Supplementary text

Protocol for whole-genome sequencing

Whole genome sequencing was performed with Illumina MiSeq (Illumina, Inc., San Diego, CA). Library preparation for Illumina MiSeq sequencing was performed with QIAseq FX DNA Library Kit (Qiagen, Tokyo, Japan). Libraries were sequenced on a MiSeq system for 600 cycles (300-bp paired-end reads). Raw reads generated by Miseq were quality trimmed with Trimmomatic tool (version 0.38) and assembled using SPAdes (version 3.12.0).

Identification of bacterial species

Bacterial species were identified using KmerFinder 2.5 by adopting the species of the sequences showing highest scores (1). Inaccurate registrations of the bacterial species of reference sequences were corrected during the process (Acccession No. CP024784: *K. quasipneumoniae* registered as *Enterobacteriaceae*; CP024784: *K. quasipneumoniae* registered as *K. pneumoniae*) (2).

Capsular genotyping

Capsular genotyping was performed using Kaptive web and only results showing match confidence of Perfect, Very high, High, or Good were adopted (3).

Protocol for Single-Nucleotide Polymorphism (SNP) Identification

Isolates belonging to K1-ST23 or K62-ST36 were selected for SNP identification. *Klebsiella pneumoniae* NTUH-2044 isolate belonging to K1-ST23 (GenBank accession number AP006725.1) was selected for the reference isolate (4). The MiSeq sequencing reads were aligned to the genomic sequence of the reference isolate using the Burrows-Wheeler Aligner (BWA) with ‘MEM’ option (5). We constructed a provisional core genome alignment for each

group (K1-ST23 and K62-ST36) using SAMtools (version 1.1) mpileup (6) and VarScan (version 2.3.7) mpileup2cns (7) and then a maximum-likelihood tree using PhyML (8). Using this as the starting tree, we inferred homologous recombination sites, which should be excluded from the core-genomes, with ClonalFrameML (9). SNPs of the confirmed core genomes (4,983,362 bp (93.6% of the genomic sequence of NTUH-2044) for K1-ST23 isolates and 4,821,982 bp (90.6% of the genomic sequence of NTUH-2044) for K62-ST36 isolates) were identified with snp-dists (https://github.com/tseemann/snp-dists).

References

1. Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Pontén T, Aarestrup FM, Ussery DW, Lund O. 2014. Benchmarking of methods for genomic taxonomy. J Clin Microbiol 52: 1529-1539.
2. Kitajima M, Ishizaki S, Jang J, Ishii S, Okabe S. 2018. Complete Genome Sequence of *Klebsiella quasipneumoniae* Strain S05, a Fouling-Causing Bacterium Isolated from a Membrane Bioreactor. Genome Announc 6: e00471-18.
3. Wick RR, Heinz E, Holt KE, Wyres KL. 2018. Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes. J Clin Microbiol 56: e00197-18.
4. Wu KM, Li LH, Yan JJ, Tsao N, Liao TL, Tsai HC, Fung CP, Chen HJ, Liu YM, Wang JT, Fang CT, Chang SC, Shu HY, Liu TT, Chen YT, Shiau YR, Lauderdale TL, Su IJ, Kirby R, Tsai SF. 2009. Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. J Bacteriol 191:4492-4501.
5. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

6. Li H1, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
7. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22: 568-576.
8. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307-321.
9. Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol 11:e1004041.