Genomic surveillance of antimicrobial-resistant *Escherichia coli* in fecal sludge and sewage in Uganda

Ryota Gomi[a, *], Yasufumi Matsumura[b], Masaki Yamamoto[b], Mai Tanaka[a], Allan John Komakech[c], Tomonari Matsuda[a], Hidenori Harada[d, *]

[a]Department of Environmental Engineering, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, 615-8540, Kyoto, Japan

[b]Department of Clinical Laboratory Medicine, Graduate School of Medicine, Kyoto University, 54 Shogoin-kawahara-cho, Sakyo-ku, 606-8507, Kyoto, Japan

[c]Department of Agricultural and Biosystems Engineering, College of Agricultural and Environmental Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda

[d]Graduate School of Asian and African Area Studies, Kyoto University, Yoshida-Shimoadachi, Sakyo-ku, 606-8501, Kyoto, Japan

*Corresponding authors

E-mail address: gomi.ryota.34v@kyoto-u.jp; Tel: +81-75-383-3354

E-mail address: harada.hidenori.8v@kyoto-u.ac.jp; Tel: +81-75-753-7830

ABSTRACT

The global increase of antimicrobial resistance (AMR) is a major public health concern. An effective AMR surveillance tool is needed to track the emergence and spread of AMR. Wastewater surveillance has been proposed as a resource-efficient tool for monitoring AMR carriage in the community. Here, we performed genomic surveillance of antimicrobial-resistant *Escherichia coli* obtained from fecal sludge and sewage in Uganda to gain insights into *E. coli* epidemiology and AMR burden in the underlying population. Selective media containing different antibiotic combinations (cefotaxime, ciprofloxacin, cefotaxime + ciprofloxacin + gentamicin) were used to obtain antimicrobial-resistant *E. coli* from fecal sludge and sewage. Short-read sequencing was performed for the obtained isolates, and a subset of isolates (selected from predominant sequence types (STs)) was also subjected to long-read sequencing. Genomic analysis of the obtained *E. coli* isolates (n = 181) revealed the prevalence of ST10 complex, including ST167 (n = 43), ST10 (n = 28), ST1284 (n = 17), and ST617 (n = 4), in both fecal sludge and sewage, irrespective of antibiotics used for selection. We also detected global high-risk clones ST1193 (n = 10) and ST131 (n = 2 clade A, n = 3 subclade C1-M27, and n = 1 subclade C2). Diverse AMR determinants, including extended-spectrum β-lactamase genes (mostly $bla_{\text{CTX-M-15}}$) and mutations in *gyrA* and *parC*, were identified. Analysis of the completed genomes revealed that diverse IncF plasmids and chromosomal integration were the major contributors to the spread of AMR genes in the predominant STs. This study showed

that a combination of sewage surveillance (or fecal sludge surveillance) and whole-genome sequencing can be a powerful tool for monitoring AMR carriage in the underlying population.

Keywords: *Escherichia coli*, Antimicrobial resistance, One Health, Sewage surveillance, Whole-genome sequencing

43 **1. Introduction**

44 The increase of antimicrobial resistance (AMR) is a global health concern. AMR can lead to
45 adverse consequences such as increased mortality rates, prolonged hospital admissions, and
46 higher medical costs.[1] Information on AMR collected through surveillance is needed to guide
47 treatment, monitor the effectiveness of interventions, and develop polices and regulations to
48 counter AMR.[2] However, surveillance of AMR in resource-limited settings is challenging,
49 especially when the surveillance targets populations.

50 Sewage contains pooled urine and feces from a large population, and one sewage sample can
51 represent bacteria from multiple people in the community. For this reason, sewage surveillance
52 (or wastewater-based epidemiology) has been proposed as a resource-efficient tool for
53 monitoring AMR carriage in the underlying population.[3-5] Sewage surveillance is attractive in
54 many aspects; for example, (i) sewage provides a sample from a large and mostly healthy
55 population, which otherwise would not be easily monitored, and (ii) it does not collect data on
56 individuals and thus reduces ethical risks.[6, 7] Although it is a promising approach, sewage
57 surveillance of AMR cannot be directly applied in areas with no connection to centralized
58 (sewered) systems. A previous study reported at least 1.8 billion people in low- and middle-
59 income countries use on-site sanitation systems like pit latrines.[8] In such areas, fecal sludge
60 from on-site systems has been proposed to be used as pooled community samples for
61 surveillance purposes.[9-11] Sewage surveillance (or fecal sludge surveillance) of AMR can be

62 either culture-dependent or culture-independent. Culture-independent methods (e.g., short-read
63 metagenomics) have some advantages over culture-dependent methods, such as the ability to
64 detect thousands of AMR genes in a single sample. However, it provides limited information
65 on the host or the genetic context of AMR genes.[6] Culture-dependent methods, on the other
66 hand, can link AMR genes to specific host bacteria, provide information on phylogenetic
67 characteristics on the host, and may even provide information on the location (chromosome or
68 plasmid) and genetic context of AMR genes.

69 *Escherichia coli* is a commensal bacterium in humans and animals and is also known to cause
70 serious infections.[12] Although *E. coli* is intrinsically susceptible to almost all clinically relevant
71 antibiotics, it can accumulate AMR genes via, e.g., horizontal gene transfer.[13] *E. coli* has been
72 suggested as an indicator for AMR surveillance in the environment,[14] and previous studies have
73 shown that *E. coli* is a good target organism for sewage surveillance of AMR.[15-20] WHO also
74 selected extended-spectrum β-lactamase (ESBL)-producing *E. coli* as an indicator for "One
75 Health" AMR surveillance (including sewage surveillance) in the Tricycle protocol.[21] The
76 protocol recommends usage of whole-genome sequencing (if possible) to determine the
77 sequence type (ST), virulence genotype, plasmid type, carriage of acquired AMR genes, and
78 phylogenetic characteristics of the obtained isolates, which contribute to the understanding of
79 the epidemiology of pathogens.

80 Uganda is a country experiencing an extremely high prevalence of AMR.[22] Previous studies
81 have reported high prevalence of multidrug resistance (MDR) among *E. coli*.[23-25] However,
82 most studies were based on phenotypic characterization or detection of a limited set of AMR
83 genes/virulence genes, and thus their genomic characteristics, including the clonal composition
84 and resistome, are largely unknown. In the present study, we isolated antimicrobial-resistant *E.*
85 *coli* from fecal sludge and sewage in Uganda and performed genomic analysis on the obtained
86 isolates, aiming to gain insights into *E. coli* epidemiology and AMR burden in the underlying
87 population.

88

89 **2. Materials and methods**

90 **2.1. Sample collection and isolation of *E. coli***

91 Samples were collected at a fecal sludge treatment plant (FSTP) and a wastewater treatment
92 plant (WWTP) in Kampala, Uganda, in October–November, 2018. We obtained samples from
93 both sources to cover a wider population in this region. For fecal sludge sampling, three grab
94 samples were collected at different times on each of five days from effluent of the primary
95 sedimentation tank in the FSTP; in total 15 samples were collected (these samples are hereafter
96 referred to simply as fecal sludge samples). For sewage sampling, up to three grab samples
97 were collected at different times over six days from influent of the WWTP, totaling 16 samples.

98 The samples were collected and stored in sterile sampling bottles and transported on ice to the
99 laboratory. The samples were then cultured, within 12 hours after collection, using a specific
100 enzyme substrate medium (CHROMagar ECC, Kanto Chemical, Tokyo, Japan) containing
101 four combinations of three antibiotics: (a) no antibiotics, (b) cefotaxime (CTX, 1 mg/L), (c)
102 ciprofloxacin (CIP, 1 mg/L), (d) CTX (1 mg/L) + CIP (1 mg/L) + gentamicin (GM, 4 mg/L).
103 The samples were serially diluted with PBS for colony counting. Up to three colonies with an
104 *E. coli* profile (blue colony) were picked from each of plate (b), (c), and (d) (i.e., we obtained
105 up to nine isolates from a single sample), and each was transferred into a screw tube containing
106 LB agar. The tubes were incubated overnight at 37 °C, stored at 4 °C, and then transported to
107 the laboratory in Japan. *E. coli* in each screw tube was again incubated on a plate with
108 CHROMAgar ECC containing the same antibiotic(s) as at the time of isolation. A colony with
109 an *E. coli* profile was picked and spread again on a plate with CHROMAgar ECC, and then a
110 single colony was picked from the plate, cultured in LB broth, and stored at –85 °C in 35%
111 glycerol.

112 **2.2. Genome sequencing and assembly**

113 A total of 192 presumptive antimicrobial-resistant *E. coli* isolates was subjected to genome
114 sequencing as below. For each isolate, a bacterial pellet was prepared from the glycerol stock.
115 DNA was extracted from each isolate using a QIAamp DNA Mini Kit (Qiagen, Hilden,
116 Germany) for Illumina sequencing. Illumina libraries were prepared using a Nextera XT DNA

117 Library Preparation Kit (Illumina, San Diego, CA) and sequenced on the NextSeq550 platform,
118 generating paired-end reads of 151 bp each. For some isolates, read depth was not sufficient,
119 so we sequenced the library again on the MiniSeq platform to obtain additional paired-end
120 reads of the same length. Illumina reads were processed using fastp (v0.23.2)[26] and assembled
121 using Unicycler (v0.4.8)[27] with the --no_correct option. The quality of each genome assembly
122 was ensured by using the quality control (QC) criteria of (i) contig number: ≤ 800; (ii) genome
123 size: 3.7 Mbp to 6.4 Mbp; and (iii) N50: >20 kb, as recommended by EnteroBase.[28]

124 For selected isolates (n = 26), we also performed Oxford Nanopore Technologies (ONT) long-
125 read sequencing to generate complete genomes. DNA was extracted from each isolate using a
126 DNeasy PowerSoil Pro Kit (Qiagen). The long-read sequencing library was prepared using the
127 SQK-LSK109 kit (Oxford Nanopore Technologies, Oxford, UK) and sequenced on the
128 MinION with a FLO-MIN106 flow cell. ONT reads were filtered using Filtlong (v0.2.1,
129 https://github.com/rrwick/Filtlong) to remove reads with low quality or short length. Hybrid
130 assembly of Illumina short reads and ONT long reads was performed by employing a long-
131 read-first approach. Briefly, filtered ONT reads were assembled using Flye (v2.9.1-b1780),[29]
132 and then the assemblies were polished with medaka (v1.7.1,
133 https://github.com/nanoporetech/medaka) using filtered ONT reads and Polypolish (v0.5.0)[30]
134 using processed Illumina reads. Hybrid assembly was also performed using a short-read-first
135 approach, i.e., Unicycler (v0.5.0) hybrid assembly followed by polishing with Polypolish

136 (v0.5.0) using processed Illumina reads. If the long-read-first approach generated a complete
137 genome for an isolate, we chose the polished Flye assembly as the final assembly (note that an
138 assembly generated by Flye can sometimes lack small plasmids, so we recovered small
139 plasmids in the Unicycler assembly that were not present in the Flye assembly[31]). If the long-
140 read-first approach did not complete the genome but the short-read-first approach did, we chose
141 the polished Unicycler assembly as the final assembly.

142 **2.3. Genomic analysis**

143 *E. coli* status was confirmed by performing an average nucleotide identity (ANI) analysis using
144 each genome sequenced in the present study and the genome sequence of the *E. coli* type strain
145 ATCC 11775 (= DSM 30083 = U5/41).[32] ANI analysis was performed using FastANI (v1.33)
146 with the threshold of 95%.[33] *E. coli* phylogroups were determined using a combination of
147 EzClermont (v0.7.0) and ClermonTyping.[34, 35] Multilocus sequence typing (MLST) was
148 performed using MLST (v2.19.0) (https://github.com/tseemann/mlst). *fimH* types, which are
149 used for high-resolution subtyping of MLST-based *E. coli* clonal groups, were determined
150 using FimTyper 1.0.[36] AMR genes were detected using a combination of ResFinder 4.1[37] and
151 Abricate (v1.0.1) (https://github.com/tseemann/abricate) with the ncbi database. Genetic
152 contexts of AMR genes were analyzed using the RAST server,[38] the BLAST algorithm
153 (https://blast.ncbi.nlm.nih.gov/Blast.cgi), and ISfinder.[39] Mutations in quinolone resistance-
154 determining regions (QRDRs) in *gyrA* (GyrA codons 83 and 87) and *parC* (ParC codons 80

155 and 84) were analyzed according to Aoike et al.[40] Virulence genes were detected using the
156 VirulenceFinder database.[41, 42] *E. coli* pathotypes, which include extraintestinal pathogenic *E.*
157 *coli* (ExPEC) and several categories of intestinal pathogenic *E. coli* (InPEC), were defined
158 based on the presence of specific virulence genes as follows: two or more of *papA* and/or *papC*,
159 *sfa-focDE*, *afa-draBC*, *iutA*, and *kpsM* II for ExPEC; *stx1* or/and *stx2* and *eae* for
160 enterohemorrhagic *E. coli* (EHEC); *eae* without *stx1* or *stx2* for enteropathogenic *E. coli*
161 (EPEC); *stx1* or/and *stx2* for Shiga toxin-producing *E. coli* (STEC); *ipaH* for enteroinvasive *E.*
162 *coli* (EIEC); *elt* or/and *est* for enterotoxigenic *E. coli* (ETEC); and *aggR* for enteroaggregative
163 *E. coli* (EAEC).[43] K1 and K5 capsular types of ST1193 isolates were determined based on
164 blastn analysis using M57382 (K1) and X53819 (K5) as references.[44] Plasmid replicons were
165 detected using PlasmidFinder 2.1, and plasmids were typed using pMLST 2.0.[45]

166 A whole-genome SNP-based tree was built for isolates belonging to phylogroup A using
167 kSNP3 based on SNPs occurring in at least 90% of the genomes.[46] To confirm the topology of
168 the tree, we also built a tree for the phylogroup A genomes using another software, parsnp
169 (v1.7.4).[47] For isolates belonging to ST1193 and ST131, we separately built a recombination-
170 free phylogenetic tree for each ST. To place our isolates in broader phylogenetic contexts, we
171 also included isolates analyzed in other studies which defined clades/subclades within each ST,
172 i.e., isolates analyzed by Johnson et al. for ST1193[48] and isolates analyzed by Matsumura et al.
173 for ST131.[49] Isolates from these studies were assembled and subjected to QC criteria as

174 described above, and genomes passing the QC criteria were used for tree building. For each
175 ST, trimmed reads were mapped to a reference genome (MCJCHV-1 chromosome for
176 ST1193[50] and EC958 chromosome for ST131[51]) using snippy (v4.6.0)
177 (https://github.com/tseemann/snippy). Recombinant regions were filtered from the snippy
178 core-full alignment using Gubbins (v3.2.1).[52] A maximum likelihood (ML) phylogenetic tree
179 was built using IQ-TREE (v2.0.3) from the resulting alignment using a GTR+G+ASC model.[53]
180 The trees were visualized using iTOL.[54]

181 **2.4. Accession numbers**

182 Genome assemblies were deposited in GenBank under BioProject PRJNA956323 (also see
183 **Table S1** for the accession number of each genome and **Table S2** for the accession number of
184 each replicon for the completed genomes).

185

186 **3. Results and discussion**

187 **3.1. Detection and isolation of antimicrobial-resistant *E. coli***

188 During the study period, we collected 15 fecal sludge samples and 16 sewage samples and
189 measured the concentrations of total *E. coli* and antimicrobial-resistant *E. coli* (**Table S3**). The
190 mean proportions of antimicrobial-resistant *E. coli* in the fecal sludge samples were 30.0% for
191 CTX, 26.9% for CIP, and 10.7% for CTX + CIP + GM. The mean proportions of antimicrobial-

192 resistant *E. coli* in the sewage samples were 39.2% for CTX, 21.0% for CIP, and 5.1% for CTX
193 + CIP + GM. These proportions were higher than those reported previously for *E. coli* in
194 wastewater (usually less than 10% resistance for CTX and CIP),[16, 18, 19] potentially highlighting
195 the extremely high prevalence of AMR in Uganda.[22]

196 In total, we obtained 192 presumptive antimicrobial-resistant *E. coli* isolates from the samples
197 and performed short-read sequencing on these isolates. After discarding contigs shorter than
198 100 bp, 187 genomes met the assembly QC criteria. Species identification of the 187 genomes
199 by the ANI analysis revealed that 181 belonged to *E. coli*, while 6 belonged to non-*E. coli*
200 species such as *Citrobacter freundii*. These 181 *E. coli* genomes, comprising 91 from fecal
201 sludge samples (31 from CTX-containing media, 31 from CIP-containing media, and 29 from
202 CTX + CIP + GM-containing media) and 90 from sewage samples (32 from CTX-containing
203 media, 26 from CIP-containing media, and 32 from CTX + CIP + GM-containing media), were
204 further analyzed in the present study.

205 **3.2. Clonal composition**

206 In total, 36 STs from six different phylogroups were detected among the 181 isolates (**Table**
207 **S1**). Phylogroup A was the most prevalent (n = 127), followed by B1 (n = 17), B2 (n = 17), C
208 (n = 10), D (n = 8), and F (n = 2). ST167 (n = 43), ST10 (n = 28), ST46 (n = 19), and ST1284
209 (n = 17), all of which belong to phylogroup A, were prevalent in both fecal sludge isolates and
210 sewage isolates (**Figure 1a**). In total, 144 isolates (79.6%) were from the top ten STs, while 37

211 isolates (20.4%) were from minor STs (i.e., STs that are not in the top ten and detected among
212 three or fewer isolates) (**Figures 1b** and **1c**). The top ten STs were generally associated with
213 two or three types of media and both fecal sludge and sewage, with a few exceptions (e.g.,
214 ST131 was exclusively found among fecal sludge isolates). Lineages associated with
215 extraintestinal infections were also among the top ten STs and detected in fecal sludge isolates
216 (n = 9 ST1193, n = 5 ST410, n = 6 ST131, n = 2 ST69) and sewage isolates (n = 1 ST1193, n
217 = 4 ST410, n = 2 ST69).[12] Thirty-four isolates (18.8%), including those from the above-
218 mentioned STs associated with extraintestinal infections, were classified as ExPEC based on
219 the virulence gene contents (**Figure 1**). On the other hand, none of the isolates were classified
220 as InPEC. ExPEC strains are facultative pathogens and part of the intestinal microflora of a
221 fraction of the healthy population,[55] which might explain the relatively high prevalence of
222 ExPEC among our *E. coli* collection.
223 We next compared the clonal composition of *E. coli* in this study and those reported previously
224 for clinical *E. coli* in Uganda. However, the clonal composition of *E. coli* in clinical settings is
225 largely unknown in Uganda (a search of human-derived *E. coli* genomes from Uganda in
226 EnteroBase returned no hits as of May 2023), and we could find only one study reporting the
227 ST composition of clinical *E. coli* isolates in Uganda. That study sequenced 24 Ugandan *E.*
228 *coli* isolates causing urinary tract infections (most were MDR) and reported the prevalence of
229 ST10 (n = 4, 15.4%) as well as the occurrence of ST410 (n = 2, 8.3%), ST131 (n = 1, 4.2%),

230 ST167 (n = 1, 4.2%), ST617 (n = 1, 4.2%), and ST1193 (n = 1, 4.2%).[56] These STs were also
231 among the top ten STs in the present study, indicating that sewage surveillance may serve as a
232 possible resource-efficient complement to clinical surveillance, though further studies are
233 needed to confirm this.
234 Previous studies have suggested that fecal sludge and sewage can be used for community
235 surveillance, but fecal sludge includes waste from fewer individuals and hence may be less
236 divergent than sewage.[10] However, the number of unique STs was larger for fecal sludge
237 isolates than sewage isolates (26 unique STs vs 23 unique STs) (**Table S1**). Moreover,
238 Simpson's diversity index was higher for fecal sludge isolates than sewage isolates when
239 isolates from the same type of media were compared (**Figure 1a**). This might be because of
240 the fecal sludge sampling strategy employed in the present study: we obtained samples from
241 effluent of the primary sedimentation tank in a FSTP, which probably represents fecal sludge
242 from a large number of individuals. This might have contributed to the relatively high diversity
243 of *E. coli* isolates from fecal sludge samples, suggesting sampling from FSTPs or otherwise
244 designated fecal sludge dumping sites (rather than individual pit latrines) may better capture
245 the population-scale information in fecal sludge surveillance, when such places exist in the
246 region.
247

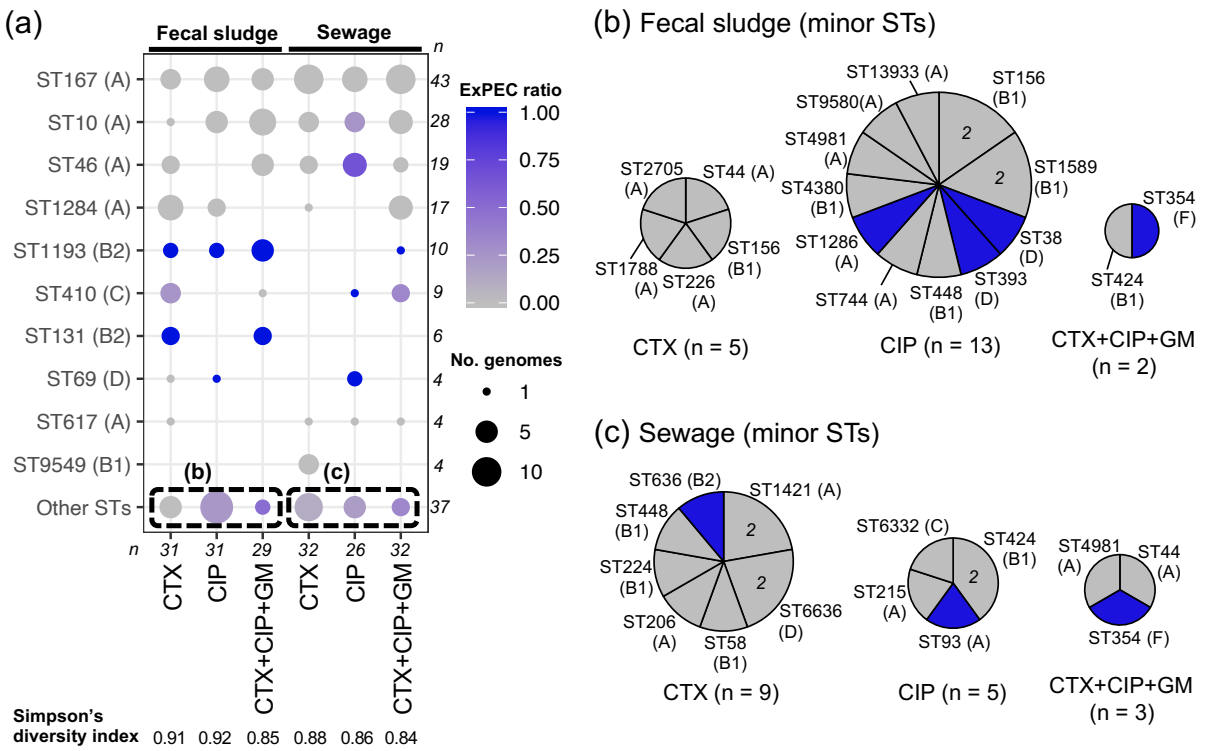


Figure 1. (a) Clonal composition of 181 *E. coli* isolates. STs are shown separately for each media type (CTX, CIP, and CTX + CIP + GM) and each sample type (fecal sludge and sewage). The top ten STs are shown in descending order. The remaining STs, all of which were with three or fewer isolates, are combined into "other STs". Phylogroup corresponding to each ST is shown in parentheses. Circle size indicates the number of genomes, and intensity of circle shading indicates the proportion of genomes classified as ExPEC (as per inset legends). Clonal composition of minor STs (i.e., STs classified as "other STs" in (a)) from (b) fecal sludge and (c) sewage. Phylogroup corresponding to each ST is shown in parentheses, and STs classified as ExPEC are colored blue.

3.3. Antimicrobial resistance determinants

260 Six different ESBL genes, namely $bla_{\text{CTX-M-3}}$, $bla_{\text{CTX-M-14}}$, $bla_{\text{CTX-M-15}}$, $bla_{\text{CTX-M-27}}$, $bla_{\text{CTX-M-55}}$, and

261 $bla_{\text{TEM-169}}$, were detected among the isolates, with $bla_{\text{CTX-M-15}}$ being the predominant type

262 (**Figure 2a**). Based on the short-read assemblies, there were 21 isolates with two different

263 ESBL genes (n = 18 $bla_{\text{CTX-M-14}}$ + $bla_{\text{CTX-M-15}}$, n = 1 $bla_{\text{CTX-M-15}}$ + $bla_{\text{CTX-M-27}}$, n = 1 $bla_{\text{CTX-M-3}}$ +

264 $bla_{\text{CTX-M-27}}$, and n = 1 $bla_{\text{TEM-169}}$ + $bla_{\text{CTX-M-15}}$) (note that isolates carrying multiple copies of the

265 same ESBL gene were not considered here because those genes can be collapsed into one

266 contig in short-read assemblies). As expected, all but one isolate obtained from CTX-

267 containing media and CTX + CIP + GM-containing media carried at least one ESBL gene.

268 Importantly, many isolates obtained from CIP-containing media carried ESBL genes (41.9%

269 for fecal sludge isolates and 42.3% for sewage isolates), indicating the convergence of ESBL

270 genes and fluoroquinolone resistance.

271 All but four isolates obtained from CIP-containing media and CTX + CIP + GM-containing

272 media carried three or four mutations in QRDRs (**Figure 2b**). In all isolates with three

273 mutations in QRDRs, two were in codons 83 and 87 in GyrA, and the remaining mutation was

274 in ParC codon 80 or 84. These are congruent with previous studies reporting that two mutations

275 in the QRDR of *gyrA* and one or two mutations in the QRDR of *parC* are the main mechanism

276 of ciprofloxacin resistance in *E. coli*.[57, 58] Importantly, most isolates obtained from CTX-

277 containing media also carried three or four mutations in QRDRs (87.1% for fecal sludge

278 isolates and 75.0% for sewage isolates), again highlighting the convergence of ESBL genes

279 and fluoroquinolone resistance determinants.

280 Gentamicin resistance in isolates obtained from CTX + CIP + GM-containing media could be

281 mostly explained by the presence of *aac(3)-II* genes (**Table S1**). Other resistance genes not

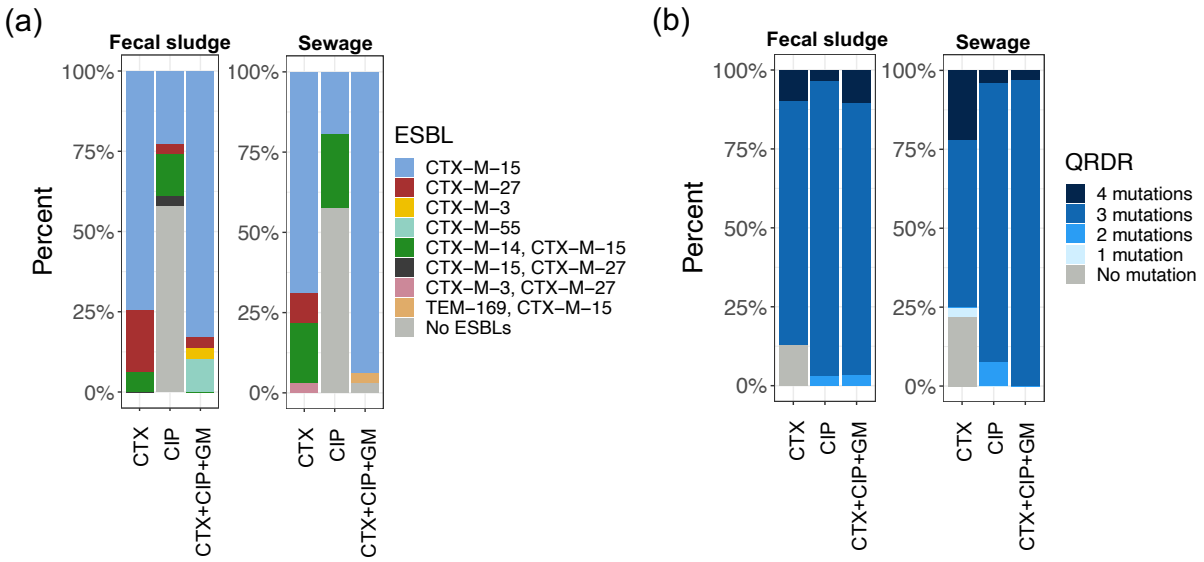
282 associated with antibiotics used in the selective media were also prevalent in the isolates; for

283 example, *sul* genes were detected in 175 isolates (96.7%) and *tet* genes were detected in 161

284 isolates (89.0%). However, genes conferring resistance to last-resort antibiotics, such as

285 carbapenems and colistin, were not detected in any isolates (**Table S1**).

286



287

288 **Figure 2**. (a) ESBL genes and (b) QRDR mutations in *E. coli* isolates. AMR determinants are

289 shown separately for each media type (CTX, CIP, and CTX + CIP + GM) and each sample

290 type (fecal sludge and sewage). This figure is based on the analysis of short-read assembly data

291 and isolates with multiple copies of the same ESBL gene can be reported to carry a single copy.

292 For QRDRs, mutations in GyrA codons 83 and 87 and ParC codons 80 and 84 were counted.

293

294 **3.4. Genomic characteristics of phylogroup A isolates**

295 Whole genome-based phylogenetic analysis was performed using kSNP3 to gain insights into

296 clonal relationships of 127 isolates belonging to phylogroup A, which was the dominant

297 phylogroup in the present study (**Figure 3**). The tree was largely separated according to STs,

298 though we observed separate clusters for ST10, ST46, and ST167. These separate clusters were

299 also confirmed in a tree generated using different software, parsnp (**Figure S1**). The ST10 and

300 ST167 clusters were characterized by distinct *fimH* types, namely *fimH54*, *fimH435*, and

301 *fimH789* for ST10, and *fimH5*, *fimH29*, and *fimH54* for ST167. ST46 isolates from both

302 clusters did not carry *fimH*. According to EnteroBase,[28] ST167, ST10, ST1284, and ST617 are

303 all classified into ST10 complex (STC10). Classifying all isolates from a monophyletic clade

304 containing these four STs (the bottom clade in **Figure 3**) into STC10 identified that 99 isolates

305 (54.7% of isolates in the present study) belong to this lineage, indicating STC10 poses a

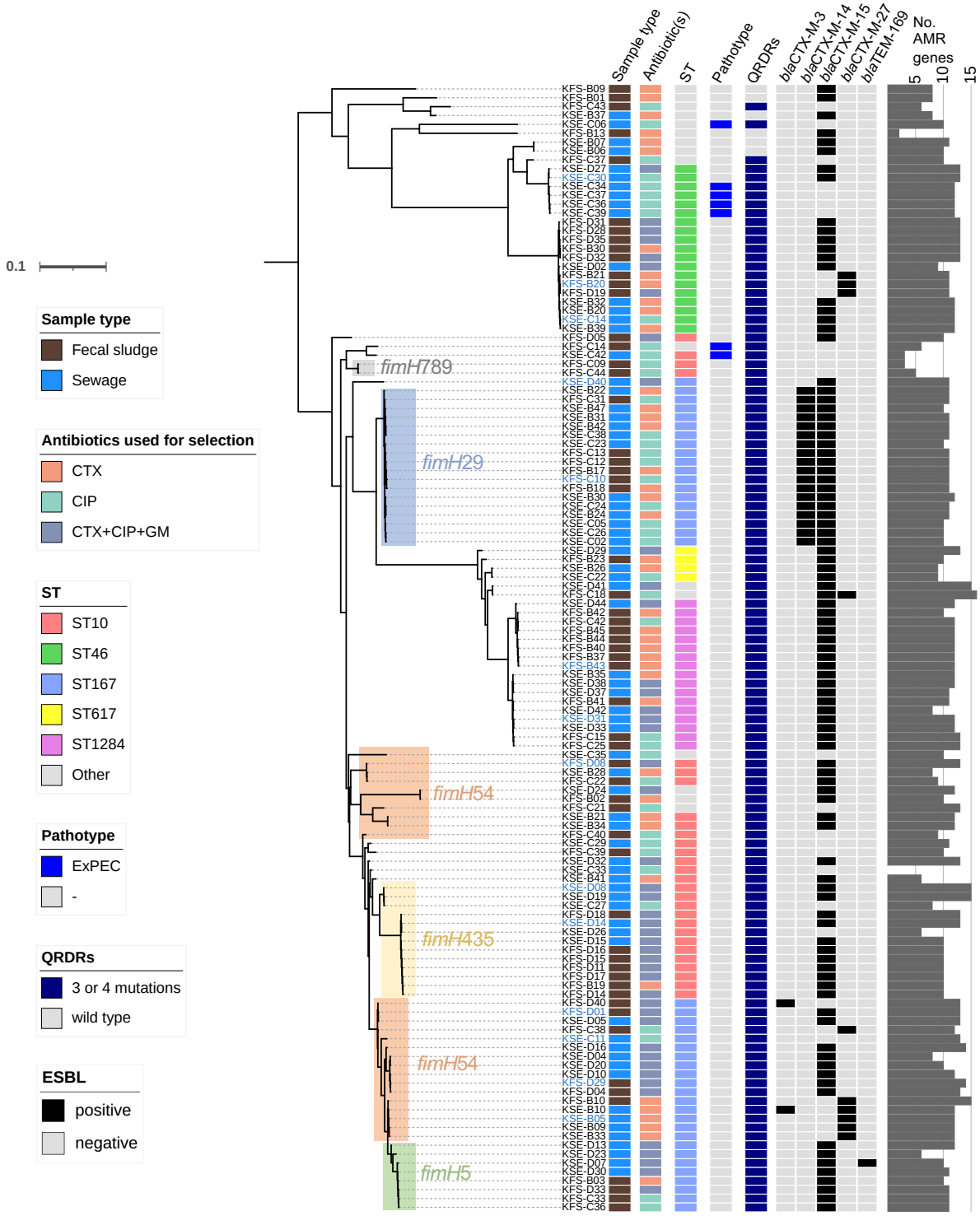
306 significant AMR burden in the underlying population. STC10 has been globally reported as a

307 resident of the intestinal tract of humans and other animals, and carriage of AMR genes

308 (including ESBL genes) has also been widely reported in this lineage.[59] Although there were

309 some small clusters specific to fecal sludge isolates (e.g., ST10 with *fimH789*) and sewage
310 isolates (e.g., a cluster containing six sewage isolates within ST46), fecal sludge isolates and
311 sewage isolates were generally intermingled in the tree. This trend was also seen for media
312 types; isolates selected with different antibiotics were generally intermingled in the tree. Most
313 phylogroup A isolates (except six isolates from minor STs obtained using CTX-containing
314 media) carried three or four mutations within QRDRs. As mentioned above, ESBL genes were
315 also prevalent even in isolates obtained using CIP-containing media (n = 21/40, 52.5%).
316 Notably, all ST167 isolates with *fimH*29 (n = 18) carried both $bla_{\text{CTX-M-14}}$ and $bla_{\text{CTX-M-15}}$.
317 Analysis of the completed genome of KFS-C10 from this clade identified that $bla_{\text{CTX-M-14}}$ was
318 integrated into the chromosome while $bla_{\text{CTX-M-15}}$ was carried on an IncF[F31:F36:A4:B1]
319 plasmid (**Table S2**). There were seven isolates with ExPEC status (n = 4 ST46, n = 1 ST10, n
320 = 1 ST93, n = 1 ST1286). All seven isolates were obtained from CIP-containing media and
321 lacked ESBL genes, but they carried 3 to 12 non-ESBL AMR genes in addition to QRDR
322 mutations.

323



324

325 **Figure 3**. kSNP3 whole-genome SNP-based phylogenetic tree of 127 phylogroup A isolates.

326 The tree was constructed using 54,632 SNP-loci, which occurred in at least 90% of the genomes.

327 A k-mer length of 19 was used for SNP identification. The presence of traits is indicated by

328 colored boxes (as per inset legends). For the ST column, major STs (i.e., ST10, ST46, ST167,
329 ST617, and ST1284) are indicated by the corresponding color boxes, while other STs are
330 represented by gray boxes. ST10 and ST167 clusters are highlighted in the tree according to
331 the *fimH* alleles, except for singletons. The names of completed genomes are colored blue.

332

333 **3.5. Genomic characteristics of ST1193 and ST131 isolates**

334 We investigated the detailed genetic characteristics of isolates belonging to ST1193 (n = 10)
335 and ST131 (n = 6), most of which were obtained from fecal sludge, because these STs are
336 recognized as established/emerging global high-risk clones and the only MDR clones that are
337 dominant among unselected ExPEC populations.[60]

338 ST1193 is derived from the ST14 clonal complex, carries *fimH64*, and harbors *gyrA* (S83L and
339 D87N) and *parC* (S80I) mutations.[60] Previous studies reported that ST1193 isolates cluster
340 together within the phylogenetic tree according to three major IncF plasmid types, namely
341 F-:A1:B1, F-:A1:B10, and F-:A1:B20, though the clusters are not strictly monophyletic.[48, 61]
342 These clusters were named by Johnson et al. as green (F-:A1:B1), red (F-:A1:B10), and pink
343 (F-:A1:B20).[48] The phylogenetic tree constructed in our study was congruent with these
344 previous studies (i.e., largely separated according to IncF types), and our isolates fell into an
345 F-:A1:B10 cluster (n = 7) and an F-:A1:B20 cluster (n = 3) (**Figure 4a**). Three isolates (KFS-

346 B24, KFS-B25, KFS-C19) and two isolates (KFS-D07, KSE-D22) with the F-:A1:B10 plasmid
347 type clustered together in the tree. Three isolates (KFS-D21, KFS-D24, KFS-D26) with the
348 F-:A1:B20 plasmid type also clustered together. ST1193 isolates with F-:A1:B:20 plasmids
349 generally contain the K5 capsular type, and those with F-:A1:B1 or F-:A1:B10 plasmids
350 generally contain the K1 capsular type.[48] However, all of our ST1193 isolates, including those
351 with F-:A1:B:20 plasmids, contained the K1 capsular type. A variety of AMR genes associated
352 with different drug classes were detected in our ST1193 isolates (**Figure 4c**). Notably, among
353 six isolates with complete genomes, three (KFS-B24, KFS-D13, KSE-D22) carried
354 chromosomal AMR genes, including $bla_{\text{CTX-M-15}}$, within different genetic contexts (discussed in
355 detail below and in **Figure S2**). This indicates chromosomal integration of AMR genes, as well
356 as plasmid carriage, can contribute to the MDR phenotype of ST1193. All ST1193 isolates in
357 this study were classified as ExPEC based on the virulence gene profiles, which potentially
358 indicates these isolates can cause extraintestinal infections.

359 ST131 is known to be divided into clades A, B, and C (containing subclades C0, C1, and C2).[62]
360 Within subclade C1 there is a clone known as C1-M27, which is responsible for the spread of
361 ESBL-producing *E. coli* in some countries.[49, 63] The ST131 isolates in the present study
362 belonged to subclade C1-M27 (n = 3), clade A (n = 2), and subclade C2 (n = 1) (**Figure 4b**).
363 Three subclade C1-M27 isolates clustered together within the tree, while two clade A isolates
364 were distantly related. The C1-M27 isolates carried *fimH*30, LNIV-type QRDRs, and $bla_{\text{CTX-M-}}$

365 $_{27}$, which are congruent with previous studies (**Figure 4c**).[49, 64] The two clade A isolates were
366 typed as *fimH41* as previously reported,[49] but carried $bla_{\text{CTX-M-15}}$ and mutations in both *gyrA* and
367 *parC* QRDRs. Although clade A *E. coli* is generally sensitive to antibiotics,[62] clade A isolates
368 with $bla_{\text{CTX-M}}$ and *gyrA*/*parC* mutations have been sporadically reported.[65] The subclade C2
369 isolate carried *fimH*30, LNIV-type QRDRs, and $bla_{\text{CTX-M-15}}$, which are the expected traits of the
370 subclade C2 *E. coli*.[62] A variety of AMR genes were detected in our ST131 isolates, and all the
371 isolates met the ExPEC criteria (**Figure 4c**).

372 Previous studies also identified the high-risk clones ST1193 and ST131 in sewage,[17, 66, 67]
373 indicating sewage surveillance can be an effective tool to monitor the occurrence of these high-
374 risk pathogens among populations. Genomic analysis, as implemented in the present study,
375 resolved fine-scale phylogenies of these high-risk clones at the clade/subclade level (**Figure**
376 **4**), showing a combination of sewage (fecal sludge) surveillance and whole-genome
377 sequencing is even more effective and informative.

378

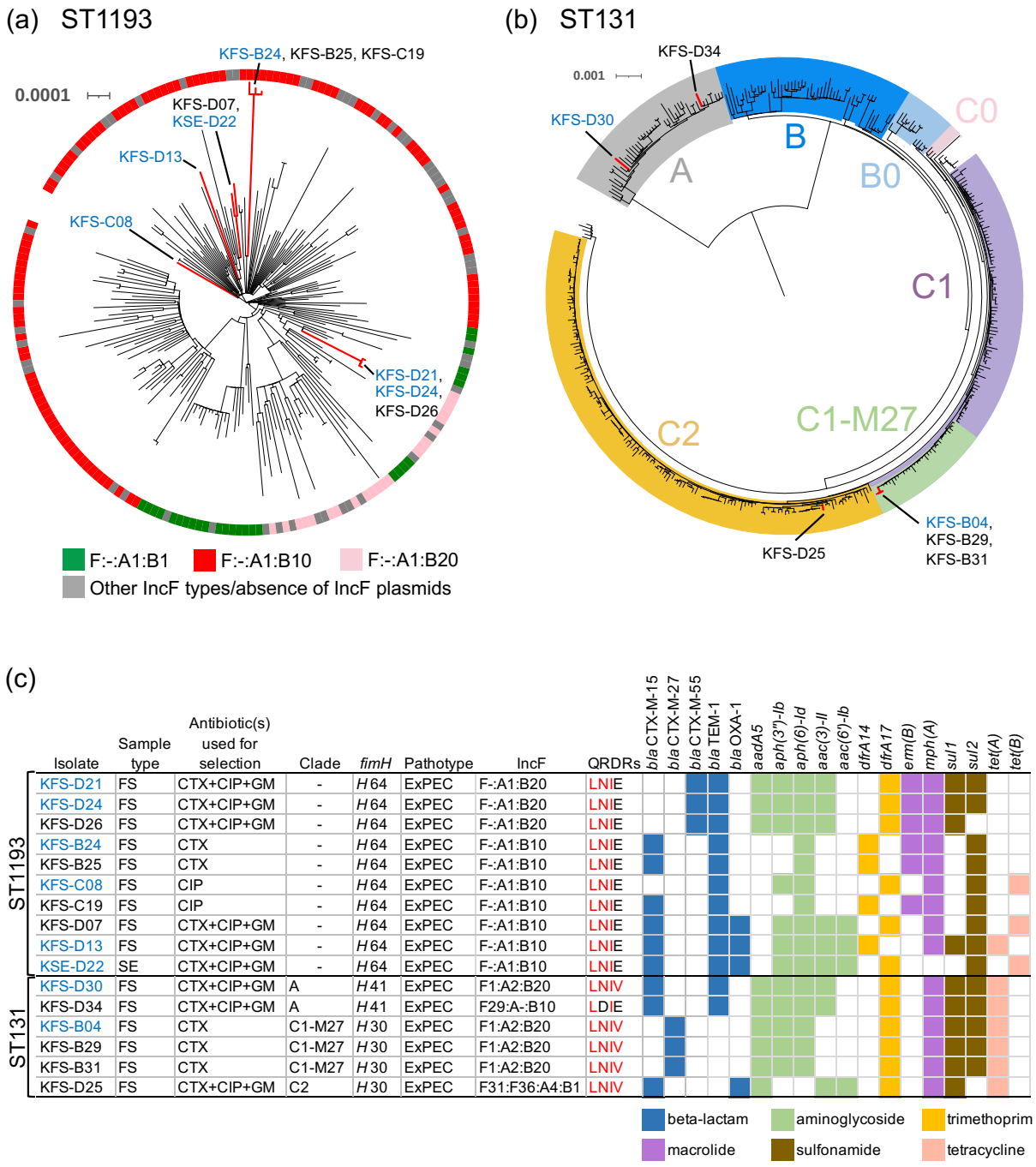


Figure 4. Recombination-free phylogenetic trees for (a) ST1193 and (b) ST131. For ST1193, we used the Assembly Dereplicator tool (v0.3.0, https://github.com/rrwick/Assembly-Dereplicator) and subsampled global isolates to remove near-identical redundant isolates and to reduce computational burden. The ST1193 tree was built using a 3,899 bp alignment

384 generated from our isolates (n = 10) and global isolates (n = 198). The ST131 tree was built
385 using a 10,712 bp alignment generated from our isolates (n = 6) and global isolates (n = 379).
386 Branches leading to our isolates are colored in red. The trees were midpoint rooted. Accession
387 numbers of global ST1193 genomes and ST131 genomes are available in **Table S4** and **Table**
388 **S5**. (c) Key genetic features of ST1193 and ST131 isolates sequenced in the present study. FS,
389 fecal sludge; SE, sewage. We did not assign clades to ST1193 isolates because the clades
390 defined by Johnson et al. are not monophyletic,[48] and the ST1193 clade definition may change
391 with larger-scale genomic analysis.[60] For QRDRs, GyrA codons 83 and 87 and ParC codons
392 80 and 84 that are different from the wild type (SDSE) are colored in red. Presence of AMR
393 genes is indicated by colored boxes (colored by drug class). All four isolates with *aac(6')-Ib*
394 carried the *aac(6')-Ib-cr* variant, which is associated with decreased susceptibility to
395 aminoglycosides and to ciprofloxacin and norfloxacin. KFS-D13 carried *aadA1*, *dfrA5*, *ere(A)*,
396 *qnrB1*, and *catA1* in addition to AMR genes shown in (c). Genomes with blue names in (a)–
397 (c) were completed.

398

399 **3.6. Locations and genetic contexts of AMR genes in isolates from major clones**

400 In total, 26 genomes were selected from the top five STs (n = 8 ST167, n = 4 ST10, n = 3 ST46,
401 n = 2 ST1284, and n = 6 ST1193) and ST131 (n = 3) and subjected to long-read sequencing to

402 elucidate the locations and genetic contexts of the AMR genes in these major STs (care was
403 taken not to select multiple closely-related isolates). Hybrid assembly using short and long
404 reads completed 22 of 26 genomes (n = 6 ST167, n = 3 ST10, n = 3 ST46, n = 2 ST1284, n =
405 6 ST1193, and n = 2 ST131). The remaining four genomes could not be completed, probably
406 due to the presence of long repeats in the genomes.

407 The completed genomes carried one to eight plasmids (median four plasmids). Among the 89
408 plasmids in the completed genomes, 38 (42.7%) carried at least one AMR gene and thus were
409 defined as AMR plasmids. IncF AMR plasmids were the most prevalent (n = 25), followed by
410 Col(pHAD28) (n = 6), IncB/O/K/Z (n = 4), IncX1 (n = 1), IncX2 (n = 1), and IncHI2-IncHI2A
411 (n = 1) (**Table 1**). We note that the completed genomes represent only a subset of our isolates,
412 and the actual AMR plasmid diversity in the whole population may be different. Nonetheless,
413 the prevalence of AMR IncF plasmids is noteworthy, and 11 (79%) of 14 plasmidic $bla_{\text{CTX-M}}$
414 genes were located on IncF plasmids. IncF plasmids are frequently detected among *E. coli* from
415 humans and can stably be maintained in commensal *E. coli* within the human gastrointestinal
416 tract without antibiotic selection pressure,[68, 69] which might explain this high prevalence. IncF
417 replicon sequence typing identified 14 different allele combinations, revealing diverse IncF
418 plasmids contributed to the spread of AMR genes within these major STs. Although AMR
419 plasmids of other replicon types were not common, two IncB/O/K/Z plasmids and one IncHI2-
420 IncHI2A plasmid contained a $bla_{\text{CTX-M}}$ gene. We found plasmids from other continents that

421 were highly similar to 13 (34%) of 38 AMR plasmids in our study under the stringent criteria
422 applied (i.e., >99% query coverage and >99% identity) (**Table1**, **Table S2**). This indicates
423 circulation of global AMR plasmids (or clones with such plasmids) in our study setting.

424 Among the 22 completed genomes, AMR genes were detected on the chromosome in eight
425 isolates (n = 3 ST167, n = 3 ST1193, n = 2 ST46). Notably, all eight isolates carried
426 chromosomal $bla_{\text{CTX-M}}$ (n = 7 $bla_{\text{CTX-M-15}}$ and n = 1 $bla_{\text{CTX-M-14}}$), and three isolates carried
427 additional chromosomal AMR genes (**Table S2**). The genetic contexts of chromosomal $bla_{\text{CTX-}}$
428 $_{\text{M}}$ genes were analyzed to determine whether different transposition events contributed to the
429 chromosomal integration of $bla_{\text{CTX-M}}$ genes within each ST (**Figure S2**). In seven of eight
430 isolates, $bla_{\text{CTX-M}}$ genes were carried on transposition units (TPU) ranging in size from 2,971
431 bp to 15,184 bp, which were flanked by different 5-bp target site duplications (TSD). In an
432 ST167 isolate KFS-D01, we could not determine the boundaries of TPU due to the lack of TSD.
433 In an ST1193 isolate KSE-D22, a 11,736 bp TPU contained four other AMR genes, and in an
434 ST1193 isolate KFS-D13 a 2,971 bp TPU was linked to four other AMR genes. A variety of
435 genetic events seem to have contributed to the diversity of the genetic contexts of chromosomal
436 $bla_{\text{CTX-M}}$, such as different transposition events of TPU (or other mobile elements containing
437 TPU) and genetic rearrangements after transposition. Detailed analyses of genetic contexts of
438 chromosomal $bla_{\text{CTX-M}}$ genes are provided in **Figure S2**. We note that recent studies have also
439 shown the frequent occurrence (more than 50%) of chromosomal $bla_{\text{CTX-M}}$ integration among

440 *E. coli* isolates using both short- and long-read sequencing.[70, 71] Chromosomal integration of
441 AMR genes might have been previously underestimated due to the inability of PCR or short-
442 read sequencing to determine the location of AMR genes. Long-read sequencing will become
443 more common in the near future, which may uncover more examples of chromosomal
444 integration of AMR genes and underline the importance of this mechanism in the stabilization
445 and maintenance of AMR genes.

446

447 **Table 1**. Basic characteristics of AMR plasmids in major STs.

Inc type (n)	Plasmid ST (n)	Chromosomal ST (n)	Size	ESBL gene (n)	No. of other AMR genes	Identical or closely related plasmids in other continents[a] (n)	Accession number(s)
Inc F[b] (25)	F-:A1:B10 (4)	ST1193 (4)	88,138 bp–108,071 bp	-	3–7	Yes (3), No (1)	CP124974, CP125028, CP125054, CP125066
	F31:F36:A4:B1 (4)	ST167 (2)	168,379 bp, 173,762 bp	$bla_{CTX-M-15}$ (2)	6–8	No (2)	CP124971, CP125044

	ST10 (1)	174,470 bp	*bla*$_{CTX-M-15}$ (1)	8	No (1)	CP125035
	ST1284 (1)	174,534 bp	*bla*$_{CTX-M-15}$ (1)	8	No (1)	CP124977
F48:A1:B49 (4)	ST167 (4)	81,682 bp–131,123 bp	*bla*$_{CTX-M-27}$ (1)	8−11	No (4)	CP124996, CP125002, CP125007, CP125040
F-:A1:B20 (2)	ST1193 (2)	90,924 bp	-	7	No (2)	CP125014, CP125021
F1:A2:B20 (2)	ST131 (clade A) (1)	146,809 bp	-	10	Yes (1)	CP125005
	ST131 (subclade C1-M27) (1)	104,558 bp	*bla*$_{CTX-M-27}$ (1)	8	Yes (1)	CP125077
F-:A-:B- (1)	ST131 (clade A) (1)	110,850 bp	*bla*$_{CTX-M-15}$ (1)	0	Yes (1)	CP125006
F-:A-:B1 (1)	ST46 (1)	97,539 bp	*bla*$_{CTX-M-27}$ (1)	10	No (1)	CP125071
F2:A-:B- (1)	ST46 (1)	73,681 bp	-	11	No (1)	CP124987
F22:F36:A1:B49 (1)	ST46 (1)	131,446 bp	-	11	No (1)	CP124991
F36:A1:B1 (1)	ST1284 (1)	122,643 bp	*bla*$_{CTX-M-15}$ (1)	9	No (1)	CP125060

	F36:A1:B20 (1)	ST10 (1)	109,736 bp	*bla*$_{CTX-M-15}$ (1)	13	No (1)	CP124993
	F40:A-:B- (1)	ST10 (1)	70,669 bp	-	5	No (1)	CP125034
	F48:A1:B1 (1)	ST10 (1)	103,762 bp	*bla*$_{CTX-M-15}$ (1)	14	No (1)	CP124985
	F77:A-:B- (1)	ST167 (1)	75,825 bp	-	7	No (1)	CP124997
Col (pHAD28) (6)	-	ST167 (3)	6,200 bp–6,222 bp	-	3	Yes (3)	CP124972, CP125041, CP125046
		ST1193 (2)	6,200 bp	-	3	Yes (2)	CP125017, CP125023
		ST1284 (1)	6,200 bp	-	3	Yes (1)	CP125063
IncB/O/K/Z (4)	-	ST1193 (3)	95,063 bp–97,447 bp	*bla*$_{CTX-M-55}$ (2)	0–3	No (3)	CP125015, CP125022, CP125065
		ST1284 (1)	95,267 bp	-	3	No (1)	CP124978
IncX1 (1)	-	ST167 (1)	37,322 bp	-	5	No (1)	CP125008
IncX2 (1)	-	ST1284 (1)	34,566 bp	-	2	No (1)	CP125061
IncHI2-IncHI2A (1)	ST1 (1)	ST1193 (1)	327,916 bp	*bla*$_{CTX-M-15}$ (1)	12	Yes (1)	CP125030

448 [a]blastn searches against the nucleotide collection (nr/nt) of the NCBI were performed to
449 identify identical/closely related plasmids (>99% query coverage and >99% identity) in May
450 2023. Also see **Table S2** for detailed information on blast hits.

451 [b]Various types of IncF replicons were detected but were represented by IncF.

452

453 **3.7. Study limitations**

454 This study has some limitations. First, the clonal composition of antimicrobial-resistant *E. coli*
455 observed in the present study might not directly reflect that of the gut flora of the underlying
456 population due to the noise caused by e.g., differential survival of *E. coli* strains in sewer
457 pipes/fecal sludge.[19] Moreover, contamination from non-human animal sources in the samples
458 cannot be excluded, though these sources should also be considered in terms of the One Health
459 concept. Inclusion of a sufficient number of fecal *E. coli* isolates from the matched human
460 population in genomic analysis will further confirm the usefulness of fecal sludge/sewage in
461 AMR surveillance. Second, we performed sampling over a relatively short period. Longitudinal
462 sampling with timeframes over one year, for example, will provide information on the change
463 of clonal composition. Third, long-read sequencing was performed only on the selected isolates
464 from major STs. Future studies employing larger-scale long-read sequencing of isolates from
465 various clones will elucidate AMR vectors also in minor STs. Nevertheless, our analyses

466 revealed detailed genomic characteristics of antimicrobial-resistant *E. coli* in fecal sludge and
467 sewage in Uganda, providing valuable insights into AMR burden in this region.

468

469 **4. Conclusions**

470 Here, we applied whole-genome sequencing to characterize antimicrobial-resistant *E. coli*
471 obtained from fecal sludge and sewage in Uganda. Our study revealed that STC10, including
472 ST167, ST10, ST1284, and ST617, were prevalent among both fecal sludge and sewage
473 isolates, indicating this lineage significantly contributed to AMR burden in the underlying
474 population. Global high-risk clones ST1193 and ST131 were also detected, adding potential
475 health risks to the population. We also revealed that many of the AMR genes were carried by
476 IncF plasmids or integrated into the chromosome by incorporating data obtained from long-
477 read sequencing. We admit that performing whole-genome sequencing in routine surveillance
478 is currently not realistic. However, as sequencing costs will continue to decrease, we believe
479 that a combination of fecal sludge/sewage surveillance and whole-genome sequencing will be
480 a powerful tool for monitoring AMR carriage in the underlying population.

481

482 **Acknowledgements**

This work was supported by the Japan Society for the Promotion of Science KAKENHI (grant numbers JP18K18881 and JP19H02274). Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics. We acknowledge the NGS core facility of the Genome Information Research Center at the Research Institute for Microbial Diseases of Osaka University for the support in DNA sequencing.

References

1. Majumder, M. A. A.; Rahman, S.; Cohall, D.; Bharatha, A.; Singh, K.; Haque, M.; Gittens-St Hilaire, M., Antimicrobial Stewardship: Fighting Antimicrobial Resistance and Protecting Global Public Health. *Infect Drug Resist* **2020,** *13*, 4713-4738.

2. WHO, Global action plan on antimicrobial resistance. **2015**.

3. Larsson, D. G. J.; Flach, C. F.; Laxminarayan, R., Sewage surveillance of antibiotic resistance holds both opportunities and challenges. *Nat Rev Microbiol* **2022**, 1-2.

4. Aarestrup, F. M.; Woolhouse, M. E. J., Using sewage for surveillance of antimicrobial resistance. *Science* **2020,** *367* (6478), 630-632.

5. Chau, K. K.; Barker, L.; Budgell, E. P.; Vihta, K. D.; Sims, N.; Kasprzyk-Hordern, B.; Harriss, E.; Crook, D. W.; Read, D. S.; Walker, A. S.; Stoesser, N., Systematic review of wastewater surveillance of antimicrobial resistance in human populations. *Environ Int* **2022,** *162*, 107171.

6. Hendriksen, R. S.; Munk, P.; Njage, P.; van Bunnik, B.; McNally, L.; Lukjancenko, O.; Roder, T.; Nieuwenhuijse, D.; Pedersen, S. K.; Kjeldgaard, J.; Kaas, R. S.; Clausen, P.; Vogt, J. K.; Leekitcharoenphon, P.; van de Schans, M. G. M.; Zuidema, T.; de Roda Husman, A. M.; Rasmussen, S.; Petersen, B.; Global Sewage Surveillance project, c.; Amid, C.; Cochrane, G.; Sicheritz-Ponten, T.; Schmitt, H.; Alvarez, J. R. M.; Aidara-Kane, A.; Pamp, S. J.; Lund, O.; Hald, T.; Woolhouse, M.; Koopmans, M. P.; Vigre, H.; Petersen, T. N.; Aarestrup, F. M., Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat Commun* **2019,** *10* (1), 1124.

7. Sims, N.; Kasprzyk-Hordern, B., Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. *Environ Int* **2020,** *139*, 105689.

514 8. Berendes, D. M.; Sumner, T. A.; Brown, J. M., Safely Managed Sanitation for All
515 Means Fecal Sludge Management for At Least 1.8 Billion People in Low and Middle Income
516 Countries. *Environ Sci Technol* **2017,** *51* (5), 3074-3083.
517 9. Chigwechokha, P.; Street, R.; Holm, R. H., Advancing the Use of Fecal Sludge for
518 Timelier and Better-Quality Epidemiological Data in Low- and Middle-Income Countries for
519 Pandemic Prevention. *Environ Sci Technol* **2022**.
520 10. Capone, D.; Berendes, D.; Cumming, O.; Knee, J.; Nala, R.; Risk, B. B.;
521 Stauber, C.; Zhu, K.; Brown, J., Analysis of Fecal Sludges Reveals Common Enteric
522 Pathogens in Urban Maputo, Mozambique. *Environ Sci Tech Let* **2020,** *7* (12), 889-895.
523 11. Capone, D.; Chigwechokha, P.; de Los Reyes, F. L., 3rd; Holm, R. H.; Risk,
524 B. B.; Tilley, E.; Brown, J., Impact of sampling depth on pathogen detection in pit latrines.
525 *PLoS Negl Trop Dis* **2021,** *15* (3), e0009176.
526 12. Denamur, E.; Clermont, O.; Bonacorsi, S.; Gordon, D., The population genetics of
527 pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2021,** *19* (1), 37-54.
528 13. Poirel, L.; Madec, J. Y.; Lupo, A.; Schink, A. K.; Kieffer, N.; Nordmann, P.;
529 Schwarz, S., Antimicrobial Resistance in *Escherichia coli*. *Microbiol Spectr* **2018,** *6* (4).
530 14. Anjum, M. F.; Schmitt, H.; Borjesson, S.; Berendonk, T. U.; network, W., The
531 potential of using E. coli as an indicator for the surveillance of antimicrobial resistance (AMR)
532 in the environment. *Curr Opin Microbiol* **2021,** *64*, 152-158.
533 15. Reinthaler, F. F.; Galler, H.; Feierl, G.; Haas, D.; Leitner, E.; Mascher, F.;
534 Melkes, A.; Posch, J.; Pertschy, B.; Winter, I.; Himmel, W.; Marth, E.; Zarfel, G.,
535 Resistance patterns of *Escherichia coli* isolated from sewage sludge in comparison with those
536 isolated from human patients in 2000 and 2009. *J Water Health* **2013,** *11* (1), 13-20.
537 16. Kwak, Y. K.; Colque, P.; Byfors, S.; Giske, C. G.; Mollby, R.; Kuhn, I.,
538 Surveillance of antimicrobial resistance among *Escherichia coli* in wastewater in Stockholm
539 during 1 year: does it reflect the resistance trends in the society? *Int J Antimicrob Agents* **2015,**
540 *45* (1), 25-32.
541 17. Raven, K. E.; Ludden, C.; Gouliouris, T.; Blane, B.; Naydenova, P.; Brown,
542 N. M.; Parkhill, J.; Peacock, S. J., Genomic surveillance of *Escherichia coli* in municipal
543 wastewater treatment plants as an indicator of clinically relevant pathogens and their resistance
544 genes. *Microb Genom* **2019,** *5* (5).
545 18. Hutinel, M.; Huijbers, P. M. C.; Fick, J.; Ahren, C.; Larsson, D. G. J.; Flach,
546 C. F., Population-level surveillance of antibiotic resistance in *Escherichia coli* through sewage
547 analysis. *Euro Surveill* **2019,** *24* (37).
548 19. Huijbers, P. M. C.; Larsson, D. G. J.; Flach, C. F., Surveillance of antibiotic resistant
549 *Escherichia coli* in human populations through urban wastewater in ten European countries.
550 *Environ Pollut* **2020,** *261*, 114200.

551 20. Grevskott, D. H.; Ghavidel, F. Z.; Svanevik, C. S.; Marathe, N. P., Resistance
552 profiles and diversity of beta-lactamases in *Escherichia coli* strains isolated from city-scale
553 sewage surveillance in Bergen, Norway mimic clinical prevalence. *Ecotoxicol Environ Saf*
554 **2021,** *226*, 112788.
555 21. WHO, WHO integrated global surveillance on ESBL-producing *E. coli* using a "One
556 Health" approach: implementation and opportunities. **2021**.
557 22. Nabadda, S.; Kakooza, F.; Kiggundu, R.; Walwema, R.; Bazira, J.; Mayito,
558 J.; Mugerwa, I.; Sekamatte, M.; Kambugu, A.; Lamorde, M.; Kajumbula, H.;
559 Mwebasa, H., Implementation of the World Health Organization Global Antimicrobial
560 Resistance Surveillance System in Uganda, 2015-2020: Mixed-Methods Study Using National
561 Surveillance Data. *JMIR Public Health Surveill* **2021,** *7* (10), e29954.
562 23. Katongole, P.; Nalubega, F.; Florence, N. C.; Asiimwe, B.; Andia, I., Biofilm
563 formation, antimicrobial susceptibility and virulence genes of Uropathogenic *Escherichia coli*
564 isolated from clinical isolates in Uganda. *BMC Infect Dis* **2020,** *20* (1), 453.
565 24. Bebell, L. M.; Ngonzi, J.; Bazira, J.; Fajardo, Y.; Boatin, A. A.; Siedner, M.
566 J.; Bassett, I. V.; Nyehangane, D.; Nanjebe, D.; Jacquemyn, Y.; van Geertruyden, J.
567 P.; Mwanga-Amumpaire, J.; Bangsberg, D. R.; Riley, L. E.; Boum, Y., 2nd,
568 Antimicrobial-resistant infections among postpartum women at a Ugandan referral hospital.
569 *PLoS One* **2017,** *12* (4), e0175456.
570 25. Odongo, I.; Ssemambo, R.; Kungu, J. M., Prevalence of *Escherichia Coli* and Its
571 Antimicrobial Susceptibility Profiles among Patients with UTI at Mulago Hospital, Kampala,
572 Uganda. *Interdiscip Perspect Infect Dis* **2020,** *2020*, 8042540.
573 26. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J., fastp: an ultra-fast all-in-one FASTQ
574 preprocessor. *Bioinformatics* **2018,** *34* (17), i884-i890.
575 27. Wick, R. R.; Judd, L. M.; Gorrie, C. L.; Holt, K. E., Unicycler: Resolving bacterial
576 genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **2017,** *13* (6),
577 e1005595.
578 28. Zhou, Z.; Alikhan, N. F.; Mohamed, K.; Fan, Y.; Agama Study, G.; Achtman,
579 M., The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis*
580 phylogeny, and *Escherichia* core genomic diversity. *Genome Res* **2020,** *30* (1), 138-152.
581 29. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P. A., Assembly of long, error-prone
582 reads using repeat graphs. *Nat Biotechnol* **2019,** *37* (5), 540-546.
583 30. Wick, R. R.; Holt, K. E., Polypolish: Short-read polishing of long-read bacterial
584 genome assemblies. *PLoS Comput Biol* **2022,** *18* (1), e1009802.
585 31. Wick, R. R.; Judd, L. M.; Holt, K. E., Assembling the perfect bacterial genome using
586 Oxford Nanopore and Illumina sequencing. *PLoS Comput Biol* **2023,** *19* (3), e1010905.

587 32. Wadley, T. D.; Jenjaroenpun, P.; Wongsurawat, T.; Ussery, D. W.; Nookaew, I.,
588 Complete Genome and Plasmid Sequences of *Escherichia coli* Type Strain ATCC 11775.
589 *Microbiol Resour Announc* **2019,** *8* (9).
590 33. Jain, C.; Rodriguez, R. L.; Phillippy, A. M.; Konstantinidis, K. T.; Aluru, S.,
591 High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
592 *Nat Commun* **2018,** *9* (1), 5114.
593 34. Waters, N. R.; Abram, F.; Brennan, F.; Holmes, A.; Pritchard, L., Easy
594 phylotyping of *Escherichia coli* via the EzClermont web app and command-line tool. *Access*
595 *Microbiol* **2020,** *2* (9), acmi000143.
596 35. Beghain, J.; Bridier-Nahmias, A.; Le Nagard, H.; Denamur, E.; Clermont, O.,
597 ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain
598 phylotyping. *Microb Genom* **2018,** *4* (7).
599 36. Roer, L.; Tchesnokova, V.; Allesoe, R.; Muradova, M.; Chattopadhyay, S.;
600 Ahrenfeldt, J.; Thomsen, M. C. F.; Lund, O.; Hansen, F.; Hammerum, A. M.;
601 Sokurenko, E.; Hasman, H., Development of a Web Tool for *Escherichia coli* Subtyping Based
602 on *fimH* Alleles. *J Clin Microbiol* **2017,** *55* (8), 2538-2543.
603 37. Bortolaia, V.; Kaas, R. S.; Ruppe, E.; Roberts, M. C.; Schwarz, S.; Cattoir,
604 V.; Philippon, A.; Allesoe, R. L.; Rebelo, A. R.; Florensa, A. F.; Fagelhauer, L.;
605 Chakraborty, T.; Neumann, B.; Werner, G.; Bender, J. K.; Stingl, K.; Nguyen, M.;
606 Coppens, J.; Xavier, B. B.; Malhotra-Kumar, S.; Westh, H.; Pinholt, M.; Anjum, M.
607 F.; Duggett, N. A.; Kempf, I.; Nykasenoja, S.; Olkkola, S.; Wieczorek, K.; Amaro,
608 A.; Clemente, L.; Mossong, J.; Losch, S.; Ragimbeau, C.; Lund, O.; Aarestrup, F. M.,
609 ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* **2020,**
610 *75* (12), 3491-3500.
611 38. Aziz, R. K.; Bartels, D.; Best, A. A.; DeJongh, M.; Disz, T.; Edwards, R. A.;
612 Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; Meyer, F.; Olsen, G. J.; Olson,
613 R.; Osterman, A. L.; Overbeek, R. A.; McNeil, L. K.; Paarmann, D.; Paczian, T.;
614 Parrello, B.; Pusch, G. D.; Reich, C.; Stevens, R.; Vassieva, O.; Vonstein, V.;
615 Wilke, A.; Zagnitko, O., The RAST Server: rapid annotations using subsystems technology.
616 *BMC Genomics* **2008,** *9*, 75.
617 39. Siguier, P.; Perochon, J.; Lestrade, L.; Mahillon, J.; Chandler, M., ISfinder: the
618 reference centre for bacterial insertion sequences. *Nucleic Acids Res* **2006,** *34* (Database issue),
619 D32-6.
620 40. Aoike, N.; Saga, T.; Sakata, R.; Yoshizumi, A.; Kimura, S.; Iwata, M.;
621 Yoshizawa, S.; Sugasawa, Y.; Ishii, Y.; Yamaguchi, K.; Tateda, K., Molecular
622 characterization of extraintestinal *Escherichia coli* isolates in Japan: relationship between
623 sequence types and mutation patterns of quinolone resistance-determining regions analyzed by
624 pyrosequencing. *J Clin Microbiol* **2013,** *51* (6), 1692-8.

625 41. Joensen, K. G.; Scheutz, F.; Lund, O.; Hasman, H.; Kaas, R. S.; Nielsen, E.
626 M.; Aarestrup, F. M., Real-time whole-genome sequencing for routine typing, surveillance,
627 and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol* **2014,** *52* (5), 1501-
628 10.
629 42. Malberg Tetzschner, A. M.; Johnson, J. R.; Johnston, B. D.; Lund, O.; Scheutz,
630 F., *In Silico* Genotyping of *Escherichia coli* Isolates for Extraintestinal Virulence Genes by
631 Use of Whole-Genome Sequencing Data. *J Clin Microbiol* **2020,** *58* (10).
632 43. Gomi, R.; Matsuda, T.; Fujimori, Y.; Harada, H.; Matsui, Y.; Yoneda, M.,
633 Characterization of Pathogenic *Escherichia coli* in River Water by Simultaneous Detection and
634 Sequencing of 14 Virulence Genes. *Environ Sci Technol* **2015,** *49* (11), 6800-7.
635 44. Johnson, J. R.; Stell, A. L., Extended virulence genotypes of *Escherichia coli* strains
636 from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* **2000,**
637 *181* (1), 261-72.
638 45. Carattoli, A.; Zankari, E.; Garcia-Fernandez, A.; Voldby Larsen, M.; Lund,
639 O.; Villa, L.; Moller Aarestrup, F.; Hasman, H., *In silico* detection and typing of plasmids
640 using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*
641 **2014,** *58* (7), 3895-903.
642 46. Gardner, S. N.; Slezak, T.; Hall, B. G., kSNP3.0: SNP detection and phylogenetic
643 analysis of genomes without genome alignment or reference genome. *Bioinformatics* **2015,** *31*
644 (17), 2877-8.
645 47. Treangen, T. J.; Ondov, B. D.; Koren, S.; Phillippy, A. M., The Harvest suite for
646 rapid core-genome alignment and visualization of thousands of intraspecific microbial
647 genomes. *Genome Biol* **2014,** *15* (11), 524.
648 48. Johnson, T. J.; Elnekave, E.; Miller, E. A.; Munoz-Aguayo, J.; Flores
649 Figueroa, C.; Johnston, B.; Nielson, D. W.; Logue, C. M.; Johnson, J. R., Phylogenomic
650 Analysis of Extraintestinal Pathogenic *Escherichia coli* Sequence Type 1193, an Emerging
651 Multidrug-Resistant Clonal Group. *Antimicrob Agents Chemother* **2019,** *63* (1).
652 49. Matsumura, Y.; Pitout, J. D. D.; Peirano, G.; DeVinney, R.; Noguchi, T.;
653 Yamamoto, M.; Gomi, R.; Matsuda, T.; Nakano, S.; Nagao, M.; Tanaka, M.;
654 Ichiyama, S., Rapid Identification of Different *Escherichia coli* Sequence Type 131 Clades.
655 *Antimicrob Agents Chemother* **2017,** *61* (8).
656 50. Nielsen, D. W.; Ricker, N.; Barbieri, N. L.; Wynn, J. L.; Gomez-Duarte, O.
657 G.; Iqbal, J.; Nolan, L. K.; Allen, H. K.; Logue, C. M., Complete Genome Sequence of
658 the Multidrug-Resistant Neonatal Meningitis *Escherichia coli* Serotype O75:H5:K1 Strain
659 mcjchv-1 (NMEC-O75). *Microbiol Resour Announc* **2018,** *7* (10).
660 51. Forde, B. M.; Ben Zakour, N. L.; Stanton-Cook, M.; Phan, M. D.; Totsika, M.;
661 Peters, K. M.; Chan, K. G.; Schembri, M. A.; Upton, M.; Beatson, S. A., The complete
662 genome sequence of *Escherichia coli* EC958: a high quality reference sequence for the globally

663 disseminated multidrug resistant *E. coli* O25b:H4-ST131 clone. *PLoS One* **2014,** *9* (8),
664 e104400.
665 52. Croucher, N. J.; Page, A. J.; Connor, T. R.; Delaney, A. J.; Keane, J. A.;
666 Bentley, S. D.; Parkhill, J.; Harris, S. R., Rapid phylogenetic analysis of large samples of
667 recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **2015,** *43* (3),
668 e15.
669 53. Minh, B. Q.; Schmidt, H. A.; Chernomor, O.; Schrempf, D.; Woodhams, M.
670 D.; von Haeseler, A.; Lanfear, R., IQ-TREE 2: New Models and Efficient Methods for
671 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **2020,** *37* (5), 1530-1534.
672 54. Letunic, I.; Bork, P., Interactive Tree Of Life (iTOL) v5: an online tool for
673 phylogenetic tree display and annotation. *Nucleic Acids Res* **2021,** *49* (W1), W293-W296.
674 55. Kohler, C. D.; Dobrindt, U., What defines extraintestinal pathogenic *Escherichia coli*?
675 *Int J Med Microbiol* **2011,** *301* (8), 642-7.
676 56. Decano, A. G.; Pettigrew, K.; Sabiiti, W.; Sloan, D. J.; Neema, S.; Bazira,
677 J.; Kiiru, J.; Onyango, H.; Asiimwe, B.; Holden, M. T. G., Pan-Resistome
678 Characterization of Uropathogenic *Escherichia coli* and *Klebsiella pneumoniae* Strains
679 Circulating in Uganda and Kenya, Isolated from 2017-2018. *Antibiotics (Basel)* **2021,** *10* (12).
680 57. van der Putten, B. C. L.; Remondini, D.; Pasquini, G.; Janes, V. A.;
681 Matamoros, S.; Schultsz, C., Quantifying the contribution of four resistance mechanisms to
682 ciprofloxacin MIC in *Escherichia coli*: a systematic review. *J Antimicrob Chemother* **2019,** *74*
683 (2), 298-310.
684 58. Matsumura, Y.; Noguchi, T.; Tanaka, M.; Kanahashi, T.; Yamamoto, M.;
685 Nagao, M.; Takakura, S.; Ichiyama, S.; th, J. B., Population structure of Japanese
686 extraintestinal pathogenic *Escherichia coli* and its relationship with antimicrobial resistance. *J*
687 *Antimicrob Chemother* **2017,** *72* (4), 1040-1049.
688 59. Reid, C. J.; DeMaere, M. Z.; Djordjevic, S. P., Australian porcine clonal complex
689 10 (CC10) *Escherichia coli* belong to multiple sublineages of a highly diverse global CC10
690 phylogeny. *Microb Genom* **2019,** *5* (3).
691 60. Pitout, J. D. D.; Peirano, G.; Chen, L.; DeVinney, R.; Matsumura, Y.,
692 *Escherichia coli* ST1193: Following in the Footsteps of *E. coli* ST131. *Antimicrob Agents*
693 *Chemother* **2022,** *66* (7), e0051122.
694 61. Wyrsch, E. R.; Bushell, R. N.; Marenda, M. S.; Browning, G. F.; Djordjevic, S.
695 P., Global Phylogeny and F Virulence Plasmid Carriage in Pandemic *Escherichia coli* ST1193.
696 *Microbiol Spectr* **2022,** *10* (6), e0255422.
697 62. Pitout, J. D. D.; Finn, T. J., The evolutionary puzzle of *Escherichia coli* ST131. *Infect*
698 *Genet Evol* **2020,** *81*, 104265.
699 63. Merino, I.; Hernandez-Garcia, M.; Turrientes, M. C.; Perez-Viso, B.; Lopez-
700 Fresnena, N.; Diaz-Agero, C.; Maechler, F.; Fankhauser-Rodriguez, C.; Kola, A.;

701 Schrenzel, J.; Harbarth, S.; Bonten, M.; Gastmeier, P.; Canton, R.; Ruiz-Garbajosa,
702 P.; Group, R. G. S., Emergence of ESBL-producing *Escherichia coli* ST131-C1-M27 clade
703 colonizing patients in Europe. *J Antimicrob Chemother* **2018,** *73* (11), 2973-2980.
704 64. Matsumura, Y.; Pitout, J. D.; Gomi, R.; Matsuda, T.; Noguchi, T.;
705 Yamamoto, M.; Peirano, G.; DeVinney, R.; Bradford, P. A.; Motyl, M. R.; Tanaka,
706 M.; Nagao, M.; Takakura, S.; Ichiyama, S., Global *Escherichia coli* Sequence Type 131
707 Clade with $bla_{\text{CTX-M-27}}$ Gene. *Emerg Infect Dis* **2016,** *22* (11), 1900-1907.
708 65. Stoesser, N.; Sheppard, A. E.; Pankhurst, L.; De Maio, N.; Moore, C. E.;
709 Sebra, R.; Turner, P.; Anson, L. W.; Kasarskis, A.; Batty, E. M.; Kos, V.; Wilson,
710 D. J.; Phetsouvanh, R.; Wyllie, D.; Sokurenko, E.; Manges, A. R.; Johnson, T. J.;
711 Price, L. B.; Peto, T. E.; Johnson, J. R.; Didelot, X.; Walker, A. S.; Crook, D. W.;
712 Modernizing Medical Microbiology Informatics, G., Evolutionary History of the Global
713 Emergence of the *Escherichia coli* Epidemic Clone ST131. *mBio* **2016,** *7* (2), e02162.
714 66. Gomi, R.; Matsuda, T.; Matsumura, Y.; Yamamoto, M.; Tanaka, M.;
715 Ichiyama, S.; Yoneda, M., Occurrence of Clinically Important Lineages, Including the
716 Sequence Type 131 C1-M27 Subclone, among Extended-Spectrum-beta-Lactamase-Producing
717 *Escherichia coli* in Wastewater. *Antimicrob Agents Chemother* **2017,** *61* (9).
718 67. Hoelle, J.; Johnson, J. R.; Johnston, B. D.; Kinkle, B.; Boczek, L.; Ryu, H.;
719 Hayes, S., Survey of US wastewater for carbapenem-resistant *Enterobacteriaceae*. *J Water*
720 *Health* **2019,** *17* (2), 219-226.
721 68. Bevan, E. R.; Jones, A. M.; Hawkey, P. M., Global epidemiology of CTX-M beta-
722 lactamases: temporal and geographical shifts in genotype. *J Antimicrob Chemother* **2017,** *72*
723 (8), 2145-2155.
724 69. Rozwandowicz, M.; Brouwer, M. S. M.; Fischer, J.; Wagenaar, J. A.;
725 Gonzalez-Zorn, B.; Guerra, B.; Mevius, D. J.; Hordijk, J., Plasmids carrying antimicrobial
726 resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* **2018,** *73* (5), 1121-1137.
727 70. Gomi, R.; Yamamoto, M.; Tanaka, M.; Matsumura, Y., Chromosomal integration
728 of $bla_{\text{CTX-M}}$ genes in diverse *Escherichia coli* isolates recovered from river water in Japan. *Curr*
729 *Res Microb Sci* **2022,** *3*, 100144.
730 71. Milenkov, M.; Proux, C.; Rasolofoarison, T. L.; Rakotomalala, F. A.;
731 Rasoanandrasana, S.; Rahajamanana, L. V.; Rafalimanana, C.; Ravaoarisaina, Z.;
732 Ramahatafandry, I. T.; Westeel, E.; Petitjean, M.; Marin, J.; Mullaert, J.; Han, L.;
733 Clermont, O.; Raskine, L.; Endtz, H.; Andremont, A.; Denamur, E.; Komurian-
734 Pradel, F.; Samison, L. H.; Armand-Lefevre, L., Tricycle surveillance in Antananarivo,
735 Madagascar: circulation of both extended-spectrum beta-lactamase producing *Escherichia coli*
736 strains and plasmids among humans, chickens and the environment. *medRXiv* **2023**.

737