Methods for *S. pneumoniae* analysis

De novo genome assembly, annotation, pangenome, and phylogenetic analysis

Each of the 616 *S. pneumoniae* genomes was independently assembled using Spades v3.13.1 (1) with default parameters. Assembled genomes were annotated with Prokka v1.14.0 (2) and default parameters. We then used Roary v3.12.0 (3) to characterize the pangenome of the population, including the '-z' parameter to generate alignments for each gene in the pangenome. Each gene alignment was aligned using MAFFT v7.407 (4). A core genome phylogenetic tree was generated using the aligned concatenation of all core genes from the pangenome and the tool RAxML v8.2.11 (5) with a general time reversible (GTR) nucleotide substitution model, four gamma categories for rate heterogeneity, and 100 bootstrap replicates. Phylogenies were visualized using the Interactive Tree of Life (6).

Detection of recombination

To identify recombination, we used fastGEAR (7) with default parameters on individual core and shared accessory genes identified by Roary. Prior to running fastGEAR, the protein specific headers in each FASTA gene alignment were replaced with a genome name using a custom script to make fastGEAR results comparable between genes. The custom script has been provided with HERO on its GitHub page (https://github.com/therealcooperpark/hero) as "sidekick.py" for reproducibility and convenience when using HERO in similar workflows.

Supplementary References

1. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol [Internet]. 2012 May [cited 2018 Jul 19];19(5):455–77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22506599

2. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.

3. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics [Internet]. 2015 Nov 15 [cited 2018 Jul 19];31(22):3691–3. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv421

4. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform [Internet]. 2017 Sep 6 [cited 2018 May 22]; Available from: http://academic.oup.com/bib/article/doi/10.1093/bib/bbx108/4106928/MAFFT-online-service-multiple-sequence-alignment

5. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics [Internet]. 2006 Nov 1 [cited 2018 May 22];22(21):2688–90. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btl446

6. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res [Internet]. 2016 Jul 8 [cited 2018 Jul 23];44(W1):W242–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27095192

7. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. Efficient Inference of Recent and Ancestral Recombination within Bacterial Populations. Mol Biol Evol. 2017 May 1;34(5):1167–82.