

S3 Table for: Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*

Kelly L Wyres^{1*}, Ryan R Wick¹, Louise M Judd¹, Roni Froumine¹, Alex Tokolyi², Claire L Gorrie³, Margaret M C Lam¹, Sebastián Duchêne², Adam Jenney⁴ and Kathryn E Holt^{1,2,5}

¹Department of Infectious Diseases, Monash University, Melbourne, Victoria, Australia

²Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria, Australia

³Department of Infectious Diseases and Microbiology Unit, The Alfred Hospital, Melbourne, Victoria, Australia

⁴Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology and Immunology, University of Melbourne, Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia

⁵London School of Hygiene and Tropical Medicine, Keppel Street, London, UK

* Corresponding author; kelly.wyres@monash.edu

S3 Table: Outcomes of statistical tests to assess the influence of sampling bias

Outcome variable	Explanatory variable		
	Sample size	Geographic diversity	Nucleotide diversity
	P (uncorrected) * significant after correction	P (uncorrected) * significant after correction	P (uncorrected) * significant after correction
Pairwise nucleotide diversity (before recombination removal, median)	0.235	0.054	n/a
Pairwise nucleotide diversity (after recombination removal, median)	0.282	0.101	n/a
r/m	0.742	0.295	0.200
# K-loci	7.12 x 10⁻⁰⁵ *	0.392	0.291
K locus diversity	0.259	0.826	0.411
# O loci	0.016 *	0.781	0.541
O-locus diversity	0.560	0.612	0.962
Gene content			
Jaccard distance (median)	0.836	0.361	0.770
Pan-genome			
Euclidean distance (median)	0.448	0.996	0.519
Pan-genome curve alpha	0.429	0.320	0.299
Accessory gene ancestral diversity	0.706	0.544	0.424
Phage Euclidean distance (median)	0.023	0.253	0.5
Plasmid replicon diversity	0.017 *	0.022	0.173
Plasmid <i>mob</i> diversity	0.187	0.926	0.533

Variables are clone-specific summary metrics tested in independent general linear models. Explanatory variables: Sample size = total number of genomes in the clone sample; Geographic diversity = effective Shannon's diversity of continent of isolate collection (as defined in S1 Table, excluding isolates from unknown geographies); Nucleotide diversity = median pairwise nucleotide diversity after removal of recombinant regions detected by Gubbins. Uncorrected p-values are shown. * indicates significance after Bonferroni multiple testing correction (within rows, P(type 1 error) <0.05).