

9. Appendices

The data listed in each appendix can be found in an enclosed CD at the back of this thesis. Alternatively, they can be accessed through the URLs provided below. Should more information is needed, please do not hesitate to contact me at cc12@sanger.ac.uk

Appendix A: Epidemiological data of Maela collection

The table contains strains, accession codes and associated epidemiological data for the Maela data set. From left to right, the columns represent strain name, accession codes associated with the data deposited in the ENA, dominant population clusters (BC1 to BC 7) used for focused analyses in chapter 4 and 5, primary BAPS cluster, secondary BAPS cluster, date of collection, serotype, sequence type, beta-lactam susceptibility and co-trimoxazole resistance.

doi:10.1038/ng.2895 (Supplementary Table 1)

<http://www.nature.com/ng/journal/v46/n3/full/ng.2895.html#supplementary-information>

Appendix B: Epidemiological data of PMEN14 collection

The table contains strains, accession codes and associated epidemiological data for the PMEN14 dataset. From left to right, the columns represent accession codes of each strain as deposited in the ENA, strain name, serotype, sequence type, region and country of origin (used for analyses in chapter 3), year of collection, and sources of diseases.

doi: 10.1093/gbe/evu120 (Supplementary Table S1)

<http://gbe.oxfordjournals.org/content/suppl/2014/06/04/evu120.DC1>

Appendix C: Associations to beta-lactam non-susceptibility co-detected in separate Maela and Massachusetts data sets

The table summarises all association statistics, linkage disequilibrium (LD) analysis, biological relevance and literature references for co-detected SNPs and associated loci in the Maela and Massachusetts data. From left to right, the columns represent coordinates from the reference genome (*S. pneumoniae* ATCC 700669), coding regions in which associations were detected, putative resistance nucleotide alleles, putative sensitivity nucleotide alleles, minor allele frequency (MAF), odds ratios (OR), synonymous and non-synonymous changes, positions on protein sequences with observed amino acid alterations, amino acid residues, alternative amino acid residues, literature reports, PubMed identifier (PMID), and linkage information.

doi:10.1371/journal.pgen.1004547.s005 (Table S1)

<http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1004547#s5>

Appendix D: Associations to beta-lactam detected in the Maela data.

The table summarises all association statistics and linkage disequilibrium (LD) analyses for variations that show significant associations in the Maela data. From left to right, the columns represent coordinate from the reference genome (*S. pneumoniae* ATCC 700669), majority detected SNPs, minor allele frequency, minor detected SNPs, odd ratios, linkage information and gene information.

doi:10.1371/journal.pgen.1004547.s006 (Table S2)

<http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1004547#s6>

Appendix E: Associations detected in the Massachusetts data

Summary of all association statistics and LD analysis for variations that show significant association in Massachusetts data. Columns were summarized using the same scheme described in Appendix C.

doi:10.1371/journal.pgen.1004547.s007 (Table S3)

<http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1004547#s7>