



## Description of *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and *Klebsiella variicola* subsp. *variicola* subsp. nov.

Carla Rodrigues, Virginie Passet, Andriniana Rakotondrasoa, Thierno Abdoulaye Diallo, Alexis Criscuolo, Sylvain Brisse

### ► To cite this version:

Carla Rodrigues, Virginie Passet, Andriniana Rakotondrasoa, Thierno Abdoulaye Diallo, Alexis Criscuolo, et al.. Description of *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and *Klebsiella variicola* subsp. *variicola* subsp. nov.. Research in Microbiology, 2019, 170 (3), pp.165-170. 10.1016/j.resmic.2019.02.003 . pasteur-03265228

**HAL Id: pasteur-03265228**

**<https://pasteur.hal.science/pasteur-03265228>**

Submitted on 25 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

**Description of *Klebsiella africanensis* sp. nov.,  
*Klebsiella variicola* subsp. *tropicalensis* subsp. nov.  
and *Klebsiella variicola* subsp. *variicola* subsp. nov.**

Carla Rodrigues<sup>a†</sup>, Virginie Passet<sup>a†</sup>, Andriniaaina Rakotondrasoa<sup>b</sup>, Thierno Abdoulaye  
Diallo<sup>c</sup>, Alexis Criscuolo<sup>d</sup> & Sylvain Brisse<sup>a,\*</sup>

<sup>a</sup> Institut Pasteur, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, France

<sup>b</sup> Institut Pasteur Madagascar, Antananarivo, Madagascar

<sup>c</sup> Institut Pasteur of Dakar, Senegal

<sup>d</sup> Institut Pasteur, Bioinformatics and Biostatistics Hub, C3BI, USR 3756 IP CNRS, Paris,  
France.

† These authors contributed equally to this work.

**E-mail addresses:** carla.parada-rodrigues@pasteur.fr (C. Rodrigues),  
virginie.passet@pasteur.fr (V. Passet), aina@pasteur.mg (A. Rakotondrasoa),  
Thierno.Diallo@pasteur.sn (T. A. Diallo), alexis.criscuolo@pasteur.fr (A. Criscuolo),  
sylvain.brisse@pasteur.fr (S. Brisse) “\*Correspondence and reprints”

## Abstract

The bacterial pathogen *Klebsiella pneumoniae* comprises several phylogenetic groups (Kp1 to Kp7), two of which (Kp5 and Kp7) have no taxonomic status. Here we show that group Kp5 is closely related to *Klebsiella variicola* (Kp3), with an average nucleotide identity (ANI) of 96.4%, and that group Kp7 has an ANI of 94.7% with Kp1 (*K. pneumoniae sensu stricto*). Biochemical characteristics and chromosomal beta-lactamase genes also distinguish groups Kp5 and Kp7 from other *Klebsiella* taxa. We propose the names *K. africanensis* for Kp7 (type strain, 200023<sup>T</sup>) and *K. variicola* subsp. *tropicalensis* for Kp5 (type strain, 1266<sup>T</sup>).

**Keywords:** *Klebsiella pneumoniae* complex; phylogeny; taxonomy; genome sequencing; beta-lactamase; human carriage

**Abbreviations:** ANI, average nucleotide identity; isDDH, *in silico* DNA-DNA hybridization

## 1. Introduction

The genus *Klebsiella* includes important human and animal pathogens and is largely distributed in animal carriage and the environment [1]. *Klebsiella pneumoniae* is the most common species isolated from infections of humans and other hosts. Brisse & Verhoef [2] reported the existence of *K. pneumoniae* phylogroups Kp1 to Kp4, and Blin *et al.* [3] reported two additional phylogroups, Kp5 and Kp6. Over the years, phylogenetic studies have demonstrated that *K. pneumoniae* phylogroups Kp1, Kp2, Kp4 and Kp3 correspond to different taxa, respectively *K. pneumoniae* (*sensu stricto*), *K. quasipneumoniae* subsp. *quasipneumoniae* and subsp. *similipneumoniae*, and *K. variicola* [4,5]. Phylogenetic analyses of Kp6 genomic sequences showed that this phylogroup corresponds to a recently proposed taxon, '*K. quasivariicola*' [6], which remains to be validly published. While screening for *K. pneumoniae* fecal carriage in Senegal, we identified strain 200023 (internal strain bank identifier, SB5857), which was not associated with any of the previous phylogroups or species, and which we here call Kp7. Together, *K. pneumoniae sensu stricto* and the other above referenced taxa and phylogroups constitute the *K. pneumoniae* complex. Typically, all members of this species complex are misidentified as *K. pneumoniae* or as *K. variicola* using standard laboratory methods, which masks their actual clinical and epidemiological significance [7–9]. The members of the *K. pneumoniae* complex can be correctly distinguished based on whole-genome sequencing (WGS) or by sequencing taxonomic marker genes such as *rpoB*. The chromosomal beta-lactamase gene also varies according to phylogroup: Kp1, Kp2, Kp3 and Kp4 harbor *bla<sub>SHV</sub>*, *bla<sub>OKP-A</sub>*, *bla<sub>LEN</sub>* and *bla<sub>OKP-B</sub>* beta-lactamase genes, respectively [10,11]. Recently, MALDI-TOF mass spectrometry protein biomarkers were described for the identification of *K. pneumoniae* complex members (Kp1 to Kp6) [12]. The aim of the present work was to define the taxonomic status of *K. pneumoniae* phylogroups Kp5 and Kp7.

## 2. Materials and methods

A total of thirty-seven strains of species, subspecies and phylogroups related to *K. pneumoniae* were included (Table 1, Fig. 1). Strain 1266<sup>T</sup> (internal strain bank identifier, SB5531) was chosen as reference strain for phylogroup Kp5, whereas strain 200023<sup>T</sup> (internal strain bank identifier, SB5857) was included as reference strain of Kp7 phylogroup. The genome (Accession Number ERS214332) of the strain 38679 [13] was included for phylogenetic comparisons even though the strain was not available at the time of this study. Strains were grown in tryptocasein soy agar (TSA) (BioRad, Marnes-La-Coquette, France). The biochemical characterization of the isolates was performed using API20E strips (BioMérieux) and Biolog phenotype microarrays plates PM1 and PM2 (Hayward, CA) following the protocol described by Blin *et al.* [3]. To extract the DNA, bacteria were plated on TSA and one colony was grown overnight with shaking at 37°C in 10 ml tryptocasein soy broth. DNA was extracted using the Wizard Genomic DNA purification kit (Promega, Charbonnières-les-Bains, France) and stored at -20°C. Genomic sequencing was performed using NextSeq-500 instrument (Illumina, San Diego, USA) with a 2 x 150 nt paired-end protocol.

Contig sequences were assembled using SPAdes v3.10.0 [14] and annotated with Prokka v1.12 [15]. JSpeciesWS [16] was used to calculate the BLAST-based average nucleotide identity (ANI). The *in silico* DNA-DNA hybridization (isDDH) was estimated using the GGDC tool (<http://ggdc.dsmz.de; formula 2>) [17]. Barrnap (<https://github.com/tseemann/barrnap>) was used to detect and extract 16S rRNA sequences from genome assemblies, whereas internal portions of genetic markers (*fusA*, *gapA*, *gyrA*, *leuS*, *rpoB*) were detected and extracted using BLAST and compared with previously reported sequences [4]. Chromosomal beta-lactamases were also extracted, and the new LEN and OKP

amino-acid sequence variants were submitted to the Institut Pasteur MLST nomenclature database (<https://bigsdb.pasteur.fr/klebsiella>) for variant number attribution, and to NCBI for accession number attribution. Phylogenetic analyses based on the 16S rRNA gene sequence, concatenation of five housekeeping genes (2,609 aligned bp in total) and chromosomal beta-lactamases were performed using MEGA v7.0 with the neighbor-joining method. Genetic distances were estimated using the Jukes-Cantor substitution model in the case of nucleotide sequences and using the Jones-Taylor-Thornton (JTT) model for beta-lactamase protein sequences. For genome-based phylogenetic analysis, a phylogenetic tree was reconstructed from the concatenation of 1,703 genes defined as core genes using Roary v3.12 [18] with a BLASTP identity cut-off of 90%. *K. oxytoca* ATCC 13182<sup>T</sup> (ERS2016112) was used as outgroup. Aligned characters affected by recombination events were detected and removed from the core gene alignment using Gubbins v2.2.0 [19]. A maximum-likelihood phylogenetic tree was inferred using FastTree v2.1.7 [20].

### 3. Results and Discussion

The nearly complete (1,462 nt) *rrs* gene sequence coding for 16S rRNA was compared among *K. pneumoniae* strains. Only a few sites showed reliable variation (Fig. S1), and all sequences were > 98.8% similar. The Kp5 reference strain 1266<sup>T</sup> differed from F2R9<sup>T</sup> (*K. variicola* type strain) by 6 (0.41%) nucleotide positions, whereas Kp7 strain 200023<sup>T</sup> differed from KPN1705 (Kp6 reference strain) and DSM 30104<sup>T</sup> (*K. pneumoniae* type strain) by 12 (0.82%) nucleotide positions in both cases. The phylogenetic information content of the highly conserved 16S rRNA sequence did not allow to reliably infer a phylogeny of 16S sequences (Fig. S1) as previously reported [21,22].

Maximum-likelihood phylogeny based on the recombination-purged nucleotide sequence alignments of 1,703 core genes (Fig. 1) showed seven highly supported branches. Phylogroup Kp5 was strongly associated with Kp3 (*K. variicola*), whereas Kp7 was associated with Kp6 and Kp1 (*K. pneumoniae*) groups. The latter result suggests that Kp6 and Kp7 share a common ancestor with *K. pneumoniae* and may therefore represent the closest relatives of the clinically most significant member of the species complex. Phylogenies based on the five concatenated housekeeping genes *fusA*, *gapA*, *gyrA*, *leuS* and *rpoB*, and of chromosomal beta-lactamases, both corroborated the distinction among the seven groups (Fig. S2 and Fig. S3). The phylogeny of chromosomal beta-lactamases was highly concordant with core genes phylogeny, with the remarkable exception of phylogroup Kp6, which harbored two different types of beta-lactamases, *bla*<sub>LEN</sub> and *bla*<sub>OKP-D</sub>, depending on the strain. This is consistent with frequent homologous recombination observed in the Kp6 phylogroup using Gubbins (data not shown) and as previously reported [23]. In order to estimate the genome-wide divergence among phylogroups, average nucleotide identity (ANI) was used. The ANI values of Kp5 (strain 1266<sup>T</sup>) with *K. variicola* F2R9<sup>T</sup> (Kp3) was 96.4%, above the cut-off value (approximately 95-96%) proposed for species distinction [24,25]. In contrast, the ANI

values of Kp7 strain 200023<sup>T</sup> with ‘*K. quasivariicola*’ KPN1705 (Kp6) and *K. pneumoniae* DSM30104<sup>T</sup> (Kp1) were 94.9% and 94.7%, respectively (Table 2). The isDDH relatedness of 1266<sup>T</sup> (Kp5) with Kp3 strains ranged from 75.0% to 75.5%, but ranged from 51.2% and 56.0% with the other phylogroups. For 200023<sup>T</sup> the isDDH relatedness with Kp1 to Kp6 phylogroups was between 50.5% and 61.9%, well below the ~70% cut-off used to define bacterial species. Therefore, both ANI and isDDH analyses lead us to consider that 1266<sup>T</sup> is a subspecies of *K. variicola* (Kp3) and that 200023<sup>T</sup> represents a novel *Klebsiella* species.

The phenotypic characteristics of the strains were compared (Table S1, Fig. 1). Using API20E strips we confirmed that all Kp5 strains and 200023<sup>T</sup> were positive for the utilization of lactose and mannitol, for the urease, malonate, lysine decarboxylase, Voges–Proskauer and ONPG tests, and reduced nitrate to nitrite, whereas they were negative for indole and ornithine decarboxylase. Furthermore, microscopy confirmed that all strains were non-motile. The use of phenotype microarrays provided insights into the metabolic profiles of the seven phylogroups and identified carbon sources that differentiate Kp5 and Kp7 strains from other groups (Fig. 1). The inability to metabolize mono-methyl succinate, D-lactic acid methyl ester and 3-O-(β-D-galactopyranosyl)-D-arabinose, and the ability to metabolize D-psicose differentiated Kp5 strains from *K. variicola*. Kp7 strain 200023<sup>T</sup> was unique in being unable to metabolize D-arabitol and was also differentiated from ‘*K. quasivariicola*’ and *K. pneumoniae*, its closest phylogenetic neighbors, by the ability to metabolize D-L-carnitine and 4-hydroxy-L-proline, respectively.

Based on the above genomic and phenotypic characteristics, we propose that Kp5 be considered as a subspecies of *K. variicola*, which we propose to name *K. variicola* subsp. *tropicalensis*; whereas Kp7 represents a novel species, which we propose to name *K. africanensis*.



### 3.1. Description of *Klebsiella variicola* subsp. *tropicalensis* subsp. nov.

(tro.pi.cal.en'sis. N.L. fem. adj. *tropicalensis*, referring to the tropical area of Earth, from where Kp5 strains have been isolated).

The description is based on six strains. Cells are gram negative, non-motile, non-spore forming, straight, rod-shaped, capsulated. Colonies are smooth, circular, white, dome-shaped, glistening. The general characteristics are as described for *K. pneumoniae*. Urease positive, ONPG positive, Voges-Proskauer test positive, indole negative. Lysine decarboxylase positive, ornithine decarboxylase negative. Distinguished from the other *K. pneumoniae* complex members by the characteristics listed in Table S1. *K. variicola* subsp. *tropicalensis* strains are able to metabolize tricarballic acid, L-galactonic acid- $\gamma$ -lactone, L-sorbose, 5-keto-D-gluconic acid, D-psicose, 4-hydroxyl-L-proline and D-arabitol. They are not able to metabolize N-acetyl-neuraminic acid, adonitol, mono-methyl succinate, D-lactic acid methyl ester, L-alaninamide and 3-O-b-galactopyranosyl-D-arabinose. *K. variicola* subsp. *tropicalensis* isolates were recovered from asymptomatic fecal carriage and from the environment.

The type strain is strain 1266<sup>T</sup> (= SB5531<sup>T</sup>, {CIP and DSMZ numbers: *in process*}), isolated in 2016 from a fecal sample of a human asymptomatic carrier in Antananarivo, Madagascar. The EMBL (GenBank/DDBJ) accession numbers of the *rrs* (coding for 16S rRNA) gene is MK040621. The genome sequence accession number is ERS2787528.

### 3.2. Description of *Klebsiella variicola* subsp. *variicola* subsp. nov. (va.ri.i'co.la.

L. adj. varius different, differing, various, L. suffix n. -cola inhabitant, N.L. fem./masc. n. *variicola* inhabitant of different places)

In accordance to Rule 46 of the Bacteriological Code, the description of *Klebsiella variicola* subsp. *tropicalensis* automatically creates a second subspecies, *Klebsiella variicola*

subsp. *variicola* subsp. nov. for which the type strain is F2R9<sup>T</sup> (= ATCC BAA-830<sup>T</sup>, DSM 15968<sup>T</sup>) [6]. In addition to the phenotypic characters consistent within all members of the *K. pneumoniae* complex, reactions common to both subspecies of *K. variicola* are the ability to metabolize tricarballic acid, L-galactonic acid- $\gamma$ -lactone, L-sorbose, 5-keto-D-gluconic acid, 4-hydroxyl-L-proline and D-arabitol, and the inability to ferment adonitol and N-acetylneuraminic acid. The ability to metabolize mono-methyl succinate, D-lactic acid methyl ester and 3-O-( $\beta$ -D-galactopyranosyl)-D-arabinose, and the inability to metabolize D-psicose differentiated *K. variicola* subsp. *variicola* from *K. variicola* subsp. *tropicalensis*.

### 3.3. Description of *Klebsiella africanensis* sp. nov. (a.fri.can.en'sis. N.L. fem. adj. *africanensis*, referring to Africa, from where Kp7 strains have been isolated).

The phenotypic description is based on one strain. Cells are gram negative, non-motile, non-spore forming, straight, rod-shaped, capsulated. Colonies are smooth, circular, white, dome-shaped, glistening. The general characteristics are as described for *K. pneumoniae*. Urease positive, ONPG positive, Voges-Proskauer test positive, indole negative. Lysine decarboxylase positive, ornithine decarboxylase negative. Distinguished from the other *K. pneumoniae* complex members by the characteristics listed in Table S1. Distinguished from all the other *K. pneumoniae* complex members by the inability to metabolize D-arabitol. *K. africanensis* type strain is able to metabolize tricarballic acid, L-galactonic acid- $\gamma$ -lactone, L-sorbose, 5-keto-D-gluconic acid, D-psicose, L-carnitine, 4-hydroxyl-L-proline, D-dulcitol, and D-tagatose. Is not able to metabolize N-acetylneuraminic acid, adonitol, D-lactic acid methyl ester, L-alaninamide and 3-O- $\beta$ -galactopyranosyl-D-arabinose and D-arabitol. So far, *K. africanensis* was recovered from asymptomatic human fecal carriage in Senegal (this study) and from human clinical samples in Kenya [13].

The type strain is strain 200023<sup>T</sup> (= SB5857<sup>T</sup>, {CIP and DSMZ numbers: *in process*}), isolated in 2016 from a fecal sample of a human asymptomatic carrier in Sokone, Senegal. The EMBL (GenBank/DDBJ) accession numbers of the *rrs* (coding for 16S rRNA) gene is MK040622. The genome sequence accession number is ERS2787533.

**3.4. Nucleotide sequence accession numbers.** The genomic sequences of strains 1266<sup>T</sup> and 200023<sup>T</sup> were submitted to the European Nucleotide Archive under accession numbers ERS2787528 and ERS2787533, respectively. 16S rRNA sequences were also submitted individually under the accession numbers MK040621 and MK040622, respectively. The genome sequence data of strains 814, 885, 1283, 1375, 01-467-2ECBU and 01-310MBV were submitted to the European Nucleotide Archive under accession numbers ERS2787526, ERS2787527 and ERS2787529 to ERS2787532, respectively (under the BioProject number PRJEB29143).

**Conflicts of interest**

The authors declare that there is no conflict of interest.

**Funding information**

This work received financial support from Institut Pasteur and from the French Government's Investissement d'Avenir program Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (grant number ANR-10-LABX-62-IBEID). CR was supported financially by the MedVetKlebs project, a component of European Joint Programme One Health EJP, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.

## 223   **Acknowledgements**

224           We thank the ‘P2M-Plateforme de Microbiologie Mutualisée’ (PIBnet) from Institut  
225   Pasteur for genomic sequencing. We acknowledge Jean-Marc Collard (Institut Pasteur of  
226   Madagascar), Bich-Tram Huynh (Institut Pasteur, France) and Raymond Bercion (Institut  
227   Pasteur of Dakar) for support in *K. pneumoniae* carriage studies. We also thank to S. Wesley  
228   Long (Houston Methodist Hospital) and David A. Rasko and J. Kristie Johnson (University of  
229   Maryland School of Medicine, Baltimore) for providing strains analyzed in this study.

230

## References

- [1] Brisse S, Grimont, F, Grimont PAD. The genus *Klebsiella*, in Dworkin M., Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (Eds), *The Prokaryotes*, Springer, New York, 2006.
- [2] Brisse S, Verhoef J. Phylogenetic diversity of *Klebsiella pneumoniae* and *Klebsiella oxytoca* clinical isolates revealed by randomly amplified polymorphic DNA, *gyrA* and *parC* genes sequencing and automated ribotyping. *Int J Syst Evol Microbiol* 2001;51:915–924.
- [3] Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. Metabolic diversity of the emerging pathogenic lineages of *Klebsiella pneumoniae*. *Environ Microbiol* 2017;19:1881–1898.
- [4] Brisse S, Passet V, Grimont PAD. Description of *Klebsiella quasipneumoniae* sp. nov., a novel species isolated from human infections, with two subspecies *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., and demonstration that *Klebsiella singaporensis* is a junior heterotypic synonym of *Klebsiella variicola*. *Int J Syst Evol Microbiol* 2014;64:3146–3152.
- [5] Rosenblueth M, Martínez L, Silva J, Martínez-Romero E. *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. *Syst Appl Microbiol* 2004;27:27–35.
- [6] Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, et al. Whole-genome sequencing of a human clinical isolate of the novel species *Klebsiella quasivariicola* sp. nov. *Genome Announc* 2017;5.
- [7] Brisse S, van Himbergen T, Kusters K, Verhoef J. Development of a rapid identification method for *Klebsiella pneumoniae* phylogenetic groups and analysis of 420 clinical isolates. *Clin Microbiol Infect* 2004;10:942–945.

- 255 [8] Seki M, Gotoh K, Nakamura S, Akeda Y, Yoshii T, Miyaguchi S, et al. Fatal sepsis  
256 caused by an unusual *Klebsiella* species that was misidentified by an automated  
257 identification system. J Med Microbiol 2013;62:801–803.
- 258 [9] Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, et al. Whole-  
259 genome sequencing of human clinical *Klebsiella pneumoniae* isolates reveals  
260 misidentification and misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*,  
261 and *Klebsiella quasipneumoniae*. MSphere 2017;2:e00290-17.
- 262 [10] Fevre C, Passet V, Weill F-X, Grimont PAD, Brisse S. Variants of the *Klebsiella*  
263 *pneumoniae* OKP chromosomal beta-lactamase are divided into two main groups, OKP-  
264 A and OKP-B. Antimicrob Agents Chemother 2005;49:5149–52.
- 265 [11] Haeggman S, Lofdahl S, Paauw A, Verhoef J, Brisse S. Diversity and Evolution of the  
266 class a chromosomal beta-lactamase gene in *Klebsiella pneumoniae*. Antimicrob Agents  
267 Chemother 2004;48:2400–2408.
- 268 [12] Rodrigues C, Passet V, Rakotondrasoa A, Brisse S. Identification of *Klebsiella*  
269 *pneumoniae*, *Klebsiella quasipneumoniae*, *Klebsiella variicola* and related phylogroups  
270 by MALDI-TOF mass spectrometry. Front Microbiol 2018;9:3000.
- 271 [13] Henson SP, Boinett CJ, Ellington MJ, Kagia N, Mwarumba S, Nyongesa S, et al.  
272 Molecular epidemiology of *Klebsiella pneumoniae* invasive infections over a decade at  
273 Kilifi County Hospital in Kenya. Int J Med Microbiol IJMM 2017;307:422–9.
- 274 [14] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes:  
275 A new genome assembly algorithm and its applications to single-cell sequencing. J  
276 Comput Biol 2012;19:455–477. doi:10.1089/cmb.2012.0021.
- 277 [15] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl  
278 2014;30:2068–9.

- 279 [16] Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server  
280 for prokaryotic species circumscription based on pairwise genome comparison.  
281 Bioinforma Oxf Engl 2016;32:929–31.
- 282 [17] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species  
283 delimitation with confidence intervals and improved distance functions. BMC  
284 Bioinformatics 2013;14:60.
- 285 [18] Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid  
286 large-scale prokaryote pan genome analysis. Bioinformatics 2015;31:3691–3.
- 287 [19] Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid  
288 phylogenetic analysis of large samples of recombinant bacterial whole genome  
289 sequences using Gubbins. Nucleic Acids Res 2015;43:e15–e15.
- 290 [20] Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees  
291 for large alignments. PloS One 2010;5:e9490.
- 292 [21] Naum M, Brown EW, Mason-Gamer RJ. Is 16S rDNA a reliable phylogenetic marker to  
293 characterize relationships below the family level in the *Enterobacteriaceae*? J Mol Evol  
294 2008;66:630–42.
- 295 [22] Boye K, Hansen DS. Sequencing of 16S rDNA of *Klebsiella*: taxonomic relations within  
296 the genus and to other *Enterobacteriaceae*. Int J Med Microbiol 2003;292:495–503.
- 297 [23] Hazen TH, Zhao L, Sahl JW, Robinson G, Harris AD, Rasko DA, et al. Characterization  
298 of *Klebsiella* sp. strain 10982, a colonizer of humans that contains novel antibiotic  
299 resistance alleles and exhibits genetic similarities to plant and clinical *Klebsiella* isolates.  
300 Antimicrob Agents Chemother 2014;58:1879–88.
- 301 [24] Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic  
302 species definition. Proc Natl Acad Sci USA 2009;106:19126–31.



303 [25] Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for  
304 prokaryotes. *Proc Natl Acad Sci USA* 2005;102:2567–72.  
305

## Figure Legends

### **Figure 1.** Maximum likelihood phylogeny and carbon sources utilization.

**Left:** Phylogenetic tree (FastTree) based on the final recombination-free alignment of concatenated nucleotide sequence alignments of 1,703 core genes. Major taxonomic groups are indicated along the branches. Branch lengths represent the number of nucleotide substitutions per site (scale, 0.1 substitution per site). Bootstrap values are indicated at major nodes. **Right:** Metabolic phenotypes for the 16 most discriminant carbon sources. Black horizontal lines delimit phylogenetic lineages. Black squares correspond to substrate utilization; white square to absence of utilization; and grey squares to intermediate values. Strain 38679 (ERS214332) was not available for phenotyping.

**Figure S1.** Phylogenetic relationships (neighbor-joining method, Jukes-Cantor correction) based on the sequence of the *rrs* gene coding for 16S rRNA and the respective multiple sequence alignment restricted to the variable positions.

The tree was rooted using *K. oxytoca* ATCC 13182<sup>T</sup> sequence. Bootstrap proportions obtained after 1000 replicates are indicated at the nodes. Branch lengths represent the number of nucleotide substitutions per site (scale, 0.001 substitution per site). Strain labels are given as Strain Bank ID (*e.g.*, SB11) followed by original strain name, followed by phylogroup. A ‘T’ after the strain name indicates type strains.

**Figure S2.** Phylogenetic relationships (neighbor-joining method, Jukes-Cantor correction) based on the concatenated sequences of the five-individual protein-coding genes *fusA*, *gapA*, *gyrA*, *leuS* and *rpoB*.

The was rooted using *K. oxytoca* ATCC 13182<sup>T</sup>. Bootstrap proportions obtained after 1000 replicates are indicated at the nodes. Branch lengths represent the number of nucleotide substitutions per site (scale, 0.01 substitution per site). Strain labels are given as Strain Bank ID (*e.g.*, SB11) followed by original strain name, followed by phylogroup. A ‘T’ after the strain name indicates type strains.

**Figure S3.** Phylogenetic tree based on the core gene nucleotide sequences (A, maximum likelihood) compared to the gene tree of the chromosomal beta-lactamase amino-acid sequences (B, neighbor-joining method, Jones-Taylor-Thornton model).

Branch lengths represent the number of nucleotide (A) or amino-acid (B) substitutions per site (scale, 0.1 substitution per site). Major taxonomic groups are indicated along the branches of the phylogenetic tree A (*Kqs*, *K. quasipneumoniae* subsp. *similipneumoniae*; *Kqq*, *K. quasipneumoniae* subsp. *quasipneumoniae*, *Kvv*, *K. variicola* subsp. *variicola*, *Kvt*, *K. variicola* subsp. *tropicalensis*; *Kpn*, *K. pneumoniae*; *Kqv*, ‘*K. quasivariicola*’, *Kaf*, *K. africanensis*). Strain labels are given as Strain Bank ID (*e.g.*, SB11) followed by original strain name, followed by phylogroup. A ‘T’ after the strain name indicates type strains. In panel B the chromosomal beta-lactamase and the corresponding accession number are indicated beside the strain names. Node labels indicate bootstrap values based on 1,000 replicates. Red stars indicate remarkable discrepancies between the two phylogenetic trees (phylogroup Kp6 presents two types of chromosomal beta-lactamases).

350 **Table 1.** Strains included in this study

<b>Taxonomic</b>	<b>PhG<sup>a</sup></b>	<b>Strain</b>	<b>Strain Name</b>	<b>Isolation</b>	<b>Host</b>	<b>Source</b>	<b>Country</b>	<b>City</b>	<b>ST<sup>c</sup></b>	<b>Accession no.</b>	<b>Intrinsic Beta-</b>
<b>Designation</b>		<b>bank (SB)</b>		<b>Year</b>							<b>lactamase</b>
		<b>ID<sup>b</sup></b>									(Accession no.)
<i>Klebsiella pneumoniae</i>	Kp1	SB1067	SB4-2	2002	Human	Feces	Netherlands	Utrecht	17	ERS2786693	SHV-11 (NG_050000)
					carriage						
<i>K. pneumoniae</i>	Kp1	SB132	ATCC 13883 <sup>T</sup>	NA	Human	Blood	NA	NA	3	GCA_000742135.1	SHV-11 (EJK91031)
<i>K. pneumoniae</i>	Kp1	SB107	MGH 78578	1994	Human	Blood	NA	NA	38	CP000647	SHV-11 (NG_050000)
<i>K. pneumoniae</i>	Kp1	SB1139	n.a.	2002	Human	Feces	Netherlands	Utrecht	37	ERS2786694	SHV-11 (NG_050000)
					carriage						
<i>K. pneumoniae</i>	Kp1	SB617	5-2	2000	Water	Natuurgebied	Netherlands	Utrecht	55	ERS2786695	SHV-11 (NG_050000)
						canal (trouble					
						water)					
<i>K. pneumoniae</i>	Kp1	SB20	04A025	1997	Human	Blood	France	Paris	15	ERS2786696	SHV-28 (NG_051877)
<i>K. pneumoniae</i>	Kp1	SB612	2-3	2000	Water	Rijnhauwen	Netherlands	Utrecht	50	ERS2786697	SHV-75 (NG_050110)
						bridge					
<i>K. pneumoniae</i>	Kp1	SB4938	Kp13	2009	Human	Blood	Brazil	Londrina	442	CP003999	SHV-110 (NG_050001)
<i>K. pneumoniae</i>	Kp1	SB3928	NTUH-K2044	NA	Human	Liver abscess	Taiwan	Taiwan	23	AP006725	SHV-11 (NG_050000)
<i>K. pneumoniae</i>	Kp1	SB4496	BJ1-GA	2011	Human	Liver abscess	France	Clichy	380	ERS2786698	SHV-207 (NG_062279)

Taxonomic Designation	PhG <sup>a</sup>	Strain bank (SB) ID <sup>b</sup>	Strain Name	Isolation Year	Host	Source	Country	City	ST <sup>c</sup>	Accession no.	Intrinsic Beta-lactamase (Accession no.)
<i>K. quasipneumoniae</i> subsp. <i>quasipneumoniae</i>	Kp2	SB11	01A030 <sup>T</sup>	1997	Human	Blood	Austria	Linz	1528	GCA_000751755.1	OKP-A-3 (NG_049363)
<i>K. quasipneumoniae</i> subsp. <i>quasipneumoniae</i>	Kp2	SB1124	n.a.	2002	Environment	Canal water	Netherlands	Baarn	2274	ERS2786699	not present
<i>K. quasipneumoniae</i> subsp. <i>quasipneumoniae</i>	Kp2	SB2110	U41	1990	Environment	Environment	Germany	NA	526	ERS2786700	OKP-A-11 (NG_049355)
<i>K. quasipneumoniae</i> subsp. <i>quasipneumoniae</i>	Kp2	SB59	18A69	1997	Human	Blood	Spain	Barcelona	1118	ERS2786701	OKP-A-2 (NG_049362)
<i>K. quasipneumoniae</i> subsp. <i>quasipneumoniae</i>	Kp2	SB98	0320584	NA	Environment	Environment	NA	NA	622	ERS2786702	OKP-A-5 (NG_049365)
<i>K. variicola</i> subsp. <i>variicola</i>	Kp3	SB1	01A065	1997	Human	Blood	Austria	Linz	2273	ERS2786703	LEN-16 (NG_049270)
<i>K. variicola</i> subsp. <i>variicola</i>	Kp3	SB48	F2R9 <sup>T</sup>	NA	Food	Banana	Mexico	Cuernavaca	2263	ERS2786704	LEN-33 (NG_049271)
<i>K. variicola</i> subsp. <i>variicola</i>	Kp3	SB579	Kp342	NA	Plant	Maize	USA	NA	146	CP000964	LEN-28 (NG_060510)

Taxonomic Designation	PhG <sup>a</sup>	Strain bank (SB) ID <sup>b</sup>	Strain Name	Isolation Year	Host	Source	Country	City	ST <sup>c</sup>	Accession no.	Intrinsic Beta-lactamase (Accession no.)
<i>K. variicola</i> subsp. <i>variicola</i>	Kp3	SB4767	At22	NA	Environment	Fungus garden	NA	NA	1220	CP001891	LEN-13 (NG_050777)
<i>K. quasipneumoniae</i> subsp. <i>similipneumoniae</i>	Kp4	SB164	09A323	1997	Human	Blood	Greece	Athens	18	ERS2786705	OKP-B-8 (NG_049384)
<i>K. quasipneumoniae</i> subsp. <i>similipneumoniae</i>	Kp4	SB203	12A476	1998	Human	Blood	Netherlands	Utrecht	384	ERS2786706	OKP-B-16 (MK161455)
<i>K. quasipneumoniae</i> subsp. <i>similipneumoniae</i>	Kp4	SB30	07A044 <sup>T</sup>	1997	Human	Blood	Germany	Freiburg	1215	GCA_000613225.1	OKP-B-1 (NG_049370)
<i>K. quasipneumoniae</i> subsp. <i>similipneumoniae</i>	Kp4	SB4697	CIP110288 <sup>T</sup> ( <i>K. alba</i> )	NA	Environment	Farmland soil	China	Nanjing	2275	ERS2786707	OKP-B-5 (NG_049381)
<i>K. quasipneumoniae</i> subsp. <i>similipneumoniae</i>	Kp4	SB610	1-1	2000	Environment	Lake kikker	Netherlands	NA	2276	ERS2786708	OKP-B-24 (MK161456)
<i>K. variicola</i> subsp. <i>tropicalensis</i>	Kp5	SB94	CDC 4241-71	NA	Environment	Environment	NA	NA	1216	ERS2786710	LEN-38 (MK161457)
<i>K. variicola</i> subsp. <i>tropicalensis</i>	Kp5	SB5387	814	2015	Human	Fecal sample	Madagascar	Antananarivo	2466	ERS2787526	LEN-39 (MK161458)

Taxonomic Designation	PhG <sup>a</sup>	Strain bank (SB) ID <sup>b</sup>	Strain Name	Isolation Year	Host	Source	Country	City	ST <sup>c</sup>	Accession no.	Intrinsic Beta-lactamase (Accession no.)
<i>K. variicola</i> subsp. <i>tropicalensis</i>	Kp5	SB5439	885	2016	Human	Fecal sample	Madagascar	Antananarivo	2486	ERS2787527	LEN-40 (MK161459)
<i>K. variicola</i> subsp. <i>tropicalensis</i>	Kp5	SB5531	1266 <sup>T</sup>	2016	Human	Fecal sample	Madagascar	Moramanga	2561	ERS2787528	LEN-43 (MK161460)
<i>K. variicola</i> subsp. <i>tropicalensis</i>	Kp5	SB5544	1283	2016	Human	Fecal sample	Madagascar	Moramanga	2605	ERS2787529	LEN-41 (MK161461)
<i>K. variicola</i> subsp. <i>tropicalensis</i>	Kp5	SB5610	1375	2016	Human	Fecal sample	Madagascar	Moramanga	2600	ERS2787530	LEN-42 (MK161462)
<i>K. quasivariicola</i>	Kp6	SB33	08A119	1997	Human	Blood	Germany	Dusseldorf	1214	ERS2786709	OKP-D-1 (MK161463)
<i>K. quasivariicola</i>	Kp6	SB6071	10982	2005	Human	Peri-rectal	USA	Maryland	1155	GCA_000523395.1	LEN-26 (NG_049281)
<i>K. quasivariicola</i>	Kp6	SB6096	KPN1705	2014	Human	Wound	USA	Houston	209	CP022823	LEN-26 (NG_049281)
<i>K. quasivariicola</i>	Kp6	SB6094	01-467-2ECBU	2015	Human	Feces	Madagascar	Antananarivo	2830	ERS2787531	OKP-D-1 (MK161463)
<i>K. quasivariicola</i>	Kp6	SB6095	01-310MBV	2013	Human	Vaginal swab	Madagascar	Antananarivo	3303	ERS2787532	LEN-26 (NG_049281)
<i>K. africanensis</i>	Kp7	SB5857	200023 <sup>T</sup>	2016	Human	Fecal sample	Senegal	Sokone	3291	ERS2787533	OKP-C-1 (MK161464)
<i>K. africanensis</i>	Kp7	n.a.	38679 <sup>d</sup>	2010	Human	NA	Kenya	Kilifi	2831	ERS214332	OKP-C-1 (MK161464)

352 <sup>a</sup>PhG, *K. pneumoniae* phylogroup; <sup>b</sup>Internal strain collection number of the Biodiversity and Epidemiology of Bacterial Pathogens unit, Institut Pasteur; <sup>c</sup>ST, Sequence-Type; <sup>d</sup>Sequence from  
353 public databases; this strain was not analysed phenotypically.

354



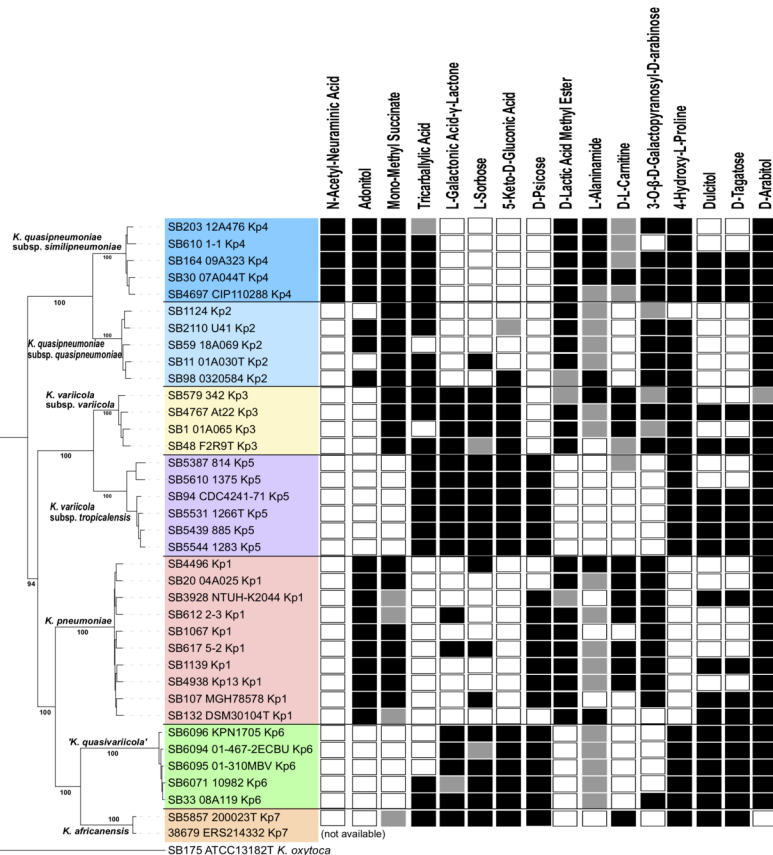
**Table 2.** Average nucleotide identity (ANI) values obtained among seven members of the *Klebsiella pneumoniae* complex

Query genome <sup>a</sup>	Size (nucleotides)	DNA G+C content (mol%)	Average nucleotide identity of test genome against query genomes						
			Kp1	Kp2	Kp3	Kp4	Kp5	Kp6	Kp7
<b>Kp1</b>	5 234 536	57.3	*	93.21	94.14	93.40	93.47	93.51	94.69
<b>Kp2</b>	5 343 479	58.1	93.21	*	93.06	96.27	92.80	92.36	92.64
<b>Kp3</b>	5 470 009	57.6	94.03	92.94	*	93.23	96.65	93.49	93.66
<b>Kp4</b>	5 084 979	58.3	93.49	96.33	93.42	*	93.03	92.52	92.82
<b>Kp5</b>	5 698 764	57.1	93.22	92.40	96.38	92.69	*	93.00	93.03
<b>Kp6</b>	5 540 188	56.9	93.50	92.37	93.58	92.50	93.27	*	94.85
<b>Kp7</b>	5 156 720	57.3	94.66	92.58	93.75	92.66	93.31	94.86	*

<sup>a</sup>Kp1, *K. pneumoniae* DSM30104<sup>T</sup>; Kp2, *K. quasipneumoniae* subsp. *quasipneumoniae* 01A030<sup>T</sup>; Kp3, *K. variicola* subsp. *variicola* F2R9<sup>T</sup>; Kp4, *K. quasipneumoniae* subsp. *similipneumoniae* 07A044<sup>T</sup>; Kp5, *K. variicola* subsp. *tropicalensis* 1266<sup>T</sup>; Kp6, *K. quasivariicola* KPN1705; Kp7, *K. africanensis* 200023<sup>T</sup>

**Respiration**

■ high  
 ■ intermediate  
 □ none



Tree scale: 0.1