nature communications

Article

https://doi.org/10.1038/s41467-024-51374-x

Clonal background and routes of plasmid transmission underlie antimicrobial resistance features of bloodstream *Klebsiella pneumoniae*

Received: 21 April 2024

Accepted: 7 August 2024

Published online: 14 August 2024

Check for updates

Odion O. Ikhimiukor[1] ✉, Nicole I. Zac Soligno[1], Ifeoluwa J. Akintayo[2], Michael M. Marcovici[1], Stephanie S. R. Souza[1], Adrienne Workman[3], Isabella W. Martin[3] & Cheryl P. Andam[1] ✉

Bloodstream infections caused by the opportunistic pathogen *Klebsiella pneumoniae* are associated with adverse health complications and high mortality rates. Antimicrobial resistance (AMR) limits available treatment options, thus exacerbating its public health and clinical burden. Here, we aim to elucidate the population structure of *K. pneumoniae* in bloodstream infections from a single medical center and the drivers that facilitate the dissemination of AMR. Analysis of 136 short-read genome sequences complemented with 12 long-read sequences shows the population consisting of 94 sequence types (STs) and 99 clonal groups, including globally distributed multidrug resistant and hypervirulent clones. In vitro antimicrobial susceptibility testing and in silico identification of AMR determinants reveal high concordance (90.44–100%) for aminoglycosides, beta-lactams, carbapenems, cephalosporins, quinolones, and sulfonamides. IncF plasmids mediate the clonal (within the same lineage) and horizontal (between lineages) transmission of the extended-spectrum beta-lactamase gene $bla_{\text{CTX-M-15}}$. Nearly identical plasmids are recovered from isolates over a span of two years indicating long-term persistence. The genetic determinants for hypervirulence are carried on plasmids exhibiting genomic rearrangement, loss, and/or truncation. Our findings highlight the importance of considering both the genetic background of host strains and the routes of plasmid transmission in understanding the spread of AMR in bloodstream infections.

The Gram-negative commensal bacterium *Klebsiella pneumoniae* typically inhabits human mucosal surfaces, including the gastrointestinal tract and upper respiratory tract[1]. It is also an opportunistic pathogen that is responsible for a wide range of nosocomial and community-acquired infections, which often lead to adverse clinical and health-economic outcomes[1,2]. The disease presentations of *K. pneumoniae* include pneumonia, bloodstream infection, meningitis, urinary tract infections, wound or soft tissue infections, and liver

[1]Department of Biological Sciences, State University of New York at Albany, Albany, NY, USA. [2]Institute for Infection Prevention and Hospital Epidemiology, Medical Centre, University of Freiburg, Freiburg, Germany. [3]Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA. ✉e-mail: oikhimiukor@albany.edu; candam@albany.edu

abscess[1]. *K. pneumoniae* can invade normally sterile sites of the body, such as the blood, through localized infectious diseases, surgical procedures, and the presence of catheters and other intravascular medical devices[3]. The entry and proliferation of bacteria in the bloodstream can lead to a strong immune response, including organ failure and dysfunction, that can be fatal if not treated promptly and appropriately[4,5]. Worldwide, *K. pneumoniae* is the third leading cause of bloodstream infections[6,7].

The burden of *K. pneumoniae* infections in hospitals is compounded by antimicrobial resistance (AMR), which greatly limits available treatment options and increases mortality rates especially for patients infected with invasive strains[8,9]. Some *K. pneumoniae* clones have become increasingly resistant to multiple antimicrobial agents and are designated as multidrug or extensively drug resistant[10,11]. Bacteria that harbor carbapenemases and extended spectrum beta-lactamases (ESBLs) are particularly notable because they render many beta-lactams ineffective[12]. The World Health Organization recognizes third-generation cephalosporin-resistant and carbapenem-resistant Enterobacteriaceae, including *K. pneumoniae*, as critical threats and for which new antibiotics are urgently needed[13]. Mortality rates of patients with bloodstream infections due to resistant *K. pneumoniae* are distressingly high, ranging from 17 to 34% for patients with ESBL-producing *K. pneumoniae*[14–16] and up to 42% for patients with carbapenem-resistant *K. pneumoniae*[17,18]. ESBL-producing *K. pneumoniae* have been reported in several hospital outbreaks[19–21] and their increasing prevalence may be linked to the overuse of expanded spectrum cephalosporins in healthcare settings[22]. A systematic analysis of combined literature reviews, hospital systems, laboratory data, and surveillance systems spanning 204 countries and territories reported over 50,000 deaths in 2019 alone that were attributed to third-generation cephalosporin resistance in *K. pneumoniae*[23]. Forecasting models suggest that if current trends in antimicrobial use continue, over half of invasive *K. pneumoniae* will become resistant to third-generation cephalosporins and carbapenems by 2030[24]. Despite the increasing threat of resistant *K. pneumoniae* and the constantly growing suite of high-risk clones, our understanding of the underlying bacterial and genetic features that promote AMR in bloodstream infections and subsequent spread within the hospital setting remains incomplete[25,26].

K. pneumoniae comprises a phenotypically and genetically diverse assemblage of clones[10]. It is broadly categorized into two pathotypes that exhibit distinct disease profiles. The so-called "classical" *K. pneumoniae* is a common cause of nosocomial infections and often harbor mobile plasmids encoding AMR genes[10,27]. Hypervirulent *K. pneumoniae* is a major cause of community-acquired infections in healthy individuals, including liver abscess, pneumonia and meningitis, endophthalmitis, and infections of the central nervous system[28]. Hypervirulent clones are often susceptible to most antimicrobial agents, but clones exhibiting both hypervirulence and multidrug resistance (referred to as convergent clones) have been described[29].

Here, we aim to elucidate the population structure of *K. pneumoniae* in bloodstream infections from a single medical center and the drivers that facilitate the dissemination of AMR. We analyzed a combination of 136 short-read genome sequences complemented with 12 long-read sequences of *K. pneumoniae* derived from surveillance of bloodstream infections at the Dartmouth-Hitchcock Medical Center (DHMC), New Hampshire, USA. Altogether, our results highlight the importance of considering both the genetic background of host strains and the routes of plasmid transmission in understanding the spread of AMR in bloodstream infections. Our findings highlight the need for improved infection prevention and control to reduce risk of healthcare-associated pathogen transmission, including the importance of genomic data surveillance to detect onward pathogen and plasmid transmission and persistence.

Results

The bloodstream *K. pneumoniae* population consists of diverse lineages

A total of 136 *K. pneumoniae* isolates from bloodstream infections in unique pediatric and adult patients at the DHMC were collected between January 2017 and January 2022 (Supplementary Data 1). The maximum likelihood phylogeny built from 188,166 single nucleotide polymorphisms (SNPs) extracted from a 3.378 Mbp sequence alignment of 3,511 core genes revealed a genetically diverse population. We identified 94 known sequence types (ST; Supplementary Data 2). The ST diversity is high (Simpson's diversity index = 0.983) and not one ST appears to dominate the population (Fig. 1A, B). Only nine STs contained at least three genomes (ST20, six genomes; ST36, ST45 and ST253, five genomes each; ST1380 and ST307, four genomes each; ST111, ST225, ST23 and ST941 with three genomes each). A total of 11 and 73 STs were represented by two and one genome(s), respectively. Thirteen genomes carried novel combinations of the 7-gene multilocus sequence typing (MLST) loci and were assigned novel ST designations by the *Klebsiella*-specific database in BIGSdb[30]: ST6357 (isolate KPB115), ST6358 (KPB126), ST6359 (KPB139), ST6360 (KPB179), ST6361 (KPB28), ST6362 (KPB29), ST6363 (KPB34), ST6364 (KPB41), ST6365 (KPB56), ST6366 (KPB79), ST6367 (KPB89), ST6368 (KPB90), and ST6369 (KPB93).

The 136 genomes can be further classified into 99 known clonal groups (CG; Simpson's diversity index = 0.986) and 76 known sublineages (SL; Simpson's diversity index = 0.971) (Fig. 1C,D and Supplementary Data 2). As defined previously[31], *K. pneumoniae* genomes belonging to the same clonal group differ by 43 allelic mismatches, while sublineages differ by 190 allelic mismatches. The most frequently detected clonal groups contained five genomes (CG45), four genomes (CG1380, CG20, CG307, CG36), and three genomes (CG10047, CG111, CG23, CG941). The most frequently detected sublineages were SL17 (consisting of 14 genomes from ST17, ST20, ST422, ST3640, ST5122, ST6364, ST6365), SL3010 (six genomes from ST1, ST5, ST6, ST8, ST9, ST10), SL45 (seven genomes from ST45 and ST987), and SL268 (six genomes from ST36 and ST268). From 2017 to 2021, less common STs, CGs and SLs comprise a large assemblage of the annual population. Overall, these results show that diverse genotypes may similarly cause bloodstream infections, thus emphasizing the opportunistic nature of *K. pneumoniae* in invasive infections.

Antimicrobial resistance is widespread in bloodstream *K. pneumoniae*

We carried out antimicrobial susceptibility testing of *K. pneumoniae* isolates against 20 antimicrobial agents from seven classes of antimicrobial drugs (Fig. 2A, Supplementary Data 1). All isolates exhibited resistance to ampicillin, which was not surprising because it is known to be intrinsic in *K. pneumoniae*[32]. Resistance was observed against ampicillin-sulbactam ($n = 16$ isolates, representing 11.76% of the population), cefuroxime ($n = 14$, 10.29%), sulphamethoxazole ($n = 14$, 10.29%), and cefazolin ($n = 12$, 8.82%). A total of 17 (12.5%), 16 (11.76%), seven (5.14%), and two (1.47%) isolates were resistant to at least one antimicrobial agent belonging to cephalosporin, beta-lactam combination agents, quinolone, and aminoglycosides sub/classes, respectively. However, all or nearly all isolates were susceptible to amikacin and meropenem ($n = 136$, 100%), ertapenem ($n = 135$, 99.26%), gentamicin ($n = 134$, 98.52%), cefoxitin ($n = 133$, 97.79%), amoxicillin-clavulanic acid and piperacillin/tazobactam ($n = 132$, 97.05%), levofloxacin ($n = 131$, 96.32). Altogether, we detected 18 unique AMR profiles, each with different combinations of resistance phenotypes (Fig. 2A). A total of 109 isolates were phenotypically resistant to ampicillin only and not to other antimicrobial agents. Isolates KPB140 and KPB97 exhibited resistance to 14 and 13 antimicrobial agents, respectively, while isolates KPB176 and KPB77 were resistant to 12 antimicrobial agents (Supplementary Data 1).

https://doi.org/10.1038/s41467-024-51374-x

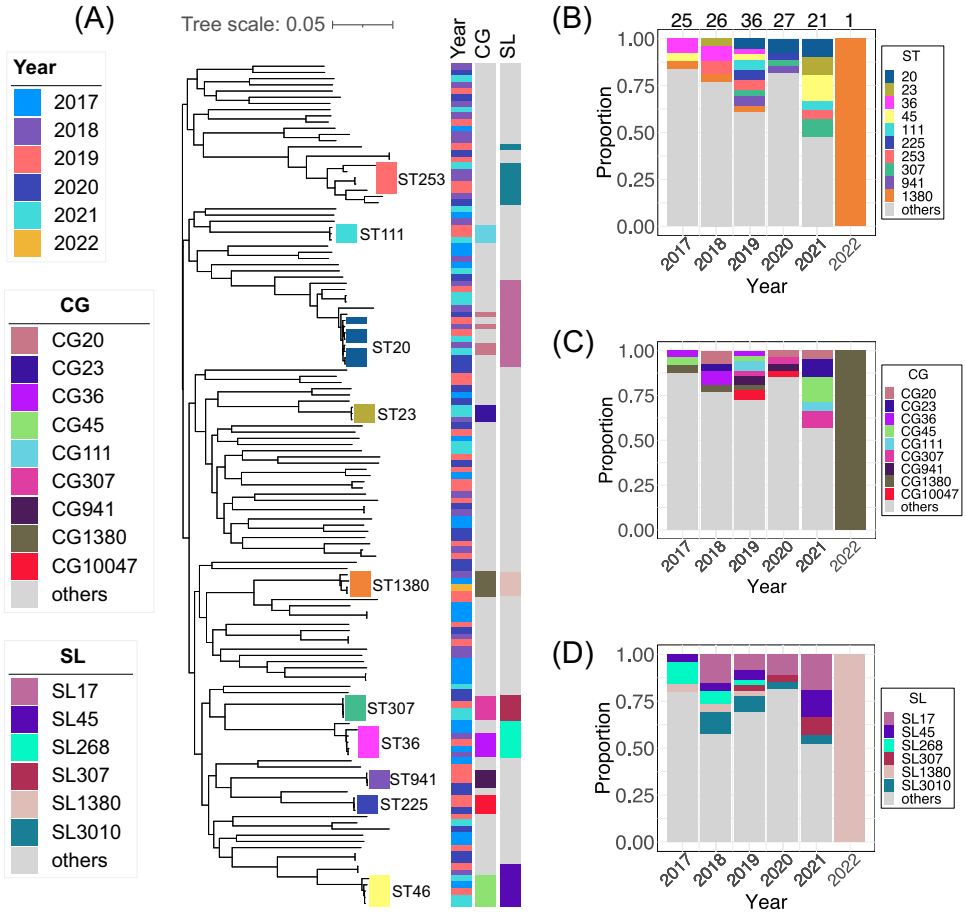

Fig. 1 | Genomic features of the 136 *K. pneumoniae* isolates from bloodstream infection. A A maximum likelihood phylogeny showing the year of isolation and the most frequent sequence types (ST), clonal groups (CG; represented by three or more genomes), and sublineages (SL; represented by four or more genomes). The midpoint-rooted tree is calculated using single nucleotide polymorphic sites from the sequence alignment of 3,511 concatenated core genes. Tree scale represents the number of substitutions per site. The barplots show the proportion of (**B**) ST (**C**) CG and (**D**) SL by sampling year. The numbers above the bars in panel B indicate the number of isolates per year. For visual clarity, only the most frequent ST, CG and SL are shown in colored blocks, while less frequent ones are grouped together as "others" in gray blocks.

Using in silico analysis of the genome sequences, we identified the presence of genes that encode AMR determinants (Fig. 2B and Supplementary Fig. S1, Data 3). Across the entire dataset, we detected a total of 64 unique genes encoding resistance to ten antimicrobial drug classes. All 136 genomes harbored at least one AMR gene that confer resistance to beta-lactams. A total of 134 (98.52%) and 123 (90.44% genomes) harbored intrinsic and chromosomally mediated genes encoding resistance to fosfomycin and phenicol/quinolone, respectively. We observed ten combinations of the phenicol/quinolone multidrug efflux pump alleles occurring in the genomes (Supplementary Data 3). The most frequently detected combinations were the *oqxA/B* ($n = 42$, 30.88%), *oqxA/B19* ($n = 27$, 19.85%) and *oqxA/B25* ($n = 15$, 11.02%). Other AMR determinants were also detected but at low frequencies (present in ≤ 21 genomes), including genes that are associated with resistance to aminoglycoside, macrolide, phenicol, quinolone, sulfonamide, tetracycline, and trimethoprim.

Among the beta-lactamases, we identified a total of 24 chromosomally encoded bla_{SHV} gene variants (Fig. 2B and Supplementary Data 3). The most frequently occurring variants were $bla_{SHV\text{-}11}$ ($n = 37$ genomes, representing 27.2% of the population), $bla_{SHV\text{-}1}$ ($n = 28$ genomes, 20.58%), and $bla_{SHV\text{-}27}$ ($n = 11$ genomes, 8.08%). The gene bla_{SHV} has been reported to have undergone robust allelic diversification in clinical *K. pneumoniae* and other *Enterobacteriaceae*, and our results were consistent with this diversity[33]. While the presence of the wild type bla_{SHV} is responsible for resistance to ampicillin, amino acid substitutions from allelic diversification may cause them to have expanded functionality such as ESBL activity and/or beta-lactamase inhibitor resistance activity[33]. The variants $bla_{SHV\text{-}38}$, $bla_{SHV\text{-}164}$, $bla_{SHV\text{-}187}$ that we detected in our dataset are known to be associated with resistance to cephalosporins[34]. However, we did not observe resistance to cephalosporins when tested in vitro in isolates harboring these genes. Furthermore, isolate KPB57 which carried the gene $bla_{SHV\text{-}41}$ exhibited resistance to first (cefazolin) and second (cefoxitin) generation cephalosporins but not to third generation cephalosporins. Detection of this *bla* variant has been associated with conflicting ESBL phenotypes[35,36].

High concordance of AMR phenotype and genotype

We sought to investigate the level of concordance between the results of the in vitro antimicrobial susceptibility testing and the in silico screening of genetic elements conferring resistance to different antimicrobial agents. We defined the following terms: (1) True positives were isolates with resistant phenotypes harboring a corresponding resistance genetic determinant; (2) True negatives were isolates with susceptible phenotypes and do not carry the corresponding AMR determinant in their genome; (3) False negatives were isolates exhibiting a resistant phenotype but with no corresponding AMR genetic element detected in the genome; and (4) False positives were isolates that exhibit a susceptible phenotype but carry the AMR genetic element in their genome. Overall, we found high concordance between

https://doi.org/10.1038/s41467-024-51374-x

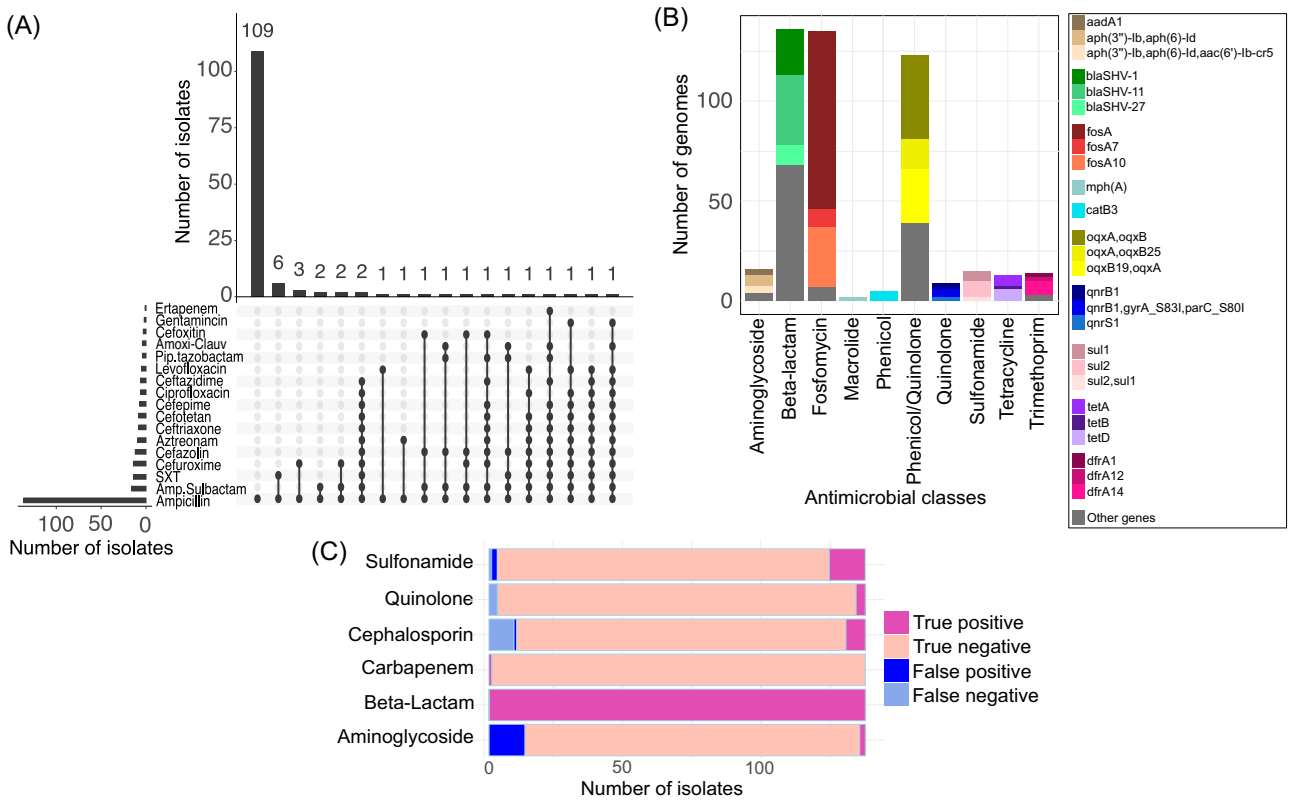

Fig. 2 | Antimicrobial susceptibility phenotypes and genotypes of the 136 *K. pneumoniae* isolates from bloodstream infection. A An UpSet plot showing the total number of isolates resistant to antimicrobials tested (left bar plot), the total number of isolates exhibiting a particular antibiogram (top bar plot) and filled dots representing the presence of an antimicrobial resistance (AMR) phenotype. Acronyms: Pip.tazobactam Piperacillin-Tazobactam, SXT Sulphamethoxazole/Trimethoprim, Amp. Sulbactam Ampicillin Sulbactam. **B** Number of genomes carrying individual AMR genes and gene combinations per antimicrobial class. The AMR genes in the color legend are grouped according to antimicrobial class. **C** Concordance analysis of resistance phenotypes and predicted genotypes.

the resistant phenotypes and the presence of the corresponding AMR genes. For the six antimicrobial classes, concordance values range from 90.44% (aminoglycosides) to 100% (carbapenems) (Fig. 2C and Supplementary Data 4).

In the phenotype-genotype concordance analysis of carbapenem resistance, the single carbapenem resistant isolate KPB97 (ESBL-producing) lacked a detectable carbapenemase gene. Examination of the gene encoding the outer membrane porin revealed a truncation in the *ompK36* gene. We observed *ompK36* in the ESBL-producer KPB97, whereas other *ompK35* and *ompK36* truncations were detected in genomes not containing ESBL (Supplementary Data 3). Previous reports have shown that truncation of *ompK36* in the presence of an ESBL is sufficient for non-susceptibility to ertapenem but not to imipenem and meropenem[37,38]. Hence, we considered this isolate as true positive. We therefore assigned perfect concordance for carbapenem resistance, with specificity of 100% (range: 97.3–100%) and sensitivity of 100% (range: 2.50–100%).

Concordance between quinolone resistance phenotype and genotype showed 97% agreement, with specificity of 100% (range: 97.18–100%) and sensitivity of 57.1% (range: 18.4–90.1%). In our study, we associated non-susceptibility to ciprofloxacin and levofloxacin with the presence of either plasmid mediated *qnrB1* or the mutations in the quinolone resistance determining regions (*gyrA*_S83I + *parC*_S80I)[39]. A total of four and 129 genomes were true positives and true negatives, respectively. We did not detect mutations in the quinolone resistance determining regions for genomes phenotypically resistant to ciprofloxacin (KPB44 and KPB68) but harbored a *qnrB1* gene. We observed no detectable mechanism of resistance in the levofloxacin resistant isolate (KPB87).

Phenotype-genotype concordance of sulfonamide resistance was 97.79%, with specificity of 98.36% (range: 94.2–99.8%%) and sensitivity of 92.85% (range: 66.1–99.8%). Cephalosporin phenotype-genotype agreement was 92.64%, with specificity of 99.16% (range: 95.44%–99.97%) and sensitivity of 43.75% (range: 19.75–70.12%). Notably, all ceftriaxone resistant isolates except isolate KPB70 (resistant to all cephalosporins) harbored the ESBL gene $bla_{\text{CTX-M-15}}$. For aminoglycosides, concordance was 90.44%, with specificity of 90.29% (range: 83.98–94.73%) and sensitivity of 100% (range: 15.81–100%).

Multidrug resistant and hypervirulent clones are present in bloodstream infections

Multidrug resistant clones are defined as those encoding acquired resistance determinants to at least three antimicrobial drug classes at high frequencies (≥56%), whereas hypervirulent clones harbored the plasmid-associated virulence genes *iuc*, *iro* and/or *rmpA*/*rmpA2* at frequencies between 31–100%[40]. We used these in silico definitions to determine whether any of our isolates belong to known multidrug resistant or hypervirulent clones. A total of 17 genomes (12.5%) in our dataset matched known multidrug resistant clones (Fig. 3). The multidrug resistant clones included CG14, CG20, CG147, CG711, CG1123, CG10094, CG10156, CG10253, CG10476 (each with one genome), CG10124 and CG10529 with two genomes each, and CG307 with four genomes. When we mapped the multidrug resistant clonal groups against the AMR gene presence and absence results determined using Kleborate[41], we found that only members of CG307 harbored resistance genes to multiple antimicrobial classes, including the ESBL gene $bla_{\text{CTX-M-15}}$ (Supplementary Data 4). However, clones that were not designated as multidrug resistant (CG10524, CG429, CG2623, CG540,

https://doi.org/10.1038/s41467-024-51374-x

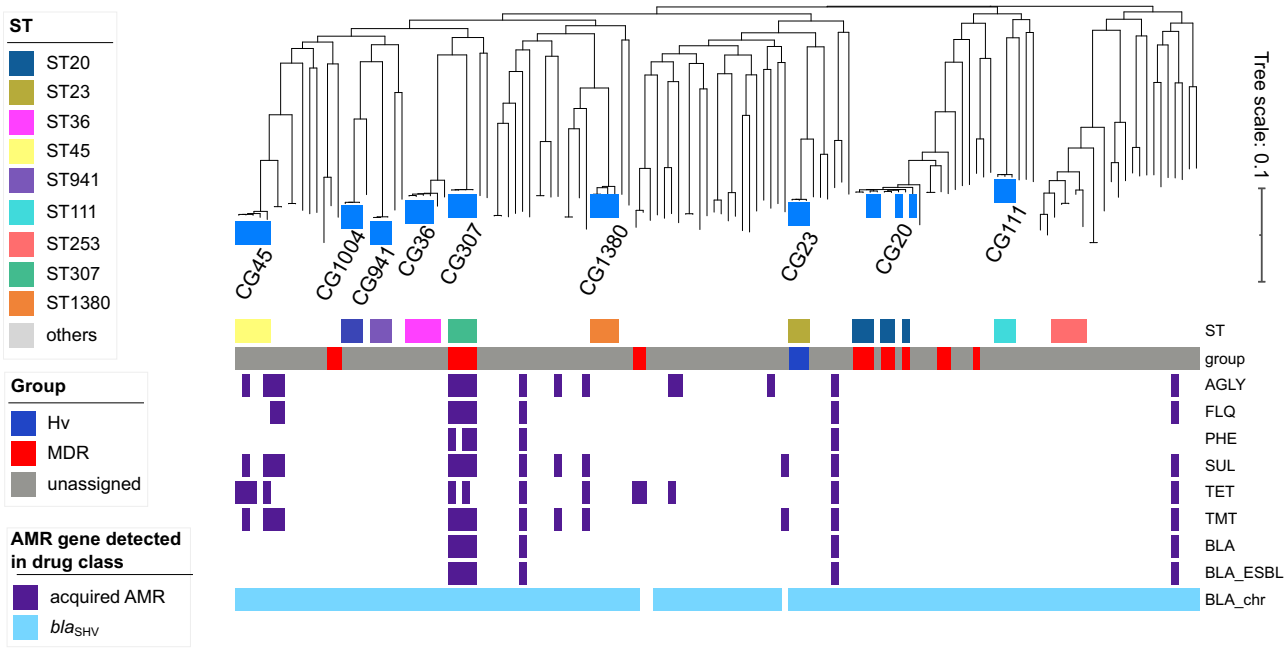

Fig. 3 | Presence of multidrug resistant and hypervirulent clones. The maximum likelihood phylogeny was built from the single nucleotide polymorphic sites of 3,511 concatenated core genes. The tree is identical to that in Fig. 1A. Only the major CG and ST are shown for visual clarity. The Group category refers to multidrug resistant (MDR; red blocks) and hypervirulent (Hv; blue blocks) *K. pneumoniae* clones defined in references[40,47] and implemented on Kleborate[41]. The matrix shows the presence (purple blocks) or absence (white) of at least one gene conferring resistance to each antimicrobial class. AGLY Acronyms, AGLY aminoglycoside, FLQ fluoroquinolones, PHE Phenicol, SUL sulfonamides, TET tetracycline, TMT trimethoprim, BLA beta-lactamase. The presence of chromosomally encoded bla_{SHV} is indicated by the light blue blocks.

CG10151, CG12251, CG45, ST6366, and ST3293) as previously reported[40] also harbored resistance genes to multiple drug classes in our study (Fig. 3, Supplementary Data 2 and Data 4). Among the unassigned clones, we detected the presence of $bla_{CTX\text{-}M\text{-}15}$ in CG10524 (ST1564), CG429 (ST429), and ST3293.

Experimental evidence from previous studies has identified the presence of key markers for hypervirulence that includes the siderophores yersiniabactin, aerobactin and salmochelin as well as hypermucoidy via capsule overproduction[28,40]. In our study, we identified only one known hypervirulent clone based on the previous classification[40], and this clone included three genomes belonging to CG23/ST23. We did not find convergent clones in our dataset, i.e., those that are both hypervirulent and multi-drug resistant[42].

Diversity and transmission of plasmid-encoded ESBL gene $bla_{CTX\text{-}M\text{-}15}$

Plasmids encoding the ESBL gene $bla_{CTX\text{-}M\text{-}15}$ often harbor resistance determinants to additional drug classes[26]. We sought to understand the epidemiology of plasmids carrying the $bla_{CTX\text{-}M\text{-}15}$ by sequencing plasmid genomes. We identified $bla_{CTX\text{-}M\text{-}15}$ in seven (5.14%) bloodstream isolates. Because multidrug resistant isolates from non-blood samples (e.g., urine) are also routinely archived by DHMC as part of patient care, we also identified another isolate (KPB102) from a urine sample that carried $bla_{CTX\text{-}M\text{-}15}$. From the long-read sequencing data of the eight isolates, we obtained 12 circular plasmid genomes (Fig. 4A and Supplementary Data 5). Of these, four isolates carried two plasmids (KBP68, KPB77, KBP97, KBP176), while the other four isolates carried only one plasmid (KBP44, KBP140, KBP166, KBP102). The eight isolates came from four STs representing four CGs, of which five isolates are members of ST307 (CG307). The gene $bla_{CTX\text{-}M\text{-}15}$ was present in either of the two replicon types IncFIB(K)/IncFII(K) ($n = 6$ plasmid genomes) or IncFIB(K) ($n = 2$ plasmid genomes). The sizes of these plasmids ranged from 131,110 bp (pKPB68_2) to 246,740 bp (pKPB176_2). Four plasmid genomes with an unknown replicon type were also detected and classified as untyped. The sizes of the untyped plasmids ranged from 3559 bp (pKPB77) to 367,083 bp (pKPB97).

Two groups of plasmid sequences sampled from different patients exhibited high sequence identity (Fig. 4B and Supplementary Fig. S2A, B). The first group consisted of plasmids pKPB102_2 and pKPB166_2, which are both of the IncFIB(K)/IncFII(K) replicon type. The two plasmid sequences are of the same size (231,545 bp) and have 100% sequence identity and sequence coverage, as well as a low pairwise mash distance of 7.14e-6 (Fig. 4B). Each plasmid was retrieved from two different isolates (KPB102 and KPB166) that both belonged to ST307, and which differed by 13 SNPs in their core genome sequence alignment. One isolate was collected from a urine sample in October 2019 and the second isolate from a blood sample in September 2021 from different patients. The high identity between the two plasmid sequences and their common strain background exemplifies a transmission pattern within the same clonal lineage (ST307).

The second group of highly similar plasmid sequences consisted of three IncFIB(K)/IncFII(K) plasmids (pKPB77_2, pKPB97_3, pKPB176_2) (Fig. 4B). The three isolates that carried these plasmids were sampled from three different patients. Plasmids pKPB77_2 (245434 bp) and pKPB97_3 (243759 bp) shared 100% sequence identity, 99% sequence coverage, and a mash distance of 3.1e-5. These two plasmids were retrieved from two distinct lineages—isolate KPB77 is a member of ST307/CG307 and KPB97 is a member of ST429/CG429. Both isolates were from blood samples collected in March 2019 and August 2019, respectively. The two isolates differed by 19,080 SNPs in their core genome alignment. ST307 and ST427 are 4-locus variants in the MLST scheme (*gapA*, *pgi*, *phoE*, *tonB*). A third isolate in this group is KPB176 (pKPB176_2), which was collected from a blood sample in November 2021. This plasmid (pKPB176_2) exhibited high identity with pKPB77_2 (99% sequence identity, 99.99% sequence coverage, mash distance of 3.1e-4). The isolates from which plasmids pKPB77_2 and pKPB176_2 were derived from both belonged to ST307 and differed by

https://doi.org/10.1038/s41467-024-51374-x

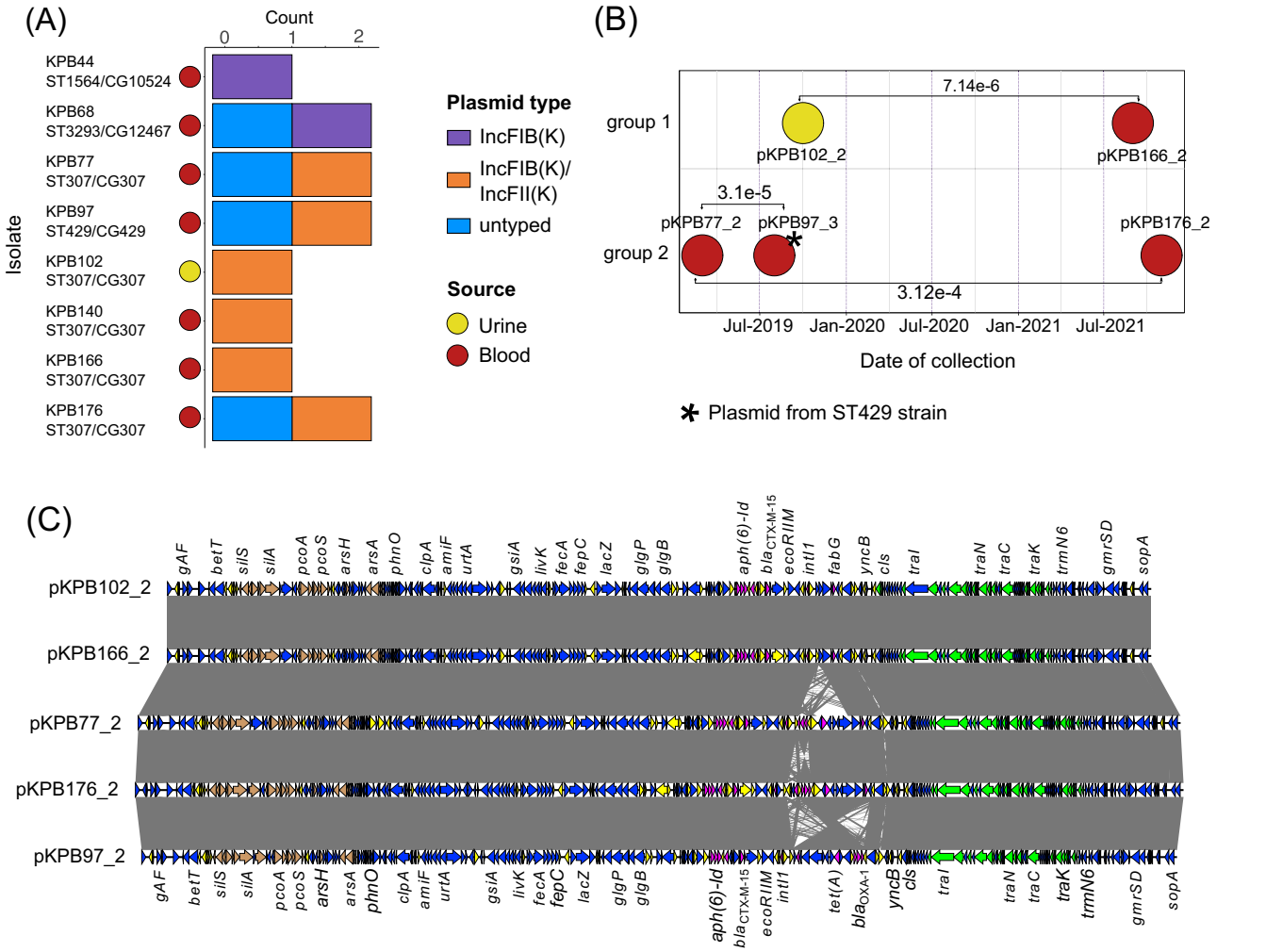

Fig. 4 | Plasmid types and transmission of plasmid carrying *bla*$_{CTX\text{-}M\text{-}15}$. **A** Bar plot showing the number of plasmid genomes and replicon types detected in the eight isolates harboring the gene *bla*$_{CTX\text{-}M\text{-}15}$. **B** Timeline of transmission of plasmids encoding *bla*$_{CTX\text{-}M\text{-}15}$ between isolates. Horizontal lines connecting circles show the mash distances between plasmid genomes in each linked group. In group 1, both isolates are from ST307/CG307. In group 2, the black asterisk (*) on pKBP97_3 indicates that the plasmid came from an ST429/CG429, while the other two isolates are members of ST307/CG307. For both panels A and B, circles are colored by clinical source (red for blood, yellow for urine). **C** Structural comparison of plasmid genomes encoding *bla*$_{CTX\text{-}M\text{-}15}$ and belonging to group 1 and group 2 in panel 2B. Gray areas between plasmids represent regions with 80–100% sequence identity. Genes are represented by arrows and are color-coded according to general function: magenta—antimicrobial resistance, brown—heavy metal resistance, green—conjugation transfer, yellow—mobile elements such as insertion sequences, transposons and integrases.

63 SNPs in their core genome sequence alignment. Such variation in their core SNPs appear to have accumulated over 2 years between March 2019 and November 2021. These results show that this plasmid type was mobilized through both horizontal transmission between phylogenetically distinct lineages (in the case of pKPB77_2 and pKPB97_3) as well as transmission within the same clonal lineage (in the case of pKPB77_2 and pKPB176_2 in ST307). The notable difference between the two groups of plasmids (i.e., group 1 and group 2 in Fig. 4B) is that group 2 contains a ~13 kb cluster of genes that includes IS6 family transposases, Tn5403 family transposase, and AMR genes *catB3*, *bla*$_{OXA\text{-}1}$, *aac(6')-Ib-cr5*, and *tetA* (Fig. 4C).

Lastly, we observed the presence of multiple resistance genes carried on different plasmids within a single genome. Multidrug resistance in isolate KPB68 (ST3293/CG12467 sampled in 2018) was mediated by resistance determinants present in two plasmids. The plasmid pKPB68_2 has a genome size of 131,110 bp, replicon type IncFIB(K), and carries *bla*$_{CTX\text{-}M\text{-}15}$, *aph(3'')-Ib*, *aph(6)-Id*, and *sul2*, *bla*$_{TEM\text{-}1}$. The plasmid pKPB68_3 has a genome size of 32,881 bp with an untyped replicon and carries *aac(6')-Ib-cr*, *qnrB1*, *catB3*, *bla*$_{OXA\text{-}1}$, *tetA*, and *dfrA14* (Supplementary Fig. S2A, B).

Antigenic diversity and hypervirulence markers

Using in silico screening of the genomes, we inspected our isolates for the presence and diversity of the surface polysaccharide capsule (KL) and lipopolysaccharide (O) that determine the antigenic serotypes. These two structures activate the human immune system during infection and are also widely used in strain typing[43,44]. We identified 53 known capsular KL locus types and four genomes with unknown KL types (Simpson's diversity index = 0.971) (Fig. 5A and Supplementary Fig. S3A, Supplementary Data 3). The most frequent KL types in our dataset were KL28 (present in nine genomes), KL24 (eight genomes), and KL102 (seven genomes). The KL1 type which is a genetically homogenous capsular type commonly associated with hypervirulent *K. pneumoniae* clones[45] was detected in three genomes belonging to ST23. We also identified 11 different O types in our dataset (Simpson's diversity index = 0.731). Three O types (O1/O2v2 = 54 genomes, O1/O2v1 = 37 genomes, O3b = 25 genomes) accounted for >85% of our dataset (Supplementary Fig. S3B). Two unknown O types were detected in ST2004 (KPB152) and ST6366 (KPB79) genomes.

We searched our short-read sequence dataset for the presence of *Klebsiella*-specific virulence factors. First, we screened the genomes for

https://doi.org/10.1038/s41467-024-51374-x

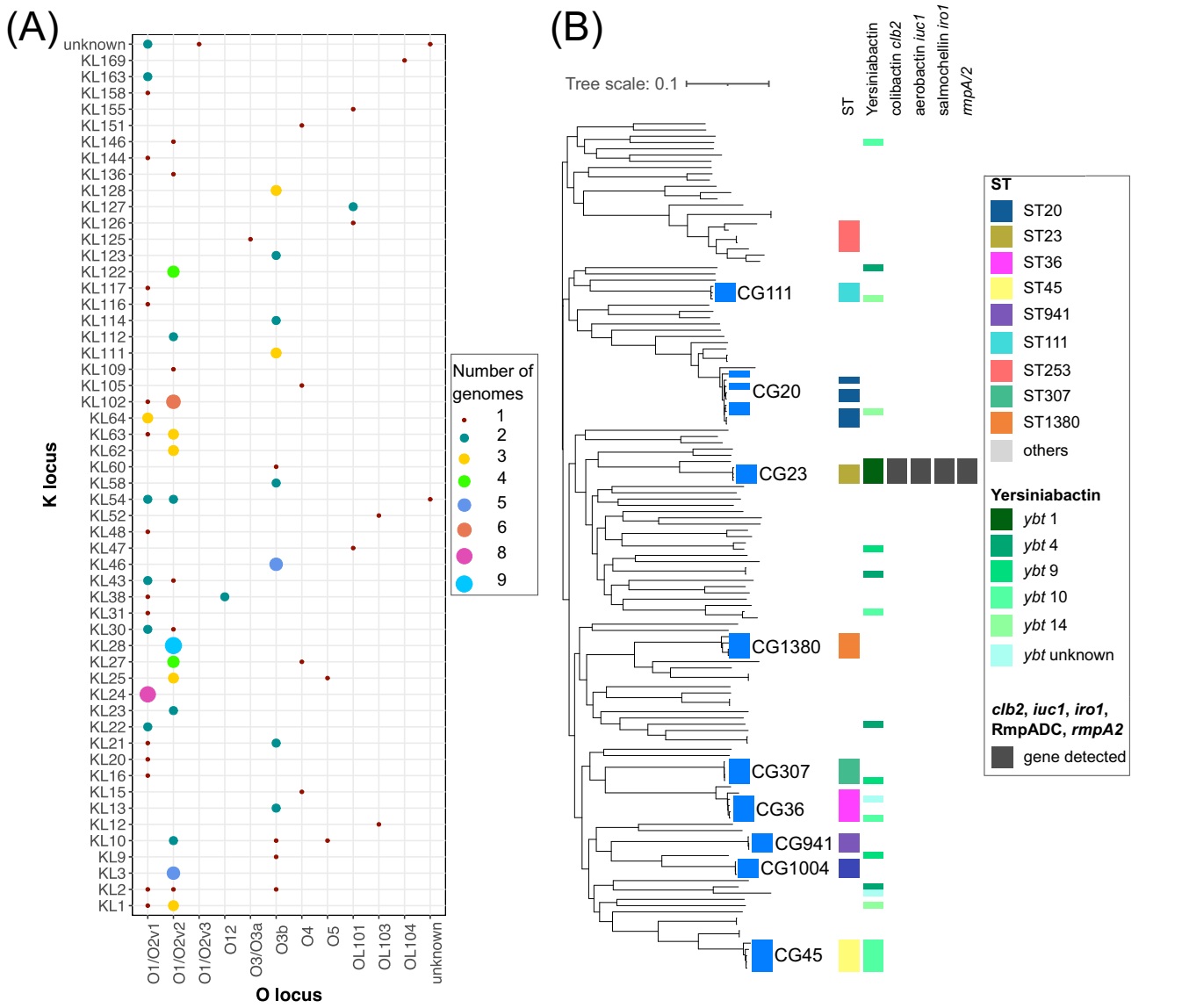

Fig. 5 | Diversity and distribution of surface polysaccharides and hypervirulence markers. A Different combinations of the surface polysaccharides capsule (KL) and lipopolysaccharide (O) loci are shown. The colors and sizes of the bubbles correspond to the number of genomes that carry unique KL and O combinations. **B** The maximum likelihood core genome phylogeny showing the phylogenetic distribution of genes encoding yersiniabactin (*ybt*), colibactin (*clb2*), aerobactin (*iuc1*), salmochelin (*iro1*), and *rmpA/2*. The tree is identical to that in Fig. 1A. Only the major CGs and STs are shown for visual clarity.

yersiniabactin, a siderophore system for sequestering iron that enhances bacteria survival and replication within the host[46]. This genetic marker is encoded by the *ybt* gene and is transferred by mobile genetic elements, in particular the integrative conjugative element ICE*kp*[46]. In our study, we detected five known *ybt* variants (*ybt* 1, 4, 9, 10 and 14) in 24 genomes representing a total of 17.64% of the population (Fig. 5B). The five *ybt* variants were widely distributed across the core genome phylogeny. The most common *ybt* variants were *ybt* 10, *ybt* 1, *ybt* 4, *ybt* 14, and *ybt* 9. Variant *ybt* 10 was present in five genomes from ST45/CG45 and one genome each belonging to ST36/CG36, ST14/CG10156 and ST1564/CG10524. *ybt* 1 was present in three genomes from ST23/CG23 and one genome from ST260/CG10671. *ybt* 4 was present in one genome each from ST37/CG10094, ST2217/CG2217, ST2.248/CG10668 and ST6360/CG12462. *ybt* 14 was present in one genome each from ST20/CG20, ST111/CG111, ST942/CG942. *ybt* 9 was present in one genome each from ST35/CG35, ST307/CG307 and ST393/CG12458. Two genomes possessed unknown *ybt* variants (KPB59 from ST36 and KPB141 from ST1554).

Using the short- and long-read hybrid assemblies, we screened for the presence of other key virulence determinants such as *clb2* (colibactin), *iuc1* (aerobactin), *iro1* (salmochelin), and *rmpA/A2* (activator for capsule biosynthesis) in plasmid genomes. These genes have been previously identified as genetic markers of hypervirulence in *K. pneumoniae*[47]. Yersiniabactin, aerobactin, and salmochelin are siderophores that promote bacterial survival in nutrient-poor environments by chelating iron[48] and is therefore particularly important in the growth, replication, and metabolism of bacteria in the blood. The RmpA/2 activator of capsule biosynthesis leads to hypermucoidy[49]. We detected the presence of these four virulence markers in only four genomes (Fig. 5B and Supplementary Data 3). Three of these four genomes are known hypervirulent clones (ST23/CG23), whereas the fourth genome belongs to clonal group ST260/CG10671, which is a double-locus variant of ST23 (in genes *pgi* and *phoE*). These four genomes clustered together in the core genome phylogeny and they all possessed the *ybt* 1 variant, KL1 capsular type, and two O types (O1/O2v1 in KPB51 and O1/O2v2 in the other three).

Analysis of long read sequencing data on these four isolates showed that the hypervirulence genes are borne on plasmids ranging in size from 178,418 bp (pKPB165_2 in isolate KPB165) to 228,574 bp (pKPB147_2 in isolate KPB147) (Supplementary Data 6). The sequences of the hypervirulence-carrying plasmids were characterized by rearrangement, loss, and/or truncation of genomic regions (Supplementary Fig. S4). For instance, a ~50 kb region present in all other hypervirulence plasmids was missing in pKPB165. This region consists of several genes including those encoding tellurium resistance (*terABCDEXZ*) and the transposase Tn3. Although the hypervirulence plasmids were frequently flanked by transposases and insertion sequences, they lack a conjugative apparatus, except for the presence of the gene encoding the TraI protein that functions in site and strand nicking[50,51]. We did not identify an origin of transfer (*oriT*) site in any of the four hypervirulence plasmids, suggesting that they are not themselves able to initiate their transfer via conjugation.

Discussion

The increasing burden caused by AMR in bloodstream infections has life-threatening consequences because it can considerably increase the rates of treatment failure and death[8,9]. Hence, understanding the mechanisms of resistance dissemination remains a key strategy for effective management and control of AMR in invasive *K. pneumoniae*. Here, we analyzed 136 short-read genome sequences of *K. pneumoniae* complemented with 12 long-read sequences to understand the drivers that facilitate the dissemination of resistant *K. pneumoniae* strains in bloodstream infections.

Our study highlights the remarkable clonal and genomic diversity of *K. pneumoniae* in bloodstream infections. Our results are consistent with those reported in other hospitals in the United States[11,52,53], Asia[42], and Europe[54]. Our findings greatly expand the number of bloodstream *K. pneumoniae* clones, indicating that the capacity to cause bloodstream infections is not restricted to one or few genetic lineages. Such diversity means that the underlying causes of resistance, other clinically relevant traits, and disease outcomes may vary considerably between *K. pneumoniae* from different patients and between settings. Among the clones detected in our study, we report the presence of known multidrug resistant high-risk pandemic clones such as ST14, ST17, ST45 ST147 and ST307[10,55,56]. Infections caused by ESBL-producing *K. pneumoniae* present a high burden associated with increased mortality, length of hospital stay, and medical costs[57,58]. This ESBL subset of *K. pneumoniae* continuously threatens the efficacies of beta-lactams antimicrobials, which are the preferred treatment options for bloodstream infections due to their broad activity spectrum and selective toxicity[59].

Lineages ST258 and ST307 are reported to be endemic in the United States and are associated with resistance to third-generation cephalosporins and carbapenems[11,52,53]. Interestingly, we did not detect ST258 in our dataset, whereas ST307 was present only at a low frequency (2.94%). All ST307 isolates in our study were multidrug resistant, as well as resistant to third-generation cephalosporins, and were involved in plasmid transmission of the ESBL gene $bla_{CTX\text{-}M\text{-}15}$. ST307 likely emerged in Europe around 1994 and rapidly spread across six continents[60], causing many outbreaks in nosocomial settings and long-term care centers[55]. The first published report of ST307 in the United States was in Texas in 2013[61]. However, a decline in the recovery of these two clones have been recorded in some parts of the United States[62,63]. A search of the *K. pneumoniae*-dedicated databases in BIGSdb[64] and Pathogenwatch[65] (as of May 2024) revealed a decline in the frequency of ST258 and ST307 in the United States after 2017, with no record of ST258 since 2022. The underlying causes for this decline are unclear. Different *K. pneumoniae* clones are known to experience repeated waves of clonal expansion and replacement in other parts of the world[66–68], and this may partly explain our findings at DHMC and the apparent decline in the frequency of STs 258 and 307 in the country. Hence, continued surveillance of *K. pneumoniae*, including antimicrobial susceptible and non-ESBL lineages that may become recipients of AMR-carrying plasmids, are critical to uncover signals of emerging STs that have the potential to replace currently known high-risk clones. Early detection and infection control efforts of new threats will help prevent their further spread.

Concordance analysis of AMR phenotype-genotype is critical for evaluating accuracies of antimicrobial susceptibility tests and correlating phenotypes to predicted genotypes. Such analyses are critical for validating phenotypic tests, uncovering novel resistance determinants, and enhancing epidemiological investigations[69,70]. While concordance for tested antimicrobials was >90%, discrepancies stemming from the inability to match resistant and susceptible phenotypes with genotypes decreased concordance sensitivity and specificities, respectively. For instance, the discrepancies we observed in cephalosporin phenotype-genotype concordance were mainly associated with the inability to sort out bla_{SHV} phenotypes. Such difficulty stems from the highly diversified SHV gene[36,71], with 182 bla_{SHV} alleles in the NCBI reference Gene Catalog (as of December 2023). Assignments of SHV allele functionality have relied mainly on varying supporting evidence, such as the assignment of ESBL phenotype without eliminating the presence of other ESBL enzymes or defining specific roles of ESBL activity in the SHV allele[33]. Future work should aim to precisely elucidate the range of SHV variation, enzyme activity and functions in genetically and ecologically diverse *K. pneumoniae*.

Mobile genetic elements such as plasmids are important vehicles for the spread of AMR and have been implicated in disease outbreaks and transmission[72,73]. Plasmids harboring ESBLs are significant contributors to the success of international and epidemic *K. pneumoniae* clones such as ST307 and ST258[10]. Our results show that plasmids belonging to the IncF family mediate the clonal (within the same lineage) and horizontal (between lineages) transmission of the ESBL gene $bla_{CTX\text{-}M\text{-}15}$. Plasmid transfer between clones and distantly related genomes has been previously reported. For example, clonal transmission of an IncHI1B/IncFIB plasmid between two *K. pneumoniae* ST11 and transmitted horizontally to *Escherichia coli* strains (ST10 and ST58) has been documented[74]. In our study, we also show two notable plasmid acquisitions that can further facilitate the stable maintenance of ESBLs in nosocomial settings. Plasmid transfer can overcome ecological barriers (e.g., urine to blood); hence, future surveillance efforts will benefit from the inclusion of non-blood isolates to improve current knowledge about the ecological range of ESBL-carrying plasmids. Acquisition of more than one plasmid within a single genome can rapidly lead to the emergence of multidrug resistance and convergent pathotypes[29]. Continuous surveillance of IncF plasmids and the topology of plasmid sharing networks across phylogenetic and spatiotemporal landscapes is therefore critical.

Hypervirulent *K. pneumoniae* are a typical cause of community-acquired infections, including liver abscesses and bacteremia[75,76]. In our study, we detected only four genomes harboring the hypervirulence genetic determinants. Three of these genomes belong to the well-established hypervirulent *K. pneumoniae* clone ST23 (CG23) commonly associated with the KL1 capsule serotype[40,76]. Intriguingly, members of this ST have evolved from drug susceptible hypervirulent phenotypes that acquired multidrug resistance plasmids[77] and carbapenemases[78]. This troubling convergence of multidrug resistance and hypervirulence creates “superbugs” and has dire consequences in controlling infections caused by this clone. The fourth hypervirulent clone in our study was an ST260 (strain KPB147 with KL1 capsule), a double locus variant of ST23. This clone may have arisen from recombination between an ST23 and another lineage of *K. pneumoniae*[76]. Our results show that pKPB147 was unlike the ST23 virulence plasmids which appear to have experienced significant genomic truncation and gene segment rearrangement. The unique characteristics of multidrug resistant and hypervirulent pathotypes

should be an important consideration in the development of clinical interventions and treatment in bloodstream infections.

Although we have performed comprehensive profiling of our bloodstream isolates, we acknowledge important limitations in our study. First, our definition of transmission groups in this study is associated with little data on epidemiological linkages other than the source and time of isolation of *K. pneumoniae* isolates. For example, we did not have information about whether patients were part of the same household, social groups or hospital wards, all of which may facilitate transmission. Second, we carried out long-read sequencing only on 11 out of the 136 blood isolates and one urine isolate. Hence, other plasmid types, plasmids not associated with ESBL carriage, and plasmids carrying other virulence determinants were left unexplored. We acknowledge that it would be ideal to sequence more plasmids, e.g., from isolates sampled in other human body sites or the hospital environment, to a gain a wider picture of transmission routes within the medical center. However, resource limitation precludes us from carrying this out. Future investigations on the ecological niches and distribution of plasmids are therefore critical in understanding their roles in invasive *K. pneumoniae* infections. Third, our relatively small isolate collection represents a short time frame and may not effectively capture long-term clonal persistence and transmission that may have occurred within the medical center. Lastly, we did not investigate other members of *Enterobacteriaceae*, which are known to play a role in interspecies transmission, horizontal gene transfer, and environmental persistence[79]. Nonetheless, our findings provide an important baseline census of the standing lineage and AMR diversity for continued surveillance of bloodstream infections. This work also presents a strong impetus to consider plasmid sharing in epidemiological surveillance within the medical center. The limitations presented here will stimulate future explorations that will define the basis for the adaptive success of *K. pneumoniae* occupying the bloodstream niche.

Our study shows that the *K. pneumoniae* in bloodstream infections can vary substantially in terms of the clonal lineages, phenotypic resistance patterns, and carriage of genes that mediate multidrug resistance and hypervirulence. The temporal persistence of resistant strains and local dissemination of ESBL genes lay in part on the rapid transmission of the IncF family of plasmids within the same bacterial lineage and between different lineages. This work will be useful in current efforts in designing effective strategies and interventions to control the spread of high-risk bacterial clones and in reducing the opportunity for pathogen persistence and onward transmission. Continued surveillance and further genomic epidemiological studies in healthcare settings are critical to determine the consequent risk to other vulnerable patients and to the wider community.

Methods

Bacterial isolation, identification, and antimicrobial susceptibility testing

Bacterial isolates from bloodstream infections in unique pediatric and adult patients were grown from clinical blood cultures submitted to the Department of Pathology and Laboratory Medicine at DHMC, New Hampshire, USA from January 2017 to January 2022. Multidrug resistant isolates from non-blood samples (e.g., urine) are also routinely archived as part of patient care. Initial species identification was carried out using either the FilmArray Blood Culture Identification (BCID) panel for isolates sampled from January 2017 to May 2021 or the FilmArray Blood Culture Identification 2 (BCID2) panel for isolates sampled from June 2021 onwards. The BCID/BCID2 panel (BioMerieux; Durham, North Carolina, USA) is a multiplexed PCR assay for rapid identification of causative pathogens from positive blood cultures[80]. Using this assay, 136 isolates were identified in the clinical laboratory as *K. pneumoniae*.

Ethical approval was granted by the Committee for the Protection of Human Subjects of DHMC and Dartmouth College. The study protocol was deemed not to be a human subjects research. Samples used in the study were subcultured bacterial isolates that had been archived in the routine course of clinical laboratory operations. No patient specimens were used, and patient protected health information was not collected.

Antimicrobial susceptibility testing (AST) was performed on the MicroScan Walkaway 96 Plus automated instrument (Beckman-Couter; La Brea, California, USA) using two FDA (Food and Drug Administration)-cleared AST panels: the NUC62 Panel from January 2017 to May 2019, then switching to the Neg MIC Panel 46 from June 2019 onwards. A verification study of each panel had been performed prior to use for testing of patient isolates. The breakpoints applied on the NUC63 Panel were those of the manufacturer at the time of FDA clearance. When transitioning to the Neg MIC Panel 46, off-label breakpoints for the carbapenems and select cephalosporins had been validated to align with the Clinical Laboratory Standards Institute M100 S28 guidelines[81]. The 20 antimicrobial agents tested represented six antimicrobial classes and sub-classes: aminoglycosides (amikacin, gentamicin); antifolate (sulfamethoxazole/trimethoprim); carbapenems (ertapenem, meropenem), cephalosporins (cefoxitin, cefazolin, cefepime, cefotaxime, cefotetan, ceftazidime, ceftriaxone, cefuroxime); fluoroquinolones (ciprofloxacin, levofloxacin); monobactam (aztreonam); penicillins (ampicillin, ampicillin-sulbactam, amoxicillin-clavulanic acid, piperacillin/tazobactam). Results of the antimicrobial susceptibility testing are presented in Supplementary Data 1. All isolates were stored in DMSO solution at −80 °C.

DNA extraction, library preparation, and whole genome sequencing

Isolates were subcultured from DMSO stocks onto commercially prepared tryptic soy agar with 10% sheep red blood cells (Remel; Lenexa, Kansas, USA) and in brain heart infusion broth (BD Difco; Franklin Lakes, New Jersey, USA) at 37 °C for 24 h. For short-read sequencing, DNA was extracted and purified from liquid cultures using the QuickDNA Fungal/Bacterial Miniprep Kit (Zymo Research; Irvine, California, USA) following manufacturer's protocol. DNA libraries of each sample was prepared using the Illumina DNA Prep Kit and IDT 10 bp UDI indices in accordance with the manufacturer's instructions. DNA samples were sequenced as multiplexed libraries on the Illumina NextSeq 2000 platform operated per the manufacturer's instructions. Illumina bcl-convert (v.3.9.3) was used for demultiplexing, quality control, and adapter trimming of reads. Sequencing resulted in 151 nt long paired end reads. For long-read sequencing, high molecular weight DNA isolates was extracted using the Quick-DNA HMW MagBead Kit (Zymo Research; Irvine, California, USA) following manufacturer's instructions. DNA samples were prepared using the Oxford Nanopore Technologies (ONT) SQK-LSK114 native barcoding kit. Sequencing was performed on the GridION platform using a FLOW-MIN114 Spot-ON Flow Cell, R10 version with a translocation speed of 400 bps. Base calling was performed on the GridION using the super-accurate base-calling model, Guppy v7.0.9. For both short- and long-read sequencing, we used Qubit fluorometer (Invitrogen; Grand Island, New York, USA) to measure DNA concentration. Illumina sequencing was carried out at SeqCenter (Pittsburgh, Pennsylvania, USA), while ONT sequencing at SeqCoast Genomics (Portsmouth, New Hampshire, USA).

Genome assembly, quality check and annotation

De novo assembly of Illumina short reads into contiguous sequences was done using shovill v1.1.0 (https://github.com/tseemann/shovill). Trimming of adapter sequences was enabled using the -trim option. Shovill includes methods for subsampling read depth down to 150X, trimming adapters, correcting sequencing errors, and assembling using SPAdes v.3.14.1[82]. For long reads obtained using ONT, adapters were removed using porechop v0.2.4 (https://github.com/rrwick/

Porechop). High quality reads were filtered using filtlong v0.2.1 (https://github.com/rrwick/Filtlong). Reads that were shorter than 1 kb were excluded. Ten percent of the worst reads were also discarded. Hybrid assembly of Illumina and ONT reads was done using unicycler v0.5.0[83].

The sequence quality of assembled *K. pneumoniae* genomes were determined using CheckM v.1.1.3[84] and QUAST v.5.0.2[85]. Using CheckM, we calculated genome completeness ranging between 98.48 and 100 % (mean = 99.87%) and contamination ranging between 0.31 and 1.82% (mean = 0.43%) for Illumina assemblies (Supplementary Data 2 and Fig. S5). For the hybrid assemblies, genome completeness ranged between 98.75 and 100 % (mean = 99.85%) and contamination ranged between 0.33 and 0.85% (mean = 0.45%). The above metrics were all within the genome quality standards recommended by CheckM (Supplementary Data 2). All 136 genome sequences were of high quality consisting of <200 contigs and N50 > 40,000 bp (Supplementary Data 2). The number of contigs in the hybrid assemblies were either two or three, except for KPB68 which had five contigs (two of which were observed not to be of plasmid origin after manual inspection and were excluded in downstream analysis). All plasmid genomes had circular topology. The draft assemblies were annotated using Prokka v.1.14.6[86]. All together, we used 136 short-read genome and 12 long-read genome sequences in all downstream analyses. Associated metadata and genome quality features of all isolates are shown in Supplementary Data 2, Data 6, Fig. S6, Fig. S7.

The species identity of all assembled genomes was confirmed using Kleborate v.2.2.0[41]. Kleborate is a *Klebsiella*-dedicated species assignment and genotyping pipeline that uses genome assemblies as input and compares them to a taxonomically curated genome dataset[41]. Kleborate confirmed all 136 isolates to be true *K. pneumoniae*.

In silico sequence typing and identification of AMR and virulence genes

The *K. pneumoniae* genome assemblies in this study were submitted to the *K. pneumoniae* database on BIGSdb (https://bigsdb.pasteur.fr/klebsiella/) to determine their 7-gene multi-locus sequence types (MLST)[87], ribosomal MLST[88], core genome (cg) MLST based on 629 previously curated core genes[89], sublineages, and clonal groups[31]. One isolate KPB111 was missing the *infB* locus and could not be placed into any known ST designation (provisionally called ST133-1LV). Genome assemblies were screened for the presence of AMR determinants using Kleborate v.2.2.0[41] and AMRFinderPlus v.3.10.23[34]. The presence of *K. pneumoniae*-specific hypervirulence factors were identified using Kleborate v.2.2.0[41].

Pan-genome estimation and phylogenetic tree reconstruction

The annotated genomes were used as input to characterize the pan-genome[90], i.e., the totality of genes of all strains in our dataset, using Roary v.3.13.0[91]. Nucleotide sequences were aligned using MAFFT v.7.471[92]. Sequence alignments of the 3,511 core genes (i.e., gene families present in 99% or 134–136 genomes) were concatenated to generate the core genome alignment. SNPs were extracted from the core genome alignment using SNP-sites v.2.5.1[93]. The core SNP alignment was used as input for building a maximum likelihood phylogenetic tree using RAxML v.8.2.12[94] with a generalized time reversible (GTR)[95] model of nucleotide substitution and gamma distribution of rate heterogeneity. The phylogenetic trees were visualized and annotated using figtree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and Interactive Tree of Life (iTOL) v.6.8.1[96].

In silico plasmid analysis

Using the ONT sequences, complete and circular plasmid sequences were clustered using Mash v.2.3[97]. Accuracy of the distance estimation was improved by implementing a minimum abundance finder --mindepth 0 to ignore likely read errors from unique kmers. Mash distance with threshold of 0.001 was used to define highly similar plasmid sequences[74]. Plasmid incompatibility types were determined using the Plasmidfinder database accessed on December 6, 2023[98] implemented on ABRicate v.1.0.1 (https://github.com/tseemann/abricate). To identify instances of genome rearrangement and gene loss, plasmid sequences were aligned, visualized and annotated using progressive MAUVE[99]. Prediction of the origin of transfer (*oriT*) site and relaxases was done using oriTfinder[100]. Plasmid assemblies were annotated using Bakta v.1.4.0[101] and Artemis[102]. Plasmid sequences were visualized and annotated using Easyfig (https://mjsull.github.io/Easyfig/) and Geneious Prime v. 2022.2.1 (http://www.geneious.com).

Statistical analysis

We determined the concordance between the results of the in vitro testing of antimicrobial susceptibility and the presence of corresponding genetic elements conferring resistance against specific antimicrobial classes by calculating the number of true positives, true negatives, false positives, and false negatives. To calculate sensitivity and specificity of a diagnostic test for AMR, we used the epi.tests function of the epiR package v.2.0.57 implemented in RStudio v.2023.12.0 + 369[103], following the protocol in ghruR (https://gitlab.com/cgps/ghru/ghrur/). AMR sensitivity and specificity were determined with 95% confidence bounds. AMR sensitivity or true positive rate measures the proportion of truly resistant strains, whereas AMR specificity or true negatives measures the proportion of truly susceptible strains at 95% confidence level. Simpson's diversity index[104] was calculated using the R package vegan[105]. All plots (bar plots, bubble plots, UpSet plots, presence/absence matrices) were generated using RStudio v.2023.12.0 + 369[103].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset supporting the conclusions of this article is included within the article and its supplementary files. Short reads data of the 137 *K. pneumoniae* isolates (136 from blood and one from urine) in this study have been deposited in the National Center for Biotechnology Information (NCBI) Short Reads Archive (SRA) under BioProject PRJNA1054115. SRA genome accession numbers are listed in Supplementary Data 2. The 137 *K. pneumoniae* Illumina-sequenced genomes are also available on PathogenWatch: https://pathogen.watch/collection/d2tfj6eocjfe-klebsiella-pneumoniae-isolated-from-invasive-infections-in-a-loc. Plasmid genome assemblies have been deposited in NCBI's BankIT database. Accession numbers for plasmid genomes are listed in Supplementary Data 7.

References

1. Paczosa, M. K. & Mecsas, J. Klebsiella pneumoniae: going on the offense with a strong defense. *Microbiol Mol. Biol. Rev.* **80**, 629–661 (2016).
2. Kang, C.-I. et al. Community-acquired versus nosocomial Klebsiella pneumoniae bacteremia: clinical features, treatment outcomes, and clinical implication of antimicrobial resistance. *J. Korean Med. Sci.* **21**, 816–822 (2006).
3. Vading, M., Nauclér, P., Kalin, M. & Giske, C. G. Invasive infection caused by Klebsiella pneumoniae is a disease affecting patients with high comorbidity and associated with high long-term mortality. *PLoS One* **13**, e0195258 (2018).
4. Marshall, J. C., Maier, R. V., Jimenez, M. & Dellinger, E. P. Source control in the management of severe sepsis and septic shock: an evidence-based review. *Crit. Care Med.* **32**, S513–S526 (2004).
5. Timsit, J.-F. et al. Treatment of bloodstream infections in ICUs. *BMC Infect. Dis.* **14**, 489 (2014).

6. Diekema, D. J. et al. The microbiology of bloodstream infection: 20 year trends from the SENTRY antimicrobial surveillance program. *Antimicrob. Agents Chemother.* **63**, e00355-19 (2019).
7. G. B. D. 2019 Antimicrobial Resistance Collaborators. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the global burden of disease Study 2019. *Lancet* **400**, 2221–2248 (2022).
8. Kohler, P. P. et al. Carbapenem resistance, initial antibiotic therapy, and mortality in Klebsiella pneumoniae bacteremia: a systematic review and meta-analysis. *Infect. Control Hosp. Epidemiol.* **38**, 1319–1328 (2017).
9. Maraolo, A. E. et al. The impact of carbapenem resistance on mortality in patients with Klebsiella pneumoniae bloodstream infection: an individual patient data meta-analysis of 1952 patients. *Infect. Dis. Ther.* **10**, 541–558 (2021).
10. Wyres, K. L., Lam, M. M. C. & Holt, K. E. Population genomics of Klebsiella pneumoniae. *Nat. Rev. Microbiol.* **18**, 344–359 (2020).
11. Kochan, T. J. et al. Genomic surveillance for multidrug-resistant or hypervirulent Klebsiella pneumoniae among United States bloodstream isolates. *BMC Infect. Dis.* **22**, 603 (2022).
12. Sawa, T., Kooguchi, K. & Moriyama, K. Molecular diversity of extended-spectrum β-lactamases and carbapenemases, and antimicrobial resistance. *J. Intensive Care* **8**, 13 (2020).
13. Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).
14. Scheuerman, O. et al. Comparison of predictors and mortality between bloodstream infections caused by ESBL-producing Escherichia coli and ESBL-producing Klebsiella pneumoniae. *Infect. Control Hosp. Epidemiol.* **39**, 660–667 (2018).
15. Sianipar, O., Asmara, W., Dwiprahasto, I. & Mulyono, B. Mortality risk of bloodstream infection caused by either Escherichia coli or Klebsiella pneumoniae producing extended-spectrum β-lactamase: a prospective cohort study. *BMC Res. Notes* **12**, 719 (2019).
16. Li, D. et al. Klebsiella pneumoniae bacteremia mortality: a systematic review and meta-analysis. *Front Cell Infect. Microbiol.* **13**, 1157010 (2023).
17. Paniagua-García, M. et al. Attributable mortality of infections caused by carbapenem-resistant Enterobacterales: results from a prospective, multinational case-control-control matched cohorts study (EURECA). *Clin. Microbiol. Infect.* **30**, 223–230 (2024).
18. Xu, L., Sun, X. & Ma, X. Systematic review and meta-analysis of mortality of patients infected with carbapenem-resistant Klebsiella pneumoniae. *Ann. Clin. Microbiol Antimicrob.* **16**, 18 (2017).
19. Haller, S. et al. What caused the outbreak of ESBL-producing Klebsiella pneumoniae in a neonatal intensive care unit, Germany 2009 to 2012? reconstructing transmission with epidemiological analysis and whole-genome sequencing. *BMJ Open* **5**, e007397 (2015).
20. Frenk, S. et al. Investigation of outbreaks of extended-spectrum beta-lactamase-producing Klebsiella pneumoniae in three neonatal intensive care units using whole genome sequencing. *Antibiotics (Basel)* **9**, 705 (2020).
21. Pruss, A. et al. Epidemiological analysis of extended-spectrum β-lactamase-producing Klebsiella pneumoniae outbreak in a neonatal clinic in Poland. *Antibiotics (Basel)* **12**, 50 (2022).
22. Aldeyab, M. A. et al. Exploring thresholds and interaction effects among antibiotic usage, covariates, and their effect on antibiotic resistance using an extended-spectrum β-lactamase-producing Klebsiella pneumoniae case. *Expert Rev. Anti Infect. Ther.* **21**, 777–786 (2023).
23. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet.* **399**, 629–655 (2022).
24. Alvarez-Uria, G., Gandra, S., Mandal, S. & Laxminarayan, R. Global forecast of antimicrobial resistance in invasive isolates of Escherichia coli and Klebsiella pneumoniae. *Int J. Infect. Dis.* **68**, 50–53 (2018).
25. Holt, K. E. et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. *Proc. Natl Acad. Sci. USA* **112**, E3574–E3581 (2015).
26. Hawkey, J. et al. ESBL plasmids in Klebsiella pneumoniae: diversity, transmission and contribution to infection burden in the hospital setting. *Genome Med.* **14**, 97 (2022).
27. Teo, T.-H. et al. Differential mucosal tropism and dissemination of classical and hypervirulent Klebsiella pneumoniae infection. *iScience* **27**, 108875 (2024).
28. Russo, T. A. & Marr, C. M. Hypervirulent Klebsiella pneumoniae. *Clin. Microbiol Rev.* **32**, e00001–e00019 (2019).
29. Kochan, T. J. et al. Klebsiella pneumoniae clinical isolates with features of both multidrug-resistance and hypervirulence have unexpectedly low virulence. *Nat. Commun.* **14**, 7962 (2023).
30. Jolley, K. A. & Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, 124 (2018).
31. Hennart, M. et al. A dual barcoding approach to bacterial strain nomenclature: genomic taxonomy of Klebsiella pneumoniae strains. *Mol. Biol. Evol.* **39**, msac135 (2022).
32. Wyres, K. L. & Holt, K. E. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr. Opin. Microbiol* **45**, 131–139 (2018).
33. Neubauer, S. et al. A genotype-phenotype correlation study of SHV β-lactamases offers new insight into SHV resistance profiles. *Antimicrob. Agents Chemother.* **64**, e02293-19 (2020).
34. Feldgarden, M. et al. AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
35. Lin, T.-L. et al. Extended-spectrum beta-lactamase genes of Klebsiella pneumoniae strains in Taiwan: recharacterization of shv-27, shv-41, and tem-116. *Micro. Drug Resist* **12**, 12–15 (2006).
36. Mulvey, M. R. et al. Ambler class A extended-spectrum beta-lactamase-producing Escherichia coli and Klebsiella spp. in Canadian hospitals. *Antimicrob. Agents Chemother.* **48**, 1204–1214 (2004).
37. Hamzaoui, Z. et al. Role of association of OmpK35 and OmpK36 alteration and blaESBL and/or blaAmpC genes in conferring carbapenem resistance among non-carbapenemase-producing Klebsiella pneumoniae. *Int J. Antimicrob. Agents* **52**, 898–905 (2018).
38. Netikul, T. & Kiratisin, P. Genetic characterization of carbapenem-resistant enterobacteriaceae and the spread of carbapenem-resistant Klebsiella pneumonia ST340 at a university hospital in Thailand. *PLoS One* **10**, e0139116 (2015).
39. Neuert, S. et al. Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal Salmonella enterica. *Front Microbiol* **9**, 592 (2018).
40. Wyres, K. L. et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae. *PLoS Genet* **15**, e1008114 (2019).
41. Lam, M. M. C. et al. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. *Nat. Commun.* **12**, 4188 (2021).
42. Wyres, K. L. et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive Klebsiella pneumoniae from South and Southeast Asia. *Genome Med* **12**, 11 (2020).
43. March, C. et al. Role of bacterial surface structures on the interaction of Klebsiella pneumoniae with phagocytes. *PLoS One* **8**, e56847 (2013).

44. Follador, R. et al. The diversity of Klebsiella pneumoniae surface polysaccharides. *Micro. Genom.* **2**, e000073 (2016).
45. Lee, I. R. et al. Differential host susceptibility and bacterial virulence factors driving Klebsiella liver abscess in an ethnically diverse population. *Sci. Rep.* **6**, 29316 (2016).
46. Lam, M. M. C. et al. Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in Klebsiella pneumoniae populations. *Micro. Genom.* **4**, e000196 (2018).
47. Russo, T. A. et al. Differentiation of hypervirulent and classical Klebsiella pneumoniae with acquired drug resistance. *mBio* **15**, e0286723 (2024).
48. Page, M. G. P. The role of iron and siderophores in infection, and the development of siderophore antibiotics. *Clin. Infect. Dis.* **69**, S529–S537 (2019).
49. Lai, Y.-C., Peng, H.-L. & Chang, H.-Y. RmpA2, an activator of capsule biosynthesis in Klebsiella pneumoniae CG43, regulates K2 cps gene expression at the transcriptional level. *J. Bacteriol.* **185**, 788–800 (2003).
50. Zatyka, M. & Thomas, C. M. Control of genes for conjugative transfer of plasmids and other mobile elements. *FEMS Microbiol Rev.* **21**, 291–319 (1998).
51. Fukuda, H. & Ohtsubo, E. Roles of TraI protein with activities of cleaving and rejoining the single-stranded DNA in both initiation and termination of conjugal DNA transfer. *Genes Cells* **2**, 735–751 (1997).
52. Kitchel, B. et al. Molecular epidemiology of KPC-producing Klebsiella pneumoniae isolates in the United States: clonal expansion of multilocus sequence type 258. *Antimicrob. Agents Chemother.* **53**, 3365–3370 (2009).
53. Shropshire, W. C. et al. Accessory genomes drive independent spread of carbapenem-resistant Klebsiella pneumoniae clonal groups 258 and 307 in Houston, TX. *mBio* **13**, e0049722 (2022).
54. David, S. et al. Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. *Nat. Microbiol* **4**, 1919–1929 (2019).
55. Peirano, G., Chen, L., Kreiswirth, B. N. & Pitout, J. D. D. Emerging antimicrobial-resistant high-risk Klebsiella pneumoniae clones ST307 and ST147. *Antimicrob. Agents Chemother.* **64**, e01148–20 (2020).
56. Turton, J. et al. Hybrid resistance and virulence plasmids in 'high-risk' clones of Klebsiella pneumoniae, including those carrying blaNDM-5. *Microorganisms* **7**, 326 (2019).
57. Tumbarello, M. et al. Bloodstream infections caused by extended-spectrum-beta-lactamase-producing Klebsiella pneumoniae: risk factors, molecular epidemiology, and clinical outcome. *Antimicrob. Agents Chemother.* **50**, 498–504 (2006).
58. Shamsrizi, P. et al. Variation of effect estimates in the analysis of mortality and length of hospital stay in patients with infections caused by bacteria-producing extended-spectrum beta-lactamases: a systematic review and meta-analysis. *BMJ Open* **10**, e030266 (2020).
59. Bush, K. & Bradford, P. A. β-lactams and β-lactamase inhibitors: an overview. *Cold Spring Harb. Perspect. Med* **6**, a025247 (2016).
60. Wyres, K. L. et al. Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae strain ST307. *J. Antimicrob. Chemother.* **74**, 577–581 (2019).
61. Castanheira, M. et al. Rapid expansion of KPC-2-producing Klebsiella pneumoniae isolates in two Texas hospitals due to clonal spread of ST258 and ST307 lineages. *Micro. Drug Resist* **19**, 295–297 (2013).
62. Iregui, A., Ha, K., Meleney, K., Landman, D. & Quale, J. Carbapenemases in New York city: the continued decline of KPC-producing Klebsiella pneumoniae, but a new threat emerges. *J. Antimicrob. Chemother.* **73**, 2997–3000 (2018).
63. Lapp, Z. et al. Distinct origins and transmission pathways of blaKPC Enterobacterales across three U.S. states. *J. Clin. Microbiol.* **61**, e0025923 (2023).
64. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinform.* **11**, 595 (2010).
65. Argimón, S. et al. Rapid genomic characterization and global surveillance of Klebsiella using Pathogenwatch. *Clin. Infect. Dis.* **73**, S325–S335 (2021).
66. Ellington, M. J. et al. Contrasting patterns of longitudinal population dynamics and antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single center. *Genome Biol.* **20**, 184 (2019).
67. Tang, N. et al. Epidemicity and clonal replacement of hypervirulent carbapenem-resistant Klebsiella pneumoniae with diverse pathotypes and resistance profiles in a hospital. *J. Glob. Antimicrob. Resist.* **32**, 4–10 (2023).
68. Heinz, E. et al. Longitudinal analysis within one hospital in sub-Saharan Africa over 20 years reveals repeated replacements of dominant clones of Klebsiella pneumoniae and stresses the importance to include temporal patterns for vaccine design considerations. *Genome Med.* **16**, 67 (2024).
69. Owen, J. R. et al. Whole-genome sequencing and concordance between antimicrobial susceptibility genotypes and phenotypes of bacterial isolates associated with bovine respiratory disease. *G3 (Bethesda)* **7**, 3059–3071 (2017).
70. Stubberfield, E. et al. Use of whole genome sequencing of commensal Escherichia coli in pigs for antimicrobial resistance surveillance, United Kingdom, 2018. *Eur. Surveill.* **24**, 1900136 (2019).
71. Chang, F. Y., Siu, L. K., Fung, C. P., Huang, M. H. & Ho, M. Diversity of SHV and TEM beta-lactamases in Klebsiella pneumoniae: gene evolution in Northern Taiwan and two novel beta-lactamases, SHV-25 and SHV-26. *Antimicrob. Agents Chemother.* **45**, 2407–2413 (2001).
72. Martin, M. J. et al. Anatomy of an extensively drug-resistant Klebsiella pneumoniae outbreak in Tuscany, Italy. *Proc. Natl Acad. Sci. USA* **118**, e2110227118 (2021).
73. Spadar, A., Perdigão, J., Campino, S. & Clark, T. G. Large-scale genomic analysis of global Klebsiella pneumoniae plasmids reveals multiple simultaneous clusters of carbapenem-resistant hypervirulent strains. *Genome Med.* **15**, 3 (2023).
74. Roberts, L. W. et al. Long-read sequencing reveals genomic diversity and associated plasmid movement of carbapenemase-producing bacteria in a UK hospital over 6 years. *Micro. Genom.* **9**, mgen001048 (2023).
75. Arena, F. et al. First case of bacteremic liver abscess caused by an ST260-related (ST1861), hypervirulent Klebsiella pneumoniae. *J. Infect.* **73**, 88–91 (2016).
76. Struve, C. et al. Mapping the evolution of hypervirulent Klebsiella pneumoniae. *mBio* **6**, e00630 (2015).
77. Shankar, C. et al. Emergence of multidrug resistant hypervirulent ST23 Klebsiella pneumoniae: multidrug resistant plasmid acquisition drives evolution. *Front Cell Infect. Microbiol.* **10**, 575289 (2020).
78. Gálvez-Silva, M. et al. Carbapenem-resistant hypervirulent ST23 Klebsiella pneumoniae with a highly transmissible dual-carbapenemase plasmid in Chile. *Biol. Res.* **57**, 7 (2024).
79. Evans, D. R. et al. Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. *Elife* **9**, e53886 (2020).
80. Berinson, B. et al. Usefulness of BioFire FilmArray BCID2 for blood culture processing in clinical practice. *J. Clin. Microbiol.* **59**, e0054321 (2021).
81. Clinical and Laboratory Standards Institute. *CLSI M100 Performance Standards for Antimicrobial Susceptibility Testing*. https://clsi.org/media/3481/m100ed30_sample.pdf (2018).

82. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
83. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* **13**, e1005595 (2017).
84. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
85. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
86. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
87. Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A. D. & Brisse, S. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. *J. Clin. Microbiol.* **43**, 4178–4182 (2005).
88. Jolley, K. A. et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiol. (Read.)* **158**, 1005–1015 (2012).
89. Bialek-Davenet, S. et al. Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups. *Emerg. Infect. Dis.* **20**, 1812–1820 (2014).
90. Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet Dev.* **15**, 589–594 (2005).
91. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
92. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
93. Page, A. J. et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Micro. Genom.* **2**, e000056 (2016).
94. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
95. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. in *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, 57–86 (Amer Mathematical Society, 1986).
96. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
97. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
98. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
99. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**, e11147 (2010).
100. Li, X. et al. oriTfinder: a web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. *Nucleic Acids Res.* **46**, W229–W234 (2018).
101. Schwengers, O. et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Micro. Genom.* **7**, 000685 (2021).
102. Rutherford, K. et al. Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
103. RStudio Team. *RStudio: Integrated Development for R*. http://www.rstudio.com/ (2020).
104. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688–688 (1949).
105. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).

Acknowledgements

The authors thank Elissa Eckhardt and Nisalda Carreiro for their operational leadership in the DHMC clinical microbiology laboratory. The authors thank the SUNY University at Albany Information Technology Services where all bioinformatics analyses were carried out. We are grateful to the Institut Pasteur team for the curation and maintenance of BIGSdb-Pasteur databases at https://bigsdb.pasteur.fr/. C.P.A. thanks Reginald and Priscilla Mae Farnsworth for insightful discussion. This research was funded by grant from the National Institutes of Health to C.P.A. (Award no. R35GM142924). The funders had no role in study design, data collection and analysis, decision to publish, and preparation of the manuscript. The findings in this study do not necessarily reflect views and policies of the authors' institutions and funders.

Author contributions

C.P.A., O.O.I. and I.W.M. designed and guided the work. O.O.I. and I.J.A performed all bioinformatics analyses. O.O.I., N.I.Z.S., M.M.M. and S.S.R.S. carried out subculturing and DNA extractions. I.W.M. oversaw bacterial sampling and phenotypic testing. A.W. and I.W.M. carried out metadata collection. C.P.A. and O.O.I. wrote the initial manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51374-x.

Correspondence and requests for materials should be addressed to Odion O. Ikhimiukor or Cheryl P. Andam.

Peer review information *Nature Communications* thanks the anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.