Genomic characterisation and context of the *bla*$_{NDM-1}$ carbapenemase in *Escherichia coli* ST101.

Melinda M. Ashcroft[1,2,3], Brian M. Forde[1,2,3], Minh-Duy Phan[1,2], Kate M. Peters[1,2], Andrew Henderson[2,4,5], Steven J. Hancock[1,2], Leah W. Roberts[1,2,3], Rhys T. White[1,2,3], Kok-Gan Chan[6,7], Teik Min Chong[6], Wai-Fong Yin[6], David L. Paterson[2,4], Timothy R. Walsh[8], Mark A. Schembri[1,2*], Scott A. Beatson[1,2,3*]

[1]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia.

[2]Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, QLD, Australia.

[3]Australian Centre for Ecogenomics, The University of Queensland, Brisbane, QLD, Australia.

[4]UQ Centre for Clinical Research, The University of Queensland, Brisbane, QLD, Australia.

[5]Infection Management Services, Princess Alexandra Hospital, Brisbane, QLD, Australia

[6]Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia.

[7]International Genome Centre, Jiangsu University, Zhenjiang, China.

[8]Department of Medical Microbiology and Infectious Disease, Cardiff University, Cardiff, United Kingdom.

*Contributed equally

***Corresponding authors:**

Scott A. Beatson, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia 4072, QLD, Australia; Telephone: +61-7-33654863; Email: s.beatson@uq.edu.au

Mark A. Schembri, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia 4072, QLD, Australia; Telephone: +61 7 336 53306; Email: m.schembri@uq.edu.au

Short title: *Escherichia coli* ST101 and the carriage of bla_{NDM-1}.

Keywords:

antimicrobial resistance; long-read sequencing; plasmid-mediated resistance; carbapenem resistance; NDM-1

Abstract

Carbapenems are last-resort antibiotics; however, the spread of plasmid-encoded carbapenemases such as the New Delhi metallo-β-lactamase 1 (NDM-1) challenges their effectiveness. The rise of NDM-1 has coincided with the emergence of extensively multidrug resistant (MDR) lineages such as *Escherichia coli* ST101. Here we present a comprehensive genomic analysis of seven *E. coli* ST101 isolates that carry the $bla_{\text{NDM-1}}$ gene. We determined the complete genomes of two isolates and the draft genomes of five isolates, enabling complete resolution of the plasmid context of $bla_{\text{NDM-1}}$. Comparisons with thirteen previously published ST101 genomes revealed a monophyletic lineage within the B1 phylogroup forming two clades (designated Clade 1 and Clade 2). Most Clade 1 strains are MDR, encoding resistance to at least 9 different antimicrobial classes, including extended spectrum cephalosporins. Additionally, we characterised different pathways for $bla_{\text{NDM-1}}$ carriage and persistence in the ST101 lineage. For IncC plasmids, carriage was associated with recombination and local transposition events within the antibiotic resistance island. In contrast, we revealed recent transfer of a large $bla_{\text{NDM-1}}$ resistance island between F-type plasmids. The complex acquisition pathways characterised here highlight the benefits of long-read Single Molecule Real Time sequencing in revealing evolutionary events that would not be apparent by short-read sequencing alone. These high-quality *E. coli* ST101 genomes will provide an important reference for further analysis of the role of mobile genetic elements in this emerging multidrug resistant lineage.

71 **Importance**

72 Carbapenem resistant *Escherichia coli* are urgent priority organisms as they are resistant to our
73 drugs of last resort. *E. coli* ST101 have been reported as carriers of the New Delhi Metallo-beta-
74 Lactamase 1 gene ($bla_{\text{NDM-1}}$), conferring resistance to carbapenems, however there is limited
75 genomic information available for this lineage. In this study we used long-read genome
76 sequencing to characterise the complete genomes of two *E. coli* ST101 strains and determine the
77 carriage of $bla_{\text{NDM-1}}$ in a collection of *E. coli* ST101 strains. We showed that carriage of $bla_{\text{NDM-1}}$ and
78 resistance determinants to eight other antimicrobial classes was confined to a single clade. We
79 also showed two different pathways for the carriage of $bla_{\text{NDM-1}}$, which was dependent on the type
80 of plasmid. Long-read sequencing allowed us to show the full complexities of these resistance
81 regions and highlighted how strains from an emerging *E. coli* lineage have become resistant to
82 nearly all available antimicrobials.

83

84 **Introduction**

85 The successful treatment of *E. coli* infections is complicated by the rising prevalence of antibiotic
86 resistance among clinical and community isolates (1). Carbapenems are considered a class of last-
87 resort antibiotics, however their effectiveness has been challenged by the emergence of bacteria
88 capable of hydrolysing carbapenems and most β-lactams. Carbapenem resistance has
89 disseminated worldwide, as genes that encode carbapenemases are easily transferred via MGEs
90 such as plasmids, transposons and integrons. In 2008, $bla_{\text{NDM-1}}$ was first reported in a *Klebsiella*
91 *pneumoniae* strain isolated from a Swedish patient who had recently travelled to India (2). Since
92 then, the $bla_{\text{NDM-1}}$ gene and its variants has been identified in Gram-negative bacteria in more than
93 fifty countries including the UK, India, Pakistan and Bangladesh, across Europe, China, the US,
94 Canada and Australia (3). NDM-positive clonal lineages can be found in multiple different
95 organisms including *K. pneumoniae, Acinetobacter baumannii* and *E. coli* (4). For example, in *K.*

96 *pneumoniae*, several of the first bla_{NDM-1} reports between 2009-2011 involved the multi-drug
97 resistant (MDR) sequence type (ST)14 clone (5). In *A. baumannii*, international clonal lineages I
98 (Clonal Complex (CC)109/CC1) and II (CC92/CC2), ST25 and ST85 are the four dominant NDM-
99 positive clonal lineages (6). In *E. coli*, bla_{NDM-1} has been detected worldwide, with isolates
100 frequently typed as ST405, ST131 and ST101 (4) and also found to carry the $bla_{CTX-M-15}$ ESBL gene
101 (7). The first characterisation of an FII plasmid carrying the bla_{NDM-1} gene was in the *E. coli* ST131
102 strain GUE (pGUE-NDM, Genbank accession: JQ364967) (8).
103
104 Other *E. coli* MDR clones such as ST131 have been well-characterised genetically (9-11), but very
105 little is known about *E. coli* ST101 at the genomic level, despite numerous reported cases of
106 carbapenemase-producing *E. coli* ST101 (4, 12-16). Mushtaq *et al*, 2011 showed that ST101 strains
107 have acquired NDM-positive plasmids of different incompatibility groups and demonstrated
108 diverse PFGE banding patterns for *E. coli* ST101 strains, with some evidence of clonal expansion
109 due to clustering of pulsotypes (12). Additionally, serotyping of five ST101 isolates indicated that
110 they were O-non-typeable:H21/42 and contained an array of virulence genes; including *fimH*, *pap*,
111 *sfa/focDE* (fimbrial genes), *iucD, iroN* (siderophore genes), *iss*, *traT* (protection genes), and *tsh*
112 (serine protease gene) (16). Other studies have reported varying levels of virulence potential in
113 ST101 isolates; however in all cases, these strains contained numerous adhesins, autotransporters
114 and siderophores normally associated with extra-intestinal pathogenic *E. coli* (4, 16, 17).
115
116 There are now several *E. coli* ST101 genomes and NDM-positive plasmid sequences available in
117 public databases, but as yet there are no published analyses of any *E. coli* ST101 complete
118 genomes. Draft assemblies provide limited information in terms of genomic context of mobile
119 elements such as insertion sequences (IS), phages, genomic islands and plasmids. The complete
120 assembly of these mobile elements is crucial for characterising the genomic context of resistance

121 elements (18, 19). Here we present a comprehensive genomic analysis of *bla*$_{NDM-1}$ carriage in *E.*
122 *coli* ST101. Using a combination of Pacific Biosciences (PacBio) Single Molecule Real Time (SMRT)
123 sequencing and Illumina sequencing, we defined the plasmid context of *bla*$_{NDM-1}$ in seven ST101
124 isolates. We also report the complete genome of two representative ST101 strains: MS6192 and
125 MS6193, including manual curation of all MGEs. Using publicly available and published ST101 draft
126 genomes, we also examined the phylogeny of ST101 and defined two major clades (1 and 2), with
127 the majority of Clade 1 strains encoding resistance to carbapenems and extended spectrum
128 cephalosporins (ESCs).
129
130 **Results**
131 **NDM-positive *E. coli* ST101 genomes.** During our molecular characterisation of 16
132 carbapenemase-producing *E. coli* collected from India and the United Kingdom in 2008-2009 (20,
133 21), we found seven that were phylogroup B1, sequence type (ST)101 and NDM-positive by PCR
134 (Supplementary Appendix and Supplementary Dataset, Table S1). To investigate the genomic
135 context of *bla*$_{NDM-1}$ and characterise ST101 at the genomic level we undertook whole genome
136 sequencing (WGS). The genomes of seven NDM-positive ST101 strains were assembled from a
137 combination of PacBio RSI or RSII long-read data and Illumina HiSeq paired-end data. In all cases
138 we could assemble a single plasmid encoding *bla*$_{NDM-1}$ together with several other antimicrobial
139 resistance (AMR) genes (see below). There were sufficient long reads to enable the assembly of
140 complete, finished quality, genome sequences for two of the strains: MS6192 and MS6193
141 (Supplementary Dataset, Table S2).
142
143 *E. coli* MS6192 is comprised of a single circular chromosome of 4,879,059 bp (Table 1). Four
144 circularised contigs of 142,890 bp, 76,661 bp, 86,336 bp and 3,608 bp represent the large *bla*$_{NDM-1}$
145 encoding MDR plasmid pMS6192A-NDM (4 replicons: FII and FII(pCoo) replicons and an FIA and

146 FIB replicon, IncF replicon sequence type (IncF RST) F36/F22:A1:B20), pMS6192B (FII, F2:A-:B-),
147 pMS6192C (IncI1, PMLST ST173) and the small cryptic plasmid pMS6192D (ColRNAI), respectively.
148 *E. coli* MS6193 is comprised of a single circular chromosome of 4,922,872 bp. Additionally, there
149 were three circularised contigs; 142,890 bp, 76,661 bp and 4,367 bp, which represent the large
150 bla_{NDM-1} encoding MDR plasmid pMS6193A-NDM (4 replicons: FII and FII(pCoo), FIA and FIB,
151 F36/F22:A1:B20), pMS6193B (FII, F2:A-:B-) and pMS6193C (untypeable).
152
153 The chromosomes of MS6192 and MS6193 were almost identical, differing by 12 single nucleotide
154 polymorphisms (SNPs). The sequence of the F-type (bla_{NDM-1}) and FII plasmids were also near-
155 identical between strains. Further investigation of the automated assembly was required to
156 unambiguously determine the sequences of pMS6192A-NDM and pMS6192C, primarily due to
157 shared regions of high nucleotide identity (Supplementary Appendix). The genomes could also be
158 distinguished by the presence of an additional prophage (Phi8) and Transposon 7 (Tn*7*)-like
159 Transposon in *E. coli* MS6193 (Supplementary Appendix, Fig S1), an additional large plasmid in
160 MS6192 (pMS6192C) and two different cryptic plasmids (pMS6192D and pMS6193C).
161
162 **Table 1. Summary of genomic information for *E. coli* ST101 strains MS6192 and MS6193.**

	MS6192 chr[1]	**MS6193 chr[1]**	**pMS6192A-NDM/ pMS6193A-NDM (FII/FII:F1A:F1B)**	**pMS6192B/ pMS6193B (FII)**	**pMS6192C[2] (IncI1)**	**pMS6193C[2] (colRNAI)**	**pMS6192D (untypeable)**
Size (bp)	4,879,059	4,922,872	142,890/ 142,890	76,661/ 76,661	86,336	4,367	3,608
No. of genes	4,834	4,876	179/173	91/84	95	6	2
No. of CDS	4,570	4,612	169/167	82/82	91	5	1
No. of tRNAs	94	94	-	-	-	-	-
No. of rRNAs	22 (7:16S, 7:23S, 8:5S)	22 (7:16S, 7:23S, 8:5S)	-	-	-	-	-
No. of IS	46	44	21/21	5/5	0	0	0
G+C	50.69	50.69	51.5/51.50	51.91/51.91	50.14	54.39	43.15

content (%)							

163 [1]chr = chromosome. [2]MS6192 has 4 plasmids (pMS6192A, pMS6192B, pMS6192C and pMS6192D); MS6193 has 3
164 plasmids (pMS6193A, pMS6193B and pMS6193C); pMS6192A-NDM and pMS6192B are near identical to pMS6193A-
165 NDM and pMS6193B, respectively; there is no genetic relationship between pMS6192C, pMS6193C or pMS6193D.
166

167 MS6194, MS6201, MS6203, MS6204 and MS6207 strains were assembled as draft genomes with
168 gapped chromosomes and circularised plasmid sequences. The five draft *E. coli* ST101 genomes
169 range in size from 5,150,883 to 5,434,929 bp with the number of contigs ranging from 16 to 55
170 (Supplementary Dataset, Table S2). Each draft genome contained a single $bla_{\text{NDM-1}}$ encoding MDR
171 plasmid and several plasmid replicons ranging from one (MS6203) to seven (MS6204).
172

173 **Phylogenomic analysis of *E. coli* ST101 reveals a monophyletic lineage within the B1 phylogroup.**
174 To explore the context of ST101 within the *E. coli* phylogeny we carried out a phylogenetic
175 reconstruction of 65 representative, publicly available, complete *E. coli* genomes together with 20
176 ST101 genomes. These included the seven ST101 genomes sequenced in the present study and 13
177 *E. coli* ST101 draft genomes that were previously published and available in GenBank or the
178 Sequence Read Archive (SRA) on October 1st, 2016 (Supplementary Dataset, Table S3). The
179 additional ST101 genomes were obtained from several clinical sources and included isolates from
180 urine (n = 5), blood (n = 4), wound (n=1) and intestinal microbiota (n = 2). A single environmental
181 isolate from untreated wastewater was also included. All 20 *E. coli* ST101 strains cluster together
182 in a single lineage within the B1 phylogroup (Figure 1A). Based on the representative complete
183 genomes included in this phylogeny, the *E. coli* ST101 lineage is most closely related to *E. coli*
184 SE11, a fecal strain of serotype O152:H28, isolated from a healthy adult (22).
185

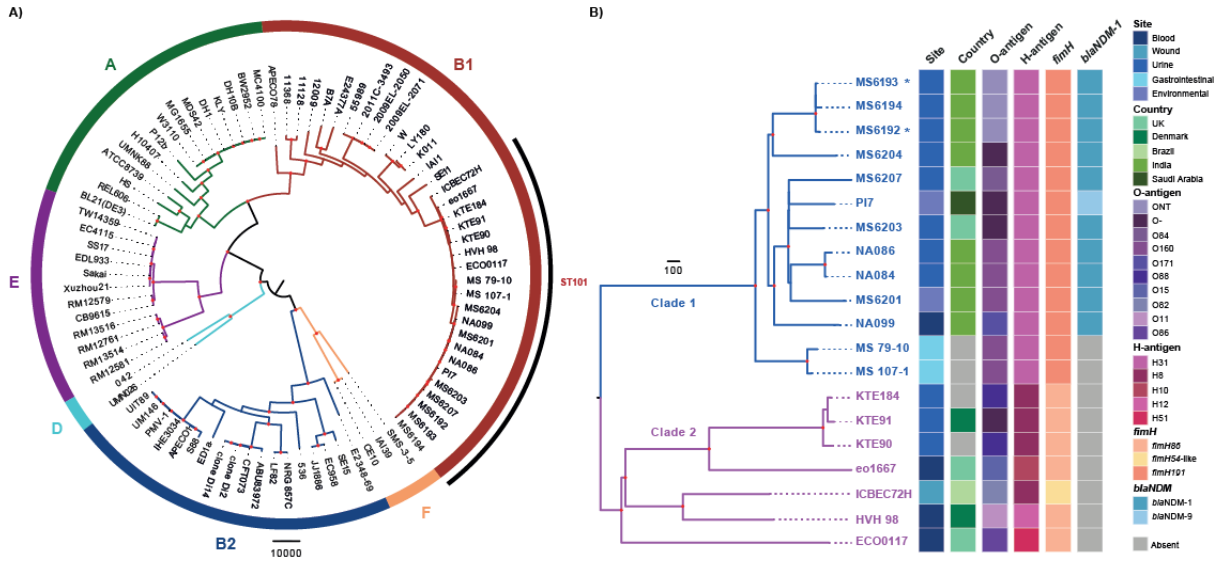
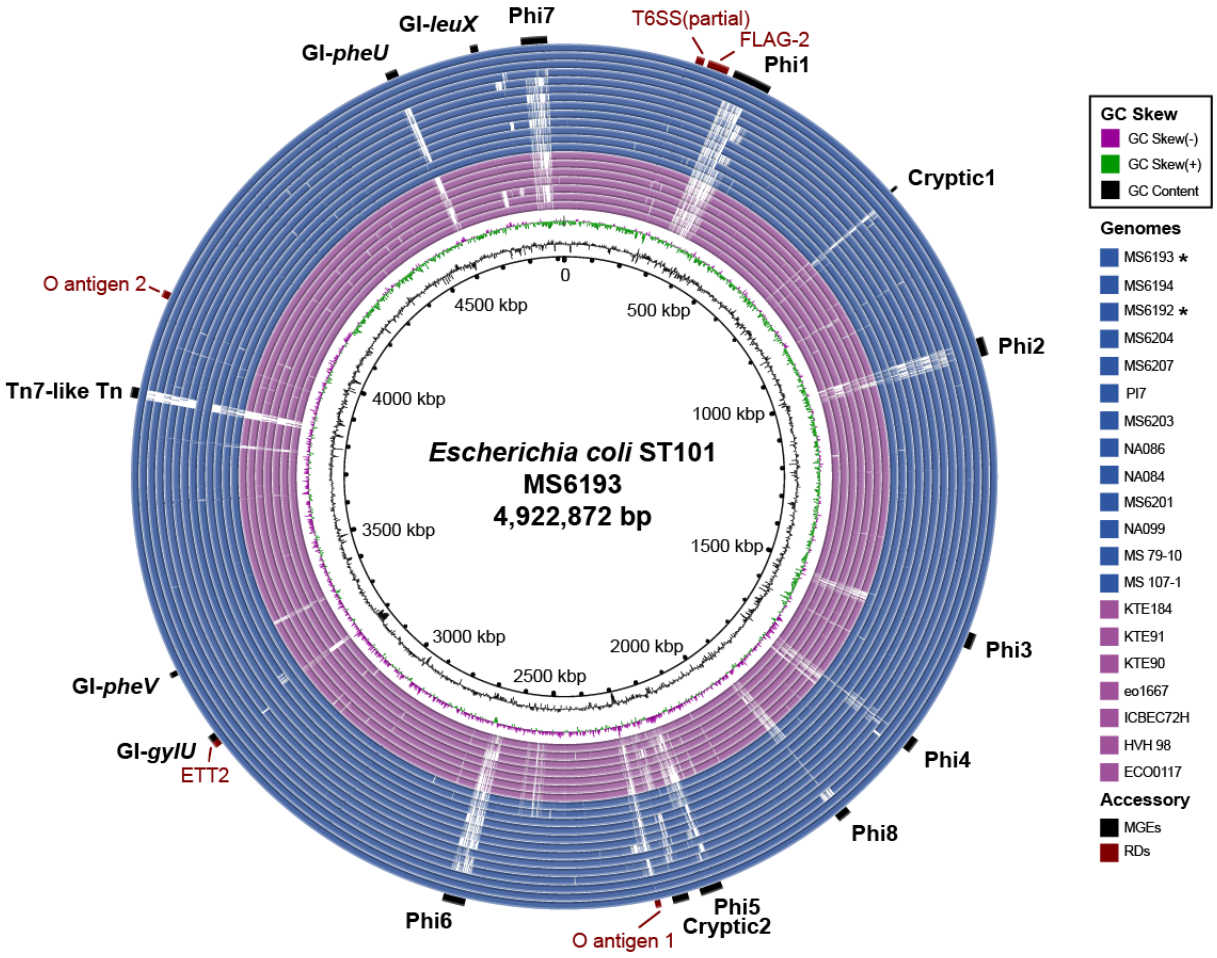


Figure 1. Phylogenetic relationship of *E. coli* ST101. A) The phylogram was built from 182,264 core-genome SNPs using Maximum Likelihood. The tree was rooted using the out-group species *E. fergusonii* ATCC35469 (not shown). The taxa labels for complete *E. coli* genomes are coloured as follows; phylogroup F: orange, phylogroup B2: dark blue, phylogroup D: aqua, phylogroup E: purple, phylogroup A: green, phylogroup B1: red (ST101 labelled). B) The mid-point rooted, recombination-filtered phylogram was built from 2,106 core-genome SNPs. Two clades can be defined. Clade 1 (blue): MS6193*, MS6194, MS6192*, MS6204, MS6207, PI7, MS6203, NA086, NA084, MS6201, NA099, MS 79-10 and MS 107-1; Clade 2 (pink): KTE184, KTE91, KTE90, eo1667, ICBEC72H, HVH 98 and ECO0117. NT: non-typeable, - indicates absence of locus. Bootstrapping support indicated by red nodes: 95-100% bootstrap support from 1000 replicates. *Denotes complete genomes.

The ST101 phylogeny formed two distinct clades, with Clade 1 showing more recent clonal expansion compared with the deeper branching Clade 2 (Fig 1B). Clade 1 is mostly comprised of urine isolates and includes all NDM-positive ST101, including the seven ST101 isolates sequenced in this study. In fact, only the two gut-derived isolates (MS 107-1 and MS 79-10) lack the $bla_{\text{NDM-1}}$ gene (or its single amino acid variant $bla_{\text{NDM-9}}$ (E152K), present in PI7). The phylogenetic position of

204 MS 107-1 and MS 79-10, basal to the remaining NDM-positive ST101 strains, suggests that *bla*$_{NDM-1}$
205 was acquired once in Clade 1 and has been inherited vertically.
206
207 To further characterise the ST101 lineage we examined key virulence and serotyping gene loci.
208 Clade 1 isolates have the *fimH*191 type 1 fimbrial adhesion allele, whereas Clade 2 isolates possess
209 the *fimH*86 allele except for ICBEC72H, which encodes a novel *fimH*54-like variant. Several
210 serogroups were predicted across both clades: O160 (MS107-1, MS79-10, MS6201, NA084 and
211 NA086), O171 (NA099), O84 (MS6207) and a novel, non-typeable O antigen region (MS6192,
212 MS6193 and MS6194), but Clade 1 is distinguished by a *fliC* gene encoding the H31 antigen.
213 Additionally, three Clade 1 strains lack part of the O antigen gene locus: MS6203, MS6204 and PI7
214 (*wzx*, *wzy*, *wzm* and *wzt*).
215
216 **Mobile genetic elements are variably conserved in *E. coli* ST101.** The major differences between
217 these ST101 genomes were MGEs, such as those defined in *E. coli* MS6192 and MS6193 (Figure 2
218 and Supplementary dataset, Table S4). MGEs in MS6192 and MS6193 include genomic islands (GI)
219 within known integration tRNA hotspots (GI-*leuX*, GI-*pheU*, GI-*pheV*, GI-*glyU*), prophage elements
220 and a Transposon-*7* (Tn*7*)-like transposon (Supplementary Dataset, Table S5). While no MGE that
221 was defined in MS6192 or MS6193 is completely conserved across the ST101 lineage, GI-*leuX* is
222 present in most strains from Clades 1 and 2, and Prophage (Phi) 3 is conserved in all Clade 1
223 strains. MS6192, MS6193 (and MS6194) differ in only two chromosomal MGEs (Phi8 and Tn7-like
224 Tn) consistent with their near identical core chromosomes. In contrast to the GI-*pheV* islands of
225 phylogroup B2 ExPEC/UPEC strains, such as the 75,054 bp locus in the *E. coli* ST131 strain EC958
226 (23, 24), the GI within the tRNA-*pheV* locus in MS6192 and MS6193 is relatively small (9,191 bp)
227 and does not contain known virulence modules. Other genomic regions of difference (RD) that are
228 known to be variable across *E. coli* are present in MS6192 and MS6193 (Supplementary Dataset,

229 Table S6) and are also conserved across the ST101 lineage. These include a remnant of the Type VI
230 secretion system (T6SS) (25), the Flag-2 lateral flagellar locus (26) and a degenerate Type III
231 secretion locus 2 (ETT2), most commonly found in intestinal pathogenic *E. coli* from the B1
232 phylogroup and rarely found in ExPEC (27). *E. coli* ST101 strains MS6192 and MS6193 also share 44
233 chromosomal insertion sequences (IS) (Supplementary Dataset, Table S7).
234

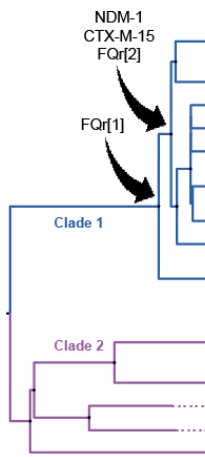


235

236 **Figure 2. Genomic map of ST101 *E. coli* showing mobile genetic elements and other notable**
237 **genomic regions.** Visualisation of the *E. coli* MS6193 genome compared to 19 *E. coli* ST101
238 complete and draft genomes. The innermost circles represent GC skew (green/purple) and GC
239 content (black) of *E. coli* MS6193. The degree of coloured shading indicates the nucleotide identity
240 between MS6193 and each *E. coli* ST101 genome. Nucleotide comparisons are coloured based on

241 identity of between 70% and 100% (dark shading = high, light shading = low). *E. coli* genomes are
242 arranged according to their previously defined phylogenetic relationship as follows from the
243 innermost ring: Clade 2 (pink): ECO0117, HVH 98, ICBEC72H, eo1667, KTE90, KTE91, KTE184; Clade
244 1 (blue): MS 107-1, MS 79-10, NA099, MS6201, NA084, NA086, MS6203, PI7, MS6207, MS6204,
245 MS6192*, MS6194, MS6193*. *Denotes complete genomes. Outer ring indicates mobile genetic
246 elements (MGEs) and select ST101 regions of difference (RDs) within the *E. coli* MS6193 genome.
247 Image prepared using BRIG.
248
249 **Most *E. coli* ST101 Clade 1 strains contain multiple antibiotic resistance genes.** ST101 strains
250 from Clade 1 possessed a high number of antibiotic resistance genes, including genes that encode
251 resistance to beta-lactams (including carbapenems), aminoglycosides, streptomycins, phenicols,
252 sulphonamides, macrolides, trimethoprims and tetracyclines (Figure 3). All NDM-positive ST101
253 Clade 1 isolates also contain at least one copy of the $bla_{\text{CTX-M-15}}$ gene that confers resistance to
254 cephalosporins. In other *E. coli*, $bla_{\text{CTX-M-15}}$ is often found on conjugative F-type plasmids as part of
255 a Tn*3*-like IS*Ecp1*-$bla_{\text{CTX-M-15}}$-orf477 mobile element (11). We could determine the genomic location
256 of $bla_{\text{CTX-M-15}}$ in all seven of the PacBio sequenced ST101 strains, showing that the $bla_{\text{CTX-M-15}}$
257 insertion site was chromosomally located in six of the seven strains (Supplementary Dataset, Table
258 S8). It appears that $bla_{\text{CTX-M-15}}$ has been acquired once and mobilised to several different genomic
259 locations by an IS*Ecp1* in Clade 1 and inherited vertically in some cases (Supplementary Appendix).
260
261 Fluoroquinolone resistance in *E. coli* is frequently associated with chromosomal point mutations in
262 the quinolone resistance-determining region (QRDR) of *gyrA* (codons 67-106) and *parC* (codons
263 56-108) (28). All Clade 1 strains had a non-synonymous mutation in the QRDR of *parC* (S80I)
264 (Figure 3). Additionally, two non-synonymous mutations in the QRDR of *gyrA* were detected in all
265 NDM-positive Clade 1 isolates, characterised by the amino acid substitutions S83L and D87N. The

266 two Clade 1 intestinal isolates MS107-1 and MS79-10 have a different single amino acid
267 substitution in *gyrA* (D87Y) but lack both bla_{NDM-1} and $bla_{CTX-M-15}$. These results suggest recent
268 development of extensive MDR in a sub-lineage of Clade 1 (hereafter referred to as Clade 1A),
269 similar to the emergence of other fluoroquinolone MDR resistant *E. coli* clonal lineages such as
270 ST131 and ST1193 (9, 29). Phenotypic antimicrobial resistance testing of the seven PacBio
271 sequenced ST101 strains in this study confirmed that phenotype could be predicted from
272 genotype in most cases (Supplementary Appendix and Supplementary Dataset, Table S9). Notably
273 Clade 2 strains did not encode QRDR mutations, with only HVH 98, ICBEC72H and eo1667
274 possessing any acquired resistance genes.



	Plasmid Replicon Types (Inc)									*In silico* typing: Fluoroquinolone resistance determinants		Colistin resistance	Beta-lactamases (*bla*)						Aminoglycosides resistance							Other resistance determinants: Streptomycins	Phenicols	Sulfonamides	Macrolides		Trimethoprims	Tetracycline
	C	FII	FIA	FIB	FIC (FII)	Colicin	Others	pMLST	Plasmid carrying bla_{NDM-1}	gyrA	parC	*mcr-1*	*CTX-M* variants	*OXA* variants	*TEM1* variants	*DHA-1*	*CMY* variants	*NDM* variants	*aac(6')Ib-cr*^	*aacA* variants	*aac(3)* variants	*aph(3')* variants	*rmtB*	*rmtC*	*aad* variants	*armA*	*strA and strB*	*cat* variants	*sul* variants	*mph* variants	*msr(E)*	*dfrA* variants	*tet* variants
MS6193 *	-	+	+	+	-	-	IncI1	F36/F22:A1:B20, F2:A-:B-, IncI1[ST-173]	FII:1A:1B	S83L & D87N	S80I	-	15	1	B	-	-	1	+	-	IIa	-	+	-	A2	-	-	B3	1	-	-	12	-
MS6194	-	+	+	+	-	-	-	F36/F22:A1:B20, F2:A-:B-	FII:1A:1B	S83L & D87N	S80I	-	15	1	B	-	-	1	+	-	IIa	-	+	-	A2	-	-	B3	1	-	-	12	-
MS6192 *	-	+	+	+	-	+	IncI1	F36/F22:A1:B20, F2:A-:B-, IncI1[ST-173]	FII:1A:1B	S83L & D87N	S80I	-	15	1	B	-	-	1	+	-	IIa	-	+	-	A2	-	-	B3	1	-	-	12	-
MS6204	-	+	+	+	-	+	IncI1 IncI2	F36/F22:A1:B20, F2:A-:B-, IncI[NT]	FII	S83L & D87N	S80I	-	15	1	B	-	-	1	+	-	IIa	-	+	-	A2	-	-	B3	1	(A)	-	12	-
MS6207	+	-	-	-	-	+	p0111	IncC[ST1]#	IncC	S83L & D87N	S80I	-	15	2	A	-	6	1	+	-	-	Ia	-	+	A2	-	+	A1	1	-	-	12 & 29	-
PI7	+	+	-	+	-	+	IncY	IncC[ST1]#, FY4:A-:B36	FIIY:1B	S83L & D87N	S80I	-	15	2	A	-	-	9	+	4	IIa	-	-	+	A2	+	+	A1	1	(E)	+	12 & 29	D
MS6203	+	-	-	-	-	-	-	IncC[ST1]#	IncC	S83L & D87N	S80I	-	15	2	A	+	6	1	+	-	IIa & IId	VIa	-	+	-	+	-	A1	1	-	-	29	-
NA086	-	+	+	+	-	+	-	F1:A2:B33	FII:1A:1B	S83L & D87N	S80I	-	15	1 & 2	-	-	-	1	+	4	IIa	VIa	-	+	-	-	-	A1 & B4	1	-	-	29	A
NA084	-	+	+	+	-	+	-	F1:A2:B33	FII:1A:1B	S83L & D87N	S80I	-	15	1 & 2	-	-	-	1	+	4	IIa	VIa	-	+	-	-	-	A1 & B4	1	-	-	29	A
MS6201	+	+	-	-	-	-	IncR	IncC[ST1]#, F2:A-:B-	IncC	S83L & D87N	S80I	-	15	2	A & B	-	4	1	-	-	IIa	-	+	-	A2	+	+	A1	1	(E)	+	12 & 29	-
NA099	+	+	+	+	-	+	IncI1	IncC[ST1]#, F22:A1:B20, IncI[NT]	FII:1A:1B	S83L & D87N	S80I	-	15	1 & 2	-	-	6	1	+	4	IIa & IId	-	-	+	-	-	+	A1 & B4	1	-	-	29	A
MS 79-10	-	-	-	-	-	-	IncI1	IncI1[ST-26]	-	D87Y	S80I	-	-	-	B	-	-	-	-	-	VIa	-	-	-	A1	-	-	A1	1	(A)	-	14	-
MS 107-1	-	-	-	-	-	-	IncI1	IncI1[ST-26]	-	D87Y	S80I	-	-	-	B	-	-	-	-	-	VIa	-	-	-	A1	-	-	A1	1	(A)	-	14	-
KTE184	-	-	-	+	+	-	-	F18:A-:B6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KTE91	-	-	-	+	+	-	-	F18:A-:B6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
KTE90	-	-	-	+	+	-	-	F18:A-:B6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
eo1667	-	-	-	+	+	+	IncI1 IncQ1	F18:A-:B8, IncI1[ST-26]	-	S83L	-	-	-	-	B	-	-	-	-	1	VIa	-	-	-	-	-	+	-	1 & 2	-	-	1	A
ICBEC72H	-	-	-	-	-	-	IncI1 IncX4	IncI1[ST-131]	-	-	-	+	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HVH 98	-	-	-	-	-	-	IncP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A1	-	-	-	1	-	-	1	-
ECO0117	-	+	-	-	-	+	IncX4	F2:A-:B-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

*Complete genomes. #Based on IncA/C plasmid typing scheme. NT: Non-typeable IncI sequence type. ^Responsible for both fluoroquinolone and aminoglycoside resistance

Figure 3. Plasmid and antimicrobial resistance typing in *E. coli* ST101. Plasmid replicon types, *in silico* typing and antimicrobial resistance determinants indicate similar patterns to the core-genome phylogeny (left) of the 20 *E. coli* ST101 strains. Black arrows indicate acquisition of fluoroquinolone resistance ([1] indicates single mutation in *gyrA*, [2] indicates double mutation), bla_{NDM-1} and $bla_{CTX-M-15}$ genes. Types of resistance genes are colour coded: colistin resistance (blue-purple) beta-lactamases (pink-purple), carbapenemases (pink), aminoglycosides (aqua), streptomycins (orange), phenicols (green), sulphonamides (blue), macrolides (red), trimethoprims (yellow) and tetracylins (gold). Grey shading indicates gene presence, white shading indicates absence, allele variants are indicated in each cell.

282 **The *bla*$_{NDM-1}$ genetic environment is highly conserved in *E. coli* ST101.** All seven PacBio
283 sequenced strains contained a plasmid-encoded *bla*$_{NDM-1}$ gene (Figure 4). In almost all cases the
284 genetic structure of the *bla*$_{NDM-1}$ module was identical and matched other previously described
285 *bla*$_{NDM-1}$ modules, consisting of a 258-287 bp fragment of IS*Aba125* containing the -35 promoter
286 region, the *bla*$_{NDM-1}$ gene, the bleomycin resistance gene (*ble*$_{MBL}$) and a truncated
287 phosphoribosylanthrantile isomerase (*iso*/*trpC*) (30). The surrounding plasmid context differed
288 between the seven strains, but like *bla*$_{CTX-M-15}$, the structural variations (inversions, transpositions,
289 and insertions/deletions) are congruent with a single *bla*$_{NDM-1}$ acquisition prior to divergence of
290 Clade 1A. For example, only pMS6203A-NDM has retained a complete IS*Aba125* element
291 upstream of *bla*$_{NDM-1}$ suggestive of horizontal transfer. Additionally, downstream of the *bla*$_{NDM-1}$
292 module, *iso*/*tat* has been truncated by the insertion of the *bla*$_{DHA-1}$ gene. In pMS6201A-NDM,
293 truncation of IS*Aba125* was mediated by the insertion of an IS*3000.* The two IS*3000* flanking the
294 *bla*$_{NDM-1}$/*ble*$_{MBL}$ region form the composite transposon Tn*3000* (31). Truncation of IS*Aba125* in
295 pMS6192A-NDM, pMS6193A-NDM, pMS6194A-NDM, pMS6204A-NDM and pMS6207A-NDM
296 appears to have resulted from the insertion of an IS*Ecp1* module, which was subsequently deleted
297 by an IS*26* composite transposon leaving only a 103 bp fragment of IS*Ecp1*. However, pMS6207A
298 differs from the other four strains with carriage of *aphA6* (amikacin resistance) instead of *aac(3')-*
299 *lla* and *tmrB* (aminoglycoside and tunicamycin resistance) on the IS*26* composite transposon.
300

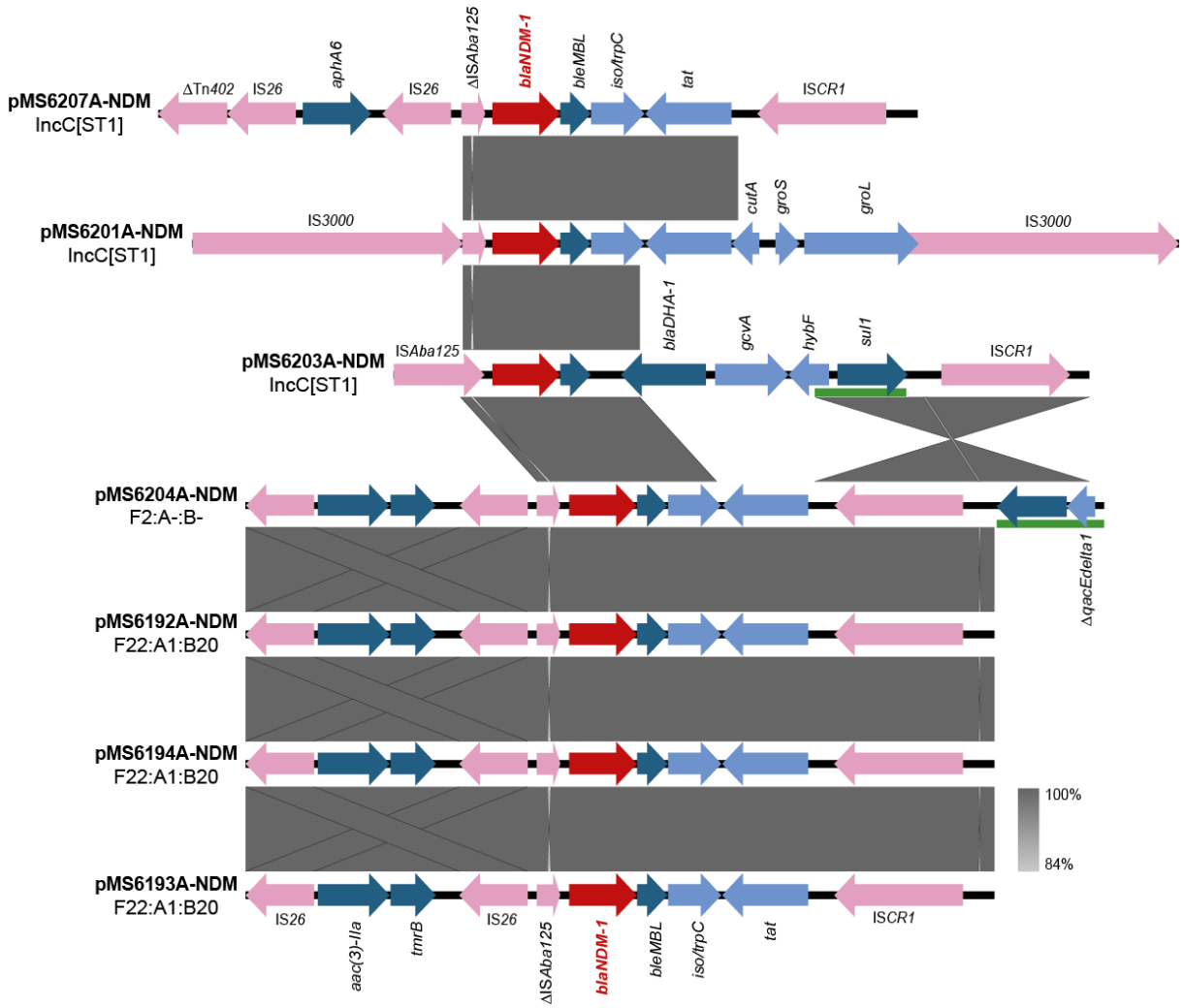


Figure 4. Comparison of the *bla*$_{NDM-1}$ genetic environment of PacBio sequenced Clade 1 ST101 isolates. Grey shading indicates nucleotide identity between sequences according to BLASTn (84-100%). pMS6204A has been reverse complemented for easier visualisation. Δ, truncated gene. IS/Tns: light pink, AMR genes: teal, CDSs: light blue. Green rectangles represent 3' conserved sequences of class 1 integrons. The *bla*$_{NDM-1}$ gene is labelled in red. Image created using EasyFig.

IncC plasmids are a vehicle for *bla*$_{NDM-1}$ carriage in *E. coli* ST101. IncC plasmids are commonly reported carriers of the *bla*$_{NDM-1}$ gene in *Enterobacteriaceae* (32-34). Three PacBio sequenced ST101 strains from this study contained IncC replicons (pMS6201A-NDM, pMS6203A-NDM and pMS6207A-NDM) and possess a similar backbone compared to other IncC plasmids such as pNDM-US-2 (Supplementary Appendix, Fig S2). However, deletions in the IncC plasmid backbone alter the

conjugation efficiency of these plasmids such that only pMS6203A is conjugative (Supplementary Appendix). *In silico* plasmid multi-locus sequence typing (pMLST) determined that the IncC plasmids reported here belong to the NDM-associated ST1 group (34), with pMS6203A and pMS6207A belonging to the cgST 1.5. Plasmid pMS6201A is missing several genes, however, is most like cgST 1.1 (Supplementary Appendix). These IncC plasmids differ greatly in their structure of the antibiotic resistance island (ARI-A) (Supplementary Appendix and Figure 5). Therefore, despite high conservation in the immediate genetic context of $bla_{\text{NDM-1}}$ and the IncC plasmid backbone in these ST101 strains, the complete plasmid sequences reveal substantial diversity within the ARI-A, consistent with mobilisation of AMR genes, including $bla_{\text{NDM-1}}$.

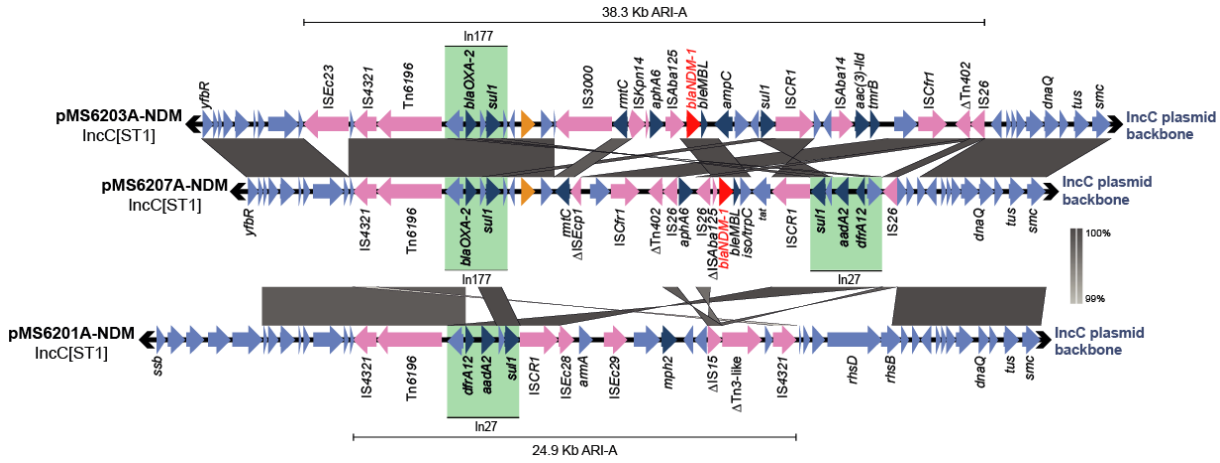
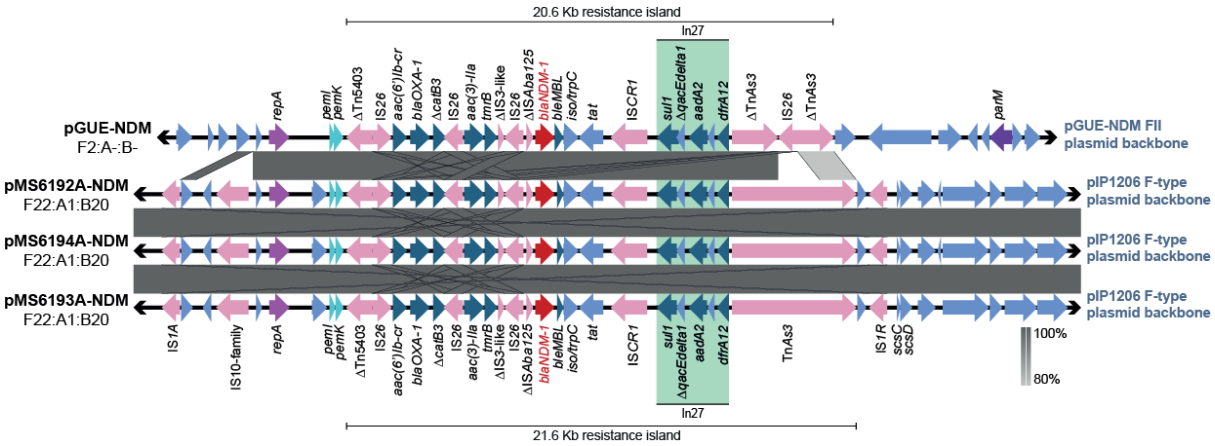


Figure 5. Major structural features of the Antibiotic Resistance Island on IncC plasmids in the PacBio sequenced Clade 1 ST101 strains. Nucleotide comparisons between the variable resistance regions of the IncC plasmid pMS6203A-NDM, pMS6207A-NDM and pMS6201A-NDM. Grey shading indicates nucleotide identity between sequences according to BLASTn (98-100%). Key genomic regions are indicated: IS elements/transposons: light pink, replication genes: dark pink, maintenance genes: purple, *pemIK* operon: aqua, MTases: orange, AMR genes: teal, other CDSs: blue, Integrons: green rectangles. The $bla_{\text{NDM-1}}$ gene is labelled in red. Image created using EasyFig.

332 ***E. coli* ST101 strains possess distinct types of *bla*$_{NDM-1}$ containing F-type plasmids.** F-type plasmids
333 have previously been associated with the carriage of *bla*$_{NDM-1}$ in *E. coli* (8, 35-37). The four *bla*$_{NDM-1}$
334 encoding F-type plasmids sequenced in this study share a near identical 21.6 kb MDR region that
335 includes the *bla*$_{NDM-1}$ gene. However, two different plasmid backbones were identified; plasmids
336 pMS6192A-NDM, pMS6193A-NDM, pMS6194A-NDM were almost identical, whereas the
337 backbone of pMS6204A-NDM differed substantially (Supplementary Appendix, Fig S3). BLAST
338 comparisons of the *bla*$_{NDM-1}$ containing plasmids pMS6192A-NDM, pMS6193A-NDM and
339 pMS6194A-NDM (all F36/F22:A1:B20), indicate that they are most similar to the F-type plasmid
340 pIP1206 (>99% nucleotide identity, 83% query coverage). Major differences include deletions in
341 the plasmid backbone, where pMS6192A-NDM, pMS6193A-NDM and pMS6194A-NDM contain an
342 incomplete conjugation region, missing *traK* and *traB*, as well as an 18,516 bp section between
343 *traU* and *traI* present in pIP1206. Resistance genes in pMS6192A-NDM, pMS6193A-NDM and
344 pMS6194A-NDM are clustered in a 21.6 Kb resistance-island, which is different to the resistance
345 island of pIP1206 and is highly similar (99.99% nucleotide identity, 91% query coverage) to the
346 resistance island in the FII plasmid pGUE-NDM (Supplementary Appendix, Fig S4).
347
348 The *bla*$_{NDM-1}$ resistance island has inserted between the *pemIK* toxin/antitoxin operon and the
349 *scsCD* operon (copper-sensitivity suppressor). It is composed of a Tn*5403* transposon (containing
350 an intact transposase and the inverted left repeat (ILR) (38)) truncated by an IS*26* (Figure 6). This is
351 followed by 3 resistance genes; *aac(6')Ib-cr* (fluoroquinolone and aminoglycoside resistance),
352 *bla*$_{OXA-1}$ (beta-lactam resistance) and a truncated *catB3* (chloramphenicol resistance). Two
353 additional resistance genes *aac(3)-IIa* (aminoglycoside resistance) and *tmrB* (tunicamycin
354 resistance) are downstream and flanked on either side by IS*26* elements. The *bla*$_{NDM-1}$ module
355 consists of an IS*Aba125* fragment containing the *bla*$_{NDM-1}$ -35 promoter region, *bla*$_{NDM-1}$
356 (carbapenem resistance), *ble*$_{MBL}$ (bleomycin resistance), *iso/trpC* (phosphoribosylanthranilate

357 isomerase), *tat* (twin-arginine translocation pathway signal protein) and an IS*CR1* element.
358 Downstream of the $bla_{\text{NDM-1}}$ module is the Class I integron In27. This contains *sul1* (sulphonamide
359 resistance) and a truncated *qacEdelta1*, followed by *aadA2* (aminoglycoside resistance) and
360 *dfrA12* (trimethoprim resistance). This MDR region ends with a truncated Tn*As3* transposon
361 (containing intact transposase and recombinase genes and the inverted right repeat (IRR) (38,
362 39)).
363



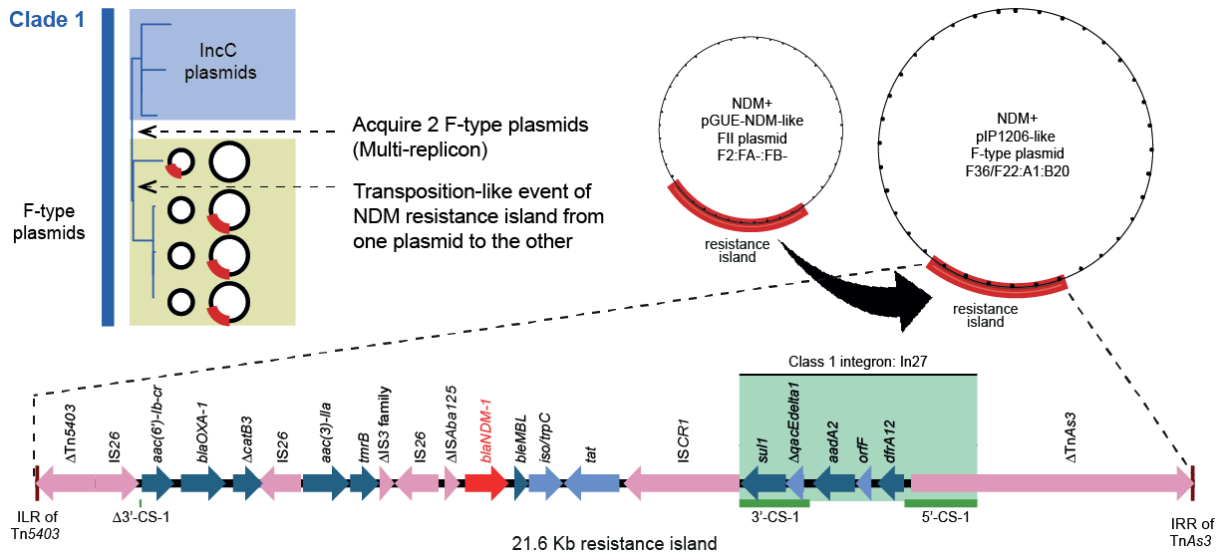
364

365 **Figure 6. Major structural features of the conserved MDR island on F-type plasmids in the PacBio**
366 **sequenced Clade 1 ST101 strains.** Nucleotide comparisons between the reference plasmid pGUE-
367 NDM (Genbank accession: JQ364967), pMS6192A-NDM, pMS6194A-NDM and pMS6193A-NDM
368 highlighting the shared resistance island and differing plasmid backbones. pGUE-NDM has been
369 reverse complemented for easier visualisation. Grey shading indicates nucleotide identity between
370 sequences according to BLASTn (80-100%). Key genomic regions are indicated: IS
371 elements/transposons: light pink, replication genes: dark pink, maintenance genes: purple, *pemIK*
372 operon: aqua, MTases: orange, AMR genes: teal, other CDSs: blue, Integrons: green rectangles.
373 The $bla_{\text{NDM-1}}$ gene is labelled in red. Image created using EasyFig.
374

375 ***bla*$_{NDM-1}$ is encoded on an FII, pGUE-NDM-like plasmid in one ST101 strain.** MS6204 also contains
376 an F-type pIP1206-like plasmid (pMS6204B; F36/F22:A1:B20). However, the resistance island in
377 pMS6204B is different to that of pIP1206 and to that of the other *bla*$_{NDM-1}$ containing plasmids
378 described above. This resistance island does not carry the *bla*$_{NDM-1}$ gene, however does encode
379 *bla*$_{CTX-M-15}$ and additional copies of the *aac(6')-Ib-cr*, *bla*$_{OXA-1}$ and *tmrB* resistance genes. In fact, for
380 MS6204, the MDR region encoding *bla*$_{NDM-1}$ is harboured by pMS6204A-NDM (F2:A-:B-), an FII
381 plasmid that is most similar to *E. coli* FII plasmid pGUE-NDM (>99% nucleotide identity, 95% query
382 coverage). Additionally, present within the MDR region of the pMS6204A-NDM plasmid is a 5,095
383 bp resistance region, which is not present in pGUE-NDM. This region is flanked on either side by
384 IS*26* and encompasses a Tn*3* transposon encoding *bla*$_{TEM-1B}$ (beta-lactam resistance) followed by
385 *rmtB* (rifampicin resistance). This association between Tn*3*, *bla*$_{TEM-1}$ and *rmtB* flanked by IS*26* has
386 been observed in several other *E. coli* and *K. pneumoniae* plasmids (Supplementary Dataset, Table
387 S10), suggesting widespread distribution of this resistance module.
388
389 **Complete plasmid sequences from closely related ST101 strains show carriage of *bla*$_{NDM-1}$ on**
390 **different F-type plasmids.** Intriguingly MS6192, MS6193 and MS6194 also contain a pGUE-NDM-
391 like plasmid, with almost identical backbones to pMS6204A-NDM (F2:A-:B-; named pMS6192B and
392 pMS6193B in our complete genomes), however, these plasmids lack the *bla*$_{NDM-1}$ encoding MDR
393 island (Supplementary Appendix, Fig S4). We therefore hypothesise that the entire island encoding
394 *bla*$_{NDM-1}$ was transferred from an FII pGUE-NDM-like plasmid (F2:A-:B-) to an F-type pIP1206-like
395 plasmid (F36/F22:A1:B20) after divergence of MS6192, MS6193 and MS6194 from MS6204
396 (Figure 7). According to our model, the most recent common ancestor (MRCA) of MS6204,
397 MS6192, MS6193 and MS6194 contained both an FII pGUE-NDM-like plasmid carrying the *bla*$_{NDM-1}$
398 resistance island and an F-type pIP1206-like plasmid carrying a different resistance island. In the
399 MRCA of MS6192, MS6193 and MS6194, the *bla*$_{NDM-1}$ resistance island appears to have been

400 transferred from the pGUE-NDM-like FII plasmid to the pIP1206-like F-type plasmid via a
401 transposition-like event with subsequent loss of *bla*$_{NDM-1}$ and most of the resistance island from
402 the donor plasmid. Although we cannot formally rule out acquisition of the 21.6 Kb *bla*$_{NDM-1}$
403 encoding resistance island from a different intermediary plasmid, we suggest a direct transfer
404 between co-existing ancestral plasmids is the most parsimonious explanation for the observed
405 plasmid sequences in this study. Whilst the ILR of Tn*5403* and IRR of Tn*As3* are present in
406 pMS6192A-NDM, pMS6193A-NDM and pMS6194A-NDM, along with intact Tn*As3* transposase and
407 recombinase genes, it remains unknown whether the MDR island in its current form is capable of
408 active transposition.

409



410

411 **Figure 7. Proposed model of transfer of NDM+ resistance island between F-type plasmids.** We
412 predict that acquisition of a pGUE-NDM-like plasmid and a pIP1206-like plasmid occurred in the
413 MRCA of MS6204, MS6192, MS6194 and MS6193. A transposition-like event of the entire
414 resistance island between both FII plasmids resulted in the transfer of *bla*$_{NDM-1}$ and surrounding
415 resistance genes. Key regions of the resistance island are indicated: Δ: truncated gene, CDS: light
416 blue, IS/Tns: light pink, AMR genes: teal, Integron: green rectangle. 5' and 3- conserved sequences

417 of class 1 integrons: dark green. The *bla*$_{NDM-1}$ gene is labelled in red. The ILR of Tn*5403* and IRR of
418 Tn*As3* are indicated at the edges of the island.
419
420 We note that the relatively small number of ST101 genomes used here may constrain our ability to
421 fully characterise the clade structure and thus verify our plasmid model. However, both clades
422 were supported in a larger analysis of 263 available ST101 draft genomes obtained from
423 Enterobase (30/07/18) (40) that were of suitable quality for phylogenetic analysis (Supplementary
424 Appendix, Fig S5). Additionally, our finding that *bla*$_{NDM-1}$ (or its variants) was confined to a single
425 sub-lineage within Clade 1A remains intact. The conservation of the pGUE-NDM-like (FII) and
426 pIP1206-like (F-type) plasmid backbones in this larger ST101 dataset were also congruent with a
427 single acquisition into a subclade of Clade 1A (with loss of the FII plasmid in some strains)
428 (Supplementary Appendix, Fig S6 and Fig S7). Complete conservation of the *bla*$_{NDM-1}$ resistance
429 island was restricted to only a few strains, with the clade structure comparable to that of our
430 smaller dataset, further supporting our model of island transfer between plasmids (Supplementary
431 Appendix, Fig S8).
432
433 **Discussion**
434 Here we used PacBio SMRT sequencing to provide the most comprehensive genomic snapshot of
435 *bla*$_{NDM-1}$ carriage in the *E. coli* ST101 lineage to date. We show that this lineage is monophyletic
436 within the B1 phylogroup with at least two distinct clades that we have labelled Clade 1 and Clade
437 2. We show that *bla*$_{NDM-1}$ carriage is confined to Clade 1 and associated with fluoroquinolone
438 resistance mutations and other resistance genes such *bla*$_{CTX-M-15}$. By determining the complete
439 sequence of two closely related ST101 strains and the complete plasmid sequences for five other
440 ST101 strains, we have characterised the genomic context of *bla*$_{NDM-1}$ and other resistance genes.
441 Notably, we revealed a mosaic region of AMR genes including *bla*$_{NDM-1}$ within the ARI-A region of

442 IncC plasmids and transfer of a large MDR resistance region encoding *bla*$_{NDM-1}$ between two
443 different F-type plasmids. These results highlight the power of long-read sequencing in revealing
444 the full complexities of mobile genetic elements.
445
446 Identifying the genomic characteristics of *E. coli* lineages is crucial to explaining the evolution and
447 global dissemination of successful clones. Carbapenem resistance resulting from the acquisition of
448 plasmids carrying *bla*$_{NDM-1}$ has been characterised in several *E. coli* clonal lineages, such as ST405,
449 ST131 and ST101 (12). However, despite numerous reports of *bla*$_{NDM-1}$ or variants within the *E. coli*
450 ST101 lineage (4, 12-14, 17, 36), there remains a paucity of whole genome studies. The use of
451 PacBio SMRT sequencing long-read technology in this study enabled the *de novo* assembly of two
452 complete ST101 genomes (MS6192 and MS6193) and five high-quality draft genomes (MS6194,
453 MS6201, MS6203, MS6204 and MS6207). The inability to completely resolve the genomes of the
454 other five isolates was due to low data output achieved using a single SMRT cell on the PacBio RSI
455 platform. Despite this impediment we could completely resolve all plasmid sequences carrying
456 *bla*$_{NDM-1}$ and define the context of AMR elements within these seven isolates.
457
458 Our comparisons of the very closely related ST101 strains MS6192, MS6193 and MS6194 revealed
459 a pattern of genome conservation consistent with a close evolutionary relationship. For example,
460 a conserved core and accessory genome with only 12 core SNP differences between the
461 chromosomes of MS6192 and MS6193, the conservation of both FII plasmid backbones (pGUE-
462 NDM-like and pIP1206-like), as well as the stepwise acquisition of Phi 8 (present in MS6194 and
463 MS6193, but not MS6192) and the Tn*7*-like transposon (present only in MS6193) are consistent
464 with epidemiologically-linked strains. This level of chromosomal similarity is consistent with long-
465 term gut colonisation by a clonal population of Enterobacteriaceae (41). In fact, we also recently
466 reported the existence of an indigenous clonal lineage (with significant plasmid diversity) amongst

467 *E. coli* ST131 isolates collected from urine and fecal samples over a five-year period from a patient
468 with recurrent urinary tract infection (42). Analysis of the core genome of the 20 ST101 isolates
469 also indicates a close relationship, however considerable diversity is observed in their MGE
470 complement. These differences are particularly evident between clades, where MGEs such as
471 prophages and genomic islands defined in our reference ST101 genomes MS6192 and MS6193, in
472 general, are not conserved. Exceptions include the GI-*leuX* locus, which is intact in most Clade 1
473 and 2 strains. Nonetheless, our study is limited by the availability of temporal and geographic
474 metadata for all sequenced strains, which restricted our ability to interpret the ST101 phylogeny.
475
476 MDR is primarily disseminated within ST101 Clade 1A by plasmids of the groups IncI1, IncC and
477 F-type. The genetic elements harbouring $bla_{\text{NDM-1}}$ can spread among Gram-negative bacteria such
478 as *Acinetobacter*, *Pseudomonas* and *Enterobacteriaceae* including *K. pneumoniae* and *E. coli* (43),
479 where IncC and FII plasmids have previously been reported as carriers of $bla_{\text{NDM-1}}$ (8, 44).
480 Antimicrobial susceptibility testing showed that the AMR phenotype closely matched the AMR
481 genotype for the seven PacBio sequenced ST101 genomes in this study. However, MS6207 had an
482 intermediate resistance phenotype to meropenem and imipenem (Supplementary Dataset 1,
483 Table S9), despite the presence of the $bla_{\text{NDM-1}}$ gene, reflecting the fact that AMR prediction from
484 WGS data in Enterobacteriaceae remains difficult (45, 46).
485
486 In this study, we characterised two pathways for $bla_{\text{NDM-1}}$ persistence in ST101 Clade 1A. For
487 strains with IncC plasmids, carriage of $bla_{\text{NDM-1}}$ was associated with recombination of class 1
488 integrons and transposition of composite transposons and IS elements, resulting in differences in
489 AMR gene repertoire. In contrast, we show carriage of the $bla_{\text{NDM-1}}$ resistance island in an FII
490 plasmid, with predicted transfer to a co-existing F-type plasmid. Transposition events mediated by
491 both IS and transposons enable the mobility of entire resistance islands, particularly in the

492 *Enterobacteriaceae* (47). For example, a recent study showed that transposition mediated by IS*26*,
493 Tn*5403* and Tn*3*-family transposons such as those found flanking this *bla*$_{\text{NDM-1}}$ resistance island
494 facilitated plasmid rearrangements in carbapenem resistant *K. pneumoniae* (39). Although we
495 cannot be certain of the precise events leading to the evolution of pMS6192A-NDM, pMS6193A-
496 NDM and pMS6194A-NDM, our proposed model suggests that pGUE-NDM-like plasmids may have
497 played significant role in early dissemination of *bla*$_{\text{NDM-1}}$ in *E. coli*. However, we note that a lack of
498 ST101 complete genomes in both major clades has constrained our ability to comprehensively
499 confirm our proposed model of *bla*$_{\text{NDM-1}}$ resistance island transfer.
500
501 Fluoroquinolone resistance in Clade 1 strains is attributed to vertically transmitted point
502 mutations within the QRDR of *gyrA* (S83L, D87N and D87Y alleles) and *parC* (S80I allele). This is
503 similar to the ST131 lineage where acquisition of QRDR alleles imparting fluoroquinolone
504 resistance occurred in the MRCA prior to the divergence of Clade C (10). With the presence of
505 chromosomally acquired fluoroquinolone resistance and the acquisition of plasmid-encoded
506 resistance genes to more than nine different antimicrobial classes, these extensively MDR Clade 1
507 strains leave few choices for treatment. Older drugs such as colistin are now being used as last-
508 resort treatments for carbapenem-resistant infections (48). In 2016 however, an ST101 strain
509 (ICBEC72H, included in this study) containing the *mcr-1* colistin resistance gene on an IncX plasmid
510 backbone was identified (49). With recent reports of both carbapenem and colistin resistant *E. coli*
511 strains appearing (50, 51), we face the very real possibility of pan-drug resistant *E. coli*. Overall,
512 these high-quality *E. coli* ST101 genomes will provide an important reference for further analysis
513 of how mobile genetic elements and antimicrobial resistance influence the evolution of this
514 emerging, MDR *E. coli* lineage.
515

516 **Methods**

517 **Bacterial genome sequencing and assembly of our seven PacBio sequenced strains.** The NDM-
518 positive *E. coli* strains used in this study have been described previously and were screened for the
519 presence of $bla_{\text{NDM-1}}$ by PCR (20, 21). Additional details regarding strain selection and DNA
520 isolation methods can be found in the Supplementary Appendix. Six of seven strains were
521 sequenced on the PacBio RSI instrument, as previously described (24), using a 10 Kb insert library.
522 *E. coli* ST101 strain MS6192 used six SMRT cells, with MS6194, MS6201, MS6203, MS6204,
523 MS6207 using two SMRT cells each. *E. coli* ST101 strain MS6193 was sequenced on the PacBio RSII
524 instrument using three SMRT cells with P4-C2 chemistry. Raw sequencing reads were *de novo*
525 assembled using the Hierarchical Genome Assembly Process (HGAP) (52) as implemented in the
526 PacBio SMRT Analysis software suite v2.3.0, with a seed read cut-off length of 6,000 bp and
527 default parameters. For *E. coli* MS6192 and MS6193, contig order and orientation was determined
528 using Contiguity and nucleotide blast of the overlapping sequences at each contig end. The
529 merged assemblies were then checked against the closely related commensal strain *E. coli* SE11
530 using Mauve v2.3.1 (53). Additionally, two small, spurious contigs were generated during the *de*
531 *novo* assembly of MS6192, consisting of rRNA and tRNA genes that mapped to multiple locations
532 with 100% nucleotide identity to the MS6192 chromosome. These were deemed chimeric contigs
533 and were not included in the final assembly. Contiguity (54) was used to visualise the assembly,
534 with overlapping contigs manually trimmed and circularised. Misassemblies were corrected by
535 aligning the reads using BLASR (55), prior to sequence polishing. Raw sequence reads were then
536 mapped to these consensus contigs using Quiver, implemented in the SMRT Analysis software
537 suite, filtering errors from the assembly. Furthermore, read pileups across all repetitive (rRNA)
538 regions were manually inspected to ensure that their position was supported by spanning reads.

539

540 Draft PacBio sequenced ST101 assemblies were screened for plasmid sequences using all-versus-
541 all nucleotide comparisons to generate inter and intra-contig pairwise alignments. All contigs were
542 then screened for overlapping sequences at the 5' and 3' ends as this is an artefact of the HGAP
543 assembly process (52) and contigs with self-similar ends likely represent fully complete, circular
544 plasmids. One end of the contig was trimmed to produce a circular sequence. Additional, putative
545 plasmid contigs were screened by nucleotide BLAST comparison against the NCBI non-redundant
546 database. As an additional quality control step, Illumina HiSeq paired-end reads were mapped
547 against each respective genome using bwa 0.7.17-r1188 (56), with Pilon v1.22 (57) used to correct
548 for small indels. Plasmid misassemblies in MS6192 were corrected using alternative assembly
549 methods (Supplementary Appendix).
550
551 **Genome annotation.** The automated annotation of all PacBio sequenced ST101 strains (2
552 complete, 5 draft assemblies) was done using Prokka v1.11 (58) using a custom *Escherichia*
553 database consisting of protein sequences from the EcoCyc website (http://ecocyc.org/) and
554 annotated NDM+ reference plasmids (Supplementary Dataset, Table S11). For the complete
555 genomes: *E. coli* MS6192 and MS6193, prophages, genomic islands and insertion sequences (IS)
556 were identified using PHAST (59), IslandViewer 3 (60) and ISSaga (61) respectively. Draft
557 annotations for MS6192 and MS6193 were then visualised using Artemis (62) and subjected to
558 manual improvement. Insertion sequence, integrons and antimicrobial resistance cassettes within
559 the complete plasmid sequences were manually curated by nucleotide comparisons against
560 ISSaga, Integrall and the Repository of Antibiotic Resistance Cassettes (61, 63, 64) databases,
561 respectively.
562
563 **Additional *E. coli* ST101 genomes included in this study.** To further characterise the ST101
564 lineage, 13 additional ST101 draft genomes were downloaded from GenBank or the SRA. Three

565 genomes (ECO0117, eo1667 and PI7) were only available as Illumina paired-end short reads and
566 were assembled *de-novo* using Spades v3.10.1 with default parameters. Strain names, accessions,
567 sources, sequencing method and available metadata are summarised in Supplementary Dataset,
568 Table S3.
569
570 **Phylogenetic analysis and recombination filtering.** To determine the phylogenetic relationship of
571 ST101 and other *E. coli* lineages, core-genome single nucleotide polymorphism (SNP) trees were
572 constructed. A core-genome SNP alignment of the 20 ST101 strains and an additional 65
573 representative complete *E. coli* genomes (Supplementary Dataset, Table S12) was produced by
574 parSNP v1.2, implemented in the Harvest suite (65) and aligned against the well-characterised
575 ExPEC complete genome *E. coli* EC958 reference (23, 24) defining 182,264 core SNPs. To integrate
576 the ST101 draft and complete genomes into our phylogenomic analyses and account for
577 differences between short and long read technologies, error-free 71-bp paired-end Illumina reads
578 were simulated from the draft assemblies and complete genomes and used as input to produce a
579 whole genome SNP alignment using Snippy v4.3.5 (https://github.com/tseemann/snippy.git)
580 against the ST101 complete genome MS6192. A pseudogenome of this ST101 SNP alignment was
581 used as input for Gubbins v2.3.1 (66), which detects recombination in closely related groups of
582 isolates. Gubbins was run using default settings in "raxml mode" generating a phylogenetic tree
583 with a generalized time-reversible model and gamma correction (GTRGAMMA). This generated a
584 recombination-filtered core-SNP alignment of 2,106 SNPs. Maximum likelihood (ML) phylogenetic
585 trees were then constructed from both core-genome SNP-based alignments with RAxML 8.1.15
586 (67) using the GTRGAMMA model and 1000 bootstrap replicates. The resulting ML phylogenetic
587 trees were visualised using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).
588

589 **Comparative genomics of *E. coli* ST101.** For all *E. coli* ST101 genomes included in this study,
590 acquired AMR genes were functionally annotated by nucleotide comparisons (a cutoff of 90%
591 identity over >=60% query cover) against ResFinder v2.1 (68). O antigen and H antigen typing were
592 determined using SeroType Finder v1.1 (69) and the EcOH database (70) within SRST2 (71), with a
593 nucleotide ID threshold of 85%. Additionally, sequences were screened for the presence of
594 mutations in the QRDR of *gyrA* and *parC*. For each isolate and the *E. coli* K12 strain MG1655
595 reference genome (Genbank accession: U00096), the amino acid sequences of GyrA and ParC
596 were aligned in MEGA v6.06 (72), using MUSCLE (73) and default parameters. All isolates also
597 underwent *fimH in silico* typing using FimTyper v1.0 (https://cge.cbs.dtu.dk/services/FimTyper/).
598 For plasmid analyses, the plasmid replicons were detected using PlasmidFinder v1.3 with a
599 nucleotide ID threshold of 90% and the plasmid Multi-Locus Sequence Typing (pMLST)
600 (*Enterobacteriaceae*) was determined using pMLST v1.4 (74). IncC pMLST was determined *in silico*
601 against the four essential IncC genes *repA*, *parA*, *parB* and *053* (34). BRIG v.0.95 (75), EasyFig
602 v2.2.2 (76) and Phandango v1.1.0 (77) were used to visually compare the genomes. Methods for
603 conjugation assays, phenotypic antimicrobial susceptibility testing and plasmid mating assays can
604 be found in the Supplementary Appendix.
605
606 **Data.** Complete genomes of MS6192 and MS6193, draft genome assemblies and PacBio and
607 Illumina sequence read data for MS6192, MS6193, MS6194, MS6201, MS6203, MS6204 and
608 MS6207 are available under the BioProjects: PRJNA580334, PRJNA580336, PRJNA580337,
609 PRJNA580338, PRJNA580339, PRJNA580341 and PRJNA580340 respectively.
610
611 **Acknowledgements**
612 Author contributions: Conceptualisation: MMA, BMF, MAS and SAB. Investigation: MMA and BMF.
613 Formal Analysis: MMA. MDP, KMP, AMH, SJH and DLP assisted in clinical/wet-lab experiments.

614 BMF, SJH, LWR and RTW assisted in data analysis. KGC, TMC and WFY performed the PacBio
615 sequencing. Illumina sequencing was performed at the Australian Genome Research Facility
616 (AGRF), The University of Queensland, Australia. Resources: TRW, KGC, MAS and SAB. Supervision:
617 BMF, MAS and SAB. Writing (Original Draft Preparation): MMA, BMF and SAB. Writing (Review and
618 Editing): MMA, MDP, AMH, SJH, KGC, MAS and SAB. All authors contributed to the final review
619 and edits.
620
621 Funding: This work was supported by grants from the Australian National Health and Medical
622 Research Council (G1033799) and from the University of Malaya High Impact Research (HIR)
623 Grants (UM-MOHE HIR Grant UM.C/625/1/HIR/MOHE/CHAN/14/1, Grants H-50001-A000027 and
624 FP022-2018A). MAS is supported by a NHMRC Senior Research Fellowship (G1106930). SAB is
625 supported by a NHMRC Career Development Fellowship (G1090456). MMA, SJH, LWR and RTW
626 were supported by an Australian Government Research Training Program Scholarship.
627
628 The authors declare no competing interests.

629 **References**

630 1. Poolman JT, Wacker M. 2016. Extraintestinal Pathogenic Escherichia coli, a Common
631 Human Pathogen: Challenges for Vaccine Development and Progress in the Field. J Infect
632 Dis 213:6-13.
633 2. Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, Walsh TR. 2009.
634 Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel
635 erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae
636 sequence type 14 from India. Antimicrob Agents Chemother 53:5046-54.
637 3. Nordmann P, Poirel L, Walsh TR, Livermore DM. 2011. The emerging NDM
638 carbapenemases. Trends Microbiol 19:588-95.
639 4. Peirano G, Mulvey GL, Armstrong GD, Pitout JD. 2013. Virulence potential and adherence
640 properties of Escherichia coli that produce CTX-M and NDM beta-lactamases. J Med
641 Microbiol 62:525-30.
642 5. Woodford N, Turton JF, Livermore DM. 2011. Multiresistant Gram-negative bacteria: the
643 role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiol Rev
644 35:736-55.
645 6. Zarrilli R, Pournaras S, Giannouli M, Tsakris A. 2013. Global evolution of multidrug-resistant
646 Acinetobacter baumannii clonal lineages. Int J Antimicrob Agents 41:11-9.
647 7. Coque TM, Novais A, Carattoli A, Poirel L, Pitout J, Peixe L, Baquero F, Canton R, Nordmann
648 P. 2008. Dissemination of clonally related Escherichia coli strains expressing extended-
649 spectrum beta-lactamase CTX-M-15. Emerg Infect Dis 14:195-200.
650 8. Bonnin RA, Poirel L, Carattoli A, Nordmann P. 2012. Characterization of an IncFII plasmid
651 encoding NDM-1 from Escherichia coli ST131. PLoS One 7:e34752.
652 9. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan MD,
653 Gomes Moriel D, Peters KM, Davies M, Rogers BA, Dougan G, Rodriguez-Bano J, Pascual A,
654 Pitout JD, Upton M, Paterson DL, Walsh TR, Schembri MA, Beatson SA. 2014. Global
655 dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci U S A
656 111:5694-9.
657 10. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-
658 Cook M, Schembri MA, Beatson SA. 2016. Sequential Acquisition of Virulence and
659 Fluoroquinolone Resistance Has Shaped the Evolution of Escherichia coli ST131. MBio
660 7:e00347-16.
661 11. Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M,
662 Chattopadhyay S, Stegger M, Andersen PS, Pearson T, Riddell K, Rogers P, Scholes D, Kahl
663 B, Keim P, Sokurenko EV. 2013. The epidemic of extended-spectrum-beta-lactamase-
664 producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx.
665 MBio 4:e00377-13.
666 12. Mushtaq S, Irfan S, Sarma JB, Doumith M, Pike R, Pitout J, Livermore DM, Woodford N.
667 2011. Phylogenetic diversity of Escherichia coli strains producing NDM-type
668 carbapenemases. J Antimicrob Chemother 66:2002-5.
669 13. Yoo JS, Kim HM, Koo HS, Yang JW, Yoo JI, Kim HS, Park HK, Lee YS. 2013. Nosocomial
670 transmission of NDM-1-producing Escherichia coli ST101 in a Korean hospital. J Antimicrob
671 Chemother 68:2170-2.
672 14. Poirel L, Savov E, Nazli A, Trifonova A, Todorova I, Gergova I, Nordmann P. 2014. Outbreak
673 caused by NDM-1- and RmtB-producing Escherichia coli in Bulgaria. Antimicrob Agents
674 Chemother 58:2472-4.

675 15. Mantilla-Calderon D, Jumat MR, Wang T, Ganesan P, Al-Jassim N, Hong PY. 2016. Isolation
676 and Characterization of NDM-Positive Escherichia coli from Municipal Wastewater in
677 Jeddah, Saudi Arabia. Antimicrob Agents Chemother 60:5223-31.
678 16. Mora A, Blanco M, Lopez C, Mamani R, Blanco JE, Alonso MP, Garcia-Garrote F, Dahbi G,
679 Herrera A, Fernandez A, Fernandez B, Agulla A, Bou G, Blanco J. 2011. Emergence of clonal
680 groups O1:HNM-D-ST59, O15:H1-D-ST393, O20:H34/HNM-D-ST354, O25b:H4-B2-ST131
681 and ONT:H21,42-B1-ST101 among CTX-M-14-producing Escherichia coli clinical isolates in
682 Galicia, northwest Spain. Int J Antimicrob Agents 37:16-21.
683 17. Ranjan A, Shaik S, Mondal A, Nandanwar N, Hussain A, Semmler T, Kumar N, Tiwari SK,
684 Jadhav S, Wieler LH, Ahmed N. 2016. Molecular Epidemiology and Genome Dynamics of
685 New Delhi Metallo-beta-Lactamase-Producing Extraintestinal Pathogenic Escherichia coli
686 Strains from India. Antimicrob Agents Chemother 60:6795-6805.
687 18. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K,
688 Song Y, Tsai YC, Boitano M, Dayal J, Brooks SY, Schmidt B, Young AC, Thomas JW, Bouffard
689 GG, Blakesley RW, Mullikin JC, Korlach J, Henderson DK, Frank KM, Palmore TN, Segre JA.
690 2014. Single-molecule sequencing to track plasmid diversity of hospital-associated
691 carbapenemase-producing Enterobacteriaceae. Sci Transl Med 6:254ra126.
692 19. Zowawi HM, Forde BM, Alfaresi M, Alzarouni A, Farahat Y, Chong TM, Yin WF, Chan KG, Li J,
693 Schembri MA, Beatson SA, Paterson DL. 2015. Stepwise evolution of pandrug-resistance in
694 Klebsiella pneumoniae. Sci Rep 5:15082.
695 20. Djoko KY, Achard MES, Phan MD, Lo AW, Miraula M, Prombhul S, Hancock SJ, Peters KM,
696 Sidjabat H, Harris PN, Mitic N, Walsh TR, Anderson GJ, Shafer WM, Paterson DL, Schenk G,
697 McEwan AG, Schembri MA. 2017. Copper ions and coordination complexes as novel
698 carbapenem adjuvants. Antimicrob Agents Chemother doi:10.1128/aac.02280-17.
699 21. Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, Chaudhary U,
700 Doumith M, Giske CG, Irfan S, Krishnan P, Kumar AV, Maharjan S, Mushtaq S, Noorie T,
701 Paterson DL, Pearson A, Perry C, Pike R, Rao B, Ray U, Sarma JB, Sharma M, Sheridan E,
702 Thirunarayan MA, Turton J, Upadhyay S, Warner M, Welfare W, Livermore DM, Woodford
703 N. 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the
704 UK: a molecular, biological, and epidemiological study. Lancet Infect Dis 10:597-602.
705 22. Oshima K, Toh H, Ogura Y, Sasamoto H, Morita H, Park SH, Ooka T, Iyoda S, Taylor TD,
706 Hayashi T, Itoh K, Hattori M. 2008. Complete genome sequence and comparative analysis
707 of the wild-type commensal Escherichia coli strain SE11 isolated from a healthy adult. DNA
708 Res 15:375-86.
709 23. Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, Bachmann N, Szubert M, Sidjabat HE,
710 Paterson DL, Upton M, Schembri MA. 2011. Insights into a multidrug resistant Escherichia
711 coli pathogen of the globally disseminated ST131 lineage: genome analysis and virulence
712 mechanisms. PLoS One 6:e26578.
713 24. Forde BM, Ben Zakour NL, Stanton-Cook M, Phan MD, Totsika M, Peters KM, Chan KG,
714 Schembri MA, Upton M, Beatson SA. 2014. The complete genome sequence of Escherichia
715 coli EC958: a high quality reference sequence for the globally disseminated multidrug
716 resistant E. coli O25b:H4-ST131 clone. PLoS One 9:e104400.
717 25. Alteri CJ, Mobley HL. 2016. The Versatile Type VI Secretion System. Microbiol Spectr 4.
718 26. Ren CP, Beatson SA, Parkhill J, Pallen MJ. 2005. The Flag-2 locus, an ancestral gene cluster,
719 is potentially associated with a novel flagellar system from Escherichia coli. J Bacteriol
720 187:1430-40.
721 27. Zhou M, Guo Z, Duan Q, Hardwidge PR, Zhu G. 2014. Escherichia coli type III secretion
722 system 2: a new kind of T3SS? Vet Res 45:32.

723 28. Hopkins KL, Davies RH, Threlfall EJ. 2005. Mechanisms of quinolone resistance in
724 Escherichia coli and Salmonella: recent developments. Int J Antimicrob Agents 25:358-73.
725 29. Johnson TJ, Elnekave E, Miller EA, Munoz-Aguayo J, Flores Figueroa C, Johnston B, Nielson
726 DW, Logue CM, Johnson JR. 2019. Phylogenomic Analysis of Extraintestinal Pathogenic
727 Escherichia coli Sequence Type 1193, an Emerging Multidrug-Resistant Clonal Group.
728 Antimicrob Agents Chemother 63.
729 30. Dortet L, Nordmann P, Poirel L. 2012. Association of the emerging carbapenemase NDM-1
730 with a bleomycin resistance protein in Enterobacteriaceae and Acinetobacter baumannii.
731 Antimicrob Agents Chemother 56:1693-7.
732 31. Campos JC, da Silva MJF, dos Santos PRN, Barros EM, Pereira MdO, Seco BMS, Magagnin
733 CM, Leiroz LK, de Oliveira TGM, de Faria-Júnior C, Cerdeira LT, Barth AL, Sampaio SCF,
734 Zavascki AP, Poirel L, Sampaio JLM. 2015. Characterization of Tn3000, a Transposon
735 Responsible for bla(NDM-1) Dissemination among Enterobacteriaceae in Brazil, Nepal,
736 Morocco, and India. Antimicrobial Agents and Chemotherapy 59:7387-7395.
737 32. Marquez-Ortiz RA, Haggerty L, Olarte N, Duarte C, Garza-Ramos U, Silva-Sanchez J, Castro
738 BE, Sim EM, Beltran M, Moncada MV, Valderrama A, Castellanos JE, Charles IG, Vanegas N,
739 Escobar-Perez J, Petty NK. 2017. Genomic Epidemiology of NDM-1-Encoding Plasmids in
740 Latin American Clinical Isolates Reveals Insights into the Evolution of Multidrug Resistance.
741 Genome Biol Evol 9:1725-1741.
742 33. Sugawara Y, Akeda Y, Sakamoto N, Takeuchi D, Motooka D, Nakamura S, Hagiya H,
743 Yamamoto N, Nishi I, Yoshida H, Okada K, Zin KN, Aye MM, Tonomo K, Hamada S. 2017.
744 Genetic characterization of blaNDM-harboring plasmids in carbapenem-resistant
745 Escherichia coli from Myanmar. PLoS One 12:e0184720.
746 34. Hancock SJ, Phan MD, Peters KM, Forde BM, Chong TM, Yin WF, Chan KG, Paterson DL,
747 Walsh TR, Beatson SA, Schembri MA. 2016. Identification of IncA/C Plasmid Replication and
748 Maintenance Genes and Development of a Plasmid Multi-Locus Sequence-Typing Scheme.
749 Antimicrob Agents Chemother doi:10.1128/aac.01740-16.
750 35. Fiett J, Baraniak A, Izdebski R, Sitkiewicz I, Zabicka D, Meler A, Filczak K, Hryniewicz W,
751 Gniadkowski M. 2014. The first NDM metallo-beta-lactamase-producing
752 Enterobacteriaceae isolate in Poland: evolution of IncFII-type plasmids carrying the
753 bla(NDM-1) gene. Antimicrob Agents Chemother 58:1203-7.
754 36. Toleman MA, Bugert JJ, Nizam SA. 2015. Extensively drug-resistant New Delhi metallo-
755 beta-lactamase-encoding bacteria in the environment, Dhaka, Bangladesh, 2012. Emerg
756 Infect Dis 21:1027-30.
757 37. Wailan AM, Paterson DL, Kennedy K, Ingram PR, Bursle E, Sidjabat HE. 2015. Genomic
758 Characteristics of NDM-Producing Enterobacteriaceae Isolates in Australia and Their
759 blaNDM Genetic Contexts. Antimicrob Agents Chemother 60:136-41.
760 38. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference
761 centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-6.
762 39. He S, Chandler M, Varani AM, Hickman AB, Dekker JP, Dyda F. 2016. Mechanisms of
763 Evolution in High-Consequence Drug Resistance Plasmids. MBio 7.
764 40. Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. 2018. A genomic overview of the population
765 structure of Salmonella. PLoS Genet 14:e1007261.
766 41. Conlan S, Park M, Deming C, Thomas PJ, Young AC, Coleman H, Sison C, Weingarten RA,
767 Lau AF, Dekker JP, Palmore TN, Frank KM, Segre JA. 2016. Plasmid Dynamics in KPC-Positive
768 Klebsiella pneumoniae during Long-Term Patient Colonization. MBio 7.
769 42. Forde BM, Roberts LW, Phan MD, Peters KM, Fleming BA, Russell CW, Lenherr SM, Myers
770 JB, Barker AP, Fisher MA, Chong TM, Yin WF, Chan KG, Schembri MA, Mulvey MA, Beatson

771 SA. 2019. Population dynamics of an Escherichia coli ST131 lineage during recurrent urinary
772 tract infection. Nat Commun 10:3643.
773 43. Diene SM, Rolain JM. 2014. Carbapenemase genes and genetic platforms in Gram-negative
774 bacilli: Enterobacteriaceae, Pseudomonas and Acinetobacter species. Clin Microbiol Infect
775 20:831-8.
776 44. Hudson CM, Bent ZW, Meagher RJ, Williams KP. 2014. Resistance determinants and mobile
777 genetic elements of an NDM-1-encoding Klebsiella pneumoniae strain. PLoS One 9:e99209.
778 45. Ingle DJ, Levine MM, Kotloff KL, Holt KE, Robins-Browne RM. 2018. Dynamics of
779 antimicrobial resistance in intestinal Escherichia coli from children in community settings in
780 South Asia and sub-Saharan Africa. Nat Microbiol 3:1063-1073.
781 46. Su M, Satola SW, Read TD. 2019. Genome-Based Prediction of Bacterial Antibiotic
782 Resistance. J Clin Microbiol 57.
783 47. Toleman MA, Walsh TR. 2011. Combinatorial events of insertion sequences and ICE in
784 Gram-negative bacteria. FEMS Microbiol Rev 35:912-35.
785 48. Yamamoto M, Pop-Vicas AE. 2014. Treatment for infections with carbapenem-resistant
786 Enterobacteriaceae: what options do we still have? Crit Care 18:229.
787 49. Fernandes MR, McCulloch JA, Vianello MA, Moura Q, Perez-Chaparro PJ, Esposito F, Sartori
788 L, Dropa M, Matte MH, Lira DP, Mamizuka EM, Lincopan N. 2016. First Report of the
789 Globally Disseminated IncX4 Plasmid Carrying the mcr-1 Gene in a Colistin-Resistant
790 Escherichia coli Sequence Type 101 Isolate from a Human Infection in Brazil. Antimicrob
791 Agents Chemother 60:6415-7.
792 50. Zheng B, Dong H, Xu H, Lv J, Zhang J, Jiang X, Du Y, Xiao Y, Li L. 2016. Coexistence of MCR-1
793 and NDM-1 in Clinical Escherichia coli Isolates. Clin Infect Dis 63:1393-1395.
794 51. Zhong LL, Zhang YF, Doi Y, Huang X, Zhang XF, Zeng KJ, Shen C, Patil S, Xing Y, Zou Y, Tian
795 GB. 2017. Coproduction of MCR-1 and NDM-1 by Colistin-Resistant Escherichia coli Isolated
796 from a Healthy Individual. Antimicrob Agents Chemother 61.
797 52. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A,
798 Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome
799 assemblies from long-read SMRT sequencing data. Nat Methods 10:563-9.
800 53. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with
801 gene gain, loss and rearrangement. PLoS One 5:e11147.
802 54. Sullivan MJ, L. BZN, M. FB, Stanton-Cook M, Beatson SA. 2015. Contiguity: Contig adjacency
803 graph construction and visualisation. PeerJ PrePrints 3:e1273.
804 55. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local
805 alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics
806 13:238.
807 56. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
808 transform. Bioinformatics 25:1754-60.
809 57. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
810 Wortman J, Young SK, Earl AM. 2014. Pilon: An Integrated Tool for Comprehensive
811 Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE 9:e112963.
812 58. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.
813 59. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool.
814 Nucleic Acids Res 39:W347-52.
815 60. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N,
816 McArthur AG, Langille MG, Brinkman FS. 2015. IslandViewer 3: more flexible, interactive
817 genomic island discovery, visualization and analysis. Nucleic Acids Res 43:W104-8.

818 61. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of
819 web-based methods for high throughput identification and semi-automatic annotation of
820 insertion sequences in prokaryotic genomes. Genome Biol 12:R30.
821 62. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated
822 platform for visualization and analysis of high-throughput sequence-based experimental
823 data. Bioinformatics 28:464-9.
824 63. Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. 2009. INTEGRALL: a
825 database and search engine for integrons, integrases and gene cassettes. Bioinformatics
826 25:1096-8.
827 64. Tsafnat G, Copty J, Partridge SR. 2011. RAC: Repository of Antibiotic resistance Cassettes.
828 Database (Oxford) 2011:bar054.
829 65. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-
830 genome alignment and visualization of thousands of intraspecific microbial genomes.
831 Genome Biol 15:524.
832 66. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR.
833 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome
834 sequences using Gubbins. Nucleic Acids Res 43:e15.
835 67. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
836 thousands of taxa and mixed models. Bioinformatics 22:2688-90.
837 68. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM,
838 Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob
839 Chemother 67:2640-4.
840 69. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and Easy In
841 Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. J
842 Clin Microbiol 53:2410-26.
843 70. Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M, Stinear T, Levine MM, Robins-
844 Browne RM, Holt KE. 2016. In silico serotyping of E. coli from short read data identifies
845 limited novel O-loci but extensive diversity of O: H serotype combinations within and
846 between pathogenic lineages. Microbial Genomics 2.
847 71. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014.
848 SRST2: Rapid genomic surveillance for public health and hospital microbiology labs.
849 Genome Med 6:90.
850 72. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary
851 Genetics Analysis version 6.0. Mol Biol Evol 30:2725-9.
852 73. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and
853 space complexity. BMC Bioinformatics 5:113.
854 74. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller
855 Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using
856 PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother
857 58:3895-903.
858 75. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator
859 (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402.
860 76. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer.
861 Bioinformatics 27:1009-10.
862 77. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. 2018.
863 Phandango: an interactive viewer for bacterial population genomics. Bioinformatics
864 34:292-293.
865