Supplement for "Revised interpretation of the Hain Lifescience GenoType MTBC to differentiate *Mycobacterium canettii* and members of the *Mycobacterium tuberculosis* complex"

Methods

894 genomes were included in this study (851 genomes previously used for Brites et al.[1] with data from several earlier studies[2-15]; 21 previously published *M. canettii* genomes from Blouin et al.[16] and Tagliani et al.[17]; 14 BCG variants from Abdallah et al.[18]; three *M. caprae* genomes from Orloski et al.[19]; three unpublished *M. canettii* genomes from the Research Center Borstel[20]; one *M. canettii* genome from the New York State Department of Health; and finally an unpublished *M. bovis* genome from the Friedrich Loeffler Institute). The accession numbers for all genomes can be found in Table S1. Table S3 provides an analysis of the pyrazinamide resistance genes (i.e. *panD*, *pncA*, and *rpsA*) of the 26 *M. canettii* genomes.

Raw genomic reads were processed as described in Menardo et al.[21] Briefly, the reads were trimmed with Trimmomatic v0.33.[22] Only reads larger than 20 bp were kept for the downstream analysis. The software SeqPrep (https://github.com/jstjohn/SeqPrep) was used to identify and merge any overlapping paired-end reads. The resulting reads were aligned to the reconstructed ancestral sequence of the *M. tuberculosis* complex using the mem algorithm of BWA v0.7.13.[2,23] Duplicated reads were marked using the MarkDuplicates module of Picard v2.9.1 (https://github.com/broadinstitute/picard). The RealignerTargetCreator and IndelRealigner modules of GATK v 3.4.0 were used to perform local realignment of reads around InDels.[24] Finally, SNPs were called with Samtools v1.2 mpileup[25] and VarScan v2.4.1[26] using the following thresholds: minimum mapping quality of 20, minimum base quality at a position of 20, minimum read depth at a position of 7X, maximum strand bias for a position 90%. SNPs were annotated using snpEff v4.1144, using the *M. tuberculosis* H37Rv reference annotation (NC_000962.3).[27] A custom python script was used to type *in-silico* all genomes of the dataset, with the SNP/deletion markers from the assay (Figure S2 and Table S2).

A maximum likelihood phylogeny was inferred with RAxML (v.8.2.8) using an alignment containing only polymorphic sites and excluding the variable positions in drug resistance-related genes. The phylogeny was inferred using the general time-reversible model of sequence evolution and *Mycobacterium canettii* (SRR011186) was used as an outgroup to root the phylogeny (Figure S1).

Figure S1. MTBC phylogeny.

Maximum likelihood phylogeny of 571 of the 894 genomes included in this study (i.e. redundant genomes were removed[21]). The branch lengths are proportional to nucleotide substitutions and the tree is rooted with *Mycobacterium canettii* SRR011186. The node supports correspond to bootstrap values. Genomes with pattern 6 in Figure 1 (i.e. four *M. bovis* genomes without the *pncA* His57Asp mutation, which confers intrinsic resistance to pyrazinamide[28], and three *M. caprae* genomes) are highlighted in light pink. The BCG genomes (banding pattern 8 in Figure 1) are indicated in brown. The RD1^{BCG} deletion (Figure S2) and five *gyrB* mutations (Table S2) interrogated by the Hain Lifescience GenoType MTBC are indicated with a yellow triangle and yellow stars, respectively (e.g. band 8 develops if the Ala403Ser mutation is not present, whereas the reverse is the case for band 10). Notably, mutation Tyr144Tyr is homoplasic (i.e. it is present in all *M. microti* genomes (pattern 4 in Figure 1) as well as some *M. canettii* genomes, which are not included in this tree (pattern 1 in Figure 1)). The previously described *gyrB* Ala329Ala marker (1113 G>A at position 6109) could be added to the Hain assay to differentiate *M. orygis* from the remaining genotypes.[29] The *lepA* V242V marker for *M. caprae* is indicated by a red star.[30] RD1^{mic}, which is discussed in Figure S2, is shown by a red triangle.

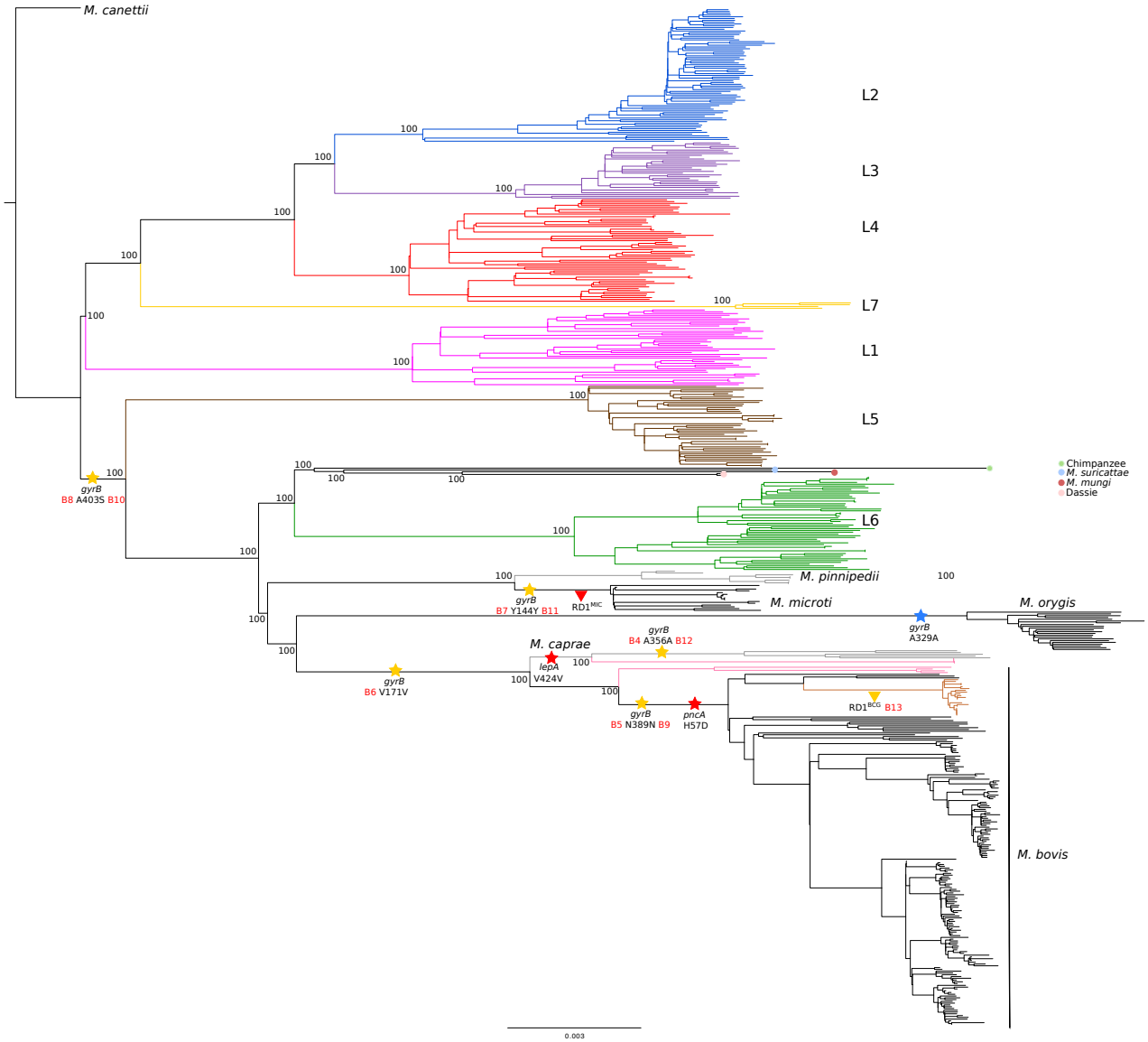

Figure S2. Amplicon and probe design for RD1^{BCG} deletion.

Diagrams adapted from Brodin et al. showing the differences in the region comprising *Rv3863* and *Rv3881c* between the deletion in BCG RD1^{BCG} (A) and *M. microti* RD1^{mic} (C) relative to *M. tuberculosis* H37Rv (B).[31] BCG is not detected by the lack of binding of a probe that falls within the RD1^{BCG} deletion, which is shared by all BCG strains. Instead, the probe in question is just upstream of that deletion and is thus also present in *M. bovis*. However, because the corresponding primers are located upstream and downstream of RD1^{BCG}, they are only sufficiently close to enable exponential amplification when the deletion is present (i.e. 52 bp vs. 9508 bp). This design also ensures that the fact that the probe and forward primer are deleted in *M. microti* is irrelevant.

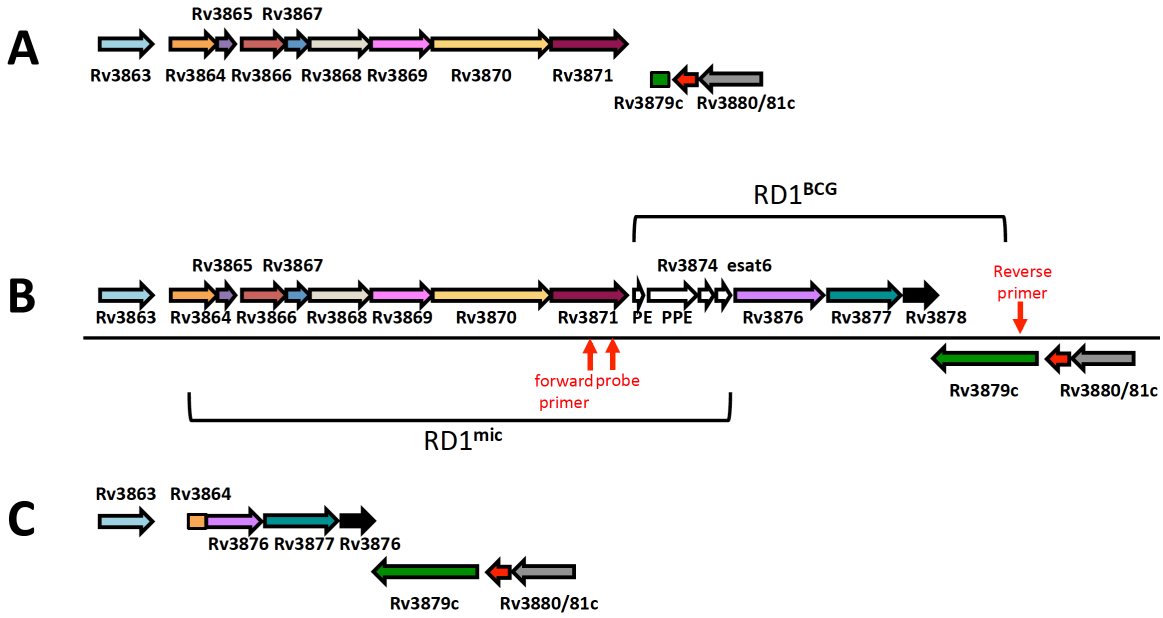

Table S2. Probe design.

Band number	Probe ID in patent	Probe sequences	Position in Huard et al.[29]	Genomic position	Mutation
3	21	GGAGTTCTGGGGCTG	NA	NA	NA
4	15	CTGGCCGCTGTGATCTC	1311	6307	wild-type
5	17	GTCTGTAACGAACAGCT	1410	6406	wild-type
6	13	CAGAACCGCACCGTTG	756	5752	wild-type
7	11	ACGGGTACGAGTGGTC	675	5671	wild-type
8	19	GACTTTCGCGTCGGTG	1450	6446	wild-type
9	18	GTCTGTAATGAACAGCT	1410 C>T	6406	Asn389Asn
10	20	GACTTTCGAGTCGGTG	1450 G>T	6446	Ala403Ser
11	12	ACGGGTATGAGTGGTC	675 C>T	5671	Tyr144Tyr
12	16	CTGGCCGCGGTGATCTC	1311 T>G	6307	Ala356Ala
13	10	CGTGGTGGAGCGGA	NA	NA	NA

References

1. Brites, D. *et al.* A new phylogenetic framework for the animal-adapted *Mycobacterium tuberculosis* complex. *Front Microbiol* **9**, 2820 (2018).
2. Comas, I. *et al.* Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* **45**, 1176-82 (2013).
3. Coscolla, M. *et al.* Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. *Emerg Infect Dis* **19**, 969-76 (2013).
4. Walker, T.M. *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* **13**, 137-46 (2013).
5. Bos, K.I. *et al.* Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494-7 (2014).
6. Casali, N. *et al.* Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* **46**, 279-86 (2014).
7. Dippenaar, A. *et al.* Whole genome sequence analysis of *Mycobacterium suricattae*. *Tuberculosis (Edinb)* **95**, 682-8 (2015).
8. Feuerriegel, S. *et al.* PhyResSE: web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. *J Clin Microbiol* **53**, 1908-14 (2015).
9. Guerra-Assunção, J.A. *et al.* Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *Elife* **4**(2015).
10. Walker, T.M. *et al.* Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis* **15**, 1193-202 (2015).
11. Alexander, K.A., Larsen, M.H., Robbe-Austerman, S., Stuber, T.P. & Camp, P.M. Draft genome sequence of the *Mycobacterium tuberculosis* complex pathogen *M. mungi*, identified in a banded mongoose (*Mungos mungo*) in Northern Botswana. *Genome Announc* **4**(2016).
12. Trewby, H. *et al.* Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis. *Epidemics* **14**, 26-35 (2016).
13. Crispell, J. *et al.* Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. *BMC Genomics* **18**, 180 (2017).
14. Malm, S. *et al.* New *Mycobacterium tuberculosis* complex sublineage, Brazzaville, Congo. *Emerg Infect Dis* **23**, 423-429 (2017).
15. Marcos, L.A. *et al. Mycobacterium orygis* lymphadenitis in New York, USA. *Emerg Infect Dis* **23**, 1749-1751 (2017).

16. Blouin, Y. *et al.* Progenitor "*Mycobacterium canettii*" clone responsible for lymph node tuberculosis epidemic, Djibouti. *Emerg Infect Dis* **20**, 21-8 (2014).
17. Tagliani, E. *et al.* Culture and next-generation sequencing-based drug susceptibility testing unveil high levels of drug-resistant-TB in Djibouti: results from the first national survey. *Sci Rep* **7**, 17672 (2017).
18. Abdallah, A.M. *et al.* Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations. *Sci Rep* **5**, 15443 (2015).
19. Orloski, K., Robbe-Austerman, S., Stuber, T., Hench, B. & Schoenbaum, M. Whole genome sequencing of *Mycobacterium bovis* isolated from livestock in the United States, 1989-2018. *Front Vet Sci* **5**, 253 (2018).
20. Feuerriegel, S., Köser, C.U., Richter, E. & Niemann, S. *Mycobacterium canettii* is intrinsically resistant to both pyrazinamide and pyrazinoic acid. *J Antimicrob Chemother* **68**, 1439-40 (2013).
21. Menardo, F. *et al.* Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics* **19**, 164 (2018).
22. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).
23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
24. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
25. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-93 (2011).
26. Koboldt, D.C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568-76 (2012).
27. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
28. Scorpio, A. & Zhang, Y. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat Med* **2**, 662-7 (1996).

29. Huard, R.C. *et al.* Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. *J Bacteriol* **188**, 4271-87 (2006).
30. Domogalla, J. *et al.* Region of difference 4 in Alpine *Mycobacterium caprae isolates* indicates three variants. *J Clin Microbiol* **51**, 1381-8 (2013).
31. Brodin, P. *et al.* Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Immun* **70**, 5568-78 (2002).