Transmission of Carbapenem-Resistant *Klebsiella pneumoniae* in US Hospitals

Courtney L. Luterbach,[1,2] Liang Chen,[3] Lauren Komarow,[4] Belinda Ostrowsky,[5] Keith S. Kaye,[6] Blake Hanson,[7,8] Cesar A. Arias,[7,8,9] Samit Desai,[10] Jason C. Gallagher,[11] Elizabeth Novick,[11] Stephen Pagkalinawan,[11] Ebbing Lautenbach,[12] Glenn Wortmann,[13] Robert C. Kalayjian,[14] Brandon Eilertson,[15] John J. Farrell,[16] Todd McCarty,[17] Carol Hill,[18] Vance G. Fowler Jr,[18] Barry N. Kreiswirth,[3] Robert A. Bonomo,[19,20,21,22] and David van Duin,[1] for the Multi-Drug Resistant Organism Network Investigators Network Investigators and the Antibacterial Resistance Leadership Group

[1]Division of Infectious Diseases, University of North Carolina, Chapel Hill, North Carolina, USA; [2]Division of Pharmacotherapy and Experimental Therapeutics, University of North Carolina, Chapel Hill, North Carolina, USA; [3]Center for Discovery and Innovation, Hackensack Meridian Health, Nutley, New Jersey, USA; [4]Biostatistics Center, George Washington University, Rockville, Maryland, USA; [5]Division of Infectious Diseases, Department of Medicine, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York, USA; [6]Division of Infectious Diseases, Rutgers Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA; [7]Division of Infectious Diseases and Center for Antimicrobial Resistance and Microbial Genomics, UTHealth, McGovern School of Medicine at Houston, Houston, Texas, USA; [8]Center for Infectious Diseases, UTHealth School of Public Health, Houston, Texas, USA; [9]Molecular Genetics and Antimicrobial Resistance Unit–International Center for Microbial Genomics, Universidad El Bosque, Bogota, Columbia; [10]Division of Infectious Diseases, Hackensack University Medical Center, Hackensack, New Jersey, USA; [11]Temple University School of Pharmacy, Philadelphia, Pennsylvania, USA; [12]Division of Infectious Diseases, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [13]Section of Infectious Diseases, MedStar Washington Hospital Center, Washington, District of Columbia, USA; [14]Department of Medicine, MetroHealth Medical Center, Cleveland, Ohio, USA; [15]Division of Infectious Diseases, Department of Medicine, State University of New York Downstate, Brooklyn, NY, USA; [16]Division of Infectious Disease, Department of Internal Medicine, OSF Saint Francis Medical Center, University of Illinois College of Medicine at Peoria, Peoria, Illinois, USA; [17]Division of Infectious Disease, University of Alabama at Birmingham, Birmingham, Alabama, USA; [18]Duke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina, USA; [19]Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, Ohio, USA; [20]Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA; [21]Departments of Pharmacology, Molecular Biology and Microbiology, Biochemistry, and Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA; and [22]Case Western Reserve University-Cleveland Veterans Affairs Medical Center for Antimicrobial Resistance and Epidemiology (Case VA CARES), Cleveland, Ohio, USA

Background. Carbapenem-resistant *Klebsiella pneumoniae* (CR*Kp*) is the most prevalent carbapenem-resistant Enterobacterales in the United States. We evaluated CR*Kp* clustering in patients in US hospitals.

Methods. From April 2016 to August 2017, 350 patients with clonal group 258 CR*Kp* were enrolled in the Consortium on Resistance Against Carbapenems in *Klebsiella* and other *Enterobacteriaceae*, a prospective, multicenter, cohort study. A maximum likelihood tree was constructed using RAxML. Static clusters shared ≤21 single-nucleotide polymorphisms (SNP) and a most recent common ancestor. Dynamic clusters incorporated SNP distance, culture timing, and rates of SNP accumulation and transmission using the R program TransCluster.

Results. Most patients were admitted from home (n = 150, 43%) or long-term care facilities (n = 115, 33%). Urine (n = 149, 43%) was the most common isolation site. Overall, 55 static and 47 dynamics clusters were identified involving 210 of 350 (60%) and 194 of 350 (55%) patients, respectively. Approximately half of static clusters were identical to dynamic clusters. Static clusters consisted of 33 (60%) intrasystem and 22 (40%) intersystem clusters. Dynamic clusters consisted of 32 (68%) intrasystem and 15 (32%) intersystem clusters and had fewer SNP differences than static clusters (8 vs 9; $P = .045$; 95% confidence interval [CI]: −4 to 0). Dynamic intersystem clusters contained more patients than dynamic intrasystem clusters (median [interquartile range], 4 [2, 7] vs 2 [2, 2]; $P = .007$; 95% CI: −3 to 0).

Conclusions. Widespread intrasystem and intersystem transmission of CR*Kp* was identified in hospitalized US patients. Use of different methods for assessing genetic similarity resulted in only minor differences in interpretation.

Keywords. carbapenem-resistant Enterobacterales; *Klebsiella pneumoniae*; transmission clusters.

Carbapenem-resistant Enterobacterales (CRE) remain an important threat. The Centers for Disease Control and Prevention (CDC) has estimated that 13 100 cases of CRE occurred in hospitalized patients in 2017 [1]. We recently reported on the clinical and molecular epidemiology of CRE in the United States [2]. We estimated that CRE are isolated from a clinical culture in approximately 57 per 100 000 US hospital admissions. Carbapenem-resistant *Klebsiella pneumoniae* (CR*Kp*) is the most common CRE species in the United States. Globally, *K. pneumoniae* sequence type (ST) 258 is the most encountered type within carbapenemase-producing CR*Kp* [3]. Common families of carbapenemases include *K. pneumoniae* carbapenemase (KPC), oxacillinase (OXA)-48–like carbapenemases, and metallo-β-lactamases such as New Delhi metallo-β-lactamase, Verona integron-encoded metallo-β-lactamase, and active on imipenem carbapenemases [4]. Previous studies have traced CR*Kp* infections in hospitalized patients within

Received 04 May 2022; ed itorial d ecision 12 September 2022; published online 29 September 2022

Correspondence: David van Duin, Division of Infectious Diseases, CB 7030, University of North Carolina, 130 Mason Farm Road, Chapel Hill, NC 27599, USA (david_vanduin@med.unc.edu).

Clinical Infectious Diseases® 2023;76(2):229–37

https://doi.org/10.1093/cid/ciac791

and between hospitals, skilled nursing facilities (SNFs), and long-term acute care (LTAC) hospitals [5, 6]. Limiting the spread of CR*Kp* between healthcare settings is an important goal, but defining thresholds of genetic relatedness between isolates for epidemiological investigations has been challenging [7–9]. Inconsistencies in defining bacterial clusters may steer epidemiologic investigations toward reaching different conclusions regarding likely transmission pathways. Whole-genome sequencing has improved the granularity of grouping bacterial strains into genetically related clusters. Selecting a static single-nucleotide polymorphism (SNP) cutoff has been a traditional approach to define closely related isolates. Alternatively, dynamic cluster assignment using a combination of sampling times, genetic distance, and rates of SNP accumulation and transmission may better represent epidemiological links between genetically similar isolates. This dynamic approach to clustering has previously been applied to define clusters for *Mycobacterium tuberculosis* and coronavirus disease 2019 but has not been evaluated in gram-negative bacteria [10, 11].

Here, we evaluated hospitalized patients with CG258 CR*Kp* from the Consortium on Resistance Against Carbapenems in *Klebsiella* and other Enterobacterales (CRACKLE-2) study [2]. To better understand transmission in hospitals, we used static and dynamic methods to determine the degree of clustering in CR*Kp* isolates from these patients and compared differences between the 2 approaches using a combination of whole-genome sequencing data and probability of recent transmission.

METHODS

Study Design

The CRACKLE-2 study has previously been described [2]. Briefly, CRACKLE-2 is a multicenter, prospective, observational cohort study of hospitalized patients with at least 1 clinical culture of CRE, as defined by the CDC [1]. The CDC defines CRE as Enterobacterales that phenotypically test resistant to any carbapenem (ie, minimum inhibitory concentrations of ≥4 μg/mL for doripenem, meropenem, or imipenem or ≥2 μg/mL for ertapenem), or harbor a gene encoding a carbapenemase, or are positive for carbapenemase production. Here, we constructed a cohort of patients with CG258 CR*Kp* nested within the CRACKLE-2 cohort who were enrolled in the United States from 30 April 2016 until 31 August 2017. One patient was excluded as an outlier as the collected isolate had, on average, $>10^4$ pairwise SNP differences from the other 350 isolates. Dates of admission, discharge, room transfer, and room location were acquired from the electronic healthcare records for 1 hospital. This hospital was selected as it had multiple clusters identified using both static and dynamic methods. While other hospitals also met these criteria, it was not feasible to perform secondary data collection at all centers. Institutional review board approval was obtained at all participating centers.

Microbiology and Whole-Genome Sequencing

DNA isolation and genome assembly were performed as part of the CRACKLE-2 study [2]. Briefly, single colonies for each isolate were selected for sequencing on lysogenic broth agar plates supplemented with 0.5 mg/L ertapenem or imipenem. Genomic DNA was extracted via the Wizard Genomic DNA Purification (Promega) or DNeasy Blood and Tissue (QIAGEN) kits and prepared for sequencing using the Illumina Nextera XT DNA sample preparation kit (Illumina, San Diego, CA). Trimmomatic v0.36 was used to trim low-quality sequences and remove Illumina Nextera indexes [12]. Draft genomes were assembled using SPAdes v3.11.1 [13] and evaluated using Quast v4.6.2 [14]. Species were confirmed using StrainSeeker v1.5 [15]. Multilocus sequence typing (MLST) was performed using the program MLST [16]; capsular polysaccharide gene clusters and *wzi* allele typing was performed using Kleborate v0.1.0 [17]; and resistance genes were called by ABRicate [18] and ARIBA using the National Center for Biotechnology Information (NCBI) Bacterial Antimicrobial Resistance Reference Gene Database and ResFinder [19, 20]. Inconsistent results between the 2 programs were manually curated.

Phylogenetic Analysis

Trimmed, paired-end sequences from each draft genome were mapped to the reference NJST258_2 genome using Snippy [21]. DNA regions masked from the alignment included prophages (PHASTER [22]), repeated regions (MUMmer [23]), and areas of recombination (Gubbins [24]). A maximum likelihood phylogenetic tree from the concatenated core genome SNP sites was constructed using RAxML v8.2.4 with a general time-reversible model of nucleotide substitution and 4 discrete gamma categories of rate heterogeneity (GTRGAMMA) [25]. The phylogenetic tree was annotated with the R packages ggtree [26]. Genomes are publicly available on NCBI (PRJNA658369).

Cluster Definitions

Static clusters were defined as strains that shared a most recent common ancestor (MRCA) based on phylogenetic analysis and a fixed cutoff of ≤21 pairwise SNP differences with every isolate within the cluster [5]. Pairwise whole-genome SNP differences (ie, includes invariant sites) were calculated using Snippy with variant calling performed using FreeBayes with default minimum coverage and quality cutoffs. In cases where an isolate could be grouped into 2 clusters, the isolate was assigned based on the smaller SNP difference between the nearest neighbors in each cluster.

Dynamic clusters were identified using the R program TransCluster [10], which uses a probabilistic methodology that combines the rate of SNP accumulation (λ), timing of CR*Kp* detection, and an estimated transmission rate (β) to model the likelihood of isolates being linked by a transmission threshold (T) (ie, maximum number of transmission events). Pairwise SNP differences identified via Snippy were used to

Table 1. Comparison of Baseline Patient Characteristics and Clinical Outcomes by Cluster Type

Characteristic[a]	Overall (n = 350)	Static Clusters			Dynamic Clusters		
		Not Clustered (n = 140)	Clustered (n = 210)	*P* Value (95% CI[b])	Not Clustered (n = 156)	Clustered (n = 194)	*P* Value (95% CI[b])
Age, y	66 (56–76)	62 (56–73)	67 (56–77)	.02 (–7 to –1)	62 (56–74)	66 (56–77)	…
Female sex	168 (48)	63 (45)	105 (50)	…	72 (46)	96 (49)	…
Ethnicity	…	…	…	…	…	…	…
Hispanic or Latino	39 (11)	15 (11)	24 (11)	…	20 (13)	19 (10)	…
Not Hispanic or Latino	254 (73)	100 (71)	154 (73)	…	110 (71)	144 (74)	…
Not reported/Unknown	57 (16)	25 (18)	32 (15)	…	26 (17)	31 (16)	…
Time from admission to culture, d	2 (0–12)	1 (0–10)	2 (0–13)	…	1 (0–11)	2 (0–13)	…
Time from admission to discharge/death, d	17 (8–34)	15 (7–32)	18 (8–35)	…	17 (8–34)	16 (8–33)	…
Culture source	…	…	…	…	…	…	…
Blood	47 (13)	17 (12)	30 (14)	…	21 (13)	26 (13)	…
Nonwound abdominal	8 (2)	4 (3)	4 (2)	…	4 (3)	4 (2)	…
Respiratory	89 (25)	38 (27)	51 (24)	…	44 (28)	45 (23)	…
Urine	149 (43)	61 (44)	88 (42)	…	63 (40)	86 (44)	…
Wound	43 (12)	15 (11)	28 (13)	…	18 (12)	25 (13)	…
Other	14 (4)	5 (4)	9 (4)	…	6 (4)	8 (4)	…
Disease status	…	…	…	…	…	…	…
Colonization	204 (58)	87 (62)	117 (56)	…	95 (61)	109 (56)	…
Infection	146 (42)	53 (38)	93 (44)	…	61 (39)	85 (44)	…
Charlson score	3 (1–5)	3 (1–5)	3 (1–5)	…	3 (1–5)	3 (1–5)	…
Pitt score	3 (2–6)	3 (2–6)	3 (2–6)	…	3 (2–6)	3 (2–6)	…
Clinical response	110 (31)	45 (32)	65 (31)	…	53 (34)	57 (29)	…
Mortality	…	…	…	…	…	…	…
30-d	79 (23)	35 (25)	44 (21)	…	37 (24)	42 (22)	…
90-d	102 (29)	45 (32)	57 (27)	…	48 (31)	54 (28)	…
Readmission within 90 d	115 (33)	36 (26)	79 (38)	.02 (1.1 to 2.9)	41 (26)	74 (38)	.02 (1.1 to 2.8
Preadmission origin	…	…	…	…	…	…	…
Home	150 (43)	68 (49)	82 (39)	…	76 (49)	74 (38)	…
Long-term acute care	29 (8)	12 (9)	17 (8)	…	14 (9)	15 (8)	…
Long-term care	115 (33)	39 (28)	76 (36)	…	42 (27)	73 (38)	…
Transfer from other hospital	53 (15)	19 (14)	34 (16)	…	22 (14)	31 (16)	…
Transfer from outside United States	1 (0)	0 (0)	1 (0)	…	0 (0)	1 (1)	…
Unknown	2 (1)	2 (1)	0 (0)	…	2 (1)	0 (0)	…

Abbreviation: CI, confidence interval.

[a]Data presented as either n (%) or median (interquartile range) unless otherwise stated.

[b]95% CI of the median difference.

measure genetic differences between isolates. The λ was set at 10.1 substitutions/genome/year as previously estimated using paired longitudinal samples of KPC-*Kp* [27]. The date of CR*Kp* culture was collected from electronic healthcare records. The β represents the estimated number of transmissions per year and is defined as the rate at which intermediate cases occur in the total time elapsed between the MRCA of 2 sampled hosts and the sampling events [10, 11]. A β value of 5.8 was calculated using an average generation time of 62.7 days derived from epidemiologic investigation of nosocomial *Kp* [28]. The threshold for T is defined as the number of intermediate transmissions ≤T between 2 isolates with a probability of 80%. Based on cluster characteristics at various thresholds, a threshold of T = 5 was selected (Supplementary Figure 1*A* and 1*B*).

Intrasystem clusters were defined as containing only isolates collected from the same healthcare system. In contrast, intersystem clusters have at least 2 isolates from different healthcare systems but may also contain a subset of isolates linked within a single healthcare system. Intrasystem and intersystem clusters are mutually exclusive; each isolate can only belong to a single cluster, and each cluster is designated as either intrasystem or intersystem. Each healthcare system was comprised of either a single hospital or an organization of related hospitals.

Statistical Analyses

To compare the average nucleotide identity (ANI) between genetically nearest neighbors (gNN), a 2-tailed Student *t* test was performed. Distributions of SNP difference and continuous variables were compared using the Mann–Whitney test. Distributions across groups for categorical variables were compared using a Fisher exact test. *P* values ≤.05 were considered

Table 2. Genetic Characterization of Carbapenem-Resistant *Klebsiella pneumoniae* Isolates Grouped by Cluster Type

Characteristic[a]	Overall (n = 350)	Static Clusters			Dynamic Clusters		
		Not Clustered (n = 140)	Clustered (n = 210)	*P* Value	Not Clustered (n = 156)	Clustered (n = 194)	*P* Value
CPE status	…	…	…	…	…	…	.030
CPE	339 (97)	132 (94)	207 (99)	…	147 (94)	192 (99)	…
Non-CP CRE	7 (2)	5 (4)	2 (1)	…	6 (4)	1 (1)	…
Unconfirmed CRE	4 (1)	3 (2)	1 (0)	…	3 (2)	1 (1)	…
bla_{KPC}	…	…	…	.013	…	…	<.001
KPC-2	186 (53)	61 (44)	125 (60)	…	65 (42)	121 (62)	…
KPC-3	152 (44)	71 (51)	81 (39)	…	82 (53)	70 (36)	…
KPC-8	1 (0)	0 (0)	1 (0)	…	0 (0)	1 (1)	…
$bla_{OXA-48-like}$	…	…	…	…	…	…	…
OXA-232	2 (1)	0 (0)	2 (1)	…	0 (0)	2 (1)	…
Multilocus sequence typing sequence type	…	…	…	…	…	…	…
CG258[b]	350 (100)	140/350 (40)	210/350 (60)	…	156/350 (45)	194/350 (55)	…
Tn4401 type	…	…	…	.047	…	…	.013
Tn4401a[c]	166 (47)	56 (40)	110 (52)	…	61 (39)	105 (54)	…
Tn4401b	66 (19)	31 (22)	35 (17)	…	33 (21)	33 (17)	…
Tn4401d	106 (30)	45 (32)	61 (29)	…	53 (34)	53 (27)	…
Undetermined	12 (3)	8 (6)	4 (2)	…	9 (6)	3 (2)	…
wzi capsule type[d]	…	…	…	<.001	…	…	<.001
154	168 (48)	80 (57)	88 (42)	…	93 (60)	75 (39)	…
29	112 (32)	41 (29)	71 (34)	…	43 (28)	69 (36)	…
168	26 (7)	11 (8)	15 (7)	…	11 (7)	15 (8)	…
50	23 (7)	0 (0)	23 (11)	…	0 (0)	23 (12)	…
Other	21 (6)	8 (6)	13 (6)	…	9 (6)	12 (6)	…

Abbreviations: CPE, carbapenemase-producing Enterobacterales; CRE, carbapenem-resistant Enterobacterales.

[a]Data presented as N (%).

[b]Includes single-locus variants of sequence type 258 (n = 7).

[c]Tn4401 isoforms containing a 35 bp (n = 1) and 210 DNA (n = 1) deletion upstream of bla_{KPC} gene.

[d]Includes imperfect allele matches for *wzi*-154 (n = 1), *wzi*-29 (n = 3), and *wzi*-168 (n = 1). "Other" includes *wzi*-150 (n = 3), *wzi*-174 (n = 5), *wzi*-83 (n = 5), and undetermined *wzi* type (n = 8).

statistically significant. Analyses were performed using either R (version 4.0.0) or GraphPad Prism (version 9.2.0).

RESULTS

The study cohort consisted of 350 hospitalized patients with a single, positive culture for CG258 *K. pneumoniae* across 25 US healthcare systems (with 42 hospitals) in 13 states and the District of Columbia. Baseline characteristics of patients and isolates are summarized in Tables 1 and 2, respectively. Most patients were admitted from home (n = 150, 43%) or a long-term care facility (n = 115, 33%). The majority of CR*Kp* isolates were from urine (n = 149, 43%) followed by the respiratory tract (n = 89, 25%). Overall, isolates were predominantly carbapenemase-producing (n = 339, 97%). Within carbapenemase-producing (CP)-CR*Kp* isolates, bla_{KPC-2} (n = 186, 53%) and bla_{KPC-3} (n = 152, 44%) were the most identified carbapenemase genes.

The phylogenetic tree is shown in Figure 1. *wzi*-29 (32%) and *wzi*-154 (48%) were the most common wzi types. Core SNP differences between all 350 isolates ranged from 0 to 259 (median, 74; interquartile range [IQR], 60–87). The overall ANI between gNNs was 99.66% (range, 96.65%–100%). Within the same healthcare system, the ANI between gNN was slightly higher compared with gNN recovered from patients at different healthcare systems (median, 99.7 vs 99.5; $P < .0001$; 95% confidence interval [CI]: .09 to .22).

Clusters

Overall, 55 static and 47 dynamic clusters were identified and incorporated 210 of 350 (60%) and 194 of 350 (55%) patients, respectively (Figure 2, Table 3). Most clusters were identified at healthcare systems within the Northeast region (static: 29 of 55, 53%; dynamic: 25 of 47, 53%). Static clusters were identified at 19 of 25 (76%) healthcare systems and were comprised of 33 (60%) intrasystem and 22 (40%) intersystem clusters. Similarly, dynamic clusters were identified across 19 of 25 (76%) healthcare systems and consisted of 32 of 47 (68%) intrasystem and 15 of 47 (32%) intersystem clusters. Overall, 29 of 55 (53%) static clusters were identical in size and isolate composition to the dynamic clusters. There were 20 and 4 isolates specific to only static clusters or dynamic clusters, respectively. Static clusters ranged in size from 2 to 21 patients (median, 2), and dynamic clusters ranged in size from 2 to 28 patients (median, 2). Overall,

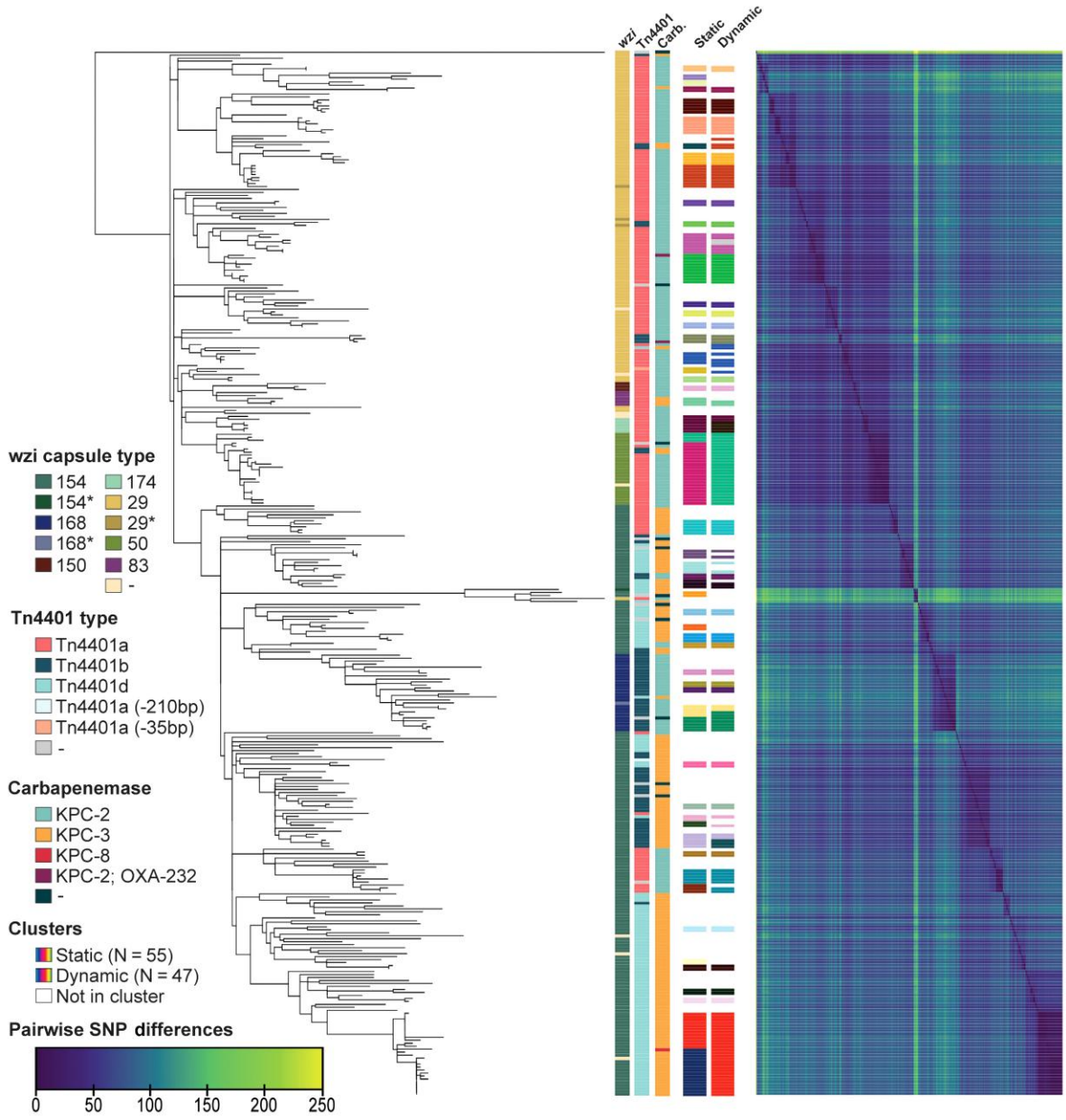

Figure 1. Phylogenetic population structure of *Klebsiella pneumoniae* CG258 isolates. A maximum likelihood tree indicates the genetic relationships between isolates. Additional metadata (from left to right) for each isolate include the *wzi* type, Tn*4401* type, detected Carb. genes, static cluster assignment, and dynamic cluster assignment. Individual clusters are coded as unique colors. The right-most panel represents a heat map of the pairwise core SNP distances between isolates with purple to yellow signifying an increasing number of SNP differences. Abbreviations: –, not detected/unknown; Carb, carbapenemase; KPC, *Klebsiella pneumoniae* Carbapenemase; OXA, oxacillinase; SNP, single-nucleotide polymorphism.

dynamic clusters had fewer SNP differences between each pair of isolates within individual clusters than static clusters (median, 8 and IQR, 4–11] vs median, 9 and IQR, 5–15; $P = .045$; 95% CI: −4 to 0). Dynamic intersystem clusters generally contained more patients than intrasystem clusters (median, 4 and IQR, 2–7 vs median, 2 and IQR, 2–2; $P = .007$; 95% CI: 0 to 3).

We evaluated the impact of adjusting the transmission threshold (T) and β parameter on dynamic clustering. Selection of higher T thresholds (ie, allowing for more transmission links between patients within a cluster) increased the overall number of patients incorporated into clusters with a maximum number of patients (n = 260) reached at T = 11 (Supplementary Figure 1*A*). Conversely, the number of clusters peaked at T = 5 but also reached a plateau at T = 11 (Supplementary Figure 1*B*). We observed that as T increased, the intrasystem clusters were incorporated into larger and more diverse intersystem clusters. Next, we evaluated the impact of varying the β parameter on dynamic clustering. The

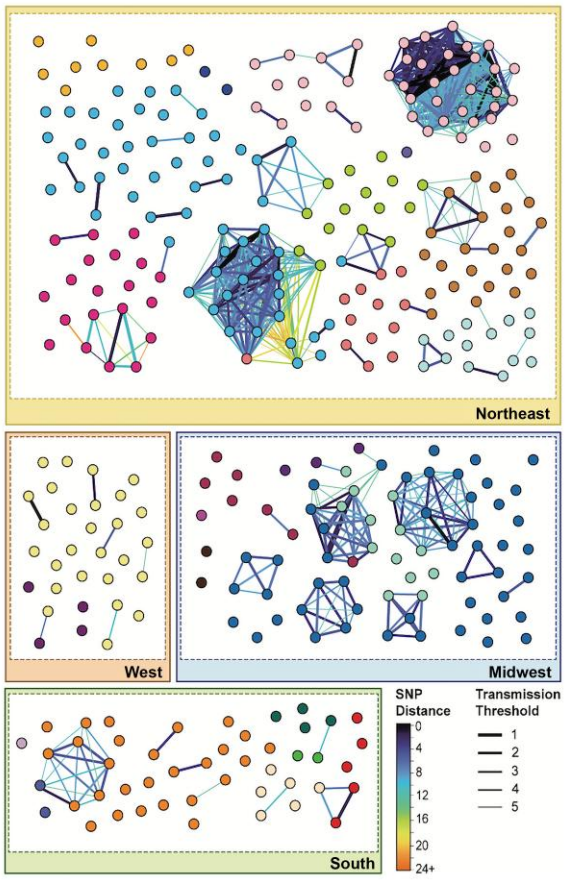

Figure 2. Structure of dynamic clusters of *Klebsiella pneumoniae* CG258 isolates. Isolates (circles) are color-coded by healthcare system (n = 25) and grouped by region within the United States. Isolates were assigned to the same dynamic cluster (connected lines) if they shared an 80% probability of being within a threshold of 5 putative transmissions. Lines are weighted by the transmission threshold linking 2 isolates and also color-coded by pairwise SNP distances. Abbreviation: SNP, single-nucleotide polymorphism.

overall number of strains within clusters and total number of clusters were reduced at the higher β values, which represents a faster transmission potential (Supplementary Figure 1*C* and 1*D*, respectively). The highest number of clusters (n = 47) was identified at β = 5.8. In alignment with Stimson et al, increasing the transmission rate β resulted in the same clusters but at higher transmission cutoffs [10].

Investigation of Spatial Clustering

We further investigated the dynamic cluster assignment compared with the spatial location between 14 unique patients within a single hospital (Supplementary Figure 2). Patient room data were grouped by hospital floor. Overall, 4 dynamic clusters (3 intrasystem and 1 intersystem) were previously identified using the spatial distance at the healthcare-system level. Intrasystem clusters contained either 2 patients (n = 2) or 3 patients (n = 1). The intersystem cluster consisted of 2 patients, with 1 patient at a different healthcare system but within the same geographic region. Most (5 of 9, 56%) of the isolates were collected from the same floor (floor C); however, only 1 intrasystem cluster had all isolates contained to a single floor. Application of the static clustering method identified the same clusters.

DISCUSSION

In this prospective cohort of consecutively enrolled patients, we show evidence of extensive nosocomial transmission and spread of CG258 CR*Kp* in US hospitals. Most patients had a CR*Kp* isolate that could be genetically linked to an isolate of at least 1 other patient, regardless of the method used to assign clusters. Most patients were part of clusters with patients from the same healthcare system; however, about one-third of clusters showed evidence of transmission across healthcare systems. The occurrence of intersystem clusters may indicate involvement of other healthcare sites (eg, SNFs, LTAC hospitals) as well as the community in perpetuating CR*Kp* spread [29, 30]. These observations emphasize that successful control of multidrug-resistant organisms requires infection prevention measures at both local and regional levels.

As a group, patients who were part of clusters were not clinically different from those who were not part of clusters. We did observe that patients within clusters had a higher rate of readmission within 90 days. Longitudinal surveillance culturing of patients would assist in better understanding the factors that drive cluster formation. When we evaluated bacterial genetics, isolates within clusters were more likely to carry the carbapenemase gene $bla_{\text{KPC-2}}$ as well as the Tn*4401a* isoform. bla_{KPC} genes are generally located within the Tn3-based transposon Tn*4401* [31]. We also observed high, but not exclusive, carriage of $bla_{\text{KPC-2}}$ among clade I isolates (characterized by the wzi-29 allele), as noted in other studies [6, 32]. Control of the presence of antibiotic-resistant bacteria in the nosocomial environment can break transmission chains and decrease the likelihood of horizontal transfer of genes associated with antibiotic resistance to other bacteria [33].

Clonal dissemination of ST258 throughout US healthcare systems has been linked to several nosocomial outbreaks [34–36]. The endemicity of ST258 makes it challenging to distinguish independent introduction events from ongoing transmission in the hospital setting [6, 35]. Isolates that are part of clusters share a high degree of genetic similarity, which suggests that they are a result of recent transmission [37]. Static clustering uses a pairwise minimum SNP difference between the core genome of isolates to indicate the likelihood of recent transmission events. The main limitation of this approach is that it requires selection of a cutoff for the SNP threshold. Differences in the methodology for genome assembly and variant calling as well as biological variability between isolates can alter the

Table 3. Comparison of Static and Dynamic Clusters

Characteristic[a]	Static Clusters				Dynamic Clusters			
	Total (n = 55)	Intrasystem (n = 33)	Intersystem (n = 22)	*P* Value (95% CI[b])	Total (n = 47)	Intrasystem (n = 32)	Intersystem (n = 15)	*P* Value (95% CI[b])
Number of strains within clusters	210/350 (60)	112/210 (53)	98/210 (47)	…	194/350 (55)	106/194 (55)	88/194 (45)	…
Cluster size	2 (2–4)	2 (2–3.5)	3 (2–5)	…	2 (2–4)	2 (2–2)	4 (2–7)	.007 [−3–0]
Pairwise single-nucleotide polymorphism distance[c]	9 (5–15)	7 (4–15)	12 (8–15)	…	8 (4–11)	5 (2–11)	9 (7–11)	…
Site location of cluster[d]								
Midwest	12 (22)	4 (12)	8 (36)	…	9 (19)	5 (16)	4 (27)	…
Northeast	29 (53)	19 (58)	10 (45)	…	25 (53)	18 (56)	7 (47)	…
South	9 (16)	5 (15)	4 (18)	…	7 (15)	4 (13)	3 (20)	…
West	9 (16)	5 (15)	4 (18)	…	6 (13)	5 (16)	1 (7)	…
Time between culture dates, d	51 (23–86)	57 (18–100)	40 (28–83)	…	50 (23–94)	54 (20–87)	41 (27–102)	…
Time from admission to culture date, d	3 (1–9)	5 (1–11)	1 (0–8)	…	2 (1–8)	2 (1–9)	1 (1–8)	…

Abbreviation: CI, confidence interval.

[a]Data presented as either N (%) or median (interquartile range) unless otherwise stated.

[b]95% CI of the median difference.

[c]Median single-nucleotide polymorphism distance between each strain within a cluster.

[d]Several intersystem static clusters (n = 4) spanned across multiple regions.

relative pairwise SNP differences [5, 8, 9, 38]. Additionally, for isolates close to the SNP threshold, there is unlikely to be a large difference in transmission likelihood based on a single additional SNP difference leading to inclusion or exclusion from clusters. Indeed, David et al identified a false-positive and false-negative rate of 14.6% and 11.7%, respectively, even at the optimal threshold of 21 SNPs to discriminate ST258/512 clusters in hospitals [5]. Furthermore, it is unlikely that a single SNP threshold would perform equally well in different settings. For example, Ferrari et al identified a threshold of <16 SNPs to define *K. pneumoniae* KPC isolates as part of the same transmission cluster based on the distribution of core SNPs among isolates [9]. Conversely, using phylogenetic analysis, Hassoun-Kheir et al identified a cutoff of ≤80 SNPs that separated *K. pneumoniae* KPC STs and a stricter cutoff of ≤6 SNPs that defined 60.5% of isolates as being both genetically linked and sharing high epidemiological support [8]. Misidentification of transmission pathways and outbreak sources may result in misdirection of time and resources. Thus, threshold-free approaches may be a useful complement to current approaches to group genetically similar isolates.

Dynamic clustering is a threshold-free approach that incorporates genetic distance, spatial distance, culture timing, the rate of SNP acquisition, and the number of transmission events over time [10]. However, several parameters also need to be set based on previously collected epidemiological and genetic data, which may be incomplete. Applying the same transmission parameters may provide a unified clustering method across bacteria with different SNP accumulation rates [10]. Overall, using the parameter settings outlined, static and dynamic clusters were similar in isolate composition and numbers of patients per cluster. We applied a SNP threshold of ≤21 when we defined static clusters, which performed similarly for dynamic clustering using a threshold of 5 transmission events. More detailed exploration toward applying model simulations with a comparison to real-world surveillance data is warranted to optimize the threshold for detecting true transmission events.

Identification of genetically similar isolates within a hospital can support epidemiological data and guide infection preventionists during outbreak investigations and implementation of prevention control measures [28]. By either method, we found that a majority of CR*Kp* isolates were clustered with at least 1 other isolate, with small clusters consisting of 2 patients within the same healthcare system being the most commonly detected cluster type. Adjusting the transmission threshold for grouping clusters may be a useful approach for tailoring the scope and precision of epidemiological investigations. Similar to setting a more restrictive SNP threshold, a lower cutoff for transmission events identifies the most similar isolates and thus those most likely to be part of an outbreak. Likewise, selecting a higher threshold results in larger clusters with a greater likelihood of there being a linked transmission event occurring within each cluster.

The model parameters for the dynamic clusters were informed based on data from previous publications. It is possible that the use of different parameters may generate different outcomes and perform better for a given dataset. For example, the generation time between *Klebsiella* infections is mostly determined through studies of hospital outbreaks and varies depending on the context of infection and the extent of culture

surveillance [28, 35, 39, 40]. Therefore, an alternative β value may better represent patients linked by strong epidemiological support. Additionally, asymptomatic screening cultures were not included in CRACKLE-2. Long-term gastrointestinal colonization [41] and environmental reservoirs (eg, plumbing, surfaces) [42, 43] can contribute to the spread of CR*Kp* among healthcare facilities. Therefore, the degree of clustering may be underrepresented. Additionally, as hospitals were based on the interest of the investigators, the identified clusters do not comprehensively represent the extent of CG258 spread between facilities. As expected, several healthcare systems were relatively overrepresented with high numbers of patients with CR*Kp*. However, this represents the variability in CRE incidence between healthcare systems and regions [2, 44, 45].

In summary, we identified widespread nosocomial transmission of CR*Kp* in hospitalized patients in the United States. Evaluation of genetic similarity is an important tool for epidemiological studies that requires further standardization. More precise determination of genetic similarity and associated likelihood of transmission will help identify hot spots of transmission for more in-depth phylogenetic analyses and direct resources for control intervention methods.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Disclaimer. The content presented here is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH) or the Department of Veterans Affairs. The National Institute of Allergy and Infectious Diseases (NIAID) had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication, veto publication, or control the journal to which the manuscript was submitted.

Financial support. This work was supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the NIH (UM1AI104681 and R21AI114508). Research reported in this publication was also supported by the National Institute of General Medical Sciences (T32GM086330 to C. L. L.) and in part by the NIAID (R01AI143910 to D. v. D.; R01AI090155 and R21AI135250 to B. N. K.; R21AI117338 to L. C. and R01AI100560, R01AI063517, and R01AI072219 to R. A. B.). This study was supported in part by funds or facilities provided by the Cleveland Department of Veterans Affairs to R. A. B. from the Biomedical Laboratory Research and Development, Veterans Affairs Office of Research and Development (1I01BX001974) and the Geriatric Research Education and Clinical Center Veterans Integrated Service Network (VISN) 10 (R. A. B.). K. S. K. is supported by the Division of Microbiology and Infectious Diseases, NIAID and Infectious Diseases (protocol 10–0065 and R01AI119446).

Potential conflicts of interest. K. S. K. has been a consultant and grant investigator for and has received speaker's bureau honoraria, consulting fees, grants, and speaker honorarium from Allergan; receives grants from Merck; and is a consultant for Merck, Shionogi, Qpex Biopharma, Micurx Pharmaceuticals, Xellia Pharmaceuticals, and Achaogen. J. C. G. has been a consultant for Qpex Biopharma, Shionogi, Merck, Achaogen, Tetraphase Pharmaceuticals, Accelerate Diagnostics, Astellas Pharma, Melinta Therapeutics, Nabriva Therapeutics, Paratek Pharma, scPharmaceuticals, Spero Therapeutics, Tetraphase Pharmaceuticals, and Melinta Therapeutics; received speaker honoraria from Allergan, Melinta Therapeutics, and Merck; received a grant from Merck; received royalties from Jones and Barlett Learning; was a speaker for Astellas Pharma, Melinta Therapeutics, Merck, and Shionogi; received support for attending meetings from Merck; and is the editor-in-chief for *Contagion*. R. A. B. received grants or research support from Achaogen, Allecra Therapeutics, Entasis Therapeutics, Merck, Roche, Shionogi, VenatoRx Pharmaceuticals, and Wockhardt; has patents with Case Western Reserve University (CWRU); and served on a data and safety monitoring board (DSMB) as a logistics associate for the Division of Microbiology and Infectious Diseases Clinical Research Operations and Management Support (DMID-CROMS), Safety Oversight Committee Support, and Technical Resources International, Inc. D. v. D. has served on advisory boards for Allergan, Achaogen, Qpex Biopharma, Shionogi, Tetraphase Pharmaceuticals, Sanofi-Pasteur, T2 Biosystems, NeuMedicine, Roche, Pfizer, MedImmune, Astellas Pharma, and Merck; has received grant support from NIH, Shionogi, and Merck; has received consulting fees from Actavis, Tetraphase Pharmaceuticals, Sanofi-Pasteur, Medimmune, Astellas Pharma, Merck, Allergan, T2 Biosystems, Roche, Achaogen, Neumedicine, Shionogi, Pfizer, Entasis Therapeutics, Qpex Biopharma, Wellspring, Karius, and Utility Therapeutics; has received payment or honoraria for lectures, presentations, speaker's bureaus, manuscript writing, or educational events from Pfizer; and has played a leadership or fiduciary role for other board, society, committee, or advocacy group for the British Society for Antimicrobial Chemotherapy (BSAC). V. G. F. reports grants to his institution from NIH, MedImmune, Allergan, Pfizer, Advanced Liquid Logics, Theravance Biopharma, Novartis, Merck, Medical Biosurfaces, Locus Biosciences, Affinergy, Contrafect, Karius, Genentech, Regeneron Pharmaceuticals, Basilea, and Janssen Pharmaceuticals; consulting fees from Novartis, Debiopharm, Genentech, Achaogen, Affinium Pharmaceuticals, Medicines Co, MedImmune, Bayer, Basilea, Affinergy, Janssen Pharmaceuticals, Contrafect, Regeneron Pharmaceuticals, Destiny Pharma, Amphliphi Biosciences, Integrated Biotherapeutics, C3J Therapeutics, Armata Pharmaceuticals, Valanbio Therapeutics, Akagera Medicines, Aridis Pharmaceuticals, Roche, and Pfizer; royalties from UpToDate; received support for attending meetings and/or travel from Contrafect; has patents planned, issued, or pending from Sepsis Diagnostics; has stock options in Arc Bio and Valanbio Therapeutics; and received honoraria for his role as associate editor of *Clinical Infectious Diseases*. B. H. reports grants or contracts from NIH/NIAID . C. A. A. reports grants or contracts from NIH/NIAID, MeMed Diagnostics Ltd, Entasis Therapeutics, Merck, and Harris County Public Health; royalties from UpToDate; support for attending meetings and/or travel from the Infectious Diseases Society of America, American Society for Microbiology, Society of Hospital Epidemiology of America, European Society for Clinical Microbiology and Infectious Diseases, Merieux Foundation, Sociedad Argentina de Infectologia, Sociedad Chilena de Infectologia, Sociedad Colombiana de Infectologia, Panamerican Society for Infectious Diseases, and Brazilian Society for Infectious Diseases; participation on a DSMB or advisory board for the World Health Organization, Antibacterial Pipeline Advisory Group and NIH Grant Review Study Sections; being a member of the Infectious Diseases Society of America Board of Directors; and being the editor-in-chief for *Antimicrobial Agents and Chemotherapy*. T. M. reports grants from Amplyx Pharmaceuticals, Pfizer, T2 Biosystems, Cidara Therapeutics, GenMark Diagnostics, and Scynexis; consulting fees from T2 Biosystems; and honoraria for an educational event from GenMarck Diagnostics. All remaining authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2019. Atlanta, GA: US Department of Health and Human Services, **2019**.
2. 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. Lancet Infect Dis **2020**; 20:731–41.
3. Peirano G, Bradford PA, Kazmierczak KM, etal. Importance of clonal complex 258 and $IncF_{K2\text{-}like}$ plasmids among a global collection of *Klebsiella pneumoniae* with bla_{KPC}. Antimicrob Agents Chemother **2017**; 61:e02610–16.
4. van Duin D, Doi Y. The global epidemiology of carbapenemase-producing *Enterobacteriaceae*. Virulence **2017**; 8:460–9.
5. 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.
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. Perdigao J, Caneiras C, Elias R, et al. Genomic epidemiology of carbapenemase producing *Klebsiella pneumoniae* strains at a northern Portuguese hospital enables the detection of a misidentified *Klebsiella variicola* KPC-3 producing strain. Microorganisms **2020**; 8:1986.
8. Hassoun-Kheir N, Snitser O, Hussein K, et al. Concordance between epidemiological evaluation of probability of transmission and whole genome sequence relatedness among hospitalized patients acquiring *Klebsiella pneumoniae* carbapenemase-producing *Klebsiella pneumoniae*. Clin Microbiol Infect **2021**; 27:468 e1–.e7.
9. Ferrari C, Corbella M, Gaiarsa S, et al. Multiple *Klebsiella pneumoniae* KPC clones contribute to an extended hospital outbreak. Front Microbiol **2019**; 10:2767.
10. Stimson J, Gardy J, Mathema B, etal. Beyond the SNP threshold: identifying outbreak clusters using inferred transmissions. Mol Biol Evol **2019**; 36:587–603.
11. Hamilton WL, Tonkin-Hill G, Smith ER, et al. Genomic epidemiology of COVID-19 in care homes in the east of England. Elife **2021**; 10:e64618.
12. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **2014**; 30:2114–20.
13. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **2012**; 19:455–77.
14. Gurevich A, Saveliev V, Vyahhi N, etal. QUAST: quality assessment tool for genome assemblies. Bioinformatics **2013**; 29:1072–5.
15. Roosaare M, Vaher M, Kaplinski L, et al. Strainseeker: fast identification of bacterial strains from raw sequencing reads using user-provided guide trees. PeerJ **2017**; 5:e3353.
16. Seemann T. *mlst*, **Github** https://github.com/tseemann/mlst.
17. 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:e000196.
18. Seemann T. *Abricate*, **Github** https://github.com/tseemann/abricate.
19. Hunt M, Mather AE, Sanchez-Buso L, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom **2017**; 3:e000131.
20. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **2012**; 67:2640–4.
21. Seemann T. *snippy*, **Github** https://github.com/tseemann/snippy.
22. Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res **2016**; 44:W16–21.
23. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. Genome Biol **2004**; 5:R12.
24. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res **2015**; 43:e15.
25. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **2014**; 30:1312–3.
26. Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinformatics **2020**; 69:e96.
27. Mathers AJ, Stoesser N, Sheppard AE, et al. *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome sequencing. Antimicrob Agents Chemother **2015**; 59: 1656–63.
28. Campbell F, Strang C, Ferguson N, etal. When are pathogen genome sequences informative of transmission events? PLoS Pathog **2018**; 14:e1006885.
29. Lin MY, Lyles-Banks RD, Lolans K, et al. The importance of long-term acute care hospitals in the regional epidemiology of *Klebsiella pneumoniae* carbapenemase-producing *Enterobacteriaceae*. Clin Infect Dis **2013**; 57:1246–52.
30. van Duin D, Cober E, Richter SS, et al. Residence in skilled nursing facilities is associated with tigecycline nonsusceptibility in carbapenem-resistant *Klebsiella pneumoniae*. Infect Control Hosp Epidemiol **2015**; 36:942–8.
31. Naas T, Cuzon G, Villegas MV, etal. Genetic structures at the origin of acquisition of the beta-lactamase bla KPC gene. Antimicrob Agents Chemother **2008**; 52: 1257–63.
32. Diago-Navarro E, Chen L, Passet V, et al. Carbapenem-resistant *Klebsiella pneumoniae* exhibit variability in capsular polysaccharide and capsule associated virulence traits. J Infect Dis **2014**; 210:803–13.
33. Sheppard AE, Stoesser N, Wilson DJ, et al. Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene *blaKPC*. Antimicrob Agents Chemother **2016**; 60:3767–78.
34. Marsh JW, Mustapha MM, Griffith MP, et al. Evolution of outbreak-causing carbapenem-resistant *Klebsiella pneumoniae* ST258 at a tertiary care hospital over 8 years. mBio **2019**; 10:01945–19.
35. Snitkin ES, Zelazny AM, Thomas PJ, et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med **2012**; 4:148ra116.
36. 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.
37. Poon AF. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. Virus Evol **2016**; 2:vew031.
38. Hawken SE, Yelin RD, Lolans K, et al. Threshold-free genomic cluster detection to track transmission pathways in health-care settings: a genomic epidemiology analysis. Lancet Microbe **2022**; 3:e652–62.
39. Cella E, Ciccozzi M, Lo Presti A, et al. Multi-drug resistant *Klebsiella pneumoniae* strains circulating in hospital setting: whole-genome sequencing and Bayesian phylogenetic analysis for outbreak investigations. Sci Rep **2017**; 7:3534.
40. Sypsa V, Psichogiou M, Bouzala GA, etal. Transmission dynamics of carbapenemase-producing *Klebsiella pneumoniae* and anticipated impact of infection control strategies in a surgical unit. PLoS One **2012**; 7:e41068.
41. Feldman N, Adler A, Molshatzki N, et al. Gastrointestinal colonization by KPC-producing *Klebsiella pneumoniae* following hospital discharge: duration of carriage and risk factors for persistent carriage. Clin Microbiol Infect **2013**; 19: E190–6.
42. Mathers AJ, Vegesana K, German Mesner I, et al. Intensive care unit wastewater interventions to prevent transmission of multispecies *Klebsiella pneumoniae* carbapenemase-producing organisms. Clin Infect Dis **2018**; 67:171–8.
43. Hota B. Contamination, disinfection, and cross-colonization: are hospital surfaces reservoirs for nosocomial infection? Clin Infect Dis **2004**; 39:1182–9.
44. Livorsi DJ, Chorazy ML, Schweizer ML, et al. A systematic review of the epidemiology of carbapenem-resistant *Enterobacteriaceae* in the United States. Antimicrob Resist Infect Control **2018**; 7:55.
45. Bulens SN, Reses HE, Ansari UA, et al. Carbapenem-Resistant Enterobacterales in individuals with and without health care risk factors—emerging infections program, United States, 2012–2015. Am J Infect Control **2022**; 7:S0196–6553(22) 00336–4.