Supplementary Information

Table S1. Good's coverage estimation for concatenated MLSA sequences and individual loci sampled across the 13 sites in which they were found. Values are calculated for unique sequences or for operational taxonomic units defined at a 0.01 nucleotide dissimilarity cutoff.

Phylogroup	**concatenated**		***atpD***		***gyrB***		***recA***		***rpoB***		***trpB***	
	unique	0.01	unique	0.01	unique	0.01	unique	0.01	unique	0.01	unique	0.01
MAN125	0.64	1	1	1	0.87	1	0.91	1	1	1	0.96	1
MAN196	0.62	1	0.98	1	0.91	1	0.98	1	1	1	0.96	1
WA1063	0.47	0.95	0.95	1	0.95	0.95	0.89	0.95	0.79	1	0.84	0.95
MS200	0.3	1	0.89	1	0.67	1	0.67	0.96	0.67	1	0.67	1
MS152	0.18	0.94	0.88	1	0.71	0.94	0.47	0.94	0.82	1	0.41	0.88
F34	0.3	1	0.83	0.97	0.78	1	0.97	1	0.89	1	0.78	0.97

Table S2. Individual MLSA loci have incongruent phylogenetic histories, as indicated by the Shimodaira-Hasegawa test for phylogenetic incongruence. Rows represent maximum likelihood topologies and columns the nucleotide alignments used to evaluate incongruence. Corresponding phylogenic trees were constructed as defined in the methods and are depicted in Figure S3.

	atpD	*gyrB*	*recA*	*rpoB*	*trpB*	concatenated
atpD	--	0	0	0	0	0
gyrB	7.0E-04	--	1.8E-03	4.8E-02	2.3E-03	2.3E-03
recA	0	3.3E-03	--	7.5E-03	0	0
rpoB	0	0	0	--	0	0
trpB	4.0E-04	3.5E-03	1.7E-03	5.8E-02	--	3.3E-03
concatenated	1.6E-01	4.5E-01	3.8E-01	3.4E-01	4.7E-01	--

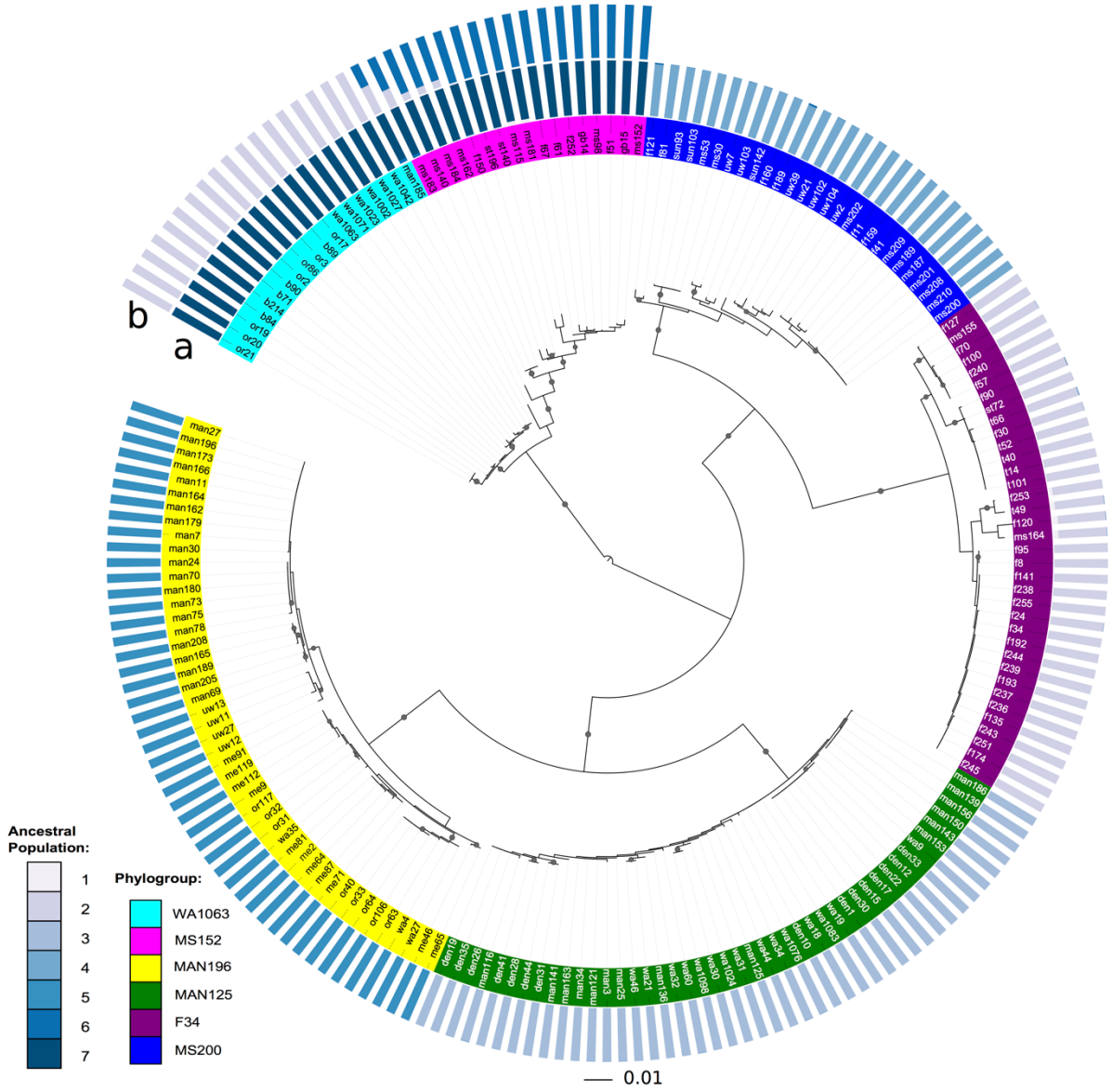


Figure S1. Structure analysis was performed on all sequences, irrespective of phylogroup membership, revealing that the genetic ancestry of each group is consistent with phylogroup boundaries as defined by ANI cutoffs. (This approach differs from that depicted in Figure 1 which shows the results of independent Structure analyses performed on each separate phylogroup.) The colored bars in outer rings indicate genetic contributions from different ancestral populations as indicated in the legend (a). Since WA1063 and MS200 share an ancestral population, we performed an additional Structure analysis that included only sequences from these phylogroups, and this analysis reveals clear subdivision between these phylogroups (b). Taxa names in the inner ring are colored to indicate the phylogroup designation of each strain as defined in the legend. The tree was constructed from concatenated MLSA loci nucleotide sequences using maximum likelihood with a GTRGAMMA evolution model. Scale bar represents nucleotide substitutions per site. The root was defined by *Mycobacterium smegmatis*. Nodes with bootstrap confidences > 80 are indicated with gray circles, and precise bootstrap values are found in Figure S3A.

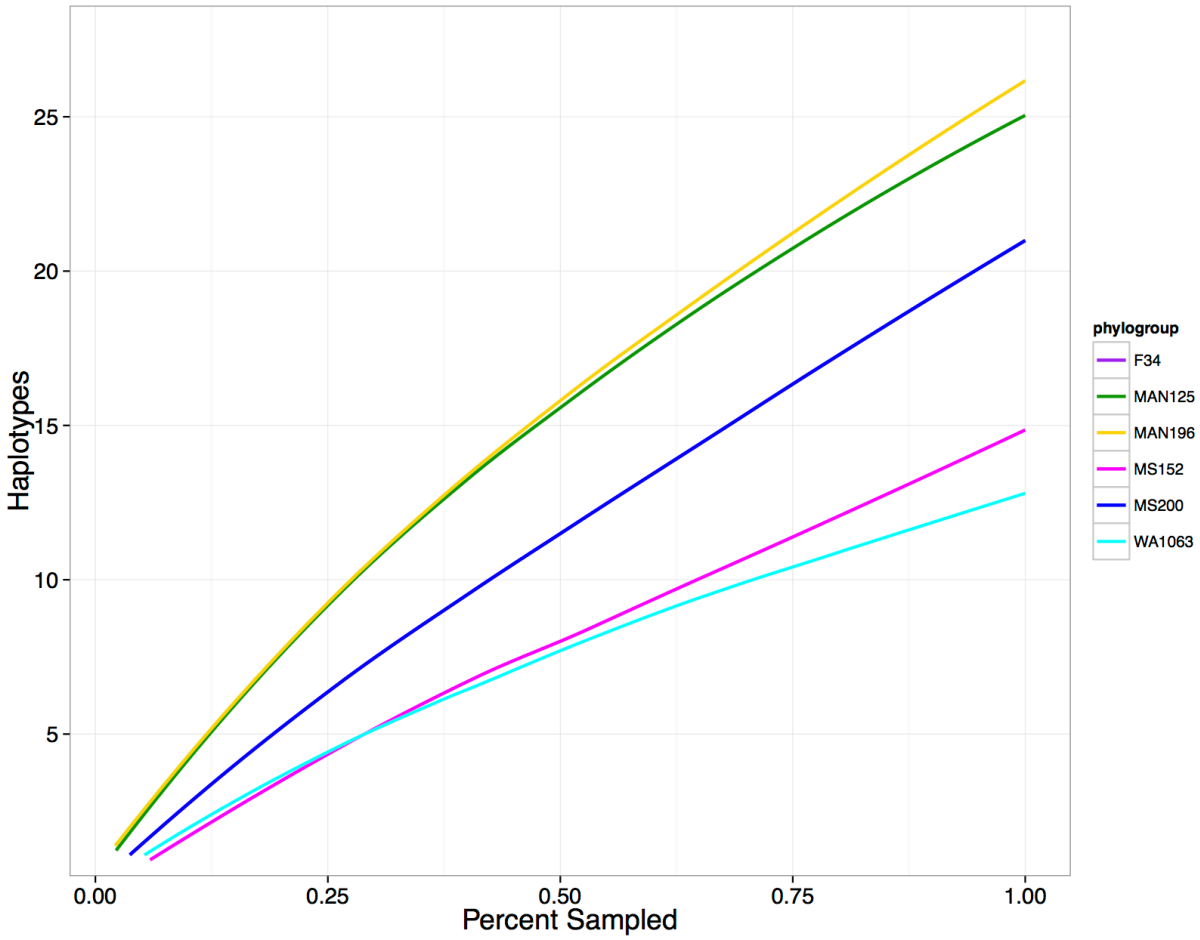
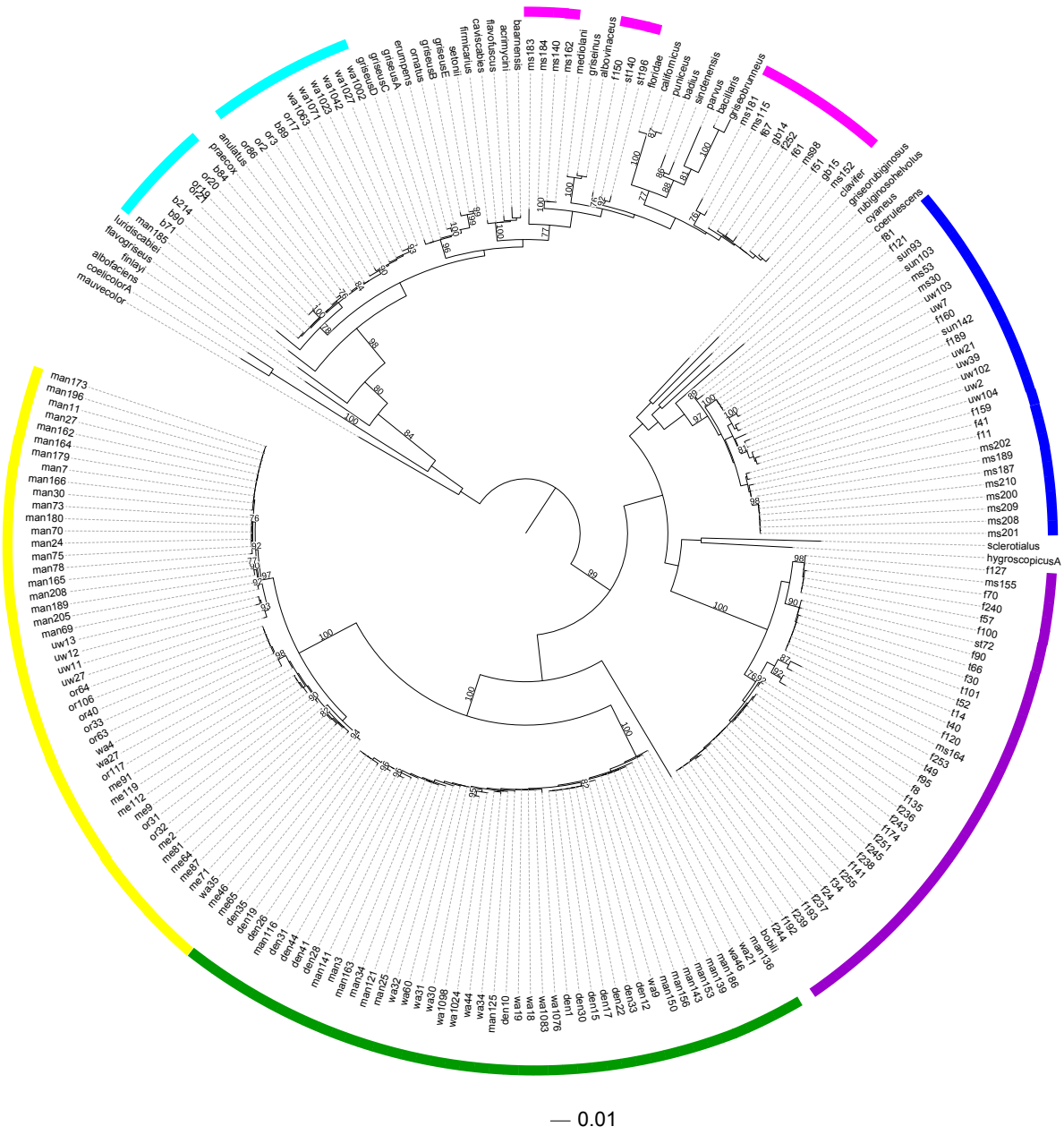
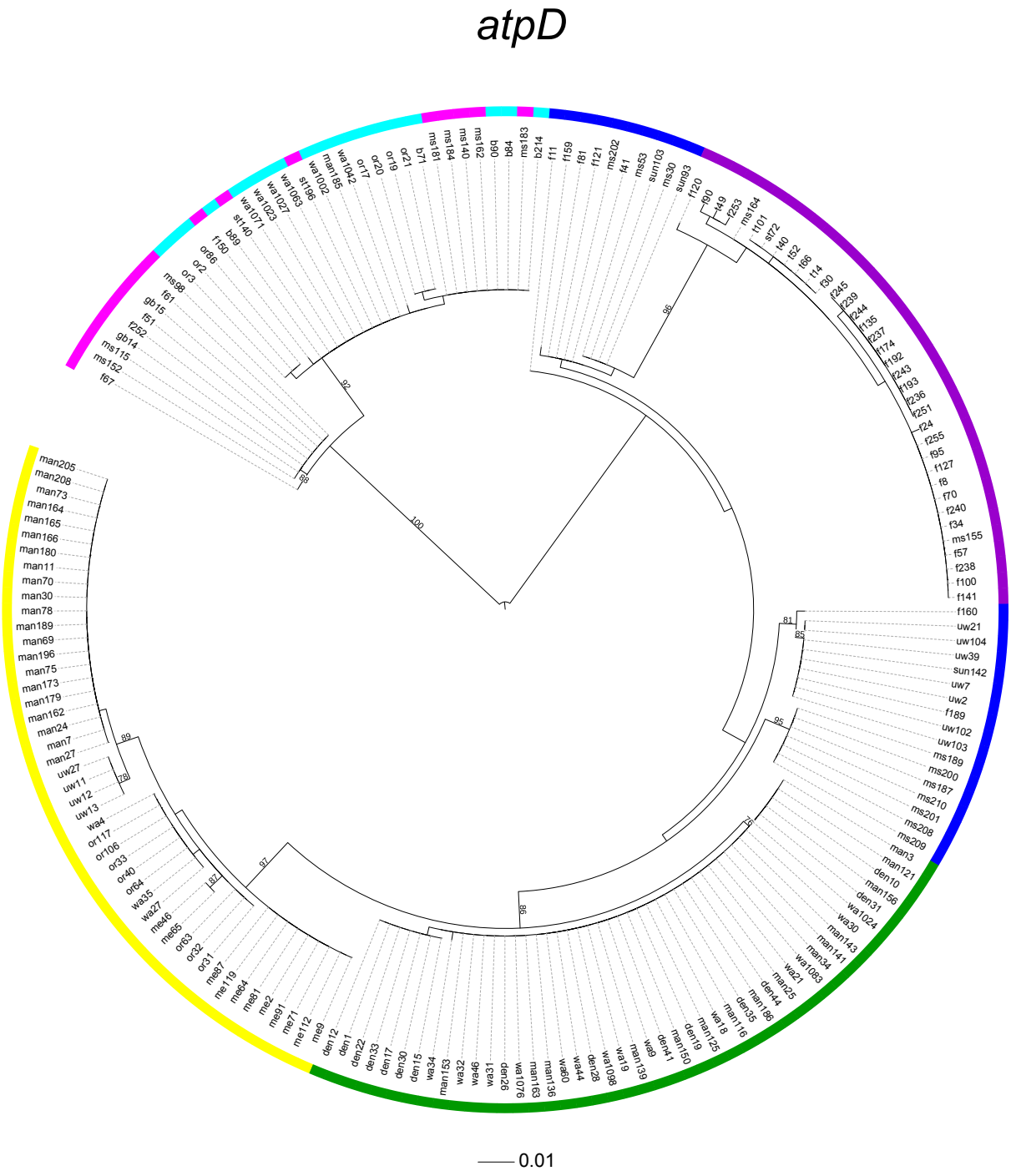


Figure S2. Rarefaction indicates that the recovery of unique MLSA haplotypes for each phylogroup is unsaturated, indicating there is substantial un-sampled haplotype diversity. To normalize sampling efforts between phylogroups, the number of unique MLSA haplotypes is plotted against the percent of total isolates sampled for MLSA (Table 2). The rarefaction curves for phylogroups MS200 and F34 are identical. Good's coverage estimations for each phylogroup can be found in Table S1.

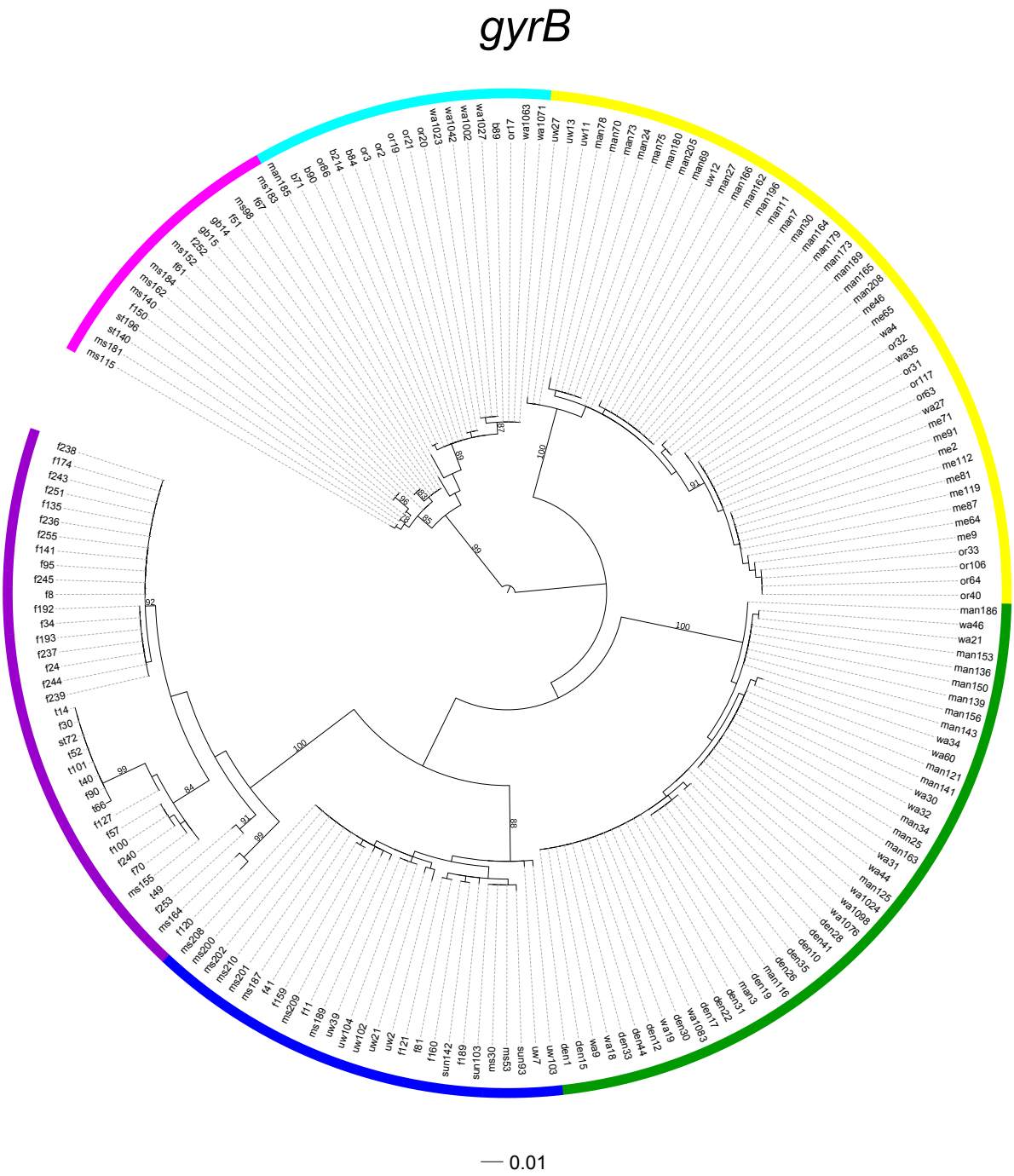
A



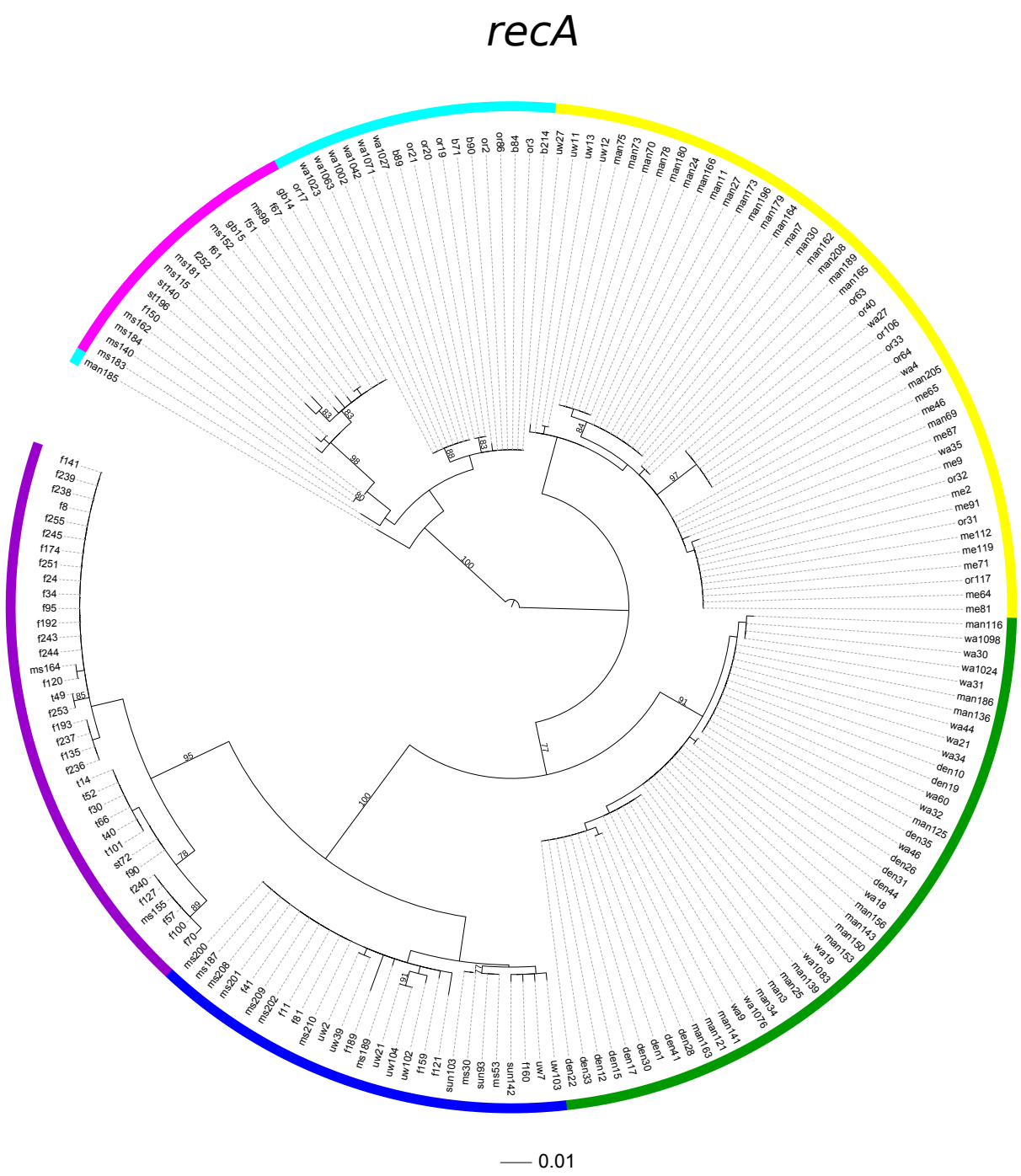
B



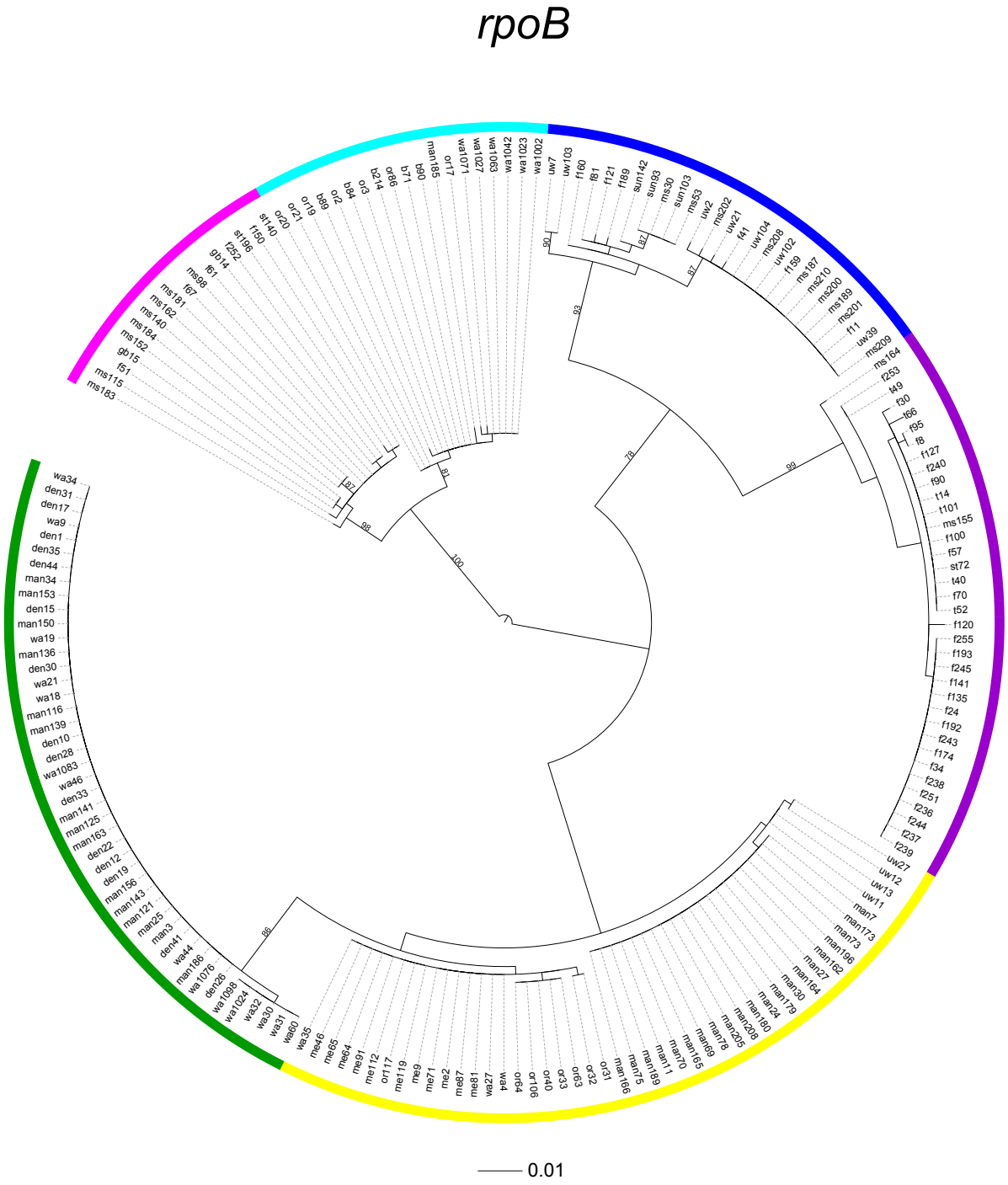
C



D



E


rpoB
wa1002
uw7
uw103
f160
f81
f121
f189
sun142
sun93
ms30
sun103
ms53
uw2
ms202
uw21
f41
uw104
ms208
uw102
f159
ms187
ms210
ms200
ms189
ms201
f11
uw39
ms209
ms164
f253
t49
f30
f66
f95
f8
f127
f240
f90
t14
t101
ms155
f100
f57
st72
t40
f70
t52
f120
f255
f193
f245
f141
f135
f24
f192
f243
f174
f34
f238
f251
f236
f244
f237
f239
uw27
uw12
uw13
uw11
man7
man173
man73
man196
man162
man27
man164
man30
man179
man24
man180
man208
man205
man78
man69
man165
man70
man11
man189
man75
man166
or31
or32
or63
or33
or40
or106
or64
wa4
wa27
me81
me87
me2
me71
me9
me119
or117
me112
me91
me64
me65
me46
wa35
wa60
wa31
wa30
wa832
wa1024
wa1098
den26
wa1076
man186
wa44
den41
man3
man25
man121
man143
man156
den19
den12
den22
man163
man125
man141
den33
wa46
wa1083
den28
den10
man139
man116
wa18
wa21
den30
man136
wa19
man150
den15
man153
man34
den44
den35
den1
wa9
den17
den31
wa34
ms183
ms115
f51
gb15
ms152
ms184
ms140
ms162
ms181
f67
ms98
f61
gb14
f252
st196
st150
or20
or21
or19
b89
or2
b84
or3
b214
or86
b71
b90
man185
or17
wa1071
wa1027
wa1063
wa1042
wa1023
0.01

F

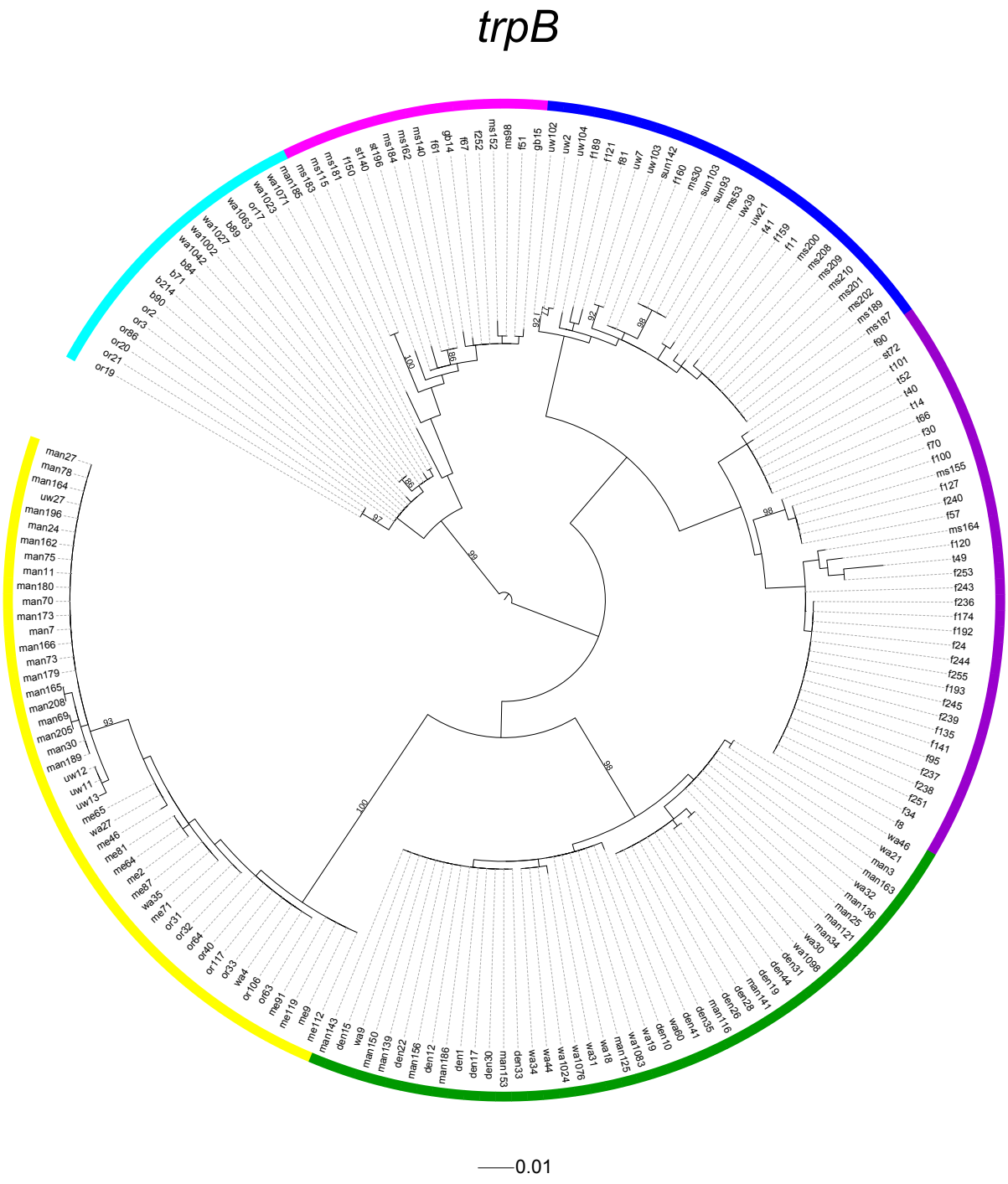


Figure S3. Phylogenetic trees were constructed from nucleotide sequences of concatenated MLSA loci (A), *atpD* (B), *gryB* (C), *recA* (D), *rpoB* (E), *trpB* (F) by maximum likelihood with a GTRGAMMA evolution model. The scale bar represents nucleotide substitutions per site. Branches with bootstrap support values > 75 are labeled. For Figure S3A, representative isolates from PubMLST were included to orient phylogroup taxonomy to existing *Streptomyces* type strains. The root was defined by *Mycobacterium smegmatis* in each case. The outer color strip depicts phylogroup membership consistent with Figures 2 and 3 [green (MAN125), yellow (MAN196), teal (WA1063), blue (MS200), pink (MS152), purple (F34)]. The key for phylogroup isolate names can be found in Table 1.

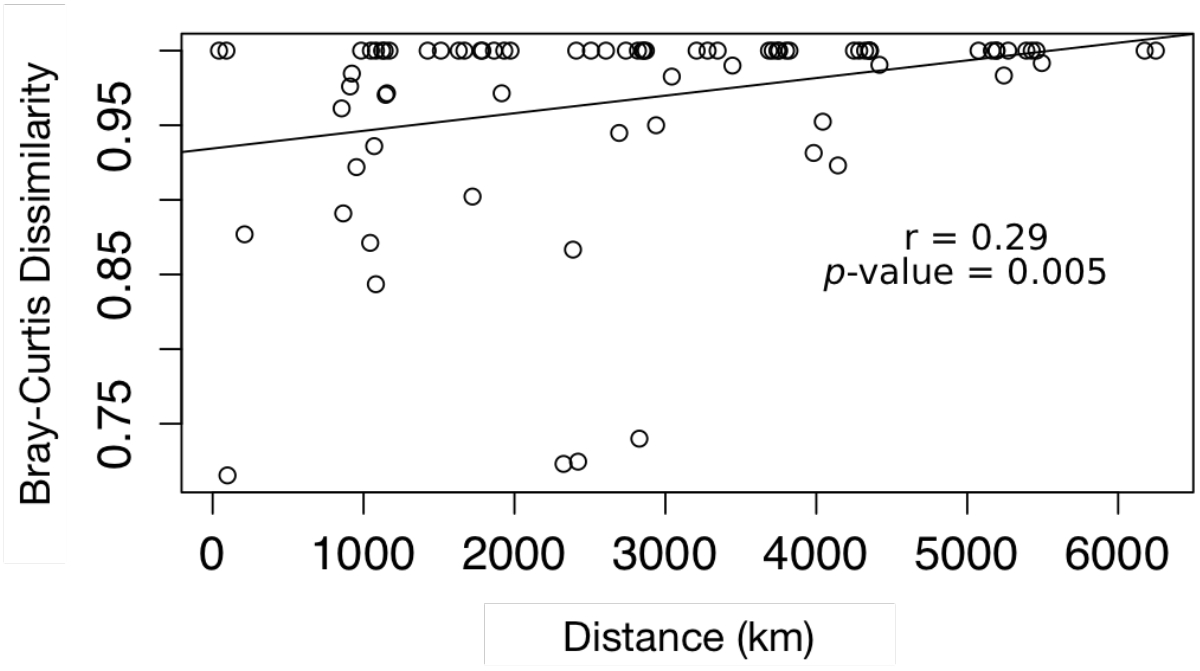


Figure S4. Gene flow limitation results in a significant relationship between geographic distance and allele composition between sites, determined using the Mantel test, such that sites closer together share more alleles. Bray-Curtis dissimilarity was calculated from unique allele composition across sites. The Mantel coefficient (r) and *p*-value under the null hypothesis of no correlation between the matrixes are reported.