Title:

A novel threshold-independent approach to genomic cluster analysis discloses persistent routes of KPC+ *Klebsiella pneumoniae* transmission in a long-term acute care hospital

Authors:

Shawn E. Hawken, PhD[1]

Rachel D. Yelin, MPH[3]

Karen Lolans, BS[3]

Robert A. Weinstein, MD[3]

Michael Y. Lin, MD[3]

Mary K. Hayden, MD[3]

Evan S. Snitkin, PhD[1,2]

and for the CDC prevention Epicenters program

Affiliations:

Departments of Microbiology & Immunology[1] and Internal Medicine Division of Infectious Diseases[2] University of Michigan Medical School, Ann Arbor, MI, USA. Department of Internal Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, IL, USA[3]

Co-corresponding authors:

Evan Snitkin, PhD

1520D MSRB I

1150 W. Medical Center Dr.

Ann Arbor, MI, 48109-5680

Tel- (734) 647-6472

Fax- 734-615-5534

Email- esnitkin@umich.edu

Mary Hayden, MD

Rush University Medical Center

1653 W. Congress Pkwy

Chicago, IL 60612

Phone: (312) 942-8727

Fax: (312) 942-6787

Email- mhayden@rush.edu

Abstract

Background

Carbapenem-resistant Enterobacteriaceae have been recognized as an urgent antibiotic resistance threat for more than a decade. Despite this attention, their prevalence has remained steady or increased in some settings, suggesting that transmission pathways remain uncontrolled by current prevention strategies. We hypothesized that these transmission pathways, and hence targets for improved prevention, could be elucidated through comprehensive patient sampling, followed by integration of whole-genome sequencing (WGS) and epidemiological data.

Methods

Longitudinal KPC+ *Klebsiella pneumoniae* (KPC-Kp) surveillance cultures were collected from 94% of patients in a long-term acute care hospital (LTACH) during a one-year bundled intervention to reduce KPC-Kp prevalence. WGS of 462 KPC-Kp isolates from 256 patients, and associated surveillance data were integrated using a distance threshold-free approach to identify transmission clusters that grouped patients acquiring KPC-Kp in the LTACH with the admission-positive 'index' patients that imported their strain into the facility. Plausible transmission pathways within clusters were identified using patient location data.

Findings

Transmission clusters (N=49) had between 2-14 patients, capturing KPC-Kp acquisitions from 100 (80%) patients who first acquired KPC-Kp in the LTACH. Within-cluster genetic diversity varied from 0-154 (median 9) single-nucleotide variants (SNVs), with elevated diversity being driven by prolonged asymptomatic colonization and evolution of hypermutator strains. Transmission between patients in clusters could be explained by spatiotemporal overlap in patient rooms (14%), wards (66%), or facility (81%). Sequential exposure to the same patient

room was the only epidemiological link for one patient, indicating that residual environmental contamination of rooms after patient discharge contributed little to transmission. Persistent, modifiable routes of transmission were associated with lapses in patient cohorting, transmission between cohort and non-cohort locations and clusters propagating due to false-negative surveillance.

Interpretation

Integration of comprehensive surveillance and WGS data using a SNV threshold-free approach disclosed specific instances where improved patient and healthcare worker cohorting, reducing exposures to common locations outside of patient rooms, and improved KPC-Kp colonization detection could reduce transmission. Overall, results highlight the potential for WGS to monitor and improve infection prevention and the importance of combining rigorous sampling with appropriate analytical strategies to generate actionable hypotheses.

Funding

CDC U54 CK000481, CDC U54 CK00016 04S2. S.E.H was supported by the University of Michigan NIH Training Program in Translational Research T32-GM113900 and the University of Michigan Rackham pre-doctoral fellowship.

Research in context

Evidence before this study

We searched PubMed for studies published before August 1, 2020, with no start date restriction, with the search terms KPC* OR carbapenem-resistant AND long-term acute care OR whole genome sequencing. We identified 16 reports that used whole genome sequencing and phylogenetic analysis to investigate nosocomial transmission of KPC-producing *Klebsiella*

pneumoniae (KPC-Kp) in an endemic setting or during a prolonged outbreak (≥12 months.) Between 8 and 100 patients were included in each publication. Four studies used a single nucleotide variant (SNV) threshold (range, 16-80 SNVs) to define transmission clusters. Most studies did not seek environmental reservoirs of KPC-Kp, but 3 studies (2 by the same author) linked KPC-Kp transmission to contaminated sink drains in patient rooms.

Added value of this study

We leveraged a large, comprehensive, longitudinal sample of KPC-Kp to demonstrate the value of whole genome sequencing using a SNV threshold-free approach in the investigation of transmission pathways within a long-term acute care hospital over one year. Patients in this endemic, high prevalence setting were colonized with genetically diverse strains that would have been misclassified epidemiologically if a strict SNV threshold were applied. We established that genomic diversity was driven by prolonged asymptomatic carriage of KPC-Kp and hypermutator strains. By integrating phylogenetic and spatiotemporal data, we discovered apparent failures in infection control processes that were associated with transmission of KPC-Kp, thereby identifying opportunities for process improvement. Sequential room exposure was rarely associated with transmission, indicating little role for residual environmental contamination in KPC-Kp transmission.

Implications of all the available evidence

KPC-Kp continue to be a global public health challenge. Future analysis of comprehensive, carefully curated collections of isolates in different healthcare and community settings is needed to improve our understanding of genomic diversity within and between patients and elucidate underlying mechanisms. Integration of phylogenetic and traditional epidemiological methods is a

powerful means of revealing nosocomial transmission pathways and opportunities for infection prevention.

Introduction

Healthcare-associated infections (HAIs) affect one in 31 hospitalized patients on any given day in the U.S. and are a major threat to patient safety.[1] Despite increased attention to infection prevention in healthcare settings, cross-transmission between hospitalized patients still occurs, suggesting that there remain poorly understood pathways of nosocomial transmission.[2] Integration of genomics with traditional hospital epidemiological investigations has proved powerful in the identification of routes of HAI transmission, primarily in the acute care setting where genomic data is typically used to rule out genetically implausible transmission links.[3–5]

The majority of genomic investigations of healthcare transmission have applied single-nucleotide variant (SNV) thresholds to identify cases of likely cross-transmission within a healthcare facility.[6–11] The application of SNV thresholds has proved effective in outbreak settings, where clonal relationships and short time periods make a single SNV threshold a reasonable approach to group patients linked by transmission. However, this strategy has clear limitations in endemic and high-transmission settings where imposing a single threshold is likely to lead to both false positive and false-negative transmission inferences. One challenge that contributes to misclassification is that the prevalence of prominent HAI-causing pathogens is often geographically dominated by successful epidemic clonal lineages, making it challenging to distinguish between transmission that occurred within a study facility versus at a connected healthcare facility.[12–16]

Long-term acute care hospitals (LTACHs) are healthcare settings that have high colonization pressure for HAI-causing pathogens and are places where multi-drug resistant organisms (MDROs) that cause HAIs are often endemic.[17,18] Carbapenem resistant Enterobacteriaceae (CRE) are MDROs that are resistant to nearly all antibiotics and that cause an

estimated 13,100 HAIs and 1,100 deaths in the U.S. annually.[19] Recent work points to LTACHs having a disproportionately high prevalence of CRE and suggests that these facilities contribute to regional CRE transmission when LTACH patients are transferred between facilities in connected healthcare networks.[20–22] Given the role of LTACHs as reservoirs for MDROs like CRE, it is imperative to improve our understanding of how transmission occurs in these facilities in order to develop better methods to prevent spread.

We recently reported the effectiveness of a bundled intervention that reduced HAIs due to a particular type of CRE, *Klebsiella pneumoniae* that carry the KPC carbapenemase (KPC-Kp), in an LTACH with high KPC-Kp prevalence.[23] However, despite the overall success of the intervention, some patients still acquired KPC-Kp, suggesting that some routes of cross-transmission remained uncontrolled. Here, we took advantage of the comprehensive longitudinal surveillance culturing performed during the aforementioned intervention and the molecular resolution provided by WGS to discern the routes of transmission within the LTACH that persisted.[23] The near-complete knowledge of which patients imported and acquired KPC-Kp and the timing of these events allowed us to employ an SNV threshold-independent method to group patients related by cross-transmission of KPC-Kp in the facility (transmission clusters). Genetic analysis of transmission clusters clearly demonstrated how the imposition of an SNV-threshold would have resulted in extensive false-positive and false-negative transmission inferences. Epidemiological analysis of transmission clusters identified factors contributing to ongoing transmission, including lapses in cohorting, transmission between cohort and non-cohort locations and putative false-negative surveillance cultures leading to propagation of transmission clusters.

Methods

Study design, clinical setting, and sample collection

The current study was reviewed and approved by the Rush University Medical Center institutional review board. Informed consent was waived. Detailed information regarding the study design, intervention bundle and data collection are available in Hayden et. al 2015.[23] Briefly, the 1-year intervention to prevent KPC-Kp colonization and infection took place from 2012-2013 in a Chicago LTACH with KPC-Kp colonization prevalence of 32%. The intervention included rectal surveillance swab culture-based screening using the direct ertapenem disk method,[23,24] of all LTACH patients for KPC-Kp rectal colonization at LTACH admission and every two weeks thereafter until a patient received a positive test (94% adherence), physical separation of KPC-Kp-positive and KPC-Kp-negative patients by placing KPC-Kp-positive patients in ward cohorts (91% adherence), daily chlorhexidine bathing of all patients in the LTACH and a hand hygiene campaign.[23–25]

Patient surveillance categories

Patients were grouped into categories based on surveillance data. Patients who were either positive at the start of the study or within three days of LTACH admission were considered potential sources of KPC-Kp importation and onward transmission within the LTACH. We have termed these potential transmission sources 'Index' patients for their potential to start transmission clusters in the LTACH. 'Convert' patients converted from KPC-Kp-negative to KPC-Kp-positive after day 3 of admission. If a patient was in the facility for greater than three days before their first surveillance sampling, they were considered a convert patient for the purposes of the transmission cluster detection algorithm (see below). When an admission-

positive patient acquired an additional KPC-Kp strain (as evidenced by multi-locus sequence type (MLST) inferred from WGS data) during their stay this was termed 'Index with secondary acquisition', and such isolates from index patients were eligible to be included as convert isolates for transmission cluster detection.

Whole genome sequencing & genome processing

Glycerol stocks containing unique morphologies of KPC-Kp isolates were stored at -80°C prior to cultivation on LB agar for DNA isolation.[23,24] DNA was extracted with the MoBio PowerMag Microbial DNA kit and prepared for sequencing on an Illumina MiSeq instrument using the NEBNext Ultra kit and sample-specific barcoding. Library preparation and sequencing were performed at the Center for Microbial Systems at the University of Michigan or the University of Michigan Sequencing Core. Quality of reads was assessed with FastQC,[26] and Trimmomatic[27] was used for trimming adapter sequences and low-quality bases. In total, 462 isolates were sequenced, with the 435 isolates from 256 unique patients passing QC being used in downstream analyses. Sequence data are available under BioProject PRJNA603790.

Identification of single nucleotide variants

SNV calling was performed as in Han *et al*.[20] The variant calling pipeline can be found at https://github.com/Snitkin-Lab-Umich/variant_calling_pipeline. To summarize, raw reads were mapped to the MLST specific reference genomes listed in **Supplementary Table 1** using bwa and variant calling was performed with samtools.[28,29] MLST specific reference genomes were chosen in order to maximize the number of potential shared variants detected among genomes in the study.

Whole-genome sequence analyses

Whole-genome sequence alignments containing core and non-core variant positions were used to generate pairwise (genome by genome) single-nucleotide variant (SNV) matrices and shared-variant matrices, interrogate mutational biases, query SNVs and indels in mismatch repair genes, and construct phylogenetic trees for transmission cluster detection and descriptions of genomic variants. PanIsa was used to detect insertion sequences in bam files containing WGS alignments.[30] All whole-genome sequence analyses were performed in R version 3.6.1.

Transmission cluster detection

Transmission cluster detection using a SNV threshold-free approach was performed on isolates from MLSTs that were present in at least two patients including at least one convert patient, as this represents molecularly plausible cross-transmission within the LTACH during the study (**Table 1**). Whole-genome sequence alignments including core and non-core genome variant positions were used to generate maximum parsimony phylogenetic trees, pairwise shared variant matrices and SNV distance matrices for each MLST-specific alignment. Transmission clusters were detected by probing phylogenetic trees for the maximum subtree (**Figure 2A)** containing admission or study-start isolates from a single index patient that was collected prior to or at the same time as acquisition isolates. We required that cluster-defining subtrees to be supported by at least one unique subtree-defining variant that was not found elsewhere in the phylogeny. Multiple index patients were permitted in clusters if they shared at least one unique variant with the other cluster members. Clusters with no index isolates (acquisition isolate only clusters) were permitted if no subtree existed that included an index isolate. Only clusters that contained

isolates from at least two patients and at least one acquisition isolate were considered valid transmission clusters for downstream analyses.

Analysis of location data

Location data were abstracted from patient bed traces, i.e. patient bed and room location(s) over time. Spatiotemporal overlap explanations for cross-transmission between patients in clusters were defined as patients being in the same location (e.g. facility, ward or room) at the same time during the period between when a putative donor patient in the cluster was last negative for the isolate up until and including the day the recipient tested positive for the isolate. The last-negative date was chosen as a conservative bound for the earliest time acquisition could have occurred in order to account for acquisitions occurring between biweekly sampling dates. Sequential exposure was evaluated for the same timeframe, but restricted to patients in the same location separated by time, where the putative donor had been in a location first and the recipient later occupied the same location while they were converting from negative to positive for the isolate, and no spatiotemporal exposure between donors in the cluster and the recipient could explain the recipients' acquisition.

Statistical analysis

Two-sample Kolomogorov-Smirnov tests were used to test for a statistical difference in pairwise SNV distance distribution between admission and acquisition isolates. Multinomial tests were used to determine significant biases in mutational frequencies in transmission cluster isolates compared to overall frequencies of mutation types among all isolates of the same MLST collected in the study. The Wilcoxon rank sum test was used to detect differences in intra-patient

and intra-cluster SNV distances between admission and acquisition isolates. Permutation tests were used to evaluate enrichment in spatiotemporal and sequential exposures between patients in transmission clusters.

Role of the funding source

The funding source had no role in study design; data collection, analysis, and interpretation; or report writing. All authors had full access to all data in the study and final responsibility for the decision to submit for publication.

Results

KPC-Kp endemicity in the LTACH was due to extensive importation and acquisition within the facility of both sporadic strains and established clonal lineages

On the first day of the yearlong intervention, KPC-Kp colonization was determined for all patients in the LTACH via rectal surveillance cultures, which identified 51 patients who were positive for KPC-Kp (61 isolates). During the intervention period, admission and bi-weekly surveillance detected another 77 patients (105 isolates) who were positive within three days of first admission, and therefore presumed to have imported KPC-Kp into the facility ("index patients"). An additional 128 patients (234 isolates) were presumed to have acquired colonization in the LTACH either due to having at least one negative surveillance culture prior to a positive > 3 days after admission ("convert acquisition patients"), or having been in the facility for more than three days before having a surveillance culture taken ("putative acquisition patients"). Index patients who had additional secondary KPC-Kp isolates collected later during their stays (N=28) contributed 62 isolates, bringing the total number of isolates identified after

patients had been in the LTACH for more than three days to 296 isolates from 156 unique patients. While acquisition and importation fluctuated over time **(Figure 1B)** the overall colonization prevalence was consistently high, averaging 32% during the intervention **(Figure 1A)**.

Strain diversity and surveillance data indicate multiple distinct transmission chains in the LTACH

Examination of strains by MLST inferred from genome sequences revealed that 62% of the isolates obtained during the study belonged to ST258, the major epidemic lineage of KPC-Kp in the U.S. (**Table 1**), although other lineages were also present in smaller frequencies (**Table 1, Figure S1**). There were 424 (96% of all isolates collected) isolates from 7 different MLSTs that were found to colonize at least two patients, including at least one patient who first acquired colonization during their stay in the LTACH, demonstrating ongoing transmission of multiple distinct lineages throughout the study. The 7 lineages with putative in-LTACH transmission links were imported between 1 and 83 times each, and were the source of between 2 and 104 acquisitions over the course of the study (**Table 1**). Patients harboring these strains had extensive shared time in the facility (**Figure 2, Figure S2**), demonstrating the complexity of deciphering transmission chains in the facility.

Application of a SNV-threshold is inadequate to identify KPC-Kp cross-transmission links between patients in the LTACH with endemic KPC-Kp

We next applied the increased resolution of genomics to discern which patients were linked by cross-transmission. First, we examined the potential of applying an SNV threshold to

identify patients with isolates linked by cross-transmission that occurred in the LTACH during the study. Robust surveillance data enabled us to distinguish patients who were positive on admission, and presumably imported their colonizing strains into the facility, and patients who were negative on admission, and presumably acquired KPC-Kp via transmission from other patients in the LTACH. If an SNV threshold could be used to distinguish isolates from patients who acquired KPC-Kp colonization in the LTACH from patients who acquired colonization outside the LTACH, the isolate genetic distances among admission-positive patients (e.g. pairs not related by transmission within the facility) should be greater than the isolate genetic distances between patients who acquired KPC-Kp and their closest admission-positive patient (e.g. their putative intra-facility transmission source). However, examination of the pairwise SNV distances comparing isolates imported by admission positive patients to each other and to acquired isolates from patients who converted from KPC-Kp-negative to KPC-Kp-positive in the LTACH revealed no SNV threshold that distinguished these distributions (**Figure 3** Kolmogorov-Smirnov test p-value=0.82). These results demonstrate that an SNV threshold could not distinguish intra-LTACH transmission from importation and therefore an alternative approach is required to identify transmission links in this endemic setting that has both extensive importation of closely related strains and high rates of transmission.

Transmission clusters detected with an SNV-threshold free approach link the majority of KPC-Kp acquisitions to importation by admission positive patients

To circumvent challenges associated with applying an SNV threshold to infer transmission linkages, we took advantage of our comprehensive knowledge of which patients imported and acquired KPC-Kp and applied an algorithm whereby each acquisition isolate was

grouped with the admission isolate with which it shared the most variants. In essence, this approach groups each acquisition with the admission isolate with which it shares a most recent common ancestor (Figure S4). Application of this genomic cluster detection method yielded 49 putative transmission clusters grouping a median of 3 (range 2-14) patients into clusters representing at least one acquisition event and at least two patients (**Figure 4**). Overall, transmission clusters detected with this method grouped at least one isolate from 151 (60%) of the KPC-Kp-positive patients in the study, including 100 (80%) convert patients who acquired KPC-Kp colonization in the LTACH during the study. There were 22 convert patients who although they harbored isolates from MLSTs with plausible in-LTACH transmission, were not grouped into transmission clusters (N= 37 isolates: ST 13, 6 isolates; ST 14, 4 isolates; ST 16, 4 isolates; ST 20, 4 isolates; ST 258, 11 isolates; ST 327, 5 isolates). Transmission clusters that included these presumably acquired isolates were not detected because there was either no plausible index patient donor (isolates from all patients with ST 327, and 2 patients with ST 14 isolates) or the phylogenetic relationships between these isolates and other isolates of the same MLST did not meet the cluster detection criteria (see Methods). There were 18 (14%) study start and importation index patients whose imported isolates were not associated with onward transmission of KPC-Kp. Transmission clusters that were traced back to importation events by a single index patient (N=23 clusters) included 41 patients who acquired colonization in the LTACH, representing 26% of acquisitions captured overall in clusters. The remaining acquisitions were grouped into 26 clusters with uncertain sources of importation including 27 (17.3%) linked to multiple index patients, 42 (27%) linked to no index patient, and 26 (16.6%) linked to epidemiologically implausible index patients who were first in the facility after the convert patient had already acquired KPC-Kp colonization (**Figure 4**). Categorizing clusters

based on their overall epidemiologic coherence revealed cases where there was: 1) importation by and index patient and spatiotemporal overlap for all patients in the cluster (N = 24), 2) missing intermediate source patients (N = 11), 3) putative false-negative surveillance of an admission positive patient (N = 7), 4) multiple strain colonization by index patients (N = 5), as well as 5) unidentified source patient (N = 2, **Table 2, Figure S4**).

Transmission clusters detected with SNV-threshold independent approach range in genetic diversity

We then sought to understand the magnitude and drivers of genetic variation within clusters. Calculation of the intra-cluster diversity revealed that the maximum number of SNVs separating isolates in identified clusters ranged from 0 to 154, with a median 7 SNVs. While the majority of clusters varied by small genetic distances, 9 clusters (18 %) had larger SNV distances (greater than 30 SNVs) (**Figure 5A**). One source of large SNV distances could be the improper inclusion of index patients who are not the true source of the transmission cluster (false-positive transmission links). Indeed, we identified multiple index patients in two clusters (**Figure 5A**, cluster 258_21, 258_108) where the intra-cluster genetic diversity was >30 SNVs, suggesting that one of the admission positive patients was incorrectly included in the cluster. A second source of elevated inter-patient SNV distances could be the accumulation of genetic variation during prolonged asymptomatic colonization, and potential propagation of this variation via transmission. In support of this, we observed a distribution of intra-patient diversity among both index and convert patients who contributed multiple isolates to a cluster (**Figure 5C**). Moreover, we observed a significantly greater intra-patient diversity among admission-positive patients, (Wilcoxon rank sum test p-value < 0.03), supporting the role of prolonged colonization driving

intra-patient diversity (**Figure 5C**). In our examination of intra-patient diversity, we also observed several cases of extreme SNV distances which were inconsistent with previously reported evolutionary rates for KPC-Kp. We hypothesized that these large distances could be due to the emergence of hyper-mutator phenotypes, as has been reported for other commensal and pathogenic bacteria.[31] Genomic signatures of hypermutators include specific mutational biases as well as disruption of DNA mismatch repair genes, which can lead to greater than expected number of mutations in a given time. Analyses of these genomic signatures revealed that the transmission clusters with the largest numbers of intra cluster SNVs (cluster 16_16 - 153 SNVs and cluster 258_117 - 58 SNVs), **Figure 5B**) had clear skews in their mutational frequencies (multinomial test, p-value < 0.05) as well as a large insertion in MutS (cluster 258_117).

Two-thirds of acquisitions in clusters could be explained by spatiotemporal exposure between patients in shared ward or rooms

Then, we evaluated shared space and time relationships among patients in clusters to identify potential routes of cross-transmission in the LTACH. Transmission clusters spanned a median of 93 days (time between importation or first positive isolate and the first isolate from the last patient to be included in cluster), but the duration of clusters varied widely from 0 (all isolates being collected at the same time within a single 14-day surveillance window) to 334 days. Examining shared exposures among patients revealed that the vast majority of transmissions within clusters could be explained by spatiotemporal overlap. Considering overlap at the levels of facility, ward and room we found that 81%, 66% and 8.5% of KPC-Kp acquisitions, respectively, could be explained by overlap with another cluster member. Compared to random groups of patients of the same size and patient type distribution (e.g.

numbers of admission-positive and acquisition patients), actual transmission clusters were strongly enriched for these spatiotemporal overlaps between patients (**Figure 6** permutation tests, P-value < 0.001, all locations).

Sequential exposure to common locations was not enriched in transmission clusters

Next, we examined whether sequential exposure to common locations could explain transmission links for patients that lack spatiotemporal overlap explanations. Transmission links among patients whose exposure to a location separated by time could be a signal of prolonged contamination of the environment or other unidentified reservoir for KPC-Kp transmission that persists after a patient's stay in the LTACH. However, we found minimal support for this, as sequential exposures among patients who did not have any spatiotemporal overlap were infrequent, with only 8.5% of acquisitions across clusters explained by sequential exposure in the facility, 4.7% for sequential exposure to a ward, and 0.78% for sequential exposure to a room. The lack of explanatory power for sequential exposures was supported by permutation tests, which in contrast to spatiotemporal overlaps, were not statistically enriched compared to random clusters (permutation tests, p-value >0.60, all locations, **Figure 6**). Of special note given previous reports of sinks as a vehicle for longitudinal transmission, examination of sequential exposures to common rooms, among individual cluster patients revealed only a single patient (**Figure S5** cluster 258_175, patient 174) whose sequential exposure to a room previously occupied by another patient from their cluster was the only exposure detected that could explain their acquisition.

Genomic epidemiologic transmission cluster detection reveals testable hypotheses for how to reduce transmission that persisted during the intervention

Lastly, we sought to examine transmission clusters more holistically to gain insight into generalizable principles regarding how transmission persisted in the context of the bundled intervention. Visual inspection of transmission clusters revealed several themes that manifested across multiple clusters (**Figure 7**) that help account for persistent transmission during the intervention. These themes included: i) transmission between cohort and non-cohort locations (**Figure 7A),** ii) lapses in cohorting e.g. transmission due to housing a known positive patient in the same location as a negative patient (**Figure 7B),** iii) apparent false-negative surveillance cultures allowing clusters to propagate undetected (**Figure 7C**), iv) missing patient or environmental sources as evidenced by temporal gaps in clusters (**Figure 7D**) and v) likely exposure between cluster patients that occurred in an outside facility (**Figure 7E**). The plausible routes of transmission illustrated in these vignettes are not mutually exclusive of one another and evidence supporting multiple routes occurred in several clusters.

Discussion

Whole-genome sequencing has become the gold-standard for molecular epidemiological investigations of transmission, although there is still a lack of consensus on best practices for its application.[32] The most commonly employed approach for identifying transmission links using WGS data is to apply an SNV threshold based either on the evolutionary rate of an organism or empiric observations of the SNV distances among epidemiologically supported transmission pairs.[3,33] However, while straightforward in outbreak situations, in endemic settings with high colonization pressure—such as with KPC-Kp in LTACHs—identifying transmission links by

applying an SNV threshold could lead to erroneous inferences. [31,33,34] False-positive transmission inferences may be made when closely related isolates identified among patients in a healthcare facility are actually linked by recent transmission at a connected healthcare facility. Conversely, false-negative transmission links could be inferred due to two patients isolates harboring more variation than would be expected based on the known evolutionary rate and time between collection of isolates. To gain insight into best practices for the application of WGS to study transmission in healthcare settings we employed an SNV-threshold free approach, that relied on comprehensive surveillance culturing to group patients acquiring KPC-Kp in an endemic LTACH with the imported isolate with which it shared the most variants. Examining transmission clusters identified through this approach revealed a spectrum of genetic diversity among patients linked by transmission, and demonstrated that no single SNV threshold could accurately identify intra-facility transmission linkages. By applying our threshold-free approach to group together patients linked by transmission we were able to understand the transmission pathways that persisted despite the intervention, and generate specific hypotheses for how interventions might be improved in the future.

Our study has several strengths. First, there was comprehensive admission and subsequent in-LTACH surveillance culturing to detect KPC-Kp positive patients, which enabled our use of a SNV threshold-free approach to detect putative transmission links capturing 80% of KPC-Kp acquisitions in an endemic setting. Applying this method allowed us to gain an unbiased assessment of the genetic diversity observed among patients linked by transmission, as well as more precise knowledge of when common exposures had to occur to facilitate transmission in the LTACH. For example, we observed that while 81% of acquisitions in

clusters could be explained by spatiotemporal exposure between patients in the facility, only 66% could be explained by patients being on the same ward at the same time. This suggests that 15% of transmissions were due to exposures that originated outside of a patient's room, and points towards the need for improved infection control practices when patients visit common locations in the LTACH, or are cared for by staff members who work on more than one ward. Second, by sequencing multiple isolates from the subset of patients from whom multiple isolates were collected, we were able to gain insight into the intra-patient diversity in an endemic setting. Third, since our approach enabled detection of transmission clusters with a range of genetic diversity, we were able to evaluate cluster diversity and detect transmission links involving hypermutators, which would have been ruled-out by a SNV-threshold based approach. Fourth, the enrollment of virtually all patients and access to detailed location data over time allowed us to comprehensively explore when and where transmission in the LTACH was occurring.

Finally, and perhaps most significantly, our data collection and analytic approach provided a comprehensive assessment that disclosed the ways in which transmission pathways persisted despite an intervention that included active surveillance and patient cohorting. Specifically, we saw evidence of transmission due to lapses in the practice of cohorting, where negative patients were either housed on or moved to a cohort ward prior to KPC-Kp acquisition. And we observed several transmission clusters that included patients who were separated by time and space, indicating that there were unidentified intermediate sources of transmission. One potential source was persistent environmental contamination, such as hospital surfaces and plumbing infrastructure that have been hypothesized as potential reservoirs for MDROs including KPC-Kp.[35–40] However, our analysis revealed that only a single patient acquisition could be explained by sequential exposure to a common room that was previously occupied by a

donor in their cluster, importantly suggesting that persistent environmental reservoirs in patient rooms contributed minimally to transmission in this setting. While our lack of detailed exposure data outside of patient room and ward prevented us from identifying alternative reservoirs, we note that our approach, which identifies groups of patients with unexplained transmission, is powerful in highlighting patients between whom to look for additional common exposures. Future studies, if applied in real-time, could employ a similar approach to understand the role of devices, shared time in a common area such as physical therapy or procedure rooms, and shared exposures to staff members that care for patients in multiple locations, such as respiratory therapists and physicians. Finally, our identification of several large clusters that lacked an observed index patient supports a potential role for false-negative surveillance culturing in the propagation of transmission clusters. Thus, efforts to improve the sensitivity of surveillance cultures could improve the effectiveness of similar interventions.

Our study has several limitations related to biases in sampling. First, although we sampled 94% of patients in the facility during the yearlong study, only a single or small number of colonies (representative unique morphologies) were collected and sequenced per patient, and patients were not re-surveilled systematically once identified to carry KPC-Kp. Therefore, we may have missed cases where a patient imported multiple strains into the facility, or where a patient acquired a secondary KPC-Kp strain later in their hospital stay. Either of these limitations could potentially account for some of the cases where we were unable to identify a cluster index patient. Second, our lack of knowledge of where patients were prior to admission to the LTACH prevents us from understanding how transmission at connected healthcare facilities influenced grouping of patients in clusters. We hypothesize that these transmission events outside of the facility accounted for cases where admission-positive patients were not the first members of their

cluster to test positive. Third, there is an inherent limit of detection of surveillance culturing, which is likely associated with variation in the density of KPC-Kp colonization in the gut. Thus, false-negative surveillance cultures could have resulted in patients who imported KPC-Kp into the LTACH either being inferred to have acquired it in the facility, or remaining undetected over the course of the study. These cryptic cases could account for clusters lacking index patients and/or be the cause of clusters where not all transmission could be explained by spatiotemporal exposures between patients. Fourth, we studied patients in a single LTACH with high prevalence of KPC-Kp. Results may not be generalizable to other lower prevalence settings.

Overall, our results highlight the potential for WGS to monitor and improve infection prevention when combined appropriate sampling and analytical strategies that are jointly tailored to generate actionable hypotheses. The threshold-free approach applied here could be deployed with only admission and discharge surveillance culturing, although higher resolution sampling would facilitate more precise delineation of transmission pathways within clusters. Importantly, by relying on shared variants, as opposed to genetic distances, inferences should be agnostic to the specific MDRO species or strain, thereby circumventing the need for constant refinement of discriminatory criteria and facilitating clearer interpretation by healthcare epidemiologists.

Acknowledgments

We thank the patients and staff of the long-term acute-care hospital (LTACH) for their gracious participation in this study; Ali Pirani for bioinformatics support and members of the Snitkin lab and the Rush University/University of Michigan genomics working group for critical review of the manuscript.

Funding source

This work was supported by CDC U54 CK00016 04S2 and CDC U54 CK000481.

S.E.H was supported by the University of Michigan NIH Training Program in Translational Research T32-GM113900 and the University of Michigan Rackham pre-doctoral fellowship.

References

1 Magill SS, O'Leary E, Janelle SJ, *et al.* Changes in Prevalence of Health Care–Associated Infections in U.S. Hospitals. *N Engl J Med* 2018; **379**: 1732–44.

2 Healthcare-associated Infections (HAI) Progress Report | HAI | CDC. http://www.cdc.gov/hai/surveillance/progress-report/index.html (accessed April 5, 2016).

3 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**: 148ra116-148ra116.

4 Köser CU, Holden MTG, Ellington MJ, *et al.* Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. *N Engl J Med* 2012; **366**: 2267–75.

5 Stoesser N, Sheppard AE, Shakya M, *et al.* Dynamics of MDR Enterobacter cloacae outbreaks in a neonatal unit in Nepal: insights using wider sampling frames and next-generation sequencing. *J Antimicrob Chemother* 2015; **70**: 1008–15.

6 Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clin Microbiol Infect* 2018; **24**: 350–4.

7 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* 2020; published online April 28. DOI:10.1016/j.cmi.2020.04.017.

8 Fontana C, Angeletti S, Mirandola W, *et al.* Whole genome sequencing of carbapenem-resistant Klebsiella pneumoniae: evolutionary analysis for outbreak investigation. *Future Microbiol* 2020; **15**: 203–12.

9 Ferrari C, Corbella M, Gaiarsa S, *et al.* Multiple Klebsiella pneumoniae KPC Clones Contribute to an Extended Hospital Outbreak. *Front Microbiol* 2019; **10**. DOI:10.3389/fmicb.2019.02767.

10 van Dorp L, Wang Q, Shaw LP, *et al.* Rapid phenotypic evolution in multidrug-resistant Klebsiella pneumoniae hospital outbreak strains. *Microb Genomics* 2019; **5**. DOI:10.1099/mgen.0.000263.

11 Jiang Y, Wei Z, Wang Y, Hua X, Feng Y, Yu Y. Tracking a hospital outbreak of KPC-producing ST11 Klebsiella pneumoniae with whole genome sequencing. *Clin Microbiol Infect* 2015; **21**: 1001–7.

12 Chen L, Mathema B, Pitout JDD, DeLeo FR, Kreiswirth BN. Epidemic Klebsiella pneumoniae ST258 Is a Hybrid Strain. *mBio* 2014; **5**: e01355-14.

13 Oliver A, Mulet X, López-Causapé C, Juan C. The increasing threat of Pseudomonas aeruginosa high-risk clones. *Drug Resist Updat* 2015; **21**: 41–59.

14 Moradigaravand D, Boinett CJ, Martin V, Peacock SJ, Parkhill J. Recent independent emergence of multiple multidrug-resistant Serratia marcescens clones within the United Kingdom and Ireland. *Genome Res* 2016; **26**: 1101–9.

15 Adler A, Hussein O, Ben-David D, *et al.* Persistence of Klebsiella pneumoniae ST258 as the predominant clone of carbapenemase-producing Enterobacteriaceae in post-acute-care hospitals in Israel, 2008–13. *J Antimicrob Chemother* 2014; : dku333.

16 Kawalec M, Pietras Z, Daniłowicz E, *et al.* Clonal Structure of Enterococcus faecalis Isolated from Polish Hospitals: Characterization of Epidemic Clones. *J Clin Microbiol* 2007; **45**: 147–53.

17 Chitnis AS, Caruthers PS, Rao AK, *et al.* Outbreak of Carbapenem-Resistant Enterobacteriaceae at a Long-Term Acute Care Hospital: Sustained Reductions in Transmission through Active Surveillance and Targeted Interventions. *Infect Control Hosp Epidemiol* 2012; **33**: 984–92.

18 Haverkate MR, Weiner S, Lolans K, *et al.* Duration of Colonization With Klebsiella pneumoniae Carbapenemase-Producing Bacteria at Long-Term Acute Care Hospitals in Chicago, Illinois. *Open Forum Infect Dis* 2016; **3**. DOI:10.1093/ofid/ofw178.

19 CDC. The biggest antibiotic-resistant threats in the U.S. Cent. Dis. Control Prev. 2019; published online Nov 14. https://www.cdc.gov/drugresistance/biggest-threats.html (accessed Dec 19, 2019).

20 Han JH, Lapp Z, Bushman F, *et al.* Whole-Genome Sequencing To Identify Drivers of Carbapenem-Resistant Klebsiella pneumoniae Transmission within and between Regional Long-Term Acute-Care Hospitals. *Antimicrob Agents Chemother* 2019; **63**: e01622-19.

21 Toth DJA, Khader K, Slayton RB, *et al.* The potential for interventions in a long-term acute care hospital to reduce transmission of carbapenem-resistant Enterobacteriaceae in affiliated healthcare facilitiesLTACH interventions reduce regional CRE. *Clin Infect Dis* DOI:10.1093/cid/cix370.

22 Snitkin ES, Won S, Pirani A, *et al.* Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant Klebsiella pneumoniae in a regional outbreak. *Sci Transl Med* 2017; **9**: eaan0093.

23 Hayden MK, Lin MY, Lolans K, *et al.* Prevention of Colonization and Infection by Klebsiella pneumoniae Carbapenemase–Producing Enterobacteriaceae in Long-term Acute-Care Hospitals. *Clin Infect Dis* 2015; **60**: 1153–61.

24 Lolans K, Calvert K, Won S, Clark J, Hayden MK. Direct Ertapenem Disk Screening Method for Identification of KPC-Producing Klebsiella pneumoniae and Escherichia coli in Surveillance Swab Specimens. *J Clin Microbiol* 2010; **48**: 836–41.

25 Haverkate MR, Bootsma MCJ, Weiner S, *et al.* Modeling Spread of KPC-Producing Bacteria in Long-Term Acute Care Hospitals in the Chicago Region, USA. *Infect Control Hosp Epidemiol* 2015; **36**: 1148–54.

26 Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed Aug 4, 2018).

27 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–20.

28 Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.

29 Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; **25**: 1754–60.

30 Treepong P, Guyeux C, Meunier A, Couchoud C, Hocquet D, Valot B. panISa: ab initio detection of insertion sequences in bacterial genomes from short read sequence data. *Bioinformatics* 2018; **34**: 3795–800.

31 Couce A, Caudwell LV, Feinauer C, *et al.* Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc Natl Acad Sci* 2017; **114**: E9026–35.

32 Gerner-Smidt P, Besser J, Concepción-Acevedo J, *et al.* Whole Genome Sequencing: Bridging One-Health Surveillance of Foodborne Diseases. *Front Public Health* 2019; **7**. DOI:10.3389/fpubh.2019.00172.

33 Stimson J, Gardy J, Mathema B, Crudu V, Cohen T, Colijn C. Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions. *Mol Biol Evol* 2019; **36**: 587–603.

34 Köser CU, Holden MTG, Ellington MJ, *et al.* Rapid Whole-Genome Sequencing for Investigation of a Neonatal MRSA Outbreak. *N Engl J Med* 2012; **366**: 2267–75.

35 Weingarten RA, Johnson RC, Conlan S, *et al.* Genomic Analysis of Hospital Plumbing Reveals Diverse Reservoir of Bacterial Plasmids Conferring Carbapenem Resistance. *mBio* 2018; **9**: e02011-17.

36 Gordon K, E A, Mathers AJ, *et al.* The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections—A Systematic Review of the Literature. *Clin Infect Dis* 2017; **64**: 1435–44.

37 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.

38 Feng Y, Wei L, Zhu S, *et al.* Handwashing sinks as the source of transmission of ST16 carbapenem-resistant Klebsiella pneumoniae, an international high-risk clone, in an intensive care unit. *J Hosp Infect* 2020; **104**: 492–6.

39 Yan Z, Zhou Y, Du M, *et al.* Prospective investigation of carbapenem-resistant Klebsiella pneumonia transmission among the staff, environment and patients in five major intensive care units, Beijing. *J Hosp Infect* 2019; **101**: 150–7.

40 Kaiser T, Finstermeier K, Häntzsch M, *et al.* Stalking a lethal superbug by whole-genome sequencing and phylogenetics: Influence on unraveling a major hospital outbreak of carbapenem-resistant Klebsiella pneumoniae. *Am J Infect Control* 2018; **46**: 54–9.

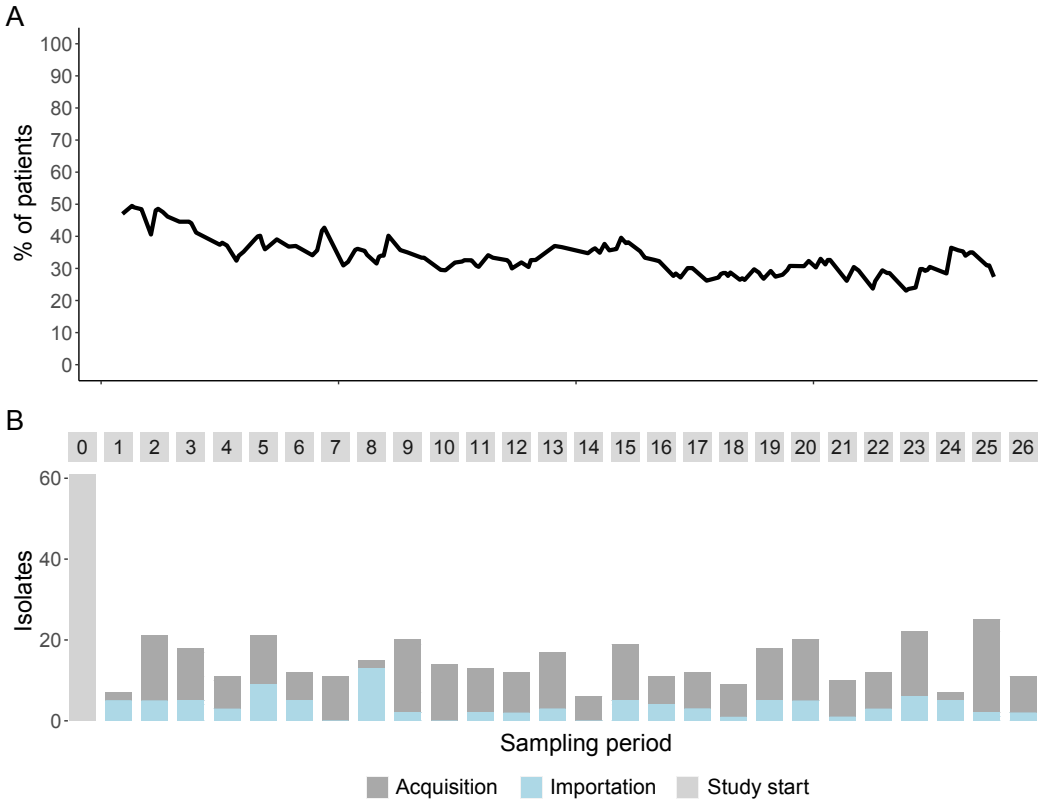

Figure 1: Endemicity of KPC-Kp in the LTACH is due to extensive importation and acquisition
A. Prevalence (black line) defined as number of patients presently in the LTACH who are or ever had been surveillance positive for at least one KPC-Kp isolate during the study divided by the number of patients in the facility throughout the 1-year study. **B.** Isolates obtained through bi-weekly rectal surveillance culturing of LTACH patients. Grey boxes indicate the study start (time 0) and every two 14-day surveillance periods (28 days) throughout the study. Bars indicate the KPC-Kp isolates collected at the beginning of the study (light grey, study start), within 3 days of the patient first entering the facility (blue, importation), or after negative surveillance or >3 days after ever being in the LTACH during the study (dark grey, acquisition).

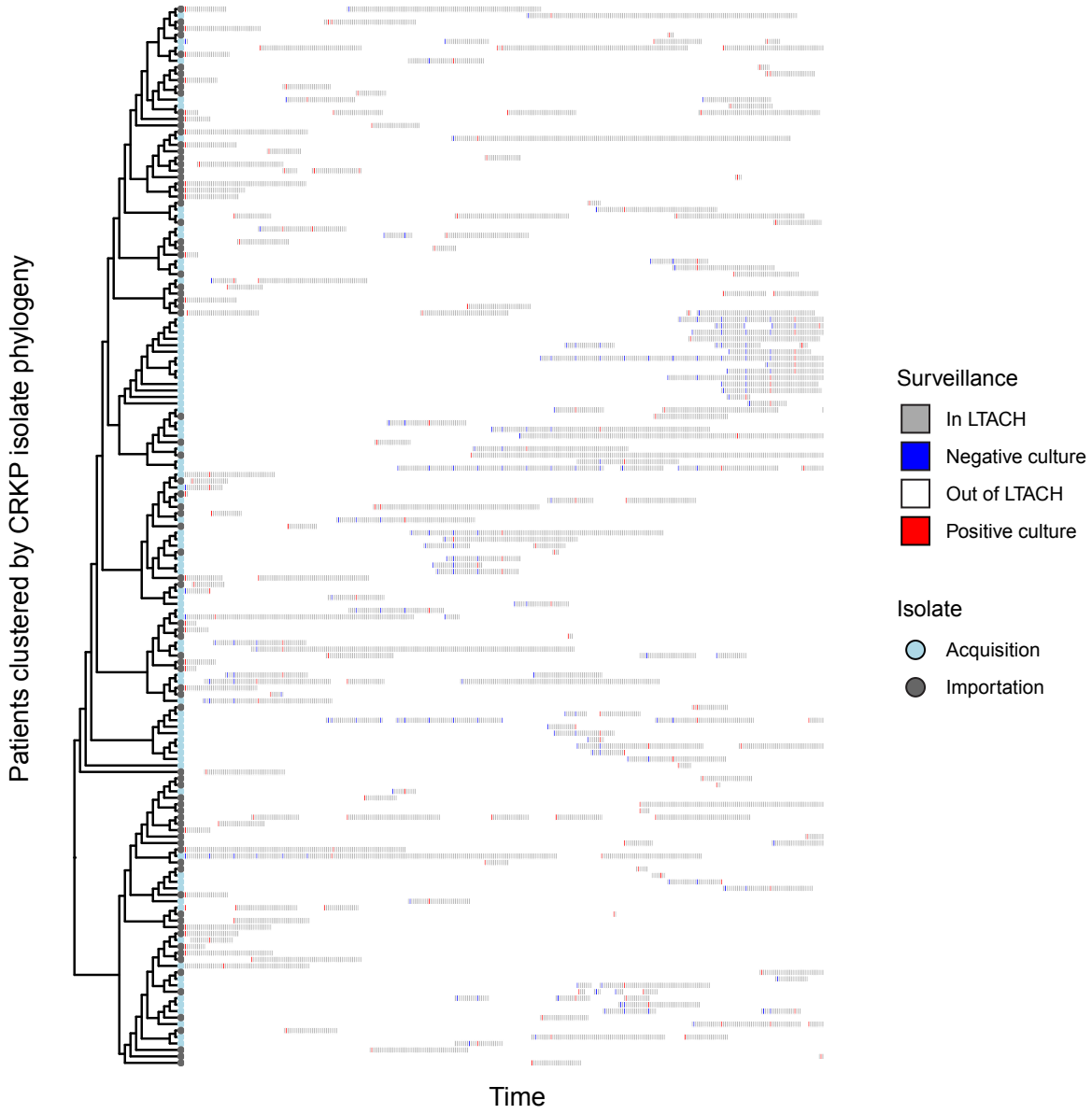

Figure 2: Tracing transmission links within the LTACH is complicated by both extensive importation and acquisition of related strains of KPC-Kp, and patients with shared time exposures in the LTACH. Patient bed trace showing surveillance culture data for patients who tested positive for KPC-Kp strain ST258, at any point during the study. The order of patients on the Y-axis is by the phylogenetic relationship between isolates collected from them. Grey bars indicate patients are in the LTACH, white indicates the patient is out outside of the facility, red indicates positive surveillance culture, blue indicates negative surveillance culture dates. Plots illustrating other MLSTs detected during the study are shown in **Figure S2**.

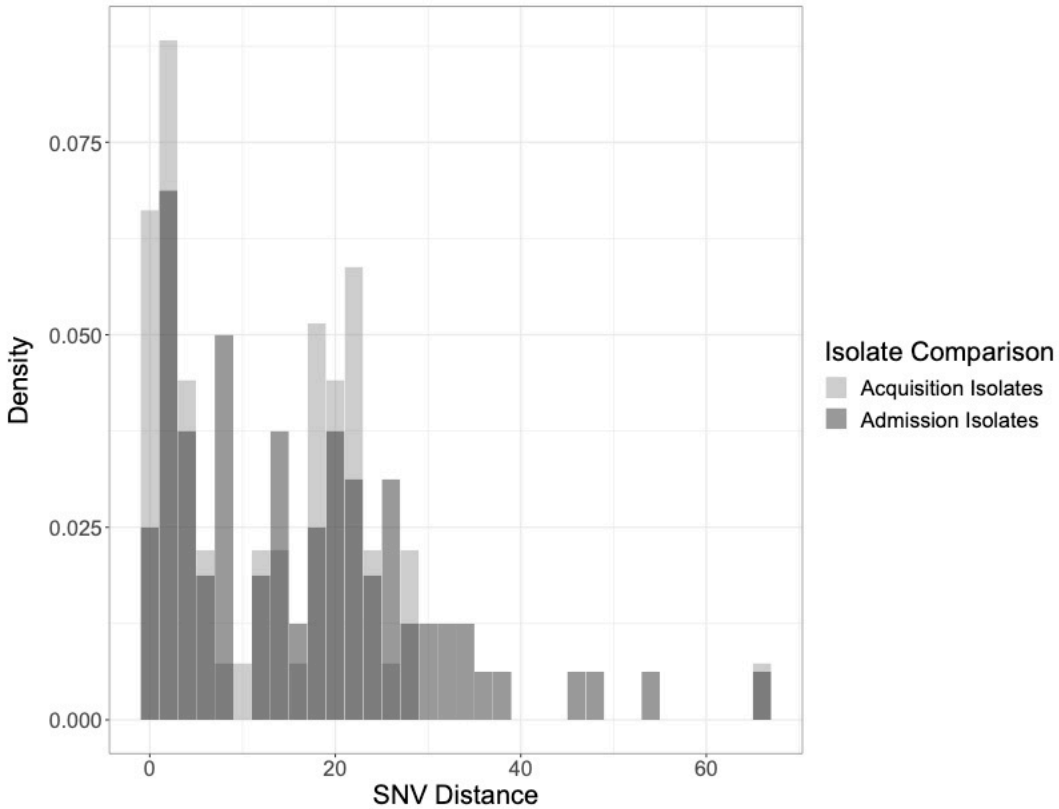

Figure 3: There is no single-nucleotide variant threshold that distinguishes isolates acquired in the LTACH from isolates that are imported by admission-positive patients.
Comparison of minimum pairwise SNV distances between the closest related imported isolate and acquired or imported isolates. X-axis indicates SNV distance, Y-axis indicates density of KPC-Kp isolates from ST258. Light grey bars indicate the minimum distance between isolates collected from patients who acquired KPC-Kp colonization after being in the LTACH >3 days during the admission positive isolates. Dark grey bars indicate the minimum distance between isolates collected from patients who were positive upon admission to the LTACH. Note that study-start positive patients who were KPC-Kp-positive on the first day of the study, who represent a mixture of recent and prior colonization, were considered admit positive for this analysis to allow for acquisitions derived from those transmission chains could be linked. Two-sample Kolmogorov-Smirnov test for differences in the distribution of pairwise SNV distances, p-value = 0.277 (ST258), and p-value=0.39 (non-ST258), 0.81 (all MLSTs combined).

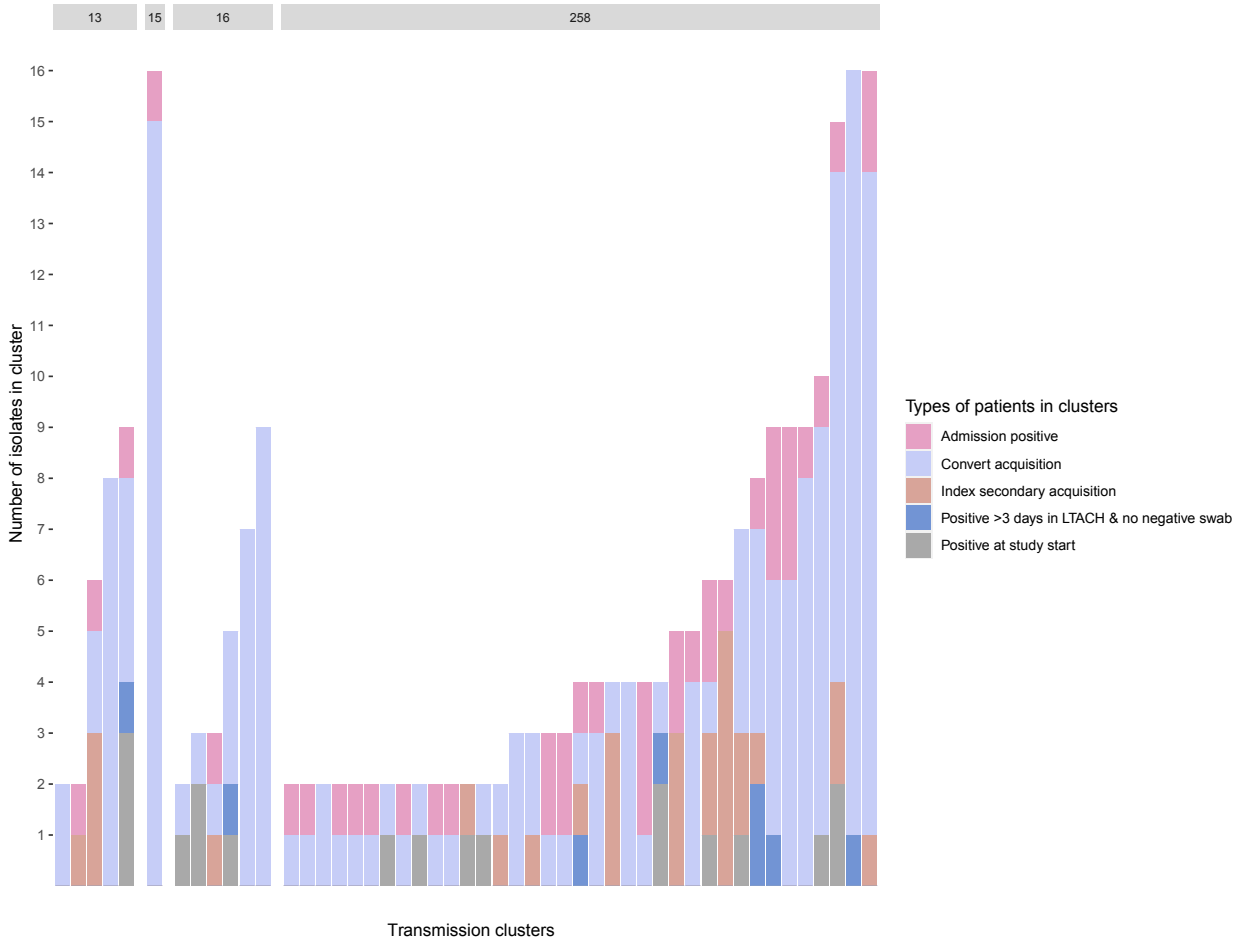

Figure 4: Transmission clusters detected with method based on shared genomic variants and robust surveillance data links the majority of KPC-Kp acquisitions during the study. Distribution of isolates and patients in the 49 transmission clusters detected with genomic method described in supplemental Figure 4. Each column represents isolates from one cluster. Admission positive patients (pink) are patients whose isolate in the clusters was obtained within 3 days of first admission to the facility. Periwinkle indicates isolates obtained from convert patients who first acquired KPC-Kp colonization more than 3 days after first admission to the LTACH. Tan indicates isolates from index patients that were collected >3 days after admission to the LTACH, indicating either prolonged colonization or secondary strain acquisition in the LATCH. Blue indicates patients who were first positive after being in the LTACH for >3 days, but from whom no negative swab was collected prior to first KPC-Kp detection. Grey bars across indicate patients who were positive on the first day of the study.

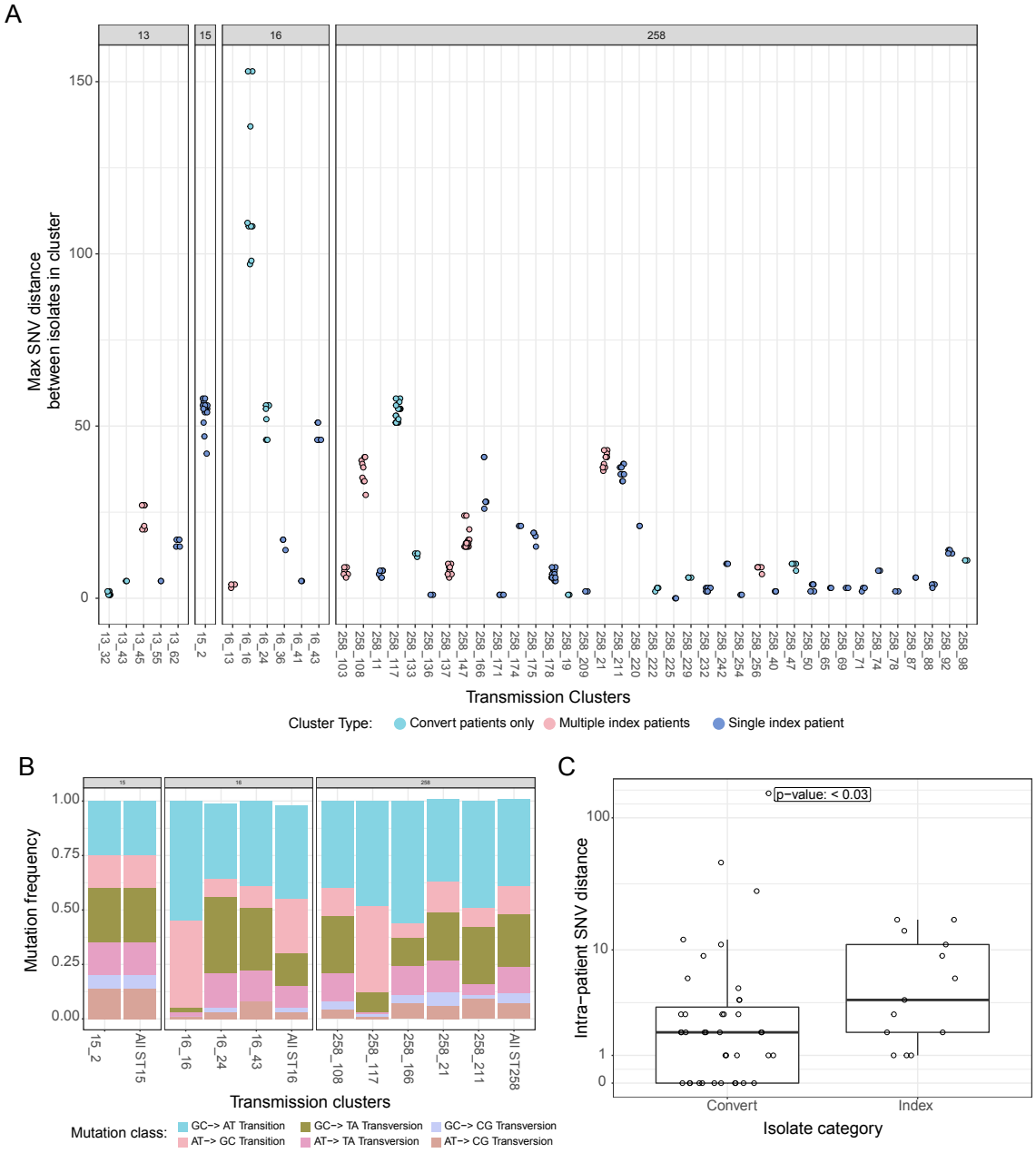

Figure 5: Elevated genetic diversity in transmission clusters is attributable to prolonged colonization and emergence of hypermutator strains. Grey bars indicate the MLST of the isolates in transmission clusters. **A.** Maximum pairwise SNV distance distinguishing isolates from the same cluster. Colors indicate whether the cluster has only convert patients, multiple index patients or a single admission-positive index patient. **B.** Observed frequencies in mutational classes across isolates included in each transmission cluster among clusters with a maximum pairwise SNV distance of 30 SNV or greater. Bars on the right of each MLST group indicate the overall population frequency of mutational classes among members of that MLST in the study population. Statistically significant skews in mutational frequencies were observed in clusters with large SNV distances (16_16 and 258_117), supporting the role of mismatch repair mutations (e.g. hypermutators). **C.** Maximum intra-patient intra-cluster genetic diversity among index and convert patients. Intra-patient intra-cluster genetic diversity is greater among isolates from index patients (Wilcoxon rank sum test, p-value < 0.03).

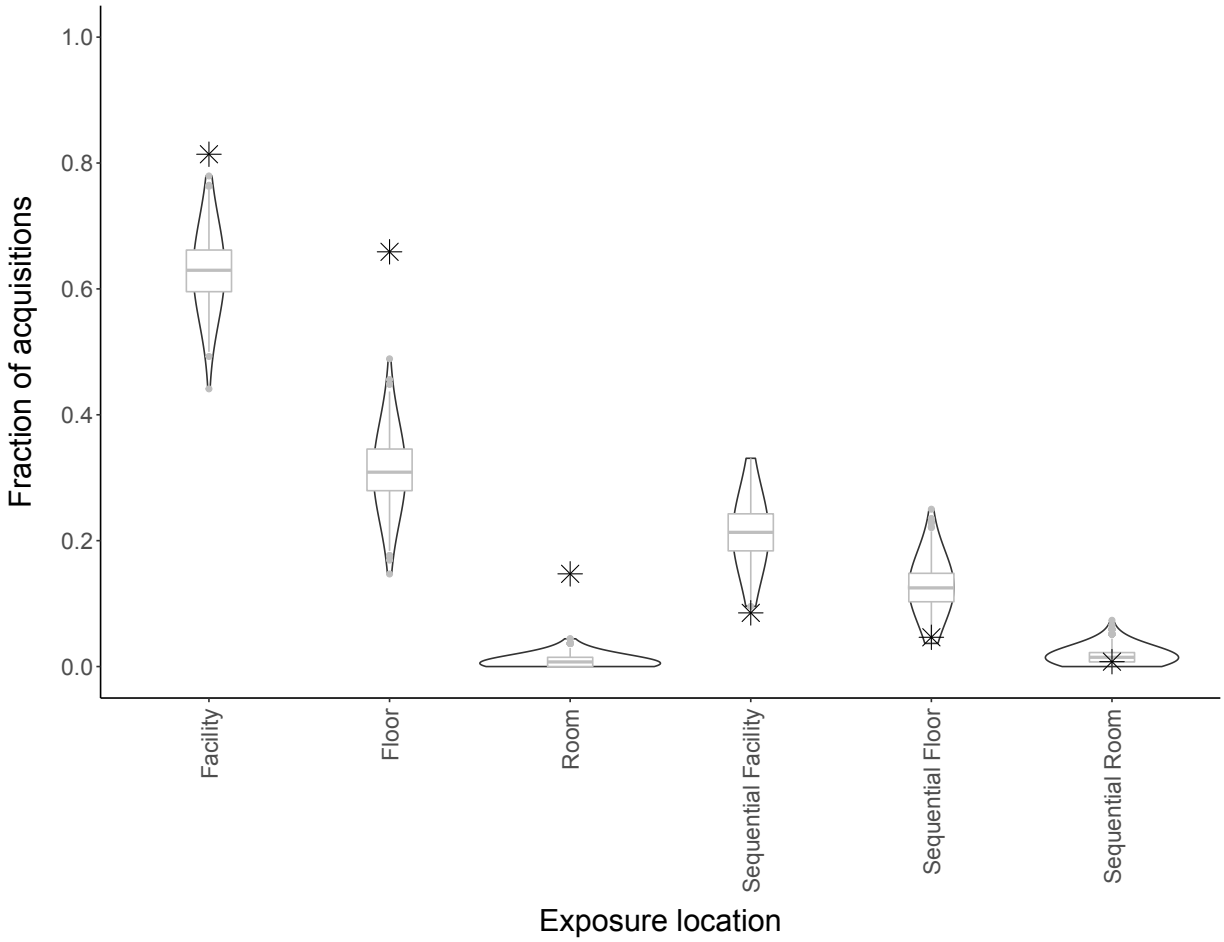

Figure 6: Epidemiologic exposures within transmission clusters point to frequent acquisition of an isolate from outside of a patient's room location (i.e., ward or facility) and infrequent acquisition linked to sequential occupation of any common location. X-axis indicates locations, Y axis indicates fraction of acquisitions in transmission clusters that could be attributed to putative donor and recipient (acquisition) patients in the cluster being in the same place at the same time (spatiotemporal exposure) or in the same place separated by time after a donor had left that location (sequential exposure). Stars indicate observed values, violins indicate exposures among permuted random transmission clusters. Spatiotemporal exposure is enriched in transmission clusters compared to permuted groups of patients of the same size and patient makeup (index and convert patients) as the observed clusters (permutation tests, $p<0.001$ for all locations). Sequential exposure is not enriched in transmission clusters compared to random clusters (permutation tests, $p>0.60$ for all locations.)

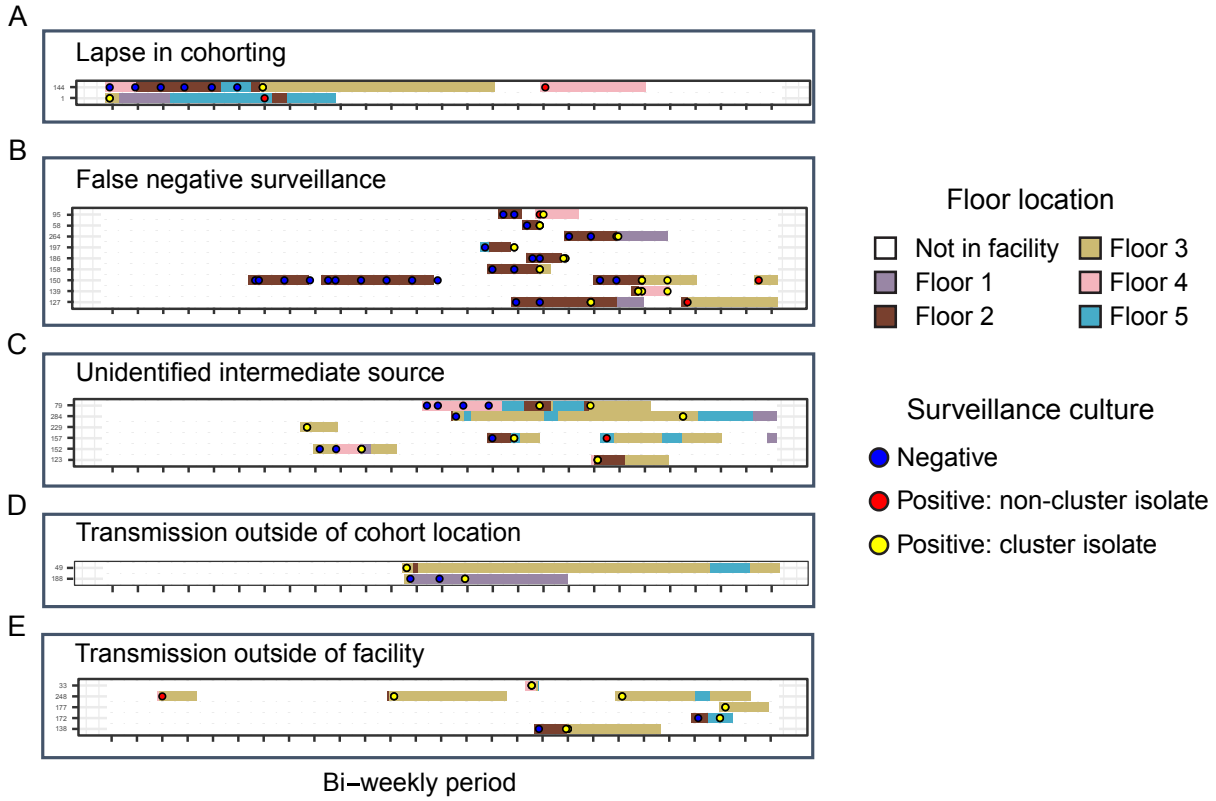

Figure 7: Descriptive vignettes from transmission clusters detected through the integration of genomic and surveillance data illustrate putative routes of uncontrolled transmission that persisted throughout the study. Patients are indicated on the y-axis and time is on the x-axis. Putative routes of transmission within each cluster is indicated in the text above the cluster. Surveillance culturing information is indicated by the circles and floor location in the LTACH is indicated by the colored rectangles. **A.** Transmission between positive index patient 1 to negative convert patient 144. Both were on the teal ward while patient 1 was positive and patient 144 was negative. **B.** No admission positive patient precedes several convert patients in this cluster, therefore false negative surveillance of a patient in the cluster or a patient not captured in the study is likely source. **C.** Lack of spatiotemporal exposures between several patients indicate missing intermediate source patient undetected by surveillance culturing. **D.** Transmission between two patients who did not reside on the same ward indicates potential escape from cohort location, or transmission at a common location, or via an unidentified healthcare worker source in the facility. **E.** Multiple index patients and lack of spatiotemporal exposures in the LTACH indicates potential transmission outside of the facility.

Table 1: Distribution of KPC-Kp strains isolated from colonized LTACH patients.

MLST	Number of isolates	Number of patients	Number of index isolates	Number of patients with index isolates	Number of convert isolates	Number of convert patients	Number of index patients with acquisitions occurring afterwards	Number of convert patients potentially acquiring colonization from a plausible index source	Potential cross-transmission link in the LTACH
13	63	37	25	18	38	22	18	15	yes
14	7	5	2	1	5	4	1	2	yes
15	17	11	1	1	16	10	1	7	yes
16	47	32	15	15	32	20	15	17	yes
20	6	6	1	1	5	5	1	4	yes
36	1	1	1	1	0	0	0	0	no
134	1	1	1	1	0	0	0	0	no
193	2	2	2	2	0	0	0	0	no
258	271	177	101	83	170	104	83	86	yes
327	6	4	0	0	6	4	0	0	yes
834	2	1	2	1	0	0	0	0	no
874	7	5	4	3	3	2	2	1	yes
883	1	1	1	1	0	0	0	0	no
950	3	1	0	0	3	1	0	0	no
Novel	1	1	1	1	0	0	0	0	no

*Possible cross-transmission link in LTACH during study inferred by at least two patients with isolate of the same MLST and at least one patient converting from negative to positive for colonization with an isolate of the same MLST.
†Isolate total represents isolates with quality WGS data; 27 samples were excluded from the 462 total isolates obtained due to poor sequence quality.

Table 2. Epidemiologic categorization of transmission clusters

Cluster category	Description	Number of clusters (percentage of clusters) *
Patient to patient transmission	Importation by index and spatiotemporal overlap explanations for all converts	24 (49)
Missing intermediate source	Lacking spatiotemporal exposure explanation for at least one convert	11 (22)
Multiply colonized index patient	Index patient is the first to test positive for an isolate that is not included in the cluster, and spatiotemporal overlap explanations for all except one convert.	5(10)
False negative surveillance	Index patient is not the first patient to test positive in the cluster, but all but one convert has spatiotemporal exposure explanations. One convert is permitted to account for false negative surveillance.	7 (14)
Missing source	Clusters with unclear source, that do not fit into above categories	2(4)
Total Transmission clusters		49

* all transmissin clusters are shown in PDF figure S6.