

1 **Title. An ST131 clade and a Phylogroup A clade bearing a novel *Escherichia coli* O-antigen**
2 **cluster predominate among bloodstream *E. coli* isolates from southwest Nigeria hospitals.**

3
4 Ayorinde O. Afolayan¹, Aaron O. Aboderin², Anderson O. Oaikhen¹, Erkison Ewomazino
5 Odih¹, Veronica O. Ogunleye³, Adeyemi T. Adeyemo⁴, Abolaji T. Adeyemo⁵, Oyeniyi S.
6 Bejide¹, Anthony Underwood⁶, Silvia Argimón⁶, Monica Abrudan⁶, Abiodun Egwuenu⁷, Chikwe
7 Ihekweazu⁷, David M. Aanensen⁶ and Iruka N. Okeke^{1*}.

8
9 ¹Global Health Research Unit for the Genomic Surveillance of Antimicrobial Resistance,
10 Department of Pharmaceutical Microbiology, Faculty of Pharmacy, University of Ibadan, Oyo
11 State, Nigeria.

12 ²Department of Medical Microbiology and Parasitology, Obafemi Awolowo University Teaching
13 Hospitals Complex, Ile-Ife, Osun State, Nigeria.

14 ³Department of Medical Microbiology and Parasitology, University College Hospital, Ibadan,
15 Oyo State, Nigeria.

16 ⁴Department of Medical Microbiology and Parasitology, Obafemi Awolowo University Teaching
17 Hospitals Complex, Ile-Ife, Osun State, Nigeria.

18 ⁵Department of Medical Microbiology and Parasitology, University of Osun Teaching Hospital,
19 Osogbo, Osun State, Nigeria.

20 ⁶Centre for Genomic Pathogen Surveillance, Big Data Institute, University of Oxford, Old Road
21 Campus, Oxford, United Kingdom and Wellcome Genome Campus, Hinxton, UK.

22 ⁷Nigeria Centre for Disease Control, Jabi, Abuja, Nigeria.

23
24 ***Corresponding author:** Iruka N. Okeke. Department of Pharmaceutical Microbiology,
25 Faculty of Pharmacy, University of Ibadan, Oyo State, Nigeria. +234 805 328 1714.
26 iruka.n.okeke@gmail.com

27 **Abstract**

28 *Escherichia coli* bloodstream infections are typically attributed to a limited number of lineages
29 that carry virulence factors associated with invasion and, in recent years, invasive *E. coli* are
30 increasingly multiply antimicrobial resistant. In Nigeria, *E. coli* is a common cause of
31 bloodstream infections but the identity of circulating clones is largely unknown and surveillance
32 of their antimicrobial resistance has been limited. We verified and whole genome-sequenced 68
33 bloodstream *E. coli* isolates recovered between 2016 and 2018 at three sentinel sites in
34 southwestern Nigeria and susceptibility tested 67 of them. Resistance to antimicrobials
35 commonly used in Nigeria was high, with 67(100%), 62 (92.5%), 53 (79%) and 37(55%)
36 showing resistance to trimethoprim, ampicillin, ciprofloxacin and aminoglycosides, respectively.
37 All the isolates were susceptible to carbapenems and colistin. The strain set included isolates
38 from globally disseminated high risk clones including those belonging to ST12 (n=2), ST131
39 (n=12) and ST648 (n = 4). Twenty-three (33.82%) of the isolates clustered within two clades.
40 The first of these consisted of ST131 strains, comprised of O16:H5 and O25:H4 sub-lineages.
41 The second was an ST10-ST167 complex clade comprised of strains carrying capsular genes that
42 may have originated in *Klebsiella*. We additionally determined that four temporally-associated
43 ST90 strains from one sentinel were closely related enough to suggest that at least some of them
44 represented a retrospectively detected outbreak cluster. Our data demonstrate that a broad
45 repertoire of invasive *E. coli* isolates cause bloodstream infections in southwest Nigeria. In
46 addition to pandemic lineages, particularly ST131, these include a previously undescribed
47 lineage carrying an O-antigen cluster previously only reported from *Klebsiella*. Genomic
48 surveillance is valuable for tracking these and other clones and for outbreak identification.

49
50 **Keywords:** *Escherichia coli*, Genomic Surveillance, ST131, antimicrobial resistance, *cpsACP*,
51 Nigeria

53 **Data Summary**

54 Phylogenetic tree, clinical data, and epidemiological data were visualized using Microreact
55 ([https://microreact.org/project/hmj3KwxS1dmmFPCKFx6qeA-invasive-escherichia-coli-sw-](https://microreact.org/project/hmj3KwxS1dmmFPCKFx6qeA-invasive-escherichia-coli-sw-nigeria-2016-2018)
56 [nigeria-2016-2018](https://microreact.org/project/hmj3KwxS1dmmFPCKFx6qeA-invasive-escherichia-coli-sw-nigeria-2016-2018)). All the sequence data have been deposited in the ENA under the project ID
57 PRJEB29739 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB29739>). Accessions can be found
58 in Table S6.

59
60

61 **Introduction**

62 Extra-intestinal pathogenic *Escherichia coli* (ExPEC) are responsible for the majority of
63 infections in the blood, urine, meninges, prostate, and other normally sterile sites [1–3].
64 Although typically initially acquired within the gastrointestinal tract (and less commonly via
65 genital route), these strains of *E. coli* differ from commensal and diarrhoeogenic *E. coli* in
66 possession of factors associated with systemic virulence [4, 5], allowing them to survive in
67 different extra-intestinal niches. Some of the virulence genes associated with ExPEC include
68 adhesins (*fim*, *pap*, *sfa*, *afa*), invasins (*ibeA*), iron acquisition genes (*ybt*, *iro*, *iuc*), toxins (*clb*,
69 *cnf*, *hly*, SPATE genes) and protectins (*traT*, *ompT*, *kpsMT*), among others [4, 5]. ExPEC also
70 often have K, O and H antigens that make them recognizable to *E. coli* experts and, in some
71 instances, assist them in evading immune responses. Virulence and AMR determinants, as well
72 as negative frequency-dependent selection, likely influence the stability of the most dominant
73 ExPEC groups, which are often multiply resistant [6], thereby sustaining the occurrence of extra-
74 intestinal diseases globally. In Africa, available data reveal that there is an upward trend in the
75 prevalence of globally dominant ExPEC lineages in humans [7] and animals [8], painting a grim
76 picture for disease and antimicrobial resistance.

77 Robust surveillance is urgently needed to tackle antimicrobial resistance in a more robust and
78 consistent manner within each country. Whole genome sequencing (WGS) has helped to provide
79 a clearer picture on the epidemiology of infectious diseases caused by ExPEC and has identified
80 a number of pandemic clones. Incorporation of WGS with existing epidemiological frameworks
81 in national public health institutes is critical for providing genomic context for prospective
82 surveillance and for designing and implementing AMR-eliminating strategies.

83 Although studies conducted in Africa and other low- and middle-income countries (LMICs) have
84 shown the abundance of invasive *E. coli* [9, 10], these studies are too few and far between so that
85 ExPEC and AMR epidemiology are poorly understood [7, 11]. In Nigeria, there is sparse
86 molecular information on ExPEC, but a few studies point to likely clonal expansion of resistant
87 lineages and local presence of pandemic clones of concern [12–16]. These studies provide
88 valuable information but represent an insufficient picture of ExPEC clones in Nigeria with few
89 data available from the South. To extend the body of knowledge on genomic epidemiology of
90 ExPEC in south western Nigeria, including AMR patterns and mechanisms, we leveraged on our

91 existing genomic surveillance of bacterial AMR efforts by characterizing the genomes of
92 invasive isolates from tertiary hospitals in South-west Nigeria.

93

94 **Methodology**

95

96 **Species Validation and Antimicrobial Susceptibility Tests**

97 Sentinel hospital laboratories referred anonymized clinical data and bloodstream *E. coli* isolates
98 collected between the years 2016 to 2018 to the National Reference Laboratory for verification
99 of isolate identity and antimicrobial susceptibility tests (AST). Isolate identity and AST were
100 achieved using the VITEK2 Instrument. Drugs tested include amikacin (AMK), gentamicin
101 (GEN), ampicillin (AMP), amoxicillin/clavulanic acid (AMC), piperacillin/tazobactam (TZP),
102 cefuroxime (CXM), cefuroxime axetil (CXMA), cefepime (FEP), ceftriaxone (CRO),
103 cefoperazone/sulbactam (SFP), nitrofurantoin (NIT), nalidixic acid (NAL), ciprofloxacin (CIP),
104 ertapenem (ERT), imipenem (IPM), meropenem (MEM), and trimethoprim-sulfamethoxazole
105 (SXT). Resistance profiles were generated from VITEK2 AST data. Antibiotic susceptibility
106 results were interpreted in line with the CLSI 2019 standards [17]. Minimum Inhibitory
107 Concentration values were converted to RIS values and the bug-drug combination table was
108 generated using the AMR R package (v1.4.0; <https://github.com/msberends/AMR>) [18].

109

110 **Biofilm Assay**

111 Following Wakimoto's procedure [19], we sub-cultured invasive *E. coli* strains overnight in LB
112 broth at 37°C with shaking at 160 rpm. Afterwards, we measured 190 µL of Dulbecco's
113 Modified Eagle's Medium (DMEM) containing 0.45 % glucose using a pipette into each well of
114 a 96-well plate (different plates were used for each time point (3h, 6h, 12h, 24 h)). Five µL of
115 each isolate was inoculated in triplicate per time point into a 96-well plate from the stock plate
116 and incubated at 37°C without shaking until each time point was reached. Absorbance at 595 nm
117 was taken on completion of time point.

118 We pipetted spent media out of the 96-well plates. Each well was washed with PBS three times,
119 fixed (10 mins, 200 µL of 75% ethanol), dried, and stained with 195 µL of 0.5 % crystal violet

120 for 5 minutes. This was followed by washing and drying of the plates. We added 200 μ L of 95%
121 ethanol to each well, allowed the wells to stand for 20 minutes at room temperature, and
122 determined absorbance with an enzyme linked immunosorbent assay (ELISA) plate reader at 570
123 nm. Biofilm index was defined using the average of the values for the optical density (OD) at
124 570 nm and 595 nm, and was calculated by dividing the OD values for each strain at a given
125 timepoint by the OD values of the negative control at the given timepoint [20].
126 Enteroaggregative *E. coli* strain 042 was used as positive control while *E. coli* K-12 DH5 α was
127 used as negative control. The magnitude of biofilm formed by different genetically defined
128 subgroups of isolates was compared using a two-tailed Mann-Whitney test at $p < 0.05$.

129

130 **Haemolysis Test**

131 Overnight LB cultures of the isolates were spotted onto the surface of the blood agar. Alpha
132 haemolysis and beta haemolysis is indicated by a green colouration and a clear zone around
133 bacteria colonies, respectively. Uropathogenic *E. coli* isolate ATCC 11175 and *stx*-deleted *E.*
134 *coli* O157 isolate ATCC 7000728 were used as controls.

135

136 **DNA Extraction, Library Preparation, and Whole Genome Sequencing**

137 Genomic DNA was extracted following the protocol outlined in a previous report [21]. Briefly,
138 the Wizard DNA extraction kit (Promega; Wisconsin, USA; Cat. No: A1125) was used in the
139 extraction of genomic DNA of all isolates. A dsDNA Broad Range quantification assay was used
140 in the quantification of DNA extracts (Invitrogen; California, USA; Cat. No: Q32853). DNA
141 libraries were prepared and sequenced using the NEBNext Ultra II FS DNA library kit (New
142 England Biolabs, Massachusetts, USA; Cat. No: E6617L) and Illumina HiSeq X10 instrument
143 (Illumina, CA, USA), respectively.

144

145 **Genome Assembly and Speciation**

146 Raw sequence reads were assembled using the GHRU pipeline
147 (<https://gitlab.com/cgps/ghru/pipelines/assembly>) which is summarily explained by the GHRU

148 *de novo* assembly protocol [22] . Speciation was possible using Bactinspector (v0.1.3;
149 <https://gitlab.com/antunderwood/bactinspector/>), implemented within the GHRU pipeline.

150

151 **Single Nucleotide Polymorphism (SNP) Analysis and Phylogenetic Tree Generation**

152 The complete chromosome sequence of *Escherichia coli* strain EC958 (accession
153 GCF_000285655.3) (https://www.ncbi.nlm.nih.gov/assembly/GCF_000285655.3/) was used to
154 infer a whole-genome alignment of the sequence reads and identify SNP positions, which were,
155 in turn used to infer a maximum-likelihood phylogenetic tree as per the GHRU mapping-based
156 phylogeny protocol [22] which summarizes the steps implemented within the GHRU SNP
157 phylogeny pipeline (https://gitlab.com/cgps/ghru/pipelines/snp_phylogeny).

158 Isolates that clustered closely within the phylogenetic tree, belonged to the same ST, and shared
159 similar resistance profiles, plasmid profiles, geography and dates of isolation, were investigated
160 further for clue on potential outbreaks. We selected the closest reference genome to the outbreak
161 isolates using Bactinspector, aligned the outbreak isolates (ST90) to the reference genome of *E.*
162 *coli* strain D3 (accession NZ_CP010140.1;
163 https://www.ncbi.nlm.nih.gov/assembly/GCF_001900635.1/), used Gubbins [23] to mask
164 recombinant regions, and calculated pairwise SNP distances from the pseudo-genome alignment
165 using FastaDist (v0.0.7; <https://gitlab.com/antunderwood/fastadist>), SNP-dists (v0.8.2;
166 <https://github.com/tseemann/snp-dists>), and/or the R package harrietr (v0.2.4;
167 <https://github.com/andersgs/harrietr>). The adegenet R package (v2.1.5;
168 <https://github.com/thibautjombart/adegenet/>) was used to generate a pseudogenome alignment in
169 DNA.bin format (one of the input files accepted by harrietr) from a pseudogenome alignment
170 file. Phylogenetic tree, epidemiological data and in silico data were visualized in the web-based
171 viewers Microreact [24], and the interactive Tree Of Life (iTOL) [25].

172

173 **AMR gene, Virulence gene, and Plasmid Replicon Prediction**

174 The program SRST2 (v0.2.0; <https://github.com/katholt/srst2/>) [26] and the ARGannot database
175 (https://raw.githubusercontent.com/katholt/srst2/master/data/ARGannot_r3.fasta) were used to
176 screen raw sequence reads for the presence of acquired resistance genes. We also validated the

177 report/output by utilizing the GHRU AMR Pipeline
178 (https://gitlab.com/cgps/ghru/pipelines/dsl2/pipelines/amr_prediction) as explained in this
179 protocol [22]. *ampC1* and *ampC2* were excluded from downstream analysis and visualization as
180 they are beta-lactamase genes present in almost all *E. coli* isolates and are unlikely to confer
181 antibiotic resistance in *E. coli* [27]. Overall plot of resistance determinants was constructed using
182 the *upset* function from the UpSetR package (v1.4.0) [28]. Plots of AMR genes stratified by
183 sequence type and sentinel site were constructed using the ggupset package (v0.3.0). Association
184 between virulence genes and phylogroups of *E. coli* was determined using the *fisher_test*
185 function from the rstatix package (v0.7.0). The calculation is based on single genes (including
186 those that make up an operon), not operons. The associations were corrected for multiple testing
187 using the Bonferroni method offered by rstatix. Bar plots were visualized using the *ggplot*
188 function from the Tidyverse package (v1.3.1) in R.

189 The raw reads were screened for virulence genes with the GHRU pipeline, which utilizes
190 ARIBA [29] and the VFDB database [30]. Plasmid replicon types were determined with same
191 GHRU pipeline but using the PlasmidFinder database instead [31].

192

193 **Multilocus Sequence Typing (MLST) profiling**

194 Following the Achtman MLST scheme [32], multi-locus sequence types (STs) were determined
195 using the SRST2 program. We confirmed the results by using the GHRU MLST pipeline, which
196 also implements the Achtman scheme. as summarized in the aforementioned GHRU protocol
197 [22].

198

199 ***In silico* Serotyping and Phylogroup Determination**

200 The O and H serotypes of *Escherichia coli* were determined using the SRST2 program and the
201 EcOH database (<https://raw.githubusercontent.com/katholt/srst2/master/data/EcOH.fasta>). This
202 result was validated using ECTyper (v1.0.0; https://github.com/phac-nml/ecoli_serotyping) and
203 SerotypeFinder (<https://bitbucket.org/genomicepidemiology/serotypefinder/src/master/>).
204 SerotypeFinder utilizes KMA [33] and BLAST+ [34] to predict *E. coli* serotypes from the

205 alignment of raw and assembled reads, respectively, against the SerotypeFinder database.
206 *Escherichia coli* genomes were assigned into phylogroups using ClermonTyping (v20.03) [35].

207

208 **Concordance**

209 The agreement between phenotypic and genotypic antimicrobial resistance was determined for
210 beta-lactams, cephalosporins, amikacin/kanamycin/gentamicin, trimethoprim, and the
211 quinolones. Metrics such as sensitivity, specificity, true positives, true negatives, false positives,
212 and false negatives were determined using the R script (<https://gitlab.com/-/snippets/2050300>,
213 first used in a previous report) which utilizes the *epi.tests* function within the epiR package
214 (v2.0.26) for each antimicrobial tested. Here, “Sensitive” and “Intermediate” values were
215 combined, taking cognizance of the arguments for and against the use of the term “Intermediate”
216 in clinical settings [36].

217

218 **Genome Annotation, Comparative Genomics, and gene location prediction.**

219 Functional annotation of Onovel32 clade genomes was performed using Bakta v1.0.4 [37] .
220 Genomes were compared and visualized using Artemis v18.1.0 [38], Artemis Comparison Tool
221 v18.1.0 [39], and Clinker v0.0.21 [40]. The mlplasmid web tool [41] was used to predict whether
222 clinically-relevant virulence genes were borne on plasmids or on chromosomes.

223

224

225 **Results**

226 **Epidemiology and Species Identification**

227 Three hospital laboratories in South-west Nigeria submitted retrospective bloodstream
228 *Escherichia coli* isolates with clinical and epidemiological data between the years 2016 to 2018.
229 Available data showed that isolates were collected from patients aged 1 day to 71 years, with 17
230 (25%) of the isolates recovered from children under 60 days old (range 1- 20 days, median 9
231 days), who would be characterized as neonates. The isolates were submitted from the University
232 College Hospital (UCH; n = 22), Obafemi Awolowo University (OAU) Teaching Hospitals

233 Complex (n = 18), and Osun State University Teaching Hospital, Teaching Hospital (until
234 recently known as Ladoke Akintola University) (LAU; n = 28).

235 Of the 68 invasive isolates confirmed as *E. coli* by whole-genome sequencing (WGS), 48
236 (70.6%) and 64 (94.1%) were correctly identified as *E. coli* by the sentinel biochemical testing
237 and reference laboratory VITEK2, respectively. *Escherichia coli* isolates were often
238 misidentified as *Klebsiella pneumoniae* (n = 8) or *Citrobacter freundii* (n = 6) at the sentinel
239 laboratories, while the VITEK2 system misidentified *E. coli* as *Klebsiella pneumoniae* (n = 3) or
240 *Enterobacter aerogenes* (n = 1).

241

242 **Phylogroups, serotypes, and sequence types of *E. coli* bloodstream isolates**

243 *E. coli* sent from all three hospital sentinel laboratories spanned all *E. coli* phylogroups, with 18,
244 15, 19, 8, 2, 1, and 5 *E. coli* genomes classified within phylogroups A, B1, B2, C, D, E, and F,
245 respectively. The most common STs among the 33 identified include; ST131 (n = 12), ST156 (n
246 = 5), and 4 each of ST10, ST167, ST410, ST648, and ST90. These 7 STs accounted for 54% of
247 the *E. coli* isolates. Of these, only STs 131 and 167 were found across the 3 sentinel sites (Figure
248 1, Table S4). Diverse lineages were recovered from all three sentinel sites (OAU = 10 STs; LAU
249 = 18 STs, UCH = 15 STs) (Table 1). While ST131 genomes made up 63% of genomes within
250 phylogroup B2 (the second most common phylogroup), ST10 and ST167 genomes accounted for
251 44% of genomes within phylogroup A (the most common phylogroup).

252 While none of the isolates were submitted as suspected outbreak strains, , four isolates recovered
253 from LAU within one month (January 2017) were typed as ST90, serotype O8:H9, carried the
254 same plasmid profile (IncFI; IncFIA; IncFIB_AP001918; IncQ1), AMR gene profile, and similar
255 but not identical resistance profile (Table S4). These four isolates were resistant to trimethoprim,
256 the quinolones, gentamicin, cefuroxime axetil, and ampicillin (Table S3b). Comparison of
257 pairwise SNP differences revealed that two of the four isolates were identical while the other two
258 isolates differed from these by 11 SNPs and 72 SNPs. The current literature supports a
259 clustering threshold between the range of 0 to 17 SNPs for suspected outbreak-related *E. coli*
260 blood isolates [42–44]. Although these isolates were correctly identified within this sentinel site,
261 the cluster was only recognized retrospectively, likely due to different beta lactam resistance
262 profiles.

263

264 ST131 was the most common sequence type detected and the 12 isolates belonging to this ST
265 clustered into two distinct lineages defined by serotype H5 (n = 5; O16:H5) and the serotype H4
266 lineage (n = 7; O25:H4), and henceforth referred to as the ST131 lineage 1 and ST131 lineage 2,
267 respectively (Figure 1).

268 A total of 38 unique serotypes and 25 O-groups were identified. Of note was the most common
269 O-type, which was novel and identified in eleven isolates belonging to phylogroup A. These
270 ONovel:32 strains belonged to ST10 or ST167, or were single or double locus variants of these
271 STs. They included six ONovel32:H10 isolates as well as three H21 and one H4 –flagellin-
272 encoding strains. Along with one ONT:H10 strain that also belonged to ST10, they formed a
273 distinct clade on the phylogenetic tree (Figure 1). Strains belonging to this cluster were
274 submitted from all three hospitals. Two of them were originally misclassified as *Klebsiella*
275 *pneumoniae* by VITEK2 at the reference laboratory level. Irrespective of whether the ST90
276 clade is discounted, the ONovel:32 clade and the ST131 clade were the most abundant. Together
277 they accounted for 32.4 % of the isolates and both clades were found in all three hospitals
278 (Figure 1).

279

280 **Virulence Factor Profiles of the Bloodstream *E. coli* Isolates**

281 Diverse virulence genes were observed among the ExPEC genomes: 159 virulence-associated
282 (VAG) genes were found at least once in the 68 bloodstream isolate genomes. Enterobactin
283 genes (*entB*, *entC*, *entE_1*, *entS*) and ferrienterobactin precursors and proteins (*fep* operon genes
284 (ABCDEG) and *fes_1*), were found in more than 95% of the isolates. The outer membrane hemin
285 receptor (*chu*), siderophores (*fyuA*, *irp*, *ybt*), intimin-like adhesin (*fdeC*), haemolysin (*hly*),
286 aerobactin (*iuc*), polysialic acid transport protein *kpsM_1*, pyelonephritis-associated pili *pap*,
287 SPATE genes (*sat*, *vat*), and plasmid-encoded *Shigella* enterotoxin *senB* were more abundant
288 and significantly more associated with phylogroup B2 than with phylogroups A, B1, and C
289 (Fisher's exact Test, $p < 0.05$; Figure 2a). Figure 2b shows that a wide range of biofilm-forming
290 capacities were seen in the subset of isolates tested with moderate or strong biofilm-formers
291 being most common in phylogroups B1 and B2 as well as the ST90 outbreak strain-containing

292 phylogroup C. No significant difference in biofilm-formation between the phylogroups was
293 observed.

294

295 Phylogroup B2 isolates, comprised largely of ST131 strains, carried the highest number of
296 virulence-associated genes (n = 86 VAGs). Thirty-four VAGs were significantly more common
297 in ST131 isolates (n = 12) than in non-ST131 isolates (n = 56). Of these the genes encoding the
298 outer membrane haem receptor (*chu*), yersiniabactin (*ybt*), and pyelonephritis-associated pili
299 (*pap*) were found in more than 84% of the ST131 isolates (Fisher's exact Test, $p < 0.05$; Fig. 3a).
300 On the other hand, 6 VAGs were significantly more common in non-ST131 isolates than in
301 ST131 isolates, including 2 genes (*gspK* and *gspL*) present in more than 85% of the non-ST131
302 isolates, but in only about 50% of ST131 isolates. The VAGs *entD*, *espL1*, *espX1*, and *espL5*
303 were absent in ST131 isolates but were present in at least 32% of non-ST131 isolates, notably
304 Phylogroup A ONovel32 strains (Fig. 3a).

305 Both lineages of ST131 share common VAGs, including; adhesins, yersiniabactin, aerobactin,
306 enterotoxin, and transport-associated genes but some genes were only seen, or predominantly
307 seen in one of the two lineages. For example, the haemolysin (*hly*) carried by two isolates within
308 ST131 lineage 2 were absent in ST131 lineage 1. Also, two isolates within ST131 lineage 2
309 carried the cytotoxic necrotizing factor *cnf1*. (Figure 3a and 3b). The median pairwise distance
310 between isolates in ST131 lineage 1 is 2857.5 SNPs (range: 0 – 2866), while the median pairwise
311 distance between isolates within ST131 lineage 2 is 499 SNPs (range: 0 – 827). The inter-clade
312 SNP distance is 13619 (range: 12900 – 14146) (Figure S1).

313

314 None of the ONovel32 clade isolates carried alpha haemolysin or pyelonephritis associated pili,
315 which were present in ST131 and some of the other lineages. However, some virulence-
316 associated genes were predominant among or restricted to ONovel32 strains (Figure 4a and 4b).
317 These include the *esp4L* type III secretion effector. We also searched independently for type III
318 secretion systems and identified a cluster 97.5% identical to *E. coli* type III secretion system 2,
319 previously associated with virulence in septicemic *E. coli* [45], in all *esp4L*-positive and one
320 *esp4L*-negative Onovel32 isolate (GI8584164). *esp4L* was more common among these isolates
321 than other phylogroup A strains as well as multiple genes encoding capsular modification

322 enzymes. All the ONovel32 strains and one associated ONT strain carried a *ugd* gene. The only
323 other isolates in this collection with this gene were three isolates belonging to the globally
324 disseminated high risk clone ST648 [46], known to possess biofilm-associated features that
325 enhance pathogen emergence and persistence in both the human body and the environment. The
326 *ugd* gene is associated with hypermucoviscosity and invasive virulence and the ONovel32 allele
327 is 96% identical to that from *K. pneumoniae* NTUH-K2044, a hypervirulent *K. pneumoniae*
328 isolate [47] and *K. variicola* (Accession number CP079802.1) capsular cluster *ugd* genes. Six of
329 the ONovel:32 strains (but not the ONT:H10 strain in the same clade) carried *cpsACP*, a
330 chromosomally-borne gene, which is predicted to encode a phosphatidic acid phosphatase (PAP2
331 Pfam 01569) family gene. PAP2 phosphatases replace phosphate groups on lipid A with amine
332 groups resulting in a positively charged lipid A that confers resistance to cationic peptides [48].
333 PAP2 phosphatases have been known to be transmitted horizontally solitarily or as part of
334 capsular clusters [49]. A BLAST search revealed that the ONovel32 PAP2 allele is 99.6%
335 identical to endogenous PAP2 genes from *Klebsiella variicola* (Accession number CP079802.1).
336 As shown in Figure 5, depicting the region for ONovel32 ST1284 strain OAU-VOA-056,
337 *cpsACP* is located within a capsular gene cluster identical to a *K. variicola* cluster and flanked by
338 a 5' IS3 transposase and a 3' IS1 protein InsB-encoding gene. At the opposite end of the cluster
339 is the *ugd* gene. The cluster shows G+C content and other base-pattern signatures that depart
340 from the *E. coli* flanking sequence (Figure 5a and 5b).

341
342 Observing that ONovel:32 cluster isolates, some of which were originally misclassified as
343 *Klebsiella*, showed mucoidity upon plate culture, we sought to determine whether these strains
344 had distinctive colonization-associated capacities. The median A_{570} at six hours was 0.178 for
345 phylogroup A strains lacking *cpsACP* and 0.216 for those with the gene (Figure 2), but these
346 differences were not significant. All the isolates were also negative in the string test for
347 hypermucoviscosity in *Klebsiella* [50].

348
349
350 **Resistance Profiles (RP) and Concordance with Predicted Antimicrobial Resistance**

351 Susceptibility testing of 67 out of the 68 isolates showed that, of the 16 antibiotics tested,
352 resistance to trimethoprim-sulfamethoxazole (n = 67; 100%), ampicillin (n = 62; 92.5%),
353 nalidixic acid (n = 57; 85.1%), and ciprofloxacin (n = 53; 79.1%) was commonly observed (Fig.
354 6a, Fig. 6b, Table S2). On the other hand, resistance to cefoperazone.sulbactam (n = 3; 4.5%)
355 was less common. All isolates were susceptible to amikacin, ertapenem, , and meropenem (Table
356 S2, Fig. S1a and S1b). One *bla*_{CTX-M-15} –positive isolate from LAU (ST11025, B1, O51:H12) was
357 categorized as showing intermediate to imipenem but did not carry any carbapenemase-
358 producing gene. Resistance profiles were remarkably similar among the three sentinels (Figure
359 6b). Resistance to trimethoprim/sulfamethoxazole among isolates belonging to all 33 STs could
360 be explained by the possession of *dfrA* (n = 54/67) and *dfrB* (n = 2/67) genes. Ciprofloxacin
361 resistance (23 STs) was largely attributable to mutation in the quinolone resistance determining
362 regions (QRDR) of *gyrA* (D87N, S83L), *parC* (E84A, E84G, E84K, S57T, S80I), and/or *parE*
363 (E460D, I355T, I529L, I529L, I464F, L416F, L445H, S458A, S458T), with or without the
364 presence of plasmid quinolone resistance genes (*qnrS*, *qnrVC4*, *qepA*, *aac*-(6′)-*Ib-cr*). About
365 45% (n = 24/53) of isolates phenotypically resistant to ciprofloxacin carried a combination of
366 *aac*-(6′)-*Ib-cr5* gene and mutation in the QRDR regions (*gyrA*, *parC*, and *parE*) (Figure 7a and
367 7b). The most common quinolone resistance gene profile observed is the
368 “*gyrA*_D87N,*gyrA*_S83L,*parC*_S80I,*parE*_S458A”, as observed in a quarter (n = 14/53) of
369 isolates showing phenotypic resistance to ciprofloxacin.

370

371 Ampicillin resistance among isolates belonging to 29 STs (n = 62) could be explained by the
372 carriage of a range of beta-lactamase genes including *bla*_{TEM-1} (36/62 of isolates), *bla*_{TEM-40} (3
373 isolates), *bla*_{TEM-84} (2 isolates), *bla*_{TEM-135} (2 isolates), , or *bla*_{OXA-2} (co-occurring with *bla*_{TEM-90}; 1
374 isolate). Extended-spectrum beta-lactam resistance likely resulted from *bla*_{CTX-M-15}, *bla*_{CTX-M-27},
375 and *bla*_{CMY-42} (Figure S2a and S2b) co-occurring largely with the plasmids Inc types
376 FIB_AP001918, FIA, IncFI, IncQ1, IncFII_p, and the col plasmid Col156.

377 Of the 5 drug classes tested, the highest concordance between phenotypic resistance and
378 predicted antimicrobial resistance was observed for trimethoprim (100% concordance, TP =
379 67/67, Sensitivity = 100%, Specificity = NA) and the lowest concordance for drugs within the

380 aminoglycoside class (concordance = 55.22%, TP = 37/67, Sensitivity = 100%, Specificity = 0
381 %) (Table S3a).

382

383 **Multidrug Resistance (MDR)**

384 We observed a total of 23 resistance profiles (RP) (Table S1) with resistance profiles in many
385 cases associated with specific STs (Table 1 and Table S4). Fifty-nine (88.06%) of the isolates
386 were resistant to at least one agent within at least three classes of antibiotics, fitting the multi-
387 drug resistant (MDR) definition of the international AMR community [51] (Table S1). Isolates
388 belonging to the most frequent sequence types form the bulk of invasive isolates carrying ESBL
389 genes or mutations in the quinolone-resistance determining regions (*gyrA*, *parC*, *parE*), and the
390 plasmid-mediated quinolone resistance gene *aac(6')-Ib-cr*, as well as genes conferring resistance
391 to trimethoprim. Carbapenemase genes were conspicuously absent, in agreement with the AMR
392 phenotype. Almost all isolates carrying ESBL genes (n = 32/34) are multidrug resistant. About
393 30% of these showed the most frequent resistance profile (RP-1) (Figure S3). Furthermore,
394 multi-drug resistant bloodstream isolates do not carry the same resistance determinants and
395 plasmid replicons, apart from *gyrA/parC/parE* mutations (Figure S3).

396

397 Resistance gene profiles differed significantly between the two ST131 clades (Figure 3c).
398 Unlike ST131 lineage 1 isolates, ST131 lineage 2 isolates did not carry any beta-lactamase gene
399 besides *ampC*. Within the ST131 lineage 1, two isolates did not carry genes conferring
400 resistance/reduced susceptibility to the aminoglycosides, phenicols, macrolides, and quaternary
401 ammonium compounds. Furthermore, the absence of the Col156, IncB_O_K_Z, and IncFI
402 plasmids in these two isolates seem to have been compensated by the possession of IncFIA,
403 IncFII_p, and IncI1 plasmids, noted to have been absent in the other three ST131 isolates within
404 the ST131 lineage 1.

405

406 All the ONovel32 strains (but not the co-clustering ONT strain) carried one ESBL gene, *bla_{CTX-M-}*
407 *15*, and these strains also carried IncF plasmids, common among phylogroup A strains (Table S4,
408 Table S5). Every one of them carried the most common four resistance-conferring mutations in

409 the QRDRs (*gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A*) and seven, including the ONT
410 strain, carried *qepA4*. Four ONovel32 isolates additionally carried *aac-(6')-Ib-cr*, conferring
411 aminoglycoside and ciprofloxacin resistance, as well as horizontally-transmitted genes
412 conferring resistance to trimethoprim, chloramphenicol and tetracyclines were common
413 (Microreact link, Table S5).

414
415 Common to both the ST131 and the ONovel:32 clades is the high prevalence of *qacEdelta1* (n =
416 9 and 8, respectively) conferring resistance to quaternary ammonium compounds, commonly
417 used for disinfecting hospital surfaces and associated with class 1 integrons. Altogether, these
418 clades comprised 17 (42.50%) of 40 *E. coli* isolates carrying this gene.

419

420

421 **Discussion**

422 This study characterized 68 bloodstream *E. coli* isolates as an important first step in
423 understanding their epidemiology within south-west Nigeria. In this small collection, we
424 identified multiple clones of pandemic importance, and found two predominant clades. One of
425 these, comprised of two ST131 lineages is globally disseminated and this study illustrates its
426 importance in Nigerian health facilities. The second predominant clade does not feature in
427 present discourse on international ExPEC clones and represents strains belonging to ST10,
428 ST167 and related STs, which predominantly encode genes that could confer a novel O-antigen
429 type. In addition to these prominent clades, we identified strains belonging to major pandemic
430 ExPEC lineages, including ST12, ST73, and ST648, and their single locus variants. ST69,
431 ST95, and ST405 lineages were not detected but our sample is not very large and therefore our
432 findings are insufficient to rule them epidemiologically insignificant in our setting. We
433 additionally identified in the collection five ST90/410 strains, all from a single facility. Two
434 ST90 isolates are identical, differ from a third by only 11 SNPs, and likely represent an outbreak.
435 Our findings add to information that is chronicling ExPEC lineages of importance within Nigeria
436 [14–16, 52, 53], other parts of Africa [16, 54], and other low- middle-income countries [16, 55].

437

438 Among ST131, we found both globally-disseminated lineages within our isolate collection. The
439 majority of haemolytic and *pap* gene-bearing phylogroup B2 strains belonged to this ST.
440 Biofilm formation among these strains was common and associated with the *kpsD* gene, a known
441 contributor to biofilm formation [56]. The ST10, ST167 and related strains that comprised the
442 ONovel32 clade were distinguished by the presence of one of two variations of a capsular island
443 that has been well described in *Klebsiella* with biofilm formation among them being more
444 pronounced among strains carrying the *cpsACP*-containing portion of the island. While the
445 virulence of ST131 has been well described, features of this clade that cause it to predominate
446 remain unknown and further studies are required to understand its pathogenicity and selective
447 success in our setting.

448
449 High levels of phenotypic resistance to antibiotics within the antifolate (trimethoprim. co-
450 trimoxazole), quinolone (nalidixic acid), fluoroquinolone (ciprofloxacin), cephalosporin
451 (cefuroxime axetil), and the aminoglycosides (gentamicin) classes observed among invasive
452 isolates in this study corroborates reports from previous studies [12, 52, 53] and pose a serious
453 concern for clinical therapeutics. *In silico* data confirm the abundance of genes conferring
454 resistance, with many isolates carrying more than one gene conferring resistance to a specific
455 antimicrobial or class. Before the turn of the millennium, resistance to many of the
456 aforementioned antimicrobials was rare in Africa. However, steady increase in the availability
457 and use of these agents in the empiric treatment of ExPEC-related infections has inevitably
458 selected for AMR. The increased rate of fluoroquinolone resistance in diarrhoeagenic *E. coli* and
459 other enterobacteriales, for instance, coincided with increased fluoroquinolones use in Nigeria
460 [57] and other parts of Africa [7, 58]. While findings from a study conducted in Nigeria more
461 than a decade ago [13] concluded that nalidixic acid was still an effective antimicrobial, we
462 observed in this study that resistance to nalidixic acid is now common.

463 Similarly, cephalosporin resistance emerged and expanded much more in Africa than in other
464 parts of the world as these agents became the drug of choice for treating multi-drug resistant
465 pathogens [58]. Fluoroquinolone resistance has also been previously associated with the presence
466 of ESBL genes because ESBL genes are often borne on transferable large plasmids that co-host
467 some of the PMQR genes [59]. We observed the co-carriage of ESBL and fluoroquinolone

468 resistance genes in more than a quarter of the ExPEC isolates, and particularly in over-
469 represented lineages.

470

471 We find that both predominant lineages we have highlighted in this report show multiple
472 resistance. This has important implications for patients with life-threatening bloodstream
473 infections and provides a plausible explanation for their evolutionary success in our setting.
474 Resistance to antimicrobials used intensively in Nigeria (trimethoprim, aminopenicillins and
475 ciprofloxacin) was rife and resistance to the agents most frequently used empirically when blood
476 stream infections are suspected – second- and third-generation cephalosporins and
477 aminoglycosides was also worryingly common. While none of these clades showed carbapenem
478 or colistin resistance, these reserve antimicrobial classes are out of the reach of most patients
479 attending the three hospitals from which the strains were obtained.

480 Next-generation sequencing has emerged as a promising complement to clinical bacteriology as
481 it provides answers to medical conundrums as well as a more robust picture of the epidemiology
482 of infectious diseases. It can also reveal, as in this case, circulation of hitherto unrecognized
483 clones of concern. Although its integration into clinical diagnosis and patient care gaining
484 ground many parts of the world, its adoption in LMICs (and the rest of the globe) is still hindered
485 by lack of infrastructure, cost of implementing WGS, limited bioinformatics expertise and as yet
486 mildly inaccurate prediction of antimicrobial resistance [60, 61]. In our study, we observed
487 perfect concordance (100%) with phenotypic AST data for trimethoprim, but not for the
488 cephalosporins, the quinolones, or the aminoglycosides. This therefore shows that further
489 understanding of resistance mechanisms and routine AMR database update is needed to enhance
490 the feasibility of gradual and sustained integration of WGS into routine diagnosis.

491 Another advantage of next-generation sequencing is its potential in rapid detection of outbreaks
492 either retrospectively, or in real-time. We report a likely retrospective ST90 outbreak in LAU,
493 which would not have been detected using traditional diagnostic methods in the sentinel
494 laboratories. ST90 strains have variously been highlighted for their zoonotic potential and
495 association with device-related hospital outbreaks [62–64] and its epidemiology in our setting
496 remains to be understood ST90, and related ST410, were only seen at one facility and the
497 representation of ST90 at this facility may have been amplified by an outbreak. The adoption of

498 genomic surveillance in diagnostic laboratories within Nigeria will ensure that outbreaks of this
499 and other clones can be detected in real time, while definite tracking and containment of the
500 spread of such clones will be achieved before lives are lost.

501 This study has a few limitations. Blood culture is infrequently performed in Nigerian hospitals
502 and until recently, most isolates were not archived. Therefore, these isolates represent but a
503 fraction of the ExPEC likely to have infected patients in the three hospitals and may not be
504 representative. Our short-read data makes it impossible to accurately determine whether the
505 resistance genes are located on the bacterial chromosome or plasmids (or other mobile genetic
506 elements). In future, we will incorporate long read sequencing into our prospective surveillance
507 efforts in order to correctly identify plasmid-borne AMR genes.

508 In conclusion, this study provided hospital-specific information on the population structure of
509 ExPEC lineages needed to track pandemic lineages and guide infection disease control practices
510 in line with Nigeria's national action plan on antimicrobial resistance.

511

512 **Author Statements**

513 **Conceptualization:** David M. Aanensen, Iruka N. Okeke, Chikwe Ihekweazu; **Data curation:**
514 Ayorinde O. Afolayan, Anthony Underwood; **Formal Analysis:** Ayorinde O. Afolayan, Anthony
515 Underwood, Oyeniyi S. Bejide, Iruka N. Okeke; **Funding acquisition:** David M. Aanensen,
516 Aaron O. Aboderin, Iruka N. Okeke; **Investigation:** Ayorinde O. Afolayan, Abiodun Egwuenu,
517 Erkison Ewomazino Odih, Oyeniyi S. Bejide, Aaron O. Aboderin, Iruka N. Okeke;
518 **Methodology:** Ayorinde O. Afolayan, Aaron O. Aboderin, Anderson O. Oaikhena, Erkison
519 Ewomazino Odih, Oyeniyi S. Bejide, Veronica O. Ogunleye, Adeyemi T. Adeyemo, Abolaji T.
520 Adeyemo, Anthony Underwood, Silvia Argimón, Monica Abrudan, David M. Aanensen, Iruka
521 N. Okeke; **Project administration:** Abiodun Egwuenu, Chikwe Ihekweazu, Aaron O. Aboderin,
522 Iruka N. Okeke; **Resources:** Anthony Underwood, Silvia Argimón, Veronica O. Ogunleye,
523 Adeyemi T. Adeyemo, Abolaji T. Adeyemo, David M. Aanensen, Iruka N. Okeke; **Software:**
524 Ayorinde O. Afolayan, Silvia Argimón, Anthony Underwood; **Supervision:** Aaron O. Aboderin,
525 Anthony Underwood, Silvia Argimón, Chikwe Ihekweazu, David M. Aanensen, Iruka N. Okeke;
526 **Validation:** All authors; **Visualization:** Ayorinde O. Afolayan, Iruka N. Okeke; **Writing –**
527 **original draft:** Ayorinde O. Afolayan; **Writing – review & editing:** All authors.

528

529 **Acknowledgments/Funding**

530 We thank Damilola Q. Olaoye, Chinenye Ekemezie, Ifeoluwa Akintayo and Ifeoluwa Komolafe
531 for excellent technical assistance.

532 This work was supported by Official Development Assistance (ODA) funding from the National
533 Institute of Health Research [grant number 16_136_111] and the Wellcome Trust grant number
534 206194. INO was an African Research Leader supported by the UK Medical Research Council
535 (MRC) and the UK Department for International Development (DFID) under the MRC/DFID
536 Concordat agreement that is also part of the EDCTP2 program supported by the European
537 Union. The funders had no role in the content, crafting or submission of this paper.

538

539 **Ethics**

540 Isolates were obtained as part of the surveillance efforts in line with Nigeria's national action
541 plan and/or for bloodstream pathogen and *Escherichia coli* research. In Nigeria. Ethical
542 approval for research using these isolates was obtained from the University of Ibadan/University
543 College Hospital and the Obafemi Awolowo University (OAU) Teaching Hospitals complex
544 ethics committees. Respective IRB registration numbers are UI/EC/15/093 and ERC/2017/05/06.

545

546 **Conflicts of Interest**

547 The authors have no conflicts of interest to declare.

548 **References**

- 549 1. **de Kraker MEA, Jarlier V, Monen JCM, Heuer OE, van de Sande N, et al.** The
550 changing epidemiology of bacteraemias in Europe: Trends from the European
551 antimicrobial resistance surveillance system. *Clin Microbiol Infect* 2013;19:860–868.
- 552 2. **Foxman B.** The epidemiology of urinary tract infection. *Nature Reviews Urology*
553 2010;7:653–660.
- 554 3. **Manges AR, Geum HM, Guo A, Edens TJ, Fibke CD, et al.** Global extraintestinal

- 555 pathogenic escherichia coli (Expec) lineages. *Clin Microbiol Rev*;32. Epub ahead of print
556 1 July 2019. DOI: 10.1128/CMR.00135-18.
- 557 4. **Johnson JR, Russo TA.** Molecular Epidemiology of Extraintestinal Pathogenic
558 Escherichia coli. *EcoSal Plus*;8. Epub ahead of print 8 February 2018. DOI:
559 10.1128/ecosalplus.esp-0004-2017.
- 560 5. **Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, Frej-Madrzak M, Ksiazczyk M, et**
561 **al.** Virulence factors, prevalence and potential transmission of extraintestinal pathogenic
562 Escherichia coli isolated from different sources: Recent reports. *Gut Pathogens*
563 2019;11:10.
- 564 6. **McNally A, Kallonen T, Connor C, Abudahab K, Aanensen DM, et al.** Diversification
565 of colonization factors in a multidrug-resistant escherichia coli lineage evolving under
566 negative frequency- dependent selection. *MBio*;10. Epub ahead of print 1 March 2019.
567 DOI: 10.1128/mBio.00644-19.
- 568 7. **Chattaway MA, Aboderin AO, Fashae K, Okoro CK, Opintan JA, et al.**
569 Fluoroquinolone-resistant enteric bacteria in sub-saharan Africa: Clones, implications and
570 research needs. *Frontiers in Microbiology* 2016;7:558.
- 571 8. **Alonso CA, Zarazaga M, Ben Sallem R, Jouini A, Ben Slama K, et al.** Antibiotic
572 resistance in Escherichia coli in husbandry animals: the African perspective. *Letters in*
573 *Applied Microbiology* 2017;64:318–334.
- 574 9. **Khairy RM, Mohamed ES, Abdel Ghany HM, Abdelrahim SS.** Phylogenic
575 classification and virulence genes profiles of uropathogenic E. coli and diarrhegenic E.
576 coli strains isolated from community acquired infections. *PLoS One* 2019;14:e0222441.
- 577 10. **Forson AO, Tsidi WB, Nana-Adjei D, Quarchie MN, Obeng-Nkuramah N.**
578 Escherichia coli bacteriuria in pregnant women in Ghana: Antibiotic resistance patterns
579 and virulence factors. *BMC Res Notes* 2018;11:901.
- 580 11. **Huynh BT, Padget M, Garin B, Herindrainy P, Kermorvant-Duchemin E, et al.**
581 Burden of bacterial resistance among neonatal infections in low income countries: How
582 convincing is the epidemiological evidence? *BMC Infect Dis* 2015;15:127.
- 583 12. **Seni J, Peirano G, Okon KO, Jibrin YB, Mohammed A, et al.** The population structure

- 584 of clinical extra-intestinal *Escherichia coli* in a teaching hospital from Nigeria. *Diagn*
585 *Microbiol Infect Dis* 2018;92:46–49.
- 586 13. **Okonko IO, Donbraye-Emmanuel OB, Ijandipe LA, Ogun AA, Adedeji AO, et al.**
587 Antibiotics Sensitivity and Resistance Patterns of Uropathogens to Nitrofurantoin and
588 Nalidixic Acid in Pregnant Women with Urinary Tract Infections in Ibadan, Nigeria.
589 *Middle-East J Sci Res* 2009;4:105–109.
- 590 14. **Ogbolu DO, Piddock LJV, Webber MA.** Opening Pandora’s box: High-level resistance
591 to antibiotics of last resort in Gram-negative bacteria from Nigeria. *J Glob Antimicrob*
592 *Resist* 2020;21:211–217.
- 593 15. **Duru C, Olanipekun G, Odili V, Kocmich N, Rezac A, et al.** Molecular characterization
594 of invasive Enterobacteriaceae from pediatric patients in Central and Northwestern
595 Nigeria. *PLoS One* 2020;15:e0230037.
- 596 16. **Sands K, Carvalho MJ, Portal E, Thomson K, Dyer C, et al.** Characterization of
597 antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low-
598 and middle-income countries. *Nat Microbiol* 2021 64 2021;6:512–523.
- 599 17. **CLSI.** M100 Performance Standards for Antimicrobial Susceptibility Testing A CLSI
600 supplement for global application. Performance Standards for Antimicrobial Susceptibility
601 Testing Performance Standards for Antimicrobial Susceptibility Testing.
602 [https://www.academia.edu/41587236/M100_Performance_Standards_for_Antimicrobial_](https://www.academia.edu/41587236/M100_Performance_Standards_for_Antimicrobial_Susceptibility_Testing_A_CLSI_supplement_for_global_application_28th_Edition)
603 [Susceptibility_Testing_A_CLSI_supplement_for_global_application_28th_Edition](https://www.academia.edu/41587236/M100_Performance_Standards_for_Antimicrobial_Susceptibility_Testing_A_CLSI_supplement_for_global_application_28th_Edition) (2020,
604 accessed 6 August 2020).
- 605 18. **Berends MS, Luz CF, Friedrich AW, Sinha BNM, Albers CJ, et al.** AMR - An R
606 package for working with antimicrobial resistance data. *bioRxiv* 2019;810622.
- 607 19. **Wakimoto N, Nishi J, Sheikh J, Nataro JP, Sarantuya J, et al.** Quantitative biofilm
608 assay using a microtiter plate to screen for enteroaggregative *Escherichia coli*. *Am J Trop*
609 *Med Hyg* 2004;71:687–690.
- 610 20. **Naves P, Prado G Del, Huelves L, Gracia M, Ruiz V, et al.** Measurement of biofilm
611 formation by clinical isolates of *Escherichia coli* is method-dependent. *J Appl Microbiol*
612 2008;105:585–590.

- 613 21. **Afolayan AO, Oaikhena AO, Aboderin AO, Olabisi OF, Amupitan AA, et al.** Clones
614 and Clusters of Antimicrobial-Resistant *Klebsiella* from Southwestern Nigeria. *bioRxiv*
615 2021;2021.06.21.449255.
- 616 22. **Underwood A.** GHRU (Genomic Surveillance of Antimicrobial Resistance)
617 Retrospective 1 Bioinformatics Methods. [https://www.protocols.io/view/ghru-genomic-](https://www.protocols.io/view/ghru-genomic-surveillance-of-antimicrobial-resista-bpn6mmhe)
618 [surveillance-of-antimicrobial-resista-bpn6mmhe](https://www.protocols.io/view/ghru-genomic-surveillance-of-antimicrobial-resista-bpn6mmhe) (2020, accessed 9 January 2022).
- 619 23. **Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, et al.** Rapid phylogenetic
620 analysis of large samples of recombinant bacterial whole genome sequences using
621 Gubbins. *Nucleic Acids Res* 2015;43:e15.
- 622 24. **Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, et al.** Microreact:
623 visualizing and sharing data for genomic epidemiology and phylogeography. *Microb*
624 *genomics* 2016;2:e000093.
- 625 25. **Letunic I, Bork P.** Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
626 tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.
- 627 26. **Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, et al.** SRST2: Rapid genomic
628 surveillance for public health and hospital microbiology labs. *Genome Med*
629 2014;6:006627.
- 630 27. **Ingle DJ, Levine MM, Kotloff KL, Holt KE, Robins-Browne RM.** Dynamics of
631 antimicrobial resistance in intestinal *Escherichia coli* from children in community settings
632 in South Asia and sub-Saharan Africa. *Nat Microbiol* 2018;3:1063–1073.
- 633 28. **Conway JR, Lex A, Gehlenborg N.** UpSetR: An R package for the visualization of
634 intersecting sets and their properties. *Bioinformatics* 2017;33:2938–2940.
- 635 29. **Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, et al.** ARIBA: Rapid
636 antimicrobial resistance genotyping directly from sequencing reads. *Microb Genomics*;3.
637 Epub ahead of print 1 October 2017. DOI: 10.1099/mgen.0.000131.
- 638 30. **Chen L, Yang J, Yu J, Yao Z, Sun L, et al.** VFDB: a reference database for bacterial
639 virulence factors. DOI: 10.1093/nar/gki008.
- 640 31. **Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, et al.** In Silico

- 641 detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence
642 typing. *Antimicrob Agents Chemother* 2014;58:3895–3903.
- 643 32. **Wirth T, Falush D, Lan R, Colles F, Mensa P, et al.** Sex and virulence in *Escherichia*
644 *coli*: An evolutionary perspective. *Mol Microbiol* 2006;60:1136–1151.
- 645 33. **Clausen PTL, Aarestrup FM, Lund O.** Rapid and precise alignment of raw reads
646 against redundant databases with KMA. *BMC Bioinformatics* 2018;19:1–8.
- 647 34. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al.** BLAST+:
648 Architecture and applications. *BMC Bioinformatics*;10. Epub ahead of print 15 December
649 2009. DOI: 10.1186/1471-2105-10-421.
- 650 35. **Beghain J, Bridier-Nahmias A, Nagard H Le, Denamur E, Clermont O.**
651 ClermonTyping: An easy-to-use and accurate in silico method for *Escherichia* genus strain
652 phylotyping. *Microb Genomics* 2018;4:e000192.
- 653 36. **Kahlmeter G, Giske CG, Kirn TJ, Sharp SE.** Point-counterpoint: Differences between
654 the European Committee on Antimicrobial susceptibility testing and Clinical and
655 Laboratory standards institute recommendations for reporting antimicrobial susceptibility
656 results. *J Clin Microbiol*;57. Epub ahead of print 1 September 2019. DOI:
657 10.1128/JCM.01129-19.
- 658 37. **Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, et al.** Bakta: rapid and
659 standardized annotation of bacterial genomes via alignment-free sequence identification.
660 *Microb Genomics* 2021;7:000685.
- 661 38. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al.** Artemis: sequence
662 visualization and annotation. *Bioinformatics* 2000;16:944–945.
- 663 39. **Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, et al.** ACT:
664 the Artemis comparison tool. *Bioinformatics* 2005;21:3422–3423.
- 665 40. **Gilchrist CLM, Chooi Y-H.** clinker & clustermap.js: automatic generation of gene
666 cluster comparison figures. *Bioinformatics* 2021;37:2473–2475.
- 667 41. **Arredondo-Alonso S, Rogers MRC, Braat JC, Verschuuren TD, Top J, et al.**
668 mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences

- 669 for single species. *Microb Genomics* 2018;4:e000224.
- 670 42. **Roer L, Hansen F, Frølund Thomsen MC, Knudsen JD, Hansen DS, et al.** WGS-
671 based surveillance of third-generation cephalosporin-resistant *Escherichia coli* from
672 bloodstream infections in Denmark. *J Antimicrob Chemother* 2017;72:1922–1929.
- 673 43. **Schürch AC, Arredondo-Alonso S, Willems RJL, Goering R V.** Whole genome
674 sequencing options for bacterial strain typing and epidemiologic analysis based on single
675 nucleotide polymorphism versus gene-by-gene–based approaches. *Clin Microbiol Infect*
676 2018;24:350–354.
- 677 44. **Ludden C, Coll F, Gouliouris T, Restif O, Blane B, et al.** Defining nosocomial
678 transmission of *Escherichia coli* and antimicrobial resistance genes: a genomic
679 surveillance study. *The Lancet Microbe* 2021;2:e472.
- 680 45. **Ideses D, Gophna U, Paitan Y, Chaudhuri RR, Pallen MJ, et al.** A degenerate type III
681 secretion system from septicemic *Escherichia coli* contributes to pathogenesis. *J Bacteriol*
682 2005;187:8164–8171.
- 683 46. **Schaufler K, Semmler T, Wieler LH, Trott DJ, Pitout J, et al.** Genomic and functional
684 analysis of emerging virulent and multidrug-resistant *Escherichia coli* lineage sequence
685 type 648. *Antimicrob Agents Chemother*;63. Epub ahead of print 1 June 2019. DOI:
686 10.1128/AAC.00243-19.
- 687 47. **Fang CT, Chuang YP, Shun CT, Chang SC, Wang JT.** A Novel Virulence Gene in
688 *Klebsiella pneumoniae* Strains Causing Primary Liver Abscess and Septic Metastatic
689 Complications. *J Exp Med* 2004;199:697–705.
- 690 48. **Touzé T, Tran AX, Hankins J V, Mengin-Lecreulx D, Trent MS.** Periplasmic
691 phosphorylation of lipid A is linked to the synthesis of undecaprenyl phosphate. *Mol*
692 *Microbiol* 2008;67:264.
- 693 49. **Forde BM, Zowawi HM, Harris PNA, Roberts L, Ibrahim E, et al.** Discovery of mcr-
694 1 -Mediated Colistin Resistance in a Highly Virulent *Escherichia coli* Lineage .
695 *mSphere*;3. Epub ahead of print 31 October 2018. DOI: 10.1128/MSPHERE.00486-18.
- 696 50. **Hagiya H, Watanabe N, Maki M, Murase T, Otsuka F.** Clinical utility of string test as
697 a screening method for hypermucoviscosity-phenotype *Klebsiella pneumoniae*. *Acute*

- 698 *Med Surg* 2014;1:245–246.
- 699 51. **Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, et al.** Multidrug-
700 resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert
701 proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect*
702 2012;18:268–281.
- 703 52. **Aboderin OA, Abdu LR, Odetoyin BW, Lamikanra A.** Antimicrobial resistance in
704 *Escherichia coli* strains from urinary tract infections. *J Natl Med Assoc* 2009;101:1268–
705 1273.
- 706 53. **Onanuga A, Mahindroo J, Singh S, Taneja N.** Phenotypic and molecular
707 characterization of antimicrobial resistant *Escherichia coli* from urinary tract infections in
708 port-harcourt, nigeria. *Pan Afr Med J*;34. Epub ahead of print 2019. DOI:
709 10.11604/pamj.2019.34.144.18182.
- 710 54. **Tegha G, Ciccone EJ, Krysiak R, Kaphatika J, Chikaonda T, et al.** Genomic
711 epidemiology of *Escherichia coli* isolates from a tertiary referral center in Lilongwe,
712 Malawi. *Microb Genomics* 2020;7:000490.
- 713 55. **Lim C, Takahashi E, Hongsuwan M, Wuthiekanun V, Thamlikitkul V, et al.**
714 Epidemiology and burden of multidrug-resistant bacterial infection in a developing
715 country. *Elife*;5. Epub ahead of print 6 September 2016. DOI: 10.7554/eLife.18082.
- 716 56. **Russo TA, Wenderoth S, Carlino UB, Merrick JM, Lesse AJ.** Identification, genomic
717 organization, and analysis of the group III capsular polysaccharide genes *kpsD*, *kpsM*,
718 *kpsT*, and *kpsE* from an extraintestinal isolate of *Escherichia coli* (CP9, O4/K54/H5). *J*
719 *Bacteriol* 1998;180:338–349.
- 720 57. **Lamikanra A, Crowe JL, Lijek RS, Odetoyin BW, Wain J, et al.** Rapid evolution of
721 fluoroquinolone-resistant *Escherichia coli* in Nigeria is temporally associated with
722 fluoroquinolone use. *BMC Infect Dis*;11. Epub ahead of print 7 November 2011. DOI:
723 10.1186/1471-2334-11-312.
- 724 58. **Musicha P, Cornick JE, Bar-Zeev N, French N, Masesa C, et al.** Trends in
725 antimicrobial resistance in bloodstream infection isolates at a large urban hospital in
726 Malawi (1998–2016): a surveillance study. *Lancet Infect Dis* 2017;17:1042–1052.

- 727 59. **Rawat D, Nair D.** Extended-spectrum β -lactamases in gram negative bacteria. *J Glob*
728 *Infect Dis* 2010;2:263.
- 729 60. **Mitchell SL, Simner PJ.** Next-Generation Sequencing in Clinical Microbiology: Are We
730 There Yet? *Clinics in Laboratory Medicine* 2019;39:405–418.
- 731 61. **Rossen JWA, Friedrich AW, Moran-Gilad J.** Practical issues in implementing whole-
732 genome-sequencing in routine diagnostic microbiology. *Clinical Microbiology and*
733 *Infection* 2018;24:355–360.
- 734 62. **Borges CA, Tarlton NJ, Riley LW.** Escherichia coli from Commercial Broiler and
735 Backyard Chickens Share Sequence Types, Antimicrobial Resistance Profiles, and
736 Resistance Genes with Human Extraintestinal Pathogenic Escherichia coli. *Foodborne*
737 *Pathog Dis* 2019;16:813–822.
- 738 63. **Gauthier L, Dortet L, Cotellon G, Creton E, Cuzon G, et al.** Diversity of
739 carbapenemase-producing Escherichia coli isolates in France in 2012-2013. *Antimicrobial*
740 *Agents and Chemotherapy*;62. Epub ahead of print 1 August 2018. DOI:
741 10.1128/AAC.00266-18.
- 742 64. **Sartori L, Fernandes MR, Ienne S, de Souza TA, Gregory L, et al.** Draft genome
743 sequences of two fluoroquinolone-resistant CTX-M-15-producing Escherichia coli ST90
744 (ST23 complex) isolated from a calf and a dairy cow in South America. *Journal of Global*
745 *Antimicrobial Resistance* 2017;11:145–147.
- 746
747

748 **Figures and Tables**

749 **Table 1: Distribution of Sequence Types (ST) and Resistance Profiles (RP) within Sentinel**
 750 **Sites**

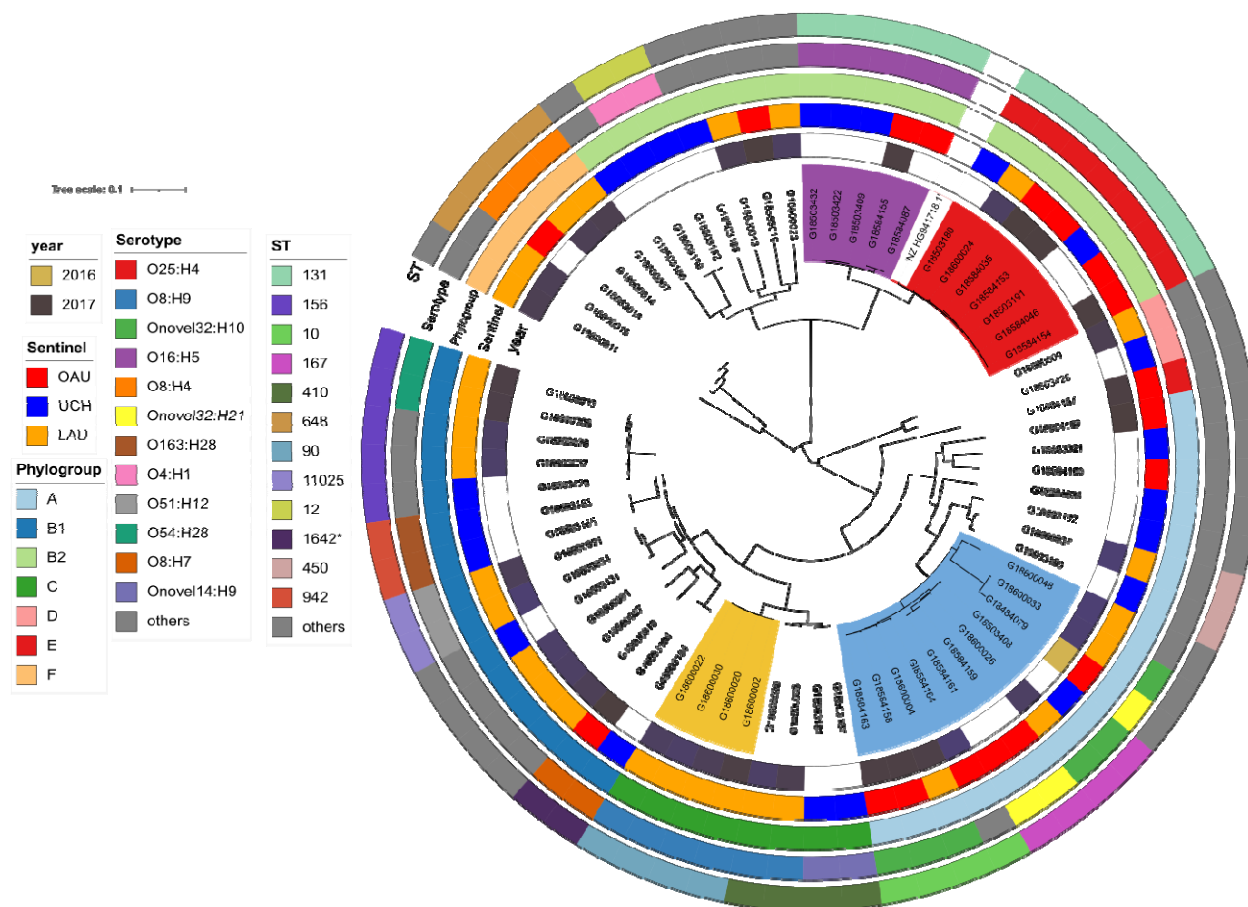
Sentinel Site	No. of Isolates	No. of Observed STs	No. of Most Prevalent ST (number)	No. of Observed Resistance Profile	No. of Most Prevalent Resistance Profile (total number, STs (number))
OAU	18	10	ST131 (6)	9	RP-1: AMP CIP CRO CXM CXMA GEN NAL SXT (n = 5; ST131 (n = 3))
UCH	22	15	ST131 (5)	14	RP-5: AMP SXT (n = 4, ST12 (n = 2), ST942 (n = 2))
LAU	28	18	ST156 (4) & ST90 (4)	14	RP-2: AMP CIP CXMA GEN NAL SXT (Total = 7, ST156 (n = 3), ST90 (n = 3))

751

752

753 A

754



755

756

757 Figure 1: Maximum likelihood SNP tree of bloodstream *E. coli* isolates sequenced for this study.

758 The purple-coloured clade represent ST131 lineage 1 (O16:H5 serotype); the red clade represent

759 the ST131 lineage 2 (O25:H4 serotype) ; the blue clade represent the ST10-ST167 clade ; The

760 light brown clade mark the cluster from a single site representing likely outbreak clone of ST90

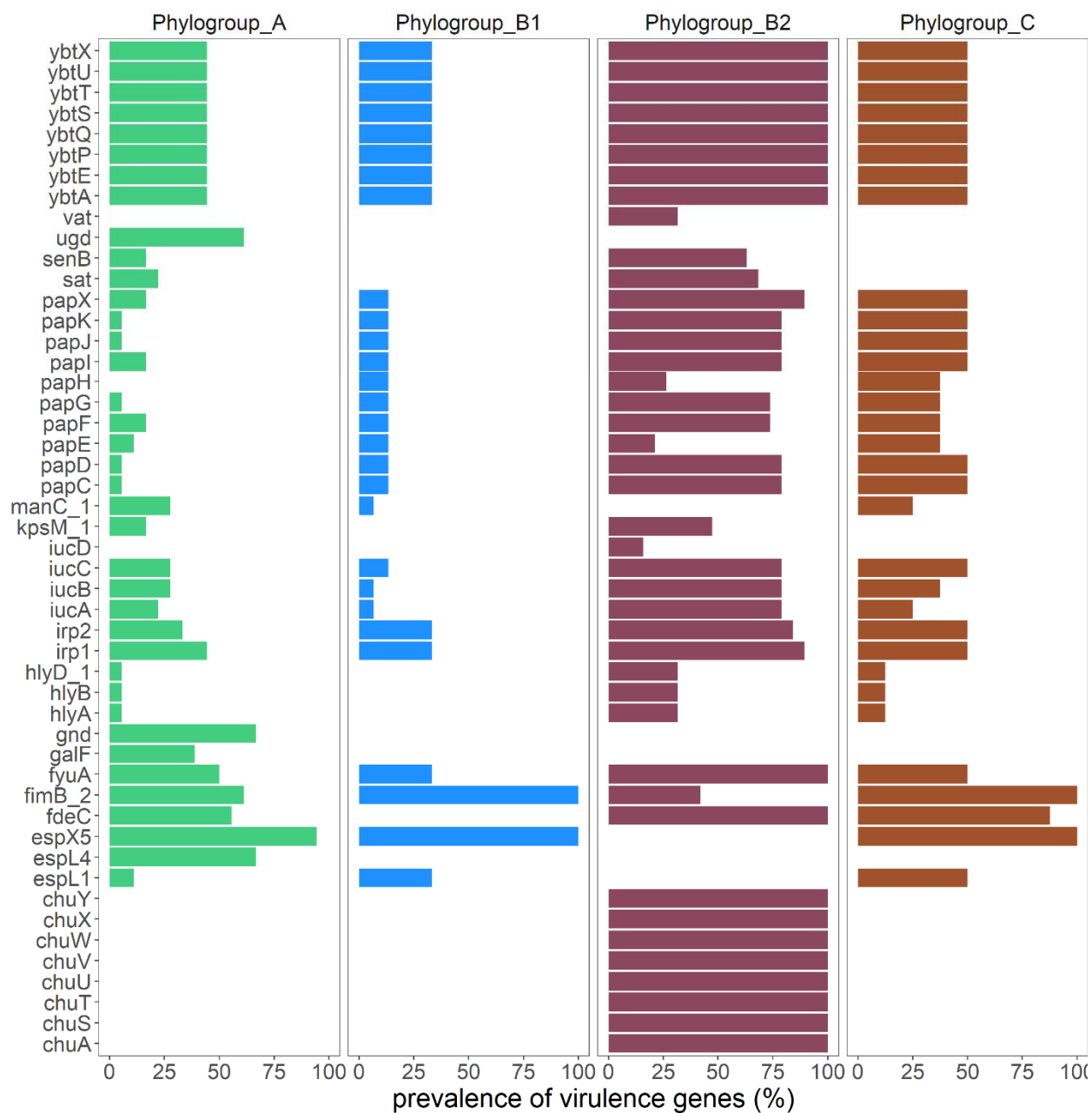
761 (b) The ST90 likely outbreak cluster of O8:H9 strains, showing resistance and plasmid profiles

762

763

764

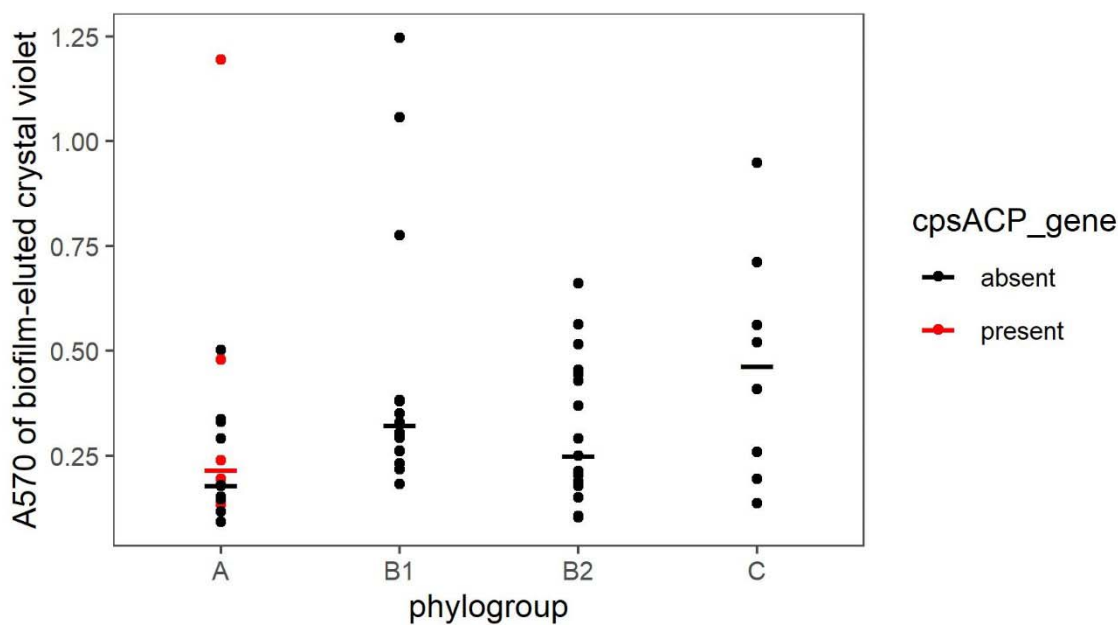
765 A.



766
767

768

769 B



770

771

772 Fig. 2: a. Comparison of virulence genes among isolates clustered within phylogroups A-C.
773 Graph shows the prevalence of virulence factors in isolates within phylogroups containing more
774 than five isolates: A (n = 18), phylogroup B1 (n = 15), phylogroup B2 (n = 19), and phylogroup
775 C (n = 8). Using Fisher's test ($p < 0.05$), only the VAGs significantly more prevalent in at least
776 one phylogroup are shown. B. Biofilm formation in 67 strains, measured as A570 nm of crystal
777 violet eluted from fixed and stained six-hour biofilms. Each dot represents data from a single
778 strain belonging to the phylogroup listed on the X axis. Horizontal bars mark the median for each
779 phylogroup, outliers inclusive due to the small number of tested strains in each phylogroup. The
780 A570 nm values for isolates carrying the cpsACP gene were represented as red dots.

781

782 A

783



784

785

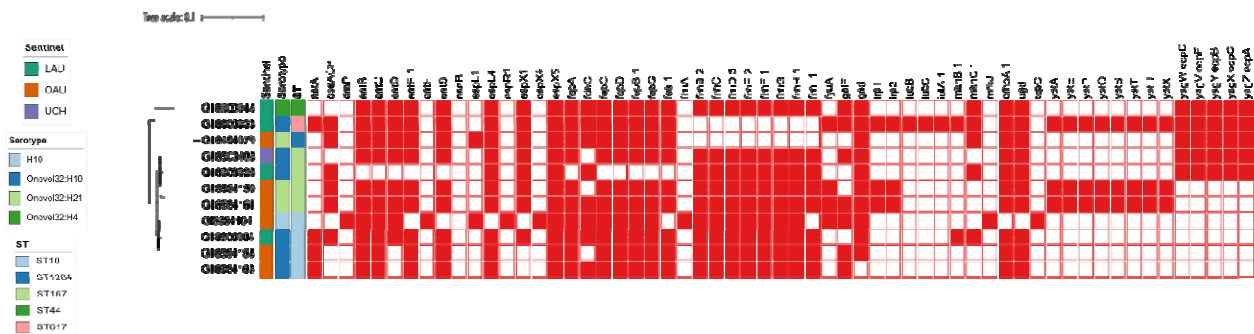
800 A



801

802

803 B



804

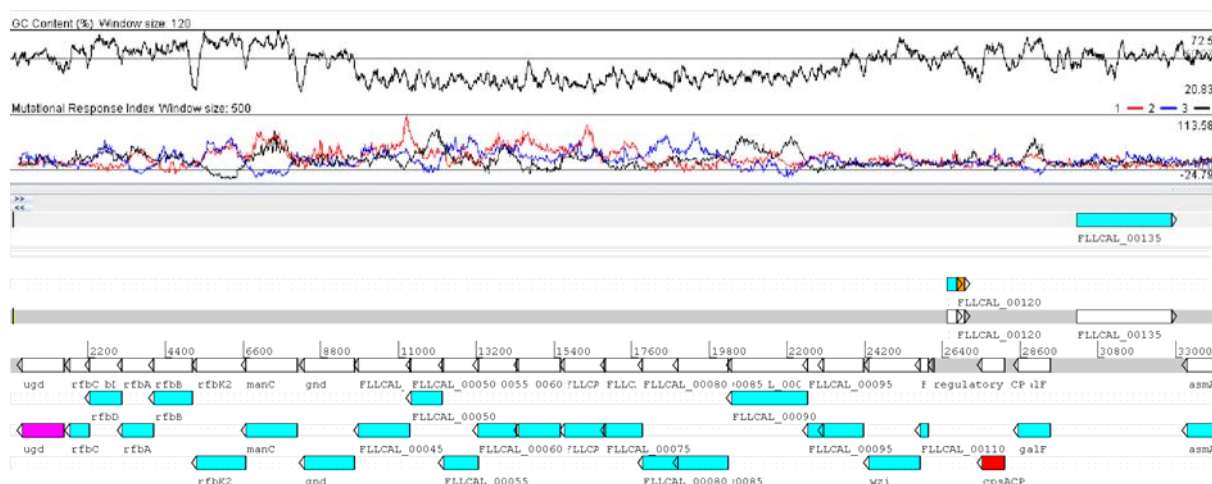
805

806 Fig. 4a: Comparison of virulence genes between Onovel32 (n = 11) and non-Onovel32 (n = 57)
807 clade isolates. Using Fisher's test ($p < 0.05$, Bonferroni-corrected), only the virulence genes
808 significantly more prevalent are shown. The cpsACP gene and the ugd gene bars were coloured
809 red and magenta, respectively.

810 b: Onovel32 clade isolates belonging to the sequence types ST10, ST167, and their locus variants
811 belong to phylogroup A. Leaf tip colours differentiate the sentinel sites; OAU (Red), UCH
812 (Blue), LAU (Yellow).

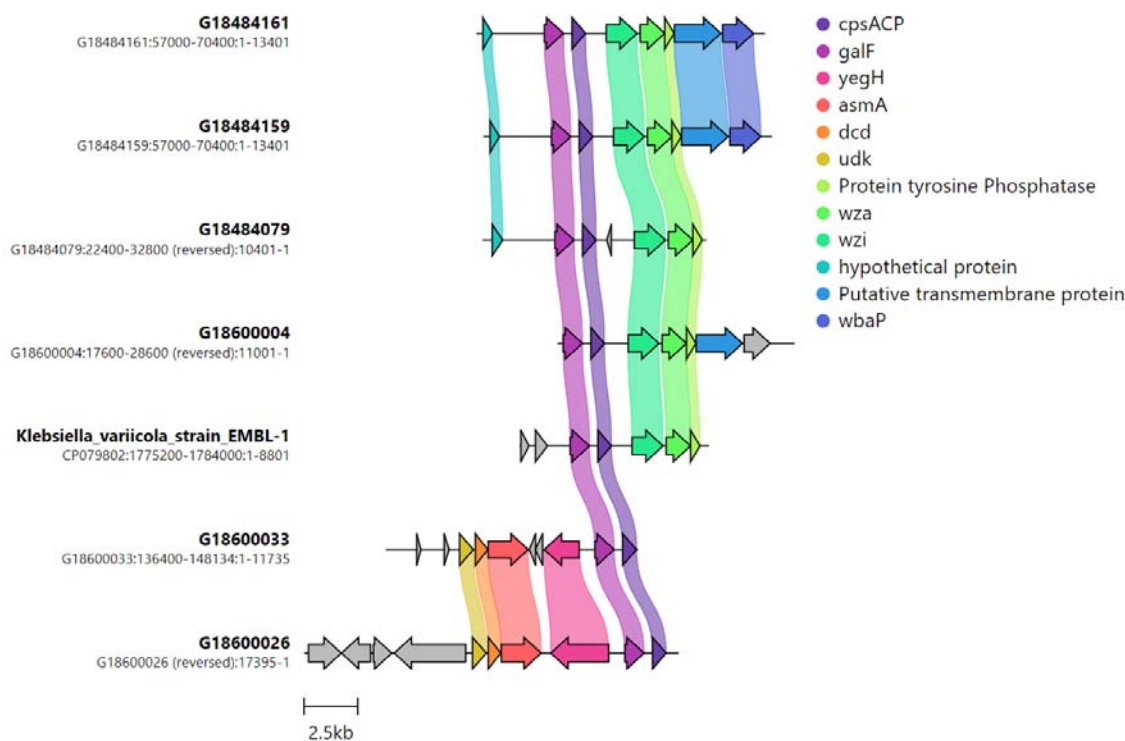
813

814 A



815

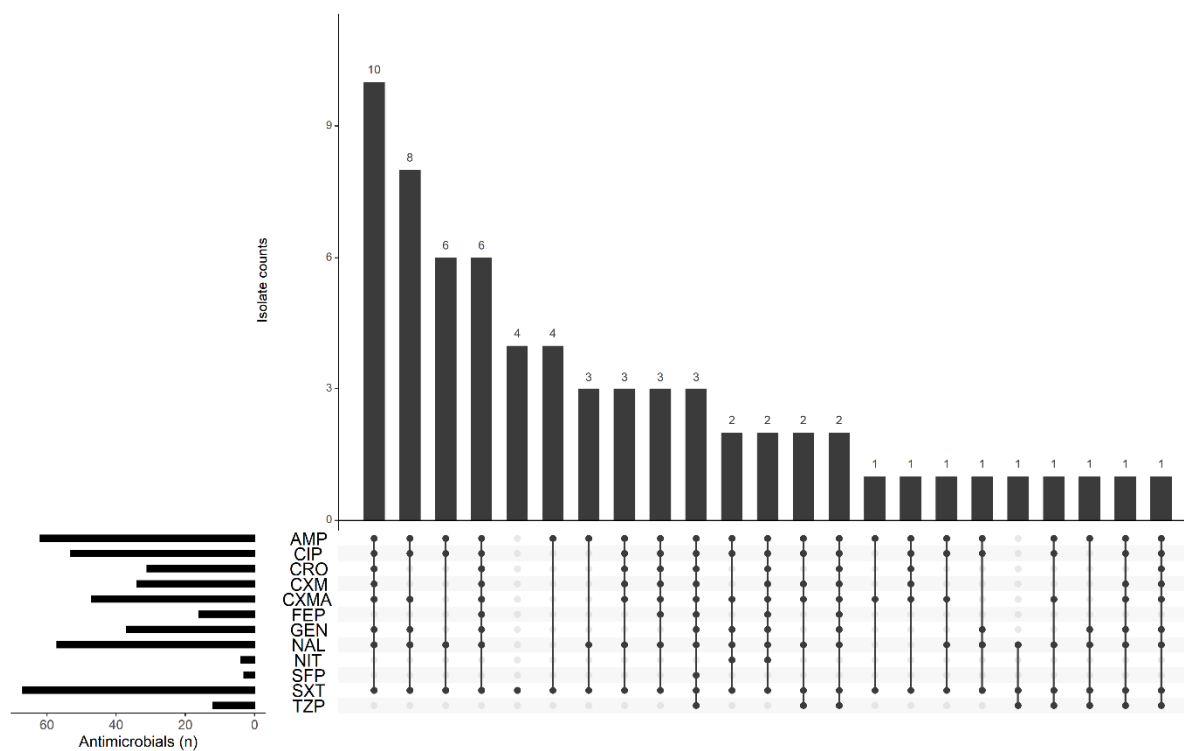
816 B



817

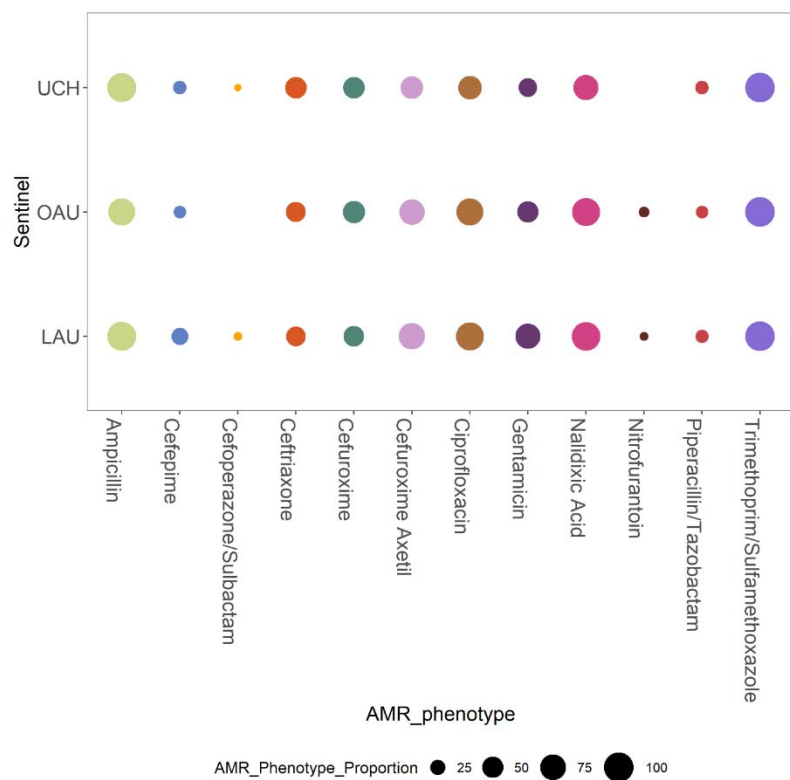
818 Figure 5a: Schematic cluster in ST1294 ONovel32 chromosome that includes the *ugd* and
 819 *cpsACP* genes. The genes within the cluster are syntenic and identical with a cluster from
 820 *Klebsiella variicola* (Accession number: [CP079802.1](https://doi.org/10.1093/nar/gkab111)). Above the schematic depiction of genes
 821 is a G+C content plot and mutational response index plot. (b) comparison of the ONovel32
 822 cluster with analogous cluster in *Klebsiella variicola* strain EMBL-1

823 A



824
825

B



826

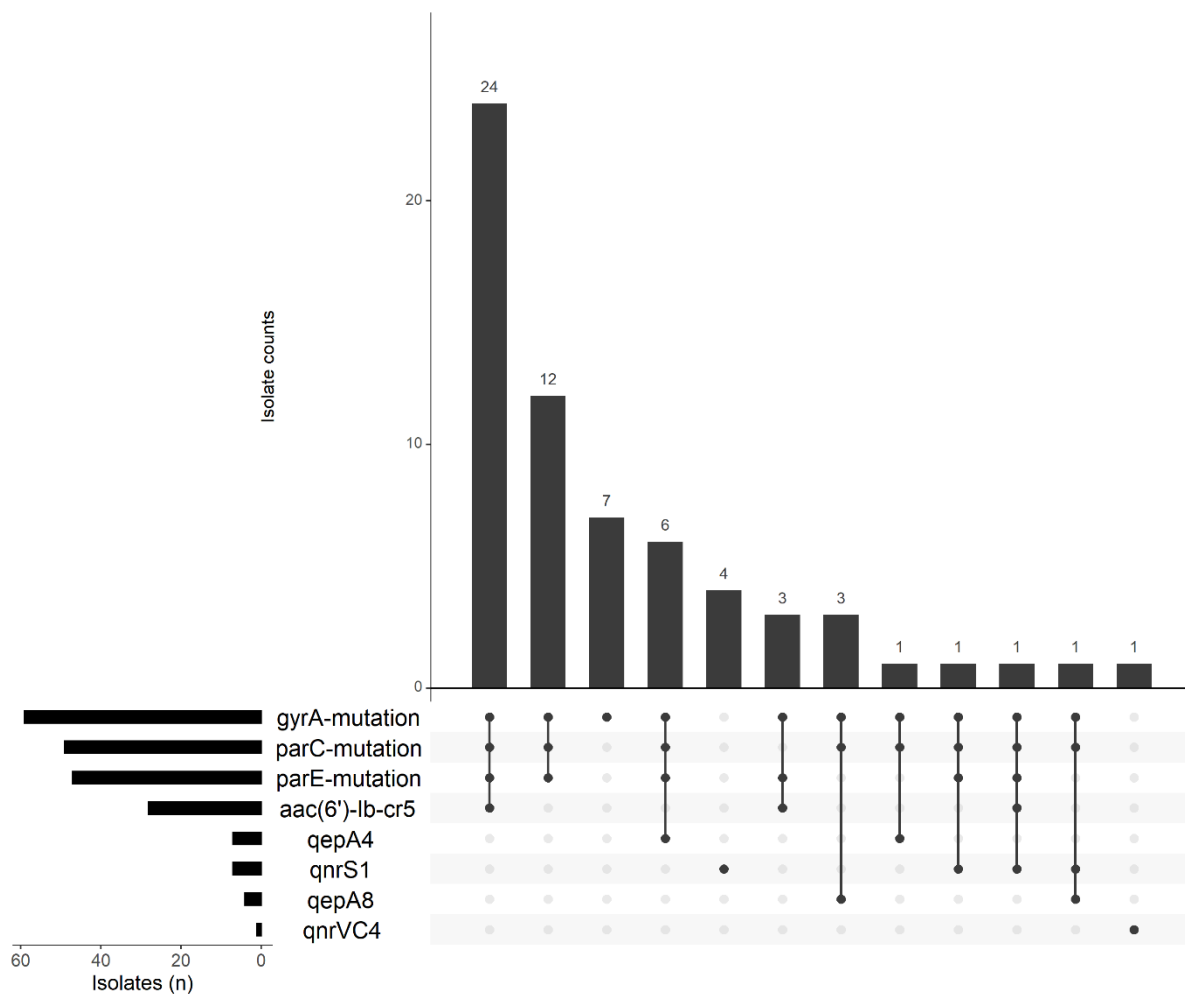
827 Fig. 6a: Resistance Profile of ExPEC isolates. The main bar chart demonstrates the number of
828 ExPEC isolates with each combination of resistance to tested antibiotics, and is ordered in

829 descending order by the frequency of resistance profiles observed among 67 ExPEC isolates. The
830 side bar chart demonstrates the number of isolates that are resistant to each of the named
831 antibiotics. The dots and lines between dots at the base of the main bar chart (and the right side
832 of the side bar chart) show the co-resistance status of the ExPEC isolates. All isolates were
833 susceptible to Amikacin, Meropenem, Imipenem, and Ertapenem.

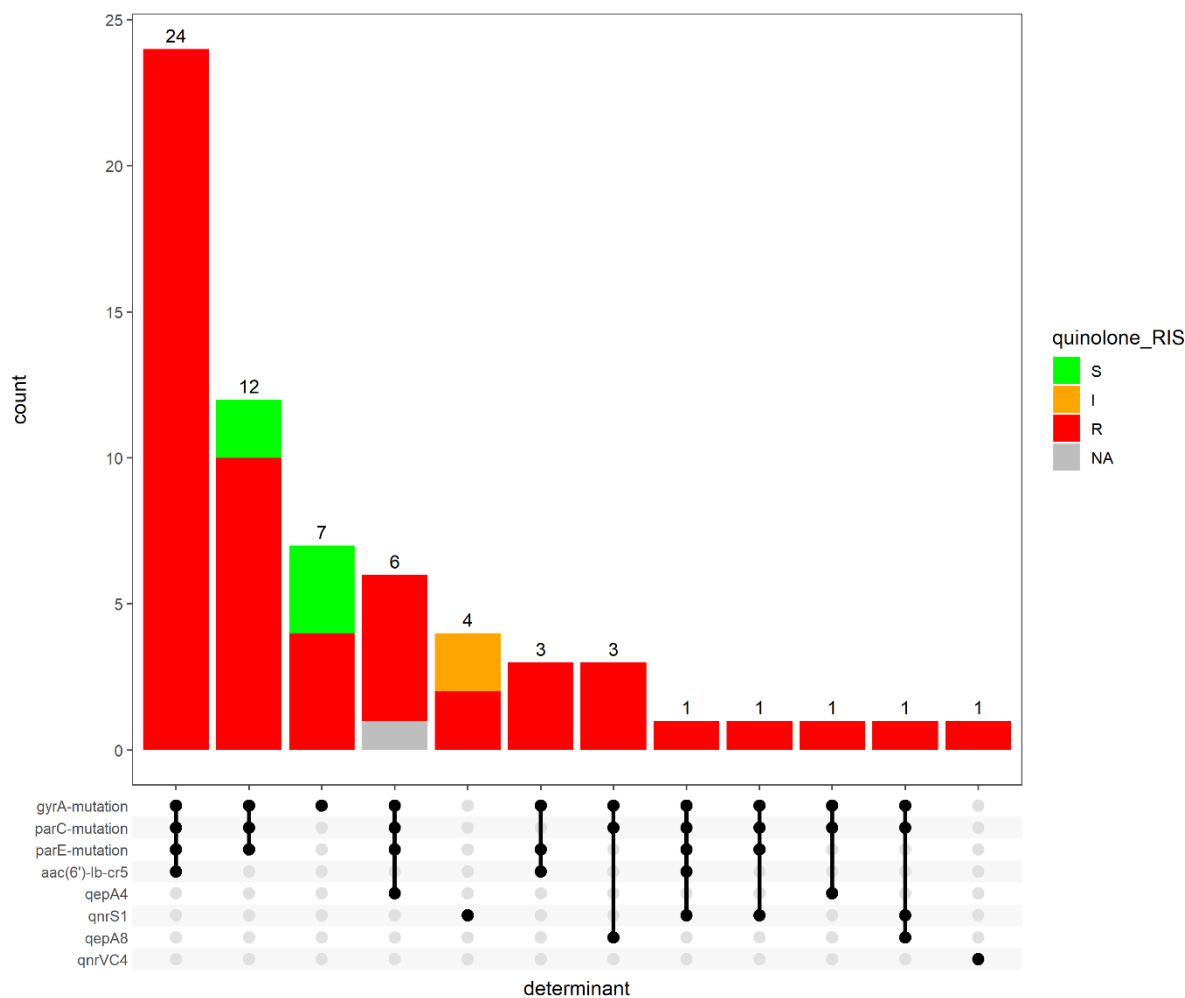
834 AMP = Ampicillin; CIP = Ciprofloxacin; CRO = Ceftriaxone; CXM = Cefuroxime; CXMA =
835 Cefuroxime Axetil; FEP = Cefepime; GEN = Gentamicin; NAL = Nalidixic Acid; NIT =
836 Nitrofurantoin; SFP = Cefoperazone/Sulbactam; SXT = Trimethoprim/Sulfamethoxazole; TZP =
837 Piperacillin/Tazobactam

838 Fig. 6b: AMR phenotypes of ExPEC Isolates, stratified by Sentinel Site (LAU (n = 28), OAU (n
839 = 22), UCH (n = 21)). The size of the coloured circles represent the proportion of isolates
840 recovered from each sentinel site that demonstrated resistance to the tested antibiotics.

841
842



843



844
 845 Fig. 7: Quinolone resistance gene combinations in ExPEC isolates. The upset plots show (a, b)
 846 the number of ExPEC isolates carrying each combination of genes conferring resistance to the
 847 quinolones (b), and is coloured by the proportion of observed phenotypic antimicrobial
 848 susceptibility, and is ordered in descending order by the frequency of resistance gene profiles
 849 observed. The side bar chart demonstrates the number of isolates that carry each of the resistance
 850 genes. The dots and lines between dots at the base of the main bar chart (and the right side of the
 851 side bar chart) show the co-resistance gene profile of the ExPEC isolates. From the legend of the
 852 coloured upset plot, NA means that phenotypic antimicrobial susceptibility data for the isolate is
 853 absent.

854
 855
 856
 857

858 **Supplementary Tables**

859 **Table S1: Resistance profiles of ExPEC isolates**

Resistance Profile (number)	Resistance Profile code
AMP CIP CRO CXM CXMA GEN NAL SXT (10)	RP-1 (MDR)
AMP CIP CXMA GEN NAL SXT (8)	RP-2 (MDR)
AMP CIP CRO CXM CXMA FEP GEN NAL SXT (6)	RP-3 (MDR)
AMP CIP NAL SXT (6)	RP-4 (MDR)
AMP SXT (4)	RP-5
SXT (4)	RP-6
AMP CIP CRO CXM CXMA FEP GEN NAL SFP SXT TZP (3)	RP-7 (MDR)
AMP CIP CRO CXM CXMA FEP NAL SXT (3)	RP-8 (MDR)
AMP CIP CRO CXM CXMA NAL SXT (3)	RP-9 (MDR)
AMP NAL SXT (3)	RP-10 (MDR)
AMP CIP CRO CXM CXMA FEP GEN NAL NIT SXT (2)	RP-11 (MDR)
AMP CIP CRO CXM CXMA FEP GEN NAL SXT TZP (2)	RP-12 (MDR)
AMP CIP CXM CXMA NAL SXT TZP (2)	RP-13 (MDR)
AMP CIP CXMA GEN NAL NIT SXT (2)	RP-14 (MDR)
AMP CIP CRO CXM CXMA GEN NAL SXT TZP (1)	RP-other (MDR)
AMP CIP CRO CXM CXMA SXT (1)	RP-other (MDR)
AMP CIP CXM CXMA GEN NAL SXT TZP (1)	RP-other (MDR)
AMP CIP CXMA NAL SXT (1)	RP-other (MDR)
AMP CIP CXMA NAL SXT TZP (1)	RP-other (MDR)
AMP CIP GEN NAL SXT (1)	RP-other (MDR)
AMP CXMA SXT (1)	RP-other (MDR)
AMP GEN NAL SXT TZP (1)	RP-other (MDR)
NAL SXT TZP (1)	RP-other (MDR)

860

861

862 **Table S2: Summary of Antibiotic Susceptibility Test data showing the number of *E. coli***
 863 **isolates that were resistant to the named antibiotics.**

Drug class	Drug	Resistance % (resistance (n)/species (n))
Aminoglycosides	Gentamicin (GEN)	55.20% (37/67)
Beta-lactams/penicillins	Ampicillin (AMP)	92.50% (62/67)
	Piperacillin/tazobactam (TZP)	17.90% (12/67)
Cephalosporins (2nd gen.)	Cefuroxime (CXM)	50.70% (34/67)
	Cefuroxime Axetil (CXMA)	70.10% (47/67)
Cephalosporins (3rd gen.)	Cefoperazone.Sulbactam (SFP)	4.50% (3/67)
	Ceftriaxone (CRO)	46.30% (31/67)
Cephalosporins (4th gen.)	Cefepime (FEP)	23.90% (16/67)
Other antibacterials	Nitrofurantoin (NIT)	6.00% (4/67)
Quinolones	Ciprofloxacin (CIP)	79.10% (53/67)
	Nalidixic acid (NAL)	85.10% (57/67)
Trimethoprim	Trimethoprim/sulfamethoxazole (SXT)	100.0% (67/67)

Table S3: Degree of Agreement between phenotypic antimicrobial susceptibility tests and genotypic predicted antimicrobial resistance

drug	Tests (N)	concordance (%)	TP	FP	TN	FN	specificity	Specificity lower limits	Specificity upper limits	sensitivity	Sensitivity lower limits	Sensitivity upper limits
Aminoglycosides	67	55.22	37	30	0	0	0	0	0.115703308	1	0.905109413	1
Beta-lactams	67	71.64	45	1	3	18	0.75	0.19412045	0.993690537	0.714285714	0.586538974	0.821070793
Cephalosporins	67	68.66	46	20	0	1	0	0	0.168433471	0.978723404	0.887062283	0.999461468
Quinolones	67	86.57	56	7	2	2	0.222222222	0.028144973	0.600093574	0.965517241	0.880922689	0.995796353
Trimethoprim	67	100	67	0	0	0	NA	0	1	1	0.946430346	1

(Table S3b). **Degree of Agreement between phenotypic antimicrobial susceptibility tests and genotypic predicted antimicrobial resistance in isolates belonging to the ST131, ST90, and ST10-167 clades.**

id	ST	Aminoglycosides AMR	Aminoglycosides AST	Betalactam AMR	Betalactam AST	Cephalosporins AMR	Cephalosporins AST	Quinolones AMR	Quinolones AST	Trimethoprim AMR	Trimethoprim AST
G18484079	1284	R	S	S	R	R	R	R	R	R	R
G18503180	131	R	R	S	R	R	R	R	R	R	R
G18503191	131	R	S	S	R	R	R	R	R	R	R
G18503409	131	R	R	R	R	R	R	R	R	R	R
G18503422	131	R	R	R	R	R	R	R	R	R	R
G18503432	131	R	R	R	R	R	R	R	R	R	R
G18584035	131	R	R	S	R	R	R	R	R	R	R
G18584046	131	R	R	S	R	R	R	R	R	R	R
G18584087	131	R	S	R	R	R	S	R	R	R	R
G18584153	131	R	R	S	R	R	R	R	R	R	R
G18584154	131	R	R	S	R	R	R	R	R	R	R
G18584155	131	R	S	R	S	R	S	R	S	R	R
G18584158	10	R	R	R	R	R	R	R	R	R	R
G18584159	167	R	R	R	R	R	R	R	R	R	R

G18584161	167	R	R	R	R	R	R	R	R	R	R
G18584163	10	R	R	R	R	R	R	R	R	R	R
G18600004	10	R	S	R	R	R	R	R	R	R	R
G18600024	131	R	R	S	R	R	R	R	R	R	R
G18600026	167	R	R	R	R	R	R	R	R	R	R
G18600033	617	R	R	S	R	R	R	R	R	R	R
G18600048	44	R	R	S	R	R	R	R	R	R	R
GI8584164	10	R	S	R	R	S	R	R	R	R	R
G18600002	90	R	R	R	R	R	R	R	R	R	R
G18600020	90	R	R	R	R	R	R	R	R	R	R
G18600022	90	R	R	R	R	R	R	R	R	R	R
G18600030	90	R	R	R	R	R	R	R	R	R	R

Table S3c: List of genes predicted to confer resistance to antibiotics belonging to the drug classes in Table S3a

Drug Classes	AMR genes
Aminoglycosides	aac(6')-Ie/aph(2'')-Ia, aac(6')-Ib-cr5, catB8/aac(6')-Ib', aac(6')-30/aac(6')-Ib', aac(3)-Iie, aac(3)-IId, aac(6')-Ib4, ant(2'')-Ia, aph(3')-Ia, aadA2, aph(3'')-Ib, aph(6)-Id, aadA5, aadA13, aadA1, aadA8
Beta-lactams	blaTEM-1, blaOXA-2, blaTEM-90, blaTEM-57, blaTEM-40, blaTEM-84, blaZ, blaR1, blaTEM-135
Cephalosporins	blaCTX-M-15, blaEC-15, blaEC-5, blaOXA-1, blaEC-16, blaEC-18, blaCTX-M-27, blaEC-8, blaCMY-2, blaEC-19, blaCMY-42, blaTEM-187, blaVEB-1
Quinolones	gyrA, parC, and pare mutations, qepA4, qnrS1, qepA8, qnrVC4
Trimethoprim	dfrA12, dfrA17, dfrA7, dfrA14, dfrA29, dfrF, dfrB4, dfrA5, dfrA1

Table S4: Genotypic characteristics of the most common STs in the invasive microbial population.

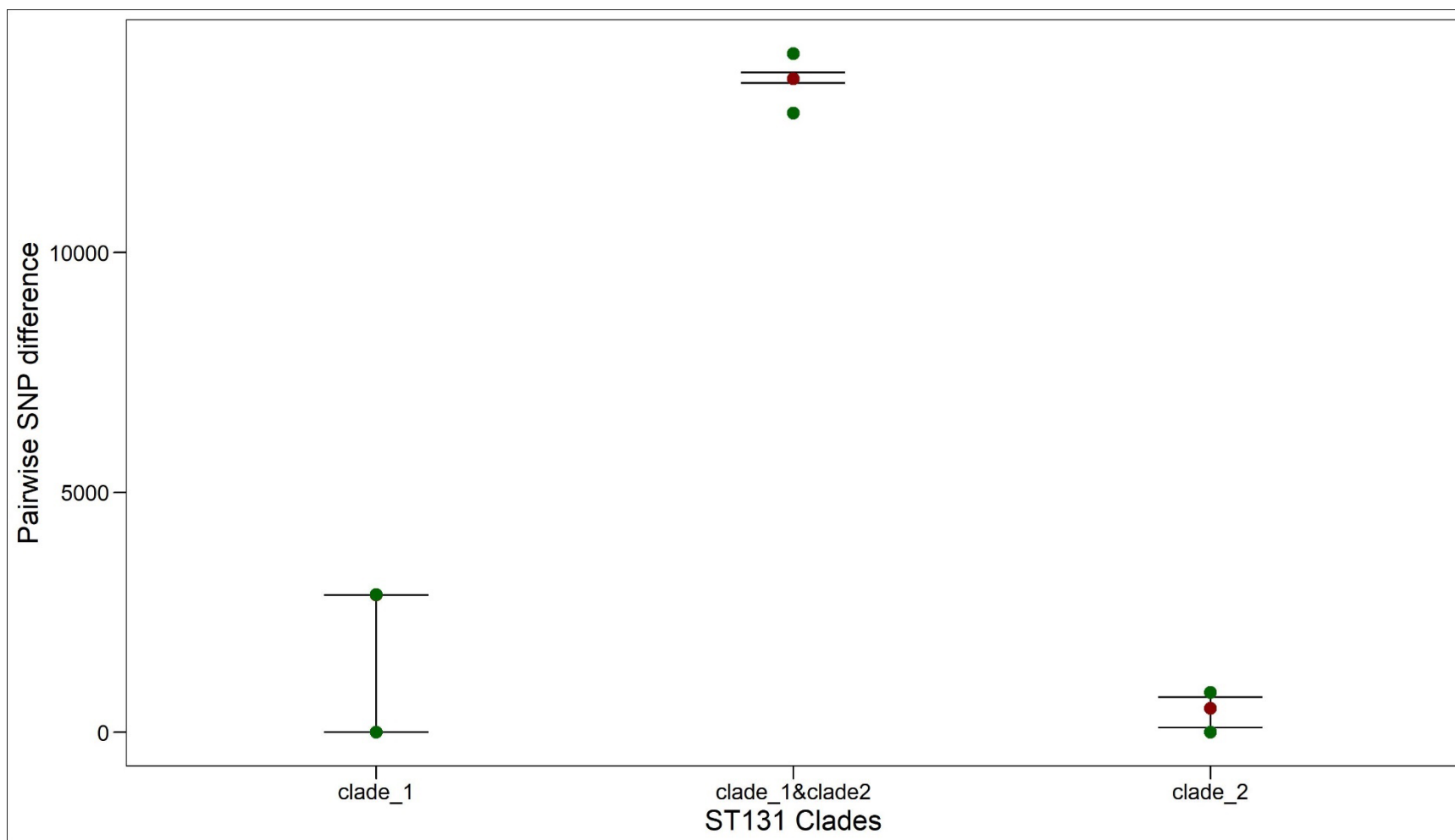
ST (number of strains)	Phylo- group	Sentinel Site (number)	Serotype (number)	Most Prevalent Resistance Profile (number)	Most Prevalent AMR gene profile (number)	Most prevalent plasmid replicon profile (number)
ST131 (12)	B2	OAU (6) UCH (5) LAU (1)	O16:H5 (5), O25:H4 (7)	RP-1: AMP CIP CRO CXM CXMA GEN NAL SXT (5)	aac(6')-Ib-cr5; Aac3-IIa*,Aac6Ib-cr*,AadA5; blaCTX-M-15,blaEC-5,blaOXA-1; CatB4_602; DfrA17; gyrA_D87N,gyrA_S83L,parC_E84V,parC_S80I,parE_I529L; MphA_1663; qacEdelta1; SulI; TetA (n = 3) and aac(6')-Ib-cr5; Aac3-IIId*,Aac6Ib-cr*,AadA5,StrA*,StrB*; blaCTX-M-15,blaEC-5,blaOXA-1; blaTEM-40; CatB4; DfrA17; gyrA_S83L,parE_I529L; MphA; qacEdelta1; SulII,SulI; TetA (3)	Col156; IncB_O_K_Z; IncFI; IncFIB_AP001918 (3) and Col156; IncFI; IncFIA; IncFIB_AP001918 (3)
ST10 (4)	A	OAU (3) LAU (1)	Onovel32:H10 (3), H10 (1)	RP-14: AMP CIP CXMA GEN NAL NIT SXT (2)	Each isolate had a unique AMR gene profile	IncFI; IncFIA; IncFIB_AP001918; IncQ1 (2)
ST156 (5)	B1	LAU (4) UCH (1)	H45 (1), O174:H28 (1), O54:H28 (2),	RP-2: AMP CIP CXMA GEN NAL SXT (3)	Aac3-IIId*,StrA*,StrB; blaEC-18; blaTEM-1; DfrA17; gyrA_D87N,gyrA_S83L,parC_S80I; qepA8; SulII*; TetB* (3)	IncFIA; IncFIB; IncQ1 (3)

		O76:H30 (1)							
ST167 (4)	A	OAU (2) LAU (1) UCH (1)	Onovel32:H 10 (2), Onovel32:H 21 (2)	Each isolate had a unique AST profile	Aac3-IId*; blaCTX-M-15,blaEC-15; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; TetB* (2)	blaTEM-1;	IncFI; IncFIA; IncFIB_AP001918 (3)		
ST410 (4)	C	LAU (2) UCH (2)	O8:H9 (2), Onovel14:H 9 (2)	RP-9: AMP CIP CRO CXM CXMA NAL SXT (2)	Each isolate had a unique AMR gene profile		Col_MG828; IncFI; IncFIA; IncFIB_AP001918 (2)		
ST648 (4)	F	LAU (3) OAU (1)	O153var1:H 6 (1), O8:H4 (3)	RP-7: AMP CIP CRO CXM CXMA FEP GEN NAL SFP SXT TZP (2)	aac(6')-Ib-cr5; Aac3-IIa*,Aac6Ib-cr*,AadA5; 42,blaCTX-M-15,blaEC-19,blaOXA-1; CatA1*,CatB4; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; MphA; qacEdelta1; SullI; TetB* (2).	blaCMY- blaTEM-1;	Col_BS512; Col_BS512__1; IncFIA; IncFIB_AP001918; IncFII_p; IncI1 (2)		
ST90 (4)	C	LAU (4)	O8:H9 (4)	RP-2: AMP CIP CXMA GEN NAL SXT (3)	aac3-IId, aac6Ib-cr, strA, strB; blaOXA-1; gyrA_D87N, gyrA_S83L, parC_S80I, parE_S458A; dfrA17; mphA; sullI; tetB (4).	blaTEM-1; blaEC-16,	IncFI; IncFIA; IncFIB_AP001918; IncQ1 (4)		

Table S5: Genotypic characteristics of Onovel32 genomes in the invasive microbial population.

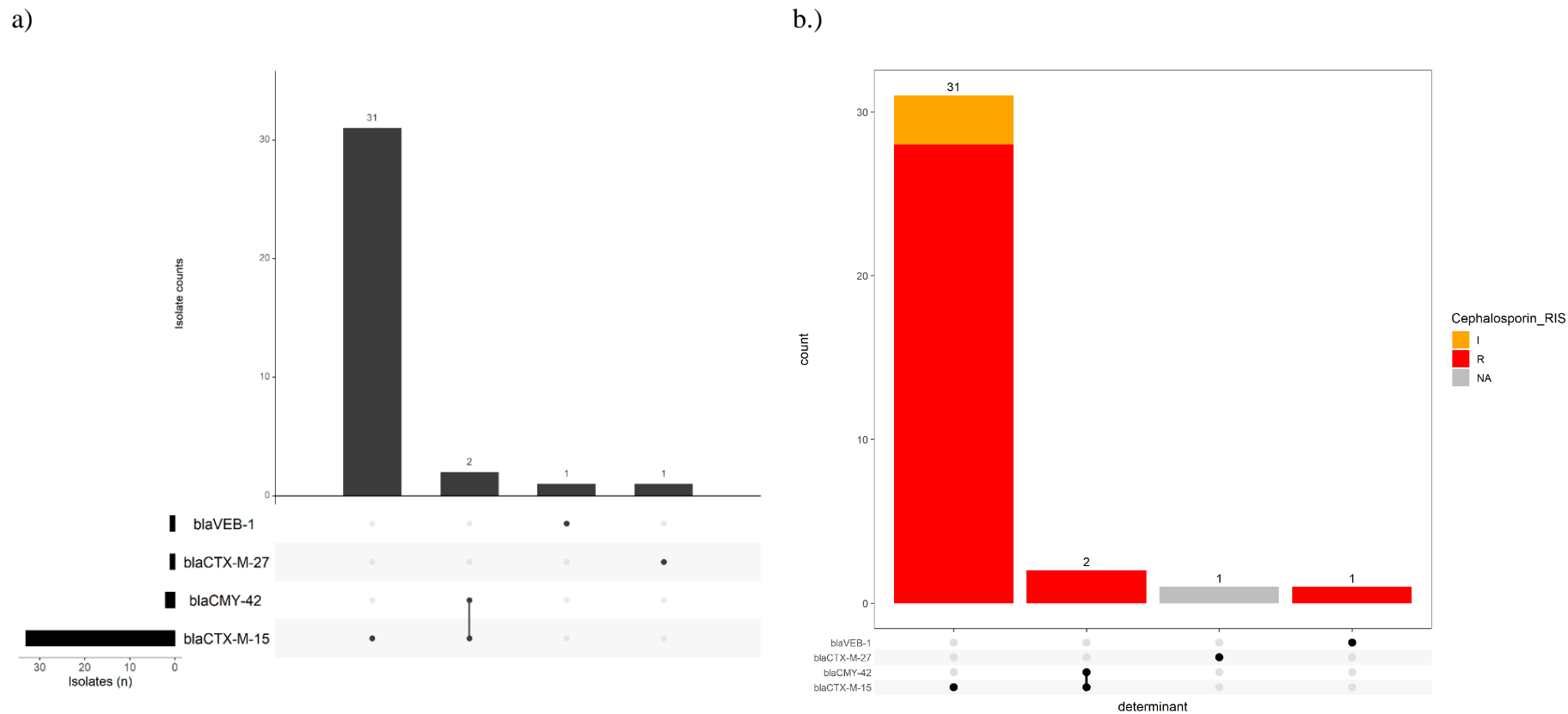
id	Sentinel	ST	Serotype	Resistance gene profile	plasmid_profile
G18484079	OAU	1284	Onovel32:H21	AadA2*, StrA*?,StrB; blaCTX-M-15,blaEC-15; CatA2*; DfrA12; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; MphA; qacEdelta1; qepA4; SulII, SullI; TetB*	IncFI; IncFIA; IncFIB_AP001918; IncX1; IncX1_1
G18503408	UCH	167	Onovel32:H10	AadA2*,StrA*,StrB*; blaCTX-M-27, blaEC-15; CatA1*,CatA2*; DfrA12; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; MphA*?; qacEdelta1; qepA4; SulII, SullI; TetA,TetB*,TetM*	IncFI; IncFIA; IncFIB_AP001918; IncX2
G18584158	OAU	10	Onovel32:H10	aac(6)-Ib-cr5; Aac3-IId*,Aac6Ib-cr*,AadA5,StrA*,StrB; blaOXA-1; blaTEM-84; CatA1*,CatB4; DfrA17; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; MphA; qacEdelta1; SulII*; TetB*	IncFI; IncFIA; IncFIB_AP001918; IncQ1
G18584159	OAU	167	Onovel32:H21	Aac3-IId*; blaCTX-M-15, blaEC-15; blaTEM-1; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; TetB*	IncFI; IncFIA; IncFIB_AP001918
G18584161	OAU	167	Onovel32:H21	Aac3-IId*; blaCTX-M-15, blaEC-15; blaTEM-1; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; TetB*	IncFI; IncFIA; IncFIB_AP001918
G18584163	OAU	10	Onovel32:H10	aac(6)-Ib-cr5; Aac3-IId*,Aac6Ib-cr*,AadA5,StrA*,StrB; blaOXA-1; blaTEM-84; CatA1*,CatB4*; DfrA17; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A;	IncFI; IncFIA; IncFIB_AP001918; IncQ1

				MphA; qacEdelta1; SulII*; TetB*	
G18600004	LAU	10	Onovel32:H10	AadA1*?,StrA*,StrB*; blaOXA-1; blaTEM-1; DfrA14; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; qepA4; SulII	IncFI; IncFIA; IncFIB_AP001918
G18600026	LAU	167	Onovel32:H10	Aac3-IIId*; blaEC-15; blaTEM-1; CatA1*; DfrA12; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; MphA; qacEdelta1; qepA4; SulI; TetB*	IncFI; IncFIA; IncFIB_AP001918
G18600033	LAU	617	Onovel32:H10	aac(6')-Ib-cr5; Aac3-IIa*,Aac6Ib-cr*,AadA5,StrA*?,StrB; blaCTX-M-15,blaEC-15,blaOXA-1; CatB4; DfrA17; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; MphA; qacEdelta1; SulII, SulI; TetB*	IncFI; IncFIA; IncFIB_AP001918
G18600048	LAU	44	Onovel32:H4	aac(6')-Ib-cr5; Aac3-IIa*,Aac6Ib-cr*,AadA5; blaCTX-M-15,blaOXA-1; CatB4; DfrA17; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458T; qacEdelta1; SulI; TetB*	Col_MG828; IncFI; IncFIB_AP001918; IncX3
GI8584164	OAU	10	H10	AadA2*,StrA*,StrB,CatA1*; blaTEM-1; DfrA12; gyrA_D87N,gyrA_S83L,parC_S80I,parE_S458A; MphA,QepA2*; qacEdelta1; qepA4; SulII*,SulI; TetB*	ColRI; Col_BS512; IncP_6



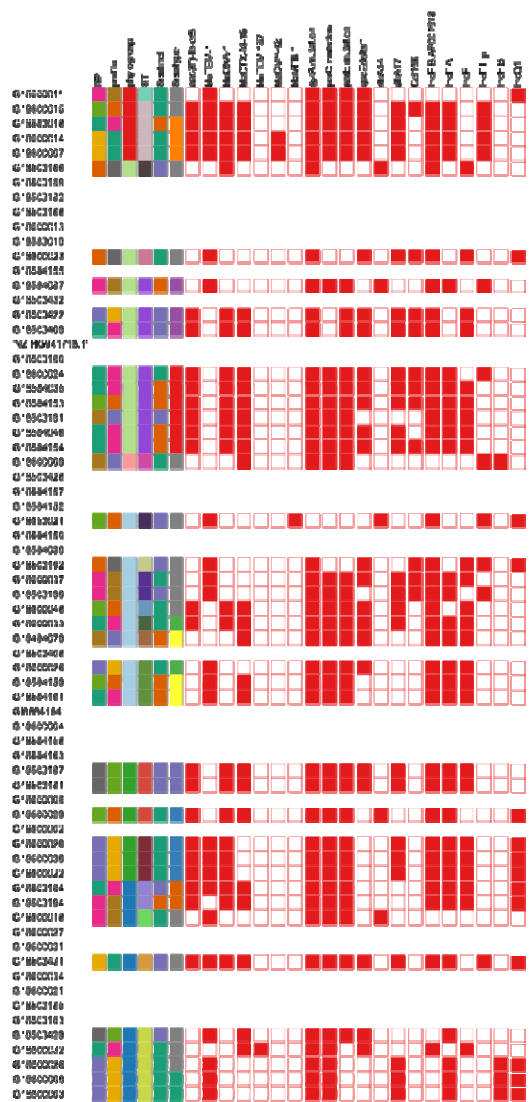
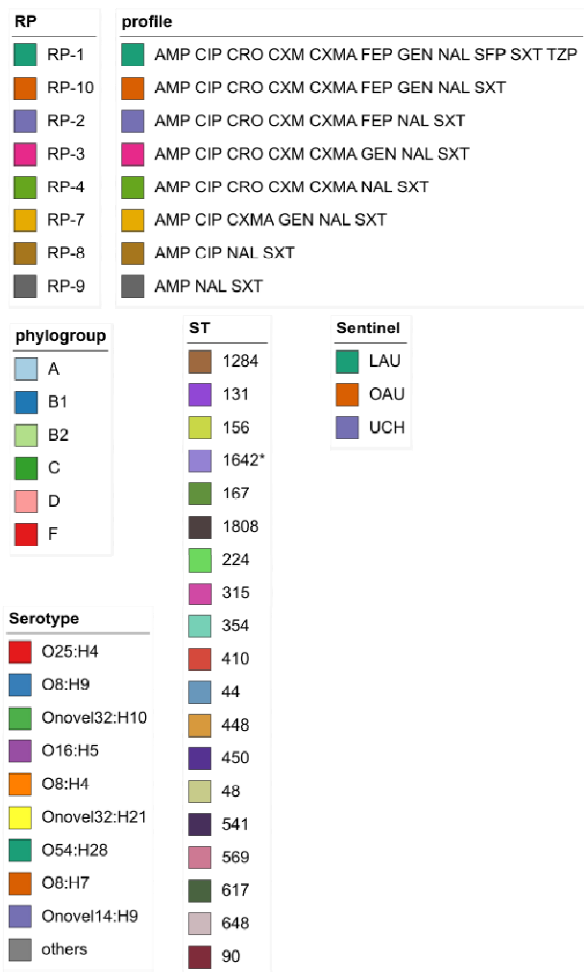
Supplementary Figure 1

Interclade (ST131 clade 1 and ST131 clade 2) distance between lineage ST131 isolates is greater than 10000 SNPs. The “whisker” plot shows the 25th percentile (lower point; green), median (red point), and the 75th percentile (upper point; green) of the pairwise SNP distances within and between lineages.



Supplementary Figure. 2: ESBL genes in ExPEC isolates. The upset plots demonstrate (a, b) the number of ExPEC isolates carrying each combination of genes conferring resistance to cephalosporins and (b), is coloured by the proportion of observed phenotypic antimicrobial susceptibility, and is ordered in descending order by the frequency of resistance gene profiles observed. The side bar chart (a) demonstrates the number of isolates that carry each of the resistance genes. The dots and lines between dots at the base of the main bar chart (and the right side of the side bar chart) show the co-resistance gene profile of the ExPEC isolates. From the legend of the coloured upset plot, NA means that there is no available data on phenotypic antimicrobial susceptibility data for the isolate.

Tree scale: 0.1



Supplementary Figure 3: Maximum likelihood SNP tree of bloodstream *E. coli* isolates sequenced for this study. Metadata for the 8 most common resistance profiles are shown in this figure. For the AMR genes and plasmid replicons, red colour indicates presence of the gene.