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

Hang T.T. Phan^{*1,2}, Zoie Aiken^{3†}, Oluwafemi Akinremi^{2,4}, Julie Cawthorne³,

Andrew Dodgson^{3,4}, Angela Downes³, Matthew Ellington^{2,5}, 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

⁵Antimicrobial Resistance and Healthcare Associated Infections Reference Unit (AMRHAI), Public 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 *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*_{KPC}, 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*_{KPC} and/or difficult to classify by MALDI-ToF. Four previously undescribed *bla*_{KPC}-carrying species were identified from the hospital environment, including a putative, novel *Enterobacter* species.

1 Main text (1000/1000 words)

2 Carbapenems are used to treat serious infections resistant to other antimicrobial
3 agents. *Klebsiella pneumoniae* carbapenemases (KPC), encoded by *bla*_{KPC} alleles , are
4 a major carbapenemase family, conferring resistance to most beta-lactams. *bla*_{KPC},
5 first identified retrospectively in *K. pneumoniae* dating from 1996, has since been
6 observed in other Enterobacteriaceae, Pseudomonadaceae and *Acinetobacter* spp.[1].
7 The host species range of *bla*_{KPC} has not been defined, and may be more diverse than
8 originally thought, given its presence in a number of environmental reservoirs[2-4].

9
10 Accurate species identification is important for epidemiological and clinical reasons.
11 Species identification methods (e.g. biochemical profiling, Matrix Assisted Laser
12 Desorption Ionization Time-of-Flight [MALDI-ToF]), or partial sequence-based
13 typing (e.g. 16S rRNA gene sequencing) can be inconsistent, particularly for
14 Enterobacteriaceae[5, 6]. Whole genome sequencing (WGS) has an advantage over
15 these methods in using the complete genetic content of an organism for
16 identification[7].

17
18 Here, WGS was used to confirm species identification of KPC-producers cultured
19 during environmental infection prevention and control investigations undertaken on
20 wards with KPC-positive patients in two hospitals in Manchester, UK (May 2012-
21 March 2013, April 2015). High-touch surfaces, toilet edges and sinks were sampled
22 with EZ-Reach sponges squeezed out into neutralizing buffer; water samples were
23 collected from sink drain P-traps with plastic tubing and a syringe. Samples were
24 centrifuged, and pellets sub-cultured overnight in tryptic soy broth (5mls) with
25 ertapenem 10µg discs. Broths were sub-cultured on chromID CARBA plates

(bioMérieux, Marcy-l'Étoile, France); different Enterobacteriaceae colony morphotypes were confirmed as *bla*_{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). *De novo* assemblies were created using SPAdes 3.6[8], and assessed for integrity using REAPR[9]. BLASTn-based analyses against assemblies confirmed the presence of *bla*_{KPC} in all 142 isolates.

Species identification was consistent by both MALDI-ToF and WGS for 78 *Klebsiella pneumoniae* isolates, 11 *Enterobacter cloacae*, five *Escherichia coli*, five *Citrobacter freundii* and four *Klebsiella oxytoca*. However, limited sequence matches with the NCBI nr reference database were observed for two isolates characterized as *Escherichia vulneris* and/or *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 *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 *Pluralibacter gergoviae* by MALDI-ToF and had a good sequence match to the NCBI nr reference database; however, this species had only been associated with *bla*_{KPC} in two previous disease-causing (i.e. not environmental) isolates (in Brazil and the USA)[10, 11]. We therefore further investigated the species identification of isolates 9885_1, 9885_2, 9885_4 (representative of the putative *P. agglomerans* group) and 1613625 using the genomic data.

51 Querying the assemblies against the Silva 16S database[12](Table 1) demonstrated
 52 that each isolate had several hits above the suggested threshold for 16S-based species
 53 identification (>99.5% similarity). However, in all cases the first and second hits were
 54 divergent. Species identification was therefore undertaken using genome-wide
 55 average nucleotide identity (ANI) and alignment fraction (AF) scores, where ANI and
 56 AF scores 96-96.5% and 0.6 respectively are considered species-specific thresholds[7,
 57 13]. Sequence assemblies were first annotated with PROKKA[14] and coding
 58 sequences were fed into the ANICalculator[13]. ANI scores for isolate assemblies and
 59 top hit species found by 16S analysis (where available as reference genomes or
 60 assemblies in GenBank; n=34) were calculated. These identified isolate 9885_1 as
 61 *vulneris*, 9885_4 as *Pantoea anthophila*, 1613625 as *P. gergoviae*, and 9885_2 as a
 62 presumptive new *Enterobacter* species (Table 1).

63

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

72

73 *E. vulneris* has been associated with both *Escherichia* and *Enterobacter* genera by
 74 DNA hybridization[17], perhaps explaining the initial identification of 9885_2 as
 75 either *E. vulneris*/*E. cancerogenus* by MALDI-ToF. *E. vulneris* has been described in

76 wound infections, bacteraemia, peritonitis, urosepsis and meningitis[17-20], and its
 77 novel association with *bla*_{KPC} in this study is therefore potentially of clinical
 78 significance. *Pantoea* spp. are recognized as opportunistic pathogens and whilst *P.*
 79 *agglomerans* has been associated with *bla*_{KPC} previously[21], *P.anthophila*, found in
 80 lakes and diseased plants[22, 23], has not. Worryingly, in this survey, it was one of
 81 the most widespread *bla*_{KPC}-positive environmental species (36 isolates from 8 sites
 82 from one ward, including clinical areas, staff areas and ward kitchens). *P. anthophila*
 83 may represent a successful intermediate species for the wider transmission of *bla*_{KPC}
 84 amongst more pathogenic strains within/between non-clinical/clinical sites. *P.*
 85 *gergoviae* has been isolated from a wide range of sources, including foodstuffs and
 86 insects[24, 25], but has not previously been identified with *bla*_{KPC} in environmental
 87 isolates. Of particular concern is the organism's innate resistance to parabens,
 88 enabling it to exist as a contaminant in a variety of personal care products, including
 89 soaps and shampoos[25].

90

91 We have identified four species not previously associated with *bla*_{KPC} in the hospital
 92 environment, one a putative, novel *Enterobacter* sp. Accurate species identification of
 93 *bla*_{KPC}-carrying environmental isolates has implications for understanding the wider
 94 reservoirs and transmission dynamics of this resistance gene family, assessing the
 95 pathogenic potential of these organisms, and for accurate susceptibility testing in
 96 isolates that have the potential to become clinically relevant. Although WGS is not
 97 currently widely used in routine diagnostics, the development of bench-top
 98 sequencers and falling costs may make this more feasible[26]. Species identification
 99 can also be hampered by incomplete reference genome databases, which is being
 100 addressed by the large-scale sequencing of type strain collections[27]. Our

101 identification of *bla*_{KPC} in diverse species in the hospital environment supports
 102 previous data suggesting that this reservoir plays a role in the dissemination of *bla*_{KPC}
 103 across species and genera[2].
 104

105 **Nucleotide sequence data**

106 These have been deposited in GenBank under project number PRJNA324191 , study
 107 number SRP076320 and accession numbers SRR3654271, SRR3654272,
 108 SRR3654273, SRR3654274.
 109

110 **ACKNOWLEDGEMENTS**

111 The Transmission of Carbapenemase-producing Enterobacteriaceae (TRACE) study
 112 investigators are (alphabetical): Zoie Aiken, Oluwafemi Akinremi, Julie Cawthorne,
 113 Paul Cleary, Derrick Crook, Valerie Decraene, Andrew Dodgson, Matthew Ellington,
 114 Ryan George, Katie Hopkins, Rachel Jones, Cheryl Lenney, Amy Mathers, Ginny
 115 Moore, Sarah Neilson, Tim Peto, Hang Phan, Mark Regan, Anna C. Seale, Nicole
 116 Stoesser, Stephanie Thomas, Jay Turner-Gardner, Vicky Watts, Jimmy Walker, Sarah
 117 Walker, David Wyllie, William Welfare and Neil Woodford.
 118

119 We are grateful to the Infection Control Teams and Microbiology Laboratory staff at
 120 the University Hospital of South Manchester NHS Foundation Trust and the Central
 121 Manchester University Hospitals NHS Foundation Trust; the Sequencing hub at the
 122 Wellcome Trust Center for Human Genetics, Oxford; and the Modernizing Medical
 123 Microbiology Informatics Group (MMMIG), Oxford.
 124

125 **FUNDING**

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

131

132 REFERENCES

- 133 1. Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican
 134 M, Cornaglia G, Garau J, Gniadkowski M, Hayden MK, et al: **Clinical
 135 epidemiology of the global expansion of *Klebsiella pneumoniae*
 136 carbapenemases**. *Lancet Infect Dis* 2013, **13**:785-796.
- 137 2. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES,
 138 Clark TA, Luong K, Song Y, et al: **Single-molecule sequencing to track
 139 plasmid diversity of hospital-associated carbapenemase-producing
 140 Enterobacteriaceae**. *Sci Transl Med* 2014, **6**:254ra126.
- 141 3. Xu G, Jiang Y, An W, Wang H, Zhang X: **Emergence of KPC-2-producing
 142 *Escherichia coli* isolates in an urban river in Harbin, China**. *World J
 143 Microbiol Biotechnol* 2015, **31**:1443-1450.
- 144 4. Montezzi LF, Campana EH, Correa LL, Justo LH, Paschoal RP, da Silva IL,
 145 Souza Mdo C, Drolshagen M, Picao RC: **Occurrence of carbapenemase-
 146 producing bacteria in coastal recreational waters**. *Int J Antimicrob Agents*
 147 2015, **45**:174-177.
- 148 5. Brady C, Cleenwerck I, Venter S, Coutinho T, De Vos P: **Taxonomic
 149 evaluation of the genus *Enterobacter* based on multilocus sequence
 150 analysis (MLSA): proposal to reclassify *E. nimipressuralis* and *E.***

- 151 **amnigenus into Lelliottia gen. nov. as Lelliottia nimipressuralis comb.**
 152 **nov. and Lelliottia amnigena comb. nov., respectively, E. gergoviae and E.**
 153 **pyrinus into Pluralibacter gen. nov. as Pluralibacter gergoviae comb. nov.**
 154 **and Pluralibacter pyrinus comb. nov., respectively, E. cowanii, E.**
 155 **radicincitans, E. oryzae and E. arachidis into Kosakonia gen. nov. as**
 156 **Kosakonia cowanii comb. nov., Kosakonia radicincitans comb. nov.,**
 157 **Kosakonia oryzae comb. nov. and Kosakonia arachidis comb. nov.,**
 158 **respectively, and E. turicensis, E. helveticus and E. pulveris into**
 159 **Cronobacter as Cronobacter zurichensis nom. nov., Cronobacter**
 160 **helveticus comb. nov. and Cronobacter pulveris comb. nov., respectively,**
 161 **and emended description of the genera Enterobacter and Cronobacter.**
 162 *Syst Appl Microbiol* 2013, **36**:309-319.
- 163 6. Richter SS, Sercia L, Branda JA, Burnham CA, Bythrow M, Ferraro MJ,
 164 Garner OB, Ginocchio CC, Jennemann R, Lewinski MA, et al: **Identification**
 165 **of Enterobacteriaceae by matrix-assisted laser desorption/ionization time-**
 166 **of-flight mass spectrometry using the VITEK MS system.** *Eur J Clin*
 167 *Microbiol Infect Dis* 2013, **32**:1571-1578.
- 168 7. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K,
 169 Kyrpides NC, Pati A: **Microbial species delineation using whole genome**
 170 **sequences.** *Nucleic Acids Res* 2015, **43**:6761-6771.
- 171 8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS,
 172 Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al: **SPAdes: a new**
 173 **genome assembly algorithm and its applications to single-cell sequencing.**
 174 *J Comput Biol* 2012, **19**:455-477.

- 175 9. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD: **REAPR:**
176 **a universal tool for genome assembly evaluation.** *Genome Biol* 2013,
177 **14**:R47.
- 178 10. Almeida AC, de Castro KK, Fehlberg LC, Gales AC, Vilela MA, de Moraes
179 MM: **Carbapenem-resistant *Enterobacter gergoviae* harbouring blaKPC-**
180 **2 in Brazil.** *Int J Antimicrob Agents* 2014, **44**:369-370.
- 181 11. Satlin MJ, Jenkins SG, Chen L, Helfgott D, Feldman EJ, Kreiswirth BN,
182 Schuetz AN: **Septic shock caused by *Klebsiella pneumoniae***
183 **carbapenemase-producing *Enterobacter gergoviae* in a neutropenic**
184 **patient with leukemia.** *J Clin Microbiol* 2013, **51**:2794-2796.
- 185 12. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J,
186 Glockner FO: **The SILVA ribosomal RNA gene database project:**
187 **improved data processing and web-based tools.** *Nucleic Acids Res* 2013,
188 **41**:D590-596.
- 189 13. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje
190 JM: **DNA-DNA hybridization values and their relationship to whole-**
191 **genome sequence similarities.** *Int J Syst Evol Microbiol* 2007, **57**:81-91.
- 192 14. Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics*
193 2014, **30**:2068-2069.
- 194 15. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M,
195 Falush D, Keane JA, Parkhill J: **Roary: rapid large-scale prokaryote pan**
196 **genome analysis.** *Bioinformatics* 2015, **31**:3691-3693.
- 197 16. Federhen S, Rossello-Mora R, Klenk H-P, Tindall BJ, Konstantinidis KT,
198 Whitman WB, Brown D, Labeda D, Ussery D, Garrity GM, et al: **Meeting**

- 199 **report: GenBank microbial genomic taxonomy workshop (12–13 May,**
200 **2015). *Standards in Genomic Sciences* 2016, **11**:15.**
- 201 17. Brenner DJ, McWhorter AC, Knutson JK, Steigerwalt AG: **Escherichia**
202 **vulneris: a new species of Enterobacteriaceae associated with human**
203 **wounds. *J Clin Microbiol* 1982, **15**:1133-1140.**
- 204 18. Senanayake SN, Jadeer A, Talaulikar GS, Roy J: **First reported case of**
205 **dialysis-related peritonitis due to Escherichia vulneris. *J Clin Microbiol***
206 **2006, **44**:4283-4284.**
- 207 19. Awsare SV, Lillo M: **A case report of Escherichia vulneris urosepsis. *Rev***
208 ***Infect Dis* 1991, **13**:1247-1248.**
- 209 20. Mohanty S, Chandra SP, Dhawan B, Kapil A, Das BK: **Meningitis due to**
210 **Escherichia vulneris. *Neurol India* 2005, **53**:122-123.**
- 211 21. Tavares CP, Pereira PS, Marques EdA, Faria Jr C, de Souza MdPAH, de
212 Almeida R, Alves CdFM, Asensi MD, Carvalho-Assef APDA: **Molecular**
213 **epidemiology of KPC-2–producing Enterobacteriaceae (non–Klebsiella**
214 **pneumoniae) isolated from Brazil. *Diagnostic Microbiology and Infectious***
215 ***Disease* 2015, **82**:326-330.**
- 216 22. Brady CL, Venter SN, Cleenwerck I, Engelbeen K, Vancanneyt M, Swings J,
217 Coutinho TA: **Pantoea vagans sp. nov., Pantoea eucalypti sp. nov., Pantoea**
218 **deleyi sp. nov. and Pantoea anthophila sp. nov. *Int J Syst Evol Microbiol***
219 **2009, **59**:2339-2345.**
- 220 23. Wan X, Hou S, Phan N, Malone Moss JS, Donachie SP, Alam M: **Draft**
221 **Genome Sequence of Pantoea anthophila Strain 11-2 from Hypersaline**
222 **Lake Laysan, Hawaii. *Genome Announc* 2015, **3**.**

- 223 24. Chan KG, Tee KK, Yin WF, Tan JY: **Complete Genome Sequence of**
224 **Pluralibacter gergoviae FB2, an N-Acyl Homoserine Lactone-Degrading**
225 **Strain Isolated from Packed Fish Paste.** *Genome Announc* 2014, **2**.
- 226 25. Periamé M, Pages JM, Davin-Regli A: **Enterobacter gergoviae adaptation**
227 **to preservatives commonly used in cosmetic industry.** *Int J Cosmet Sci*
228 2014, **36**:386-395.
- 229 26. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J,
230 Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, et al: **Rapid,**
231 **comprehensive, and affordable mycobacterial diagnosis with whole-**
232 **genome sequencing: a prospective study.** *Lancet Respir Med* 2016, **4**:49-58.
- 233 27. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Goker M, Parker CT,
234 Amann R, Beck BJ, Chain PS, Chun J, et al: **Genomic encyclopedia of**
235 **bacteria and archaea: sequencing a myriad of type strains.** *PLoS Biol*
236 2014, **12**:e1001920.
- 237

238 **Table 1.** Top 16S and genome wide species identification results of study *bla*_{KPC}-positive isolates

Isolate id	First hit species (% identity)	Second hit species (% identity)	Average nucleotide identity (ANI) species classification			Final species classification
			Best species match	ANI (bidirectional)	Alignment Fraction (AF) (bidirectional)	
9885_1	<i>Escherichia vulneris</i> (99.74)	<i>Enterobacter hormaechei</i> (99.73)	<i>Escherichia vulneris</i>	96.38 96.37	0.79 0.89	<i>Escherichia vulneris</i>
9885_2	<i>Citrobacter murlinae</i> (99.80)	<i>Enterobacter asburiae</i> LI (99.42)	<i>Enterobacter sp</i> MGH24	86.69 86.68	0.74 0.80	Novel <i>Enterobacter</i> sp.
9885_4	<i>Pantoea anthophila</i> (99.87)	<i>Pantoea agglomerans</i> (99.87)	<i>Pantoea anthophila</i>	98.75 98.75	0.74 0.96	<i>Pantoea anthophila</i>
1613625	<i>Enterobacter</i> sp. <i>ETH-2</i> (100.00)	<i>Pluralibacter gergoviae</i> (99.81)	<i>Pluralibacter gergoviae</i>	98.91 98.91	0.90 0.89	<i>Pluralibacter gergoviae</i>

239

240 **Figure 1.** Core genome phylogenies of *bla*_{KPC}-positive isolates investigated in detail
241 in this study and of closely related reference genomes from NCBI/GenBank. (a)
242 phylogeny of 9885_1, 9885_2 and 1613625 and reference *Enterobacter* spp.
243 genomes; (b) phylogeny of 9885_4 and reference *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

(a)

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

(b)