Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene bla\textsubscript{KPC}

Anna E. Sheppard, Nicole Stoesser, Daniel J. Wilson, Robert Sebra, Andrew Kasarskis, Luke W. Anson, Adam Giess, Louise J. Pankhurst, Alison Vaughan, Christopher J. Grim, Heather L. Cox, Anthony J. Yeh, the Modernising Medical Microbiology (MMM) Informatics Group, Costi D. Sifiri, A. Sarah Walker, Tim E. Petro, Derrick W. Crook, Amy J. Mathers

Modernizing Medical Microbiology Consortium, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford University, Oxford, United Kingdom; Icahn Institute and Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York, New York, USA; Food and Drug Administration, Laurel, Maryland, USA; Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA; Office of Hospital Epidemiology, University of Virginia Health System, Charlottesville, Virginia, USA; Public Health England, Microbiology Services, London, United Kingdom; Clinical Microbiology, Department of Pathology, University of Virginia Health System, Charlottesville, Virginia, USA

The recent widespread emergence of carbapenem resistance in Enterobacteriaceae is a major public health concern, as carbapenems are a therapy of last resort against 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 unknown. Using a combination of short- and long-read whole-genome sequencing of 281 bla\textsubscript{KPC}-positive Enterobacteriaceae isolates from a single hospital over 5 years, we demonstrate rapid dissemination of this carbapenem resistance gene to multiple species, strains, and plasmids. Mobility of bla\textsubscript{KPC} occurs at multiple nested genetic levels, with transmission of bla\textsubscript{KPC} between strains during individuals, frequent transfer of bla\textsubscript{KPC} plasmids between strains/species, and frequent transposition of bla\textsubscript{KPC} 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 expanded 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 by the use of reference-based methods. We also demonstrate that, as a consequence of the genetic mobility observed in this study, 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 nonclonal resistance gene dissemination can be extremely rapid, presenting significant challenges for public health surveillance and achieving effective control of antibiotic resistance.

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 that 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 to 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 bla\textsubscript{KPC} involve K. pneumoniae multilocus sequence type 258 (ST258) (9, 17), which has been found globally, indicating that clonal dissemination of this resistant lineage has been an important factor in the spread of bla\textsubscript{KPC} (9, 17–20). Nev-
ertheless, blanKPC has also been observed in other K. pneumoniae lineages, as well as other species of Enterobacteriaceae, suggesting that blanKPC HGT has also played a role in resistance dissemination (21–25). As blanKPC 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, blanKPC is usually present as part of the 10-kb Tn3-based mobile transposon Tn4401, which has been identified in various different plasmids, implicating Tn4401 transposition as another mechanism contributing to blanKPC spread (28, 29).

While Tn4401 transposition and plasmid conjugation have been measured in the laboratory (28, 30, 31), the frequencies with which these processes occur within real-world ecosystems are 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 the host strain and the 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 blanKPC mobility within outbreak settings.

Traditional approaches to 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 because of 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 prohibitively expensive for many applications. Several studies have utilized 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 approaches 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).

At our institution, blanKPC was first identified in 2007 in a patient simultaneously colonized with blanKPC-positive K. pneumoniae and Klebsiella oxytoca harboring blanKPC plasmids pKPC_UVA01 and pKPC_UVA02, respectively (36, 37). Since then, we have prospectively screened extended-spectrum cephalosporin-resistant/carbapenem-nonsusceptible isolates of all Enterobacteriaceae species for blanKPC despite national guidelines that recommend screening for carbapenem-nonsusceptible Klebsiella species and Escherichia coli (38–41). Here we describe the genetic basis of nonclonal blanKPC emergence in a single hospital setting by using a combination of short- and long-read whole-genome sequencing methods to provide genomic characterization of 281 Enterobacteriaceae isolates from the first 5 years of this multispecies blanKPC outbreak.

MATERIALS AND METHODS

Isolate collection and Illumina sequencing. Isolates were prospectively collected from August 2007 to December 2012 through the Clinical Microbiology Laboratory of the University of Virginia Health System, which serves a 619-bed tertiary care hospital, outpatient clinics in central Virginia, and since August 2010, a 40-bed long-term acute care hospital. From April 2009, weekly surveillance by perirectal swab was performed in all inpatient units with historically high transmission or where there was a patient who was known to be colonized or infected with carbapenemase-producing Enterobacteriaceae (CPE) by previously described methods (40, 42, 43). Enterobacteriaceae isolates from nonsurveillance clinical samples that were flagged as possible extended-spectrum β-lactamase (ESBL) producing or had an ertapenem MIC of ≥1 μg/ml by automated susceptibility profiling (VITEK2; bioMérieux, Durham, NC) underwent carbapenemase phenotypic testing by the modified Hodge test (August 2007 to June 2008) or the indirect carbapenemase test (July 2008 to December 2012). Isolates with a positive carbapenemase phenotypic test and/or a meropenem or imipenem MIC of ≥1 μg/ml underwent blanKPC PCR analysis as previously described (39).

A subset of 37 K. pneumoniae isolates, with corresponding sequence data, have been previously described (37). For the rest of the study isolates, Illumina sequencing, de novo assembly, mapping, and variant calling were performed as previously described (37), with some exceptions (see the supplemental material), and including the use of additional, species-specific references for mapping (see Table S3 in the supplemental material). A total of 281 isolates from 182 patients were available for analysis; for the exclusion criteria used for additional isolates, see the supplemental material.

Classification to the species level. Classification to the species level was performed by microbiological and sequenced-based methods (see the supplemental material for details).

Phylogenetic analysis and strain classification. There were 52 patients with multiple isolates of the same species. One of these (patient FK) carried two strains of K. pneumoniae that were highly divergent from each other (>20,000 chromosomal single-nucleotide variants [SNVs]), clearly representing a separate acquisition of blanKPC by each strain. Excluding this divergent strain pair, the remaining cases had differences ranging from 0 to 60 SNVs (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 because of repeated patient sampling. Phylogenetic analysis was then performed separately for each species using PheMyL (44) (see the supplemental material). Chromosomally distinct strains were defined by partitioning each phylogeny with a cutoff of ~500 SNVs (see the supplemental material). On the basis of the molecular clock of Enterobacteriaceae (1 to 20 SNVs/chromosome/year) (6, 37, 45), we can be relatively confident that isolates belonging to distinct strains will not have a shared ancestor within the time frame of blanKPC dispersal, and the number of distinct strains thus provides a conservative estimate of the number of distinct blanKPC 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 the supplemental material.

Since the isolates used for PacBio sequencing were randomly chosen from the set of all Illumina-sequenced isolates, some of them represented within-patient strain duplicates (see the previous section on phylogenetic analysis) and were therefore not included in the phylogenetic reconstruction. For display purposes (Fig. 1), the blanKPC structure(s) determined by long-read PacBio sequencing for each of these isolates is shown alongside...
FIG 1 Diversity of bacterial species, strains, plasmids, and Tn4401 variants. For each species, a phylogeny was generated from mapping to a species-specific chromosomal reference, after the deduplication of closely related isolates from the same patient (see Materials and Methods). Distinct strains are defined by a cutoff of \( \leq 500 \) SNVs (see Materials and Methods); strains found in more than one patient are shaded gray. Circles show plasmid “presence” as determined from Illumina data, with the fill color indicating uncertainty about whether the plasmid contains \( \text{bla}_{\text{KPC}} \). Boxes show plasmid structures determined from long-read PacBio sequencing of 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, the plasmid structure of the corresponding closely related isolate from the same patient is shown. Tn4401 and \( \text{bla}_{\text{KPC}} \) variants (Table 2) are indicated by large and small squares, respectively. The likely sources of \( \text{bla}_{\text{KPC}} \) acquisition, as determined from epidemiological data, are indicated by text color.
the representative isolate of the same strain from the same patient. In all of the 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 blaKPC plasmids pKPC_UVA01 and pKPC_UVA02, together with the additional nine distinct blaKPC plasmids identified by long-read PacBio sequencing, were used as references to determine plasmid presence profiles for each isolate on the basis of the Illumina data. Plasmid presence was defined as ≥99% sequence identity over ≥80% of the length of the reference sequence, as determined by BLASTn comparison of each isolate’s de novo assembly with 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 time frame of the outbreak), while the more permissive length cutoff allows for some rearrangement. It should be noted that the method does not take structural continuity into account.

**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 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 de novo assembly. The identification of novel Tn4401 insertion sites was also based on the de novo assemblies. These methods are described in detail in the supplemental material.

**Variation in Tn4401.** Tn4401 isoform classification was performed by comparing each isolate’s de novo assembly with the previously described isoform b reference sequence from EU176013.1 (29) using BLASTn to identify structural variation. SNVs were 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 nonreference 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 deidentification. Routine perirectal surveillance cultures for silent colonization began in April 2009 (38, 40). Cases were classified as “imported” if they did not have any prior admission to the University of Virginia Medical Center/Long-term Acute Care Hospital (UVaMC) and either had a blaKPC-positive Enterobacteriaceae isolated within 48 h 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. In the remaining cases, the source of blaKPC acquisition was classified as “local.” The 48-h cutoff is arbitrary and may result in some misclassification if patients either acquire blaKPC within the first 48 h of admission or if blaKPC carriage/infection remains undetected for >48 h; however, this is expected to be minimal (see the supplemental material). Charts and patient contacts were reviewed by using bed tracing data and the electronic medical record. This 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 blaKPC isolates. The analysis was performed separately for two different levels of genetic relatedness (strain or Tn4401 variant). This is described in detail in the supplemental material.

**Nucleotide sequence accession number.** Sequence data obtained in this study have been deposited at the National Center for Biotechnology Information under BioProject no. PRJNA246471.

**RESULTS**

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

**blaKPC is found in many different host strains, indicating frequent HGT.** There were 13 different species carrying blaKPC, (Fig. 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 of the species, there were a total of 62 distinct strains (>500 chromosomal SNVs; see Materials and Methods). Of these, 18 strains were identified in multiple patients and 44 were seen in only a single patient (Fig. 1), with new strains continuing to appear throughout the study period. The very recent emergence of blaKPC on an evolutionary time scale (15) implies that each strain likely acquired blaKPC independently, demonstrating frequent HGT between different strains and species.

**The blaKPC plasmids pKPC_UVA01 and pKPC_UVA02 are widely dispersed.** We hypothesized that the spread of blaKPC could be due to conjugative transfer of the index blaKPC plasmids, pKPC_UVA01 and pKPC_UVA02. With plasmid presence defined as ≥99% sequence identity over ≥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 (Fig. 1). Although the wide dispersal of these plasmids supports the plasmid-mediated outbreak hypothesis, short-read data can be limited in providing structural inferences when repetitive sequences are present, and for many isolates, it was not possible to confirm that blaKPC was actually colocalized within pKPC_UVA01 or pKPC_UVA02 (Fig. 1).

**blaKPC is found in many different plasmids, indicating frequent Tn4401 transposition.** To further investigate blaKPC plasmid structures, we performed long-read PacBio sequencing of 17 isolates that were chosen at random from the 281 available, yielding closed blaKPC structures in all of the cases. Fifteen isolates had a single blaKPC plasmid, and two isolates had two blaKPC plasmids, giving a total of 19 blaKPC 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 of these 17 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 blaKPC, (Fig. 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 blaKPC, as the “presence” of pKPC_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 pKPC_UVA01-like plasmids described


<table>
<thead>
<tr>
<th>Isolate</th>
<th>Species</th>
<th>Patient*</th>
<th>Date</th>
<th>Size (bp)</th>
<th>Group†</th>
<th>Within-group genetic change(s)‡</th>
<th>Tn4401 variant</th>
<th>Flanking sequences§</th>
<th>Tn2-like element¶</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>NA</td>
<td>Tn4401b-1</td>
<td>GTTCT...GTCT</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>43,631</td>
<td>pKPC_UVA01</td>
<td>1 SNV</td>
<td>Tn4401b-2</td>
<td>GTTCT...GTCT</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>NA</td>
<td>Tn4401b-2</td>
<td>AGATA...AGATA</td>
<td>No</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>pKPC_UVA02</td>
<td>0 SNVs</td>
<td>Tn4401b-1</td>
<td>ATGCA...GGCCA</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>pKPC_UVA02</td>
<td>0 SNVs</td>
<td>Tn4401b-1</td>
<td>ATGCA...GGCCA</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>Tn4401b-1</td>
<td>GTTCT...GTCT</td>
<td>Yes</td>
</tr>
<tr>
<td>CAV1043</td>
<td>E. aerius</td>
<td>L</td>
<td>Mar 2008</td>
<td>pKPC_CAV1043</td>
<td>59,138</td>
<td>NA</td>
<td>Tn4401b-2</td>
<td>GTTCT...GTCT</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>pKPC_CAV1176</td>
<td>0 SNVs</td>
<td>Tn4401novel-1</td>
<td>GTTCT...GTCT</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>pKPC_CAV1176</td>
<td>0 SNVs</td>
<td>Tn4401novel-1</td>
<td>GTTCT...GTCT</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>pKPC_CAV1176</td>
<td>1 SNV, 40-kb inversion</td>
<td>Tn4401novel-1</td>
<td>GTTCT...GTCT</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>pKPC_CAV1176</td>
<td>40-kb inversion</td>
<td>Tn4401novel-1</td>
<td>GTTCT...GTCT</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>pKPC_CAV1176</td>
<td>1 SNV, 188-bp deletion</td>
<td>Tn4401novel-1</td>
<td>GTTCT...GTCT</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>pKPC_UVA01</td>
<td>1,225-bp insertion</td>
<td>Tn4401b-1</td>
<td>GTTCT...GTCT</td>
</tr>
<tr>
<td>CAV1741</td>
<td>C. freundii</td>
<td>ER</td>
<td>Oct 2012</td>
<td>pKPC_CAV1741</td>
<td>129,196</td>
<td>pKPC_UVA01</td>
<td>14,960-bp duplication, 70,615-bp insertion</td>
<td>Tn4401b-1</td>
<td>GTTCT...GTCT</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>pKPC_UVA01</td>
<td>0 SNVsf</td>
<td>Tn4401b-1</td>
<td>GTTCT...GTCT</td>
</tr>
<tr>
<td>CAV1320</td>
<td>E. aerogenes</td>
<td>EL</td>
<td>Nov 2010</td>
<td>pKPC_CAV1320</td>
<td>13,981</td>
<td>NA</td>
<td>Tn4401b-1</td>
<td>TTGTT...TTGTT</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>Tn4401b-8</td>
<td>TTTTTTTTTTT</td>
<td>No</td>
</tr>
</tbody>
</table>

* 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 time frame. 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 of the others shown.

† Difference(s) from the reference sequence of that plasmid group, as specified in the previous column.

‡ Sequences immediately flanking Tn4401; generally expected to be identical because of 5-bp target site duplication during transposition (28).

§ Tn4401 integrated into the tnpA gene of a Tn2-like element.

¶ No evidence of target site duplication.

** Two copies.

† It is noteworthy that this plasmid from K. intermedia CAV1151 is exactly identical to pKPC_UVA01 from K. pneumoniae CAV1016, with isolation dates 2 years apart.

# Anonymized patient identifiers are used; they do not represent initials or any other personal information.

0 NA, not applicable.

Tn4401 is present within a Tn2-like element in many different plasmids. In 7 of the 11 distinct, fully characterized blakPC plasmids, Tn4401 was surrounded by a sequence element related to the blatem1-containing transposon Tn2 (Fig. 3). In all of the cases, the Tn4401 insertion site within the tnpA gene of Tn2 was the same, 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, that contains the additional resistance genes blakOXA9, aadA1, and aac(6')-IIb and has been seen as a prior Tn4401 insertion site (46).

Tn4401 variation. There were five different structural variants of Tn4401 (Table 2). The majority of the 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 blakPC (n = 39) or one of two truncated Tn4401 structures (n = 4). At the nucleotide level, there were seven sites that were variable within Tn4401b. Of these, there were three sites located within blakPC, giving rise to three different blakPC alleles, blakPC-2 (n = 179), blakPC-3 (n = 44), and blakPC-4 (n = 5). All non-Tn4401b isolates contained blakPC-2. Taking all structural and nucleotide variations into account, there were a total of 12 different
Tn\textsubscript{4401} variants. However, most of these were very rare, with seven found only in single patients.

\textit{bla}\textsubscript{KPC} mobility has occurred within the hospital. On the basis of prior health care exposure, the \textit{bla}\textsubscript{KPC} acquisition source was classified as “imported” (likely acquisition prior to admission to our institution) for 15/182 (8%) patients and “local” (likely acquisition within our institution) for 167/182 (92%) patients (Fig. 1; see Materials and Methods). Imports were more likely to be infected/colonized with \textit{K. pneumoniae}, particularly ST258 (see Table S3 in the supplemental material), consistent with previous reports of this strain being the dominant \textit{bla}\textsubscript{KPC} carrier in the United States (9, 47). Thus, most host strain variation likely originated within the hospital via \textit{bla}\textsubscript{KPC} HGT. In support of this, 15/16 (94%) patients infected/colonized with multiple strains/species shared Tn\textsubscript{4401} variants within the patient (see Table S4 in the supplemental material), suggesting recent \textit{bla}\textsubscript{KPC} HGT. Notably, this included one patient with two different species carrying Tn\textsubscript{4401}b-3, which is not found in any other patient.

There was also some evidence of recent within-strain Tn\textsubscript{4401} transposition. From the isolates that were randomly chosen for long-read sequencing, 4/17 (24%) had multiple Tn\textsubscript{4401} copies (Table 1). If we assume that this randomly chosen subset is representative, this extrapolates to approximately 66/281 isolates across the whole data set. However, only 2/281 isolates had multiple Tn\textsubscript{4401} variants (Tn\textsubscript{4401}b-5; Table 2), indicating that many isolates likely had multiple copies of the same Tn\textsubscript{4401} variant, consistent with recent Tn\textsubscript{4401} transposition.

Taken together, these results indicate that much of the genetic diversity observed is due to recent \textit{bla}\textsubscript{KPC} mobility, likely within the hospital ecosystem over the described 5-year outbreak.

Direct patient-to-patient transmission does not explain \textit{bla}\textsubscript{KPC} acquisition. To further investigate the \textit{bla}\textsubscript{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 Tn\textsubscript{4401} variant. Considering only “local” acquisitions (see above), 48/167 (29%) patients had ward contact with another patient carrying the same \textit{bla}\textsubscript{KPC}-positive strain (Fig. 4, top). A greater proportion of the patients, 106/167 (63%), had ward contact with another patient carrying the same Tn\textsubscript{4401} variant. However, as Tn\textsubscript{4401}b-1 is very common (66% of the 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 Tn\textsubscript{4401} variant (Fig. 4, bottom). Therefore, both genetic levels (strain or Tn\textsubscript{4401} variant) demonstrated plausible transmissions for only a minority of the patients, indicating that direct patient-to-patient transmission is not the dominant mode of \textit{bla}\textsubscript{KPC} acquisition or that there are many silently colonized patients below the limit of detection by our surveillance methods (40, 48).

**DISCUSSION**

Here we have demonstrated high levels of genetic diversity in KPC-producing \textit{Enterobacteriaceae} within a single institution over 5 years. This diversity occurs at multiple genetic levels, revealing a complex evolutionary history of the \textit{bla}\textsubscript{KPC} gene involving many different host strains and plasmids.

In 7/11 distinct \textit{bla}\textsubscript{KPC} plasmids identified through long-read sequencing, Tn\textsubscript{4401} was located within a Tn2-like element. As these Tn2-like elements differed substantially from each other...
(Fig. 3), it is unlikely that this arose via the transposition of a composite \(\text{Tn}4401\)-\(\text{Tn}2\)-like structure. Instead, it suggests that \(\text{Tn}4401\) has been repeatedly incorporated into pre-existing \(\text{Tn}2\)-like elements, which are known to be widespread, and genetically divergent, in \(\text{Enterobacteriaceae}\) (49, 50). However, the insertion site was identical in all of the cases, yet \(\text{Tn}4401\) has been reported to have no insertion site specificity (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 \(\text{Tn}2\)-like elements following an initial integration event, as recently suggested for another multidrug resistance gene, \(\text{bla}\text{CTX-M-15}\) (51). This implies that \(\text{Tn}4401\) mobility may have been enhanced via integration into a second, already widely dispersed, transposon. As the \(\text{Tn}4401\)-\(\text{Tn}2\)-like structure was present in the index case isolate (CAV1016, August 2007), we presume that the initial transposition of \(\text{Tn}4401\) into a \(\text{Tn}2\)-like element occurred prior to entry into our hospital system. In support of this, one particular \(\text{Tn}2\)-like element, \(\text{Tn}1331\), has been previously reported to contain \(\text{Tn}4401\) (in exactly the same position within the \(\text{tnp}A\) gene as that described here) (21, 46, 52, 53), including one report describing a \(\text{K. pneumoniae}\) isolated in 2005, which predates \(\text{bla}_{\text{KPC}}\) in our institution (46). We are not aware of any previous reports describing \(\text{Tn}4401\) within a non-\(\text{Tn}1331\) \(\text{Tn}2\)-like element.

The prevalence of \(\text{Tn}4401\) insertions within \(\text{Tn}2\)-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 \(\text{Tn}4401\) element, as well as a PCR method to assay the presence of what we had wrongly assumed was a single plasmid, \(\text{pKPC}_{\text{UVA01}}\). This PCR assay targeted the immediate \(\text{Tn}4401\) insertion site within a \(\text{Tn}2\)-like element (54), which we have demonstrated here is present in many different plasmids, highlighting that PCR assay results, and indeed, those of any partial typing methods, need to be interpreted with a great deal of caution. We were further misled by the analysis of short-read whole-genome sequencing data that indicated the presence of \(\text{pKPC}_{\text{UVA01}}\) in the majority of our isolates. Taking these findings together, it was tempting to conclude that horizontal transfer of \(\text{pKPC}_{\text{UVA01}}\) was responsible for the great majority of the \(\text{bla}_{\text{KPC}}\) carriage at our institution. However, long-read sequencing refuted this, revealing a far more complex picture.

More generally, this highlights certain limitations of plasmid reconstruction from short-read data. To illustrate by way of example, there were five isolates where long-read sequencing revealed \(\text{pKPC}_{\text{UVA01}}\)-like plasmids that were identical to the reference \(\text{pKPC}_{\text{UVA01}}\) sequence apart from the absence of \(\text{Tn}4401\) and the associated 5-bp target site duplication (Fig. 2). We presume that in these lineages, \(\text{bla}_{\text{KPC}}\) may have been initially ac-
TABLE 2 Tn\(^{4401}\) variations

<table>
<thead>
<tr>
<th>Tn(^{4401}) variant</th>
<th>Structural isoform (29)</th>
<th>SNV(s) (d)</th>
<th>bla(^\text{KPC}) variant</th>
<th>No. of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tn(^{4401})-1'</td>
<td>b</td>
<td>8015C(\rightarrow)T(d)</td>
<td>bla(^{\text{KPC}})-2</td>
<td>121</td>
</tr>
<tr>
<td>Tn(^{4401})-2</td>
<td>b</td>
<td>8015C(\rightarrow)T, 9621T(\rightarrow)C</td>
<td>bla(^{\text{KPC}})-3</td>
<td>40</td>
</tr>
<tr>
<td>Tn(^{4401})-3</td>
<td>b</td>
<td>7197T(\rightarrow)A, 8015C(\rightarrow)T, 9621T(\rightarrow)C</td>
<td>bla(^{\text{KPC}})-3</td>
<td>1</td>
</tr>
<tr>
<td>Tn(^{4401})-4</td>
<td>b</td>
<td>8015N()</td>
<td>bla(^{\text{KPC}})-3</td>
<td>2</td>
</tr>
<tr>
<td>Tn(^{4401})-6</td>
<td>b</td>
<td>7509C(\rightarrow)G, 7917T(\rightarrow)G</td>
<td>bla(^{\text{KPC}})-4</td>
<td>1</td>
</tr>
<tr>
<td>Tn(^{4401})-7</td>
<td>b</td>
<td>6800T(\rightarrow)C, 7509C(\rightarrow)G, 7917T(\rightarrow)G</td>
<td>bla(^{\text{KPC}})-4</td>
<td>1</td>
</tr>
<tr>
<td>Tn(^{4401})-8</td>
<td>b</td>
<td>9663T(\rightarrow)C</td>
<td>bla(^{\text{KPC}})-2</td>
<td>1</td>
</tr>
<tr>
<td>Tn(^{4401})-a</td>
<td>a (del 7020-7118)</td>
<td>800N()</td>
<td>bla(^{\text{KPC}})-2</td>
<td>8</td>
</tr>
<tr>
<td>Tn(^{4401})-novel-1</td>
<td>Novel (del 6919-7106)</td>
<td>800N()</td>
<td>bla(^{\text{KPC}})-2</td>
<td>28</td>
</tr>
<tr>
<td>Tn(^{4401})-trunc-1</td>
<td>Truncated (del 1-6564)</td>
<td>800N()</td>
<td>bla(^{\text{KPC}})-2</td>
<td>2</td>
</tr>
<tr>
<td>Tn(^{4401})-trunc-2</td>
<td>Truncated (del 1-6727)</td>
<td>6800N()</td>
<td>bla(^{\text{KPC}})-2</td>
<td>1</td>
</tr>
</tbody>
</table>

\(a\) Variants are named such that letters indicate previously described structural isofoms and numbers indicate nucleotide level variations (SNVs) within an isoform (apart from the truncated Tn\(^{4401}\) structures, where numbers are used to indicate different truncation locations).

\(b\) With respect to Tn\(^{4401}\)-b, which was considered the reference Tn\(^{4401}\) sequence in this study.

\(c\) These two substitutions convert C to T (i.e., Tn\(^{4401}\)-b 5 differs from the reference Tn\(^{4401}\)-b isoform by the following SNVs: 4939C\(\rightarrow\)T, 7917T\(\rightarrow\)C).

\(d\) This substitution converts bla\(^{\text{KPC}}\)-2 to bla\(^{\text{KPC}}\)-4.

\(e\) Quality filters failed at this position because of a lack of reads supporting C and T (i.e., Tn\(^{4401}\)-b 5 actually represents a mixture of Tn\(^{4401}\)-b 1 and Tn\(^{4401}\)-b 2).

\(f\) These two substitutions convert bla\(^{\text{KPC}}\)-2 to bla\(^{\text{KPC}}\)-4.

\(g\) Quality filters failed at this position because of a lack of reads mapped in the reverse direction. All of the reads mapped in the forward direction supported a reference (T) call.

Required via HGT of pKPC_UVA01, with subsequent homologous recombination transferring Tn\(^{4401}\) from pKPC_UVA01 to a different plasmid containing a Tn2-like element. In each of these five isolates, there are multiple Tn2-like elements that have 100% sequence identity over approximately 1 kb on either side of the Tn\(^{4401}\) 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\(^{\text{KPC}}\) by 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 the 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 by 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 are 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 are also consistent with an alternative structure. As several recent studies have utilized reference-based approaches for plasmid assembly/inference (33, 34), our results indicate that results obtained by any such methods should be interpreted with extreme caution.

Across the bla\(^{\text{KPC}}\)-positive patients, there was large variation in both host strains and bla\(^{\text{KPC}}\) plasmids, with Tn\(^{4401}\) 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 Tn\(^{4401}\). However, we found limited variation within the transposon, as Tn\(^{4401}\) 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 bla\(^{\text{KPC}}\) 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 are several limitations to this study. Because of the cost and effort involved in long-read sequencing, we were able to resolve only a minority of bla\(^{\text{KPC}}\) plasmids. This means that although we have a compelling indicator of the diversity created by KPC plasmids, with Tn\(^{4401}\)-like elements, but future investigations could be used to illuminate the effect of this composite structure on Tn\(^{4401}\) mobility. Another issue is the limit of detection of the culture-based screening methods and phenotypic tests used to identify bla\(^{\text{KPC}}\)-positive clinical isolates. No single perirectal screening method to capture asymptomatically colonized patients is perfect (40, 48, 55), including the method used here, which has a sensitivity of \(~86%\) (40). In recognition of the fact that bla\(^{\text{KPC}}\) expression and carbapenem susceptibility may be variable in different host species, and in the context of additional resistance mechanisms such as porin alterations, we lowered our surveillance thresholds to include all of the possible ESBL-producing organisms that subsequently tested positive in phenotypic carbapenemase tests. Even with this mitigation strategy, we anticipate that we have missed a proportion of the KPC acquisition events occurring in the early stages of antimicrobial resistance.
gene emergence in a single institution identifies several distinct processes occurring at high frequency (Fig. 5). First, the presence of shared bla\textsubscript{KPC}-containing strains in different patients reflects traditional (clonal) outbreak models. Second, bla\textsubscript{KPC} mobility between strains/species is facilitated by promiscuous bla\textsubscript{KPC} plasmids such as pKPC\textsubscript{UVA01}. Third, bla\textsubscript{KPC} transfer between plasmids is likely enhanced by homologous recombination between Tn\textsubscript{2}-like elements, facilitating the movement of Tn\textsubscript{4401} from one plasmid to another. Finally, bla\textsubscript{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, with mobility occurring at multiple nested genetic
levels, analogous to a Russian doll set. This has resulted in a high level of diversity in KPC-producing Enterobacteriaceae, at multiple genetic levels. As blaKPC prevalence continues to increase, so will this genetic diversity, inevitably resulting in a wider variety of more pathogenic strains carrying blaKPC.

Our results indicate that the current standard practice of screening only specific species for blaKPC carriage is likely to hamper surveillance efforts by grossly underestimating its 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.

ACKNOWLEDGMENTS

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.

The Modernising Medical Microbiology (MMM) Informatics Group includes Jim Davies, Charles Crichton, Milind Acharya, and Carlos del Ojo Elias.

FUNDING INFORMATION

This publication presents independent research commissioned by the Health Innovation Challenge Fund (grants HICF-T5-358 and WT098615/Z/12/Z), a parallel funding partnership of 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 the University of Oxford, and the 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 funders and not necessarily those of the funders. D.W.C. and T.E.P. are NIHR senior investigators. N.S. was supported by a Wellcome Trust University of Oxford research fellowship during this work. D.J.W. is a Sir Henry Dale Fellow jointly funded by the Wellcome Trust and the Royal Society (grant 101237/Z/13/Z).

REFERENCES


FIG 5 blaKPC spreads at multiple genetic levels, resulting in a high level of diversity in blaKPC-positive Enterobacteriaceae isolates.


