

Supplementary data

Table S1. PCR primers used to confirm mutations detected by comparative genome hybridization

Gene	FW primer	RV primer	Start position	End position	Length (bp)
MT0208	ATTCTGCGTCTGCTCGCGGGGT	GGTTGGCGGTTCGATCCGCACTCC	234602	234775	174
MT0516	GAAGGTCTCATCTTGCGCAAAC	TTACCGGAGCTCTCTGTTTCGG	588602	588780	169
MT1253	GAAGCTACCGCAGACCAGCATC	CTACCGCGACTTGGCCTTTCCC	1358101	1358302	200
Δ MT1928 - 1976	GAGGTGGTTCGTCTTGAACAGG	CTGAGGTATTGGGCGCGGCAGG	2127002	2176282	245

Primers are depicted in the 5'-3' direction, the start and end positions refer to the beginning and end of the PCR product and length refers to the length of the PCR product. Gene ID and start and end positions are based on the annotated genome sequence of strain CDC1551 (GenBank AE000516).

Table S2. Genes that are (partially) deleted in strain RB14H5

Gene	Name/function
MT1928	hypothetical protein (partially deleted)
MT1929	P450 haem-thiolate protein
MT1930	lipoprotein antigen
MT1931	short chain dehydrogenase
MT1931.1	hypothetical protein
MT1932	hypothetical protein
MT1933	chorismate mutase
MT1934	putative esterase/antigen 85-B
MT1935	hypothetical protein
MT1936	hypothetical protein
MT1937	hypothetical protein
MT1940	hypothetical protein
MT1941	hypothetical protein
MT1942	hypothetical protein
MT1943	hypothetical protein
MT1944	hypothetical protein
MT1945	hypothetical protein
MT1946	zinc-binding dehydrogenase
MT1947	hypothetical protein
MT1948	D-tyrosyl-tRNA deacylase
MT1949	hypothetical protein
MT1950	hypothetical protein
MT1951	hydrolase alpha/beta family
MT1952	competence damage inducible protein A
MT1953	sugar transporter family
MT1954	hypothetical protein
MT1955	anti-anti-sigma factor
MT1956	D-amino acid oxidase
MT1957	hypothetical protein
MT1958	hypothetical protein
MT1959	catalase peroxidase (<i>katG</i>)
MT1960	ferric uptake regulator protein
MT1961	phosphatidylethanolamine-binding protein
MT1962	phosphatidylethanolamine-binding protein
MT1963	quinone oxidoreductase
MT1964	hypothetical protein
MT1965	hypothetical protein
MT1966	isocitrate lyase
MT1968	PPE family protein
MT1969	PPE family protein
MT1970	hypothetical protein
MT1971	acyltransferase family protein
MT1972	lipoprotein
MT1973	peptidase putative
MT1974	esterase putative
MT1975	hypothetical protein
MT1976	acyl-CoA synthase (partially deleted)

Gene ID and name/function of deleted genes are based on the annotated genome sequence of strain CDC1551 (GenBank AE000516).

Figure S1. Diagram from comparative genome hybridization data depicting the 49 kbp deletion. The deleted fragment spans 47 genes (see Table S2), including *katG*, coding for the only known catalase(-peroxidase) present in the *M. tuberculosis* genome.

