ARCTIC
VOL. 70, NO. 4 (DECEMBER 2017) P. S1 – S5
https://doi.org/10.14430/arctic4683

# A Case Study: Was Private William Braine of the 1845 Franklin Expedition a Victim of Tuberculosis?

Jannine Forst[1,2] and Terence A. Brown[1,3]

## APPENDIX 1: NGS DATA ANALYSIS

The NGS data were analysed using the following procedure and programs. The Illumina adapters were trimmed and overlapping reads merged using AdapterRemoval v.2.1.3 (standard parameters, as well as –trimns and –collapse; Lindgreen, 2012; Schubert et al., 2016). The sequence reads were aligned to five reference genomes: *Mycobacterium tuberculosis* (H37Rv clinical strain GCA_000195955.2), *Mycobacterium tusciae* (JS617, GCA_000243415.3), *Streptococcus pneumoniae* (R6, GCA_000007045.1), human nuclear (hg38, GRCh38/GCA_000001405.15, with the mitochondrial DNA removed) and human mitochondrial (hg38 Chr M, GRCh38/GCA_000001405.15). Alignments were performed using BWA v.0.7.12-r1039, the Burrows-Wheeler Aligner (Li and Durbin, 2009). The parameters used were changed slightly to take account of the typical features of aDNA, with seed length disabled (-l 1024) and the maximum number of differences set to 0.1 (-n 0.1). For the human genome, the maximum number of differences was left at the (less stringent) standard setting of -n 0.04. The BWA output was then cleaned using Picard Tools v.1.95, sorted by co-ordinate (Picard Tools), and filtered using Samtools v.1.3 so unmapped or badly mapped (quality score < 30) reads were discarded (Li et al., 2009). Duplicates were removed using aweSam as the script takes 3′ coordinates into consideration when determining if reads are duplicates, unlike duplicate removal using Picard Tools which removes reads with identical 5′ coordinates, regardless of whether the 3′ coordinate differs and thus the read is not a true duplicate read. For the unmerged reads, singletons (with which only one of the two paired reads aligned to a genome) were also discarded.

The file format was converted to fasta using bedtools v.2.25.0 and seqtk v.1.0-r31 and the reads compared with the NCBI nucleotide database using BLASTn (Altschul et al., 1990; Quinlan and Hall, 2010). The standard parameters were used, with the exception that a minimum E-value of $10^{-6}$, minimum percent identity of 90%, and minimum word size of 15 bp were applied. Finally, the BLASTn output was assessed using MEGAN6 (Huson et al., 2007). The standard MEGAN6 parameters were changed to match those used for BLASTn (v.2.4.0+). Namely, the minimum score was changed to 0, maximum expected to $10^{-6}$, top percentage to 0, minimum support percentage turned off, minimum support to 2, and LCA percent as 100. The 23 reads identified as being specific to MTBC are shown in Appendix Table S1, along with the genes that those sequences match and the particular MTBC species giving the closest match. The species identifications were based on the species that were most frequent in the top 100 matches. In most cases all these matches had equal confidence from their E values, identity, and other attributes.

The human reads were also analyzed using GATK (v.3.6; McKenna et al., 2010) for variant calling, according to the GATK best practices recommendations for small data sets, with slight modifications for the new GATK version (DePristo et al., 2011; Van der Auwera et al., 2013). For this data set, however, the single nucleotide polymorphisms (SNPs) that were called had confidence values that were too low to report because of the extremely low coverage and lack of read depth. The mitochondrial DNA haplogroup was determined manually using Geneious v.7.1.9 (Kearse et al., 2012; http://www.geneious.com) and HaploGrep 2 (Weissensteiner et al., 2016), because the coverage was too low to use GATK. The haplogroup was assigned with low confidence. This is because SNPs from multiple haplogroups were present, indicating a considerable amount of contamination with modern human DNA. When looked at individually, many of the polymorphisms were either present at low coverage (i.e., only one sequence covered each SNP) or sporadic. One example is the set of three sequences obtained for mitochondrial DNA position 8860. One of these displayed an SNP while the other two sequences showed no polymorphisms when compared to the revised Cambridge Reference Sequence. All three of these sequences also showed potential aDNA damage (Appendix Fig. S1). This area was subsequently excluded from analysis using HaploGrep 2. However, this means that a mixed haplogroup could have been obtained since many of the other areas were covered by only one sequence so the polymorphisms (or lack thereof) could not be validated. This possibility is further shown by the fact that after analysis using GATK's Unified Genotyper, only one SNP was determined with any confidence—position 2706, where three sequences agreed that a G is found instead of the A of the reference genome. Unfortunately, many haplogroups display this polymorphism.

[1] Manchester Institute of Biotechnology, School of Earth and Environmental Sciences, University of Manchester, Manchester M1 7DN, United Kingdom
[2] Present address: Department of Anthropology, University of California Santa Cruz, Santa Cruz, California 95064, USA
[3] Corresponding author: terry.brown@manchester.ac.uk

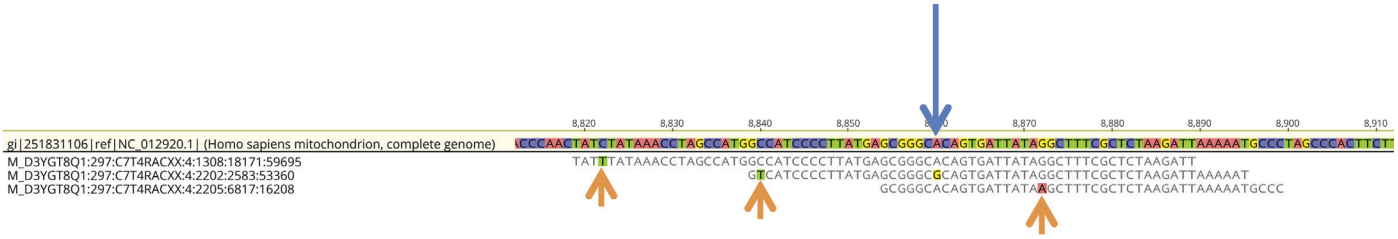


FIG. S1. Example of the difficulties in assigning a mitochondrial DNA haplogroup to William Braine. The reference mitochondrial DNA sequence from positions 8809 to 8915 is shown, along with three sequence reads from William Braine. The blue arrow indicates position 8860, where two genotypes can be seen: one that matches the reference (A) and one where an SNP is present (G instead of A). The orange arrows indicate the polymorphisms that may be either the result of DNA degradation or private SNPs (not haplogroup defining) from various individuals.

## REFERENCES


Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215(3):403–410.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire J.R., Hartl, C., Philippakis, A.A., et al. 2011. A framework for variation discovery ad genotyping using next-generation DNA sequencing data. Nature Genetics 43(5):491–498. https://doi.org/10.1038/ng.806

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. 2007. MEGAN analysis of metagenomic data. Genome Research 17(3):377–386. https://doi.org/10.1101/gr.5969107

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Lew, J.M., Kapopoulou, A., Jones, L.M., and Cole, S.T. 2011. TubercuList – 10 years after. Tuberculosis (Edinburgh, Scotland) 91(1):1–7. https://doi.org/10.1016/j.tube.2010.09.008

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindgreen, S. 2012. AdapterRemoval: Easy cleaning of next generation sequencing reads. BMC Research Notes 5: 337. https://doi.org/10.1186/1756-0500-5-337

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernystsky, A., Garimella, K., et al. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20(9):1297–1303. https://doi.org/10.1101/gr.107524.110

Quinlan, A.R., and Hall, I.M. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033

Schubert, M., Lindgreen, S., and Orlando, L. 2016. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. BMC Research Notes 9: 88. https://doi.org/10.1186/s13104-016-1900-2

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., et al. 2013. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. Current Protocols in Bioinformatics 43: Unit 11.10:1–33. https://doi.org/10.1002/0471250953.bi1110s43

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. 2016. HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Research 44(W1):W58–63. https://doi.org/10.1093/nar/gkw233

TABLE S1. Identifications for the 23 NGS sequences from William Braine's rib that were specific to the MTBC.

| Gene | Gene product role[1] | Species comprising the top 100 matches | Sequence of matched gene (with mismatches to corresponding William Braine sequence in red) |
|---|---|---|---|
| Sequence 1: | CGCAAGGCGCTGTCGATGGGGGCCGACAAGGCCG | | |
| fixA | Probable protein subunit which transfers electrons to the main respiratory chain | *Mycobacterium sinense* | CGCAAGGCGCTGTCGATGGGGGCCGACAAGGCCG |
| | | *Mycobacterium canettii* | CGCAAGGCGCTGTCGATGGGGGCCGACAAGGCCG |
| Sequence 2: | CCGCGGTTGGCGTCGATCAGCGCCGTGAAGC | | |
| echA20 | Probable enoyl hydrase that oxidizes fatty acids using specific components | *Mycobacterium tuberculosis* | CCGCGGTTGGCGTCGATCAGCGCCGTGAA |
| | | *Mycobacterium bovis* | CCGCGGTTGGCGTCGATCAGCGCCGTGAA |
| | | *Mycobacterium africanum* | CCGCGGTTGGCGTCGATCAGCGCCGTGAA |
| | | *Mycobacterium canettii* | CCGCGGTTGGCGTCGATCAGCGCCGTGAAGC |
| | | *Mycobacterium microti* | CCGCGGTTGGCGTCGATCAGCGCCGTGAA |
| Sequence 3: | CCGAAGCCCGGCTCACCCCGTTGGCGATGGAGATGC | | |
| gyrA | DNA gyrase subunit A that manipulates double-stranded DNA | *Mycobacterium tuberculosis* | CCGAAGCCCGGCTGACCCCGTTGGCGATGGAGATGC |
| | | *Mycobacterium bovis* | CCGAAGCCCGGCTGACCCCGTTGGCGATGGAGATGC |
| | | *Mycobacterium africanum* | CCGAAGCCCGGCTGACCCCGTTGGCGATGGAGATGC |
| Sequence 4: | CCCGTGACCCGACGCGCGCGGATCTTGCCGCG | | |
| rpsR1 | 30S ribosomal protein thought to play a functional role in polypeptide synthesis | *Mycobacterium tuberculosis* | CCCGTGACCCGACGCGCGCGGATCTTGCCGCG |
| | | *Mycobacterium bovis* | CCCGTGACCCGACGCGCGCGGATCTTGCCGCG |
| | | *Mycobacterium africanum* | CCCGTGACCCGACGCGCGCGGATCTTGCCGCG |
| | | *Mycobacterium canettii* | CCCGTGACCCGACGCGCGCGGATCTTGCCGCG |
| | | *Mycobacterium microti* | CCCGTGACCCGACGCGCGCGGATCTTGCCGCG |
| Sequence 5: | GCGGCAAGATCCGCGCGCGTCGGGTCACGGGCAAC | | |
| rpsR1 | 30S ribosomal protein thought to play a functional role in polypeptide synthesis | *Mycobacterium tuberculosis* | GCGGCAAGATCCGCGCGCGTCGGGTCACGGGCAAC |
| | | *Mycobacterium bovis* | GCGGCAAGATCCGCGCGCGTCGGGTCACGGGCAAC |
| | | *Mycobacterium africanum* | GCGGCAAGATCCGCGCGCGTCGGGTCACGGGCAAC |
| | | *Mycobacterium canettii* | GCGGCAAGATCCGCGCGCGTCGGGTCACGGGCAAC |
| | | *Mycobacterium microti* | GCGGCAAGATCCGCGCGCGTCGGGTCACGGGCAAC |
| Sequence 6: | ACACAGTTGCCCGTGACCCGACGCGCGCGGATCTTGCCG | | |
| rpsR1 | 30S ribosomal protein thought to play a functional role in protein synthesis | *Mycobacterium tuberculosis* | ACGCAGTTGCCCGTGACCCGACGCGCGCGGATCTTGCCG |
| | | *Mycobacterium bovis* | ACGCAGTTGCCCGTGACCCGACGCGCGCGGATCTTGCCG |
| | | *Mycobacterium africanum* | ACGCAGTTGCCCGTGACCCGACGCGCGCGGATCTTGCCG |
| | | *Mycobacterium canettii* | ACGCAGTTGCCCGTGACCCGACGCGCGCGGATCTTGCCG |
| | | *Mycobacterium microti* | ACGCAGTTGCCCGTGACCCGACGCGCGCGGATCTTGCCG |
| Sequence 7: | GCTCGTCTTCCATGCGCAGCAGCTTGCCGGTCTCG | | |
| clpB | Probable subunit of an ATP-binding heat shock protein, predicted to be essential for in vivo survival and pathogenicity | *Mycobacterium tuberculosis* | GCTCGTCTTCCATGCGCAGCAGCTTGGCGGTCTCG |
| | | *Mycobacterium bovis* | GCTCGTCTTCCATGCGCAGCAGCTTGGCGGTCTCG |
| | | *Mycobacterium africanum* | GCTCGTCTTCCATGCGCAGCAGCTTGGCGGTCTCG |
| | | *Mycobacterium canettii* | GCTCGTCTTCCATGCGCAGCAGCTTGGCGGTCTCG |
| | | *Mycobacterium microti* | GCTCGTCTTCCATGCGCAGCAGCTTGGCGGTCTCG |
| Sequence 8: | ACCCGCATGGGTTCGGGCAAGGGCTCGCCGAAGTGG | | |
| rplP | 50S ribosomal protein, potentially with a role in protein subunit assembly | *Mycobacterium tuberculosis* | ACCCGCATGGGTTCGGGCAAGGGCTCGCCGGAGTGG |
| | | *Mycobacterium bovis* | ACCCGCATGGGTTCGGGCAAGGGCTCGCCGGAGTGG |
| | | *Mycobacterium africanum* | ACCCGCATGGGTTCGGGCAAGGGCTCGCCGGAGTGG |
| | | *Mycobacterium canettii* | ACCCGCATGGGTTCGGGCAAGGGCTCGCCGGAGTGG |
| | | *Mycobacterium microti* | ACCCGCATGGGTTCGGGCAAGGGCTCGCCGGAGTGG |
| Sequence 9: | AAGAGTACGGTGGCGGCGGCGCAGGCATGTACGAA | | |
| fadE12 | Involved in lipid degradation but with an unknown function | *Mycobacterium tuberculosis* | AAGAGTACGGCGGCGGCGGCGCAGGCATGTACGAA |
| | | *Mycobacterium bovis* | AAGAGTACGGCGGCGGCGGCGCAGGCATGTACGAA |
| | | *Mycobacterium africanum* | AAGAGTACGGCGGCGGCGGCGCAGGCATGTACGAA |
| | | *Mycobacterium canettii* | AAGAGTACGGCGGCGGCGGCGCAGGCATGTACGAA |
| | | *Mycobacterium microti* | AAGAGTACGGCGGCGGCGGCGCAGGCATGTACGAA |

TABLE S1. Identifications for the 23 NGS sequences from William Braine's rib that were specific to the MTBC – *continued*:

| Gene | Gene product role[1] | Species comprising the top 100 matches | Sequence of matched gene (with mismatches to corresponding William Braine sequence in red) |
|---|---|---|---|
| Sequence 10: TTCATCACCGACGCCAGCCACGAGCTGCGTACGC | | | |
| trcS | Sensor portion of a sensor histidine kinase, usually involved in environmental information processing | *Mycobacterium tuberculosis* | TTCATCACCGACGCCAGCCACGAGCTGCGTACGC |
| | | *Mycobacterium bovis* | TTCATCACCGACGCCAGCCACGAGCTGCGTACGC |
| | | *Mycobacterium africanum* | TTCATCACCGACGCCAGCCACGAGCTGCGTACGC |
| | | *Mycobacterium canettii* | TTCATCACCGACGCCAGCCACGAGCTGCGTACGC |
| | | *Mycobacterium microti* | TTCATCACCGACGCCAGCCACGAGCTGCGTACGC |
| Sequence 11: CCGCGAAGTCGAGCACCCGCGGGTAGGCCGACCACC | | | |
| glyA | DNA gyrase subunit A that manipulates double-stranded DNA | *Mycobacterium tuberculosis* | CCGCGAAGTCGAGCACCCGCGGGTAGGCCGACCACC |
| | | *Mycobacterium bovis* | CCGCGAAGTCGAGCACCCGCGGGTAGGCCGACCACC |
| | | *Mycobacterium africanum* | CCGCGAAGTCGAGCACCCGCGGGTAGGCCGACCACC |
| | | *Mycobacterium canettii* | CCGCGAAGTCGAGCACCCGCGGGTAGGCCGACCACC |
| | | *Mycobacterium microti* | CCGCGAAGTCGAGCACCCGCGGGTAGGCCGACCACC |
| Sequence 12: GTCGGCCTACCCGCGGGTGCTCGACTTCGCGGCCTT | | | |
| glyA | DNA gyrase subunit A that manipulates double-stranded DNA | *Mycobacterium tuberculosis* | GTCGGCCTACCCGCGGGTGCTCGACTTCGCGGC**G**TT |
| | | *Mycobacterium bovis* | GTCGGCCTACCCGCGGGTGCTCGACTTCGCGGC**G**TT |
| | | *Mycobacterium africanum* | GTCGGCCTACCCGCGGGTGCTCGACTTCGCGGC**G**TT |
| | | *Mycobacterium canettii* | GTCGGCCTACCCGCGGGTGCTCGACTTCGCGGC**G**TT |
| | | *Mycobacterium microti* | GTCGGCCTACCCGCGGGTGCTCGACTTCGCGGC**G**TT |
| Sequence 13: CATGAGCGACAACATCCGCACCGTGCTCTACGGCGC | | | |
| lysA | Product involved in intermediary metabolism and respiration through biosynthesis of lysine | *Mycobacterium tuberculosis* | CATGAGCGACAACATCCGCACCG**C**GCTCTACGGCGC |
| | | *Mycobacterium bovis* | CATGAGCGACAACATCCGCACCG**C**GCTCTACGGCGC |
| | | *Mycobacterium africanum* | CATGAGCGACAACATCCGCACCG**C**GCTCTACGGCGC |
| | | Mycobacterium *canettii* | CATGAGCGACAACATCCGCACCG**C**GCTCTACGGCGC |
| | | *Mycobacterium microti* | CATGAGCGACAACATCCGCACCG**C**GCTCTACGGCGC |
| Sequence 14: ATAGGTGTTGTGGAAGCCGGGCCGCTTGCACCCGACCG | | | |
| | Putative mono-oxygenase gene<br>Possibly involved in metabolic pathways | *Mycobacterium tuberculosis* | ATAGGTGTTGTGGAAGCCGGGCCGCTTGCACCC**C**ACCG |
| | | *Mycobacterium bovis* | ATAGGTGTTGTGGAAGCCGGGCCGCTTGCACCC**C**ACCG |
| | | *Mycobacterium africanum* | ATAGGTGTTGTGGAAGCCGGGCCGCTTGCACCC**C**ACCG |
| | | *Mycobacterium canettii* | ATAGGTGTTGTGGAAGCCGGGCCGCTTGCACCC**C**ACCG |
| | | *Mycobacterium microti* | ATAGGTGTTGTGGAAGCCGGGCCGCTTGCACCC**C**ACCG |
| Sequence 15: GGTGTGCGCTATTTCGGGCCGTACTCGCACG | | | |
| uvrC | Probable excinuclease subunit, involved in nucleotide excision repair | *Mycobacterium tuberculosis* | GGTGTGCGCTATTTCGGGCCGTACTCGCACG |
| | | *Mycobacterium bovis* | GGTGTGCGCTATTTCGGGCCGTACTCGCACG |
| | | *Mycobacterium africanum* | GGTGTGCGCTATTTCGGGCCGTACTCGCACG |
| | | *Mycobacterium canettii* | GGTGTGCGCTATTTCGGGCCGTACTCGCACG |
| | | *Mycobacterium microti* | GGTGTGCGCTATTTCGGGCCGTACTCGCACG |
| | | *Mycobacterium avium* | GGTGTGCGCTATTTCGGGCCGTACTCGCACG |
| Sequence 16: TGGGTCAAGGGCGCGACGTCCGGCCACACCCAGCATG | | | |
| hisI | Probable product involved in histidine biosynthesis pathway, essential gene for in vitro growth of TB | *Mycobacterium tuberculosis* | TGGGTCAAGGGCGCGACGTCCGGCCACACCCAGCA |
| | | *Mycobacterium bovis* | TGGGTCAAGGGCGCGACGTCCGGCCACACCCAGCA |
| | | *Mycobacterium africanum* | TGGGTCAAGGGCGCGACGTCCGGCCACACCCAGCA |
| | | *Mycobacterium canettii* | TGGGTCAAGGGCGCGACGTCCGGCCACACCCAGCA |
| | | *Mycobacterium microti* | TGGGTCAAGGGCGCGACGTCCGGCCACACCCAGCA |
| Sequence 17: AGACGGTGCGCCCGGTGGAGCGCTTCTGCAG | | | |
| uvrA | Probable excinuclease subunit, involved in nucleotide excision repair | *Mycobacterium tuberculosis* | AGACGGTGCGCCCGGTGGAGCGCTTCTGCAG |
| | | *Mycobacterium bovis* | AGACGGTGCGCCCGGTGGAGCGCTTCTGCAG |
| | | *Mycobacterium africanum* | AGACGGTGCGCCCGGTGGAGCGCTTCTGCAG |
| | | *Mycobacterium canettii* | AGACGGTGCGCCCGGTGGAGCGCTTCTGCAG |
| | | *Mycobacterium microti* | AGACGGTGCGCCCGGTGGAGCGCTTCTGCAG |

TABLE S1. Identifications for the 23 NGS sequences from William Braine's rib that were specific to the MTBC – *continued*:

| Gene | Gene product role[1] | Species comprising the top 100 matches | Sequence of matched gene (with mismatches to corresponding William Braine sequence in red) |
|---|---|---|---|
| Sequence 18: CGACGCCGGTGATGGTGCCACCGGTGCCGACT | | | |
| cysK1 | Cysteine synthase involved in cysteine biosynthesis | *Mycobacterium tuberculosis* | CGACGCCGGTGATGGTGCCACCGGTGCCGACT |
| | | *Mycobacterium bovis* | CGACGCCGGTGATGGTGCCACCGGTGCCGACT |
| | | *Mycobacterium africanum* | CGACGCCGGTGATGGTGCCACCGGTGCCGACT |
| | | *Mycobacterium canettii* | CGACGCCGGTGATGGTGCCACCGGTGCCGACT |
| | | *Mycobacterium microti* | CGACGCCGGTGATGGTGCCACCGGTGCCGACT |
| Sequence 19: CTTCGGCTCCGGGTTGAGCCAGTGCGCGTGCCGGCT | | | |
| | Hypothetical protein | *Mycobacterium tuberculosis* | CTT**G**GGCTCCGGGTTGAGCCAGTGCGCGTGCCGGCT |
| | A conserved hypothetical protein | *Mycobacterium bovis* | CTT**G**GGCTCCGGGTTGAGCCAGTGCGCGTGCCGGCT |
| | | *Mycobacterium africanum* | CTT**G**GGCTCCGGGTTGAGCCAGTGCGCGTGCCGGCT |
| | | *Mycobacterium canettii* | CTT**G**GGCTCCGGGTTGAGCCAGTGCGCGTGCCGGCT |
| | | *Mycobacterium microti* | CTT**G**GGCTCCGGGTTGAGCCAGTGCGCGTGCCGGCT |
| Sequence 20: ATTGCGACCGCGACCGATGTCGACGAACGCCGCCT | | | |
| rne | Possible ribonuclease thought to be involved in several cellular processes | *Mycobacterium tuberculosis* | ATTGCGACCGCG**C**CCGATGTCGACGA ACGCCGCCT |
| | | *Mycobacterium bovis* | ATTGCGACCGCG**C**CCGATGTCGACGA ACGCCGCCT |
| | | *Mycobacterium africanum* | ATTGCGACCGCG**C**CCGATGTCGACGA ACGCCGCCT |
| | | *Mycobacterium canettii* | ATTGCGACCGCG**C**CCGATGTCGACGA ACGCCGCCT |
| | | *Mycobacterium microti* | ATTGCGACCGCG**C**CCGATGTCGACGA ACGCCGCCT |
| Sequence 21: CCTTGCAGCGCAAAGAGAATCACGATGGTGAA | | | |
| arsC | Involved in export of arsenic compounds across the membrane | *Mycobacterium tuberculosis* | CCTTGCAGCGCAAAGAGAATCACGATGGTGAA |
| | | *Mycobacterium bovis* | CCTTGCAGCGCAAAGAGAATCACGATGGTGAA |
| | | *Mycobacterium africanum* | CCTTGCAGCGCAAAGAGAATCACGATGGTGAA |
| | | *Mycobacterium canettii* | CCTTGCAGCGCAAAGAGAATCACGATGGTGAA |
| | | *Mycobacterium microti* | CCTTGCAGCGCAAAGAGAATCACGATGGTGAA |
| Sequence 22: TCGGTTCGCCCAAGCGGCGTGAGATCGGGCACGGCG | | | |
| gpsI | Involved in mRNA degradation | *Mycobacterium tuberculosis* | TCGGTTCGCCCAAGCGGCGTGAGATCGGGCACGGCG |
| | | *Mycobacterium bovis* | TCGGTTCGCCCAAGCGGCGTGAGATCGGGCACGGCG |
| | | *Mycobacterium africanum* | TCGGTTCGCCCAAGCGGCGTGAGATCGGGCACGGCG |
| | | *Mycobacterium canettii* | TCGGTTCGCCCAAGCGGCGTGAGATCGGGCACGGCG |
| | | *Mycobacterium microti* | TCGGTTCGCCCAAGCGGCGTGAGATCGGGCACGGCG |
| Sequence 23: TAGTCGACGATGTGGTCGGTGTGCAGGTGGGTGAT | | | |
| | Hypothetical protein | *Mycobacterium tuberculosis* | TAGTCGACGATGTGGTCGGTGTGCA**A**GTGGGTGAT |
| | A conserved hypothetical protein | *Mycobacterium bovis* | TAGTCGACGATGTGGTCGGTGTGCA**A**GTGGGTGAT |
| | | *Mycobacterium africanum* | TAGTCGACGATGTGGTCGGTGTGCA**A**GTGGGTGAT |
| | | *Mycobacterium canettii* | TAGTCGACGATGTGGTCGGTGTGCA**A**GTGGGTGAT |
| | | *Mycobacterium microti* | TAGTCGACGATGTGGTCGGTGTGCA**A**GTGGGTGAT |

[1] The gene descriptions were obtained from the TubercuList knowledge base (Lew et al., 2011)