Identification of A Novel CG307 Sub-clade in Third Generation Cephalosporin Resistant *Klebsiella pneumoniae* Causing Invasive Infections in the United States

Selvalakshmi Selvaraj Anand[1], Chin-Ting Wu[1], Jordan Bremer[2], Micah Bhatti[3], Todd J Treangen[4], Awdhesh Kalia[1], Samuel A Shelburne[2,5#], William C Shropshire[2]

[1]Graduate Program in Diagnostic Genetics and Genomics, MD Anderson Cancer Center School of Health Professions, Houston, TX, USA

[2]Department of Infectious Diseases, Infection Control, and Employee Health, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[3]Department of Laboratory Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[4]Department of Computer Science, Rice University, Houston, TX, USA

[5]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

#Address for correspondence: Samuel A Shelburne, MD, PhD, 1901 East Rd., South Campus Research Building, 4SCR5.1042, Houston, TX 77054; phone: (713) 792-7629, e-mail: sshelburne@mdanderson.org

Keywords. *3rd generation cephalosporin resistant Klebsiella pneumoniae, clonal group 307, multi-drug resistant K. pneumoniae surveillance, nosocomial transmission, accessory genome*

ABSTRACT

Despite the notable clinical impact, recent molecular epidemiology regarding third-generation cephalosporin-resistant *Klebsiella pneumoniae* (3GC-R*Kp*) in the United States remains limited. We performed whole genome sequencing of 3GC-R*Kp* bacteremia isolates collected from March 2016 to May 2022 at a tertiary care cancer center in Houston, TX using Illumina and Oxford Nanopore Technologies platforms. A comprehensive comparative genomic analysis was performed to dissect population structure, transmission dynamics, and pan-genomic signatures of our 3GC-R*Kp* population. Of the 194 3GC-R*Kp* bacteremias that occurred during our study timeframe, we were able to analyze 153 (79%) bacteremia isolates, 126 initial and 27 recurrent isolates respectively. While isolates belonging to the widely prevalent clonal group (CG) 258 were rarely observed, the predominant clonal group, CG307, accounted for 37 (29%) index isolates and displayed a significant correlation (Pearson correlation test *P*-value = 0.03) with the annual frequency of 3GC-R*Kp* bacteremia. Within our CG307 cohort, 89% (33/37) of our isolates belong to the global rather than previously described Texas-specific clade. Strikingly, we identified a new CG307 sub-clade (*i.e.,* cluster 1 isolates) comprised of 18 isolates characterized by the chromosomally-encoded $bla_{\text{SHV-205}}$ and unique accessory genome content. This CG307 sub-clade was detected in various United States regions, with genome sequences from 24 additional strains becoming recently available in the NCBI SRA database. Collectively, this study underscores the emergence and dissemination of a distinct CG307 sub-clade that is a prevalent cause of 3GC-R*Kp* bacteremia among cancer patients seen in Houston, TX and has recently been isolated throughout the United States.

DATA SUMMARY

WGS data sequenced during this study period was submitted to NCBI and can be accessed within BioProject PRJNA648389. WGS data from previous study of carbapenem non-

susceptible *Enterobacterales* can be accessed from BioProject PRJNA836696. Assembly information and BioAccession numbers are provided in Table S1.

IMPACT STATEMENT

Infections due to 3rd generation cephalosporin resistant *Klebsiella pneumoniae* (3GC-R*Kp*) are considered among the most urgent public health threats. However, molecular epidemiology studies on 3GC-R*Kp* in the United States are limited. Our analysis indicates a preponderance of genetically diverse 3GC-R*Kp* isolates harboring the key antimicrobial resistance determinant $bla_{\text{CTX-M-15}}$ at our institution. Importantly, however, we detected evidence of long duration transmission of highly genetically related CG307 and CG29 specific clusters at our institution. Interestingly, we rarely detected the pandemic CG258 lineage in our cohort and did not detect more than two genetically related CG258 isolates from this lineage. We found that 90% of our isolates from the most prevalent clonal group, CG307, belonged to a novel, nested-population of a “global” CG307 clade in contrast to the more commonly detected “Texas-specific” clade that has circulated in our region. We searched the NCBI SRA database using genomic markers of the novel CG307 clade and found evidence of this clade causing recent invasive infections in other locations across the United States. Our study highlights the shifting population dynamics of *K. pneumoniae* causing invasive infections and the necessity to continue AMR surveillance in order to identify emerging high-risk populations.

INTRODUCTION

The necessity to address antimicrobial resistant (AMR) infections as a global public health threat has become increasingly apparent during the 21st century. A recent study estimated that 4.95 million deaths were associated with AMR infections worldwide with *Klebsiella pneumoniae* infections being the third most common pathogen associated with mortality [1]. *K. pneumoniae*, a member of the *Enterobacterales* family, is an opportunistic pathogen that has the capacity to develop resistance to multiple classes of antibiotics. A recent meta-analysis estimated about 33% of nosocomial *K. pneumoniae* infections are caused by MDR strains highlighting the impact of these pathogens in the healthcare setting [2].

MDR *K. pneumoniae* clonal populations can acquire β-lactamase encoding genes such as extended-spectrum β-lactamase (ESBL) or AmpC encoding genes via horizontal gene transfer (HGT), which results in third-generation cephalosporin resistance (3GC-R). These HGT plasmid vectors can also harbor carbapenemase encoding genes such as the $bla_{KPC,}$ which encodes the *Klebsiella pneumoniae* carbapenemase (KPC) that effectively hydrolyzes most classes of β-lactams [3-12]. One of the most concerning MDR *K. pneumoniae* clonal populations is the pandemic group of strains known as clonal group 258 (CG258), which include sequence types 258 (ST258), ST512, and ST11, that are all closely related (*i.e.,* the CG258 average pairwise SNP distance is ~214) and strongly associated with worldwide bla_{KPC} dissemination [6, 13-17]. ST258 strains are the predominant MDR *K. pneumoniae* detected in the United States over the past two decades [5, 7, 16-19].

While ST258 prevalence has remained high in certain regions of the United States, there is growing evidence of an emergent CG307 lineage co-circulating with CG258 in the Houston, TX region [7, 9, 20, 21]. When performing comparative genomics between the two lineages, a discerning factor is the strong association of the ESBL encoding gene $bla_{CTX\text{-}M\text{-}15}$ that is present

in >90% of CG307 isolates whereas *bla*$_{CTX-M-15}$ is detected in <10% of CG258 isolates [9, 10, 20]. Recent molecular epidemiological investigations of CG307 isolates have identified two predominant CG307 clades with a strong phylogenetic signal that distinguishes isolates of Texas origin from those from the more globally disseminated clade [10, 20]. When comparing the two ST307 lineages, the Texas-specific clade has a stable chromosomal insertion of two IS*Ecp1-bla*$_{CTX-M-15}$ transposition units whereas the more globally disseminated CG307 lineage harbors *bla*$_{CTX-M-15}$ primarily on large, multireplicon, F-type conjugative plasmids [8, 10, 20]. Furthermore, recent AMR surveillance studies from Europe have highlighted how CG307 strains are the primary drivers of 3GC-RKp (3GC-R*Kp*) infections within hospitals in part due to their strong association with *bla*$_{CTX-M-15}$ carriage and transmission [22-24].

Given the high prevalence of CG307 in Houston and surrounding areas [7, 9, 20, 21, 25, 26] as well as a lack of current US-based 3GC-R*Kp* surveillance studies, we sought to use WGS to determine the molecular epidemiology of 3GC-R*Kp* causing bacteremia in cancer patients at our institution within a five-year timeframe. We found that CG307 strains were the most common CG detected whereas the previously highly prevalent CG258 was rarely identified. Interestingly, 90% of CG307 bacteremias were caused by strains of the global clade; in particular, we identified that a previously unidentified, global sub-clade with distinct genomic signatures caused approximately 50% of the CG307 bacteremias at our institution. Search of publicly available WGS data identified isolates of this sub-clade recently collected from geographically diverse sites throughout the United States. These data expand upon the rapidly changing epidemiology of 3GC-R*Kp*, including the increasing importance of emerging, diverse CG307 sub-clades.

METHODS

Study design and sample collection

The study included all K. *pneumoniae* bacteremias that occurred from March 1st, 2016, chosen coincident with the date of implementation of the Epic Electronic Health Record system at the University of Texas MD Anderson Cancer Center (MDACC) in Houston, TX, to May 30th, 2022. We defined an index *K. pneumoniae* bacteremia isolate as first occurrence of a positive blood culture and a recurrent *K. pneumoniae* bacteremia isolate as a positive blood culture that occurred at least 14 days from a previous *K. pneumoniae* bacteremia isolate. Antibiotic susceptibility testing of all *K. pneumoniae* bacteremia isolates was performed by the MDACC microbiology laboratory using the Accelerate PhenoTest™ BC (Accelerate Diagnostics), ETEST® (bioMérieux), and VITEK® 2 (bioMérieux) with susceptibility interpretations based on CLSI guidelines [27]. *K. pneumoniae* bacteremia isolates were considered as extended-spectrum cephalosporin resistant (3GC-R) if they had a ceftriaxone (CRO) minimum inhibitory concentration (MIC) $\geq$ 4 μg/mL and/or a predicted ESBL phenotype. Out of the 194 3GC-R*Kp* causing bacteremia during the study period, there were 79% (161/194) 3GC-R*Kp* isolates available for sequencing. Further information on sampling is included in the results section.

Short-read and long-read sequencing

3GC-R*Kp* isolates were stocked in thioglycolate broth supplemented with 40% glycerol as part of an ongoing surveillance of bacteremia isolates. This surveillance study has received ethical approval through the University of Texas MD Anderson Cancer Center (MDACC) Institutional Review Board (Protocol ID: PA13-0334). MicroBank (MB) numbers are unique to patient infections and deidentified. Samples were plated onto Trypticase Soy Agar (TSA with Sheep Blood). These plates were subsequently incubated overnight at 37°C and 4% CO_2. After 12-24 hours, a single colony was inoculated in autoclaved Miller's LB Broth and incubated at 37°C for

2 hours with mild agitation (225 rpm). Two mL of the inoculated media were aliquoted into 15 mL Eppendorf tubes and spun down to pellet cells. Genomic DNA (gDNA) was extracted according to the protocol specified on QIAGEN “Blood and Cell Culture DNA Kit.” (Cat No. 69581).

ESC-R-*K. pneumoniae* isolates were sequenced by Illumina NovaSeq 6000 as previously described [21]. There were 126 3GC-R*Kp* isolates with 150 bp paired-end reads that passed quality control as assessed using fastqc-v0.11.9 (https://github.com/s-andrews/FastQC). We performed long-read sequencing on nine isolates of interest that were part of transmission clusters to obtain their complete genomes using the Oxford Nanopore Technologies (ONT) MinION platform as described previously [21]. Briefly, genomic DNA extracted for the purpose of short-read sequencing was used as input for ONT long-read sequencing. Library preparation was accomplished using the Rapid Sequencing Kit 96 V10 (SQK-RBK110.96). The input gDNA was normalized to 50 ng to ensure even distributions of libraries across pooled samples. The prepared libraries were loaded onto the R9.4.1 flowcell (FLO-MIN106D) for sequencing with MinKNOW software to generate fast5 files. Guppy-v6.4.6 basecaller was used to perform basecalling from the fast5 files to obtain fastq files using super high accuracy model (SUP).

Short and long-read sequencing data analysis

Quality assessed, paired-end 150 bp reads were assembled via SPAdes-v3.15.5 [28] with default parameters and the inclusion of the isolate option. The quality assessment, assignment of sequence types, clonal groups, and capsule type of isolates was assessed using Kleborate-v2.2.0 with Kaptive-v2.0.0 activated [29, 30]. One genome assembly had Kleborate QC warnings due to ambiguous bases, genome size > 7.5 Mbp or < 4.5 Mbp, or N50< 10 Kbp and was removed from analysis. The COpy Number Variant quantifICation Tool (convict-v1.0) was used to identify antimicrobial resistance (AMR) from the ResFinder database (Accessed 2021-

11-09) and estimate gene copy numbers as described previously (Shropshire, W convict GitHub: https://github.com/wshropshire/convict) [21]. CONVICT as part of its pipeline employs the KmerResistance-2.0 bioinformatic tool that uses short-read k-mer alignment (KMA) to identify homologous AMR genes from redundant databases [31, 32]. The NCBI AMRFinderPlus command-line tool (v3.11.14) was used to confirm AMR gene identification form CONVICT using the NCBI curated database (version 2023-04-17.1).

A short- and long-read assembly pipeline (Shropshire, W flye_hybrid_assembly_pipeline GitHub: https://github.com/wshropshire/flye_hybrid_assembly_pipeline) was used to close complete genomes of ONT sequenced data as described previously [21]. Incomplete assemblies were re-assembled using Unicycler-v0.5.0 and manually curated for errors using short- and long-read pileups and visualizing with the integrated genome browser (IGV-v2.14.1) [33]. The high-performance computing (HPC) cluster, Seadragon, that is hosted through MDACC was used to perform genomic analyses. Sequencing QC, 3GC-R gene determinants, and antimicrobial susceptibility testing for isolates sequenced for this project in addition to previous MDACC studies is available in Table S1.

Core Gene and Genome Analyses of 3GC-R*Kp* Isolates

Genomes from index isolates (n=126) were annotated using Prokka-v1.14.5 and the annotation files in generalized feature format (GFF3) were used for the subsequent pan-genome analysis [34]. The gff3 files from Prokka were used to make a core gene alignment through Roary-v3.13 using MAFFT-v7.4 [35, 36]. The index 3GC-R*Kp* core gene length was 3720547 bases. The pairwise SNP distances from the core gene alignment were identified using snp-dists-v0.8.4 (https://github.com/tseemann/snp-dists). IQ-TREE-v2.0.6 was used to create a maximum likelihood phylogeny using the core gene alignment to determine the population structure [37]. Bootstrap analysis was performed using UFBoot approximation and SH-like approximate

likelihood test with 1000 replicates though IQ-TREE-v2.0.6 [37, 38]. The tree visualization was performed using the R package ggtree-v3.9.1.

PopPUNK-v2.6.0 was used with index 3GC-R*Kp* draft assemblies (n=126) to identify potential cluster networks based on the core and accessory genome [39]. Following assignment, a core SNP phylogeny for each PopPUNK group was performed with a complete genome reference isolate using snippy-v4.6.0 (https://github.com/tseemann/snippy). The recombination regions were masked using Gubbins-v2.3.4 [36] and the filtered alignment file was used as the input file to obtain a core SNP maximum likelihood phylogeny using IQ-TREE-v2.0.6 [37]. The filtered polymorphic sites were used to generate pairwise SNP matrices to determine genetic relatedness within each specific PopPUNK group.

For CG307 population structure analysis, we performed a convenience sampling of publicly available CG307 isolates and oversampled for CG307 isolates harboring *bla*$_{\text{SHV-205}}$ alleles. CG307 short-reads were downloaded from NCBI using the sratoolkit-v2.10.9 through the fasterq-dump function (https://github.com/ncbi/sra-tools). SRA accession numbers and metadata for CG307 isolates are available on Table S2. A core SNP phylogeny for the CG307 isolates was performed using Kp616 (GenBank Accession Number: GCA_003076555.1) as a reference isolate using snippy-v4.6.0 as described in the paragraphs above. The Kp616 isolate was selected as a reference as it was previously used by Wyres et al. for comparing CG307 population structure in addition to the genome being closed, completely resolved, and collected in 2009 [10]. The recombination regions were masked using Gubbins-v2.3.4 and a core SNP maximum likelihood phylogeny using IQ-TREE-v2.0.6 was performed as previously mentioned [37]. The per branch statistics output from Gubbins was analyzed to look for regions of high recombination.

A Bayesian analysis of population structure was performed using the core genome alignment through rhierBAPS-v1.0.1 to identify CG307 specific clusters[40]. Bayesian dating of the nodes of the CG307 phylogenetic tree was performed using BactDating-v1.1.0 (https://xavierdidelot.github.io/BactDating/). The recombination free core SNP phylogenetic tree output from Gubbins-v2.3.4 was used to perform root-to-tip regression analysis prior to molecular dating (Fig. S1). The output timed tree was further annotated and visualized using R package ggtree-v3.9.1.

Accessory genome analysis

We used a subset of the accessory genome excluding low (<5%) and high frequency (>95%) genes for agglomerative hierarchical clustering of the gene presence/absence matrix from Roary to analyze the accessory genomes of CG307. Comparison of plasmid vectors with BLAST ring image generator (BRIG) was performed using the Proksee webserver [41]. Phage content in the genomes were characterized using the PHASTER webserver (https://phaster.ca/). Replicon types were identified using the "PlasmidFinder" database (Accessed 2023-07-01) with ABRicate-v1.0.0 (https://github.com/tseemann/abricate). Data visualization was generated using R-v4.0.4 packages or Geneious Prime software (2023.2.1).

Statistical analysis

Time series data were assessed using the Mann-Kendall Trend test. Comparisons between rates of *K. pneumoniae* bacteremias by 6-month groups were performed using Student's *t*-test. Chi-squared tests were used to compare group proportions between *K. pneumoniae* bacteremias. All statistics was computed on R-v4.0.4 or Prism 9 with a type one error rate (α=0.05) used across all hypothesis testing.

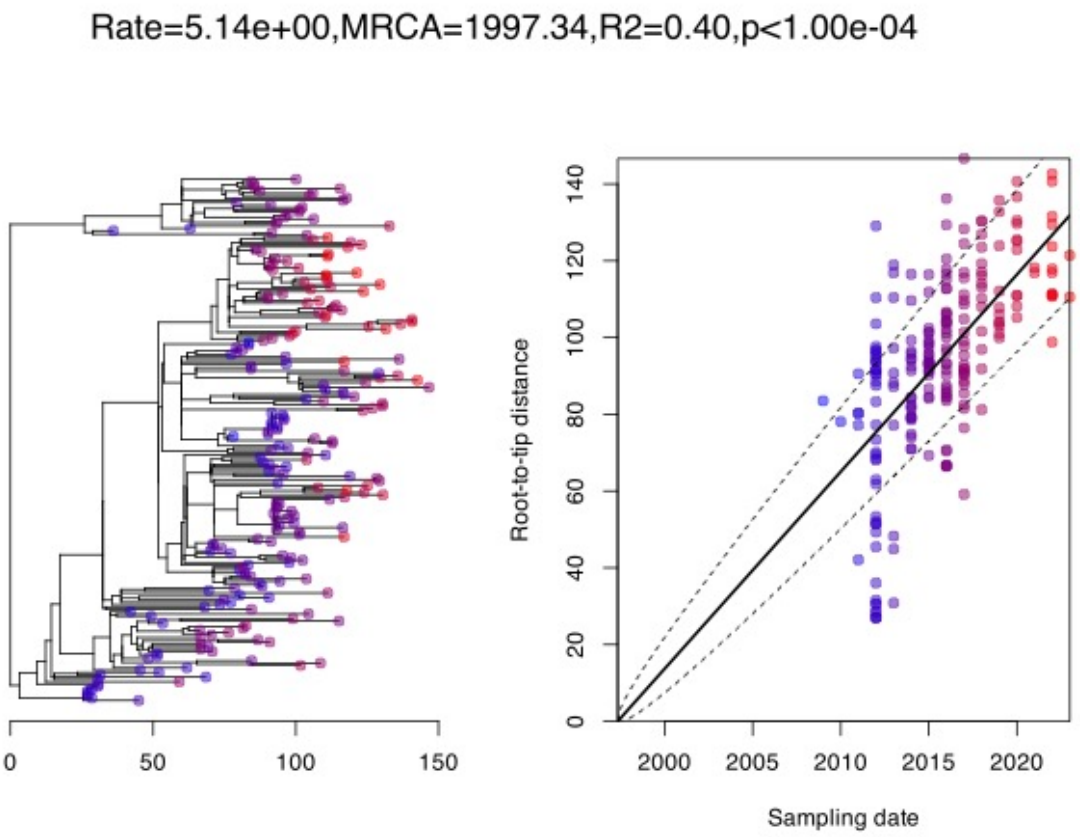
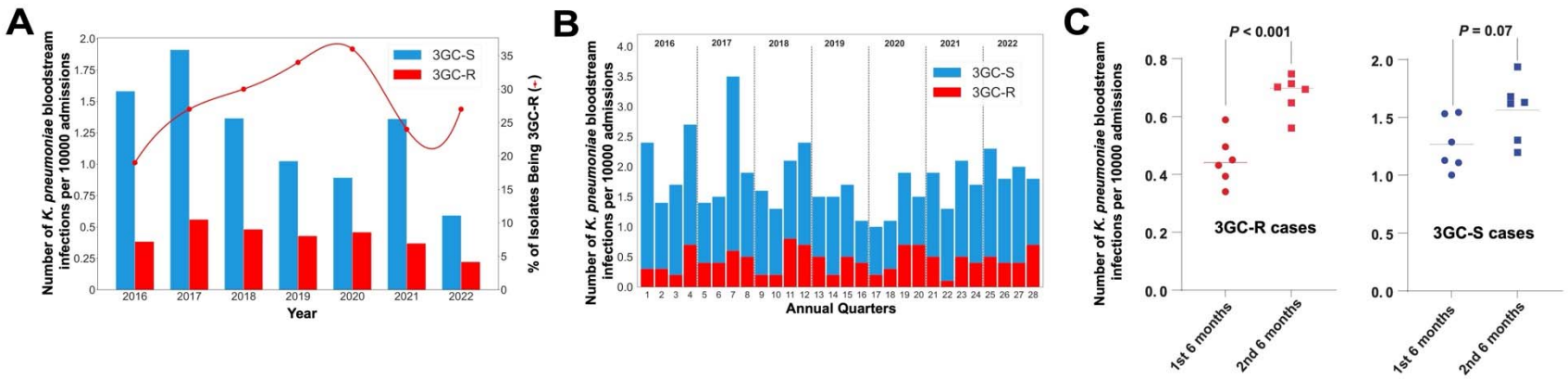


Fig. S1. Initial root-to-tip correlation analysis of 224 CG307 isolates. Initial test of Gubbins phylogeny with an estimate mu=5.14 substitutions/genome, an MRCA = 1997, and significant test of temporal signal. Model testing with a null temporal signal (*i.e.,* all sampling dates equal) phylogeny indicates temporal signal model (model=carc) was significant. We found a strong correlation between the sampling dates and the root-to-tip distances in addition to the significant temporal signal (*P*-value<0.05). We selected for 10000 Markov Chain Monte Carlo simulations using a 'carc' model to obtain the dated phylogeny.

RESULTS

Identification of higher 3GC-R*Kp* bacteremia prevalence in latter half of annual periods

Out of 668 *K. pneumoniae* bacteremias that occurred from March 2016 to May 2022, 83% (554/668) were index infections (*i.e.*, the first occurrence of patient *K. pneumoniae* bacteremia episode). There were 114 (17%) index bacteremias with one or more recurrent infections (*i.e.*, bacteremia occurring ≥ 14 days following an index bacteremia episode). 3GC-R*Kp* infections accounted for 29% (194/668) of the *K. pneumoniae* bacteremias, of which 72% (139/194) were index infections. The proportion of index bacteremias that had recurrent episodes were significantly higher for 3GC-R*Kp* (39%, 55/139) compared to third generation cephalosporin susceptible *K. pneumoniae* (3GC-S*Kp*) infections (14%, 59/415, χ^2 *P*-value < 0.001). When normalized by the number of admissions to account for fluctuations in patient volume, the rates of index 3GC-R*Kp* bacteremias were not significantly different over the course of the study (Mann Kendall trend test *P*-value > 0.05) (Fig. 1A). Similar to our recent study of 3GC-R *Escherichia coli* bacteremia [42], 3GC-R*Kp* bacteremia prevalence was significantly higher in the last six months of the year relative to the first six months (0.67 vs. 0.45 bacteremias per month per 1,000 patient admissions respectively; Student's *t*-test *P*-value < 0.001) (Fig. 1B, 1C). No statistically significant difference in 3GC-S*Kp* bacteremias stratified by time of year was observed (1.6 vs. 1.3 bacteremias per month per 1,000 patient admissions in 2nd half of year vs. 1st half of year, respectively; Students *t*-test *P*-value = 0.07, Fig. 1C). These data indicate that for the past several years the rates of 3GC-R*Kp* bacteremia have not been significantly changing and that environmental factors may contribute to 3GC-R*Kp* infections.



253

254 **Fig. 1.** Epidemiology of *K. pneumoniae* bacteremias. (A) Index *K. pneumoniae* bacteremias by year. The x-axis shows the year, and
255 the y-axis (left) shows the normalized number of bacteremias with blue bars representing the 3GC-S*Kp* and red bars representing
256 the 3GC-R*Kp* infections, respectively. The y-axis (right) shows the proportion of 3GC-R*Kp* isolates amongst total *K. pneumoniae*
257 bacteremias (red-line). (B) *K. pneumoniae* bacteremias stratified by quarters of the year. The x-axis represents quarters of the year,
258 and the y-axis represents the normalized number of *K. pneumoniae* bacteremias with 3GC-R*Kp* (red) and 3GC-S*Kp* (blue) labelled
259 respectively. (C) Comparison of 3GC-R*Kp* and 3GC-S*Kp* bacteremia in the first half and the second half of the year. The red dots on
260 the left represent the average number 3GC-R*Kp* bacteremia for each month (*e.g.*, January, February, etc.), and the blue dots on the
261 right represent the average number of 3GC-S*Kp* bacteremias by month. Student's *t*-test *P*-values are labelled accordingly.

CG307 strains caused more 3GC-R*Kp* bacteremias during study timeframe compared to other 3GC-R*Kp* genotypes

Out of the 194 3GC-R*Kp* causing bacteremia during the study period, there were 79% (161/194) 3GC-R*Kp* isolates available (Fig. 2A) that met our study's inclusion criteria with 27 3GC-R*Kp* isolates included from a previous study [21]. One index 3GC-R*Kp* isolate had an assembly with genome size >7.5 Mbp and was excluded from the study. Further, we identified seven index 3GC-R*Kp* isolates (4%) that belonged to *K. pneumoniae* species complex taxa that were not *K. pneumoniae sensu stricto* and were thus excluded from analysis [4]. For the remainder of the study, all references to *K. pneumoniae* refer to *K. pneumoniae sensu stricto* strains. Thus, we had a total of 153 isolates with WGS data available for analysis with 126 index and 27 recurrent 3GC-R*Kp* isolates respectively (Fig. 2A). Among the 126 index isolates, there were a total of 52 unique STs detected highlighting the genomic diversity of our cohort. When grouping isolates by sharing at least five of seven loci from the Pasteur MLST scheme (*i.e.,* clonal group designations), a total of 19 unique CGs were identified (Fig. 2B). The most predominant clonal group identified from our index 3GC-R*Kp* isolates was CG307 (29%; 37/126) with all 37 isolates sharing identical ST307 schema. Interestingly, we only detected 6 index 3GC-R*Kp* isolates (5%) that belonged to the pandemic CG258 lineage with ST258 (n=2), ST11 (n=2), and ST395 (n=2) identified. Additional to CG258 and CG307 isolates, we only identified clonal groups with 5 or more index 3GC-R*Kp* isolates for CG29 (9/126; 7%), CG2947 (9/126; 7%), CG15 (8/126; 6%), and CG392 (6/126; 5%) with the remaining 40% (51/126) index 3GC-R*Kp* isolates belonging to rarely identified clonal groups (*i.e.,* <5 matching clonal group isolates detected). When stratified by full calendar years in our sampling frame (Fig. 2C), CG307 remained the most frequent CG detected per year except in 2021 when CG392 (n=4) was the most common CG detected (*N.B.,* 2016 and 2022 were excluded due to incomplete annual sampling). We observed a statistically significant positive correlation (Pearson correlation coefficient r=0.91; P-value=0.03; Fig. 2D) between annual frequency of CG307 isolates collected and total 3GC-R*Kp* infections detected

per year indicating that CG307 isolates may be the primary clonal group contributing to 3GC-R*Kp* bacteremia prevalence.

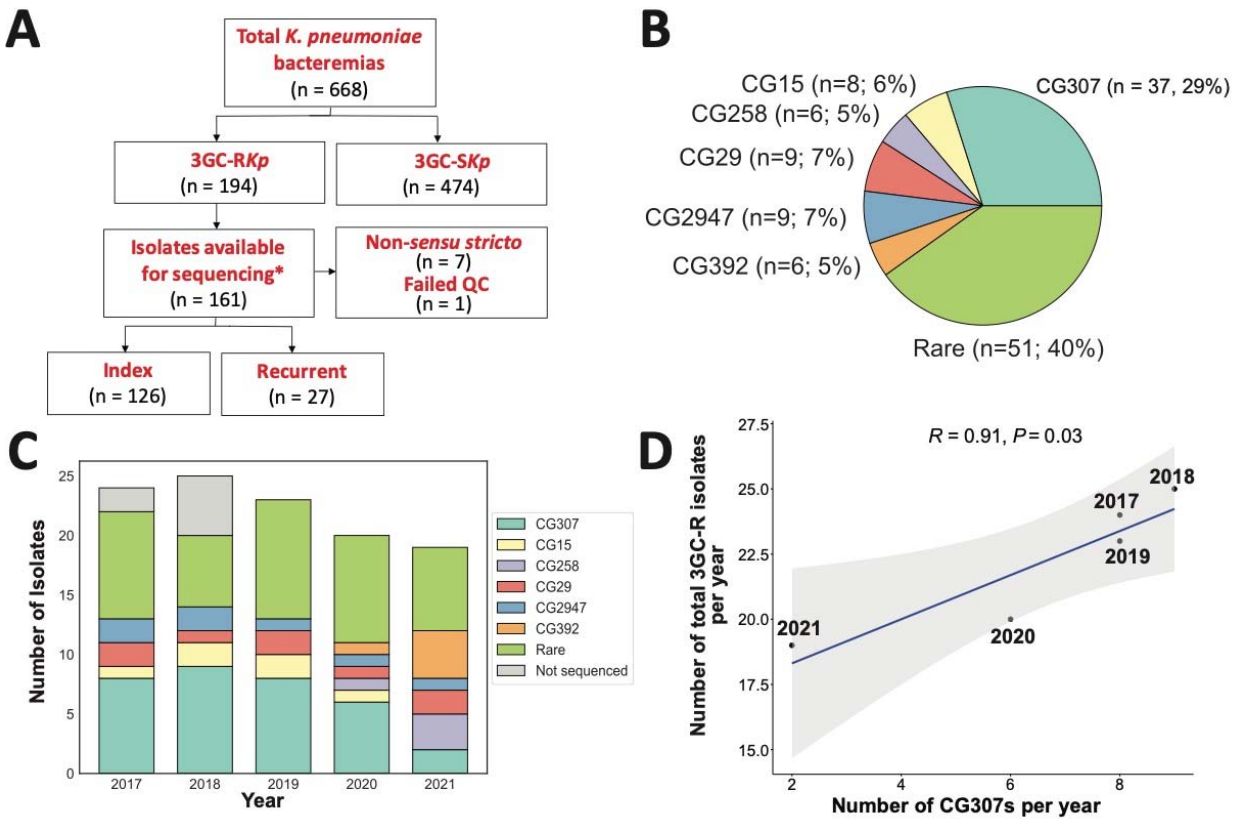


Fig. 2. Whole genome sequencing workflow, clonal group distributions, and trends of 3GC-R*Kp* isolates. (A) Workflow of WGS inclusion/exclusion criteria for *K. pneumoniae* bacteremia isolates. *27 isolates from a previous study (BioProject Accession #: PRJNA836696) are included in our analyses. (B) Pie chart showing the clonal group distribution of index 3GC-R*Kp* isolates. (C) Clonal groups of index 3GC-R*Kp* isolates stratified by year. Legend indicates clonal groups by color with gray representing index 3GC-R*Kp* isolates that were not sequenced. (D) Positive correlation between sequenced CG307 isolates per year (x-axis) and the total number of index 3GC-R*Kp* isolates from 2017-2021 (y-axis). Pearson's correlation coefficient (*r*) with associated correlation test *P*-value reported in figure.

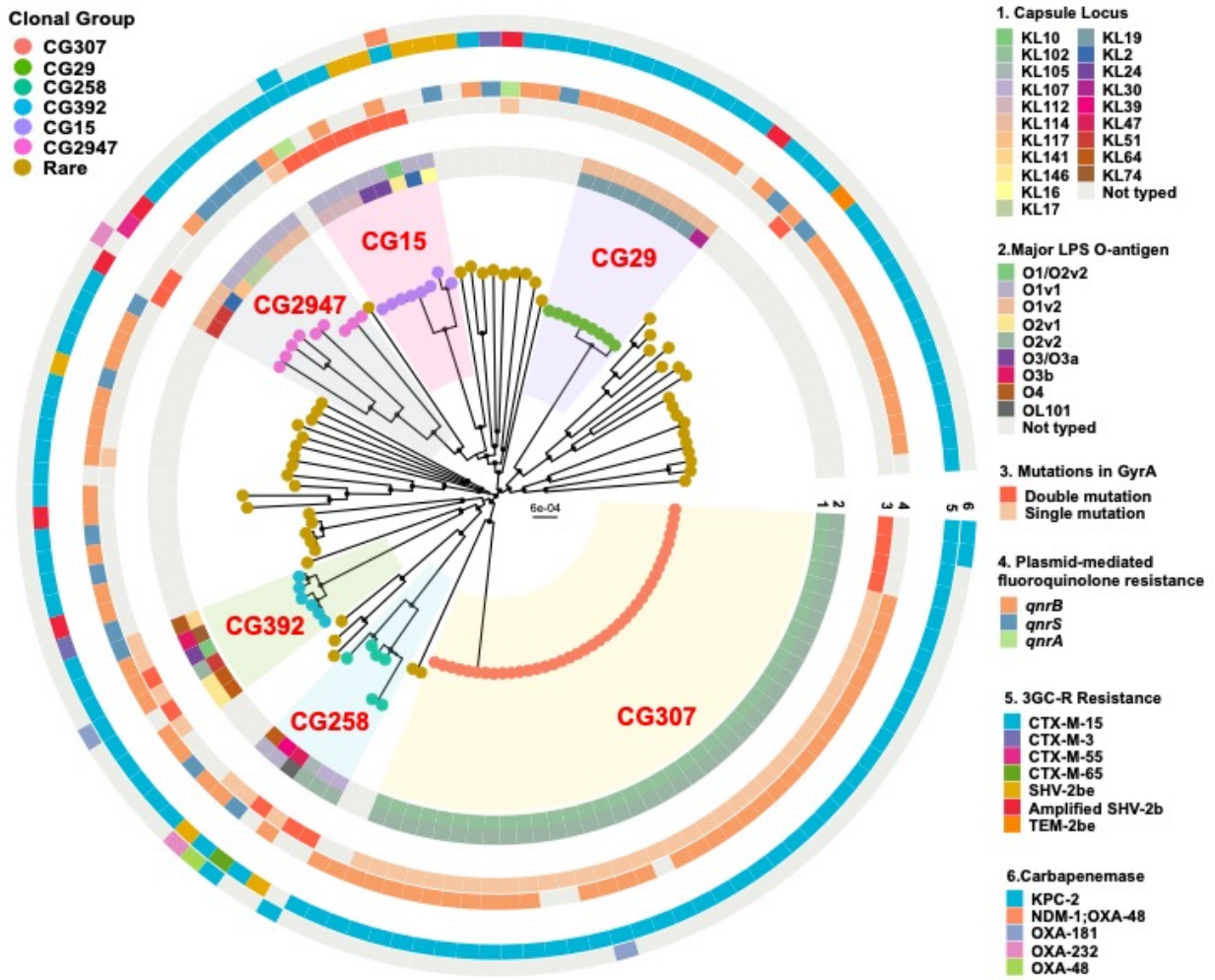
High genomic diversity within index 3GC-R*Kp* core gene population structure

To gain further insights into 3GC-R*Kp* population structure, we performed a maximum-likelihood phylogeny inferred by a core gene alignment of index 3GC-R*Kp* isolates (n=126) and identified diverse, deep-branching lineages often observed in MDR *K. pneumoniae* [4]. Using a core gene threshold definition of gene presence in ≥99% of cohort, 16% (3873/24209) of the pangenome was 'core' gene content reflecting substantial pangenome diversity. The median pairwise nucleotide divergence was 0.58% (median pairwise SNP distance [MPSD] = 21402) comparable to previous estimates [3, 20]. Although most CGs evidenced significant genetic diversity, there were clear clusters of closely related CG307 (MPSD = 84) and CG29 (MPSD = 15) isolates (Fig. 3). Both the CG307 (KL102) and CG29 (KL19) isolates had highly conserved capsule locus (KL) genes compared to the other CGs (Fig. 3). Similar to the KL locus, the O-locus was highly conserved in CG307s (O2v2), CG29 (O1v2), and CG15 (O1v1) when compared to the other CGs. When analyzing the primary determinants of hypervirulent *K. pneumoniae* (hv*KP*) [29, 43], the siderophore yersiniabactin locus (*ybt*) was present in 31% of our 3GC-R*Kp* population, consistent with the findings of Lam et al. 2018 [43]. Less commonly detected hv*KP* determinants were the genotoxin colibactin (*clb*, n=4), siderophore aerobactin (*iuc*, n=4), siderophore salmochelin (*iro*, n=1), and hypermucoidy *rmpADC* operon (n=1). One particularly interesting isolate belonged to ST412 (*i.e.,* MB3190), which is a member of hypervirulent-associated CG23, that harbored the *rmpADC* operon in addition to *iro1* and *iuc1*, often carried on the FIB_k virulence plasmid KpVP-1 [43].

When focusing on the genetic determinants associated with 3GC-R*Kp*, 94% (119/126) harbored AMR genes encoding enzymes that are associated with a 3GC-R*Kp* phenotype. The overwhelming majority of 3GC-R*Kp* isolates (n=109) harbored the ESBL encoding gene, $bla_{CTX\text{-}M}$ with 95% (104/109) harboring the $bla_{CTX\text{-}M\text{-}15}$ variant. We detected $bla_{SHV\text{-}2be}$ variants in eight strains, only two of which were identified using the NCBI AMRFinderPlus database with our

genomic assemblies; the remainder required the use of a short-read, k-mer mapping to database approach with KmerResistance to identify homologous *bla*$_{\text{SHV-2b}}$ and *bla*$_{\text{SHV-2be}}$ variants in the same strains. Additionally, CONVICT identified five strains with amplification of *bla*$_{\text{SHV-2b}}$ variants (*i.e.*, gene copy number of ≥ 2.0×), with similar amplifications of penicillinase encoding genes associated with a 'false ESBL phenotype' [44]. A carbapenemase encoding gene was observed in 11 strains (9%) (*bla*$_{\text{KPC-2}}$, n =5; *bla*$_{\text{OXA-48}}$, n=2; *bla*$_{\text{OXA-181}}$, n=2; *bla*$_{\text{OXA-232}}$, n=2) (Fig. 3). In total, a putative genetic mechanism for the 3GC-R phenotype was present in 125/126 index 3GC-R*Kp* isolates (99%).

Given the extensive use of fluoroquinolones (FQs) as prophylaxis for neutropenic patients at our institution in addition to associated FQ resistance (FQR) and ESBL producing Gram-negative organisms [45], we next analyzed fluoroquinolone susceptibility and FQR mechanisms. Ciprofloxacin non-susceptibility was observed in 73% (92/126) of index 3GC-R*Kp* isolates. Quinolone resistance determining region (QRDR) mutations in *gyrA* and *parC* were observed in 50% (64/126) and 41% (52/126) of the strains, respectively, whereas the most commonly observed FQR mechanism in our cohort was the plasmid mediated FQR gene *qnr,* primarily *qnrB1*, observed in 79% (101/126) of the strains. Single and double QRDR mutation for *gyrA* were almost exclusively observed in the major CGs such as CG307, CG258, CG392, and CG15, whereas rare CGs typically lacked *gyrA* QRDR mutations but did contain *qnr* genes (Fig. 3). The double mutation pattern *gyrA-83+87* was present in 14 strains including four CG307s (Fig. 3).



345

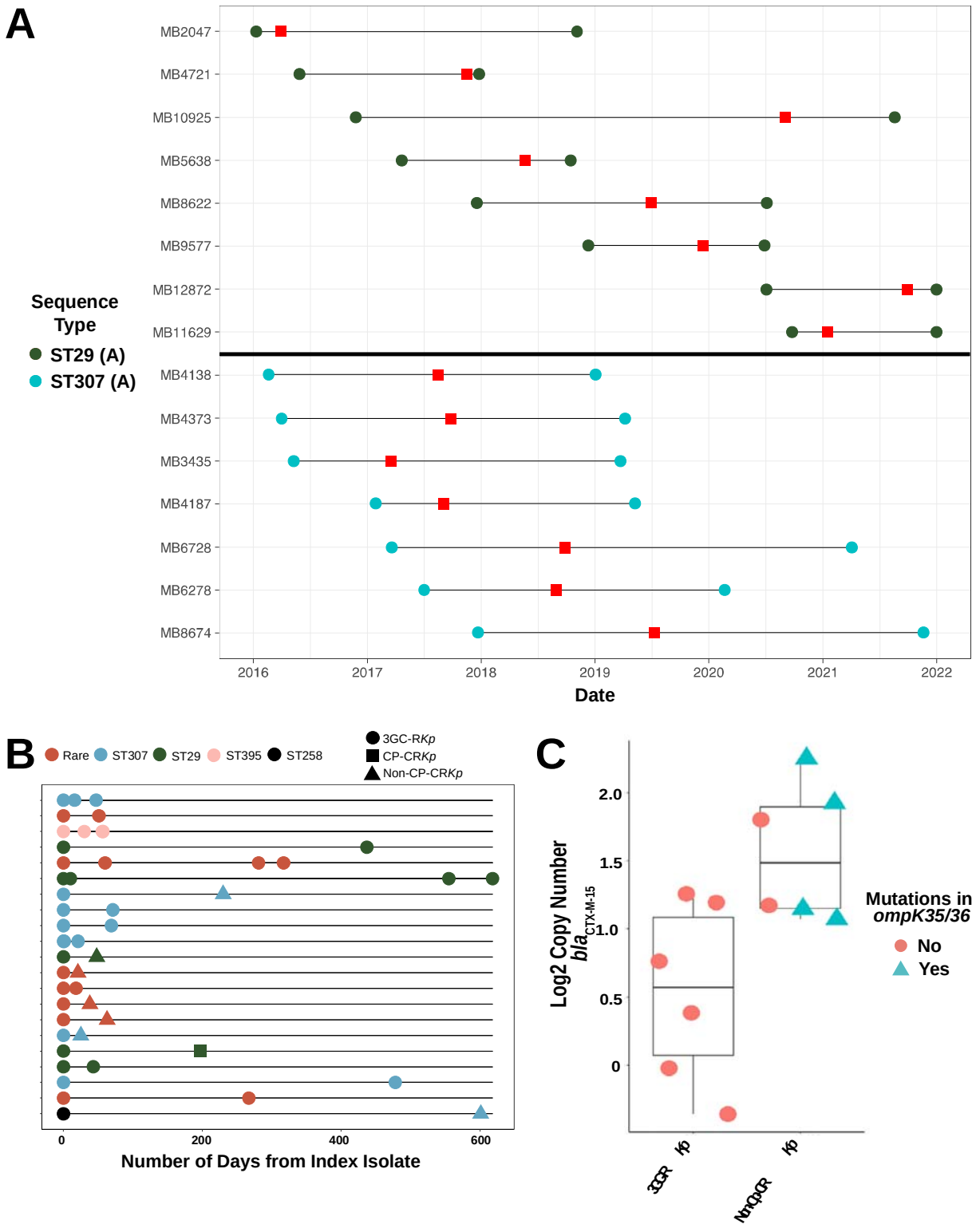
346 **Fig. 3.** Core gene maximum-likelihood inferred phylogenetic analysis of index 3GC-R*Kp* strains.
347 Phylogenetic tree is midpoint rooted with circular internal nodes representing 95% UFBoot
348 confidence values. Branch tips are labelled by clonal group with each CG highlighted within tree
349 structure. Rings are labelled as follows: 1) Capsule type; 2) O-antigen; 3) GyrA polymorphisms;
350 4) Plasmid-mediated fluoroquinolone resistance genes; 5) 3GC-R resistance genes; and 6)
351 Carbapenemase encoding genes.

3GC-R*Kp* transmission and recurrence dynamics driven primarily by CG307 and CG29 clades

Given the known capacity of 3GC-R*Kp* to be transmitted in healthcare settings [2, 46], we used PopPUNK to identify nested populations of 3GC-R*Kp* and subsequently performed PopPUNK cluster specific, core genome alignment inferred phylogenies masked for recombination to potentially identify transmission networks. We were able to identify 56 PopPUNK groups of which 14 included two or more isolates (*i.e.*, 42 isolates were uniquely divergent from the full cohort, consistent with our clonal group population structure). From these 14 PopPUNK groups, we identified 28 3GC-R*Kp* isolates from unique patients that differed by <25 pairwise SNPs (*i.e.,* a SNP threshold that has been previously used to define potential *K. pneumoniae* transmission [47]) with a minumum of one other unique patient 3GC-R*Kp* (Table S3). Potential strain transmission cluster sizes varied from 2-8 isolates with a median pairwise SNP distance of 8 and range of 0-24 SNPs. Putative transmission clusters with genetically related pairwise isolates involved four distinct STs, namely ST307 (five clusters, 16 total isolates), ST29 (one cluster, eight isolates), ST152 (one cluster, 2 isolates), and ST280 (one cluster, 2 isolates). Interestingly, 97% (27/28) of index isolates from putative transmission clusters caused infections in patients with hematologic maligancy compared to 66% (65/98) non-clustered strains (χ^2 *P*-value < 0.001). The two largest transmission clusters (ST307 [pp8-1], n = 7; ST29, n = 8 [pp65-1]) both occurred over prolonged periods of time, 28 months for ST307 pp8-1 cluster and nearly 60 months for the ST29 pp65-1 cluster in patients with multiple admissions and extensive contact with the healthcare system (Fig. 4A).

During the study timeframe, 12% (21/126) of index *b*acteremia patients had a recurrent 3GC-R*Kp* bacteremia occurring an average of 94 days following the initial 3GC-R*Kp* bacteremia episode (Fig. 4B). The same ST as the index isolate caused 91% of the recurrent infections. Consistent with carbapenems being the primary treatment for 3GC-R*Kp*, 38% (8/21) of isolates

causing recurrent infections developed carbapenem resistance (Fig. 4B). The majority of strains (n=7) developed carbapenem resistance in the absence of a carbapenemase gene (*i.e.*, were non-carbapenemase producing carbapenem resistant *K. pneumoniae* or non-CP-CR*Kp*), whereas a single ST29 strain acquired bla_{KPC-2} (Fig. 4BB). Through a combination of ONT long-read sequencing in addition to CONVICT gene copy number estimates, we found that six out of the seven non-CP-CRE recurrent isolates had evidence of $bla_{CTX-M-15}$ gene amplification (Fig. 4C). Additionally, mutations in *ompK*35 (n=1) and *ompK*36 (n=4) were observed in 5/7 (71%) of the non-CP-CR*Kp* isolates (Fig. 4C). These data indicate that recurrence of 3GC-R*Kp* bacteremia primarily occurs due to re-infection by the same strain which has often developed carbapenem resistance through non-carbapenemase mechanisms.



387

388 **Fig. 4.** 3GC-R*Kp* transmission and recurrence dynamics. (A) Genetically related (*i.e.*, <25

389 pairwise SNPs) isolates collected over time from unique patients for CG29 (top; dark green;

transmission cluster pp8-1) and CG307 (bottom; light blue; transmission cluster pp65-1) clades. Circles represent start and end of healthcare interactions at our institution for each respective patient with horizontal black line representing the duration of healthcare services. Red rectangles indicate collection date for each respective unique isolate. Note that patient-sample pair for MB12872 and MB11629 had end dates that went past 2021/12/31. (B) Index 3GC-R*Kp* isolates with paired recurrent (*i.e.*, *K. pneumoniae* bacteremia isolate collected >14 days from antecedent index isolate). Duration of days from index to recurrent isolate is shown on the x-axis. Colors denote sequence type and shape indicates phenotypes as shown in legend. 3GC-R*Kp* = third-generation cephalosporin resistant *K. pneumoniae*; CP-CR*Kp* = carbapenemase producing carbapenem-resistant *K. pneumoniae*; Non-CP-CR*Kp* = non-carbapenemase producing carbapenem-resistant *K. pneumoniae* (C) $bla_{\text{CTX-M}}$ copy numbers and *ompK35/ompK36* mutation status in paired 3GC-R*Kp* (left) and recurrent Non-CP-CR*Kp* (right) isolates. Y-axis shows log2 transformed $bla_{\text{CTX-M-15}}$ copy number estimates. Shapes and colors show *ompK35/ompK36* status as delineated in the legend.

Identification of a novel lineage of CG307 causing bacteremia at our institution

Given the stable detection of CG307 within the Houston region [7, 9, 20], including previously at our own institution [21], we further analyzed CG307 isolates from our cohort (n=37) along with a worldwide distribution of geographically diverse CG307 isolates (n=187) publicly available on NCBI. A Bayesian dated, recombination masked, phylogeny of 224 CG307s inferred from a core genome SNP alignment is presented on Fig. 5. Initial root-to-tip analysis of our CG307 cohort indicated a strong temporal signal (Fig. S1). Consistent with previous data, the CG307 splits into two primary clades consistent with the previously described "Texas-specific" (n =74) and "global" clades (n = 150). Within these respective CG307 clades, there were two nested sub-populations within each clade with clusters 3 and 4 belonging to the Texas-specific lineage and clusters 1 and 2 as part of the Global lineage [10, 20]. The Texas-specific clade harbored double GyrA-83I-87N mutations and 2× copies of $bla_{\text{CTX-M-15}}$ that have previously been characterized (Fig. 5) [10]. Interestingly, only four CG307 isolates in our cohort clustered with the Texas-specific clade. In contrast, the majority (n=33) of our CG307 isolates belonged to the global clade with an approximate even split between nested populations (18 and 15 isolates in clusters 1 and 2 respectively).

Our Bayesian dating analysis of our phylogeny estimates CG307 emerged in 1995 (95% confidence interval [95%CI]: 1991 – 1997), consistent with a previous estimation of CG307 emergence in 1994 [10]. The date of divergence between the Texas-specific and global CG307 clades is estimated to have occurred in 1997 (95%CI: 1995 – 2000) shortly after CG307 had emerged, suggesting that CG307 may have originated from the Houston region as previously hypothesized [10]. When focusing on the global clade, cluster 1 strains were estimated to emerge from cluster 2 strains approximately 1999 (95%CI: 1998-2001) with cluster 1 isolates beginning to clonally expand near 2008 (95%CI: 2006-2011) (Fig. 5). A conserved genomic feature of cluster 1 CG307 strains was the stable vertical transmission of $bla_{\text{SHV-205}}$, a single

amino acid variant of the chromosomal *bla*$_{SHV-1}$ gene (Fig. 5). Consistent with our estimation of cluster 1 strains emerging near 2008, all publicly available cluster 1 strains with metadata were isolated from 2016 onwards with 15/25 isolated since 2020. All but one strain (Paraguay 2020) came from the United States including such geographically diverse locales as Utah, Chicago, and New York and even included two strains isolated from dogs in Texas. Given that MDA has a broad geographic patient distribution, we analyzed the home location for each patient with a CG307 infection at our institution and found that regardless of the cluster of infecting isolates, most patients (31/37, 82%) were from Texas and the majority were from the Houston area (54%) (Fig. S1A). There were three patients from the United States who resided outside Texas all of whom were infected by strains of the global lineage as were the three patients from outside the United States (Fig. S1B). These data indicate that the global lineage of CG307 caused the majority of CG307 infections even among our patients from Houston and the rest of Texas.

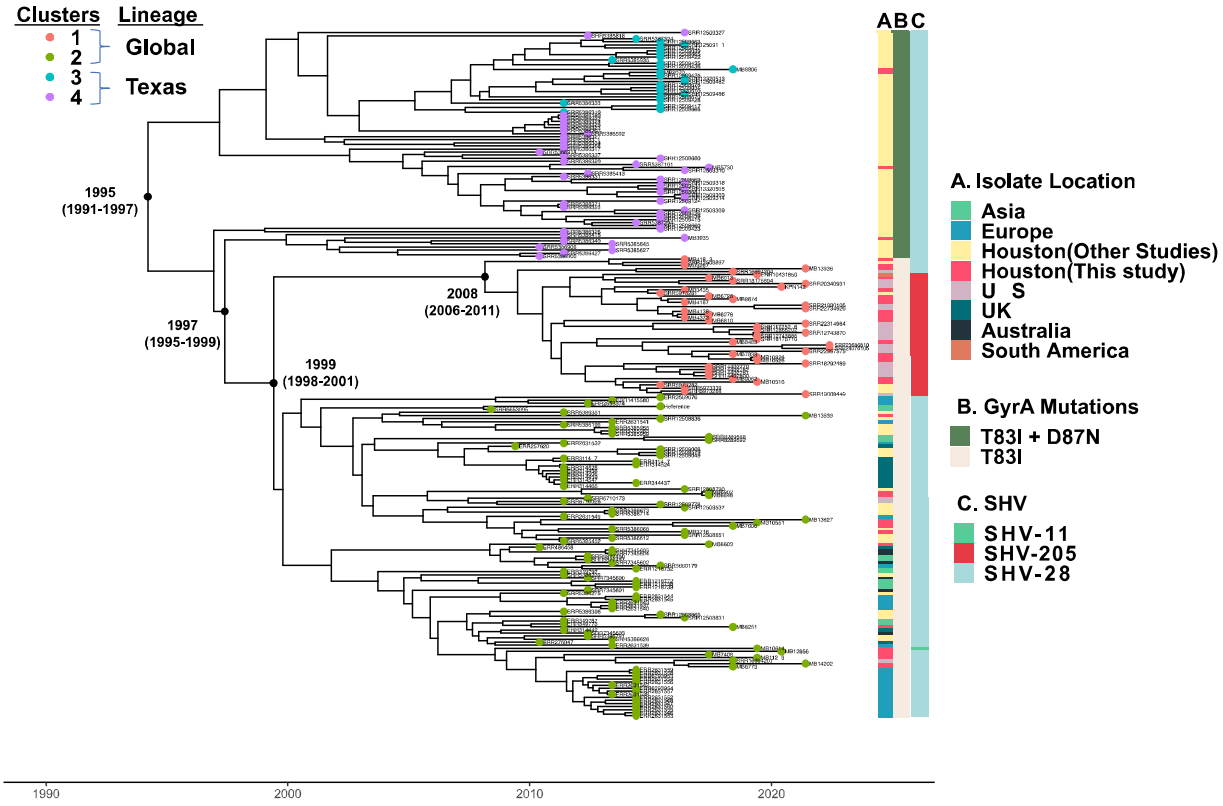
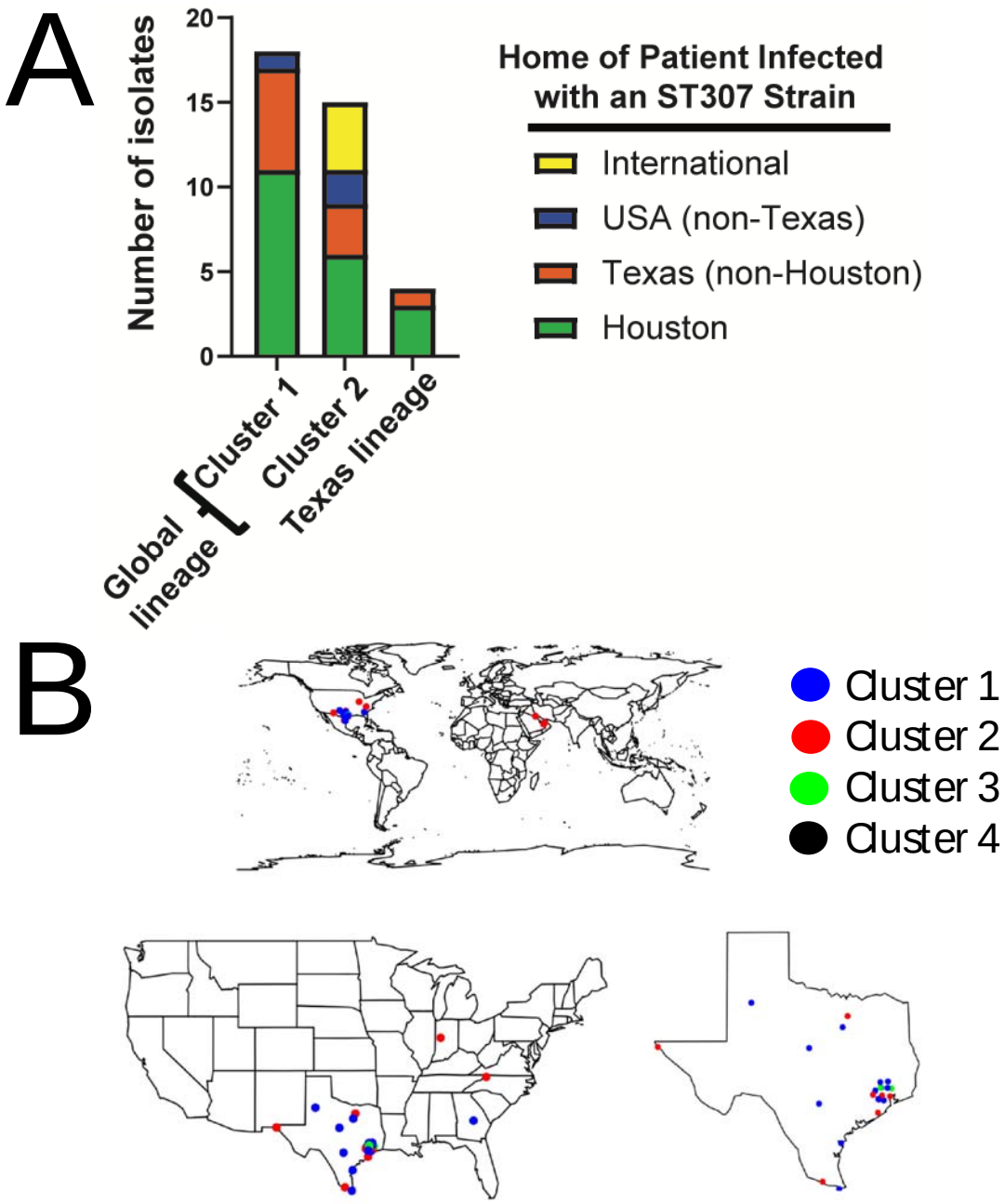


Fig. 5. Identification of unique CG307 sub-clade (*i.e.,* cluster 1) causing bacteremia at our institution. Presented is a Bayesian dated, core SNP recombination masked, phylogeny of 224 CG307 strains using isolate Kp616 (BioSample Accession Number: SAMN08391417) as reference. Tip colors indicate hierBAPS predicted taxa with clusters 1 and 2 belonging to the previously identified global clade and clusters 3 and 4 being part of the Texas-specific clade. Years indicate predicted dates of divergence from internal nodes of interest as estimated using BactDating. Dates in parentheticals represent the 95% confidence interval. Metadata presented on right columns include A) geographic location of isolate collection; B) presence of *gyrA* mutations; and C) chromosomal bla_{SHV} isoforms.



452

453 **Supplemental Fig. 2.** Geographic location of patient origin in which CG307 strains caused
454 bacteremia at our institution. (A) X-axis stratifies isolates by clusters identified using hierarchical
455 clustering. Colors of patient origin as shown in legend. (B) Geographical distribution from a
456 world, country, and state perspective. Isolates are colored by cluster as indicated in legend.

Differing accessory genome content across each of the CG307 clades

To gain further insight into the various CG307 clusters, we used our Illumina data in conjunction with ONT long-read sequencing to dissect accessory genome content and determine genomic context of 3GC-R*Kp* determinants. We subset all accessory genome content shared in greater than 5% but less than 95% of the CG307 population to look for gene presence/absence signals present across our four clades and observed a demarcation across the four clusters (Fig. 6A). We noted that there were global and Texas-specific clade differences in replicon type detection, which has been described in previous literature (Fig. 6B) [20]. The global clade had an enrichment of IncFII (84%) and $IncFIB_K$ (97%) whereas there were very few cluster 3 isolates with either replicon detected (IncFII [36%]; $IncFIB_K$ [20%]) (Fig. 6B). These multireplicon F-type plasmid differences correlates with $bla_{CTX\text{-}M\text{-}15}$ carriage differences between global and Texas-specific clades as previously described [10, 20]. As shown in Fig. S3A, both MB5730 (Cluster 4) and MB8806 (Cluster 3) from the Texas-specific clade harbored 2 copies of $bla_{CTX\text{-}M\text{-}15}$ inserted in the chromosome. Conversely, isolates MB7606 (Cluster 2) and MB8674 (Cluster 1) from the global clade harbored $bla_{CTX\text{-}M\text{-}15}$ in association with IS*26* pseudo-compound transposons made of two or more IS*26* units with directly flanking IS*26* transposases on $IncFIB_K$ plasmids (Fig. S3B).

We focused on accessory genome differences between the global cluster 1 and cluster 2 isolates to determine what was driving the potential emergence of cluster 1 in our region. We found that cluster 1 strains had a very high recombination to mutation (r/m) ratio of 16 suggesting that recombination events were driving diversity in this clade. We found two recombination blocks in cluster 1 isolates that appeared to be regions that differed with cluster 2 isolates (Fig. 6C). First, we found that, relative to cluster 2 strains, cluster 1 strains uniquely contain a 40 Kbp prophage-like region which contains numerous putative regulatory proteins as well as proteins of unknown function (Fig. 6C). Second, although both cluster 1 and cluster 2 strains contain a type 6 secretion system (T6SS), cluster 1 strains lack a set of T6SS

effector/immunity proteins which are present in cluster 2 strains (Fig. 6D). Taken together, we conclude that the cluster 1 strains identified in our study are a distinct genetic branch of the global CG307 clade that have unique accessory gene content as well as are causing disease throughout the United States.

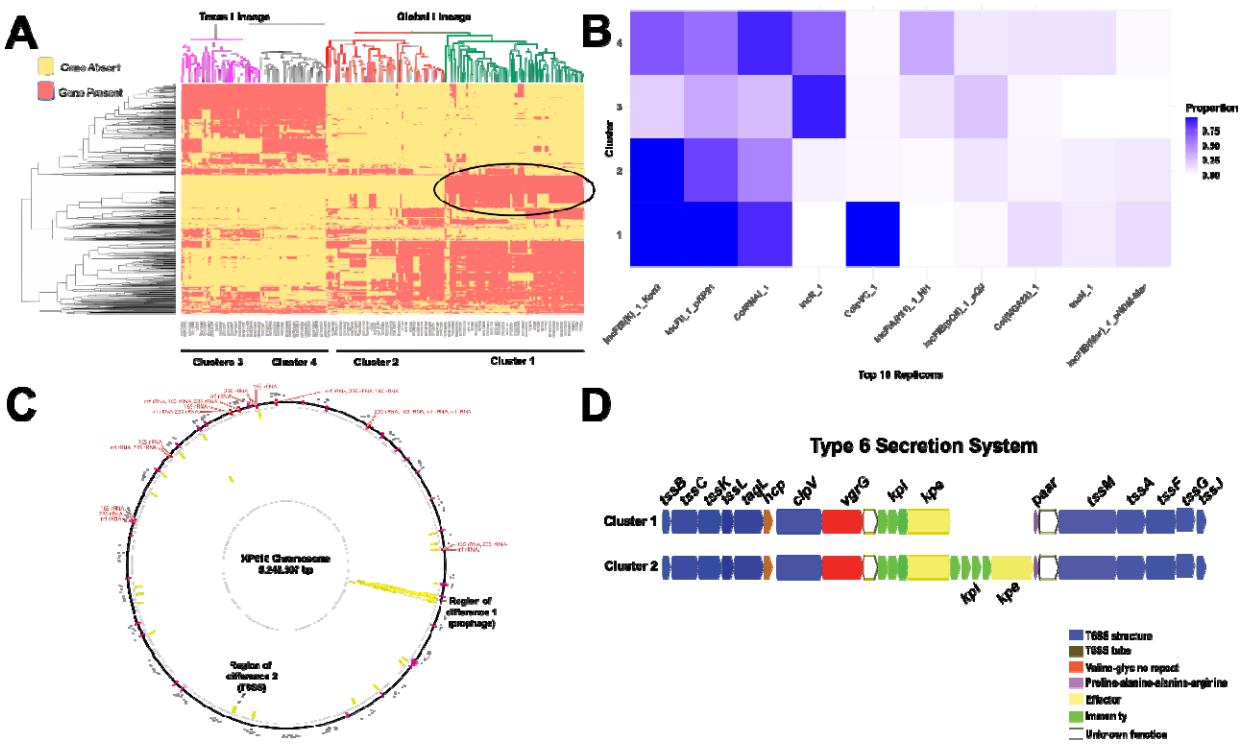
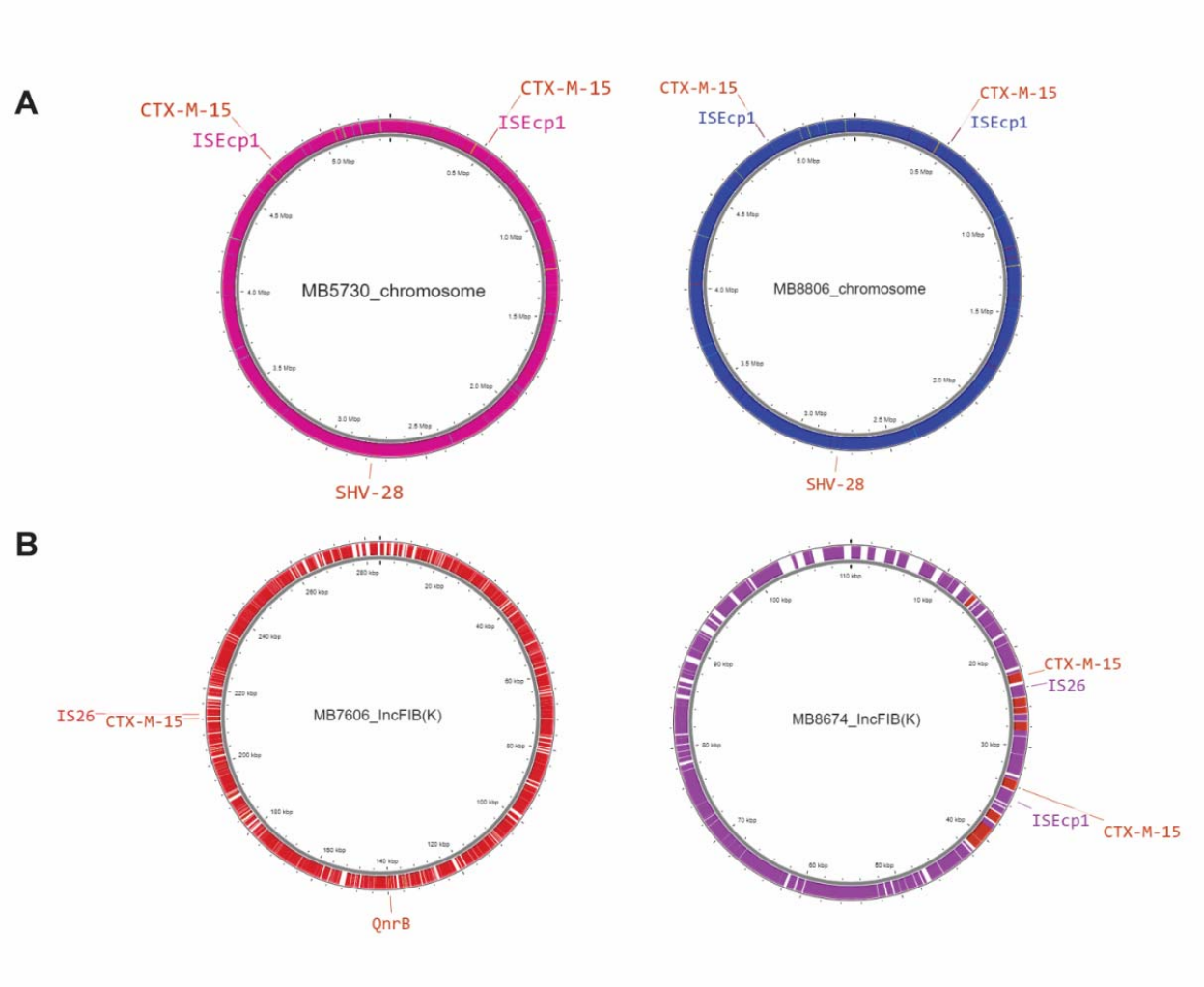


Fig. 6. Assessment of accessory genome differences between CG307 strains. (A) Gene presence/absence heatmap representing the agglomerative hierarchical clustering of accessory genomes of CG307 strains. Circled region indicates genetic content unique to cluster 1 isolates. (B) Proportion heatmap of the most common replicon types identified using PlasmidFinder stratified by CG307 cluster. (C) Recombination hotspots (yellow blocks) of CG307 isolates using Kp616 as a reference and Gubbins predicted recombination sites output. Regions labelled on chromosome indicate genomic context we identified that differed between cluster 1 and 2. (D) Region of difference 2 between CG307 cluster 1 and cluster 2 strains involving a type 6 secretion system (T6SS). Labeling system for T6SS proteins is taken from a previous study [48] with description of color of putative gene function shown in the legend. Note that cluster 1 strains lack of a set of immunity/effector proteins present in cluster 2 strains.



498

499 **Supplemental Fig. 3.** Analysis of CG307 bla$_{CTX\text{-}M\text{-}15}$ copy number and genomic context from
500 MDA 3GC-R*Kp* isolates. (A) CG307 Texas-specific clade isolates (*i.e.,* cluster 3 and cluster 4)
501 that have two copies of *bla*$_{CTX\text{-}M\text{-}15}$ present on the chromosome in association with IS*Ecp1* as
502 previously described [49]. (B) CG307 global clade isolates (*i.e.,* cluster 1 and cluster 2) that
503 have *bla*$_{CTX\text{-}M\text{-}15}$ present in one or more copies on F-type multireplicon plasmids.

DISCUSSION

K. pneumoniae is a major contributor to healthcare associated infections and can acquire resistance to a diverse array of antimicrobials thereby rendering treatment problematic [1, 50, 51]. Given the limited molecular epidemiology of 3rd generation cephalosporin resistant *K. pneumoniae* (3GC-R*Kp*) in the United States, we characterized 126 index and 27 recurrent 3GC-R*Kp* bacteremia isolates collected from 2016 to 2022 at our institution (Fig. 2A). Consistent with other worldwide MDR *K. pneumoniae* surveillance studies, we observed long-branching, genetically diverse 3GC-R*Kp* isolates [22, 23, 47, 52-55]; however, there were clusters of "global problem clones" including CG307, CG29, and CG15 isolates with evidence of limited transmission [4]. Moreover, through analysis of our patients and publicly available genomes, we have identified a previously uncharacterized *K. pneumoniae* CG307 sub-clade (*i.e.,* cluster 1) that caused 14% of our index 3GC-R*Kp* infections and has been isolated recently from multiple locations across the United States [26, 55].

A major impetus for our study were reports of a recent expansion of 3GC-R*Kp* in the United States [12, 56]; however, prospective data from active 3GC-R*Kp* surveillance is limited, and thus increases in incident 3GC-R*Kp* infections remain unclear [57]. During our study timeframe, we observed fairly stable absolute frequencies as well as admission adjusted prevalence of 3GC-R*Kp* bacteremias (Fig 1). It has been suggested that increases in 3GC-R *Enterobacterales* infections in the United States have been largely driven by community-onset cases [50]. The genetic diversity of 3GC-R*Kp* isolates (Fig. 3) in conjunction with seasonal prevalence differences of 3GC-R*Kp* infections (Figs. 1B, 1C) suggest a possible community-wide dissemination in our region. These data are similar to those recently described for a hospital network in Australia in which comprehensive WGS of *K. pneumoniae* found that the majority of infections were caused by widespread circulation in the community rather than hospital transmission [47]. Consistent with our observation 3GC-R*Kp* infections tended to occur in the

warmer second half of the calendar year compared to the first half of the year, a multi-site surveillance study across several continents found that environmental factors associated with warmer months may increase incidence rates of *K. pneumoniae* infections [58]. Nevertheless, similar to a recent analysis on 3GC-R *E. coli* prevalence at our institution [42], we observed a statistically significant association only for 3GC-R*Kp* but not 3GC-S*Kp* infections. Thus, our results taken together with previous studies suggest that limiting 3GC-R*Kp* acquisition outside of the hospital is likely to be critical to mitigation efforts.

Although we had a long-branching, genetically diverse cohort of 3GC-R*Kp* isolates (Figs. 2, 3), we also observed numerous instances of potential hospital-based 3GC-R*Kp* transmission (Fig. 4A). It is worth noting that given we only sampled 3GC-R*Kp* isolates collected from bacteremia infections, we are likely underestimating the true scope of transmission that could be occurring at our institution. Interestingly, transmission clusters almost exclusively were due to the "problem clones" CG307 and CG29 (Fig. 4A) and involved patients with hematologic malignancy, likely because of the propensity of such individuals to develop bacteremias following initial colonization [59]. Outbreaks of these two "global problem clones" have been described [60] although whether such transmission is happening due to direct human contact or being acquired from the hospital environment is not currently known. Interestingly, for both CG307 and CG29 strains, we observed highly genetically related bacteria causing infections over a several years' time-frame (Fig. 4A), suggesting potential environmental sources as has been observed for *Pseudomonas* spp. [61] and vancomycin resistant *Enterococcus faecium* [62]. The lack of clear epidemiologic links between infected patients means that standard infection control approaches would be unlikely to detect such transmission events emphasizing the potential complementary role that WGS could play in infection control efforts if real-time data were available [63].

Recurrence is a major concern following treatment of 3GC-R*Kp* infections [64, 65], and indeed we observed an approximate 20% recurrence rate when just considering bacteremias. The recurrent strain was nearly always the same ST and highly genetically related to the index isolate suggesting re-infection rather than acquisition of a new 3GC-R*Kp* strain. Interestingly, the recurrent organism could either retain a 3GC-R, carbapenem-susceptible phenotype or transition to carbapenem resistance, usually without acquisition of a carbapenemase (Fig. 4B). Although data are limited, a few studies have found that intravenous carbapenem administration does not reliably eradicate ESBL positive *Enterobacterales* from the intestinal tract, presumably due to low carbapenem stool concentrations [66, 67]. Alternatively, it has been recently demonstrated that *Enterobacter* spp. can enter into a cell-well deficient spheroplast capable of surviving carbapenem exposure and returning to normal growth once carbapenem pressure is removed, and thus carbapenem "tolerance" could have accounted for strain persistence [68]. 3GC-R*Kp* isolates can also develop outright resistance to carbapenems in the absence of carbapenemase by limiting drug influx through porin mutations along with hyperproduction of $bla_{\text{CTX-M}}$ encoding enzymes, which we were able to demonstrate in multiple serial isolates using a long-read sequencing approach [21]. Remarkably, we collected many recurrent isolates months or years from the index isolate collection date (Fig. 4B) with minimal pairwise SNP differences (*i.e.,* < 10 SNPs), suggesting the long duration of colonization achievable by 3GC-R*Kp* [69]. We envision that a strategy of targeted *K. pneumoniae* gastrointestinal tract decolonization, such as using CRISPR-based technologies, could be part of the future treatment of 3GC-R*Kp* infections [70].

We were intrigued by the paucity of Texas-specific CG307 isolates in contrast to the Global CG307 clade collected in our cohort in addition to the relatively low number of CG258 isolates detected (Fig. 2). A study from a Houston-wide hospital system analyzing ESBL *K. pneumoniae* strains collected from 2011-2015 found similar numbers of CG307 and CG258 strains, an

unexpected result given the predominant prevalence of CG258 in most United States investigations [5, 9, 19]. A comparative genomics analysis of CG307 isolates found that the Texas-specific clade was genetically distinct from other CG307 strains circulating worldwide, with multiple chromosomal insertions of *bla*$_{\text{CTX-M-15}}$ and a fixed double GyrA mutation as a predominant genomic feature of the Texas-specific CG307 clade [10]. Another Houston based study found divergent accessory genomes between CG258 and CG307 with greater plasmid content in the latter, suggesting that CG307 isolates had potential greater capacity to adapt to selective pressures [20]. Nevertheless, this study found very few of the global CG307 clade circulating in their particular hospital system [20]. The reason for our finding that only a small fraction of CG307 isolates belong to the Texas lineage is not clear but might be due to temporal variation given our most recent sampling timeframe. Support for this hypothesis comes from the relatively recent isolation of CG307 strains from across multiple United States locales including Central Texas [26] and Mississippi that cluster together with the majority of our Global CG307 isolates. These isolates have a distinct chromosomal SHV, lack the multiple copies of chromosomal *bla*$_{\text{CTX-M-15}}$ hypothesized to be important to the success of the Texas-specific lineage, and contain a unique accessory genome relative to other CG307 strains (Figs. 5, 6). Our findings mirror similar, recent clonal expansions of CG307 strains from the global lineage recently observed in Wales and Norway, suggesting that the United States may be in the early stages of dissemination of a CG307 Global sub-clade [22, 23].

We identified two major events distinguishing the Global CG307 cluster 1 and cluster 2 isolates. The cluster 1 strains contained a prophage that did not contain any clear virulence factors, although numerous open reading frames encoded proteins of unknown function and acquisition of exogenous DNA has been shown to influence the transcriptome of distant genes possibly through the presence of transcriptional regulators in the acquired DNA [71, 72]. Additionally, relative to cluster 1, cluster 2 strains had a distinct set of effector/immunity proteins from a type

6 secretion system (Fig. 6D). T6SS function as injection machinery that can directly puncture both eukaryotic and prokaryotic membranes and have been identified as contributing to *K. pneumoniae* pulmonary infection, gastrointestinal colonization, bacterial competition, and liver abscess formation [73-76]. Future work will be necessary to further functionally characterize potential differences in virulence and resistance factors contributing to each CG307 clades' respective success.

Although our study had many strengths, there are some limitations. First, our isolates were recovered from a single center and thus the generalizability of our findings are not known; however, our cancer center draws from a large geographic region, and Houston has been a notable center of 3GC-R*Kp* infections [9, 20]. Second, we only analyzed bloodstream infections, which limits our results' generalizability when it comes to other 3GC-R*Kp* infectious sources; nevertheless, given the invasive nature of bacteremia, this reduces likelihood of identifying false positive infections. Finally, the genetic diversity of 3GC-R*Kp* isolates meant that we could only analyze in-depth a few clonal groups even with a sufficient large sample size of over 150 3GC-R*Kp* samples sequenced.

In summary, we present a contemporary, comparative genomics-based analysis of 3GC-R*Kp* causing invasive disease in immunocompromised patients. We found that a previously unrecognized CG307 sub-clade is causing significant disease both in Houston and in diverse locations across the United States suggesting ongoing spread of an MDR *K. pneumoniae* strain with unique accessory genome content. Our findings emphasize the continued need for 3GC-R *Enterobacterales* surveillance in addition to the efforts made to track carbapenem resistant strains.

ACKNOWLEDGEMENTS

We would like to thank the MDACC clinical microbiology lab for all their hard work in identifying, handling, and transferring these pathogenic strains of interest to us for our research projects. Core grant CA016672(ATGC) and NIH 1S10OD024977-01 grant provide funding for the Advanced Technology Genomics Core (ATGC) sequencing facility at MD Anderson Cancer Center. SSA was supported by Dell Family Fund for the School of Health Professional scholarship and CTW is supported by Peter and Cynthia Hu scholarship. WCS is supported through the National Institute of Allergy and Infectious Diseases (NIAID) T32 AI141349 Training Program in Antimicrobial Resistance. Support for this study was also provided by the NIAID R21AI151536 and P01AI152999 for S.A.S. The authors acknowledge the support of the High-Performance Computing for research facility at the University of Texas MD Anderson Cancer Center for providing computational resources that have contributed to the research results reported in this paper.

REFERENCES

1. Murray CJ, Ikuta KS, Sharara F, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet **2022**; 399:629-55.
2. Mohd Asri NA, Ahmad S, Mohamud R, et al. Global Prevalence of Nosocomial Multidrug-Resistant Klebsiella pneumoniae: A Systematic Review and Meta-Analysis. Antibiotics **2021**; 10:1508.
3. Holt KE, Wertheim H, Zadoks RN, 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 U S A **2015**; 112:E3574-81.
4. Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. Nat Rev Microbiol **2020**; 18:344-59.
5. Kitchel B, Rasheed JK, Patel JB, et al. Molecular epidemiology of KPC-producing Klebsiella pneumoniae isolates in the United States: clonal expansion of multilocus sequence type 258. Antimicrob Agents Chemother **2009**; 53:3365-70.
6. Bowers JR, Kitchel B, Driebe EM, et al. Genomic Analysis of the Emergence and Rapid Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic. PLoS One **2015**; 10:e0133727.
7. Castanheira M, Farrell SE, Wanger A, Rolston KV, Jones RN, Mendes RE. Rapid expansion of KPC-2-producing Klebsiella pneumoniae isolates in two Texas hospitals due to clonal spread of ST258 and ST307 lineages. Microbial Drug Resistance **2013**; 19:295-7.
8. Villa L, Feudi C, Fortini D, et al. Diversity, virulence, and antimicrobial resistance of the KPC-producing Klebsiella pneumoniae ST307 clone. Microb Genom **2017**; 3:e000110.

9. Long SW, Olsen, R. J., Eagar, T. N., Beres, S. B., Zhao, P., Davis, J. J., ... & Musser, J. M. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing Klebsiella pneumoniae isolates, Houston, Texas: unexpected abundance of clonal group 307. MBio **2017**; 8:e00489-17.
10. Wyres KL, Hawkey J, Hetland MAK, et al. Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae strain ST307. J Antimicrob Chemother **2018**; 74:577-81.
11. Queenan AM, Bush K. Carbapenemases: the Versatile β-Lactamases. Clinical Microbiology Reviews **2007**; 20:440-58.
12. Castanheira M, Kimbrough JH, DeVries S, Mendes RE, Sader HS. Trends of beta-Lactamase Occurrence Among Escherichia coli and Klebsiella pneumoniae in United States Hospitals During a 5-Year Period and Activity of Antimicrobial Agents Against Isolates Stratified by beta-Lactamase Type. Open Forum Infect Dis **2023**; 10:ofad038.
13. Rojas LJ, Weinstock GM, De La Cadena E, et al. An Analysis of the Epidemic of Klebsiella pneumoniae Carbapenemase-Producing K. pneumoniae: Convergence of Two Evolutionary Mechanisms Creates the "Perfect Storm". J Infect Dis **2017**; 217:82-92.
14. Wang M, Earley M, Chen L, et al. Clinical outcomes and bacterial characteristics of carbapenem-resistant Klebsiella pneumoniae complex among patients from different global regions (CRACKLE-2): a prospective, multicentre, cohort study. The Lancet Infectious Diseases **2021**; 22:401-12.
15. David S, Reuter S, Harris SR, et al. Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. Nat Microbiol **2019**; 4:1919-29.
16. Cerqueira GC, Earl AM, Ernst CM, et al. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A **2017**; 114:1135-40.
17. Satlin MJ, Chen L, Patel G, et al. Multicenter Clinical and Molecular Epidemiological Analysis of Bacteremia Due to Carbapenem-Resistant Enterobacteriaceae (CRE) in the CRE Epicenter of the United States. Antimicrobial Agents and Chemotherapy **2017**; 61:AAC.02349-16.
18. van Duin D, Arias CA, Komarow L, et al. Molecular and clinical epidemiology of carbapenem-resistant Enterobacterales in the USA (CRACKLE-2): a prospective cohort study. The Lancet Infectious Diseases **2020**; 20:731-41.
19. Deleo FR, Chen L, Porcella SF, et al. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae. Proceedings of the National Academy of Sciences **2014**; 111:4988-93.
20. Shropshire WC, Dinh AQ, Earley M, et al. Accessory genomes drive independent spread of carbapenem-resistant Klebsiella pneumoniae clonal groups 258 and 307 in Houston, TX. MBio **2022**; 13:e00497-22.
21. Shropshire W, Konovalova A, McDaneld P, et al. Systematic Analysis of Mobile Genetic Elements Mediating β-Lactamase Gene Amplification in Noncarbapenemase-Producing Carbapenem-Resistant Enterobacterales Bloodstream Infections. Msystems **2022**; 7:e00476-22.
22. Fostervold A, Hetland MAK, Bakksjo R, et al. A nationwide genomic study of clinical Klebsiella pneumoniae in Norway 2001-15: introduction and spread of ESBLs facilitated by clonal groups CG15 and CG307. J Antimicrob Chemother **2022**; 77:665-74.
23. David S, Mentasti M, Sands K, et al. Genomic surveillance of multidrug-resistant Klebsiella in Wales reveals persistent spread of Klebsiella pneumoniae ST307 and adaptive evolution of pOXA-48-like plasmids. Microbial Genomics **2023**; 9.
24. Heiden SE, Hübner N-O, Bohnert JA, et al. A Klebsiella pneumoniae ST307 outbreak clone from Germany demonstrates features of extensive drug resistance, hypermucoviscosity, and enhanced iron acquisition. Genome Medicine **2020**; 12:1-15.
25. Black CA, So W, Dallas SS, et al. Predominance of Non-carbapenemase Producing Carbapenem-Resistant Enterobacterales in South Texas. Frontiers in Microbiology **2021**; 11:3629.

26. Parker JK, Gu R, Estrera GA, et al. Carbapenem-Resistant and ESBL-Producing Enterobacterales Emerging in Central Texas. Infection and Drug Resistance **2023**; Volume 16:1249-61.
27. CLSI. Performance Standards for Antimicrobial Susceptibility Testing. 33rd ed. ed. Wayne, PA: Clinical and Laboratory Standards Institute, **2023**.
28. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics **2020**; 70:e102.
29. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nature Communications **2021**; 12:4188.
30. Wyres KL, Wick, R. R., Gorrie, C., Jenney, A., Follador, R., Thomson, N. R., & Holt, K. E. . Identification of Klebsiella capsule synthesis loci from whole genome data. Microbial genomics **2016**; 2:e000102.
31. Clausen PTLC, Zankari E, Aarestrup FM, Lund O. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. Journal of Antimicrobial Chemotherapy **2016**; 71:2484-8.
32. Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics **2018**; 19:1-8.
33. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nature biotechnology **2011**; 29:24-6.
34. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics **2014**; 30:2068-9.
35. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics **2015**; 31:3691-3.
36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **2013**; 30:772-80.
37. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol **2020**; 37:1530-4.
38. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution **2018**; 35:518-22.
39. Lees JA, Harris SR, Tonkin-Hill G, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Research **2019**; 29:304-16.
40. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Res **2018**; 3:93.
41. Grant JR, Enns E, Marinier E, et al. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Research **2023**; 51:W484-W92.
42. Shropshire WC, Strope B, Anand SS, et al. Temporal dynamics of genetically heterogeneous extended-spectrum cephalosporin-resistant Escherichia coli bloodstream infections. mSphere **2023**; 0:e00183-23.
43. Lam MMC, Wick RR, Wyres KL, et al. Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in Klebsiella pneumoniae populations. Microb Genom **2018**; 4.
44. Beceiro A, Maharjan S, Gaulton T, et al. False extended-spectrum beta-lactamase phenotype in clinical isolates of Escherichia coli associated with increased expression of OXA-1 or TEM-1 penicillinases and loss of porins. J Antimicrob Chemother **2011**; 66:2006-10.
45. Wiener ES, Heil EL, Hynicka LM, Johnson JK. Are Fluoroquinolones Appropriate for the Treatment of Extended-Spectrum β-Lactamase-Producing Gram-Negative Bacilli? Journal of Pharmacy Technology **2016**; 32:16-21.
46. Weber A, Neffe L, Diaz LAP, et al. Analysis of transmission-related third-generation cephalosporin-resistant Enterobacterales by electronic data mining and core genome multi-locus sequence typing. J Hosp Infect **2023**; 140:96-101.

47. Gorrie CL, Mirceta M, Wick RR, et al. Genomic dissection of Klebsiella pneumoniae infections in hospital patients reveals insights into an opportunistic pathogen. Nat Commun **2022**; 13:3017.
48. Li W, Liu X, Tsui W, et al. Identification and Comparative Genomic Analysis of Type VI Secretion Systems and Effectors in Klebsiella pneumoniae. Front Microbiol **2022**; 13:853744.
49. Shropshire WC, Konovalova A, McDaneld P, et al. Systematic Analysis of Mobile Genetic Elements Mediating Beta-Lactamase Gene Amplification in Noncarbapenemase-Producing Carbapenem-Resistant Enterobacterales Bloodstream Infections. Msystems **2022**; 7:e00476-22.
50. Jernigan JA, Hatfield KM, Wolford H, et al. Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012–2017. New England Journal of Medicine **2020**; 382:1309-19.
51. Arcari G, Carattoli A. Global spread and evolutionary convergence of multidrug-resistant and hypervirulent Klebsiella pneumoniae high-risk clones. Pathog Glob Health **2023**; 117:328-41.
52. Ljungquist O, Haldorsen B, Pontinen AK, et al. Nationwide, population-based observational study of the molecular epidemiology and temporal trend of carbapenemase-producing Enterobacterales in Norway, 2015 to 2021. Euro Surveill **2023**; 28:2200774.
53. Garcia-Gonzalez N, Fuster B, Tormo N, Salvador C, Gimeno C, Gonzalez-Candelas F. Genomic analysis of the initial dissemination of carbapenem-resistant Klebsiella pneumoniae clones in a tertiary hospital. Microb Genom **2023**; 9:001032.
54. Thorpe HA, Booton R, Kallonen T, et al. A large-scale genomic snapshot of Klebsiella spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings. Nature Microbiology **2022**; 7:2054-67.
55. Kochan TJ, Nozick SH, Medernach RL, et al. Genomic surveillance for multidrug-resistant or hypervirulent Klebsiella pneumoniae among United States bloodstream isolates. BMC Infect Dis **2022**; 22:603.
56. McDanel J, Schweizer M, Crabb V, et al. Incidence of Extended-Spectrum beta-Lactamase (ESBL)-Producing Escherichia coli and Klebsiella Infections in the United States: A Systematic Literature Review. Infect Control Hosp Epidemiol **2017**; 38:1209-15.
57. Duffy N, Karlsson M, Reses HE, et al. Epidemiology of extended-spectrum β-lactamase–producing Enterobacterales in five US sites participating in the Emerging Infections Program, 2017. Infection Control & Hospital Epidemiology **2022**:1-9.
58. Anderson DJ, Richet H, Chen LF, et al. Seasonal variation in Klebsiella pneumoniae bloodstream infection on 4 continents. J Infect Dis **2008**; 197:752-6.
59. de la Court JR, Woudt SHS, Schoffelen AF, et al. Third-generation cephalosporin resistant gram-negative bacteraemia in patients with haematological malignancy; an 11-year multi-centre retrospective study. Ann Clin Microbiol Antimicrob **2022**; 21:54.
60. Gorrie CL, Mirceta M, Wick RR, et al. Antimicrobial-Resistant Klebsiella pneumoniae Carriage and Infection in Specialized Geriatric Care Wards Linked to Acquisition in the Referring Hospital. Clin Infect Dis **2018**; 67:161-70.
61. Diorio-Toth L, Wallace MA, Farnsworth CW, et al. Intensive care unit sinks are persistently colonized with multidrug resistant bacteria and mobilizable, resistance-conferring plasmids. Msystems **2023**; 8:e00206-23.
62. El Haddad L, Hanson BM, Arias CA, et al. Emergence and Transmission of Daptomycin and Vancomycin-Resistant Enterococci Between Patients and Hospital Rooms. Clin Infect Dis **2021**; 73:2306-13.
63. Sundermann AJ, Chen J, Kumar P, et al. Whole-Genome Sequencing Surveillance and Machine Learning of the Electronic Health Record for Enhanced Healthcare Outbreak Detection. Clin Infect Dis **2022**; 75:476-82.

64. Mert D, Iskender G, Kolgelier S, Ertek M. Evaluation of risk factors, causative pathogens, and treatment in recurrent percutaneous nephrostomy catheter-related urinary tract infections in cancer patients. Medicine (Baltimore) **2023**; 102:e33002.
65. Refay SM, Ahmed EH, Abd ELzaher AR, Morsy AM, Yasser MM, Mahmoud AM. Risk of Drug Resistance and Repeated Infection with Klebsiella pneumoniae and Escherichia coli in Intensive Care Unit Cancer Patients. Combinatorial Chemistry & High Throughput Screening **2022**; 25:324-34.
66. Djukovic A, González-Barberá EM, Sanz J, et al. High heterogeneity of multidrug-resistant Enterobacteriaceae fecal levels in hospitalized patients is partially driven by intravenous β-lactams. Antimicrobial agents and chemotherapy **2020**; 64:e01415-19.
67. Gundes S, Arisoy AE, Kolayli F, et al. An outbreak of SHV-5 producing Klebsiella pneumoniae in a neonatal intensive care unit; meropenem failed to avoid fecal colonization. The new microbiologica **2005**; 28:231-6.
68. Murtha AN, Kazi MI, Schargel RD, et al. High-level carbapenem tolerance requires antibiotic-induced outer membrane modifications. PLOS Pathogens **2022**; 18:e1010307.
69. Herrera S, Torralbo B, Herranz S, et al. Carriage of multidrug-resistant Gram-negative bacilli: duration and risk factors. Eur J Clin Microbiol Infect Dis **2023**; 42:631-8.
70. Campos-Madueno EI, Moradi M, Eddoubaji Y, et al. Intestinal colonization with multidrug-resistant Enterobacterales: screening, epidemiology, clinical impact, and strategies to decolonize carriers. Eur J Clin Microbiol Infect Dis **2023**; 42:229-54.
71. Roshika R, Jain I, Medicielo J, Wachter J, Danger JL, Sumby P. The RD2 Pathogenicity Island Modifies the Disease Potential of the Group A Streptococcus. Infect Immun **2021**; 89:e0072220.
72. Vega LA, Sanson MA, Cubria MB, et al. The integrative conjugative element ICESpyM92 contributes to pathogenicity of emergent antimicrobial-resistant emm92 group A Streptococcus. Infection and Immunity **2022**; 90:e00080-22.
73. Wong Fok Lung T, Charytonowicz D, Beaumont KG, et al. Klebsiella pneumoniae induces host metabolic stress that promotes tolerance to pulmonary infection. Cell Metab **2022**; 34:761-74 e9.
74. Merciecca T, Bornes S, Nakusi L, et al. Role of Klebsiella pneumoniae Type VI secretion system (T6SS) in long-term gastrointestinal colonization. Sci Rep **2022**; 12:16968.
75. Storey D, McNally A, Astrand M, et al. Klebsiella pneumoniae type VI secretion system-mediated microbial competition is PhoPQ controlled and reactive oxygen species dependent. PLoS Pathog **2020**; 16:e1007969.
76. Wang H, Guo Y, Liu Z, Chang Z. The Type VI Secretion System Contributes to the Invasiveness of Liver Abscess Caused by Klebsiella pneumoniae. J Infect Dis **2023**; 228:1127-36.