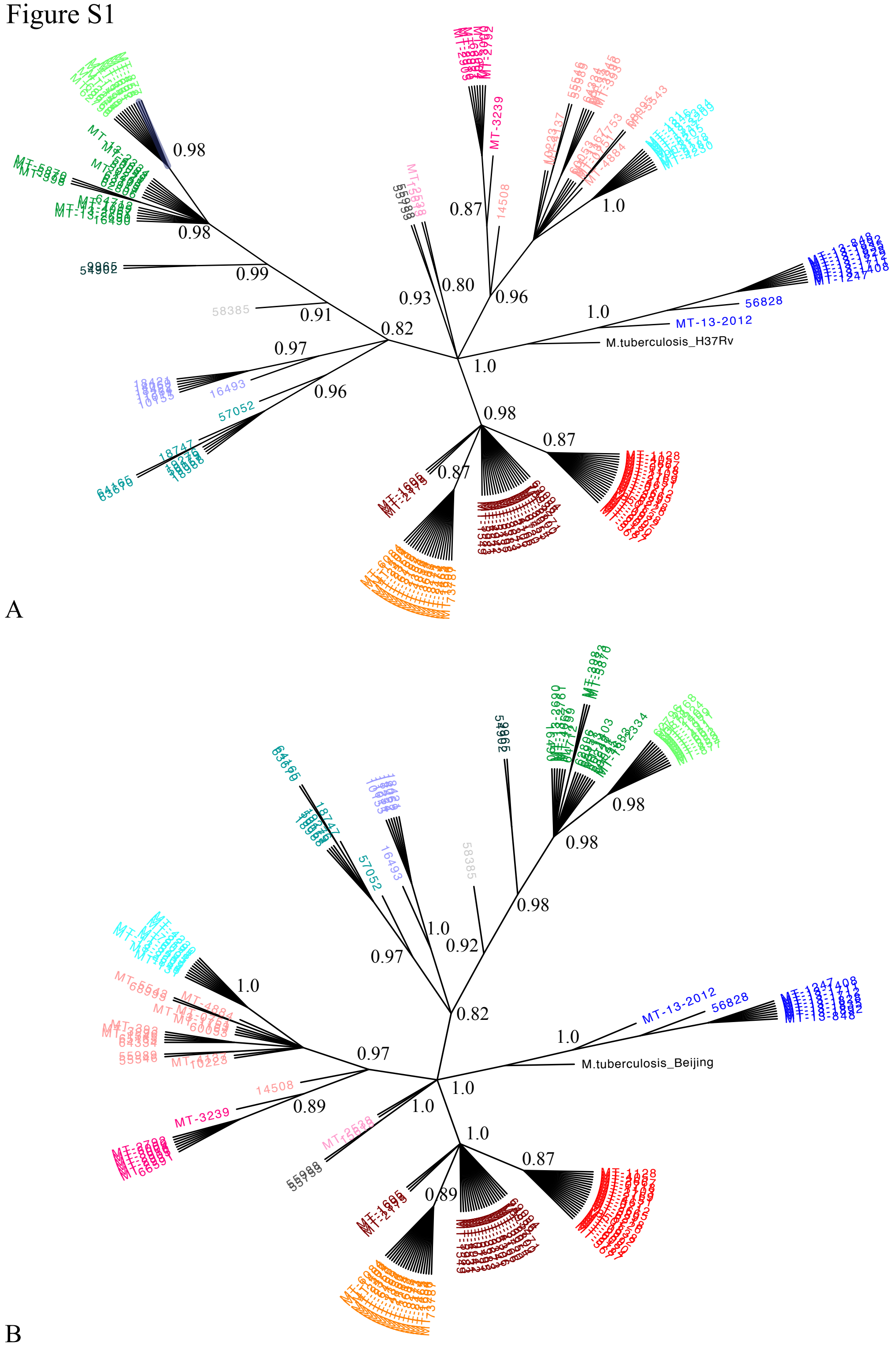
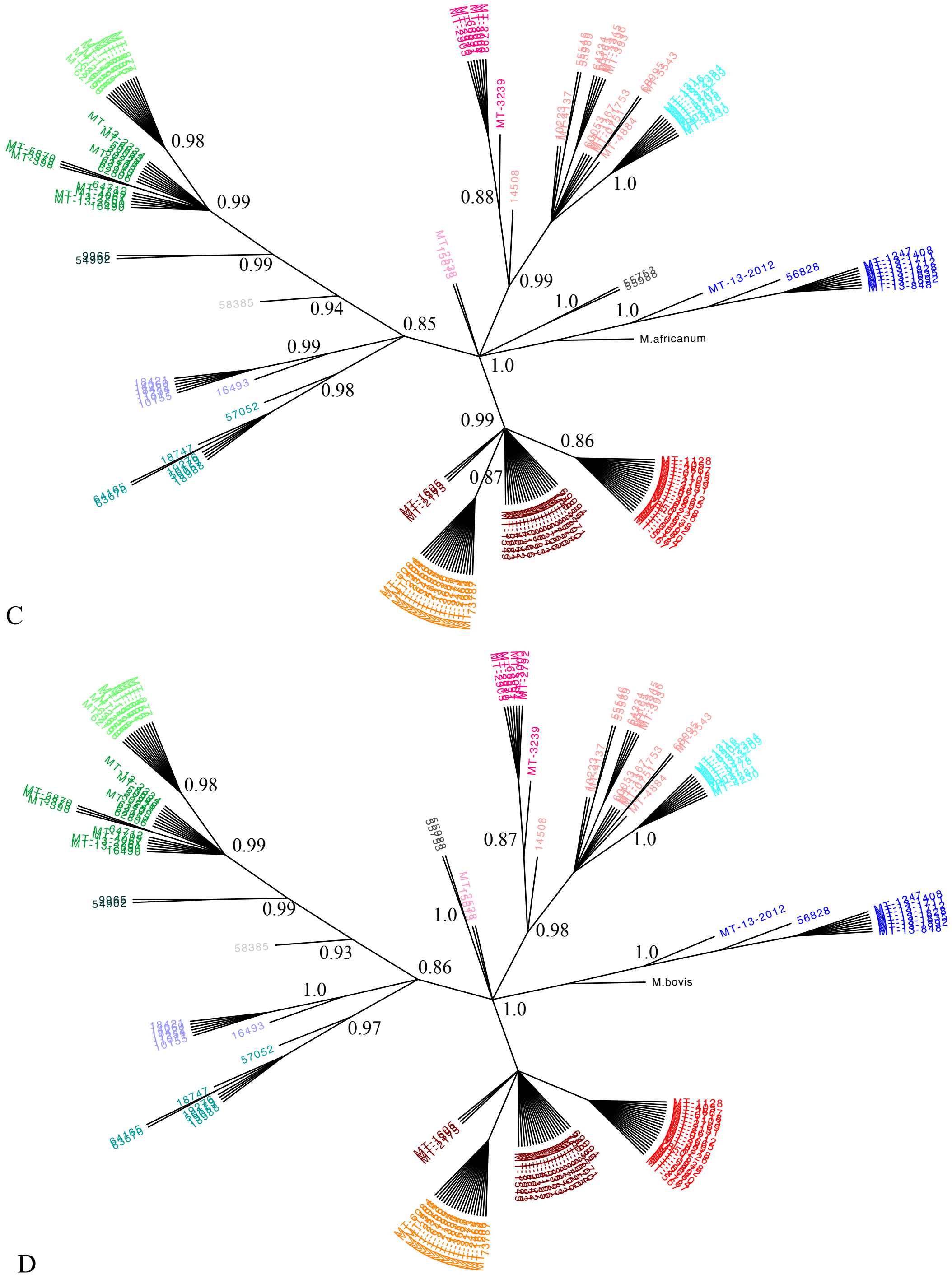
Figure S1



A
M.tuberculosis_H37Rv
MT-13-2012
56828
MT-1247408
MT-3239
14508
MT-2538
55988
58385
16493
57052
18747
MT-4884
MT-1128
MT-1695
MT-2195
0.98
0.98
0.99
0.91
0.82
0.97
0.96
0.93
0.80
0.87
0.96
1.0
1.0
1.0
0.98
0.87
0.87
B
M.tuberculosis_Beijing
MT-13-2012
56828
MT-1247408
MT-13-848
58385
16493
57052
18747
MT-4884
14508
MT-3239
MT-2538
55988
MT-1128
MT-1695
MT-2195
0.98
0.98
0.98
1.0
0.92
0.97
0.82
1.0
1.0
0.97
0.89
1.0
1.0
1.0
0.87
0.89



Legend. Maximum likelihood trees with 1000 bootstrap replicates, with branches below an 80% bootstrap threshold collapsed (branch lengths are therefore not to scale). For clarity, bootstrap p values are indicated only for the most proximal node defining each cluster. Each tree was rooted on its respective reference. Isolates were coloured for their respective clusters identified according to CDC1551 (and H37Rv (1)). Isolates were then kept the same colour across all panels, to facilitate quick comparison between the new reference analysis and CDC1551. See Table S4 for cluster names. A – Reference *M. tuberculosis* Lineage 4 H37Rv, using the Tamura 3-parameter model (2) of nucleotide substitution with 1,405 SNP loci. B – Reference *M. tuberculosis* Lineage 2 CCDC5079, using the General Time Reversible (3) model of nucleotide substitution with 2,048 SNP loci. C – Reference *M. africanum* (Lineage 1), using the Tamura 3-parameter model of nucleotide substitution with 2,721 SNP loci. D – Reference *M. bovis*, using the GTR model of nucleotide substitution with 2,803 SNP loci.