 Information Fig. 3). Of note, the evaluation was based on comparing the detection via long reads  
212 with the detection via contigs assembled from short reads, rather than directly comparing the raw  
213 reads generated by both sequencing methods. Several previous studies used raw short reads  
214 without assembly and identified putative ARG hosts through correlation analysis that compared  
215 the abundance of ARGs and taxonomical markers<sup>57,58</sup>. However, using raw reads to assign ARG  
216 hosts for wastewater surveillance has several challenges. First, this approach relies heavily on  
217 statistical correlation analysis which requires multiple sample replicates. Obtaining and  
218 processing multiple replicate samples significantly increases the amount of work required for  
219 sample collection, preparation, and sequencing, as well as the time and cost involved in routine

220 surveillance. Most importantly, the raw read approach is prone to introduce false positives when  
221 detecting ARG-host linkages<sup>59</sup>. Thus, while using assembly is a relatively conservative means to  
222 identify ARG-host linkages as compared to using raw short-reads, it is less likely to generate  
223 false positives, which is crucial in wastewater surveillance and risk assessment. Furthermore, the  
224 substantial processing time and computational memory requirements of read assembly of short-  
225 read sequencing data can be massively reduced by using long-reads. Recent studies have shown  
226 that ONT, one of the leading long-read sequencing technologies, can rapidly and reliably detect  
227 resistomes and pathogens in one hour in wastewater<sup>50</sup>.

228

### 229 **3.2. EpicPCR detected more ARG hosts as compared to long-read sequencing**

230 Since long-read sequencing revealed more ARG hosts than short-read sequencing, we  
231 next compared the ARG hosts detected using long-read sequencing to the host range of three  
232 ARG targets (*sull*, *ermB*, and *tetO*) detected by epicPCR. We found that epicPCR detected a  
233 greater number of host species for the three ARG targets than long-read sequencing (Table 1). As  
234 expected, epicPCR was more sensitive than long-read sequencing for ARG host detection as it is  
235 a targeted method that includes a PCR amplification step that enhances the signal of the target  
236 ARGs and bacterial host marker genes. An additional reason why epicPCR may be more  
237 sensitive than metagenomic sequencing is because it includes plasmid-associated linkages that  
238 may have been overlooked by metagenomics. With epicPCR, as long as the target ARG is  
239 present in the cell, it can be fused with the taxonomical marker (i.e., 16S rRNA gene) via PCR  
240 for host classification regardless of whether the ARG is located on a plasmid or chromosome.  
241 However, metagenomics analysis pipelines generally only classify hosts for ARGs that are  
242 located on the chromosome because the analysis requires the presence of taxonomic markers co-

243 located on the ARG read, which are more frequently found in the chromosome than on plasmids.

244 Thus, epicPCR should generate a more comprehensive host profile than metagenomics, because

245 ARGs are widely distributed on plasmids<sup>28,42,60</sup>.

246

247 **Table 1.** The number of ARG-associated reads and host species detected by epicPCR and long-read  
 248 sequencing in WWTP influent and effluent samples (n=3). The comprehensive list of hosts detected by  
 249 epicPCR can be found in Supplementary Information Table 6.

WWTP influent						
Method	EpicPCR	Long-read metagenomics				Both
	Host species	Total reads	Host species via classifying total reads	Chromosomal reads	Host species via classifying chromosomal reads	Host species
<i>sulI</i>	311	124	32	3	3	2
<i>ermB</i>	57	18	8	3	1	1
<i>tetO</i>	104	30	20	8	4	4
WWTP effluent						
Method	EpicPCR	Long-read metagenomics				Both
	Host species	Total reads	Host species via classifying total reads	Chromosomal reads	Host species via classifying chromosomal reads	Host species
<i>sulI</i>	323	8	7	0	-	-
<i>ermB</i>	14	0	-	0	-	-
<i>tetO</i>	35	0	-	0	-	-

250  
 251 The significantly lower number of hosts identified via long-read sequencing as compared  
 252 to epicPCR was likely due to the low fraction of chromosomal reads among the total ARG-  
 253 associated reads. To further investigate, we selected all long reads that were found to carry *sulI*,  
 254 *ermB*, and *tetO* disregarding whether they were on chromosomes. As expected, the vast majority  
 255 of ARG-carrying reads were not classified as chromosomal reads for the three ARG targets  
 256 (Table 1). Unfortunately, it is not feasible to classify hosts using metagenomics for non-  
 257 chromosomal ARGs, such as plasmid-associated ARGs. This is because phylogenetic analysis of  
 258 plasmids is extremely challenging, due to the hardship to reconstruct the potentially shared



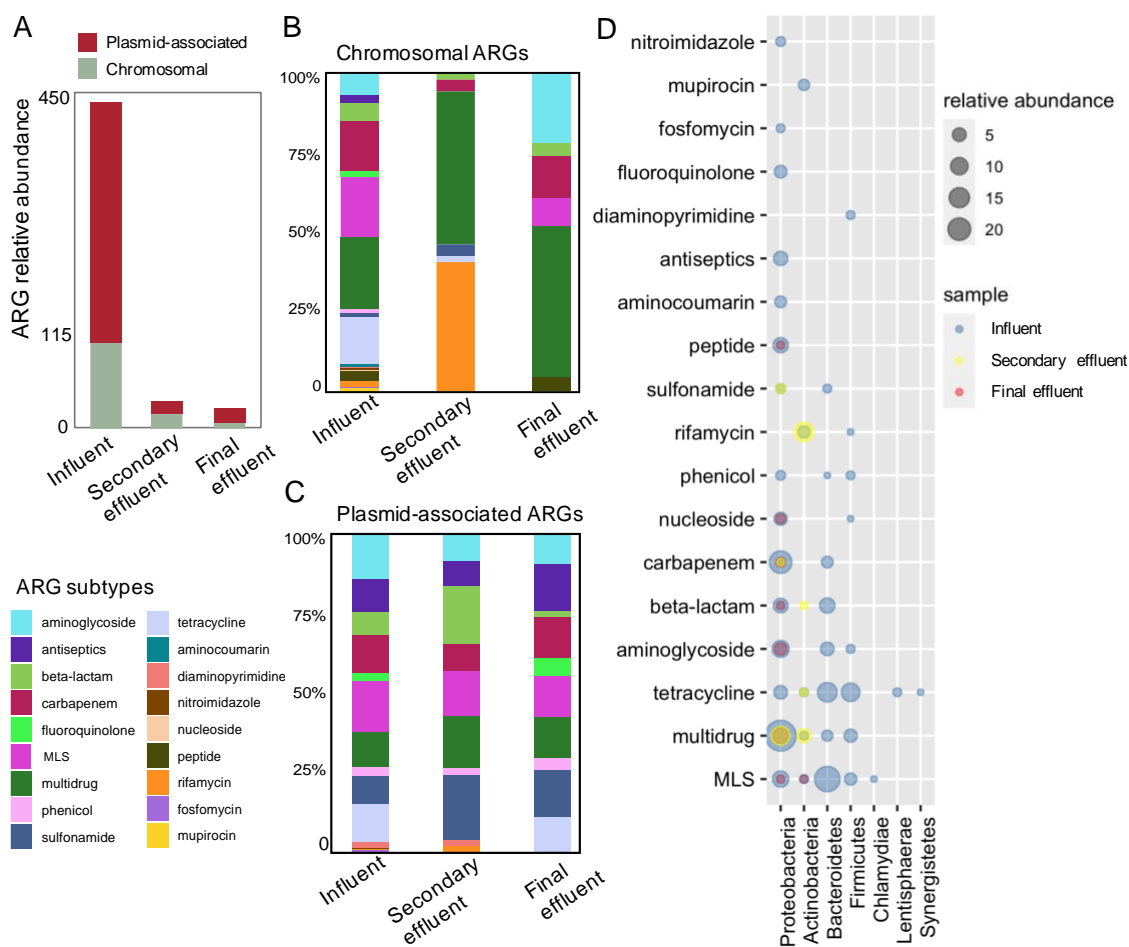
259 “core” genes<sup>61</sup> by plasmid subtypes<sup>62,63</sup>. In addition, plasmids can be horizontally transferred  
260 among different hosts and the transfer pattern is still not fully unveiled.

261 Despite epicPCR’s improved detection sensitivity over long-read sequencing, one  
262 undeniable value of long-read sequencing is its ability to capture the genetic context of an ARG.  
263 For instance, long-read sequencing identified the associations between *ermB* and the gene  
264 encoding conjugative transposon proteins in *Clostridioides difficile* (data not shown),  
265 highlighting *ermB*’s potential to be horizontally transferred. EpicPCR results in only a short,  
266 fused product of the target gene and taxonomic marker gene, retaining no additional contextual  
267 information in the sequenced amplicons. Thus, there is a tradeoff between sensitivity and  
268 contextual information that should be considered when deciding whether to use epicPCR versus  
269 long-read sequencing for ARG detection and risk assessment in environmental monitoring.

270

### 271 **3.3. The ARG-host phylum linkages were relatively consistent across WWTP influent and** 272 **effluent, whereas new ARG-host species linkages appeared in the WWTP effluent**

273 ARGs were efficiently removed via conventional activated sludge treatment followed by  
274 chlorination disinfection, where we observed a 93.6% removal rate of all ARGs based on the  
275 relative abundance of total ARGs across the WWTP (Fig. 2a). The removal rate across the



**Fig. 2.** Dynamics of resistomes and ARG hosts across the WWTP revealed by long read sequencing. A. The relative abundance of ARGs across the WWTP (influent, secondary effluent and final effluent,  $n = 3$  for each sampling location). ARGs were grouped by their location (red: plasmids, green: chromosome). B. The composition of chromosomal ARGs across samples. ARGs are colored by drug class subtype. C. The composition of plasmid-associated ARGs across samples. ARGs are colored by drug class subtype. D. The ARG host phyla across samples. ARGs are grouped by subtype (y-axis) and hosts are grouped by phyla (x-axis). The size of dots represents the relative abundance of ARGs corresponding to the specific subtype and host phylum. Dot colors indicate sampling location.

276 activated sludge treatment process (91.8%) was comparable to those reported in previous  
 277 studies<sup>14,30,64</sup>. The chlorination process further removed chromosomal ARGs, but resulted in a  
 278 slight increase in the relative abundance of plasmid-associated ARGs, leading to limited removal  
 279 of total ARG relative abundance (21.0%) from the secondary effluent (Fig. 2a). Although the  
 280 role of chlorination remains under debate with respect to its impact on antibiotic resistance<sup>13</sup>,

281 several studies have shown that chlorination has a limited or even negative effect on the removal  
282 of ARGs from secondary effluent<sup>19,65–67</sup>.

283 To further understand the dynamics of resistomes across the treatment processes, the  
284 composition of ARGs on chromosomes and plasmids was assessed separately with respect to  
285 ARG subtypes (Fig. 2b). Chromosomal and plasmid-associated ARGs shared 16 ARG subtypes,  
286 whereas ARGs encoding resistance to fosfomycin and mupirocin were only detected on  
287 chromosomes, and ARGs conferring resistance to colistin and glycopeptide were only found on  
288 plasmids. The distribution of ARGs across chromosomes and plasmids is likely influenced by  
289 their resistance mechanism. We observed that ARGs causing antibiotic inactivation,  
290 replacement, or protection of the antibiotic's target, were more frequently associated with  
291 plasmids than chromosomes, while ARGs associated with efflux pumps were more frequently  
292 located on chromosomes (Supplementary Information Fig. 4). This distribution pattern is  
293 consistent with a recent study investigating the distribution of ARGs across chromosomes and  
294 plasmids in major groups of *Enterobacteriaceae*<sup>68</sup>.

295 In general, the ARG subtypes present in secondary effluent and final effluent samples  
296 were a subset of those in influent samples (Fig. 2b,c). However, chromosomal ARGs (Fig. 2b)  
297 demonstrated a less consistent composition profile across the treatment processes as compared to  
298 plasmid-associated ARGs (Fig. 2c). Among the chromosomal ARGs, a spike of rifamycin  
299 resistance genes (i.e., *rpoB2*, *RbpA*, and *efpA*) in was observed in the secondary effluent and was  
300 likely attributed to the growth of their putative host Actinobacteria (Fig. 2d), whose relative  
301 abundance increased substantially in secondary effluent as compared to in the influent (data not  
302 shown). Similarly, the fraction of multidrug resistance genes (MDRs) increased in the secondary  
303 effluent (Fig. 2b), and these MDRs were also carried by Actinobacteria (Fig. 2d). The growth of

304 Actinobacteria bacteria, which are common aerobes<sup>69</sup>, was likely the result of the presence of  
305 high dissolved oxygen concentrations in the activated sludge treatment process. In contrast, the  
306 relative abundance of the obligate anaerobic Bacteroidetes and the facultatively anaerobic  
307 Firmicutes decreased in the secondary effluent (Supplementary Information Table 5).  
308 Consequently, associations between Bacteroidetes or Firmicutes bacteria with ARGs were only  
309 observed in the influent samples (Fig. 2d). Proteobacteria was the dominant host phylum for  
310 ARGs across the entire wastewater treatment process (Fig. 2d), which is consistent with previous  
311 studies<sup>70-72</sup>. Our results indicate that the shift in the microbial community, and specifically the  
312 growth and decay of certain phyla drove changes in the resistome across the WWTP (Fig. 2b,c).

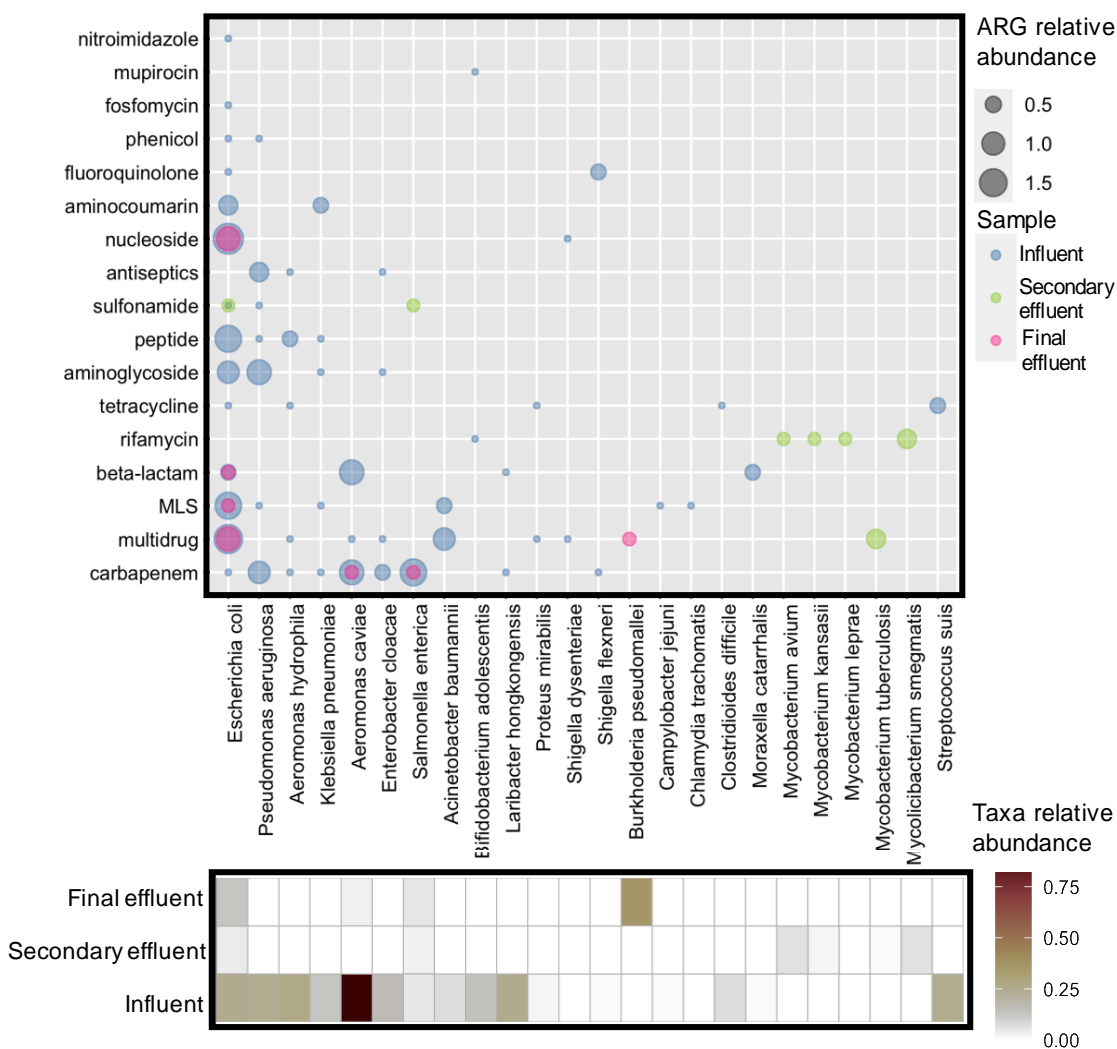
313 WWTP influent and effluent hosts were similar at the phylum level, as shown by both  
314 epicPCR and long-read sequencing (Supplementary Information Table 6 & Fig. 5), which is  
315 consistent with another study that used Nanopore sequencing for ARG host detection in WWTP  
316 influent and activated sludge<sup>42</sup>. However, at the species level, ARG hosts in the WWTP effluent  
317 were not entirely a subset of those in the WWTP influent due to the emergence of new hosts in  
318 the effluent (Supplementary Information Fig. 5). To gain a deeper understanding of the  
319 mechanisms responsible for the removal and selection of ARG hosts by different wastewater  
320 treatment unit processes, future research should focus on understanding the relative importance  
321 of horizontal gene transfer versus vertical propagation of ARGs via the growth and decay of  
322 ARG host, as well as the impact of environmental and operational variables on ARG propagation  
323 mechanisms<sup>73</sup>.

324

325 **3.4. ARGs associated with pathogens and mobile genetic elements were present in the final**  
326 **effluent**

327           We narrowed our focus to ARGs that were most likely to pose a public health risk using  
328 the following criteria: 1) presence in the final effluent, 2) association with an MGE, and 3)  
329 association with a pathogenic host species. Given the strong performance of long-read  
330 sequencing (i.e., high detection sensitivity on resistomes, hosts and MGEs), we performed this  
331 analysis using information obtained via long-read sequencing results. Notably, among all ARG-  
332 carrying reads in the effluent, 41.3% of them contained MDRs. ARGs associated with pathogens  
333 were abundant in the influent, which contained high abundances of enteric bacteria and as well  
334 as a diverse array of ARGs and ARG-carrying pathogens (Fig. 3).

335 We observed several ARG-carrying pathogens in the secondary effluent that were not detected in  
 336 the influent, such as *Mycobacterium* species carrying rifamycin-resistance genes (Fig. 3).  
 337 *Mycobacterium* is ubiquitous in wastewater and activated sludge and is considered a scavenger  
 338 of insoluble compounds in wastewater<sup>74,75</sup>. One of the detected putative *Mycobacterium*  
 339 pathogens, *Mycobacterium tuberculosis* (TB), was found to carry *efpA* which encodes an efflux  
 340 pump system capable of extruding the isoniazid to the exterior of the cell<sup>76</sup>. This is particularly  
 341 concerning because isoniazid is a drug commonly used in TB therapy. However, ARG-carrying



**Fig. 3.** ARG-carrying putative pathogens detected in the influent, secondary effluent, and final effluent. ARGs are grouped by drug class subtype on y-axis; pathogenic species are shown on the x-axis. Dot size indicates the relative abundance of ARGs. Dot color indicates the sample location. The heatmap shows the taxa relative abundance of pathogenic species in each sample.

342 *Mycobacterium* species were not found in the final effluent, indicating its effective removal in  
343 the disinfection process. *Burkholderia pseudomallei*, which can cause the disease melioidosis<sup>77</sup>,  
344 was an ARG host detected in the final effluent, but not in the influent or secondary effluent. It  
345 was associated with *MuxB*, a resistance-nodulation-cell division (RND) antibiotic efflux pump  
346 gene that significantly reduces susceptibility to macrolide, beta-lactams, and fluoroquinolones in  
347 bacteria.

348 Overall, six out of seven ARG-carrying putative pathogens present in the final effluent  
349 were also detected in the influent (Fig. 3). It is worth noting that the relative abundance of ARGs  
350 carried by *E. coli*, especially those encoding resistance against multidrug, beta-lactam, and  
351 nucleoside, were persistent across the entire treatment process (Fig. 3). In addition, the estimated  
352 relative abundance of total *E. coli* decreased significantly from influent to final effluent (Fig. 3).  
353 Together, these results suggest that the chlorination process may have selected for resistant *E.*  
354 *coli*. Previous studies have also observed that multidrug-resistant *E. coli* was persistent during  
355 wastewater treatment<sup>78</sup> and was capable of escaping the oxidation by disinfectants<sup>79</sup>.

356 A variety of ARG-associated MGEs including *IntI*s, recombinases, transposases, and  
357 integrases were frequently observed in the effluent samples (Supplementary Information Table  
358 7), which suggests they may be involved in the HGT of ARGs among bacteria<sup>80-82</sup>. Recent  
359 studies revealed the striking prevalence of insertion sequences (IS) in resistant pathogens and the  
360 relatively consistent linkages between certain IS and specific ARGs across highly diverse  
361 bacterial genotypes, indicating the role of IS in mediating the HGT of these ARGs<sup>83,84</sup>. Similarly,  
362 we also found diverse IS families that were associated with ARGs across samples. Particularly,  
363 the IS6 family transposase was found to be frequently associated with two macrolide resistance  
364 genes across samples, namely, *msrE* and *mphE* (Supplementary Information Table 7). In

365 addition, this specific *mphE/msrE*-the IS6 family transposase association was found on a  
366 conjugative plasmid equipped with the T4SS and MOBQ machinery, highlighting its potential  
367 for HGT via the interaction of IS and a conjugative plasmid.

368

#### 369 **4. Conclusions**

370 The bacterial host and genetic context of an ARG present in our water and wastewater  
371 systems is critical to assessing its potential risk to human health. Specifically, ARGs of highest  
372 priority for further study are those hosted by pathogenic bacteria and/or with the potential to be  
373 horizontally transferred to pathogens (i.e., associated with an MGE). In this study, we evaluated  
374 and compared long and short-read metagenomics as well as epicPCR for identifying ARG hosts  
375 and associations with MGEs. We found that long-read sequencing outperformed short-read  
376 sequencing by generating a higher relative abundance of ARGs, especially of ARGs associated  
377 with MGEs, as well as a more diverse ARG host profile. Moreover, long-read sequencing  
378 generally yielded a greater number of reads supporting ARG-host linkages compared to the  
379 number of contigs assembled from short reads. EpicPCR outperformed long-read sequencing in  
380 terms of the breadth of hosts detected for three ARG targets (*ermB*, *sull*, and *tetO*), however, it  
381 does not provide any additional contextual information (e.g., whether the ARG is associated with  
382 an MGE). When we applied these methods to understand ARG host dynamics across the WWTP,  
383 we observed consistent trends using long-read sequencing and epicPCR. Overall, the linkages of  
384 ARGs and host phyla in the WWTP effluent resembled those in the WWTP influent. However,  
385 at the species level, ARG hosts in the WWTP effluent were no longer a subset of those in the  
386 WWTP influent, which reinforces the need for more and longer-term surveillance of emerging



387 effluent ARG hosts, and the importance of understanding the mechanisms of removal and  
388 selection of ARG hosts across treatment.

389         These results suggest that for environmental surveillance, long-read sequencing has many  
390 advantages as a tool for ARG detection and host tracking due to its high sequencing efficiency  
391 and because it does not require assembly. However, if any clinically-relevant ARG targets, such  
392 as MCRs (colistin resistance genes), are of particular concern to public health, epicPCR assays  
393 could be developed and applied to capture a more comprehensive host profile to complement  
394 routine metagenomic screening. Future studies should focus on evaluating standardized methods  
395 for wastewater-based surveillance of antibiotic resistance, developing guidelines for better  
396 reproducibility, and establishing a risk estimation framework for ARGs in the environment.

397

#### 398 **Declaration of Competing Interest**

399 The authors declare no competing financial interest.

400

#### 401 **Acknowledgments**

402 This research was supported by funds from the National Institute of Food and Agriculture (grant  
403 no. 2016-68007-25044), the National Science Foundation (CBET 1805901 and 2029025, EF-  
404 2126387), a Johnson & Johnson WiSTEM2D 2D award, seed funds from Rice University, and a  
405 NIH grant from NIAID (P01-AI152999). We thank the operators at the West University City  
406 Place WWTP for helping with accessing wastewater samples.

407

#### 408 **Data Availability**

409 The metagenomic sequencing analysis pipeline and the epicPCR analysis pipeline are  
410 deposited at [https://gitlab.com/treangenlab/wasterwater\\_arg\\_metagenomics](https://gitlab.com/treangenlab/wasterwater_arg_metagenomics). All sequencing data  
411 can be found at NCBI SRA (project accession number: PRJNA842493).

412

## 413 References

- 414 1. Murray, C. J. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.  
415 *The Lancet* 399, 629–655 (2022).
- 416 2. McEwen, S. A. & Collignon, P. J. Antimicrobial Resistance: A One Health Perspective.  
417 *Microbiology Spectrum* 6, 6.2.10 (2018).
- 418 3. Guo, J., Li, J., Chen, H., Bond, P. L. & Yuan, Z. Metagenomic analysis reveals wastewater treatment  
419 plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Research* 123,  
420 468–478 (2017).
- 421 4. Rizzo, L. et al. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and  
422 genes spread into the environment: A review. *Science of The Total Environment* 447, 345–360  
423 (2013).
- 424 5. Kappell, A. D. et al. Removal of antibiotic resistance genes in an anaerobic membrane bioreactor  
425 treating primary clarifier effluent at 20 °C. *Environ. Sci.: Water Res. Technol.* 4, 1783–1793 (2018).
- 426 6. Lou, E. G., Harb, M., Smith, A. L. & Stadler, L. B. Livestock manure improved antibiotic resistance  
427 gene removal during co-treatment of domestic wastewater in an anaerobic membrane bioreactor.  
428 *Environ. Sci.: Water Res. Technol.* 6, 2832–2842 (2020).
- 429 7. Du, J. et al. Variation of antibiotic resistance genes in municipal wastewater treatment plant with  
430 A2O-MBR system. *Environ Sci Pollut Res* 22, 3715–3726 (2015).
- 431 8. Munir, M., Wong, K. & Xagorarakis, I. Release of antibiotic resistant bacteria and genes in the  
432 effluent and biosolids of five wastewater utilities in Michigan. *Water Research* 45, 681–693 (2011).
- 433 9. Li, N., Sheng, G. P., Lu, Y. Z., Zeng, R. J. & Yu, H. Q. Removal of antibiotic resistance genes from  
434 wastewater treatment plant effluent by coagulation. *Water Research* 111, 204–212 (2017).
- 435 10. Cheng, H. & Hong, P. Y. Removal of Antibiotic-Resistant Bacteria and Antibiotic Resistance Genes  
436 Affected by Varying Degrees of Fouling on Anaerobic Microfiltration Membranes. *Environ. Sci.*  
437 *Technol.* 51, 12200–12209 (2017).
- 438 11. Li, Z. H., Yuan, L., Gao, S. X., Wang, L. & Sheng, G. P. Mitigated membrane fouling and enhanced  
439 removal of extracellular antibiotic resistance genes from wastewater effluent via an integrated pre-  
440 coagulation and microfiltration process. *Water Research* 159, 145–152 (2019).
- 441 12. Li, H. et al. Electrochemical disinfection of secondary effluent from a wastewater treatment plant:  
442 Removal efficiency of ARGs and variation of antibiotic resistance in surviving bacteria. *Chemical*  
443 *Engineering Journal* 392, 123674 (2020).
- 444 13. Lin, W., Zhang, M., Zhang, S. & Yu, X. Can chlorination co-select antibiotic-resistance genes?  
445 *Chemosphere* 156, 412–419 (2016).
- 446 14. Mao, D. et al. Prevalence and proliferation of antibiotic resistance genes in two municipal wastewater  
447 treatment plants. *Water Research* 85, 458–466 (2015).
- 448 15. McConnell, M. M. et al. Removal of antibiotic resistance genes in two tertiary level municipal  
449 wastewater treatment plants. *Science of The Total Environment* 643, 292–300 (2018).
- 450 16. Rafrat, I. D. et al. Abundance of antibiotic resistance genes in five municipal wastewater treatment  
451 plants in the Monastir Governorate, Tunisia. *Environmental Pollution* 219, 353–358 (2016).

- 452 17. Bueno, I. et al. Role of wastewater treatment plants on environmental abundance of Antimicrobial  
453 Resistance Genes in Chilean rivers. *International Journal of Hygiene and Environmental Health* 223,  
454 56–64 (2020).
- 455 18. Mukherjee, M., Laird, E., Gentry, T. J., Brooks, J. P. & Karthikeyan, R. Increased Antimicrobial and  
456 Multidrug Resistance Downstream of Wastewater Treatment Plants in an Urban Watershed. *Front*  
457 *Microbiol* 12, 657353 (2021).
- 458 19. Yuan, Q. B., Guo, M. T. & Yang, J. Fate of Antibiotic Resistant Bacteria and Genes during  
459 Wastewater Chlorination: Implication for Antibiotic Resistance Control. *PLOS ONE* 10, e0119403  
460 (2015).
- 461 20. Kneis, D., Berendonk, T. U., Forslund, S. K. & Hess, S. Antibiotic Resistance Genes in River  
462 Biofilms: A Metagenomic Approach toward the Identification of Sources and Candidate Hosts.  
463 *Environ. Sci. Technol.* (2022) doi:10.1021/acs.est.2c00370.
- 464 21. Chu, B. T. T. et al. Metagenomics Reveals the Impact of Wastewater Treatment Plants on the  
465 Dispersal of Microorganisms and Genes in Aquatic Sediments. *Appl Environ Microbiol* 84, e02168-  
466 17, /aem/84/5/e02168-17.atom (2017).
- 467 22. Quintela-Baluja, M. et al. Spatial ecology of a wastewater network defines the antibiotic resistance  
468 genes in downstream receiving waters. *Water Research* 162, 347–357 (2019).
- 469 23. Sabri, N. A. et al. Prevalence of antibiotics and antibiotic resistance genes in a wastewater effluent-  
470 receiving river in the Netherlands. *Journal of Environmental Chemical Engineering* 8, 102245 (2020).
- 471 24. Zhang, A. N. et al. An omics-based framework for assessing the health risk of antimicrobial  
472 resistance genes. *Nat Commun* 12, 4765 (2021).
- 473 25. Zhang, Z. et al. Assessment of global health risk of antibiotic resistance genes. *Nat Commun* 13,  
474 1553 (2022).
- 475 26. Ben, Y. et al. Human health risk assessment of antibiotic resistance associated with antibiotic  
476 residues in the environment: A review. *Environmental Research* 169, 483–493 (2019).
- 477 27. Rice, E. W., Wang, P., Smith, A. L. & Stadler, L. B. Determining Hosts of Antibiotic Resistance  
478 Genes: A Review of Methodological Advances. *Environ. Sci. Technol. Lett.* 7, 282–291 (2020).
- 479 28. Che, Y. et al. Mobile antibiotic resistome in wastewater treatment plants revealed by Nanopore  
480 metagenomic sequencing. *Microbiome* 7, 44 (2019).
- 481 29. Ma, L. et al. Metagenomic Assembly Reveals Hosts of Antibiotic Resistance Genes and the Shared  
482 Resistome in Pig, Chicken, and Human Feces. *Environ. Sci. Technol.* 50, 420–427 (2016).
- 483 30. Yin, X. et al. An assessment of resistome and mobilome in wastewater treatment plants through  
484 temporal and spatial metagenomic analysis. *Water Research* 209, 117885 (2022).
- 485 31. Fuhrmeister, E. R. et al. Surveillance of potential pathogens and antibiotic resistance in wastewater  
486 and surface water from Boston, USA and Vellore, India using long-read metagenomic sequencing.  
487 2021.04.22.21255864 <https://www.medrxiv.org/content/10.1101/2021.04.22.21255864v1> (2021)  
488 doi:10.1101/2021.04.22.21255864.
- 489 32. Majeed, H. J. et al. Evaluation of Metagenomic-Enabled Antibiotic Resistance Surveillance at a  
490 Conventional Wastewater Treatment Plant. *Frontiers in Microbiology* 12, (2021).
- 491 33. Riquelme, M. V. et al. Wastewater Based Epidemiology Enabled Surveillance of Antibiotic  
492 Resistance. medRxiv 2021.06.01.21258164 (2021) doi:10.1101/2021.06.01.21258164.
- 493 34. Ju, F. et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic  
494 exchange, and upregulated expression in the effluent microbiomes. *The ISME Journal* 13, 346–360  
495 (2019).
- 496 35. Yin, X. et al. Exploration of the antibiotic resistome in a wastewater treatment plant by a nine-year  
497 longitudinal metagenomic study. *Environment International* 133, 105270 (2019).
- 498 36. Vollmers, J., Wiegand, S. & Kaster, A. K. Comparing and Evaluating Metagenome Assembly Tools  
499 from a Microbiologist’s Perspective - Not Only Size Matters! *PLOS ONE* 12, e0169662 (2017).
- 500 37. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile  
501 metagenomic assembler. *Genome Res.* 27, 824–834 (2017).

- 502 38. Brown, C. L. et al. Critical evaluation of short, long, and hybrid assembly for contextual analysis of  
503 antibiotic resistance genes in complex environmental metagenomes. *Scientific Reports* 11, 3753  
504 (2021).
- 505 39. Juraschek, K. et al. Outcome of Different Sequencing and Assembly Approaches on the Detection of  
506 Plasmids and Localization of Antimicrobial Resistance Genes in Commensal *Escherichia coli*.  
507 *Microorganisms* 9, 598 (2021).
- 508 40. Maguire, F. et al. Metagenome-assembled genome binning methods with short reads  
509 disproportionately fail for plasmids and genomic Islands. *Microbial Genomics*, 6, e000436 (2020).
- 510 41. Arango-Argoty, G. A. et al. NanoARG: a web service for detecting and contextualizing antimicrobial  
511 resistance genes from nanopore-derived metagenomes. *Microbiome* 7, 88 (2019).
- 512 42. Dai, D. et al. Long-read metagenomic sequencing reveals shifts in associations of antibiotic  
513 resistance genes with mobile genetic elements from sewage to activated sludge. *Microbiome* 10, 20  
514 (2022).
- 515 43. Liu, Z. et al. Metagenomic and metatranscriptomic analyses reveal activity and hosts of antibiotic  
516 resistance genes in activated sludge. *Environment International* 129, 208–220 (2019).
- 517 44. Zhao, R. et al. Deciphering the mobility and bacterial hosts of antibiotic resistance genes under  
518 antibiotic selection pressure by metagenomic assembly and binning approaches. *Water Research* 186,  
519 116318 (2020).
- 520 45. Spencer, S. J. et al. Massively parallel sequencing of single cells by epicPCR links functional genes  
521 with phylogenetic markers. *ISME J* 10, 427–436 (2016).
- 522 46. Leggett, R. M. et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant  
523 pathogens. *Nat Microbiol* 5, 430–442 (2020).
- 524 47. Brandt, C., Bongcam-Rudloff, E. & Müller, B. Abundance Tracking by Long-Read Nanopore  
525 Sequencing of Complex Microbial Communities in Samples from 20 Different Biogas/Wastewater  
526 Plants. *Applied Sciences* 10, 7518 (2020).
- 527 48. Govender, K. N. & Eyre, D. W. Benchmarking taxonomic classifiers with Illumina and Nanopore  
528 sequence data for clinical metagenomic diagnostic applications. 2022.01.11.475979 Preprint at  
529 <https://doi.org/10.1101/2022.01.11.475979> (2022).
- 530 49. Sevim, V. et al. Shotgun metagenome data of a defined mock community using Oxford Nanopore,  
531 PacBio and Illumina technologies. *Sci Data* 6, 285 (2019).
- 532 50. Yang, Y. et al. Rapid absolute quantification of pathogens and ARGs by nanopore sequencing.  
533 *Science of The Total Environment* 809, 152190 (2022).
- 534 51. Abera, B., Kibret, M. & Mulu, W. Extended-Spectrum beta ( $\beta$ )-Lactamases and Antibiogram in  
535 Enterobacteriaceae from Clinical and Drinking Water Sources from Bahir Dar City, Ethiopia. *PLOS*  
536 *ONE* 11, e0166519 (2016).
- 537 52. Lee, K. et al. Mobile resistome of human gut and pathogen drives anthropogenic bloom of antibiotic  
538 resistance. *Microbiome* 8, 2 (2020).
- 539 53. Ngbede, E. O. et al. Concurrent Resistance to Carbapenem and Colistin Among Enterobacteriaceae  
540 Recovered From Human and Animal Sources in Nigeria Is Associated With Multiple Genetic  
541 Mechanisms. *Front Microbiol* 12, 740348 (2021).
- 542 54. Castanheira, M., Simner, P. J. & Bradford, P. A. Extended-spectrum  $\beta$ -lactamases: an update on their  
543 characteristics, epidemiology and detection. *JAC-Antimicrobial Resistance* 3, dlab092 (2021).
- 544 55. Li, L. et al. Extended-Spectrum  $\beta$ -Lactamase and Carbapenemase Genes are Substantially and  
545 Sequentially Reduced during Conveyance and Treatment of Urban Sewage. *Environ. Sci. Technol.*  
546 55, 5939–5949 (2021).
- 547 56. Søråas, A., Sundsfjord, A., Sandven, I., Brunborg, C. & Jenum, P. A. Risk Factors for Community-  
548 Acquired Urinary Tract Infections Caused by ESBL-Producing Enterobacteriaceae –A Case–Control  
549 Study in a Low Prevalence Country. *PLOS ONE* 8, e69581 (2013).
- 550 57. Jia, S. et al. Fate of antibiotic resistance genes and their associations with bacterial community in  
551 livestock breeding wastewater and its receiving river water. *Water Research* 124, 259–268 (2017).

- 552 58. Li, B., Ju, F., Cai, L. & Zhang, T. Profile and Fate of Bacterial Pathogens in Sewage Treatment  
553 Plants Revealed by High-Throughput Metagenomic Approach. *Environ. Sci. Technol.* 49, 10492–  
554 10502 (2015).
- 555 59. Deshpande, A. S. & Fahrenfeld, N. L. Abundance, diversity, and host assignment of total,  
556 intracellular, and extracellular antibiotic resistance genes in riverbed sediments. *Water Research* 217,  
557 118363 (2022).
- 558 60. Rozwandowicz, M. et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae.  
559 *Journal of Antimicrobial Chemotherapy* 73, 1121–1137 (2018).
- 560 61. Fondi, M. et al. Exploring the evolutionary dynamics of plasmids: the *Acinetobacter* pan-  
561 plasmidome. *BMC Evol Biol* 10, 59 (2010).
- 562 62. Orlek, A. et al. Plasmid Classification in an Era of Whole-Genome Sequencing: Application in  
563 Studies of Antibiotic Resistance Epidemiology. *Frontiers in Microbiology* 8, (2017).
- 564 63. Tazzyman, S. J. & Bonhoeffer, S. Why There Are No Essential Genes on Plasmids. *Molecular*  
565 *Biology and Evolution* 32, 3079–3088 (2015).
- 566 64. Yang, Y., Li, B., Zou, S., Fang, H. H. P. & Zhang, T. Fate of antibiotic resistance genes in sewage  
567 treatment plant revealed by metagenomic approach. *Water Research* 62, 97–106 (2014).
- 568 65. Cheng, X., Xu, J., Smith, G. & Zhang, Y. Metagenomic insights into dissemination of antibiotic  
569 resistance across bacterial genera in wastewater treatment. *Chemosphere* 271, 129563 (2021).
- 570 66. Liu, S. S. et al. Chlorine disinfection increases both intracellular and extracellular antibiotic  
571 resistance genes in a full-scale wastewater treatment plant. *Water Research* 136, 131–136 (2018).
- 572 67. Yuan, Q. B. et al. Redistribution of intracellular and extracellular free & adsorbed antibiotic  
573 resistance genes through a wastewater treatment plant by an enhanced extracellular DNA extraction  
574 method with magnetic beads. *Environment International* 131, 104986 (2019).
- 575 68. Wang, Y., Batra, A., Schulenburg, H. & Dagan, T. Gene sharing among plasmids and chromosomes  
576 reveals barriers for antibiotic resistance gene transfer. *Philosophical Transactions of the Royal*  
577 *Society B: Biological Sciences* 377, 20200467 (2022).
- 578 69. Barka, E. A. et al. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol*  
579 *Biol Rev* 80, 1–43 (2015).
- 580 70. Azli, B. et al. Metagenomics Insights Into the Microbial Diversity and Microbiome Network Analysis  
581 on the Heterogeneity of Influent to Effluent Water. *Frontiers in Microbiology* 13, (2022).
- 582 71. Gu, Q. et al. Characteristics of Antibiotic Resistance Genes and Antibiotic-Resistant Bacteria in Full-  
583 Scale Drinking Water Treatment System Using Metagenomics and Culturing. *Frontiers in*  
584 *Microbiology* 12, (2022).
- 585 72. Liu, Z. et al. Bacterial Hosts and Genetic Characteristics of Antibiotic Resistance Genes in  
586 Wastewater Treatment Plants of Xinjiang (China) Revealed by Metagenomics. *Applied Sciences* 12,  
587 3100 (2022).
- 588 73. Barancheshme, F. & Munir, M. Strategies to Combat Antibiotic Resistance in the Wastewater  
589 Treatment Plants. *Front. Microbiol.* 8, (2018).
- 590 74. Guo, F. et al. Mycobacterial species and their contribution to cholesterol degradation in wastewater  
591 treatment plants. *Sci Rep* 9, 836 (2019).
- 592 75. Radomski, N. et al. Mycobacterium Behavior in Wastewater Treatment Plant, A Bacterial Model  
593 Distinct From *Escherichia coli* and Enterococci. *Environ. Sci. Technol.* 45, 5380–5386 (2011).
- 594 76. Rodrigues, L., Machado, D., Couto, I., Amaral, L. & Viveiros, M. Contribution of efflux activity to  
595 isoniazid resistance in the Mycobacterium tuberculosis complex. *Infection, Genetics and Evolution*  
596 12, 695–700 (2012).
- 597 77. Draper, A. D. K. et al. Association of the Melioidosis Agent *Burkholderia pseudomallei* with Water  
598 Parameters in Rural Water Supplies in Northern Australia. *Applied and Environmental Microbiology*  
599 76, 5305–5307 (2010).
- 600 78. Aslan, A., Cole, Z., Bhattacharya, A. & Oyibo, O. Presence of Antibiotic-Resistant *Escherichia coli*  
601 in Wastewater Treatment Plant Effluents Utilized as Water Reuse for Irrigation. *Water* 10, 805  
602 (2018).

- 603 79. Mounaouer, B. & Abdennaceur, H. Modeling and kinetic characterization of wastewater disinfection  
604 using chlorine and UV irradiation. *Environ Sci Pollut Res* 23, 19861–19875 (2016).
- 605 80. Gillings, M. R. et al. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution.  
606 *The ISME Journal* 9, 1269–1279 (2015).
- 607 81. Knapp, C. W. et al. Indirect Evidence of Transposon-Mediated Selection of Antibiotic Resistance  
608 Genes in Aquatic Systems at Low-Level Oxytetracycline Exposures. *Environ. Sci. Technol.* 42,  
609 5348–5353 (2008).
- 610 82. Ma, L., Li, A. D., Yin, X. L. & Zhang, T. The Prevalence of Integrons as the Carrier of Antibiotic  
611 Resistance Genes in Natural and Man-Made Environments. *Environ. Sci. Technol.* 51, 5721–5728  
612 (2017).
- 613 83. Che, Y. et al. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer  
614 of antimicrobial resistance genes. *Proceedings of the National Academy of Sciences* 118,  
615 e2008731118 (2021).
- 616 84. Che, Y. et al. High-resolution genomic surveillance elucidates a multilayered hierarchical transfer of  
617 resistance between WWTP- and human/animal-associated bacteria. *Microbiome* 10, 16 (2022).
- 618 85. Spencer, S. et al. epicPCR (Emulsion, Paired Isolation, and Concatenation PCR). *Protocol Exchange*  
619 (2015) doi:10.1038/protex.2015.094.
- 620 86. Stalder, T., Press, M. O., Sullivan, S., Liachko, I. & Top, E. M. Linking the resistome and  
621 plasmidome to the microbiome. *The ISME Journal* 13, 2437–2446 (2019).
- 622 87. Wei, Z. et al. High-Throughput Single-Cell Technology Reveals the Contribution of Horizontal Gene  
623 Transfer to Typical Antibiotic Resistance Gene Dissemination in Wastewater Treatment Plants.  
624 *Environ. Sci. Technol.* (2021) doi:10.1021/acs.est.1c01250.
- 625 88. De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing  
626 and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669 (2018).
- 627 89. Curry, K. D. et al. Emu: species-level microbial community profiling of full-length 16S rRNA  
628 Oxford Nanopore sequencing data. *Nat Methods* 19, 845–853 (2022).
- 629 90. Yin, X. et al. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for  
630 enhancement characterization and quantification of antibiotic resistance genes in environmental  
631 metagenomes. *Bioinformatics* 34, 2263–2270 (2018).
- 632 91. Alcock, B. P. et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic  
633 resistance database. *Nucleic Acids Res* 48, D517–D525 (2020).
- 634 92. Krawczyk, P. S., Lipinski, L. & Dziembowski, A. PlasFlow: predicting plasmid sequences in  
635 metagenomic data using genome signatures. *Nucleic Acids Res* 46, e35 (2018).
- 636 93. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: rapid and sensitive classification  
637 of metagenomic sequences. *Genome Res.* 26, 1721–1729 (2016).
- 638 94. WHO. WHO publishes list of bacteria for which new antibiotics are urgently needed.  
639 [https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-](https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed)  
640 [antibiotics-are-urgently-needed](https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed) (2017).
- 641 95. Robertson, J. & Nash, J. H. E. Y. MOB-suite: software tools for clustering, reconstruction and typing  
642 of plasmids from draft assemblies. *Microbial Genomics* 4, e000206.
- 643