

1 **Resistance mechanisms and population structure of highly drug resistant**
2 ***Klebsiella* in Pakistan during the introduction of the carbapenemase**
3 **NDM-1**

4 Eva Heinz^{a*}

5 Hasan Ejaz^{b,c,d}

6 Josefin Bartholdson Scott^e

7 Nancy Wang^c

8 Shruti Gujran^c

9 Derek Pickard^a

10 Jonathan Wilksch^c

11 Hanwei Cao^c

12 Ikram-ul-Haq^f

13 Gordon Dougan^{a,e}

14 Richard A Strugnell^{c*}

15

16 a: Parasites and Microbes, Wellcome Trust Sanger Institute, Hinxton, CB10
17 1SA, UK

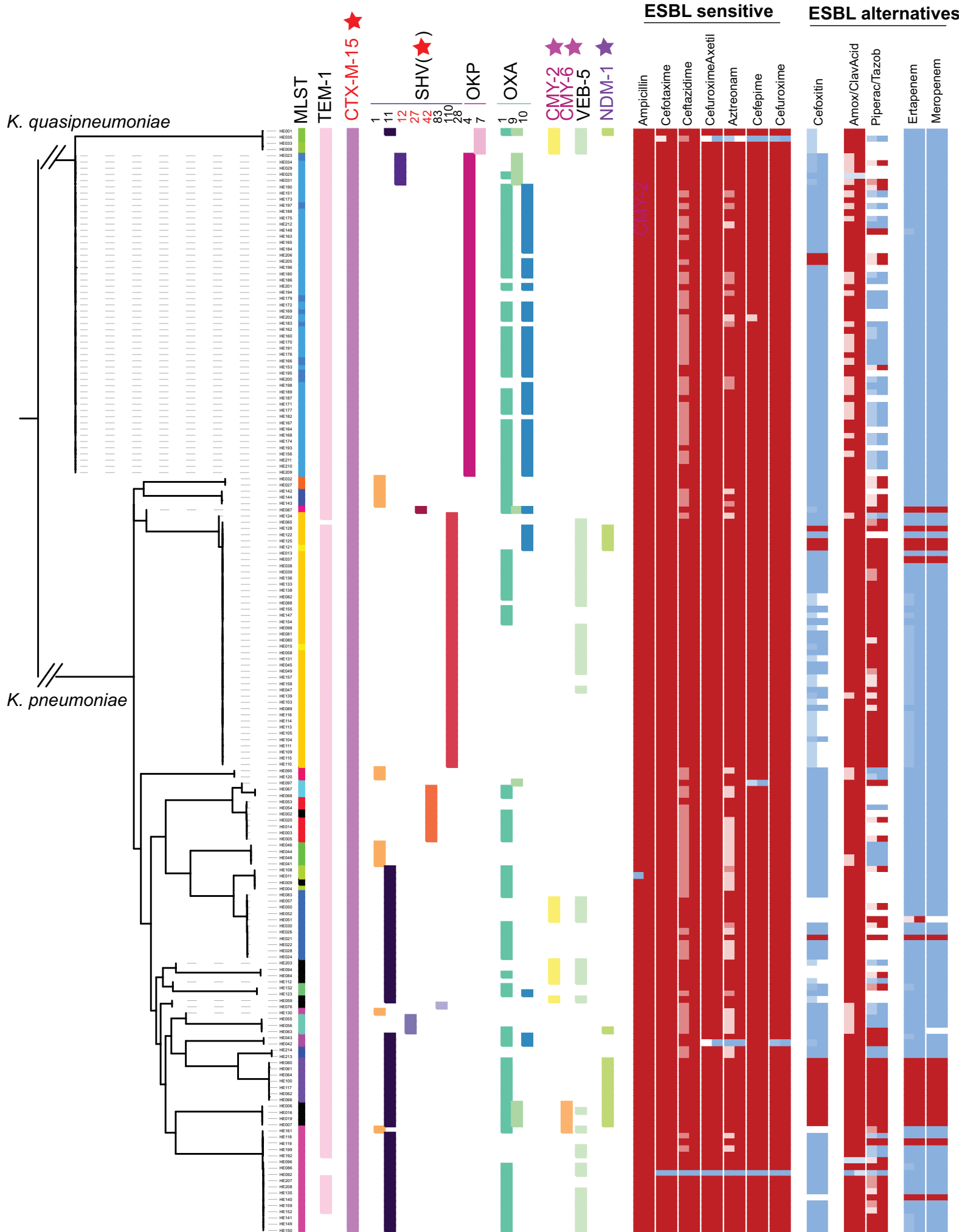
18 b: Department of Clinical Laboratory Sciences, CAMS, Jouf University, Al-Jouf,
19 Kingdom of Saudi Arabia

20 c: Department of Microbiology and Immunology, The University of Melbourne,
21 at Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

22 d: Department of Microbiology, The Children's Hospital & The Institute of Child
23 Health, Lahore, Pakistan

24 e: Department of Medicine, University of Cambridge, Cambridge, UK

25 f: Institute of Industrial Biotechnology, GC University, Lahore, Pakistan



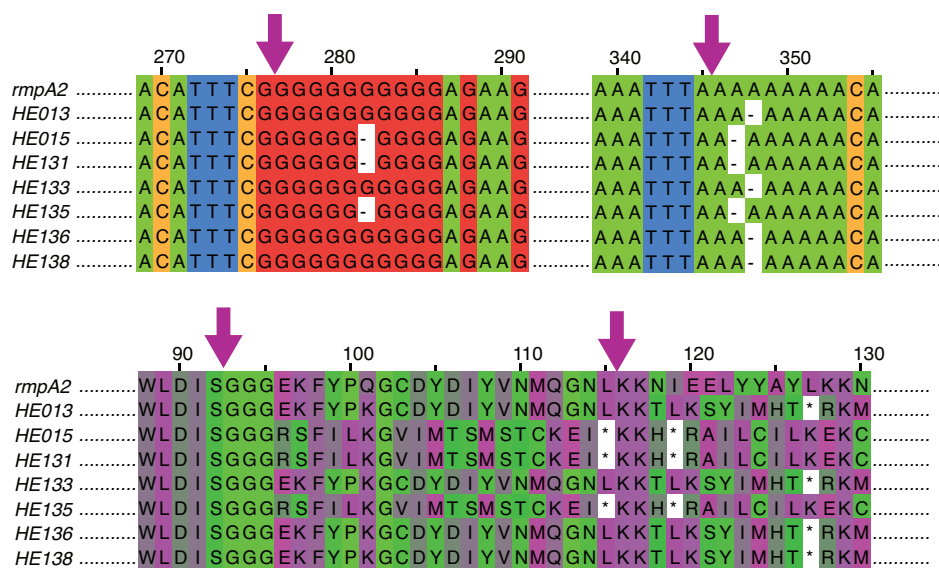
★ ESBL

★ Carbapenemase if porins deactivated

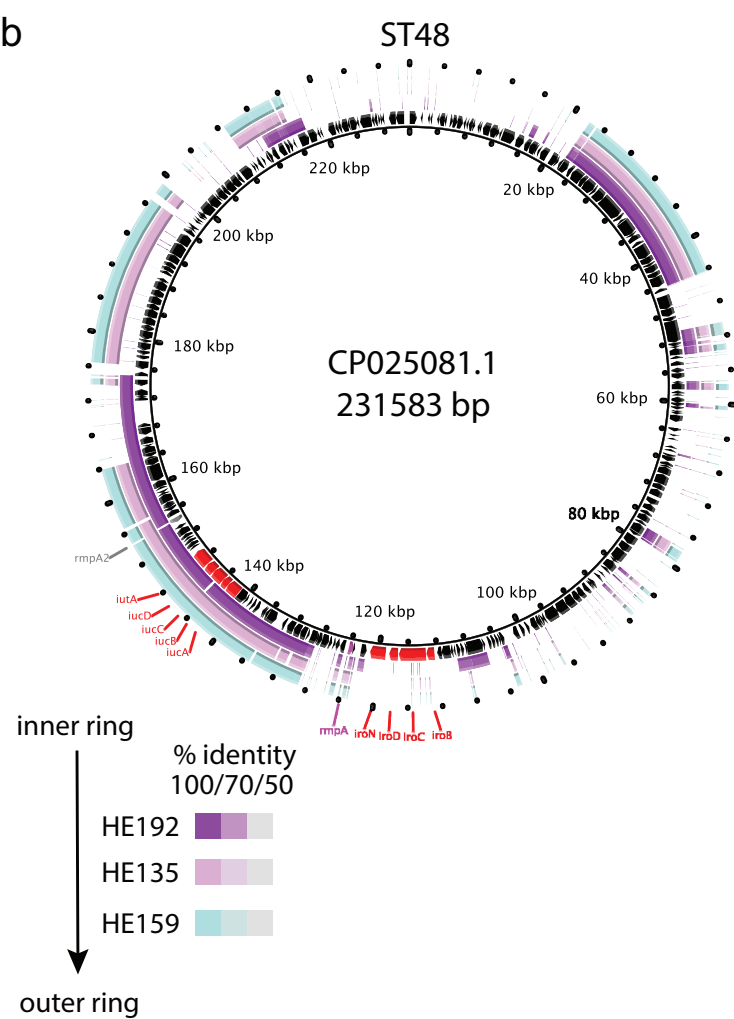
★ Carbapenemase

Fig. S1: Details of predicted beta-lactamase enzymes and resistance profiles to treatment alternatives for ESBL organisms. The guidance tree is as in Fig. 1. All isolates with TEM or CTX genes encode for the broad-spectrum beta-lactamase TEM-1 and the ESBL enzyme CTX-M-15, respectively, and only the VEB-5 and NDM-1 allele, and two alleles for the AmpC CMY (2 and 6) could be identified. *Bla*-SHV can have different activity spectrums; the red stars indicate extended-spectrum beta-lactamase activity. Pink indicates weak carbapenemase activity which needs additional mutations (e.g. porin deactivation) to confer resistance, dark violet denotes carbapenemase activity. The allele assignment was controlled at the beta-lactamase online database (<http://www.laced.uni-stuttgart.de>); activity assignments are according to⁹⁷. Vitek measurements are compared with the predicted known resistance-conferring genes for aminoglycosides, fluoroquinolones, sulfonamides and beta-lactams (including carbapenems). Red indicates resistant, white indicates intermediate, blue indicates sensitive measurement; details are given in Table S2.

a



b



c

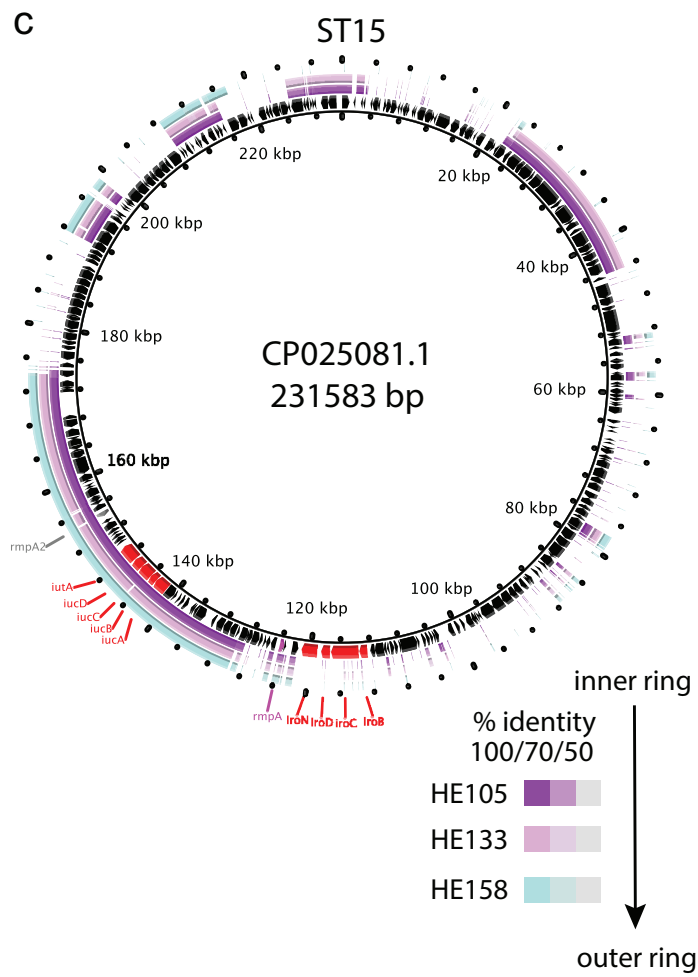


Fig. S2: Comparison of isolates with *rmpA2* pseudogenes to *Klebsiella* virulence plasmids. **(a)** The sequences retrieved from ariba predicted as *rmpA2* pseudogenes were retrieved and aligned to a reference sequence (amino acids NP_943354.1 and the respective CDS). Two homopolymer regions, a poly-G and a poly-A repeat, show mutations that lead to subsequent frameshifts and stop codons in the translated sequence. The arrows mark the corresponding positions in the amino acid and nucleotide alignments to facilitate orientation. Colours, nucleotides (upper panel); helix propensity (lower panel); as implemented in JalView⁹⁸ **(b)** The relevant Illumina contigs were extracted using blastn and abacas, and mapped against the reference plasmid pSGH10^{70,48} using BRIG⁴⁹ to further illustrate the partial conversation of the virulence plasmid also in the assembled data.

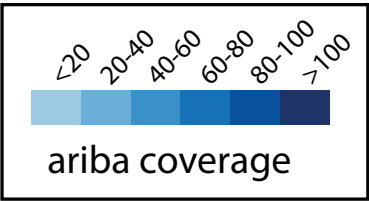
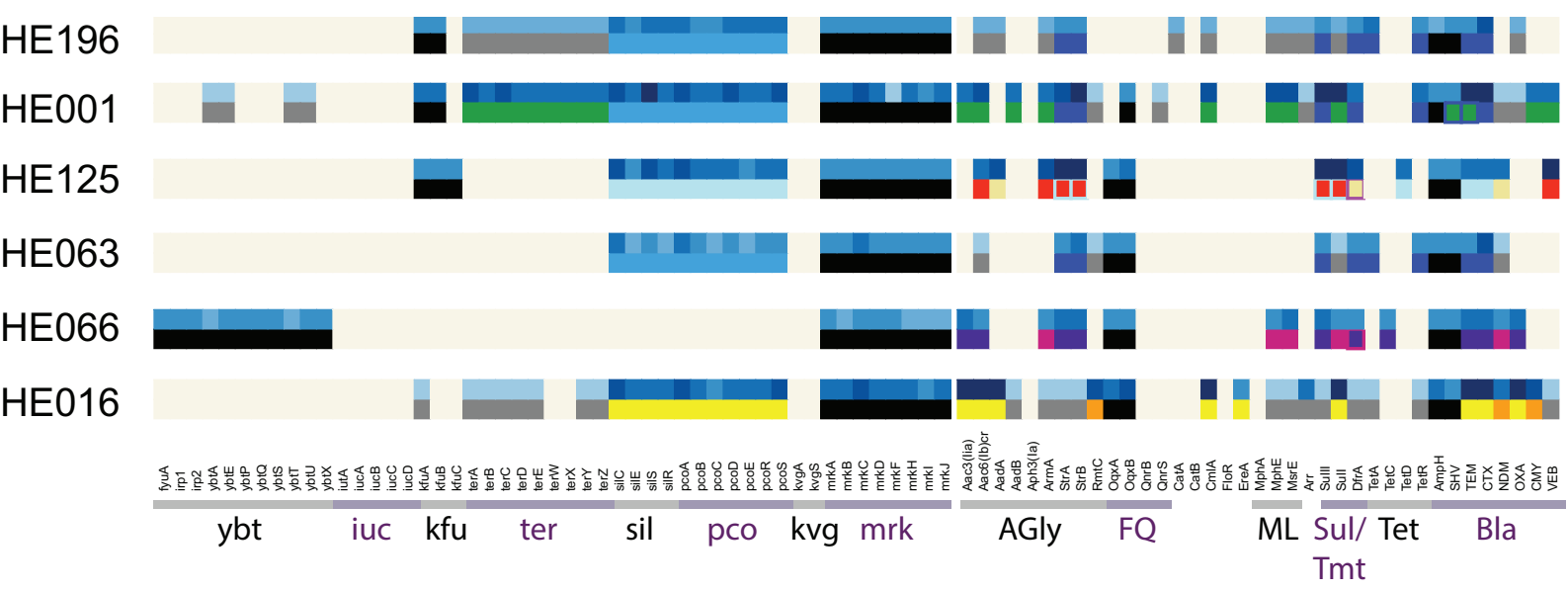


Fig. S3: The plasmid diversity for representative isolates. The contigs obtained from PacBio for selected strains are shown; plasmid replicons as predicted by the PlasmidFinder web server (<https://cge.cbs.dtu.dk/services/PlasmidFinder/>)⁹³. An overview of the resistance genes as predicted by ariba is shown on a colour gradient per strain (upper layer) displaying the read coverage of the hit; and the layer below these indicate on which plasmid the respective genes were found (colours according to the cell overviews, lower panel). Black indicates chromosomal, dark grey indicates gene only found in Illumina reads but not in the PacBio data. The presence of genes was investigated by blastn, only complete perfect (100% identity) hits were considered.

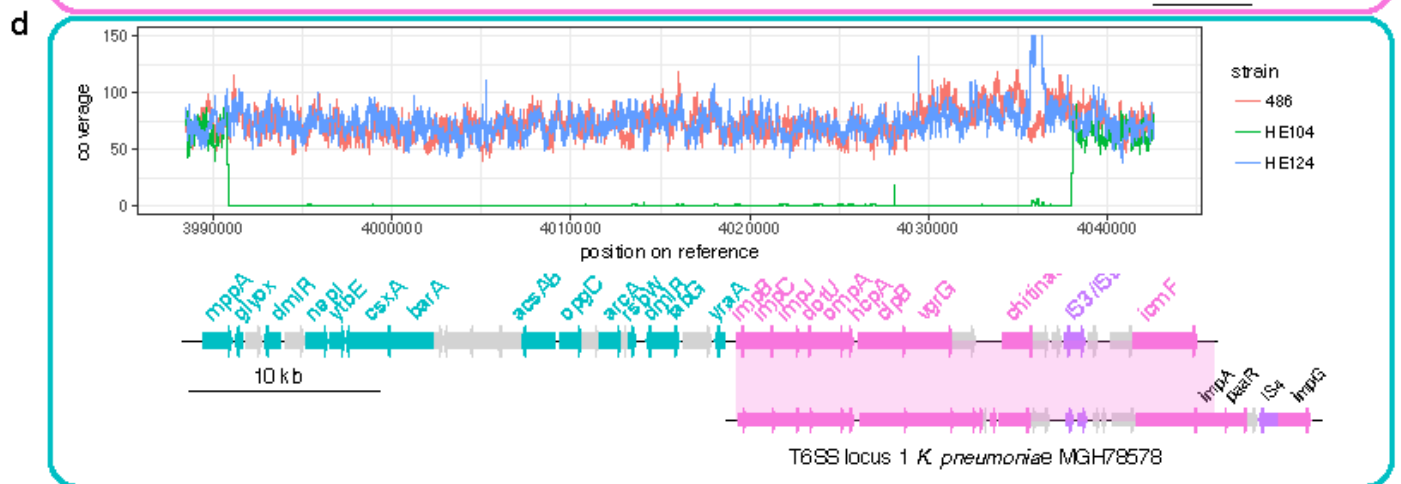
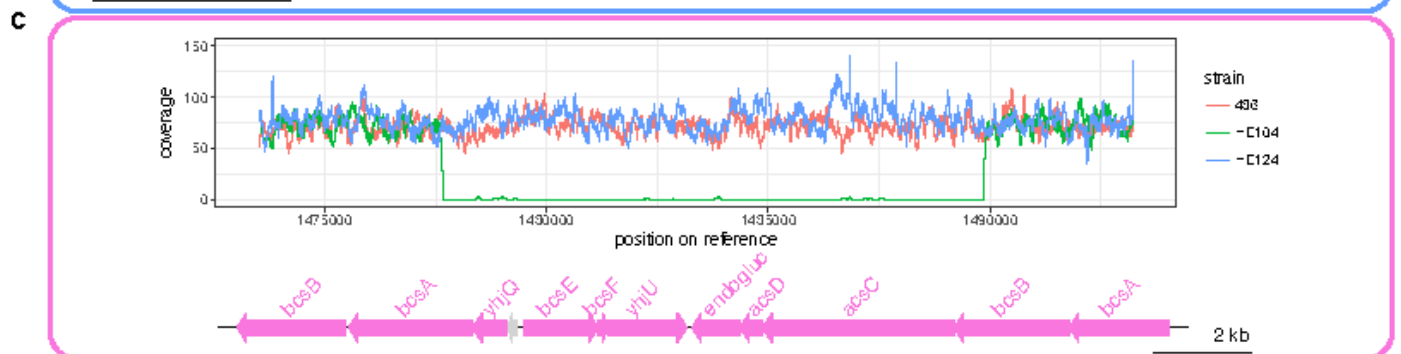
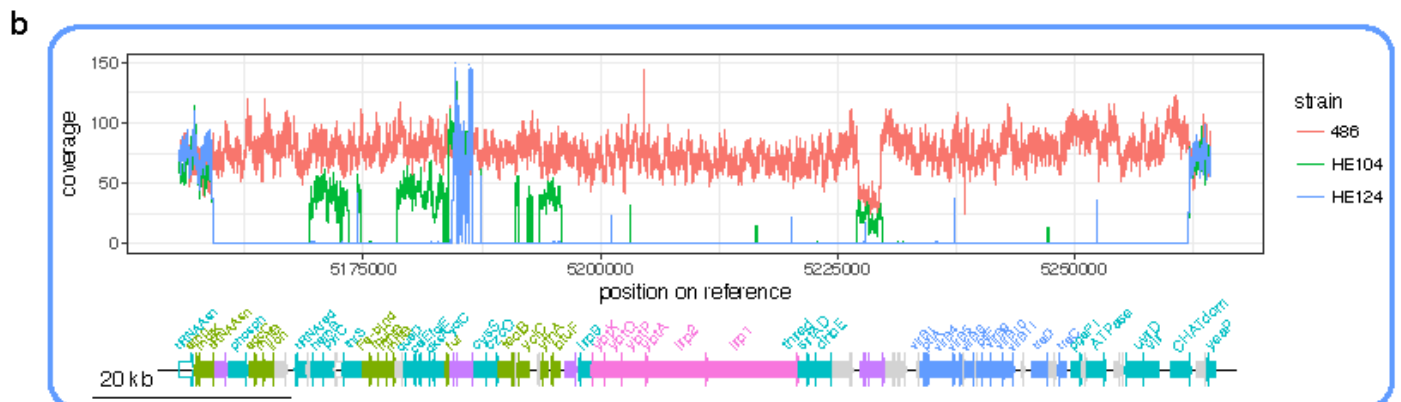
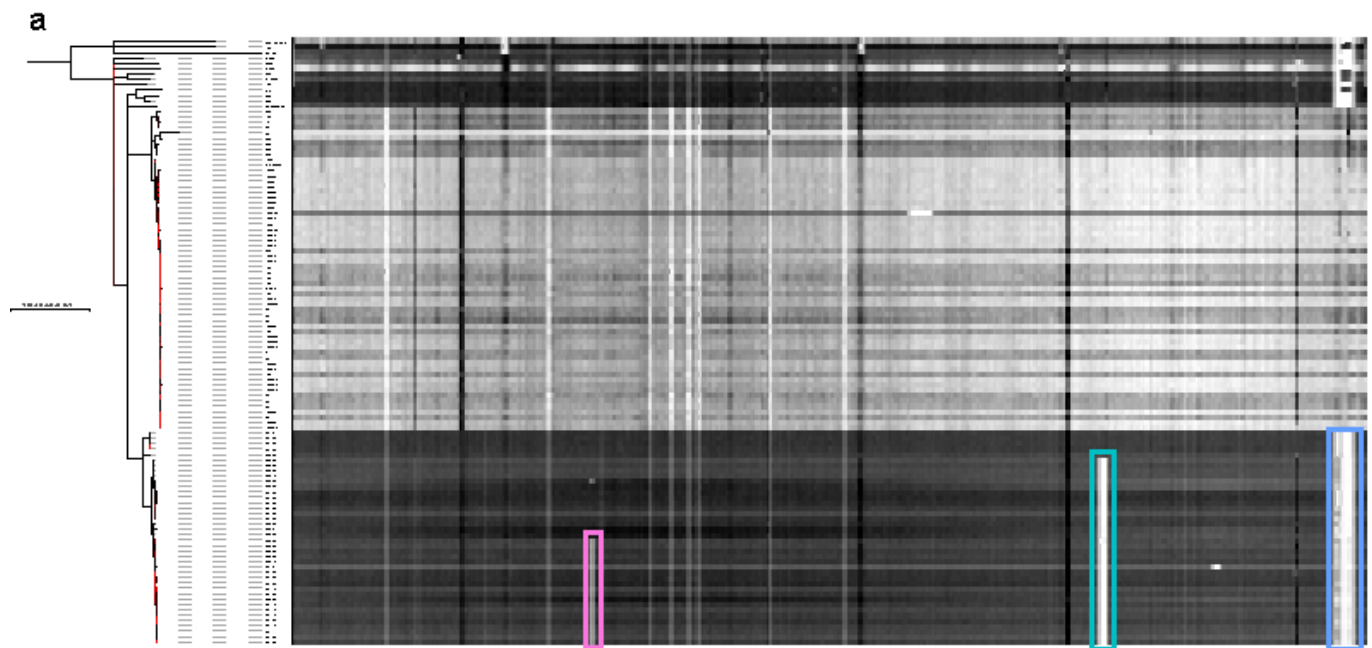


Fig. S4: In-depth comparison of expanded lineages. (a) The isolates belonging to ST15 of this study was compared to the ST15 outbreak lineage from Nepal^{59,60} by mapping against the completed genome of an isolate from the Nepal outbreak⁵⁹, which also carried four plasmids. The heatmap shows the mapping coverage across the chromosome before removal of recombination from white (low coverage) to black (high coverage), emphasizing the large deletions and acquisitions in the Pakistan and Nepal line, respectively. Even coverage of the genome could be observed, with three genetic islands missing in different subsets of our isolates, indicating the dynamic within this closely related lineage. The tree was constructed after removing recombination by gubbins, bootstrap values are depicted as colour gradient; from red (0) to black (100). **(b)** A detailed view of two regions in the genome absent in groups within the isolates from this study (plotted areas: 1473529-1493215, region 1; 3988478-4042608, region 2; coordinates from CP008929.1), but present in the lineage from Nepal, as well as other isolates from the same ST but not part of the clonal lineages as outgroups. In addition, the Nepal lineage acquired the chromosomal locus including yersiniabactin⁴²; plotted region 5155581-5264456); these three regions were also removed by gubbins before the tree construction.

Fig. S5: Comparison of the plasmid profiles between the Pakistan and Nepal ST15 lineage. The full genome sequence of a representative strain from the Nepal outbreak⁵⁹ was used as reference, and all five plasmid contigs from HE125, a representative of the related ST15 lineage from Nepal, were compared against the four plasmids as annotated (plasmid A = CP008930.1, B = CP008931.1, C = CP008932.1, D = CP008933.1). The legend indicates similarity values for the different plasmid contigs from the HE125 strain.

Table S2: Details on the strains from Pakistan.

Table S1: Vitek measurements.

Table S3: Additional strains included in the detailed analysis of ST15.