1   **Nested Russian Doll-like Genetic Mobility Drives Rapid Dissemination of the**

2   **Carbapenem Resistance Gene _bla_$_{KPC}$**

3   Running title: Multiple genetic levels of _bla_$_{KPC}$ mobility

4   Anna E. Sheppard[a]*, Nicole Stoesser[a], Daniel J. Wilson[a], Robert Sebra[b], Andrew Kasarskis[b], Luke

5   W. Anson[a], Adam Giess[a]†, Louise J. Pankhurst[a], Alison Vaughan[a], Christopher J. Grim[c], Heather

6   L. Cox[d], Anthony J. Yeh[d]‡, the Modernising Medical Microbiology (MMM) Informatics Group[a],

7   Costi D. Sifri[d,e], A. Sarah Walker[a], Tim E. Peto[a], Derrick W. Crook[a,f], Amy J. Mathers[d,g]*

8   [a] Modernizing Medical Microbiology Consortium, Nuffield Department of Clinical Medicine,

9   John Radcliffe Hospital, Oxford University, Oxford, United Kingdom.

10   [b] Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of Medicine,

11   Mount Sinai, New York, New York, USA.

12   [c] Food and Drug Administration, Laurel, Maryland, USA.

13   [d] Division of Infectious Diseases and International Health, Department of Medicine, University

14   of Virginia Health System, Charlottesville, Virginia, USA.

15   [e] Office of Hospital Epidemiology, University of Virginia Health System, Charlottesville, Virginia,

16   USA.

17   [f] Public Health England, Microbiology Services, London, United Kingdom.

18   [g] Clinical Microbiology, Department of Pathology, University of Virginia Health System,

19   Charlottesville, Virginia, USA.

20   * Correspondence to: anna.sheppard@ndm.ox.ac.uk or ajm5b@virginia.edu.

21   † Present address: Department of Informatics, University of Bergen, Bergen, Norway.

22   ‡ Present address: Pathogen Molecular Genetics Section, Laboratory of Human Bacterial

23   Pathogenesis, National Institute of Allergy and Infectious Diseases, National Institutes of Health,

24   Bethesda, Maryland, USA.

25

26   Abstract: 244 words

27   Main text: 4376 words

28   **Abstract**

29   The recent widespread emergence of carbapenem resistance in *Enterobacteriaceae* is a major

30   public health concern, as carbapenems are a therapy of last resort in this family of common

31   bacterial pathogens. Resistance genes can mobilize via various mechanisms including

32   conjugation and transposition, however the importance of this mobility in short-term evolution,

33   such as within nosocomial outbreaks, is currently unknown. Using a combination of short- and

34   long-read whole genome sequencing of 281 bla_{KPC}-positive *Enterobacteriaceae* isolated from a

35   single hospital over five years, we demonstrate rapid dissemination of this carbapenem

36   resistance gene to multiple species, strains, and plasmids. Mobility of bla_{KPC} occurs at multiple

37   nested genetic levels, with transmission of bla_{KPC} strains between individuals, frequent transfer

38   of bla_{KPC} plasmids between strains/species, and frequent transposition of the bla_{KPC} transposon

39   Tn*4401* between plasmids. We also identify a common insertion site for Tn*4401* within various

40   Tn*2*-like elements, suggesting that homologous recombination between Tn*2*-like elements has

41   enhanced the spread of Tn*4401* between different plasmid vectors. Furthermore, while short-

42   read sequencing has known limitations for plasmid assembly, various studies have attempted to

43   overcome this with the use of reference-based methods. We also demonstrate that as a

44   consequence of the genetic mobility observed herein, plasmid structures can be extremely

45   dynamic, and therefore these reference-based methods, as well as traditional partial typing

46   methods, can produce very misleading conclusions. Overall, our findings demonstrate that non-

47   clonal resistance gene dissemination can be extremely rapid, presenting significant challenges

48   for public health surveillance and achieving effective control of antibiotic resistance.

3

49

Importance

51   Increasing antibiotic resistance is a major threat to human health, as highlighted by the recent

52   emergence of multi-drug resistant "superbugs". Here, we tracked how one important multi-

53   drug resistance gene spread in a single hospital over five years. This revealed high levels of

54   resistance gene mobility to multiple bacterial species, which was facilitated by various different

55   genetic mechanisms. The mobility occurred at multiple nested genetic levels, analogous to a

56   Russian doll set where smaller dolls may be carried along inside larger dolls. Our results

57   challenge traditional views that drug-resistance outbreaks are due to transmission of a single

58   pathogenic strain. Instead, outbreaks can be "gene-based", and we must therefore focus on

59   tracking specific resistance genes and their context rather than only specific bacteria.

60

61   **Introduction**

62   Although antibiotic resistance genes have been identified in ancient bacterial DNA (1), much of

63   the recent, alarming increase in pathogen antimicrobial resistance is attributable to the

64   dissemination of resistance genes via horizontal gene transfer (HGT), in response to selection

65   imposed by widespread antibiotic use in medicine and agriculture (2, 3). Many resistance genes

66   are located on plasmids, which can be transferred between different bacterial strains or

67   species, thus facilitating HGT (4). Furthermore, resistance gene mobility can be enhanced by

68   integration into transposable elements, which are short stretches of DNA (several kilobases)

69   that can autonomously mobilize between different genomic locations (5). However, the

70   importance of HGT in short-term evolution is unclear, as capturing the processes in real-time is

71   challenging, and outbreaks in health care settings are often thought to be dominated by clonal

72   transmission (6-9).

73   Carbapenem resistance in *Enterobacteriaceae* has been recognized as a key threat to modern

74   medicine (10, 11), as carbapenems often represent the therapy of last resort for serious

75   infections (12, 13). One of the most prevalent carbapenem resistance genes is the *Klebsiella*

76   *pneumoniae* carbapenemase (KPC) gene, bla_{KPC}, first identified in 1996 and now endemic in

77   many regions of the world (14). KPC is a beta-lactamase capable of hydrolyzing all beta-lactams,

78   including penicillins, monobactams, cephalosporins and carbapenems (15), leaving few

79   treatment options for infected vulnerable hospitalized patients and resulting in worse

80   treatment outcomes (16).

81   Most reports of bla_{KPC} involve *K. pneumoniae* multi-locus sequence type (ST)258 (9, 17), which

82   has been found globally, indicating that clonal dissemination of this resistant lineage has been

83   an important factor in the spread of bla_{KPC} (9, 17-20). Nevertheless, bla_{KPC} has also been

84   observed in other *K. pneumoniae* lineages, as well as other species of *Enterobacteriaceae*,

85   suggesting that bla_{KPC} HGT has also played a role in resistance dissemination (21-25). As bla_{KPC} is

86   often found on conjugative plasmids, some of which have been identified in multiple strains or

87   species, this provides a likely mechanism for HGT (21, 26, 27). In addition, bla_{KPC} is usually

88   present as part of the 10 kb composite Tn*3*-based mobile transposon Tn*4401*, which has been

89   identified in various different plasmids, implicating Tn*4401* transposition as another mechanism

90   contributing to bla_{KPC} spread (28, 29).

91   While Tn*4401* transposition and plasmid conjugation have been measured in the laboratory

92   (28, 30, 31), the frequencies with which these processes occur within real-world ecosystems is

93   not fully understood. In clinical contexts, it is often assumed that short-term evolution is

94   dominated by clonal propagation, such that transmission chains generally involve a single

95   pathogenic strain. However, if HGT is frequent relative to transmission (e.g. a "plasmid

96   outbreak"), then linked patients may show variation in strain composition. If transposition is

97   also frequent, then both host strain and resistance plasmid may show high variability within a

98   single outbreak. As current surveillance strategies tend to focus on the host strain, it is

99   important to establish the relevance of bla_{KPC} mobility within outbreak settings.

100   Traditional approaches for plasmid investigation, such as PCR-based replicon typing, are limited

101   in resolution. Next-generation sequencing has been successfully applied to molecular

102   epidemiological investigation of a number of pathogens at the host strain level, however the

103   application and limitations of this technology for transmission chains involving HGT are

104   relatively unexplored. Whole genome sequencing using short-read technologies (e.g. Illumina)

105   has become cheap and accessible, but is not ideal for plasmid analysis due to *de novo* assembly

106   limitations, as it is often not possible to accurately reconstruct the genomic context

107   surrounding repeated sequences (21, 32). Long-read sequencing (e.g. PacBio) can largely

108   overcome this, often providing single-contig plasmid assemblies, but it is currently prohibitively

109   expensive for many applications. Several studies have utilised reference-based methods for

110   plasmid assembly or inference of plasmid structures using short-read data (33, 34), however

111   these approaches make the implicit assumption that plasmid structures are relatively stable. It

112   will be important to understand the potential shortcomings of these assumptions in relation to

113   mobile genetic elements, which may frequently be involved in plasmid rearrangements.

114   Understanding when and how to successfully apply short and/or long-read sequencing

115   technologies to molecular epidemiology tracking will be important to the field as the incidence

116   of HGT is increasingly recognized (35).

117   In our institution, bla_{KPC} was first identified in 2007 in a patient simultaneously colonized with

118   bla_{KPC}-positive *K. pneumoniae* and *Klebsiella oxytoca*, harbouring the bla_{KPC} plasmids

119   pKPC_UVA01 and pKPC_UVA02 respectively (36, 37). Since then, we have been prospectively

120   screening all *Enterobacteriaceae* species for bla_{KPC} carriage despite national guidelines which

121   recommend screening of *Klebsiella* species and *Escherichia coli* only (38-41). Here we describe

122   the genetic basis of non-clonal bla_{KPC} emergence in a single hospital setting, using a

123   combination of short- and long-read whole genome sequencing to provide genomic

7

124   characterization of 281 *Enterobacteriaceae* isolates from the first five years of this multi-species

125   *bla*$_{KPC}$ outbreak.

126

127   **Results**

128   There were 204 patients infected/colonized with *bla*$_{KPC}$-positive *Enterobacteriaceae* during the

129   prospective sampling period, based on clinical and surveillance sampling. We performed short-

130   read Illumina sequencing on all 294 available isolates; 13 of these were excluded due to quality

131   issues (see Methods), leaving 281 isolates, from 182/204 (89%) patients, for analysis (Table S1).

132   In all 281 isolates, *bla*$_{KPC}$ was carried within a complete or partial Tn*4401* structure.

133   **bla$_{KPC}$ is found in many different host strains, indicating frequent HGT**

134   There were 13 different species carrying *bla*$_{KPC}$ (Figure 1). The four most prevalent species were

135   *Enterobacter cloacae* (96 isolates from 80 patients), *K. pneumoniae* (94 isolates from 55

136   patients), *Klebsiella oxytoca* (35 isolates from 20 patients), and *Citrobacter freundii* (30 isolates

137   from 25 patients), each of which showed substantial genetic diversity. Across all species, there

138   were a total of 62 distinct strains (>500 chromosomal SNVs; see Methods). Of these, 18 strains

139   were identified in multiple patients, and 44 were only seen in a single patient (Figure 1), with

140   new strains continuing to appear throughout the study period. The very recent emergence of

141   *bla*$_{KPC}$ on an evolutionary timescale (15) implies that each strain likely acquired *bla*$_{KPC}$

142   independently, demonstrating frequent HGT between different strains and species.

8

143   **The bla_{KPC} plasmids pKPC_UVA01 and pKPC_UVA02 are widely dispersed**

144   We hypothesised that the spread of bla_{KPC} could be due to conjugative transfer of the index

145   bla_{KPC} plasmids, pKPC_UVA01 and pKPC_UVA02. Defining plasmid presence as ≥99% sequence

146   identity over ≥80% of the plasmid length, 121 (66%) and 32 (18%) patients had isolates carrying

147   pKPC_UVA01 and pKPC_UVA02, respectively, corresponding to 39 and 5 distinct strains from 10

148   and 4 species, respectively (Figure 1). Although the wide dispersal of these plasmids supports

149   the plasmid-mediated outbreak hypothesis, short-read data is limited in the structural

150   inferences it can provide when repetitive sequences are present, and for many isolates it was

151   not possible to confirm that bla_{KPC} was actually co-located within pKPC_UVA01 or pKPC_UVA02

152   (Figure 1).

153   **bla_{KPC} is found in many different plasmids, indicating frequent Tn*4401* transposition**

154   To further investigate bla_{KPC} plasmid structures, we performed long-read PacBio sequencing on

155   17 isolates that were chosen at random from the 281 available, yielding closed bla_{KPC} structures

156   in all cases. Fifteen isolates had a single bla_{KPC} plasmid and two isolates had two bla_{KPC} plasmids,

157   giving a total of 19 bla_{KPC} plasmids from the 17 isolates (Table 1). One isolate additionally had a

158   chromosomal insertion of Tn*4401*.

159   From the analysis of Illumina data described above, 11/17 of these isolates contained

160   pKPC_UVA01. As expected, the PacBio assemblies revealed a pKPC_UVA01-like plasmid in each

161   of these isolates. However, only five of these pKPC_UVA01-like plasmids actually contained

162   bla_{KPC} (Figure 2). The other six pKPC_UVA01-like plasmids lacked the entire Tn*4401* element,

9

163   which was present on a different plasmid in these isolates. Importantly, this demonstrates that

164   plasmid presence (as defined by Illumina sequencing) is an unreliable indicator of the mobile

165   unit carrying bla_{KPC}, as the "presence" of pKPC_UVA01 was misleading in 55% (6/11) of the

166   randomly selected PacBio isolates.

167   After accounting for multiple variants of the same plasmid backbone (e.g. the pKPC_UVA01-like

168   plasmids described above), the 19 bla_{KPC} plasmids identified through long-read sequencing

169   represented 11 distinct plasmid structures (Table 1, Figure S1). These consisted of five

170   pKPC_UVA01-like plasmids, two pKPC_UVA02-like plasmids, four pKPC_CAV1176-like plasmids,

171   and eight bla_{KPC} plasmids that were each present in only a single PacBio-sequenced isolate.

172   Using Illumina data to assess the presence of each of these 11 distinct bla_{KPC} plasmids across

173   the entire set of isolates as described above revealed varied patterns of plasmid presence

174   (Figure 1). However, in the majority of cases it was not possible to determine from Illumina

175   data whether these plasmids contained bla_{KPC}, so the precise details regarding distribution of

176   bla_{KPC}-containing plasmids across the 281 isolates remains elusive.

177   Taken together, these results demonstrate a great deal of bla_{KPC} plasmid diversity, as 11 distinct

178   bla_{KPC} plasmids were identified through long-read sequencing of 17 isolates. Given that these

179   isolates were randomly chosen, the total number of distinct bla_{KPC} plasmids across the entire

180   set of 281 isolates is likely to be much greater than this. Additional Tn4401 insertion sites were

181   identified from the subset of isolates where flanking sequences could be adequately assembled

182   using short-read data, further supporting this hypothesis (Table S2). Therefore, HGT of the

183   index bla_{KPC} plasmids (pKPC_UVA01 and pKPC_UVA02) only partially explains bla_{KPC} spread, and

184   the large number of distinct bla_{KPC} plasmids indicates high levels of Tn4401 mobility.

185   **Tn4401 is present within a Tn2-like element in many different plasmids**

186   In 7 of the 11 distinct, fully characterised, bla_{KPC} plasmids, Tn4401 was surrounded by a

187   sequence element related to the bla_{TEM-1}-containing transposon Tn2 (Figure 3). In all cases, the

188   insertion site of Tn4401 within the $tnpA$ gene of Tn2 was identical, with approximately 1 kb of

189   flanking sequence on either side of Tn4401 showing 100% identity, but the remainder of these

190   Tn2-like elements showed substantial variation. For example, while the sequence surrounding

191   Tn4401 in pKPC_CAV1176 was identical to the reference Tn2* sequence, the Tn2-like element

192   in pKPC_CAV1043 was truncated. Additionally, pKPC_CAV1344 and pKPC_CAV1596-78

193   contained a Tn2 derivative, Tn1331, which contains the additional resistance genes bla_{OXA-9},

194    $aadA1$, and $aac(6')$-Ib and has been seen as a prior site of insertion for Tn4401 (42).

195   **Tn4401 variation**

196   There were five different structural variants of Tn4401 (Table 2). The majority of isolates,

197   230/281 (82%), had the Tn4401b isoform, with the remaining isolates containing Tn4401a

198   (n=8), a novel Tn4401 isoform with a 188 bp deletion upstream of bla_{KPC} (n=39), or one of two

199   truncated Tn4401 structures (n=4). At the SNV level, there were seven sites that were variable

200   within Tn4401b. Three of these were located within bla_{KPC}, giving rise to three different bla_{KPC}

201   alleles, bla_{KPC-2} (n=179), bla_{KPC-3} (n=44), and bla_{KPC-4} (n=5). All non-Tn4401b isolates contained

202   bla_{KPC-2}. Taking all structural and SNV variation into account, there were a total of 12 different

11

203   Tn*4401* variants. However, most of these were very rare, with seven only found in a single

204   patient.

205   **bla_{KPC} mobility has occurred within the hospital**

206   Based on prior healthcare exposure, bla_{KPC} acquisition source was classified as "imported"

207   (likely acquisition prior to admission at our institution) for 15/182 (8%) patients and "local"

208   (likely acquisition within our institution) for 167/182 (92%) patients (Figure 1; see Methods).

209   Imports were more likely to be infected/colonised with *K. pneumoniae*, particularly ST258

210   (Table S3), consistent with previous reports of this strain being the dominant bla_{KPC} carrier in

211   the US (9, 43). Thus, most host strain variation likely originated within the hospital via bla_{KPC}

212   HGT. In support of this, 15/16 (94%) patients infected/colonised with multiple strains/species

213   had shared Tn*4401* variants within the patient (Table S4), suggesting recent bla_{KPC} HGT.

214   Notably, this included one patient with two different species carrying Tn*4401*-6, which is not

215   found in any other patient.

216   There was also some evidence for recent within-strain Tn*4401* transposition. From the isolates

217   that were randomly chosen for long-read sequencing, 4/17 (24%) had multiple Tn*4401* copies

218   (Table 1). If we assume that this randomly chosen subset is representative, this extrapolates to

219   approximately 66/281 isolates across the whole dataset. However, only 2/281 isolates had

220   multiple Tn*4401* variants (Tn*4401*-11; Table 2), indicating that many isolates likely had multiple

221   copies of the same Tn*4401* variant, consistent with recent Tn*4401* transposition.

222   Taken together, these results indicate that much of the genetic diversity observed is due to

223   recent bla_{KPC} mobility, likely within the hospital ecosystem over the described five year

224   outbreak.

225   **Direct patient-to-patient transmission does not explain bla_{KPC} acquisition**

226   To further investigate bla_{KPC} acquisition source, we combined epidemiological and genetic data

227   to trace possible transmission chains, at two different genetic levels. We considered possible

228   transmission events where the donor and recipient were on the same ward at the same time,

229   and carried the same host strain or Tn*4401* variant. Considering only "local" acquisitions (see

230   above), 48/167 (29%) patients had ward contact with another patient carrying the same bla_{KPC}-

231   positive strain (Figure 4, top panel). A greater proportion, 106/167 (63%), of patients had ward

232   contact with another patient carrying the same Tn*4401* variant. However, as Tn*4401*-1 is very

233   common (66% of patients), these inferred transmissions may be spurious. With patients

234   carrying this common variant excluded, only 15/50 (30%) had ward contact with another

235   patient carrying the same Tn*4401* variant (Figure 4, bottom panel). Therefore, both genetic

236   levels (strain or Tn*4401* variant) demonstrated plausible transmissions for only a minority of

237   patients, indicating that direct patient-to-patient transmission is not the dominant mode of

238   bla_{KPC} acquisition.

239

240

241   **Discussion**

242   Here we have demonstrated high levels of genetic diversity in KPC-producing

243   *Enterobacteriaceae* within a single institution over five years. This diversity occurs at multiple

244   genetic levels, revealing a complex evolutionary history of the bla_{KPC} gene involving many

245   different host strains and plasmids.

246   In 7/11 distinct bla_{KPC} plasmids identified through long-read sequencing, Tn*4401* was located

247   within a Tn*2*-like element. As these Tn*2*-like elements differed substantially from each other

248   (Figure 3), it is unlikely that this arose via transposition of a composite Tn*4401*-Tn*2*-like

249   structure. Instead, it suggests that Tn*4401* has been repeatedly incorporated into pre-existing

250   Tn*2*-like elements, which are known to be widespread, and genetically divergent, in

251   *Enterobacteriaceae* (44, 45). However, the insertion site was identical in all cases, yet Tn*4401*

252   has been reported to have no insertion site specificity (28), suggesting that this was not

253   facilitated by a standard transposition mechanism. Therefore, we suggest that this is most likely

254   mediated by homologous recombination with other Tn*2*-like elements following an initial

255   integration event, as recently suggested for another multi-drug resistance gene, $bla_{CTX-M-15}$ (46).

256   This implies that Tn*4401* mobility may have been enhanced via integration into a second,

257   already widely dispersed, transposon. As the Tn*4401*-Tn*2*-like structure was present in the

258   index case isolate (CAV1016, Aug 2007) we presume that the initial transposition of Tn*4401* into

259   a Tn*2*-like element occurred prior to entry into our hospital system. In support of this, one

260   particular Tn*2*-like element, Tn*1331*, has been previously described to contain Tn*4401* (in

261   exactly the same position within the *tnpA* gene as described here) (21, 42, 47, 48), including

14

262   one report describing a _K. pneumoniae_ isolated in 2005, which predates bla_{KPC} in our institution

263   (42). We are not aware of any previous reports describing Tn_4401_ within a non-Tn_1331_ Tn_2_-like

264   element.

265   The prevalence of Tn_4401_ insertions within Tn_2_-like elements also has important implications

266   with regard to plasmid tracking. We previously published a method for arbitrary PCR to track

267   the flanking regions around the Tn_4401_ element, as well as a PCR method to assay presence of

268   what we had wrongly assumed was a single plasmid, pKPC_UVA01. This PCR assay targeted the

269   immediate Tn_4401_ insertion site within a Tn_2_-like element (49), which we have here

270   demonstrated is present in many different plasmids, highlighting that PCR assays, and indeed

271   any partial typing methods, need to be interpreted with a great deal of caution. We were

272   further mislead by the analysis of short-read whole genome sequencing data, which indicated

273   presence of pKPC_UVA01 in the majority of isolates. Taken together it was tempting to

274   conclude that horizontal transfer of pKPC_UVA01 was responsible for the vast majority of bla_{KPC}

275   carriage in our institution. However, long-read sequencing refuted this, revealing a far more

276   complex picture.

277   More generally, this highlights certain limitations for plasmid reconstruction from short-read

278   data. To illustrate by way of example, there were five isolates where long-read sequencing

279   revealed pKPC_UVA01-like plasmids that were identical to the reference pKPC_UVA01

280   sequence apart from the absence of Tn_4401_ and associated 5 bp target site duplication (Figure

281   2). We presume that in these lineages, bla_{KPC} may have been initially acquired via HGT of

282   pKPC_UVA01, with subsequent homologous recombination transferring Tn_4401_ from

15

283   pKPC_UVA01 to a different plasmid containing a Tn2-like element. In each of these five isolates,

284   multiple Tn2-like elements are present, which have 100% sequence identity over approximately

285   1 kb on either side of the Tn4401 insertion site. As this is longer than the fragment length used

286   for paired-end sequencing, it is not possible to resolve the plasmid context of bla_{KPC} using short-

287   read data. Importantly, any reference-based method for plasmid reconstruction (e.g. in this

288   case using the pKPC_UVA01 reference sequence to infer presence of the plasmid in each

289   isolate) is liable to produce misleading results. More generally, it is exactly the repetitive

290   regions that cannot be resolved using short-read data that could be expected to be involved in

291   plasmid rearrangements, either through homologous recombination as suggested here, or by

292   virtue of the fact that transposable elements are often present in multiple copies. Therefore,

293   having short-read data that is consistent with a known plasmid structure, even within the same

294   outbreak, should not be sufficient to conclude that that structure is present, if the data is also

295   consistent with an alternative structure. As several recent studies have utilised reference-based

296   approaches for plasmid assembly / inference (33, 34), our results indicate that any such

297   methods should be interpreted with extreme caution.

298   Across the bla_{KPC}-positive patients, there was large variation in both host strains and bla_{KPC}

299   plasmids, with Tn4401 being the largest genetic unit that was consistently present. Therefore,

300   surveillance strategies aimed at tracking individual strains or plasmids could be misleading, and

301   it may be more appropriate to focus on Tn4401. However, we found limited variation within the

302   transposon, as Tn4401 sequences from 121/182 (66%) patients were identical to the index case

303   (Table 2). This lack of variation implies that even the highest resolution genetic methods may be

304   insufficient for determining specific transmission routes. Even so, we have demonstrated that

305   only a minority of bla_{KPC} acquisition events can be explained by direct patient-to-patient

306   transmission. Future studies should therefore contemporaneously investigate the possible

307   involvement of unsampled reservoirs (e.g. environmental or silent colonization by additional

308   carriers).

309   There several limitations to this study. Because of the cost and effort involved in long-read

310   sequencing, we were only able to resolve a minority of bla_{KPC}-plasmids. This means that

311   although we have a compelling indicator of the diversity created by mobile genetic elements

312   within a single hospital over a five year period, we are limited in the ability to genetically

313   resolve pathways of bla_{KPC} mobility between host strains and plasmid vectors, even within a

314   single patient. We also speculate about the effect of Tn*4401* insertion into Tn*2*-like elements,

315   but future *in vitro* studies could be used to illuminate the effect of this composite structure on

316   Tn*4401* mobility.

317   In conclusion, our detailed genetic analysis of the evolutionary events occurring in the early

318   stages of antimicrobial resistance gene emergence in a single institution identifies several

319   distinct processes occurring at high frequency (Figure 5). First, the presence of shared bla_{KPC}-

320   containing strains in different patients reflects traditional (clonal) outbreak models. Second,

321   bla_{KPC} mobility between strains/species is facilitated by promiscuous bla_{KPC} plasmids such as

322   pKPC_UVA01. Third, bla_{KPC} transfer between plasmids is likely enhanced by homologous

323   recombination between Tn*2*-like elements, facilitating the movement of Tn*4401* from one

324   plasmid to another. Finally, bla_{KPC} mobility is also enabled by standard Tn*4401* transposition.

325   Rather than a single process dominating, resistance dissemination is driven by a combination of

326   these factors, resulting in a high level of diversity in KPC-producing *Enterobacteriaceae*, at

327   multiple genetic levels. As bla_{KPC} prevalence continues to increase, so will this genetic diversity,

328   inevitably resulting in a wider variety of more pathogenic strains carrying bla_{KPC}.

329   Our results indicate that the current standard practice of only screening specific species for

330   bla_{KPC} carriage is likely to hamper surveillance efforts by grossly underestimating true

331   prevalence. Instead of the traditional view of an outbreak involving a single pathogenic strain,

332   we propose that for KPC-producing *Enterobacteriaceae*, and possibly more generally, we should

333   instead adopt the view of a "gene-based outbreak", with surveillance strategies tracking the

334   resistance gene itself rather than a specific host strain.

Methods

Isolate collection and Illumina sequencing

A subset of *K. pneumoniae* isolates, with corresponding sequence data, have been previously described (37). Isolate collection, *de novo* assembly, mapping and variant calling were performed as previously described (37), however here we used species-specific references for mapping (Table S5). Illumina sequencing was also performed as previously described (37), with some exceptions (see Supplementary Methods). In total, 281 isolates from 182 patients were available for analysis; exclusion criteria for additional isolates is described in Supplementary Methods.

Species classification

Species classification was performed using microbiological and sequenced-based methods (see Supplementary Methods for details).

Phylogenetic analysis and strain classification

There were 52 patients with multiple isolates of the same species. One of these (patient FK) involved two strains of *K. pneumoniae* that were highly divergent from each other (>20,000 chromosomal SNVs), clearly representing a separate bla_{KPC} acquisition by each strain. Excluding this divergent strain pair, the remaining cases had SNV differences ranging from 0 to 60 (median 2 SNVs). As these could plausibly represent clonal evolution within the patient, we conservatively chose to include only a single representative (the earliest isolate) for phylogenetic reconstruction, in order to avoid artificially inflating genetic clusters due to

19

355   repeated patient sampling. Phylogenetic analysis was then performed separately for each

356   species using PhyML (50) (see Supplementary Methods). Chromosomally distinct strains were

357   defined by partitioning each phylogeny with a cutoff of ~500 SNVs (see Supplementary

358   Methods). Based on the molecular clock of *Enterobacteriaceae* (1-20 SNVs/chromosome/year)

359   (6, 37, 51), we can be relatively confident that isolates belonging to distinct strains will not have

360   a shared ancestor within the timeframe of bla_{KPC} dispersal, and the number of distinct strains

361   thus provides a conservative estimate of the number of distinct bla_{KPC} acquisition events.

362   **Long-read PacBio sequencing**

363   For long-read sequencing, 17 isolates were randomly chosen from the entire set of sequenced

364   isolates (i.e. including patient duplicates). Long-read sequencing and initial *de novo* assembly

365   were performed as previously described (37). Refinement of assemblies and closure of

366   plasmid/chromosomal sequences was performed as described in Supplementary Methods.

367   Since the isolates for PacBio sequencing were randomly chosen from the set of all Illumina

368   sequenced isolates, some of them represented within-patient strain duplicates (see previous

369   section on phylogenetic analysis), and were therefore not included in phylogenetic

370   reconstruction. For display purposes (in Figure 1), the bla_{KPC} structure(s) determined from long-

371   read PacBio sequencing for each of these isolates is shown alongside the representative isolate

372   of the same strain from the same patient. In all cases, the representative isolate has the same

373   short-read plasmid profile and Tn*4401* variant as the PacBio sequenced isolate.

374   **Plasmid presence / absence classification**

20

375   The index bla_{KPC} plasmids pKPC_UVA01 and pKPC_UVA02, together with the additional nine

376   distinct bla_{KPC} plasmids identified though long-read PacBio sequencing, were used as references

377   to determine plasmid presence profiles for each isolate based on the Illumina data. Plasmid

378   presence was defined as ≥99% sequence identity over ≥80% of the length of the reference

379   sequence, as determined by BLASTn comparisons between each isolate's *de novo* assembly and

380   the reference plasmid. The high identity cutoff was chosen to reduce false positives from

381   sequences that are only distantly related (and therefore unlikely to have a common ancestor

382   within the timeframe of the outbreak), while the more permissive length cutoff allows for some

383   rearrangement. It should be noted that the method does not take any account of structural

384   continuity.

385   **Analysis of Tn*4401* flanking sequences**

386   Where a plasmid was classified as being present in a particular isolate, it was not always certain

387   to contain Tn*4401*. The plasmid presence classification was further refined as: "containing

388   Tn*4401*" if the isolate's *de novo* assembly supported Tn*4401* being present within the expected

389   sequence context of that plasmid, "not containing Tn*4401*" if the plasmid was assembled

390   without Tn*4401*, or "uncertain" if structure could not be determined from the *de novo*

391   assembly. The identification of novel Tn*4401* insertion sites was also based on the *de novo*

392   assemblies. These methods are described in detail in Supplementary Methods.

393   **Variation in Tn*4401***

394   Tn*4401* isoform classification was performed by comparing each isolate's *de novo* assembly

395   with the previously described isoform b reference sequence from EU176013.1 (29) using

21

396   BLASTn, to identify structural variation. SNV variation was determined by mapping to a

397   reference consisting of pKPC_UVA01 plus a species-specific chromosome as described above,

398   followed by extraction of the Tn*4401* region. Variation is reported for all sites where at least

399   one isolate had a non-reference call, including any ambiguity at that site in other isolates.

400   Ambiguity at non-variable sites is not reported, which may result in an underestimate of true

401   variation. However, any resulting underestimation is likely to be very minor, as the proportion

402   of called sites, excluding deleted regions described above, was >96% for all isolates.

403   **Epidemiological classification**

404   For epidemiologic analysis, patients were assigned a one or two letter code for de-

405   identification. Routine peri-rectal surveillance cultures for silent colonization began in April

406   2009 (38, 40). Patients were classified as "imported" if they did not have any prior admission to

407   University of Virginia Medical Center/Long-term Acute Care Hospital (UVaMC) and either had a

408   *bla*$_{KPC}$-positive *Enterobacteriaceae* isolated within 48 hours of admission, or had a carbapenem-

409   resistant *Enterobacteriaceae* culture before transfer to UVaMC with a subsequent isolate at

410   UVaMC confirmed as *bla*$_{KPC}$ PCR positive. The index case was also classified as imported. For the

411   remaining patients, the source of *bla*$_{KPC}$ acquisition was classified as "local". The 48h cutoff is

412   arbitrary and may result in some misclassification if patients either acquire *bla*$_{KPC}$ within the

413   first 48h of admission, or if *bla*$_{KPC}$ carriage/infection remains undetected for >48h, however this

414   is expected to be minimal (see Supplementary Methods). Charts and patient contacts were

415   reviewed using bed tracing data and the electronic medical record. The study was approved by

416   the University of Virginia Institutional Review Board (protocol # 13558).

417 **Transmission analysis**

418 Possible patient-to-patient transmission events were determined on the basis of having

419 overlapping stays on the same ward, as well as genetically-related bla_{KPC} isolates. The analysis

420 was performed separately for two different levels of genetic relatedness (strain or Tn*4401*

421 variant). This is described in detail in Supplementary Methods.

422

423 **Funding Information**

433

434 **Acknowledgements**

437

Modernising Medical Microbiology (MMM) informatics group

439   Jim Davies, Charles Crichton, Milind Acharya, Carlos del Ojo Elias

440

Additional Information

442   Sequence data has been deposited with the National Centre for Biotechnology Information

443   (NCBI) under BioProject PRJNA246471.

444

445 **References**

446 1.   **D'Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, Froese D, Zazula G, Calmels F,**

447   **Debruyne R, Golding GB, Poinar HN, Wright GD.** 2011. Antibiotic resistance is ancient. Nature

448   **477**:457-461.

449 2.   **Datta N, Hughes VM.** 1983. Plasmids of the same Inc groups in Enterobacteria before and after

450   the medical use of antibiotics. Nature **306**:616-617.

451 3.   **Hughes VM, Datta N.** 1983. Conjugative plasmids in bacteria of the 'pre-antibiotic' era. Nature

452   **302**:725-726.

453 4.   **Carattoli A.** 2013. Plasmids and the spread of resistance. Int J Med Microbiol **303**:298-304.

454 5.   **Partridge SR.** 2011. Analysis of antibiotic resistance regions in Gram-negative bacteria. FEMS

455   Microbiol Rev **35**:820-855.

456 6.   **Stoesser N, Giess A, Batty EM, Sheppard AE, Walker AS, Wilson DJ, Didelot X, Bashir A, Sebra**

457   **R, Kasarskis A, Sthapit B, Shakya M, Kelly D, Pollard AJ, Peto TE, Crook DW, Donnelly P,**

458   **Thorson S, Amatya P, Joshi S.** 2014. Genome Sequencing of an Extended Series of NDM-

459   Producing Klebsiella pneumoniae Isolates from Neonatal Infections in a Nepali Hospital

460   Characterizes the Extent of Community- versus Hospital-Associated Transmission in an Endemic

461   Setting. Antimicrob Agents Chemother **58**:7347-7357.

462 7.   **Poirel L, Savov E, Nazli A, Trifonova A, Todorova I, Gergova I, Nordmann P.** 2014. Outbreak

463   caused by NDM-1- and RmtB-producing Escherichia coli in Bulgaria. Antimicrob Agents

464   Chemother **58**:2472-2474.

465 8.   **Garza-Ramos U, Barrios H, Reyna-Flores F, Sánchez-Pérez A, Tamayo-Legorreta E, Ibarra-**

466   **Pacheco A, Salazar-Salinas J, Núñez-Ceballos R, Silva-Sanchez J.** 2014. Characteristics of KPC-2-

467   producing Klebsiella pneumoniae (ST258) clinical isolates from outbreaks in 2 Mexican medical

468   centers. Diagn Microbiol Infect Dis **79**:483-485.

469   9.   **Kitchel B, Rasheed JK, Patel JB, Srinivasan A, Navon-Venezia S, Carmeli Y, Brolund A, Giske CG.**

470   2009. Molecular epidemiology of KPC-producing *Klebsiella pneumoniae* isolates in the United

471   States: clonal expansion of multilocus sequence type 258. Antimicrob Agents Chemother

472   **53**:3365-3370.

473   10.   **World Health Organization.** 2014. Antimicrobial Resistance A Global Report on Surviellence.

474   World Health Organization, Press W, 20 Avenue Appia, 1211 Geneva 27, Switzerland.

475   11.   **Centers for Disease Control and Prevention (CDC).** 2013. *Antibiotic resistance threats in the*

476   *United States.* (CDC) CfDCaP, Atlanta, GA USA.

477   12.   **Patel G, Huprikar S, Factor SH, Jenkins SG, Calfee DP.** 2008. Outcomes of carbapenem-resistant

478   Klebsiella pneumoniae infection and the impact of antimicrobial and adjunctive therapies. Infect

479   Control Hosp Epidemiol **29**:1099-1106.

480   13.   **Falagas ME, Tansarli GS, Karageorgopoulos DE, Vardakas KZ.** 2014. Deaths attributable to

481   carbapenem-resistant enterobacteriaceae infections. Emerg Infect Dis **20**:1170-1175.

482   14.   **Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, Cornaglia G,**

483   **Garau J, Gniadkowski M, Hayden MK, Kumarasamy K, Livermore DM, Maya JJ, Nordmann P,**

484   **Patel JB, Paterson DL, Pitout J, Villegas MV, Wang H, Woodford N, Quinn JP.** 2013. Clinical

485   epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. Lancet Infect

486   Dis **13**:785-796.

487   15.   **Yigit H, Queenan AM, Anderson GJ, Domenech-Sanchez A, Biddle JW, Steward CD, Alberti S,**

488   **Bush K, Tenover FC.** 2001. Novel carbapenem-hydrolyzing beta-lactamase, KPC-1, from a

489   carbapenem-resistant strain of Klebsiella pneumoniae. Antimicrob Agents Chemother **45**:1151-

490   1161.

491   16.   **Tzouvelekis LS, Miriagou V, Kotsakis SD, Spyridopoulou K, Athanasiou E, Karagouni E, Tzelepi**

492   **E, Daikos GL.** 2013. KPC-Producing, Multi-Drug Resistant Klebsiella pneumoniae ST258 as a

493   Typical Opportunistic Pathogen. Antimicrob Agents Chemother.

494   17.   **Leavitt A, Carmeli Y, Chmelnitsky I, Goren MG, Ofek I, Navon-Venezia S.** 2010. Molecular

495   epidemiology, sequence types, and plasmid analyses of KPC-producing Klebsiella pneumoniae

496   strains in Israel. Antimicrob Agents Chemother **54:**3002-3006.

497   18.   **Adler A, Hussein O, Ben-David D, Masarwa S, Navon-Venezia S, Schwaber MJ, Carmeli Y,**

498   **Group obotP-A-CHC-REW.** 2014. Persistence of Klebsiella pneumoniae ST258 as the

499   predominant clone of carbapenemase-producing Enterobacteriaceae in post-acute-care

500   hospitals in Israel, 2008-13. J Antimicrob Chemother.

501   19.   **Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, Chavda KD, Jacobs MR,**

502   **Mathema B, Olsen RJ, Bonomo RA, Musser JM, Kreiswirth BN.** 2014. Molecular dissection of

503   the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*.

504   Proc Natl Acad Sci U S A **111:**4988-4993.

505   20.   **Chen L, Mathema B, Chavda KD, DeLeo FR, Bonomo RA, Kreiswirth BN.** 2014. Carbapenemase-

506   producing Klebsiella pneumoniae: molecular and genetic decoding. Trends Microbiol **22:**686-

507   696.

508   21.   **Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song**

509   **Y, Tsai YC, Boitano M, Dayal J, Brooks SY, Schmidt B, Young AC, Thomas JW, Bouffard GG,**

510   **Blakesley RW, Mullikin JC, Korlach J, Henderson DK, Frank KM, Palmore TN, Segre JA, Program**

511   **NCS.** 2014. Single-molecule sequencing to track plasmid diversity of hospital-associated

512   carbapenemase-producing Enterobacteriaceae. Sci Transl Med **6:**254ra126.

27

513 22. **Ruiz-Garbajosa P, Curiao T, Tato M, Gijon D, Pintado V, Valverde A, Baquero F, Morosini MI,**

514 **Coque TM, Canton R.** 2013. Multiclonal dispersal of KPC genes following the emergence of non-

515 ST258 KPC-producing Klebsiella pneumoniae clones in Madrid, Spain. J Antimicrob Chemother.

516 23. **Ocampo AM, Chen L, Cienfuegos AV, Roncancio G, Chavda KD, Kreiswirth BN, Jiménez JN.**

517 2015. High frequency of non-CG258 clones of carbapenem-resistant *Klebsiella pneumoniae* with

518 distinct clinical characteristics: A two-year surveillance in five Colombian tertiary care hospitals.

519 Antimicrob Agents Chemother.

520 24. **Tavares CP, Pereira PS, Marques EeA, Faria C, de Souza MaP, de Almeida R, Alves CeF, Asensi**

521 **MD, Carvalho-Assef AP.** 2015. Molecular epidemiology of KPC-2-producing Enterobacteriaceae

522 (non-*Klebsiella pneumoniae*) isolated from Brazil. Diagn Microbiol Infect Dis **82:**326-330.

523 25. **Bonura C, Giuffrè M, Aleo A, Fasciana T, Di Bernardo F, Stampone T, Giammanco A, Palma DM,**

524 **Mammina C, Group M-GW.** 2015. An Update of the Evolving Epidemic of bla_{KPC} Carrying

525 *Klebsiella pneumoniae* in Sicily, Italy, 2014: Emergence of Multiple Non-ST258 Clones. PLoS One

526 **10:**e0132936.

527 26. **Chen L, Chavda KD, Melano RG, Jacobs MR, Koll B, Hong T, Rojtman AD, Levi MH, Bonomo RA,**

528 **Kreiswirth BN.** 2014. Comparative genomic analysis of KPC-encoding pKpQIL-like plasmids and

529 their distribution in New Jersey and New York Hospitals. Antimicrob Agents Chemother **58:**2871-

530 2877.

531 27. **Tijet N, Muller MP, Matukas LM, Khan A, Patel SN, Melano RG.** 2015. Lateral dissemination and

532 inter-patient transmission of bla_{KPC-3}: role of a conjugative plasmid in spreading carbapenem

533 resistance. J Antimicrob Chemother.

534 28. **Cuzon G, Naas T, Nordmann P.** 2011. Functional characterization of Tn*4401*, a Tn*3*-based

535 transposon involved in bla_{KPC} gene mobilization. Antimicrob Agents Chemother **55:**5370-5373.

536   29.   **Naas T, Cuzon G, Villegas MV, Lartigue MF, Quinn JP, Nordmann P.** 2008. Genetic structures at

537   the origin of acquisition of the beta-lactamase bla KPC gene. Antimicrob Agents Chemother

538   **52**:1257-1263.

539   30.   **Siu LK, Lin JC, Gomez E, Eng R, Chiang T.** 2012. Virulence and plasmid transferability of KPC

540   Klebsiella pneumoniae at the Veterans Affairs Healthcare System of New Jersey. Microb Drug

541   Resist **18**:380-384.

542   31.   **Cai JC, Zhou HW, Zhang R, Chen GX.** 2008. Emergence of Serratia marcescens, Klebsiella

543   pneumoniae, and Escherichia coli Isolates possessing the plasmid-mediated carbapenem-

544   hydrolyzing beta-lactamase KPC-2 in intensive care units of a Chinese hospital. Antimicrob

545   Agents Chemother **52**:2014-2018.

546   32.   **Brouwer MS, Tagg KA, Mevius DJ, Iredell JR, Bossers A, Smith HE, Partridge SR.** 2015. Incl

547   shufflons: Assembly issues in the next-generation sequencing era. Plasmid **80**:111-117.

548   33.   **Lanza VF, de Toro M, Garcillán-Barcia MP, Mora A, Blanco J, Coque TM, de la Cruz F.** 2014.

549   Plasmid flux in *Escherichia coli* ST131 sublineages, analyzed by plasmid constellation network

550   (PLACNET), a new method for plasmid reconstruction from whole genome sequences. PLoS

551   Genet **10**:e1004766.

552   34.   **Pecora ND, Li N, Allard M, Li C, Albano E, Delaney M, Dubois A, Onderdonk AB, Bry L.** 2015.

553   Genomically Informed Surveillance for Carbapenem-Resistant Enterobacteriaceae in a Health

554   Care System. MBio **6**:e01030.

555   35.   **Partridge SR.** 2015. Resistance mechanisms in Enterobacteriaceae. Pathology **47**:276-284.

556   36.   **Mathers AJ, Cox HL, Bonatti H, Kitchel B, Brassinga AK, Wispelwey B, Sawyer RG, Pruett TL,**

557   **Hazen KC, Patel JB, Sifri CD.** 2009. Fatal cross infection by carbapenem-resistant Klebsiella in

558   two liver transplant recipients. Transpl Infect Dis **11**:257-265.

559   37.   **Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, Didelot X, Turner SD, Sebra**

560   **R, Kasarskis A, Peto T, Crook D, Sifri CD.** 2015. *Klebsiella pneumoniae* carbapenemase (KPC)-

561   producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome

562   sequencing. Antimicrob Agents Chemother **59**:1656-1663.

563   38.   **Enfield KB, Huq NN, Gosseling MF, Low DJ, Hazen KC, Toney DM, Slitt G, Zapata HJ, Cox HL,**

564   **Lewis JD, Kundzins JR, Mathers AJ, Sifri CD.** 2014. Control of simultaneous outbreaks of

565   carbapenemase-producing enterobacteriaceae and extensively drug-resistant Acinetobacter

566   baumannii infection in an intensive care unit using interventions promoted in the Centers for

567   Disease Control and Prevention 2012 carbapenemase-resistant Enterobacteriaceae Toolkit.

568   Infect Control Hosp Epidemiol **35**:810-817.

569   39.   **Mathers AJ, Carroll J, Sifri CD, Hazen KC.** 2013. Modified Hodge Test versus Indirect

570   Carbapenemase Test: Prospective Evaluation of a Phenotypic Assay for Detection of Klebsiella

571   pneumoniae Carbapenemase (KPC) in Enterobacteriaceae. J Clin Microbiol **51**:1291-1293.

572   40.   **Mathers AJ, Poulter M, Dirks D, Carroll J, Sifri CD, Hazen KC.** 2014. Clinical Microbiology Costs

573   for Methods of Active Surveillance for Klebsiella pneumoniae Carbapenemase-Producing

574   Enterobacteriaceae. Infect Control Hosp Epidemiol **35**:350-355.

575   41.   **CDC.** 2009. Guidance for control of infections with carbapenem-resistant or carbapenemase-

576   producing Enterobacteriaceae in acute care facilities. MMWR Morb Mortal Wkly Rep **58**:256-

577   260.

578   42.   **Chen L, Chavda KD, Al Laham N, Melano RG, Jacobs MR, Bonomo RA, Kreiswirth BN.** 2013.

579   Complete nucleotide sequence of a blaKPC-harboring IncI2 plasmid and its dissemination in New

580   Jersey and New York hospitals: a hidden threat. Antimicrob Agents Chemother **57**:6.

581   43.   **Kitchel B, Sundin DR, Patel JB.** 2009. Regional dissemination of KPC-producing Klebsiella

582   pneumoniae. Antimicrob Agents Chemother **53**:4511-4513.

583 44. **Partridge SR, Hall RM.** 2005. Evolution of transposons containing blaTEM genes. Antimicrob

584 Agents Chemother **49:**1267-1268.

585 45. **Bailey JK, Pinyon JL, Anantham S, Hall RM.** 2011. Distribution of the blaTEM gene and blaTEM-

586 containing transposons in commensal Escherichia coli. J Antimicrob Chemother **66:**745-751.

587 46. **Zong Z, Ginn AN, Dobiasova H, Iredell JR, Partridge SR.** 2015. Different IncI1 plasmids from

588 Escherichia coli carry ISEcp1-blaCTX-M-15 associated with different Tn2-derived elements.

589 Plasmid.

590 47. **Rice LB, Carias LL, Hutton RA, Rudin SD, Endimiani A, Bonomo RA.** 2008. The KQ element, a

591 complex genetic region conferring transferable resistance to carbapenems, aminoglycosides,

592 and fluoroquinolones in *Klebsiella pneumoniae*. Antimicrob Agents Chemother **52:**3427-3429.

593 48. **Martínez T, Vázquez GJ, Aquino EE, Martínez I, Robledo IE.** 2014. ISEcp1-mediated

594 transposition of blaKPC into the chromosome of a clinical isolate of Acinetobacter baumannii

595 from Puerto Rico. J Med Microbiol **63:**1644-1648.

596 49. **Mathers AJ, Cox HL, Kitchel B, Bonatti H, Brassinga AK, Carroll J, Scheld WM, Hazen KC, Sifri**

597 **CD.** 2011. Molecular dissection of an outbreak of carbapenem-resistant enterobacteriaceae

598 reveals Intergenus KPC carbapenemase transmission through a promiscuous plasmid. MBio

599 **2:**e00204-00211.

600 50. **Guindon S, Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large

601 phylogenies by maximum likelihood. Syst Biol **52:**696-704.

602 51. **Reeves PR, Liu B, Zhou Z, Li D, Guo D, Ren Y, Clabots C, Lan R, Johnson JR, Wang L.** 2011. Rates

603 of mutation and host transmission for an Escherichia coli clone over 3 years. PLoS One **6:**e26907.

604

605

31

606   **Figure Legends**

607

608   **Figure 1. Diversity in bacterial species, strains, plasmids and Tn*4401* variants.** For each

609   species, a phylogeny was generated from mapping to a species-specific chromosomal

610   reference, after deduplication of closely related isolates from the same patient (see Methods).

611   Distinct strains are defined by a cutoff of ~500 SNVs (see Methods); strains found in more than

612   one patient are indicated by grey background shading. Circles show plasmid "presence" as

613   determined from Illumina data, with fill colour indicating uncertainty in whether the plasmid

614   contains bla_{KPC}. Boxes show plasmid structure determined from long-read PacBio sequencing

615   for 17 randomly chosen isolates, as well as the previously sequenced isolates from index

616   patient B (37). Where the PacBio-sequenced isolate was excluded from the phylogeny as a

617   patient duplicate, plasmid structure is shown for the corresponding closely related isolate from

618   the same patient. Tn*4401* and bla_{KPC} variants (Table 2) are indicated by large and small squares

619   respectively. The likely source of bla_{KPC} acquisition as determined from epidemiological data is

620   indicated by text colour.

621

622   **Figure 2. pKPC_UVA01-like plasmids identified through long-read PacBio sequencing.** The

623   reference pKPC_UVA01 sequence is shown, together with all 11 pKPC_UVA01-like plasmids

624   identified through long-read PacBio sequencing, including the six that do not contain bla_{KPC}.

625   Arrows indicate predicted open reading frames; Tn*4401* is shown in purple. Pink shading

626   indicates regions of identity between adjacent sequences, with SNVs indicated by red lines.

627

628 **Figure 3. Tn*4401* is commonly integrated into a Tn*2*-like element.** The Tn*4401* and surrounding

629 region (i.e. partial plasmid sequence, except for pKPC_CAV1320) is shown for each distinct

630 *bla*$_{KPC}$ plasmid. Variants of the same plasmid backbone (see Table 1) are not shown. Arrows

631 indicate predicted open reading frames; Tn*4401* is shown in purple. Pink shading indicates

632 regions of identity between adjacent sequences, with SNVs indicated by red lines and short

633 indels (1-2 bp) by blue lines. The top panel shows the Tn*2** reference sequence from AY123253

634 (44).

635

636 **Figure 4. Ward contacts between patients with genetically related isolates.** Each horizontal

637 line represents a different strain (top) or Tn*4401* variant (bottom). Filled circles indicate

638 patients that had previous ward contact with another patient on the same horizontal line (i.e.

639 possible patient-to-patient transmission). As Tn*4401*-1 is present in two-thirds of patients,

640 many coincidental ward contacts may be expected to occur, resulting in a substantial

641 overestimate of transmission number. Therefore, for Tn*4401*, the total number of acquisitions

642 explainable by direct ward contact is indicated, as well as with Tn*4401*-1 patients excluded. The

643 vertical line indicates onset of routine patient screening.

644

645 **Figure 5. *bla*$_{KPC}$ spreads at multiple genetic levels, resulting in a high level of diversity in**

646 ***bla*$_{KPC}$-positive *Enterobacteriaceae*.**

Table 1. *bla*$_{KPC}$-containing structures ascertained from long-read PacBio sequencing of 17 randomly chosen isolates

Isolate	Species	Patient	Date	*bla*$_{KPC}$ plasmid	Size (bp)	Group[a]	Within-group genetic changes[b]	Tn*4401* variant	Flanking sequence[c]	Tn2-like element[d]
CAV1344	*K. pneumoniae*	EP	Dec-2010	pKPC_CAV1344	176,497	Singleton	NA	Tn*4401*-1	GTTCT…GTTCT	Yes
CAV1392	*K. pneumoniae*	EU	Mar-2011	pKPC_CAV1392	43,621	pKPC_UVA01	1 SNV	Tn*4401*-5	GTTCT…GTTCT	Yes
				NA (chromosomal)	NA	NA	NA	Tn*4401*-5	AGATA…AGATA	No
CAV1596	*K. pneumoniae*	FK	Apr-2012	pKPC_CAV1596-78	77,801	Singleton	NA	Tn*4401*-5	GTTCT…GTTCT	Yes
				pKPC_CAV1596-97	96,702	Singleton	NA	Tn*4401*-5	TATCG…TATCG	No
CAV1099	*K. oxytoca*	AU	Apr-2009	pKPC_CAV1099	113,105	pKPC_UVA02	0 SNVs	Tn*4401*-1	ATGCA…GGCCA[e]	No
CAV1335	*K. oxytoca*	EQ	Dec-2010	pKPC_CAV1335	113,105	pKPC_UVA02	0 SNVs	Tn*4401*-1	ATGCA…GGCCA[e]	No
CAV1374	*K. oxytoca*	ED	Aug-2010	pKPC_CAV1374	332,956	Singleton	NA	Tn*4401*-1	GTTCT…GTTCT	Yes
CAV1043	*E. asburiae*	L	Mar-2008	pKPC_CAV1043	59,138	Singleton	NA	Tn*4401*-5	GTTCT…GTTCT	Yes
CAV1176	*E. cloacae*	DN	May-2010	pKPC_CAV1176	90,452	pKPC_CAV1176	0 SNVs	Tn*4401*-3	GTTCT…GTTCT	Yes
CAV1311	*E. cloacae*	EO	Jan-2011	pKPC_CAV1311	90,452	pKPC_CAV1176	0 SNVs	Tn*4401*-3	GTTCT…GTTCT	Yes
CAV1411	*E. cloacae*	FC	Jun-2011	pKPC_CAV1411	90,452	pKPC_CAV1176	1 SNV, 40 kb inversion	Tn*4401*-3	GTTCT…GTTCT	Yes
CAV1669	*E. cloacae*	HV	Aug-2012	pKPC_CAV1669	90,452	pKPC_CAV1176	40kb inversion	Tn*4401*-3	GTTCT…GTTCT	Yes
CAV1668	*E. cloacae*	HQ	Aug-2012	pKPC_CAV1668	43,433	pKPC_UVA01	1 SNV, 188 bp deletion	Tn*4401*-3	GTTCT…GTTCT	Yes
CAV1321	*C. freundii*	EG	Nov-2010	pKPC_CAV1321-45	44,846	pKPC_UVA01	1,225 bp insertion	Tn*4401*-1	GTTCT…GTTCT	Yes
				pKPC_CAV1321-244	243,709	Singleton	NA	Tn*4401*-1	GTTCT…GTTCT	Yes
CAV1741	*C. freundii*	ER	Oct-2012	pKPC_CAV1741	129,196	pKPC_UVA01	14,960 bp duplication, 70,615 bp insertion	Tn*4401*-1[f]	GTTCT…GTTCT	Yes
CAV1151	*K. intermedia*	CD	Sep-2009	pKPC_CAV1151	43,621	pKPC_UVA01	0 SNVs[g]	Tn*4401*-1	GTTCT…GTTCT	Yes
CAV1320	*E. aerogenes*	EL	Nov-2010	pKPC_CAV1320	13,981	Singleton	NA	Tn*4401*-1	TTGTT…TTGTT	No
CAV1492	*S. marcescens*	GL	Dec-2011	pKPC_CAV1492	69,158	Singleton	NA	Tn*4401*-8	TTTTT…TTTTT	No

648 [a] Plasmids are defined as belonging to the same group if the sequences are largely identical, allowing for a small number of substitutions and/or rearrangements that may be

649 expected to occur within the outbreak timeframe. Different groups have very limited homology outside the Tn*4401* region, indicative of independent integrations into distinct

650 plasmid structures. 'Singleton' indicates a plasmid backbone that is distinct from all others shown

651 [b] Differences relative to the reference sequence for that plasmid group, as specified in the previous column

652 [c] Sequences immediately flanking Tn*4401*; generally expected to be identical due to 5 bp target site duplication during transposition (28)

653 [d] Tn*4401* integrated into the *tnpA* gene of a Tn2-like element

654 [e] No evidence of target site duplication

655 [f] 2 copies

656 [g] It is noteworthy that this plasmid from CAV1151 (*K. intermedia*) is exactly identical to pKPC_UVA01 from CAV1016 (*K. pneumoniae*), with isolation dates 2 years apart

657 **Table 2. Tn*4401* variation**

Tn*4401* variant	Structural isoform(29)	SNVs[a]	*bla*$_{KPC}$ variant	Patients	Isolates	Strains
Tn*4401*-1[b]	b	-	*bla*$_{KPC-2}$	121	176	42
Tn*4401*-2	a (del 7020-7118)	-	*bla*$_{KPC-2}$	5	8	1
Tn*4401*-3	novel (del 6919-7106)	-	*bla*$_{KPC-2}$	28	39	2
Tn*4401*-4	truncated (del 1-6654)	-	*bla*$_{KPC-2}$	2	3	1
Tn*4401*-5	b	8015 C→T[c]	*bla*$_{KPC-3}$	22	40	19
Tn*4401*-6	b	8015 C→T, 9621 T→C	*bla*$_{KPC-3}$	1	3	2
Tn*4401*-7	b	7199 T→A, 8015 C→T, 9621 T→C	*bla*$_{KPC-3}$	1	1	1
Tn*4401*-8	b	9663 T→C	*bla*$_{KPC-2}$	1	3	1
Tn*4401*-9	b	7509 C→G, 7917 T→G[d]	*bla*$_{KPC-4}$	1	1	1
Tn*4401*-10	b	6800 T→C, 7509 C→G, 7917 T→G	*bla*$_{KPC-4}$	1	4	1
Tn*4401*-11	b	8015 N[e]	*bla*$_{KPC-2}$ / *bla*$_{KPC-3}$	1	2	1
Tn*4401*-12	truncated (del 1-6727)	6800 N[f]	*bla*$_{KPC-2}$	1	1	1

658 [a] With respect to Tn*4401*-1, which is considered as the reference Tn*4401* sequence for this study

659 [b] Tn*4401*-1 differs from the reference isoform b sequence in EU176013.1 by 14 SNVs, as follows: 4939 C→G, 4989 C→T, 5099

660 A→T, 5131 A→G, 5154 T→G, 5185 G→C, 5255 C→A, 5361 G→C, 5375 C→G, 5390 A→C, 5996 G→A, 5998 G→C, 8112 C→A,

661 8113 A→C

662 [c] This substitution converts *bla*$_{KPC-2}$ to *bla*$_{KPC-3}$

663 [d] These two substitutions convert *bla*$_{KPC-2}$ to *bla*$_{KPC-4}$

664 [e] Quality filters failed at this position due to a mixture of reads supporting C and T (i.e. Tn*4401*-11 actually represents a mixture

665 of Tn*4401*-1 and Tn*4401*-5)

666 [f] Quality filters failed at this position due to a lack of reads mapped in the reverse direction. All reads mapped in the forward

667 direction supported a reference (T) call

668

669 **Supplementary Legends**

670

671 **Figure S1. Distinct *bla*$_{KPC}$ plasmids identified through long-read PacBio sequencing.** Variants of

672 the same plasmid backbone (see Table 1) are not shown. Arrows indicate predicted open

673 reading frames; Tn*4401* is shown in purple.
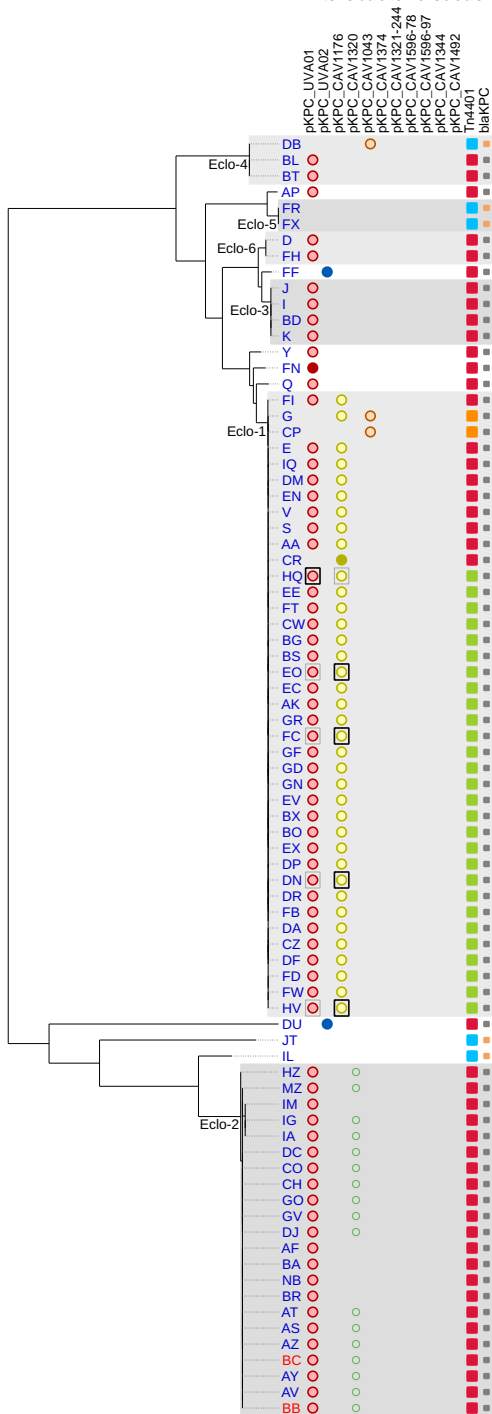
674

675 **Table S1. Details of sequenced isolates.**

676

677 **Table S2. Additional Tn*4401* insertion sites ascertained from short-read Illumina data**

678

679 **Table S3. Association of importation status with *K. pneumoniae* and the epidemic *bla*$_{KPC}$ *K.***

680 ***pneumoniae* strain ST258**

681

682 **Table S4. Patients with multiple *bla*$_{KPC}$-positive strains or species**

683

684 **Table S5. Chromosomal references used for mapping**

685

pKPC_CAV1741

pKPC_CAV1321-45

pKPC_CAV1668

pKPC_CAV1392

pKPC_CAV1151

pKPC_UVA01

pCAV1669-34

pCAV1411-34

pCAV1311-34

pCAV1176-34

pCAV1374-34

pCAV1344-40

blaKPC

pKPC_UVA01-like plasmids
with blaKPC

pKPC_UVA01-like plasmids
without blaKPC

Antibiotic resistance
Transposition / Recombination
Conjugation / Type IV secretion
Replication
Hypothetical protein
Other
Tn4401

10 kb

Legend:
- Antibiotic resistance (red)
- Transposition / Recombination (green)
- Conjugation / Type IV secretion (light blue)
- Replication (blue)
- Hypothetical protein (yellow)
- Other (gray)
- Tn4401 (purple)

Plasmids shown: pKPC_CAV1176, pKPC_CAV1321-244, pKPC_UVA01, pKPC_CAV1374, pKPC_CAV1344, pKPC_CAV1596-78, pKPC_CAV1043, pKPC_CAV1596-97, pKPC_CAV1320, pKPC_UVA02, pKPC_CAV1492

Gene labels: Tn2*, blaTEM-1, tnpR, tnpA, lstA, lstB, blaKPC

Bracket annotations: Tn4401 integrated into Tn2-like element; Distinct Tn4401 integration sites

Scale: 1 kb

Host strain

Patients with ward contact
48/167 (29%)

Tn4401

Patients with ward contact
All local acquisitions: 106/167 (63%)
Excluding Tn4401-1: 15/50 (30%)

Tn4401-8
Tn4401-10
Tn4401-9
Tn4401-11
Tn4401-7
Tn4401-6
Tn4401-3
Tn4401-12
Tn4401-2
Tn4401-5
Tn4401-4
Tn4401-1

1 Jan 2008   1 Jan 2009   1 Jan 2010   1 Jan 2011   1 Jan 2012   1 Jan 2013

Klebsiella pneumoniae
Klebsiella oxytoca
Enterobacter cloacae
Citrobacter freundii
Other

Imported
Ward contact, same Tn4401 variant
Ward contact, different Tn4401 variant
No ward contact

Klebsiella pneumoniae carbapenemase (KPC) resistance gene
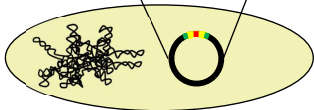
bla_{KPC}

Transposition of Tn*4401* into different plasmids

Mobile transposon Tn*4401*

Tn*4401* nested within mobile Tn*2*-like element

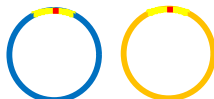Proposed mobilization of Tn*4401* between Tn2-like elements in different plasmids, **via homologous recombination**

Transposon located within mobile extra-chromosomal plasmid DNA
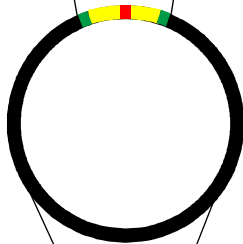
Bacterial cell with mobile plasmid and immobile host chromosome (not to scale)
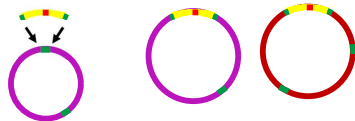
Plasmid movement via conjugation into different bacterial strains and species