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[4,5,2], Steven J. Hancock[1,2], Leah W. Roberts[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.

*** Corresponding authors:**

Scott A. Beatson, Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane 4072, QLD, Australia;

Telephone: +61-7-33654863; Email: s.beatson@uq.edu.au

OR

Melinda M. Ashcroft, affiliation is now: The Department of Immunology and Microbiology, The Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Melbourne, VIC, 3000. Telephone: +61-3-83443353; Email: melinda.ashcroft@unimelb.edu.au

Keywords:

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

Word count (main text): 5592 (including in-text citations)

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_{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_{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 (ESCs). Additionally, we characterised different pathways for bla_{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 (ARI-A). In contrast, we revealed recent transfer of a large bla_{NDM-1} resistance island between F-type plasmids. The complex acquisition pathways characterised here highlight the benefits of long-read SMRT 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 MGEs in this emerging MDR lineage.

Abstract word count: 224

Introduction

The successful treatment of *E. coli* infections is complicated by the rising prevalence of antibiotic resistance among clinical and community isolates (Poolman and Wacker 2016). Carbapenems are considered a class of last-resort antibiotics, however their effectiveness has been challenged by the emergence of bacteria capable of hydrolysing carbapenems and most β-lactams. Carbapenem resistance has disseminated worldwide, as genes that encode carbapenemases are easily transferred via MGEs such as plasmids, transposons and integrons. In 2008, $bla_{\text{NDM-1}}$ was first reported in a *Klebsiella pneumoniae* strain isolated from a Swedish patient who had recently travelled to India (Yong et al. 2009). Since then, the $bla_{\text{NDM-1}}$ gene and its variants has been identified in Gram-negative bacteria in more than fifty countries including the UK, India, Pakistan and Bangladesh, across Europe, China, the US, Canada and Australia (Nordmann et al. 2011). NDM-positive clonal lineages can be found in multiple different organisms including *K. pneumoniae, Acinetobacter baumannii* and *E. coli* (Peirano et al. 2013). For example, in *K. pneumoniae*, several of the first $bla_{\text{NDM-1}}$ reports between 2009-2011 involved the multi-drug resistant (MDR) sequence type (ST)14 clone (Woodford et al. 2011). In *A. baumannii*, international clonal lineages I (Clonal Complex (CC)109/CC1) and II (CC92/CC2), ST25 and ST85 are the four dominant NDM-positive clonal lineages (Zarrilli et al. 2013). In *E. coli*, $bla_{\text{NDM-1}}$ has been detected worldwide, with isolates frequently typed as ST405, ST131 and ST101 (Peirano et al. 2013) and also found to carry the $bla_{\text{CTX-M-15}}$ ESBL gene (Coque et al. 2008). The first characterisation of an FII plasmid carrying the $bla_{\text{NDM-1}}$ gene was the *E. coli* ST131 strain GUE (pGUE-NDM, Genbank accession: JQ364967) (Bonnin et al. 2012).

Other *E. coli* MDR clones such as ST131 have been well-characterised genetically (Price et al. 2013; Petty et al. 2014; Ben Zakour et al. 2016), but very little is known about *E. coli* ST101 at the genomic level, despite numerous reported cases of carbapenemase-producing *E. coli* ST101 (Mora et al. 2011; Mushtaq et al. 2011; Peirano et al. 2013; Yoo et al. 2013; Poirel et al. 2014; Mantilla-Calderon

et al. 2016). Mushtaq *et al*, 2011 showed that ST101 strains have acquired NDM-positive plasmids of different incompatibility groups and demonstrated diverse PFGE banding patterns for *E. coli* ST101 strains, with some evidence of clonal expansion due to clustering of pulsotypes (Mushtaq et al. 2011). Additionally, serotyping of five ST101 isolates indicated that they were O-non-typeable:H21/42 and contained an array of virulence genes; including *fimH*, *pap*, *sfa/focDE* (fimbrial genes), *iucD, iroN* (siderophore genes), *iss*, *traT* (protection genes), and *tsh* (serine protease gene) (Mora et al. 2011). Other studies have reported varying levels of virulence potential in ST101 isolates; however in all cases, these strains contained numerous adhesins, autotransporters and siderophores normally associated with extra-intestinal pathogenic *E. coli* (Mora et al. 2011; Peirano et al. 2013; Ranjan et al. 2016).

There are now several *E. coli* ST101 genomes and NDM-positive plasmid sequences available in public databases, but as yet there are no analyses of any *E. coli* ST101 complete genomes that have been published. Draft assemblies provide limited information in terms of genomic context of mobile elements such as insertion sequences (IS), phages, genomic islands and plasmids. The complete assembly of these mobile elements is crucial for characterising the genomic context of resistance elements (Conlan et al. 2014; Zowawi et al. 2015). Here we present a comprehensive genomic analysis of $bla_{\text{NDM-1}}$ carriage in *E. coli* ST101. Using a combination of PacBio SMRT sequencing and Illumina sequencing, we defined the plasmid context of bla_{NDM-1} in seven ST101 isolates. We also report the complete genome of two representative ST101 strains: MS6192 and MS6193, including manual curation of all MGEs. Using publicly available ST101 draft genomes, we also examined the phylogeny of ST101 and defined two major clades (1 and 2), with the majority of Clade 1 strains encoding resistance to carbapenems and extended spectrum cephalosporins (ESCs).

Results

NDM-positive *E. coli* ST101 genomes. During our molecular characterisation of 16 carbapenemase producing *E. coli* collected from India and the United Kingdom (Kumarasamy et al. 2010; Djoko et al. 2017) we found seven that were phylogroup B1, sequence type (ST)101 and NDM-positive by PCR (Supplementary Appendix and Supplementary Dataset, Table S1). To investigate the genomic context of $bla_{\text{NDM-1}}$ and characterise ST101 at the genomic level we undertook whole genome sequencing (WGS). The genomes of seven $bla_{\text{NDM-1}}$-positive ST101 strains were assembled from a combination of PacBio RSI or RSII long-read data and Illumina HiSeq paired-end data. In all cases we could assemble a single plasmid encoding $bla_{\text{NDM-1}}$ together with several other antimicrobial resistance (AMR) genes (see below). There were sufficient long-reads to enable the assembly of complete, finished quality, genome sequences for two of the strains: MS6192 and MS6193 (Supplementary Dataset, Table S2).

E. coli MS6192 is comprised of a single circular chromosome of 4,879,059 bp (Table 1). Four circularised contigs of 142,820 bp, 76,661 bp, 86,357 bp and 3,608 bp represent the large $bla_{\text{NDM-1}}$ encoding MDR plasmid pMS6192A-NDM (4 replicons: FII and FII(pCoo) replicons and an FIA and FIB replicon, IncF replicon sequence type (IncF RST) F36/F22:A1:B20), pMS6192B (FII, F2:A-:B-), pMS6192C (Incl1, PMLST ST173) and the small cryptic plasmid pMS6192D (ColRNAI), respectively. *E. coli* MS6193 is comprised of a single circular chromosome of 4,922,872 bp. Additionally, there were three circularised contigs; 142,890 bp, 76,661 bp and 4,367 bp, which represent the large $bla_{\text{NDM-1}}$ encoding MDR plasmid pMS6193A-NDM (4 replicons: FII and FII(pCoo), FIA and FIB, F36/F22:A1:B20), pMS6193B (FII, F2:A-:B-) and pMS6193C (untypeable).

The chromosomes and large MDR plasmids of MS6192 and MS6193 were almost identical, differing by 12 single nucleotide polymorphisms (SNPs) within the substitution-only core-chromosome. The genomes could also be distinguished by the presence of an additional prophage (Phi8) and Transposon 7 (Tn*7*)-like transposon in *E. coli* MS6193 (Supplementary Appendix, Fig S1), an additional large plasmid in MS6192 (pMS6192C) and the presence of two different cryptic plasmids (pMS6192D and pMS6193C).

Table 1. Summary of genomic information for *E. coli* ST101 strains MS6192 and MS6193.

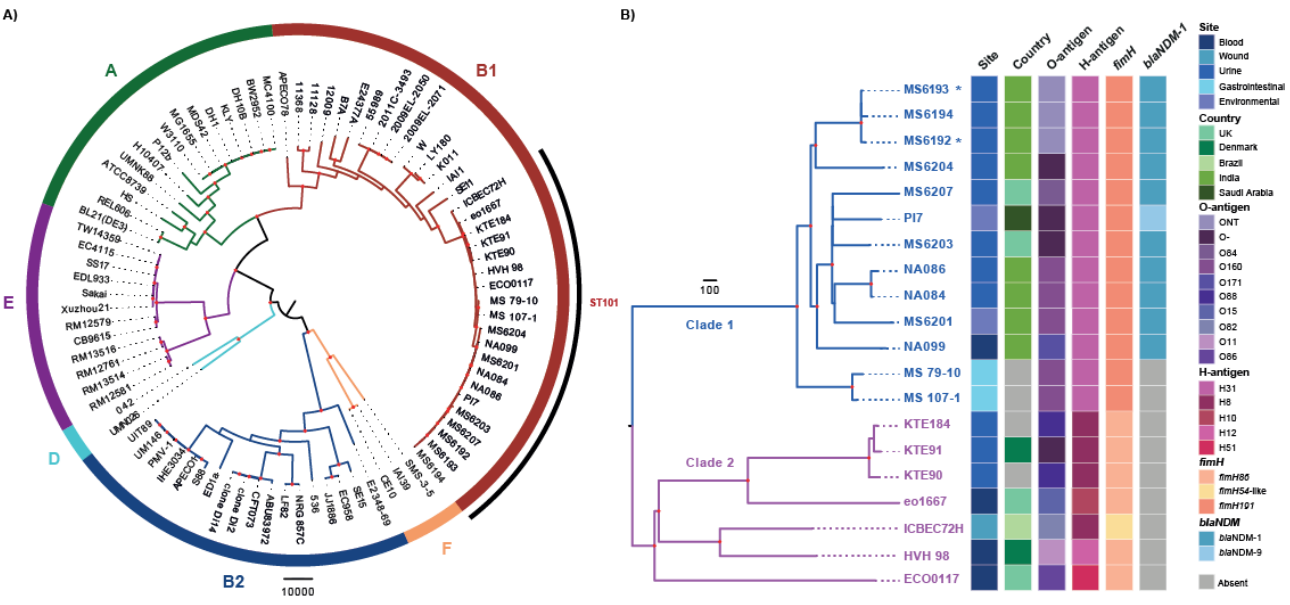
	MS6192 chr[1]	MS6193 chr[1]	pMS6192A-NDM/ pMS6193A-NDM (FII/FII:F1A:F1B)	pMS6192B/ pMS6193B (FII)	pMS6192C[2] (IncI1)	pMS6193C[2] (colRNAI)	pMS6192D (untypeable)
Size (bp)	4,879,059	4,922,872	142,820/ 142,890	76,661/ 76,661	86,357	4,367	3,608
No. of genes	4,834	4,876	174/173	84/84	94	6	2
No. of CDS	4,570	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 content (%)	50.69	50.69	51.5/51.50	51.91/51.91	50.14	54.39	43.15

[1]chr = chromosome

[2]MS6192 has 4 plasmids (pMS6192A, pMS6192B, pMS6192C and pMS6192D); MS6193 has 3 plasmids (pMS6193A, pMS6193B, pMS6193C); There is no genetic relationship between pMS6192C and pMS6193C.

MS6194, MS6201, MS6203, MS6204 and MS6207 strains were assembled as draft genomes with gapped chromosomes and circularised plasmid sequences. The five draft *E. coli* ST101 genomes range in size from 5,150,883 to 5,434,929 bp with the number of contigs ranging from 16 to 55

157 (Supplementary Dataset, Table S2). Each draft genome contained a single *bla*$_{NDM-1}$ encoding MDR
158 plasmid and several plasmid replicons ranging from one (MS6203) to seven (MS6204).
159
160 **Phylogenomic analysis of *E. coli* ST101 reveals a monophyletic lineage within the B1 phylogroup.**
161 To explore the context of ST101 within the *E. coli* phylogeny we carried out a phylogenetic
162 reconstruction of 65 representative, publicly available, complete *E. coli* genomes together with 20
163 ST101 genomes. These included the seven ST101 genomes sequenced in the present study and 13
164 *E. coli* ST101 draft genomes that were previously published and available in Genbank or the
165 Sequence Read Archive (SRA) on October 1st, 2016 (Supplementary Dataset, Table S3). The
166 additional ST101 genomes were obtained from several clinical sources and included isolates from
167 urine (n = 5), blood (n = 4), wound (n=1) and intestinal microbiota (n = 2). A single environmental
168 isolate from untreated wastewater was also included. All 20 *E. coli* ST101 strains cluster together in
169 a single lineage within the B1 phylogroup (Figure 1A). Based on the representative complete
170 genomes included in this phylogeny, the *E. coli* ST101 lineage is most closely related to *E. coli* SE11,
171 a fecal strain of serotype O152:H28, isolated from a healthy adult (Oshima et al. 2008).
172



173

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 MS 107-1 and MS 79-10, basal to the remaining NDM-positive ST101 strains, suggests that $bla_{\text{NDM-1}}$ was acquired once in Clade 1 and has been inherited vertically.

To further characterise the ST101 lineage we examined key virulence and serotyping gene loci. Clade 1 isolates have the *fimH*191 type 1 fimbrial adhesion allele, whereas Clade 2 isolates possess the *fimH*86 allele except for ICBEC72H, which encodes a novel *fimH*54-like variant. Several serogroups were predicted across both clades: O160 (MS107-1, MS79-10, MS6201, NA084 and NA086), O171 (NA099), O84 (MS6207) and a novel, non-typeable O antigen region (MS6192, MS6193 and MS6194), but Clade 1 is distinguished by a *fliC* gene encoding the H31 antigen. Additionally, three Clade 1 strains lack part of the O antigen gene locus: MS6203, MS6204 and PI7 (*wzx*, *wzy*, *wzm* and *wzt*).

Mobile genetic elements are variably conserved in *E. coli* ST101. The major differences between these ST101 genomes were MGEs, such as those defined in *E. coli* MS6192 and MS6193 (Figure 2

and Supplementary dataset, Table S4). MGEs in MS6192 and MS6193 include genomic islands (GI) within known integration tRNA hotspots (GI-*leuX*, GI-*pheU*, GI-*pheV*, GI-*glyU*), prophage elements and a Transposon-*7* (Tn*7*)-like transposon (Supplementary Dataset, Table S5). While no MGE that was defined in MS6192 or MS6193 is completely conserved across the ST101 lineage, GI-*leuX* is present in most strains from Clades 1 and 2, and Prophage (Phi) 3 is conserved in all Clade 1 strains. MS6192, MS6193 (and MS6194) differ in only two chromosomal MGEs (Phi8 and Tn7-like GI) consistent with their near identical core chromosomes. In contrast to the GI-*pheV* islands of phylogroup B2 ExPEC/UPEC strains, such as the 75,054 bp locus in the *E. coli* ST131 strain EC958 (Totsika et al. 2011; Forde et al. 2014), the GI within the tRNA-*pheV* locus in MS6192 and MS6193 is relatively small (9,191 bp) and does not contain known virulence modules. Other genomic regions of difference (RD) that are known to be variable across *E. coli* are present in MS6192 and MS6193 (Supplementary Dataset, Table S6) and are also conserved across the ST101 lineage. These include a remnant of the Type VI secretion system (T6SS) (Alteri and Mobley 2016), the Flag-2 lateral flagellar locus (Ren et al. 2005) and a degenerate Type III secretion locus 2 (ETT2), most commonly found in intestinal pathogenic *E. coli* from the B1 phylogroup and rarely found in ExPEC (Zhou et al. 2014). *E. coli* ST101 strains MS6192 and MS6193 also share 44 chromosomal insertion sequences (IS) (Supplementary Dataset, Table S7).

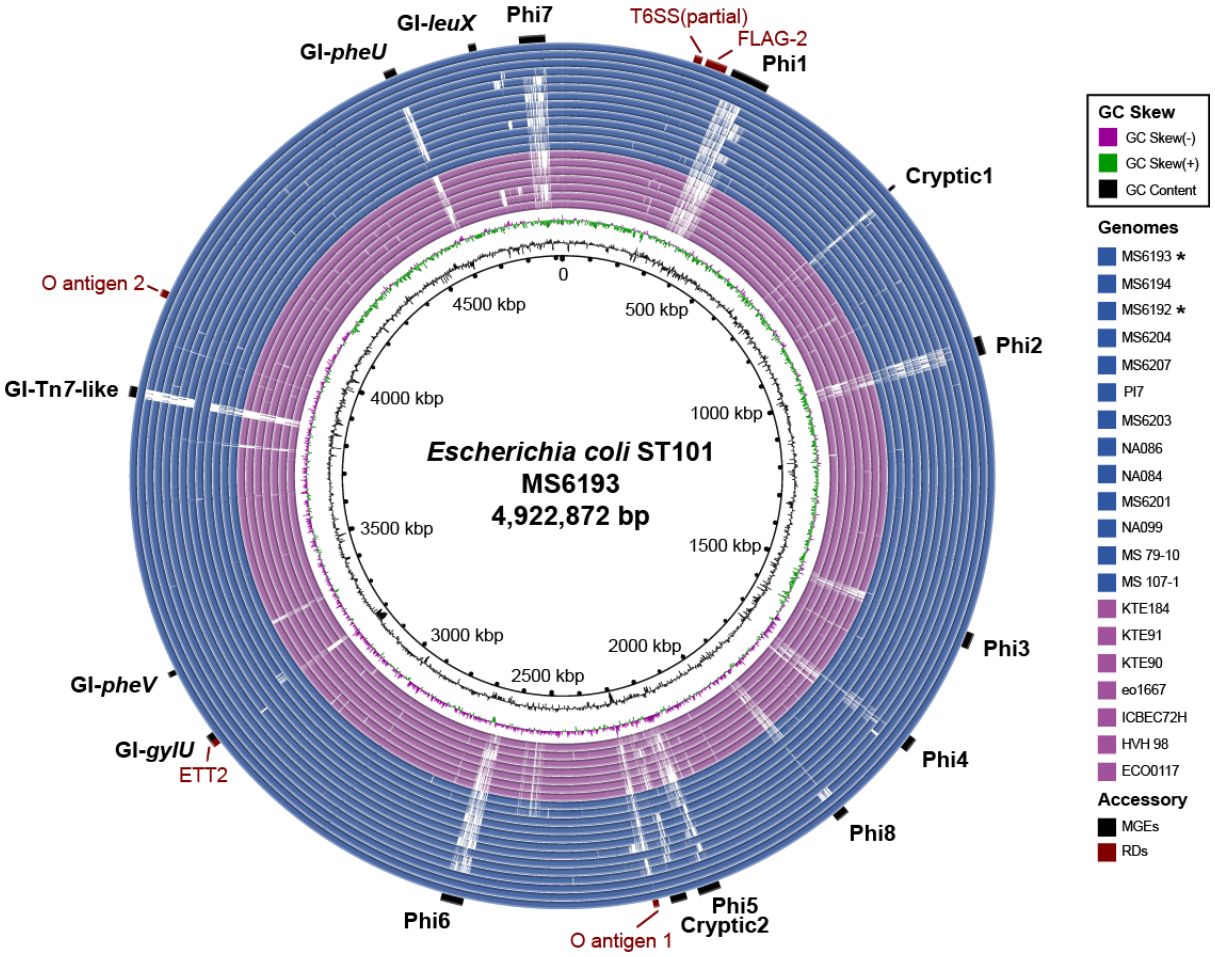
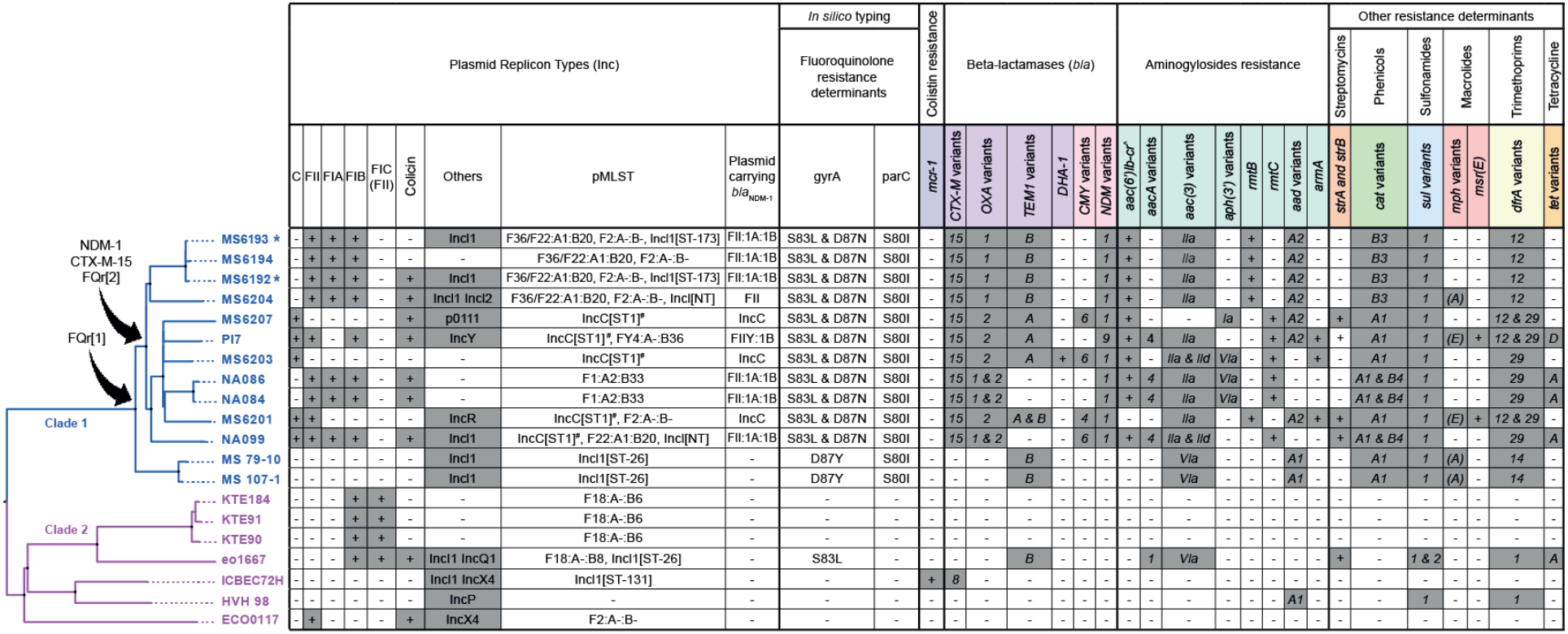


Figure 2. Genomic map of ST101 *E. coli* showing mobile genetic elements and other notable genomic regions. Visualisation of the *E. coli* MS6193 genome compared to 19 *E. coli* ST101 complete and draft genomes. The innermost circles represent GC skew (green/purple) and GC content (black) of *E. coli* MS6193. The degree of coloured shading indicates the nucleotide identity between MS6193 and each *E. coli* ST101 genome. Nucleotide comparisons are coloured based on identity of between 70% and 100% (dark shading = high, light shading = low). *E. coli* genomes are arranged according to their previously defined phylogenetic relationship as follows from the innermost ring: Clade 2 (pink): ECO0117, HVH 98, ICBEC72H, eo1667, KTE90, KTE91, KTE184; Clade 1 (blue): MS 107-1, MS 79-10, NA099, MS6201, NA084, NA086, MS6203, PI7, MS6207, MS6204, MS6192*, MS6194, MS6193*. *Denotes complete genomes. Outer ring indicates mobile genetic elements (MGEs) and select ST101 regions of difference (RDs) within the *E. coli* MS6193 genome. Image prepared using BRIG.

Most *E. coli* ST101 Clade 1 strains contain multiple antibiotic resistance genes. ST101 strains from Clade 1 possessed a high number of antibiotic resistance genes, including genes that encode resistance to beta-lactams (including carbapenems), aminoglycosides, streptomycins, phenicols, sulphonamides, macrolides, trimethoprims and tetracyclines (Figure 3). All NDM-positive ST101 Clade 1 isolates also contain at least one copy of the *bla*$_{\text{CTX-M-15}}$ gene that confers resistance to cephalosporins. In other *E. coli*, *bla*$_{\text{CTX-M-15}}$ is often found on conjugative F-type plasmids as part of a

241 Tn*3*-like IS*Ecp1*-*bla*$_{CTX-M-15}$-orf477 mobile element (Price et al. 2013). We could determine the
242 genomic location of *bla*$_{CTX-M-15}$ in all seven of the PacBio sequenced ST101 strains, showing that the
243 *bla*$_{CTX-M-15}$ insertion site was chromosomally located in six of the seven strains (Supplementary
244 Dataset, Table S8). It appears that *bla*$_{CTX-M-15}$ has been acquired once and mobilised to several
245 different genomic locations by an IS*Ecp1* in Clade 1 and inherited vertically in some cases
246 (Supplementary Appendix).
247
248 Fluoroquinolone resistance in *E. coli* is frequently associated with chromosomal point mutations in
249 the quinolone resistance-determining region (QRDR) of *gyrA* (codons 67-106) and *parC* (codons 56-
250 108) (Hopkins et al. 2005). All Clade 1 strains had a non-synonymous mutation in the QRDR of *parC*
251 (S80I) (Figure 3). Additionally, two non-synonymous mutations in the QRDR of *gyrA* were detected
252 in all NDM-positive Clade 1 isolates, characterised by the amino acid substitutions S83L and D87N.
253 The two Clade 1 intestinal isolates MS107-1 and MS79-10 have a different single amino acid
254 substitution in *gyrA* (D87Y) but lack both *bla*$_{NDM-1}$ and *bla*$_{CTX-M-15}$. These results suggest recent
255 development of extensive MDR in a sub-lineage of Clade 1 (hereafter referred to as Clade 1A),
256 similar to the emergence of other fluoroquinolone MDR resistant *E. coli* clonal lineages such as
257 ST131 and ST1193 (Petty et al. 2014; Johnson et al. 2019). Phenotypic antimicrobial resistance
258 testing of the seven PacBio sequenced ST101 strains in this study confirmed that phenotype could
259 be predicted from genotype in most cases (Supplementary Appendix and Supplementary Dataset,
260 Table S9). Notably Clade 2 strains did not encode QRDR mutations, with only HVH 98, ICBEC72H and
261 eo1667 possessing any acquired resistance genes.



	Plasmid Replicon Types (Inc)									*In silico* typing		Colistin resistance	Beta-lactamases (*bla*)						Aminoglycosides resistance								Other resistance determinants					
										Fluoroquinolone resistance determinants																					
	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*	Streptomycins	Phenicols	Sulfonamides	Macrolides	Trimethoprims	Tetracycline
																											strA and strB	*cat* variants	*sul* variants	*mph* variants	*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.

The *bla*$_{NDM-1}$ genetic environment is highly conserved in *E. coli* ST101. All seven PacBio sequenced strains contained a plasmid-encoded *bla*$_{NDM-1}$ gene (Figure 4). In almost all cases the genetic structure of the *bla*$_{NDM-1}$ module was identical and matched other previously described *bla*$_{NDM-1}$ modules, consisting of a 258-287 bp fragment of IS*Aba125* containing the -35 promoter region, the *bla*$_{NDM-1}$ gene, the bleomycin resistance gene (*ble*$_{MBL}$) and a truncated phosphoribosylanthrantile isomerase (*iso*/*trpC*) (Dortet et al. 2012). The surrounding plasmid context differed between the seven strains, but like *bla*$_{CTX-M-15}$, the structural variations (inversions, transpositions and insertions/deletions) are congruent with a single *bla*$_{NDM-1}$ acquisition prior to divergence of Clade 1A. For example, only pMS6203A-NDM has retained a complete IS*Aba125* element upstream of *bla*$_{NDM-1}$ suggestive of horizontal transfer. Additionally, downstream of the *bla*$_{NDM-1}$ module, *iso*/*tat* has been truncated by the insertion of the *bla*$_{DHA-1}$ gene. In pMS6201A-NDM, truncation of IS*Aba125* was mediated by the insertion of an IS*3000.* The two IS*3000* flanking the *bla*$_{NDM-1}$/*ble*$_{MBL}$ region form the composite transposon Tn*3000* (Campos et al. 2015). Truncation of IS*Aba125* in pMS6192A-NDM, pMS6193A-NDM, pMS6194A-NDM, pMS6204A-NDM and pMS6207A-NDM appears to have resulted from the insertion of an IS*Ecp1* module, which was subsequently deleted by an IS*26* composite transposon leaving only a 103 bp fragment of IS*Ecp1*. However, pMS6207A differs from the other four strains with carriage of *aphA6* (amikacin resistance) instead of *aac(3')-lla* and *tmrB* (aminoglycoside and tunicamycin resistance) on the IS*26* composite transposon.

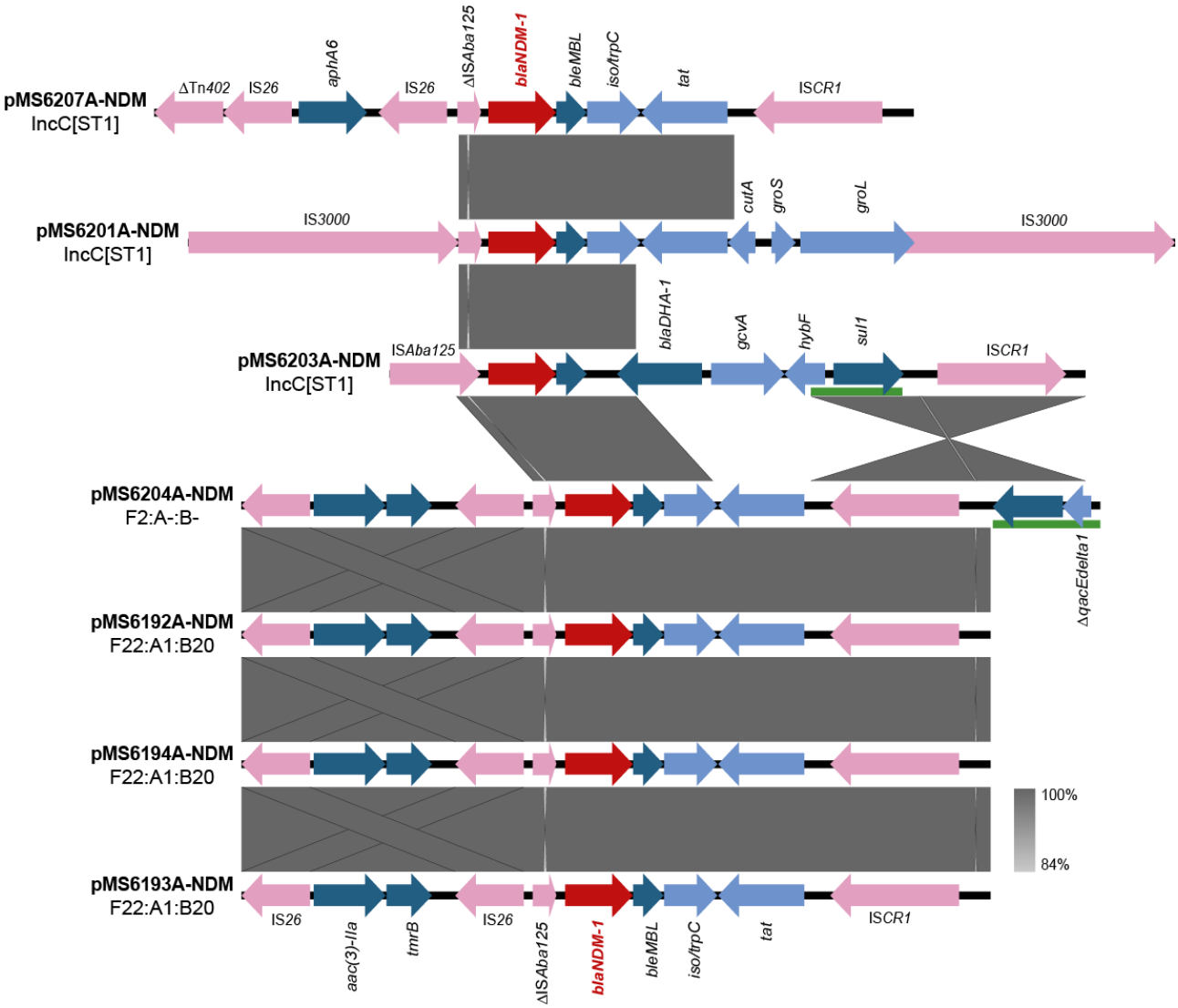
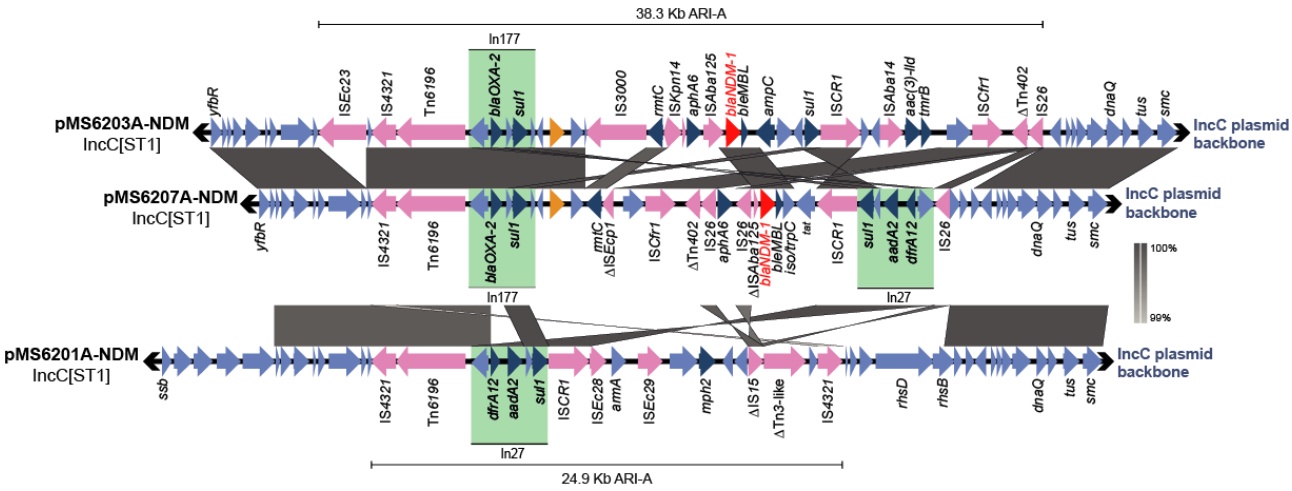


Figure 4. Comparison of the *bla*$_{NDM-1}$ genetic environment of PacBio sequenced Clade 1 ST101 isolates. Grey shading indicates nucleotide identity between sequences according to BLASTn (84-100%). pMS6204A has been reverse complemented for easier visualisation. Δ, truncated gene. IS/Tns: light pink, 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* (Hancock et al. 2016; Marquez-Ortiz et al. 2017; Sugawara et al. 2017). 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

303 sequence typing (pMLST) determined that the IncC plasmids reported here belong to the NDM-
304 associated ST1 group (Hancock et al. 2016), with pMS6203A and pMS6207A belonging to the cgST
305 1.5. Plasmid pMS6201A is missing several genes, however, is most like cgST 1.1 (Supplementary
306 Appendix). These IncC plasmids differ greatly in their structure of the antibiotic resistance island
307 (ARI-A) (Supplementary Appendix and Figure 5). Therefore, despite high conservation in the
308 immediate genetic context of *bla*$_{NDM-1}$ and the IncC plasmid backbone in these ST101 strains, the
309 complete plasmid sequences reveal substantial diversity within the ARI-A, consistent with
310 mobilisation of AMR genes, including *bla*$_{NDM-1}$.
311



312

313 **Figure 5. Major structural features of the Antibiotic Resistance Island on IncC plasmids in the PacBio**
314 **sequenced Clade 1 ST101 strains.** Nucleotide comparisons between the reference plasmid pGUE-NDM
315 (Genbank accession: JQ364967), pMS6192A-NDM, pMS6194A-NDM and pMS6193A-NDM highlighting the
316 shared resistance island and differing plasmid backbones. pGUE-NDM has been reverse complemented for
317 easier visualisation. Grey shading indicates nucleotide identity between sequences according to BLASTn (80-
318 100%). Key genomic regions are indicated: IS elements/transposons: light pink, replication genes: dark pink,
319 maintenance genes: purple, *pemIK* operon: aqua, MTases: orange, AMR genes: teal, other CDSs: blue,
320 Integrons: green rectangles. The *bla*$_{NDM-1}$ gene is labelled in red. Image created using EasyFig.
321

322 ***E. coli* ST101 strains possess distinct types of *bla*$_{NDM-1}$ containing F-type plasmids.** F-type plasmids
323 have previously been associated with the carriage of *bla*$_{NDM-1}$ in *E. coli* (Bonnin et al. 2012; Fiett et
324 al. 2014; Toleman et al. 2015; Wailan et al. 2015). The four *bla*$_{NDM-1}$ encoding F-type plasmids
325 sequenced in this study share a near identical 21.6 kb MDR region that includes the *bla*$_{NDM-1}$ gene.
326 However, two different plasmid backbones were identified; plasmids pMS6192A-NDM, pMS6193A-

NDM, pMS6194A-NDM were almost identical, whereas the backbone of pMS6204A-NDM differed substantially (Supplementary Appendix, Fig S3). BLAST comparisons of the *bla*$_{NDM-1}$ containing plasmids pMS6192A-NDM, pMS6193A-NDM and pMS6194A-NDM (all F36/F22:A1:B20), indicate that they are most similar to the F-type plasmid pIP1206 (>99% nucleotide identity, 83% query coverage). Major differences include deletions in the plasmid backbone, where pMS6192A-NDM, pMS6193A-NDM and pMS6194A-NDM contain an incomplete conjugation region, missing *traK* and *traB*, as well as an 18,516 bp section between *traU* and *traI* present in pIP1206. Resistance genes in pMS6192A-NDM, pMS6193A-NDM and pMS6194A-NDM are clustered in a 21.6 Kb resistance-island, which is different to the resistance island of pIP1206 and is highly similar (99.99% nucleotide identity, 91% query coverage) to the resistance island in the FII plasmid pGUE-NDM (Supplementary Appendix, Fig S4).

The *bla*$_{NDM-1}$ resistance island has inserted between the *pemIK* toxin/antitoxin operon and the *scsCD* operon (copper-sensitivity suppressor). It is composed of a Tn*5403* transposon (containing an intact transposase and the inverted left repeat (ILR) (Siguier et al. 2006)) truncated by an IS*26* (Figure 6). This is followed by 3 resistance genes; *aac(6')Ib-cr* (fluoroquinolone and aminoglycoside resistance), *bla*$_{OXA-1}$ (beta-lactam resistance) and a truncated *catB3* (chloramphenicol resistance). Two additional resistance genes *aac(3)-IIa* (aminoglycoside resistance) and *tmrB* (tunicamycin resistance) are downstream and flanked on either side by IS*26* elements. The *bla*$_{NDM-1}$ module consists of an IS*Aba125* fragment containing the *bla*$_{NDM-1}$ -35 promoter region, *bla*$_{NDM-1}$ (carbapenem resistance), *ble*$_{MBL}$ (bleomycin resistance), *iso*/*trpC* (phosphoribosylanthranilate isomerase), *tat* (twin-arginine translocation pathway signal protein) and an IS*CR1* element. Downstream of the *bla*$_{NDM-1}$ module is the Class I integron In27. This contains *sul1* (sulphonamide resistance) and a truncated *qacEdelta1*, followed by *aadA2* (aminoglycoside resistance) and *dfrA12* (trimethoprim resistance). This MDR region ends with a truncated Tn*As3* transposon (containing

intact transposase and recombinase genes and the inverted right repeat (IRR) (Siguier et al. 2006; He et al. 2016)).

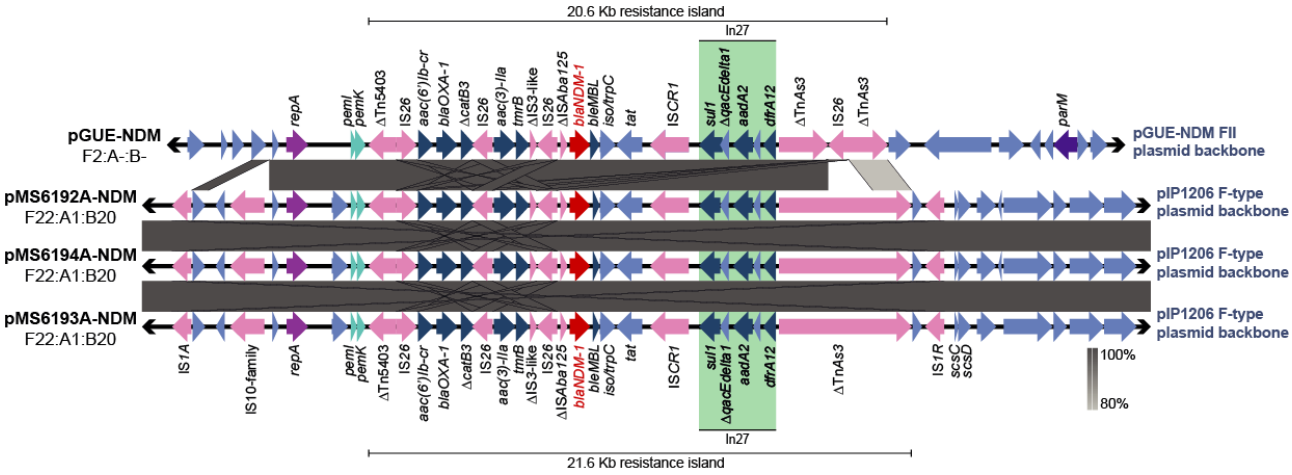


Figure 6. Major structural features of the conserved MDR island on F-type plasmids in the PacBio sequenced Clade 1 ST101 strains. Nucleotide comparisons between the variable resistance regions of the IncC plasmid pMS6203A-NDM, pMS6207A-NDM and pMS6201A-NDM. Grey shading indicates nucleotide identity 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*$_{NDM-1}$ gene is labelled in red. Image created using EasyFig.

***bla*$_{NDM-1}$ is encoded on an FII, pGUE-NDM-like plasmid in one ST101 strain.** MS6204 also contains an F-type pIP1206-like plasmid (pMS6204B; F36/F22:A1:B20). However, the resistance island in pMS6204B is different to that of the pIP1206 reference plasmid and to that of the other *bla*$_{NDM-1}$ containing plasmids described above. This resistance island does not carry the *bla*$_{NDM-1}$ gene, however does encode *bla*$_{CTX-M-15}$ and additional copies of the *aac(6')-Ib-cr*, *bla*$_{OXA-1}$ and *tmrB* resistance genes. In fact, for MS6204, the MDR region encoding *bla*$_{NDM-1}$ is harboured by pMS6204A-NDM (F2:A-:B-), an FII plasmid that is most similar to *E. coli* FII plasmid pGUE-NDM (>99% nucleotide identity, 95% query coverage). Additionally, present within the MDR region of the pMS6204A-NDM plasmid is a 5,095 bp resistance region, which is not present in pGUE-NDM. This region is flanked on either side by IS*26* and encompasses a Tn*3* transposon encoding *bla*$_{TEM-1B}$ (beta-lactam resistance) followed by *rmtB* (rifampicin resistance). This association between Tn*3*, *bla*$_{TEM-1}$ and

375 *rmtB* flanked by IS*26* has been observed in several other *E. coli* and *K. pneumoniae* plasmids
376 (Supplementary Dataset, Table S10), suggesting widespread distribution of this resistance module.
377
378 **Complete plasmid sequences from closely related ST101 strains show carriage of *bla*$_{NDM-1}$ on**
379 **different F-type plasmids.** Intriguingly MS6192, MS6193 and MS6194 also contain a pGUE-NDM-
380 like plasmid, almost identical to pMS6204A-NDM (F2:A-:B-; named pMS6192B and pMS6193B in our
381 complete genomes), however, these plasmids lack the *bla*$_{NDM-1}$ encoding MDR island
382 (Supplementary Appendix, Fig S4). We therefore hypothesise that the entire island encoding
383 *bla*$_{NDM-1}$ was transferred from an FII pGUE-NDM-like plasmid (F2:A-:B-) to an F-type pIP1206-like
384 plasmid (F36/F22:A1:B20) prior to the divergence of MS6192, MS6193 and MS6194 from MS6204
385 (Figure 7). The transposition-like rearrangements that led to this new plasmid are likely complex.
386 According to our model, the most recent common ancestor (MRCA) of MS6204, MS6192, MS6193
387 and MS6194 contained both an FII pGUE-NDM-like plasmid carrying the *bla*$_{NDM-1}$ resistance island
388 and an F-type pIP1206-like plasmid carrying a different resistance island. In the MRCA of MS6192,
389 MS6193 and MS6194, the *bla*$_{NDM-1}$ resistance island was transferred from the pGUE-NDM-like FII
390 plasmid to the pIP1206-like F-type plasmid via a transposition-like event. Although we compared
391 the sequences of plasmids pMS6192B and pMS6193B with pMS6204B at the site of the possible
392 transposition-like event, no duplications consistent with transposition were identified, as this region
393 has been replaced with a truncated Tn*3*. Nonetheless, while the ILR of Tn*5403* and IRR of Tn*As3* are
394 present, along with intact transposase and recombinase genes, it remains unknown whether the
395 MDR island in its current form is capable of active transposition.
396

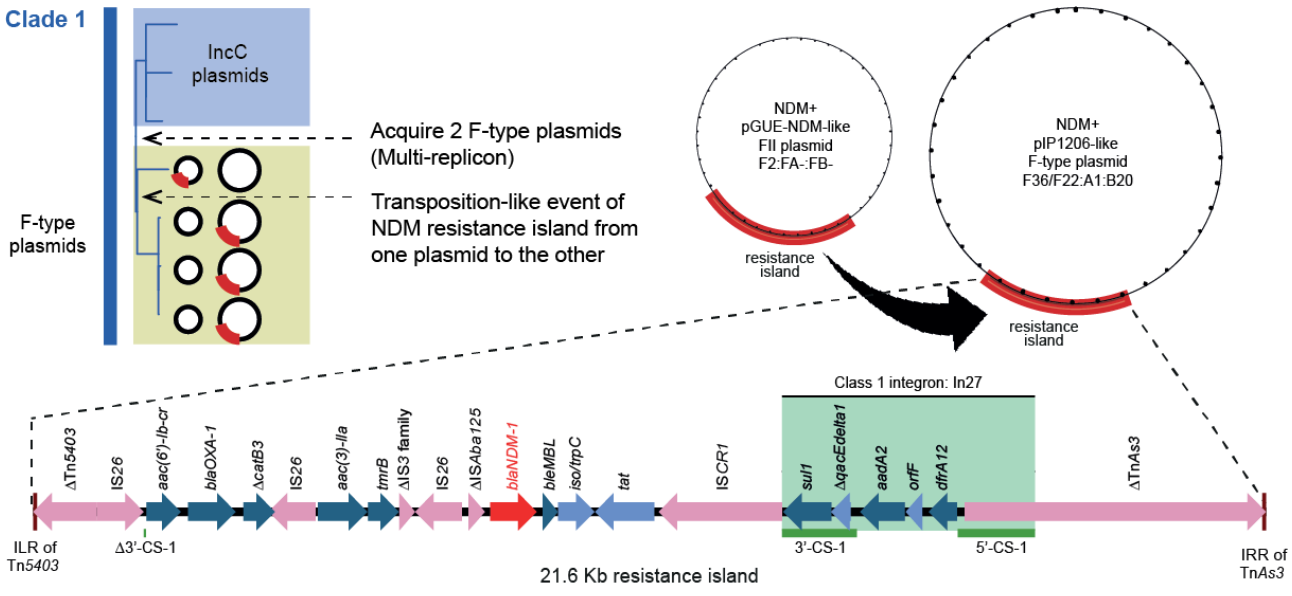


Figure 7. Proposed model of transfer of NDM+ resistance island between F-type plasmids. The acquisition of a pGUE-NDM-like plasmid and a pIP1206-like plasmid (Steps 1 & 2) occurred in the MRCA of MS6204, MS6192, MS6194 and MS6193. A transposition-like event of the entire resistance island between both FII plasmids resulted in the transfer of *bla*$_{NDM-1}$ and surrounding resistance genes. Key regions of the resistance island are indicated: Δ: truncated gene, CDS: light blue, IS/Tns: light pink, AMR genes: teal, Integron: green rectangle. 5' and 3- conserved sequences of class 1 integrons: dark green. The *bla*$_{NDM-1}$ gene is labelled in red. The ILR of Tn*5403* and IRR of Tn*As3* are indicated at the edges of the island.

Discussion

Here we used PacBio SMRT sequencing to provide the most comprehensive genomic snapshot of *bla*$_{NDM-1}$ carriage in the *E. coli* ST101 lineage to date. We show that this lineage is monophyletic within the B1 phylogroup with at least two distinct clades that we have labelled Clade 1 and Clade 2. We show that *bla*$_{NDM-1}$ carriage is confined to Clade 1 and associated with fluoroquinolone resistance mutations and other resistance genes such *bla*$_{CTX-M-15}$. By determining the complete sequence of two closely related ST101 strains and the complete plasmid sequences for five other ST101 strains, we have characterised the genomic context of *bla*$_{NDM-1}$ and other resistance genes. Notably, we revealed a mosaic region of AMR genes including *bla*$_{NDM-1}$ within the ARI-A region of IncC plasmids and transfer of a large MDR resistance region encoding *bla*$_{NDM-1}$ between two different F-type plasmids. These results highlight the power of long-read sequencing in revealing the full complexities of mobile genetic elements.

Identifying the genomic characteristics of *E. coli* lineages is crucial to explaining the evolution and global dissemination of successful clones. Carbapenem resistance resulting from the acquisition of plasmids carrying bla_{NDM-1} has been characterised in several *E. coli* clonal lineages, such as ST405, ST131 and ST101 (Mushtaq et al. 2011). However, despite numerous reports of bla_{NDM-1} or variants within the *E. coli* ST101 lineage (Mushtaq et al. 2011; Peirano et al. 2013; Yoo et al. 2013; Poirel et al. 2014; Toleman et al. 2015; Ranjan et al. 2016), there remains a paucity of whole genome studies. The use of PacBio SMRT sequencing long-read technology in this study enabled the *de novo* assembly of two complete ST101 genomes (MS6192 and MS6193) and five high-quality draft genomes (MS6194, MS6201, MS6203, MS6204 and MS6207). The inability to completely resolve the genomes of the other five isolates was due to low data output achieved using a single SMRT cell on the PacBio RSI platform. Despite this impediment we could completely resolve all plasmid sequences carrying bla_{NDM-1} and define the context of AMR elements within these seven isolates.

Our comparisons of the very closely related ST101 strains MS6192, MS6193 and MS6194 revealed a pattern of genome conservation consistent with a close evolutionary relationship. For example, a conserved core and accessory genome with only 12 core SNP differences between the chromosomes of MS6192 and MS6193, the conservation of both FII plasmid backbones (pGUE-NDM-like and pIP1206-like), as well as the stepwise acquisition of Phi 8 (present in MS6194 and MS6193, but not MS6192) and the Tn*7*-like transposon (present only in MS6193) are consistent with epidemiologically-linked strains. This level of chromosomal similarity is consistent with long-term gut colonisation by a clonal population of Enterobacteriaceae (Conlan et al. 2016). In fact, we also recently reported the existence of an indigenous clonal lineage (with significant plasmid diversity) amongst *E. coli* ST131 isolates collected from urine and fecal samples over a five-year period from a patient with recurrent urinary tract infection (Forde et al. 2019). Analysis of the core genome of the 20 ST101 isolates also indicates a close relationship, however considerable diversity is observed in

444 their MGE complement. These differences are particularly evident between clades, where MGEs
445 such as prophages and genomic islands defined in our reference ST101 genomes MS6192 and
446 MS6193, in general, are not conserved. Exceptions include the GI-*leuX* locus, which is intact in most
447 Clade 1 and 2 strains.
448
449 Nonetheless, our study is limited by the availability of temporal and geographic metadata for all
450 sequenced strains, which restricts our ability to interpret the ST101 phylogeny. We also note that
451 the relatively small number of ST101 complete and draft genomes that were both publicly available
452 and published, constrained our ability to fully characterise the clade structure of this lineage.
453 However, both clades were supported in a larger analysis of 263 available ST101 draft genomes
454 obtained from EnteroBase (30/07/18) (Alikhan et al. 2018) that were of suitable quality for
455 phylogenetic analysis (Supplementary Appendix, Fig S5). Perhaps more importantly, our finding that
456 *bla*$_{\text{NDM-1}}$ (or its variants) was confined to a single sub-lineage within Clade 1 (Clade 1A) remains
457 intact.
458
459 MDR is primarily disseminated within ST101 Clade 1A by plasmids of the groups IncI1, IncC and F-
460 type. The genetic elements harbouring *bla*$_{\text{NDM-1}}$ can spread among Gram-negative bacteria such as
461 *Acinetobacter*, *Pseudomonas* and *Enterobacteriacea*e including *K. pneumoniae* and *E. coli* (Diene
462 and Rolain 2014), where IncC and FII plasmids have previously been reported as carriers of *bla*$_{\text{NDM-1}}$
463 (Bonnin et al. 2012; Hudson et al. 2014). Antimicrobial susceptibility testing showed that the AMR
464 phenotype closely matched the AMR genotype for the seven PacBio sequenced ST101 genomes in
465 this study. However, MS6207 had an intermediate resistance phenotype to Meropenem and
466 Imipenem (Supplementary Dataset 1, Table S9), despite the presence of the *bla*$_{\text{NDM-1}}$ gene, reflecting
467 the fact that AMR prediction from WGS data in Enterobacteriaceae is difficult (Ingle et al. 2018; Su
468 et al. 2019).

In this study, we characterised two pathways for *bla*$_{NDM-1}$ persistence in ST101 Clade 1A. For strains with IncC plasmids, carriage of *bla*$_{NDM-1}$ was associated with recombination of class 1 integrons and transposition of composite transposons and IS elements, resulting in differences in AMR gene repertoire. For the F-type plasmids however, we show carriage of the *bla*$_{NDM-1}$ resistance island in one F-type plasmid, with possible transfer to the other. Transposition events mediated by both IS and transposons enable the mobility of entire resistance islands, particularly in the *Enterobacteriaceae* (Toleman and Walsh 2011). For example, a recent study showed that transposition mediated by IS*26*, Tn*5403* and Tn*3*-family transposons such as those found flanking this *bla*$_{NDM-1}$ resistance island facilitated plasmid rearrangements in carbapenem resistant *K. pneumoniae* (He et al. 2016).

Fluoroquinolone resistance in Clade 1 strains is attributed to vertically transmitted point mutations within the QRDR of *gyrA* (S83L, D87N and D87Y alleles) and *parC* (S80I allele). This is similar to the ST131 lineage where acquisition of QRDR alleles imparting fluoroquinolone resistance occurred in the MRCA prior to the divergence of Clade C (Ben Zakour et al. 2016). With the presence of chromosomally acquired fluoroquinolone resistance and the acquisition of plasmid encoded resistance genes to more than nine different antimicrobial classes, these extensively MDR Clade 1 strains leave few choices for treatment. Older drugs such as colistin are now being used as last-resort treatments for carbapenem-resistant infections (Yamamoto and Pop-Vicas 2014). In 2016, however an ST101 strain (ICBEC72H, included in this study) containing the *mcr-1* colistin resistance gene on an IncX plasmid backbone was identified (Fernandes et al. 2016). With recent reports of both carbapenem and colistin resistant *E. coli* strains appearing (Zheng et al. 2016; Zhong et al. 2017), we face the very real possibility of pan-drug resistant *E. coli*. Overall, these high-quality *E. coli*

493 ST101 genomes will provide an important reference for further analysis of how mobile genetic
494 elements and antimicrobial resistance influence the evolution of this emerging, MDR *E. coli* lineage.
495
496 **Methods**
497 **Bacterial genome sequencing and assembly of our seven PacBio sequenced strains.** The *bla*$_{NDM-1}$-
498 positive *E. coli* strains used in this study have been described previously and were screened for the
499 presence of *bla*$_{NDM-1}$ by PCR (Kumarasamy et al. 2010; Djoko et al. 2017). Additional details regarding
500 strain selection and DNA isolation methods can be found in the Supplementary Appendix. Six of
501 seven strains were sequenced on the PacBio RSI instrument, as previously described (Forde et al.
502 2014), using a 10 Kb insert library. *E. coli* ST101 strain MS6192 used six SMRT cells, with MS6194,
503 MS6201, MS6203, MS6204, MS6207 using two SMRT cells each. *E. coli* ST101 strain MS6193 was
504 sequenced on the PacBio RSII instrument using three SMRT cells with P4-C2 chemistry. Raw
505 sequencing reads were *de novo* assembled using the Hierarchical Genome Assembly Process (HGAP)
506 (Chin et al. 2013) as implemented in the PacBio SMRT Analysis software suite v2.3.0, with a seed
507 read cut-off length of 6,000 bp and default parameters. For *E. coli* MS6192 and MS6193, contig
508 order and orientation was determined using Contiguity and nucleotide blast of the overlapping
509 sequences at each contig end. The merged assemblies were then checked against the closely related
510 commensal strain *E. coli* SE11 using Mauve v2.3.1 (Darling et al. 2010). Additionally, two small,
511 spurious contigs were generated during the *de novo* assembly of MS6192, consisting of rRNA and
512 tRNA genes that mapped to multiple locations with 100% nucleotide identity to the MS6192
513 chromosome. These were deemed chimeric contigs and were not included in the final assembly.
514 Contiguity (Sullivan et al. 2015) was used to visualise the assembly, with overlapping contigs
515 manually trimmed and circularised. Misassemblies were corrected by aligning the reads using BLASR
516 (Chaisson and Tesler 2012), prior to sequence polishing. Raw sequence reads were then mapped to
517 these consensus contigs using Quiver, implemented in the SMRT Analysis software suite, filtering

errors from the assembly. Furthermore, read pileups across all repetitive (rRNA) regions were manually inspected to ensure that their position was supported by spanning reads.

Draft PacBio sequenced ST101 assemblies were screened for plasmid sequences using all-versus-all nucleotide comparisons to generate inter and intra-contig pairwise alignments. All contigs were then screened for overlapping sequences at the 5' and 3' ends as this is an artefact of the HGAP assembly process (Chin et al. 2013) and contigs with self-similar ends likely represent fully complete, circular plasmids. One end of the contig was trimmed to produce a circular sequence. Additional, putative plasmid contigs were screened by nucleotide BLAST comparison against the NCBI non-redundant database. As an additional quality control step, Illumina HiSeq paired-end reads were mapped against each respective genome using bwa 0.7.17-r1188 (Li and Durbin 2009), with Pilon v1.22 (Walker et al. 2014) used to correct for small indels. Plasmid misassemblies in MS6192 were corrected using alternative assembly methods (Supplementary Appendix).

Genome annotation. The automated annotation of all PacBio sequenced ST101 strains (2 complete, 5 draft assemblies) was done using Prokka v1.11 (Seemann 2014) using a custom *Escherichia* database consisting of protein sequences from the EcoCyc website (http://ecocyc.org/) and annotated NDM+ reference plasmids (Supplementary Dataset, Table S11). For the complete genomes: *E. coli* MS6192 and MS6193, prophages, genomic islands and insertion sequences (IS) were identified using PHAST (Zhou et al. 2011), IslandViewer 3 (Dhillon et al. 2015) and ISSaga (Varani et al. 2011) respectively. Draft annotations for MS6192 and MS6193 were then visualised using Artemis and subjected to manual improvement. Insertion sequence, integrons and antimicrobial resistance cassettes within the complete plasmid sequences were manually curated by nucleotide comparisons against ISSaga, Integrall and the Repository of Antibiotic Resistance Cassettes (Moura et al. 2009; Tsafnat et al. 2011; Varani et al. 2011) databases, respectively.

Additional *E. coli* ST101 genomes included in this study. To further characterise the ST101 lineage, 13 additional ST101 draft genomes were downloaded from Genbank or the SRA. Three genomes (ECO0117, eo1667 and PI7) were only available as Illumina paired-end short reads and were assembled *de-novo* using Spades v3.10.1 with default parameters. Strain names, accessions, sources, sequencing method and available metadata are summarised in Supplementary Dataset, Table S3.

Phylogenetic analysis and recombination filtering. To determine the phylogenetic relationship of ST101 and other *E. coli* lineages, core-genome single nucleotide polymorphism (SNP) trees were constructed. A core-genome SNP alignment of the 20 ST101 strains and an additional 65 representative complete *E. coli* genomes (Supplementary Dataset, Table S12) was produced by parSNP v1.2, implemented in the Harvest suite (Treangen et al. 2014) and aligned against the well-characterised ExPEC complete genome *E. coli* EC958 reference (Totsika et al. 2011; Forde et al. 2014) defining 182,264 core SNPs. To integrate the ST101 draft and complete genomes into our phylogenomic analyses and account for differences between short and long read technologies, error-free 71-bp paired-end Illumina reads were simulated from the draft assemblies and complete genomes and used as input to produce a whole genome SNP alignment using Snippy v4.3.5 (https://github.com/tseemann/snippy.git) against the ST101 complete genome MS6192. A pseudogenome of this ST101 SNP alignment was used as input for Gubbins v2.3.1 (Croucher et al. 2015), which detects recombination in closely related groups of isolates. Gubbins was run using default settings in “raxml mode” generating a phylogenetic tree with a generalized time-reversible model and gamma correction (GTRGAMMA). This generated a recombination-filtered core-SNP alignment of 2,106 SNPs. Maximum likelihood (ML) phylogenetic trees were then constructed from both core-genome SNP-based alignments with RAxML 8.1.15 (Stamatakis 2006) using the

GTRGAMMA model and 1000 bootstrap replicates. The resulting ML phylogenetic trees were visualised using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Comparative genomics of *E. coli* ST101. For all *E. coli* ST101 genomes included in this study, acquired AMR genes were functionally annotated by nucleotide comparisons (a cutoff of 90% identity over >=60% query cover) against ResFinder v2.1 (Zankari et al. 2012). O antigen and H antigen typing were determined using SeroType Finder v1.1 (Joensen et al. 2015) and the EcOH database (Ingle et al. 2016) within SRST2 (Inouye et al. 2014), with a nucleotide ID threshold of 85%. Additionally, sequences were screened for the presence of mutations in the QRDR of *gyrA* and *parC*. For each isolate and the *E. coli* K12 strain MG1655 reference genome (Genbank accession: U00096), the amino acid sequences of GyrA and ParC were aligned in MEGA v6.06 (Tamura et al. 2013), using MUSCLE (Edgar 2004) and default parameters. All isolates also underwent *fimH in silico* typing using FimTyper v1.0 (https://cge.cbs.dtu.dk/services/FimTyper/). For plasmid analyses, the plasmid replicons were detected using PlasmidFinder v1.3 with a nucleotide ID threshold of 90% and the plasmid Multi-Locus Sequence Typing (pMLST) (*Enterobacteriaceae*) was determined using pMLST v1.4 (Carattoli et al. 2014). IncC pMLST was determined *in silico* against the four essential IncC genes *repA*, *parA*, *parB* and *053* (Hancock et al. 2016). BRIG v.0.95 (Alikhan et al. 2011), EasyFig v2.2.2 (Sullivan et al. 2011) and Phandango v1.1.0 (Hadfield et al. 2018) were used to visually compare the genomes. Methods for conjugation assays, phenotypic antimicrobial susceptibility testing and plasmid mating assays can be found in the Supplementary Appendix.

Data. Complete genomes of MS6192 and MS6193, draft genome assemblies and PacBio and Illumina sequence read data for MS6192, MS6193, MS6194, MS6201, MS6203, MS6204 and MS6207 are available under the BioProjects: PRJNA580334, PRJNA580336, PRJNA580337, PRJNA580338, PRJNA580339, PRJNA580341 and PRJNA580340 respectively.

Acknowledgements

Author contributions: MMA, BMF, MAS and SAB designed the study. MMA and BMF performed most of the experiments and data analysis. KMP, MDP, SJH and DLP assisted in clinical/wet-lab experiments. LWR and SJH assisted in data analysis. KGC, TMC and WFY performed the PacBio sequencing. Illumina sequencing was performed at the Australian Genome Research Facility (AGRF), The University of Queensland, Australia. TRW provided the strains. MMA. BMF and SAB wrote the manuscript. BMF, MAS, MDP, KGC, TRW, DLP and SAB contributed substantial edits. All authors contributed to the final review and edits.

Funding.

This work was supported by grants from the Australian National Health and Medical Research Council (G1033799) and from the University of Malaya High Impact Research (HIR) Grants (UM-MOHE HIR Grant UM.C/625/1/HIR/MOHE/CHAN/14/1, Grant No. H-50001-A000027 and FP022-2018A). MAS is supported by a NHMRC Senior Research Fellowship (G1106930). SAB is supported by a NHMRC Career Development Fellowship (G1090456). MMA, LWR and SJH were supported by The Australian Government Research Training Program Scholarship.

Conflict of Interest.

None to declare.

References

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics* **12**: 402.

Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. 2018. A genomic overview of the population structure of Salmonella. *PLoS Genet* **14**: e1007261.

Alteri CJ, Mobley HL. 2016. The Versatile Type VI Secretion System. *Microbiology spectrum* **4**.

Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M, Schembri MA, Beatson SA. 2016. Sequential Acquisition of Virulence and Fluoroquinolone Resistance Has Shaped the Evolution of Escherichia coli ST131. *mBio* **7**: e00347-00316.

Bonnin RA, Poirel L, Carattoli A, Nordmann P. 2012. Characterization of an IncFII plasmid encoding NDM-1 from Escherichia coli ST131. *PLoS One* **7**: e34752.

Campos JC, da Silva MJF, dos Santos PRN, Barros EM, Pereira MdO, Seco BMS, Magagnin CM, Leiroz LK, de Oliveira TGM, de Faria-Júnior C et al. 2015. Characterization of Tn3000, a Transposon Responsible for bla(NDM-1) Dissemination among Enterobacteriaceae in Brazil, Nepal, Morocco, and India. *Antimicrobial agents and chemotherapy* **59**: 7387-7395.

Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial agents and chemotherapy* **58**: 3895-3903.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC bioinformatics* **13**: 238.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature methods* **10**: 563-569.

Conlan S, Park M, Deming C, Thomas PJ, Young AC, Coleman H, Sison C, Weingarten RA, Lau AF, Dekker JP et al. 2016. Plasmid Dynamics in KPC-Positive Klebsiella pneumoniae during Long-Term Patient Colonization. *mBio* **7**.

Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song Y et al. 2014. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Science translational medicine* **6**: 254ra126.

Coque TM, Novais A, Carattoli A, Poirel L, Pitout J, Peixe L, Baquero F, Canton R, Nordmann P. 2008. Dissemination of clonally related Escherichia coli strains expressing extended-spectrum beta-lactamase CTX-M-15. *Emerging infectious diseases* **14**: 195-200.

Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research* **43**: e15.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N, McArthur AG, Langille MG et al. 2015. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic acids research* **43**: W104-108.

Diene SM, Rolain JM. 2014. Carbapenemase genes and genetic platforms in Gram-negative bacilli: Enterobacteriaceae, Pseudomonas and Acinetobacter species. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **20**: 831-838.

659 Djoko KY, Achard MES, Phan MD, Lo AW, Miraula M, Prombhul S, Hancock SJ, Peters KM, Sidjabat
660 H, Harris PN et al. 2017. Copper ions and coordination complexes as novel carbapenem
661 adjuvants. *Antimicrobial agents and chemotherapy* doi:10.1128/aac.02280-17.
662 Dortet L, Nordmann P, Poirel L. 2012. Association of the emerging carbapenemase NDM-1 with a
663 bleomycin resistance protein in Enterobacteriaceae and Acinetobacter baumannii.
664 *Antimicrobial agents and chemotherapy* **56**: 1693-1697.
665 Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space
666 complexity. *BMC bioinformatics* **5**: 113.
667 Fernandes MR, McCulloch JA, Vianello MA, Moura Q, Perez-Chaparro PJ, Esposito F, Sartori L,
668 Dropa M, Matte MH, Lira DP et al. 2016. First Report of the Globally Disseminated IncX4
669 Plasmid Carrying the mcr-1 Gene in a Colistin-Resistant Escherichia coli Sequence Type 101
670 Isolate from a Human Infection in Brazil. *Antimicrobial agents and chemotherapy* **60**: 6415-
671 6417.
672 Fiett J, Baraniak A, Izdebski R, Sitkiewicz I, Zabicka D, Meler A, Filczak K, Hryniewicz W,
673 Gniadkowski M. 2014. The first NDM metallo-beta-lactamase-producing
674 Enterobacteriaceae isolate in Poland: evolution of IncFII-type plasmids carrying the
675 bla(NDM-1) gene. *Antimicrobial agents and chemotherapy* **58**: 1203-1207.
676 Forde BM, Ben Zakour NL, Stanton-Cook M, Phan MD, Totsika M, Peters KM, Chan KG, Schembri
677 MA, Upton M, Beatson SA. 2014. The complete genome sequence of Escherichia coli
678 EC958: a high quality reference sequence for the globally disseminated multidrug resistant
679 E. coli O25b:H4-ST131 clone. *PLoS One* **9**: e104400.
680 Forde BM, Roberts LW, Phan MD, Peters KM, Fleming BA, Russell CW, Lenherr SM, Myers JB,
681 Barker AP, Fisher MA et al. 2019. Population dynamics of an Escherichia coli ST131 lineage
682 during recurrent urinary tract infection. *Nature communications* **10**: 3643.
683 Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. 2018. Phandango: an
684 interactive viewer for bacterial population genomics. *Bioinformatics (Oxford, England)* **34**:
685 292-293.
686 Hancock SJ, Phan MD, Peters KM, Forde BM, Chong TM, Yin WF, Chan KG, Paterson DL, Walsh TR,
687 Beatson SA et al. 2016. Identification of IncA/C Plasmid Replication and Maintenance
688 Genes and Development of a Plasmid Multi-Locus Sequence-Typing Scheme. *Antimicrobial*
689 *agents and chemotherapy* doi:10.1128/aac.01740-16.
690 He S, Chandler M, Varani AM, Hickman AB, Dekker JP, Dyda F. 2016. Mechanisms of Evolution in
691 High-Consequence Drug Resistance Plasmids. *mBio* **7**.
692 Hopkins KL, Davies RH, Threlfall EJ. 2005. Mechanisms of quinolone resistance in Escherichia coli
693 and Salmonella: recent developments. *International journal of antimicrobial agents* **25**:
694 358-373.
695 Hudson CM, Bent ZW, Meagher RJ, Williams KP. 2014. Resistance determinants and mobile
696 genetic elements of an NDM-1-encoding Klebsiella pneumoniae strain. *PLoS One* **9**:
697 e99209.
698 Ingle DJ, Levine MM, Kotloff KL, Holt KE, Robins-Browne RM. 2018. Dynamics of antimicrobial
699 resistance in intestinal Escherichia coli from children in community settings in South Asia
700 and sub-Saharan Africa. *Nature microbiology* **3**: 1063-1073.
701 Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M, Stinear T, Levine MM, Robins-Browne RM,
702 Holt KE. 2016. In silico serotyping of E. coli from short read data identifies limited novel O-
703 loci but extensive diversity of O: H serotype combinations within and between pathogenic
704 lineages. *Microbial Genomics* **2**.
705 Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2:
706 Rapid genomic surveillance for public health and hospital microbiology labs. *Genome*
707 *medicine* **6**: 90.

Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *Journal of clinical microbiology* **53**: 2410-2426.

Johnson TJ, Elnekave E, Miller EA, Munoz-Aguayo J, Flores Figueroa C, Johnston B, Nielson DW, Logue CM, Johnson JR. 2019. Phylogenomic Analysis of Extraintestinal Pathogenic Escherichia coli Sequence Type 1193, an Emerging Multidrug-Resistant Clonal Group. *Antimicrobial agents and chemotherapy* **63**.

Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, Chaudhary U, Doumith M, Giske CG, Irfan S et al. 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *The Lancet Infectious diseases* **10**: 597-602.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**: 1754-1760.

Mantilla-Calderon D, Jumat MR, Wang T, Ganesan P, Al-Jassim N, Hong PY. 2016. Isolation and Characterization of NDM-Positive Escherichia coli from Municipal Wastewater in Jeddah, Saudi Arabia. *Antimicrobial agents and chemotherapy* **60**: 5223-5231.

Marquez-Ortiz RA, Haggerty L, Olarte N, Duarte C, Garza-Ramos U, Silva-Sanchez J, Castro BE, Sim EM, Beltran M, Moncada MV et al. 2017. Genomic Epidemiology of NDM-1-Encoding Plasmids in Latin American Clinical Isolates Reveals Insights into the Evolution of Multidrug Resistance. *Genome biology and evolution* **9**: 1725-1741.

Mora A, Blanco M, Lopez C, Mamani R, Blanco JE, Alonso MP, Garcia-Garrote F, Dahbi G, Herrera A, Fernandez A et al. 2011. Emergence of clonal groups O1:HNM-D-ST59, O15:H1-D-ST393, O20:H34/HNM-D-ST354, O25b:H4-B2-ST131 and ONT:H21,42-B1-ST101 among CTX-M-14-producing Escherichia coli clinical isolates in Galicia, northwest Spain. *International journal of antimicrobial agents* **37**: 16-21.

Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. 2009. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics (Oxford, England)* **25**: 1096-1098.

Mushtaq S, Irfan S, Sarma JB, Doumith M, Pike R, Pitout J, Livermore DM, Woodford N. 2011. Phylogenetic diversity of Escherichia coli strains producing NDM-type carbapenemases. *The Journal of antimicrobial chemotherapy* **66**: 2002-2005.

Nordmann P, Poirel L, Walsh TR, Livermore DM. 2011. The emerging NDM carbapenemases. *Trends in microbiology* **19**: 588-595.

Oshima K, Toh H, Ogura Y, Sasamoto H, Morita H, Park SH, Ooka T, Iyoda S, Taylor TD, Hayashi T et al. 2008. Complete genome sequence and comparative analysis of the wild-type commensal Escherichia coli strain SE11 isolated from a healthy adult. *DNA research : an international journal for rapid publication of reports on genes and genomes* **15**: 375-386.

Peirano G, Mulvey GL, Armstrong GD, Pitout JD. 2013. Virulence potential and adherence properties of Escherichia coli that produce CTX-M and NDM beta-lactamases. *Journal of medical microbiology* **62**: 525-530.

Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, Phan MD, Gomes Moriel D, Peters KM, Davies M et al. 2014. Global dissemination of a multidrug resistant Escherichia coli clone. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 5694-5699.

Poirel L, Savov E, Nazli A, Trifonova A, Todorova I, Gergova I, Nordmann P. 2014. Outbreak caused by NDM-1- and RmtB-producing Escherichia coli in Bulgaria. *Antimicrobial agents and chemotherapy* **58**: 2472-2474.

Poolman JT, Wacker M. 2016. Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. *The Journal of infectious diseases* **213**: 6-13.

Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, Nordstrom L, Billig M, Chattopadhyay S, Stegger M et al. 2013. The epidemic of extended-spectrum-beta-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. *mBio* **4**: e00377-00313.

Ranjan A, Shaik S, Mondal A, Nandanwar N, Hussain A, Semmler T, Kumar N, Tiwari SK, Jadhav S, Wieler LH et al. 2016. Molecular Epidemiology and Genome Dynamics of New Delhi Metallo-beta-Lactamase-Producing Extraintestinal Pathogenic Escherichia coli Strains from India. *Antimicrobial agents and chemotherapy* **60**: 6795-6805.

Ren CP, Beatson SA, Parkhill J, Pallen MJ. 2005. The Flag-2 locus, an ancestral gene cluster, is potentially associated with a novel flagellar system from Escherichia coli. *Journal of bacteriology* **187**: 1430-1440.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research* **34**: D32-36.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* **22**: 2688-2690.

Su M, Satola SW, Read TD. 2019. Genome-Based Prediction of Bacterial Antibiotic Resistance. *Journal of clinical microbiology* **57**.

Sugawara Y, Akeda Y, Sakamoto N, Takeuchi D, Motooka D, Nakamura S, Hagiya H, Yamamoto N, Nishi I, Yoshida H et al. 2017. Genetic characterization of blaNDM-harboring plasmids in carbapenem-resistant Escherichia coli from Myanmar. *PLoS One* **12**: e0184720.

Sullivan MJ, L. BZN, M. FB, Stanton-Cook M, Beatson SA. 2015. Contiguity: Contig adjacency graph construction and visualisation. *PeerJ PrePrints* **3**: e1273.

Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)* **27**: 1009-1010.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30**: 2725-2729.

Toleman MA, Bugert JJ, Nizam SA. 2015. Extensively drug-resistant New Delhi metallo-beta-lactamase-encoding bacteria in the environment, Dhaka, Bangladesh, 2012. *Emerging infectious diseases* **21**: 1027-1030.

Toleman MA, Walsh TR. 2011. Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS microbiology reviews* **35**: 912-935.

Totsika M, Beatson SA, Sarkar S, Phan MD, Petty NK, Bachmann N, Szubert M, Sidjabat HE, Paterson DL, Upton M et al. 2011. Insights into a multidrug resistant Escherichia coli pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PLoS One* **6**: e26578.

Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome biology* **15**: 524.

Tsafnat G, Copty J, Partridge SR. 2011. RAC: Repository of Antibiotic resistance Cassettes. *Database : the journal of biological databases and curation* **2011**: bar054.

Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome biology* **12**: R30.

802 Wailan AM, Paterson DL, Kennedy K, Ingram PR, Bursle E, Sidjabat HE. 2015. Genomic
803 Characteristics of NDM-Producing Enterobacteriaceae Isolates in Australia and Their
804 blaNDM Genetic Contexts. *Antimicrobial agents and chemotherapy* **60**: 136-141.
805 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J,
806 Young SK et al. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant
807 Detection and Genome Assembly Improvement. *PLOS ONE* **9**: e112963.
808 Woodford N, Turton JF, Livermore DM. 2011. Multiresistant Gram-negative bacteria: the role of
809 high-risk clones in the dissemination of antibiotic resistance. *FEMS microbiology reviews*
810 **35**: 736-755.
811 Yamamoto M, Pop-Vicas AE. 2014. Treatment for infections with carbapenem-resistant
812 Enterobacteriaceae: what options do we still have? *Critical care (London, England)* **18**: 229.
813 Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, Walsh TR. 2009. Characterization of a
814 new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene
815 carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from
816 India. *Antimicrobial agents and chemotherapy* **53**: 5046-5054.
817 Yoo JS, Kim HM, Koo HS, Yang JW, Yoo JI, Kim HS, Park HK, Lee YS. 2013. Nosocomial transmission
818 of NDM-1-producing Escherichia coli ST101 in a Korean hospital. *The Journal of*
819 *antimicrobial chemotherapy* **68**: 2170-2172.
820 Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen
821 MV. 2012. Identification of acquired antimicrobial resistance genes. *The Journal of*
822 *antimicrobial chemotherapy* **67**: 2640-2644.
823 Zarrilli R, Pournaras S, Giannouli M, Tsakris A. 2013. Global evolution of multidrug-resistant
824 Acinetobacter baumannii clonal lineages. *International journal of antimicrobial agents* **41**:
825 11-19.
826 Zheng B, Dong H, Xu H, Lv J, Zhang J, Jiang X, Du Y, Xiao Y, Li L. 2016. Coexistence of MCR-1 and
827 NDM-1 in Clinical Escherichia coli Isolates. *Clinical infectious diseases : an official*
828 *publication of the Infectious Diseases Society of America* **63**: 1393-1395.
829 Zhong LL, Zhang YF, Doi Y, Huang X, Zhang XF, Zeng KJ, Shen C, Patil S, Xing Y, Zou Y et al. 2017.
830 Coproduction of MCR-1 and NDM-1 by Colistin-Resistant Escherichia coli Isolated from a
831 Healthy Individual. *Antimicrobial agents and chemotherapy* **61**.
832 Zhou M, Guo Z, Duan Q, Hardwidge PR, Zhu G. 2014. Escherichia coli type III secretion system 2: a
833 new kind of T3SS? *Veterinary research* **45**: 32.
834 Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic*
835 *acids research* **39**: W347-352.
836 Zowawi HM, Forde BM, Alfaresi M, Alzarouni A, Farahat Y, Chong TM, Yin WF, Chan KG, Li J,
837 Schembri MA et al. 2015. Stepwise evolution of pandrug-resistance in Klebsiella
838 pneumoniae. *Scientific reports* **5**: 15082.
839