

Capsular type K54, clonal group 29 and virulence plasmids: an analysis of K54 and non-K54 closely related isolates of *Klebsiella pneumoniae*

Original Paper

Cite this article: Turton JF, Payne Z, Micah K, Turton JA (2018). Capsular type K54, clonal group 29 and virulence plasmids: an analysis of K54 and non-K54 closely related isolates of *Klebsiella pneumoniae*. *Epidemiology and Infection* **146**, 1813–1823. <https://doi.org/10.1017/S0950268818001826>

Received: 21 March 2018
Revised: 29 May 2018
Accepted: 9 June 2018
First published online: 26 July 2018

Key words:

Klebsiella; typing

Author for correspondence:

J. F. Turton, E-mail: jane.turton@phe.gov.uk

J. F. Turton¹, Z. Payne¹, K. Micah¹ and J. A. Turton²

¹Antimicrobial Resistance and Healthcare Associated Infections (AMRHA) Reference Unit, National Infection Service, Public Health England, London NW9 5EQ, UK and ²London NW9 0TA, UK

Abstract

Capsular type K54 of *Klebsiella pneumoniae* is associated with hypervirulence and we sought to discover the basis for this among isolates submitted to the UK reference laboratory between 2012 and 2017. Isolates were typed by variable number tandem repeat analysis, and capsular type and virulence elements sought by PCR. The most prevalent type found (15/31 isolates) corresponded to clonal group (CG) 29 and included five representatives carrying *rmpA*, *rmpA2* (regulators of mucoid phenotype), *iutA* and *iroD* (from the aerobactin and salmochelin siderophore clusters) associated with virulence plasmids. These included isolate KpvK54, recovered from pus. The remaining isolates did not carry a virulence plasmid. We also noted 11 further related isolates, including NCTC 9159, not of capsular type K54, but nevertheless sometimes associated with sepsis and abscesses. Whole-genome sequencing showed that KpvK54 carried a large virulence plasmid and an ICEKp3-like structure carrying the yersinia-bactin cluster, absent in NCTC 9159. Comparative chromosomal analysis with an additional four genomes showed that KpvK54 shared further genes with K1-ST23 hypervirulent isolates, and with LS358, a K54-ST29 isolate from liver abscess puncture fluid. While CG29 isolates displayed varying degrees of virulence, some, especially those with the virulence plasmid (all K54), were clearly associated with hypervirulence.

The polysaccharide capsule of *Klebsiella pneumoniae* is regarded as a virulence factor, protecting it against the host immune response. Seventy-nine capsular types have been identified [1–3]. Some of these capsular types have been associated with invasive disease, largely because of particular ‘hypervirulent’ strains belonging to them, most notably clonal groups (CG) 23 (of capsular type K1), 86, 375 and 380 (capsular type K2) all of which have been strongly associated with sepsis and liver abscesses, often in otherwise healthy people [4, 5]. Compared with those belonging to capsular types K1 and K2, isolates belonging to other capsular types have been relatively little studied. Nevertheless, capsular type K54 has also been associated with hypervirulence, particularly representatives of sequence type (ST) 29 [6–10]; the presence of a large virulence plasmid carrying capsule-upregulation genes *rmpA/rmpA2* and the aerobactin siderophore cluster is also important. We sought to investigate *K. pneumoniae* submitted to the UK national reference laboratory and positive for the K54 capsular type for evidence of hypervirulence. In doing so, it quickly became clear that one type (with variants), corresponding to CG29, dominated and that we had both K54 and non-K54 representatives of it, the latter including the reference strain NCTC 9159 [11], the whole-genome sequence of which had been determined in the NCTC 3000 project [12]. In addition, whole-genome sequences were available of two isolates of capsular type K54 and of ST29; one from liver abscess puncture fluid from China (LS358, CP025629) and the other from the environment in Australia (KP-1, CP012883). This provided a rare opportunity to be able to compare these isolate sets, with the aim of providing potential insights into the hypervirulent phenotype and the importance or otherwise of CG and capsular type.

Bacterial isolates

Between 2012 and the end of June 2017, 5566 isolates of *K. pneumoniae*, including those from 4670 patients, were typed by the laboratory, of which 42 were of capsular type K54 and/or of variable number tandem repeat (VNTR) type 7,3,1/7,5,1,2,3/4,1,2/3/4,2,4/5 (corresponding to CG29) and were included in this study. Clinical details were obtained from information from the request form completed by the sending hospital accompanying each isolate.

Typing and PCR detection of capsular types and virulence genes

Isolates were typed by VNTR analysis at 11 loci (A, E, H, J, K, D, N1, N2, 52, 45, 60 in that order) as described previously [13]; those no longer in our archives had been typed at nine loci, of which the first eight were in common with those in the 11-locus scheme. Detection of capsular types K1, K2, K5, K20, K54 and K57 and of *rmpA*, *rmpA2* and *wcaG* was by multiplex PCR [13, 14]. Virulence genes *iutA*, *allS*, *rmpA*, *mrkD* and *ybtS* were sought by a further multiplex PCR using the primers and conditions of Compain *et al.* [15], together with *iroD* (using primers *iroDF*, 5'-GTATGCGCGGCAATTGAC-3' and *iroDR*, 5'-GAACGGTCTGCCACTG) and *clbR* (using primers *clbRF*, 5'-CATGCTGGCACAAGGTGA and *clbRR*, 5'-CGTGATTCGT-ATCCGAGC-3') giving products of 920, 764, 461, 340, 242, 194 and 122 bp, respectively. The *iroD* and *clbR* primers were designed taking into account all the alleles of these genes available from the multilocus sequence typing (MLST) 'download allele sequences' page [16] and were present at a final concentration of 0.1 pmol/μl. These additional targets did not affect the performance of the PCR, which gave the same results for the *iutA*, *allS*, *rmpA*, *mrkD* and *ybtS* targets as the 5-plex alone. The 7-plex PCR was validated against a panel of 14 isolates carrying various combinations of these genes, and which had previously undergone whole-genome sequencing and detection of virulence genes from the sequences. DNA extracted from isolate KpvST23_OXA-48, which was positive for all seven targets, was used as positive control [13]. MLST was carried out as described by Diancourt *et al.* [17].

Nanopore sequencing of isolate KpvK54

Genomic DNA was extracted from KpvK54 using a GeneJet kit (ThermoFisher, Loughborough, UK). The whole-genome sequence was determined on an Mk1B minION using a MIN107 flow cell and an SQK-LSK308 1D² sequencing kit (R9.5 chemistry) (Oxford Nanopore technologies). Library preparation was carried out following the protocol described on the nanoporetech community pages with no additional shearing step. Briefly, 1 μg genomic DNA was end-repaired/dA tailed using New England Biolabs (NEB) Ultra II End-prep reaction buffer/enzyme mix (E7546), ligated with 1D² adapter using NEB Blunt/TA Ligase Mix (M0367) and adapter ligated using Ampure XP beads to purify the DNA between each step. DNA was eluted in a volume of 15 μl to give the pre-sequencing mix, which was then mixed with running buffer with fuel mix and Library Loading Bead kit (Oxford Nanopore technologies) and loaded onto a primed flow cell. Sequencing was carried out using the NC_48Hr_Sequencing_Run_FLOMIN107_SQK-LSK308 script on minKNOW (version 1.2.8).

Reads were basecalled using Albacore (ont-albacore-1.1.2-amd64) and assembled with miniasm 0.2-R159-dirty and then corrected with three iterations of Racon [18], using minimap 0.2-R124-dirty to map the reads to the assembly [19]. Following the initial analysis, a new version of Albacore (ont-albacore-2.0.2-amd64) was released and the basecalling was repeated; this analysis gave 1D² pass reads, which were used to correct the previous analysis, again with three iterations of Racon. Additionally, assembly using Canu (version 1.5 +102 changes) was performed [20]. The sequences were deposited in GenBank under accession numbers CP023134–CP023137. The original albacore basecalled

reads fastq file, together with the 1D² pass reads, is available under STUDY: PRJNA400335 (SRP118927), SAMPLE: KpvK54 (SRS2542913), EXPERIMENT: KpvK54_minION (SRX3218008) RUN: K54.fastq.gz (SRR6078568) and can be accessed at <https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP118927>.

Whole-genome sequence of NCTC 9159

The whole-genome sequence of NCTC 9159 was obtained from <http://www.sanger.ac.uk/resources/downloads/bacteria/nctc/> (ftp://ftp.sanger.ac.uk/pub/project/pathogens/NCTC3000/datalinks_manual/ERS686710.gff). Virulence genes and *wzc* and *wzi* alleles downloaded from the bigsdbMLST database [16] or from NCBI nucleotide were detected from whole-genome sequences using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch).

Identification of integrative conjugative element (ICEKp) in KpvK54

The ICEKp was identified by comparing against the 14 sequences available on 'Kleborate' [21] (in Kleborate/build/lib.linux-x86_64-2.7/kleborate/ICEKp-references).

Comparison of gene contents of KpvK54, NCTC 9159, LS358, KP-1, NTUH-K2044 and RJF999

The coding sequences from the chromosome sequence of KpvK54 (utg000001c, CP023134) were identified using Prodigal version 2.6.1 [22]. These, together with the annotated coding sequences from the chromosome sequences of NCTC 9159, LS358 (CP025629), KP-1 (CP012883), NTUH-K2044 (AP006725) and RJF999 (CP014010), were put through cd-hit version 4.7 [23] to cluster similar genes together using the default setting of 90% similarity. To ensure that genes were not missed in any of the genomes, a representative sequence from each cluster was compared against the entire chromosome sequences using a local BLAST database (BLAST 2.2.29+) and any matches added to the relevant cluster. A matrix of presence or absence for each coding sequence in the six genomes was made and genes put into sets depending on which combinations of the genomes in which they were found.

Six genes identified as potentially associated with virulence from this approach were further sought by PCR in a subset of isolates representing all three groups described in Table 1, using the primers detailed in Supplementary Table S1. In addition, K54 isolates of other types were included. PCRs were carried out using the QIAGEN *Taq* core kit with denaturation at 95 °C for 3 min, followed by 35 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min, followed by a final extension at 72 °C for 10 min.

Results

We had recorded 31 non-duplicate isolates belonging to capsular type K54 on our database, of which the largest group (15) were of the types 7,3,1/7,5,1,2,3/4,1,2/3/4,2,4/5, with most (≥9) having the profile 7,3,1,5,1,2,4,1,3/4,2,4; there was clear evidence for hypervirulence among isolates of this type (blood isolates, *rmpA*/*rmpA2*-positive, clinical details of septicaemia, abscesses, query of hypervirulence from requesting hospitals) (Table 1A) (Fig. 1). All of these were also PCR-positive for *wcaG*, always found in K54 isolates. We also had 11 K54-negative (and *wcaG*-negative) isolates of this type on our database (Table 1C),

Table 1. Isolates of VNTR type 7,3,1/7,5,1,2,3/4,1,2/3/4,2,4/5 received from UK hospitals between 2012 and end of June 2017

Isolate ^a	VNTR profile	Hospital ^b	Carbapenemase	<i>rmpA</i>	<i>rmpA2</i>	Virulence gene PCR <i>ybtS</i> , <i>mrkD</i> , <i>allS</i> , <i>iutA</i> , <i>iroD</i> , <i>clbR</i>	Isolation site	Clinical details
A. <i>rmpA/rmpA2</i>-positive representatives of capsular type K54								
P1_L1_K54v_03.12	7,3,1,5,1,2,4,1,(1) ^c	L1	–	+	nt ^d	nt	Abscess	Liver abscess
P2_L2_K54v_42.15	7,3,1,5,1,2,4,1,3,2,4	L2	–	+	+	<i>ybtS</i> , <i>mrkD</i> , <i>iutA</i> , <i>iroD</i>	Blood culture	Pyrexial
P3_L2_K54v_52.15	7,3,1,5,1,2,4,1,2,2,4	L2	–	+	+	<i>ybtS</i> , <i>mrkD</i> , <i>iutA</i> , <i>iroD</i>	Blood	Sepsis
P4_L3_K54v_16.16	7,3,1,5,1,2,4,1,3,2,4	L3	–	+	+	<i>ybtS</i> , <i>mrkD</i> , <i>iutA</i> , <i>iroD</i>	Blood culture	'Hypervirulent <i>Klebsiella</i> testing' requested
P5_L1_K54v_18.17 (KpvK54)	7,3,1,5,1,2,4,1,3,2,4	L1	–	+	+	<i>ybtS</i> , <i>mrkD</i> , <i>iutA</i> , <i>iroD</i>	Pus from neck	'Hypervirulent <i>Klebsiella</i> testing' requested
B. <i>rmpA/rmpA2</i>-negative isolates of capsular type K54								
P6_NE1_K54_42.13	7,3,7,5,-,2,3,1,4,2,4	NE1	–	–	–	<i>mrkD</i>	Blood	Bacteraemia
P7_NE2_K54_50.12	7,3,7,5,-,2,3,1,4,2,4	NE2	–	–	–	<i>mrkD</i>	Blood	Biliary sepsis
P8_SE1_K54_26.15	7,3,1,5,1,2,4,1,(1) ^c	SE1	NDM, OXA-48-like	–	–	<i>mrkD</i>	Wound swab	Burns patient
P9_L2_K54_32.15	7,3,1,5,1,2,4,1,4,2,4	L2	–	–	–	<i>ybtS</i> , <i>mrkD</i>	Blood culture	Bacteraemia; neonate
P10_L2_K54_32.15	7,3,1,-,1,2,4,1,4,2,4	L2	–	–	–	<i>ybtS</i> , <i>mrkD</i>	Urine	None given (clinical); neonate
P11_SW1_K54_46.15	7,3,1,5,1,2,4,1,4,2,4	SW1	–	–	–	<i>mrkD</i>	Sputum	None given (ICU patient, clinical)
P12_L2_K54_03.17	7,3,1,5,1,2,3,1,4,2,4	L2	–	–	–	<i>mrkD</i>	Blood	Septicaemia
P13_NW1_K54_11.17	7,3,1,5,1,-,4,3,-,-,4	NW1	–	–	–	<i>mrkD</i>	Rectal swab	Screening
P14_L2_K54_12.17	7,3,1,5,1,2,3,1,4,2,4	L2	–	–	–	<i>mrkD</i>	Blood	None given
P15_M1_K54_17.17	7,3,1,5,1,2,4,1,4,2,4	M1	NDM	–	–	<i>mrkD</i>	Skin wound	None given
C. Non-K54 isolates								
P16_L1_30.12	7,3,1,5,1,2,4,1,4,2,4	L1	NDM	–	–	<i>mrkD</i>	Urine	Urinary tract infection
P17_M2_32.12	7,3,1,5,1,2,4,1,(1) ^c	M2	–	–	nt	nt	Blood	? Bacteraemia (burns patient)
P18_SE2_24.13	7,3,1,5,1,2,4,1,(1) ^c	SE2	–	nt	–	nt	Blood	Bacteraemia, pneumonia
P19_M2_13.14	7,3,1,5,1,2,4,1,4,2,4	M2	OXA-48-like	–	–	<i>mrkD</i>	Blood	Bacteraemia
P20_NE3_27.14	7,3,1,5,1,2,4,1,4,2,4	NE3	–	–	–	<i>ybtS</i> , <i>mrkD</i>	Pus	Abscess following kidney transplant
NCTC 9159	7,3,1,5,1,2,4,1,4,2,4	NCTC	–	–	–	<i>mrkD</i>	Urine	None given
P21_NE4_29.15	7,3,1,-,1,2,4,1,5,2,4	NE4	OXA-48-like	–	–	<i>mrkD</i>	Faeces	Not given

(Continued)

Table 1. (Continued.)

Isolate ^a	VNTR profile	Hospital ^b	Carbapenemase	<i>rmpA</i>	<i>rmpA2</i>	Virulence gene PCR <i>ybtS</i> , <i>mrkD</i> , <i>dllS</i> , <i>iutA</i> , <i>iroD</i> , <i>clbR</i>	Isolation site	Clinical details
P22_SE3_05.16	7,3,1,5,1,2,4,1,4,2,4	SE3	–	–	–	<i>mrkD</i>	Swab	Not given
P23_L4_28.16	7,3,1,5,1,2,4,1,4,2,4	L4	OXA-48-like	–	–	<i>mrkD</i>	Pleural fluid	Post-op sepsis
P24_NW2_09.17	7,3,1,5,1,2,4,1,4,2,5	NW2	OXA-48-like	–	–	<i>ybtS</i> , <i>mrkD</i>	Rectal swab	None given
P25_L4_12.17	7,3,1,5,1,2,4,1,4,2,4	L4	OXA-48-like	–	–	<i>mrkD</i>	Rectal swab	Screen

^aIsolates were labelled by patient (P1–25), hospital, whether K54-positive, or K54-positive with evidence of virulence plasmid (K54v) and date received (week/year).

^bHospitals were labelled according to the region and number within that region; L = London, NE = North East England, NW = North West England, SE = South East England, M = Midlands.

^cIsolates no longer in our archives and typed only with a nine-locus VNTR scheme, of which the first eight loci are in common with the 11-locus scheme used for the remainder.

^dnt; not tested.

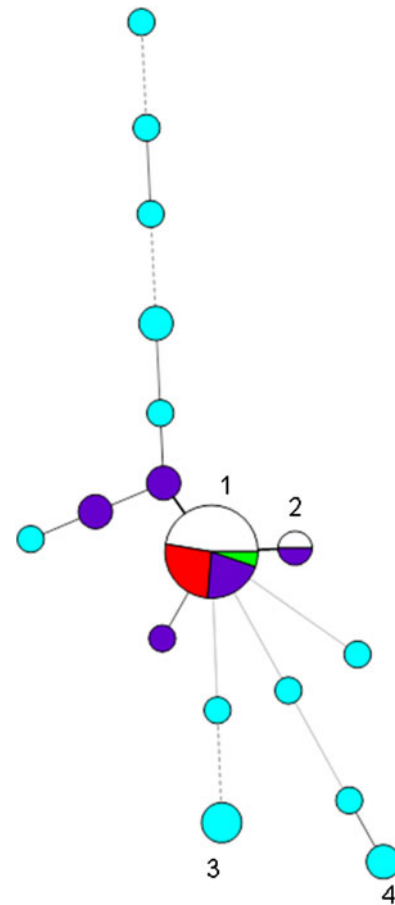


Fig. 1. Minimum spanning tree of variable number tandem repeat (VNTR) profiles at the first eight loci of the 42 isolates belonging to capsular type K54 and/or of profile 7,3,1/7,5,1,2,3/4,1,2/3/4,2,4/5 (corresponding to CG29) found during 2012–2017. Each circle represents a type and the size of the circle is proportional to the number of isolates of that type. Circles 1 and 2 denote profiles beginning 7,3,1,5,1,2,4,1 (19 isolates) and 7,3,1,-,1,2,4,1 (two isolates), respectively; segments marked in red denote K54 isolates of CG29-positive for *rmpA/rmpA2*, those in dark blue K54 isolates attributed to CG29 in which *rmpA/rmpA2* were not detected, and uncoloured/green segments those of CG29-negative for K54 and negative for *rmpA/rmpA2*; NCTC 9159 is denoted by the green segment. K54 isolates not of CG29 are denoted by turquoise circles, e.g. circles 3 and 4 denote isolates with profiles beginning 4,1,3,2,2,2,3 and 1,-,2,4,0,1,2,5. Thick solid lines join single-locus variants, while thinner, dashed or dotted lines join multilocus variants (up to three, four and six loci different, respectively).

with many of these associated with carbapenemase genes; none were PCR-positive for *rmpA/rmpA2* or *iutA*, indicating lack of a virulence plasmid; only three of the 11 isolates were from blood, but they also included those from pus (from an abscess) and pleural fluid. One of these was a reference isolate, NCTC 9159 [11], of capsular type 39 and ST985 (a single locus variant (SLV) of ST29), isolated from urine in 1952 in Denmark, the whole-genome sequence of which had been determined in the NCTC 3000 project [12] (shown in green in Fig. 1).

These 26 isolates (15 K54-positive and 11 K54-negative) sharing similar profiles (7,3,1/7,5,1,2,3/4,1,2/3/4,2,4/5) could therefore be divided into three groups: (i) K54-positive isolates that were PCR-positive for *rmpA*, *rmpA2*, *iutA* and *iroD*, indicating the presence of a virulence plasmid carrying both the aerobactin and salmochelin siderophore clusters, and for *ybtS* from the yersiniabactin siderophore cluster (5); (ii) K54-positive isolates that were PCR-negative for *rmpA*, *rmpA2*, *iutA* and *iroD* and

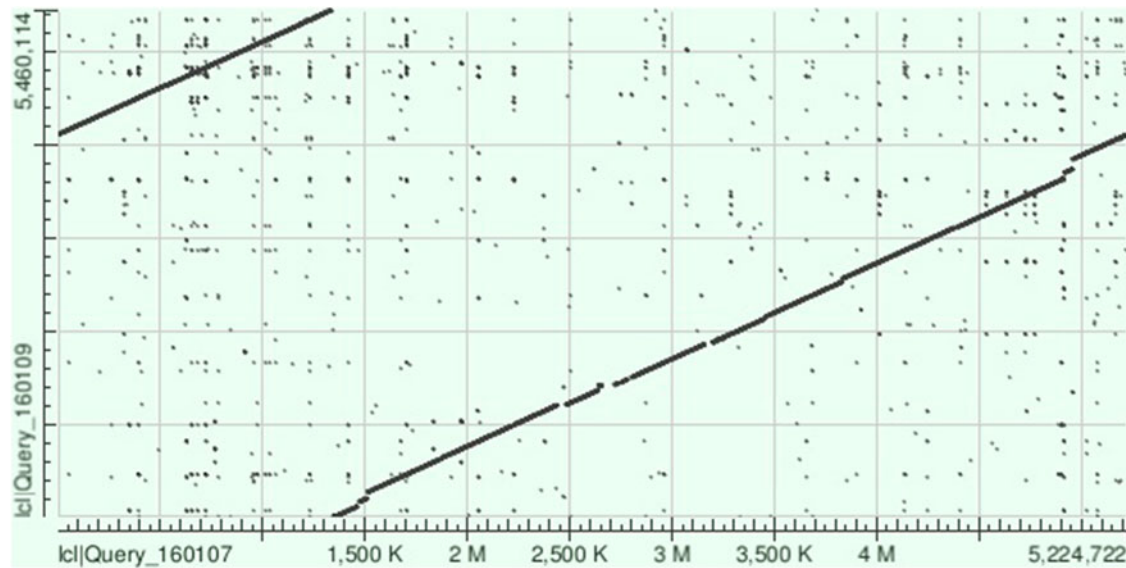


Fig. 2. Dot plot obtained following aligning of the sequences of the chromosomes of KpvK54 (Query_160107) and NCTC 9159 (Query_160109) on BLAST showing the strong similarity between the two sequences and highlighting those relatively small regions of the sequence of the KpvK54 chromosome not present in the NCTC 9159 chromosome sequence (breaks in the line between the two sequences).

clbR, indicating lack of a virulence plasmid, with *ybtS* variably present (10) (Table 1B); and (iii) K54-negative isolates lacking evidence of a virulence plasmid, with *ybtS* variably present (11). Isolates with the virulence plasmid had 2 or 3 repeats at the ninth VNTR locus, where determined, while the other isolates mostly had a repeat of 4 or 5 at that locus (16/17). The non-K54 representatives were more likely to have a carbapenemase gene (6/11) than the K54 examples (2/15) ($P = 0.04$ by two-tailed Fisher's exact test); three of these isolates were from rectal/faecal screens. We did not detect *clbR* (from the colibactin cluster) in any of the isolates.

The whole-genome sequence of isolate KpvK54, which belonged to the first group (i; group A in Table 1) and was isolated in 2017 from pus, was determined by nanopore sequencing. The sequence was assembled from 363 659 reads (total 4 176 296 833 bases) with an average read length of 11 484 bases and 4156 reads longer than 46 879 bases; the longest read was 122 975 bases. The whole-genome sequence confirmed the capsular type of K54 (*wzc*₅₄). The sequence obtained was at very high coverage (783-fold), and therefore provided complete assemblies, but was not entirely accurate (99% according to BLAST searches and comparison with Sanger sequencing of the MLST housekeeping genes) mainly because of systematic problems with homopolymer runs (runs of the same base), which are a recognised problem with minION sequences. Correction of the initial analysis with the 1D² pass reads basecalled with the later version of Albacore improved the accuracy. The final miniasm/Racon-corrected assembly gave four contigs of 5 224 722 bp (utg000001c, the chromosome, CP023134), 211 454 bp (utg000002c, the virulence plasmid, CP023135), 31 074 bp (utg000003 l, CP023136), which has 79% homology with an extrachromosomal sequence described in *K. pneumoniae* CAV1217 (CP018673) and 12 361 bp (utg000004c, CP023137), which matched parts of a plasmid found in *Escherichia coli* (pEC15II, KU932031.1 and pDKX1-TEM-52, JQ269336). The Canu assembly (from the original basecalled reads) gave two contigs of length 5 230 099 bp (tig0000001) and 263 869 bp, neither of which was circular; we chose to use the miniasm/Racon-corrected assembly.

Sanger sequencing confirmed the MLST type (ST2849) as a double locus variant of ST29 of allelic profile 158,3,2,2,6,4,160 (with alleles differing from ST29 underlined) and a SLV of NCTC 9159. The most similar isolate on the MLST database (isolate id 769) belongs to ST714, a SLV from a patient from Asia with bacteraemia [24].

KpvK54 had a large virulence plasmid (211 kb) carrying the aerobactin (*iutA*, *iucABCD*) and salmochelin (*iroBCDN*) acquired siderophore systems as well as *rmpA* and *rmpA2* and lead (*pbrABCR*), copper (*pcoABCDERS*), silver (*silCERS*) and tellurium (*terABCDEWXYZ*) resistance genes, typical of hypervirulent types (Supplementary Table S2). The yersiniabactin cluster (*ybtAEPQSTUX* and *fyuA*, *irp1*, *irp2*) was also present on the chromosome, so that the isolate carried three acquired siderophore systems, strongly correlated with hypervirulence. It also contained other virulence genes sought, both on the plasmid (hemin, SAM-dependent methyltransferase, *luxR*, *pagO*, *ibrB*, *fecA*, *fecI*, *fecR*, Fe 3+ citrate ABC transporter, *fur*, *shiF*, lysozyme inhibitor, *cobW*) and on the chromosome (*mrkABCDFHIJ* (type 3 fimbrial genes)).

In common with most hypervirulent isolates, which are typically susceptible to most antibiotics, isolate KpvK54 contained few resistance genes: with *bla*_{SHV-83} (penicillinase non-extended spectrum β -lactamase), *oqxA* and *oqxB* (fluoroquinolone) (core genes in *K. pneumoniae*) and *fosA* (fosfomycin) detected on the chromosome.

The NCTC 9159 chromosome sequence (called 'unitig_0_quiver[quiver]') covered 96% of that of KpvK54 with 99% identity (BLAST total score of 9.984×10^6), highlighting the similarity between these two isolates, despite isolation 65 years apart (Fig. 2). On BLAST analysis, the KpvK54 chromosome was also closely similar to those of LS358 (CP025629) (98% coverage and 99% identity, BLAST total score 9.627×10^6) and KP-1 (CP012883) (96% coverage and 99% identity, BLAST total score of 1.006×10^7); both isolates are also of capsular type K54 and belong to ST29 with LS358 having been isolated from liver abscess puncture fluid, while KP-1 was from the environment. Other best matches appearing on BLAST comparisons were

with NTUH-K2044 (AP006725) (total score of 9.482×10^6) and RJF999 (CP014010) (total score of 9.316×10^6) which had 92% and 93% coverage, respectively, and 98% identity with the KpvK54 chromosome. Both NTUH-K2044 and RJF999 belong to K1-ST23, the type most associated with hypervirulence.

However, NCTC 9159 had no virulence plasmid and lacked almost all the virulence genes detected in KpvK54, with the exception of the *mrkABCDHIJ* type 3 fimbrial genes (found in most isolates of *K. pneumoniae*); there was some evidence for lead resistance genes *pbrA* and *pbrR* on the chromosome, but these did not have complete coverage and had low identity (80–89%) with the downloaded version from the MLST website; there was a ferric uptake *fur* gene. It contained an acquired resistance gene not found in KpvK54 (*strAB*) but was otherwise similarly devoid of most resistance determinants, with only *bla*_{SHV-83}, *oqxA* and *oqxB* in addition, as in KpvK54. The whole-genome sequence confirmed the capsular type as K39 (*wzc*₃₉) and provided the ST of 985, a SLV of ST29 (allelic profile 10,3,2,2,6,4,4 with the *gapA* allele (underlined) differing from that of ST29). Neither KpvK54 nor NCTC 9159 carried *pld1*, encoding a phospholipase D family protein, which has also been identified as a virulence factor [25].

Only the chromosome sequence of isolate LS358, of ST29, was available, so we cannot comment on whether it possesses a virulence plasmid or elements such as the aerobactin cluster and tellurite resistance genes associated with them. The only plasmid sequence described of KP-1, also of ST29, (NZ_CP012884) contains none of the virulence elements sought and it therefore likely does not have a virulence plasmid. Similarly to NCTC 9159, the chromosome of the environmental isolate KP-1 lacked most of the virulence genes described among hypervirulent isolates, with the exception of *mrkABCDHJI*; it did not have the yersiniabactin cluster. By contrast, the chromosome of LS358, isolated from liver abscess puncture fluid, carried the yersiniabactin cluster (*ybtAEPQSTUX* and *fyuA*, *irp1*, *irp2*), but not the *virB* genes encoding type IV secretion proteins, suggesting that the yersiniabactin cluster was not in an integrative conjugative element (ICE), which requires a type IV secretion system. Surprisingly, we did not detect *wcaG* in this isolate. Both isolates carried *bla*_{SHV-83}, *oqxA* and *oqxB* resistance genes. As was the case with KpvK54 and NCTC 9159, neither LS358 nor KP-1 carried *pld-1*.

There were four main regions where the chromosome sequence of KpvK54 had no significant similarity with that of NCTC 9159. Perhaps the most significant of these are the regions of nt 2 656 985–2 718 902, encoding an ‘integrase/retron-type reverse transcriptase, putative ATPase’ (according to the annotations from matches in BLAST) and the virulence-associated yersiniabactin loci (*ybtS*, *ybtX*, *ybtQ*, *ybtP*, *ybtA*, *irp2*, *irp1*, *ybtU*, *ybtT*, *ybtE*, *fyuA*), and nt 2 775 659–2 799 701 encoding the capsular polysaccharide (*cps*) gene synthesis cluster, including *wzc*₅₄ and *wzi*₁₁₅. There are three instances of the CCAGTCAGAG-GAGCCAA 17 bp flanking direct repeat sequence of *K. pneumoniae* ICEs (ICEKp) (at nt 2 646 067–2 64 083, 2 656 808–2 656 824 and 271 886–2 718 902) that are associated with mobilisation of yersiniabactin [26] and it is likely that *irp1*, *irp2*, *fyuA* and *ybtAEPQSTUX* are all within an ICEKp element bounded between nt 2 650 808 and 2 718 902, which covers 96% of the sequence of ICEKp3 and includes an integrase gene and type IV secretory system, essential components of an ICE.

NCTC 9159 was isolated from urine (rather than a site indicating invasive disease) and lacks characteristics associated generally with hypervirulence (including a virulence plasmid), and therefore

arguably would not be considered hypervirulent. Similarly, the environmental isolate KP-1 has not been associated with hypervirulence, so the differences between KpvK54 and NCTC 9159 and KP-1 may be significant in this context, particularly where they are also shared with NTUH-K2044 and RJF999, also associated with hypervirulence, and perhaps also with LS358, isolated from liver abscess puncture fluid. The gene matrix showing the number of genes common to each combination of these isolates revealed only 13 coding sequences found in KpvK54, LS358, NCTC-K2044 and RJF999, but not in KP-1 and NCTC 9159, all but two of which belong to the virulence-associated yersiniabactin cluster (*ybtS*, *ybtX*, *ybtQ*, *ybtP*, *ybtA*, *irp2*, *irp1*, *ybtU*, *ybtT*, *ybtE*, *fyuA*) (Supplementary Table S3). The remaining genes were an integrase gene (part of the ICEKp element carrying the yersiniabactin cluster) and K54_4379379_4380137 (which partially covers C0077_23090 and RJF9_26025 coding for α -galactosidase/glycoside hydrolase); however, 84% of this coding sequence (corresponding to nt 4 379 379–4 379 969 and nt 4 380 086–4 380 137 of the chromosome of KpvK54) was also present in NCTC 9159 and KP-1. Table 2 gives a list of coding sequences common to KpvK54 and NTUH-K2044 and/or RJF999, but not in NCTC 9159; some of these were also present in LS358 and/or KP-1. NTUH-K2044 and KpvK54 alone of the six genomes shared *mobB*, *mobC*, *virB11* and *ardC*. Interestingly, the type IV secretion system genes *virB1*, *virB2*, *virB3-4*, *virB5*, *virB8* and *virB10* were common only to KpvK54, NTUH-K2044 and RJF999, and were not found in LS358, nor was *alpA2* (encoding a putative transcriptional regulator/DNA-binding protein). There were also a further 26 genes common only to KpvK54 and RJF999, and a further 15 common to KpvK54 and LS358 only (Supplementary Table S3). KpvK54 shared in common with RJF999, a region corresponding to RJF9_19730 to RJF9_19760, which was not present in any of the other genomes. *potD*, *rhaP* and eight other genes were shared by all the genomes except NCTC 9159. There were 335 genes common to NTUH-K2044 and RJF999 (K1-ST23 isolates) that were not present in KpvK54, LS358, KP-1 or NCTC 9159 (the CG29 isolates), which included *allABCSR* (the allantoin cluster), *acoABCDK*, *cas1*, *cas2*, *cas3*, *cas5e*, *cse1*, *cse2*, *cse3*, *cse4*, *magA*, *fdxA*, *gatABCDYZ*, *pflCD*, *sfuABC* and *ylbCEF*. Among the genes found only in K54 isolates, 15 were found in KpvK54 and LS358 that were not present (or in some cases (e.g. C0077_05160) were only partially present) in KP-1 (Supplementary Table S4); there were a further 15 genes found in KpvK54, LS358 and KP-1 that were not present in any of the other genomes (e.g. C0077_09035 and C0077_05470, both coding for hypothetical proteins), only some of which were genes in the capsular polysaccharide synthesis cluster. There were also significant numbers of coding sequences exclusive to each isolate among the six genomes; in the case of the 79 predicted genes found only in KpvK54 these were mainly found together in regions (e.g. 2 709 483–2 717 651 (in the ICE element)) and correspond almost exclusively to ‘hypothetical proteins’ on BLAST searches.

PCR testing showed that RJF9_19740, RJF9_19755 and RJF9_19640 were present exclusively in KpvK54 among the 15 isolates in which they were sought; they were not detected in any of the other three *rmpA/rmpA2*-positive isolates of K54 (Table 3). C0077_05300, C0077_05255 and C0077_04790 were detected in all four *rmpA/rmpA2*-positive isolates of K54, with C0077_05255 not being found in any of the other isolates tested. These results are consistent with the association of these targets with virulent isolates. Both C0077_05255 and C0077_04790 have few matches on BLAST, indicating that they are not widely

Table 2. Chromosomal genes found in KpvK54 (CP023134) and NUTH-K2044 (AP006725) or RJF999 (CP014010), but not found in NCTC 9159. Those in bold are in an ICE*Kp* element that includes the yersiniabactin cluster. LS358 (CP025629) and KP-1 (CP012883) are representatives of K54-ST29 from liver abscess puncture fluid and the environment, respectively

Gene(s)	Position in KpvK54	NTUH-K2044 locus tag	RJF999 locus tag	LS358/KP-1 locus tags	Gene function/product
<i>rhaP</i>	4 574 680–4 576 908	KP1_0067	RJF9_00325	C0077_24045/ KLP1_13090	Branched-chain amino acid ABC transporter permease
KP1_2125/RJF9_10060	1 338 807–1 340 526	KP1_2125	RJF9_10060	C0077_08940/ KLP1_22995	Aldehyde dehydrogenase
<i>potD</i>	1 340 572–1 341 567	KP1_2127	RJF9_10065	C0077_08945/ KLP1_23000	Spermidine/putrescine ABC transporter substrate-binding protein
RJF9_10370	1 361 832–1 362 892	–	RJF9_10370	–/KLP1_23090	Diguanylate cyclase AdrA
RJF9_10375	1 363 247–1 363 943	–	RJF9_10375	–/KLP1_23095	Diguanylate phosphodiesterase
KP1_2290 RJF9_10965	c1 468 509–1 469 412	KP1_2290	RJF9_10965	C0077_09495/ KLP1_23680	Transcriptional regulator
KP1_2291 RJF9_10970	1 469 541–1 470 847	KP1_2291	RJF9_10970	KLP1_23685 C0077_09500	Hypothetical protein
KP1_2292 RJF9_10975	1 470 862–1 471 949	KP1_2292	RJF9_10975	KLP1_23690 C0077_09505	Tartrate dehydrogenase
KP1_2293 RJF9_10980	1 471 952–1 473 208	KP1_2293	RJF9_10980	KLP1_23695 C0077_09510	Hydroxypyruvate reductase
KP1_2400 RJF9_11475 K54_1556841_1558466	1 556 820–1 558 572	KP1_2400	RJF9_11475	–/–	Type IV secretion protein Rhs
KP1_3565 RJF9_16915	c2640860–2 641 249	KP1_3565	RJF9_16915	KLP1_04010 C0077_15075	Enamine deaminase RidA/putative translation initiation inhibitor
Integrase ^a	2 656 985–2 658 245	KP1_3581	RJF9_17075, RJF9_17225	C0077_15155/–	Integrase
<i>ybtAEPQSTUX</i> and <i>fyuA</i> , <i>irp1</i> , <i>irp2</i>	2 658 439–2 685 067	KP1_3587, 3592, 3586, 3585, 3583, 3591, 3590, 3584 and 3593, 3589, 3588	RJF9_17250, 17275, 17245, 17240, 17230, 17270, 17265, 17235 and 17280, 17260, 17255	C0077_15160, 15165, 15170, 15175, 15180, 15185,15190,15195, 159200, 15205, 15210/–	Yersiniabactin cluster
KP1_3595 RJF9_17285	2 688 095–2 689 022	KP1_3595	RJF9_17285	–/–	Conserved hypothetical protein
<i>alpA2</i>	2 689 163–2 689 350	KP1_3596	RJF9_17290	–/–	Putative transcriptional regulator/ DNA-binding protein
KP1_3597 RJF9_17295	2 689 539–2 690 120	KP1_3597	RJF9_17295	–/–	Hypothetical protein
RFJ9_17300	–	–	RFJ9_17300	–/–	Transcriptional regulator
KP1_3625 RJF9_17305	2 691 258–2 691 490	KP1_3625	RJF9_17305	–/–	ATP-binding protein (next to virB1)
virB1 (2 691 563–2 692 273 ^b), virB2 (2 692 273–2 692 565), virB3-4 (2 692 578–2 695 309), virB5 (2 695 327– 2 696 034), virB6 (2 696 350–2 697 358), virB8 (2 697 579–2 698 258), virB9 (2 698 255–2 699 154), virB10 (2 699 203–2 700 450), virB11 (2 700 440–2 701 461)	2 691 563–2 701 461	KP1_3626, 3627, 3628, 3629, 3633, 3634, 3635, 3636	RJF9_17310, 17315, 17320, 17325, 17340, 17345, 17350	–/– –/– –/– –/– –/– –/– –/– –/– –/–	Type IV secretion proteins

(Continued)

Table 2. (Continued.)

Gene(s)	Position in KpvK54	NTUH-K2044 locus tag	RJF999 locus tag	LS358/KP-1 locus tags	Gene function/product
KP1_3640	2 702 653–2 702 955	KP1_3640	3 551 246–3 551 548 and c3 452 758–3 453 056	–/–	Hypothetical protein
<i>mobB</i> , <i>mobC</i>	2 703 225–2703785, 2705129–2 705 875	KP1_3641, 3642	–	–/–	Mobilisation proteins
K54_2704497_2705135	2 704 497–2 705 135	–	–	–/–	
<i>ardC</i>	c ² 706 067–2 707 010	KP1_3643	Not in RJF999	–/–	Antirestriction protein
RJF9_17330	2 690 563–2 690 691	–	RJF9_17330	–/–	Conjugal transfer protein TrbL
RJF9_17335	2 697 432–2 697 549	3 455 097–3 455 215	RJF9_17335	–/–	Hypothetical protein
RJF9_17475, 17480	c2 741 976–2742243, 2741333–2 741 491	–	RJF9_17475, 17480	–/KLP1_04225 –/–	Transposases
<i>wcaG</i> K54_2784254–2785072	2 784 106–2 785 072	KP1_3709	RJF9_17775	–/KLP1_04405	GDP-fucose synthetase
KP1_3725 RJF9_17825 K54_2797636_2798265	2 703 786–2 705 115	KP1_3725	RJF9_17825	KLP1_04450 C0077_15460	Hypothetical protein/putative acid phosphatase
RJF9_19510	c3 158 954–3 159 430	–	RJF9_19510	–/–	Endopeptidase
RFJ9_19515	c3 160 061–3 160 350	–	RFJ9_19515	–/–	Holin
RJF9_19530, 19540, 19545, 19550	3 163 444–3 163 737, c3 164 024–3 164 320, c3 164 518–3 166 988, c3 166 992–3 168 795	–	RJF9_19530, 19540, 19545, 19550	–/– –/– –/–	Hypothetical proteins
RFJ9_19635, 19640	3 181 869–3 183 341 c3 183 338–3 183 931	–	RFJ9_19635, 19640	–/– –/–	Terminases
RFJ9_19670, 19690, 19710	c2 471 693–2 471 899, c3 187 383–3 187 726, c3 189 746–3 189 979	–	RJF9_19670, 19690, 19710	–/– –/– –/–	Hypothetical proteins
RFJ9_19715	3 190 136–3 190 715	–	RFJ9_19715	–/–	Repressor
RFJ9_19730	3 191 377–3 192 196	–	RFJ9_19730	–/–	Exodeoxyribonuclease VIII
RJF9_19735 (only 81% coverage)	3 192 193–3 192 909	–	RFJ9_19735	–/–	Recombinase RecT
RJF9_19740	3 193 122–3 193 368	–	RJF9_19740	–/–	Transcriptional regulator
RJF9_19745	3 193 478–3 193 771	–	RJF9_19745	–/–	PerC transcriptional activator family protein
RJF9_19750	3 194 575–3 195 165	–	RJF9_19750	–/–	Adenine methylase
RJF9_19755	3 195 162–3 195 351	–	RJF9_19755	–/–	Cell division protein
RJF9_19760	c3 195 372–3 196 619	–	RJF9_19760	–/–	Integrase

^ant 2 656 808–2 718 919 in the KpvK54 chromosome corresponds to a ICEKp element closest to ICEKp3.

^bNucleotide positions in KpvK54.

Table 3. Results of PCR testing for six targets identified as potentially associated with virulence among 11 clonal group 29 (CG29) isolates described in Table 1 (groups A, B and C) and four further isolates of K54 not belonging to CG29

Isolate	Group	C0077_0530 ^a	C0077_05255	C0077_04790	RJF9_19740 ^b	RJF9_19755	RJF9_19640
P2_L2_K54v_42.15	A <i>rmpA/rmpA2</i> K54	+	+	+	–	–	–
P3_L2_K54v_52.15	A <i>rmpA/rmpA2</i> K54	+	+	+	–	–	–
P4_L3_K54v_16.16	A <i>rmpA/rmpA2</i> K54	+	+	+	–	–	–
KpvK54	A <i>rmpA/rmpA2</i> K54	+	+	+	+	+	+
P6_N1_K54_42.13	B non- <i>rmpA/rmpA2</i> K54	+	–	+	–	–	–
P9_L2_K54_32.15	B non- <i>rmpA/rmpA2</i> K54	–	–	+	–	–	–
P10_L2_K54_32.15	B non- <i>rmpA/rmpA2</i> K54	–	–	+	–	–	–
P14_L2_K54_11.17	B non- <i>rmpA/rmpA2</i> K54	+	–	–	–	–	–
P23_L4_28.16	C non-K54	–	–	+	–	–	–
P24_NW2_09.17	C non-K54	–	–	+	–	–	–
NCTC 9159	C non-K54	–	–	–	–	–	–
K54 isolates; other types (four different types ^c)	Non-CG29 K54	–	–	–	–	–	–

^aC0077_0530, C0077_05255 and C0077_04790 were found in the LS358 chromosome, but not in those of KP-1, NCTC 9159, NTUH-K2044 and RJF999.

^bRJF9_19740, RJF9_19755 and RJF9_19640 were found in the RJF999 chromosome, but not in those of LS358, KP-1, NCTC 9159 and NTUH-K2044.

^cVNTR profiles 4,1,3,2,2,2,3,5,2,4 (corresponding to ST530); 6,3,4,2,1,2,3,4,3,3 (corresponding to ST240); 1,8,2,4,0,2,2,5,3,2,3 and 1,7,4,15,2,2,1,4,5,2,3.

found. None of these genes were detected in the four non-CG29 K54 isolates tested.

None of the representatives of the 12 other VNTR types of K54 on our database (from 16 patients) were PCR-positive for *rmpA/rmpA2*, although they did include seven isolates from blood. Most of the other types were from only one or two patients with the next largest group in the K54 set (after CG29) being from three patients only (in each case from urine) and corresponding to ST530 (VNTR profile 4,1,3,2,2,2,3,5,2,4).

Discussion

It is likely that the K54 isolates of CG29 with the virulence plasmid (VNTR profiles 7,3,1,5,1,2,4,1,2/3,2,4) in this study were associated with the hypervirulent phenotype, all being isolated from blood or pus, with clinical details of pyrexia, sepsis or liver abscesses (5/5). Where determined, these isolates all had a repeat number of 2 or 3 at the ninth VNTR locus, in contrast to those without the virulence plasmid, which exclusively had a repeat of 4 at that locus (Table 1). Some of the K54 isolates of CG29 without a virulence plasmid were also isolated from blood and associated with bacteraemia or sepsis (5/10). Similarly, among the non-K54 representatives, there were examples of blood/pus/pleural fluid isolates associated with bacteraemia/abscesses/sepsis, suggesting that the K54 capsular type in itself may not be of great importance, other than that it alerts the investigator to a potentially hypervirulent clonal type. However, it was only isolates of the K54 capsular type that carried the virulence plasmid. While host factors and the clinical procedures carried out will undoubtedly also have been important, these results suggest that the clonal type is a major factor, with the virulence plasmid being an important contributory (but not necessarily essential) factor. It must also be borne in mind that the gastro-intestinal tract is the likely source of the organism in hypervirulent infections, with individuals being colonised prior

to infection, so that isolation from screening swabs is not necessarily inconsistent with a hypervirulent type [2]. Similarly, it has previously been observed that virulence plasmid-cured Kp52.145 (a hypervirulent strain of capsular type K2) remained more virulent than other strains not carrying the plasmid [27], again suggesting that higher pathogenicity is correlated with clonal type, even when the virulence plasmid is absent. However, this goes against observations that suggest that the virulence plasmid is a requirement for hypervirulence, *rmpA/rmpA2* and aerobactin, in particular being strongly associated with the hypervirulent syndrome [10, 28, 29]. *RmpA* can be found chromosomally, but it is unlikely that this was the case in this study, since we failed to detect it in our PCRs using two different sets of primers among those isolates with no evidence of a virulence plasmid, nor was it detected in the whole-genome sequences, even searching using the chromosomal-*rmpA*-specific primers described by Hsu *et al.* [10].

However, it would not be true to say that all representatives of CG29, whether of capsular type K54 or not, are associated with virulence, NCTC 9159 (isolated from urine) and KP-1 perhaps being good examples of this, and we therefore set out to compare KpvK54 with NCTC 9159, KP-1, LS358 and hypervirulent representatives of K1-ST23 NTUH-K2044 and RJF999. Despite being remarkably similar to KpvK54, NCTC 9159 lacked the very factors (virulence plasmid, yersiniabactin cluster) that have been identified as important to hypervirulence. Indeed, among the chromosomal genes found in the six isolates we compared, it was almost exclusively the yersiniabactin cluster genes that were common to those most associated with virulence (KpvK54, NTUH-K2044, RJF999 and LS358), but not present in NCTC 9159 and the environmental isolate KP-1. While the yersiniabactin cluster has long been associated with virulence, it is now found commonly in clinical isolates of *K. pneumoniae* [13, 26], so one might expect that other chromosomal elements may also be important. KpvK54 also shared other genes with hypervirulent K1-ST23 isolates (e.g. the *mobB* and

mobC genes encoding mobilisation proteins in the ICE element and a region corresponding to RJF9_19730 to RJF9_19760), not present in NCTC 9159, which may be important. There were also genes found only among the K54 isolates of CG29 among the six genomes compared, some only found among the two K54 isolates associated with virulence (KpvK54 and LS358) (e.g. C0077_05255), which may play a role and provide further explanation for the association of K54 with virulence. Notably, these were all also present in isolate KP10 (CP025091), a further representative of K54-ST29 carrying the yersiniabactin cluster, and were also mostly present in KP14 (CP025093), even when there were few other matches/near-matches. However, we do not know any clinical information concerning these isolates. All three such genes sought by PCR (C0077_05300 (encoding lysozyme), C0077_05255 and C0077_04790) were detected in all four of our *rmpA/rmpA2*-positive isolates of K54 available for testing. According to InterPro, C0077_05255 is an Ead/Ea22-like protein, belonging to a family containing phage proteins and bacterial proteins likely to represent integrated phage proteins, while C0077_04790 may be involved in bacteriophage resistance. Our approach concentrated on the presence or absence of coding sequences, but changes in those sequences and indeed in non-coding sequences may also have a profound effect on virulence, as evidenced by recent findings with *Salmonella enterica* serovar Typhimurium ST313 [30].

Our observations add to the already existing body of evidence identifying CG29 as a hypervirulent type that could explain the association of capsular type K54 with virulence. K54-positive representatives of CG29 were found in similar numbers among our isolates as the hypervirulent ST86 type belonging to K2 (from 14 to 15/4670 patients), although representatives with a virulence plasmid were rarer (from 5/4670 patients). However, the situation is complex; CG29 isolates displayed a range of virulence, from environmental through to hypervirulence, with those most consistently associated with virulence being of capsular type K54 and possessing a virulence plasmid containing both the aerobactin and salmochelin-acquired siderophore systems, showing that it is important to at least characterise isolates for these elements. These all shared a VNTR profile of 7,3,1,5,1,2,4,1,2/3,2,4. Our work suggests that they may also be recognised by the detection of the C0077_05255 marker. ‘Switching’ of capsular type was observed, highlighting that these types may be missed if screened by capsular typing alone.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268818001826>

Acknowledgements. The authors thank the team of the curators of the Institut Pasteur MLST system (Paris, France) for importing novel alleles, profiles and/or isolates at <http://bigsd.db.pasteur.fr>, Oxford Nanopore Technologies for providing a MIN107 flow cell and an SQK-LSK308 1D² sequencing kit and staff in UK hospitals for sending isolates to the authors. They thank colleagues in AMRHAI for carbapenemase detection and colleagues in the Genomic Services and Development Unit for Sanger sequencing. The authors are also grateful to the NCTC 3000 project team for providing the whole-genome sequence of NCTC 9159.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

References

- Pan YJ *et al.* (2015) Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella* spp. *Scientific Reports* **5**, 15573.
- Paczosa MK and Meccas J (2016) *Klebsiella pneumoniae*: going on the offense with a strong defense. *Microbiology and Molecular Biology Reviews* **80**, 629–661.
- Hsu CR *et al.* (2016) Identification of a capsular variant and characterization of capsular acetylation in *Klebsiella pneumoniae* PLA-associated type K57. *Scientific Reports* **6**, 31946.
- Bialek-Davenet S *et al.* (2014) Genomic definition of hypervirulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. *Emerging Infectious Diseases* **20**, 1812–1820.
- Holt KE *et al.* (2015) Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proceedings of the National Academy of Sciences of the USA* **112**, E3574–3581.
- Iwasaki Y *et al.* (2017) Bacterial meningitis caused by hypervirulent *Klebsiella pneumoniae* capsular genotype K54 with development of granuloma-like nodal enhancement in the brain during the subacute phase. *Internal Medicine* **56**, 373–376.
- Chuang C *et al.* (2016) The emergence of *Klebsiella pneumoniae* liver abscess in non-diabetic patients and the distribution of capsular types. *Gut Pathogens* **8**, 46.
- Yan Q *et al.* (2016) Hypervirulent *Klebsiella pneumoniae* induced ventilator-associated pneumonia in mechanically ventilated patients in China. *European Journal of Clinical Microbiology & Infectious Diseases* **35**, 387–396.
- Lin Y-T *et al.* (2015) Community-onset *Klebsiella pneumoniae* pneumonia in Taiwan: clinical features of the disease and associated microbiological characteristics of isolates from pneumonia and nasopharynx. *Frontiers in Microbiology* **6**, 122.
- Hsu CR *et al.* (2011) The role of *Klebsiella pneumoniae rmpA* in capsular polysaccharide synthesis and virulence revisited. *Microbiology* **157**, 3446–3457.
- Public Health England Culture Collections NCTC 9159. Available at <http://www.phe-culturecollections.org.uk/products/bacteria/detail.jsp?refId=NCTC+9159&collection=nctc> (Accessed 25 July 2017).
- Public Health England Culture Collections NCTC 3000 Project. Available at <https://www.phe-culturecollections.org.uk/collections/nctc-3000-project.aspx> (Accessed 25 July 2017).
- Turton JF *et al.* (2018) Virulence genes in isolates of *Klebsiella pneumoniae* from the UK during 2016, including among carbapenemase gene-positive hypervirulent K1-ST23 and ‘non-hypervirulent’ types ST147, ST15 and ST383. *Journal of Medical Microbiology* **67**, 118–128.
- Turton JF *et al.* (2010) PCR characterization and typing of *Klebsiella pneumoniae* using capsular type-specific, variable number tandem repeat and virulence gene targets. *Journal of Medical Microbiology* **59**, 541–547.
- Compain F *et al.* (2014) Multiplex PCR for detection of seven virulence factors and K1/K2 capsular serotypes of *Klebsiella pneumoniae*. *Journal of Clinical Microbiology* **52**, 4377–4380.
- Pasteur MLST *Klebsiella* Sequence Typing Download allele sequences Available at http://bigsd.db.pasteur.fr/perl/bigsd/db/bigsd.pl?db=pubmlst_klebsiella_seqdef_public&page=downloadAlleles (Accessed 6 July 2017).
- Diancourt L *et al.* (2005) Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *Journal of Clinical Microbiology* **43**, 4178–4182.
- Vaser R *et al.* (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* **27**, 737–746.
- Li H (2016) Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics (Oxford, England)* **32**, 2103–2110.
- Koren S *et al.* (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27**, 722–736.
- Kleborate. Available at <https://github.com/katholt/Kleborate> (Accessed 4 August 2017).
- Hyatt D *et al.* (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119.
- Fu L *et al.* (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)* **28**, 3150–3152.
- Klebsiella* sequence typing database. Available at <http://bigsd.db.pasteur.fr/klebsiella/klebsiella.html> (Accessed 15 August 2017).

25. **Lery LM *et al.*** (2014) Comparative analysis of *Klebsiella pneumoniae* genomes identifies a phospholipase D family protein as a novel virulence factor. *BMC Biology* **12**, 41.
26. **Lam MMC *et al.*** Frequent emergence of pathogenic lineages of *Klebsiella pneumoniae* via mobilisation of yersiniabactin and colibactin. bioRxiv preprint first posted online 4 January 2017. doi: <http://dx.doi.org/10.1101/098178>.
27. **Nassif X and Sansonetti PJ** (1986) Correlation of the virulence of *Klebsiella pneumoniae* K1 and K2 with the presence of a plasmid encoding aerobactin. *Infection and Immunity* **54**, 603–608.
28. **Tang HL *et al.*** (2010) Correlation between *Klebsiella pneumoniae* carrying pLVPK-derived loci and abscess formation. *European Journal of Clinical Microbiology & Infectious Diseases* **29**, 689–698.
29. **Russo TA *et al.*** (2015) Aerobactin, but not yersiniabactin, salmochelin, or enterobactin, enables the growth/survival of hypervirulent (hypermucoviscous) *Klebsiella pneumoniae* ex vivo and in vivo. *Infection and Immunity* **83**, 3325–3333.
30. **Hammarlöf DL *et al.*** (2018) Role of a single noncoding nucleotide in the evolution of an epidemic African clade of *Salmonella*. *Proceedings of the National Academy of Sciences of the USA* **115**, E2614–E2623.