



Published in final edited form as:

*Sci Transl Med.* 2014 September 17; 6(254): 254ra126. doi:10.1126/scitranslmed.3009845.

## Single molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae

Sean Conlan<sup>1</sup>, Pamela J. Thomas<sup>2</sup>, Clayton Deming<sup>1</sup>, Morgan Park<sup>2</sup>, Anna F. Lau<sup>3</sup>, John P. Dekker<sup>3</sup>, Evan S. Snitkin<sup>1</sup>, Tyson A. Clark<sup>4</sup>, Khai Luong<sup>4</sup>, Yi Song<sup>4</sup>, Yu-Chih Tsai<sup>4</sup>, Matthew Boitano<sup>4</sup>, Jyoti Gupta<sup>2</sup>, Shelise Y. Brooks<sup>2</sup>, Brian Schmidt<sup>2</sup>, Alice C. Young<sup>2</sup>, James W. Thomas<sup>2</sup>, Gerard G. Bouffard<sup>2</sup>, Robert W. Blakesley<sup>2</sup>, NISC Comparative Sequencing Program<sup>2</sup>, James C. Mullikin<sup>2</sup>, Jonas Korlach<sup>4</sup>, David K. Henderson<sup>3</sup>, Karen M. Frank<sup>3,\*</sup>, Tara N. Palmore<sup>3,\*</sup>, and Julia A. Segre<sup>1,\*</sup>

<sup>1</sup>National Human Genome Research Institute, Bethesda, MD

<sup>2</sup>National Institutes of Health Intramural Sequencing Center (NISC), Bethesda, MD

<sup>3</sup>National Institutes of Health Clinical Center, Bethesda, MD

<sup>4</sup>Pacific Biosciences, Menlo Park, CA

### Abstract

Public health officials have raised concerns that plasmid transfer between Enterobacteriaceae species may spread resistance to carbapenems, an antibiotic class of last resort, thereby rendering common healthcare-associated infections nearly impossible to treat. We performed comprehensive surveillance and genomic sequencing to identify carbapenem-resistant Enterobacteriaceae in the NIH Clinical Center patient population and hospital environment in order to articulate the diversity of carbapenemase-encoding plasmids and survey the mobility of and assess the mobility of these plasmids between bacterial species. We isolated a repertoire of carbapenemase-encoding Enterobacteriaceae, including multiple strains of *Klebsiella pneumoniae*, *Klebsiella oxytoca*, *Escherichia coli*, *Enterobacter cloacae*, *Citrobacter freundii*, and *Pantoea* species. Long-read

\*To whom correspondence should be addressed. karen.frank@nih.gov (K.M.F); tpalmore@mail.nih.gov (T.N.P.); jsegre@nhgri.nih.gov (J.A.S.).

#### Supplementary Materials

Materials and Methods

Fig. S1. Plasmid composition of three KPC<sup>+</sup> *K. pneumoniae* from 2013.

Table S1. Assembly metrics.

Table S2. Sequence Data Summary.

Table S3. PacBio and MiSeq comparison.

Table S4. Plasmid copy number.

**Author contributions:** SC, JCM, RWB, JK, DKH, KMF, TNP and JAS conceived of the study. DKH and TNP oversaw hospital epidemiology. AFL, JPD and KMF oversaw clinical microbiology. PJT, MP, TAC, KL, YS, Y-CT, MB, JG, SYB, BS, ACY, JWT, GGB, RWB, JCM, JK and NISC performed genome sequencing and analysis. CD performed molecular analyses. SC and ESS performed data analysis. SC, DKH, KMF, TSP and JAS wrote the manuscript.

**Competing interests:** TAC, KL, YS, Y-CT, MB and JK are employees of Pacific Biosciences, a company commercializing DNA sequencing technologies.

**Data and materials availability:** All sequence data can be retrieved associated with NCBI BioProject PRJNA251756. Isolates can be requested from KMF.

genome sequencing with full end-to-end assembly revealed that these organisms carry the carbapenem-resistance genes on a wide array of plasmids. *Klebsiella pneumoniae* and *Enterobacter cloacae* isolated simultaneously from a single patient harbored two different carbapenemase-encoding plasmids, overriding the epidemiological scenario of plasmid transfer between organisms within this patient. We did, however, find evidence supporting horizontal transfer of carbapenemase-encoding plasmids between *Klebsiella pneumoniae*, *Enterobacter cloacae* and *Citrobacter freundii* in the hospital environment. Our comprehensive sequence data, with full plasmid identification, challenges assumptions about horizontal gene transfer events within patients and identified wider possible connections between patients and the hospital environment. In addition, we identified a new carbapenemase-encoding plasmid of potentially high clinical impact carried by *Klebsiella pneumoniae*, *Escherichia coli*, *Enterobacter cloacae* and *Pantoea* species, from unrelated patients and the hospital environment.

---

## Introduction

Carbapenem-resistant *Enterobacteriaceae* (CRE) are formidable Gram-negative bacterial pathogens that pose a serious triple threat to hospitalized patients around the globe. First, CRE are resistant to most, if not all, antibiotics, with investigations reporting as high as a 40–80% mortality rate from infection (1–4). Second, the incidence of CRE in the United States has quadrupled over the past decade, with isolates reported from nearly every state and detected in 3.9% of hospitals and 17.8% of long-term acute care facilities (5). CRE strains are transmitted easily in healthcare settings from patients who have asymptomatic intestinal colonization (4). Third, CRE have the potential to spread antibiotic resistance through plasmid transfer to other bacterial species, including common human flora and potential pathogens such as *Escherichia coli* (1, 4).

In 1928, Griffith demonstrated that a non-encapsulated pneumococcal strain could be transformed into a virulent, encapsulated type by incubation with heat killed encapsulated pneumococci (6), a phenomenon later confirmed to occur through DNA transfer. In the 86 years since this discovery, plasmids have been recognized as important mediators of horizontal gene transfer between bacteria and have been exploited extensively as tools for molecular biology. Whereas the mechanism of plasmid transfer in model organisms and laboratory strains has been elucidated, much less is known about clinically relevant plasmid “trafficking” in the hospital setting where the environment, patients, and healthcare personnel may each serve as reservoirs. Understanding horizontal gene transfer among *Enterobacteriaceae* is particularly important since the majority of carbapenem resistance genes are found on plasmids (7). Carbapenemase-mediated resistance in the United States is most commonly acquired through the plasmid-encoded beta lactamase *bla<sub>KPC</sub>* (KPC) gene. Although evidence-based guidelines to prevent transmission of CRE exist (8), they do not explicitly take into account the mobility and potential spread of the *bla<sub>KPC</sub>* gene through plasmid transfer, which may complicate tracking nosocomial transmission.

Comparative genomics has been utilized to track the dissemination and transmission of a wide variety of antibiotic-resistant bacterial pathogens (9–13). SNV analysis has been used to characterize the global spread of pathogens (14, 15) as well as localized outbreaks (4, 16–

18). However, the vast majority of these studies were based on shotgun assembled sequence data or read mapping alone, and thus were not able to discriminate and resolve plasmid from chromosomal encoded genes.

In 2011, a patient colonized with KPC+ *Klebsiella pneumoniae* was transferred to the NIH Clinical Center, after which 17 additional patients became colonized, and six severely ill patients died from bloodstream infections due to this multidrug-resistant organism. We traced this hospital outbreak back to three separate transmissions from the index patient (4). We deduced the most likely transmission of KPC+ *K. pneumoniae* amongst these 18 patients, integrating epidemiological data (patient location records) and genomic data (single nucleotide variants or SNVs). Co-inheritance of SNVs provided the resolution to infer transmission, as the evolutionary rate of these molecular markers is on the same timescale as nosocomial spread. SNVs in patient isolates were identified from high-quality genomic data that did not require fully assembled genomes. Although the outbreak ended in December 2011, we have continued to conduct active surveillance for CRE among patients and the hospital environment (7).

During 2012–2013, KPC+ Enterobacteriaceae belonging to six species were cultured from NIH Clinical Center patient and environmental surveillance cultures. Here, we present the genomic detective work required to identify and track plasmids within these isolates in a hospital setting, revealing a complex pattern of introduction and environmental spread of carbapenemase-encoding plasmids.

## Results

### Overview of hospital surveillance of patients and surfaces

During 2012–2013, perirectal swabs and/or throat/groin swabs were collected twice weekly from all patients in the intensive care unit and a high-risk medical ward. Once each month, surveillance cultures were ordered on every inpatient except those housed in locked mental health wards. Beginning July 2012, all patients transferred from other healthcare facilities underwent surveillance cultures upon admission for two consecutive days, and were placed on empiric contact isolation until cultures demonstrated no CRE growth. Beginning in September 2013, admission surveillance cultures were implemented for every patient admitted to non-mental health wards. Patients colonized with KPC+ organisms were cohoused in a single ward with cohorted care (19). Environmental cultures were performed when a KPC+ organism was identified in a patient in whom nosocomial acquisition was suspected. A total of 416 environmental samples and 14204 surveillance cultures from 1086 patients were collected.

Throughout 2012 and 2013, surveillance cultures identified ten patients colonized with KPC+ Enterobacteriaceae in our institution (Fig. 1 and Table 1), as well as persistent colonization among patients who had acquired the 2011 outbreak strain. Patients A–K are denoted with letters to symbolize independent introduction, whereas patient 19 is given a number to denote linkage with the 18 patients of our institution's 2011 outbreak (evidence presented below). Colonization was detected in six patients prior to or shortly after admission and in four patients after weeks in the hospital, raising concern about possible

nosocomial transmission. Two additional isolates were included in this study – these were obtained from patients D and E (Table 1) who had been admitted to our hospital, discharged, and subsequently sought medical treatment elsewhere and were found at those facilities to be colonized with KPC<sup>+</sup> *Enterobacteriaceae*. Environmental surveillance identified six additional KPC<sup>+</sup> *Enterobacteriaceae*. In total, this study includes twenty isolates: two from the 2011 NIH outbreak, ten from 2012–2013 NIH patients, two from NIH patients identified at outside institutions, and six from environmental surveillance.

### **Single molecule, real-time sequencing of 20 KPC<sup>+</sup> carbapenemase-encoding *Enterobacteriaceae* isolates**

Whole genome shotgun data, generated as ‘short-reads’ of 100 to 500 base pairs (bp) with Illumina or 454/Roche platforms, have been utilized by our group to characterize a number of hospital-associated pathogens including *K. pneumoniae* and *Acinetobacter baumannii* (4, 20). However, preliminary analysis of short-read genome sequences quickly proved the difficulty of assembling plasmids, replete with repeats and mobile genetic elements (21). Recombination, driven by mobile genetic elements, made assisted assembly of plasmids based on published references more difficult. And yet, these plasmids are 50–200 kb, too large for simple small plasmid DNA purification methods. To characterize the diversity of *bla*<sub>KPC</sub>-encoding plasmids and investigate possible horizontal transfer, we explored the utility of single molecule, real-time (SMRT) sequencing from Pacific Biosciences. SMRT sequencing long-read technology (22) produces highly accurate fully-contiguous genomes (23), both critical requirements to infer phylogenetic relationships. SMRT sequencing generates kilobase-long DNA sequences as seeds to which shorter reads are aligned and used for error correction. The resulting highly accurate, long DNA sequences are further assembled to obtain fully contiguous genomes.

In total, we generated 20 finished genome sequences for KPC<sup>+</sup> *Enterobacteriaceae*, including *Klebsiella pneumoniae*, *Klebsiella oxytoca*, *Enterobacter cloacae*, *Escherichia coli*, *Citrobacter freundii* and *Pantoea* sp. (Tables 1, 2 and S1). These 20 bacterial genomes varied from 3.9 to 6.2 Mb in chromosome size and carried between 1 and 5 plasmids ranging from 9.3 to 379 kb. Most organisms had a single copy of the *bla*<sub>KPC</sub> gene. However, KONIH1 (patient C) carried a plasmid with 2 copies of the *bla*<sub>KPC</sub> gene, and ECNIH2 (environmental) carried 3 different plasmids, each encoding the *bla*<sub>KPC</sub> gene. KPNIH33 (patient G) had 3 copies of the *bla*<sub>KPC</sub> gene, one carried on a plasmid and two chromosomal copies. Multiple copies of the *bla*<sub>KPC</sub> gene have been found on plasmids previously and no gene dosage effect was observed (24). The genomic architecture of these isolates with multiple *bla*<sub>KPC</sub> genes is clinically relevant, but technically difficult to extract from fragmented genome data.

We independently validated assembly and nucleotide error-rates for SMRT sequencing using data from the OpGen Argus and Illumina MiSeq platforms (see Supplemental Materials) for 10 of these finished genomes. Briefly, OpGen physical maps, produced from ordered maps of restriction enzyme sites, identified one chromosomal inversion in the PacBio genome assemblies caused by repetitive flanking sequence. Manual inspection identified reads supporting the correct orientation and final chromosome sequences showed

complete concordance. To evaluate accuracy at the base pair level, independently generated MiSeq fragment reads from each isolate were aligned to the respective PacBio sequence, followed by analysis with Most Probable Genotype software (25). The final accuracy for each genome exceeded 99.9999% (Table S2).

### Genome and plasmid sequencing to investigate possible nosocomial transmission

Our 2011 outbreak strain, a sequence type (ST) 258 *K. pneumoniae*, carried the KPC-3 allele of the *bla<sub>KPC</sub>* gene inside the Tn4401a transposon on a 113.6 kb plasmid nearly identical to the published sequence of pKpQIL from a 2006 Israeli isolate (NC\_014016) (26). Aligning these two plasmid sequences, we detect only two high-confidence base pair changes, C40547A and T72669C. The C40547A change occurred in an intergenic region between convergently transcribed genes. The T72669C resulted in a silent mutation (TGT to TGC) in the 3' end of a hypothetical protein. These two SNVs were stably inherited and found in every patient isolate obtained in the 2011 outbreak. Thus, whereas these genomic changes may not be evolutionarily selected, they do provide genetic 'fingerprints' of our outbreak strain. Our dominant 2011 KPC<sup>+</sup> *K. pneumoniae* strain contained two additional plasmids, the 15.1 kb pAAC154 plasmid (27) and a 243.8 kb plasmid similar to pKN-LS6 (28). The genomes from the index patient and patient 5 provided accurate and complete references for investigating links between the 2011 outbreak and CREs subsequently detected at our institution.

In addition to possible transmission of KPC<sup>+</sup> *K. pneumoniae* between patients within our hospital, we expanded our investigations to explore whether carbapenem resistance was acquired through nosocomial *bla<sub>KPC</sub>* plasmid transfer between bacteria of the same or different species. Through these sequencing efforts we assessed the diversity and complexity of *bla<sub>KPC</sub>* carbapenem-resistance encoding plasmids. To identify hospital practices associated with increased risk of transmission, we incorporated genomic sequence data with epidemiologic data such as time from admission, previous negative surveillance cultures, and ward location. First, we discuss five cases of patients colonized with KPC<sup>+</sup> *Klebsiella* sp. in which genomic data confirmed or refuted nosocomial transmission.

Patient A was identified as colonized with KPC<sup>+</sup> *K. pneumoniae* 70 days into his hospital stay after 19 negative perirectal surveillance cultures. Colonization was identified while the patient was in the intensive care unit after receiving prolonged broad-spectrum antibiotics, just two weeks after we detected the last patient of the 2011 outbreak. Surveillance data put this case into a category for high suspicion of nosocomial transmission. The isolate from patient A (KPNIH27) was ST 34, and hundreds of distributed SNVs distinguished this genome from our outbreak strain (ST 258). This isolate's carbapenem resistance was encoded by a *bla<sub>KPC-2</sub>* gene, carried on a 268.3 kb plasmid (pKEC-dc3) that bears no similarity to pKpQIL, and was not flanked by a Tn4401 transposon (Table 2). Genomic data did not support nosocomial transmission from an outbreak patient, and yet the origin of this patient's *bla<sub>KPC</sub>* encoding plasmid remains unknown. This case may represent (1) low-level patient colonization on admission that eluded the many perirectal surveillance cultures until selected out by broad-spectrum antibiotics, (2) transmission from the environment, or (3) possible transmission from individuals interacting with the patient, such as visitors or

healthcare personnel, who are not included in routine surveillance in our institution. This patient's isolate could have been characterized with techniques less powerful than complete genome sequencing, as multiple molecular diagnostics demonstrate KPNIH27's dissimilarity to the dominant outbreak strain. However, each sequenced isolate expands the currently limited public database of plasmid sequences and enables the development of future, more accurate molecular diagnostics.

Patient B had intestinal colonization with KPC<sup>+</sup> *Klebsiella pneumoniae* detected by a surveillance culture collected three days after transfer to our facility from another hospital. As this patient was new to our institution, epidemiologic evidence pointed to acquisition of the isolate at another healthcare facility. However, because of concern for possible transmission of the 2011 outbreak strain, we investigated this strain with complete genome sequencing. The outbreak strain (KPNIH1) and this patient's strain (KPNIH24) shared some similarities, such as classification as the predominant ST 258 (Fig. 2, A and B). Nonetheless, major differences existed at the genomic level: the bacterial chromosomes of patient 1 and patient B have distinct capsular biosynthetic locus sizes of 19,637 and 24,931 bp, respectively. In addition, there were 155 high-confidence SNVs distributed along the bacterial chromosomes of patient 1 and patient B, further supporting that the bacteria were independently introduced. Patient B's isolate contained the *bla*<sub>KPC-2</sub> gene encoded on an 85.5 kb plasmid (pKPC-484, Fig. 2B), arguing against plasmid transfer from the 2011 strain. However, molecular diagnostics targeted to the KPC gene region might have yielded conflicting data since patient B's *bla*<sub>KPC</sub> gene is flanked by ~10 kb Tn4401a transposon, embedded in a 30 kb region that is 99.9% identical to pKpQIL (Fig. 2, A and B). The case of patient B demonstrates the value of accurate whole genome sequencing to exclude definitively nosocomial transmission of a bacterial isolate and *bla*<sub>KPC</sub> plasmid.

Patient E, who received treatment at our institution during the 2011 outbreak, had multiple negative surveillance cultures at the time. He was admitted subsequently to another facility and was identified as KPC<sup>+</sup> *Klebsiella pneumoniae* colonized during that hospitalization. We requested that the outside facility send the isolate for genomic sequencing so that we could determine whether the organism might have originated in our institution. We envisioned two distinct possibilities: our surveillance culture missed low-level colonization with our outbreak strain, or the patient acquired a new organism after leaving our hospital. Similar to our outbreak strain (KPNIH1), this patient's isolate (KPR0928) is *K. pneumoniae* ST 258 with the *bla*<sub>KPC</sub> gene encoded on the pKpQIL plasmid (Fig. 2C). Standard classification and typing strategies to differentiate bacterial strains, such as pulsed-field gel electrophoresis, multilocus sequence typing, and PCR of plasmid targets, even combined with organism identification and susceptibility results, might not be able to conclude definitively whether these isolates were related. However, full genome analysis revealed that isolates from patient 1 (the index patient in the outbreak) and patient E differ at the capsule biosynthetic locus, with distinct locus sizes of 19,637 and 23,731 bp. In addition, we identified 173 high-confidence SNVs (recombinant regions excluded) between the chromosomes of KPNIH1 (patient 1) and KPR0928 (patient E). Moreover, the pKpQIL plasmid from patient E lacks the C40547A and T72669C SNVs (relative to NC\_014016) specific to our institution's outbreak, and contains an additional eight SNVs (including the

KPC-2 allele) and unique insertion/deletions in the pAAC154 plasmid. Based on these differences, we conclude that the *K. pneumoniae* strain and pKpQIL plasmid from patient E are distinct from those of our institution's 2011 outbreak.

An open question remains whether patient E's strain is circulating at the other hospital. Supporting this possibility is the genomic similarity of isolates from patient E and patient I, who was transferred to our facility in 2013 from the same medical system as patient E. Patient I was colonized with a nearly identical strain of *K. pneumoniae* (KPNIH30; Fig. 2D) as patient E, which both contained the identical plasmids pKpQIL-531 and pAAC154-a50 as well as a new plasmid encoding additional antibiotic resistance genes. The example of patients E and I highlights the value of detailed sequencing for strain tracking, and the potential of this technology to draw connections across time and space. This case exemplifies how detailed sequencing was critical for excluding nosocomial acquisition in one hospital and supporting nosocomial, but not likely direct, transmission in another.

Patient C had seven negative surveillance swabs over two months prior to detection of KPC<sup>+</sup> *K. oxytoca* from an eighth surveillance swab. This isolate (KONIH1) was detected on the 50<sup>th</sup> day of her second hospitalization. KPC<sup>+</sup> *K. oxytoca* had not been isolated previously from any other patient or environmental source in our institution. KONIH1 contains a pair of *bla*<sub>KPC-2</sub> genes in opposite orientations on a 205.6 kb plasmid (Fig. 2E), which share little sequence similarity or genomic architecture with other sequenced clinical or environmental CRE isolates. A genome assembly based on short-read Illumina MiSeq data for this isolate collapsed both copies of the *bla*<sub>KPC</sub> gene into a single chimeric 20.8 kb contig, demonstrating the greater resolution of SMRT sequencing. As in the case of patient A, the plasmid sequence ruled out horizontal gene transfer from any known colonized patients. No other patients have been identified with KPC<sup>+</sup> *K. oxytoca*. In the setting of extensive patient and environmental surveillance the KPC<sup>+</sup> *K. oxytoca* was likely present on admission and undetected by serial surveillance cultures, possibly due to bacterial titer or relative insensitivity of screening technique, or was due to an undefined transmission as discussed for patient A above. This case highlights potential limitations of surveillance techniques, and encourages careful consideration of how frequently surveillance cultures should be collected.

Patient 19 was a hospitalized stem cell transplant recipient who was identified as carrying KPC<sup>+</sup> *K. pneumoniae* from a perirectal surveillance culture in July 2013, following 26 negative cultures over 9 months. The genomic sequence of this patient's isolate was phylogenetically linked to our institute's 2011 outbreak strain (KPNIH1). Specifically, patient 19's KPC<sup>+</sup> *K. pneumoniae* isolate contained the genetic signature of SNVs that had evolved during the long-term colonization of index patient 1 and during our institute's subsequent outbreak. This genetic data clearly indicated nosocomial transmission and follow-up sequencing identified the source as a chronically colonized outbreak patient who was concomitantly receiving treatment at our hospital. An identical isolate cultured from the handrail outside the patient's room provided additional epidemiologic support for nosocomial transmission and suggested a role for contaminated hands as a possible route of transmission. Patient 19 later died of complications related to KPC<sup>+</sup> *K. pneumoniae* infection. This case is the only example from the five suspected cases that resulted in a

definitive conclusion that nosocomial transmission had occurred, again underscoring the power of genomic sequencing in the clinical setting (4).

These five cases illustrate the resolution achieved with whole genome and plasmid sequencing, and the ability to use these data in combination with clinical data to resolve individual healthcare epidemiology puzzles that have great implications for allocation of hospital infection control resources. In the single case of documented nosocomial transmission (patient 19), our institution launched an extensive investigation and heightened infection control measures aimed at containing spread from known colonized patients, such as expanding adherence monitoring and increasing the frequency of surveillance cultures. Other cases, in which sequencing demonstrated that a new isolate did not match a previously identified strain, prompted efforts to improve admission surveillance rather than tracking transmission. Given the limited resources in all healthcare facilities, sequence data can help focus and direct the use of allotted funds for infection control interventions, thus providing the best patient care.

### Transmission of plasmids in the hospital environment

Whereas the KPC<sup>+</sup> *K. pneumoniae* isolated from patient A was not linked to other patients, we did find evidence of plasmid transfer with bacteria in the hospital environment, specifically KPC<sup>+</sup> isolates identified in two sinks in this patient's room. Four days after this patient's KPC<sup>+</sup> *K. pneumoniae* isolate was identified, KPC<sup>+</sup> *Citrobacter freundii* and KPC<sup>+</sup> *Enterobacter cloacae*, were cultured from a faucet aerator and sink drain, respectively, within the patient's room. Full-genome assembly of long read SMRT sequences clarified the chromosomal and plasmid diversity of these isolates.

The *bla*<sub>KPC-2</sub> encoding p55-like plasmids from patient A (KPNIH27; pKEC-dc3) and the two sinks (CFNIH1; pKEC-ac3 and ECNIH2; pKEC-39c) differ only by a very small number of insertion/deletion events (Fig. 2, F, G and H, Fig. 3), consistent with horizontal plasmid transfer between different genera of *Enterobacteriaceae*. Relative to pKEC-dc3, the *C. freundii* *bla*<sub>KPC-2</sub> encoding plasmid pKEC-ac3 has an 11 bp deletion upstream of the *stbA* gene and an expanded class 1 integron. The *E. cloacae* *bla*<sub>KPC-2</sub> encoding plasmid pKEC-39c carries the same genetic alterations as the *C. freundii* pKEC-ac3 and an additional 47.7 kb insertion encoding over 80 genes, including an iron transport system. Acquisition of a similar set of genes, including an iron uptake system, has been observed in other plasmids (29) where it has been postulated to improve fitness by facilitating iron uptake in the human host or environment. No SNVs exist across the aligned portions of these plasmid genomes, precluding definitive conclusions about the transmission order between the patient and the environment. The simplest explanation would be that the KPNIH27 isolate was a transient member of the complex microbial community of the sink biofilms.

Epidemiologic data would suggest that the patient's isolate colonized the environment, and there is no evidence that these sink biofilm isolates ever spread further to the hospital environment or other humans. Extensive surveillance carried out in this room and surrounding areas after plumbing was removed and cleaned identified no additional KPC<sup>+</sup> organisms.



One troubling observation is that the *E. cloacae* sink isolate ECNIH2 carries three different *bla*<sub>KPC</sub>-encoding plasmids (Fig. 2H, Table 2). In addition to the pKEC-39c plasmid described above, ECNIH2 carries the *bla*<sub>KPC-3</sub> gene embedded within a Tn4401b transposon in a 282.4 kb IncIH2 plasmid related to pEC-IMP and a previously undescribed 47.3 kb plasmid. All three plasmids are predicted to encode proteins involved in the stable maintenance of plasmids including restriction/anti-restriction proteins, methylases and components of plasmid addiction systems. These plasmid maintenance systems, along with other genes that improve bacterial fitness, are barriers to clearance of antibiotic resistance plasmids. As noted before, with KONIH1, resolution of multiple *bla*<sub>KPC</sub> genes was not possible with assembled short-read data where only a single complete *bla*<sub>KPC</sub> gene was found; however, SMRT sequencing was able to distinguish the three *bla*<sub>KPC</sub> genes on three distinct plasmids. The two additional *bla*<sub>KPC</sub>-encoding plasmids contained within the *E. cloacae* sink isolate ECNIH2 do not match any other patient or environmental isolates from our institution, and contribute genetic diversity to the public database of KPC-encoding plasmids.

### Genomic evidence discordant with horizontal plasmid transfer

Patient 5 became colonized during the 2011 outbreak at our institution with his first KPC<sup>+</sup> *K. pneumoniae* isolate (KPNIH10) detected in August 2011. He had undergone no previous perirectal surveillance cultures. A follow-up culture two weeks later indicated continued colonization with KPC<sup>+</sup> *K. pneumoniae* and a culture 3 weeks subsequent identified the simultaneous presence of KPC<sup>+</sup> *E. cloacae* (ECNIH3). Given the patient's persistent colonization with the KPC<sup>+</sup> *K. pneumoniae* outbreak strain, we hypothesized horizontal transfer of the *bla*<sub>KPC</sub> gene-encoding plasmid from the outbreak strain to the *E. cloacae* had occurred. However, genome analysis comparing the sequences of the *bla*<sub>KPC</sub> gene encoding plasmids in the *K. pneumoniae* and *E. cloacae* strains demonstrated that this was not the case (Fig. 4, A and B). The KPC<sup>+</sup> *E. cloacae* harbored a 50.3 kb incompatibility group N (IncN) plasmid with a *bla*<sub>KPC-2</sub> gene flanked by an IS26 and partial Tn4401 transposon. In contrast, the KPC<sup>+</sup> *K. pneumoniae* isolate harbored pKpQIL, a 113.6 kb IncF plasmid carrying the *bla*<sub>KPC-3</sub> gene within a Tn4401a transposon. These data rule out both transfer of a plasmid between bacteria and also transposition of the Tn4401 element. The IncN plasmid (pKPC-47e) is ST 6, but rearranged compared to the closest reference "plasmid 12" (FJ223605) (Fig. 5). In addition, the KPC<sup>+</sup> *E. cloacae* isolate belonged to a new sequence type (ST 97). Taken together, these data suggest that patient 5 may have been colonized with the KPC<sup>+</sup> *E. cloacae* isolate when he was admitted to our hospital and then became co-colonized with the KPC<sup>+</sup> *K. pneumoniae* outbreak strain via nosocomial transmission.

### Circulation of newly emerging strains of CRE

Plasmid sequencing has also enabled us to identify epidemiological patterns that may otherwise escape notice. When analyzing surveillance data, the possibility that regional strains are circulating amongst hospitals should be considered as an alternative hypothesis to nosocomial transmission between patients with no discernible epidemiological connection. KPC<sup>+</sup> *K. pneumoniae* ST 258 with *bla*<sub>KPC</sub> gene on pKpQIL is known to be a dominant strain in U.S. hospitals, as demonstrated by the epidemiologically unrelated, genetically

similar isolates obtained from patients 1 and E. Next we explore the use of complete genomes to characterize and detect a newly emerging KPC<sup>+</sup> *E. cloacae* strain.

Patient D was admitted to our institution in 2012 and was discharged shortly after a single negative perirectal surveillance culture. She was subsequently admitted at another hospital where KPC<sup>+</sup> *E. cloacae* (ECR091) grew from a urine culture. Comparison of this isolate's genome to our existing references showed striking resemblance to ECNIH3 from patient 5. ECR091 and ECNIH3 are the same sequence type and differ by only 12 high confidence SNVs across the 4.6 Mb chromosomes. These two KPC<sup>+</sup> *E. cloacae* isolates differ in their total plasmid composition (Fig. 4, B and C) but they both carry the pKPC-47e plasmid, which encodes the *bla*<sub>KPC</sub> gene. In addition, they each carry an IncHI2 plasmid related to pEC-IMP, with insertion/deletions unique to each isolate. However, there was no discernible epidemiological link between patient 5 and patient D. The KPC<sup>+</sup> *E. cloacae* isolate from patient D was identified over a year after that of patient 5, and three months after his most recent hospitalization. The two patients were from disparate geographic areas with no common exposure at another healthcare facility. In this case, the combination of the close genetic sequence similarity and the lack of discernable epidemiologic link suggests that the KPC<sup>+</sup> *E. cloacae* ST97 is a newly recognized strain circulating in U.S. hospitals.

### A broad-host range KPC<sup>+</sup> IncN plasmid

After identifying the IncN plasmid, pKPC-47e, in patient 5 and patient D (above), we began to screen additional KPC<sup>+</sup> isolates for the presence of an IncN, *bla*<sub>KPC</sub> plasmid with a multiplex PCR assay using IncN replicon typing primers (30). Between 2011 and 2013, we identified IncN, *bla*<sub>KPC</sub> plasmids (Fig. 1, Fig. 4) carried by two additional patient isolates (ECONIH1 and KPNIH29) and four environmental isolates (ECNIH4, ECNIH5, PSNIH1, and PSNIH2). The host range of this plasmid includes *E. cloacae*, *Pantoea* sp., *K. pneumoniae*, and *E. coli*. Given the potential clinical impact of pKPC-47e, we characterized further this plasmid. pKPC-47e is 99.8% identical over 75% of its length to the published sequence of *K. pneumoniae* plasmid 12 (NC\_011385), a Tn4401b-*bla*<sub>KPC-3</sub> encoding plasmid (Fig. 5). The pKPC-47e plasmid has undergone substantial rearrangement compared with plasmid 12 and differed by an expanded iteron region (direct repeats that play a role in plasmid maintenance), insertion of a region encoding *ccgCD*, deletion of an arsenic resistance operon, and deletion of a large antimicrobial resistance region (Fig. 5). Deletion of the antimicrobial resistance region disrupts the Tn4401b transposon, replacing the region upstream of the *bla*<sub>KPC</sub> gene with an IS26 element. IS26 elements have been associated with mobilization of antibiotic resistance genes and have been shown to modify expression of flanking genes (31–33).

In addition to the dominant pKPC-47e plasmid, two major variants of this plasmid were also detected. The Tn4401b element is intact in both variants, suggesting that pKPC-47e may be a product of a deletion event from one of these plasmids or a third unobserved plasmid. The *E. coli* plasmid, pKPC-629, differed from pKPC-47e by a 30 kb region, upstream of the *bla*<sub>KPC</sub> gene, encoding a Tn1331 transposon and a mercury resistance operon (Fig. 5B). The presence of this KPC<sup>+</sup> plasmid in the ST 648 *E. coli* from patient H was of concern because this sequence type is commonly associated with extraintestinal pathogenic *E. coli* infections.

A second variant of pKPC-47e was identified in *K. pneumoniae* (Fig. 5C) that is similar to the *E. coli* plasmid but is missing the mercury resistance operon.

The pKPC-47e backbone was represented frequently in our pool of KPC<sup>+</sup> organisms, occurring in eight of the twenty isolates. Whereas the presence of a common backbone suggests a common source, no epidemiological links were identified among the patients or environmental sites in which these isolates were found. Finally, while determining the distribution of the IncN plasmid in isolates from 2013, three additional KPC<sup>+</sup> *K. pneumoniae* isolates were sequenced, two of which carried new KPC<sup>+</sup> plasmids (Fig S1). Notably, two of these isolates (KPNIH33, KPNIH31) carry chromosomally integrated Tn4401-KPC cassettes.

## Discussion

The NIH Clinical Center is the nation's largest clinical research hospital, with approximately 1500 clinical research studies currently in progress. The patient population includes many severely ill and immunocompromised individuals who are particularly vulnerable to healthcare-associated infections. In addition, patients referred to the NIH Clinical Center are often receiving treatment at multiple healthcare facilities, increasing their risk of exposure to nosocomial pathogens. The motivation for the current work is to improve the detection and tracking of carbapenemase-producing organisms, to enable clinicians to better assess the threats posed by emerging antibiotic-resistant organisms and to target infection control efforts to relevant instances of transmission.

Sequencing can elucidate the landscape of bacterial transmission, allowing hospitals to target funds and personnel for infection control interventions that provide the best patient care. On the basis of our sequence data demonstrating that only one of the 10 cases represented nosocomial spread, we directed our resources toward surveillance at admission for carbapenemase-producing organisms. In the case of patient 19 in whom hospital transmission was confirmed, we targeted resources to improve adherence to barrier precautions and hand hygiene on a high-risk inpatient ward. The cost of whole genome sequencing is dwarfed by the others costs associated with outbreaks and their investigations, including the human and financial toll (34) and the loss of patient confidence in the healthcare facility. Detailed microbial sequencing thus provides actionable data that can improve patient outcomes and utilization of hospital resources.

Our 2012 KPC<sup>+</sup> *K. pneumoniae* study at NIH exploited the synergy between epidemiologic trace data and bacterial genetic analyses to produce an integrated transmission map. Here, we extended this concept of tracking antibiotic-resistant bacterial transmission by investigating the next layer of complexity, the plasmid transmission network. While examples of shared plasmids were detected, the number of independent introductions observed was surprisingly high and indicated a complex network of plasmids with incredible diversity. CRE isolates carrying an average of three plasmids for a total of 62 plasmids across 20 isolates. After clustering of similar plasmids, there are 48 plasmids that differ from available reference sequences, resulting in >1% growth in the size of the non-redundant plasmid database. These diverse *bla*<sub>KPC</sub> encoding plasmid sequences can inform the

development of advanced molecular diagnostics, such as PCR-based assays to differentiate strains and plasmids (35).

We observed only a few examples of horizontal plasmid transfer, which is somewhat surprising given the selective advantage of antibiotic resistance. Deterrents for plasmid transmission include host range specificity, plasmid incompatibility, clustered regularly interspaced short palindromic repeat (CRISPR) loci, and restriction modification systems. The 22 KPC<sup>+</sup> plasmids (ECNIH2 had 3 KPC<sup>+</sup> plasmids) sequenced as part of this study belonged to several classes of replicon incompatibility groups, including IncF, IncA/C, IncN and IncHI2. Some of these replicons have been reported to have a broad host range, such as IncN and IncA/C, and agree with trends we see across isolates. CRISPR systems, which act to prevent phage infection and plasmid transfer (36), were detected in seven of the sequenced isolates. DNA restriction and modification systems are another barrier to plasmid transfer and one which is uniquely addressable by SMRT sequencing technology (37). In this study, pKpQIL plasmid presence is associated with N<sup>6</sup>-methyladenine present on both the chromosomal and plasmid DNA in the sequence context 5'-<sup>m6</sup>ACGNNNNNNCTG-3' (methylated A on opposite DNA strand of underlined T). In the end, the low frequency of transmission is likely the result of molecular barriers to transmission combined with infection control interventions specified by Centers for Disease Control and Prevention (CDC) guidelines.

We observed three instances in which genomic data linked epidemiologically unrelated carbapenemase-encoding plasmids. The first example is a pair of KPC<sup>+</sup> *K. pneumoniae* isolates carrying identical pKpQIL-531 plasmids, from patients E and I. These patients did not overlap in our institution but both had previously received care at the same healthcare complex, pointing to a possible link across time and space. The second example of a newly emerging strain is a pair of KPC<sup>+</sup> *E. cloacae* isolates from patients with no evident epidemiological connection and nearly identical genomes (12 SNVs), *bla*<sub>KPC-2</sub> containing plasmids, and a second shared plasmid that differs by a single, large deletion. Our assessment of independent introduction is supported by the detection of this same *bla*<sub>KPC</sub>-containing plasmid in a broader range of *Enterobacteriaceae*, including KPC<sup>+</sup> *K. pneumoniae*, *E. coli* and *Pantoea* sp. We are actively screening CREs from patients and the environment for IncN plasmids to document the evolution of this plasmid.

The third example of shared carbapenemase-encoding plasmids was detected across three different bacterial genera (*K. pneumoniae*, *E. cloacae*, *C. freundii*) with plausible epidemiologic links. The broad host range of a pR55-like plasmid and possibility that it persists in environmental reservoirs like sink drain biofilms, represents a challenge for hospital infection control. In this case, plumbing was removed, disinfected, and replaced, but the extent to which environmental bacteria represent a reservoir for transmission and the optimal method for remediation remain unknown. Several reports have implicated hospital sinks in the nosocomial transmission of multidrug-resistant Gram-negative bacteria, but none have been able to prove definitively whether the sinks were the source or the recipients (38–40).

One limitation of this study is that our highly detailed plasmid sequence data cannot explain how these common environmental commensals acquired the plasmids, but nonetheless raise concern that KPC plasmid dissemination may occur silently in the hospital environment in the absence of any known colonized patients. Identification of the pKPC-47e backbone in both patient and environmental samples is disquieting for infection control personnel because some host organisms (e.g., two nonclonal strains of *Pantoea* sp.) are capable of establishing a foothold on environmental surfaces, while others are human commensals and/or potential pathogens. *E. coli*, for instance, are a common cause of community-acquired infection, and carbapenemase producing strains could potentially affect increasing numbers of patients outside healthcare facilities (41).

A second limitation is the sensitivity of perirectal surveillance cultures. In several cases, KPC<sup>+</sup> organisms were detected in patients who had produced multiple negative surveillance cultures prior to a positive culture. For patient 19, genomic and epidemiologic data pointed to nosocomial transmission of KPC<sup>+</sup> *K. pneumoniae*. In all other cases, genomic and epidemiological data did not support nosocomial transmission. And, these KPC<sup>+</sup> isolates were not observed again even with extensive patient and environmental surveillance performed in our institution, suggesting strong adherence to infection control measures. Ultimately, the genetic and epidemiologic evidence led us to question whether these patients who converted from negative to positive likely represent low-level gastrointestinal colonization on admission not detected by surveillance cultures. Perirectal surveillance swabs, which only sample flora of the perianal area, can produce false negative results due to poor sampling technique or insufficiently sensitive culture methods. Even the most sensitive method cannot detect KPC<sup>+</sup> isolates present at low levels in areas of the gastrointestinal tract not accessed by the surveillance swab. Administration of broad-spectrum antibiotics may then have selected for the antimicrobial-resistant KPC<sup>+</sup> organisms, yielding a positive culture. Although undetected acquisition is not entirely ruled out, the absence of any matching isolates in a setting of extensive surveillance among a highly immunocompromised patient population points to likely acquisition predating hospital admission.

Over the last decade there has been a steady and alarming increase in antibiotic-resistant bacteria (42); a trend that poses a serious threat to the U.S. medical system. At the same time, development of antimicrobial drugs has nearly ground to a halt, with only two new antibiotics approved since 2009 (43). In the face of a dwindling selection of drugs to fight healthcare-associated infections, prevention is critical. In addition to implementing recommended infection control measures such as surveillance, hand hygiene, and barrier precautions, we must find more sophisticated methods to detect, track and eradicate multidrug-resistant bacteria. Collaboration among physicians who have expertise in healthcare epidemiology, microbiologists who have expertise in diagnostics, and scientists who have expertise in genomics is critical to take advantage of emerging technologies and translate them into improved patient care.

## Materials and Methods

### Environmental and Patient CRE Surveillance

Patients were screened on admission for risk factors of CRE carriage, including hospitalization in the United States in the past week or abroad in the past six months. These higher-risk patients underwent surveillance peri-rectal cultures on two consecutive days, and were placed on empiric contact isolation until cultures demonstrated no growth of carbapenem-resistant bacteria. Patients lacking prior hospitalization risk factors had one surveillance culture upon admission. Patient surveillance swabs were collected as described (19).

Environmental surface swabs were collected using 2"× 2" squares of sterile gauze moistened with fastidious broth (Hardy Diagnostics) then, with sterile gloves, applied systematically to surfaces of interest. The gauze was immersed in broth and incubated overnight before culturing. Sink drain cultures were collected by inserting a culturette swab through the drain sieve and rubbing once around the interior walls of the drain.

Specimens were cultured onto RambaCHROM KPC (Gibson Bioscience). Organisms were identified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using the MicroFlex LT instrument (Bruker Daltonics Inc.) with spectra acquired as referenced (44). Organism identifications were confirmed or refined by comparing marker gene sequences, specifically the 16S rRNA and *rpoB* genes, to curated databases and other fully sequenced genomes. Carbapenemase production was assessed by Taqman real-time PCR with KPC-targeted primers adapted from the CDC (45) and as referenced (46). Antibiotic susceptibility testing was performed by broth microdilution (Phoenix instrument; Becton, Dickinson and Co.) and Kirby-Bauer disk diffusion.

### Genomic DNA Sequencing and analysis

Genomic DNA was sheared to 10–15 kb using Covaris g-tubes (Covaris), and converted into SMRTbell template libraries (47). The library was subsequently subjected to DNA size selection using the BluePippin (Sage Science) to select the longest DNA fragments (lower size cutoff ~5 kb). Sequencing was performed on the PacBio RSII using P4 polymerase binding and C2 sequencing kits with magnetic bead loading and 120 minute acquisition. Genome assemblies were performed using HGAP and Quiver (23) as part of SMRTAnalysis version 2.0. SMRT-based assemblies were validated using optical maps and MiSeq read data. A full description of the independent sequence analysis is found in Supplemental Materials. Briefly, Nextera libraries were generated from the same genomic DNA and sequenced using a paired-end 200 base dual index run on an Illumina MiSeq to generate 1–2 million read pairs per library. Base calling accuracy was determined with the MPG software (25). MiSeq read alignments were visualized with IGV (48), consed (49) and Hawkeye (50) software. Genome annotation was done using NCBI's Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP: [http://www.ncbi.nlm.nih.gov/genome/annotation\\_prok/](http://www.ncbi.nlm.nih.gov/genome/annotation_prok/)).

## Typing

Multilocus sequence typing was used to classify isolates from their complete genomes. Alleles for each typing method were downloaded from their respective repositories: *K. pneumoniae* (Institut Pasteur), *E. cloacae* (PubMLST), *E. coli* (University of Warwick), IncN replicon typing (PubMLST). Newly detected alleles or combinations of alleles were submitted to the appropriate databases for inclusion in future typing schemes. The *bla*<sub>KPC</sub> allele was identified by BLAST alignment to reference sequences from NC\_011382 (KPC-2) and NC\_014016 (KPC-3). The diversity of capsular polysaccharide regions was assessed by determining the size between conserved flanking genes *galF* and *ugd* (51) using *in silico* PCR with the following primers: *cpsF*, CGCTTCATAATGCCCTGAAT; and *cpsR*, CAAAGGCAATTCCAAAGGAG.

## SNV characterization

High-confidence SNVs were identified essentially as described in (4). Briefly, SNVs were extracted from pairwise alignments (NUCmer, v 3.06) of chromosome or plasmid sequences. SNVs were filtered if they were in repetitive regions, flanked by homopolymer runs, within two bases of a neighboring SNV or 20 bases of the end of a sequence. We also filtered out single-base insertions and deletions. For SNVs detected between chromosomes, recombinant regions were removed with a sliding window to detect regions of high SNV density. These filtering rules were originally developed to address systematic errors inherent in 454 and other sequencing platforms. Whereas SMRT sequencing has not been shown to have these sorts of systematics errors, we elected for conservative filtering since we are using genomes generated from assembled Sanger, 454 or Illumina data as references.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank staff of the NIH Clinical Center Hospital Epidemiology and Microbiology Services for their many contributions to the infection control methods described in this report. Segre lab members provided thoughtful reading of the manuscript. Julia Fekecs provided graphic design assistance with figures.

**Funding:** Supported by the National Human Genome Research Institute and NIH Clinical Center Intramural Research Programs and by an NIH Director's Challenge Award. ESS is supported by a Pharmacology Research Associate Training Fellowship, NIGMS. Sequencing was performed at Pacific Biosciences, NISC and Leidos Biomedical Research, Frederick National Laboratory for Cancer Research.

## References

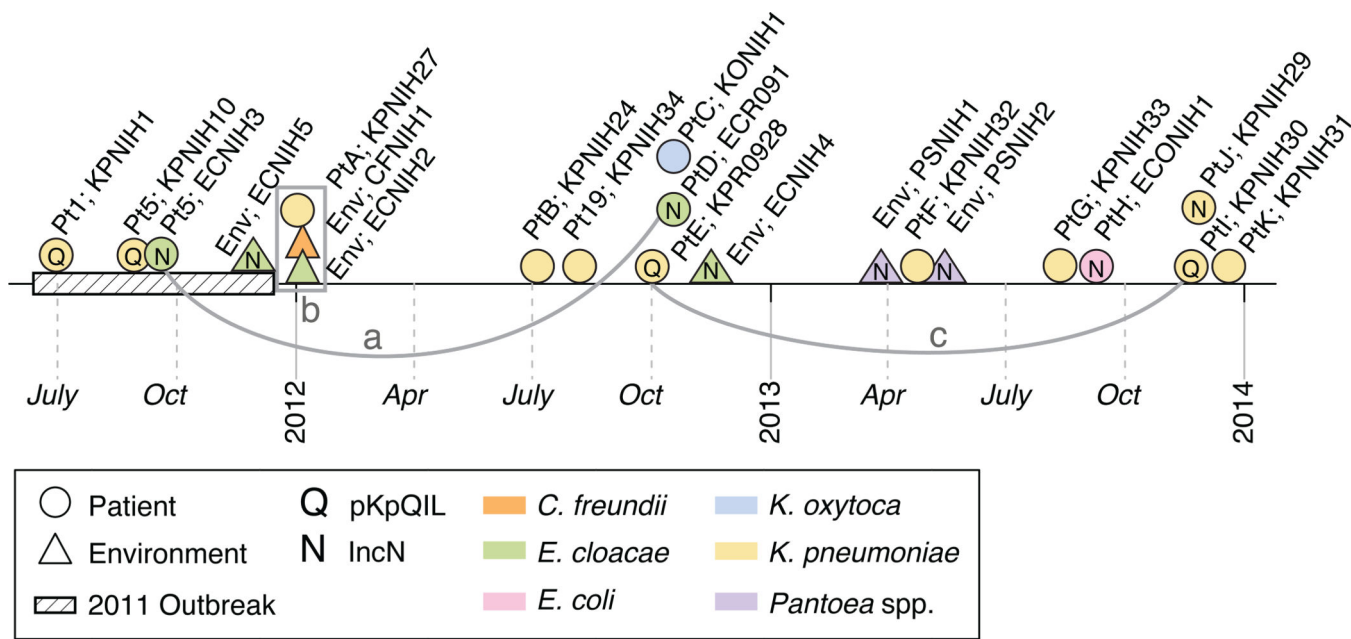
1. Mathers AJ, et al. Molecular dissection of an outbreak of carbapenem-resistant enterobacteriaceae reveals Intergenous KPC carbapenemase transmission through a promiscuous plasmid. *MBio*. 2011; 2:e00204–e00211. [PubMed: 22045989]
2. Satlin MJ, et al. Emergence of carbapenem-resistant Enterobacteriaceae as causes of bloodstream infections in patients with hematologic malignancies. *Leuk. Lymphoma*. 2013; 54:799–806. [PubMed: 22916826]
3. Ben-David D, et al. Outcome of carbapenem resistant *Klebsiella pneumoniae* bloodstream infections. *Clin. Microbiol. Infect.* 2012; 18:54–60. [PubMed: 21722257]

4. Snitkin ES, et al. Tracking a Hospital Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae* with Whole-Genome Sequencing. *Sci Transl Med*. 2012; 4 148ra116.
5. CDC Vital Signs. available at <http://www.cdc.gov/vitalsigns/hai/cre/index.html>
6. Griffith F. The Significance of Pneumococcal Types. *J Hyg (Lond)*. 1928; 27:113–159. [PubMed: 20474956]
7. Nordmann P, Dortet L, Poirel L. Carbapenem resistance in Enterobacteriaceae: here is the storm! *Trends Mol Med*. 2012; 18:263–272. [PubMed: 22480775]
8. Guidance for Control of Carbapenem-Resistant Enterobacteriaceae, 2012 CRE Tooklit - CRE-guidance-508.pdf. available at <http://www.cdc.gov/hai/pdfs/cre/CRE-guidance-508.pdf>
9. Fricke WF, Rasko DA. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat. Rev. Genet*. 2014; 15:49–55. [PubMed: 24281148]
10. Ehrlich GD, Post JC. The time is now for gene- and genome-based bacterial diagnostics: “you say you want a revolution.”. *JAMA Intern Med*. 2013; 173:1405–1406. [PubMed: 23857368]
11. Köser CU, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog*. 2012; 8:e1002824. [PubMed: 22876174]
12. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet*. 2012; 13:601–612. [PubMed: 22868263]
13. Peacock S. Health care: Bring microbial sequencing to hospitals. *Nature*. 2014; 509:557–559. [PubMed: 24877182]
14. Harris SR, et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science*. 2010; 327:469–474. [PubMed: 20093474]
15. Mutreja A, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*. 2011; 477:462–465. [PubMed: 21866102]
16. Harris SR, et al. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect Dis*. 2013; 13:130–136. [PubMed: 23158674]
17. Rasko DA, et al. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N. Engl. J. Med*. 2011; 365:709–717. [PubMed: 21793740]
18. Gardy JL, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med*. 2011; 364:730–739. [PubMed: 21345102]
19. Palmore TN, Henderson DK. Managing transmission of carbapenem-resistant enterobacteriaceae in healthcare settings: a view from the trenches. *Clin. Infect. Dis*. 2013; 57:1593–1599. [PubMed: 23934166]
20. Snitkin ES, et al. Genome-wide recombination drives diversification of epidemic strains of *Acinetobacter baumannii*. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108:13758–13763. [PubMed: 21825119]
21. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biol*. 2013; 14:405. [PubMed: 23822731]
22. Eid J, et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science*. 2009; 323:138. 133.
23. Chin C-S, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods*. 2013
24. Gootz TD, et al. Genetic organization of transposase regions surrounding blaKPC carbapenemase genes on plasmids from *Klebsiella* strains isolated in a New York City hospital. *Antimicrob. Agents Chemother*. 2009; 53:1998–2004. [PubMed: 19258268]
25. Teer JK, et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res*. 2010; 20:1420–1431. [PubMed: 20810667]
26. Leavitt A, Chmelnitsky I, Carmeli Y, Navon-Venezia S. Complete nucleotide sequence of KPC-3-encoding plasmid pKpQIL in the epidemic *Klebsiella pneumoniae* sequence type 258. *Antimicrob. Agents Chemother*. 2010; 54:4493–4496. [PubMed: 20696875]
27. Warburg G, et al. A carbapenem-resistant *Klebsiella pneumoniae* epidemic clone in Jerusalem: sequence type 512 carrying a plasmid encoding aac(6′)-Ib. *J. Antimicrob. Chemother*. 2012; 67:898–901. [PubMed: 22287232]

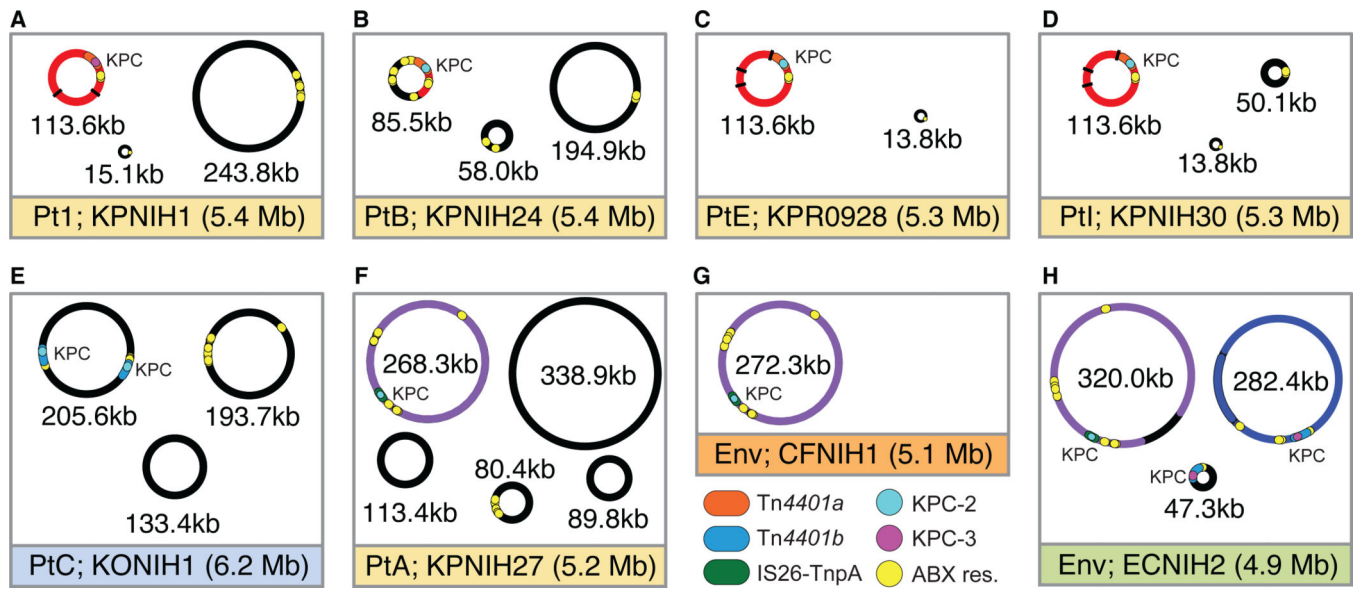


28. Villa L, et al. Reversion to susceptibility of a carbapenem-resistant clinical isolate of *Klebsiella pneumoniae* producing KPC-3. *J. Antimicrob. Chemother.* 2013; 68:2482–2486. [PubMed: 23800906]
29. García-Fernández A, et al. *Klebsiella pneumoniae* ST258 producing KPC-3 identified in Italy carries novel plasmids and OmpK36/OmpK35 porin variants. *Antimicrob. Agents Chemother.* 2012; 56:2143–2145. [PubMed: 22252815]
30. García-Fernández A, et al. Multilocus sequence typing of IncN plasmids. *J. Antimicrob. Chemother.* 2011; 66:1987–1991. [PubMed: 21653604]
31. Skipper KA, Andersen PR, Sharma N, Mikkelsen JG. DNA transposon-based gene vehicles - scenes from an evolutionary drive. *J. Biomed. Sci.* 2013; 20:92. [PubMed: 24320156]
32. Miriagou V, Carattoli A, Tzelepi E, Villa L, Tzouvelekis LS. IS26-associated In4-type integrons forming multiresistance loci in enterobacterial plasmids. *Antimicrob. Agents Chemother.* 2005; 49:3541–3543. [PubMed: 16048979]
33. Lee KY, Hopkins JD, Syvanen M. Direct involvement of IS26 in an antibiotic resistance operon. *J. Bacteriol.* 1990; 172:3229–3236. [PubMed: 2160941]
34. Spearing NM, Jensen A, McCall BJ, Neill AS, McCormack JG. Direct costs associated with a nosocomial outbreak of *Salmonella* infection: an ounce of prevention is worth a pound of cure. *Am J Infect Control.* 2000; 28:54–57. [PubMed: 10679138]
35. Yang JY, et al. Pan-PCR, a computational method for designing bacterium-typing assays based on whole-genome sequence data. *J. Clin. Microbiol.* 2013; 51:752–758. [PubMed: 23254127]
36. Hatoum-Aslan A, Marraffini LA. Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens. *Curr. Opin. Microbiol.* 2014; 17C:82–90. [PubMed: 24581697]
37. Davis BM, Chao MC, Waldor MK. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Curr. Opin. Microbiol.* 2013; 16:192–198. [PubMed: 23434113]
38. Kotsanas D, et al. “Down the drain”: carbapenem-resistant bacteria in intensive care unit patients and handwashing sinks. *Med. J. Aust.* 2013; 198:267–269. [PubMed: 23496403]
39. Hota S, et al. Outbreak of multidrug-resistant *Pseudomonas aeruginosa* colonization and infection secondary to imperfect intensive care unit room design. *Infect Control Hosp Epidemiol.* 2009; 30:25–33. [PubMed: 19046054]
40. Lowe C, et al. Outbreak of extended-spectrum  $\beta$ -lactamase-producing *Klebsiella oxytoca* infections associated with contaminated handwashing sinks(1). *Emerging Infect. Dis.* 2012; 18:1242–1247. [PubMed: 22841005]
41. Bratu S, et al. Detection and spread of *Escherichia coli* possessing the plasmid-borne carbapenemase KPC-2 in Brooklyn, New York. *Clin. Infect. Dis.* 2007; 44:972–975. [PubMed: 17342651]
42. Lockhart SR, et al. Antimicrobial resistance among Gram-negative bacilli causing infections in intensive care unit patients in the United States between 1993 and 2004. *J. Clin. Microbiol.* 2007; 45:3352–3359. [PubMed: 17715376]
43. Boucher HW, et al. 10 x ‘20 Progress--Development of New Drugs Active Against Gram-Negative Bacilli: An Update From the Infectious Diseases Society of America. *Clin. Infect. Dis.* 2013; 56:1685–1694. [PubMed: 23599308]
44. Lau AF, Drake SK, Calhoun LB, Henderson CM, Zelazny AM. Development of a clinically comprehensive database and a simple procedure for identification of molds from solid media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 2013; 51:828–834. [PubMed: 23269728]
45. Multiplex Real-Time PCR Detection of KPC and NDM-1 genes - KPC-NDM-protocol-2011.pdf. available at <http://www.cdc.gov/HAI/pdfs/labSettings/KPC-NDM-protocol-2011.pdf>
46. Hindiyeh M, et al. Rapid detection of blaKPC carbapenemase genes by real-time PCR. *J. Clin. Microbiol.* 2008; 46:2879–2883. [PubMed: 18614657]
47. Travers KJ, Chin C-S, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* 2010; 38:e159. [PubMed: 20571086]

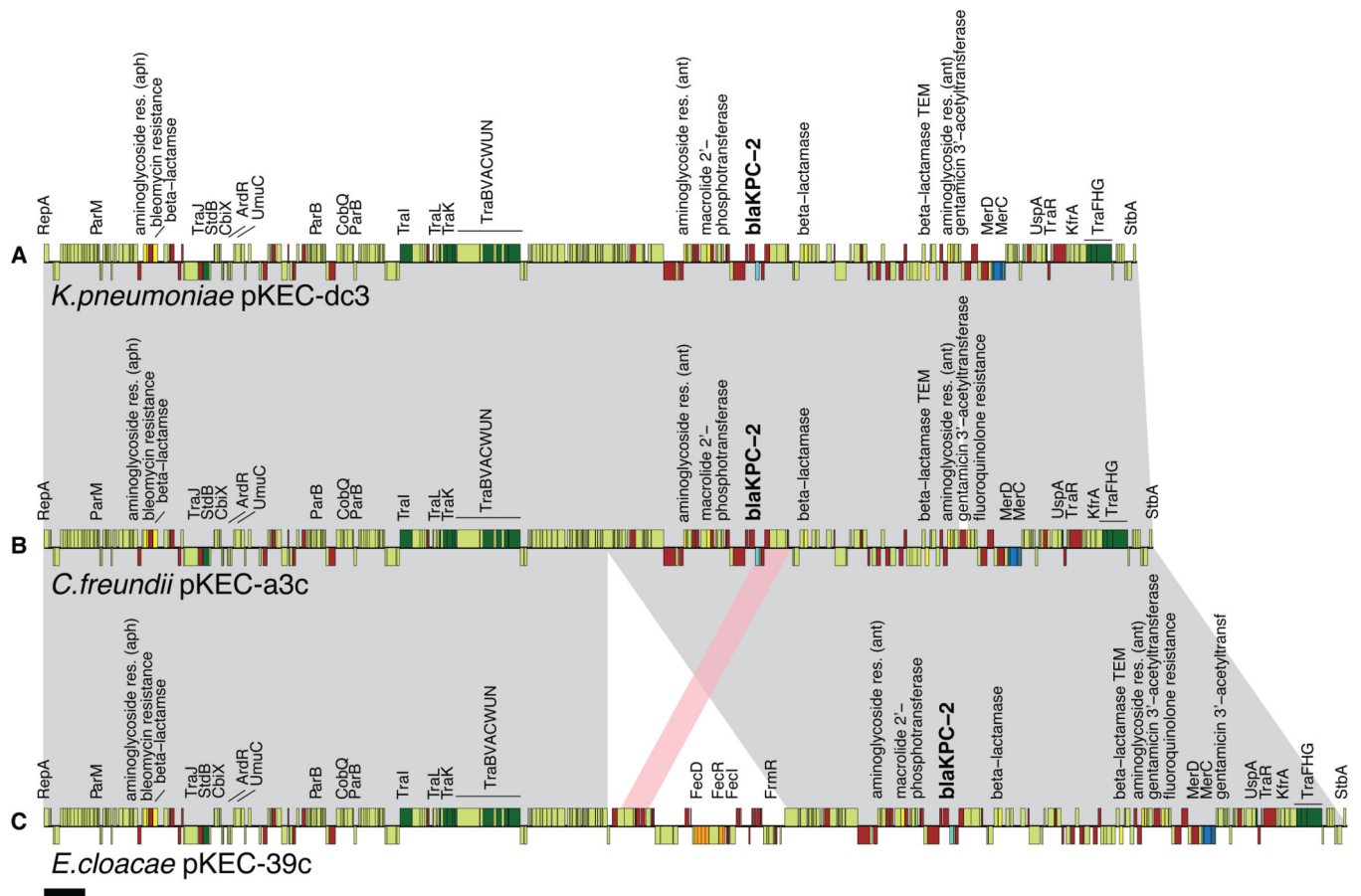
48. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics*. 2013; 14:178–192. [PubMed: 22517427]
49. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res*. 1998; 8:195–202. [PubMed: 9521923]
50. Schatz MC, et al. Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. *Brief. Bioinformatics*. 2013; 14:213–224. [PubMed: 22199379]
51. Shu H-Y, et al. Genetic diversity of capsular polysaccharide biosynthesis in *Klebsiella pneumoniae* clinical isolates. *Microbiology (Reading, Engl.)*. 2009; 155:4170–4183.



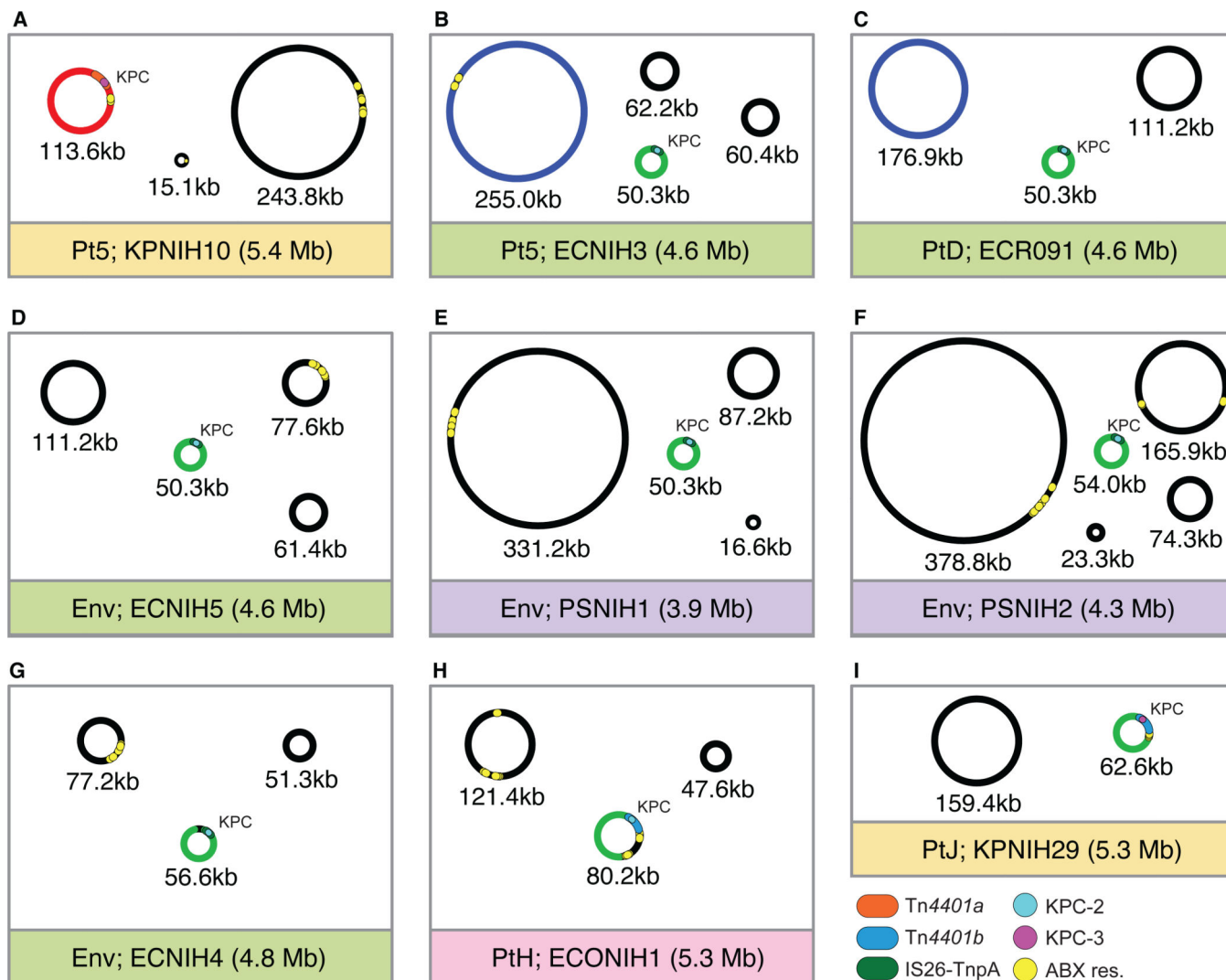
**Fig. 1.** Timeline of KPC<sup>+</sup> Enterobacteriaceae organisms identified and sequenced from 2011–2013. Patient code, genus, species and strain denoted in label; for example, Pt1; KPNIH1 is Patient 1; *Klebsiella pneumoniae* NIH1 strain and Env; ECNIH5 is Environmental; *Enterobacter cloacae* NIH5 strain. Detailed data are presented in Table 1. The isolate source is denoted by shape and the bacterial species is denoted by color. Isolates carrying a variant of the plasmid pKpQIL are marked with a “Q”. Isolates carrying the pKPC-47e backbone (IncN, this study) are marked with an “N”. Isolates connected by lines a and c share genetic similarity, without epidemiologic linkage. Isolates in box b show evidence of an inter-genus exchange of a pR55-like plasmid. The 2011 KPC<sup>+</sup> *K. pneumoniae* outbreak is denoted with a hashed line.

**Fig. 2.**

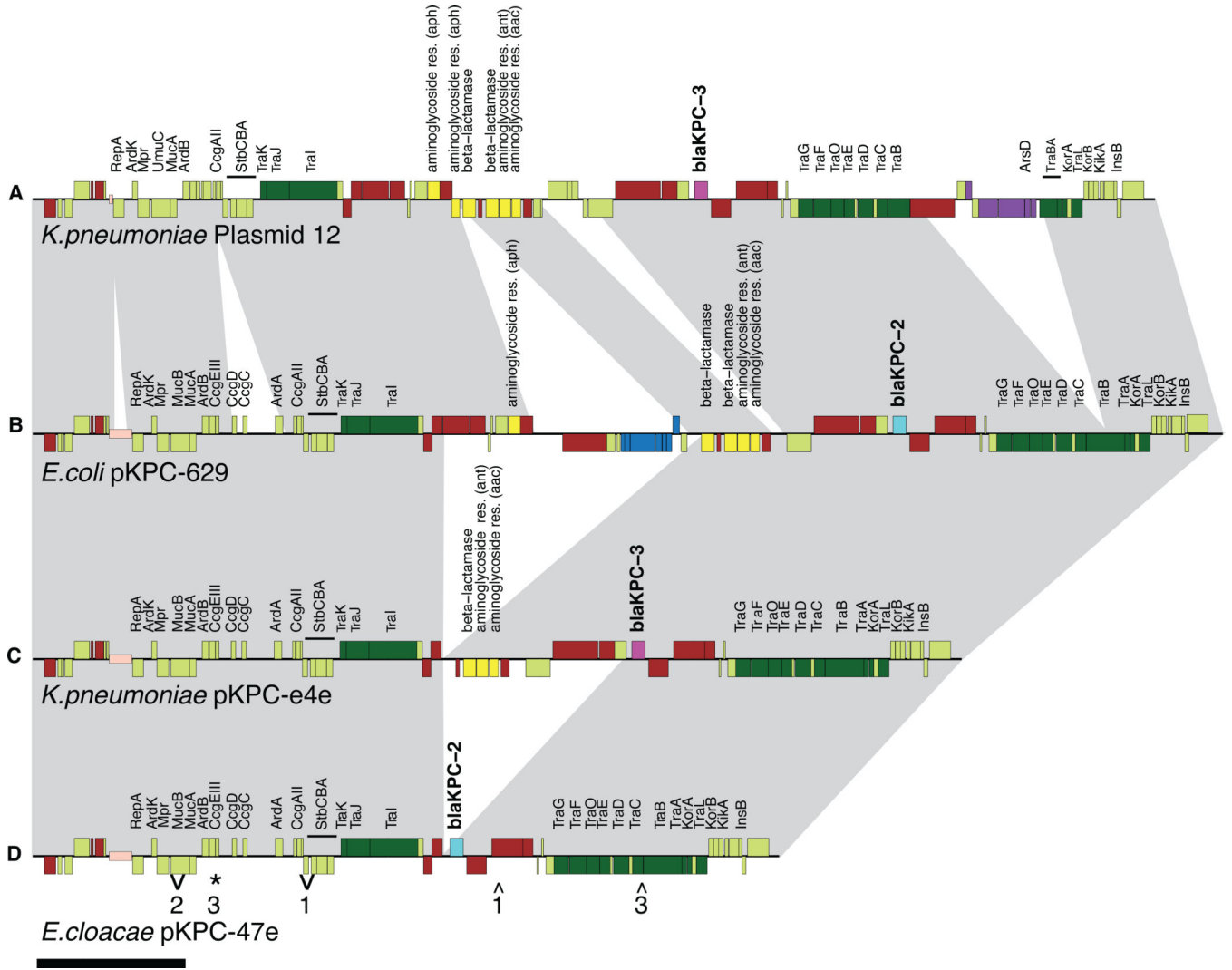
Genome sequencing and plasmid composition of *Klebsiella* isolates (A–E) and isolates associated with an inter-genus plasmid transfer (F–H). The color of title bar corresponds to the bacterial species from Fig. 1. Patient code or environmental isolate code and chromosome size denoted in each title bar. Detailed patient code, genus and species described in Table 1. Plasmids are colored to highlight three common backbones across isolates (red, pKpQIL; blue, pEC-IMP-like; purple, pR55-like). The *bla*<sub>KPC</sub> genes are colored by allele type, KPC-2 is cyan and KPC-3 is magenta. Flanking sequences are colored to differentiate the *bla* gene context (red, Tn4401a; blue, Tn4401b; green, IS26-TnpA). Additional antibiotic resistance genes are marked in yellow. The positions of SNVs in pKpQIL plasmids, relative to the reference (NC\_014016), are marked as black tick marks in panels A, C and D.



**Fig. 3.** Genetic tracking of plasmid transfer between Enterobacteriaceae genera. (A) KPC<sup>+</sup>*K. pneumoniae* plasmid pKEC-dc3 from patient A. (B) *C. freundii* plasmid pKEC-a3c from sink aerator. (C) *E. cloacae* plasmid pKEC-39c from sink drain. Transposase and resolvase genes are brown, *tra* genes are dark green, *bla*<sub>KPC-2</sub> is cyan, other antibiotic resistance genes are yellow, mercury resistance genes are blue and iron acquisition genes are orange. Shared regions of homology are indicated in gray shading with a duplication event indicated in pink. Scale bar is 10 kb.



**Fig. 4.** Genome sequencing of isolates with a broad host range IncN plasmid. Color of title bar corresponds to the bacterial species from Fig. 1. Patient code or environmental isolate code and chromosome size denoted in each title bar. Detailed patient code, genus and species described in Table 1. Plasmids are colored to highlight three common backbones across isolates (red, pKpQIL; blue, pEC-IMP-like; green, pKPC-47e). The *bla*<sub>KPC</sub> genes are colored by allele type, KPC-2 is cyan and KPC-3 is magenta. Flanking sequences are colored to differentiate the *bla* gene context (red, Tn4401a; blue, Tn4401b; green, IS26-TnpA). Additional antibiotic resistance genes are marked in yellow.



**Fig. 5.** Genetic similarity of IncN plasmids identified across KPC<sup>+</sup>*E. coli*, *K. pneumoniae*, *E. cloacae* and *Pantoea* sp. Gray ribbons between panels mark regions of >99% sequence similarity. (A) Reference KPC<sup>+</sup>*K. pneumoniae* plasmid 12. (B) KPC<sup>+</sup>*E. coli*, ECONIH1, plasmid pKPC-629. (C) KPC<sup>+</sup>*K. pneumoniae*, KPNIH29, plasmid pKPC-e4e. (D) KPC<sup>+</sup>*E. cloacae* plasmid pKPC-47e. ECR091, ECNIH3, and ECNIH5 have identical sequences. Small insertion and deletions are marked with arrows for ECNIH4 (1), PSNIH1(2), and PSNIH2(3). SNVs are marked with \*. Transposase and resolvase genes are colored brown, *tra* genes are dark green, *bla*<sub>KPC-2</sub> is cyan, *bla*<sub>KPC-3</sub> is magenta, other antibiotic resistance genes are yellow, mercury resistance genes are blue and arsenic resistance genes are purple. The expanded iteron repeat is indicated in beige. Scale bar is 10 kb.

**Table 1**

Specimen source data for KPC<sup>+</sup> isolates.

Patient #	Date of Isolation	Genus and species	Strain	Source/ Culture Type	Underlying Condition
1	Jun-11	<i>Klebsiella pneumoniae</i>	KPNIH1	groin	organ transplant
5	Aug-11	<i>Klebsiella pneumoniae</i>	KPNIH10	groin	solid tumor
5	Sep-11	<i>Enterobacter cloacae</i>	ECNIH3	tracheal aspirate	solid tumor
Env	Nov-11	<i>Enterobacter cloacae</i>	ECNIH5	sink drain	NA
A	Jan-12	<i>Klebsiella pneumoniae</i>	KPNIH27	groin	stem cell transplant
Env	Jan-12	<i>Citrobacter freundii</i>	CFNIH1	sink aerator	NA
Env	Jan-12	<i>Enterobacter cloacae</i>	ECNIH2	sink drain	NA
B	Jul-12	<i>Klebsiella pneumoniae</i>	KPNIH24	throat/groin	immunodeficiency
19	Jul-12	<i>Klebsiella pneumonia</i>	KPNIH34	perirectal	stem cell transplant
D	Oct-12	<i>Enterobacter cloacae</i>	ECR091	urine	solid tumor
C	Oct-12	<i>Klebsiella oxytoca</i>	KONIH1	perirectal	hematologic malignancy
E	Oct-12	<i>Klebsiella pneumoniae</i>	KPR0928	sputum	immunodeficiency
Env	Nov-12	<i>Enterobacter cloacae</i>	ECNIH4	sink drain	NA
Env	Mar-13	<i>Pantoea</i>	PSNIH1	shelf	NA
F	Apr-13	<i>Klebsiella pneumoniae</i>	KPNIH32	perirectal	hematologic malignancy
Env	May-13	<i>Pantoea</i>	PSNIH2	hand rail	NA
G	Aug-13	<i>Klebsiella pneumoniae</i>	KPNIH33	urine	solid tumor
H	Sep-13	<i>Escherichia coli</i>	ECONIH1	perirectal	stem cell transplant
I	Nov-13	<i>Klebsiella pneumoniae</i>	KPNIH30	perirectal	hematologic malignancy
J	Nov-13	<i>Klebsiella pneumoniae</i>	KPNIH29	perirectal	organ transplant
K	Dec-13	<i>Klebsiella pneumoniae</i>	KPNIH31	urine	hematologic malignancy

Patients associated with the 2011 *K. pneumoniae* clonal strain are numbered (1–19), additional patients are indicated by letters (A–K) and environmental samples are indicated by Env. NA, not applicable.



**Table 2**

Summary of genomic data for *bla*<sub>KPC</sub> carrying plasmids identified in this study.

	Genus and species	Strain	Genome Size	ST	Plasmid Count	Plasmid Name	Plasmid Size (kb)	Replicon (Plasmid)	KPC allele	Context
1	<i>Klebsiella pneumoniae</i>	KPNIH1	5.4 Mb	258	3	pKpQIL-6e6	113.6	IncF (Q)	3	Tn4401a
2	<i>Klebsiella pneumoniae</i>	KPNIH10	5.4 Mb	258	3	pKpQIL-6e6	113.6	IncF (Q)	3	Tn4401a
3	<i>Enterobacter cloacae</i>	ECNIH3	4.6 Mb	97*	4	pKPC-47e	50.3	IncN (N)	2	IS26-TnpA
4	<i>Enterobacter cloacae</i>	ECNIH5	4.6 Mb	97*	4	pKPC-47e	50.3	IncN (N)	2	IS26-TnpA
5	<i>Klebsiella pneumoniae</i>	KPNIH27	5.2 Mb	34	5	pKEC-dc3	268.3	IncA/C	2	IS26-TnpA
6	<i>Citrobacter freundii</i>	CFNIH1	5.1 Mb	n.d.	1	pKEC-a3c	272.3	IncA/C	2	IS26-TnpA
7	<i>Enterobacter cloacae</i>	ECNIH2	4.9 Mb	93*	3	pKEC-39c	320.0	IncA/C	2	IS26-TnpA
8	<i>Klebsiella pneumoniae</i>	KPNIH24	5.4 Mb	258	3	pKPC-491	47.3	n.d.	3	Tn4401b
9	<i>Enterobacter cloacae</i>	ECR091	4.6 Mb	97*	3	pKPC-272	282.4	IncHI2	3	Tn4401b
10	<i>Klebsiella oxytoca</i>	KONIH1	6.2 Mb	n.d.	3	pKPC-484	85.5	n.d.	2	Tn4401a
11	<i>Klebsiella pneumoniae</i>	KPR0928	5.3 Mb	258	2	pKpQIL-531	113.6	IncF (Q)	2	Tn4401a
12	<i>Enterobacter cloacae</i>	ECNIH4	4.8 Mb	191*	3	pKPC-860	56.6	IncN (N)	2	IS26-TnpA
13	<i>Pantoea</i>	PSNIH1	3.9 Mb	n.d.	4	pKPC-1c5	50.3	IncN (N)	2	IS26-TnpA
14	<i>Klebsiella pneumoniae</i>	KPNIH32	5.5 Mb	258	3	pKPC-def	115.3	IncF	3	Tn4401d
15	<i>Pantoea</i>	PSNIH2	4.3 Mb	n.d.	5	pKPC-c29	54.0	IncN (N)	2	IS26-TnpA
16	<i>Klebsiella pneumoniae</i>	KPNIH33	5.6 Mb	258	3	pKPC-63d	75.6	IncF	3	Tn4401d
					1	chromosome	n.a	n.a	3	Tn4401d
					1	chromosome	n.a	n.a	3	Tn4401d
17	<i>Escherichia coli</i>	ECONIH1	5.3 Mb	648	3	pKPC-629	80.2	IncN (N)	2	Tn4401b
18	<i>Klebsiella pneumoniae</i>	KPNIH30	5.3 Mb	258	3	pKpQIL-531	113.6	IncF (Q)	2	Tn4401a
19	<i>Klebsiella pneumoniae</i>	KPNIH29	5.3 Mb	1518*	2	pKPC-e4e	62.6	IncN (N)	3	Tn4401b
20	<i>Klebsiella pneumoniae</i>	KPNIH31	5.2 Mb	392	3	chromosome	n.a.	n.a.	2	Tn4401a

Genome size corresponds to chromosome. ST indicates the sequence type if a database exists for this species. Sequence types marked by a (\*) are new. The number of different plasmids found in the isolate is indicated in the plasmid count column. The plasmid replicon was identified by the presence of typing genes. Plasmids that are simple variants (SNVs or small indels) of the published sequences (pKpQIL; NC\_014016) are named accordingly. Two common plasmid backbones are indicated by letters in parenthesis after replicon and correspond to Fig. 1. The last two columns indicate whether the KPC-2 or KPC-3 allele is present and its sequence context. Missing values are indicated by n.a. (not applicable) or n.d. (no data).