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

Hang T.T. Phan*[1,2], Zoie Aiken[3†], Oluwafemi Akinremi [2,4], Julie Cawthorne[3], Andrew Dodgson[3,4], Angela Downes[3], Matthew Ellington[2,5], Ryan George[3], Amy Mathers[6], Nicole Stoesser[1], Stephanie Thomas[7], on behalf of the TRACE Investigators' Group

Authors' affiliations:

[1]Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom

[2]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University

[3]Central Manchester University Hospitals NHS Foundation Trust, Oxford Road, Manchester, M13 9WL, UK

[4]Public Health England, London, UK

[5]Antimicrobial Resistance and Healthcare Associated Infections Reference Unit (AMRHAI), Public Health England, London, UK

[6] Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia Health System, Charlottesville, Virginia, USA

[7]Department 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.

Main text (1000/1000 words)

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

(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.

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 *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 bla_{KPC} in this study is therefore potentially of clinical significance. *Pantoea* spp. are recognized as opportunistic pathogens and whilst *P. agglomerans* has been associated with bla_{KPC} previously[21], *P.anthophila*, found in lakes and diseased plants[22, 23], has not. Worryingly, in this survey, it was one of the most widespread bla_{KPC}-positive environmental species (36 isolates from 8 sites from one ward, including clinical areas, staff areas and ward kitchens). *P. anthophila* may represent a successful intermediate species for the wider transmission of bla_{KPC} amongst more pathogenic strains within/between non-clinical/clinical sites. *P. gergoviae* has been isolated from a wide range of sources, including foodstuffs and insects[24, 25], but has not previously been identified with bla_{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 bla_{KPC} in the hospital environment, one a putative, novel *Enterobacter* sp. Accurate species identification of bla_{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*$_{KPC}$ in diverse species in the hospital environment supports previous data suggesting that this reservoir plays a role in the dissemination of *bla*$_{KPC}$ 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

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

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

1. Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, Cornaglia G, Garau J, Gniadkowski M, Hayden MK, et al: **Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases.** *Lancet Infect Dis* 2013, **13:**785-796.
2. Conlan S, Thomas PJ, Deming C, Park M, Lau AF, Dekker JP, Snitkin ES, Clark TA, Luong K, Song Y, et al: **Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae.** *Sci Transl Med* 2014, **6:**254ra126.
3. Xu G, Jiang Y, An W, Wang H, Zhang X: **Emergence of KPC-2-producing Escherichia coli isolates in an urban river in Harbin, China.** *World J Microbiol Biotechnol* 2015, **31:**1443-1450.
4. Montezzi LF, Campana EH, Correa LL, Justo LH, Paschoal RP, da Silva IL, Souza Mdo C, Drolshagen M, Picao RC: **Occurrence of carbapenemase-producing bacteria in coastal recreational waters.** *Int J Antimicrob Agents* 2015, **45:**174-177.
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. *Syst Appl Microbiol* 2013, **36:**309-319.

6. Richter SS, Sercia L, Branda JA, Burnham CA, Bythrow M, Ferraro MJ, Garner OB, Ginocchio CC, Jennemann R, Lewinski MA, et al: **Identification of Enterobacteriaceae by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using the VITEK MS system.** *Eur J Clin Microbiol Infect Dis* 2013, **32:**1571-1578.

7. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A: **Microbial species delineation using whole genome sequences.** *Nucleic Acids Res* 2015, **43:**6761-6771.

8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al: **SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.** *J Comput Biol* 2012, **19:**455-477.

9. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD: **REAPR: a universal tool for genome assembly evaluation.** *Genome Biol* 2013, **14:**R47.

10. Almeida AC, de Castro KK, Fehlberg LC, Gales AC, Vilela MA, de Morais MM: **Carbapenem-resistant Enterobacter gergoviae harbouring blaKPC-2 in Brazil.** *Int J Antimicrob Agents* 2014, **44:**369-370.

11. Satlin MJ, Jenkins SG, Chen L, Helfgott D, Feldman EJ, Kreiswirth BN, Schuetz AN: **Septic shock caused by Klebsiella pneumoniae carbapenemase-producing Enterobacter gergoviae in a neutropenic patient with leukemia.** *J Clin Microbiol* 2013, **51:**2794-2796.

12. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.** *Nucleic Acids Res* 2013, **41:**D590-596.

13. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM: **DNA-DNA hybridization values and their relationship to whole-genome sequence similarities.** *Int J Syst Evol Microbiol* 2007, **57:**81-91.

14. Seemann T: **Prokka: rapid prokaryotic genome annotation.** *Bioinformatics* 2014, **30:**2068-2069.

15. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J: **Roary: rapid large-scale prokaryote pan genome analysis.** *Bioinformatics* 2015, **31:**3691-3693.

16. Federhen S, Rossello-Mora R, Klenk H-P, Tindall BJ, Konstantinidis KT, Whitman WB, Brown D, Labeda D, Ussery D, Garrity GM, et al: **Meeting**

report: GenBank microbial genomic taxonomy workshop (12–13 May, 2015). *Standards in Genomic Sciences* 2016, **11:**15.

17. Brenner DJ, McWhorter AC, Knutson JK, Steigerwalt AG: **Escherichia vulneris: a new species of Enterobacteriaceae associated with human wounds.** *J Clin Microbiol* 1982, **15:**1133-1140.

18. Senanayake SN, Jadeer A, Talaulikar GS, Roy J: **First reported case of dialysis-related peritonitis due to Escherichia vulneris.** *J Clin Microbiol* 2006, **44:**4283-4284.

19. Awsare SV, Lillo M: **A case report of Escherichia vulneris urosepsis.** *Rev Infect Dis* 1991, **13:**1247-1248.

20. Mohanty S, Chandra SP, Dhawan B, Kapil A, Das BK: **Meningitis due to Escherichia vulneris.** *Neurol India* 2005, **53:**122-123.

21. Tavares CP, Pereira PS, Marques EdA, Faria Jr C, de Souza MdPAH, de Almeida R, Alves CdFM, Asensi MD, Carvalho-Assef APDA: **Molecular epidemiology of KPC-2–producing Enterobacteriaceae (non–Klebsiella pneumoniae) isolated from Brazil.** *Diagnostic Microbiology and Infectious Disease* 2015, **82:**326-330.

22. Brady CL, Venter SN, Cleenwerck I, Engelbeen K, Vancanneyt M, Swings J, Coutinho TA: **Pantoea vagans sp. nov., Pantoea eucalypti sp. nov., Pantoea deleyi sp. nov. and Pantoea anthophila sp. nov.** *Int J Syst Evol Microbiol* 2009, **59:**2339-2345.

23. Wan X, Hou S, Phan N, Malone Moss JS, Donachie SP, Alam M: **Draft Genome Sequence of Pantoea anthophila Strain 11-2 from Hypersaline Lake Laysan, Hawaii.** *Genome Announc* 2015, **3**.

24. Chan KG, Tee KK, Yin WF, Tan JY: **Complete Genome Sequence of Pluralibacter gergoviae FB2, an N-Acyl Homoserine Lactone-Degrading Strain Isolated from Packed Fish Paste.** *Genome Announc* 2014, **2**.

25. Periame M, Pages JM, Davin-Regli A: **Enterobacter gergoviae adaptation to preservatives commonly used in cosmetic industry.** *Int J Cosmet Sci* 2014, **36:**386-395.

26. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, et al: **Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study.** *Lancet Respir Med* 2016, **4:**49-58.

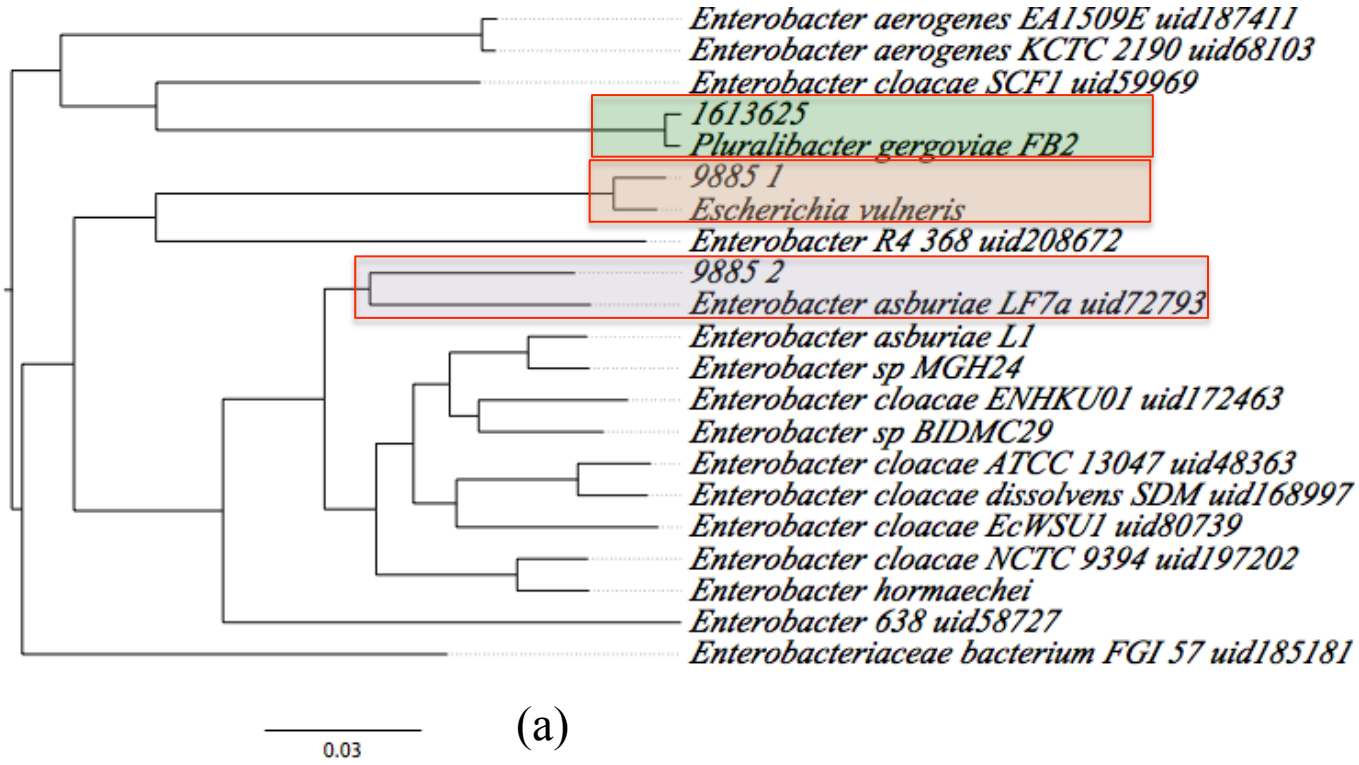
27. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Goker M, Parker CT, Amann R, Beck BJ, Chain PS, Chun J, et al: **Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains.** *PLoS Biol* 2014, **12:**e1001920.

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	*Escherichia vulneris* (99.74)	*Enterobacter hormaechei* (99.73)	*Escherichia vulneris*	96.38|96.37	0.79|0.89	*Escherichia vulneris*
9885_2	*Citrobacter murliniae* (99.80)	*Enterobacter asburiae L1* (99.42)	*Enterobacter sp MGH24*	86.69|86.68	0.74|0.80	Novel *Enterobacter* sp.
9885_4	*Pantoea anthophila* (99.87)	*Pantoea agglomerans* (99.87)	*Pantoea anthophila*	98.75|98.75	0.74|0.96	*Pantoea anthophila*
1613625	*Enterobacter sp. ETH-2(100.00)*	*Pluralibacter gergoviae* (99.81)	*Pluralibacter gergoviae*	98.91|98.91	0.90|0.89	*Pluralibacter gergoviae*

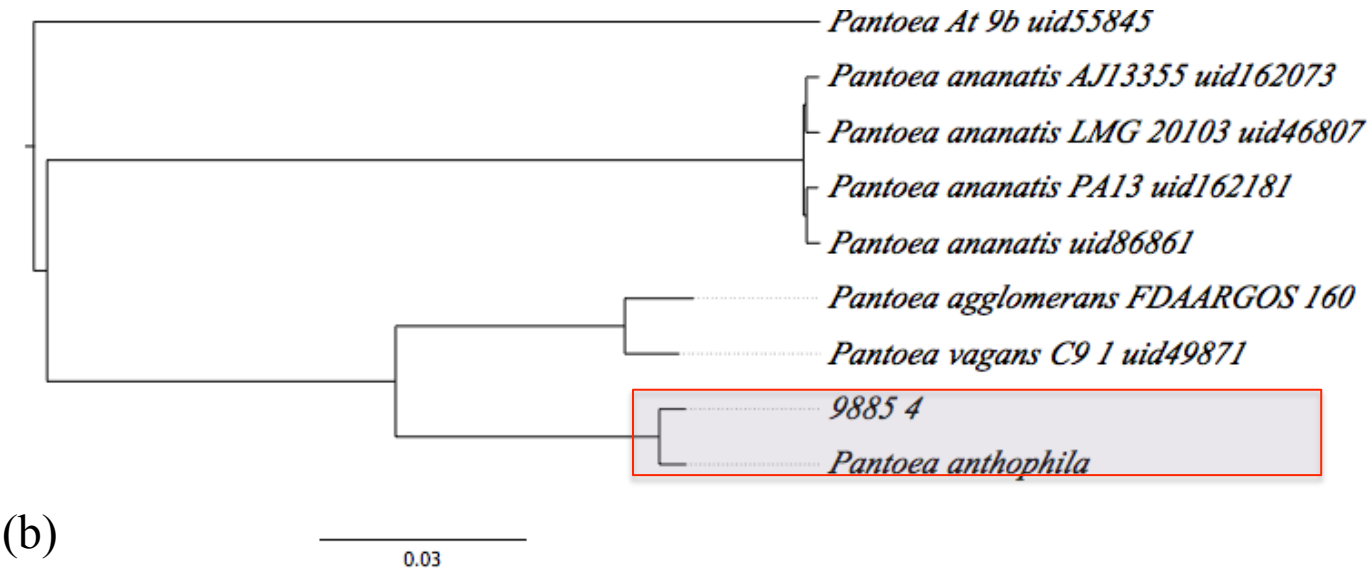
239

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)