1 **Genomic epidemiology of complex, multi-species, plasmid-borne *bla*$_{KPC}$**

2 **carbapenemase in Enterobacterales in the UK, 2009-2014**

3

4 Stoesser N*[#a, b]/Phan HTT*[b,c], Anna C. Seale[a,d], Zoie Aiken[e], Stephanie Thomas[e],

5 Matthew Smith[e], David Wyllie[a], Ryan George[e], Robert Sebra[f], Amy J Mathers[g,h],

6 Alison Vaughan[a], Tim EA Peto[a,b,i], Matthew J Ellington[j], Katie L Hopkins[j], Derrick

7 W Crook[a,b,i], Alex Orlek[a,b], William Welfare[k], Julie Cawthorne[e], Cheryl Lenney[e],

8 Andrew Dodgson[e, l], Neil Woodford[j], A Sarah Walker[a,b,i], and the TRACE

9 Investigators' Group

10 * These authors contributed equally to this work. NS coordinated the study, and is

11 therefore listed first.

12 # Corresponding author

13

14 [a] Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK

15 [b] NIHR Health Protection Research Unit in Healthcare Associated Infections and

16 Antimicrobial Resistance, University of Oxford, Oxford, UK

17 [c] Faculty of Medicine, University of Southampton, Southampton, UK

18 [d] Department of Infectious Disease Epidemiology, London School of Hygiene and

19 Tropical Medicine, UK

20 [e] Manchester University NHS Foundation Trust, Manchester, UK

21 [f] Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of

22 Medicine, Mount Sinai, New York, USA

23 [g] Division of Infectious Diseases and International Health, Department of Medicine,

24 University of Virginia Health System, Charlottesville, Virginia, USA

25 [h] Clinical Microbiology Laboratory, Department of Pathology, University of Virginia
26 Health System, Charlottesville, Virginia, USA
27 [i] NIHR Biomedical Research Centre, University of Oxford, Oxford, UK
28 [j] Antimicrobial Resistance and Healthcare Associated Infections (AMRHAI)
29 Reference Unit, National Infection Service, Public Health England, London, UK
30 [k] Public Health England North West, Manchester, UK
31 [l] Public Health Laboratory, Public Heath England, Manchester, UK
32
33 **Running title**: Genomic epidemiology of bla_{KPC} in the UK
34
35 **Keywords**: carbapenemase, Enterobacterales, KPC, whole genome sequencing, short-
36 read sequencing, long-read sequencing,
37
38 **Corresponding author**: Nicole Stoesser
39 Email: nicole.stoesser@ndm.ox.ac.uk
40 Tel: +44 (0) 1865 220856

41 **ABSTRACT**

42 Carbapenem resistance in Enterobacterales is a public health threat. *Klebsiella*
43 *pneumoniae* carbapenemase (encoded by alleles of the bla_{KPC} family) is one of the
44 commonest transmissible carbapenem resistance mechanisms worldwide. The
45 dissemination of bla_{KPC} has historically been associated with distinct *K. pneumoniae*
46 lineages (clonal group 258 [CG258]), a particular plasmid family (pKpQIL), and a
47 composite transposon (Tn*4401)*. In the UK, bla_{KPC} has represented a large-scale,
48 persistent, management challenge for some hospitals, particularly in North-West
49 England. The dissemination of bla_{KPC} has evolved to be polyclonal and poly-species,
50 but the genetic mechanisms underpinning this evolution have not been elucidated in
51 detail; this study used short-read whole genome sequencing of 604 bla_{KPC}-positive
52 isolates (Illumina) and long-read assembly (PacBio)/polishing (Illumina) of 21
53 isolates for characterisation. We observed the dissemination of bla_{KPC} (predominantly
54 bla_{KPC-2}; 573/604 [95%] isolates) across eight species and more than 100 known
55 sequence types. Although there was some variation at the transposon level (mostly
56 Tn*4401*a, 584/604 (97%) isolates; predominantly with ATTGA-ATTGA target site
57 duplications, 465/604 [77%] isolates), bla_{KPC} spread appears to have been supported
58 by highly fluid, modular exchange of larger genetic segments amongst plasmid
59 populations dominated by IncFIB (580/604 isolates), IncFII (545/604 isolates) and
60 IncR replicons (252/604 isolates). The subset of reconstructed plasmid sequences (21
61 isolates, 77 plasmids) also highlighted modular exchange amongst non-bla_{KPC} and
62 bla_{KPC} plasmids, and the common presence of multiple replicons within bla_{KPC}
63 plasmid structures (>60%). The substantial genomic plasticity observed has important
64 implications for our understanding of the epidemiology of transmissible carbapenem

65 resistance in Enterobacterales, for the implementation of adequate surveillance
66 approaches, and for control.

INTRODUCTION

Antimicrobial resistance (AMR) in Enterobacterales is a critical public health threat. Carbapenem resistance is of particular concern, and its evolution and spread in multiple species of Enterobacterales (i.e. carbapenemase-producing Enterobacterales [CPE]) is increasingly reported(1-4). Exchange of AMR genes, including carbapenem resistance genes, happens at multiple genetic levels(5), and is often facilitated by their presence on plasmids [circular DNA structures of variable size (2kb~>1Mb)], and/or other smaller mobile genetic elements (MGEs) such as transposons and insertion sequences (IS), that form part of the accessory genome.

Whole genome sequencing (WGS) has significantly improved our understanding of infectious diseases epidemiology and is used in both community-associated and nosocomial transmission analyses(6, 7). Although useful for delineating transmission routes in clonal, strain-based outbreaks, standard phylogenetic approaches and comparative analyses have been more difficult where multiple bacterial strains/species and transmissible resistance genes are involved(5). Reconstruction of the genetic structures of plasmids carrying relevant antimicrobial resistance genes using long-read sequencing has improved our understanding of the genetic complexity of the spread of important resistance genes, but has been difficult to undertake on a large scale.

Although approximately 40 *Klebsiella pneumoniae* carbapenemase (KPC; encoded by bla_{KPC}) variants have now been described (as per NCBI's AMR reference gene catalogue, available at https://www.ncbi.nlm.nih.gov/pathogens/isolates#/refgene/), only two have been most widely reported globally, namely KPC-2 and KPC-3

92 (H272Y with respect to KPC-2; single nucleotide difference in *bla*$_{KPC}$ [C814T])(8, 9).
93 In the UK, the first KPC isolate identified was a KPC-4-containing *Enterobacter* sp.
94 isolated in Scotland in 2003(10), with subsequent identification of KPC-3 in isolates
95 in the UK in 2007. From 2007, increasing numbers of suspected KPC isolates were
96 referred to Public Health England (PHE's) Antimicrobial Resistance and Healthcare
97 Associated Infections (AMRHAI) Reference Unit, with the majority of confirmed
98 KPC-producers (>95%) being KPC-2 and from hospitals in North-West England, first
99 recognised in 2008-2009(11). These isolates were predominantly *bla*$_{KPC}$-positive
100 Enterobacterales cultured from patients in the Central Manchester University
101 Hospitals NHS Foundation Trust (CMFT; now part of Manchester University NHS
102 Foundation Trust)(12). *bla*$_{KPC}$ is thought to have been introduced into the region via a
103 pKpQIL-like plasmid(13, 14), a plasmid backbone previously associated with the
104 global dissemination of *bla*$_{KPC}$ in *K. pneumoniae* clonal group 258, and already
105 observed in other *K. pneumoniae* sequence types (STs) and species in an analysis of
106 44 UK KPC-Enterobacterales from 2008-2010(14).
107
108 We used WGS to undertake a large-scale retrospective study of this multi-species,
109 polyclonal, dissemination of *bla*$_{KPC}$ in major Manchester hospitals in North-West
110 England from 2009, generating complete genome structures, including *bla*$_{KPC}$
111 plasmids, for a subset of isolates. We contextualised our analysis of strains in
112 Manchester by sequencing a subset of isolates from the local region (North-West
113 England) and other hospitals in the UK collected through a national *bla*$_{KPC}$
114 surveillance programme, with the goal of understanding the genetic structures
115 associated with the regional emergence of *bla*$_{KPC}$ in this setting.
116

117 **RESULTS**

118 Of 742 isolates identified for sequencing, 60 (8%) were not retrievable or cultivable
119 from the laboratory archives. After de-duplicating by taking the first bla_{KPC}-positive
120 Enterobacterales (KPC-E) per patient, and excluding sequencing failures, any
121 sequences without bla_{KPC} (assumed lost in culture), and mixtures (identified from
122 genomic data analysis, see Methods), 604 evaluable isolate sequences were included.
123 These represented: 327 archived isolates (54%) from inpatients in the early stages of
124 the observed introduction of bla_{KPC} in the two Manchester hospitals studied (2009-
125 2011), of which 309 and 18 isolates were from CMFT and the University Hospital of
126 South Manchester NHS Foundation Trust (UHSM; now part of Manchester
127 University NHS Foundation Trust) respectively; 78 (13%) later isolates from
128 CMFT/UHSM (2012-2014); 119 (20%) isolates from other hospitals (n=15 hospitals)
129 in North-West England (2009-2014, excluding CMFT and UHSM, up to the first 25
130 consecutive KPC-E isolates per hospital); 72 (12%) isolates from UK and Irish
131 hospitals (n=72 locations [n=4 from Ireland]) outside the North-West (2009-2014)
132 (first KPC-E isolate per hospital); and 8 (1%) isolates from English
133 outpatient/primary care settings (7 from the North-West region, 1 from a southern UK
134 location). The geographic and numerical distribution of isolates is depicted in
135 Supplementary Data S1.
136
137 Consistent with increasing numbers of bla_{KPC}-positive Enterobacterales reported
138 nationally to the reference laboratory, cases in CMFT/UHSM also began to rise from
139 2009. Anecdotally the first cases were reported in 2009, with 63 carbapenem-resistant
140 Enterobacterales cultured from 18630 microbiological specimens processed (0.3%),
141 with a ten-fold increase by 2014 (988/29593 [3%]) (Supplementary Data S2).

142 *bla_{KPC-2} dominates in the UK, but highly variable bla_{KPC} copy number and associated*
143 *resistance marker profiles suggest a diverse, flexible accessory genome underpinning*
144 *its spread*
145 Although three *bla*$_{KPC}$ variants were observed in the 604 included isolates, *bla*$_{KPC-2}$
146 dominated (n=573, 95%); *bla*$_{KPC-3}$ [n=27, 4%] and *bla*$_{KPC-4}$ [n=4, 1%]) were also
147 observed. Two isolates (0.3%; trace524, trace534) showed evidence of mixed
148 populations of *bla*$_{KPC-2}$ and *bla*$_{KPC-3}$.
149
150 The median *bla*$_{KPC}$ copy number estimate was 1.8 (IQR: 1.6-2.1), with a maximum of
151 8.2. *bla*$_{KPC}$ copy number estimates were strongly associated with meropenem
152 minimum inhibitory concentrations (MICs) for 588 isolates for which both were
153 available (Supplementary Data S3; estimated +0.56 higher doubling dilution (95%
154 +0.40,+0.72) per copy number higher, $p<0.0001$), and also between approaches
155 deriving estimates from short-read assemblies versus from mapping to reconstructed
156 genomes (see Methods; Supplementary Data S4; Pearson's correlation co-
157 efficient=0.97 [$p=0.0001$]). Across the three main species, *bla*$_{KPC}$ copy numbers were
158 higher in *K. pneumoniae* (n=525 [87%], median 1.8 [IQR: 1.6-2.1]), than *E. coli* (40
159 [7%]: 1.7 [1.5-1.9]) or *E. cloacae* (26 [4%], 1.6 [1.4-2.0]) (Kruskal-Wallis; $p=0.0003$;
160 Fig.1A). Amongst common STs, copy number was highest in *K. pneumoniae* ST258
161 (n=65 [11%], median 2.4 [IQR: 1.8-2.9]) versus other species/STs (n=531 [89%],
162 median 1.8 [1.6, 2.0]) (Kruskal-Wallis; $p=0.0001$; Fig.1B, Supplementary Data S5).
163 Of note, *bla*$_{KPC}$ copy number estimates represent an average across all individual cells
164 sequenced; *bla*$_{KPC}$ copy number estimates <1 in a small number of isolates (20/604
165 [3%]) (median=0.82 [IQR: 0.40-0.96]) suggest that a proportion of cells in the
166 populations sequenced may have lost their *bla*$_{KPC}$-harbouring plasmid.

167

168 There were 364 distinct resistance marker profiles in isolates, with only 12% (74/604)
169 of isolates sharing exactly the same profile as >10 other isolates (Supplementary
170 dataset SD1). Other broad or extended-spectrum beta-lactamase genes were also
171 commonly present across isolates, including: *bla*$_{TEM}$ (n=452, all *bla*$_{TEM-1}$), *bla*$_{OXA}$
172 (n=492; *bla*$_{OXA-9}$ [n=425], *bla*$_{OXA-1}$ [n=138]), *bla*$_{SHV}$ (n=497) and *bla*$_{CTX-M}$ (n=89;
173 *bla*$_{CTX-M-15}$ [n=57], *bla*$_{CTX-M-9}$ [n=28]). Aminoglycoside resistance genes were also
174 widely prevalent: *aac* (n=243), *aph* (n=196), *ant* (n=93) and *aadA* (n=280). In terms
175 of acquired quinolone resistance, 160 isolates contained *qnr* variants, and 137 isolates
176 contained *aac(6')-Ib-cr*; no *qep* variants were seen.

177

178 *bla*$_{KPC}$ *in the UK is a multi-species, largely polyclonal phenomenon*

179 In contrast to the almost uniform presence of *bla*$_{KPC-2}$ in isolates, species and lineage
180 diversity amongst our entire isolate collection was substantial, with eight different
181 species amongst sequenced isolates. For species with developed MLST schemes, this
182 represented a total of 102 different known species-ST combinations and 26 additional
183 unknown species-ST combinations, namely: *K. pneumoniae* (n=525 isolates, 70
184 known STs, 20 unknown STs), *E. coli* (n=40, 20 known STs, 1 unknown ST),
185 *Enterobacter cloacae* (n=26, 9 known STs, 2 unknown STs), *Klebsiella oxytoca* (n=6,
186 3 known STs, 3 unknown STs), *Raoultella ornithinolytica* (n=4), *Enterobacter*
187 *aerogenes* (n=2), *Serratia marcescens* (n=1) and *Kluyvera ascorbata* (n=1). The most
188 common STs were all *K. pneumoniae*, including ST258 (n=66), ST11 (n=35), ST491
189 (n=31), ST1162 (n=29) and ST54 (n=27) (Fig.2). Amongst these five common STs,
190 the distribution of pairwise single nucleotide variant (SNV) distances between isolates
191 suggested that some isolates could be considered highly genetically related at the

192 strain-level; however overall, when considering SNV distances alongside specific
193 accessory genome differences, isolates from these STs were not clonal
194 (Supplementary Data S6-10). Notably, 42% (252/604) of isolates were found in STs
195 represented by ≤10 isolates only (Fig.2).
196
197 Although some of the earliest sequenced isolates in the collection were KPC-*K.*
198 *pneumoniae* ST258 and ST11 (both in 2009) [two major KPC strains from CG258
199 circulating globally and in China at the time(8, 15)] and although KPC-producing *K.*
200 *pneumoniae* ST258 appears to have been one of the earliest strains observed in the
201 two Manchester hospitals (CMFT and UHSM), multiple diverse STs and species were
202 clearly subsequently rapidly recruited specifically in these two hospital settings in
203 2010 and 2011, with bla_{KPC} emerging in at least 30 new species-ST groups per year
204 (76% and 59% of first-per-patient carbapenem-resistant Enterobacterales culture-
205 positives sequenced, respectively; Fig.3)
206
207 *Most* bla_{KPC} *in the UK is supported by a conserved Tn4401a unit with uniform target*
208 *site sequences, suggesting that direct Tn4401 transposition is not the main mode of*
209 bla_{KPC} *transmission in this context*
210 In the absence of evidence of significant clonal spread by "high-risk" bacterial
211 lineages, we explored the diversity amongst mobile genetic features. Tn*4401* is a
212 ~10kb transposon that has been the major transposable context for bla_{KPC} to date, and
213 is flanked by 5bp signatures of transposition (target site duplications [TSD] or target
214 site sequences [TSS](16, 17), with no known target site specificity(16). A
215 predominant Tn*4401* isoform was associated with both $bla_{\text{KPC-2}}$ and $bla_{\text{KPC-3}}$ in this
216 study, namely Tn*4401*a(16), which occurred in 584/604 (97%) isolates (Fig.4). Other

217 known variants included Tn*4401*b (n=7) and Tn*4401*d (n=3). Only 20/584 (3%)
218 isolates demonstrated evidence of SNV-level variation in Tn*4401*a (homozygous calls
219 at 6 positions; heterozygous calls [i.e. mixed populations] at 3 positions).
220
221 *bla*$_{KPC-2}$-Tn*4401*a (n=539 isolates) was predominantly flanked by a 5-bp TSD
222 ATTGA, with 465/604 (77%; 465/539 [86%] of this sub-type) isolates with this
223 Tn*4401*/TSD combination throughout the study period (Fig.4). In 74 other *bla*$_{KPC-2}$-
224 Tn*4401*a isolates, the Tn*4401*a was flanked by other TSS combinations, consistent
225 with additional transposition events. Thirty-two of these were TSDs (16 AATAT-
226 AATAT, 16 AGTTG-AGTTG), which have been described as more consistent with
227 inter-plasmid transposition of Tn*4401*(18), and 35 were non-duplicate TSS
228 combinations (ATTGA with either ATATA, TGGTA, CTGCC, AATAA, AGGAT),
229 described as more consistent with intra-plasmid transposition. Evidence of multiple
230 TSSs around *bla*$_{KPC-2}$-Tn*4401*a within single isolates was seen in 6 cases (i.e. multiple
231 right and/or left Tn*4401* TSSs); 1 case had a right TSS present, but no left TSS
232 identified.
233
234 *Plasmid replicon typing demonstrates diverse plasmid populations present in bla*$_{KPC}$-
235 *positive isolates in the UK, but with combinations of IncF, IncR, ColRNAI and IncX3*
236 *replicons predominating*
237 The 604 isolates contained 91 unique combinations of plasmid replicon family types,
238 a crude proxy of plasmid populations present. However, it was not possible to
239 determine co-localisation of specific replicon types on plasmid structures, or direct
240 associations with *bla*$_{KPC}$ using this approach and short-read sequencing data. No
241 isolate was replicon negative. However, there were seven predominant replicon

242 family combinations (Fig.5) represented in 443/604 (73%) isolates, and these
243 included six major replicon family types, namely IncF (FIB [found in n=580 isolates],
244 FII [n=545]), FIA (n=103), IncR (n=252), ColRNAI (n=86), and IncX3 (n=60).
245 IncFIB/IncFII and IncFIB/IncFII/IncR populations were most widely distributed
246 across species-STs (Supplementary Data S11), geographical regions and over time
247 (Fig.5). The diversity of plasmid backgrounds present in these isolates may facilitate
248 opportunities for bla_{KPC} exchange amongst different plasmid families.
249

250 For many of the plasmid families, several different reference replicon sequences exist
251 in the PlasmidFinder database, with a degree of homology amongst sequences in the
252 same family, making it difficult to establish robustly which sub-type of replicon is
253 present. However, restricting to 90% matches to reference replicon types (with
254 matches being a composite of percentage sequence identity x percentage reference
255 sequence coverage) for these common families, top matches found in more than 10%
256 of isolates included: IncFIB(K)_1_Kpn3_JN233704 (n=490 isolates);
257 IncFIB(pQil)_JN233705 (n=299); IncR_1_DQ449578 (n=253);
258 IncFII(K)_1_CP000648 (n=89; plasmid MLST $IncFII_{K1}$); ColRNAI_1_DQ298019
259 (n=86); IncFIA(HI1)_1_HI1_AF250878 (n=80); IncFII_1_pKP91_CP000966 (n=90;
260 plasmid MLST $IncFII_{K4}$); and IncX3_1_JN247852 (n=61). At this more detailed
261 typing resolution, we found 183 plasmid replicon sub-type profiles amongst our
262 isolates, only ten of which were shared across ≥10 isolates, and two across more than
263 10% of isolates. The two most widespread profiles were IncFIB(pQil)_JN233705 +
264 IncFIB(K)_1_Kpn3_JN233704, and IncFIB(K)_1_Kpn3_JN233704 +
265 IncR_1__DQ449578 (Supplementary Data S12).
266

267 *Reference-, short-read based bla$_{KPC}$ plasmid typing demonstrates that pKpQIL-like*

268 *plasmids have been key in disseminating bla$_{KPC}$ in the UK, but that no known bla$_{KPC}$*

269 *plasmid vector was responsible in almost a third of cases*

270 We then tried to identify the most likely *bla*$_{\mathrm{KPC}}$-associated plasmid candidate, as

271 opposed to broadly profiling plasmid populations within isolates by plasmid replicon

272 typing. Attempts to identify complete plasmids from short-read data by comparison to

273 a reference plasmid database have been estimated as being correct in only ~45%-85%

274 of cases in previous studies(5, 19). However, 13/14 (93%) of isolates for which we

275 had hybrid assemblies (see Methods) with only one completely reconstructed *bla*$_{\mathrm{KPC}}$

276 plasmid had the correct top match using this *bla$_{KPC}$* plasmid typing method

277 (Supplementary Dataset SD2). We therefore compared all short-read sequences with

278 our reference *bla*$_{\mathrm{KPC}}$ plasmid database (see Methods), recognising that any complete

279 plasmid typing approach from short-read data is sub-optimal; matches to one or more

280 reference *bla*$_{\mathrm{KPC}}$ plasmid sequences were identified in 554/604 (92%) isolates.

281 Filtering the single match with the highest score at the predefined ≥0.80 threshold left

282 a subset of 428/554 (77%) for evaluation. These 428 isolates had matches to 12

283 *bla*$_{\mathrm{KPC}}$ plasmid clusters (Fig.6).

284

285 Whilst the majority of isolates appeared to contain pKpQIL-like plasmids (323/604

286 [53%]), no significant matches to any reference plasmid were found in 162/604 (27%)

287 isolates, suggesting that the genetic background supporting *bla*$_{\mathrm{KPC}}$ in these isolates

288 has diversified substantially and rapidly (Fig.6). *bla*$_{\mathrm{KPC}}$ plasmid cluster assignations

289 were shared across a median (IQR) of 3 (1-6) STs, with pKpQIL-like plasmids being

290 most widespread across species/STs (7 species, 75 STs; Fig.6), and clearly playing a

291 major role in the dissemination of *bla*$_{\mathrm{KPC}}$ in Manchester, North-West England, and

292 nationally over time (Fig.7). Other plasmid types identified as top-matches across the
293 entire dataset included those fully resolved by long-read sequencing performed within
294 this study, some of which were seen in ≥5% of study isolates (e.g. pKPC-trace75 [a
295 non-typeable replicon]), and in non-North-West UK settings, likely reflecting
296 recombination and generation of new *bla*$_{KPC}$ plasmid variants in North-West England
297 and their subsequent dissemination (Fig.7). Of note, amongst the 26 *E. cloacae*
298 isolates sequenced, 12 (46%; five STs) had a match to the pKPC-272-like cluster,
299 suggesting perhaps that *E. cloacae* is a particularly suitable host for this plasmid
300 group (Fig.6).
301
302 *Reconstruction of a subset of genomes using long-read sequencing data demonstrates*
303 *a diverse population of plasmids with genetic rearrangement observed in both bla*$_{KPC}$*-*
304 *positive and bla*$_{KPC}$*-negative cases*
305 In addition to short-read data, to resolve genetic structures fully we obtained long-
306 read PacBio data for 23 isolates, chosen to maximise the *bla*$_{KPC}$ plasmid diversity
307 assayed and focussing on isolates collected from the two main Manchester hospitals
308 (12 CMFT isolates, 5 UHSM; plus 2 from other hospitals in North-West England, 4
309 from other UK locations). These included the two earliest available *bla*$_{KPC}$ isolates
310 from both CMFT and UHSM, as well as isolates sharing the same species/ST but with
311 different plasmid replicon combinations or from North West regional versus national
312 locations, same-species isolates with different STs, and isolates of different species.
313 One PacBio sequencing dataset represented a clear isolate mixture (trace597 [UHSM]
314 of *E. cloacae* ST133 and *K. pnemoniae* ST258), and for one isolate (trace457
315 [CMFT]), there were discrepancies between the short-read and long-read sequencing
316 datasets, suggesting a laboratory error (*E. cloacae* ST45 long-read, *E. coli* ST88

short-read). These two assemblies were excluded, leaving 21 assemblies for further analysis (Table S1).

Of the 153 contigs from these 21 assemblies, 30 were clearly chromosomal, 77 plasmid, one chromosomal with an integrated plasmid, and 45 with unclear provenance (i.e. possibly phage, plasmid, or chromosomal). Overall 78/153 [51%] contigs were circularised, including 56/77 [73%] clear plasmid sequences. Thirty-one contigs (21 [68%] circularised) harboured *bla*$_{\text{KPC}}$, of which 21 were clearly plasmids (14/21 [67%] circularised) and one (trace552, *K. pneumoniae* ST11) had *bla*$_{\text{KPC}}$ integrated into the chromosome (not circularised). For nine other *bla*$_{\text{KPC}}$-positive contigs, we were not able to clearly ascertain whether these were chromosomal or plasmid (Supplementary Dataset SD3).

We observed *bla*$_{\text{KPC}}$ in multiple plasmid backgrounds (Fig.8), including a majority of *bla*$_{\text{KPC}}$ plasmids with multiple replicons (13/21 [60%] clear plasmid contigs, as represented in Fig.8), particularly with IncFIB/IncFII and/or IncR, consistent with replicon patterns in the isolates overall (Fig.5). For the IncFII group, for which we had 11 complete, reconstructed plasmid sequences with an IncFII(K)_CP000648-like replicon (plasmidFinder match; 3 *bla*$_{\text{KPC}}$-negative [i.e. not represented in Fig.8] and 8 *bla*$_{\text{KPC}}$-positive), there was evidence of exchange and rearrangement of plasmid components between both *bla*$_{\text{KPC}}$-positive and *bla*$_{\text{KPC}}$-negative plasmids, as well as sharing between STs and species (Fig.9). For example, between the *bla*$_{\text{KPC}}$-negative IncFII(K) plasmid isolated from a *Raoultella ornithinolytica* isolate and a *bla*$_{\text{KPC}}$-positive *K. pneumoniae* ST1828 isolate, >70kb of sequence was ≥90% identical at the nucleotide level, as well as being largely syntenic (Fig.9, annotation (i)). A

342 genetically related IncFII(K)/IncR plasmid was shared across species/lineages (*K.*
343 *pneumoniae* [novel ST], *E. coli* [ST372], *K. pneumoniae* [ST883]); and was also
344 similar to an IncFII(K)/IncFIB(pQIL) plasmid (found in a *K. pneumoniae* ST1828 and
345 a *K. pneumoniae* ST588) (Fig.9, annotation (ii)).
346
347 In addition to their plasticity, part of the success of these bla_{KPC} plasmids may also be
348 attributable to the presence of toxin-antitoxin plasmid addiction systems (*ccdA*/*ccdB*
349 n=4 bla_{KPC} plasmids; *higA* n=6; *vapB/vapC* n=11); anti-restriction mechanisms (*klcA*
350 n=16, previously shown to promote bla_{KPC} dissemination(20)); and heavy metal
351 resistance (*terB* [tellurite] n=3; *ars* operon [arsenicals] n=3; chromate resistance n=1;
352 *cop* operon/*pcoC*/*pcoE* [copper] n=7; *mer* operon [mercury] n=10).
353
354 **DISCUSSION**
355 We present the largest WGS-based analysis of bla_{KPC}-positive isolates (n=604) to our
356 knowledge, focused on assessing genetic diversity around the carbapenemase gene
357 itself rather than limiting the analysis based on species type, and incorporating a
358 sampling frame from UK regional and national collections, over five years. bla_{KPC}
359 remains one of the three most common carbapenemases observed in the UK,
360 accounting for ~11% of cases referred to the AMRHAI Reference Unit in 2018
361 (OXA-48-like=52%, NDM=27%)(21), and presenting a significant challenge to
362 hospitals in North-West England, including Manchester, where it accounted for >97%
363 of carbapenem resistance through 2015(22).
364
365 Our study provides an interesting context in which to consider the findings of a
366 recently published pan-European survey of carbapenem non-susceptible *K.*

367 *pneumoniae* (the EuSCAPE study; 6 months, 2013-2014; 244 hospitals, 32
368 countries)(23). In EuSCAPE, 684 carbapenemase-producing *Klebsiella* spp. isolates
369 were sequenced using Illumina technology, and similar to our study, most cases were
370 healthcare-exposed (<2% from outpatients). EuSCAPE carbapenemase-producing
371 isolates were also predominantly bla_{KPC} (~45%, n=311 isolates), but mostly bla_{KPC-3}
372 (232/311 [75%] versus 27/604 [5%] in our study), and ST258/ST512 (226/311 [73%]
373 versus 107/525 (20%) of *K. pneumoniae* overall in our study). Based on identifying
374 genetic "nearest-neighbours" in their data, the EuSCAPE team found 51% of bla_{KPC}-
375 *K. pneumoniae* were most closely related to another isolate from the same hospital.
376 The authors concluded that there was strong geographic structuring of strains, and that
377 the expansion of a handful of clonal lineages was predominantly responsible for the
378 spread of carbapenemases in *K. pneumoniae* in Europe, with onward nosocomial
379 transmission. Like bla_{KPC-3} in EuSCAPE, bla_{KPC-2} has also been linked with the clonal
380 expansion of ST258 in Australia(24), where 48% of 176 *K. pneumoniae* isolates
381 sequenced were bla_{KPC-2}-containing ST258.
382
383 However, instead of clonal expansion as found in EuSCAPE, in our study we found
384 rapid dissemination of mobile backgrounds supporting bla_{KPC-2}, similar to
385 observations from sequencing of other polyclonal bla_{KPC} scenarios reported
386 elsewhere, including the US(5, 25). Tn*4401*a, associated with high levels of bla_{KPC}
387 expression(26), has been previously predominantly seen in *K. pneumoniae*, and in
388 isolates from the US, Israel and Italy, and similarly most commonly with an ATTGA-
389 ATTGA TSD(9). Thus our findings are consistent with the importation of the
390 predominant bla_{KPC-2}-Tn*4401*a-ATTGA-ATTGA motif into CMFT/North-West
391 England and subsequent horizontal spread. Notably, as in EuSCAPE, 46/72 (64%)

392 singleton isolates we sampled from UK hospitals were also CG258, but our much
393 more detailed sampling across species reflected a very different local/regional
394 molecular epidemiology. Although the EuSCAPE study is large and impressive, its
395 breadth may have been limiting in understanding regional diversity - for example, the
396 subset of bla_{KPC}-*K. pneumoniae* from the UK that were analysed in EuSCAPE
397 consisted of 11 isolates submitted from six centres
398 (https://microreact.org/project/EuSCAPE_UK). The focus was also more on analysing
399 species-specific clonal relationships, with no analysis of other species or MGEs.
400
401 Although in our study diversification occurred at all genetic levels (Tn*4401*+TSSs,
402 plasmids, plasmid populations, strains, species), there was more limited variation
403 observed within the Tn*4401* transposon and its flanking regions, and the spread of
404 bla_{KPC} appears to have been supported by highly plastic modular exchange of larger
405 genetic segments within a distinct plasmid population, particularly IncFIB/IncFII
406 (found in 580 and 545 of the 604 isolates respectively) and IncR replicons (252/604
407 isolates). A previous study, in which 11 transformed bla_{KPC} plasmids from the UK
408 (2008-2010) were sequenced (Roche 454/assembly, PCR+sequencing based gap
409 closure), identified a UK variant of the pKpQIL plasmid, designated pKpQIL-UK
410 ($IncFII_{K2}$ by plasmid MLST), that was highly similar to pKpQIL (maximum 32 SNVs
411 diversity), and several other $IncFII_{K2}$ pKpQIL-like plasmids, but with novel
412 segmental genetic rearrangements (gains/losses; pKpQIL-D1, pKpQIL-D2)(14). Our
413 data support the importance of $IncFII_{K2}$-like plasmids in bla_{KPC} dissemination too, but
414 also that other $IncFII_{K}$-like plasmids (e.g. $IncFII\text{-}_{K1, -K4, -K7, -K15}$) and replicons (IncFIB,
415 IncR) have been a significant feature. In addition to their plasticity, the plasmids
416 identified frequently harboured AMR genes other than bla_{KPC} which might offer a

417 selective advantage, alongside heavy metal resistance genes, and plasmid toxin-
418 antitoxin addiction systems. The plasticity and association of $IncFII_K$ plasmids with
419 resistance genes and IncFIB replicons has been supported by findings of a recent
420 analysis of $IncFII_K$ plasmids(27).
421
422 The problem of accurately classifying plasmid populations from short-read data was
423 exemplified in this analysis, and highlighted by our smaller long-read/short-read
424 hybrid assembly-based analysis, which demonstrated significant diversity within
425 structures assigned as similar by short-read based typing approaches. With this caveat,
426 it was interesting that even with relatively relaxed thresholds, 29% of isolates did not
427 have a match to our reference $\mathit{bla}_{\mathrm{KPC}}$ plasmid database (based on clustering of all
428 publicly available reference sequences, as in Methods), consistent with rapid
429 diversification in the plasmid backgrounds supporting Tn*4401*/$\mathit{bla}_{\mathrm{KPC-2}}$ in this setting.
430
431 Our findings demonstrated that it is also important to consider plasmids without the
432 resistance gene of interest in a population, as these may be relevant to a wider
433 understanding of the transmission and evolution of smaller mobile genetic elements
434 harbouring resistance genes (Fig.9). This was also shown to be relevant in a previous
435 analysis of a large KPC-*E. coli* outbreak in the same setting in 2015-2016, in which a
436 circulating $\mathit{bla}_{\mathrm{KPC}}$-negative plasmid, pCAD3 (IncFIB/FII), acquired Tn*4401* from a
437 IncHI2/HI2A $\mathit{bla}_{\mathrm{KPC}}$-positive plasmid, and went on to dominate within a clonal *E. coli*
438 lineage(22). Most studies in general however tend to focus on analysing AMR
439 plasmids of interest. Fortunately, long-read sequencing is becoming increasingly low
440 cost and high-throughput, and hybrid assembly is able to reconstruct plasmid
441 sequences in Enterobacterales(28, 29). New developments in large-scale comparative

442 genomics of complete genomes, including plasmid structures, are essential for future
443 large-scale analysis of AMR gene transmission.
444
445 There are several limitations to our study. We were only able to long-read sequence a
446 small number of isolates, and the reconstructed genomes generated using long-read
447 PacBio data remained incomplete (49% of all contigs uncircularised). Improvements
448 in long-read technology and assembly approaches will likely overcome this(28). Our
449 short-read and long-read datasets were generated from the same frozen stocks of
450 isolates, but from separate sub-cultures (because we used the short-read data to inform
451 selection for long-read sequencing); ideally they would have been generated from the
452 same DNA extract. PacBio sequencing library preparation incorporates size selection,
453 and this may have led to short plasmid sequences (<15kb) being lost. Our
454 interpretation of the evolution of backgrounds supporting bla_{KPC} was limited by the
455 diversity present, and the inability to capture sequential evolutionary events, even
456 with this large study. We restricted our WGS reference-based plasmid typing to
457 analysing top matches to our bla_{KPC} plasmid database because of the uncertainty in
458 using short-read data for plasmid typing(19); we may therefore have underestimated
459 the diversity of bla_{KPC} plasmids present and missed cases with >1 bla_{KPC} plasmid.
460 Lastly, very limited epidemiological data linked to the isolates were available,
461 meaning that we were unable to ascertain any epidemiological drivers which might be
462 contributing to the enormous heterogeneity of bla_{KPC} transmission over apparently
463 short timeframes; the latter finding also precluded the useful application of standard
464 phylogenetic approaches based on identifying variants core to and within species. In
465 addition, the collection of isolates by PHE as part of regional and national
466 surveillance was dictated by referral patterns of isolates from the hospitals surveyed,

467 and we do not have any denominator information on cultures (either *bla*$_{KPC}$-positive
468 or *bla*$_{KPC}$-negative) to corroborate details on the robustness of this referral process, or
469 to determine what proportion of all UK *bla*$_{KPC}$-positive Enterobacterales over the
470 relevant timeframe we have sequenced.
471
472 In conclusion, our large analysis highlights the difficulty and complexity of bacterial
473 transmission networks once important AMR genes have "escaped" the genetic
474 confines of particular mobile genetic elements and bacterial species/lineages, with
475 important implications for surveillance. These include the need to consider multiple
476 bacterial species and plasmids as potential hosts of *bla*$_{KPC}$, and invest resource in
477 sequencing approaches to adequately reconstruct genetic structures and avoid
478 misinterpreting the molecular epidemiology. It also demonstrates that regional
479 differences in AMR gene epidemiology may be quite marked, which may affect the
480 generalizability of control methods. Finally, it is important to consider the wider
481 genetic background of host strains and plasmids in understanding the evolution and
482 dissemination of important AMR genes, as AMR gene transfer between plasmid
483 backgrounds within bacteria may occur over short timescales, and the interaction of
484 several plasmids (i.e. not just those harbouring the AMR gene of interest at any given
485 time) in a population may be highly relevant to the persistence and dissemination of
486 the AMR gene itself.
487
488 **MATERIAL AND METHODS**
489 *Study isolates and setting*
490 We sequenced archived carbapenem-resistant Enterobacterales isolates from two
491 large teaching hospitals in Manchester (formerly known as CMFT and UHSM),

492 aiming to include all inpatient isolates archived following the observed introduction
493 of *bla*$_{KPC}$-positive Enterobacterales (KPC-E) in this hospital system in 2009-2011. We
494 also sequenced a subset of (KPC-E) isolates archived and sequenced as part of
495 regional and national surveillance of carbapenemases undertaken by Public Health
496 England (PHE, 2009-2014). The PHE set included: (i) a further random set of isolates
497 referred from CMFT/UHSM from 2012-2014; (ii) up to the first 25 consecutive KPC-
498 E isolates from any hospital in North-West England (2009-2014) and referred to the
499 PHE reference laboratory (2009-2014); (ii) the first KPC-E isolate from any other
500 hospital in the UK and Ireland referred to PHE (2009-2014); and, (iii) any KPC-E
501 isolates from outpatient/primary care settings in the UK referred to PHE (2009-2014).
502 For the UHSM/CMFT isolate subset, we were able to determine sampling and clinical
503 sample culture-positivity denominators from an anonymised database of linked
504 electronic bacteriology and patient administration records going back to 2009(22).
505
506 Ethical approval was not required as only bacterial isolates were sequenced, and their
507 collection was part of infection control investigation and management.
508
509 *DNA extraction and sequencing*
510 For short-read Illumina sequencing (HiSeq 2500, 150bp PE reads), DNA was
511 extracted using Quickgene (Fujifilm, Japan), with an additional mechanical lysis step
512 following chemical lysis (FastPrep, MP Biomedicals, USA). Sequencing libraries
513 were constructed using the NEBNext Ultra DNA Sample Prep Master Mix Kit (NEB)
514 with minor modifications and a custom automated protocol on a Biomek FX
515 (Beckman). Ligation of adapters was performed using Illumina Multiplex Adapters,
516 and ligated libraries were size-selected using Ampure magnetic beads (Agencourt).

517 Each library was PCR-enriched with custom primers (Index primer plus dual index
518 PCR primer) (30). Enrichment and adapter extension of each preparation was
519 obtained using 9ul of size-selected library in a 50ul PCR reaction. Reactions were
520 then purified with Ampure beads (Agencourt/Beckman) on a Biomek NXp after 10
521 cycles of amplification (as per Illumina recommendations). Final size distributions of
522 libraries were determined using a Tapestation 1DK system (Agilent/Lab901), and
523 quantified by Qubit fluorometry (Thermofisher).
524
525 For long-read sequencing (PacBio [n=28]), DNA was extracted using the Qiagen
526 Genomic tip 100/G kit (Qiagen, Netherlands). DNA extracts were initially sheared to
527 an average length of 15kb using g-tubes, as specified by the manufacturer (Covaris).
528 Sheared DNA was used in SMRTbell library preparation, as recommended by the
529 manufacturer. Quantity and quality of the SMRTbell libraries were evaluated using
530 the High Sensitivity dsDNA kit and Qubit Fluorimeter (Thermo Fisher Scientific) and
531 DNA 12000 kit on the 2100 Bioanalyzer (Agilent). To obtain the longest possible
532 SMRTbell libraries for sequencing (as recommended by the manufacturer), a further
533 size selection step was performed using the PippinHT pulsed-field gel electrophoresis
534 system (Sage Science), enriching for the SMRTbell libraries >15kb for loading onto
535 the instrument. Sequencing primer and P6 polymerase were annealed and bound to
536 the SMRTbell libraries, and each library was sequenced using a single SMRT cell on
537 the PacBio RSII sequencing system.
538
539 *Sequence data processing and assembly*
540 We used Kraken(31) to assign species to sequenced isolates from short read
541 (Illumina) data. SPAdes(32) v3.6 was used to *de novo* assemble short reads (default

542 options; subsequent removal of contigs shorter than 500bp and assembly coverage
543 <2X). Isolates with short read sequence assemblies >6.5Mb were excluded to ensure
544 that potentially mixed sequences were not included in the analyses. MLST was
545 derived in silico from short read assemblies by blasting de novo these against publicly
546 available MLST databases for *E. coli* (http://mlst.warwick.ac.uk/mlst/dbs/Ecoli), *K.*
547 *pneumoniae, E. cloacae and K. oxytoca* (https://pubmlst.org/). Isolates with mixed
548 MLST outputs were excluded. Antimicrobial resistance (AMR) genes, plasmid
549 replicon (Inc) types and insertion sequences (IS) were identified from short read data
550 using resistType (https://github.com/hangphan/resistType_docker; curated AMR gene
551 database as in(33), plasmid replicon reference sequences from PlasmidFinder(34), and
552 ISs from the ISFinder platform(35); ≥80% identity used as a threshold). *bla*$_{KPC}$ copy
553 number (per bacterial genome) for each isolate was estimated from short read
554 assemblies by dividing coverage of the contig containing *bla*$_{KPC}$ by the average
555 coverage for the assembly (weighted by contig length). Confirmation that this was a
556 biologically meaningful estimate was obtained by estimating the association between
557 copy number estimates and meropenem minimum inhibitory concentrations (MICs)
558 from routine clinical laboratory antimicrobial susceptibility testing using interval
559 regression (outcome $\log_2$(MIC), left and right censored to reflect the actual MIC in
560 the extreme categories being unobserved but only within a range) as the dependent
561 variable and KPC copy number as the independent variable. *bla*$_{KPC}$ copy number
562 estimates were also validated by mapping Illumina reads to the reconstructed
563 genomes derived from the PacBio data (using bwa-mem(36) [bwa-0.7.12-r1034], and
564 only where both chromosome and *bla*$_{KPC}$ plasmid structures were deemed complete),
565 and then by calculating mean coverage for the *bla*$_{KPC}$ gene versus the chromosomal
566 contig, and comparing this with the *bla*$_{KPC}$ copy number estimate derived from the

567 short-read assembly described above. Plasmid MLST for common family types
568 identified in short read data was confirmed by 100% sequence matches to reference
569 alleles for families catalogued in the plasmidMLST website
570 (https://pubmlst.org/plasmid/; IncA/C, IncHI1/2, IncN).
571
572 For the smaller subset of isolates on which long read (i.e. PacBio) sequencing was
573 performed, long read data were assembled using the HGAP pipeline(37), and polished
574 with the corresponding Illumina datasets using Pilon (version 1.18, default
575 parameters)(38). Chromosomal sequences and plasmid sequences were then manually
576 curated where possible to create complete, closed, circular structures by using
577 BLASTn to identify overlaps at the end of assembled contigs. Those with overlapping
578 ends larger than 1000bp with sequence identity >99% were considered
579 circularised/complete, and trimmed appropriately for resolution. Complete sequences
580 were annotated using PROKKA (version 1.11)(39); annotations were used to
581 determine genes known to encode toxin-antitoxin systems, heavy metal resistance,
582 and anti-restriction mechanisms. Plasmid MLST was confirmed for these assemblies
583 as above.
584
585 *Tn4401 typing*
586 Tn*4401* typing was performed using TETyper(9), using the Tn*4401*, SNP and
587 structural profile reference files included with the package
588 (https://github.com/aesheppard/TETyper, version 1.1), and a flanking length of 5bp,
589 representative of the known target site signature sequence indicative of Tn*4401*
590 transposition(40).
591

592 *Plasmid database for bla*$_{KPC}$ *plasmid typing*

593 A reference *bla*$_{KPC}$ plasmid sequence database composed of *bla*$_{KPC}$-harbouring

594 contigs/plasmids from long-read sequencing of isolates in this study and all complete

595 *bla*$_{KPC}$ plasmids from (41-43) (August 2018) was used for *bla*$_{KPC}$ plasmid typing

596 within this study. To construct this database, all 279/6018 evaluable plasmid

597 sequences carrying *bla*$_{KPC}$ were first compared using *dnadiff*(44) to obtain the

598 pairwise similarity between any two plasmid sequences p_i and p_j. The similarity was

599 defined as a function of their lengths l_i, l_j, and the aligned bases l_{ij}, l_{ji} as reported by:

$$(p_i, p_j) = \frac{1}{2}\left(\frac{l_{ij}}{l_i} + \frac{l_{ji}}{l_j}\right) \times \min\left(\frac{l_i}{l_j}, \frac{l_j}{l_i}\right)$$

600

601 The score was designed to penalise differences in length of the compared sequences,

602 i.e. to make sequences of different lengths proportionately more different. The

603 resulting similarity matrix was used to perform clustering of plasmid sequences using

604 the affinity propagation clustering technique, suitable for graph clustering problems

605 with sparse similarity matrix and uneven cluster size and cluster number(45), and

606 resulted in 34 clusters of 1-43 plasmids per cluster (Supplemental Dataset SD4). The

607 largest cluster was the set of pKpQIL-like plasmids comprising 43 related sequences.

608 Representative sequences of each *bla*$_{KPC}$ plasmid cluster in this network were chosen

609 randomly, to generate a set (*KPC-pDB*) of plasmids ranging from 7,995bp

610 (NC_022345.1; plasmid pAP-2) to 447,095bp (NZ_CP029436.1; plasmid

611 pKPC_CAV2013) in the final database used for *bla*$_{KPC}$ plasmid typing in this study.

612

613 Subsequently, *bla*$_{KPC}$ plasmid typing for each study isolate sequence was performed

614 as follows: (1) assembled sequences for each isolate were BLASTed (BLASTn)

615 against *KPC-pDB*; (2) any >1kb contig with >90% nucleotide identity and >80% total

616 coverage match to sequences in *KPC-pDB* was retained; (3) for any sequence p_i in
617 *KPC-pDB*, a score s_i was calculated by dividing the total matched bases of all contigs
618 matched to p_i by p_i's length; and (4) an isolate was assumed to plausibly carry p_i if s_i
619 $\geq$0.80. An isolate could have several *bla*$_{KPC}$ plasmid matches; we restricted to the top
620 match for each isolate in our analyses.
621
622 *Phylogenetic reconstruction for isolates from commonly represented STs*
623 To ensure that for the most common STs (*K. pneumoniae* STs 258, 11 and 491) we
624 were essentially not characterising a single, highly clonal set of sequences,
625 recombination-corrected phylogenies were reconstructed for each ST following
626 mapping to the *K. pneumoniae* reference MGH78578 (GenBank accession:
627 CP000647.1), as previously described(46). In brief, following read mapping, variant
628 calling, and generation of consensus fasta sequences of variants called, IQtree
629 followed by ClonalFrameML were run using a python wrapper available at:
630 https://github.com/davideyre/runListCompare (accessed Jan 2020).
631
632 *Statistical analysis and data visualisation*
633 Statistical analysis (Kruskal-Wallis tests, interval regression, Pearson's correlation)
634 was carried out in Stata 14.2. Plots for figures 1-7, and supplementary figures were
635 generated using the ggplot2 and ape packages in R (version 1.1.463). Figure 9 was
636 generated using the GenomeDiagram package(47) in Biopython(48).
637
638 *Sequencing data availability*
639 Sequencing data have been deposited in the NCBI (BioProject Accession:
640 PRJNA564424). PacBio/Pilon assemblies are available at:

641 https://doi.org/10.6084/m9.figshare.11777631.v1. Typing results and metadata for
642 each isolate are available in Supplemental dataset SD1.
643
644
645 **Acknowledgements and funding:** We are grateful to and acknowledge the sharing of
646 isolates by microbiology and clinical teams from contributing UK hospitals, and from
647 Martin Cormican, the National Reference Laboratory in Galway, Ireland, and the
648 contributing laboratories in Ireland. We are also grateful to the microbiology
649 laboratory staff and infection control teams at Manchester University NHS
650 Foundation Trust (formerly CMFT and UHSM); the staff of the Manchester Medical
651 Microbiology Partnership; and the research laboratory, informatics and project
652 management teams working as part of the Modernising Medical Microbiology
653 consortium, Oxford.
654
655 Contemporaneous investigation by CMFT, UHSM and PHE was undertaken as part
656 of routine activity. The retrospective investigation was funded by the National
657 Institute for Health Research Health Protection Research Unit (NIHR HPRU) in
658 Healthcare Associated Infections and Antimicrobial Resistance at Oxford University
659 in partnership with Public Health England (PHE) [grant HPRU-2012-10041] and
660 supported by the NIHR Biomedical Research Centre, Oxford. The views expressed in
661 this publication are those of the authors and not necessarily those of the NHS, the
662 National Institute for Health Research, the Department of Health or Public Health
663 England. NS is funded by a PHE/University of Oxford Academic Clinical
664 Lectureship. TEAP, DWC and ASW are NIHR Senior Investigators.
665

666 The Transmission of Carbapenemase-producing Enterobacteriaceae (TRACE) study
667 investigators are listed alphabetically, and include several of the authors also listed by
668 name in the main author list: Zoie Aiken, Oluwafemi Akinremi, Aiysha Ali, Julie
669 Cawthorne, Paul Cleary, Derrick W. Crook, Valerie Decraene, Andrew Dodgson,
670 Michel Doumith, Matthew J. Ellington, Ryan George, John Grimshaw, Malcolm
671 Guiver, Robert Hill, Katie L. Hopkins, Rachel Jones, Cheryl Lenney, Amy J.
672 Mathers, Ashley McEwan, Ginny Moore, Mark Neilson, Sarah Neilson, Tim E.A.
673 Peto, Hang T.T. Phan, Mark Regan, Anna C. Seale, Nicole Stoesser, Jay Turner-
674 Gardner, Vicky Watts, A. Sarah Walker, Jimmy Walker, David Wyllie, William
675 Welfare and Neil Woodford.

676

677 None of the authors has any conflicts of interest to declare.

678 **References**

679 1. Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, Didelot
680 X, Turner SD, Sebra R, Kasarskis A, Peto T, Crook D, Sifri CD. 2015.
681 Klebsiella pneumoniae Carbapenemase (KPC)-Producing K-pneumoniae at a
682 Single Institution: Insights into Endemicity from Whole-Genome Sequencing.
683 Antimicrobial Agents and Chemotherapy 59:1661-1668.
684 2. Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M,
685 Chapman SB, Reis-Cunha JL, Shea TP, Young S, Zeng Q, Delaney ML, Kim
686 D, Peterson EM, O'Brien TF, Ferraro MJ, Hooper DC, Huang SS, Kirby JE,
687 Onderdonk AB, Birren BW, Hung DT, Cosimi LA, Wortman JR, Murphy CI,
688 Hanage WP. 2017. Multi-institute analysis of carbapenem resistance reveals
689 remarkable diversity, unexplained mechanisms, and limited clonal outbreaks.
690 Proc Natl Acad Sci U S A 114:1135-1140.

691 3. Ruiz-Garbajosa P, Curiao T, Tato M, Gijon D, Pintado V, Valverde A,
692 Baquero F, Morosini MI, Coque TM, Canton R. 2013. Multiclonal dispersal of
693 KPC genes following the emergence of non-ST258 KPC-producing Klebsiella
694 pneumoniae clones in Madrid, Spain. J Antimicrob Chemother 68:2487-92.
695 4. Martin J, Phan H, Findlay J, Stoesser N, Pankhurst L, Navickaite I, De Maio
696 N, Eyre D, Toogood G, Orsi N, Kirby A, Young N, Turton J, Hill R, Hopkins
697 K, Woodford N, Peto T, Walker A, Crook D, Wilcox M. 2017. Covert
698 dissemination of carbapenemase-producing Klebsiella pneumoniae (KPC) in a
699 successfully controlled outbreak: long and short-read whole-genome
700 sequencing demonstrate multiple genetic modes of transmission. Journal of
701 Antimicrobial Chemotherapy.
702 5. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW,
703 Giess A, Pankhurst LJ, Vaughan A, Grim CJ, Cox HL, Yeh AJ, Modernising
704 Medical Microbiology Informatics G, Sifri CD, Walker AS, Peto TE, Crook
705 DW, Mathers AJ. 2016. Nested Russian Doll-Like Genetic Mobility Drives
706 Rapid Dissemination of the Carbapenem Resistance Gene blaKPC.
707 Antimicrob Agents Chemother 60:3767-78.
708 6. Quainoo S, Coolen JPM, van Hijum S, Huynen MA, Melchers WJG, van
709 Schaik W, Wertheim HFL. 2017. Whole-Genome Sequencing of Bacterial
710 Pathogens: the Future of Nosocomial Outbreak Analysis. Clin Microbiol Rev
711 30:1015-1063.
712 7. Walker TM, Monk P, Smith EG, Peto TE. 2013. Contact investigations for
713 outbreaks of Mycobacterium tuberculosis: advances through whole genome
714 sequencing. Clin Microbiol Infect 19:796-802.

715 8. Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican
716 M, Cornaglia G, Garau J, Gniadkowski M, Hayden MK, Kumarasamy K,
717 Livermore DM, Maya JJ, Nordmann P, Patel JB, Paterson DL, Pitout J,
718 Villegas MV, Wang H, Woodford N, Quinn JP. 2013. Clinical epidemiology
719 of the global expansion of Klebsiella pneumoniae carbapenemases. Lancet
720 Infect Dis 13:785-96.
721 9. Sheppard AE, Stoesser N, German-Mesner I, Vegesana K, Sarah Walker A,
722 Crook DW, Mathers AJ. 2018. TETyper: a bioinformatic pipeline for
723 classifying variation and genetic contexts of transposable elements from short-
724 read whole-genome sequencing data. Microb Genom 4.
725 10. Palepou MFW, N.; Hope, R.; Colman, M.; Glover, J.; Kaufmann, M.; Lafong,
726 C.; Reynolds, R.; Livermore, D. M. 2005. Novel class A carbapenemase,
727 KPC-4, in an Enterobacter isolate from Scotland, abstr abstr. 1134_01_20.
728 Prog. Abstr. 15th Eur. Cong. Clin. Microbiol. Infect. Dis., Copenhagen,
729 Denmark.
730 11. Public Health England. 2011. Carbapenemase-producing Enterobacteriaceae:
731 laboratory confirmed cases, 2003 to 2013.
732 https://www.gov.uk/government/publications/carbapenemase-
733 producing-enterobacteriaceae-laboratory-confirmed-
734 cases/carbapenemase-producing-enterobacteriaceae-laboratory-
735 confirmed-cases-2003-to-2013. Accessed 02/09/2016.
736 12. Donker T, Henderson KL, Hopkins KL, Dodgson AR, Thomas S, Crook DW,
737 Peto TEA, Johnson AP, Woodford N, Walker AS, Robotham JV. 2017. The
738 relative importance of large problems far away versus small problems closer

739 to home: insights into limiting the spread of antimicrobial resistance in
740 England. BMC Med 15:86.
741 13. Findlay J, Hopkins KL, Doumith M, Meunier D, Wiuff C, Hill R, Pike R, Loy
742 R, Mustafa N, Livermore DM, Woodford N. 2016. KPC enzymes in the UK:
743 an analysis of the first 160 cases outside the North-West region. J Antimicrob
744 Chemother 71:1199-206.
745 14. Doumith M, Findlay J, Hirani H, Hopkins KL, Livermore DM, Dodgson A,
746 Woodford N. 2017. Major role of pKpQIL-like plasmids in the early
747 dissemination of KPC-type carbapenemases in the UK. J Antimicrob
748 Chemother 72:2241-2248.
749 15. Qi Y, Wei Z, Ji S, Du X, Shen P, Yu Y. 2011. ST11, the dominant clone of
750 KPC-producing Klebsiella pneumoniae in China. J Antimicrob Chemother
751 66:307-12.
752 16. Cuzon G, Naas T, Nordmann P. 2011. Functional characterization of Tn4401,
753 a Tn3-based transposon involved in blaKPC gene mobilization. Antimicrob
754 Agents Chemother 55:5370-3.
755 17. Cuzon G, Naas T, Truong H, Villegas MV, Wisell KT, Carmeli Y, Gales AC,
756 Venezia SN, Quinn JP, Nordmann P. 2010. Worldwide diversity of Klebsiella
757 pneumoniae that produce beta-lactamase blaKPC-2 gene. Emerg Infect Dis
758 16:1349-56.
759 18. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, Dyda F.
760 2015. Insertion Sequence IS26 Reorganizes Plasmids in Clinically
761 Isolated Multidrug-Resistant Bacteria by Replicative Transposition. mBio
762 6:e00762-15.

763 19. Arredondo-Alonso S, Willems RJ, van Schaik W, Schurch AC. 2017. On the
764 (im)possibility of reconstructing plasmids from whole-genome short-read
765 sequencing data. Microb Genom 3:e000128.
766 20. Liang W, Xie Y, Xiong W, Tang Y, Li G, Jiang X, Lu Y. 2017. Anti-
767 Restriction Protein, KlcAHS, Promotes Dissemination of Carbapenem
768 Resistance. Front Cell Infect Microbiol 7:150.
769 21. Public Health England. 2018. English Surveillance Programme for
770 Antimicrobial Utilisation and Resistance (ESPAUR), Report 2018.
771 22. Decraene V, Phan HTT, George R, Wyllie DH, Akinremi O, Aiken Z, Cleary
772 P, Dodgson A, Pankhurst L, Crook DW, Lenney C, Walker AS, Woodford N,
773 Sebra R, Fath-Ordoubadi F, Mathers AJ, Seale AC, Guiver M, McEwan A,
774 Watts V, Welfare W, Stoesser N, Cawthorne J, Group TI. 2018. A Large,
775 Refractory Nosocomial Outbreak of Klebsiella pneumoniae Carbapenemase-
776 Producing Escherichia coli Demonstrates Carbapenemase Gene Outbreaks
777 Involving Sink Sites Require Novel Approaches to Infection Control.
778 Antimicrob Agents Chemother 62.
779 23. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab
780 K, Goater R, Giani T, Errico G, Aspbury M, Sjunnebo S, Eu SWG, Group ES,
781 Feil EJ, Rossolini GM, Aanensen DM, Grundmann H. 2019. Epidemic of
782 carbapenem-resistant Klebsiella pneumoniae in Europe is driven by
783 nosocomial spread. Nat Microbiol doi:10.1038/s41564-019-0492-8.
784 24. Sherry NL, Lane CR, Kwong JC, Schultz M, Sait M, Stevens K, Ballard S,
785 Goncalves da Silva A, Seemann T, Gorrie CL, Stinear TP, Williamson DA,
786 Brett J, van Diemen A, Easton M, Howden BP. 2019. Genomics for Molecular

787 Epidemiology and Detecting Transmission of Carbapenemase-Producing
788 Enterobacterales in Victoria, Australia, 2012 to 2016. J Clin Microbiol 57.
789 25. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, Lau AF,
790 Khil P, Odom RT, Deming C, Park M, Thomas PJ, Program NCS, Henderson
791 DK, Palmore TN, Segre JA, Frank KM. 2018. Genomic Analysis of Hospital
792 Plumbing Reveals Diverse Reservoir of Bacterial Plasmids Conferring
793 Carbapenem Resistance. MBio 9.
794 26. Cheruvanky A, Stoesser N, Sheppard AE, Crook DW, Hoffman PS, Weddle
795 E, Carroll J, Sifri CD, Chai W, Barry K, Ramakrishnan G, Mathers AJ. 2017.
796 Enhanced Klebsiella pneumoniae Carbapenemase Expression from a Novel
797 Tn4401 Deletion. Antimicrob Agents Chemother 61.
798 27. Bi D, Zheng J, Li JJ, Sheng ZK, Zhu X, Ou HY, Li Q, Wei Q. 2018. In Silico
799 Typing and Comparative Genomic Analysis of IncFIIK Plasmids and Insights
800 into the Evolution of Replicons, Plasmid Backbones, and Resistance
801 Determinant Profiles. Antimicrob Agents Chemother 62.
802 28. De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, Wick
803 R, AbuOun M, Stubberfield E, Hoosdally SJ, Crook DW, Peto TEA, Sheppard
804 AE, Bailey MJ, Read DS, Anjum MF, Walker AS, Stoesser N, On Behalf Of
805 The Rehab C. 2019. Comparison of long-read sequencing technologies in the
806 hybrid assembly of complex bacterial genomes. Microb Genom
807 doi:10.1099/mgen.0.000294.
808 29. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome
809 assemblies with multiplex MinION sequencing. Microb Genom 3:e000132.
810 30. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-
811 Fahmawi B, Piazza P. 2013. Improved workflows for high throughput library

812 preparation using the transposome-based nextera system. BMC Biotechnology
813 13:104.
814 31. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence
815 classification using exact alignments. Genome Biol 15:R46.
816 32. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS,
817 Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV,
818 Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new
819 genome assembly algorithm and its applications to single-cell sequencing. J
820 Comput Biol 19:455-77.
821 33. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C,
822 Johnson JR, Walker AS, Peto TE, Crook DW. 2013. Predicting antimicrobial
823 susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using
824 whole genomic sequence data. J Antimicrob Chemother 68:2234-44.
825 34. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa
826 L, Moller Aarestrup F, Hasman H. 2014. In silico detection and typing of
827 plasmids using PlasmidFinder and plasmid multilocus sequence typing.
828 Antimicrob Agents Chemother 58:3895-903.
829 35. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the
830 reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-
831 6.
832 36. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-
833 Wheeler transform. Bioinformatics 25:1754-60.
834 37. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A,
835 Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013.

836 Nonhybrid, finished microbial genome assemblies from long-read SMRT
837 sequencing data. Nat Methods 10:563-9.
838 38. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo
839 CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool
840 for comprehensive microbial variant detection and genome assembly
841 improvement. PLoS One 9:e112963.
842 39. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation.
843 Bioinformatics 30:2068-9.
844 40. Naas T, Cuzon G, Villegas MV, Lartigue MF, Quinn JP, Nordmann P. 2008.
845 Genetic structures at the origin of acquisition of the beta-lactamase bla KPC
846 gene. Antimicrob Agents Chemother 52:1257-63.
847 41. Villa L, Feudi C, Fortini D, Brisse S, Passet V, Bonura C, Endimiani A,
848 Mammina C, Ocampo AM, Jimenez JN, Doumith M, Woodford N, Hopkins
849 K, Carattoli A. 2017. Diversity, virulence, and antimicrobial resistance of the
850 KPC-producing Klebsiella pneumoniae ST307 clone. Microbial Genomics 3.
851 42. Orlek A, Phan H, Sheppard AE, Doumith M, Ellington M, Peto T, Crook D,
852 Walker AS, Woodford N, Anjum MF, Stoesser N. 2017. A curated dataset of
853 complete Enterobacteriaceae plasmids compiled from the NCBI nucleotide
854 database. Data in Brief 12:423-426.
855 43. Stoesser N, Sheppard AE, Peirano G, Anson LW, Pankhurst L, Sebra R, Phan
856 HTT, Kasarskis A, Mathers AJ, Peto TEA, Bradford P, Motyl MR, Walker
857 AS, Crook DW, Pitout JD. 2017. Genomic epidemiology of global Klebsiella
858 pneumoniae carbapenemase (KPC)-producing Escherichia coli. Sci Rep
859 7:5917.

860 44. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C,
861 Salzberg SL. 2004. Versatile and open software for comparing large genomes.
862 Genome Biol 5:R12.
863 45. Frey BJ, Dueck D. 2007. Clustering by passing messages between data points.
864 Science 315:972-6.
865 46. Kizny Gordon A, Phan HTT, Lipworth SI, Cheong E, Gottlieb T, George S,
866 Peto TEA, Mathers AJ, Walker AS, Crook DW, Stoesser N. 2020. Genomic
867 dynamics of species and mobile genetic elements in a prolonged blaIMP-4-
868 associated carbapenemase outbreak in an Australian hospital. J Antimicrob
869 Chemother doi:10.1093/jac/dkz526.
870 47. Pritchard L, White JA, Birch PR, Toth IK. 2006. GenomeDiagram: a python
871 package for the visualization of large-scale genomic data. Bioinformatics
872 22:616-7.
873 48. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I,
874 Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. 2009. Biopython: freely
875 available Python tools for computational molecular biology and
876 bioinformatics. Bioinformatics 25:1422-3.
877

878 **Figure legends**

879 **Figure 1. Estimated *bla*$_{KPC}$ copy number distributions (derived from Illumina**
880 **assemblies) within major species (Fig.1A), and the top nineteen commonest**
881 **species/ST combinations (Fig.1B) observed within the study (other ST/species**
882 **combinations assigned as "Other" or "Other species/ST" respectively).** Dots
883 represent estimated copy number for single isolates; boxplots represent median
884 estimated *bla*$_{KPC}$ copy number +/- 1.58*IQR/sqrt(n). Boxplots are ordered by most
885 common species and species/ST categories, left-to-right, except for the "Other",
886 "Other species/ST", assigned to the right of the plots. For species assignations,
887 "Kpne" = *Klebsiella pneumoniae*, "Ecol" = *Escherichia coli*, and "Eclo" =
888 *Enterobacter cloacae*.
889

890 **Figure 2. Incidence plot of species-ST by year-month and geography.** Dots are
891 coloured by location of isolate collection, as defined in Methods, and scaled by the
892 number of isolates; the eight out-patient isolates have not been plotted.
893

894 **Figure 3. Incidence curve of species-ST in Manchester hospitals CMFT/UHSM,**
895 **2010-2012.** Sequencing ascertainment of first-per-patient carbapenem-resistant
896 Enterobacterales was 76% 2010 and 59% in 2011.
897

898 **Figure 4. Incidence plot of Tn*4401* type and target site sequences by year-month**
899 **and geography.** Dots are coloured by location of isolate collection, as defined in
900 Methods, and scaled by the number of isolates; the eight out-patient isolates have not
901 been plotted.
902

903 **Figure 5. Incidence plot of plasmid populations identified in isolates (replicon**
904 **typing) by year-month and geography.** Dots are coloured by location of isolate
905 collection, as defined in Methods, and scaled by the number of isolates; the eight out-
906 patient isolates have not been plotted. The most predominant combinations are
907 highlighted in yellow.
908

909 **Figure 6. Distribution of *bla*$_{KPC}$ plasmid types by species-ST.** Dots are coloured by
910 Tn4401/target site sequence type, and scaled by the number of isolates.
911

912 **Figure 7. Incidence plot of *bla*$_{KPC}$ plasmid types identified by year-month and**
913 **geography.** Dots are coloured by location of isolate collection, as defined in Methods,
914 and scaled by the number of isolates; the eight out-patient isolates have not been
915 plotted.
916

917 **Figure 8.** Schematic of *bla*$_{KPC}$ plasmid types and sizes identified from long-
918 read/short-read hybrid sequencing approach by species/ST and year of collection (NB
919 only 21 contigs clearly designated as plasmid are represented). Closed circles denote
920 circularised contigs (i.e. complete plasmids); replicons are denoted by coloured
921 triangles in their approximate positions in the structure. Triangle colours denote
922 replicon types assigned to each plasmid sequence (i.e. multiple coloured triangles
923 represent multi-replicon plasmids). Plasmids from isolates from the wider UK
924 collection (i.e. collected through the national reference laboratory) are denoted with a
925 "*".
926

927 **Figure 9**. Alignments of complete plasmid sequences harbouring an

928 IncFII(K)_1_CP000648-like replicon, including *bla*$_{KPC}$-negative and *bla*$_{KPC}$-positive

929 sequences. All sequences were re-orientated to start at IncFII for the purposes of

930 alignment visualization. Loci of interest have been coloured and annotated as shown.

931 Shading between sequences denotes regions of homology, with light pink shading

932 denoting areas ≥90% nucleotide identity, dark pink areas ≥50% nucleotide identity,

933 and light blue areas ≥90% nucleotide identity in reverse orientation. The order of

934 sequences is adjusted to highlight genetic overlap between sequences, but not to

935 imply any specific direct exchange events. Annotations (i) and (ii) denote specific

936 features highlighted in the main text.

Estimated blaKPC copy number

0 1 2 3 4 5 6 7 8

Kpne Ecol Eclo Other

Species

Estimated blaKPC copy number

0 1 2 3 4 5 6 7 8

Kpne−ST−258 Kpne−ST−11 Kpne−ST−491 Kpne−ST−1162 Kpne−ST−54 Kpne−ST−588 Kpne−ST−2938 Kpne−ST−321 Kpne−ST−1828 Kpne−ST−492 Kpne−ST−468 Kpne−ST−48 Kpne−ST−252 Kpne−ST−34 Kpne−ST−12 Ecol−ST−131 Kpne−ST−107 Kpne−ST−25 Kpne−ST−37 Other species/ST

Species and MLST

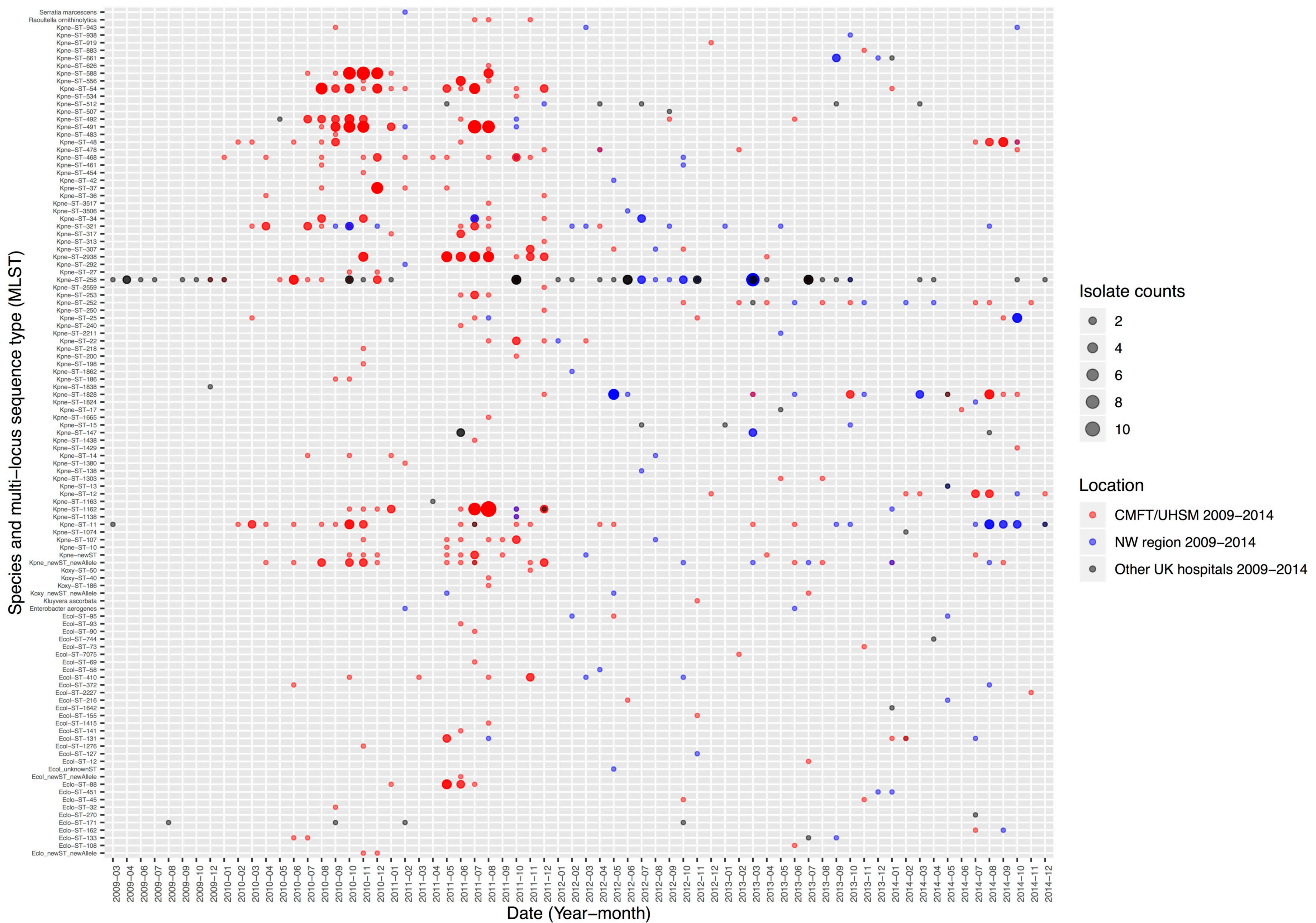
Species

Kpne

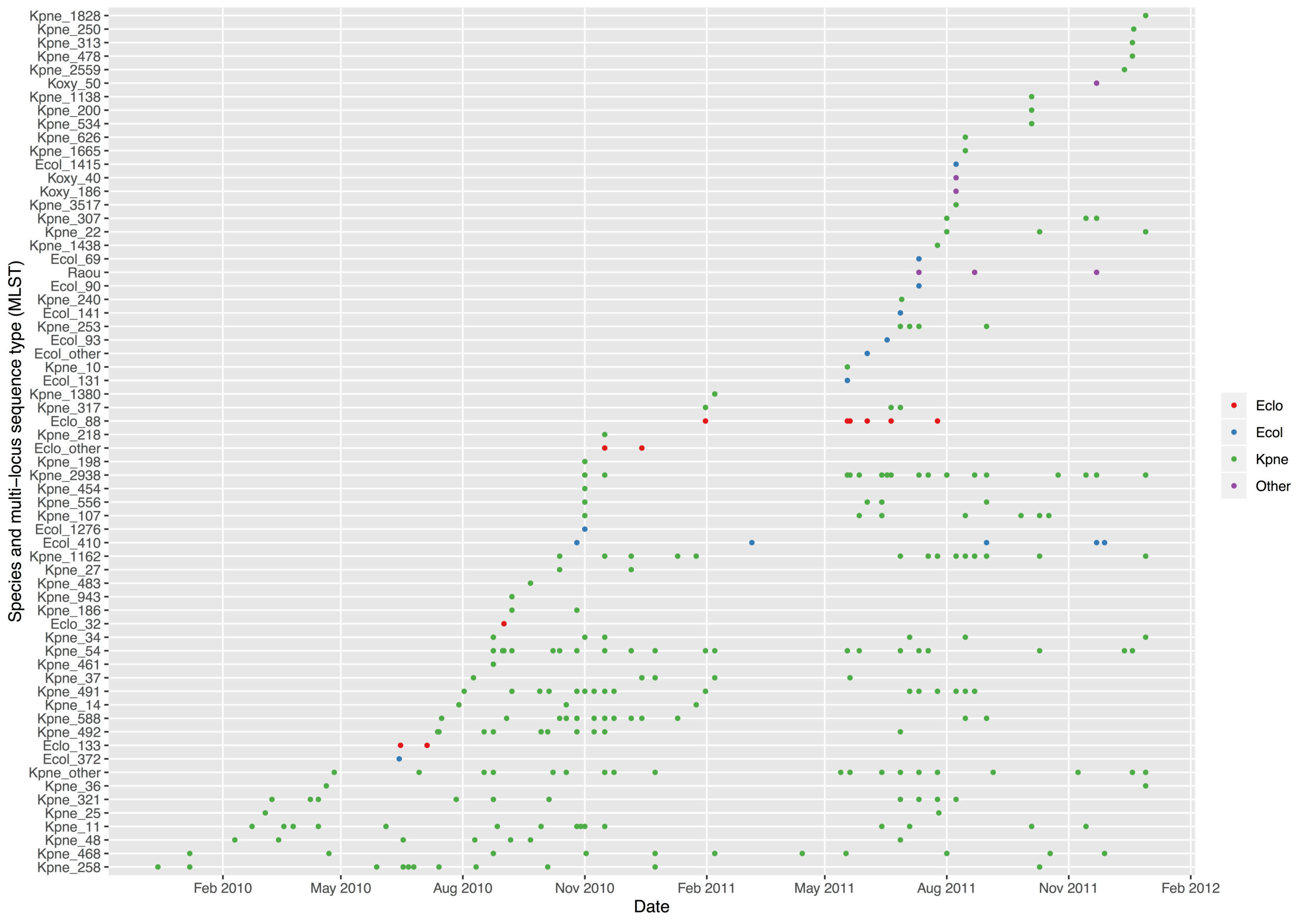
Ecol

Eclo

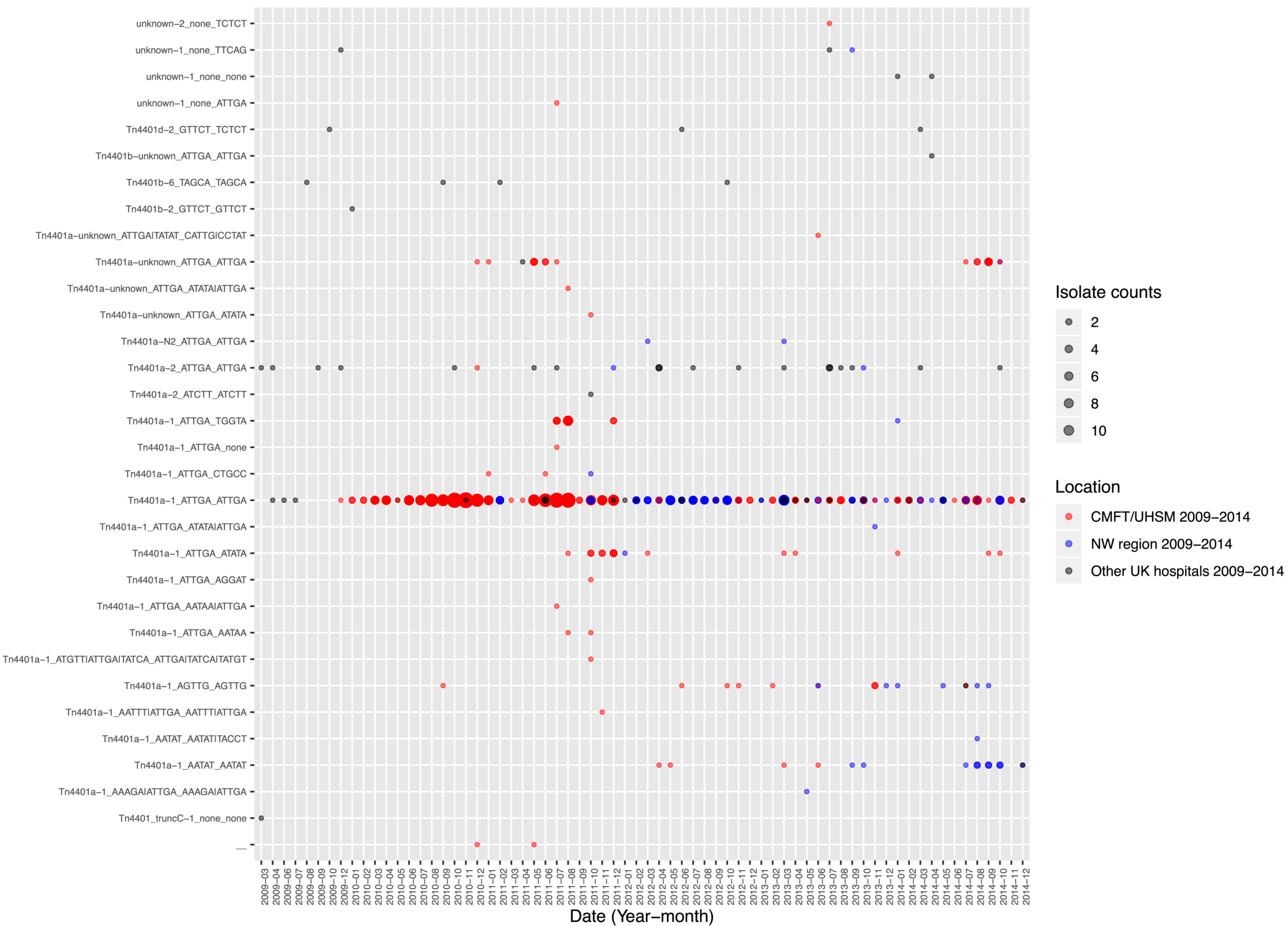
Other



Species and multi–locus sequence type (MLST)
Serratia marcescens
Raoultella ornithinolytica
Kpne–ST–943
Kpne–ST–938
Kpne–ST–919
Kpne–ST–883
Kpne–ST–661
Kpne–ST–626
Kpne–ST–588
Kpne–ST–556
Kpne–ST–54
Kpne–ST–534
Kpne–ST–512
Kpne–ST–507
Kpne–ST–492
Kpne–ST–491
Kpne–ST–483
Kpne–ST–48
Kpne–ST–478
Kpne–ST–468
Kpne–ST–461
Kpne–ST–454
Kpne–ST–42
Kpne–ST–37
Kpne–ST–36
Kpne–ST–3517
Kpne–ST–3506
Kpne–ST–34
Kpne–ST–321
Kpne–ST–317
Kpne–ST–313
Kpne–ST–307
Kpne–ST–2938
Kpne–ST–292
Kpne–ST–27
Kpne–ST–258
Kpne–ST–2559
Kpne–ST–253
Kpne–ST–252
Kpne–ST–250
Kpne–ST–25
Kpne–ST–240
Kpne–ST–2211
Kpne–ST–22
Kpne–ST–218
Kpne–ST–200
Kpne–ST–198
Kpne–ST–1862
Kpne–ST–186
Kpne–ST–1838
Kpne–ST–1828
Kpne–ST–1824
Kpne–ST–17
Kpne–ST–1665
Kpne–ST–15
Kpne–ST–147
Kpne–ST–1438
Kpne–ST–1429
Kpne–ST–14
Kpne–ST–1380
Kpne–ST–138
Kpne–ST–1303
Kpne–ST–13
Kpne–ST–12
Kpne–ST–1163
Kpne–ST–1162
Kpne–ST–1138
Kpne–ST–11
Kpne–ST–1074
Kpne–ST–107
Kpne–ST–10
Kpne–newST
Kpne_newST_newAllele
Koxy–ST–50
Koxy–ST–40
Koxy–ST–186
Koxy_newST_newAllele
Kluyvera ascorbata
Enterobacter aerogenes
Ecol–ST–95
Ecol–ST–93
Ecol–ST–90
Ecol–ST–744
Ecol–ST–73
Ecol–ST–7075
Ecol–ST–69
Ecol–ST–58
Ecol–ST–410
Ecol–ST–372
Ecol–ST–2227
Ecol–ST–216
Ecol–ST–1642
Ecol–ST–155
Ecol–ST–1415
Ecol–ST–141
Ecol–ST–131
Ecol–ST–1276
Ecol–ST–127
Ecol–ST–12
Ecol_unknownST
Ecol_newST_newAllele
Eclo–ST–88
Eclo–ST–451
Eclo–ST–45
Eclo–ST–32
Eclo–ST–270
Eclo–ST–171
Eclo–ST–162
Eclo–ST–133
Eclo–ST–108
Eclo_newST_newAllele
2009–03
2009–04
2009–06
2009–07
2009–08
2009–09
2009–10
2009–12
2010–01
2010–02
2010–03
2010–04
2010–05
2010–06
2010–07
2010–08
2010–09
2010–10
2010–11
2010–12
2011–01
2011–02
2011–03
2011–04
2011–05
2011–06
2011–07
2011–08
2011–09
2011–10
2011–11
2011–12
2012–01
2012–02
2012–03
2012–04
2012–05
2012–06
2012–07
2012–08
2012–09
2012–10
2012–11
2012–12
2013–01
2013–02
2013–03
2013–04
2013–05
2013–06
2013–07
2013–08
2013–09
2013–10
2013–11
2013–12
2014–01
2014–02
2014–03
2014–04
2014–05
2014–06
2014–07
2014–08
2014–09
2014–10
2014–11
2014–12
Date (Year–month)
Isolate counts
2
4
6
8
10
Location
CMFT/UHSM 2009–2014
NW region 2009–2014
Other UK hospitals 2009–2014



Species and multi-locus sequence type (MLST)
Kpne_1828
Kpne_250
Kpne_313
Kpne_478
Kpne_2559
Koxy_50
Kpne_1138
Kpne_200
Kpne_534
Kpne_626
Kpne_1665
Ecol_1415
Koxy_40
Koxy_186
Kpne_3517
Kpne_307
Kpne_22
Kpne_1438
Ecol_69
Raou
Ecol_90
Kpne_240
Ecol_141
Kpne_253
Ecol_93
Ecol_other
Kpne_10
Ecol_131
Kpne_1380
Kpne_317
Eclo_88
Kpne_218
Eclo_other
Kpne_198
Kpne_2938
Kpne_454
Kpne_556
Kpne_107
Ecol_1276
Ecol_410
Kpne_1162
Kpne_27
Kpne_483
Kpne_943
Kpne_186
Eclo_32
Kpne_34
Kpne_54
Kpne_461
Kpne_37
Kpne_491
Kpne_14
Kpne_588
Kpne_492
Eclo_133
Ecol_372
Kpne_other
Kpne_36
Kpne_321
Kpne_25
Kpne_11
Kpne_48
Kpne_468
Kpne_258
Feb 2010
May 2010
Aug 2010
Nov 2010
Feb 2011
May 2011
Aug 2011
Nov 2011
Feb 2012
Date
Eclo
Ecol
Kpne
Other



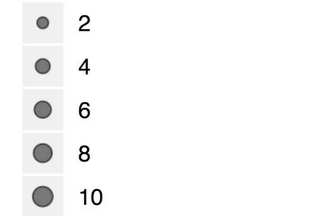
unknown-2_none_TCTCT
unknown-1_none_TTCAG
unknown-1_none_none
unknown-1_none_ATTGA
Tn4401d-2_GTTCT_TCTCT
Tn4401b-unknown_ATTGA_ATTGA
Tn4401b-6_TAGCA_TAGCA
Tn4401b-2_GTTCT_GTTCT
Tn4401a-unknown_ATTGAITATAT_CATTGICCTAT
Tn4401a-unknown_ATTGA_ATTGA
Tn4401a-unknown_ATTGA_ATATAIATTGA
Tn4401a-unknown_ATTGA_ATATA
Tn4401a-N2_ATTGA_ATTGA
Tn4401a-2_ATTGA_ATTGA
Tn4401a-2_ATCTT_ATCTT
Tn4401a-1_ATTGA_TGGTA
Tn4401a-1_ATTGA_none
Tn4401a-1_ATTGA_CTGCC
Tn4401a-1_ATTGA_ATTGA
Tn4401a-1_ATTGA_ATATAIATTGA
Tn4401a-1_ATTGA_ATATA
Tn4401a-1_ATTGA_AGGAT
Tn4401a-1_ATTGA_AATAAIATTGA
Tn4401a-1_ATTGA_AATAA
Tn4401a-1_ATGTTIATTGAITATCA_ATTGAITATCAITATGT
Tn4401a-1_AGTTG_AGTTG
Tn4401a-1_AATTTIATTGA_AATTTIATTGA
Tn4401a-1_AATAT_AATATITACCT
Tn4401a-1_AATAT_AATAT
Tn4401a-1_AAAGAIATTGA_AAAGAIATTGA
Tn4401_truncC-1_none_none
Tn4401 type and target site sequences (TSS)
Date (Year-month)
Isolate counts
2
4
6
8
10
Location
CMFT/UHSM 2009-2014
NW region 2009-2014
Other UK hospitals 2009-2014

Plasmid populations (replicon typing)

IncP|IncR
IncHI2|IncHI2A|IncR
IncHI2|IncHI2A|IncN
IncHI2|IncHI2A
IncFII|IncR|IncX1
IncFII|IncR
IncFII|IncHI2|IncHI2A|IncR
IncFII
IncFIB|IncR
IncFIB|IncI2|IncQ1|IncR|IncY
IncFIB|IncHI2|IncHI2A|IncR
IncFIB|IncHI2|IncHI2A
IncFIB|IncHI2
IncFIB|IncFII|pSL483
IncFIB|IncFII|IncY
IncFIB|IncFII|IncX3
IncFIB|IncFII|IncR|IncY
IncFIB|IncFII|IncR|IncX3
IncFIB|IncFII|IncR
IncFIB|IncFII|IncQ1
IncFIB|IncFII|IncN3|IncR
IncFIB|IncFII|IncN|IncR
IncFIB|IncFII|IncN|IncQ1
IncFIB|IncFII|IncN
IncFIB|IncFII|IncL/M|IncR
IncFIB|IncFII|IncI2|IncQ1
IncFIB|IncFII|IncHI2|IncHI2A|IncR
IncFIB|IncFII|IncHI2|IncHI2A|IncN
IncFIB|IncFII|IncHI2|IncHI2A
IncFIB|IncFII|IncHI1B|IncR|IncY
IncFIB|IncFII|IncHI1B|IncI1|IncR
IncFIB|IncFII|IncHI1B
IncFIB|IncFII
IncFIB|IncFIC|IncFII|IncI1|p0111
IncFIB
IncFIA|IncR
IncFIA|IncFII|IncHI1A|IncHI1B
IncFIA|IncFIB|IncHI1B|IncR
IncFIA|IncFIB|IncHI1B
IncFIA|IncFIB|IncFII|IncR
IncFIA|IncFIB|IncFII|IncN
IncFIA|IncFIB|IncFII|IncI1|IncQ1|IncR
IncFIA|IncFIB|IncFII|IncHI2|IncHI2A|IncR
IncFIA|IncFIB|IncFII|IncHI2|IncHI2A
IncFIA|IncFIB|IncFII|IncHI1B|IncR|IncY
IncFIA|IncFIB|IncFII|IncHI1B|IncR
IncFIA|IncFIB|IncFII|IncHI1B
IncFIA|IncFIB|IncFII|IncHI1A|IncHI1B
IncFIA|IncFIB|IncFII
IncFIA|IncFIB
IncA/C2|IncQ1|p0111
IncA/C2|IncFIB|IncFII|IncHI1B|IncR
ColRNAI|IncFIB|IncFII|IncX3
ColRNAI|IncFIB|IncFII|IncR|IncX3
ColRNAI|IncFIB|IncFII|IncR
ColRNAI|IncFIB|IncFII|IncN
ColRNAI|IncFIB|IncFII|IncI2|IncR
ColRNAI|IncFIB|IncFII|IncHI2|IncHI2A
ColRNAI|IncFIB|IncFII|IncHI1B|IncX3
ColRNAI|IncFIB|IncFII
ColRNAI|IncFIB
ColRNAI|IncFIA|IncFIB|IncFII|IncX3
ColRNAI|IncFIA|IncFIB|IncFII|IncR
ColRNAI|IncFIA|IncFIB|IncFII
Col8282|IncFIB|IncFIC|IncFII|IncX1
Col8282|IncFIA|IncFIB|IncFII|IncX1
Col8282|IncFIA|IncFIB|IncFII
Col8282|ColE10|IncFIB|IncFII|IncHI1B|IncY
Col156|IncFIB|IncHI2|IncHI2A
Col156|IncFIB|IncFII|IncR|p0111
Col156|IncFIB|IncFII|IncI1
Col156|IncFIB|IncFII|IncHI2|IncHI2A
Col156|IncFIB|IncFII
Col156|IncFIB
Col156|Col8282|IncFIA|IncFIB|IncFII
Col|IncR
Col|IncFII|IncHI2|IncHI2A|IncR
Col|IncFIB|IncFII|IncY
Col|IncFIB|IncFII|IncR
Col|IncFIB|IncFII|IncHI2|IncHI2A|IncR
Col|IncFIB|IncFII
Col|IncFIA|IncFIB|IncFII|IncHI2|IncHI2A|IncR
Col|IncFIA|IncFIB|IncFII
Col|ColRNAI|IncFIB|IncFII|IncX4
Col|ColRNAI|IncFIB|IncFII|IncR|IncX3
Col|ColRNAI|IncFIB|IncFII|IncHI1B
Col|Col8282|IncFIB|IncFII|IncR
Col|Col8282|IncFIA|IncFIB|IncFII
Col|Col156|IncFIB|IncFII|IncHI2|IncHI2A|IncR

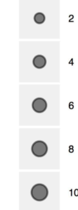
2009–03 2009–04 2009–06 2009–07 2009–08 2009–09 2009–10 2009–12 2010–01 2010–02 2010–03 2010–04 2010–05 2010–06 2010–07 2010–08 2010–09 2010–10 2010–11 2010–12 2011–01 2011–02 2011–03 2011–04 2011–05 2011–06 2011–07 2011–08 2011–09 2011–10 2011–11 2011–12 2012–01 2012–02 2012–03 2012–04 2012–05 2012–06 2012–07 2012–08 2012–09 2012–10 2012–11 2012–12 2013–01 2013–02 2013–03 2013–04 2013–05 2013–06 2013–07 2013–08 2013–09 2013–10 2013–11 2013–12 2014–01 2014–02 2014–03 2014–04 2014–05 2014–06 2014–07 2014–08 2014–09 2014–10 2014–11 2014–12

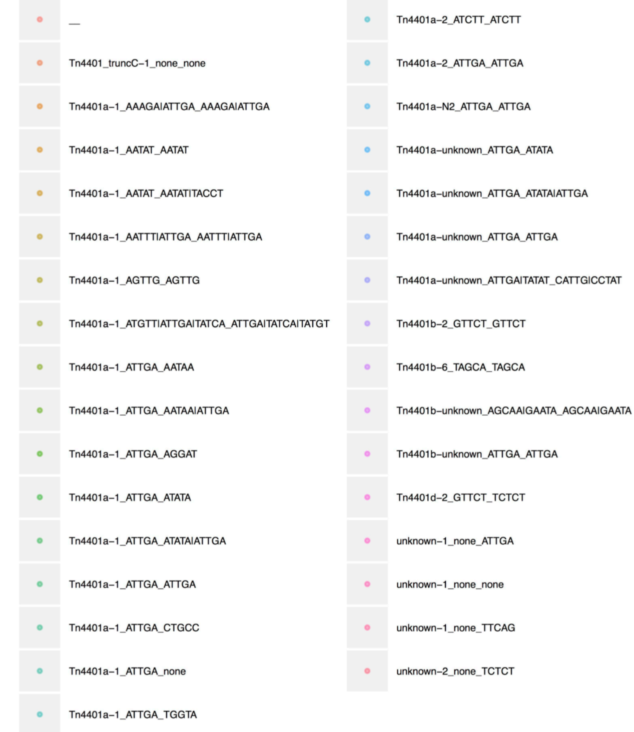
Date (Year–month)



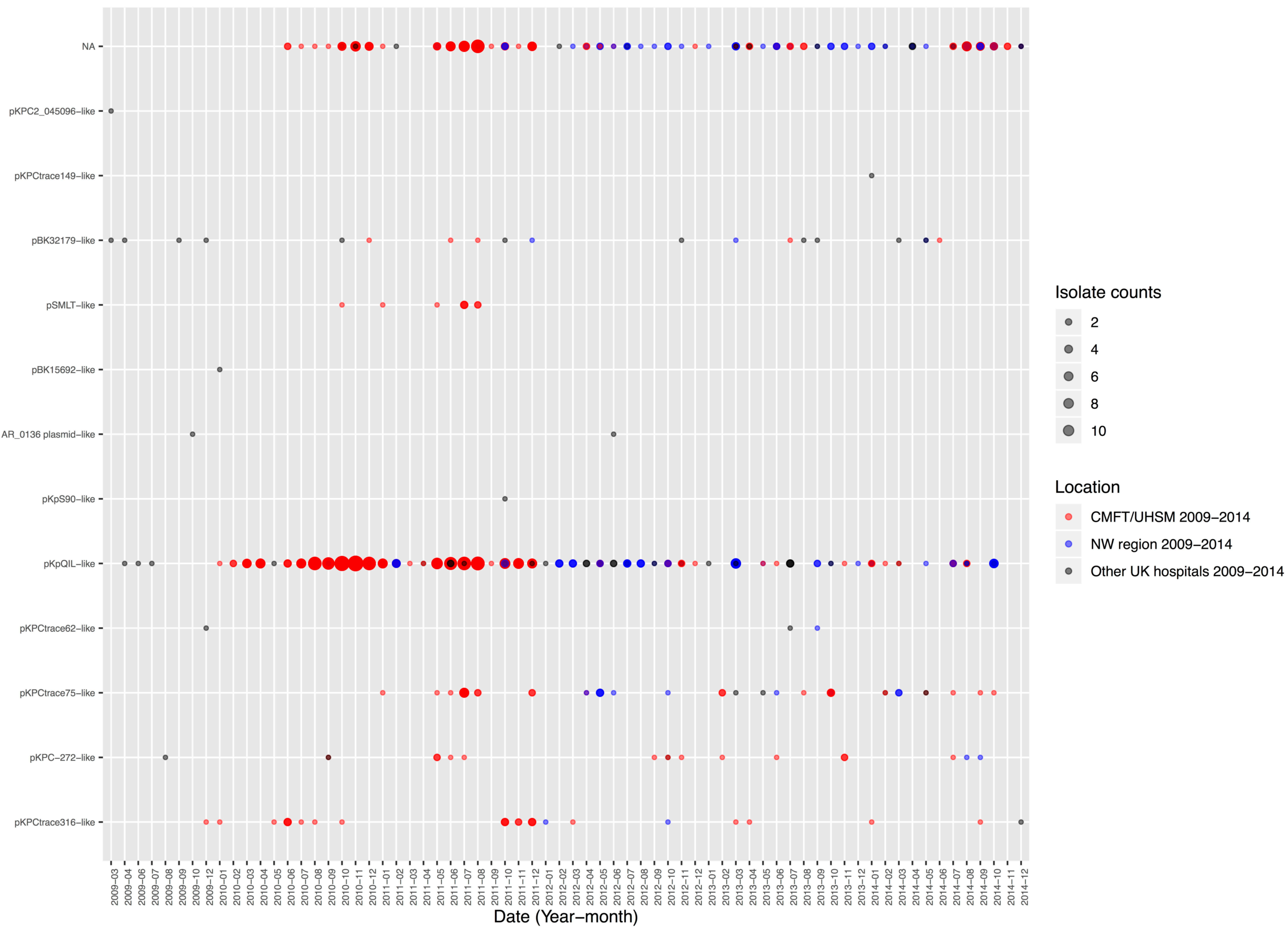
blaKPC plasmid cluster

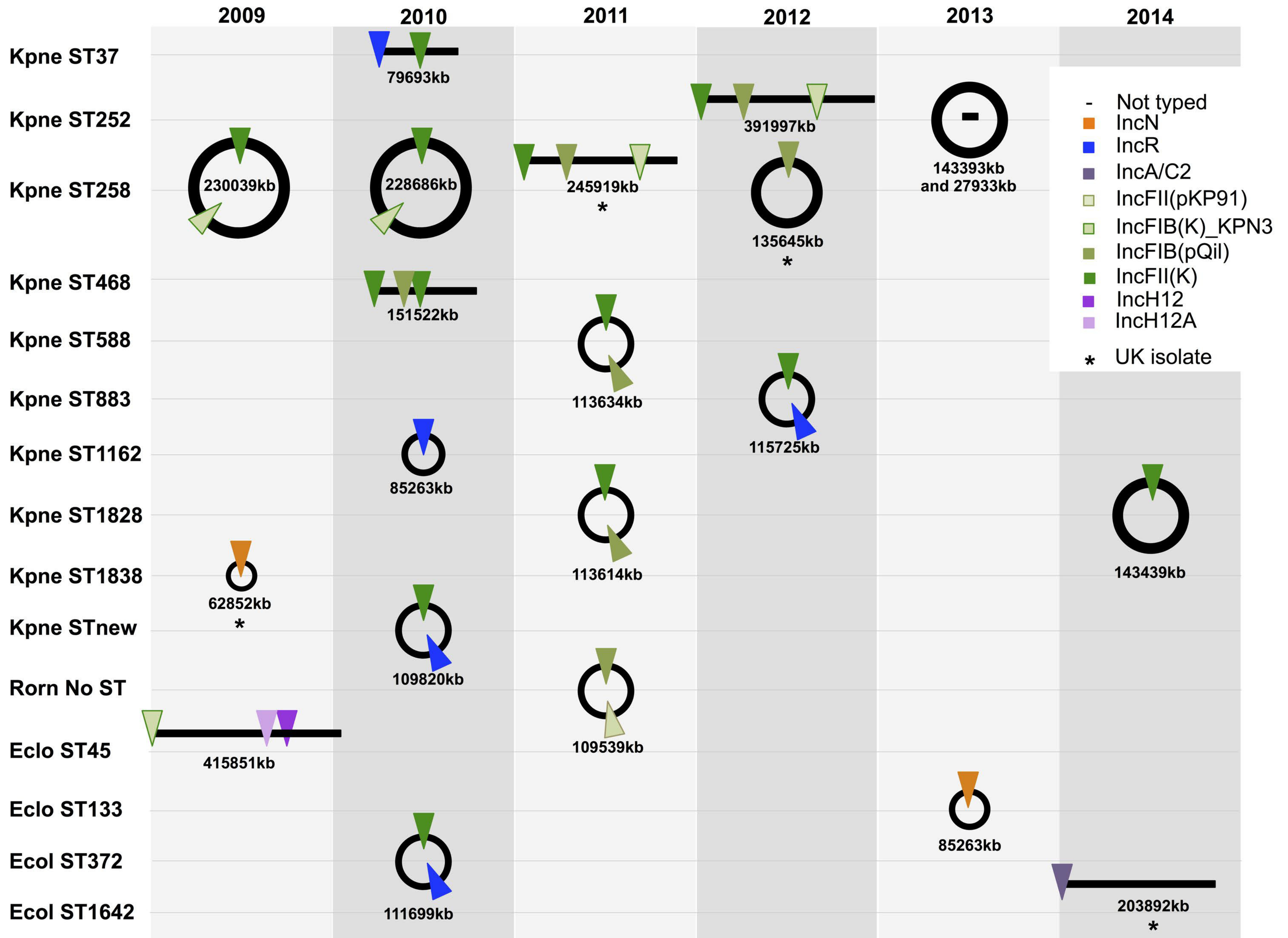
NA
pKPC2_045096–like
pKPCtrace149–like
pBK32179–like
pSMLT–like
pBK15692–like
AR_0136 plasmid–like
pKpS90–like
pKpQIL–like
pKPCtrace62–like
pKPCtrace75–like
pKPC–272–like
pKPCtrace316–like

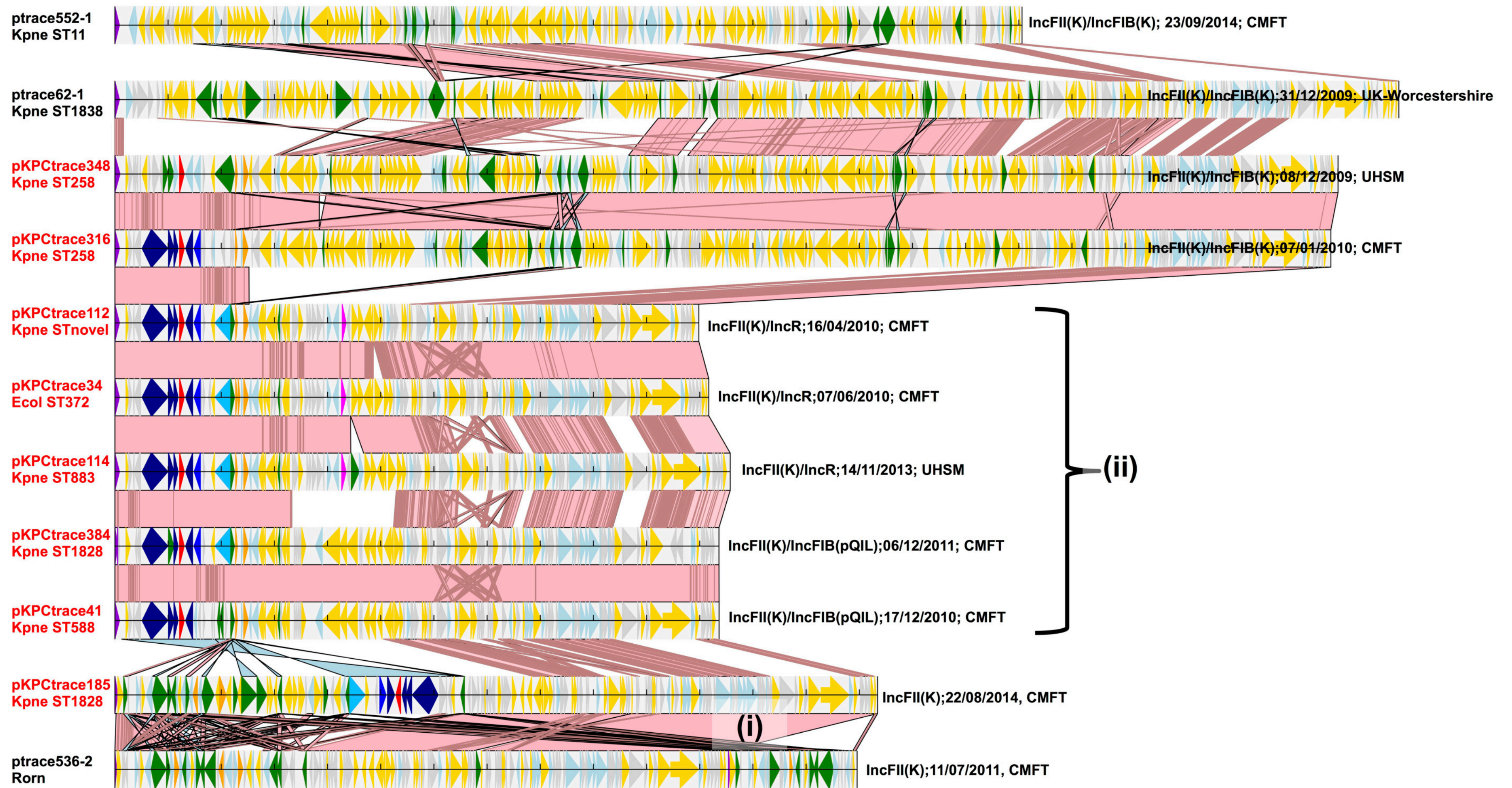




Species and multi–locus sequence type (MLST)


blaKPC plasmid cluster
NA
pKPC2_045096–like
pKPCtrace149–like
pBK32179–like
pSMLT–like
pBK15692–like
AR_0136 plasmid–like
pKpS90–like
pKpQIL–like
pKPCtrace62–like
pKPCtrace75–like
pKPC–272–like
pKPCtrace316–like
Date (Year–month)
Isolate counts
2
4
6
8
10
Location
CMFT/UHSM 2009–2014
NW region 2009–2014
Other UK hospitals 2009–2014


2009
2010
2011
2012
2013
2014
Kpne ST37
Kpne ST252
Kpne ST258
Kpne ST468
Kpne ST588
Kpne ST883
Kpne ST1162
Kpne ST1828
Kpne ST1838
Kpne STnew
Rorn No ST
Eclo ST45
Eclo ST133
Ecol ST372
Ecol ST1642
79693kb
391997kb
143393kb
and 27933kb
230039kb
228686kb
245919kb
*
135645kb
*
151522kb
113634kb
115725kb
85263kb
113614kb
143439kb
62852kb
*
109820kb
109539kb
415851kb
85263kb
111699kb
203892kb
*
- Not typed
IncN
IncR
IncA/C2
IncFII(pKP91)
IncFIB(K)_KPN3
IncFIB(pQil)
IncFII(K)
IncH12
IncH12A
* UK isolate



ptrace552-1
Kpne ST11
IncFII(K)/IncFIB(K); 23/09/2014; CMFT
ptrace62-1
Kpne ST1838
IncFII(K)/IncFIB(K);31/12/2009; UK-Worcestershire
pKPCtrace348
Kpne ST258
IncFII(K)/IncFIB(K);08/12/2009; UHSM
pKPCtrace316
Kpne ST258
IncFII(K)/IncFIB(K);07/01/2010; CMFT
pKPCtrace112
Kpne STnovel
IncFII(K)/IncR;16/04/2010; CMFT
pKPCtrace34
Ecol ST372
IncFII(K)/IncR;07/06/2010; CMFT
pKPCtrace114
Kpne ST883
IncFII(K)/IncR;14/11/2013; UHSM
pKPCtrace384
Kpne ST1828
IncFII(K)/IncFIB(pQIL);06/12/2011; CMFT
pKPCtrace41
Kpne ST588
IncFII(K)/IncFIB(pQIL);17/12/2010; CMFT
(ii)
pKPCtrace185
Kpne ST1828
IncFII(K);22/08/2014, CMFT
(i)
ptrace536-2
Rorn
IncFII(K);11/07/2011, CMFT
bla_{KPC}
Tn4401
Tn5403
ISKpn31
Other transposons/IS
Other
Conjugation/mobilisation
Hypothetical protein
$IncFII_K$
IncR