PHYLOGENETIC INFERENCE OF THE TRANSMISSION DIRECTION OF PNEUMOCOCCAL INFECTION, A VALIDATION STUDY

Hackman J,[a] Sheppard C,[b] Sobkowiak B,[c] Phelan J,[c] Shah S,[c] Litt D,[b] Fry NK,[b,d] Hibberd ML,[c] Miller L,[a] Flasche S,[a] Hué S [a]

BACKGROUND

- To ameliorate the cost of PCVs, vaccine strategies may be targeted towards mitigating transmission through herd immunity.
- A better understanding of pneumococcal transmission pathways and their role in herd immunity is therefore needed.

AIM

- To validate current phylogenetic inference methods for detecting the occurrence and direction of pneumococcal transmission.

METHODS

- Longitudinal nasopharyngeal swabs from 10 transmission pairs were cultured and isolates were sequenced from whole plate sweeps using Illumina MiSeq next-generation sequencing.
- All analyses was conducted on sequences alone and blinded to epidemiological data.
- Maximum likelihood phylogenetic trees were reconstructed from consensus sequences with 1000 bootstrap replicates to identify source-recipient pairs.
- Samples were assessed for multiple carriage using LoFreq v2.1.4[(1)] and major and minor strains were parsed and analyzed independently.
- Direction of transmission was inferred and compared to inference from epidemiological linkage based on timing. Phyloscanner v1.4.7[(2)] was used and possible scenarios were:

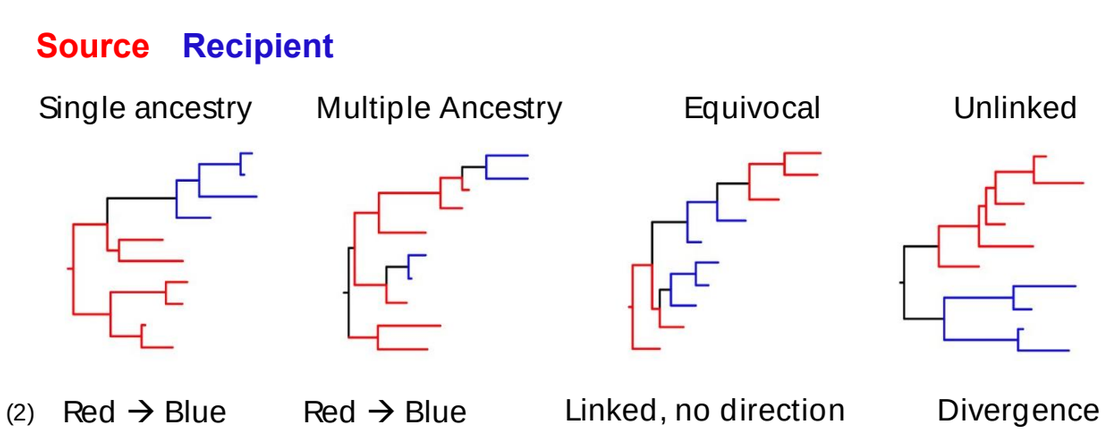

Fig 1. Phyloscanner scans across the genome with specified window size and minimum number of single nucleotide polymorphism (SNPs) and reconstructs trees from these regions. The possible scenarios are red infects blue (correct), vice versa, blue infects red (not shown, incorrect), linked, and unlinked. The most occurred scenario across the whole genome is the inferred transmission direction.

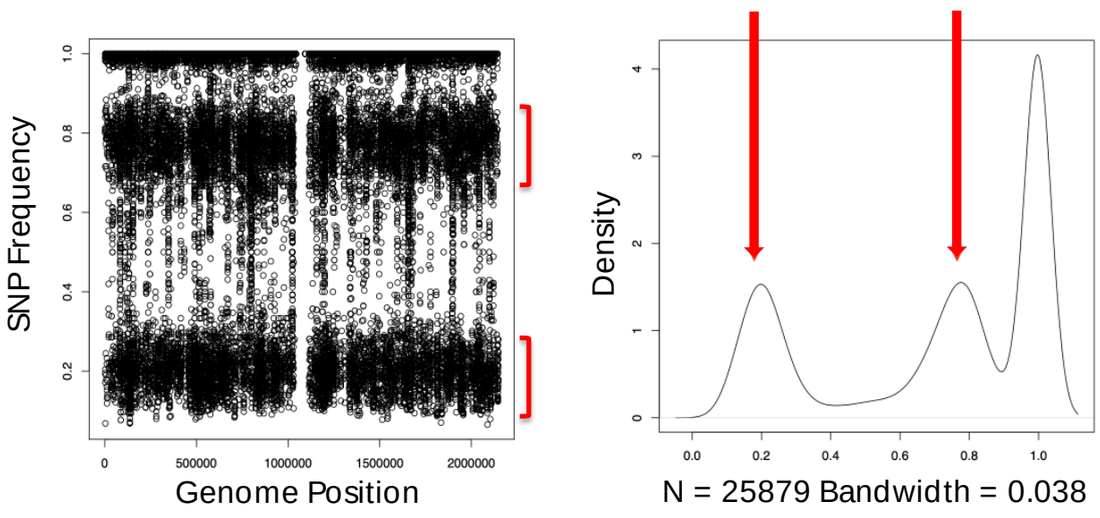

Fig 2. The SNP frequencies across the genome revealed a mixed *S. pneumoniae* population from S07 with 80% (major) and 20% (minor). Reads were parsed based on the frequencies and consensus sequences were generated & analyzed for both strains independently.

RESULTS

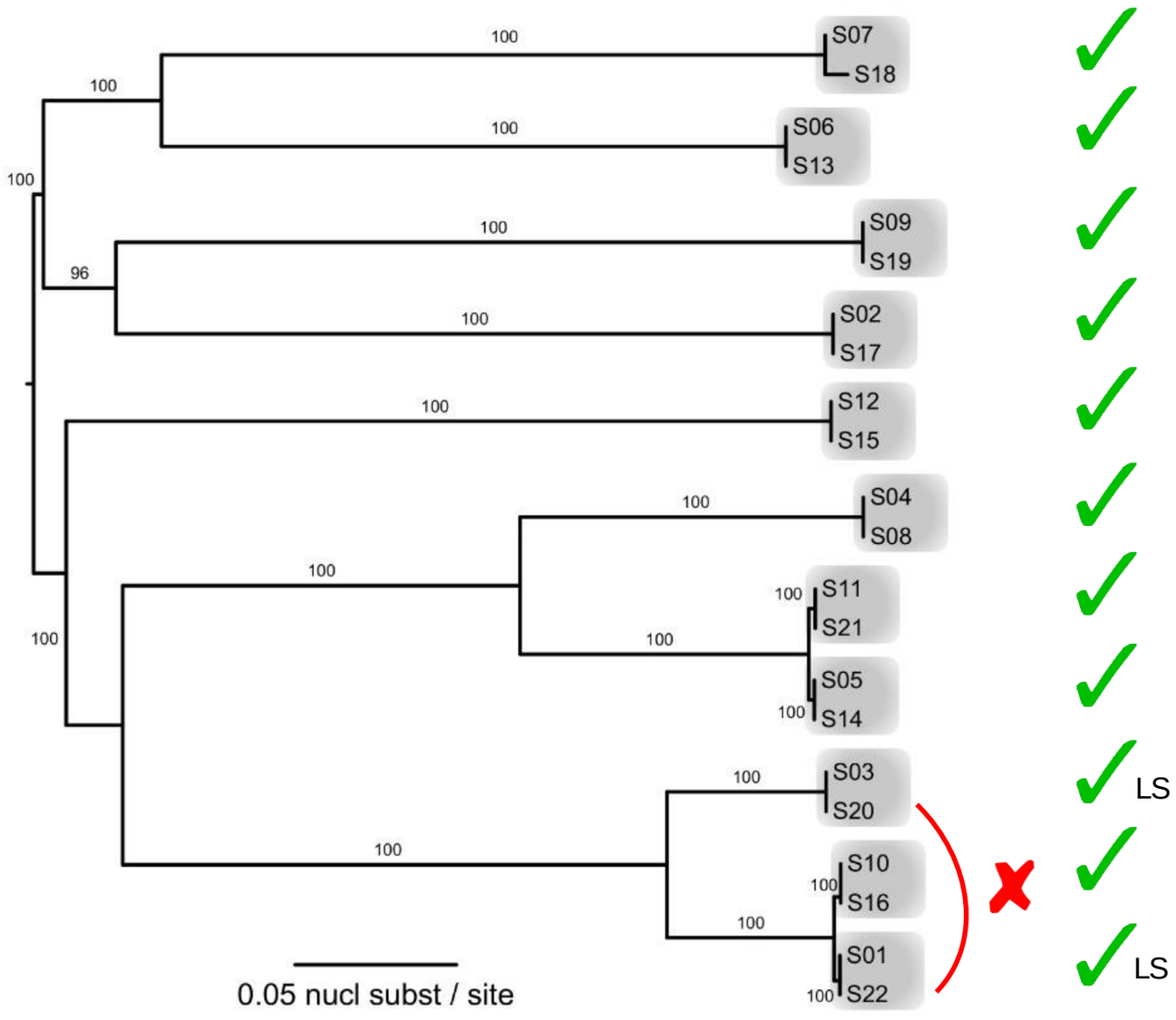

Fig 3. Maximum-likelihood tree revealed genetic linkage amongst pairs based on branch topology and support (≥90%). Longitudinal sets (LS) were included (S03 | S20 and S01 | S22). 9/10 pairs were correctly matched while 1/10 pairs (S20 | S22) was undetected due to the confounding longitudinal samples. All 10 pairs were assessed for transmission direction, 4/10 was correctly inferred, 4/10 were incorrect, and 2/10 were equivocal.

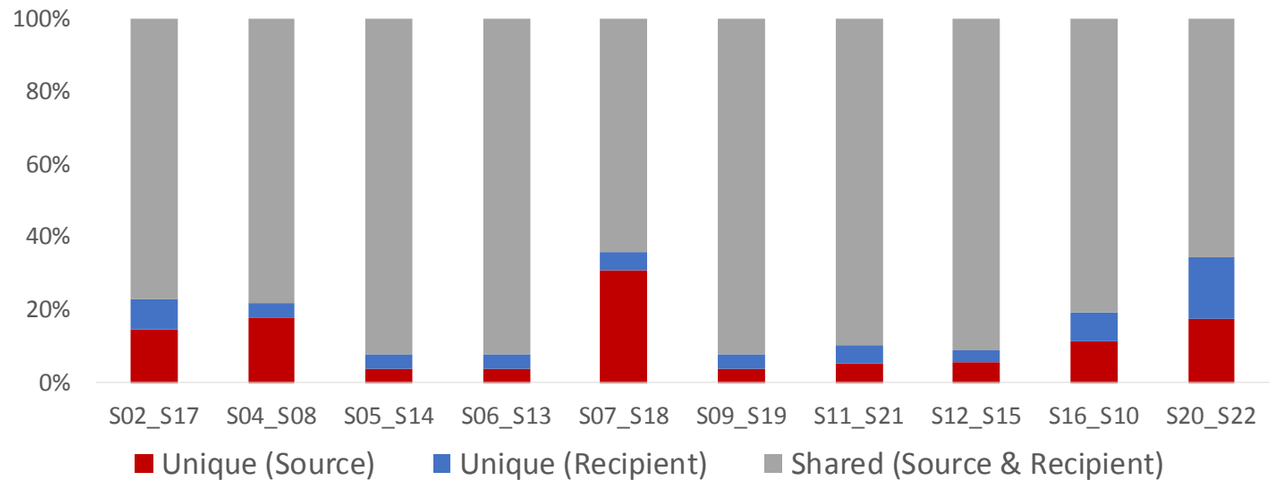

Fig 4. Proportional abundances of unique SNPs detected amongst transmission pairs revealed larger abundance of unique SNPs associated with the source compared to the recipient; this trend was not observed in random within-serotype pairs (data not shown).

SUMMARY

- Mixed within-host populations were detected and variants were parsed & reconstructed for transmission analysis.
- Transmission pairs were correctly identified from the genomic data alone. Longitudinal samples were confounders in detecting pairs.
- Proportional abundances of unique SNPs revealed possible transmission bottlenecks and could be used to identify source recipient relationships.
- Direction of transmission was correctly inferred for 4/10 pairs and increased to 7/10 after unblinding and parameter optimization.
- Further work is required to validate this method.

Incorrect calls could be due to:

- False negative samples/test through transport/culturing/serotyping.
- Potential unsampled intermediary transmission partners.
- Phylogenetic uncertainties & limitations of short read fragments.

[a] Faculty of Epidemiology and Population Health, LSHTM, London, UK
[b] Vaccine Preventable Bacteria Section, PHE, London, UK
[c] Faculty of Infectious and Tropical Diseases, LSTHM, London, UK
[d] Immunisation & Countermeasures Division, PHE, London, UK

[1] Wilm et al. LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012; 40(22):11189-201.
[2] Wymant et al. Phyloscanner: Inferring transmission from within- and between-host pathogen genetic diversity. Mol Biol Evol 2018; 35:719–733.