Multi- and Extensively-Drug Resistant *Mycobacterium tuberculosis* in South Africa: A Molecular Analysis of Historical Isolates

Nontuthuko E. Maningi,[a] Luke T. Daum,[b] John D. Rodriguez,[b] Halima M. Said,[c] Remco P.H. Peters,[a,d] John Osei Sekyere[e], Gerald W. Fischer,[b] James P. Chambers,[f] P. Bernard Fourie[a]#

[a]University of Pretoria, Department of Medical Microbiology, Pretoria, South Africa;

[b]Longhorn Vaccines & Diagnostics, San Antonio, Texas, USA;

[c]Centre for Tuberculosis, National TB Reference Laboratory, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa;

[d]Anova Health Institute, Johannesburg, South Africa;

[e]Kwame Nkrumah University of Science and Technology, Kumasi, Ghana;

[f]University of Texas at San Antonio, San Antonio, Texas, USA

Running Head: Molecular analysis of historical MDR-TB isolates

Keywords: Drug-resistant tuberculosis; South Africa; historical isolates; whole genome sequencing; *Mycobacterium tuberculosis.*

#Address correspondence to P Bernard Fourie, bernard.fourie@up.ac.za

ABSTRACT (242 words)

Modern advances in genomics provide an opportunity to re-interpret historical bacterial culture collections. In this study, genotypic antibiotic resistance profiles of *Mycobacterium tuberculosis* isolates from an historical 20-year-old multidrug-resistant tuberculosis (MDR-TB) culture collection in South Africa are described. DNA samples extracted from the phenotypically MDR-TB isolates (n=240) were assayed by Hain line probe assay (LPA) for the confirmation of MDR-TB and by Illumina Miseq whole genome sequencing (WGS) for the characterization of mutations in eight genes (*rpo*B, *kat*G, *inh*A, *rps*L, *pnc*A, *emb*B, *gyr*A and *rrs*) that are known to code for resistance to commonly used anti-TB agents.

LPA identified 71.3% of the TB isolates as MDR-TB, 18.3% as rifampicin (RIF) mono-resistant, 2% as isoniazid (INH) mono-resistant, and 8.3% as susceptible to both RIF and INH. In a subset of 42 randomly selected isolates designated RIF+INH-resistant by Löwenstein-Jensen (LJ) culture in 1993, LPA and WGS results confirmed MDR-TB. In all five INH mono-resistant isolates by LPA, and in all but one (wild-type) of the 34 successfully sequenced RIF mono-resistant isolates, WGS revealed matching mutations. Only 26% of isolates designated as susceptible by LPA, however, were found to be wild-type by WGS. Novel mutations were found in the *rpo*B (Thr480Ala, Gln253Arg, Val249Met, Val251Tyr; Val251Phe), *kat*G (Trp477STOP; Gln88STOP; Trp198STOP; and Trp412STOP), *emb*B (Thr11Xaa, Gln59Pro) and *pnc*A (Thr100Ile, Thr159Ala, Ala134Arg; Val163Ala, Thr153Ile, DelGpos7, Phe106Ser) genes. Three MDR-TB isolates showed mutations in both the *gyr*A and *rrs* genes, suggesting that extensively drug-resistant tuberculosis existed in South Africa well-before its formal recognition in 2006.

INTRODUCTION

South Africa has one of the highest multidrug-resistant tuberculosis (MDR-TB) burdens globally, with an estimated 10 000 total number of MDR or rifampicin-resistant (RR) TB cases in 2015 (1). Furthermore, extensively drug-resistant tuberculosis (XDR-TB) was first defined in 2006 by the World Health Organization (WHO) after an outbreak in Tugela Ferry, KwaZulu-Natal province (KZNP), South Africa. By the end of 2015, XDR-TB had been reported in 117 WHO member states, which is 9.5% of the global 480 000 MDR cases (1,2). However, detailed knowledge on drug resistance patterns and mutations associated with the MDR-TB epidemic in South Africa is limited, with only a few studies providing genomic data on MDR-TB strains in the country. These are mainly from the Western Cape and KwaZulu-Natal provinces (3,4).

At the time of establishing the strain collection used in this study, the gold standard method for identification of MDR-TB strains was the agar proportion method in Löwenstein-Jensen (LJ) solid agar culture medium, a time-consuming process requiring several weeks to obtain adequate results. Currently, the liquid culture Bactec 960 MGIT system is widely recommended for the detection of *Mycobacterium tuberculosis* (*M.tb*) in clinical specimens and for drug sensitivity testing of first-line and certain second-line drugs. It is recommended as a reliable alternative with faster turn-around times compared to agar proportion methods (5,6). Unfortunately, liquid culture methods are prone to higher rates of contamination, and consequently lost results, than solid agar methods. To overcome this problem, molecular assays are increasingly being employed for rapid detection of mutations associated with susceptibility in MDR-TB and XDR-TB.

In 2008 and 2016 respectively, the WHO endorsed the MTBDR*plus*® and MTBDR*sl* line probe assay (Hain Lifescience, Germany) for detecting common mutations in regions of genes conferring drug-resistance in *M.tb*. These molecular assays detect mutations in the "hot-spot"

region of genes that confer resistance to rifampicin (RIF), isoniazid (INH), ethambutol (EMB), quinolones and aminoglycosides (7).

With the recent development of next-generation sequencing (NGS) techniques, however, large-scale genome sequencing projects have become possible. Furthermore, NGS is gradually entering the diagnostic laboratory for the detection of *M.tb* and its resistance to anti-TB drugs, promising decreased cost of sequencing and increased throughput. (8)

Given the emerging molecular technologies that are becoming available recently, allowing for specific investigations into mutations associated with anti-TB drug resistance at the genomic level, we re-analysed the drug resistance profiles of an historical South African MDR-TB strain collection dating from the period 1993 to 1995. The aim was to specifically describe the mutations associated with drug resistance in each isolate, and to investigate whether any of these 20-year-old isolates harboured combinations of mutations that defined XDR-TB before this profile was formally reported in 2006. We also wanted to gain insights into the genetic diversity of MDR-TB strain types per geographical location (Western Cape Province in the South and Gauteng Province in the North).

MATERIALS AND METHODS

Collection of samples: A collection of 625 MDR-TB cultures, originally derived from sputum samples submitted to two routine TB laboratories (samples were collected from public hospital patients) in the Western Cape Province (WCP) and Gauteng Province (GP) between 1993 and 1995, was centrally stored at the South African Medical Research Council (SAMRC) laboratories in Pretoria, South Africa, after a confirmation of their MDR-TB status. From this collection, 240 isolates (120 each from WCP and GP) were randomly selected in 2014 for DNA extraction and molecular analysis. Care was taken not to include more than one isolate from each

patient represented in the collection. The demographic characteristics of patients represented by the 240 isolates were largely comparable. GP had slightly more males (77%) than females (23%) and a higher median age (38.1 years) than WCP with 67% males, 33% females and a median age of 33.3 years.

At the time of collection, sputum specimens were cultured on LJ medium and subjected to the proportion method as recommended by the WHO/International Union against Tuberculosis and Lung Disease (IUATLD) (9) for drug susceptibility testing of RIF and INH, and routinely also for streptomycin (SM), EMB and the second-line drugs kanamycin (KAN), capreomycin (CPM), cycloserine (CLS), and ofloxacin (OFX). Drug concentrations used were as follows: SM 4 μg/mL, INH 0.2 μg/mL, RIF 4 μg/mL, EMB 2 μg/mL, pyrazinamide (PZA) 100 μg/mL, KAN 20 μg/mL, CPM 2.5 μg/mL, CLS 30 μg/mL, and OFX 2 μg/mL.

Extraction of DNA from non-viable stored *M. tuberculosis* cultures

Preparation of suspensions of *M.tb* colonies from the stored specimens depended on the state of preservation of the original cultures. For intact slopes, colonies were gently scraped off using an inoculation loop and further suspended (washed down) in 1 mL of sterile water in the original culture bottle. The suspension was then pipetted-off and transferred into a 1.5 mL Eppendorf tube. For dried-out media and cultures, the dry substrate was crushed with an inoculation loop and a portion was transferred into a 1.5 mL Eppendorf tube. This was then gently crushed into finer fragments with a glass rod prior to adding 1 mL of sterile water and shaking. Eppendorf tubes containing suspensions obtained from either the intact or dried-out samples were centrifuged at 13000 × *g* for 5 mins and the supernatant discarded. DNA extraction was successfully achieved by applying the NucleoSpin®Tissue Kit (Macherey-Nagel, Germany) for extraction of genomic DNA as per the manufacturer's instructions. This procedure allowed

for complete digestion of remnant protein from the growth media, resulting in high-quality, pure DNA. DNA extraction was successful in all isolates included in this study.

Molecular drug resistance testing

First-line drug susceptibility testing for RIF and INH was done using the GenoType® version 2 MTBDR*plus* assay (Hain Lifescience, Germany). Amplification reactions were undertaken using a 35 µL primer nucleotide mix, 10 µL polymerase mix (Hain Lifescience, Germany), and 5µL of genomic DNA. PCR and hybridization were performed and interpreted as per the manufacturer's instructions using GTBlot and GenoScan machines (Hain Lifescience, Germany). The PCR products were denatured in a special solution at room temperature, followed by hybridization with provided buffer at 45°C for 30 min in a shaking water bath. After stringent washing of the PCR products, hybridization was detected by colorimetric reaction to control for cross contamination. Water, in lieu of the DNA template, was used as the negative control. Second-line drug susceptibility testing by similar methods as described above was carried out using the GenoType® MTBDR*sl* version 1 assay to detect resistance to fluoroquinolones (FQLs), amikacin (AMK), KAN, CPM and EMB.

Library preparation

Whole genome sequencing (WGS) was carried out using the Illumina Miseq (Illumina Inc., USA) to detect mutations in eight genes that confer resistance to anti-TB drugs (*rpo*B, *kat*G, *inh*A, *emb*B, *rps*L, *pnc*A, *gyr*A, *rrs*). The DNA used for WGS was quantified using a Qubit (Invitrogen, Eugene, USA). A concentration of 1-5 ng/µL genomic DNA from 111 samples was prepared and used with the Nextera XT Sample Prep Kit (Illumina, San Diego, CA, USA). Illumina whole genome sequencing (WGS) libraries were indexed into groups of 24 and pooled, as per the manufacturer's instructions (Illumina, San Diego, CA, USA). MiSeq NGS was

performed using the MiSeq Reagent Kit (V3) with 600 cycles. The sequencing was done as paired-end 150 base reads. The sequencing coverage was between 33-152x.

Bioinformatics analyses

Sequence output quality control was achieved by using the trimmomatric software for the trimming of adaptors and artifact to get the true DNA sequence of the genome. The same software was used to remove poor sequences with low base quality scores. Bioinformatics including contig assembly, mutational analysis, and multiple sequence alignments were performed using SeqMan NGen (V4) and LaserGene (V10) Core Suite software (DNAStar, Inc., Madison, WI, USA) with reference mapping.

Genotyping

Genotyping was performed on all 240 MDR isolates using a commercial spoligotyping kit (Ocimum BioSolutions, India), following procedures previously described (10). Resulting spoligotypes were compared with the TB Drug Resistance Mutation Database (11) and with the updated version SITVIT 2 from the Pasteur Institute in Guadeloupe. Spoligotype clusters with more than 9 isolates were selected for characterization using the 24-loci MIRU-VNTR quadruplex typing kit (GenoScreen, France).

PCR amplification was done using a 96-well plate (LASEC, South Africa) in a total PCR reaction volume of 12 µL consisting of 8 µL of MIRU-VNTR quadruplex mix and 2 µL of template DNA. The thermocycling profile included a 15-minute denaturation step at 95°C, followed by 40 cycles of 1 min at 94°C, 1 min at 59°C, and 1.5 min at 72°C, with a final extension step at 72°C for 10 min. The 24-loci MIRU-VNTR typing method was subjected to electrophoresis using an ABI 3130 genetic analyser (Applied Biosynthesis, USA). Sizing of PCR fragments and assignments of the alleles of the 24-loci was done using the MIRU-VNTR

calibration kit (GenoScreen, France) and Gene Mapper software version 4.0 (Applied Biosystem, USA).

Dendrograms were constructed using the Unweighted Pair Group Method with Arithmetic Averages (UPGMA) algorithm for spoligotyping combined with 24-loci MIRU-VNTR typing, and the categorical coefficient was used to calculate the distance matrix.

RESULTS

Resistance profiles as determined by Line Probe Assay

Of the 240 MDR-TB isolates investigated by LPA, 171 (71%) were confirmed by MTBDR*plus* as MDR-TB (RIF+INH resistant); 98 were from WCP (n=120) and 73 from GP (n=120). Overall, 44/240 (18.0%) isolates were mono-resistant to RIF (GP=30, WCP=14), 5/240 (2.1%) were mono-resistant to INH (GP=3, WCP=2), and 20/240 (8.0%) showed no resistance to RIF or INH (14 GP, 6 WCP) (Table 1).

For three MDR-TB isolates from the GP (n=120), no results were obtained for OFX or for the aminoglycosides. Of the remaining 237 isolates, three showed resistance to both OFX and an aminoglycoside by second-line LPA (GP=1, WCP=2). In three more isolates, one was phenotypically resistant to OFX but susceptible to the aminoglycosides (WCP), one was susceptible to OFX (GP) and the 3rd (WCP) had no phenotypic results recorded for OFX. A further three of the MDR-TB isolates (1.0%) were mono-resistant to OFX (1 GP, 2 WCP). None of the isolates showed mono-resistance to AMK or KAN. Fifty isolates were resistant to EMB, of which 42 were originally also detected by LJ drug susceptibility testing (Table 1 and Figure 1).

Illumina Miseq Whole Genome Sequencing

Funding limitations prevented sequencing of all 240 isolates. Whole-genome sequencing was performed on 42 randomly selected isolates with RIF+INH resistance originally detected by the LJ proportion method and confirmed by LPA, as well as on all 69 isolates that showed discrepant results for RIF and INH between the original LJ phenotypic DST results and the LPA results (44 mono-resistant to RIF, five mono-resistant to INH, and 20 isolates susceptible to both RIF and INH). Fifteen sequencing runs (10 in the RIF mono-resistant group and five in the susceptible group) failed, however, leaving a total of 96 (86.5%) selected isolates having results interpretable by WGS. All 42 RIF+INH isolates confirmed by LPA were also confirmed by WGS. Of the 34 RIF mono-resistant isolates identified by LPA and successfully sequenced, all except one (wild-type) were confirmed by WGS as resistant to RIF only. All five of the INH mono-resistant isolates by LPA were confirmed by WGS. WGS was successfully conducted in 15 of the 20 isolates susceptible to RIF and INH by LPA. Four were shown to be wild-type by WGS, two were mono-resistant to INH, one was mono-resistant to RIF and eight were resistant to RIF+INH. Eleven of the RIF-susceptible group of isolates had mutations compatible with RIF+INH resistance (Figure 1 and Table 2a).

Description of mutations

Resistance to rifampicin: The *rpo*B gene was successfully sequenced in 76 isolates designated as RIF resistant by LJ and LPA. Resistance was confirmed by WGS in all isolates, except one (WT) (Table 2a). In 15 successfully sequenced RIF+INH isolates found to be resistant by LJ but susceptible by LPA, 10 isolates (66.7%) showed mutations in *rpo*B by WGS. Of the total 86 isolates with *rpo*B gene mutations by WGS, amino acid positions 531 (n=47; 54.7%), 526 (n=24; 27.9%) and 516 (n=6; 7.0%) were most commonly affected. WGS also detected five novel mutations (Thr480Ala, Gln253Arg, Val249Met, Val251Tyr; Val251Phe) and one deletion at position 526 (Table 2a).

Resistance to isoniazid: Eighty-four isolates had *kat*G mutations (GP=30, WCP=54) by LPA. INH resistance was mostly due to mutations in the *kat*G gene at position 315 (44; 52.4%). Four novel stop codons mutations in the *kat*G gene were found in this study [Trp477*; Gln88*; Trp198*; and Trp412*] (Table 2a). Three isolates had mixed mutations. Some isolates had double or triple mutations per gene. Six indels were observed in this study. There were 11 wild-type *kat*G isolates. Only two mutations were found in the *inh*A gene and both mutations were novel and contributed to INH resistance. Both these mutations were from the GP isolates (Ser100Ala; Ile200Thr). No mutations were found in the promoter region of the *inh*A gene.

Resistance to pyrazinamide: Thirty-three of the 96 isolates with interpretable results had single mutations in the *pnc*A gene (GP=14, WCP=21) and seven had double mutations (GP=3, WCP=4). Three hetero-resistant (mixed isolates with both resistant and susceptible strains) mutations were detected in the *pnc*A gene (GP: Gln10Xaa; WCP: Thr168Xaa, His71Xaa). We also detected seven indels and seven novel *pnc*A mutations (WCP: Thr100Ile, Thr159Ala, Ala134Arg; GP: Val163Ala, Thr153Ile, Del G pos 7, Phe106Ser). Only one isolate (WCP) had a mutation in the promoter region of the *pnc*A gene (-11A>R) (Table 2b).

Resistance to ethambutol: Of the 96 isolates sequenced (as per Table 2a), a total of 34 isolates had EMB mutations (GP=10, WCP=24). EMB resistance was due to Met306Val and Met306Ile mutations in 21 isolates (GP=4, WCP=17). Two novel mutations were found in WCP isolates (Thr11Xaa and Gln51Pro) (Table 2b).

Resistance to fluoroquinolones (ofloxacin) and injectable aminoglycosides (streptomycin, kanamycin, amikacin): We have observed three isolates with mutations in both the *gyr*A and *rrs* genes (WCP: Ser91Pro with 1484G>T, Ala90Xaa with 1401A>G; GP: Asp94Ala with 1484G>T), which would define XDR-TB in the clinical context. WCP had eight

*rps*L mutations (Lys43Arg) and GP had three (one Lys43Arg and two Lys88Arg). Isolates resistant to SM were not also resistant to AMK or KAN.

Spoligotyping results

Spoligotyping results of all 240 isolates included in this study are provided in Table 3. Of these, 159 (GP=81, WCP=78) clustered into 28 Shared International Types (SIT) (GP=11, WCP=17), with each cluster consisting of two to 17 isolates. The remaining 40 isolates showed unique spoligotype SIT. Forty-one isolates were orphans (GP=18, WCP=23). Most of the isolates belonged to the following six lineages: T, LAM, Beijing, X, S and Harlem. The other minor lineages were MANU2, U, EAI1_SOM and H37Rv (Table 3).

The three largest groups were SIT1 (Beijing lineage), consisting of 34 isolates (GP=17, WCP=17), followed by SIT33 (LAM3) with 21 isolates (GP=11, WCP=10) and SIT53 (ill-defined T sub-lineage) consisting of 19 isolates (GP=10, WCP=9). Other minor groups were SIT34 (S family) (GP=10, WCP=4), SIT92 (X3sub-lineage) (GP=8, WCP=4), SIT60 and SIT811 (LAM4sub-lineage) (GP=6, WCP=1), SIT119 (X1 sub-lineage) (WCP=4) and SIT44 (T5 sub-lineage) (WCP=4). Other less frequent SITs were represented by three isolates, and with a spoligotyping pattern resembling H37Rv. The 24-loci MIRU-VNTR-typing was conducted on 78 isolates that had the largest SIT clusters of nine to 17 isolates (ST1, ST33, ST53 and ST34), representing both GP and WCP. There were 64 unique and seven clustered strains. The clusters of 16 isolates ranged between two to four isolates. Forty-three isolates were split as individual patterns in this study. When combining 24-loci MIRU-VNTR and spoligotyping, there were 26 number types detected with 64 unique isolates, 14 isolates clustered into six clusters. The range of a cluster size was two to four isolates (Table 4).

DISCUSSION

The emergence of drug-resistant tuberculosis has complicated *M.tb* eradication, in particular multidrug-resistant TB. This urgently calls for additional control measures such as characterization of resistant markers, new diagnostic methods, better drugs for treatment, and a more effective vaccine.

In this study, we re-analyzed 240 isolates collected between 1993 and 1995 from the Gauteng and Western Cape Provinces of South Africa that were phenotypically defined as multidrug-resistant. Using current molecular methods in order to retrospectively assess the genotypic characteristics of the isolates, we aimed to confirm the original diagnosis and to specifically define any associated mutations.

Apart from confirming a large proportion of the earlier phenotypic results obtained from solid media susceptibility testing in this study, LPA and NGS methods also revealed several misclassifications in the earlier dataset compared with the WGS results in this study, and additionally identified a few new putative resistance-conferring mutations that have not been reported elsewhere.

Mutations in the *rpo*B (3519 bp) gene were most frequent at positions 531, 526 and 516. These results are in line with other studies in South Africa and other countries in the world that reported the same trends (12). Mutation Ser531Leu, which is common in XDR-TB isolates (13), was also identified in two of the three XDR-TB isolates included in this study. Another mutation described as associated with XDR-TB and pre-XDR-TB (mutation 526), was seen in 24 MDR-TB isolates in our study, but not in the XDR-TB or pre-XDR-TB isolates. In this study, five novel mutations were identified in the *rpo*B gene (Thr480Ala, Gln253Arg, Val249Met, Val251Tyr; Val251Phe), all of which were outside the rifampicin resistance-determining region

(RRDR). Two of these mutations were responsible for RIF resistance, and two (Thr480Ala+Ser531Leu and Gln253Arg+His526Tyr) were paired with mutations within the RRDR of *rpo*B. It is not clear if these two mutations are involved in RIF resistance or if they represent compensatory mutations.

Resistance to INH was mostly due to mutations in the *kat*G gene, particularly at position Ser315Thr. This correlated with published findings from South Africa and elsewhere (12,14). In 96 isolates with phenotypic INH resistance in this study, 14 isolates had mutation Arg463Leu, and this mutation was found as a single mutation in the *kat*G gene in seven of the isolates, suggesting an association with INH resistance. These seven observations applied to GP isolates only. In the WCP isolates, Arg463Leu mutations were also observed in seven isolates, all of which were also associated with Ser315Thr mutations in *kat*G. Thirteen of the 14 isolates with an Arg463Leu mutation were of the Beijing strain type, and one was an X3 strain (GP). These findings seem to support observations from China that Arg463Leu mutations in *kat*G are associated with INH resistance, particularly low-level resistance (15). If that is the case, then the Arg463Leu mutation might need to be taken into consideration as a target when developing rapid molecular diagnostic assays for INH resistance. However, it is possible that the resistance to isoniazid might have originated from the intergenic region between *oxy*R and *ahpC* or between Rv1482c and fabG1 (16).

We identified four novel stop codons, signifying an evolution in *kat*G resistance genetics (Trp477*; Gln88*; Trp198S*and Trp412*P). As for the Arg463Leu mutation, the stop codon mutations are not included in rapid genotyping assays; therefore, they might offer additional discriminatory power if included as targets.

Two novel mutations (Ser100Ala and Ile200Thr) were also identified in the *inh*A gene from INH-resistant GP isolates. The isolates lacked *kat*G mutations, and therefore, these two

mutations could be regarded as contributing to INH resistance. These mutations cannot be identified by the line probe assay since it only detects mutations on the promoter regions (-8, -15 and -16) of the *inh*A gene. No mutations were found in the promoter region of the *inh*A gene in any of the isolates in this study.

PZA mutations were scattered throughout the *pnc*A gene, including its promoter region. We noted seven novel *pnc*A mutations (WCP: Thr100Ile, Thr159Ala, Ala134Gly; GP: Val163Ala, Thr153Ile, Del G pos 7, and Phe106Ser). Amongst these, only one mutation (FPhe106Ser) showed resistance to PZA phenotypically. More functional studies are needed to define the role of the other six mutations in phenotypic PZA resistance. The promoter -11A>R mutation was from a WCP isolate and is known to contribute to PZA's phenotypic resistance. This mutation was first identified in the USA in 2007 (17), and has subsequently been reported in other parts of the world, including South Africa (18, 19). Mutation Leu35Arg, seen in four isolates in this study, was first reported in South Africa and later confirmed in America, where it reportedly contributes significantly to PZA resistance (17-18, 20). Mutation Asp8Asn was found in one isolate in this study, albeit it has been previously identified in South Africa (18) and described as a significant contributor to PZA phenotypic resistance in the country.

Streptomycin mutations detected in this study in the *rps*L gene were Lys43Arg (1 GP, 8 WCP) and Lys88Arg (two isolates). These mutations have previously been shown to be involved in SM resistance (21, 22).

Resistance to EMB was mostly due to the codon 306 of the *emb*B gene (3 GP, 17 WCP). Two novel mutations were identified in codons 11 and 59. Isolates with a mutation in codon 11 contributed to EMB resistance phenotypically. GP had three isolates with Gln497Arg mutations. Mutations in codons 306, 497 and 406 are known to increase EMB's minimum inhibition concentrations (MICs) in *M.tb* (23). One-third of the isolates did not carry mutations in codon

306 (used by rapid molecular test to detect polymorphisms in the codon), but on codons 497 and 406, mutations are not detectable by rapid molecular methods. This explains the low sensitivity and specificity of the line probe assay in the detection of ethambutol resistance. These two mutations were confirmed as hot spot mutation in a study conducted in China (24). The inclusion of these two positions can improve the molecular diagnostic assays.

Isolates resistant to fluoroquinolones (FLQ) had mutations at codons 90, 91 and 94 on the quinolone resistant determining region (QRDR) of the *gyr*A gene, which is implicated in 70-90% of all FLQ-resistance mutations (25, 26). Fifteen isolates (WCP=11, GP=4) had Ser95Thr and Gly668Asp mutations. These mutations do not contribute to resistance, but serve as evolutionary markers for classifying M.tb strains into the three principal genetic groups (PGGs) (27).

Resistance-conferring mutations in the *rrs* gene were due to 1484G>T and 1401A>R in this study. Mutations 1484G>T and 1401A>G cause high-level resistance to AMK, CPM, and KAN (28). Fifteen isolates had a C492T mutation (GP=5, WCP=10); however, this mutation does not confer resistance to AMK, CPM, or KAN (29). In this study, we also found two isolates from WCP with 514A>C and four isolates (2 GP, 2 WCP) with 517C>T mutations. These mutations have previously been reported in other studies and they cause cross-resistance between AMK, CPM, and KAN (30). Other mutations identified in this study were at positions 1260, 1278 and 1445, but also did not seem to influence AMK, CPM, or KAN resistance.

Three of the MDR isolates were XDR-TB, suggesting that XDR-TB was already present in South Africa 10 years ahead of the first XDR-TB cases reported in Kwazulu-Natal province in 2006 (2). In the 1990's however, there was no reason for screening strains routinely for resistance to second-line TB drugs as per the current-day XDR-TB definition, and if these were to be pursued, it would have required culture-based methods to detect phenotypic resistance to the range of drugs available for treating MDR-TB. Such assays were at best available in

specialised reference or research laboratories at the time. Two of these XDR-TB strains were Beijing and one represented the X3 lineage. One Beijing XDR-TB strain harboured mutations in seven out of eight genes tested in this study (*rpo*B, *kat*G, *emb*B, *rps*L, *pnc*A, *gyr*A and *rrs*). This means that the strain was incurable and would not have responded to any antibiotics used to treat drug-resistant TB at the time.

Spoligotyping and 24 MIRU-VNTR were used to describe the diversity of *M.tb* strains that were circulating in the WCP and GP between the years 1993 to 1995. The T, LAM, and Beijing lineages were the largest genotypes that were circulating in both GP and WCP of South Africa and were the driving forces behind drug-resistant TB in the early 1990s. These genotypes were evenly distributed in both provinces.

The predominance of these genotypes in the two South African provinces reported here reflects the genotypes that are widely found in Africa and other parts of the world and concur with the current distribution of genotypes in South Africa. Generally, the even distribution of the Beijing family between the GP and the WCP in this study differs from reports suggesting that the genotype is more commonly found in the WCP than in other provinces in South Africa.

Furthermore, the S and LAM4/F15/KZN genotypes were present in the GP. These two genotypes are known to be dominant in the KZNP. Strain exchange between GP and KZNP is not unexpected, because of traditional migrant labour opportunities in the GP that tended to draw labour from the KZNP to the GP.

MIRU-VNTR 24-loci analysis was specifically performed on a subgroup of strains representing the major spoligotype clusters. MIRU-VNTR further broke down the isolates into seven clusters of 16 isolates. The combination of spoligotying and 24-loci MIRU-VNTR

provided a better resolution of the large spoligotyping clusters, resulting in a final definition of six clusters of 14 isolates (Table 4).

In conclusion, Illumina Miseq WGS showed better concordance with the LJ phenotypic assay than with LPA. WGS detected mutations (*rpo*B=13, *katG*=32, *inh*A=1, *emb*B=10) that were missed by the LPA assay. The majority of these mutations were located outside the "hot spot" region normally detected by the LPA assay. Some of these mutations are novel and have not been reported elsewhere in the literature. It points to a need for confirming discordant results between phenotypic DST and LPA by WGS. However, this study also showed high concordance between LPA and WGS in detecting common mutations, confirming that the LPA assay is a reliable screening assay for drug resistance.

Observations from this study also highlight the diagnostic advantages of WGS over rapid genotyping methods, as most conventional molecular assays are unable to detect mutations outside of known "hot spot" regions. Our findings point to new opportunities for rapid molecular diagnostic test developers to improve on the current range of targets included in current assays.

Specifically, the R463L mutation on the *kat*G gene is not currently included in rapid genotyping assays. Together with the novel mutations in the *pnc*A gene, additional discriminatory power might be achieved if these targets are considered for inclusion in new rapid molecular tests for improved detection of INH- and PZA-resistant *M.tb* strains.

The limitations of the study include the following: WGS was not applied to all 240 isolates in the study to obtain a full picture of drug-resistant patterns in the two provinces because of the cost involved in employing the technology on such an extensive collection of isolates. Also, PZA and second-line drug susceptibility testing was not initially conducted on all the WCP isolates, as opposed to the GP isolates, raising the possibility that results from this

province might provide a biased picture of the mutation pattern. Finally, the cultures from which the genomic DNA was extracted for this study were non-viable as they had only been stored at a moderately low temperature (4ºC) for decades and could not be re-cultured for repeat phenotypic DST.

ETHICS

Ethical approval was obtained from the University of Pretoria Ethics Committee (Ethics code 56/2013)

STUDY SITE

The study was conducted at the laboratories of the South African Medical Research Council (MRC) in Pretoria, the Department of Medical Microbiology of the University of Pretoria, and for whole genome sequencing, at the Research Laboratories of Longhorn Vaccines & Diagnostics, San Antonio, Texas, United States of America. The MIRU-VNTR was conducted and National Institute of Communicable Diseases, TB reference laboratory.

ACKNOWLEDGEMENTS

NEM received PhD study bursaries from the University of Pretoria, the South African Medical Research Council, and the National Research Foundation of South Africa. She was also awarded a Global Infectious Diseases Research Training Fogarty Fellowship in 2013 which enabled access to next-generation sequencing laboratory facilities and training with Drs. Luke T. Daum (Longhorn Vaccines and Diagnostics, San Antonio, Texas, USA), and Dr James P. Chambers (University of Texas at San Antonio, San Antonio, Texas, USA).

REFERENCES

1. World Health Organization. *Global Tuberculosis Report 2015*. World Health Organization, Geneva: 2016.
2. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, Zeller K, Andrews J and Friedland G. 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet **368(9547**):1575-80.
3. Cohen KA , Abeel T, McGuire AM, Desjardins CA, Munsamy V, Shea TP, Walker JB, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY. 2015. Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of *Mycobacterium tuberculosis* Isolates from KwaZulu-Natal. PLOS Medicine DOI:10.1371/journal.pmed.1001880.
4. Loerger TR, Koo S, No E-G, Chen X, Larsen MH, Jacobs Jr WR, Pillay M, Sturm AW, Sacchettini JC. 2009. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. PLoS One **4**: e7778.
5. Cambau E, Viveiros M, Machado D, Raskine L, Ritter C, Tortoli E, Matthys V, Hoffner S, Richter E, Perez Del Molino ML, Cirillo DM. 2014. Revisiting susceptibility testing in MDR-TB by a standardized quantitative phenotypic assessment in a European multicentre study. J Antimicrob Chemother **11**.
6. Gomathi NS, Kumar V.2014. Reliability of Mycobacterial Growth Indicator Tube (MGIT) 960 for the detection of isoniazid resistance in a tuberculosis endemic setting. Indian J Med Res **139(3)**:471-3.
7. Barnard M, Albert H, Coetzee G, O'Brien R and Bosman ME. 2008. Rapid molecular screening for multidrug-resistant tuberculosis in a high-volume public health laboratory in South Africa. Am. J. Respir. Crit. Care Med. 177:787-792.

8. Metzker ML. 2010. Sequencing technologies - the next generation. Nat Rev Genet **11(1)**:31-46.
9. WHO/International Union Against Tuberculosis and Lung Diseases (IUATLD), guidelines for drug suceptibility testing 2001.
10. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, and van Embden J. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol **35(4)**:907-14.
11. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. 2009. Tuberculosis drug resistance mutation database. PLoS Med **6**:e2. (database accessed at http://www.tbdreamdb.com on 18 December 2017).
12. Evans J, Stead MC, Nicol MP, Segal H. 2009. Rapid genotypic assays to identify drug-resistant *Mycobacterium tuberculosis* in South Africa. J Antimicrob Chemother **63(1)**:11-6.
13. Dymova MA, Cherednichenko AG, Alkhovik OI, Khrapov EA, Petrenko TI, Filipenko ML. 2014. Characterization of extensively drug-resistant *Mycobacterium tuberculosis* isolates circulating in Siberia. BMC Infect Dis **14**:478.
14. Daum LT, Rodriguez JD, Worthy SA, Ismail NA, Omar SV, Dreyer AW, Fourie PB, Hoosen AA, Chambers JP and Fischer GW. 2012. Next-generation Ion Torrent sequencing of drug resistance mutations in *Mycobacterium tuberculosis* strains. J Clin Microbiol **50(12)**:3831-7.
15. Zhang M, Yue J, Yang YP, Zhang HM, Lei JQ, Jin RL, Zhang XL, Wang HH. 2005. Detection of mutations associated with isoniazid resistance in *Mycobacterium tuberculosis* isolates from China. J Clin Microbiol **43(11)**:5477-82.
16. Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y. 2013. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet **45**: 1255–60.

17. Somoskovi A, Dormandy J, Parsons LM, Kaswa M, Goh KS Rastogi N, Salfinger M. 2007. Sequencing of the pncA gene in members of the *Mycobacterium tuberculosis* complex has important diagnostic applications: Identification of a species-specific pncA mutation in "Mycobacterium canettii" and the reliable and rapid predictor of pyrazinamide resistance. J Clin Microbiol **45(2)**:595-9.
18. Mphahlele M, Syre H, Valvatne H, Stavrum R, Mannsåker T, Muthivhi T, Weyer K, Fourie PB, Grewal HM. 2008. Pyrazinamide resistance among South African multidrug-resistant *Mycobacterium tuberculosis* isolates. J Clin Microbiol **46(10)**:3459-64.
19. Maningi NE, Daum LT, Rodriguez JD, Mphahlele M, Peters RPH, Fischer GW, Chambers JP, Fourie PB. 2015. Improved detection by next-generation sequencing of pyrazinamide resistance in *Mycobacterium tuberculosis* isolates. J Clin Microbiol **53(12)**:3779-83.
20. Louw GE, Warren RM, Donald PR, Murray MB, Bosman M, van Helden PD, Young DB and Victor TC. 2006. Frequency and implications of Pyrazinamide resistance in managing previously treated tuberculosis patients. Int.J.Tuberc.Lung Dis **10(7)**:802-7
21. Nair J, Rouse DA, Bai GH, Morris SL. 1993. The rpsL gene and streptomycin resistance in single and multiple drug-resistant strains of *Mycobacterium tuberculosis*. Mol Microbiol **10(3)**:521-7.
22. Brossier F, Veziris N, Aubry A, Jarlier V, Sougakoff W. 2010. Detection by GenoType MTBDRsl test of complex mechanisms of resistance to second-line drugs and ethambutol in multidrug-resistant Mycobacterium tuberculosis complex isolates. J Clin Microbiol **48(5)**:1683-9.
23. Safi H, Fleischmann RD, Peterson SN, Jones MB, Jarrahi B, Alland D. 2010. Allelic exchange and mutant selection demonstrate that common clinical *emb*CAB gene mutations only modestly increase resistance to ethambutol in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother **54(1)**:103-8.

24. Shi W, Zhang X, Jiang X, Yuan H, Lee JC, Barry CE III, Wang H, Zhang W, Zhang Y. 2011. Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. Science **333**:1630–1632

25. Antonova OV, Gryadunov DA, Lapa SA, Kuz'min AV, Larionova EE, Smirnova TG, Nosova EY, Skotnikova OI, Chernousova LN, Moroz AM, Zasedatelev AS. 2008. Detection of mutations in *Mycobacterium tuberculosis* genome determining resistance to fluoroquinolones by hybridization on biological microchips. Bull Exp Biol Med **145(1)**:108-13.

26. Hoshide M, Qian L, Rodrigues C, Warren R, Victor T, Evasco HB 2nd, Tupasi T, Crudu V, Douglas JT. 2014. Geographical differences associated with single-nucleotide polymorphisms (SNPs) in nine gene targets among resistant clinical isolates of Mycobacterium tuberculosis. J Clin Microbiol **52(5)**:1322-9.

27. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A **94(18)**:9869-74.

28. Du Q, Dai G, Long Q, Yu X, Dong L, Huang H, Xie J. 2013. *Mycobacterium tuberculosis* rrs A1401G mutation correlates with high-level resistance to kanamycin, amikacin, and capreomycin in clinical isolates from mainland China. Diag Microb and Infecs Dis **77**:138-142

29. Engström A, Perskvist N, Werngren J, Hoffner SE, Jureen P. 2011. Comparison of clinical isolates and in vitro selected mutants reveals that tlyA is not a sensitive genetic marker for capreomycin resistance in *Mycobacterium tuberculosis*. J. Antimicrob. Chemother. **66**:1247–1254

30. Jugheli L, Bzekalava N, de Rijk P, Fissette K, Portaels F, Rigouts L. 2009. High level of cross-resistance between kanamycin, amikacin, and capreomycin among *Mycobacterium tuberculosis* isolates from Georgia and a close relation with mutations in the rrs gene. Antimicrob Agents Chemother **53(12)**:5064-8.

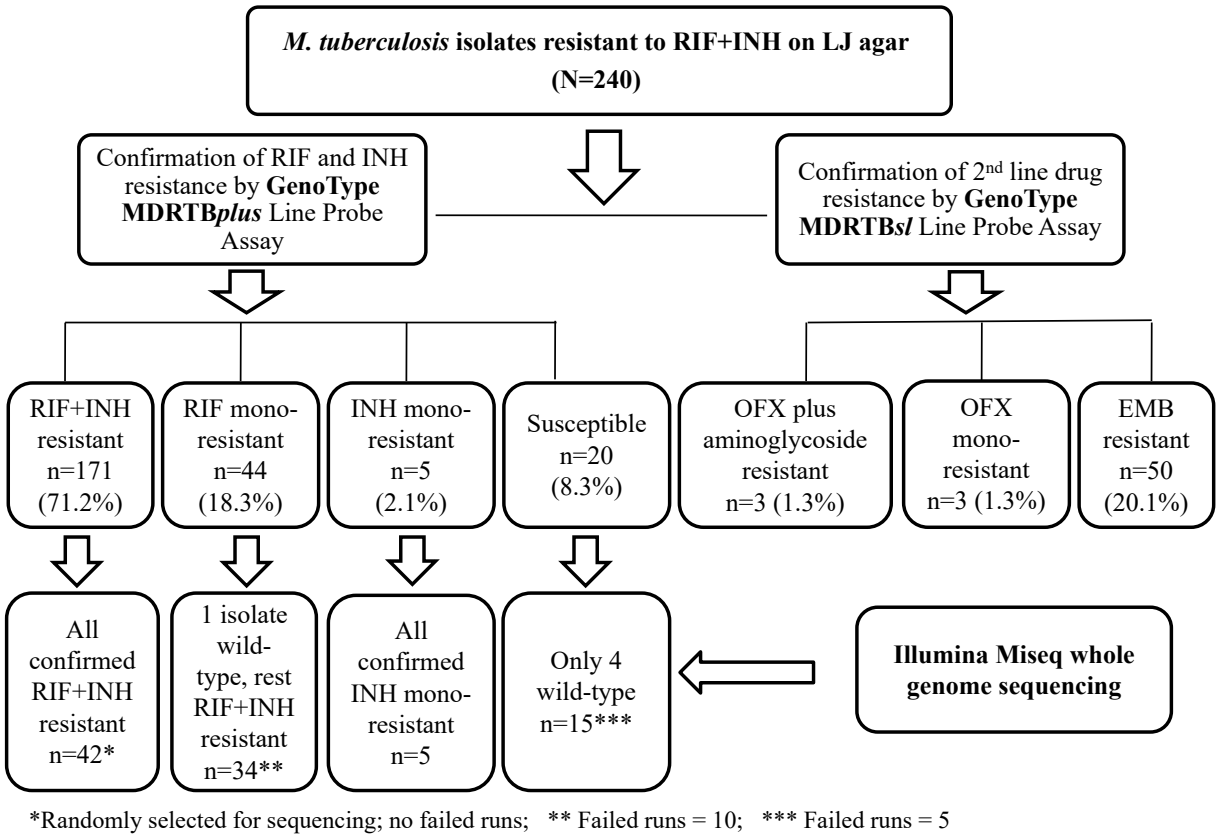

Figure 1: Confirmation of drug resistance status by line probe assay and whole-genome sequencing of 240 *Mycobacterium tuberculosis* isolates originally designated as multi-drug resistant in 1993 by proportion method drug susceptibility testing in Löwenstein-Jensen agar medium

Table 1: Comparison of resistance profiles by (1) original phenotypic drug susceptibility testing in 1993-95; and (2) Genotype MTBDR*plus* or MTBDR*sl* on extracted DNA in 2014

Description*	**Gauteng Province (n=120)**	**Western Cape Province (n=120)**
Phenotypic MDR-TB profile (1993-1995)		
RIF	120 (100%)	120 (100%)
INH	120 (100%)	120 (100%)
PZA	42 (35%)	0
EMB	7 (6%)	35 (29%)
ETM	14 (12%)	45 (37%)
SM	17 (14%)	45 (37%)
OFX	1 (<1%)	1 (<1%)
KAN	1 (<1%)	0
Genotypic resistance profile (Genotype MTBDRplus or MTBDRsl LPA) on extracted DNA (2014)		
RIF + INH	73 (61%)	98 (82%)
RIF mono	30 (25%)	14 (12%)
INH mono	3 (2%)	2 (2%)
EMB	11 (9%)	39 (34%)
OFX + AMG**	1 (<1%)	2 (2%)
OFX mono**	1 (<1%)	2 (2%)
Susceptible	14 (12%)	6 (5%)

*RIF=rifampicin; INH=isoniazid; EMB=ethambutol; PZA=pyrazinamide; ETM=ethionamide; SM=streptomycin; OFX=ofloxacin; KAN=kanamycin; AMG=aminoglycoside

**n=117 for Gauteng Province

Table 2a. MTBDR*plus* line-probe assay (LPA) results and associated mutations observed by Illumina MiSeq whole genome sequencing (WGS) for confirmation of rifampicin-isoniazid resistance as originally defined by LJ culture in 20-year-old *M. tuberculosis* isolates (n=240)

LPA results		Number selected for WGS	Description of mutations by WGS**		Number of isolates
Category	Number tested		*rpo*B	*Kat*G	
RIF+INH resistant	171 (71.3%)	42/171*	Ser531Leu	Ser315Thr	15
			His526Try	Ser315Thr	7
			Ser531Leu	Ser315Thr, Arg463Leu	3
			Ser526Asp/TyrSTOP	Ser315Thr	2
				Ser315Ile	1
			Ser531Leu	Ser315Thr/SerSTOP	1
			Asp516Val	Ser315Thr	1
			Ser531Thr	Ser315Asn	1
			Asp516Val	DelCpos526, Ser315Thr	1
			Thr525Pro	Ser315Ile	1
			Asp516Val	Ser315Asn	1
			His526Asp	Ser315Asn	1
			Leu533Pro	Trp91Arg, Arg463Leu	1
			His526Leu	Ser315Thr	1
			Thr480Ala,	Ser315Arg	1
			Ser531Leu	Ser315Thr, Arg463Leu,	1
			His526Tyr	Ala464Leu	1
			Ser531Leu	Trp91Arg, Ser140Gly,	1
			His526Tyr	Arg463Leu	1
			Gln409Arg,	Arg463Leu	
			Ser450Leu	**Ile200Thr** (*inh*A gene)	
			His 526 Tyr		
RIF mono-resistant	44 (18.3%)	44 (Sequencing failed in 10 isolates, leaving 34 with WGS results***)	His531Leu	No mutations	4
			His531Leu	Arg463Leu	2
			His531Leu	Del2185G	1
			His531Leu	Pro131Arg, Arg463Leu	1
			His531Leu	Tyr98Cys	1
			His531Leu	Ser315Thr	1
			His531Leu	Asp381Glu	1
			His531Leu	Pro232Arg	1
			His531Leu	Asp1426Arg	1
			His531Leu	del 1426G (Missense)	1
			His531Leu	Pro29Ser, Met420Thr	1
			His531Leu	Pro232Ala	1
			His531Leu	**Ser100Ala** (*inh*A gene)	1
			His526Tyr	Arg463Leu	2
			His526Tyr	Ins19A	1
			His526Tyr	Phe594Ser	1
			His526Tyr	**Gln88STOP**	1
			His526Asp	Gly699Glu	1
			His526Asp	Leu378Arg	1
			Ser531Trp	Ala122Gly/Ala	1
			Ser531Trp	**Trp477STOP**	1
			Ser522Leu	**Trp412STOP**	1

			Asp516Val	Arg463Leu	1
			Asp516Val	Ins576A	1
			Gln253Arg,	No mutations	1
			His256Arg	**Trp198STOP**	1
			His526Arg	Gly699GLU	1
			Leu511Pro/LeuSTOP	Arg463Leu, **Trp477STOP**	1
			DelC526	Arg463Leu, Del entire codon pos 480	1
			Wild type		
INH mono-resistant	5 (2%)	5	**Val251Phe**	Thr380Ile	2
			Ser522Leu	Ser315Thr, Arg463Leu	1
			Val249Met, Val251Tyr	Ser315Thr, Arg463Leu	1
			Ser531Leu	Ser315Thr	1
RIF + INH Susceptible	20 (8.3%)	20 (Sequencing failed in 5 isolates, leaving 15 with WGS results***)	Wild type	Wild type	4
			Ser531Leu	No mutations	2
			No mutations	Ser315Thr, W198STOP	1
			His526Leu/His (mixed)	Ser315Thr	1
				Pro232Ala	1
			Ser531Leu	Glu553Lys	1
			Ser531Leu	Gly299Ser	1
			His526Leu/Ala	Arg104Trp	1
			Ile572Phe/IleSTOP	DelGpos1284	1
			Asp516Val	Gly299Arg/GlySTOP	1
			Ile572Phe/IleSTOP	Arg463Leu	1
			Ser531Leu/SerSTOP		
Total	240	111			96

*Randomly selected; ** Novel mutations present by WGS shown in bold type; *** DNA concentration too low for WGS

Table 3: Distribution of study isolates per the various shared-types and corresponding lineages/sub-lineages as defined by spoligotyping for the Gauteng and Western Cape Provinces in South Africa (N=240)

Lineage	No of lineages (N=240)	Gauteng Province	Western Cape Province
T	58	25	33
LAM	43	29	14
Beijing	34	17	17
X	28	9	19
S	19	11	8
H	8	3	5
MANU2	3	0	3
U	4	0	4
EAI1_SOM	1	1	0
H37Rv	1	0	1
Orphans	41	18	23
TOTAL	**240**	**113**	**127**

Table 4. Type, unique isolates, clustered isolates, cluster and cluster range by spoligotyping and MIRU-VNTR analysis

Method	No. of types	No. of unique isolates	No. of clustered isolates	No of clusters	Cluster size range
Spoligotyping	4	0	78	4	2 to 17
24-loci MIRU-VNTR	26	62	16	7	2 to 4
Spoligotyping + MIRU-VNTR	26	64	14	6	2 to 4