

Supplementary Information

Supplementary Note 1: Data and software availability

The software of analysis pipelines and data are provided in

(<http://www.genomicsresearch.org/public/researcher/npAnalysis/>. A script for each analysis (species detection, MLTS typing, strain typing and antibiotic resistance profile is provided). Parameters for the tools are also included in these scripts. All data is deposited in the Sequence Read Archive:

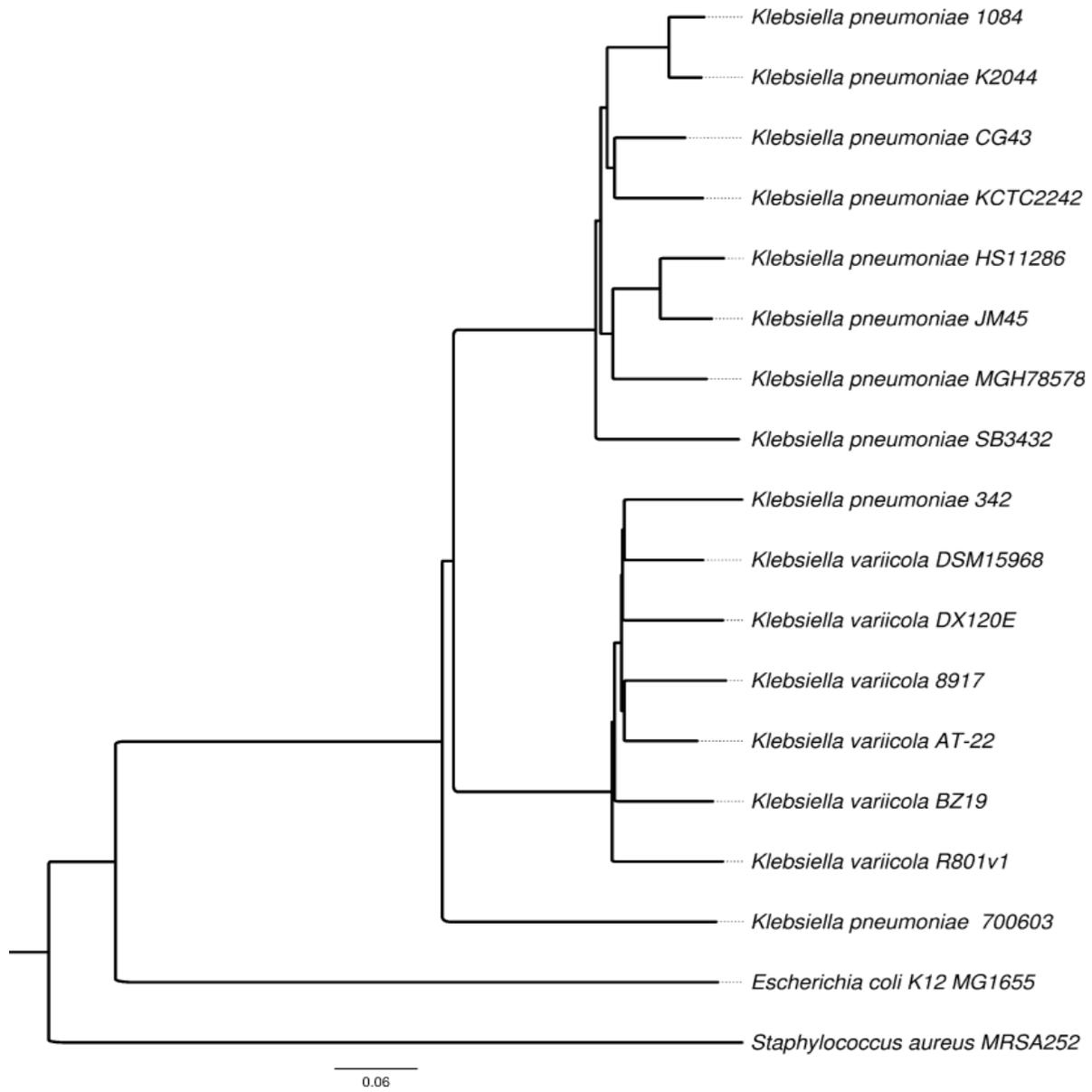
<http://www.ebi.ac.uk/ena/data/view/ERP010377>. The source code of the software is publicly available in github repository japsa (<https://github.com/mdcao/japsa>).

Supplementary Note 2: Phylogenetic analysis of *Klebsiella pneumoniae* and *Klebsiella variicola* strains

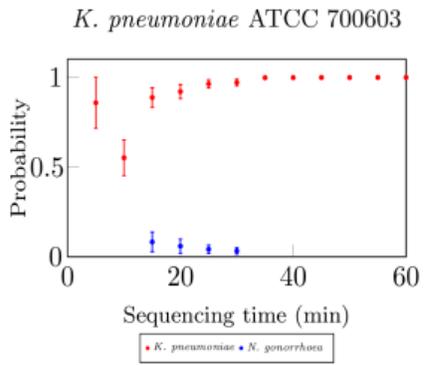
We examined evolutionary relationship of the *K.pneumoniae* strain ATCC-700603 with a number of *K.pneumoniae* and *Klebsiella variicola* strains. We collected the assembly of the *K.pneumoniae* strain ATCC-700603, together with the assemblies of nine *K.pneumoniae* and five *K.variicola* strains from Genbank (See Supplementary Table 1). We included the genomes of an *Escherichia coli* strain and a *Staphylococcus aureus* strain as the out-groups. We then used XMDistance¹ to estimate the genetic distance matrix of these strains and generated the phylogenetic tree using neighbour joining method² from PHYLIP package³. In this phylogeny, the eight *K.pneumoniae* strains (1084, CG43, HS11286, JM45, KTCT2242, MGH 78578, NTUH K2044 and SB3432) were grouped together into one cluster, while strain 342 was grouped together with the six *K.variicola* strains, which is consistent with the recent suggestion that this strain was a misclassification⁴. Our analysis also showed that strain ATCC-700603 was an ancestor of the other *K.pneumoniae* and *K.variicola* strains which explains the shared identity of its genome.

Supplementary Table 1: List of bacterial strains used in the phylogenetic analysis of *Klebsiella pneumoniae* strain ATCC-700603.

Strain	Assembly ID	Assembly status
<i>Klebsiella pneumoniae</i> ATCC-700603	GCA_000349245.1	Contig
<i>Klebsiella pneumoniae</i> 1084	GCA_000294365.1	Complete
<i>Klebsiella pneumoniae</i> 342	GCA_000019565.1	Complete
<i>Klebsiella pneumoniae</i> CG43	GCA_000474015.1	Complete
<i>Klebsiella pneumoniae</i> HS11286	GCA_000240185.2	Complete
<i>Klebsiella pneumoniae</i> JM45	GCA_000445405.1	Complete
<i>Klebsiella pneumoniae</i> KCTC 2242	GCA_000220485.1	Complete
<i>Klebsiella pneumoniae</i> MGH 78578	GCA_000016305.1	Complete
<i>Klebsiella pneumoniae</i> NTUH K2044	GCA_000009885.1	Complete
<i>Klebsiella pneumoniae</i> SB3432	GCA_000967845.1	Complete
<i>Klebsiella variicola</i> AT-22	GCA_000025465.1	Complete
<i>Klebsiella variicola</i> DX120E	GCA_000812205.1	Complete
<i>Klebsiella variicola</i> DSM 15968	GCA_000828055.1	Complete
<i>Klebsiella variicola</i> BZ19	GCA_000742915.1	Contig
<i>Klebsiella variicola</i> 801	GCA_000826565.1	Contig
<i>Klebsiella variicola</i> 8917	GCA_000941635.1	Contig
<i>Escherichia coli</i> str. K-12 substr. MG1655	GCA_000005845.2	Complete
<i>Staphylococcus aureus</i> subsp. aureus MRSA252	GCA_000011505.1	Complete



Supplementary Figure 1: Phylogenetic tree of *Klebsiella pneumoniae* and *Klebsiella variicola* clades. The *Escherichia coli* K12 MG1655 and *Staphylococcus aureus* MRSA252 strains were used as the outgroups.



Supplementary Figure 2: Species typing result of *Klebsiella pneumoniae* ATCC-700603 strain when its reference assembly was included into the species database.

Supplementary Note 3: Examining the profiles of most probable MLST types

In MLST typing of strain BAA-2146 using Nanopore sequencing data, we found the correct type (ST-11) tied with types ST-751 and ST-864 as the highest log-likelihood score (1985.47). By examining the gene allele profiles of these types, we found each of the types ST-751 and ST-864 differed to type ST-11 by only one gene allele (alleles rpoB_46 and rpoB_1 and alleles tonB_2 and tonB_4, See Supplementary Table 2). We aligned these gene alleles, and found that these differences were only by one SNP (See Supplementary Data.txt). We further examined the fourth and fifth highest score types (ST-1080 and ST-1680) and noticed that they also differed to the correct type by one gene allele. These alleles were respectively determined by a deletion of 2 bases (phoE_164) and an insertion of four bases (tonB_111). The ability to discriminate strains based on only a small difference clearly showed the usability of Nanopore sequencing data.

Supplementary Table 2: Profiles of top five five probable MLTS types probable MLTS types from typing strain BAA-2146. Each type differs to the correct type by only one gene allele.

ST	gapA	infB	mdh	pgi	phoE	rpoB	tonB	Log-likelihood
11	3	3	1	1	1	1	4	1985.47
751	3	3	1	1	1	46	4	1985.47
864	3	3	1	1	1	1	2	1985.47
1080	3	3	1	1	164	1	4	1984.46
1680	3	3	1	1	1	1	111	1982.62

REFERENCES:

- 1 Cao, M. D., Lloyd, A. & Trevor, D. I. in *Proceedings of the 22rd Australian Joint Conference on Artificial Intelligence* 71-80 (Springer Berlin Heidelberg, Melbourne, Australia, 2009).
- 2 Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-425 (1987).
- 3 PHYLIP (Phylogeny Inference Package) (Department of Genome Sciences, University of Washington, Seattle, 2005).
- 4 Kumar, V. *et al.* Comparative genomics of *Klebsiella pneumoniae* strains with different antibiotic resistance profiles. *Antimicrobial agents and chemotherapy* **55**, 4267-4276, doi:10.1128/AAC.00052-11 (2011).