

Complete genomic characterization of two *Escherichia coli* lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital

Zong, Zhiyong; Fenn, Samuel; Connor, Christopher; Feng, Yu; McNally, Alan

DOI:
[10.1093/jac/dky210](https://doi.org/10.1093/jac/dky210)

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zong, Z, Fenn, S, Connor, C, Feng, Y & McNally, A 2018, 'Complete genomic characterization of two *Escherichia coli* lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital', *Journal of Antimicrobial Chemotherapy*. <https://doi.org/10.1093/jac/dky210>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Journal of Antimicrobial Chemotherapy* following peer review. The version of record Zhiyong Zong, Samuel Fenn, Christopher Connor, Yu Feng, Alan McNally; Complete genomic characterization of two *Escherichia coli* lineages responsible for a cluster of carbapenem-resistant infections in a Chinese hospital, *Journal of Antimicrobial Chemotherapy*, , dky210, <https://doi.org/10.1093/jac/dky210> is available online at: [10.1093/jac/dky210](https://doi.org/10.1093/jac/dky210)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

1 **Complete genomic characterisation of two *Escherichia coli* lineages**
2 **responsible for a cluster of carbapenem resistant infections in a Chinese**
3 **hospital**

4

5 Zhiyong ZONG, Samuel FENN, Christopher CONNOR, Yu FENG, Alan McNALLY*

6 ¹Centre for Infectious Diseases, West China Hospital of Sichuan University,
7 Chengdu, China

8 ²Institute of Microbiology and Infection, College of Medical and Dental Science,
9 University of Birmingham, Birmingham, United Kingdom, B15 2TT

10

11 *Corresponding author: Dr Alan McNally, Institute of Microbiology and Infection,
12 College of Medical and Dental Science, University of Birmingham, Birmingham B15
13 2TT. 0044 121 4158433. a.mcnally.1@bham.ac.uk

14

15 Running title: Carbapenem resistant clones of *E. coli*

16

17 **Abstract**

18 Objectives: The increase in infections as a result of multi-drug resistant strains of
19 *Escherichia coli* is a global health crisis. The emergence of globally disseminated
20 lineages of *E. coli* carrying ESBL genes has been well characterised. An increase in
21 strains producing carbapenemase enzymes and mobile colistin resistance is now
22 being reported, but to date there is little genomic characterisation of such strains.

23 Methods: Routine screening of patients within an ICU of West China Hospital
24 identified a number of *E. coli* carrying the *bla*_{NDM-5} carbapenemase gene, found to be
25 two distinct clones, *E. coli* ST167 and ST617.

26 Results: Interrogation of publically available data shows isolation of ESBL and
27 carbapenem resistant strains of both lineages from clinical cases across the world.
28 Further analysis of a large collection of publically available genomes shows that
29 ST167 and ST617 have emerged in distinct patterns from the ST10 clonal complex
30 of *E. coli*, but share evolutionary events involving switches in LPS genetics,
31 intergenic regions and anaerobic metabolism loci.

32 Conclusions: The identification of these lineages of *E. coli* and their shared genetic
33 traits suggest there may be evolutionary events which underpin the emergence of
34 carbapenem resistance plasmid carriage in *E. coli*.

35

36 **Introduction**

37 Infections from multi-drug resistant (MDR) *Escherichia coli* are a significant global
38 health care threat.¹ MDR in *E. coli* is largely confined to strains capable of causing
39 extra-intestinal infections (ExPEC) such as urinary tract infections (UTI) and
40 bacteraemia.¹⁻⁴ As many as 50% of *E. coli* strains isolated from UTI and
41 bacteraemia cases may exhibit resistance to three or more classes of antibiotic,
42 termed MDR. This resistance is primarily driven by the acquisition of large plasmids
43 containing multiple resistance genes.² The rapid global dissemination of MDR *E. coli*
44 is associated with carriage of plasmids containing genes encoding extended-
45 spectrum β -lactamases (ESBL) which confer resistance to third-generation
46 cephalosporins.⁵ The carriage of MDR plasmids containing ESBL genes renders *E.*
47 *coli* susceptible only to the carbapenem class of antibiotics and the antimicrobial
48 compound colistin.⁵ However strains of *E. coli* are now being reported with plasmids
49 containing β -lactamases conferring resistance to carbapenems (carbapenemases)
50 and the *mcr-1* colistin resistance gene.⁶⁻⁹

51 The global dissemination of ESBL *E. coli* is attributable to the rapid dispersal of a
52 small number of *E. coli* lineages. The most dominant of these is the ST131 lineage
53 which is predominantly associated with carriage of the *bla*_{CTX-M-15} ESBL gene.²
54 ST131 is an ExPEC lineage and the most common cause of UTI and bacteraemia in
55 the developed world.² Other dominant lineages of ESBL *E. coli* are ST73, ST95, and
56 ST648 which are also ExPEC.^{3,4} ESBL carriage can also be found transiently in
57 strains belonging the ST10 clonal complex of *E. coli*.³ ST10 complex strains are
58 host generalist *E. coli* which are frequently found as intestinal commensal inhabitants
59 of mammals and avian species,¹⁰ and are devoid of the virulence-associated genes
60 known to be required for pathogenesis.¹¹ Our knowledge of the genomic landscape

61 of carbapenemase production in *E. coli* is far less developed, with the vast majority of
62 reports being genomes of individual clinical isolates sporadically distributed across
63 the globe. Just one significant publication exists reporting a specifically designed
64 genomic analysis of a temporal collection of carbapenem resistant *E. coli* which
65 showed very wide dissemination of carbapenem resistance across species and
66 within-species lineages of the enterobacteriaceae.¹²

67 Here we report the isolation of *E. coli* containing the carbapenem-resistance gene
68 *bla*_{NDM-5} in an ICU ward in West China Hospital, Chengdu. The isolates do not
69 belong to one of the dominant MDR lineages of ExPEC, but to ST167 and ST617,
70 both members of the ST10 clonal complex. Genomic data supports the long-term
71 presence of these bacteria in the ICU with repeated dissemination from a central
72 reservoir. Contextualisation of the Chinese strains with a collection of publically
73 available genomes shows isolation of MDR ST167 and ST617 strains from clinical
74 episodes across the world, and in the case of ST167 frequent occurrence of carriage
75 of both ESBL and carbapenemase genes. By comparing these lineages to a large
76 number of publically available ST10 genomes we identify potentially significant
77 events in their evolutionary trajectories, including mutations in the LPS biosynthesis
78 locus which truncate LPS. We also find evidence of compensatory mutations in
79 intergenic regions as found in *E. coli* ST131 as well as mutations in anaerobic
80 metabolism loci. Our findings support the need for a more concerted global
81 surveillance effort focussing on identifying frequently occurring lineages of
82 carbapenem resistant *E. coli*.

83 **Methods**

84 **Bacterial isolation and characterisation**

85 Strain 0215 was recovered from a rectal swab of a 75-year-old male patient on
86 September 2013 in a 50-bed medical ICU at West China Hospital, Chengdu, during
87 routine screening that is performed as standard in the ICU on all new admissions.
88 During a 7-month period from May to November 2014, *bla*_{NDM-5} positive *E. coli* were
89 recovered from the rectal swabs of 8 different patients (Supplementary Table S1)
90 from a total of 560 patients admitted to the ICU during this period. Furthermore, one
91 of the 8 patients developed bacteraemia during his ICU stay and an *E. coli* was
92 recovered from his blood and included in the study. During the study period, two
93 additional *E. coli* clinical isolates carrying *bla*_{NDM-5} were recovered in the hospital,
94 from two patients on admission. Rectal swabs were collected from patients within 2
95 days of admission to the ICU and within the 3 days prior to ICU discharge for those
96 patients with a length of stay of 3 days or more. Swabs were transferred to the
97 laboratory in transport media and were screened for carbapenem-resistant
98 Enterobacteriaceae using the CHROMAgar Orientation agar plates containing 2
99 µg/mL meropenem.

100 **Ethics**

101 This study was conducted in accordance with the amended Declaration of Helsinki
102 and was approved, under a waiver of consent, by the Ethics Committee of West
103 China Hospital. Rectal swabs were collected from patients within 2 days of
104 admission to the ICU and within the 3 days prior to ICU discharge for those patients
105 with a length of stay of 3 days or more.

106 **Genome sequencing**

107 The ST167 and ST617 strains isolated in Chengdu were cultured in LB broth at 37°C
108 overnight. DNA was extracted using QIAamp[®] DNA Mini Kit (QIAGEN) and 150 bp
109 paired-end libraries of each strain prepared and sequenced using the Illumina HiSeq

110 X-Ten platform (raw data accession numbers Table S2 and S3). Genomes were
111 assembled using SPAdes¹³ and annotated using Prokka.¹⁴ The MLST sequence
112 type of the strains was determined using the in silico prediction tool MLSTFinder.¹⁵
113 The *E. coli* genome database Enterobase (www.enterobase.warwick.ac.uk) was
114 interrogated on 1st December 2016 and all available ST167 and ST617 genomes
115 were downloaded (Table S2 and S3) and annotated using Prokka. A further 256
116 ST10 genomes were selected to represent the geographical, temporal, and source
117 attribution diversity present in the database (Table S4) and were downloaded and
118 annotated using Prokka. To select these genomes a phylogenetic tree was inferred
119 from the assembled genome of every ST10 on Enterobase using Parsnp.¹⁶ From
120 this phylogeny 500 genomes were chosen to span the entire phylogenetic diversity,
121 and then the final selection made to represent the full ST10 diversity as described.
122 The antibiotic resistance gene profile of all isolates was determined using Abricate
123 (<https://github.com/tseemann/abricate>).

124 **High-resolution SNP analysis**

125 We created a closed genome sequence for a Chinese ST167 strain 1237 by
126 combining our Illumina sequence data with data generated on the Minlon sequencer.
127 Raw Minlon reads were converted into fastQ format (accession number
128 PRJNA422975) using Poretools¹⁷ and assembled using Canu,¹⁸ resulting in a
129 single contig chromosome and four distinct single contig plasmids. The raw illumina
130 data was then used to polish the genome assembly via five iterative rounds of
131 polishing with Pilon.¹⁹ The ST167 and ST617 genomes from Chengdu were
132 analysed by mapping raw reads against the hybrid assembled ST167 genome.
133 Mapping was performed using Snippy (<https://github.com/tseemann/snippy>) and the
134 resulting SNP profiles were used to create a consensus sequence for each genome

135 which was aligned using the parsnp alignment tool in Harvest.¹⁶ Analysis of the
136 plasmid containing the *bla*_{NDM-5} gene revealed that it was a 47-kb IncX3 plasmid and
137 there were no antibiotic resistant genes other than *bla*_{NDM-5} located on the plasmid.
138 Specific mapping of the raw Illumina data against the pNDM5 plasmid was
139 performed for all strains as described above.

140 **Phylogenetic analysis**

141 Pan-genomes were constructed for the ST167, ST617, ST10, and combined
142 datasets using Roary²⁰ with the --e --mafft setting to create a concatenated
143 alignment of core CDS. The alignments were used to infer ST167, ST617, ST10, and
144 combined phylogenies using RaxML²¹ with the GTR-Gamma model of site
145 heterogeneity and 100 bootstrap iterations. Carriage of ESBL and carbapenemase
146 genes was annotated on the trees using Phandango
147 (<https://jameshadfield.github.io/phandango/>), and geographical source was
148 annotated using iTOL.²²

149 **Detection of lineage specific genetic traits**

150 Microbial GWAS was performed using two approaches. First the combined data set
151 pan-genome matrix was used as input for Scoary²³ searching for loci unique to
152 ST167, ST617, and both ST167 and ST617 versus ST10. In parallel we also used
153 SEER²⁴ to detect kmers significantly associated with ST167, ST617, or both
154 combined versus ST10. The results of both approaches were combined to identify
155 coding loci associated with the emergence of ST167 and ST617. In silico serotyping
156 was performed using two independent methods, SRST2 and SerotypeFinder.^{25,26}
157 Both methods utilise WGS data to specific O and H antigens to strains. Intergenic
158 regions (IGRs) were investigated using Piggy²⁷ to search for IGRs which had
159 switched²⁸ in ST617, ST167, or both compared to ST10. This data was combined

160 with SEER data to identify high-confidence IGR switches associated with the
161 emergence of ST167 and ST617.

162 **Results**

163 **Presence of *E. coli* ST167 and ST617 strains containing the NDM-5** 164 **carbapenemase resistance gene in an ICU ward in West China Hospital.**

165 A total of ten isolates of *E. coli* containing *bla*_{NDM-5} were obtained during the
166 investigation. Nine of these isolates belonged to sequence types ST167/617 (Table
167 S1), which are members of the ST10 complex of *E. coli* most commonly associated
168 with mammalian intestinal commensal carriage. Three ST167 isolates (0215, 243
169 and 25) were obtained from swabs or clinical samples collected on admission to
170 hospital, suggesting that they were introduced from external sources. The three
171 patients were all citizens of Chengdu city but they were admitted to different local
172 hospitals before transferring to West China hospital. The remaining ST167 isolates
173 were recovered from swabs or samples collected at least 3 days after admission to
174 the ICU of West China hospital, from patients whose initial swabs were CRE
175 negative, indicating that they were acquired during their ICU stay. ST167 *E. coli*
176 carrying *bla*_{NDM-5} caused infections (bacteremia and abdominal infection) in only two
177 patients but colonised the others. Both ST617 *E. coli* carrying *bla*_{NDM-5} only colonised
178 patients. All patients colonised or infected with *E. coli* carrying *bla*_{NDM-5} of ST167 or
179 ST617 had received carbapenems before the recovery of the isolates.

180 **SNP analysis suggests continued dissemination of strains from a central** 181 **reservoir and sharing of resistance plasmid between lineages.**

182 To determine the level of relatedness between all isolated strains we mapped reads
183 of all the strains against a closed ST167 strain (strain 1237) generated by a
184 combination of Illumina and Minlon sequence data. The resulting high-resolution

185 SNP alignment showed the distance between the ST167 and ST617 strains to be
186 over 25,000 SNPs, confirming they are distinct lineages, with the two ST617 isolates
187 separated by just 7 SNPs. Deeper analysis of the ST167 cluster of strains showed
188 diversity ranging from 5 to 799 SNPs (Fig 1). Strains 936 and 1222 (both carriage
189 isolates) are the most closely related isolates with just 5 SNPs difference between
190 them, with both strains being acquired by patients in the ICU within one month of
191 each other. However these strains are 73 SNPs different from a strain isolated the
192 exact same month on the ICU from a strain (1237) that was acquired in the ICU. This
193 is almost double the genetic distance (46 SNPs) from a strain acquired (442 and 57,
194 isolated from the same patient) in the ICU two months earlier. These distances are
195 also larger than those for any isolate to the first two strains brought into the ICU,
196 strain 0215 and strain 243, which differ from all other isolates by around 30 SNPs,
197 and from each other by 15 SNPs. Such an observation suggests a potential
198 combination of patient-to-patient transmission in the affected ICU,²⁹ along with the
199 continued dissemination of the strain from a central reservoir where there is an
200 accumulation of diversity.^{29,30} Genomic analysis also allows us to identify a second
201 introgression of an ST167 strain (25) from the community, which is over 700 SNPs
202 different from the other isolates. Mapping of the raw sequence data against the 43kb
203 IncX3 plasmid containing *bla_{NDM-5}* also confirmed that the plasmid present in the
204 ST617 strains was identical to that in all of the ST167 strains with just two detectable
205 SNPs difference across the isolates.

206 **MDR ST167 and ST617 *E. coli* have been isolated across the world.**

207 We sought to contextualise the wider relevance of our Chengdu isolates by
208 investigating the wider prevalence of ST167 and ST617 strains. We searched the
209 Enterobase *E. coli* database and recovered a total of 87 genomes of ST167 (table

210 S2) and 86 genomes of ST617 (table S3), isolated from across the world. A core
211 CDS-based phylogeny of both lineages showed a diverse set of genomes with
212 around 17,000 SNPs in ST167 and around 15,000 SNPs in ST617. Annotation of the
213 ST617 phylogeny with β -lactamase gene carriage shows a high prevalence of the
214 *bla*_{CTX-M-15} ESBL gene in characterised isolates (Fig 2A). Annotation of the ST167
215 phylogeny with β -lactamase gene carriage (Fig 2B) shows a pattern of resistance
216 gene carriage, with multiple independent acquisitions of carbapenemase across the
217 phylogeny including *bla*_{NDM-1}, *bla*_{NDM-5}, *bla*_{NDM-7}, *bla*_{OXA-181}, and *bla*_{KPC-3}. For both
218 phylogenies there is clear evidence of isolation of strains from across the globe.

219 **Evolutionary genomic analysis correlates switches in LPS gene content with**
220 **the emergence of the ST167/ST617 lineage**

221 Both ST167 and ST617 are single locus variants of the ST10 lineage of *E. coli*. ST10
222 is the most abundant lineage of *E. coli* represented in the Enterobase database and
223 contains isolates ranging from drug susceptible environmental and human
224 commensal strains, to multi-drug resistant strains isolated from human clinical UTI
225 and bacteraemia infections. We selected 256 ST10 genomes from Enterobase
226 (Table S4) to represent the known spectrum of ST10 diversity present in the
227 database, and merged this data set with our publically available ST167/ST617
228 genome data set to create a larger ST10 complex phylogeny (Fig S1). The resulting
229 phylogeny shows that ST167 and ST617 are sister clades with respect to ST10, with
230 ST617 emerging as a nested clade from a single outlying ST167 genome, though
231 the distance between ST167 and ST617 is around 18,000 SNPs.

232 Given the phylogenetic pattern of ST167 and ST617 with respect to ST10, we sought
233 to determine if their emergence from ST10 is associated with defined evolutionary
234 events. We used a combined GWAS approach to compare the ST167/617 genomes

235 with ST10, using both SEER and SCOARY analysis of a pangenome matrix. Only
236 loci considered to be significantly associated with one lineage over the other by both
237 methods were further investigated (Dataset S1). Most striking was the absence of
238 the *wzzB* gene and *wca* biosynthetic cluster in ST167/ST617 whilst the majority of
239 the ST10 genomes contained both (Figure S2). These genes are involved in LPS
240 biosynthesis with *wzzB* being the master controller of O antigen chain length in the
241 *wzx/wzy* pathway, whilst *wca* genes are responsible for colonic acid biosynthesis.³¹
242 In silico *E. coli* serotyping³² established that ST167 and ST617 demonstrate the
243 exact same O antigenic type (O32novel) with similarity also seen in H antigen type
244 (H9 or H10) (Figure S2), whilst the SerotypeFinder database identified the strains as
245 O89.

246 Our combined GWAS analysis also identified another ~90 CDS which were present
247 across the entire data set, but which had distinct alleles in the ST167/ST617
248 genomes compared to those in ST10 (Fig 3, Dataset S2). Many of these CDS
249 encode dehydrogenase enzymes involved in anaerobic metabolism, or are part of
250 the *cob/pdu/eut* operons known to be involved in anaerobic respiration during
251 intestinal inflammation.³³ This would appear to suggest differential evolutionary
252 events in key genes involved in anaerobic metabolism in the formation of the
253 ST167/ST617 lineage. Also present were unique alleles in core CDS involved in acid
254 and bile salt tolerance, and a number of fimbrial-like proteins. In conjunction these
255 data would suggest differential evolutionary forces acting on loci involved in
256 mammalian colonisation in ST167/617 in comparison to ST10. Furthermore a
257 combined SEER and Piggy approach identified unique sequences in 17 intergenic
258 regions (IGRs) upstream of core CDS in ST167/617 that were distinct from ST10,

259 including IGRs upstream of anaerobic metabolic loci also present in the
260 SEER/SCOARY analysis (Dataset S1).

261 **Discussion**

262 Our data presented here provide a comprehensive genomic analysis of two lineages
263 of carbapenem resistant *E. coli* infecting multiple patients within the ICU of West
264 China hospital. Both these lineages, ST167 and ST617, are members of the larger
265 ST10 complex of *E. coli*, which is ubiquitously found in environmental, human
266 clinical, and mammalian intestinal commensal sampling. Our analysis is the first
267 genome level characterisation of strains belonging to ST167 or ST617, despite a
268 number of single site reports of clinical infections with both lineages existing in the
269 literature.

270 Our analysis shows that the diversity which accumulates in the genome of the ST167
271 isolates during the course of the investigation is not mirrored by diversity in the
272 plasmid carrying the *bla_{NDM-5}* gene. Only 1 SNP difference existed between the
273 sequence of this plasmid in the ST167 isolates, and only 2 SNPs difference between
274 the ST167 and ST617 isolates. As a result it is impossible to tell if the IncX3 plasmid
275 associated with dissemination of *bla_{NDM-5}* in China³⁴ was transferred between ST167
276 and ST617 in the hospital, or if the plasmid is highly stable with only deleterious
277 mutations occurring and quickly purged from the population. Clearly there is a need
278 for more thorough and detailed analysis of various resistance plasmids within and
279 between hospitals, such as was done recently for NDM-1 plasmids in Latin America.

280 ³⁵

281 The lack of appropriately designed isolate collection and sequencing strategy means
282 it is impossible to conduct any form of genomic epidemiological analyses of these *E.*
283 *coli* lineages beyond our Chinese investigation. However the ready availability of a

284 large number of good-quality, curated genome assemblies in the Enterobase
285 genome database do allow us to delve deeper into the evolutionary history of *E. coli*
286 ST167 and ST617. Whilst data generated and uploaded to Enterobase is prone to a
287 bias towards clinical MDR strains, it is still clear that ESBL and carbapenem resistant
288 strains of both these lineages have been isolated from across the world over the past
289 20 or so years (Tables S1 and S2).

290 Comparative genomic analysis and GWAS for traits specific to ST167 and ST617
291 compared to ST10 also support emergence along a shared evolutionary branch. Key
292 among these is the complete loss of the *wca* operon encoding colanic acid
293 biosynthesis in the LPS biosynthesis pathway. The majority of *E. coli* produce their
294 LPS utilising the O-unit translocation pathway encoded for by *wzx* and *wzy*.³¹ This
295 method utilises glycosyltransferases to assemble the O antigen in units at the
296 cytoplasmic membrane. These units are then translocated by Wzx and polymerized
297 by Wzy until the O antigen chain length is reached. This mechanism is utilised by the
298 majority of the ST10 isolates, however genomic analysis shows that ST167 and
299 ST617 utilise an alternative *wzm/wzt* ATP transporter pathway. This biosynthetic
300 pathway assembles the entire O-antigen on the cytoplasmic face before Wzt
301 transports the O-chain across,³¹ resulting in an O-antigen with truncated chain
302 length. O-antigen chain length plays a major role in pathogenicity of Gram negative
303 organisms, and it has been demonstrated that loss of long O-antigen chains in
304 *Salmonella* optimizes immune evasion and allows successful colonisation.³⁶

305 Alongside the LPS genetic changes, we also observed unique alleles of anaerobic
306 metabolism genes and genes potentially involved in host colonisation in ST167/617
307 compared to ST10. Recent modelling data has shown that any factor influencing the

308 ability of a bacterium to colonise a host will also influence its likelihood of evolving
309 antimicrobial resistance.³⁷

310 **Conclusions**

311 We provide data for the first ever, single hospital genomic analysis of clinical isolates
312 of carbapenem resistant *E. coli* belonging to the ST167/617 lineage. Our data
313 presented here provide evidence for evolutionary events that would affect microbial
314 interaction with a mammalian host underpinning the emergence of the ST167/617
315 lineage from ST10. There is also evidence for lineage specific alterations in
316 intergenic regions in ST167/617, a phenomenon which has already been described
317 as underpinning the emergence of MDR plasmid-containing *E. coli* ST131 strains.²⁸

318 Clearly there is now a need for a fully designed genomic epidemiological
319 investigation of lineages of *E. coli* associated with carriage of carbapenem resistance
320 plasmids arising from the ST10 clade, both in China and internationally. Such a
321 study will fully inform us of any potential parallelism in the evolution of MDR lineages
322 of *E. coli*, and of the true nature and scope of their prevalence and global
323 dissemination.

324 **Funding section**

325 This work was funded by a Royal Society Newton Advanced Fellowship project
326 (NA150363) and a grant from the National Natural Science Foundation of China
327 (project no. 8151101182) awarded to ZZ and AM. SF was funded by the Wellcome
328 Antimicrobial Resistance doctoral training project at UoB, and CC by the Wellcome
329 MIDAS doctoral training program at UoB.

330 **Transparency declaration**

331 None to declare

332

333 **References**

- 334 1. de Kraker MEA, Jarlier V, Monen JCM, et al. The changing epidemiology of
335 bacteraemias in Europe: trends from the European Antimicrobial Resistance
336 Surveillance System. *Clin. Microbiol. Infect.* 2013;**19**:860–8.
- 337 2. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and
338 international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae.
339 *Clin. Microbiol. Rev.* 2015;**28**:565–91.
- 340 3. Alhashash F, Weston V, Diggle M, McNally A. Multidrug-Resistant *Escherichia coli*
341 Bacteremia. *Emerg. Infect. Dis.* 2013;**19**:1699–701.
- 342 4. Croxall G, Hale J, Weston V, et al. Molecular epidemiology of extraintestinal
343 pathogenic *Escherichia coli* isolates from a regional cohort of elderly patients
344 highlights the prevalence of ST131 strains with increased antimicrobial resistance in
345 both community and hospital care settings. *J. Antimicrob. Chemother.*
346 2011;**66**:2501–8.
- 347 5. Livermore DM, Hawkey PM. CTX-M: changing the face of ESBLs in the UK. *J.*
348 *Antimicrob. Chemother.* 2005;**56**:451–4.
- 349 6. Feng Y, Yang P, Xie Y, et al. *Escherichia coli* of sequence type 3835 carrying
350 blaNDM-1, blaCTX-M-15, blaCMY-42 and blaSHV-12. *Sci. Rep.* 2015;**5**:12275.
- 351 7. Zhang L, Xue W, Meng D. First report of New Delhi metallo-beta-lactamase 5
352 (NDM-5)-producing *Escherichia coli* from blood cultures of three leukemia patients.
353 *Int. J. Infect. Dis.* 2016;**42**:45–6.
- 354 8. Cuzon G, Bonnin RA, Nordmann P. First identification of novel NDM
355 carbapenemase, NDM-7, in *Escherichia coli* in France. *PLoS One.* 2013;**8**:e61322.
- 356 9. Zheng B, Dong H, Xu H, et al. Coexistence of MCR-1 and NDM-1 in Clinical
357 *Escherichia coli* Isolates. *Clin. Infect. Dis.* 2016;**63**:1393–5.

- 358 10. Leflon-Guibout V, Blanco J, Amaqdouf K, et al. Absence of CTX-M Enzymes but
359 High Prevalence of Clones, Including Clone ST131, among Fecal *Escherichia coli*
360 Isolates from Healthy Subjects Living in the Area of Paris, France. *J. Clin. Microbiol.*
361 2008;**46**:3900–5.
- 362 11. Kohler C-D, Dobrindt U. What defines extraintestinal pathogenic *Escherichia*
363 *coli*? *Int. J. Med. Microbiol.* 2011;**301**:642–7.
- 364 12. Cerqueira GC, Earl AM, Ernst CM, et al. Multi-institute analysis of carbapenem
365 resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal
366 outbreaks. *Proc. Natl. Acad. Sci.* 2017;**114**:1135–40.
- 367 13. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly
368 algorithm and its applications to single-cell sequencing. *J. Comput. Biol.*
369 2012;**19**:455–77.
- 370 14. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.*
371 2014;**30**:2068–9.
- 372 15. Larsen M V, Cosentino S, Rasmussen S, et al. Multilocus sequence typing of
373 total-genome-sequenced bacteria. *J. Clin. Microbiol.* 2012;**50**:1355–61.
- 374 16. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid
375 core-genome alignment and visualization of thousands of intraspecific microbial
376 genomes. *Genome Biol.* 2014;**15**:524.
- 377 17. Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence
378 data. *Bioinformatics.* 2014;**30**:3399–401.
- 379 18. Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read
380 assembly via adaptive k-mer weighting and repeat separation. *Genome Res.*
381 2017;**27**:722–36.
- 382 19. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive

383 microbial variant detection and genome assembly improvement. *PLoS One*.
384 2014;**9**:e112963.

385 20. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan
386 genome analysis. *Bioinformatics*. 2015;**31**:3691–3.

387 21. Stamatakis A Ludwig T MH. RAxML-III: a fast program for maximum likelihood-
388 based inference of large phylogenetic trees. *Bioinformatics*. 2005;**21**:456.

389 22. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of
390 phylogenetic trees made easy. *Nucleic Acids Res*. 2011;**39**:W475-8.

391 23. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in
392 microbial pan-genome-wide association studies with Scoary. *Genome Biol*.
393 2016;**17**:238.

394 24. Lees JA, Vehkala M, Valimaki N, et al. Sequence element enrichment analysis to
395 determine the genetic basis of bacterial phenotypes. *Nat. Commun*. 2016;**7**:12797.

396 25. Inouye M, Conway TC, Zobel J, Holt KE. Short read sequence typing (SRST):
397 multi-locus sequence types from short reads. *BMC Genomics*. 2012;**13**:338.

398 26. Joensen KG, Tetzschner AMM, Iguchi A, et al. Rapid and Easy In Silico
399 Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data.
400 *J. Clin. Microbiol*. 2015;**53**:2410–26.

401 27. Thorpe HA, Bayliss SC, Sheppard SK, Feil EJ. Piggy: A Rapid, Large-Scale Pan-
402 Genome Analysis Tool for Intergenic Regions in Bacteria. *Gigascience*. 2018;**7**:1-11.

403 28. McNally A, Oren Y, Kelly D, et al. Combined Analysis of Variation in Core,
404 Accessory and Regulatory Genome Regions Provides a Super-Resolution View into
405 the Evolution of Bacterial Populations. *PLoS Genet*. 2016;**12**:e1006280.

406 29. Köser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for
407 investigation of a neonatal MRSA outbreak. *N. Engl. J. Med*. 2012;**366**:2267–75.

408 30. Quick J, Cumley N, Wearn CM, et al. Seeking the source of *Pseudomonas*
409 *aeruginosa* infections in a recently opened hospital: an observational study using
410 whole-genome sequencing. *BMJ Open*. 2014;**4**:e006278.

411 31. Iguchi A, Iyoda S, Kikuchi T, et al. A complete view of the genetic diversity of the
412 *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res*. 2015;**22**:101–7.

413 32. Ingle DJ, Valcanis M, Kuzevski A, et al. In silico serotyping of *E. coli* from short
414 read data identifies limited novel O-loci but extensive diversity of O:H serotype
415 combinations within and between pathogenic lineages. *Microb. genomics*.
416 2016;**2**:e000064.

417 33. McNally A, Thomson NR, Reuter S, Wren BW. “Add, stir and reduce”: *Yersinia*
418 *spp.* as model bacteria for pathogen evolution. *Nat. Rev. Microbiol*. 2016;**14**:177–90.

419 34. Yang P, Xie Y, Feng P, Zong Z. blaNDM-5 carried by an IncX3 plasmid in
420 *Escherichia coli* sequence type 167. *Antimicrob. Agents Chemother*. 2014;**58**:7548–
421 52.

422 35. Marquez-Ortiz RA, Haggerty L, Olarte N, et al. Genomic Epidemiology of NDM-1-
423 Encoding Plasmids in Latin American Clinical Isolates Reveals Insights into the
424 Evolution of Multidrug Resistance. *Genome Biol. Evol*. 2017;**9**:1725–41.

425 36. Crawford RW, Wangdi T, Spees AM, et al. Loss of very-long O-antigen chains
426 optimizes capsule-mediated immune evasion by *Salmonella enterica* serovar Typhi.
427 *MBio*. 2013;**4**: e00232-13.

428 37. Lehtinen S, Blanquart F, Croucher NJ, et al. Evolution of antibiotic resistance is
429 linked to any genetic mechanism affecting bacterial duration of carriage. *Proc. Natl.*
430 *Acad. Sci*. 2017;**114**:1075–80.

431
432

433 Figure 1: Maximum likelihood phylogenetic tree of *E. coli* ST167 strains isolated from
434 the ICU of West China hospital. The phylogeny is inferred from a SNP alignment
435 obtained by mapping raw data against a MinIon/Illumina hybrid complete assembly
436 of isolate 1237. The annotation denotes the presence of ESBL and CPE associated
437 β -lactamases as determined by Abricate.

438

439 Figure 2: Maximum likelihood phylogenetic trees of a global collection of (A) ST617
440 and (B) ST167 strains. The phylogeny is inferred from an alignment of concatenated
441 core CDS sequences as determined by Roary, and is mid-point rooted. The
442 annotation denotes the presence of ESBL and CPE associated β -lactamases as
443 determined by Abricate.

444

445 Figure 3: Manhattan skyline plot showing position of kmers identified by GWAS
446 analysis as being significantly associated with ST167/617 compared to ST10. The x
447 axis indicates the position on the WCHEC1237 complete genome assembly, whilst
448 the Y axis indicates the numbers of statistically significant kmers mapping at that
449 position. Hits indicated in red are either intergenic regions (labelled IGR) identified as
450 being unique by both Piggy and SEER analysis, or anaerobic metabolism loci
451 identified as significantly different by both SEER and Scoary.

452

453

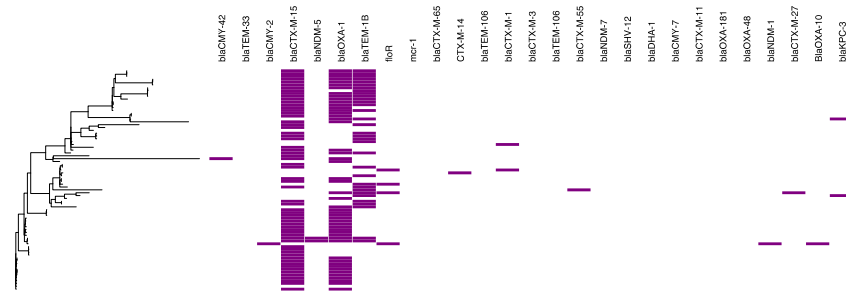


SNP Distance Matrix	1222	215	243	25	442	57	936	1237
1222	0	30	39	725	45	43	5	74
215	30	0	15	705	25	23	31	100
243	39	15	0	714	34	32	40	113
25	725	705	714	0	720	718	726	799
442	45	25	34	720	0	6	46	113
57	43	23	32	718	6	0	44	117
936	5	31	40	726	46	44	0	73
1237	74	100	113	799	113	117	73	0

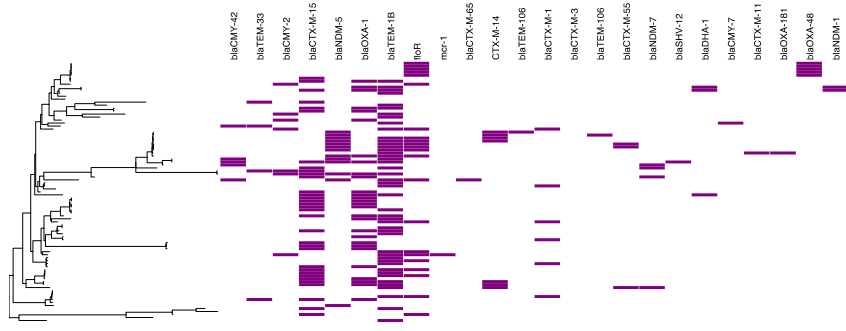
454
455

Figure 1

456



A

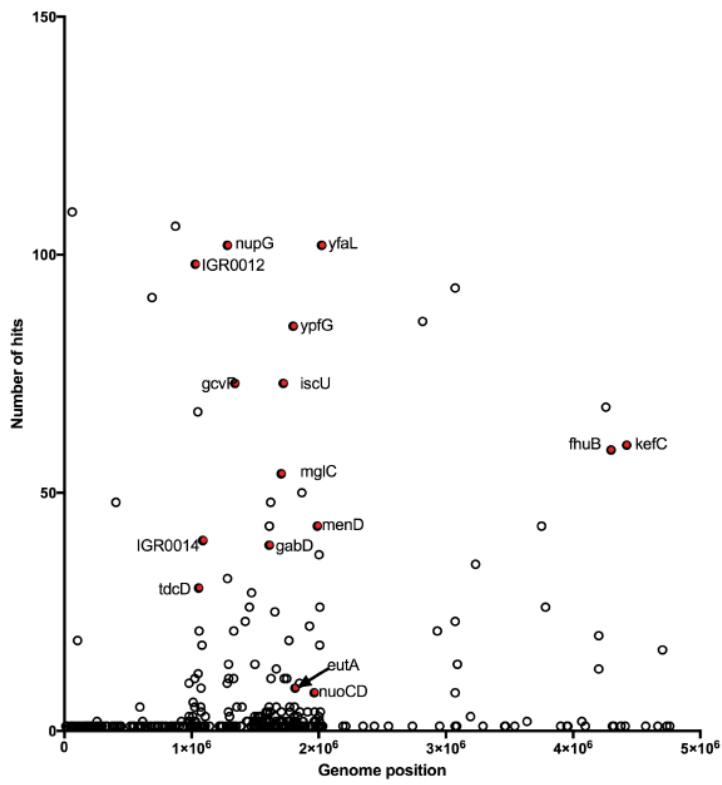


B

457

458 Figure 2

459



460

461 Figure 3

462