HOKKAIDO UNIVERSITY

Title	First insight into the genetic population structure of Mycobacterium tuberculosis isolated from pulmonary tuberculosis patients in Egypt
Author(s)	Diab, Hassan Mahmoud; Nakajima, Chie; Kotb, Saber A. et al.
Citation	Tuberculosis, 96, 13-20 https://doi.org/10.1016/j.tube.2015.11.002
Issue Date	2016-01
Doc URL	https://hdl.handle.net/2115/68587
Rights	©2015, Elsevier. This manuscript version is made available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/
Rights(URL)	https://creativecommons.org/licenses/by-nc-nd/4.0/
Type	journal article
File Information	MS_Diab-1st_TUBE_2015_252_Main text_2015.10.30_Submitted to Tuberculosis-2.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

First insight into the genetic population structure of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis patients in Egypt

Hassan Mahmoud Diab [a, b], Chie Nakajima [b, c], Saber A. Kotb [d], Alaa Mokhtar [e], Nagwa F.M. Khder[f], Ahmed S.A. Abdelaal [f], Azza Hegazy [f], Ajay Poudel [b], Yogendra Shah [b] and Yasuhiko Suzuki [b, c, *].

[a] Department of Animal Hygiene, Faculty of Veterinary Medicine, South Valley University, Qena, Egypt

[b] Division of Bioresources, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan

[c] Hokkaido University The Global station for Zoonosis Control, Sapporo, Japan

[d] Department of Animal Hygiene, Faculty of Veterinary Medicine, Assiut University, Egypt

[e] National Tuberculosis Control Program, Ministry of Health and Population, Egypt

[f] TB Supranational Reference Laboratory, Central Public Health Laboratories, Clinical Microbiology Department, Ministry of Health and Population, Egypt

* Correspondence to: Yasuhiko Suzuki, Division of Bioresources, Hokkaido University Research Center for Zoonosis Control, Kita 20-Nishi 10, Kita-ku, Sapporo 001-0020, Japan

Tel.: +81 11 706 9503; fax: +81 11 706 7310.

E-mail address: suzuki@czc.hokudai.ac.jp

ABSTRACT

The present study aimed to assess the population structure of *Mycobacterium tuberculosis* (MTB) isolates from Egypt. A total of 230 MTB isolates were analysed using spoligotyping, large sequence polymorphism (LSPs), mycobacterial interspersed repetitive unit–variable number tandem repeat (MIRU-VNTR) typing and multi-locus sequence typing (MLST). The majority of isolates (93.0%) belonged to lineage 4, including 44.3, 13.4 and 10.8% of the ill-defined T clade, LAM and Haarlem families, respectively, and lineage 3 was identified in 7.0 % of the isolates. MIRU-VNTRs typing allowed efficient discrimination of the spoligotype-defined clusters, including spoligo-international types (SIT) 53, 34, and 4, into 56 patterns, including 13 clusters and 43 unique patterns. A new SNP at position 311614 was identified in all six isolates to form the biggest MIRU-VNTR cluster, which suggested a recent clonal expansion. This SNP could possibly be used as a genetic marker for robust discriminations of Egyptian MTB isolates belonging to SIT53. The combination of spoligotyping, 12 MIRU-VNTRs loci and MLST provided insight into the genetic diversity and transmission dynamics of the Egyptian MTB genotypes and could be a key to implementation of effective control measures by public health authorities.

Keywords: Genotyping, MTB Lineages, Egypt.

1. Introduction

Tuberculosis (TB) remains a major global health problem, causing 9 million new cases and 1.5 million deaths in 2013. Most of TB cases are reported from Asia (56%) and Africa (29%), whilst only a small number of cases (8%) occurred in the Eastern Mediterranean Region (EMR) [1]. Egypt is a transcontinental, economically diversified middle-income country with 82 million inhabitants. It is bordered by the Mediterranean Sea to the north, and categorized as an EMR member of the World Health Organization. In 2013, Egypt ranked amongst mid-level TB-burdened countries, with reported prevalence and incidence rates (includes HIV+TB) of 27 and 16 per 100,000 inhabitants, respectively [1].

Over the last decades, many genotyping tools such as spoligotyping [2], mycobacterial interspersed repetitive unit–variable number tandem repeat (MIRU-VNTR) typing [3] and multilocus sequence typing (MLST) [4,5] have provided insights into the genetic diversity, transmission dynamics and phylogenetic analysis of the *Mycobacterium tuberculosis* complex (MTBC) lineages, including biomedical and epidemiological characteristics specific to each phylogenetic lineage strain [6-8].

The MTBC associated to human infections consists of 7 main lineages, each defined by lineage-specific markers. Lineage 3 includes the spoligotype-defined Central Asian (CAS) family, which belongs to PGG1 and spreads mainly across North India and East Africa. Lineage 4 is characterized by the deletion of spacers 33–36 in the direct repeat (DR) locus and classified into 5 clades (T, Haarlem, LAM, S and X). It is a modern phylogenetic group belonging to the principle genetic group 2-3 (PGG2-3) and is the prevalent lineage in Europe and the Americas, although it has spread across different regions of Africa and the Middle East [9-11].

The few investigations that have addressed the genetic diversity of MTBC lineages in Egypt have indicated the predominance of lineage 4 amongst Egyptian isolates [12-14]. Nonetheless, the epidemiology and molecular characterisation of MTBC strains in Egypt remain largely unknown. The current study aimed to use molecular genotyping tools such as spoligotyping, MIRU-VNTR

and MLST typing for elucidating the genetic diversity and characterising the prevalent lineages of *Mycobacterium tuberculosis* (MTB) strains infecting Egyptian patients. It also aimed to contribute better understanding of the population structure, phylogenetic relations and the transmission dynamics of MTB strains in Egypt and surrounding EMR countries.

2. Material and Methods

2.1. Samples collection

During the period 2012-2014, MTB isolates from Egyptian patients were processed and cultured either on site or at intermediate laboratories and sent to the TB Supranational Reference Laboratory (TB-SNRL), Central Public Health Laboratories, Ministry of Health and Population, Egypt. From this MTB isolate pool, 230 were randomly selected for the present study. In addition, patients' epidemiological and demographic data were obtained from medical records at TB-SNRL (Table 1). These patients were from 16 Egyptian governorates, namely, Cairo, Alexandria, El Giza, El Fayoum, El Dakahlia, El Gharbia, El Minya, Suez, Kafr elsheik, Damietta, El Behira, Helwan, El Sharqia, El Ismailia , Qena and North Sinai (Figure 1).

2.2. Sample processing and MTBC identification

Sputum samples were decontaminated and cultured on Löwenstein–Jensen medium (LJ). The clinical isolates were identified as members of the MTBC by biochemical and cultural characteristics. DNA was extracted by the boiling method. Afterwards, EDTA was added at the final concentration of 1 mM, and kept at -20 °C until used for further molecular analysis.

To confirm the species as a member of the MTBC, *rpoB* gene sequencing and comparison with the type strain (H37Rv) sequence were performed according to the procedure described by Poudel *et al.* (2012) [15].

2.3. Spoligotyping

All MTB isolates were analysed by spoligotyping, as reported by Kamerbeek *et al* (1997) [2]. Briefly, a PCR-based reverse-hybridisation technique was used in which the DR region was

amplified with a primer pair. The PCR products were hybridized to a set of 43 oligonucleotide probes corresponding to each spacer and covalently bound to the membranes.

2.4. LSPs

Isolates with orphan spoligotype patterns were identified with a PCR-based method using specific primers for the expected Regions of Difference (RD) for each lineage, as carried out by Gagneux *et al* (2006) [9].

2.5. MIRU-VNTR typing

MIRU-VNTR typing was conducted according to Supply *et al* (2006) [3], and the number of repeats for each locus was determined. MIRU-VNTR loci, namely, MIRU (10, 16, 26 and 40), Mtub (4, 21, 30 and 39), ETR-A and QUB (11b, 26 and 4156) were selected from amongst spoligotype-defined clusters SIT53 (n= 59), SIT34 (n=15) and SIT4 (n=11) for MIRU-VNTR typing after a comprehensive literature review of strains belonging to lineage 4 [16-21].

2.6. MLST

Classification of isolates into PGG2 or PGG3 was carried out based on polymorphism at *gyrA*95 codon, as reported by Sreevatsan *et al* (1997) [4]. Detailed procedures for PCR amplification of *gyrA* gene and sequence comparison were according to Poudel *et al* (2013) [22]. Differentiation of the isolates belonging to PPG3-SNP cluster group 6 (SCG -6) into SCG-6a and SCG-6b was carried out by sequencing based on the SNP at position 311611 (T/G) of the H37Rv chromosome [5]. PCR primers (311611-Fw: CCTGCACAGTGCGGTCGACG and 311611-Rv: GTTCAAAGCAGCCGGCCACG) were designed to amplify a 400 bp product including the SNP position. PCR and sequencing procedures were same as Poudel *et al* (2013) [22] with a primer set described above.

2.7. Data management and analysis

Major MTBC clade/subclade spoligotype signatures and Spoligotyping International Type numbers (SIT) were assigned according to the SITVIT WEB database [23] and are available online at http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/. Hunter–Gaston Diversity Index

(HGDI) was used to determine the diversity of each MIRU–VNTR locus and the discriminatory power of the typing schemes [24]. A cluster was defined as a group of two of more isolates sharing the same spoligotype or MIRU-VNTR pattern, and the clustering was calculated as a ratio of clustered isolates. Phylogenetic analysis and the drawing of dendrograms showing isolates clustering were generated based on a clustering algorithm by the hierarchic unweighted pair group method analysis (UPGMA) using an online MIRU-VNTR*plus* database application available at: http://www.miru-vntrplus.org/ [25].

3. Results

3.1. Sociodemographic data analysis

Most of isolates (83%) were recovered from treated pulmonary TB patients, and the majority (68%) collected from Cairo and Alexandria. Other epidemiological data are shown in Table 1.

3.2. Spoligotyping and LSPs

The distribution of lineages/clades of MTB strains across Egyptian governorates are shown in Table 2 and Supplementary Figure 1. All MTB isolates exhibited 76 spoligotype patterns, with 203 isolates belonging to 55 shared types and 27 showing orphan spoligotype~~s~~ patterns. Only two lineages, lineage 4 and lineages 3, were identified. The predominant lineage was lineage 4 with 214 isolates, while lineage 3 was detected in 16 isolates. The overall repartition of strains according to major phylogenetic MTB clades was carried out based on signatures found in the SITVIT WEB database: ill-defined T clade 102/230 (44.3%) > LAM 31/230 (13.4%) > Haarlem 25/230 (10.8%) > S clades 17/230 (7.4%) and CAS1_Delhi 13/230 (5.6%). About 40% of all typed isolates were found to be associated with SIT53, SIT34, SIT4, SIT25 and SIT50. LSPs showed that most orphan strains belonged to lineage 4, and that only two were attributed to lineage 3. Phylogenetic analysis revealed the existence of 27 clusters, the largest comprising 59 strains displaying SIT53 spoligotype patterns, followed by two of 15 and 11 strains displaying SIT34 and SIT4 spoligotype patterns, respectively.

3.3. MIRU-VNTR result analysis

We subjected SIT53, SIT34 and SIT4 to MIRU-VNTR typing using the 12 selected loci. Eight specimens showed two or more MIRU-VNTR bands either in a single locus or several loci (excluded from the phylogenetic analysis), but only 77 isolates exhibiting single MIRU-VNTR band/locus were considered for further analysis. Phylogenetic analysis revealed that, of the 52 isolates belonging to spoligotype-defined cluster SIT53, 27 fitted into 10 clusters and the remaining 25 exhibited unique MIRU-VNTR patterns with a clustering rate of 52%. Out of the 15 isolates belonging to SIT34, 7 were grouped into 3 clusters, and 8 showed different patterns with a clustering rate of 47%. All isolates of the SIT4/unknown clade yielded diverse MIRU-VNTR patterns (Figure 2). Amongst the 13 MIRU-VNTR clusters obtained, the biggest cluster consisted of six isolates belonging to the SIT53/T1 clade isolated from patients from the governorate EL Gharbia.

Individual and cumulative HGDI, detailed allelic diversity and the clustering rate for each MIRU-VNTR locus are shown in Table 3 and Supplementary Table 1. The results obtained from the analysis of SIT53T1, SIT34/S and SIT4 revealed that QUB-26, ETR-A and Mtub39 showed high discriminatory power (HGDI >0.6), whilst of the remaining loci, seven had moderate (0.3≤ HGDI ≤0.6) and two had poor (HGDI<0.3) discriminatory power. In SIT53, Mtub39, QUB-26 and QUB-11b yielded high discriminatory power, whilst 2 and 7 loci showed moderate and poor discriminatory power, respectively. The overall discriminatory power of the MIRU-VNTR typing was HGDI= 0.9376 in all analysed isolates versus HGDI = 0.9145 in SIT53.

3.4. MLST

A total of 52, 15 and 10 isolates belonging to SIT53, SIT34 and SIT4, respectively, were subjected to SNP analysis of codon 95 in *gyrA*. The results showed that three isolates belonging to SIT53 yielded overlapped peaks and hence, were excluded from further analysis. SIT53 was differentiated into 47 AGC (Ser) and 2 ACC (Thr) corresponding to PGG3 and PGG2, respectively (Figure 2A), whilst all isolates belonging to SIT34 and SIT4 were classified as PGG2 (Figure 2B).

SNP typing at position 311611 subdivided isolates belonging to SIT53-SCG6 into 46 isolates with T (SCG6a) and 1 with G (SCG6b) [5]. A previously unknown SNP at position 311614 subdivided isolates belonging to SIT53-SCG6a into 2 sub-clusters, 40 isolates with G and 6 with C (Figure 2A).

4. Discussion

4.1. Spoligotyping and LSPs data interpretation

The results of the present work indicated that the majority of TB patients (93%) were infected with strains belonging to lineage 4, which reflected the actual predominance of this lineage in Egypt. Lineage 4 belongs to modern PGG2-3 and is known to be prevalent in Europe and America. Comparative analysis of prevalence of these strain families between Egypt and neighbouring EMR and other North African countries [12-14, 17, 26-31] highlights an important observation that certain strain families seem to be prevalent and widely spread in defined geographical areas. For example, Ill-defined T clade was endemic in Egypt and Syria, the Harlem family in Tunisia and the LAM family in Morocco. This observation implies that MTB exhibits clonal and phylogeographical population structures that importantly defines the geographical distribution of relevant strain-specific phenotypic traits including differences in virulence and immunogenicity [7]. Detailed comparative data are shown in Table 4.

Previous studies conducted in Egypt have described the MTB lineages and spoligotypes circulating amongst the local population (Table 4). Cooksey *et al.*, (2002) analysed 67 clinical isolates from cerebrospinal fluid (CSF) of meningitis patients during the period 1998-2000 [12], and Abbadi *et al.,* (2009) analysed a total of 45 MTBC strains isolated from the sputum of pulmonary TB patients in the Suez Canal region of Egypt [13]. Both studies found lineage 4 to be predominant. Our results on pulmonary tuberculosis are in agreement with the above-mentioned findings regarding the prevalence of that lineage over time and geographical distribution throughout Egypt. In addition, Helal *et al.,* (2009) examined a total of 151 MTB strains from patients collected

over a period of one year (2005) at three different chest hospitals in Egypt [14]. However, their study showed a high proportion of the ancestral MANU genotype of MTB (27.2%), which contrasts with our work, as we were unable to identify any isolate belonging to MANU clade. The cause of this discrepancy maybe due to MANU, a well-known spoligotype clade, can be constructed with a mixture of more than two spoligotypes [32]. According to those studies, the predominance of lineage 4 and the endemicity of T clade in Egyptian MTB isolates over a period of time (15 year or more) indicate that these strains are strongly established, adapted and associated with the host.

According to our results, strains belonging to lineage 3 were found to be accounting for 7% of TB morbidity in Egypt (Table 4). It was also found that about 5.6% of total isolates belonged to CAS1_Delhi. An important factor affecting the prevalence of the CAS/CAS1_Delhi family in Egyptian TB patients is that Egypt has deeply rooted geographical, historical and social relations with many African and Asian countries like Sudan, Saudi Arabia and Iraq, from which previous studies reported the predominance of CAS/CAS1_Delhi at 49% [26], 22.5% [27] and 24% [30]**,** respectively.

4.2. MIRU-VNTR and clustering analysis

A total of 85 isolates belonging to the largest spoligotype-defined clusters SIT53, SIT34 and SIT4 were subjected to further analysis using the selected 12 loci MIRU-VNTR scheme. The SIT53 cluster was efficiently discriminated and subdivided into 35 MIRU-VNTR patterns. The largest MIRU-VNTR-defined cluster 315422212352/SIT53 comprised of six isolates, which, according to the epidemiological data, were all recovered from patients from the governorate of EL Gharbia. Furthermore, the second largest cluster, 325422231452/SIT53, included three isolates and also recovered from patients from the same governorate. It is worth mentioning that MTB strains circulating in this governorate were seemingly transmitted from patient to patient. In contrast, all isolates belonging to the spoligotype-defined SIT4 were completely discriminated from each other and exhibited diverse MIRU-VNTR patterns. This observation can be explained by the fact that strains belonging to the SIT4/unknown clade might have independently evolved by acquiring

genetic diversity over a long period of time that resulted in successively established strains stably associating with the Egyptian population [9].

Comparative analyses were performed between our data and those reported from others countries based on MIRU-VNTR typing of MTB strains belonging to SIT53 (Table 5) [18-21, 33]. Locus Mtub21 was highly conserved in Egyptian MTB strains belonging to SIT53 and showed poor discrimination with HGDI = 0.0384 in the same manner as the isolates belonging to SIT34 and SIT4 with one repeat unit. In contrast, the same locus was variable and had high and moderate discriminatory power in the MTB isolates from other countries (Table 5). This result has prompted for the selection and optimisation of suitable MIRU-VNTR typing schemes to achieve better characterizations, discriminations and phylogenetic examinations of MTB strains based on local characteristics in each geographical region. The current study is the first to optimise suitable MIRU-VNTR schemes for genotyping of lineage 4, which is endemic in Egypt. According to the results of the discriminatory power of the 12 selected loci shown in Table 3, the MIRU-VNTR scheme that includes QUB-26, ETR-A, Mtub39, QUB-11b, MIRU16, MIRU10, Mtub04, Mtub21, and MIRU40 is proposed for the analysis of isolates belonging to lineage 4, including SIT53, SIT34 and SIT4. For the prevalent SIT53, the MIRU-VNTR set that includes loci Mtub39, QUB-26, QUB-11b, ETR-A, MIRU40, MIRU26, MIRU10 and MIRU16 is recommended for efficient discrimination.

4.3. MLST and PGG phylogenetic analysis.

Combined data from spoligotyping, MIRU-VNTR and MLST indicate that isolates determined as SIT53 were highly diverse. Two of them were belonging to a genetically distinct group, PGG2, although majority were PGG3 (Figure 2). Among SIT53-PGG3 (defined as SCG6 in Filliol *et al*'s SNP cluster group [5]), vast majority were SCG6a, however, one isolate was revealed as SCG6b. Similar observations on the high diversity of SIT53 have been reported [5, 10, 11] and this may be attributed to the definition of the spoligotype. SIT53 is defined as "all spacer present except 33-36" [34, 35], i.e. the prototypic form of the lineage 4 including ill-defined T clade. Thus,

this classical, worldwide prevalent spoligotype strains could have higher variety in their genome than other spoligotype strains emerged later. SIT53 is also known to be produced by a mixed infection with different genotypes as a false spoligotype [20, 32]; however, we carefully checked and excluded the samples showed multiple bands in MIRU-VNTR analysis to eliminate the possible misidentification. The poor discrimination power of spoligotyping for ill-defined T clade should be compensated with MIRU-VNTR and/or MLST, especially in the area where the majority of the isolates show the same spoligotype like SIT53 in Egypt.

The isolates belonging to the biggest MIRU-VNTR cluster, 315422212352/SIT53-PGG3-SCG6a, showed the same characteristic SNP C at position 311614 (Figure 2). This cluster consisted of six isolates recovered from patients from the governorate of EL Gharbia, which suggested that this clone seems to be actively circulating amongst the population of that region.

5. Conclusions

Our analysis demonstrated that the MTB population structure exclusively belonged to the modern evolutionary phylogenetic group PGG2-3. Lineage 4 was predominant, and T family was the most endemic genotype among MTB strains isolated from Egypt in 2012-2014. The genotyping protocol established in the current study will contribute to efficient characterizations, discriminations and phylogenetic clusters examinations based on local characteristics of MTB strains circulating in Egypt. In addition, based on the results from the present study, analysis of more representative number of isolates from Egypt is recommended to further elucidate the prevalence, evolutionary history and actual distribution of newly found SNP at position 311614 in Egyptian MTB isolates and ascertain the possible use of this SNP as a genetic marker for rapid and robust discriminations of MTB isolates belonging to PGG3/SCG-6. This additional study could contribute to the implementation of more effective control measures by Egyptian public health authorities, which may help prevent spread of TB infection.

Acknowledgements

This work was supported in part by the Japan Initiative for Global Research Network on Infectious Diseases from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (MEXT); in part by a grant for the Establishment of International Collaboration Centers for Zoonosis Control, Hokkaido University from MEXT; in part by a grant for the Joint Research Program of the Research Center for Zoonosis Control, Hokkaido University from MEXT.

References

[1] WHO. Global tuberculosis report. Geneva, Switzerland: World Health Organization; 2014. WHO/HTM/TB/2014.08.

[2] Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.

[3] Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 2006;44:4498–510. doi:10.1128/JCM.01392-06.

[4] Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 1997;94:9869–74. doi:10.1073/pnas.94.18.9869.

[5] Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbo MH, Bobadilla M, et al. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and

recommendations for a minimal standard SNP set. J Bacteriol 2006;188:759–72. doi: 10.1128/ JB.188. 2.759

[6] Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN. Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 2006;19:658–85. doi:10.1128/CMR. 00061-05.

[7] Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect Dis 2007;7:328–37. doi:10.1016/S1473-3099(07)70108-1.

[8] Bouakaze C, Keyser C, De Martino SJ, Sougakoff W, Veziris N, Dabernat H, et al. Identification and genotyping of *Mycobacterium tuberculosis complex* species by use of a SNaPshot minisequencing-based assay. J Clin Microbiol 2010;48:1758–66. doi:10.1128/ JCM.02255-09.

[9] Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 2006;103:2869–73. doi:10.1073/pnas.0511240103.

[10] Rindi L, Lari N, Garzelli C. Large Sequence Polymorphisms of the Euro-American lineage of *Mycobacterium tuberculosis*: A phylogenetic reconstruction and evidence for convergent evolution in the DR locus. Infect Genet Evol 2012;12:1551–7. doi:10.1016/j.meegid.2012. 06.008.

[11] Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj S a, et al. *Mycobacterium tuberculosis complex* genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 2006;6:23. doi:10.1186/1471-2180-6-23.

[12] Cooksey RC, Abbadi SH, Woodley CL, Sikes D, Wasfy M, Crawford JT, et al. Characterization of *Mycobacterium tuberculosis complex* isolates from the cerebrospinal fluid of meningitis patients at six fever hospitals in Egypt. J Clin Microbiol 2002;40:1651–5. doi:10.1128/JCM.40.5.1651-1655.2002.

[13] Abbadi S, El Hadidy G, Gomaa N, Cooksey R. Strain differentiation of *Mycobacterium tuberculosis complex* isolated from sputum of pulmonary tuberculosis patients. Int J Infect Dis

2009;13:236–42. doi:10.1016/j.ijid.2008.06.020.

[14] Helal ZH, Ashour MSED, Eissa S a., Abd-Elatef G, Zozio T, Babapoor S, et al. Unexpectedly high proportion of ancestral manu genotype *Mycobacterium tuberculosis* strains cultured from tuberculosis patients in Egypt. J Clin Microbiol 2009;47:2794–801. doi:10.1128/JCM. 00360-09.

[15] Poudel A, Nakajima C, Fukushima Y, Suzuki H, Pandey BD, Maharjan B, Suzuki Y. Molecular characterization of multidrug-resistant Mycobacterium tuberculosis isolated in Nepal. Antimicrob Agents Chemother. 2012;56:2831-6. doi: 10.1128/AAC.06418-11.

[16] Asante-Poku A, Nyaho MS, Borrell S, Comas I, Gagneux S, Yeboah-Manu D. Evaluation of customised lineage-specific sets of MIRU-VNTR loci for genotyping *Mycobacterium tuberculosis complex* isolates in Ghana. PLoS One 2014;9. doi:10.1371/journal.pone. 0092675.

[17] Lahlou O, Millet J, Chaoui I, Sabouni R, Filali-Maltouf A, Akrim M, et al. The Genotypic Population Structure of *Mycobacterium tuberculosis* Complex from Moroccan Patients Reveals a Predominance of Euro-American Lineages. PLoS One 2012;7:1–10. doi:10.1371/journal.pone.0047113.

[18] Rindi L, Medici C, Bimbi N, Buzzigoli A, Lari N, Garzelli C. Genomic Variability of *Mycobacterium tuberculosis* Strains of the Euro-American Lineage Based on Large Sequence Deletions and 15-Locus MIRU-VNTR Polymorphism. PLoS One 2014;9:e107150. doi:10.1371/journal.pone.0107150.

[19] Mulenga C, Shamputa IC, Mwakazanga D, Kapata N, Portaels F, Rigouts L. Diversity of *Mycobacterium tuberculosis* genotypes circulating in Ndola, Zambia. BMC Infect Dis 2010;10:177. doi:10.1186/1471-2334-10-177.

[20] Stavrum R, Mphahlele M, Øvreås K, Muthivhi T, Fourie PB, Weyer K, et al. High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates. J Clin Microbiol 2009;47:1848–56. doi:10.1128/JCM. 02167-08.

[21] Vasconcellos SEG, Acosta CC, Gomes LL, Conceição EC, Lima KV, de Araujo MI, et al. Strain Classification of *Mycobacterium tuberculosis* Isolates in Brazil Based on Genotypes

352 Obtained by Spoligotyping, Mycobacterial Interspersed Repetitive Unit Typing and the Presence of
353 Large Sequence and Single Nucleotide Polymorphism. PLoS One 2014;9:e107747.
354 doi:10.1371/journal.pone.0107747.
355 [22] Poudel A, Maharjan B, Nakajima C, Fukushima Y, Pandey BD, Beneke A, Suzuki Y.
356 Characterization of extensively drug-resistant Mycobacterium tuberculosis in Nepal. Tuberculosis
357 (Edinb). 2013;93:84-8. doi: 10.1016/j.tube.2012.10.007.
358 [23] Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. SITVITWEB - A publicly
359 available international multimarker database for studying *Mycobacterium tuberculosis* genetic
360 diversity and molecular epidemiology. Infect Genet Evol 2012;12:755–66.
361 doi:10.1016/j.meegid.2012.02.004.
362 [24] Hunter PR, Gaston M a. Numerical index of the discriminatory ability of typing systems: an
363 Application of Simpson's Index of Diversity. J Clin Microbiol 1988;26:2465–6.
364 doi:0095-1137/88/112465-0202.00/0.
365 [25] Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: A web tool for
366 polyphasic genotyping of *Mycobacterium tuberculosis complex* bacteria. Nucleic Acids Res
367 2010;38:326–31. doi:10.1093/nar/gkq351.
368 [26] Sharaf Eldin GS, Fadl-Elmula I, Ali MS, Ali AB, Salih ALG, Mallard K, et al. Tuberculosis in
369 Sudan: a study of *Mycobacterium tuberculosis* strain genotype and susceptibility to
370 anti-tuberculosis drugs. BMC Infect Dis 2011;11:219. doi:10.1186/1471-2334-11-219.
371 [27] Al-Hajoj S a M, Zozio T, Al-Rabiah F, Mohammad V, Al-Nasser M, Sola C, et al. First insight
372 into the population structure of *Mycobacterium tuberculosis* in Saudi Arabia. J Clin Microbiol
373 2007;45:2467–73. doi:10.1128/JCM.02293-06.
374 [28] Varghese B, Supply P, Allix-Béguec C, Shoukri M, Al-Omari R, Herbawi M, et al. Admixed
375 Phylogenetic Distribution of Drug Resistant *Mycobacterium tuberculosis* in Saudi Arabia. PLoS
376 One 2013;8:1–8. doi:10.1371/journal.pone.0055598.
377 [29] Bedrossian N, Rahmo A, Karam W, Hamze M. *Mycobacterium tuberculosis* spoligotypes

circulating in the Syrian population: A retrospective study. Int J Mycobacteriology 2013;2:141–7. doi:10.1016/j.ijmyco.2013.05.004.

[30] Merza M a., Salih AM. First insight into the genetic diversity of *Mycobacterium tuberculosis* strains from patients in Duhok, Iraq. Int J Mycobacteriology 2012;1:13–20. doi:10.1016/ j.ijmyco.2012.01.003.

[31] Namouchi A, Karboul A, Mhenni B, Khabouchi N, Haltiti R, Hassine R Ben, et al. Genetic profiling of *Mycobacterium tuberculosis* in Tunisia: Predominance and evidence for the establishment of a few genotypes. J Med Microbiol 2008;57:864–72. doi:10.1099/jmm. 0.47483-0.

[32] Lazzarini LCO, Rosenfeld J, Huard RC, Hill V, Lapa e Silva JR, DeSalle R, et al. *Mycobacterium tuberculosis* spoligotypes that may derive from mixed strain infections are revealed by a novel computational approach. Infect Genet Evol 2012;12:798–806. doi:10.1016/j.meegid.2011.08.028.

[33] Ouassa T, Borroni E, Loukou GY, Faye-Kette H, Kouakou J, Menan H, et al. High Prevalence of Shared International Type 53 among *Mycobacterium tuberculosis Complex* Strains in Retreated Patients from Côte d'Ivoire. PLoS One 2012;7:1–7. doi:10.1371/journal. pone.0045363.

[34] Filliol I, Driscoll JR, Van Soolingen D, Kreiswirth BN, Kremer K, Valétudie G, et al. Global distribution of Mycobacterium tuberculosis spoligotypes. Emerg Infect Dis. 2002;8:1347-9.

[35] Sebban M, Mokrousov I, Rastogi N, Sola C. A data-mining approach to spacer oligonucleotide typing of Mycobacterium tuberculosis. Bioinformatics. 2002;18:235-43.

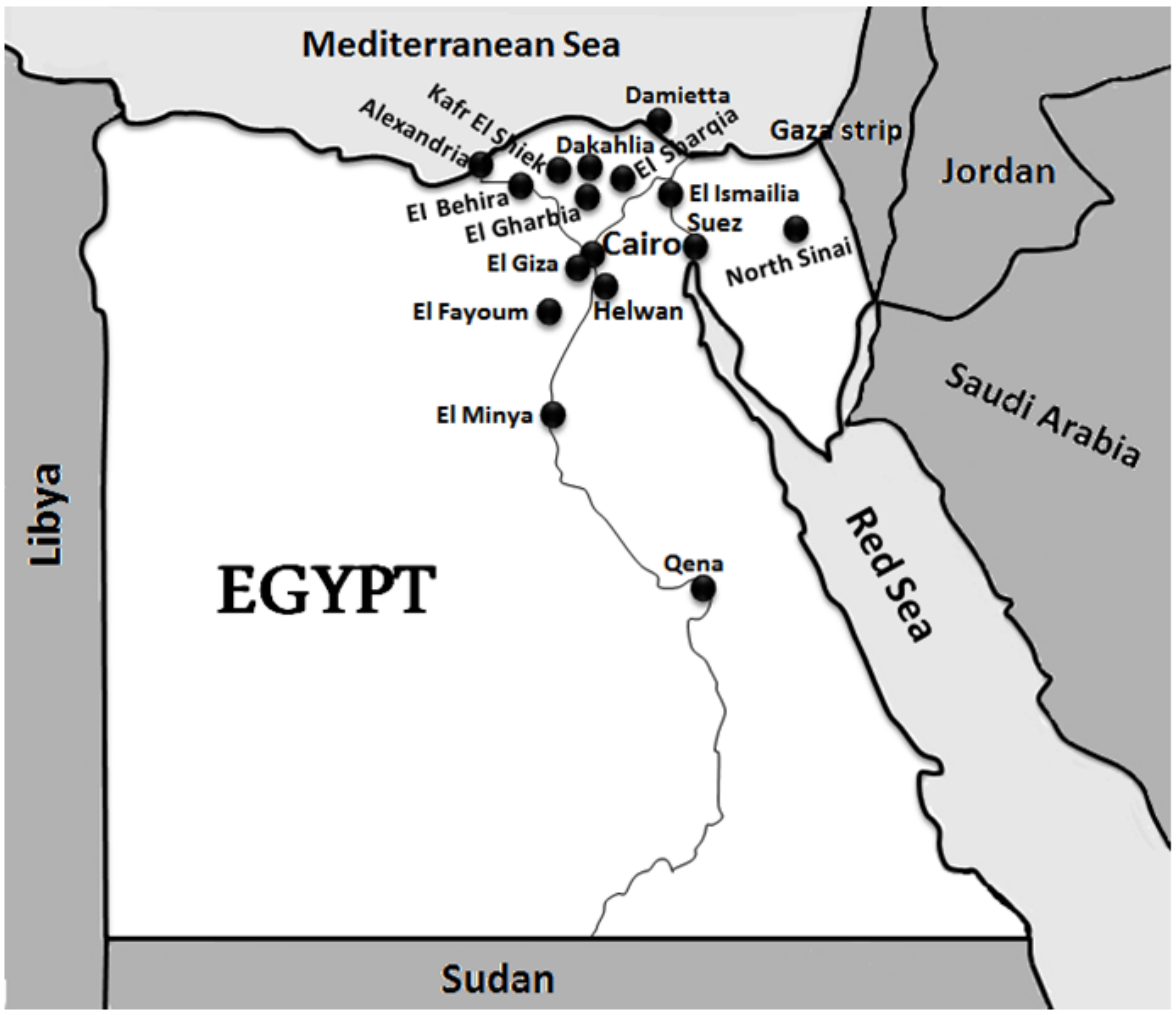


Fig 1. Geographical distribution of *M. tuberculosis* isolates from the 16 Egyptian governorates used in the present study.

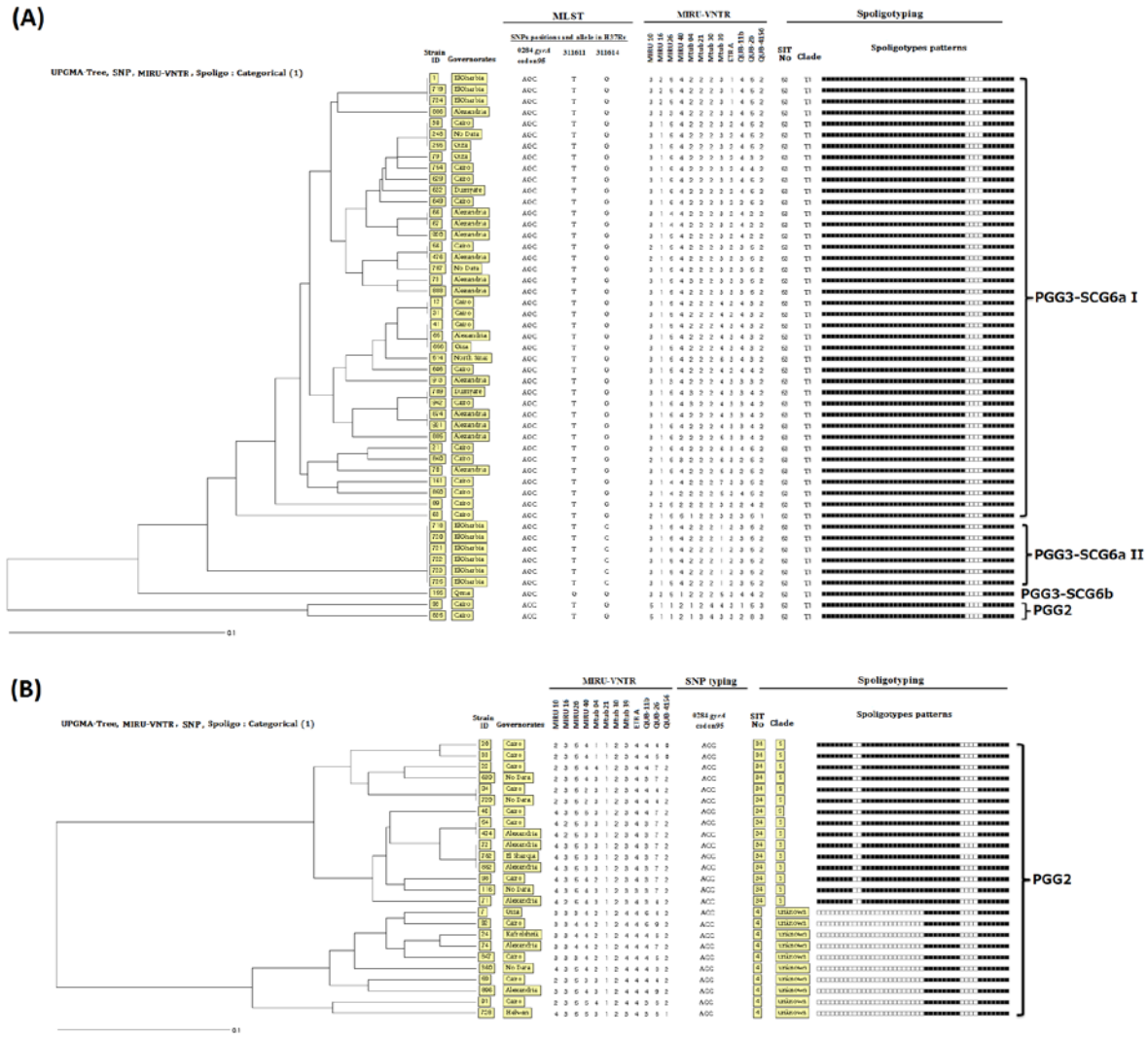


Figure 2. UPGMA-tree dendrograms. **(A)** Phylogenetic analysis based on Multi-Locus Sequence Typing (MLST), 12 MIRU-VNTR loci and spoligotyping patterns of *Mycobacterium tuberculosis* isolates representing the largest spoligotype-defined cluster SIT53/T1. **(B)** Strains diversity based on combined results of 12 MIRU-VNTR loci, SNP typing, and spoligotyping analysis of isolates representing the spoligotype-defined clusters included (SIT34/S and SIT4/unknown clade). The categorical-based UPGMA tree was generated by an MIRU-VNTR*plus* database application available online (http://www.miru-vntrplus.org/).

Table 1. Patient epidemiological and demographic data of the present study

Variants		Number	Ratio
Sex	Male	169	73
	Female	45	20
	No data	16	7
Governorates (patients origin)	Cairo	112	49
	Alexandria	44	19
	Others	60	26
	No data	14	6
Treatment history	Treated cases	190	83
	New cases	24	10
	No data	16	7

Table 2. Frequency distribution of major lineages/subclades of *M. tuberculosis* in Egyptian isolates.

Global lineage/Subclades*			No. of isolates	% of isolates
Lineage 1		-	0	0
Lineage 2		-	0	0
Lineage 3	CAS		14	6.1
	Orphan strains		2	0.87
Lineage 4	T superfamily	T1	89	38.7
		T2	2	0.87
		T3	6	2.6
		T4	5	2.2
	Haarlem		1	0.43
		H1	5	2.2
		H3	19	8.3
	LAM	LAM	7	3
		LAM1	1	0.43
		LAM3	2	0.87
		LAM5	1	0.43
		LAM6	1	0.43
		LAM7_TUR	8	3.5
		LAM9	9	3.9
		LAM10_CAM	2	0.87
	S		17	7.4
	Orphan strains		25	10.87
	Unknown		14	6.1
Lineage 5		-	0	0
Lineage 6		-	0	0
Lineage 7		-	0	0
Total			230	100

*Lineages classifications based on [9], [11] and [23].

Table 3 Individual and cumulative HGDI of MIRU-VNTR loci in [SIT53T1 + SIT34/S + SIT4/Unknown] versus independent SIT53/T1.

SIT53T1 + SIT34/S + SIT4/unknown clade [a]									
locus	Alias	No. of patterns	No. of clusters	No. of clustered isolates	Size of clusters	Unique patterns	Clustering rate (%)	Individual HGDI [c]	Cumulative HGDI [c]
4052	QUB-26								
2165	ETR-A	16	13	74	2 to 16	3	96.1	0.6992	0.9056
3690	Mtub39	30	19	66	2 to 9	11	85.7	0.6651	0.9586
2163b	QUB-11b	43	16	50	2 to 7	27	64.9	0.5971	0.9761
1644	MIRU16	45	16	48	2 to 6	29	62.3	0.5386	0.9798
960	MIRU10	49	14	42	2 to 6	35	54.5	0.5311	0.9829
424	Mtub04	54	14	37	2 to 6	40	48.1	0.4788	0.9870
1955	Mtub21	54	14	37	2 to 6	40	48.1	0.4617	0.9870
802	MIRU40	55	13	35	2 to 6	42	45.5	0.4392	0.9874
2996	MIRU26	56	13	34	2 to 6	43	44.2	0.3014	0.9880
4156	QUB-4156	56	13	34	2 to 6	43	44.2	0.1729	0.9880
2401	Mtub30	56	13	34	2 to 6	43	44.2	0.0513	0.9880
SIT53/T1									
3690	Mtub39							0.7375	0.7376
4052	QUB-26	15	11	48	2 to 15	4	92.3	0.6305	0.8824
2163b	QUB-11b	23	9	38	2 to 8	14	73.1	0.6041	0.9397
2165	ETR-A	27	13	38	2 to 6	14	73.1	0.5618	0.9661
802	MIRU40	29	12	35	2 to 6	17	67.3	0.3115	0.9691
2996	MIRU26	31	11	32	2 to 6	20	61.5	0.2805	0.9713
960	MIRU10	33	10	29	2 to 6	23	55.8	0.2745	0.9736
1644	MIRU16	33	10	29	2 to 6	23	55.8	0.2149	0.9736
424	Mtub04	35	10	27	2 to 6	25	51.9	0.2149	0.9774
4156	QUB-4156	35	10	27	2 to 6	25	51.9	0.1478	0.9774
2401	Mtub30	35	10	27	2 to 6	25	51.9	0.0754	0.9774
1955	Mtub21	35	10	27	2 to 6	25	51.9	0.0385	0.9774

[a] SIT: Spoligotyping International Type number.

[b] : MIRU-VNTR: Mycobacterial Interspersed Repetitive Unit–Variable Number Tandem Repeat.

[c] : HGDI: Hunter–Gaston Diversity Index.

Table 4. Summary of *M. tuberculosis* lineage/clade distributions from previous studies in Egypt and other EMR[†] and North African countries

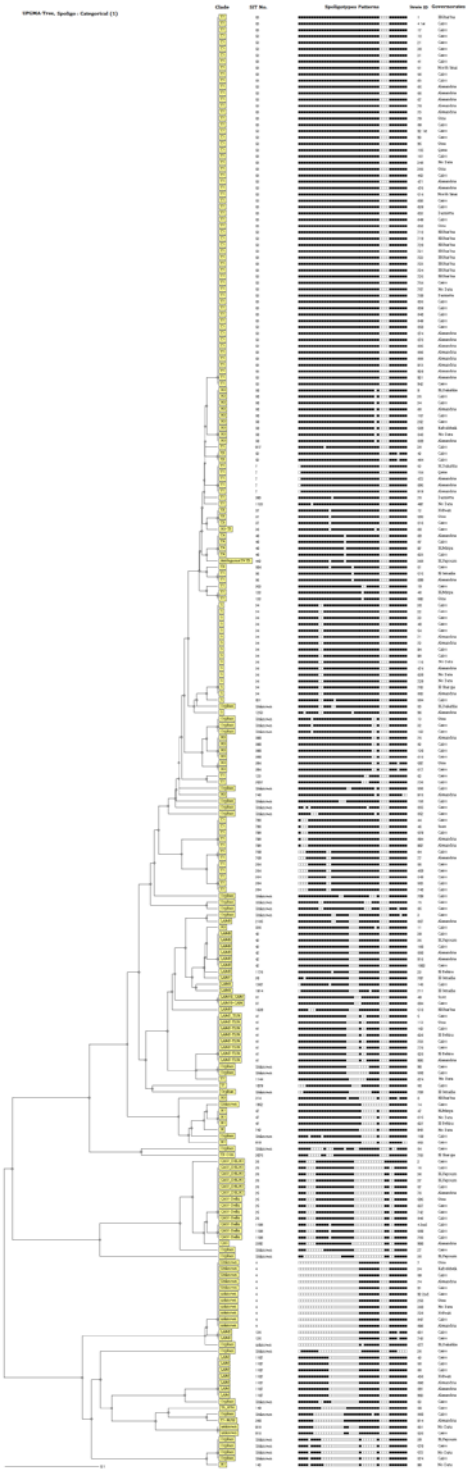
Lineages/Clades[*]		Country of isolation (%)										
		Egypt				Sudan	Saudi Arabia		Syria	Iraq	Tunisia	Morocco
		Current study 2015	Cooksey et al (2002) [12]	Abbadi et al (2009) [13]	Helal et al (2009) [14]	Sharaf et al (2011) [26]	Al-Hajoj et al (2007) [27]	Varghese et al (2013) [28]	Bedrossian et al (2013) [29]	Merza and Salih (2012) [30]	Namouchi et al (2008) [31]	Lahlou et al (2012) [17]
		N= 230	N= 67	N= 44	N= 151	N= 235	N= 1,505	N= 322	N= 96	N= 53	N= 378	N= 592
Lineage 1	Manu	0	0	0	27.2	0	2.7	0	0	0	0	0
Lineage 3	CAS	7	13.6	9	4	50.7	22.5	21.1	10.4	24	0	0
Lineage 4	T	44.4	34.8	50	53.6	8.2	19.5	0	47.8	30	16.4	20.1
	Haarlem	10.9	1.5	4.5	2	3	7.5	10.6	13.4	18	40.2	22.6
	LAM	13. 5	15.2	18.2	4.6	2.2	7.2	7.5	22.8	0	28.3	43.7
	S	7.4	6.1	4.5	1.3	0.4	0	2.45	0	14	0	3.8

[†] Eastern Mediterranean Region.
N: number of isolates
[*]Lineages classifications based on [9], [11] and [23].

Table 5 Comparison of data from the HGDI analysis based on 12 MIRU-VNTR loci used in the present study with those reported by other researchers in MTB strains belonging to SIT53/T1 [a].

MIRU-VNTR / HGDI [b, c]			Country					
			Egypt 2015 N = 52	Italy Rindi *et al* (2014) [19] N = 13	Zambia Mulenga *et al* (2010) [20] N = 16	South Africa Stavrum *et al* (2009) [21] N = 13	Brazil Vasconcellos *et al* (2014) [22] N = 26	Coˆ te d'Ivoire Ouassa *et al* (2012) [33] N = 74
MIRU-VNTR	3690	Mtub39	0.7375	0.7051	0.5809	0.8333	0.7169	0.0270
	4052	QUB-26	0.6305	0.5385	0.6912	0.7179	0.8338	0.0796
	2163b	QUB-11b	0.6041	0.7308	0.4853	0.7692	0.8031	0.1292
	2165	ETR-A	0.5618	0.5128	0.5294	0.3846	0.5815	0.0270
	802	MIRU40	0.3115	0.7308	0.75	0.6795	0.8092	0.4838
	2996	MIRU26	0.2805	0.7564	0.3235	0.5385	0.6092	0.0796
	960	MIRU10	0.2745	0.5897	0.5	0.6026	0.7969	0.0270
	1644	MIRU16	0.2149	0.8205	0.5294	0.5385	0.6246	0
	424	Mtub04	0.2149	0.7308	0.2279	0.5	0.6831	0.2217
	4156	QUB-4156	0.1478	0.1538	0.3824	0.4103	0.6092	0.0270
	2401	Mtub30	0.0754	0	0.4706	0.3846	0.7108	0.2455
	1955	Mtub21	0.0385	0.1538	0.4853	0.5256	0.6769	0.1292

[a] SIT: Spoligotyping International Type number.
[b] MIRU-VNTR: Mycobacterial Interspersed Repetitive Unit–Variable Number Tandem Repeat.
[c] HGDI: Hunter–Gaston Diversity Index.
N: number of isolates



Supplementary Figure 1. UPGMA tree dendrogram showing spoligotyping patterns and clustering analysis of the 230 *Mycobacterium tuberculosis* isolates from Egypt. The categorical-based UPGMA tree was generated by an MIRU-VNTR*plus* database application available online at (http://www.miru-vntrplus.org/).

Supplementary Table 1 Frequency distribution of MTB isolates [SIT53T1 + SIT34/S + SIT4/Unknown] versus independent SIT53/T1 in the selected 12 MIRU-VNTR loci.

			SIT53/T1 + SIT34/S + SIT4/unknown clade *									
MIRU -VNTR †		No of isolates	Number of repeats									
locus	Alias	(N=77)	0	1	2	3	4	5	6	7	8	9
960	MIRU10	77			14	50	11	2				
1644	MIRU16	77		46	6	25						
2996	MIRU26	77		2		4	7	64				
802	MIRU40	77		1	7	7	57	3	1	1		
424	Mtub04	77		5	53	17	2					
1955	Mtub21	77		25	51	1						
2401	Mtub30	77			75		2					
3690	Mtub39	77		6	2	39	21	5	3	1		
2165	ETR-A	77		3	26	24	24					
2163b	QUB-11b	77		1	5	31	38	2				
4052	QUB-26	77			3	9	16	31	3	11	2	2
4156	QUB-4156	77	2	3	70	2						
			SIT53/T1									
MIRU -VNTR †		No of isolates	Number of repeats									
locus	Alias	(N=77)	0	1	2	3	4	5	6	7	8	9
960	MIRU10	52			6	44		2				
1644	MIRU16	52		46	3	3						
2996	MIRU26	52		2		2	4	44				
802	MIRU40	52		1	5	1	43	1		1		
424	Mtub04	52		3	46	3						
1955	Mtub21	52			51	1						
2401	Mtub30	52			50		2					
3690	Mtub39	52		6	2	23	12	5	3	1		
2165	ETR-A	52		3	26	23						
2163b	QUB-11b	52		1	5	20	26					
4052	QUB-26	52			3	9	10	29			1	
4156	QUB-4156	52		2	48	2						

[a] SIT: Spoligotyping International Type number.
[b] : MIRU-VNTR: Mycobacterial Interspersed Repetitive Unit–Variable Number Tandem Repeat.
[c] : HGDI: Hunter–Gaston Diversity Index.