*Epidemiology and Infection*

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

J. F. Turton, Z. Payne, K. Micah and J. A. Turton.

Supplementary Table S1. Primers for six coding sequences sought on the basis of differential presence among KpvK54, NCTC9159, RJF999, NTUH-K2044, LS358 and KP-1.

Supplementary Table S2. Virulence factors in KpvK54 and NCTC 9159.

Supplementary Table S3. Gene matrix showing the number of genes common to each combination of isolates.

Supplementary Table S4. Coding sequences present in the chromosomes of KpvK54 ([CP023134](https://www.ncbi.nlm.nih.gov/nuccore/CP023134)) and LS358 ([CP025629](https://www.ncbi.nlm.nih.gov/nuccore/CP025629)), but not in KP-1 (K54-ST29) or NCTC 9159 (K39-CG29), NTUH-K2044 or RJF999 (K1-ST23).

Supplementary Table S1. Primers for six coding sequences sought on the basis of differential presence among KpvK54, NCTC9159, RJF999, NTUH-K2044, LS358 and KP-1.

|  |  |  |
| --- | --- | --- |
| Gene | Primers | Product size (bp) |
| C0077\_05300 (end section) | GGCAGTAACGGAAAAAGCGACTAAAGCGCAGCCCTTAAATG | 177 |
| C0077\_05255 (including both end sections) | AAGCGGCAGCAGAACACTACCCTCCACCTTGATGCCAG | 546 |
| C0077\_04790 | CCGAGATCATCTTGGTTGCCTGTCCAAGCAAGGCAGC | 565 |
| RJF9\_19740 | GTGGTGCCTGTATCGATGAGCAACCGGGCGAAATACTC | 199 |
| RJF9\_19755 | CCAGAGGAATGGTGAGATCCTCGCCTGCGAAGAATTCTC | 122 |
| RJF9\_19640 | TCACCACCCTGACGCTCTTTCCCGAAGCGCAAATCAACTC | 566 |

Supplementary Table S2. Virulence factors in KpvK54 and NCTC 9159. Entries in bold are on the virulence plasmid (utg000002c, GenBank accession number [CP023135](https://www.ncbi.nlm.nih.gov/nuccore/CP023135)) of KpvK54 (211454 bp). Utg000001c (5224722 bp, GenBank accession number [CP023134](https://www.ncbi.nlm.nih.gov/nuccore/CP023134)) and unitig\_0 (5460114 bp) refer to the chromosomes of KpvK54 and NCTC 9159, respectively. While nucleotide positions are given, the KpvK54 sequence is only 99% accurate and therefore may include/not include some bases that are not present/present. c=complement.

|  |  |  |
| --- | --- | --- |
| Gene | Position in KpvK54 | Position in NCTC 9159 |
| *allCDRS* | - | - |
| *arcC* | - | - |
| *clbABCDEFGHIJKLMNOPQR* | - | - |
| *fdrA* | - | - |
| *fyuA* | 2685198-2687215utg000001c  | - |
| *gcl* | - | - |
| *glxK* | - | - |
| *glxR* | - | - |
| *hyi* | - | - |
| ***iroB*** | **22034-23147 utg000002c**  | - |
| ***iroC*** | **23209-26933 utg000002c**  | - |
| ***iroD*** | **27038-28264 utg000002c**  | - |
| ***iroN*** | **28722-30895 utg000002c**  | - |
| *irp1* | 2672127-2681595utg000001c  | - |
| *irp2* | 2665939-2672040utg000001c  | - |
| ***iucA*** | **c135237-137025 utg000002c**  | - |
| ***iucB*** | **c134291-135236 utg000002c**  | - |
| ***iucC*** | **c132562-134291 utg000002c**  | - |
| ***iucD*** | **c131280-132558 utg000002c**  | - |
| ***iutA*** | **c129003-131199 utg000002c**  | - |
| *kfuABC* | - | - |
| KP1\_1364 | - | - |
| KP1\_1371 | - | - |
| *kvgAS* | - | - |
| *mceABCDEGHIJ* | - | - |
| *mrkA* | c3634838-3635446utg000001c  | c2335402 to 2336010 unitig\_0 |
| *mrkB* | c 3634036-3634742utg000001c  | c2334605-2335306 unitig\_0 |
| *mrkC* | c3631541-3634024 utg000001c  | c2332107-2334593unitig\_0  |
| *mrkD* | c3630557-3631550utg000001c  | c2331121-2332116 unitig\_0 |
| *mrkF* | c3629907-3630544utg000001c  | c2330472-2331107unitig\_0 |
| *mrkH* | 3647961-3648677 utg000001c  | 2328283-2328987unitig\_0 |
| *mrkI* | 3628428-3629013 utg000001c  | 2328993-2329577 unitig\_0 |
| *mrkJ* | 3629157-3629871utg000001c  | 2329721-2330437 unitig\_0 |
| *rmpA* | **c33233-33865 utg000002c**  | - |
| *rmpA2* | **123492-124125 utg000002c**  | - |
| *ybbWY* | - | - |
| *ybtA* | 2664791-2665750utg000001c  | - |
| *ybtE* | 2683491-2685067 utg000001c  | - |
| *ybtP* | c2662828-2664535 utg000001c  | - |
| *ybtQ* | c2661041-2662841utg000001c  | - |
| *ybtS* | c2658439-2659741utg000001c  | - |
| *ybtT* | 2682687-2683487utg000001c  | - |
| *ybtU* | 2681592-2682690utg000001c  | - |
| *ybtX* | c2659769-2661048 utg000001c  | - |
| *ylbEF* | - | - |
| ***pbrA*** | **178673-181070 utg000002c**  | c3783879 to 3786284unitig\_0 (89%) |
| ***pbrBC*** | **180995-182145 utg000002c**  | - |
| ***pbrR*** | **c178153-178587 utg000002c**  | 3786370 to 3786679 (71% coverage, 80% identity) |
| ***pcoA*** | **171663-173480 utg000002c**  | - |
| ***pcoB*** | **173477-174375 utg000002c (212438)** | - |
| ***pcoC*** | **174415-174794 utg000002c**  | - |
| ***pcoD*** | **174799-175724 utg000002c**  | - |
| ***pcoE*** | **c170987-171435 utg000002c**  | - |
| ***pcoR*** | **175773-176458 utg000002c** | - |
| ***pcoS*** | **176455-177854 utg000002c**  | - |
| ***silC*** | **160104-161640 utg000002c**  | - |
| ***silE*** | **c157245-157792 utg000002c**  | - |
| ***silR*** | **c159390-160070 utg000002c**  | - |
| ***silS*** | **c157926-159397 utg000002c**  | - |
| ***terA*** | **97035-98184 utg000002c**  | - |
| ***terB*** | **98207-98662 utg000002c**  | - |
| ***terC*** | **98686-99721 utg000002c**  | - |
| ***terD*** | **99760-100336 utg000002c**  | - |
| ***terE*** | **100441-101015 utg000002c**  | - |
| ***terW*** | **c89435-89893 utg000002c**  | - |
| ***terX*** | **c87647-88289 utg000002c**  | - |
| ***terY*** | **c88314-88952 utg000002c**  | - |
| ***terZ*** | **96455-97038 utg000002c**  | - |
| **hemin** | **145197-146230 utg000002c**  | - |
| **SAM-dependent methyltransferase** | **c34402-35065 utg000002c**  | - |
| ***luxR*** | **c32296-32849 utg000002c**  | - |
| ***pagO*** | **c31320-32223 utg000002c**  | - |
| ***ibrB*** | **c21045-21684 utg000002c**  | - |
| ***fecA*** | **c15254-17491 utg000002c**  | - |
| ***fecI*** | **c18415-18936 utg000002c**  | - |
| ***fecR*** | **c17898-18418 utg000002c**  | - |
| **Fe 3+ citrate ABC transporter** | **c15023-15178 utg000002c**  | - |
| ***shiF*** | **137091-138310 utg000002c**  | - |
| **lysozyme inhibitor** | **141122-141439 utg000002c**  | - |
| CCP4 like integrase | - | - |
| alginate lyase | - | - |
| ***cobW*** | **c204891-206112 utg000002c**  | - |
| ***fur*** | **c199858-200311 utg000002c**  | - but *fur* (KP1\_1659) at 4983602-4984051 unitig\_0 |
| *pld1* (phospholipase D family protein) | - | - |
| *blaSHV-83* (beta-lactamase) | c1769127-1769982 utg000001c  | c524853 to 525711unitig\_0 |
| *OqxA* and *OqxB* (Fluoroquinolone) | 3321035-3325925 utg000001c  | 2002880-2007777unitig\_0  |
| *fosA* (fosfomycin) | 5191381-5191800utg000001c  | - |
| *strA*/*strB* | - | 1863278 -1864026 unitig\_0 |
| *wzc* | wzc\_54 c2792001-2792115 utg000001c  | wzc\_39 c1492727-1492859 unitig\_0  |
| *wzc\_CR2* | wzc\_CR2\_75 c2791902-2792000 utg000001c (99% identity) | wzc\_CR2\_21 c1492627-1492726unitig\_0  |
| *wzi* | wzi\_115 (446/447 identities)c2796228-2796674utg000001c  | wzi\_39c1498032-1498478 untig\_0  |

Supplementary Table S3. Gene matrix showing the number of genes common to each combination of isolates; combinations are marked by a 1 where that isolate was included and a 0 where it was not (e.g. one gene was found only in the NTUH-K2044, LS358, RJF999 and NCTC 9159 combination, 79 genes were found in KpvK54 alone and 5081 genes were common to all six isolates). NTUH-K2044 and RJF999 belong to K1-ST23, KpvK54, LS358 and KP-1 belong to K54-CG29, while NCTC 9159 belongs to CG29, but capsular type K39. KP-1 is an environmental isolate. A 90% similarity threshold was used to cluster genes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| NTUH-K2044 | KpvK54 | KP-1 | LS358 | RJF999 | NCTC 9159 | **Count** |
| 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| 1 | 0 | 0 | 1 | 1 | 0 | 1 |
| 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 2 |
| 1 | 1 | 1 | 1 | 0 | 1 | 2 |
| 1 | 1 | 0 | 1 | 1 | 1 | 2 |
| 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| 1 | 0 | 0 | 0 | 1 | 1 | 2 |
| 0 | 0 | 1 | 1 | 0 | 0 | 3 |
| 1 | 0 | 1 | 0 | 1 | 1 | 3 |
| 0 | 1 | 1 | 0 | 1 | 0 | 4 |
| 0 | 1 | 1 | 1 | 1 | 1 | 5 |
| 1 | 0 | 1 | 1 | 1 | 1 | 5 |
| 1 | 1 | 0 | 0 | 0 | 0 | 5 |
| 0 | 0 | 1 | 0 | 1 | 0 | 8 |
| 0 | 1 | 0 | 1 | 0 | 1 | 9 |
| 1 | 1 | 1 | 1 | 1 | 0 | 10 |
| 0 | 0 | 1 | 0 | 0 | 1 | 11 |
| 1 | 1 | 0 | 1 | 1 | 0 | 13 |
| 0 | 1 | 1 | 1 | 0 | 0 | 15 |
| 0 | 1 | 0 | 1 | 0 | 0 | 15 |
| 1 | 1 | 0 | 0 | 1 | 0 | 17 |
| 0 | 1 | 0 | 0 | 1 | 0 | 26 |
| 0 | 1 | 0 | 0 | 0 | 1 | 26 |
| 0 | 1 | 1 | 0 | 1 | 1 | 28 |
| 0 | 0 | 0 | 0 | 1 | 1 | 34 |
| 0 | 1 | 1 | 0 | 0 | 1 | 41 |
| 1 | 1 | 1 | 0 | 1 | 1 | 58 |
| 1 | 0 | 0 | 0 | 0 | 0 | 59 |
| 0 | 1 | 1 | 0 | 0 | 0 | 60 |
| 0 | 1 | 0 | 0 | 0 | 0 | 79 |
| 0 | 0 | 0 | 1 | 0 | 0 | 88 |
| 0 | 0 | 1 | 0 | 0 | 0 | 92 |
| 0 | 0 | 0 | 0 | 1 | 0 | 153 |
| 0 | 1 | 1 | 1 | 0 | 1 | 183 |
| 1 | 0 | 0 | 0 | 1 | 0 | 335 |
| 0 | 0 | 0 | 0 | 0 | 1 | 431 |
| 1 | 1 | 1 | 1 | 1 | 1 | 5081 |

Supplementary Table S4. Coding sequences present in the chromosomes of KpvK54 ([CP023134](https://www.ncbi.nlm.nih.gov/nuccore/CP023134)) and LS358 ([CP025629](https://www.ncbi.nlm.nih.gov/nuccore/CP025629)) (both of capsular type K54 and clonal group 29), but not in KP-1 (K54-ST29) or NCTC 9159 (K39-CG29), NTUH-K2044 or RJF999 (K1-ST23). All the genes listed were outside the K54 capsular polysaccharide synthesis region (nt 2775659 to 2799701 in KpvK54).

|  |  |  |  |
| --- | --- | --- | --- |
| LS358 locus tag/genea | position in KpvK54 corresponding to LS358 locus tag from BLAST  | Function/product | Comments |
| C0077\_04785 K54\_531391\_531561 K54\_531862\_532479 | 531391 to 532479 | integrase | frameshifted;internal stop |
| C0077\_04790 K54\_532653\_533369 | 532550 to 533369 | hypothetical protein |  |
| C0077\_05255 K54\_3186020\_3186607 | cb 3186020 to 3186607 | hypothetical protein |  |
| C0077\_05290 | c 2466780 to 2466996 | hypothetical protein |  |
| C0077\_05295 K54\_2466340\_2466654 | c 2466340 to 2466654 | hypothetical protein |  |
| C0077\_05300 K54\_2465933\_2466337 | c 2465833 to 2466337 | lysozyme |  |
| C0077\_05305 K54\_2465411\_2465617 K54\_2465621\_2465836 | c 2465433 to 2465836 | lysis protein |  |
| C0077\_23115 K54\_4385220\_4385933 | 4385259 to 4386435 | integrase |  |
| C0077\_23120 | 4391804 to 4392410 | DUF1768 domain-containing protein |  |

a Five genes were omitted (C0077\_04415 encoding an amidohydrolase, C0077\_05160 encoding conjugal transfer protein TraR, C0077\_10035 encoding an aspartate aminotransferase, C0077\_12400 encoding an alkane 1-monooxygenase and C0077\_15605 encoding xylulokinase), since there was a significant match with KP-1 (>80 % coverage, despite sections of sequence being omitted/added). K54\_2479479\_2479613 was also not included, since the region was intergenic (between C0077\_05155 and C0077\_05160).

b c stands for complement.