



Figure S2: Percentage TIGR4 reference coding sequences identified identically within 104 pneumococcal genomes.

Points represent independent pneumococcal isolates. Isolates representing CCs for which $n \geq 3$ are labelled by CC. Isolates representing CCs for which $n < 3$ are labelled as 'Miscellaneous CCs'. Black line represents mean percentage identical coding sequences ± 2 standard deviations (gray lines).