Nested Russian Doll-like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene $bla_{KPC}$

Running title: Multiple genetic levels of $bla_{KPC}$ mobility

Anna E. Sheppard$^a$, Nicole Stoesser$^a$, Daniel J. Wilson$^a$, Robert Sebra$^b$, Andrew Kasarskis$^b$, Luke W. Anson$^a$, Adam Giess$^a$†, Louise J. Pankhurst$^a$, Alison Vaughan$^a$, Christopher J. Grim$^c$, Heather L. Cox$^d$, Anthony J. Yeh$^d$‡, the Modernising Medical Microbiology (MMM) Informatics Group$^a$,

Costi D. Sifri$^d,e$, A. Sarah Walker$^a$, Tim E. Peto$^a$, Derrick W. Crook$^a,f$, Amy J. Mather$^d,g$*

$^a$ Modernizing Medical Microbiology Consortium, Nuffield Department of Clinical Medicine,
John Radcliffe Hospital, Oxford University, Oxford, United Kingdom.

$^b$ Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of Medicine,
Mount Sinai, New York, New York, USA.

$^c$ Food and Drug Administration, Laurel, Maryland, USA.

$^d$ Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA.

$^e$ Office of Hospital Epidemiology, University of Virginia Health System, Charlottesville, Virginia,
USA.

$^f$ Public Health England, Microbiology Services, London, United Kingdom.

$^g$ Clinical Microbiology, Department of Pathology, University of Virginia Health System,
Charlottesville, Virginia, USA.
* Correspondence to: anna.sheppard@ndm.ox.ac.uk or ajm5b@virginia.edu.

† Present address: Department of Informatics, University of Bergen, Bergen, Norway.

‡ Present address: Pathogen Molecular Genetics Section, Laboratory of Human Bacterial Pathogenesis, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA.

Abstract: 244 words

Main text: 4376 words
Abstract

The recent widespread emergence of carbapenem resistance in Enterobacteriaceae is a major public health concern, as carbapenems are a therapy of last resort in this family of common bacterial pathogens. Resistance genes can mobilize via various mechanisms including conjugation and transposition, however the importance of this mobility in short-term evolution, such as within nosocomial outbreaks, is currently unknown. Using a combination of short- and long-read whole genome sequencing of 281 blaKPC-positive Enterobacteriaceae isolated from a single hospital over five years, we demonstrate rapid dissemination of this carbapenem resistance gene to multiple species, strains, and plasmids. Mobility of blaKPC occurs at multiple nested genetic levels, with transmission of blaKPC strains between individuals, frequent transfer of blaKPC plasmids between strains/species, and frequent transposition of the blaKPC transposon Tn4401 between plasmids. We also identify a common insertion site for Tn4401 within various Tn2-like elements, suggesting that homologous recombination between Tn2-like elements has enhanced the spread of Tn4401 between different plasmid vectors. Furthermore, while short-read sequencing has known limitations for plasmid assembly, various studies have attempted to overcome this with the use of reference-based methods. We also demonstrate that as a consequence of the genetic mobility observed herein, plasmid structures can be extremely dynamic, and therefore these reference-based methods, as well as traditional partial typing methods, can produce very misleading conclusions. Overall, our findings demonstrate that non-clonal resistance gene dissemination can be extremely rapid, presenting significant challenges for public health surveillance and achieving effective control of antibiotic resistance.
Importance

Increasing antibiotic resistance is a major threat to human health, as highlighted by the recent emergence of multi-drug resistant “superbugs”. Here, we tracked how one important multi-drug resistance gene spread in a single hospital over five years. This revealed high levels of resistance gene mobility to multiple bacterial species, which was facilitated by various different genetic mechanisms. The mobility occurred at multiple nested genetic levels, analogous to a Russian doll set where smaller dolls may be carried along inside larger dolls. Our results challenge traditional views that drug-resistance outbreaks are due to transmission of a single pathogenic strain. Instead, outbreaks can be “gene-based”, and we must therefore focus on tracking specific resistance genes and their context rather than only specific bacteria.
Introduction

Although antibiotic resistance genes have been identified in ancient bacterial DNA (1), much of the recent, alarming increase in pathogen antimicrobial resistance is attributable to the dissemination of resistance genes via horizontal gene transfer (HGT), in response to selection imposed by widespread antibiotic use in medicine and agriculture (2, 3). Many resistance genes are located on plasmids, which can be transferred between different bacterial strains or species, thus facilitating HGT (4). Furthermore, resistance gene mobility can be enhanced by integration into transposable elements, which are short stretches of DNA (several kilobases) that can autonomously mobilize between different genomic locations (5). However, the importance of HGT in short-term evolution is unclear, as capturing the processes in real-time is challenging, and outbreaks in health care settings are often thought to be dominated by clonal transmission (6-9).

Carbapenem resistance in Enterobacteriaceae has been recognized as a key threat to modern medicine (10, 11), as carbapenems often represent the therapy of last resort for serious infections (12, 13). One of the most prevalent carbapenem resistance genes is the Klebsiella pneumoniae carbapenemase (KPC) gene, bla\textsubscript{KPC}, first identified in 1996 and now endemic in many regions of the world (14). KPC is a beta-lactamase capable of hydrolyzing all beta-lactams, including penicillins, monobactams, cephalosporins and carbapenems (15), leaving few treatment options for infected vulnerable hospitalized patients and resulting in worse treatment outcomes (16).
Most reports of \textit{bla}\textsubscript{KPC} involve 	extit{K. pneumoniae} multi-locus sequence type (ST)258 (9, 17), which has been found globally, indicating that clonal dissemination of this resistant lineage has been an important factor in the spread of \textit{bla}\textsubscript{KPC} (9, 17-20). Nevertheless, \textit{bla}\textsubscript{KPC} has also been observed in other 	extit{K. pneumoniae} lineages, as well as other species of \textit{Enterobacteriaceae}, suggesting that \textit{bla}\textsubscript{KPC} HGT has also played a role in resistance dissemination (21-25). As \textit{bla}\textsubscript{KPC} is often found on conjugative plasmids, some of which have been identified in multiple strains or species, this provides a likely mechanism for HGT (21, 26, 27). In addition, \textit{bla}\textsubscript{KPC} is usually present as part of the 10 kb composite Tn3-based mobile transposon Tn\textsubscript{4401}, which has been identified in various different plasmids, implicating Tn\textsubscript{4401} transposition as another mechanism contributing to \textit{bla}\textsubscript{KPC} spread (28, 29).

While Tn\textsubscript{4401} transposition and plasmid conjugation have been measured in the laboratory (28, 30, 31), the frequencies with which these processes occur within real-world ecosystems is not fully understood. In clinical contexts, it is often assumed that short-term evolution is dominated by clonal propagation, such that transmission chains generally involve a single pathogenic strain. However, if HGT is frequent relative to transmission (e.g. a "plasmid outbreak"), then linked patients may show variation in strain composition. If transposition is also frequent, then both host strain and resistance plasmid may show high variability within a single outbreak. As current surveillance strategies tend to focus on the host strain, it is important to establish the relevance of \textit{bla}\textsubscript{KPC} mobility within outbreak settings.

Traditional approaches for plasmid investigation, such as PCR-based replicon typing, are limited in resolution. Next-generation sequencing has been successfully applied to molecular
epidemiological investigation of a number of pathogens at the host strain level, however the
application and limitations of this technology for transmission chains involving HGT are
relatively unexplored. Whole genome sequencing using short-read technologies (e.g. Illumina)
has become cheap and accessible, but is not ideal for plasmid analysis due to de novo assembly
limitations, as it is often not possible to accurately reconstruct the genomic context
surrounding repeated sequences (21, 32). Long-read sequencing (e.g. PacBio) can largely
overcome this, often providing single-contig plasmid assemblies, but it is currently prohibitively
expensive for many applications. Several studies have utilised reference-based methods for
plasmid assembly or inference of plasmid structures using short-read data (33, 34), however
these approaches make the implicit assumption that plasmid structures are relatively stable. It
will be important to understand the potential shortcomings of these assumptions in relation to
mobile genetic elements, which may frequently be involved in plasmid rearrangements.
Understanding when and how to successfully apply short and/or long-read sequencing
technologies to molecular epidemiology tracking will be important to the field as the incidence
of HGT is increasingly recognized (35).

In our institution, \textit{bla}_{KPC} was first identified in 2007 in a patient simultaneously colonized with
\textit{bla}_{KPC}-positive \textit{K. pneumoniae} and \textit{Klebsiella oxytoca}, harbouring the \textit{bla}_{KPC} plasmids
pKPC\textsubscript{UVA01} and pKPC\textsubscript{UVA02} respectively (36, 37). Since then, we have been prospectively
screening all \textit{Enterobacteriaceae} species for \textit{bla}_{KPC} carriage despite national guidelines which
recommend screening of \textit{Klebsiella} species and \textit{Escherichia coli} only (38-41). Here we describe
the genetic basis of non-clonal \textit{bla}_{KPC} emergence in a single hospital setting, using a
combination of short- and long-read whole genome sequencing to provide genomic
characterization of 281 *Enterobacteriaceae* isolates from the first five years of this multi-species
*bla*KPC outbreak.

**Results**

There were 204 patients infected/colonized with *bla*KPC-positive *Enterobacteriaceae* during the prospective sampling period, based on clinical and surveillance sampling. We performed short-read Illumina sequencing on all 294 available isolates; 13 of these were excluded due to quality issues (see Methods), leaving 281 isolates, from 182/204 (89%) patients, for analysis (Table S1). In all 281 isolates, *bla*KPC was carried within a complete or partial Tn4401 structure.

*bla*KPC is found in many different host strains, indicating frequent HGT

There were 13 different species carrying *bla*KPC (Figure 1). The four most prevalent species were *Enterobacter cloacae* (96 isolates from 80 patients), *K. pneumoniae* (94 isolates from 55 patients), *Klebsiella oxytoca* (35 isolates from 20 patients), and *Citrobacter freundii* (30 isolates from 25 patients), each of which showed substantial genetic diversity. Across all species, there were a total of 62 distinct strains (>500 chromosomal SNVs; see Methods). Of these, 18 strains were identified in multiple patients, and 44 were only seen in a single patient (Figure 1), with new strains continuing to appear throughout the study period. The very recent emergence of *bla*KPC on an evolutionary timescale (15) implies that each strain likely acquired *bla*KPC independently, demonstrating frequent HGT between different strains and species.
The \textit{bla}\textsubscript{KPC} plasmids pKPC_UVA01 and pKPC_UVA02 are widely dispersed

We hypothesised that the spread of \textit{bla}\textsubscript{KPC} could be due to conjugative transfer of the index \textit{bla}\textsubscript{KPC} plasmids, pKPC_UVA01 and pKPC_UVA02. Defining plasmid presence as $\geq 99\%$ sequence identity over $\geq 80\%$ of the plasmid length, 121 (66\%) and 32 (18\%) patients had isolates carrying pKPC_UVA01 and pKPC_UVA02, respectively, corresponding to 39 and 5 distinct strains from 10 and 4 species, respectively (Figure 1). Although the wide dispersal of these plasmids supports the plasmid-mediated outbreak hypothesis, short-read data is limited in the structural inferences it can provide when repetitive sequences are present, and for many isolates it was not possible to confirm that \textit{bla}\textsubscript{KPC} was actually co-located within pKPC_UVA01 or pKPC_UVA02 (Figure 1).

\textit{bla}\textsubscript{KPC} is found in many different plasmids, indicating frequent Tn4401 transposition

To further investigate \textit{bla}\textsubscript{KPC} plasmid structures, we performed long-read PacBio sequencing on 17 isolates that were chosen at random from the 281 available, yielding closed \textit{bla}\textsubscript{KPC} structures in all cases. Fifteen isolates had a single \textit{bla}\textsubscript{KPC} plasmid and two isolates had two \textit{bla}\textsubscript{KPC} plasmids, giving a total of 19 \textit{bla}\textsubscript{KPC} plasmids from the 17 isolates (Table 1). One isolate additionally had a chromosomal insertion of Tn4401.

From the analysis of Illumina data described above, 11/17 of these isolates contained pKPC_UVA01. As expected, the PacBio assemblies revealed a pKPC_UVA01-like plasmid in each of these isolates. However, only five of these pKPC_UVA01-like plasmids actually contained \textit{bla}\textsubscript{KPC} (Figure 2). The other six pKPC_UVA01-like plasmids lacked the entire Tn4401 element,
which was present on a different plasmid in these isolates. Importantly, this demonstrates that
plasmid presence (as defined by Illumina sequencing) is an unreliable indicator of the mobile
unit carrying \( \text{bla}_{\text{KPC}} \), as the “presence” of \( \text{pKPC}_{\text{UVA01}} \) was misleading in 55% (6/11) of the
randomly selected PacBio isolates.

After accounting for multiple variants of the same plasmid backbone (e.g. the \( \text{pKPC}_{\text{UVA01}} \)-like
plasmids described above), the 19 \( \text{bla}_{\text{KPC}} \) plasmids identified through long-read sequencing
represented 11 distinct plasmid structures (Table 1, Figure S1). These consisted of five
\( \text{pKPC}_{\text{UVA01}} \)-like plasmids, two \( \text{pKPC}_{\text{UVA02}} \)-like plasmids, four \( \text{pKPC}_{\text{CAV1176}} \)-like plasmids,
and eight \( \text{bla}_{\text{KPC}} \) plasmids that were each present in only a single PacBio-sequenced isolate.

Using Illumina data to assess the presence of each of these 11 distinct \( \text{bla}_{\text{KPC}} \) plasmids across
the entire set of isolates as described above revealed varied patterns of plasmid presence
(Figure 1). However, in the majority of cases it was not possible to determine from Illumina
data whether these plasmids contained \( \text{bla}_{\text{KPC}} \), so the precise details regarding distribution of
\( \text{bla}_{\text{KPC}} \)-containing plasmids across the 281 isolates remains elusive.

Taken together, these results demonstrate a great deal of \( \text{bla}_{\text{KPC}} \) plasmid diversity, as 11 distinct
\( \text{bla}_{\text{KPC}} \) plasmids were identified through long-read sequencing of 17 isolates. Given that these
isolates were randomly chosen, the total number of distinct \( \text{bla}_{\text{KPC}} \) plasmids across the entire
set of 281 isolates is likely to be much greater than this. Additional Tn\text{4401} insertion sites were
identified from the subset of isolates where flanking sequences could be adequately assembled
using short-read data, further supporting this hypothesis (Table S2). Therefore, HGT of the
index \( \text{bla}_{\text{KPC}} \) plasmids (pKPC_UVA01 and pKPC_UVA02) only partially explains \( \text{bla}_{\text{KPC}} \) spread, and the large number of distinct \( \text{bla}_{\text{KPC}} \) plasmids indicates high levels of Tn4401 mobility.

**Tn4401 is present within a Tn2-like element in many different plasmids**

In 7 of the 11 distinct, fully characterised, \( \text{bla}_{\text{KPC}} \) plasmids, Tn4401 was surrounded by a sequence element related to the \( \text{bla}_{\text{TEM-1}} \) containing transposon Tn2 (Figure 3). In all cases, the insertion site of Tn4401 within the \( \text{tnpA} \) gene of Tn2 was identical, with approximately 1 kb of flanking sequence on either side of Tn4401 showing 100% identity, but the remainder of these Tn2-like elements showed substantial variation. For example, while the sequence surrounding Tn4401 in pKPC_CAV1176 was identical to the reference Tn2* sequence, the Tn2-like element in pKPC_CAV1043 was truncated. Additionally, pKPC_CAV1344 and pKPC_CAV1596-78 contained a Tn2 derivative, Tn1331, which contains the additional resistance genes \( \text{bla}_{\text{OXA-9}}, \text{aadA1}, \) and \( \text{aac(6')-Ib} \) and has been seen as a prior site of insertion for Tn4401 (42).

**Tn4401 variation**

There were five different structural variants of Tn4401 (Table 2). The majority of isolates, 230/281 (82%), had the Tn4401b isoform, with the remaining isolates containing Tn4401a (n=8), a novel Tn4401 isoform with a 188 bp deletion upstream of \( \text{bla}_{\text{KPC}} \) (n=39), or one of two truncated Tn4401 structures (n=4). At the SNV level, there were seven sites that were variable within Tn4401b. Three of these were located within \( \text{bla}_{\text{KPC}} \), giving rise to three different \( \text{bla}_{\text{KPC}} \) alleles, \( \text{bla}_{\text{KPC-2}} \) (n=179), \( \text{bla}_{\text{KPC-3}} \) (n=44), and \( \text{bla}_{\text{KPC-4}} \) (n=5). All non-Tn4401b isolates contained \( \text{bla}_{\text{KPC-2}} \). Taking all structural and SNV variation into account, there were a total of 12 different
Tn4401 variants. However, most of these were very rare, with seven only found in a single patient.

**bla**<sub>KPC</sub> mobility has occurred within the hospital

Based on prior healthcare exposure, **bla**<sub>KPC</sub> acquisition source was classified as “imported” (likely acquisition prior to admission at our institution) for 15/182 (8%) patients and “local” (likely acquisition within our institution) for 167/182 (92%) patients (Figure 1; see Methods).

Imports were more likely to be infected/colonised with *K. pneumoniae*, particularly ST258 (Table S3), consistent with previous reports of this strain being the dominant **bla**<sub>KPC</sub> carrier in the US (9, 43). Thus, most host strain variation likely originated within the hospital via **bla**<sub>KPC</sub> HGT. In support of this, 15/16 (94%) patients infected/colonised with multiple strains/species had shared Tn4401 variants within the patient (Table S4), suggesting recent **bla**<sub>KPC</sub> HGT.

Notably, this included one patient with two different species carrying Tn4401-6, which is not found in any other patient.

There was also some evidence for recent within-strain Tn4401 transposition. From the isolates that were randomly chosen for long-read sequencing, 4/17 (24%) had multiple Tn4401 copies (Table 1). If we assume that this randomly chosen subset is representative, this extrapolates to approximately 66/281 isolates across the whole dataset. However, only 2/281 isolates had multiple Tn4401 variants (Tn4401-11; Table 2), indicating that many isolates likely had multiple copies of the same Tn4401 variant, consistent with recent Tn4401 transposition.
Taken together, these results indicate that much of the genetic diversity observed is due to recent $bla_{KPC}$ mobility, likely within the hospital ecosystem over the described five year outbreak.

**Direct patient-to-patient transmission does not explain $bla_{KPC}$ acquisition**

To further investigate $bla_{KPC}$ acquisition source, we combined epidemiological and genetic data to trace possible transmission chains, at two different genetic levels. We considered possible transmission events where the donor and recipient were on the same ward at the same time, and carried the same host strain or Tn4401 variant. Considering only “local” acquisitions (see above), 48/167 (29%) patients had ward contact with another patient carrying the same $bla_{KPC}$-positive strain (Figure 4, top panel). A greater proportion, 106/167 (63%), of patients had ward contact with another patient carrying the same Tn4401 variant. However, as Tn4401-1 is very common (66% of patients), these inferred transmissions may be spurious. With patients carrying this common variant excluded, only 15/50 (30%) had ward contact with another patient carrying the same Tn4401 variant (Figure 4, bottom panel). Therefore, both genetic levels (strain or Tn4401 variant) demonstrated plausible transmissions for only a minority of patients, indicating that direct patient-to-patient transmission is not the dominant mode of $bla_{KPC}$ acquisition.
Discussion

Here we have demonstrated high levels of genetic diversity in KPC-producing Enterobacteriaceae within a single institution over five years. This diversity occurs at multiple genetic levels, revealing a complex evolutionary history of the \textit{bla}_{KPC} gene involving many different host strains and plasmids.

In 7/11 distinct \textit{bla}_{KPC} plasmids identified through long-read sequencing, Tn4401 was located within a Tn2-like element. As these Tn2-like elements differed substantially from each other (Figure 3), it is unlikely that this arose via transposition of a composite Tn4401-Tn2-like structure. Instead, it suggests that Tn4401 has been repeatedly incorporated into pre-existing Tn2-like elements, which are known to be widespread, and genetically divergent, in \textit{Enterobacteriaceae} \cite{44, 45}. However, the insertion site was identical in all cases, yet Tn4401 has been reported to have no insertion site specificity \cite{28}, suggesting that this was not facilitated by a standard transposition mechanism. Therefore, we suggest that this is most likely mediated by homologous recombination with other Tn2-like elements following an initial integration event, as recently suggested for another multi-drug resistance gene, \textit{bla}_{CTX-M-15} \cite{46}.

This implies that Tn4401 mobility may have been enhanced via integration into a second, already widely dispersed, transposon. As the Tn4401-Tn2-like structure was present in the index case isolate (CAV1016, Aug 2007) we presume that the initial transposition of Tn4401 into a Tn2-like element occurred prior to entry into our hospital system. In support of this, one particular Tn2-like element, Tn1331, has been previously described to contain Tn4401 (in exactly the same position within the \textit{tnpA} gene as described here) \cite{21, 42, 47, 48}, including
one report describing a *K. pneumoniae* isolated in 2005, which predates *bla*KPC in our institution (42). We are not aware of any previous reports describing Tn4401 within a non-Tn1331 Tn2-like element.

The prevalence of Tn4401 insertions within Tn2-like elements also has important implications with regard to plasmid tracking. We previously published a method for arbitrary PCR to track the flanking regions around the Tn4401 element, as well as a PCR method to assay presence of what we had wrongly assumed was a single plasmid, pKPC_UVA01. This PCR assay targeted the immediate Tn4401 insertion site within a Tn2-like element (49), which we have here demonstrated is present in many different plasmids, highlighting that PCR assays, and indeed any partial typing methods, need to be interpreted with a great deal of caution. We were further mislead by the analysis of short-read whole genome sequencing data, which indicated presence of pKPC_UVA01 in the majority of isolates. Taken together it was tempting to conclude that horizontal transfer of pKPC_UVA01 was responsible for the vast majority of *bla*KPC carriage in our institution. However, long-read sequencing refuted this, revealing a far more complex picture.

More generally, this highlights certain limitations for plasmid reconstruction from short-read data. To illustrate by way of example, there were five isolates where long-read sequencing revealed pKPC_UVA01-like plasmids that were identical to the reference pKPC_UVA01 sequence apart from the absence of Tn4401 and associated 5 bp target site duplication (Figure 2). We presume that in these lineages, *bla*KPC may have been initially acquired via HGT of pKPC_UVA01, with subsequent homologous recombination transferring Tn4401 from...
pKPC_UVA01 to a different plasmid containing a Tn2-like element. In each of these five isolates, multiple Tn2-like elements are present, which have 100% sequence identity over approximately 1 kb on either side of the Tn4401 insertion site. As this is longer than the fragment length used for paired-end sequencing, it is not possible to resolve the plasmid context of bla\textsubscript{KPC} using short-read data. Importantly, any reference-based method for plasmid reconstruction (e.g. in this case using the pKPC_UVA01 reference sequence to infer presence of the plasmid in each isolate) is liable to produce misleading results. More generally, it is exactly the repetitive regions that cannot be resolved using short-read data that could be expected to be involved in plasmid rearrangements, either through homologous recombination as suggested here, or by virtue of the fact that transposable elements are often present in multiple copies. Therefore, having short-read data that is consistent with a known plasmid structure, even within the same outbreak, should not be sufficient to conclude that that structure is present, if the data is also consistent with an alternative structure. As several recent studies have utilised reference-based approaches for plasmid assembly / inference (33, 34), our results indicate that any such methods should be interpreted with extreme caution.

Across the bla\textsubscript{KPC}-positive patients, there was large variation in both host strains and bla\textsubscript{KPC} plasmids, with Tn4401 being the largest genetic unit that was consistently present. Therefore, surveillance strategies aimed at tracking individual strains or plasmids could be misleading, and it may be more appropriate to focus on Tn4401. However, we found limited variation within the transposon, as Tn4401 sequences from 121/182 (66%) patients were identical to the index case (Table 2). This lack of variation implies that even the highest resolution genetic methods may be insufficient for determining specific transmission routes. Even so, we have demonstrated that
only a minority of \textit{bla}_{\textit{KPC}} acquisition events can be explained by direct patient-to-patient transmission. Future studies should therefore contemporaneously investigate the possible involvement of unsampled reservoirs (e.g. environmental or silent colonization by additional carriers).

There several limitations to this study. Because of the cost and effort involved in long-read sequencing, we were only able to resolve a minority of \textit{bla}_{\textit{KPC}}-plasmids. This means that although we have a compelling indicator of the diversity created by mobile genetic elements within a single hospital over a five year period, we are limited in the ability to genetically resolve pathways of \textit{bla}_{\textit{KPC}} mobility between host strains and plasmid vectors, even within a single patient. We also speculate about the effect of Tn\textsubscript{4401} insertion into Tn2-like elements, but future \textit{in vitro} studies could be used to illuminate the effect of this composite structure on Tn\textsubscript{4401} mobility.

In conclusion, our detailed genetic analysis of the evolutionary events occurring in the early stages of antimicrobial resistance gene emergence in a single institution identifies several distinct processes occurring at high frequency (Figure 5). First, the presence of shared \textit{bla}_{\textit{KPC}}-containing strains in different patients reflects traditional (clonal) outbreak models. Second, \textit{bla}_{\textit{KPC}} mobility between strains/species is facilitated by promiscuous \textit{bla}_{\textit{KPC}} plasmids such as p\textit{KPC} \_\textit{UVA}01. Third, \textit{bla}_{\textit{KPC}} transfer between plasmids is likely enhanced by homologous recombination between Tn2-like elements, facilitating the movement of Tn\textsubscript{4401} from one plasmid to another. Finally, \textit{bla}_{\textit{KPC}} mobility is also enabled by standard Tn\textsubscript{4401} transposition. Rather than a single process dominating, resistance dissemination is driven by a combination of
these factors, resulting in a high level of diversity in KPC-producing Enterobacteriaceae, at multiple genetic levels. As $\text{bla}_{\text{KPC}}$ prevalence continues to increase, so will this genetic diversity, inevitably resulting in a wider variety of more pathogenic strains carrying $\text{bla}_{\text{KPC}}$.

Our results indicate that the current standard practice of only screening specific species for $\text{bla}_{\text{KPC}}$ carriage is likely to hamper surveillance efforts by grossly underestimating true prevalence. Instead of the traditional view of an outbreak involving a single pathogenic strain, we propose that for KPC-producing Enterobacteriaceae, and possibly more generally, we should instead adopt the view of a “gene-based outbreak”, with surveillance strategies tracking the resistance gene itself rather than a specific host strain.
Methods

Isolate collection and Illumina sequencing

A subset of *K. pneumoniae* isolates, with corresponding sequence data, have been previously described (37). Isolate collection, *de novo* assembly, mapping and variant calling were performed as previously described (37), however here we used species-specific references for mapping (Table S5). Illumina sequencing was also performed as previously described (37), with some exceptions (see Supplementary Methods). In total, 281 isolates from 182 patients were available for analysis; exclusion criteria for additional isolates is described in Supplementary Methods.

Species classification

Species classification was performed using microbiological and sequenced-based methods (see Supplementary Methods for details).

Phylogenetic analysis and strain classification

There were 52 patients with multiple isolates of the same species. One of these (patient FK) involved two strains of *K. pneumoniae* that were highly divergent from each other (>20,000 chromosomal SNVs), clearly representing a separate *blaKPC* acquisition by each strain. Excluding this divergent strain pair, the remaining cases had SNV differences ranging from 0 to 60 (median 2 SNVs). As these could plausibly represent clonal evolution within the patient, we conservatively chose to include only a single representative (the earliest isolate) for phylogenetic reconstruction, in order to avoid artificially inflating genetic clusters due to
repeated patient sampling. Phylogenetic analysis was then performed separately for each species using PhyML (50) (see Supplementary Methods). Chromosomally distinct strains were defined by partitioning each phylogeny with a cutoff of ~500 SNVs (see Supplementary Methods). Based on the molecular clock of Enterobacteriaceae (1-20 SNVs/chromosome/year) (6, 37, 51), we can be relatively confident that isolates belonging to distinct strains will not have a shared ancestor within the timeframe of \( \text{bla}_{\text{KPC}} \) dispersal, and the number of distinct strains thus provides a conservative estimate of the number of distinct \( \text{bla}_{\text{KPC}} \) acquisition events.

**Long-read PacBio sequencing**

For long-read sequencing, 17 isolates were randomly chosen from the entire set of sequenced isolates (i.e. including patient duplicates). Long-read sequencing and initial de novo assembly were performed as previously described (37). Refinement of assemblies and closure of plasmid/chromosomal sequences was performed as described in Supplementary Methods. Since the isolates for PacBio sequencing were randomly chosen from the set of all Illumina sequenced isolates, some of them represented within-patient strain duplicates (see previous section on phylogenetic analysis), and were therefore not included in phylogenetic reconstruction. For display purposes (in Figure 1), the \( \text{bla}_{\text{KPC}} \) structure(s) determined from long-read PacBio sequencing for each of these isolates is shown alongside the representative isolate of the same strain from the same patient. In all cases, the representative isolate has the same short-read plasmid profile and Tn4401 variant as the PacBio sequenced isolate.

**Plasmid presence / absence classification**
The index \textit{bla}_{KPC} plasmids pKPC_UVA01 and pKPC_UVA02, together with the additional nine distinct \textit{bla}_{KPC} plasmids identified though long-read PacBio sequencing, were used as references to determine plasmid presence profiles for each isolate based on the Illumina data. Plasmid presence was defined as \geq 99\% sequence identity over \geq 80\% of the length of the reference sequence, as determined by BLASTn comparisons between each isolate’s \textit{de novo} assembly and the reference plasmid. The high identity cutoff was chosen to reduce false positives from sequences that are only distantly related (and therefore unlikely to have a common ancestor within the timeframe of the outbreak), while the more permissive length cutoff allows for some rearrangement. It should be noted that the method does not take any account of structural continuity.

\textbf{Analysis of Tn4401 flanking sequences}

Where a plasmid was classified as being present in a particular isolate, it was not always certain to contain Tn4401. The plasmid presence classification was further refined as: “containing Tn4401” if the isolate’s \textit{de novo} assembly supported Tn4401 being present within the expected sequence context of that plasmid, “not containing Tn4401” if the plasmid was assembled without Tn4401, or “uncertain” if structure could not be determined from the \textit{de novo} assembly. The identification of novel Tn4401 insertion sites was also based on the \textit{de novo} assemblies. These methods are described in detail in Supplementary Methods.

\textbf{Variation in Tn4401}

Tn4401 isoform classification was performed by comparing each isolate’s \textit{de novo} assembly with the previously described isoform \textit{b} reference sequence from EU176013.1 (29) using
BLASTn, to identify structural variation. SNV variation was determined by mapping to a reference consisting of pKPC_UVA01 plus a species-specific chromosome as described above, followed by extraction of the Tn4401 region. Variation is reported for all sites where at least one isolate had a non-reference call, including any ambiguity at that site in other isolates. Ambiguity at non-variable sites is not reported, which may result in an underestimate of true variation. However, any resulting underestimation is likely to be very minor, as the proportion of called sites, excluding deleted regions described above, was >96% for all isolates.

Epidemiological classification

For epidemiologic analysis, patients were assigned a one or two letter code for de-identification. Routine peri-rectal surveillance cultures for silent colonization began in April 2009 (38, 40). Patients were classified as “imported” if they did not have any prior admission to University of Virginia Medical Center/Long-term Acute Care Hospital (UVaMC) and either had a blaKPC-positive Enterobacteriaceae isolated within 48 hours of admission, or had a carbapenem-resistant Enterobacteriaceae culture before transfer to UVaMC with a subsequent isolate at UVaMC confirmed as blaKPC PCR positive. The index case was also classified as imported. For the remaining patients, the source of blaKPC acquisition was classified as “local”. The 48h cutoff is arbitrary and may result in some misclassification if patients either acquire blaKPC within the first 48h of admission, or if blaKPC carriage/infection remains undetected for >48h, however this is expected to be minimal (see Supplementary Methods). Charts and patient contacts were reviewed using bed tracing data and the electronic medical record. The study was approved by the University of Virginia Institutional Review Board (protocol # 13558).
Transmission analysis

Possible patient-to-patient transmission events were determined on the basis of having overlapping stays on the same ward, as well as genetically-related \( \text{bla}_{\text{KPC}} \) isolates. The analysis was performed separately for two different levels of genetic relatedness (strain or Tn4401 variant). This is described in detail in Supplementary Methods.

Funding Information

This publication presents independent research commissioned by the Health Innovation Challenge Fund (grant HICF-T5-358 and WT098615/Z/12/Z), a parallel funding partnership between the Department of Health and Wellcome Trust, the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust and University of Oxford, and NIHR Oxford Health Protection Research Unit on Healthcare Associated Infection and Antimicrobial Resistance (HPRU-2012-10041). The views expressed in this publication are those of the author(s) and not necessarily those of the funders. DWC and TEP are NIHR Senior investigators. NS is supported by a Wellcome Trust University of Oxford research fellowship.

Acknowledgements

We thank the UVaMC Clinical Microbiology staff for collection of study isolates and the UVaMC Infection Prevention and Control staff for assistance with patient tracking.
Modernising Medical Microbiology (MMM) informatics group

Jim Davies, Charles Crichton, Milind Acharya, Carlos del Ojo Elias

Additional Information

Sequence data has been deposited with the National Centre for Biotechnology Information (NCBI) under BioProject PRJNA246471.
References


23. **Ocampo AM, Chen L, Cienfuegos AV, Roncancio G, Chavda KD, Kreiswirth BN, Jiménez JN.**


Figure Legends

Figure 1. Diversity in bacterial species, strains, plasmids and Tn4401 variants. For each species, a phylogeny was generated from mapping to a species-specific chromosomal reference, after deduplication of closely related isolates from the same patient (see Methods). Distinct strains are defined by a cutoff of ~500 SNVs (see Methods); strains found in more than one patient are indicated by grey background shading. Circles show plasmid “presence” as determined from Illumina data, with fill colour indicating uncertainty in whether the plasmid contains blaKPC. Boxes show plasmid structure determined from long-read PacBio sequencing for 17 randomly chosen isolates, as well as the previously sequenced isolates from index patient B (37). Where the PacBio-sequenced isolate was excluded from the phylogeny as a patient duplicate, plasmid structure is shown for the corresponding closely related isolate from the same patient. Tn4401 and blaKPC variants (Table 2) are indicated by large and small squares respectively. The likely source of blaKPC acquisition as determined from epidemiological data is indicated by text colour.

Figure 2. pKPC_UVA01-like plasmids identified through long-read PacBio sequencing. The reference pKPC_UVA01 sequence is shown, together with all 11 pKPC_UVA01-like plasmids identified through long-read PacBio sequencing, including the six that do not contain blaKPC. Arrows indicate predicted open reading frames; Tn4401 is shown in purple. Pink shading indicates regions of identity between adjacent sequences, with SNVs indicated by red lines.
**Figure 3.** Tn4401 is commonly integrated into a Tn2-like element. The Tn4401 and surrounding region (i.e. partial plasmid sequence, except for pKPC_CAV1320) is shown for each distinct blakPC plasmid. Variants of the same plasmid backbone (see Table 1) are not shown. Arrows indicate predicted open reading frames; Tn4401 is shown in purple. Pink shading indicates regions of identity between adjacent sequences, with SNVs indicated by red lines and short indels (1-2 bp) by blue lines. The top panel shows the Tn2* reference sequence from AY123253 (44).

**Figure 4.** Ward contacts between patients with genetically related isolates. Each horizontal line represents a different strain (top) or Tn4401 variant (bottom). Filled circles indicate patients that had previous ward contact with another patient on the same horizontal line (i.e. possible patient-to-patient transmission). As Tn4401-1 is present in two-thirds of patients, many coincidental ward contacts may be expected to occur, resulting in a substantial overestimate of transmission number. Therefore, for Tn4401, the total number of acquisitions explainable by direct ward contact is indicated, as well as with Tn4401-1 patients excluded. The vertical line indicates onset of routine patient screening.

**Figure 5.** blakPC spreads at multiple genetic levels, resulting in a high level of diversity in blakPC-positive Enterobacteriaceae.
Table 1. **bla**<sub>KPC</sub>-containing structures ascertained from long-read PacBio sequencing of 17 randomly chosen isolates

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Species</th>
<th>Patient</th>
<th>Date</th>
<th><strong>bla</strong>&lt;sub&gt;KPC&lt;/sub&gt; plasmid</th>
<th>Size (bp)</th>
<th>Group&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Within-group genetic changes&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Tn4401 variant</th>
<th>Flanking sequence&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Tn2-like element&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAV1344</td>
<td>K. pneumoniae</td>
<td>EP</td>
<td>Dec-2010</td>
<td>pKPC_CAV1344</td>
<td>176,497</td>
<td>Singleton</td>
<td>NA</td>
<td>Tn4401-1</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1392</td>
<td>K. pneumoniae</td>
<td>EU</td>
<td>Mar-2011</td>
<td>pKPC_CAV1392</td>
<td>43,621</td>
<td>NA (chromosomal)</td>
<td>NA</td>
<td>Tn4401-5</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1596</td>
<td>K. pneumoniae</td>
<td>FK</td>
<td>Apr-2012</td>
<td>pKPC_CAV1596-78</td>
<td>77,801</td>
<td>Singleton</td>
<td>NA</td>
<td>Tn4401-5</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1099</td>
<td>K. oxytoca</td>
<td>AU</td>
<td>Apr-2009</td>
<td>pKPC_CAV1099</td>
<td>113,105</td>
<td>NA</td>
<td>1 SNV</td>
<td>Tn4401-1</td>
<td>ATGCA...GGGCA&lt;sup&gt;e&lt;/sup&gt;</td>
<td>No</td>
</tr>
<tr>
<td>CAV1335</td>
<td>K. oxytoca</td>
<td>EQ</td>
<td>Dec-2010</td>
<td>pKPC_CAV1335</td>
<td>113,105</td>
<td>NA</td>
<td>0 SNVs</td>
<td>Tn4401-1</td>
<td>ATGCA...GGGCA&lt;sup&gt;e&lt;/sup&gt;</td>
<td>No</td>
</tr>
<tr>
<td>CAV1374</td>
<td>K. oxytoca</td>
<td>ED</td>
<td>Aug-2010</td>
<td>pKPC_CAV1374</td>
<td>332,956</td>
<td>NA</td>
<td>0 SNVs</td>
<td>Tn4401-1</td>
<td>ATGCA...GGGCA&lt;sup&gt;e&lt;/sup&gt;</td>
<td>No</td>
</tr>
<tr>
<td>CAV1043</td>
<td>E. asburiae</td>
<td>L</td>
<td>Mar-2008</td>
<td>pKPC_CAV1043</td>
<td>59,138</td>
<td>NA</td>
<td>NA</td>
<td>Tn4401-5</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1176</td>
<td>E. cloacae</td>
<td>DN</td>
<td>May-2010</td>
<td>pKPC_CAV1176</td>
<td>90,452</td>
<td>NA</td>
<td>0 SNVs</td>
<td>Tn4401-3</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1311</td>
<td>E. cloacae</td>
<td>EO</td>
<td>Jan-2011</td>
<td>pKPC_CAV1311</td>
<td>90,452</td>
<td>NA</td>
<td>0 SNVs</td>
<td>Tn4401-3</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1411</td>
<td>E. cloacae</td>
<td>FC</td>
<td>Jun-2011</td>
<td>pKPC_CAV1411</td>
<td>90,452</td>
<td>NA</td>
<td>1 SNV, 40 kb inversion</td>
<td>Tn4401-3</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1669</td>
<td>E. cloacae</td>
<td>HV</td>
<td>Aug-2012</td>
<td>pKPC_CAV1669</td>
<td>90,452</td>
<td>NA</td>
<td>40kb inversion</td>
<td>Tn4401-3</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1668</td>
<td>E. cloacae</td>
<td>HQ</td>
<td>Aug-2012</td>
<td>pKPC_CAV1668</td>
<td>43,433</td>
<td>NA</td>
<td>1 SNV, 188 bp deletion</td>
<td>Tn4401-3</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1321</td>
<td>C. freundii</td>
<td>EG</td>
<td>Nov-2010</td>
<td>pKPC_CAV1321-45</td>
<td>44,846</td>
<td>NA</td>
<td>1,225 bp insertion</td>
<td>Tn4401-1</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1321</td>
<td>C. freundii</td>
<td>ER</td>
<td>Oct-2012</td>
<td>pKPC_CAV1321-244</td>
<td>243,709</td>
<td>Singleton</td>
<td>14,960 bp duplication, 70,615 bp insertion</td>
<td>Tn4401-1&lt;sup&gt;f&lt;/sup&gt;</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1151</td>
<td>K. intermedia</td>
<td>CD</td>
<td>Sep-2009</td>
<td>pKPC_CAV1151</td>
<td>43,621</td>
<td>NA</td>
<td>0 SNVs&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Tn4401-1</td>
<td>GTTCT...GTTC</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1330</td>
<td>E. aerogenes</td>
<td>EL</td>
<td>Nov-2010</td>
<td>pKPC_CAV1330</td>
<td>13,981</td>
<td>Singleton</td>
<td>NA</td>
<td>Tn4401-1</td>
<td>TTTGTT...TTGT</td>
<td>No</td>
</tr>
<tr>
<td>CAV1492</td>
<td>S. marcescens</td>
<td>GL</td>
<td>Dec-2011</td>
<td>pKPC_CAV1492</td>
<td>69,158</td>
<td>NA</td>
<td>NA</td>
<td>Tn4401-8</td>
<td>TTTT...TTTT</td>
<td>No</td>
</tr>
</tbody>
</table>

<sup>a</sup> Plasmids are defined as belonging to the same group if the sequences are largely identical, allowing for a small number of substitutions and/or rearrangements that may be expected to occur within the outbreak timeframe. Different groups have very limited homology outside the Tn4401 region, indicative of independent integrations into distinct plasmid structures. 'Singleton' indicates a plasmid backbone that is distinct from all others shown.

<sup>b</sup> Differences relative to the reference sequence for that plasmid group, as specified in the previous column.

<sup>c</sup> Sequences immediately flanking Tn4401; generally expected to be identical due to 5 bp target site duplication during transposition (28).

<sup>d</sup> Tn4401 integrated into the tnpA gene of a Tn2-like element.

<sup>e</sup> No evidence of target site duplication.

<sup>f</sup> 2 copies.

<sup>g</sup> It is noteworthy that this plasmid from CAV1151 (K. intermedia) is exactly identical to pKPC_UVA01 from CAV1016 (K. pneumoniae), with isolation dates 2 years apart.

---

*The copyright holder for this preprint (which was not peer-reviewed online) is the author/funder. It is made available under a CC-BY 4.0 International license.*
Table 2. Tn4401 variation

<table>
<thead>
<tr>
<th>Tn4401 variant</th>
<th>Structural isoform(29)</th>
<th>SNVs&lt;sup&gt;a&lt;/sup&gt;</th>
<th>bla&lt;sub&gt;KPC&lt;/sub&gt; variant</th>
<th>Patients</th>
<th>Isolates</th>
<th>Strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tn4401-1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>b</td>
<td>-</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-2</td>
<td>121</td>
<td>176</td>
<td>42</td>
</tr>
<tr>
<td>Tn4401-2</td>
<td>a (del 7020-7118)</td>
<td>-</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-2</td>
<td>5</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Tn4401-3</td>
<td>novel (del 6919-7106)</td>
<td>-</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-2</td>
<td>28</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td>Tn4401-4</td>
<td>truncated (del 1-6654)</td>
<td>-</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-2</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Tn4401-5</td>
<td>b</td>
<td>8015 C→T&lt;sup&gt;c&lt;/sup&gt;</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-3</td>
<td>22</td>
<td>40</td>
<td>19</td>
</tr>
<tr>
<td>Tn4401-6</td>
<td>b</td>
<td>8015 C→T, 9621 T→C</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-3</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Tn4401-7</td>
<td>b</td>
<td>7199 T→A, 8015 C→T, 9621 T→C</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tn4401-8</td>
<td>b</td>
<td>9663 T→C</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-2</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Tn4401-9</td>
<td>b</td>
<td>7509 C→G, 7917 T→G&lt;sup&gt;d&lt;/sup&gt;</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tn4401-10</td>
<td>b</td>
<td>6800 T→C, 7509 C→G, 7917 T→G</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-4</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Tn4401-11</td>
<td>b</td>
<td>8015 N&lt;sup&gt;e&lt;/sup&gt;</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-2 / bla&lt;sub&gt;KPC&lt;/sub&gt;-3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Tn4401-12</td>
<td>truncated (del 1-6727)</td>
<td>6800 N&lt;sup&gt;e&lt;/sup&gt;</td>
<td>bla&lt;sub&gt;KPC&lt;/sub&gt;-2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup> With respect to Tn4401-1, which is considered as the reference Tn4401 sequence for this study


<sup>c</sup> This substitution converts bla<sub>KPC</sub>-2 to bla<sub>KPC</sub>-3

<sup>d</sup> These two substitutions convert bla<sub>KPC</sub>-2 to bla<sub>KPC</sub>-4

<sup>e</sup> Quality filters failed at this position due to a mixture of reads supporting C and T (i.e. Tn4401-11 actually represents a mixture of Tn4401-1 and Tn4401-5)

<sup>f</sup> Quality filters failed at this position due to a lack of reads mapped in the reverse direction. All reads mapped in the forward direction supported a reference (T) call

---

<sup>658</sup> Table 2. Tn4401 variation

---

<sup>659</sup> With respect to Tn4401-1, which is considered as the reference Tn4401 sequence for this study


<sup>662</sup> A→T

<sup>663</sup> A→T

<sup>664</sup> Quality filters failed at this position due to a mixture of reads supporting C and T (i.e. Tn4401-11 actually represents a mixture of Tn4401-1 and Tn4401-5)

<sup>665</sup> Quality filters failed at this position due to a lack of reads mapped in the reverse direction. All reads mapped in the forward direction supported a reference (T) call
Supplementary Legends

Figure S1. Distinct bla_{KPC} plasmids identified through long-read PacBio sequencing. Variants of the same plasmid backbone (see Table 1) are not shown. Arrows indicate predicted open reading frames; Tn4401 is shown in purple.

Table S1. Details of sequenced isolates.

Table S2. Additional Tn4401 insertion sites ascertained from short-read Illumina data

Table S3. Association of importation status with K. pneumoniae and the epidemic bla_{KPC} K. pneumoniae strain ST258

Table S4. Patients with multiple bla_{KPC}-positive strains or species

Table S5. Chromosomal references used for mapping
pKPC_CAV1741
pKPC_CAV1321-45
pKPC_CAV1668
pKPC_CAV1392
pKPC_CAV1151
pKPC_UVA01
pCAV1669-34
pCAV1411-34
pCAV1311-34
pCAV1176-34
pCAV1374-34
pCAV1344-40

pKPC_UVA01-like plasmids with blaKPC

pKPC_UVA01-like plasmids without blaKPC

Antibiotic resistance
Transposition / Recombination
Conjugation / Type IV secretion
Replication
Hypothetical protein
Other
Tn4401

10 kb
Patients with ward contact

48/167 (29%)

Host strain

Tn4401-1

Tn4401-4

Tn4401-5

Tn4401-2

Tn4401-12

Tn4401-3

Tn4401-6

Tn4401-7

Tn4401-11

Tn4401-9

Tn4401-10

Tn4401-8

Patients with ward contact

15/50 (30%)

106/167 (63%)

Patients with ward contact

48/167 (29%)

All local acquisitions 106/167 (63%)

Excluding Tn4401-1 15/50 (30%)

Tn4401-1

Klebsiella pneumoniae

Klebsiella oxytoca

Enterobacter cloacae

Citrobacter freundii

Other

Imported

Ward contact, same Tn4401 variant

Ward contact, different Tn4401 variant

No ward contact

1 Jan 2008 1 Jan 2009 1 Jan 2010 1 Jan 2011 1 Jan 2012 1 Jan 2013

CC-BY 4.0 International license peer-reviewed) is the author/funder. It is made available under a http://dx.doi.org/10.1101/033522 doi: bioRxiv preprint first posted online Dec. 2, 2015;
Klebsiella pneumoniae carbapenemase (KPC) resistance gene

Transposition of Tn4401 into different plasmids

Proposed mobilization of Tn4401 between Tn2-like elements in different plasmids, via homologous recombination

Plasmid movement via conjugation into different bacterial strains and species

Mobile transposon Tn4401

Tn4401 nested within mobile Tn2-like element

Transposon located within mobile extra-chromosomal plasmid DNA

Bacterial cell with mobile plasmid and immobile host chromosome (not to scale)