Supplementary Information

Table S1. Good's coverage estimation for concatenated MLSA sequences and individual loci. 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. Reported values are *p*-values for 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 S4.

	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	--

Table S3. MLSA PCR conditions and primer sets. PCR amplifications were performed using AmpliTaq Gold reagents (Applied Biosystems, Foster City, CA, USA) and the following 25 µl reaction: 11.75 µl H_2O, 2.5 µl 10X buffer, 3 µl 25 mM $MgCl_2$, 2 µl dNTP mixture (2.5 mM each dNTP, 10 mM total dNTPs, Promega, Madison, WI, USA), 1 µl forward primer from 10 µM stock, 1 µl reverse primer from 10 µM stock, 2.5 µl DMSO, 0.25 µl AmpliTaq Gold (5 U/µl), and 1 µl DNA template. For all MLSA primer sets, the following reaction conditions were used: 95˚C for 10 min; 35 cycles of 95˚C for 20 s, 65˚C for 30 s, 72˚C for 45 s; 72˚C for 10 min. For the 16S locus, reactions conditions were the same expect for 35 cycles of annealing at 55˚C for 90 s.

Locus	Primer	Primer sequence (5'-3')	Amplicon size (bp)	Annealing (˚C)
atpD	*atpD*-F	GTCGGCGACTTCACCAAGGGCAAGGTGTTCAACACC	998	65
	atpD-R	GTGAACTGCTTGGCGACGTGGGTGTTCTGGGACAGGAA		
gyrB	*gryB*-F	CTGGACGCGGTCCGCAAGCG	1,018	65
	gryB-R	GTCTGGCCCTCGAACTGCGGCT		
recA	*recA-F*	CCGCRCTCGCACAGATTGAACGSCAATTC	913	65
	recA-R	GCSAGGTCGGGGTTGTCCTTSAGGAAGTTGCG		
rpoB	*rpoB*-F	GAGCGCATGACCACCCAGGACGTCGAGGC	994	65
	rpoB-R	CCTCGTAGTTGTGACCCTCCCACGGCATGA		
trpB	*trpB*-F	GCGCGAGGACCTGAACCACACCGGCTCACACAAGATCAACA	822	65
	trpB-R	TCGATGGCCGGGATGATGCCCTCGGTGCGCGACAGCAGGC		
16S	8F	AGAGTTTGATCCTGGCTCAG	1,484	55
	1492R	CGGTTACCTTGTTACGACTT		

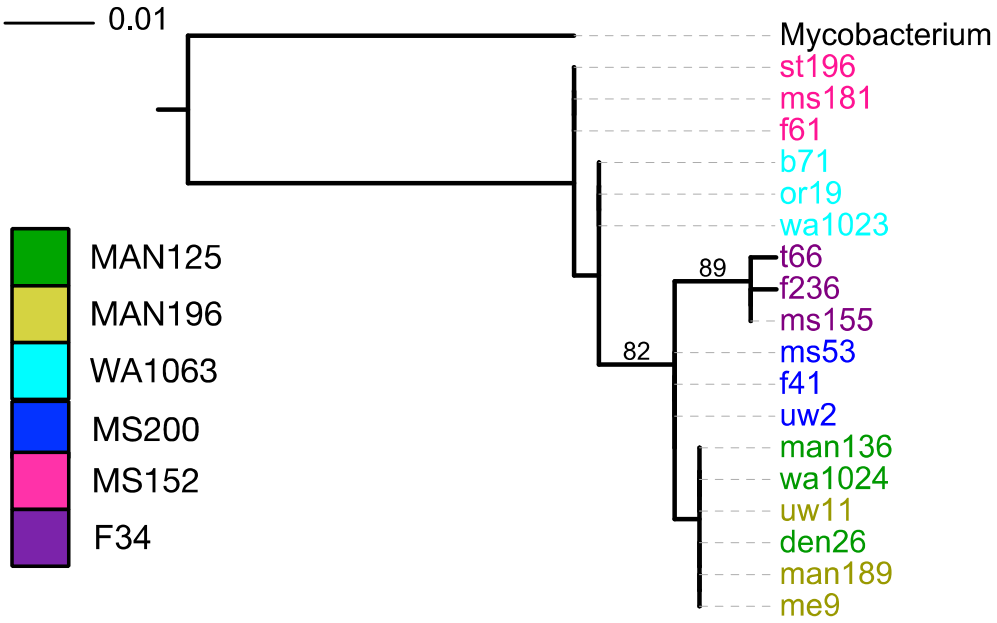


Figure S1. The maximum likelihood tree was constructed with a GTRGAMMA evolution model using partial SSU rRNA gene sequences from three representatives for each phylogroup. Scale bar represents nucleotide substitutions per site. Branches with bootstrap support values greater than 75% are labeled, and the outgroup is *Mycobacterium smegmatis*. Strains are colored to reveal phylogroup membership as indicated in the color legend provided.

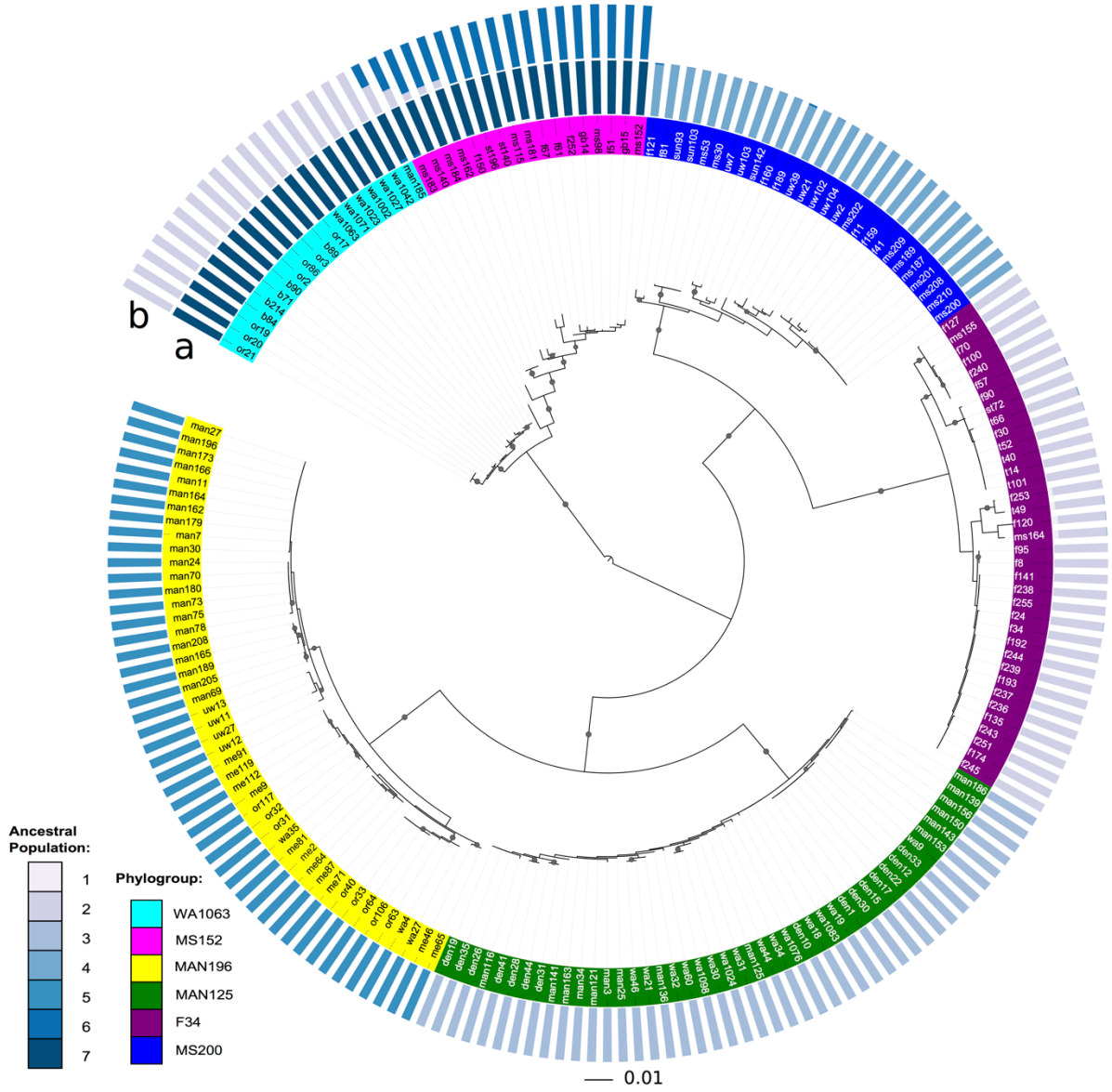


Figure S2. Structure analysis was performed on all sequences, irrespective of phylogroup membership. (This approach differs from that depicted in Figure 1 which shows the results of independent Structure analyses performed on each separate phylogroup.) The colored bars in outer rings indicate genetic 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 S4A.

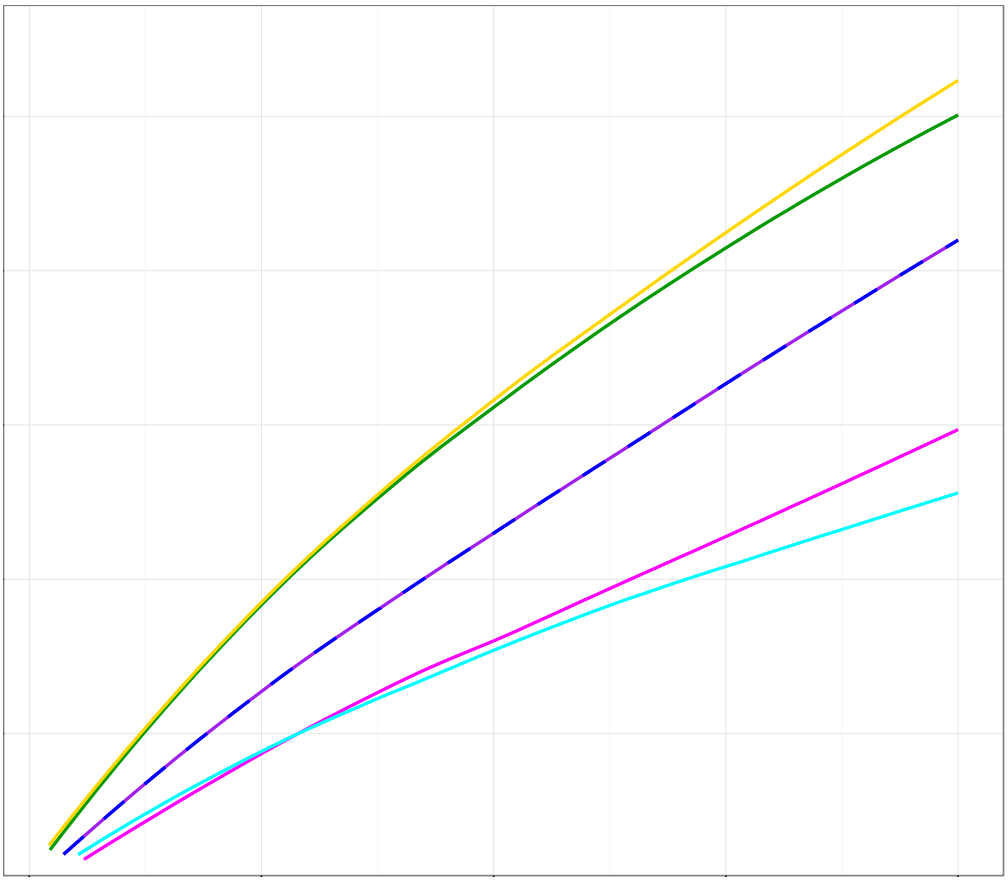
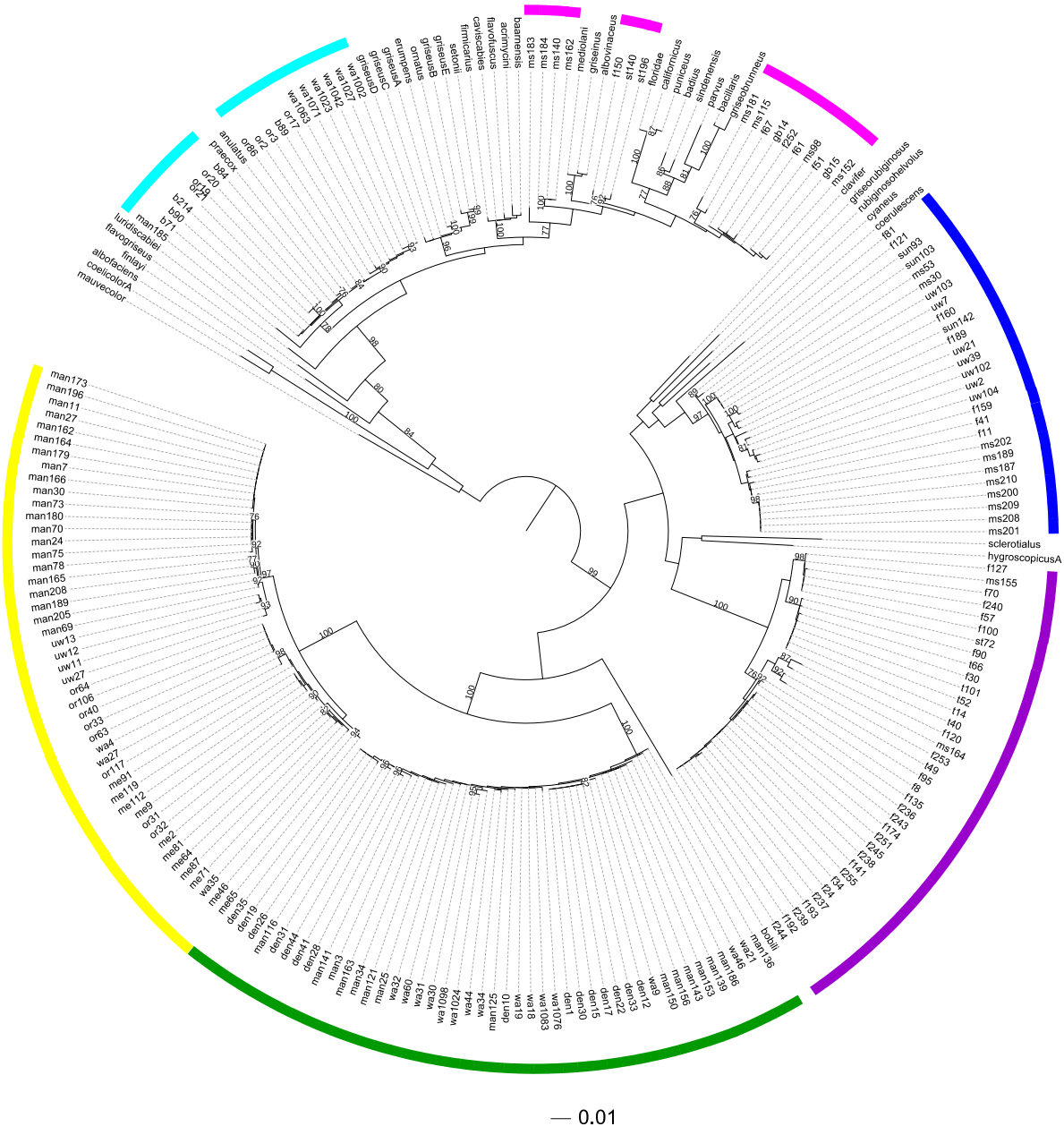
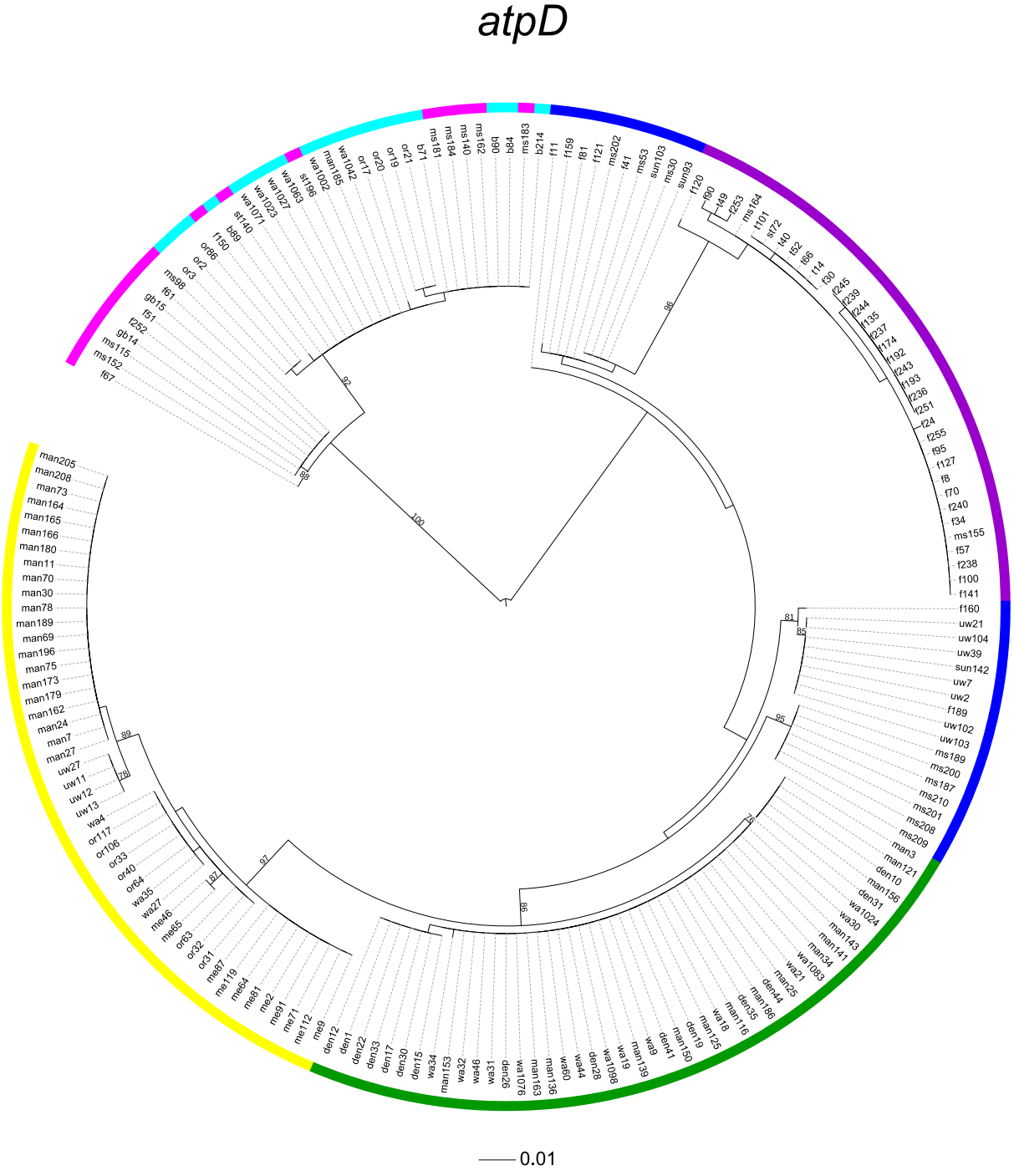


Figure S3. The number of unique MLSA haplotypes is plotted against the percent of total isolates sampled for MLSA (Table 2). The rarefaction curves for phylogroups MS200 and F34 are identical and are depicted with dashed lines. Good's coverage estimations for each phylogroup can be found in Table S1.

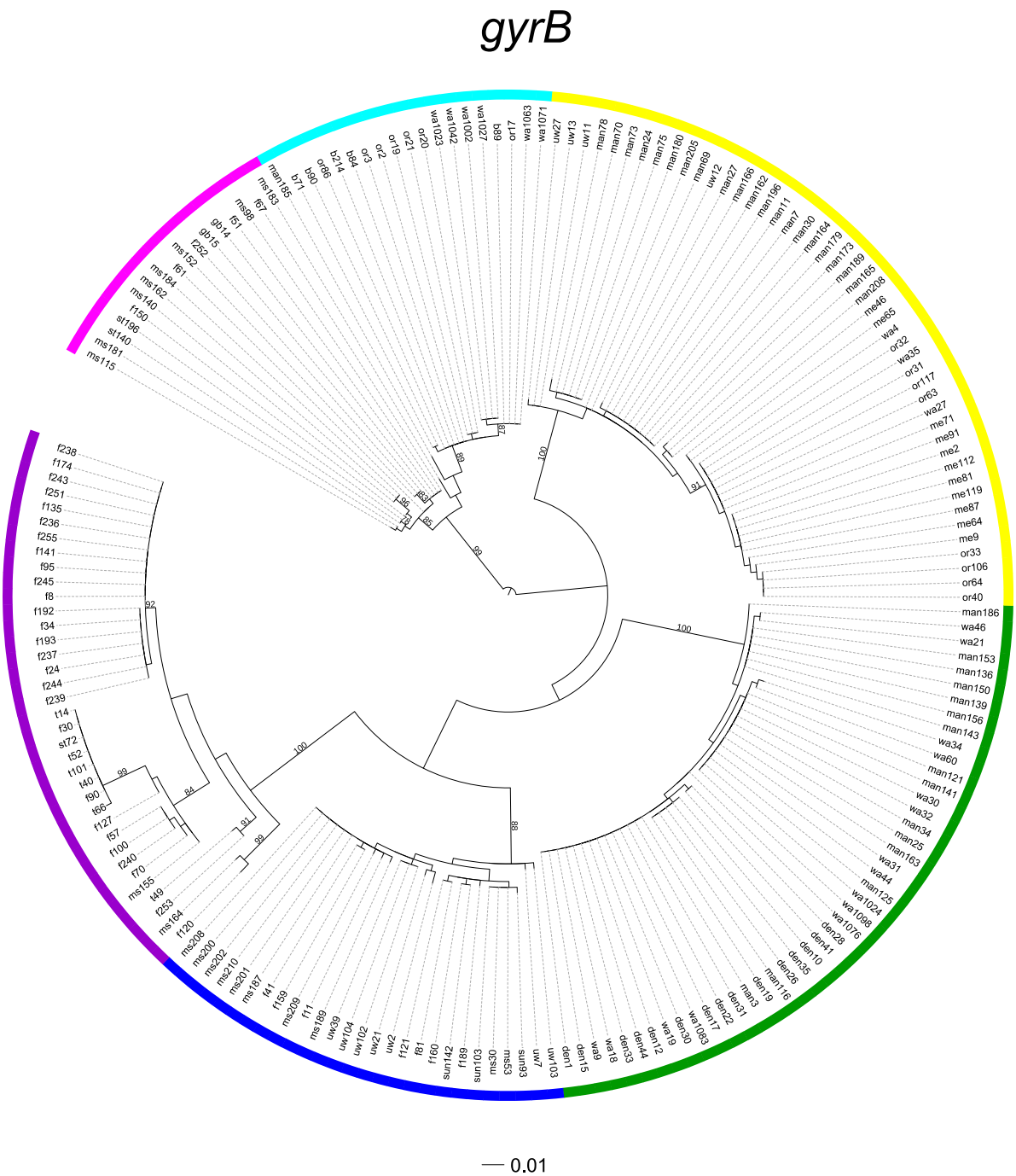
A



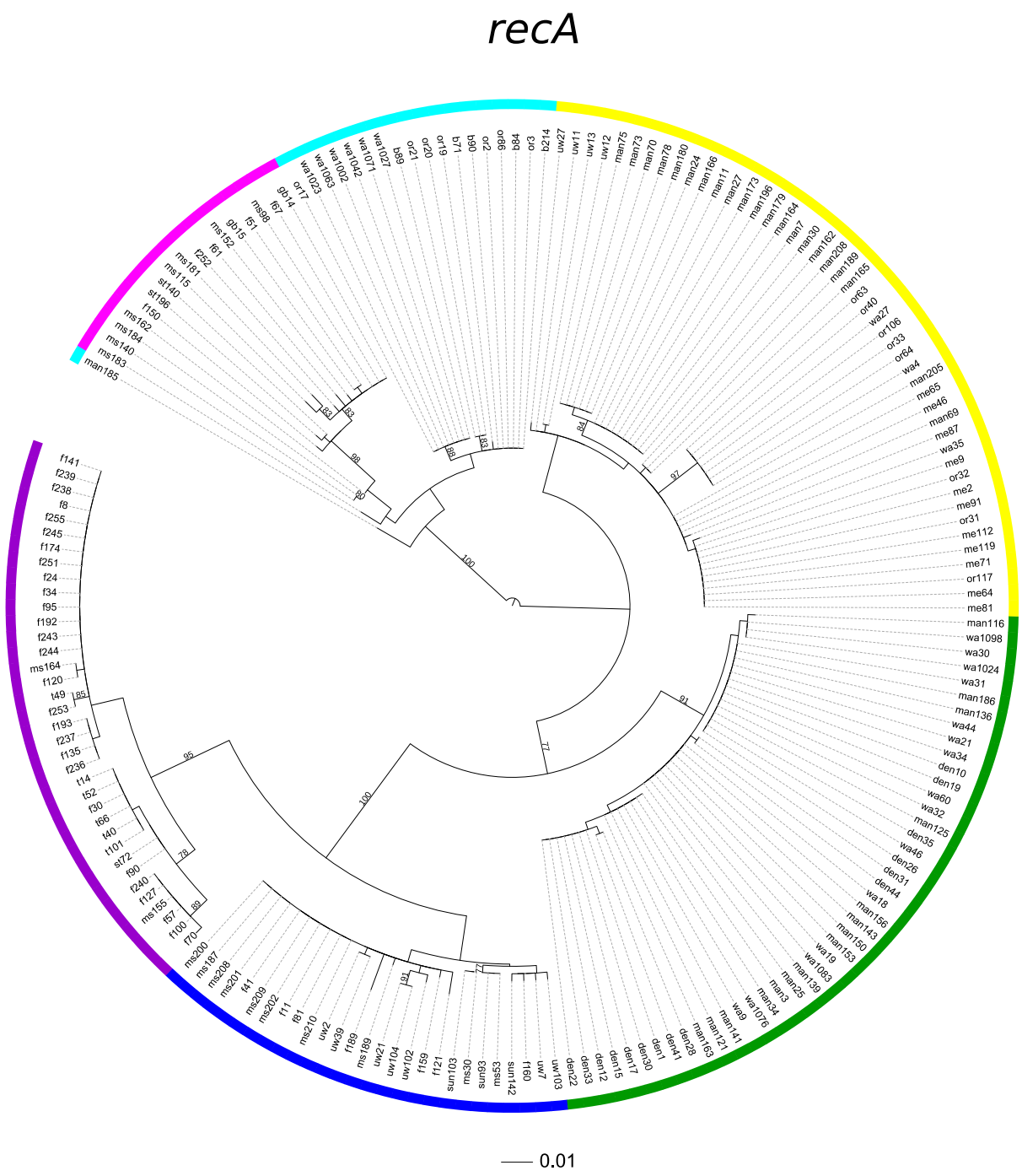
B



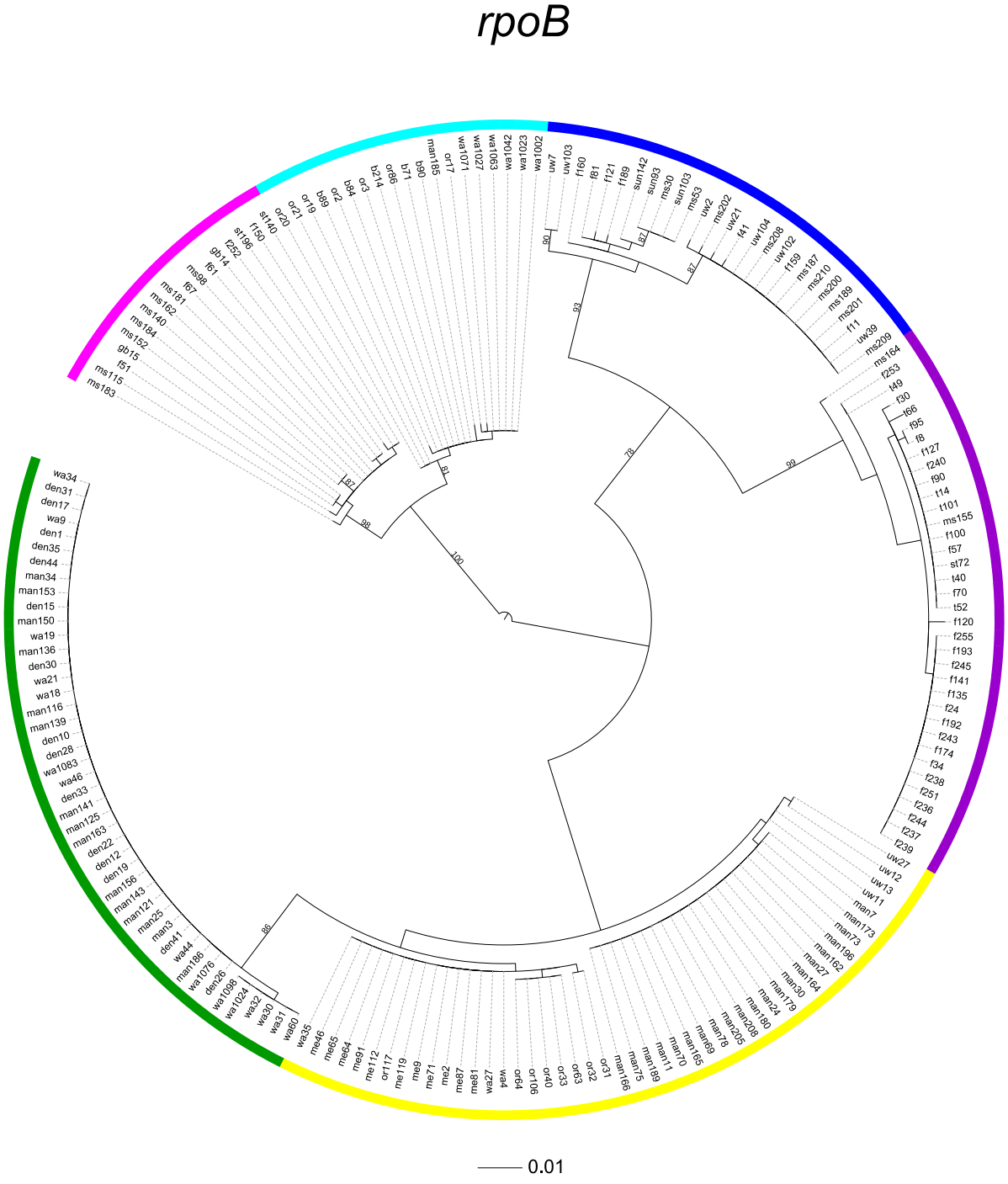
C



D



E



F

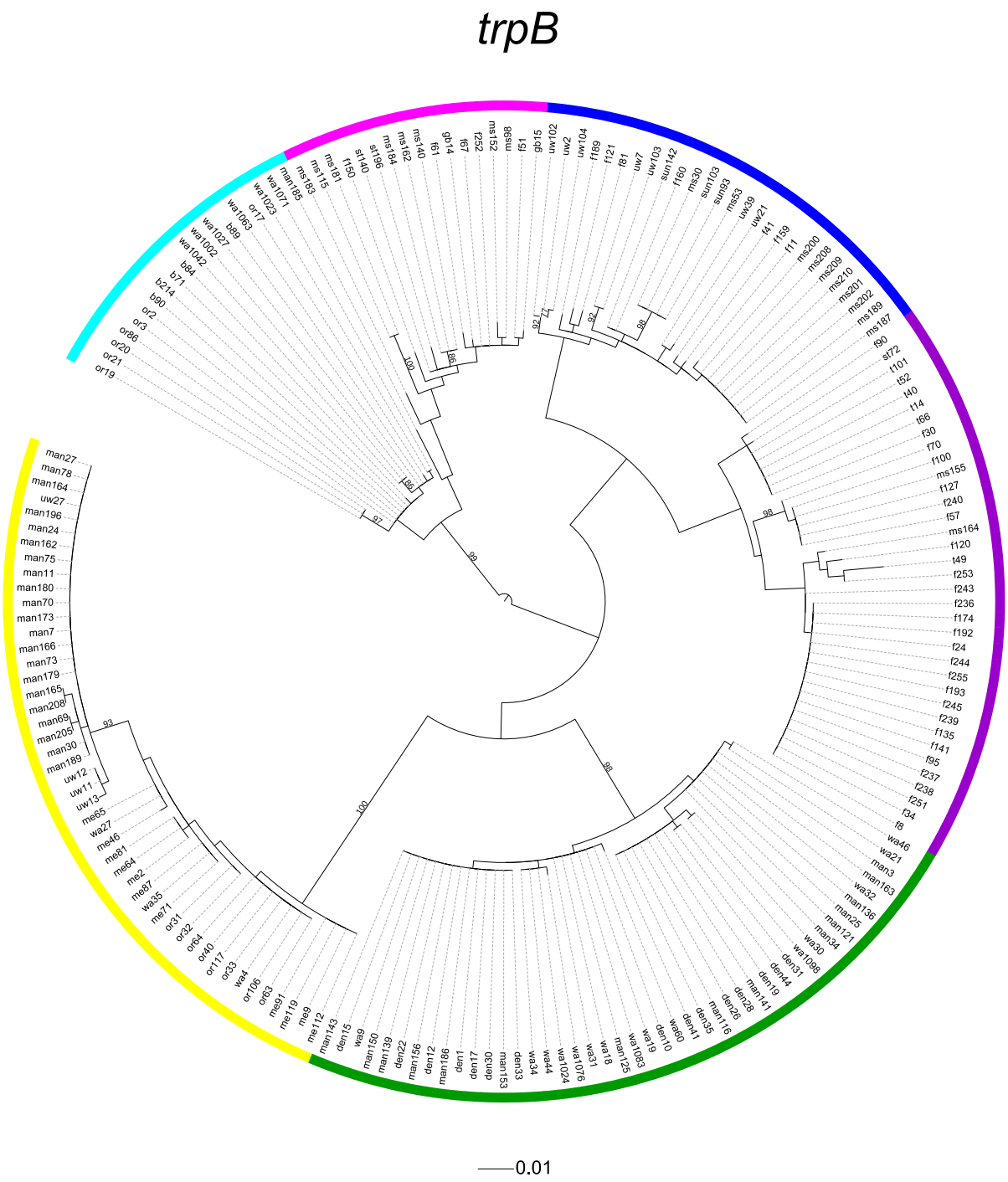


Figure S4. Phylogenetic trees were constructed from nucleotide sequences of concatenated MLSA loci (A), *atpD* (B), *gryB* (C), *recA* (D), *rpoB* (E), *trpB* (F) by maximum likelihood with a GTRGAMMA evolution model. The scale bar represents nucleotide substitutions per site. Branches with bootstrap support values > 75 are labeled. For Figure S4A, 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 Figure 2 [green (MAN125), yellow (MAN196), teal (WA1063), blue (MS200), pink (MS152), purple (F34)]. The key for phylogroup isolate names can be found in Table 1.

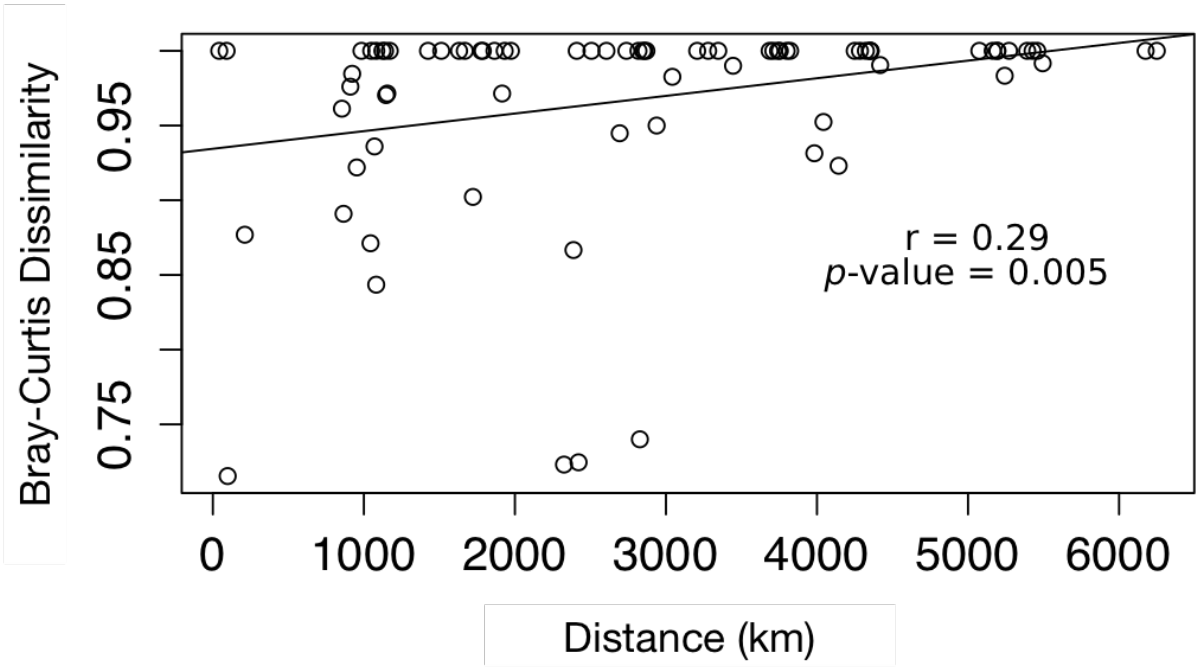


Figure S5. Circles represent pairwise comparisons of Bray-Curtis dissimilarity and geographic distance between sites for all phylogroups. Bray-Curtis dissimilarity was calculated from allele composition across sites. The Mantel coefficient (r) and *p*-value under the null hypothesis of no correlation between the matrixes are reported.