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

Kelly L Wyres¹, Ryan R Wick¹, Louise M Judd¹, Roni Froumine¹, Alex Tokolyi², Claire L Gorrie³, Margaret M C Lam¹, Sebastián Duchêne², Adam Jenney⁴ and Kathryn E Holt¹,²,⁵

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

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Sample size</th>
<th>Geographic diversity</th>
<th>Nucleotide diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P (uncorrected) * significant after correction</td>
<td>P (uncorrected) * significant after correction</td>
<td>P (uncorrected) * significant after correction</td>
</tr>
<tr>
<td>Pairwise nucleotide diversity (before recombination removal, median)</td>
<td>0.235</td>
<td>0.054</td>
<td>n/a</td>
</tr>
<tr>
<td>Pairwise nucleotide diversity (after recombination removal, median)</td>
<td>0.282</td>
<td>0.101</td>
<td>n/a</td>
</tr>
<tr>
<td>r/m</td>
<td>0.742</td>
<td>0.295</td>
<td>0.200</td>
</tr>
<tr>
<td># K-loci</td>
<td>7.12 x 10^{-5} *</td>
<td>0.392</td>
<td>0.291</td>
</tr>
<tr>
<td>K locus diversity</td>
<td>0.259</td>
<td>0.826</td>
<td>0.411</td>
</tr>
<tr>
<td># O loci</td>
<td>0.016 *</td>
<td>0.781</td>
<td>0.541</td>
</tr>
<tr>
<td>O-locus diversity</td>
<td>0.560</td>
<td>0.612</td>
<td>0.962</td>
</tr>
<tr>
<td>Gene content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jaccard distance (median)</td>
<td>0.836</td>
<td>0.361</td>
<td>0.770</td>
</tr>
<tr>
<td>Pan-genome Euclidean distance (median)</td>
<td>0.448</td>
<td>0.996</td>
<td>0.519</td>
</tr>
<tr>
<td>Pan-genome curve alpha</td>
<td>0.429</td>
<td>0.320</td>
<td>0.299</td>
</tr>
<tr>
<td>Accessory gene ancestral diversity</td>
<td>0.706</td>
<td>0.544</td>
<td>0.424</td>
</tr>
<tr>
<td>Phage Euclidean distance (median)</td>
<td>0.023</td>
<td>0.253</td>
<td>0.5</td>
</tr>
<tr>
<td>Plasmid replicon diversity</td>
<td>0.017 *</td>
<td>0.022</td>
<td>0.173</td>
</tr>
<tr>
<td>Plasmid <em>mob</em> diversity</td>
<td>0.187</td>
<td>0.926</td>
<td>0.533</td>
</tr>
</tbody>
</table>

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).