Novel \textit{bla}_{KPC}-carrying species identified in the hospital environment

Hang T.T. Phan*,1,2, Zoie Aiken3†, Oluwafemi Akinremi2,4, Julie Cawthorne3, Andrew Dodgson3,4, Angela Downes3, Matthew Ellington2,5, Ryan George3, Amy Mathers6, Nicole Stoesser1, Stephanie Thomas7, on behalf of the TRACE Investigators’ Group

Authors’ affiliations:

1Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom
2NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University
3Central Manchester University Hospitals NHS Foundation Trust, Oxford Road, Manchester, M13 9WL, UK
4Public Health England, London, UK
6Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA
7Department of Microbiology, University Hospital South Manchester

Running title: Novel \textit{bla}_{KPC}-carrying species

*corresponding author

e-mail: hang.phan@ndm.ox.ac.uk

† All authors except the first author are listed in alphabetical order
Abstract (75/75 words)

bla<sub>KPC</sub>, encoding one of five dominant global carbapenemase families, is increasingly identified in environmental species difficult to characterize using routine diagnostic methods, with epidemiological and clinical implications. During environmental hospital infection prevention and control investigations (Manchester, UK) we used whole genome sequencing to confirm species identification for isolates infrequently associated with bla<sub>KPC</sub> and/or difficult to classify by MALDI-ToF. Four previously undescribed bla<sub>KPC</sub>-carrying species were identified from the hospital environment, including a putative, novel Enterobacter species.
Carbapenems are used to treat serious infections resistant to other antimicrobial agents. *Klebsiella pneumoniae* carbapenemases (KPC), encoded by *bla*KPC alleles, are a major carbapenemase family, conferring resistance to most beta-lactams. *bla*KPC, first identified retrospectively in *K. pneumoniae* dating from 1996, has since been observed in other Enterobacteriaceae, Pseudomonadaceae and *Acinetobacter* spp.[1].

The host species range of *bla*KPC has not been defined, and may be more diverse than originally thought, given its presence in a number of environmental reservoirs[2-4].

Accurate species identification is important for epidemiological and clinical reasons. Species identification methods (e.g. biochemical profiling, Matrix Assisted Laser Desorption Ionization Time-of-Flight [MALDI-ToF]), or partial sequence-based typing (e.g. 16S rRNA gene sequencing) can be inconsistent, particularly for Enterobacteriaceae[5, 6]. Whole genome sequencing (WGS) has an advantage over these methods in using the complete genetic content of an organism for identification[7].

Here, WGS was used to confirm species identification of KPC-producers cultured during environmental infection prevention and control investigations undertaken on wards with KPC-positive patients in two hospitals in Manchester, UK (May 2012-March 2013, April 2015). High-touch surfaces, toilet edges and sinks were sampled with EZ-Reach sponges squeezed out into neutralizing buffer; water samples were collected from sink drain P-traps with plastic tubing and a syringe. Samples were centrifuged, and pellets sub-cultured overnight in tryptic soy broth (5mls) with ertapenem 10μg discs. Broths were sub-cultured on chromID CARBA plates.
different Enterobacteriaceae colony morphotypes were confirmed as \textit{bla}\textsubscript{KPC} positive by PCR and species determined by MALDI-ToF (Bruker, Billerica, Massachusetts, United States).

WGS was performed on 142 isolates from 55 samples using the Illumina HiSeq platform (150bp, paired-end reads). \textit{De novo} assemblies were created using SPAdes 3.6 \cite{8}, and assessed for integrity using REAPR \cite{9}. BLASTn-based analyses against assemblies confirmed the presence of \textit{bla}\textsubscript{KPC} in all 142 isolates.

Species identification was consistent by both MALDI-ToF and WGS for 78 \textit{Klebsiella pneumoniae} isolates, 11 \textit{Enterobacter cloacae}, five \textit{Escherichia coli}, five \textit{Citrobacter freundii} and four \textit{Klebsiella oxytoca}. However, limited sequence matches with the NCBI \textit{nr} reference database were observed for two isolates characterized as \textit{Escherichia vulneris} and/or \textit{Enterobacter cancerogenus} by MALDI-ToF (9885\_1 [81\% sequence similarity/58\% coverage with the top match]; and 9885\_2 [84\%/74\%]), and 36 isolates characterized as \textit{Pantoea agglomerans} by MALDI-ToF (all closely genetically related; only 88\%/85\% coverage with the top match). In addition, one isolate (1613625) was characterized as \textit{Pluralibacter gergoviae} by MALDI-ToF and had a good sequence match to the NCBI \textit{nr} reference database; however, this species had only been associated with \textit{bla}\textsubscript{KPC} in two previous disease-causing (i.e. not environmental) isolates (in Brazil and the USA) \cite{10, 11}. We therefore further investigated the species identification of isolates 9885\_1, 9885\_2, 9885\_4 (representative of the putative \textit{P. agglomerans} group) and 1613625 using the genomic data.
Querying the assemblies against the Silva 16S database\[12\] (Table 1) demonstrated that each isolate had several hits above the suggested threshold for 16S-based species identification (>99.5% similarity). However, in all cases the first and second hits were divergent. Species identification was therefore undertaken using genome-wide average nucleotide identity (ANI) and alignment fraction (AF) scores, where ANI and AF scores 96-96.5% and 0.6 respectively are considered species-specific thresholds\[7, 13\]. Sequence assemblies were first annotated with PROKKA\[14\] and coding sequences were fed into the ANICalculator\[13\]. ANI scores for isolate assemblies and top hit species found by 16S analysis (where available as reference genomes or assemblies in GenBank; n=34) were calculated. These identified isolate 9885_1 as *E. vulneris*, 9885_4 as *Pantoea anthophila*, 1613625 as *P. gergoviae*, and 9885_2 as a presumptive new *Enterobacter* species (Table 1).

We compared the genomes of these four isolates with those of closely related species (n=34) using a core/pan-genome approach (ROARY)\[15\], and including only those sub-groups with a core genome >200 core genes. For this analysis, contigs of plasmid origin (high sequence similarity with plasmids in the NCBI nr database) were excluded. A maximum likelihood phylogeny from core genome alignments was reconstructed; the core genome phylogenies (Figure 1) concurred with the ANI results and highlighted inconsistencies in traditional taxonomic assignations that are likely to be revised as more bacterial genome sequences become available\[16\].

*E. vulneris* has been associated with both *Escherichia* and *Enterobacter* genera by DNA hybridization\[17\], perhaps explaining the initial identification of 9885_2 as either *E. vulneris/E. cancerogenus* by MALDI-ToF. *E. vulneris* has been described in...
wound infections, bacteraemia, peritonitis, urosepsis and meningitis[17-20], and its novel association with \textit{bla}_{\textit{KPC}} in this study is therefore potentially of clinical significance. \textit{Pantoea} spp. are recognized as opportunistic pathogens and whilst \textit{P. agglomerans} has been associated with \textit{bla}_{\textit{KPC}} previously[21], \textit{P. anthophila}, found in lakes and diseased plants[22, 23], has not. Worryingly, in this survey, it was one of the most widespread \textit{bla}_{\textit{KPC}}-positive environmental species (36 isolates from 8 sites from one ward, including clinical areas, staff areas and ward kitchens). \textit{P. anthophila} may represent a successful intermediate species for the wider transmission of \textit{bla}_{\textit{KPC}} amongst more pathogenic strains within/between non-clinical/clinical sites. \textit{P. gergoviae} has been isolated from a wide range of sources, including foodstuffs and insects[24, 25], but has not previously been identified with \textit{bla}_{\textit{KPC}} in environmental isolates. Of particular concern is the organism’s innate resistance to parabens, enabling it to exist as a contaminant in a variety of personal care products, including soaps and shampoos[25].

We have identified four species not previously associated with \textit{bla}_{\textit{KPC}} in the hospital environment, one a putative, novel \textit{Enterobacter} sp. Accurate species identification of \textit{bla}_{\textit{KPC}}-carrying environmental isolates has implications for understanding the wider reservoirs and transmission dynamics of this resistance gene family, assessing the pathogenic potential of these organisms, and for accurate susceptibility testing in isolates that have the potential to become clinically relevant. Although WGS is not currently widely used in routine diagnostics, the development of bench-top sequencers and falling costs may make this more feasible[26]. Species identification can also be hampered by incomplete reference genome databases, which is being addressed by the large-scale sequencing of type strain collections[27]. Our
identification of bla\textsubscript{KPC} in diverse species in the hospital environment supports previous data suggesting that this reservoir plays a role in the dissemination of bla\textsubscript{KPC} across species and genera[2].

**Nucleotide sequence data**

These have been deposited in GenBank under project number PRJNA324191, study number SRP076320 and accession numbers SRR3654271, SRR3654272, SRR3654273, SRR3654274.

**ACKNOWLEDGEMENTS**


We are grateful to the Infection Control Teams and Microbiology Laboratory staff at the University Hospital of South Manchester NHS Foundation Trust and the Central Manchester University Hospitals NHS Foundation Trust; the Sequencing hub at the Wellcome Trust Center for Human Genetics, Oxford; and the Modernizing Medical Microbiology Informatics Group (MMMIG), Oxford.

**FUNDING**
The research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University in partnership with Public Health England (PHE). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

REFERENCES


5. Brady C, Cleenwerck I, Venter S, Coutinho T, De Vos P: Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E.
amnigenus into Lelliottia gen. nov. as Lelliottia nimipressuralis comb. nov. and Lelliottia amnigena comb. nov., respectively, E. gergoviae and E. pyrinus into Pluralibacter gen. nov. as Pluralibacter gergoviae comb. nov. and Pluralibacter pyrinus comb. nov., respectively, E. cowanii, E. radicincitans, E. oryzae and E. arachidis into Kosakonia gen. nov. as Kosakonia cowanii comb. nov., Kosakonia radicincitans comb. nov., Kosakonia oryzae comb. nov. and Kosakonia arachidis comb. nov., respectively, and E. turicensis, E. helveticus and E. pulveris into Cronobacter as Cronobacter zurichensis nom. nov., Cronobacter helveticus comb. nov. and Cronobacter pulveris comb. nov., respectively, and emended description of the genera Enterobacter and Cronobacter.


Table 1. Top 16S and genome wide species identification results of study $bla_{KPC}$-positive isolates

<table>
<thead>
<tr>
<th>Isolate id</th>
<th>First hit species (% identity)</th>
<th>Second hit species (% identity)</th>
<th>Average nucleotide identity (ANI) species classification</th>
<th>Final species classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Best species match</td>
<td>ANI (bidirectional)</td>
</tr>
<tr>
<td>9885_1</td>
<td><em>Escherichia vulneris</em> (99.74)</td>
<td><em>Enterobacter hormaechei</em> (99.73)</td>
<td><em>Escherichia vulneris</em></td>
<td>96.38</td>
</tr>
<tr>
<td>9885_2</td>
<td><em>Citrobacter murliniae</em> (99.80)</td>
<td><em>Enterobacter asburiae</em> L1 (99.42)</td>
<td><em>Enterobacter sp MGH24</em></td>
<td>86.69</td>
</tr>
<tr>
<td>9885_4</td>
<td><em>Pantoea anthophila</em> (99.87)</td>
<td><em>Pantoea agglomerans</em> (99.87)</td>
<td><em>Pantoea anthophila</em></td>
<td>98.75</td>
</tr>
<tr>
<td>1613625</td>
<td><em>Enterobacter sp. ETH-2</em> (100.00)</td>
<td><em>Pluralibacter gergoviae</em> (99.81)</td>
<td><em>Pluralibacter gergoviae</em></td>
<td>98.91</td>
</tr>
</tbody>
</table>
Figure 1. Core genome phylogenies of \( \text{bla}_{\text{KPC}} \)-positive isolates investigated in detail in this study and of closely related reference genomes from NCBI/GenBank. (a) phylogeny of 9885_1, 9885_2 and 1613625 and reference \textit{Enterobacter} spp. genomes; (b) phylogeny of 9885_4 and reference \textit{Pantoea} spp. genomes
Core genes (99 ~100%): 203
Soft core genes (95~ 99%): 56
Shell genes (15~95%): 3,066
Cloud genes (0~15%): 45,574
Total genes: 48,899

Core genes (99 ~100%): 199
Soft core genes (95~ 99%): 0
Shell genes (15~95%): 10,545
Cloud genes (0~15%): 10,453
Total genes: 21,197