

# SCIENTIFIC REPORTS



OPEN

## Genomic epidemiology of global *Klebsiella pneumoniae* carbapenemase (KPC)-producing *Escherichia coli*

N. Stoesser<sup>1</sup>, A. E. Sheppard<sup>1</sup>, G. Peirano<sup>2,3</sup>, L. W. Anson<sup>1</sup>, L. Pankhurst<sup>1</sup>, R. Sebra<sup>4</sup>, H. T. T. Phan<sup>1,5</sup>, A. Kasarskis<sup>4</sup>, A. J. Mathers<sup>6,7</sup>, T. E. A. Peto<sup>1</sup>, P. Bradford<sup>8</sup>, M. R. Motyl<sup>9</sup>, A. S. Walker<sup>1</sup>, D. W. Crook<sup>1</sup> & J. D. Pitout<sup>2,3,10,11,12</sup>

The dissemination of carbapenem resistance in *Escherichia coli* has major implications for the management of common infections. *bla*<sub>KPC</sub> encoding a transmissible carbapenemase (KPC), has historically largely been associated with *Klebsiella pneumoniae*, a predominant plasmid (pKpQIL), and a specific transposable element (Tn4401, ~10 kb). Here we characterize the genetic features of *bla*<sub>KPC</sub> emergence in global *E. coli*, 2008–2013, using both long- and short-read whole-genome sequencing. Amongst 43/45 successfully sequenced *bla*<sub>KPC</sub>-*E. coli* strains, we identified substantial strain diversity ( $n = 21$  sequence types, 18% of annotated genes in the core genome); substantial plasmid diversity ( $\geq 9$  replicon types); and substantial *bla*<sub>KPC</sub>-associated, mobile genetic element (MGE) diversity (50% not within complete Tn4401 elements). We also found evidence of inter-species, regional and international plasmid spread. In several cases *bla*<sub>KPC</sub> was found on high copy number, small Col-like plasmids, previously associated with horizontal transmission of resistance genes in the absence of antimicrobial selection pressures. *E. coli* is a common human pathogen, but also a commensal in multiple environmental and animal reservoirs, and easily transmissible. The association of *bla*<sub>KPC</sub> with a range of MGEs previously linked to the successful spread of widely endemic resistance mechanisms (e.g. *bla*<sub>TEM</sub>, *bla*<sub>CTX-M</sub>) suggests that it may become similarly prevalent.

Carbapenemases have emerged over the last 15 years as a major antimicrobial resistance threat in Enterobacteriaceae, many species of which are major human pathogens<sup>1</sup>. They are enzymes with broad-spectrum hydrolytic activity targeting beta-lactams, and commonly associated with additional resistance mechanisms producing cross-resistance to other antimicrobial classes<sup>2</sup>. The *Klebsiella pneumoniae* carbapenemase (KPC) enzyme, encoded by alleles of the *bla*<sub>KPC</sub> gene, represents one of the five major carbapenemase families, others being the VIM, IMP and NDM metallo-beta-lactamases, and the OXA-48-like oxacillinases<sup>3</sup>.

KPC, unlike the other major carbapenemases, is an Ambler class A enzyme, with a serine in the active site, which hydrolyses penicillins, cephalosporins, aztreonam and carbapenems<sup>4</sup>. At least 18 variants are known, with nucleotide mutations across 20 positions (13 amino acid substitutions), and one variant with a 6 bp deletion

<sup>1</sup>Modernising Medical Microbiology Consortium, Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford, United Kingdom. <sup>2</sup>Division of Microbiology, Calgary Laboratory Services, Calgary, Alberta, Canada. <sup>3</sup>Department of Pathology and Laboratory Medicine, University of Calgary, Alberta, Canada. <sup>4</sup>Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York, USA. <sup>5</sup>National Institute for Health Research (NIHR) Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, United Kingdom. <sup>6</sup>Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA. <sup>7</sup>Office of Hospital Epidemiology, University of Virginia Health System, Charlottesville, Virginia, USA. <sup>8</sup>AstraZeneca Pharmaceuticals LP, Waltham, Massachusetts, USA. <sup>9</sup>Clinical Microbiology, Merck and Co Inc., Rahway, New Jersey, USA. <sup>10</sup>Department of Microbiology, Immunology and Infectious diseases, University of Calgary, Alberta, Canada. <sup>11</sup>Snyder Institute for Chronic diseases, University of Calgary, Alberta, Canada. <sup>12</sup>Department of Medical Microbiology, University of Pretoria, Pretoria, South Africa. Correspondence and requests for materials should be addressed to N.S. (email: [nicole.stoesser@ndm.ox.ac.uk](mailto:nicole.stoesser@ndm.ox.ac.uk))

(KPC-14, nucleotide positions: 722–727].  $bla_{KPC}$  is typically located within a 10 kb mobile transposon (Tn4401), most often on conjugative plasmids. In publicly available sequence data,  $bla_{KPC}$  is mostly found as a single copy on any individual plasmid although it can exist in duplicate (6/133 [5%] KPC plasmid structures with Tn4401/ $bla_{KPC-2/3}$  duplications available in GenBank, March 2017 [Phan HTT, unpublished data]). It has also been described on multiple plasmids within the same isolate, and/or in multiple copies shared amongst the chromosome and plasmids<sup>5,6</sup>.

The first KPC-producer, a *K. pneumoniae* strain harbouring  $bla_{KPC-2}$ , was identified in 1996 in the eastern USA; since then, KPC-2 and KPC-3 (H272Y [C814T] with respect to KPC-2) have become widespread, and entrenched in endemic hotspots in the USA, Greece, Israel, China and Latin America<sup>7,8</sup>. KPC-3 confers a 4-fold increase in ceftazidime resistance compared with KPC-2<sup>9</sup>. The other variants remain relatively rare in published surveys. The spread of the epidemic *K. pneumoniae* lineage, ST258, is thought to have contributed significantly to global  $bla_{KPC-2}/bla_{KPC-3}$  dissemination<sup>10</sup>, although these genes have now been observed in several species of Enterobacteriaceae<sup>5,6</sup>.

Acquired carbapenem resistance in *Escherichia coli* was considered rare as recently as 2010, although the first cases of KPC-*E. coli* were observed as early as 2004–2005 in Cleveland (n = 1, KPC-2<sup>11</sup>), New York City (n = 2, KPC-2), New Jersey, USA (n = 1, KPC-3)<sup>12</sup>, and Tel Aviv, Israel (n = 4, KPC-2)<sup>13,14</sup>. No apparent epidemiological links between any of these cases were identified. Genotyping was limited at this time, but supported diversity being present in both host *E. coli* and  $bla_{KPC}$  plasmid backgrounds. Since then, direct, plasmid-mediated transfer of  $bla_{KPC}$  into *E. coli* within human hosts has been described<sup>15</sup>, and clusters of KPC-*E. coli* have been identified in several locations, from China to Puerto Rico<sup>16,17</sup>, and in the context of clinical infections<sup>16,17</sup>, asymptomatic colonization<sup>18</sup> and the environment<sup>19</sup>.

More recently there has been concern around  $bla_{KPC}$  in *E. coli* sequence type (ST) 131, a globally disseminated, clinically successful strain<sup>20–22</sup>. Notably, the H30R/C1 (fluoroquinolone-resistant) and H30Rx/C2 (fluoroquinolone and extended-spectrum cephalosporin-resistant) ST131 sub-lineages have previously emerged in association with particular drug resistance mechanisms, including the extended-spectrum beta-lactamase (ESBL) gene,  $bla_{CTX-M-15}$  (clade C2)<sup>23,24</sup>. Given the high rates of ST131 community and healthcare-associated infections<sup>25</sup>, and its capacity for asymptomatic gastrointestinal colonisation<sup>26</sup>, a stable association of ST131 with  $bla_{KPC}$  could have important consequences for the management of *E. coli* infections<sup>14</sup>.

Despite these concerns, there are limited detailed molecular epidemiological data investigating the genetic structures associated with  $bla_{KPC}$  in *E. coli* and the extent to which these may have been shared amongst Enterobacteriaceae. Here we used short-read (Illumina) and long-read (PacBio) sequencing to investigate 43  $bla_{KPC}$ -positive *E. coli* isolates obtained consecutively from global surveillance schemes (67 participating countries, 2008–2013), fully resolving the  $bla_{KPC}$ -containing plasmids in 22 cases, and comparing these data with other  $bla_{KPC}$  plasmid sequences.

## Results

### Global $bla_{KPC}$ -*E. coli* strains are diverse, even within the most prevalent ST, ST131, with evidence for local transmission.

45 isolates were obtained from 21 cities in 11 countries across four continents (2010–2013; previous laboratory typing results summarized in Table S1). One isolate was  $bla_{KPC}$ -negative on whole-genome sequencing (WGS; *ecol\_252*), potentially having lost  $bla_{KPC}$  during storage or sub-culture in the intervening time period between when the original typing was undertaken and the subsequent DNA extraction and preparation for WGS. For one isolate the WGS data were inconsistent with the lab typing results (*ecol\_451*), likely representing a laboratory mix-up; *ecol\_252* and *ecol\_451* were therefore excluded from subsequent analyses. The other 43 isolates were successfully sequenced (for quality metrics see Table S1). Amongst these 43 isolates, twenty-one different *E. coli* STs were represented (Table 1; predicted *in silico* from WGS), including: ST131 [n = 16], ST410 [n = 4], ST38 [n = 3], ST10, ST69 [n = 2 each] (remaining isolates singleton STs).

Of 16,053 annotated open reading frames (ORFs) identified across all KPC-*E. coli* isolates, only 2,950 (18.4%) were shared in all isolates (“core”), and a further 222 (1.4%) in 95- < 100% of isolates (“soft core”<sup>27</sup>). At the nucleotide level there were 213,352 single nucleotide variants (SNVs) in the core genome, consistent with the previously observed species diversity<sup>28</sup>. Resistance gene profiles also varied markedly between strains, with some harbouring several beta-lactam, aminoglycoside, tetracycline and fluoroquinolone resistance mechanisms (e.g. *ecol\_224*) and others containing  $bla_{KPC}$  only (e.g. *ecol\_584*; Fig. 1). For the 16 KPC-ST131 strains, 4,071/7,910 (51%) ORFs were core, with 6,778 SNVs across the core genome of these isolates, again consistent with previous global studies of ST131 diversity<sup>23,24</sup> (Figure S1). Accessory genomes were highly concordant for some (e.g. *ecol\_356/ecol\_276/ecol\_875*), but not all (e.g. *ecol\_AZ159/ecol\_244*) isolates that were closely related in their core genomes, supporting highly variable evolutionary dynamics between core and accessory genomes (Fig. 1). The geographic distribution of isolates closely related in both the core and accessory genomes supports local (e.g. *ecol\_AZ166, ecol\_AZ167* [ST131, Beijing, China]) transmission of particular KPC-*E. coli* strains. The homology of genetic flanking motifs around the  $bla_{KPC}$  genes in these closely related isolate pairs would also be consistent with this hypothesis, and less consistent with multiple acquisition events of  $bla_{KPC}$  within the same genetic background, especially given the diversity in  $bla_{KPC}$  flanking sequences observed across the rest of the dataset (see below).

### $bla_{KPC}$ genes appear restricted to plasmid contexts in *E. coli* at present, but may exist in multiple copies on single plasmid structures or in high copy number plasmids.

Thirty-four isolates (80%) contained  $bla_{KPC-2}$ , and nine isolates (20%)  $bla_{KPC-3}$ . Chromosomal integration of  $bla_{KPC}$  has been described in other Enterobacteriaceae, *Pseudomonas* and *Acinetobacter* spp. but remains rare<sup>5,29,30</sup>. There was no evidence of chromosomal integration of  $bla_{KPC}$  in either the 18 chromosomal structures reconstructed from long-read sequencing, or based on review of the annotations in the  $bla_{KPC}$ -containing contigs (derived from Illumina *de novo* assemblies) for the other 25 isolates.  $bla_{KPC}$  alleles were not segregated by ST.

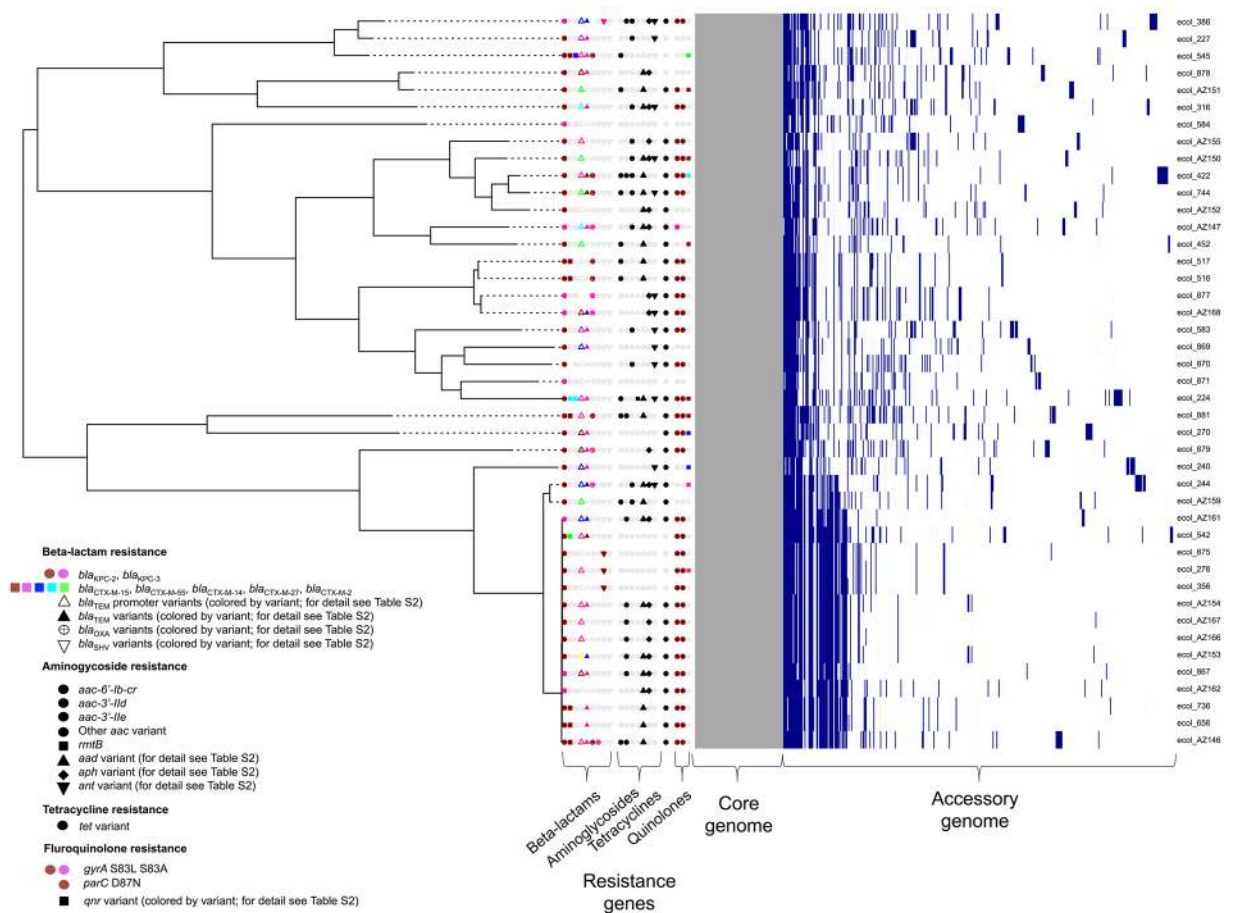
Inc type	Sequence type (number of isolates)																			Total number of isolates [number of <i>bla</i> <sub>KPC</sub> plasmids]	p		
	10 (n=2)	38 (3)	44 (1)	69 (2)	101 (1)	131 (16)	155 (1)	167 (1)	182 (1)	224 (1)	297 (1)	354 (1)	361 (1)	393 (1)	410 (4)	428 (1)	540 (1)	648 (1)	744 (1)			1193 (1)	1431 (1)
A/C2		1	1[1]						1					1[1]								4[2]	0.05
B/O/K/Z						1																1[0]	1
FIA	1	1	1		1	15		1				1		1	2			1		1		26[0]	0.001
FIB	2	1		2	1	11	1	1				1		1	4		1	1	1	1		29[0]	0.14
FII	1	2	1	1	1	12[2] <sup>ab</sup>		1				1		1	4[2]	1	1[1] <sup>c</sup>	1	1	1		30[5] <sup>abc</sup>	0.34
FIC(FII)						1	1															2[0]	0.67
HI1b	1																					1[0]	0.47
HI2+HIA2						1																1[0]	1
I1	1	2			1		1								2		1					8[0]	0.02
I2							1															1[0]	0.67
L/M			1[1]								1											2[1]	0.26
N	1[1]			1	1[1]	3[3]		1	1				1		4[2]			1[1]	1		1	16[8]	0.009
P						1[1]	1								2							4[1]	0.47
Q1				1		2[1]								1					1	1		6[1]	0.31
R						2[1]																2[1]	1
U				1[1]								1[1]										2[2]	0.21
X1																			1			1	0.37
X3						3																3[0]	1
X4						4																3[0]	0.98
Y									1													1[0]	0.37
col	1	1	1	1		12[5] <sup>d</sup>	1[1]					1[1]			4					1	1	24[7] <sup>d</sup>	0.04
p0111											1												0.37

**Table 1.** Plasmid replicon families present by ST, using the PlasmidFinder database<sup>58</sup>. Undmbers in square brackets represent the known subset of *bla*<sub>KPC</sub> plasmids in each cell. Exact test compares presence/absence of each Inc type by ST. The replicon type specifically associated with *bla*<sub>KPC</sub> could not be evaluated in 15 isolates, due to limitations of the assemblies. <sup>a</sup>one multi-replicon plasmid also containing IncFIA. <sup>b</sup>one multi-replicon plasmid also containing IncFIA and IncR. <sup>c</sup>one multi-replicon plasmid also containing IncFIB. <sup>d</sup>one multi-replicon plasmid also containing *repA*.

Estimates of *bla*<sub>KPC</sub> copy number per bacterial chromosome varied between <1 (ecol\_879, ecol\_881) and 55 (ecol\_AZ152). In nine cases this estimate was ≥10 copies of *bla*<sub>KPC</sub> per bacterial chromosome (ecol\_276, ecol\_356, ecol\_867, ecol\_869, ecol\_870, ecol\_875, ecol\_AZ150, ecol\_AZ152, ecol\_AZ159, Table S2). Six of these isolates contained *bla*<sub>KPC</sub> in a col-like plasmid context, in two cases the plasmid rep type was unknown, and in one case it was an IncN replicon. Plasmid copy number is associated with higher levels of antibiotic resistance if the relevant gene is located on a high-copy unit. Interestingly, high copy number plasmids are postulated to have higher chances of fixing in descendant cells, as they distribute more adequately by chance and without the requirement for partitioning systems<sup>31</sup>, and of being transferred in any conjugation event, either directly or indirectly<sup>32–34</sup>.

***bla*<sub>KPC</sub> and non-*bla*<sub>KPC</sub> plasmid populations across global KPC-*E. coli* strains are extremely diverse.** Plasmid Inc typing revealed the presence of a median of four plasmid replicon types per isolate (range: 1–6; IQR: 3–5), representing wide diversity (Table 1). However, IncN, col, IncFIA and IncI1 replicons were disproportionately over-represented in certain STs (p < 0.05; Table 1). Amongst the 18 isolates that underwent PacBio sequencing, we identified 53 closed, non-*bla*<sub>KPC</sub> plasmids, ranging from 1,459 bp to 289,903 bp (Table S1; at least four additional, partially complete plasmid structures were present). Of these non-*bla*<sub>KPC</sub> plasmids, 10 (size: 2,571–150,994 bp) had <70% similarity (defined by percent sequence identity multiplied by proportion of query length demonstrating homology) to other sequences available in GenBank, highlighting that a proportion of the “plasmidome” in KPC-*E. coli* remains incompletely characterized. For the other 43 plasmids, the top match in GenBank was a plasmid from *E. coli* in 35 cases, *K. pneumoniae* in 5 cases, and *Citrobacter freundii*, *Shigella sonnei*, *Salmonella enterica* in 1 case each (Table S3).

Twenty-two *bla*<sub>KPC</sub> plasmid structures were fully resolved (17 from Pacbio data only, four from Illumina data only, 1 from both PacBio and Illumina data), ranging from 14,029 bp to 287,067 bp (median = 55,590 bp; IQR: 23,499–82,765 bp). These *bla*<sub>KPC</sub>-containing plasmids, and six additional cases where *bla*<sub>KPC</sub> was identified on a replicon-containing contig, were highly diverse based on Inc typing (Table S1). IncN was the most common type (n = 8/28 type-able *bla*<sub>KPC</sub> structures; 29%), followed by small, col-like plasmids (n = 6/28 [col-like plasmids with single replicons only]; 21%). Other less common types were: A/C2, FII(k), U (all n = 2); and L/M, P, Q1 and R



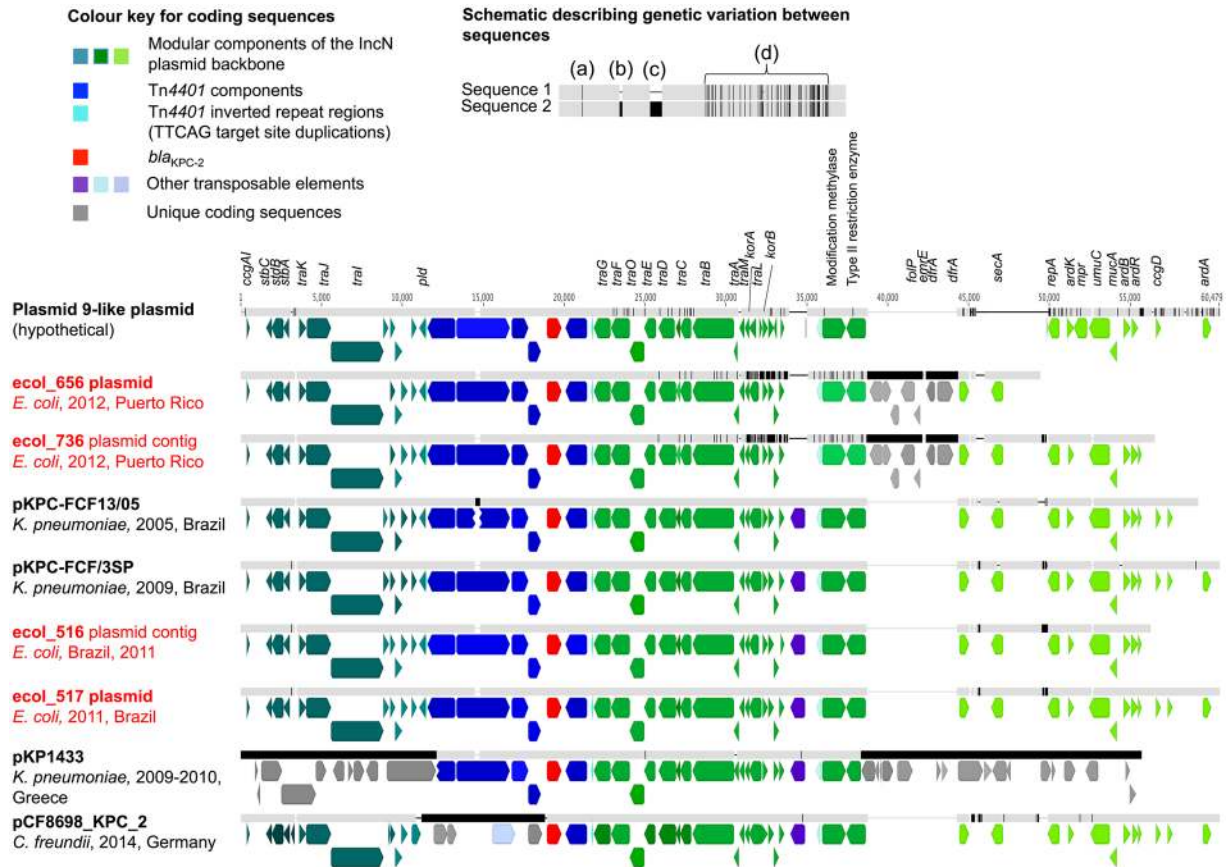
**Figure 1.** Phylogeny of KPC-*Escherichia coli* identified from global carbapenem resistance surveillance schemes, 2008–2013. Panels to the right of the phylogeny represent common resistance gene mechanisms (full details of resistance gene typing in Table S2), core and accessory genome components. For the accessory genome panel, blue represents annotated regions that are present, and white those that are absent.

(all  $n = 1$ ). Four (14%) *bla*<sub>KPC</sub> plasmids were multi-replicon constructs, namely: col/*repA*, FIB/FII, FIA/FII, and FIA/FII/R.

**Common IncN plasmid backbones have dispersed globally within *E. coli*.** From GenBank, we selected all unique, fully sequenced IncN-*bla*<sub>KPC</sub> plasmid sequences (Table S4) for comparison, dating from as early as 2005, around the time of the earliest reports of KPC-producing *E. coli*. The plasmid backbones and flanking sequences surrounding *bla*<sub>KPC</sub> in these 16 plasmid references and a subset of 12 study sequences (see “Methods”) were consistent with multiple acquisitions of two known IncN-Tn4401-*bla*<sub>KPC</sub> complexes in divergent *E. coli* STs: firstly, within a Plasmid-9 (FJ223607, 2005, USA)-like background, and secondly, within a Tn2/3-like element in a Plasmid-12 (FJ223605, 2005, USA)-like background.

In the first instance, genetic similarities were identified between Plasmid-9, pKPC-FCF/3SP, pKPC-FCF13/05, pCF8698, pKP1433 (representing a hybrid IncN), and *bla*<sub>KPC</sub> plasmids from isolates ecol\_516, ecol\_517, ecol\_656, and ecol\_736 (this study). Plasmid-9 contains duplicate Tn4401b elements in reverse orientation with four different 5 bp flanking sequences in an atypical arrangement within a group II intron<sup>35</sup>. The backbone structures of the other plasmids in this group are consistent with a separate acquisition event of a Tn4401b element between the *pld* and *traG* regions within an ancestral version of the Plasmid-9 structure, with the generation of a flanking TTCAG target site duplication (TSD) (labelled as Plasmid 9-like plasmid (hypothetical), Fig. 2). International spread followed by local evolution both within and across species would account for the differences between plasmids, including: (i) nucleotide level variation (observed in all plasmids); (ii) small insertion/deletion events (observed in all plasmids); (iii) larger insertion/deletion events mediated by transposable elements (e.g. pCF8698\_KPC\_2); and (iv) likely homologous recombination, resulting in clustered variation within a similar plasmid backbone (e.g. ecol\_656/ecol\_736), as well as more distinct rearrangements, including the formation of “hybrid” plasmids (e.g. pKP1433)(Fig. 2).

In Plasmid-12 (FJ223605), Tn4401b has inserted into a hybrid Tn2-Tn3-like element (with associated drug resistance genes including *bla*<sub>TEM-1</sub>, *bla*<sub>OXA-9</sub>, and several aminoglycoside resistance genes), albeit in the absence of target sequence duplication, possibly as the result of an intra-molecular, replicative transposition event generating mismatched target site sequences (L TSS = TATTA; R TSS = GTTCT). This complex is in turn located

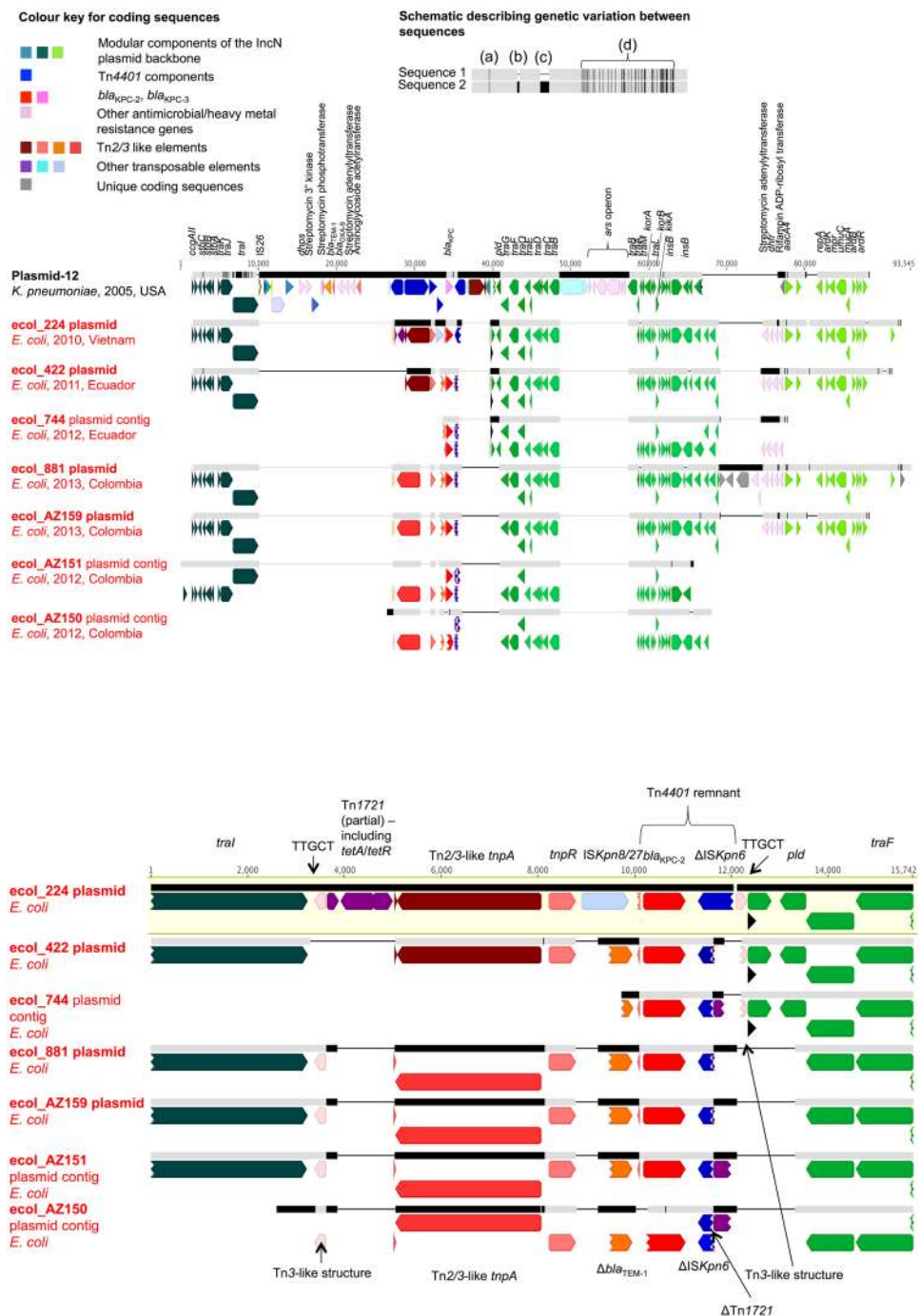


**Figure 2.** Comparison schematic of FJ223607-like (Plasmid 9-like) IncN plasmids (publicly available; this study), and their geographic origin/dates of isolation. Plasmid sequence names in red are those from this study, derived from PacBio data and closed (*ecol\_517*, *ecol\_656*) or incomplete plasmid structures (*ecol\_516*, *ecol\_736*) derived from Illumina data. Aligned bars adjacent to plasmid names represent plasmid sequences: light grey denotes regions with 100% sequence identity; black represents nucleotide diversity between sequences; and thin lines represent indels. Coding sequences are represented by fat arrows below individual sequence bars and are colour-coded as per the colour key. The inset schematic describing genetic variation between sequences depicts examples of evolutionary events identified: (a) single nucleotide level change, (b) small indels ( $\leq 100$  bp), (c) large indels ( $> 100$  bp), (d) recombination events.

between two *IS15DIV* (*IS15Δ*)/*IS26*-like elements flanked by 8 bp inverted repeats, and located between the *traI* (891 bp from 3' end) and *pld* loci (~28 Kb; Fig. 3A). The backbone components of the IncN Plasmid-12 are consistent with those seen in an NIH outbreak<sup>5</sup> and in a rearranged version in a University of Virginia outbreak (CAV1043; 2008)<sup>6</sup>. From this study, plasmids from *ecol\_224*, *ecol\_881*, *ecol\_AZ159*, *ecol\_422*, and scaffolds from *ecol\_AZ151*, *ecol\_744*, *ecol\_AZ150* all share near identical structures to Plasmid-12, with clustered nucleotide level variation present in the *traJ*-*traI* genes, consistent with a homologous recombination event affecting this region, and evidence of sporadic insertion/deletion events (Fig. 3A). However, the *bla*<sub>KPC</sub>-*Tn4401* structures in these isolates are almost entirely degraded by the presence of other mobile genetic elements (MGEs), including *Tn2*/*Tn3*-like elements, *ISKpn8/27* and *Tn1721*. In *ecol\_224*, *bla*<sub>KPC-2</sub> has been inserted into the IncN backbone as part of two repeat, inverted *Tn3*-like structures, flanked by a TTGCT TSD, and closer to *traI* (136 bp from 3' end) than the aforementioned *IS15DIV* (*IS15Δ*)/*IS26*-like complex in Plasmid-12 (Fig. 3B). Although it is not possible to accurately trace the evolutionary history of this genomic region given the available data, the presence of shared signatures of this structure in *ecol\_422*, *ecol\_744*, *ecol\_881*, *ecol\_AZ159*, *ecol\_AZ150* and *ecol\_AZ151* suggest a common acquisition, and multiple subsequent rearrangements mediated by the presence of the large number of MGEs flanking *bla*<sub>KPC-2</sub>.

**Col-like plasmids may represent an important vector of transmission for *bla*<sub>KPC</sub> in *E. coli*.** Small col-like plasmids were the second most common type of plasmid carrying *bla*<sub>KPC</sub> in *E. coli* ( $n = 5$  [plasmids with single replicons only]), but three of these were identical (*bla*<sub>KPC-2</sub>, 16,559 bp), all isolated in Pittsburgh, USA, from ST131 isolates across a two year timeframe (*ecol\_276* [PacBio; 2010], *ecol\_356* [2011], *ecol\_875* [2013]). These three isolates additionally contained FIA, FIB, FII, X3 and X4 replicons, suggesting stable persistence of a clonal strain + plasmids over time, consistent with both SNV/core and accessory genome analyses (Fig. 1, Figure S1).

The other two col-like plasmids effectively represent short stretches of DNA encoding different mobilization genes (*mbeA*/*mbeC*/*mbeD*) harnessed to *Tn4401*/*bla*<sub>KPC</sub> modules. The 5 bp sequences flanking *Tn4401*



**Figure 3.** Comparison schematic of FJ23605-like (Plasmid-12-like) IncN KPC plasmids from this study. Panel 3A. Geographic origin, dates of isolation and overall alignment of plasmid/contig structures. Plasmid sequence names in red are those from this study, derived from PacBio data and closed (ecol\_224, ecol\_422, ecol\_881, ecol\_AZ159) or incomplete plasmid structures (ecol\_744, ecol\_AZ151, ecol\_AZ150) derived from Illumina data. Aligned bars adjacent to plasmid names represent plasmid sequences: light grey denotes regions with 100% sequence homology; black represents nucleotide diversity between sequences; and thin lines represent indels. Coding sequences are represented by fat arrows below individual sequence bars and are colour-coded as per the colour key. The inset schematic describing genetic variation between sequences depicts examples of evolutionary events identified: (a) single nucleotide level change, (b) small indels ( $\leq 100$  bp), (c) large indels ( $> 100$  bp), (d) recombination events. Panel 3B. Close-up of the region between *tral* and *pld* containing *bla<sub>KPC-2</sub>* in study isolates only. Coding sequences are colour-coded as in Fig. 3A; sequence regions referred to in the text are annotated.

were consistent with direct, intermolecular transposition in both cases (ecol\_870: TGTTT-TGTTT; ecol\_867: TGTGA-TGTGA). A *col/repA* co-integrate plasmid was also observed in this dataset (ecol\_AZ161), in which Tn4401b was inserted between colE3 signature sequences and a Tn3 element (Tn4401 TSS: AGATA-GTTCT). The formation of such co-integrate plasmid structures in *E. coli* has also been previously described<sup>36</sup>, including that of a fused col/pKpQIL-like plasmid structure (pKpQIL being historically associated with *bla*<sub>KPC</sub>)<sup>37</sup>.

Col-like plasmids have been associated with KPC-producers in other smaller, regional studies<sup>21, 38</sup>. Of concern, these small vectors have been shown to be responsible for the inter-species diffusion of *qnr* genes mediating fluoroquinolone resistance, even in the absence of any obvious antimicrobial selection pressure<sup>39</sup>. The significant association of col-like plasmids with particular *E. coli* STs (predominantly ST131) in this study could be one explanation for the disproportionate representation of *bla*<sub>KPC</sub> in this lineage.

**Diverse Tn4401 5 bp target site sequences (TSSs) support high transposon mobility.** Complete Tn4401 isoforms flanking *bla*<sub>KPC-2</sub> or *bla*<sub>KPC-3</sub> were observed in only 24/43 (56%) isolates, including Tn4401a/a-like (n = 10; one isolate with a contig break upstream of *bla*<sub>KPC</sub>), Tn4401b (n = 12), and Tn4401d (n = 2) variants. Eleven different 5 bp target site sequence (TSS) pairs were identified, of which 7 (64%) were not observed in any comparison plasmid downloaded from GenBank (Table S5). Tn4401a had three different 5 bp TSSs, Tn4401b seven, and Tn4401d one. Most represented TSDs, but in three cases different 5 bp TSSs were flanking Tn4401, consistent with both direct inter- and replicative intra-molecular transposition events.

From the full set of GenBank plasmids and *in vitro* transposition experiments carried out by others, 30 different types of 5 bp TSS pairs have been characterized, seven in the experimental setting only<sup>40</sup>. The downloaded plasmids come from a range of species and time-points (2005–2014), although they may under-represent wider Tn4401 insertion site diversity as a result of sampling biases. Our data however would be consistent with significant Tn4401 mobility within *E. coli* following acquisition of diverse Tn4401 isoforms and/or represent multiple importation events into *E. coli* from other species.

**The traditional association of *bla*<sub>KPC</sub> with Tn4401 has been significantly eroded in KPC plasmids in *E. coli*.** Notably, in the other 19/43 (44%) isolates the Tn4401 structure had been degraded through replacement with MGEs, only some of which have been previously described<sup>41, 42</sup>. Two isolates had novel Tn4401Δb structures (upstream truncations by IS26 [ecol\_270] or IS26-ΔIS5075 [ecol\_584]). A Tn4401e-like structure (255 bp deletion upstream of *bla*<sub>KPC</sub>) was present in three isolates (ecol\_227, ecol\_316, ecol\_583): this was further characterized in one complete PacBio plasmid assembly (ecol\_316) and represented a rearrangement at the site of the L TSS of the ISKpn7 element. In this plasmid, a second, partial Tn4401 element was present without *bla*<sub>KPC</sub>, which would be consistent with an incomplete, replicative, intra-molecular transposition event (GGGAA = L TSS and R TSS on the two Tn4401b elements, in reverse orientation). Other motifs flanking *bla*<sub>KPC</sub> included: hybrid Tn2/Tn3 elements-ISKpn8/27-*bla*<sub>KPC</sub> (n = 1; ecol\_224); IS26-Δ*tnpR*(Tn3)-ISKpn8/27-*bla*<sub>KPC</sub>-ΔTn1721-IS26 (n = 5; ecol\_AZ153-AZ155, ecol\_AZ166, ecol\_AZ167); IS*Apu2-tnpR*(Tn3)-Δ*bla*<sub>TEM</sub>-*bla*<sub>KPC</sub>-*korC-klcA*-ΔTn1721-IS26 (n = 1; ecol\_542); IS26-*tnpR*(Tn3)-Δ*bla*<sub>TEM</sub>-*bla*<sub>KPC</sub>-*korC*-IS26 (n = 1; ecol\_545); hybrid Tn2/Tn3 elements + Δ*bla*<sub>TEM</sub>-*bla*<sub>KPC</sub>-ΔTn1721 (n = 2; ecol\_744, ecol\_422), Tn3 elements-Δ*bla*<sub>TEM</sub>-*bla*<sub>KPC</sub>-ΔTn1721 (n = 4; ecol\_881, ecol\_AZ151, ecol\_AZ159, ecol\_AZ150) and ΔTn3-Δ-ΔIS3000 (Tn3-like) (n = 1; ecol\_AZ152). We were unable to assess the flanking context of *bla*<sub>KPC</sub> in ecol\_452 due to limitations of the assembly.

This apparent diversity in independently acquired MGEs around the *bla*<sub>KPC</sub> gene extends the means by which *bla*<sub>KPC</sub> can be mobilized. Interestingly, as observed previously<sup>43</sup>, all the degraded Tn4401 sequences in this dataset were associated with variable stretches of flanking Tn2/3-like sequences, suggesting that the insertion of Tn4401 into a Tn2/Tn3-like context may have enabled the latter to act as a hotspot for the insertion of other MGEs<sup>6</sup>. A particular finding of note is the association with IS26, which has been linked to the dissemination of several other resistance genes in *E. coli*, including CTX-M ESBLs<sup>24, 44</sup>, is able to increase the expression of closely co-located resistance genes<sup>45</sup>; participates in co-integrate formation and hence plasmid rearrangement<sup>46</sup>; and enhances the occurrence of other IS26-mediated transfer events into plasmids harbouring IS26<sup>46</sup>.

## Discussion

This study of KPC-*E. coli* obtained from two global resistance surveillance schemes has demonstrated the diversity of genetic structures associated with *bla*<sub>KPC</sub> at all genetic levels, including: (i) host bacterial strain; (ii) plasmid types; (iii) associated transposable MGEs, including transposons and insertion sequences; and (iv) *bla*<sub>KPC</sub> alleles. This has previously been observed within institutional, poly-species outbreaks, particularly for non-*E. coli* Enterobacteriaceae<sup>5, 6</sup>, as well as in a more recent study of nine KPC-*E. coli* from the US<sup>47</sup>. We have identified global and regional *bla*<sub>KPC</sub> spread at the strain and plasmid levels, including signatures consistent with inter-species spread of plasmids, over short timeframes. Although the geographic reach of sampling has been more substantial than any other similar study, there are some limitations in the sampling consistency of both surveillance schemes<sup>22</sup> (e.g. isolates from China were only submitted in 2008, 2012 and 2013).

We utilized long-read sequencing on only a subset of isolates, given resource limitations, allowing us to completely resolve chromosomal and plasmid structures in less than half the isolates. Nevertheless, despite this drawback, we have highlighted the extraordinary diversity present. This study, along with other recent analyses utilizing long-read sequencing to resolve antimicrobial resistance plasmids<sup>5, 6</sup>, also demonstrates the difficulty in making evolutionary comparisons for MGEs, given the absence of effective phylogenetic methods/tools to characterize their genetic histories which commonly involve genetic rearrangements, and evolutionary events that are not restricted to single nucleotide mutations.

This study has demonstrated the particular association of *bla*<sub>KPC</sub> in *E. coli* with IncN plasmids, previously associated with the spread of other antimicrobial resistance elements<sup>48</sup>, as well as col-like plasmids, which are

small, potentially highly mobile, and generally high copy-number units. The traditional association of *bla*<sub>KPC</sub> with Tn4401 has apparently been eroded in *E. coli*, with the complete Tn4401 structure absent in 50% of strains investigated. This finding is in contrast to most global descriptions of *K. pneumoniae* where *bla*<sub>KPC</sub> has been stably associated with largely intact Tn4401 isoforms for more than a decade. Instead, other shorter MGEs, such as Tn2/Tn3-like elements and IS26, appear to be commonly involved in *bla*<sub>KPC</sub> dispersal in *E. coli*. These MGEs have been associated with the spread of multiple resistance mechanisms, such as *bla*<sub>TEM</sub> and *bla*<sub>CTX-M</sub>, and will potentially similarly contribute to the dissemination of *bla*<sub>KPC</sub> in *E. coli*. We did not undertake any functional assays investigating the dynamics of *bla*<sub>KPC</sub> transmission in *E. coli* to support this hypothesis, but this would be illuminating work for future study.

The global emergence and spread of *bla*<sub>KPC</sub> in *E. coli* has been driven by multiple mechanisms, including local and international spread of highly genetically related strains, exchange of plasmids with other Enterobacteriaceae and between *E. coli* lineages, transposition events within the species, and a breakdown of the traditional association of *bla*<sub>KPC</sub> with Tn4401. The genetic flexibility observed is impressive, especially given the timeframes and number of KPC-*E. coli* characterized. Tracking the spread of resistance genes given such multi-level genetic variability is complicated, even with a high-resolution typing method such as WGS. The association of *E. coli*, both a common pathogen and commensal in a wide range of environmental/animal reservoirs, with MGEs (col-like plasmids, IS26) that have been shown to facilitate the dissemination of other successful resistance genes even in the absence of antimicrobial selection pressures, may represent a difficult situation to control.

## Methods

**Isolate collection and sampling frames.** Isolates were obtained from two global antimicrobial resistance surveillance schemes (The Merck Study for Monitoring Antimicrobial Resistance Trends [SMART], 2008–2012; AstraZeneca global surveillance study of antimicrobial resistance, 2012–2013; 417 institutions operating in 95 countries), as previously described<sup>22</sup>. Of 55,874 isolates collected, 45 (0.08%) were positive for *bla*<sub>KPC</sub> by PCR ( $n = 7$  from 2010, 10 from 2011, 13 from 2012, 15 from 2013; Table S1). Isolates had been previously characterized using partial, sequenced-based typing methods, including multi-locus sequence typing (MLST; Achtman scheme), *fimH* typing, PCR for beta-lactamases, strain/plasmid PFGE (Table S1)<sup>22</sup>.

**DNA extraction and sequencing.** All isolates were sequenced on the Illumina MiSeq; a subset of 18 were purposively selected for PacBio sequencing, to capture potential diversity across a range of years of isolation, geographic location, standard ST, plasmid size and resistance gene content (based on laboratory typing). DNA for sequencing was extracted from sub-cultures of bacterial stocks (frozen at  $-80^{\circ}\text{C}$ ; cultured overnight on Columbia blood agar at  $37^{\circ}\text{C}$ ) using the Qiagen Genomic tip 100/G extraction kit, as per the manufacturer's instructions (Qiagen, Hilden, Germany; catalogue no: 10243).

DNA libraries for MiSeq sequencing were generated and normalized using 300 base, paired-end Nextera XT DNA library preparation kits (Illumina, San Diego, CA, USA). PacBio sequencing on the subset of strains was performed as previously described<sup>49</sup>; in these cases, the same DNA extract was used for both Illumina and PacBio sequencing approaches.

**Sequence data processing.** *Illumina (short-read data).* Mapping-based approaches: Prior to reference-based mapping to the O150:H5 SE15 *E. coli* reference genome (Genbank accession: NC\_013654), Illumina data were trimmed using cutadapt version 1.5. SE15, which is ST131, was chosen as the reference given the largest number of strains sequenced (and in the dataset) came from this ST. Repetitive regions of the reference were identified using self-self BLASTn analysis with default settings; these regions were then masked prior to mapping and base calling. Properly paired sequence reads were mapped to the reference using Stampy (v1.0.17) (Supplementary methods).

Single-nucleotide variants (SNVs) were determined across all mapped non-repetitive sites using SAMtools (version 0.1.18) mpileup. mpileup was run twice to separate high-quality base calls from low-quality base calls; variant call format (VCF) files of annotated variant sites were created using GATK (v1.4.21). Base calls derived from these two VCF files were filtered to retain only high quality calls (Supplementary methods).

Core variable sites (site called in all sequenced isolates, excluding “N” or “-” calls) derived from mapping to the SE15 reference were “padded” with invariant sites in a proportion consistent with the GC content and length of the reference genome (4.72 Mb, 51% average GC content), to generate a modified alignment of input sequences to generate phylogenies. Phylogenies were reconstructed using RaxML (Version 7.7.6)<sup>50</sup>, with a generalized time reversible model, four gamma categories (allowing for variable rates of mutation between sites), and bootstrapped 100 times.

*De novo* assemblies of Illumina data: *De novo* assemblies of short-read Illumina data for all isolates were generated using the A5-MiSeq pipeline (version 20140604; default settings)<sup>51</sup>, which includes adapter/low-quality region read trimming steps (Trimmomatic), initial contig assembly, crude scaffolding, misassembly correction and final scaffolding. We used the unscaffolded contigs file in subsequent analyses (\*.contigs.fasta).

**PacBio (long-read data).** DNA library preparation for and sequencing on the PacBio RSII were performed in accordance with the manufacturer's instructions, using P5-C3 sequencing enzyme and chemistries respectively, and following a 7–50 kbp fragment selection step (full details in ref. 49). *De novo* assemblies were constructed using HGAP3 (version 2.2.0)<sup>52</sup>, resulting in phased chromosomal and plasmid contigs, which were manually closed by resolving and trimming overlapping repeats at the contig ends. Illumina reads for the respective isolate were then mapped to the resulting closed PacBio assemblies using bwa-MEM (version 0.7.9a-r786, default settings)<sup>53</sup>. Read pileups were visualized in Geneious<sup>54</sup>; mismatches between the sequence derived from mapping and the reference PacBio assemblies were inspected manually to identify the correct structure. Finally, to capture



small plasmids that may have been filtered out due to size-selection of DNA fragments >7 kbp prior to PacBio sequencing, unmapped Illumina reads (extracted using the SAMtools view command, with the -f 12 flag) derived from this process were *de novo* assembled using the A5-MiSeq pipeline 20140604<sup>51</sup>; any assembled contigs were manually closed by assessing for and trimming overlapping repeats (100% match over a length of  $\geq 100$  bp; no match to any other contig in the assembly). The final consensus chromosomal and plasmid sequences derived from these processes were used in analyses and submitted to GenBank.

**Automated and manual annotation of *de novo* assemblies.** All plasmid structures and *de novo* assemblies were annotated using PROKKA<sup>55</sup>, with subsequent manual refinement of annotations for regions of interest using BLASTn<sup>56</sup> and the NCBI bacterial and ISFinder databases<sup>57</sup>. Alignments of sequence structures were visualized and modified in Geneious. *bla*<sub>KPC</sub>-containing contigs identified in the *de novo* assemblies derived from Illumina data were manually inspected for overlapping repeats as above; if these were present, these contigs were considered additional putative KPC plasmids and trimmed and closed as above.

**Core/accessory genome comparisons.** These were undertaken using the pangenome pipeline, ROARY<sup>27</sup>, by inputting the \*.gff files generated from the PROKKA annotation of each of the Illumina *de novo* assemblies (default settings). Comparisons were made separately for all isolates and the ST131 subset. The output gene\_presence\_absence.csv files were processed using the pheatmap function in R. Resistance genes were identified using ResistType, a command-line tool developed in-house that identifies the presence of reference loci in WGS data using both BLASTn against *de novo* assemblies and mapping-based approaches, and estimates copy number by comparing contig coverage at any given locus with median contig coverage [scripts, reference resistance gene database and manual available at: <https://github.com/hangphan/resistType>]. These features were plotted on the maximum likelihood phylogenies using the Ape package in R.

**Comparisons with publicly available KPC plasmid sequences.** All complete KPC RefSeq plasmid sequences available in GenBank in May 2015 were identified using the search terms “plasmid” + “KPC” + “complete sequence”. The resulting list was filtered manually to exclude any additional sequences present that were not complete plasmid sequences. In total 63 plasmid sequences were included (Table S5).

For the IncN plasmid comparisons, we included the following from our dataset: (i) six cases where PacBio sequencing had fully resolved the *bla*<sub>KPC</sub> IncN plasmid (ecol\_224, ecol\_422, ecol\_517, ecol\_656, ecol\_881, ecol\_AZ159); (ii) two cases where the IncN rep and *bla*<sub>KPC</sub> were co-located on the same, incomplete contig (ecol\_516, ecol\_736); and (iii) three cases where *bla*<sub>KPC</sub> was present in isolates containing an IncN rep and on contigs that showed high similarity to the IncN plasmid backbones under scrutiny (ecol\_744, ecol\_AZ150, ecol\_AZ151).

**Availability of Data and Materials.** Sequencing datasets (Illumina raw reads, PacBio assemblies) are available in GenBank/SRA (project accession: PRJNA316786 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=316786>))(Table S1).

## References

- Nordmann, P., Naas, T. & Poirel, L. Global spread of Carbapenemase-producing Enterobacteriaceae. *Emerg Infect Dis* **17**, 1791–1798, doi:10.3201/eid1710.110655 (2011).
- Doi, Y. & Paterson, D. L. Carbapenemase-producing Enterobacteriaceae. *Semin Respir Crit Care Med* **36**, 74–84, doi:10.1055/s-0035-1544208 (2015).
- Nordmann, P. & Poirel, L. The difficult-to-control spread of carbapenemase producers among Enterobacteriaceae worldwide. *Clin Microbiol Infect* **20**, 821–830, doi:10.1111/1469-0691.12719 (2014).
- Palzkill, T. Metallo-beta-lactamase structure and function. *Ann N Y Acad Sci* **1277**, 91–104, doi:10.1111/j.1749-6632.2012.06796.x (2013).
- Conlan, S. *et al.* Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci Transl Med* **6**, 254ra126, doi:10.1126/scitranslmed.3009845 (2014).
- Sheppard, A. E. *et al.* Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene *bla*<sub>KPC</sub>. *Antimicrob Agents Chemother* **60**, 3767–3778, doi:10.1128/AAC.00464-16 (2016).
- Munoz-Price, L. S. *et al.* Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. *Lancet Infect Dis* **13**, 785–796, doi:10.1016/S1473-3099(13)70190-7 (2013).
- Nordmann, P., Cuzon, G. & Naas, T. The real threat of *Klebsiella pneumoniae* carbapenemase-producing bacteria. *Lancet Infect Dis* **9**, 228–236, doi:10.1016/S1473-3099(09)70054-4 (2009).
- Mehta, S. C., Rice, K. & Palzkill, T. Natural Variants of the KPC-2 Carbapenemase have Evolved Increased Catalytic Efficiency for Ceftazidime Hydrolysis at the Cost of Enzyme Stability. *PLoS Pathog* **11**, e1004949, doi:10.1371/journal.ppat.1004949 (2015).
- Pitout, J. D., Nordmann, P. & Poirel, L. Carbapenemase-Producing *Klebsiella pneumoniae*, a Key Pathogen Set for Global Nosocomial Dominance. *Antimicrob Agents Chemother* **59**, 5873–5884, doi:10.1128/AAC.01019-15 (2015).
- Deshpande, L. M., Rhomberg, P. R., Sader, H. S. & Jones, R. N. Emergence of serine carbapenemases (KPC and SME) among clinical strains of Enterobacteriaceae isolated in the United States Medical Centers: report from the MYSTIC Program (1999–2005). *Diagn Microbiol Infect Dis* **56**, 367–372, doi:10.1016/j.diagmicrobio.2006.07.004 (2006).
- Hong, T. *et al.* *Escherichia coli*: development of carbapenem resistance during therapy. *Clin Infect Dis* **40**, e84–86, doi:10.1086/429822 (2005).
- Navon-Venezia, S. *et al.* Plasmid-mediated imipenem-hydrolyzing enzyme KPC-2 among multiple carbapenem-resistant *Escherichia coli* clones in Israel. *Antimicrob Agents Chemother* **50**, 3098–3101, doi:10.1128/AAC.00438-06 (2006).
- Mathers, A. J., Peirano, G. & Pitout, J. D. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. *Clin Microbiol Rev* **28**, 565–591, doi:10.1128/CMR.00116-14 (2015).
- Gona, F. *et al.* *In vivo* multiclonal transfer of *bla*<sub>(KPC-3)</sub> from *Klebsiella pneumoniae* to *Escherichia coli* in surgery patients. *Clin Microbiol Infect* **20**, O633–635, doi:10.1111/1469-0691.12577 (2014).
- Luo, Y. *et al.* Characterization of KPC-2-producing *Escherichia coli*, *Citrobacter freundii*, *Enterobacter cloacae*, *Enterobacter aerogenes*, and *Klebsiella oxytoca* isolates from a Chinese Hospital. *Microb Drug Resist* **20**, 264–269, doi:10.1089/mdr.2013.0150 (2014).

17. Robledo, I. E., Aquino, E. E. & Vazquez, G. J. Detection of the KPC gene in *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Acinetobacter baumannii* during a PCR-based nosocomial surveillance study in Puerto Rico. *Antimicrob Agents Chemother* **55**, 2968–2970, doi:10.1128/AAC.01633-10 (2011).
18. Mavroidi, A. *et al.* Emergence of *Escherichia coli* sequence type 410 (ST410) with KPC-2 beta-lactamase. *Int J Antimicrob Agents* **39**, 247–250, doi:10.1016/j.ijantimicag.2011.11.003 (2012).
19. Xu, G., Jiang, Y., An, W., Wang, H. & Zhang, X. Emergence of KPC-2-producing *Escherichia coli* isolates in an urban river in Harbin, China. *World J Microbiol Biotechnol* **31**, 1443–1450, doi:10.1007/s11274-015-1897-z (2015).
20. Cai, J. C., Zhang, R., Hu, Y. Y., Zhou, H. W. & Chen, G. X. Emergence of *Escherichia coli* sequence type 131 isolates producing KPC-2 carbapenemase in China. *Antimicrob Agents Chemother* **58**, 1146–1152, doi:10.1128/AAC.00912-13 (2014).
21. O'Hara, J. A. *et al.* Molecular epidemiology of KPC-producing *Escherichia coli*: occurrence of ST131-fimH30 subclone harboring pKpQIL-like IncFIIk plasmid. *Antimicrob Agents Chemother* **58**, 4234–4237, doi:10.1128/AAC.02182-13 (2014).
22. Peirano, G. *et al.* Global incidence of carbapenemase-producing *Escherichia coli* ST131. *Emerg Infect Dis* **20**, 1928–1931, doi:10.3201/eid2011.141388 (2014).
23. Petty, N. K. *et al.* Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci USA* **111**, 5694–5699, doi:10.1073/pnas.1322678111 (2014).
24. Stoesser, N. *et al.* Evolutionary History of the Global Emergence of the *Escherichia coli* Epidemic Clone ST131. *MBio* **7**, doi:10.1128/mBio.02162-15 (2016).
25. Banerjee, R. & Johnson, J. R. A new clone sweeps clean: the enigmatic emergence of *Escherichia coli* sequence type 131. *Antimicrob Agents Chemother* **58**, 4997–5004, doi:10.1128/AAC.02824-14 (2014).
26. Zhong, Y. M. *et al.* Emergence and spread of O16-ST131 and O25b-ST131 clones among faecal CTX-M-producing *Escherichia coli* in healthy individuals in Hunan Province, China. *J Antimicrob Chemother* **70**, 2223–2227, doi:10.1093/jac/dkv114 (2015).
27. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693, doi:10.1093/bioinformatics/btv421 (2015).
28. Touchon, M. *et al.* Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5**, e1000344, doi:10.1371/journal.pgen.1000344 (2009).
29. Martinez, T., Vazquez, G. J., Aquino, E. E., Martinez, I. & Robledo, I. E. *ISEcp1*-mediated transposition of *bla<sub>KPC</sub>* into the chromosome of a clinical isolate of *Acinetobacter baumannii* from Puerto Rico. *J Med Microbiol* **63**, 1644–1648, doi:10.1099/jmm.0.080721-0 (2014).
30. Chen, L. *et al.* Genome Sequence of a *Klebsiella pneumoniae* Sequence Type 258 Isolate with Prophage-Encoded *K. pneumoniae* Carbapenemase. *Genome Announc* **3**, doi:10.1128/genomeA.00659-15 (2015).
31. Norman, A., Hansen, L. H. & Sorensen, S. J. Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci* **364**, 2275–2289, doi:10.1098/rstb.2009.0037 (2009).
32. Watve, M. M., Dahanukar, N. & Watve, M. G. Sociobiological control of plasmid copy number in bacteria. *PLoS One* **5**, e9328, doi:10.1371/journal.pone.0009328 (2010).
33. Paulsson, J. Multileveled selection on plasmid replication. *Genetics* **161**, 1373–1384 (2002).
34. San Millan, A. *et al.* Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. *Antimicrob Agents Chemother* **59**, 3335–3341, doi:10.1128/AAC.00235-15 (2015).
35. Gootz, T. D. *et al.* Genetic organization of transposase regions surrounding *bla<sub>KPC</sub>* carbapenemase genes on plasmids from *Klebsiella* strains isolated in a New York City hospital. *Antimicrob Agents Chemother* **53**, 1998–2004, doi:10.1128/AAC.01355-08 (2009).
36. Peterson, B. C., Hashimoto, H. & Rownd, R. H. Cointegrate formation between homologous plasmids in *Escherichia coli*. *J Bacteriol* **151**, 1086–1094 (1982).
37. Villa, L. *et al.* Reversion to susceptibility of a carbapenem-resistant clinical isolate of *Klebsiella pneumoniae* producing KPC-3. *J Antimicrob Chemother* **68**, 2482–2486, doi:10.1093/jac/dkt235 (2013).
38. Partridge, S. R. *et al.* Emergence of *bla<sub>KPC</sub>* carbapenemase genes in Australia. *Int J Antimicrob Agents* **45**, 130–136, doi:10.1016/j.ijantimicag.2014.10.006 (2015).
39. Pallecchi, L. *et al.* Small qnrB-harboring ColE-like plasmids widespread in commensal enterobacteria from a remote Amazonian population not exposed to antibiotics. *J Antimicrob Chemother* **66**, 1176–1178, doi:10.1093/jac/dkr026 (2011).
40. Cuzon, G., Naas, T. & Nordmann, P. Functional characterization of Tn4401, a Tn3-based transposon involved in *bla<sub>KPC</sub>* gene mobilization. *Antimicrob Agents Chemother* **55**, 5370–5373, doi:10.1128/AAC.05202-11 (2011).
41. Chen, L. *et al.* Carbapenemase-producing *Klebsiella pneumoniae*: molecular and genetic decoding. *Trends Microbiol* **22**, 686–696, doi:10.1016/j.tim.2014.09.003 (2014).
42. Shen, P. *et al.* Novel genetic environment of the carbapenem-hydrolyzing beta-lactamase KPC-2 among Enterobacteriaceae in China. *Antimicrob Agents Chemother* **53**, 4333–4338, doi:10.1128/AAC.00260-09 (2009).
43. Liu, L. *et al.* *bla<sub>CTX-M-15/91</sub>* Hybrid Genes May Have Been Generated from *bla<sub>CTX-M-15</sub>* on an Inc12 Plasmid. *Antimicrob Agents Chemother* **59**, 4464–4470, doi:10.1128/AAC.00501-15 (2015).
44. Partridge, S. R., Zong, Z. & Iredell, J. R. Recombination in IS26 and Tn2 in the evolution of multiresistance regions carrying *bla<sub>CTX-M-15</sub>* on conjugative IncF plasmids from *Escherichia coli*. *Antimicrob Agents Chemother* **55**, 4971–4978, doi:10.1128/AAC.00025-11 (2011).
45. Lee, K. Y., Hopkins, J. D. & Syvanen, M. Direct involvement of IS26 in an antibiotic resistance operon. *J Bacteriol* **172**, 3229–3236 (1990).
46. Harmer, C. J., Moran, R. A. & Hall, R. M. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* **5**, e01801–01814, doi:10.1128/mBio.01801-14 (2014).
47. Chavda, K. D., Chen, L., Jacobs, M. R., Bonomo, R. A. & Kreiswirth, B. N. Molecular Diversity and Plasmid Analysis of KPC-Producing *Escherichia coli*. *Antimicrob Agents Chemother* **60**, 4073–4081, doi:10.1128/AAC.00452-16 (2016).
48. Garcia-Fernandez, A. *et al.* Multilocus sequence typing of IncN plasmids. *J Antimicrob Chemother* **66**, 1987–1991, doi:10.1093/jac/dkr225 (2011).
49. Mathers, A. J. *et al.* *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome sequencing. *Antimicrob Agents Chemother* **59**, 1656–1663, doi:10.1128/AAC.04292-14 (2015).
50. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, doi:10.1093/bioinformatics/btu033 (2014).
51. Coil, D., Jospin, G. & Darling, A. E. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* **31**, 587–589, doi:10.1093/bioinformatics/btu661 (2015).
52. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10**, 563–569, doi:10.1038/nmeth.2474 (2013).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, doi:10.1093/bioinformatics/btp324 (2009).
54. Kearse, M. M. *et al.* A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
55. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069, doi:10.1093/bioinformatics/btu153 (2014).

56. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410, doi:[10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
57. Siguiet, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**, D32–36, doi:[10.1093/nar/gkj014](https://doi.org/10.1093/nar/gkj014) (2006).
58. Carattoli, A. *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* **58**, 3895–3903, doi:[10.1128/AAC.02412-14](https://doi.org/10.1128/AAC.02412-14) (2014).

## Acknowledgements

We acknowledge the contributions of the laboratory, healthcare and administrative teams contributing to the SMART and Astra Zeneca global antimicrobial resistance surveillance programs, and the Modernising Medical Microbiology Informatics Group (MMMIG). For this study, the MMMIG consisted of Adam Giess, Carlos Del Ojo Elias, Milind Acharya, Nicholas Sanderson, Trien Do and Vasiliki Kostiou. NS is currently funded through a Public Health England/University of Oxford Clinical Lectureship; the sequencing work was also partly funded through a previous Wellcome Trust Doctoral Research Fellowship (#099423/Z/12/Z). Additional funding support was provided by a research grant from Calgary Laboratory Services (#10006465), and by the Health Innovation Challenge Fund (a parallel funding partnership between the Wellcome Trust [WT098615/Z/12/Z] and the Department of Health [grant HICF-T5-358]). This research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Center (BRC) Program, and the Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford, in partnership with Public Health England (PHE). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

## Author Contributions

N.S. and J.P. conceived of the study. Significant contributions to sample collection, laboratory processing and sequencing were made by G.P., L.W.A., L.P., P.B., M.R.M., N.S. and J.P. Short-read (Illumina) sequencing was performed by L.W.A. and L.P.; long-read (PacBio) sequencing by R.S. and A.K. Sequence data processing and analysis were performed by A.E.S., H.T.T.P. and N.S. N.S. drafted the manuscript, which was reviewed and improved by all authors, including A.S.W., T.E.A.P., D.W.C. and A.J.M.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-06256-2](https://doi.org/10.1038/s41598-017-06256-2)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017