

1 **Metatranscriptomic analysis of wastewater sites reveals a high abundance of antimicrobial**  
2 **resistance genes from hospital wastewater.**

3

4 **Sebastian Musundi<sup>1</sup>, Samuel Ng'ang'a<sup>1</sup>, Esther Wangui Wanjiru<sup>1</sup>, Kate Sagoe<sup>1</sup>, Ernest**  
5 **Wandera<sup>1 2</sup>, Bernard Kanoi<sup>1</sup>, Jesse Gitaka<sup>1\*</sup>**

6

7 <sup>1</sup> Center for Malaria Elimination, Institute of Tropical Medicine, Mount Kenya University, Thika  
8 Kenya

9 <sup>2</sup> KEMRI/Nagasaki University, Institute of Tropical Medicine, Kenya Research Station, P.O. Box  
10 19993-00202, Nairobi, Kenya

11

12 **\*Correspondence:**

13 Corresponding Author

14 [jgitaka@mku.ac.ke](mailto:jgitaka@mku.ac.ke)

15

16

17

18

19

20

21

22

23

24

25

## 26 **Abstract**

### 27 **Objective**

28 This study analyzed the metatranscriptome of wastewater samples from different sites with a focus of  
29 identifying antibiotic resistance genes and the bacterial community.

### 30 **Methods**

31 Twenty-four wastewater samples were collected from a hospital, university sewer, and the influent and  
32 effluent of a wastewater treatment plant (WWTP). The metatranscriptome was sequenced to identify  
33 antimicrobial resistance genes (ARGs), mobile genetic elements, and bacterial community  
34 composition.

### 35 **Results**

36 Metatranscriptome analysis revealed varying abundances of ARG transcripts across different sites,  
37 with 84, 27, 14, and 9 ARG transcripts identified in wastewater collected from the hospital, university,  
38 influent, and effluent of a WWTP, respectively. Notably, hospital wastewater contained clinically  
39 relevant beta-lactam ARGs, including *bla-NDM-1* and several *blaOXA* transcripts. Four core ARGs,  
40 against sulfonamides *sul1* and *sul2* and aminoglycosides *aph(6)-Ia* and *aph('3')-Ib* were also identified.  
41 The predominant bacterial community comprised of *Gammaproteobacteria*, with priority pathogens  
42 like *Neisseria gonorrhoea* and *Helicobacter pylori* present in hospital wastewater and WWTP influent.

### 43 **Conclusion**

44 These findings provide insights into the wastewater resistome and how meta-transcriptomic data can  
45 be utilized for the surveillance of antibiotic resistance. Overall, our study highlights the utility of  
46 wastewater surveillance in understanding and addressing the dissemination of antibiotic resistance and  
47 emphasizes the crucial role of proper wastewater management in protecting public and environmental  
48 health.

49 **Keywords:** wastewater surveillance, antibiotic resistance genes, hospital wastewater, wastewater  
50 treatment plants, metatranscriptomics

## 51 **1. Introduction**

52 Antimicrobial resistance (AMR) remains one of the major global health challenges of the 21<sup>st</sup> century  
53 and has been prioritized by the World Health Organization as one of the top 10 threats to humanity  
54 worldwide. In 2019, drug-resistant infections accounted for approximately 4.95 million deaths globally  
55 and disproportionately affected low- and middle-income countries in sub-Saharan Africa [1]. To tackle  
56 the threat posed by AMR effectively, continuous surveillance, which is critical for understanding trends  
57 in pathogen incidence and monitoring the effectiveness of various interventions, needs to be carried  
58 out [2]. Large networks sharing AMR data at the national, regional, and global levels, including the  
59 Global Antimicrobial Resistance Surveillance System (GLASS), Central Asian and European  
60 Surveillance of Antimicrobial Resistance (CAESAR), and European Antimicrobial Resistance  
61 Surveillance Network (EARS-Net), have been established [3]. However, these bodies primarily focus  
62 on individual-level sampling with selection bias toward antibiotic-resistant bacteria (ARB) in clinical  
63 settings and do not represent AMR within the population [4]. Furthermore, most laboratory settings  
64 within low- and middle-income countries (LMICs) lack the infrastructure for patient-level surveillance;  
65 therefore, data from these countries are underrepresented despite the region facing the greatest threat  
66 posed by AMR.

67 Wastewater surveillance (WWS) is a cost-effective approach for monitoring the presence of ARB and  
68 ARGs at the population level, as it involves the biomass of an entire population of a community [5].  
69 The success of using WWS was demonstrated during the SARS-CoV-2 pandemic, when it was used  
70 to provide real-time information about the circulating strains of the virus [6]. Currently, the use of  
71 WWS is quickly gaining traction in AMR surveillance since it provides insights into current and  
72 previous antibiotic use, real-time information about circulating ARB and an early warning system for  
73 the emergence of multidrug-resistant pathogens in the community [7]. Furthermore, information  
74 obtained from WWS can be used to evaluate the efficacy of various methods of treating wastewater  
75 before release into the environment. For example, previous research has shown that sewage treatment

76 does not eradicate ARG or antibiotics, which can enter the food chain using sewage sludge and water  
77 in agriculture [8,9]. WWS provides a suitable approach for detecting ARB and ARGs and can  
78 complement other existing surveillance methods, especially within LMICs.

79 Surveillance sites for assessing the presence of ARGs and ARB include areas where antibiotics are  
80 primarily used, such as in clinical or hospital settings and wastewater treatment plants where  
81 disinfection occurs. Wastewater treatment plants (WWTPs) are hot spots for AMR since antibiotics  
82 from different sources, including those originating from healthcare settings, agriculture, and industry,  
83 converge [10]. Failure to remove ARGs and ARB from the effluent of WWTPs results in their release  
84 and spread to the environment horizontally through mobile genetic elements [11]. Apart from WWTPs,  
85 hospital settings have the highest use of antibiotics due to their heavy use, and their effluent is expected  
86 to harbor high numbers of ARB and ARGs [12]. One common unexplored area for antibiotic use is  
87 within residential or institutional settings, especially in low- and middle-income countries where the  
88 prescription of antibiotics is not strictly regulated. Considering that antibiotic use is common due to  
89 self-medication and overall misuse, evaluating wastewater that originates not only from common hot  
90 spots such as WWTPs and hospital settings but also from residential or institutional settings may  
91 provide insights into the spread of ARGs.

92 Recent advances in high-throughput sequencing technologies have provided a better understanding of  
93 the wastewater resistome, allowing for the rapid identification of ARGs and microbial communities.  
94 One such approach is the analysis of the wastewater metatranscriptome, which provides information  
95 about the active bacterial community and ARGs by quantifying the messenger RNA of specific genes  
96 [13]. Therefore, through meta-transcriptomic analysis of wastewater samples, we can determine the  
97 abundance and expression patterns of specific ARGs within various settings and shed light on the  
98 measures needed to control ARGs from wastewater sources. In this study, we identified ARGs across  
99 four different sites, including a hospital, university wastewater and the influent and effluent of a  
100 WWTP. Our findings revealed many ARGs in hospital effluent and from universities and fewer ARGs

101 in the influent and effluent of WWTPs. Furthermore, we identified clinically relevant ARGs from  
102 hospital wastewater and identified mobile genetic elements and virulence factors that play a role in the  
103 horizontal transfer of ARGs.

## 104 **2. Materials and Methods**

### 105 **2.1 Sample collection**

106 Twenty-four wastewater samples were collected from the hospital, university hostels, influent, and  
107 effluent of a WWTP using Moore swabs, created by cutting cotton gauze into pieces of approximately  
108 90 cm long and 30 cm wide. These cut pieces were folded three times, firmly tied at the center with a  
109 nylon string, placed in the inflow stream of wastewater, and fastened with the nylon string by tying  
110 them to a hook. After 24 hours, Moore swabs were retrieved, placed in Ziploc bags, transported to the  
111 laboratory on ice in a cooler box at -4°C, and stored in a -80°C freezer until processing.

112

### 113 **2.2 RNA extraction and sequencing**

114 The Moore swabs were processed by thawing overnight at +4°C after adding 50 mL of 1X PBS. Once  
115 thawed, the swab was squeezed into a beaker, and the liquid was transferred into two 50 mL Falcon  
116 tubes. Fifteen milliliters of the squeezed-out raw wastewater was centrifuged at 10,000 × g for 10  
117 minutes. Ten milliliters of clear supernatant was retrieved and subjected to total nucleic acid extraction  
118 using the MagMAX™ Wastewater Ultra Nucleic Acid Isolation Kit with Virus Enrichment (A52610;  
119 Thermo Fisher Scientific™, Massachusetts, U.S.), following the manufacturer's instructions. The  
120 extracted nucleic acids were eluted in 50 µL of elution buffer and stored at -80°C until further  
121 processing. The RNA integrity and quality were assessed using an Invitrogen Qubit 4 Fluorometer  
122 (Thermo Fisher Scientific, Inc.). The RNA was subjected to Illumina MiSeq next-generation  
123 sequencing as described previously [14]. A 200 bp fragment library ligated with barcoded adapters was  
124 constructed using NEBNext Ultra RNA Library Prep Kit for Illumina v1.2 (New England Biolabs)  
125 according to the manufacturer's instructions. The cDNA library was purified using Agencourt AMPure

126 XP magnetic beads (Beckman Coulter). After assessing the quality and quantity of the purified cDNA  
127 library, nucleotide sequencing was performed on an Illumina MiSeq sequencer (Illumina) using a  
128 MiSeq Reagent Kit v3 (Illumina).

129

### 130 **2.3 Reads processing**

131 Paired-end reads generated from RNA sequencing were demultiplexed and retrieved from the MiSeq  
132 platform. The reads were processed using an in-house Nextflow pipeline  
133 (<https://github.com/sebymusundi/metatranscriptomics>). Briefly, the read quality was assessed using  
134 FastQC, adapter sequences were removed using FastP [15,16], and low-quality reads were removed  
135 using Trimmomatic (SLIDING WINDOW = 4:15, MINLEN = 30) [17]. The clean reads were  
136 merged using PEAR, mapped to the human reference genome Hg38 using Bowtie2, and unmapped  
137 reads were filtered out using Samtools [18]. SortmeRNA was subsequently used to remove ribosomal  
138 RNA (rRNA) from the remaining reads using default parameters against several SILVA rRNA  
139 databases: silva-euk-18s-id95.fasta, silva-bac-23s-id98.fasta, silva-bac-16s-id90.fasta, silva-arc-23s-  
140 id98.fasta, silva-arc-16s-id95.fasta, rfam-5s-database-id98.fasta, rfam-5.8s-database-id98.fasta, and  
141 silva-euk-28s-id98.fasta [19].

142

### 143 **2.4 De novo assembly and Read Quantification.**

144 The non-ribosomal reads were coassembled de novo based on their collection site using the default  
145 settings in Trinity [20]. The transcript files were indexed and mapped to individual samples using  
146 Bowtie2. The generated SAM file was converted to a BAM and subsequently sorted and indexed.  
147 Samtools idxstats was used to determine the number of mapped and unmapped reads and the length of  
148 each transcript. Relative gene abundance was calculated using a similar approach to reads per kilobase  
149 per million reads (RPKM). Briefly, relative abundance was calculated as  
150  $RNum\_Gi = \frac{Num\_Gi}{\sum_{i=1}^n Num\_Gi}$ , where Num\_Gi represents the proportion of the number of

151 mapped reads for each specific gene or transcript ( $G_i$ ) relative to the total number of mapped reads  
152 across all genes or transcripts in the sample ( $\sum_{i=1}^n(\text{Num}_{G_i})$ ) as previously described (Hu et al.,  
153 2013).

154

## 155 **2.5 Identification of antibiotic resistance genes**

156 The presence of ARGs was predicted by the Resfinder database [22]. A transcript was considered an  
157 ARG if the similarity and query coverage were greater than 80%. Transcripts associated with ARGs  
158 were retrieved, and a plot comparing the relative abundance of each class and type of ARG was  
159 generated in RNum\_Gi.

160

## 161 **2.6 Identification of mobile genetic elements and virulence genes**

162 The transcript files were further examined for the presence of different mobile genetic elements. The  
163 presence of plasmids and insertion sequences was predicted using the mobile genetic element finder  
164 version 1.0.3 [23]. Virulence factors were predicted in Abricate using the latest updated version of the  
165 VFDB database, with a percentage similarity and coverage of 80% [24].

166

## 167 **2.7 Taxonomy classification**

168 The Kraken2 database was downloaded to classify transcripts based on their taxonomy [25]. Kraken2  
169 was run on the transcripts using the standard database released in March 2023 ([https://genome-  
170 idx.s3.amazonaws.com/kraken/k2\\_standard\\_16gb\\_20231009.tar.gz](https://genome-idx.s3.amazonaws.com/kraken/k2_standard_16gb_20231009.tar.gz)). The WHO list of priority  
171 pathogens was retrieved from the generated Kraken report.

172

# 173 **3. Results**

## 174 **3.1 Identification of Antibiotic Resistance Genes**

175 All major ARGs against the antibiotic class aminoglycosides, beta-lactams, macrolides, macrolides,  
176 streptogramins, nitroimidazoles, phenicols, rifampicin, sulfonamides, tetracyclines, trimethoprim, and  
177 quinolones were detected in the hospital wastewater (Figure 1). The university wastewater had ARGs  
178 against tetracycline, sulfonamides, macrolides, beta-lactams, and aminoglycosides. The influent and  
179 effluent from the WWTP contained fewer classes of ARGs than wastewater from the hospital and  
180 university. ARGs against aminoglycosides, macrolides, sulfonamides, and tetracyclines, and  
181 aminoglycosides, macrolides, and sulfonamides, were present in the influent and effluent of the WWTP  
182 respectively (Figure 1). ARGs against tetracycline and macrolide genes were the most abundant from  
183 the referral hospital effluent, while ARGs against sulfonamides were predominant in the university  
184 wastewater.

185  
186 We detected 84, 27, 14, and 9 ARG transcripts in the wastewater from the hospital, university, WWTP  
187 influent and effluent, respectively. A high number of transcript copies were identified for *sul2*, *sul1*,  
188 *msr(E)*, *aph(3)-lb*, *aph(6)-ld*, *aac(6')-lb-Hangzhou*, *mph(E)*, *tet(Q)*, *tet(W)*, and *tet(39)*(Figure 2). The  
189 core antibiotic genes identified across all the sites included those against the sulfonamides *sul1* and  
190 *sul2* and the aminoglycosides *aph(6)-ld* and *aph(3)-lb* (Figure 2). ARG genes that were present at 3/4  
191 of the sites included *ant(2)-la*, *msr(E)*, and *tet39*. ARGs at 2/4 of the sites were found primarily in the  
192 referral hospital and university wastewater. The rare ARGs or those at individual sites included  
193 *blaCARB-4* and *aadA17* in university wastewater. The ARGs exclusively found in the hospital  
194 wastewater included *tet(G)*, *tet(Q)*, *tet(40)*, *qnrs2*, *nimE*, *mph(G)*, *floR*, *erm(T)*, *erm(A)*, *dfrG*, *dfrA1*,  
195 *catB3*, *blaOXA-97*, *blaNDM-1*, *blaGES-2*, *blaCMY-4*, *blaCARB-2*, *blaCARB-4*, *ARR-3*, *ARR-6* and  
196 *aadA1* (Figure 2). No rare ARG genes were present in the influent or effluent of the WWTP. We also  
197 identified several clinically relevant antibiotic genes in addition to the core antibiotics identified across  
198 the different wastewater sites. For example, hospital wastewater contains several relevant clinical beta-  
199 lactamase genes, such as *blaOXA* (*blaOXA-10*, *blaOXA-58*, *blaOXA-97*), *blaNDM-1*, *blaCMY-4*,



200 *blaGES-2* and *blaCARB-2*, while university effluent contains *blaOXA-10* and *blaCARB-4* (Figure 2).  
201 The tetracycline resistance genes *tet(A)*, *tet(M)*, and *tet(W)* were also found predominantly in the  
202 university effluent and referral hospitals (Figure 2).

203  
204 Analysis of the transcript abundances from wastewater collected from the referral showed high levels  
205 of *tet(G)*, *msr(E)*, and *mph(E)* and moderate expression of other ARGs associated primarily with  
206 tetracycline resistance (*tet(W)*, *tet(M)*, *tet (39)*, *tet (36)*), and beta-lactams (*cfxA5*, *cfxA4*, and *cfxA3*).  
207 Low abundances of the beta-lactam *blaOXA* genes (*blaOXA-93*, *blaOXA-58*, *blaOXA-10*), *blaNDM-1*,  
208 and *blaGES-2*, as well as *Toprj1*, *Toprj2*, and *Toprj3*, were also observed. Wastewater from the  
209 university had higher transcript levels of *sul1* and *sul2* ARGs (Figure 3).

210

### 211 **3.2 Identification of Mobile Genetic Elements and Virulence Factors**

212 Plasmids and virulence factors were detected in wastewater from the hospital and university. The  
213 plasmids detected from the transcripts from the referral hospital included *Col(MG828)*, *IncQ1*,  
214 *Col(pHAD28)*, *IncW*, and *IncP1*, while those from the university wastewater included *ColRNAI*,  
215 *Col156*, *Col(pHAD28)*, *IncFIB(AP001918)* and *Col(pHAD28)*. In addition to plasmids, insertion  
216 sequences were also identified in hospital and university wastewater, with some being located in the  
217 same transcript as ARGs. The plasmid transcript *IncQ1* present in hospital wastewater was the same  
218 as the ARG *blaCMY-4* (Table 1). Insertion sequences within the same transcripts as ARGs were  
219 predominantly found in the hospital and wastewater samples (Table 1). Most ARGs in hospital  
220 wastewater were found in IS6, which was resistant to sulfonamides, aminoglycosides, and macrolides.

221

222 Virulence factor transcripts detected from the hospital wastewater were associated with adherence  
223 (type IV pilus twitching motility protein *PilT*, twitching motility protein *PilT*, type IV pilus biogenesis  
224 protein *PilP*, and twitching motility response regulator *PilG*), motility (chemotaxis protein *CheY*,

225 flagellar export protein *FliJ*), immune modulation (outer membrane protein *OmpA*), and effector  
226 delivery systems (general secretion pathway protein G and general secretion pathway protein I). The  
227 virulence transcripts detected in the university wastewater were primarily associated with motility  
228 (flagellar basal body L-ring protein, flagellar biosynthesis protein FliO, flagellar motor switch protein  
229 FliN, and chemotaxis-specific methylesterase) and adherence (pili biogenesis protein MshG GspF-  
230 like).

231

### 232 **3.3 Taxonomy classification**

233 We also identified bacterial taxa across different wastewater sites. Across all four  
234 sites, *Gammaproteobacteria* was the predominant class of bacteria present. However, the hospital  
235 wastewater also contained *Bacteroidota* and *Bacillota*, while *Bacteroidota* was a major class in the  
236 WWTP influent (Figure 4A). At the genus level, hospital wastewater was primarily composed of  
237 *Acinetobacter*, *Aeromonas*, *Bacteroidales*, *Klebsiella*, and *Uruburuella*, while university wastewater  
238 contained *Aeromonas*, *Acinetobacter*, *Bacteroidales*, *Klebsiella*, and *Flavobacteriales*. *Aeromonas* and  
239 *Carnobacteriaceae* were the predominant genera detected in the WWTP influent, while *Acinetobacter*  
240 and *Aeromonas* composed a large proportion of the WWTP effluent (Figure 4B). We further sought to  
241 identify WHO-priority pathogens from the different wastewater treatment sites. *E. coli* made up the  
242 largest portion of WHO priority pathogens from wastewater present in the hospital, university and  
243 WWTP influent, while *Acinetobacter baumannii* had the highest proportion in the WWTP effluent  
244 (Figure 4C). Other WHO priority pathogens identified included *Helicobacter pylori*, *Neisseria*  
245 *gonorrhoeae*, *Enterococcus* spp., and *Pseudomonas aeruginosa* (Figure 4C).

246

## 247 **4. Discussion**

248 Wastewater surveillance is a valuable method for monitoring the presence of ARB and the overall  
249 resistome. Our current study sought to identify ARGs from various wastewater samples. Tracking the

250 spread of ARB and ARGs from very high-risk areas, such as wastewater from hospitals, universities  
251 and the influent and effluent of WWTPs, is critical for obtaining information used to track the spread  
252 of ARGs to the environment.

253

254 Our results revealed a greater number of ARGs in referral hospital wastewater than in the university,  
255 influent, or effluent of WWTPs. These large quantities of ARGs from hospital wastewater could be  
256 attributed to the frequent use of antibiotics within hospital settings and higher proportions of resistant  
257 bacteria and antimicrobial resistance genes [26,27]. Hospital wastewater contains a large abundance of  
258 ARGs and remains a hotspot for disseminating antimicrobial resistance genes to the environment by  
259 favoring the growth of bacteria harboring resistance genes via mobile genetic elements. These bacterial  
260 organisms survive and thrive in environments under selective pressure and can potentially transfer  
261 ARGs to other bacterial organisms that could be pathogenic to humans. Similarly, the high number of  
262 ARGs present in university settings could be linked to the misuse of antibiotics outside the hospital  
263 setting. The low number of ARGs in the influent and effluent of WWTPs could be attributed to the  
264 dilution of ARGs. Notably, the presence of various ARGs against sulfonamides, aminoglycosides, and  
265 macrolides suggested that these ARGs were not eliminated during the treatment process, as previously  
266 reported in other studies [8,10]. Our results confirm the need to continuously assess methods used to  
267 treat wastewater before it is ultimately released into the environment, considering that treated  
268 wastewater is released for use via various channels. Otherwise, there is a potential risk of introducing  
269 these ARGs into normal food chains through irrigation, further exacerbating the threat that AMR poses.

270

271 Our analysis identified a set of ubiquitous core ARGs across the different wastewater treatment sites.  
272 Sulfonamide ARGs have been detected in wastewater effluents at varying concentrations independent  
273 of the environmental conditions [28,29]. The identified core ARGs have also been previously reported  
274 when assessing the global resistosome [30]. The presence of these core ARGs, such as *sul1*, *sul2*,

275 *aph(6)-ld*, and *aph('3')-lb*, could be linked to prolonged clinical use of sulfonamides and  
276 aminoglycosides both within and outside clinical settings. Multiple studies have also reported the most  
277 abundant ARG gene identified, *tet(G)*, from hospital wastewater together with other tetracycline genes,  
278 attributing this difference to the overuse or misuse of antibiotics [31,32]. These core antibiotics could  
279 be used to monitor the spread of ARGs across different sites before they are released into the  
280 environment.

281

282 Several clinically relevant ARGs, including *blaNDM-1*, *blaOXA-10*, and *bla-OXA58*, were also  
283 detected in the hospital effluent. The presence of these clinically relevant bacteria in wastewater  
284 collected from hospitals may provide information about the increase in resistance to beta-lactam  
285 antibiotics. Moreover, these findings could lead to the use of an early warning system for the emergence  
286 or spread of multidrug-resistant strains, especially if similar ARGs are identified in clinical isolates.  
287 *blaNDM-1* shows resistance to multiple different classes of antibiotics, including aminoglycosides,  
288 beta-lactams, and fluoroquinolones, increasing its resistance to multiple agents [33]. The presence of  
289 *blaNDM-1* in hospital wastewater has also been shown in other studies [34,35].

290

291 The horizontal transfer of ARGs is a common dissemination method for antimicrobial resistance. Our  
292 analysis revealed the presence of multiple insertion sequences in the same transcripts as ARGs.  
293 Insertion sequences are mobile genetic elements present in bacterial genomes and contribute to AMR  
294 by relocating resistance genes to different locations in the genome and promoting their transfer between  
295 bacteria. IS6 family insertion sequences, including IS26 and IS6100, were identified from hospital  
296 wastewater. These insertion sequences are commonly found in gram-negative bacteria in association  
297 with ARGs and could constitute one of the methods by which ARGs are spread to the environment  
298 from wastewater [36]. These findings are consistent with previous reports on hospital wastewater,  
299 which also linked the potential role of IS6 family insertions in facilitating and expressing ARGs

300 [37,38]. Thus, the insertion sequences reported in this study could play a role in the horizontal gene  
301 transfer of ARGs against sulfonamide, aminoglycoside, and macrolide genes, while the identified  
302 plasmid could aid in the transfer of a beta-lactam ARG. In addition to mobile genetic elements, hospital  
303 and university wastewater contains virulence factors associated with adherence, immune modulation,  
304 and motility, which assist bacteria in competing for existing resources and surviving in their current  
305 environment.

306

307 Taxonomic classification allows the identification of various bacterial populations in wastewater sites.  
308 Some identified genera, such as *Aeromonas*, are part of the bacterial population predominantly found  
309 in wastewater [39]. Apart from the normal bacterial populations present in wastewater, we further  
310 identified WHO priority pathogens, including *N. gonorrhoeae*, *E. coli*, *P. aeruginosa*, *A. baumannii*,  
311 and *H. pylori*, from various wastewater sites. The presence of some of these WHO-priority bacterial  
312 strains could further spread ARGs, which are potentially harmful to human and animal life. Although  
313 short reads do not completely reveal which ARG transcripts are linked to specific bacterial genomes,  
314 recent advances in long-read sequencing are expected to fill this gap.

315 Meta-transcriptomic analysis offers a promising approach for assessing the local resistome, especially  
316 from various wastewater sites, and examining whether WWTP treatment fully eliminates ARGs before  
317 they are released into the environment. Our results identified four core antibiotic genes, *sul1*, *sul2*,  
318 *aph(6)-Ia*, and *aph(3')-Ib*, which could be used as targets for monitoring antibiotic resistance.  
319 Furthermore, we identified several clinically relevant beta-lactam resistance genes associated with  
320 multidrug resistance in hospital wastewater. Finally, we identified several ARGs in WWTP effluent,  
321 suggesting the need to improve the treatment process, especially in eliminating ARGs and other  
322 potential pathogenic bacteria. Our results highlight the potential role of using wastewater in monitoring  
323 the spread of ARGs to the environment, which could be adopted for AMR surveillance, especially  
324 within LMICs.

325 **Acknowledgment**

326 We acknowledge staff from Mount Kenya University, Thika Level 5 referral hospital and the Thika

327 Wastewater Treatment Plant for their insights into this work.

## 328 **References**

- 329 [1] Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global  
330 burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet*  
331 2022;399:629–55. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0).
- 332 [2] Masterton R. The importance and future of antimicrobial surveillance studies. *Clin Infect Dis*  
333 2008;47. <https://doi.org/10.1086/590063>.
- 334 [3] Tacconelli E, Sifakis F, Harbarth S, Schrijver R, van Mourik M, Voss A, et al. Surveillance for  
335 control of antimicrobial resistance. *Lancet Infect Dis* 2018;18:e99–106.  
336 [https://doi.org/10.1016/S1473-3099\(17\)30485-1](https://doi.org/10.1016/S1473-3099(17)30485-1).
- 337 [4] Aarestrup FM, Woolhouse MEJ. Using sewage for surveillance of antimicrobial resistance.  
338 *Science* (80- ) 2020;367:630–2. <https://doi.org/10.1126/science.aba3432>.
- 339 [5] Newton RJ, McLellan SL, Dila DK, Vineis JH, Morrison HG, Murat Eren A, et al. Sewage  
340 reflects the microbiomes of human populations. *MBio* 2015;6.  
341 <https://doi.org/10.1128/mBio.02574-14>.
- 342 [6] Ahmed W, Bivins A, Bertsch PM, Bibby K, Choi PM, Farkas K, et al. Surveillance of SARS-  
343 CoV-2 RNA in wastewater: Methods optimization and quality control are crucial for  
344 generating reliable public health information. *Curr Opin Environ Sci Heal* 2020;17:82–93.  
345 <https://doi.org/10.1016/j.coesh.2020.09.003>.
- 346 [7] Tiwari A, Kurittu P, Al-Mustapha AI, Heljanko V, Johansson V, Thakali O, et al. Wastewater  
347 surveillance of antibiotic-resistant bacterial pathogens: A systematic review. *Front Microbiol*  
348 2022;13. <https://doi.org/10.3389/fmicb.2022.977106>.
- 349 [8] Xu J, Xu Y, Wang H, Guo C, Qiu H, He Y, et al. Occurrence of antibiotics and antibiotic  
350 resistance genes in a sewage treatment plant and its effluent-receiving river. *Chemosphere*  
351 2015;119:1379–85. <https://doi.org/10.1016/j.chemosphere.2014.02.040>.
- 352 [9] Kimosop SJ, Getenga ZM, Orata F, Okello VA, Cheruiyot JK. Residue levels and discharge

- 353 loads of antibiotics in wastewater treatment plants (WWTPs), hospital lagoons, and rivers  
354 within Lake Victoria Basin, Kenya. *Environ Monit Assess* 2016;188.  
355 <https://doi.org/10.1007/s10661-016-5534-6>.
- 356 [10] Alexander J, Hembach N, Schwartz T. Evaluation of antibiotic resistance dissemination by  
357 wastewater treatment plant effluents with different catchment areas in Germany. *Sci Rep*  
358 2020;10. <https://doi.org/10.1038/s41598-020-65635-4>.
- 359 [11] Von Wintersdorff CJH, Penders J, Van Niekerk JM, Mills ND, Majumder S, Van Alphen LB,  
360 et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal  
361 gene transfer. *Front Microbiol* 2016;7. <https://doi.org/10.3389/fmicb.2016.00173>.
- 362 [12] Yuan T, Pian Y. Hospital wastewater as hotspots for pathogenic microorganisms spread into  
363 aquatic environment: A review. *Front Environ Sci* 2023;10.  
364 <https://doi.org/10.3389/fenvs.2022.1091734>.
- 365 [13] Vikram A, Lipus D, Bibby K. Metatranscriptome analysis of active microbial communities in  
366 produced water samples from the Marcellus Shale. *Microb Ecol* 2016;72:571–81.  
367 <https://doi.org/10.1007/s00248-016-0811-z>.
- 368 [14] Wandera EA, Komoto S, Mohammad S, Ide T, Bundi M, Nyangao J, et al. Genomic  
369 characterization of uncommon human G3P[6] rotavirus strains that have emerged in Kenya  
370 after rotavirus vaccine introduction, and pre-vaccine human G8P[4] rotavirus strains. *Infect*  
371 *Genet Evol* 2019;68:231–48. <https://doi.org/10.1016/j.meegid.2018.12.004>.
- 372 [15] Andrews S. FastQC. Babraham Bioinforma 2010.
- 373 [16] Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor.  
374 *Bioinformatics* 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
- 375 [17] Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data.  
376 *Bioinformatics* 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- 377 [18] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of



- 378 SAMtools and BCFtools. *Gigascience* 2021;10. <https://doi.org/10.1093/gigascience/giab008>.
- 379 [19] Kopylova E, Noé L, Touzet H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in  
380 metatranscriptomic data. *Bioinformatics* 2012;28:3211–7.  
381 <https://doi.org/10.1093/bioinformatics/bts611>.
- 382 [20] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo  
383 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference  
384 generation and analysis. *Nat Protoc* 2013;8:1494–512. <https://doi.org/10.1038/nprot.2013.084>.
- 385 [21] Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, et al. Metagenome-wide analysis of antibiotic  
386 resistance genes in a large cohort of human gut microbiota. *Nat Commun* 2013;4.  
387 <https://doi.org/10.1038/ncomms3151>.
- 388 [22] Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. ResFinder – an open  
389 online resource for identification of antimicrobial resistance genes in next-generation  
390 sequencing data and prediction of phenotypes from genotypes. *Microb Genomics* 2022;8.  
391 <https://doi.org/10.1099/mgen.0.000748>.
- 392 [23] Johansson MHK, Bortolaia V, Tansirichaiya S, Aarestrup FM, Roberts AP, Petersen TN.  
393 Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella*  
394 enterica using a newly developed web tool: MobileElementFinder. *J Antimicrob Chemother*  
395 2021;76:101–9. <https://doi.org/10.1093/JAC/DKAA390>.
- 396 [24] Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: A general classification scheme for  
397 bacterial virulence factors. *Nucleic Acids Res* 2022;50:D912–7.  
398 <https://doi.org/10.1093/nar/gkab1107>.
- 399 [25] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*  
400 2019;20. <https://doi.org/10.1186/s13059-019-1891-0>.
- 401 [26] Paulshus E, Kühn I, Möllby R, Colque P, O’Sullivan K, Midtvedt T, et al. Diversity and  
402 antibiotic resistance among *Escherichia coli* populations in hospital and community

- 403 wastewater compared to wastewater at the receiving urban treatment plant. *Water Res*  
404 2019;161:232–41. <https://doi.org/10.1016/j.watres.2019.05.102>.
- 405 [27] Hutinel M, Huijbers PMC, Fick J, Åhrén C, Larsson DGJ, Flach CF. Population-level  
406 surveillance of antibiotic resistance in *Escherichia coli* through sewage analysis.  
407 *Eurosurveillance* 2019;24. <https://doi.org/10.2807/1560-7917.ES.2019.24.37.1800497>.
- 408 [28] Duarte DJ, Zillien C, Kox M, Oldenkamp R, van der Zaan B, Roex E, et al. Characterization  
409 of urban sources of antibiotics and antibiotic-resistance genes in a Dutch sewer catchment. *Sci*  
410 *Total Environ* 2023;905. <https://doi.org/10.1016/j.scitotenv.2023.167439>.
- 411 [29] Haenelt S, Wang G, Kasmanas JC, Musat F, Richnow HH, da Rocha UN, et al. The fate of  
412 sulfonamide resistance genes and anthropogenic pollution marker *intI1* after discharge of  
413 wastewater into a pristine river stream. *Front Microbiol* 2023;14.  
414 <https://doi.org/10.3389/fmicb.2023.1058350>.
- 415 [30] Prieto Riquelme MV, Garner E, Gupta S, Metch J, Zhu N, Blair MF, et al. Demonstrating a  
416 Comprehensive Wastewater-Based Surveillance Approach That Differentiates Globally  
417 Sourced Resistomes. *Environ Sci Technol* 2022;56. <https://doi.org/10.1021/acs.est.1c08673>.
- 418 [31] Makowska N, Koczura R, Mokracka J. Class 1 integrase, sulfonamide and tetracycline  
419 resistance genes in wastewater treatment plant and surface water. *Chemosphere*  
420 2016;144:1665–73. <https://doi.org/10.1016/j.chemosphere.2015.10.044>.
- 421 [32] Stachurová T, Malachová K, Semerád J, Sterniša M, Rybková Z, Možina SS. Tetracycline  
422 induces the formation of biofilm of bacteria from different phases of wastewater treatment.  
423 *Processes* 2020;8. <https://doi.org/10.3390/PR8080989>.
- 424 [33] Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, et al.  
425 Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: A  
426 molecular, biological, and epidemiological study. *Lancet Infect Dis* 2010;10:597–602.  
427 [https://doi.org/10.1016/S1473-3099\(10\)70143-2](https://doi.org/10.1016/S1473-3099(10)70143-2).

- 428 [34] Stachurová T, Piková H, Bartas M, Semerád J, Svobodová K, Malachová K. Beta-lactam  
429 resistance development during the treatment processes of municipal wastewater treatment  
430 plants. *Chemosphere* 2021;280. <https://doi.org/10.1016/j.chemosphere.2021.130749>.
- 431 [35] Divyashree M, Mani MK, Shama Prakash K, Vijaya Kumar D, Veena Shetty A, Shetty AK, et  
432 al. Hospital wastewater treatment reduces NDM-positive bacteria being discharged into water  
433 bodies. *Water Environ Res* 2020;92:562–8. <https://doi.org/10.1002/wer.1248>.
- 434 [36] Harmer CJ, Moran RA, Hall RM. Movement of IS26-Associated antibiotic resistance genes  
435 occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to  
436 another IS26. *MBio* 2014;5. <https://doi.org/10.1128/mBio.01801-14>.
- 437 [37] Buelow E, Rico A, Gaschet M, Lourenço J, Kennedy SP, Wiest L, et al. Hospital discharges in  
438 urban sanitation systems: Long-term monitoring of wastewater resistome and microbiota in  
439 relationship to their eco-exposome. *Water Res X* 2020;7.  
440 <https://doi.org/10.1016/j.wroa.2020.100045>.
- 441 [38] Buelow E, Dauga C, Carrion C, Mathé-Hubert H, Achaibou S, Gaschet M, et al. Hospital and  
442 urban wastewaters shape the matrix and active resistome of environmental biofilms. *Water Res*  
443 2023;244. <https://doi.org/10.1016/j.watres.2023.120408>.
- 444 [39] Skwor T, Stringer S, Haggerty J, Johnson J, Duhr S, Johnson M, et al. Prevalence of  
445 potentially pathogenic antibiotic-resistant aeromonas spp. in treated urban wastewater  
446 effluents versus recipient riverine populations: A 3-year comparative study. *Appl Environ*  
447 *Microbiol* 2020;86. <https://doi.org/10.1128/AEM.02053-19>.
- 448

449 **Table 1.** Antibiotic resistance genes found in the same transcript with mobile genetic elements

<b>Site</b>	<b>Mobile Genetic Element</b>	<b>Antibiotic resistance gene</b>
<b>Referral Hospital</b>	IS6	<i>sul1, qacE, msr(E), mph(E), aac(6'')-lb-cr, aac(6')-lb-Hangzhou</i>
	IS5	<i>aph(6)-ld, aph(3'')-lb</i>
	ISVsa3	<i>sul2</i>
	IS17	<i>aph(3'')-lb, aph(6)-ld</i>
	IS6100	<i>aac(6')-lb-Hangzhou, aac(6'')-lb-cr, sul1, qacE</i>
	IncQ1	<i>blaCMY-4</i>
	IS26	<i>mph(E)</i>
<b>University</b>	IS6100	<i>qacE, sul1</i>
<b>Hostels</b>	ISVsa3	<i>sul2</i>

450

451 **Figure Legend**

452 **Figure 1.** Antibiotic classes detected across the four different sites

453 **Figure 2.** Circos plot showing the interactions of various wastewater ARGs across hospital,  
454 university and influent and effluent of a WWTP.

455 **Figure 3.** Total relative abundance of ARGs present in wastewater across different sites.

456 **Figure 4.** Taxonomic classification of bacterial organisms present across various wastewater sites. **A)**  
457 Proportion of bacterial organisms at the class level. **B)** Proportion of bacterial organisms at the genus  
458 level. **C)** Proportion of WHO list of priority pathogens identified from different wastewater sites after  
459 taxonomy classification using Kraken.













