Supplementary Material

Methods

Sanger sequencing and GenoType MTBDR*s*/ VER 2.0

All isolates to be Sanger sequenced were re-cultured from frozen bacterial stock cultures. For isolates of sample set 1, the sequence of the *eis*-promoter region in each isolate was determined by PCR amplification of thermal lysates followed by Sanger sequencing. Each isolate of sample set 1 for which Sanger sequencing detected an *eis*-promoter mutation, Sanger sequencing was repeated once to confirm the result. Isolates from sample set 2 were Sanger sequenced to confirm WGS results, using the same DNA that was used for WGS (see below). Briefly, the PCR reaction mix contained the following final concentrations of: 1x HotStartTaq *Plus* Master Mix (Qiagen, San Diego, CA, USA), 500nM of each primer (forward 5' CCATGGGACCGGTACTTGCT 3', reverse 5' ACTTCACCAGGCACCGTCAA 3'), and 1x SYTO 9 Green Fluorescent Nucleic Acid Stain (Thermo Fisher Scientific). As template, 1ul of thermal lysate (sample set 1) or purified DNA (sample set 2) was added to the reaction mix. Amplification of the *eis*-promoter region of the selected isolates was carried out using a CFX96^{TM} Real-Time System C1000 Touch Thermal Cycler (BioRad) running the following thermocycling protocol: Initial denaturation at 95°C for 5min, followed by 40 cycles of 95°C for 1min, annealing at 62°C for 1min and elongation at 72°C for 1min, followed by a final elongation at 72°C for 10min. Successful amplification was confirmed by a high-resolution melt from 80°C – 95°C with an increment of 0.5°C, each increment temperature held for 5 seconds.

Isolates which repeatedly failed to amplify were excluded from further analyses. Successfully amplified PCR products were sent to the Central DNA Sequencing Facilities of Stellenbosch University for targeted Sanger sequencing using the forward PCR Primer. The resulting chromatographs were analyzed using

BioEdit Sequence Alignment Editor v. 7.2.5 (1) comparing them to the *Mycobacterium tuberculosis* H37Rv reference genome (Accession number: AL123456).

The GenoType MTBDR*sl* VER 2.0 (MTBDR*sl*) assay was done according to the manufacturer's protocol using the same DNA used for WGS (if not indicated otherwise). Failing assays (*e.g.,* complete gene locus control or conjugate control band missing; defined as per manufacturer's protocol) were repeated once. The analytical sensitivity (limit of detection) of the MTBDR*sl* assay is 1.65 x 10^5 bacteria/ml for culture samples and 150 bacteria/ml for clinical samples (2).

Phenotypic Drug Susceptibility Testing and Minimum Inhibitory Concentration Determination

All isolates used in this study (sample sets 1 and 2), were initially subjected to routine pDST on solid Löwenstein Jensen medium against INH, RIF, AMK and ofloxacin (OFX). Susceptibility was determined according to the 1% proportion method at clinical breakpoints of 0.2ug/ml for INH, 40.0ug/ml for RIF, 30ug/ml for AMK and 2ug/ml for OFX (3, 4). MICs for KAN were subsequently determined for isolates with an *eis*-promoter mutation missed by the MTBDR*sl* (sample set 1) and for representatives of each additional (combination of) *eis*-promoter mutation(s) (sample set 2). These MICs were done using two-fold serial dilutions ranging from 10.0ug/ml to 1.25ug/ml using the BACTEC MGIT 960 system with the TB eXiST module of the EpiCentre software at Stellenbosch University (5). Susceptibility to KAN and AMK (*i.e.,* pDST at Stellenbosch University) was based on a clinical breakpoint of 2.5ug/ml for KAN and 30ug/ml for AMK as per the 1% proportion method, defined as the lowest drug concentration that inhibits > 99% of growth. One isolate showed intermediate growth (*i.e.,* growth of > 100 growth units [GU] within seven days after the growth control reached a GU of 400) at all measured drug concentrations. The bacteria that grew under KAN pressure (intermediate growth < 1%; bacteria from the 10ug/ml drug containing tube) were re-grown in KAN containing medium (*i.e.,* selective sub-culturing), and pDST and subsequent Sanger sequencing were repeated following the procedures described above.

Whole genome sequencing

For WGS each isolate was re-cultured from culture stocks and DNA was extracted by standard procedures as previously described (6). Whole genome sequencing libraries had been prepared using the standard genomic DNA sample preparation kits from Illumina (Illumina, Inc, San Diego, CA), following the manufacturer's protocol. The libraries were sequenced on an llumina HiSeq or Illumina NextGen Seq platform. The resulting sequencing reads were mapped to the *Mtb* H37*Rv* reference strain (Accession No. AL123456). Variant calling and annotation were conducted using a within-house pipeline including 3 mappers (Burrows-Wheeler Alignment tool, NovoAlign, Smalt) (7–9) and 2 variant callers (GATK Gene Analysis Tool Kit, SAMtools) (7, 10) as previously described (11). Sequences with an average coverage below 20x for 2 or 3 of the mappers and/or with mapped reads <80% for 2 or 3 of the mappers were excluded, resulting in an average coverage (*i.e.,* average across the three mappers) of 35x to 443x per isolate. Only SNVs called from all 3 alignment bam files were considered high confidence SNVs. No frequency cut-off was applied, and variants detected at a frequency ≥ 95% were considered fixed. The genotypic drug resistance profile of each isolate was determined using markers defined by Coll *et al* and Miotto *et al* (12, 13). Artemis (14) was used to visually inspect sequencing reads. Based on this visual inspection of the reads and on variant frequency analysis it was determined that none of the isolates with more than one *eis*-promoter mutation had a double mutation, *i.e.,* the different mutations were always on different WGS reads, suggesting differently evolved sub-clones within the same patient.

Raw sequencing reads of the isolates listed in Table 1 have been deposited at the European Nucleotide Archive (PRJEB41458).

Limitations

The presented study made use of two different sample sets complementing each other but also baring limitations described below.

Not all isolates of data set 1 that were typed SLID susceptible by the MTBDR*s*l were available for Sanger sequencing due to insufficient sample volumes, contamination, or loss of viability. The proportion of missed *eis*-promoter mutations could therefore be higher.

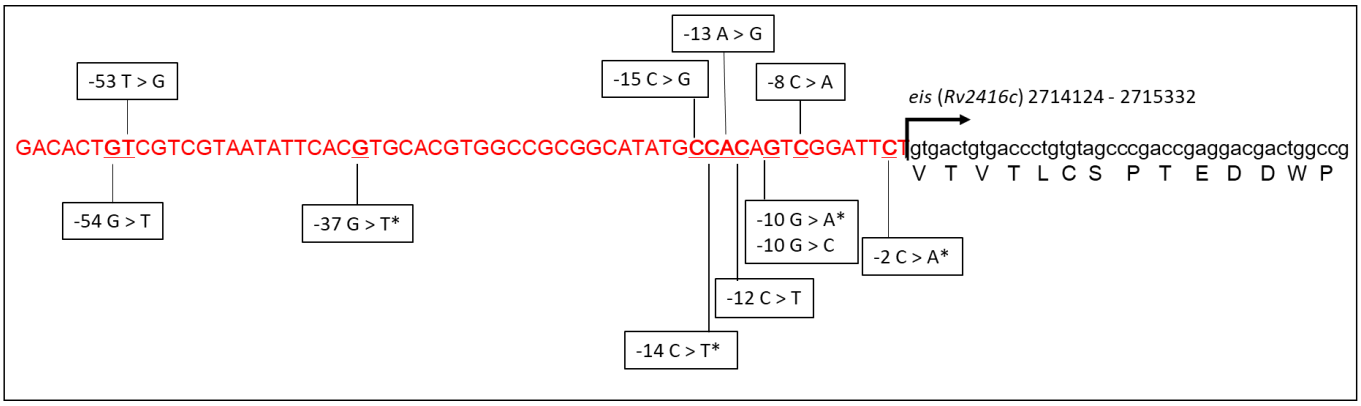
Despite analyzing data collected over the last 25 years, no conclusions about the prevalence of *eis*-promoter mutations across that period can be drawn, as sample set 2 of this study was a convenience sample of WGS isolates collected from several studies with different research questions. All WGS isolates irrespective of their phenotypic status were screened for *eis*-promoter mutations, but only representatives of those carrying such a mutation were further analyzed. However, in combination with data set 1 – which is surveillance data from one year of the rifampicin-resistant *Mtb* population in the WCP – our data provide insights on the type and frequency of *eis*-promoter mutations present in the WCP.

Unfortunately, no data on treatment outcome of the patients was available for analysis and no conclusions on the clinical impact of the detected *eis*-promoter mutations can be drawn. Similarly, sub-culturing is required to determine KAN MICs. As some of the isolates lost viability, the impact of these isolate's mutations on the MIC could not be determined.

In this study, the MIC was only determined once. For susceptible isolates with an MIC near the clinical breakpoint of 2.5ug/ml (*e.g.*, table 2) a repetition of the MIC may have resulted in a slightly elevated MIC and therefore in low-level KAN resistance.

Our results do not allow to make any conclusions on whether *eis*-promoter mutations act as steppingstones for the acquisition of high-level resistance for KAN (and/or AMK). To analyze this, further *in vitro* experiments, and analyses of serial samples would be required to investigate the development of resistance over time with and without drug pressure.

Supplementary Figures and Tables



Supplementary Figure 1: *Mycobacterium tuberculosis* H37Rv coding sequence of the gene *eis* (*Rv2416c;* black) and 60bp of its promoter region (red), with the known kanamycin resistance conferring mutations according to Coll *et al* and Miotto *et al* (12, 13). The mutations -15 C > G and -10 G > C are mutations disputed to confer resistance. The GenoType MTBDR*sl* VER 2.0 assay (Hain Lifescience, Germany) defines specific banding patterns for only the most common *eis*-promoter mutations (marked with *). The remaining *eis*-promoter mutations may however also cause a failing wild type band, which would then be interpreted as "undefined mutation detected" (2).

Supplementary Table 1: *Eis*-promoter mutations and their frequency in 2863 whole genome sequenced isolates

Supplementary Table 2: Additional information on the samples and patients of data set 2

M = male; F = female

References

1. Hall, T.A. 1999. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98.

2. Hain Lifescience. 2016. Instructions for Use for GenoType MTBDRsl VER 2.0.

3. Canetti G, Fox W, Khomenko A, Mahler HT, Menon NK, Mitchison DA, Rist N, Smelev NA. 1969. Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. Bull World Health Organ 41:21–43.

4. World Health Organization. 2018. Technical Report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis. Technical Report, Geneva, Switzerland.

5. Springer B, Lucke K, Calligaris-Maibach R, Ritter C, Böttger EC. 2009. Quantitative drug susceptibility testing of Mycobacterium tuberculosis by use of MGIT 960 and EpiCenter instrumentation. J Clin Microbiol 47:1773–1780.

6. Warren R, de Kock M, Engelke E, Myburgh R, Gey van Pittius N, Victor T, van Helden P. 2006. Safe Mycobacterium tuberculosis DNA extraction method that does not compromise integrity. J Clin Microbiol 44:254–256.

7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl 25:2078–2079.

8. NOCOCRAFT. 2014. NovoAlign Powerful tool designed for mapping of short reads onto a reference genome from Illumina, Ion Torrent, and 454 NGS plaforms.

9. Wellcome Trust Sanger Institute, Ponstingl H. SMALT (C) 2010 - 2015 Genome Research Ltd.

10. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

11. Black PA, de Vos M, Louw GE, van der Merwe RG, Dippenaar A, Streicher EM, Abdallah AM, Sampson SL, Victor TC, Dolby T, Simpson JA, van Helden PD, Warren RM, Pain A. 2015. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in Mycobacterium tuberculosis isolates. BMC Genomics 16:857.

12. Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, Kim PS, Liwski R, Zignol M, Gilpin C, Niemann S, Denkinger CM, Fleming J, Warren RM, Crook D, Posey J, Gagneux S, Hoffner S, Rodrigues C, Comas I, Engelthaler DM, Murray M, Alland D, Rigouts L, Lange C, Dheda K, Hasan R, Ranganathan UDK, McNerney R, Ezewudo M, Cirillo DM, Schito M, Köser CU, Rodwell TC. 2017. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. Eur Respir J 50.

13. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, Abdallah AM, Alghamdi S, Alsomali M, Ahmed AO, Portelli S, Oppong Y, Alves A, Bessa TB, Campino S, Caws M, Chatterjee A, Crampin AC, Dheda K, Furnham N, Glynn JR, Grandjean L, Minh Ha D, Hasan R, Hasan Z, Hibberd ML, Joloba M, Jones-López EC, Matsumoto T, Miranda A, Moore DJ, Mocillo N, Panaiotov S, Parkhill J, Penha C, Perdigão J, Portugal I, Rchiad Z, Robledo J, Sheen P, Shesha NT, Sirgel FA, Sola C, Oliveira Sousa E, Streicher EM, Helden PV, Viveiros M, Warren RM, McNerney R, Pain A, Clark TG. 2018. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet 50:307–316.

14. Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J, Rajandream M-A. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinforma Oxf Engl 24:2672–2676.