TABLE S1 Average nucleotide identity (ANI) divergence between reference genomes [a]

		Reference genomes						
		Mycobacterium tuberculosis, lineage 4 H37Rv	*Mycobacterium tuberculosis,* lineage 4 CDC1551	*Mycobacterium tuberculosis,* lineage 2 CCDC5079	*Mycobacterium africanum* GN041182	*Mycobacterium bovis* AF2122/97	*Mycobacterium canettii* CIPT 140010059	*Mycobacterium kansasii* [b] ATCC 12478
Mycobacterium tuberculosis, lineage 4	H37Rv	-	0.10 (99.37)	0.12 (99.17)	0.18 (98.82)	0.19 (98.71)	0.85 (95.95)	19.01 (54.40)
Mycobacterium tuberculosis, lineage 4	CDC1551	0.11 (99.94)	-	0.16 (99.04)	0.18 (98.91)	0.21 (98.93)	0.85 (96.06)	19.02 (54.36)
Mycobacterium tuberculosis, lineage 2	CCDC5079	0.12 (99.68)	0.15 (99.83)	-	0.22 (98.71)	0.23 (98.35)	0.88 (96.06)	19.13 (54.36)
Mycobacterium africanum	GN041182	0.14 (99.27)	0.16 (99.62)	0.18 (99.34)	-	0.15 (99.02)	0.85 (95.98)	19.02 (54.39)
Mycobacterium bovis	AF2122/97	0.16 (99.18)	0.18 (99.44)	0.20 (99.00)	0.14 (99.61)	-	0.86 (95.69)	19.00 (54.36)
Mycobacterium canettii	CIPT 140010059	0.75 (96.94)	0.75 (97.07)	0.77 (96.95)	0.75 (97.09)	0.76 (96.67)	-	19.02 (54.39)
Mycobacterium kansasii	ATCC 12478	13.69 (16.25)	13.67 (16.10)	13.69 (16.09)	13.66 (16.13)	13.66 (16.19)	13.66 (16.26)	-

[a] Percent divergence is indicated, with the average percentage of nucleotides used in each analysis in brackets. Each genome was, in turn, fragmented into consecutive 500 base-pair segments and queried against all other complete reference genomes using jSpecies (v.1.2.1, (4)). Both BLASTn (5) and MUMmer (6) algorithms were applied. BLASTn settings were as in (7), requiring ≥70% identity over ≥70% of the alignment. Default settings were used for MUMmer. For each algorithm, the mean ANI was calculated (e.g., the mean of the pair A_{500} vs $B_{COMPLETE}$ and B_{500} vs $A_{COMPLETE}$ (8)). This was then used to tabulate ANI divergence (100% - the mean of pairwise ANI, (7)). ANI divergences calculated using BLASTn are indicated in red, while those using MUMmer are indicated in blue. In brackets, the average percent of total nucleotides in the query genome that were used in the calculation has been indicated for each comparison.

[b] pMK plasmid sequence not used for alignment.