

## Supplemental material

Supplemental Table S1 - Details of all primer sequences and their regions used in this study

LSP			Ref
<b>RD</b>	<b>239</b>		(1)
	RD size	842	
	<b>Lineage 1</b>		
	Flanking region L	ACGGCCGCAT	
	Flanking region R	GCTACCGTTG	(1)
	Primer sequence L	GCCAAAGCCTGACCAGCATC	
	Primer sequence int	CCGTAGGTGATCTCGACCTC	
	Primer sequence R	GGATTCATCGACGCAATCTGACC	
	Tm	59	
	PCR product size	79+100 (absent*) / 238 (present**)	
<b>RD</b>	<b>105</b>		
	RD size	3466	
	<b>Lineage 2</b>		
	Flanking region L	CGTGGTCGCC	
	Flanking region R	GTGGGACGGG	
	Primer sequence L	ACAGCGCGGGTCATATCAC	
	Primer sequence int	TGAATCGACTTTGTTGCCGG	
	Primer sequence R	CACCGTCGTTTTCTGCCGA	
	Tm	59	
	PCR product size	185+126 (absent) / 229 (present)	
<b>RD</b>	<b>750</b>		
	RD size	790	
	<b>Lineage 3</b>		
	Flanking region L	GACGGAGTGC	
	Flanking region R	ATTGGTACTT	
	Primer sequence L	TGGCGGCCAAGATGACATTT	
	Primer sequence int	GGAGACGACGGTGTGTTTTTC	
	Primer sequence R	AAAACCTTTCGGCGGTCAGTG	
	Tm	61	

	PCR product size	106+126 (absent) / 323 (present)	
<b>Pks 15/1</b>			
<b>Lineage 4</b>	RD size	7	
	Flanking region L	GCCGCGGCC	
	Flanking region R	CCGGTGCTTT	
	Primer sequence L	CTGGGTTGGCCTGCACGTGGGCCATAA	
	Primer sequence int	-	
	Primer sequence R	GCCCCGCAGAGGCGCCGGTT	
	Tm	62	
	PCR product size	144 (absent) / 144+7 (present)	
<b>RD 181</b>			
<b>Sublineage 2</b>	RD size	710	
	Flanking region L	GGGACCAAGC	(1)
	Flanking region R	AGGTCGTACG	(1)
	Primer sequence L	CAGCACAAATCCGCCATAC	
	Primer sequence int	GTCATTGGTTGTGCGGCA	
	Primer sequence R	CGATCTGCAGGTTGGTCTTG	
	Tm	59	
	PCR product size	92+151 (absent) / 180 (present)	
<b>RD 150</b>			
<b>Sublineage 2</b>	RD size	2486	
	Flanking region L	ACGTTTCGTAT	(1)
	Flanking region R	CTAGCGCGTC	(1)
	Primer sequence L	CGGTTGGTACTTACGGTGTCT	
	Primer sequence int	TTAGCTCCTCCGCTGAGAC	
	Primer sequence R	CGAGCGGCTTGATGAACTC	
	Tm	59	
	PCR product size	57+86 (absent) / 202 (present)	
<b>RD 142</b>			
<b>Sublineage 2</b>	RD size	2850	
	Flanking region L	CATCGGCGTG	(1)

Flanking region R	GCCGTGCTGC	(1)
Primer sequence L	CACGACCCGGTAAGTGCG	
Primer sequence int	TGTGCCTATTGACGGCCTTA	
Primer sequence R	TAGCACCAGTACCGGATGTC	
Tm	59	
PCR product size	84+60 (absent) / 195 (present)	

---

\* absent = RD absent, which means identification of the specific lineage; \*\* present = RD present, which means that the specific lineage is not identified.

Reference:

1. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere Y-OL, Aman K, Kato-Maeda M, Small PM. 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci U S A* 101:4865–4870.

## Supplemental Figure S1 – Study flow

Study flow of samples analyzed in the current study. Samples were tested with a PCR based method to identify the presence or absence of RD239 (lineage 1), RD105 (lineage 2), RD750 (lineage 3) and the presence or absence of RD181, RD150 or RD142 (subsets of lineage 2). Lineage 4 was genotyped by sequencing of obtained PCR products and we could successfully determine the presence or absence of pks 15/1 in 46 cases (bold). Data from lineage 1, 2 and 3 are inconclusive since these data are not confirmed by sequencing. \*Missing results due to poor DNA quality; \*\*missing results due to limited amount of DNA or PCR not performed when PCR products of at least three different regions showed degraded DNA.

