



Strain Classification of Mycobacterium tuberculosis Isolates in Brazil Based on Genotypes Obtained by Spoligotyping, Mycobacterial Interspersed Repetitive Unit Typing and the Presence of Large Sequence and Single Nucleotide Polymorphism

Sidra E. G. Vasconcellos, Chyntia Carolina Acosta, Lia Lima Gomes, Emilyn Costa Conceicao, Karla Valéria Lima, Marcelo Ivens de Araujo, Maria de Lourdes Leite, Flavio Tannure, Paulo Cesar de Souza Caldas, Harrison M Gomes, et al.

► To cite this version:

Sidra E. G. Vasconcellos, Chyntia Carolina Acosta, Lia Lima Gomes, Emilyn Costa Conceicao, Karla Valéria Lima, et al.. Strain Classification of Mycobacterium tuberculosis Isolates in Brazil Based on Genotypes Obtained by Spoligotyping, Mycobacterial Interspersed Repetitive Unit Typing and the Presence of Large Sequence and Single Nucleotide Polymorphism. PLoS ONE, 2014, 9 (10), pp.e107747. 10.1371/journal.pone.0107747 . pasteur-01153012

HAL Id: pasteur-01153012

<https://riip.hal.science/pasteur-01153012>

Submitted on 19 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Strain Classification of *Mycobacterium tuberculosis* Isolates in Brazil Based on Genotypes Obtained by Spoligotyping, Mycobacterial Interspersed Repetitive Unit Typing and the Presence of Large Sequence and Single Nucleotide Polymorphism

Sidra E. G. Vasconcellos^{1,8}, Chyntia Carolina Acosta², Lia Lima Gomes¹, Emilyn Costa Conceição³, Karla Valéria Lima³, Marcelo Ivens de Araujo¹, Maria de Lourdes Leite⁴, Flávio Tannure⁴, Paulo Cesar de Souza Caldas⁵, Harrison M. Gomes¹, Adalberto Rezende Santos¹, Michel K. Gomgnimbou⁶, Christophe Sola⁶, David Couvin⁷, Nalin Rastogi⁷, Neio Boechat^{8,9}, Philip Noel Suffys^{1*}

1 Laboratory of Molecular Biology Applied to Mycobacteria, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Rio de Janeiro, Brazil, **2** Laboratory of Cellular Microbiology, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Rio de Janeiro, Brazil, **3** Instituto Evandro Chagas, Section of Bacteriology and Mycology, Belém, Pará, Brazil, **4** Hospital Municipal Rafael de Paula Souza, Municipal Secretary of Health, Rio de Janeiro, Rio de Janeiro, Brazil, **5** Centro de Referência Professor Hélio Fraga, Escola Nacional de Saúde Pública Sergio Arouca, FIOCRUZ, Rio de Janeiro, Rio de Janeiro, Brazil, **6** CNRS–Université Paris–Sud, Institut de Génétique et Microbiologie–Infection Genetics Emerging Pathogens Evolution Team, Orsay, France, **7** Supranational TB Reference Laboratory, Unité de la Tuberculose et des Mycobactéries, Institut Pasteur de Guadeloupe, Abymes, Guadeloupe, France, **8** Multidisciplinary Research Laboratory, University Hospital Clementino Fraga Filho – HUCFF, Federal University of Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil, **9** Graduate Program in Clinical Medicine, Faculty of Medicine, University Hospital Clementino Fraga Filho, Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Rio de Janeiro is endemic for tuberculosis (TB) and presents the second largest prevalence of the disease in Brazil. Here, we present the bacterial population structure of 218 isolates of *Mycobacterium tuberculosis*, derived from 186 patients that were diagnosed between January 2008 and December 2009. Genotypes were generated by means of spoligotyping, 24 MIRU-VNTR typing and presence of *fbpC*¹⁰³, RD^{Rio} and RD174. The results confirmed earlier data that predominant genotypes in Rio de Janeiro are those of the Euro American Lineages (99%). However, we observed differences between the classification by spoligotyping when comparing to that of 24 MIRU-VNTR typing, being respectively 43.6% vs. 62.4% of LAM, 34.9% vs. 9.6% of T and 18.3% vs. 21.5% of Haarlem. Among isolates classified as LAM by MIRU typing, 28.0% did not present the characteristic spoligotype profile with absence of spacers 21 to 24 and 32 to 36 and we designated these conveniently as “LAM-like”, 79.3% of these presenting the LAM-specific SNP *fbpC*¹⁰³. The frequency of RD^{Rio} and RD174 in the LAM strains, as defined both by spoligotyping and 24 MIRU-VNTR loci, were respectively 11% and 15.4%, demonstrating that RD174 is not always a marker for LAM/RD^{Rio} strains. We conclude that, although spoligotyping alone is a tool for classification of strains of the Euro-American lineage, when combined with MIRU-VNTRs, SNPs and RD typing, it leads to a much better understanding of the bacterial population structure and phylogenetic relationships among strains of *M. tuberculosis* in regions with high incidence of TB.

Citation: Vasconcellos SEG, Acosta CC, Gomes LL, Conceição EC, Lima KV, et al. (2014) Strain Classification of *Mycobacterium tuberculosis* Isolates in Brazil Based on Genotypes Obtained by Spoligotyping, Mycobacterial Interspersed Repetitive Unit Typing and the Presence of Large Sequence and Single Nucleotide Polymorphism. PLoS ONE 9(10): e107747. doi:10.1371/journal.pone.0107747

Editor: Igor Mokrousov, St. Petersburg Pasteur Institute, Russian Federation

Received: January 1, 2014; **Accepted:** August 21, 2014; **Published:** October 14, 2014

Copyright: © 2014 Vasconcellos et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by funding by Brazilian funding agencies PAPES/CNPq (407624/2012-0), PROEP/CNPq of the Oswaldo Cruz Institute (IOC), and CNPq PhD grant (142958/2008-5). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: psuffys@gmail.com

Introduction

Tuberculosis (TB) is an infectious disease with an effective treatment but remains an important cause of morbidity and mortality in many countries. In Brazil, 73,000 new TB cases and 4,600 cases that deceased were registered in 2011. The southeast region has 44.8% of all cases reported in the country and the state of Rio de Janeiro has the highest disease incidence (72.3/100,000) and mortality rate (5.7/100,000) in the country [1].

A couple of years ago, *Mycobacterium tuberculosis* (Mtb) strains designated as RD^{Rio} were reported as being a predominant genotype in Rio de Janeiro [2]. These strains have a deletion of 26.3 kb and seem to be restricted to the Latin American-Mediterranean (LAM) family. The RD^{Rio} strains have been associated with higher levels of recent transmission and of Multi-Drug Resistance (MDR) but there are contradictory data about their relation with disease severity [3], [4], [5], [6]. Another Region of Difference called RD174 has been described as a co-

marker of RD^{Rio} and as a marker for the LAM type [7], [8], [9], [10] and strains with the RD174 deletion were found to have an increased secondary case rate ratio in San Francisco [10].

The most commonly used genotyping methods to characterize *M. tuberculosis* Complex (MTBC) isolates are IS6110-RFLP, 24-locus MIRU-VNTR, spoligotyping, detection of Single Nucleotide Polymorphisms (SNPs) and Large Sequence Polymorphisms (LSP). Typing of MTBC by IS6110-RFLP was considered the gold-standard during more than a decade [11], but is labor intensive. Currently, the two most commonly used methods are spoligotyping and Mycobacterium Interspersed Repetitive Units – Variable Number of Tandem Repeats (MIRU-VNTR) typing, based on the variability of the direct repeat (DR) locus [12] and of minisatellites [13], [14], respectively. However, for epidemiologic studies, spoligotyping overestimates the proportion of clustered strains and should be used in combination with MIRU-VNTR typing [15]. Spoligotyping alone is also unable to reveal exact phylogenetic relationships between MTBC strains, particularly for the classification of Euro-American strains, part of the PGG2/PGG3 group [9], [16], [17], [18]. These phylogenetic subtypes include the genotype families LAM, H, T, S and X, frequently observed in Brazil, in South America and elsewhere [2], [4], [8], [18], [19], [20], [21], [22], [23].

The generalized use of spoligotyping and MIRU-VNTR resulted in the construction of large international databases of genotypes, allowing the study of global distribution and phylogenetic analysis of the distribution of *M. tuberculosis* worldwide [23], [24], [25], [26], [27], [28], [29]. Combined MIRU-VNTR typing and spoligotyping also helps in revealing epidemiologically meaningful clonal diversity of *M. tuberculosis* strain lineages and is useful to explore internal phylogenetic ramifications [8]. In addition, SNPs and LSPs represent robust markers for inferring phylogenies and for strain classification [30], [31], [32], [33], [34], [35]. Besides the presence of RDs, that of determinate SNPs is also a characteristic of LAM strains, [8], [9], [10] such as *fbpC*¹⁰³ in codon 103 (G to A) of the gene encoding antigen 85 Complex Ag85^c (Rv0129c) resulting in a Glu103Asp amino acid replacement in a protein that is involved in biosynthesis of cell wall components of *M. tuberculosis* [7], [36]. Other LAM-specific SNPs are (i) the silent C to G *ligB* mutation at genome position 3426795 [37], validated by the Abadia *et al.* [17] and (ii) the SNP at 240 codon of *mgIC* (C to T) that was discovered by Homolka *et al.* [38].

The objective of this study was to evaluate 24 MIRU-VNTR typing for alternative classification of strain families/lineages by means of phylogenetic tree building and comparison with MIRU databases (MIRU-VNTRPlus), as compared with spoligotypes-based classification (SITVITWEB and SITVIT2), and increase the consistency of the analysis by including a SNP and two LSPs. In addition, the differentiating power of both techniques in a population that is almost exclusively of the Euro-American lineages.

Materials and Methods

Study setting

The study was based on a convenience sampling of TB patients diagnosed between January 2008 and December 2009 at the “Hospital Municipal Raphael de Paula e Sousa”, Curicica, Rio de Janeiro, Brazil. Two hundred eighteen clinical isolates were obtained from 188 patients and 28 cases had more than one isolate, including 25 patients with two and three patients with three isolates. Multiple isolates from the same patient were

included to evaluate the consistency of the results of genotyping and to verify eventual multiple infections.

Ethics statement

This study was approved by the Research Ethics Committee of the Municipal Health and Civil Defense of the city of Rio de Janeiro Number 160/09 CAAE: 0182.0.314.000-9. Isolates of this study were obtained through bacteriological culture from clinical specimens of patients diagnosed at the Hospital Raphael and Paula Souza as part of routine diagnosis and drug susceptibility testing. The ethical committee stated that, as genotyping of *M. tuberculosis* isolates is complementary to routine diagnosis, there was no need for a written or verbal consent. No information other than that provided for the diagnosis was used.

Culture, DNA extraction and identification of *Mycobacterium tuberculosis* complex

Sputum samples were cultured on Löwenstein-Jensen (LJ) medium following standard microbiological laboratory procedures as a routine of the Clinical Analysis Laboratory of “Hospital Raphael de Paula e Souza”. After cultivation, bacterial mass was re-suspended in 400 µl of sterile distilled water and heat inactivated at 90°C for one hour, followed by DNA extraction and purification using the CTAB method [39].

Spoligotyping

Spoligotyping was performed either as described by Kamerbeek *et al.* (1997) [12], using commercially available kit from Ocimum Biosolutions (Hyderabad, India) according to the manufacturer’s instructions or using microbead-based hybridization assay as described by Zhang *et al.* (2009) [40].

MIRU-VNTR typing

Amplification of 24 MIRU-VNTR *loci* was performed by using a commercial typing kit (Genoscreen, Lille, France) and automated MIRU-VNTR analysis performed as previously described [14]. Fragment size of the amplicons was analyzed on a ABI 3730 DNA sequence analyzer (Applied Biosystems, California, USA) and number of copies of each locus was determined by automated assignment using the Genemapper 4.0 software (Applied Biosystems, California, USA). In the case of doubtful results, the size of the repeats was double checked by size estimation as compared to a DNA ladder (50 and 100 bp) and the positive control (H37Rv) on agarose gels and by comparing to a reference table as described [13].

PCR-RFLP of *fbpC*¹⁰³

For characterization of SNP *fbpC*¹⁰³, we adapted the procedure described by Gibson *et al.* in 2008 [7]. Amplifications were performed in 50 µl reactions containing 40 pmol each of the primers Ag85C103F (5-CTG GCT GTT GCC CTG ATA CTG CGA GGG CCA-3) and Ag85C103R (5-CGA GCA GCT TCT GCG GCC ACA ACG TT-3), 2 mM MgCl₂, 0.2 mM dNTPs, 1 U Taq DNA polymerase (Invitrogen, Brazil), buffer (10 mM Tris-HCl, 1, 5 mM MgCl₂, 50 mM KCl, pH 8.3), 5% DMSO (v/v) and 10 ng of target DNA. Amplification was performed in a Vieri Thermal Cycler (Applied Biosystems, Foster City, CA), starting with denaturation at 95°C for 5 min, followed by 45 cycles of 1 min at 95°C, 1 min at 60°C, 4 min at 72°C and final extension for 10 min at 72°C. The amplified products of 519 bp were analyzed on 2% agarose gels after staining with ethidium bromide. Fifteen microliters of the amplified products were subjected to enzymatic digestion with 1 U of *MnII* (New England

BioLabs Inc. USA) at 37°C for four hours, generating three fragments of 365 bp, 96 bp and 48 bp for the wild type allele and two bands of 461 bp and 48 bp when the SNP G103A is present. Bands were detected in 3% agarose gel and their size estimated by comparison with a 100 bp DNA ladder (Invitrogen).

RD174 deletion

For detection of RD174, we used the protocol described earlier [7], performing multiplex-PCR using two primers that anneal to the RD174 flanking regions and one to the internal sequence. For amplification, we used 40 pmol of each of the primers RD174F (5'-AGC TGC TCC GGC CGG TCG TCG TCC TTG TC-3'), RD174Fi (5'-TAT GCC GCA GCC CGG GCA TCC GTG ATT A-3') and RD174R (5'-ATC GTG AAC GCA GCG GTT TCG ACG GCA TCT-3') in a 50 µl reaction containing 2 mM MgCl₂, 0.2 mM dNTP, 1 U Taq DNA polymerase, 1× buffer, 1M of betaine and 10 ng of bacterial DNA. Amplification was performed starting with 5 min incubation at 95°C, followed by 45 cycles of 1 min at 95°C, 1 min at 60°C, 4 min at 72°C and final extension for 10 min at 72°C. To determine the size of the amplicons, 10 µl of PCR product was applied in 2% agarose gel and after electrophoresis, bands of either 300 bp (intact RD174) or 500 bp (deleted) are observed.

RD^{Rio} deletion

Detection of RD^{Rio} was performed using the multiplex PCR-protocol described by Lazzarini *et al.* (2007) [2]. For amplification, we used 20 pmol of each of the primers BridgeF (5'-CAC TCC GGC TGC CAA TCT CGT C-3'), BridgeR (5'-CAC GAG GCT CGC GAA TGA GAC C-3'), IS1561F (5'-GAC CTG ACG GCC ACA CTG C-3') and IS1561R (5'-CAC CTA CAC CGC TTC CTG CC-3') in 50 µl reactions containing 2 mM MgCl₂, 0.2 mM dNTP, 1 U Taq DNA polymerase, 1× buffer, 5% DMSO (v/v) and 10 ng of target DNA. Amplifications were performed starting with incubation for 5 min at 95°C, followed by 45 cycles of 1 min at 95°C, 1 min at 60°C, 4 min at 72°C and final extension of 10 min at 72°C. The amplified products of 1175 bp or 530 bp in the presence or absence of the deletion, respectively were analyzed in 1.5% agarose.

Classification and definition of genotype-based lineages and families

Spoligotype and MIRU patterns were defined according to the definitions in the SITVITWEB database [23], (http://www.pasteurguadeloupe.fr:8081/SITVIT_ONLINE/) in the format of November 20, 2012. In addition, spoligotypes were compared with those in SITVIT2 (proprietary database of the “Institut Pasteur de la Guadeloupe”, which is an updated in-house version of the publicly released SpolDB4 and SITVITWEB [23], [24] to be released in 2014. This comparison led to the characterization of families and lineages by spoligotyping and Spoligotype International Types (SIT), MIRU International Types (12-MIT, 15-MIT AND 24-MIT) and 5 *loci* VNTR International Types (VIT). The SIT, MIT and VIT were designated identical patterns when shared by two or more patient isolates, whereas “orphan” patterns were those observed in a single isolate that did not correspond to any of the patterns present in the SITVIT2 database. In order to classify “orphan” patterns, we used SpotClust (http://tbinsight.cs.rpi.edu/run_spotclust.html).

The 24 MIRU-VNTR profiles and spoligotypes were also compared with the MIRU-VNTR_{plus} database [28] available at (<http://www.miru-vntrplus.org/MIRU/index.faces>). The definition of lineages was done on 24 MIRU-VNTR loci by Best-match

analysis and Tree-based identification using the categorical index. The Multi-locus Variable Number Tandem Repeat analysis (MLVA) MtbC 15 and MtbC 12 were determined.

Data analysis

The number and size of genotype clusters were defined initially by introducing numerical values into an Excel table and different criteria for definition of clusters were used, being either individual identical spoligotypes or 15–24 MIRU-based genotypes, or by combining with Spoligotyping both. Excel tables were introduced into BioNumerics software (version 6.1, Applied Maths, Belgium) for construction of similarity matrices and phylogenetic trees, using the Jaccard Index (spoligotypes) and Categorical value (MIRU-VNTR) for construction of the Neighbor Joining (NJ) and Minimal Spanning Tree (MST).

The MIRU-VNTR allelic diversity (*h*) at each of the 24 loci was calculated by the equation described by Graur and Li [41] and the Diversity index was calculated as the Hunter Gaston diversity index (HGDI) [42]. Statistical analysis were performed using chi-square analysis with a confidence interval of 95% using Epi Info.Version 3.51 (Centers for Disease Control and Prevention, Atlanta, GA,USA), by using χ^2 test or Fisher exact test for the comparison of proportions. A *p* value<0.05 was considered significant.

To visualize the classification difference observed between Spoligotyping and 24 loci MIRU-VNTRs, we constructed a Confusion Matrix for the frequency of correct and incorrect predictions. From the confusion matrix, the Positive Predictive Value or precision (PPV = TP/TP+FN), Accuracy rate (AAC = TP+TN/TP+FP+TN+FN), True Positive Rate or Sensitivity (TPR = TP/TP+TN), True Negative Rate or Specificity (TNR = TN/TP+TN), Negative Predictive Value (NPV = TN/FN+TN) and Error Rate (ER = FP+FN/TP+FP+TN+FN) were calculated. (TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative).

Results

Classification by spoligotyping

The identification of spoligotype profiles as well as their definition to the family and lineage level was realized by comparison with profiles deposited in the SITVITWEB database and in SITVIT2 and compared with the classification obtained with previous version of the spoligotype database (Table S1).

Based on comparison with SITVIT2, 83 SITs were encountered in 195 isolates (89.4%), while 10 SIT (n = 15) had Unknown profiles and 23 isolates showed to have Orphan patterns. For classification of the latter two pattern types, we used Spotclust (Table S1). Ninety five isolates (43.6%) were classified as belonging to the LAM clade, 76 (34.9%) as belonging to the T clade, 40 (18.3%) as belonging to the Haarlem clade, five to the EAI5 clade (2.29%) and one each to the X clade (0.45%) and to Beijing. Overall, T1 (n = 52/23.8%), LAM9 (n = 32/14.7%) and H3 (n = 24/11%) were the most frequent subclades, while SIT53 (n = 26/50%), SIT42 (n = 18/56.25%) and SIT50 (n = 16/72.7%) were the most frequent SITs.

Table 1 illustrates the difference in frequency of SIT clusters containing three or more isolates (at least 3%) in this study, versus their worldwide distribution (macro-regions and countries), as defined in the SITVIT2 database. We observed a significantly higher frequency of genotypes T1 SIT53 (11.9% vs 5.5%; *p* = 0,027, X^2 = 3,49) and a lower frequency of LAM9 SIT42 (8.2% vs 12.5%; *p* = 0,0001; X^2 = 15.82). In addition, SIT 3907 (T1), SIT 2498 (H3) and SIT 3908 (LAM2) have so-far only been

observed in Brazil. Presently, we also observed five SITs, 136 (LAM5), 189 (T1), 268 (H3), 2107 (unknown) and 2322 (T2), that had not been reported before in Brazil. When comparing our results with the previous and the present version of the SITVIT database, we noticed some differences in classification (Table S1). Thirty-six isolates could be classified using the new version and among these nine new SITs were observed: 3125 (LAM6), 3504 (LAM6), 3904 (LAM6), 3905 (T1), 3906 (T3), 3907 (T1), 3908 (LAM2; originally designated as SIT2560 Unknown in SITVIT-WEB), 3909 (LAM8) and 3910 (T1). In addition, nineteen orphan profiles were classified and SIT268, SIT1241 and SIT2539, without classification in SITVITWEB, were now classified respectively as H3, LAM9 and LAM1. In addition, some SIT showed changes in clade and/or subclade classification (SITs 73, 102, 159, 451, 742 and 777). Further information is detailed in Table S1.

Classification based on MIRU-VNTR analysis

The 218 isolates were also submitted to 24 loci MIRU-VNTR typing with the objective to confirm or not spoligotype-based classification of existing and new sublineages and to correct eventual miss classification that could have occurred using the latter technique. We observed 208 different patterns and seven clusters with two to four strains (clustering rate = 1.37%) (Figure S1). The allelic diversity of each MIRU-VNTR locus for 218 isolates was evaluated and classified as either highly (HGDI > 0.6), moderately ($0.6 \leq 0.3$) and poorly discriminative (HGDI < 0.3), according to Sola *et al.* [15], as summarized in Table 2. Eight loci were highly discriminatory (QUB4156, QUB11, MTUB04, MIRU10, MIRU40, MIRU 23, MIRU26 and MIRU21) ten were moderate (MTUB30, MTUB39, MIRU16, ETR-C, ETR-A, QUB26, MTUB34, MIRU31, MIRU29 AND ETR-B) and six were poorly discriminative (MIRU24, MIRU29, MIRU04, MIRU27, MIRU20 and MIRU2). In the Table 2, we can observe a difference in discriminatory power of some loci according to lineage by Spoligotyping.

The patterns were classified based on the database tool that allows construction of a Neighbor-Joining based phylogenetic tree, visualizing proximity of a particular genotype with that of a set of reference strains to the genotype family level. One hundred thirty-six isolates (62.4%) were classified as LAM (127 patterns), 47 (21.5%) as Haarlem (45 patterns), 21 (9.6%) as T (21 patterns), 11 (5%) as S (11 patterns), and one isolate (0.45%) each as Beijing and EAI. Considerable differences were observed between Spoligotyping and 24 MIRU-VNTR loci-based classifications, even after excluding eventual small typing errors by repeating the assays. The differences observed between the two classifications are presented in the table 3 and in the Confusion Matrix (Figure 1). The precision, accuracy, sensitivity and error rate were respectively 0.74, 0.64, 0.82 and 0.36. The T lineage showed the highest incongruence rate related to classification (sensitivity = 0.26).

Combined analysis of 24 MIRU-VNTR and spoligotyping

The LAM family. For better understanding of the population structure of the Mtb strains, dendrograms of spoligotyping and 24 MIRU-VNTRs, either separately or as combined patterns were constructed using BioNumerics. Among the 136 (62.4%) isolates that were classified as LAM based on MIRU-VNTR analysis, 99 (72.8%) exhibited the characteristic LAM spoligotype profile with absence of spacers 21 to 24 and 33–36. Among these however, four spoligotype isolates were designated to the T clade by SITVIT2 (SITs: 306/n = 3, 1688/n = 1, 3905/n = 1) and one as T2/orphan. Another thirty-seven strains (27.2%) defined as LAM by MIRU-VNTRs did not present the lack of spacers 21–24 and

were therefore designated as “LAM-like”. Among isolates initially classified as LAM by spoligotyping (n = 95), 91 were confirmed as LAM (95.8%/TPR = 0.96) by 24 loci MIRU-VNTR (Figure 1). Six spoligotypes referred to as “Unknown” in SITVIT2 (SITs 132/n = 1; 1952/n = 1; 2107/n = 2; 2511/n = 2; 2535/n = 1 and 2548/n = 1) were also classified as LAM. Using SpotClust, SIT132 and SIT2511 were classified as EAI and SIT1952 as Haarlem. All Orphans patterns classified as LAM by SITVIT2 (n = 10) were also classified as LAM according to the MIRU-VNTR based phylogenetic tree.

Signatures based on MIRU copy number for “real” LAM (n = 99) types showed that the majority had two copies of MIRU04 (94; 95%) and two MIRU20 (n = 96; (97%), one copy of MIRU24 (n = 91; 92%), two copies of MIRU39 and ETR-A (n = 93; 94%) and 86 (87%) one copy of MTUB30. This indicates that the 221221 combination of these six alleles is representative for LAM strains, except for LAM3, presenting two copies of MTUB30 (221222).

The “LAM-like” isolates. The thirty-seven isolates that did not present the typical LAM spoligotype signature (absence of spacers 21 to 24 and 33–36) were positioned within the LAM branch in the neighbor-joining tree using MIRU-VNTR_{Plus} and were conveniently designated as “LAM-like”. These had been classified by spoligotyping earlier as T (n = 24), H (n = 10) and three “Unknown” patterns (SIT2511/n = 2 and SIT1952/n = 1; classified by SpotClust as EAI and H1 respectively). Ninety-one percent of these strains showed one copy of MIRU24 and two copies of MIRU04. Intriguingly, when including both 24 MIRU-VNTRs and spoligotypes for tree building, these isolates are localized among those classified as T and H.

The T family. Among the isolates initially classified as T by spoligotyping, only 26.3% (20/76) was confirmed by MIRU-VNTR 24 loci (TPR: 0.2631) (Table 3 and figure 1). By a Neighbor-Joining based phylogenetic tree, 31 of these isolates (40.9%) were positioned in LAM branch, 16 (21%) in the Haarlem branch and 9 (11.8%) in the S branch. The MIRU-VNTR characteristic signature for the T lineage confirmed by both techniques, was one copy of MIRU24 (95.8%), two copies each of MIRU02 (95.8%), MIRU04 (87.5%), MIRU20 (95.8%), MIRU39 (100%) and three copies each of MIRU27 (87.5%) and MTUB34 (87.5%). Unlike LAM isolates that present a single copy of MTUB30, isolates of the T family presented two or four copies of this allele.

The Haarlem family. Forty-seven isolates were classified as belonging to the H family representing 21.5% of all isolate, 85% of these isolates initially classified as Haarlem by Spoligotyping were confirmed by MIRU-VNTRs 24 loci (TPR = 0.65, TNR = 0.88) (Table 3 and figure 1). Among these isolated 16 (34%) presented spoligotype profile compatible with T family, 3 (6.4%) with LAM family and 2 (2% each) isolates, respectively compatible with X and EAI. The MIRU-VNTRs characteristic for Haarlem are two copies of MIRU02 (100%) and MIRU39 (89.1%) and three copies of MTUB34 (100%), MIRU16 (91.3%), MIRU27 (97.8%), ETR-A (95.6%) and ETR-C (91.3%). Different from LAM isolates having a single copy of MTUB30, H isolates present either three or four copies of this MIRU. Combined analysis of the MIRU-24 and spoligotyping showed that isolates of genotype H1 (absences of spacers 26–31) have two copies of MIRU04, ETR-B and MTUB29 and three copies of MIRU16, MIRU31, ETR-A and ETR-C. Isolates of genotype H3 (absences of spacer 31) share two copies of MIRU02, ETR-C, MIRU4 and MTUB29 and three copies of MIRU27, ETR-A, ETR-B and MTUB34.

Table 1. Description of clusters containing 3 or more isolates in this study and their worldwide distribution in the SITVT2 database (interrogation made on September 25th 2013).

SIT (Lineage) Octal Number; Spoligotype Description	Number (%) in study	% in study vs. database	Distribution in Regions with ≥3% of a given SITs *	Distribution in countries with ≥3% of a given SITs **
53 (T1) 77777777760771	26 (11.93)	0.41	EURO-W 15.65, AMER-N 13.48, AMER-S 12.45, EURO-S 9.41, EURO-N 7.48, ASIA-W 7.31, AFRI-S 4.97, AFRI-E 4.65, ASIA-E 4.23, AFRI-N 3.52, EURO-E 3.26, AMER-C 3.23	USA 13.2, FXX 7.88, BRA 5.54, ITA 5.33, ZAF 4.86, TUR 3.47, AUT 3.42, CHN 3.09
42 (LAM9) 777777607760771	18 (8.26)	0.54	AMER-S 29.7, AMER-N 11.85, EURO-S 11.31, EURO-W 9.46, AFRI-N 8.57, EURO-N 4.9, AMER-C 3.55, AFRI-E 3.55, AFRI-S 3.16	BRA 12.51, USA 11.85, COL 7.61, MAR 7.05, ITA 6.54, FXX 5.05, ESP 3.34, VEN 3.31, ZAF 3.16
50 (H3) 77777777720771	16 (7.34)	0.47	AMER-S 18.16, EURO-W 17.54, AMER-N 17.54, EURO-S 11.53, EURO-E 5.51, EURO-N 5.45, AFRI-N 4.25, AFRI-S 4.04, CAR 3.48, ASIA-W 3.21	USA 17.51, BRA 7.52, FXX 6.87, AUT 6.07, ITA 5.43, ESP 5.43, PER 4.72, ZAF 4.04, CZE 3.66
SIT64 (LAM6) 777777607560771	6 (2.75)	1.6	AMER-S 49.33, AMER-N 25.07, EURO-W 6.4, EURO-S 4.8, ASIA-W 3.2	BRA 38.4, USA 25.07, GUF 6.13, PRT 3.73
SIT47 (H1) 777777774020771	5 (2.29)	0.32	EURO-W 20.29, AMER-N 15.51, EURO-S 13.47, AMER-S 12.64, EURO-N 10.47, EURO-E 7.21, ASIA-W 4.02, AFRI-N 3.64	USA 15.19, ITA 8.23, AUT 8.04, BRA 7.34, FXX 6.64, FIN 6.0, CZE 3.77, ESP 3.57, SWE 3.38, PER 3.06
SIT60 (LAM4) 777777607760731	5 (2.29)	1.17	AFRI-S 38.08, AMER-S 21.5, EURO-W 7.24, AFRI-N 7.01, AFRI-W 6.31, EURO-S 5.61, AMER-N 4.4	ZAF 38.08, BRA 14.72, FXX 5.14, USA 4.44, MAR 4.44, GMB 3.97, ITA 3.74, VEN 3.51
SIT73 (T) 77777777760731	5 (2.29)	1.94	EURO-S 17.05, EURO-W 13.95, AMER-N 13.95, AMER-S 11.24, AFRI-S 10.85, AFRI-E 7.36, ASIA-E 4.65, AMER-C 3.88	ITA 14.34, USA 13.95, ZAF 10.85, FXX 7.75, BRA 6.59, CHN 4.65, MOZ 3.88, MEX 3.1
4 (Unknown) 000000007760771	4 (1.83)	1.08	EURO-S 16.49, AMER-S 14.32, AMER-N 12.43, EURO-W 10.81, ASIA-W 10.54, AFRI-S 8.11, AFRI-E 8.11, EURO-N 6.22, EURO-E 5.68	USA 9.73, ITA 8.92, BRA 8.92, ZAF 8.11, TUR 7.84, BGR 5.41, ETH 4.32, FXX 4.05, ALB 4.05
SIT17 (LAM2) 677737607760771	4 (1.83)	0.6	AMER-S 59.58, AMER-N 17.95, CAR 8.9, EURO-S 5.58	BRA 28.36, VEN 27.0, USA 17.95, ESP 4.37, HTI 3.32, GLP 3.32
SIT33 (LAM3) 776177607760771	4 (1.83)	0.35	AFRI-S 28.76, AMER-S 25.31, AMER-N 14.07, EURO-S 12.57, EURO-W 7.26, AMER-C 4.78	ZAF 28.76, USA 14.07, BRA 12.04, ESP 7.88, PER 6.02, ARG 5.04, FXX 4.69, ITA 4.07, HND 3.81

Table 1. Cont:

[illegible]

Worldwide distribution is reported for regions with more than 3% of a given SITs as compared to their total number in the SITVT2 database. The definition of macro-geographical regions and sub-regions (<http://unstats.un.org/unsd/methods/m49/m49regin.htm>) is according to the United Nations; Regions: AFRI (Africa), AMER (Americas), ASIA (Asia), EURO (Europe), and OCE (Oceania), subdivided in: E (Eastern), M (Middle), C (Central), N (Northern), S (Southern), SE (South-Eastern), and W (Western). Furthermore, CARIB (Caribbean) belongs to Americas, while Oceania is subdivided in 4 sub-regions, AUST (Australasia), MEL (Melanesia), MIC (Micronesia), and POLY (Polynesia). Note that in our classification scheme, Russia has been attributed a new sub-region by itself (Northern Asia) instead of including it among rest of the Eastern Europe. It reflects its geographical localization as well as due to the similarity of specific TB genotypes circulating in Russia (a majority of Beijing genotypes) with those prevalent in Central, Eastern and South-Eastern Asia.

doi:10.1371/journal.pone.0107747.t001

Table 2. Allelic diversity of 24 MIRU-VNTRs loci on 218 *Mycobacterium tuberculosis* strains isolated from pulmonary tuberculosis patients in Rio de Janeiro, Brazil.

MIRU-VNTR loci	Genotype families by Spoligotyping			All families
	LAM	H	T	
MIRU02	0.362	0.148	0.353	0.325
MIRU04	0.061	0.153	0.225	0.139
MIRU10	0.56	0.687	0.692	0.692
MIRU16	0.636	0.243	0.439	0.544
MIRU20	0.081	0.499	0.24	0.27
MIRU23	0.541	0.529	0.584	0.688
MIRU24	0.021	0.101	0.081	0.071
MIRU26	0.723	0.559	0.687	0.639
MIRU27	0.314	0.145	0.105	0.227
MIRU31	0.429	0.357	0.51	0.445
MIRU39	0.138	0.251	0.027	0.124
MIRU40	0.517	0.623	0.76	0.689
ETR-A	0.14	0.417	0.625	0.537
ETR-B	0.407	0.237	0.396	0.389
ETR-C	0.357	0.447	0.573	0.541
MTUB04	0.612	0.53	0.68	0.716
MTUB21	0.436	0.587	0.673	0.624
MTUB29	0.253	0.432	0.266	0.397
MTUB30	0.269	0.494	0.652	0.591
MTUB34	0.617	0.305	0.617	0.452
MTUB39	0.391	0.573	0.391	0.577
QUB11	0.697	0.702	0.697	0.737
QUB26	0.384	0.63	0.384	0.522
QUB4156	0.842	0.842	0.842	0.84

doi:10.1371/journal.pone.0107747.t002

The S family. Eleven isolates were classified as belonging to the S family and four of these exhibit SIT4, that in SpolDB4 was classified as LAM3-S Convergent, having absence of spacers 1–24 and 33–34 (The characteristic pattern of S family is absence of spacers 9–10, and 33–34). Among the others isolates, we observed SITs 53/n = 1, 102/n = 1, 378/n = 1, 2500/n = 1, 3907/n = 2 and 3909/n = 1; only two of these had the characteristic of S family (SIT3909/LAM8 and SIT2500/Unknown by SITVIT2 and H1 by Spotclust). Isolates of the S family shared their copy number in six loci, being one copy of MIRU24 and MTUB 21, two copies

each of MIRU20 and MIRU39 and three copies MIRU10 and MIRU27.

RD^{Rio}, RD174 and *fbpC*¹⁰³ analysis

The genotypes defined by the presence of RD^{Rio}, RD174 and SNP *fbpC*¹⁰³ were added to the classification based on 24-MIRU-VNTR typing (Table S2). Thirty-six isolates (16.5%) were excluded from the final analysis either because of showing genotypes suggestive for multiple infection or because of failure in at least one of the three genotype procedures. The results of 182

Table 3. Frequencies of strains according to classification by Spoligotyping (SITVIT2) and 24 loci MIRU-VNTR (MIRUVNTRPlus).

Lineage	Spoligotyping (%)		24 loci MIRU-VNTR (%)	
LAM	95	(43.58)	136	(62.39)
T	76	(34.86)	21	(9.63)
H	40	(18.35)	47	(21.56)
S	0	0	11	(5.05)
X	1	(0.46)	1	(0.46)
EAI	5	(2.29)	1	(0.46)
BEIJING	1	(0.46)	1	(0.46)

doi:10.1371/journal.pone.0107747.t003

Spoligotyping ¹									
MIRU-VNTR 24 loci ²		LAM	T	Haarlen	S	X	EAI	Beijing	Total
	LAM	91	31	11	0	0	3	0	136
	T	0	20	1	0	0	0	0	21
	Haarlem	3	16	26	0	1	1	0	47
	S	1	9	1	0	0	0	0	11
	X	0	0	1	0	0	0	0	1
	EAI	0	0	0	0	0	1	0	1
	Beijing	0	0	0	0	0	0	1	1
	Total	95	76	40	0	1	5	1	218

¹ Classification according to the SITIVIT2. For unknown Spoligotypes, we used the SpotClust.

² The patterns were classified based on MIRU-VNTRPlus tool that allows construction of a Neighbor-Joining based phylogenetic tree, visualizing proximity of a particular genotype with that of a set of reference strains to the genotype family level.

MIRU-VNTR 24 loci vs Spoligotyping		LAM	T	Haarlem
Positive Predictive Value (PPV) or Precision	0.7433	0.6691	0.9523	0.5531
Accuracy (AAC)	0.6376	0.7752	0.7385	0.8394
True positive rate (TPR) or sensitivity	0.8176	0.9578	0.2631	0.6500
True Negative Rate (TNR) or Specificity	0.3623	0.4193	0.9929	0.8820
Error Rate		0.9512	0.7157	0.9181
		0.669	0.9523	0.5531
		0.2247	0.2614	0.3500

Figure 1. Confusion Matrix comparing the classifications obtained by Spoligotyping and MIRU-VNTR. ¹ Classification according to SITIVIT2. For unknown Spoligotypes, we used SpotClust. ² Patterns were classified based on a VNTRplus database that allows construction of a Neighbor-Joining based phylogenetic tree, visualizing proximity of a particular genotype with that of a set of reference strains to the genotype family level.

doi:10.1371/journal.pone.0107747.g001

strains are summarized in Table 3 and for simplification of interpretation, we defined three groups, being LAM (n = 77), LAM-like (n = 33) and non-LAM (n = 72) as determined by spoligotyping and MIRU-VNTR typing.

The SNP *fbpC*¹⁰³ was observed in 107 (58.8%) of the isolates, including 97.3% (74/77) of LAM, 69.6% (23/29) of “LAM-like” and 13.9% (10/72) of non-LAM. The total frequency of RD^{Rio} was 11% (20/182), with 20.7% (16/77) among LAM, 6.0% (2/29) among “LAM-like” and 2.7% (2/72) among non-LAM isolates. The overall frequency of RD174 was 15.3% (28/182), being 33.7% (25/77) in the LAM genotype and 6.8% (2/29) in “LAM-like”. All LAM/RD^{Rio} isolates presented the SNP *fbpC*¹⁰³ (Table 4 and Table 5), but not all were RD174; two isolates classified as LAM2 SIT3908 were not had RD174 (isolated from different patients). The frequency of RD174 (n = 25) was therefore higher than that of RD^{Rio} (32.5% vs 20.7%; p = 0.053 and $\chi^2 = 2.69$) in LAM. Eleven LAM isolates (14.3%) presented RD174 but not RD^{Rio}, all were positive for the *fbpC*¹⁰³.

The allelic diversity of the 24 loci MIRU-VNTR loci in LAM/RD^{Rio} isolates (Figure S2) showed that the copy number of

combining of MIRU04, MIRU20, MIRU24, MIRU31, ETR-A, MTUB21 and MTUB30 loci was a signature of this genotype (2213231) (Table 4). In general, these loci present low variability in LAM, except for MTUB21, is being moderately variable in such isolates (table 2). Upon comparing 24 MIRU-VNTR signatures of LAM, LAM-like and non-LAM strains (Table 4), we observed two isolates that presented the hypothetical ancestral MIRU-VNTR signature (224226153321) for RD^{Rio} that was suggested by Lazzarini et al. (2007) [2]. One isolate (C2009) was classified as LAM1 SIT2539 and the other (C1966) as H3 SIT50 by spoligotyping and “LAM-like” by 24 MIRU-VNTR; both presented SNP *fbpC*¹⁰³ and were deleted for RD^{Rio} and RD174. Among the LAM/RD^{Rio} strains, frequency of LAM subtypes as defined by spoligotyping using SITVIT2 was 31.5% of LAM2, 25% of LAM9, 12.5% each of LAM6 and LAM5 and 6.3% each of LAM4 and LAM1. Figure 2 is a graphical representation of these three markers in the LAM strains as defined by 24 loci MIRU-VNTRs.

Table 4. Description of RD^{Rio} by 24 loci MIRU-VNTRs, *fbpC*¹⁰³ and RD174.

[illegible]

10

October 2014 | Volume 9 | Issue 10 | e107747

Rio de Janeiro is the capital of the state of Rio de Janeiro, located in the southeast of Brazil, has a population of 16 million habitants, six and a half million of these living in the capital, being the second largest city and a major touristic attraction in Brazil (Census 2010 <http://cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=330455> accessed 12/18/2013). According to the Ministry of Health, 11,639 new TB cases were recorded in 2011 in the state with an incidence of 72.3/100.000 habitants and the highest mortality rate (5.1/100.000 habitants) at the national level [1]. Rio de Janeiro city has strong economic and social contrasts, with a large portion of the population living in numerous suburbs, consisting of “communities”, urban areas where housing conditions, health, education and security are extremely precarious. These factors are directly related to the number of TB cases detected.

In the present study, we performed spoligotyping and 24 MIRU-VNTR typing and characterized a SNP in *fbpC*¹⁰³ and the status of RD174 and RD^{Rio}, to decipher the population structure and the phylogenetic relationships of the MTBC strains of TB cases in this particular population that attended a single reference center in the city of Rio de Janeiro. *Mycobacterium tuberculosis* is classified into six phylogenetic lineages, each of which can be divided into sublineages [32]. Members of the Euro- American lineages represented by the LAM, H, T, S and X Spoligotyping families, which are genetically closely related are the most common in South America [19], [20], [21] and this was confirmed recently also in Brazil [22] and in different states of the country [8], [43], [44], [45], [46], [47] [48], but differences in the frequency of these families are observed, these difference could also be due to differences in population and immigration history in each region or in period associated genotype frequencies or in differences in sample size. We here confirm the predominance of the Euro-American lineage (also known as lineage 4) with only two isolates classified as Indo-Oceanic lineage (EAI5) and East Asian lineage (Beijing), the first being more prevalent in east Africa, southeast Asia and in south India and the second in east Asia, Russia and South Africa. [32]. Recently, the frequency of such strains in Brazil was described as being less than 1 % except for the higher frequency in the state of Pará, north Brazil [22].

The most prevalent families as defined by spoligotyping were LAM (43.6%), T (34.9%) and H (18.3%); however, classification based on 24-MIRU-VNTR and Neighbor-Joining based phylogenetic tree building using the database tool (www.miruvntrplus.org) showed considerable difference with that obtained by spoligotyping, being respectively 62.4%, 9.6% and 21.6% (error rate of 0.36). Among the isolates with discordant results, mostly isolates initially classified as T (n = 31), H (n = 11) and EAI (n = 3) by spoligotyping were classified as LAM by 24 MIRU-VNTRs typing, for isolates that did not show LAM prototype (absence of spacers 21–24 and 33–36), conveniently named LAM-like. Subsequently, 78.3% these isolates confirmed to be LAM by the presence of the LAM-specific SNP *fbpC*¹⁰³ [7], [38] and when considering the presence of this SNP as an absolute marker for LAM family the frequency of this lineage is 58.9%. We also have preliminary data showing that the supposed LAM-specific marker *ligB*⁴⁰⁴ was observed in isolates that had been defined by spoligotyping as being T or H (data not shown). Very recently, Mokrousov et al. reported difference in classification of LAM strains as defined by spoligotyping and SNP analysis (*fbpC*¹⁰³ and *ligB*⁴⁰⁴) [49] although both SNPs were previously validated in different collection of clinical isolates and reference strains as to be specific for the LAM family [7], [38].

Table 5. The frequency of *fbpC*¹⁰³, RD^{Rio} and RD174 in LAM, LAM-like and Non LAM isolates, as designated by spoligotyping and 24 MIRU-VNTR typing.

	LAM (%)	LAM-Like (%)	Non LAM (%)	Total Frequency (%)
SNP and RD	(n = 77)	(n = 29)	(n = 75)	(n = 182)
<i>fbpC</i> ¹⁰³	74 (97.3)	23 (79.3)	10 (13.3%)	107 (58.8)
RD ^{Rio}	16 (20.7)	2 (6.9%)	2 (2.6%)	20 (11)
RD174	25 (32.5)	2 (6.9%)	1(1.3%)	28 (15.4)

doi:10.1371/journal.pone.0107747.t005

It is now common knowledge that spoligotyping has limitations as a tool prediction of the exact phylogenetic relationships between strains of the MTBC, particularly in modern strains (Euro American, East Asian and Indian e East African) [16], [18], [24], [48], [49], [50], [51] [52], mainly due to homoplasmy [16], [24]. The accuracy of the phylogenetic grouping by MIRU-VNTR is more exact than that of spoligotyping but depends on the number of loci included in the analysis [49] and classification errors are reduced when analyzing 24 loci [16]. Indeed, several authors have suggested that SNPs are more suitable than spoligotyping and MIRU-VNTR settings for phylogenetic classification [16], [38], [53], [54], [55], [56]. The *fbpC*¹⁰³ is described as a good marker for the LAM family [7], [38], [49], [57], [58], [59], and our original intention was to evaluate if strains RD^{Rio} was still prevalent in Rio de Janeiro as seen previously by Lazzarini in clinical isolates collected between 2002 and 2003 [2], for this

purpose we believed only a marker that could differentiate LAM and non-LAM associated with the 24 MIRU-VNTR and Spoligotyping could be sufficient, however the scenario was observed more complex.

We here present the first data comparing classification by spoligotyping and MIRU-VNTRs and *fbpC*¹⁰³ in Euro American lineage prevalent in Rio de Janeiro and Brazil. Among these different types of markers, we observed four major groups: (i) strains classified as LAM by spoligotyping and MIRU-VNTRs 24 loci without the LAM-characteristic SNP *fbpC*¹⁰³, (ii) strains not classified as LAM by spoligotyping and MIRU-VNTRs 24 loci but carrying the *fbpC*¹⁰³ SNP, (iii) strains classified by spoligotyping as non-LAM but as LAM by MIRU-VNTRs 24 loci (LAM-like) and with SNP *fbpC*¹⁰³ and (iv), strains classified by spoligotyping as non-LAM and LAM by MIRU-VNTRs 24 loci (LAM-like) but not presenting the SNP *fbpC*¹⁰³. These different scenarios could be explained by convergent evolution of spoligotypes and of MIRU-VNTRs loci (even including 24 alleles) because a limited number of loci were evaluated that might evolve rapidly and therefore susceptible to pronounced convergence [53] and/or because the existence of ancestral progenitor of Euro- American lineages containing SNP *fbpC*¹⁰³. In addition, a possible important limitation of the current classification based on MIRU-VNTRs and their similarity with genotypes present in MIRU-VNTRplus is that, despite including well characterized strains, the database contains a limited number of strains that does not reflect the real genetic diversity of isolates belonging to the MTBC; another limitation of this study is that no additional specific SNPs for H, T, S and X and T were investigated. The Whole Genome Sequencing (WGS) is superior to conventional genotyping for MTBC [60] and has been used in areas not yet studied, from global (phylogeography) for site (transmission chains and diversity of circulating strains), for single patient (clonal diversity) and the bacterium itself (evolutionary studies) [61]. We intend to compare through WGS (developing) these isolates to a better comprehension of the evolution of lineages Euro American, such as the development of new methodologies that allow a more rapid and accurate typing.

Analyzing the data obtained in this study, the spoligotypes of the T family showed the largest number of divergent results when compared to classification by 24-MIRU (TPR = 0,26/ TNR = 0,99). The prototype of the T family is characterized by the absence of spacers 33–36 only [48] and have been observed in almost every country, representing 20% of all isolates deposited in the database SITVITWEB. Despite the high frequency, this genotype family is still considered as “ill-defined” and includes non-monophyletic groups [54] [55]. In South America, the frequency of this family is 26.7% and in Brazil 18.6% (370/ 1991), with the T1 SIT53 subfamily being the most prevalent SIT [22], as observed also in this study. Here, among the 76 isolates classified as T by spoligotyping, only 21 (27.6%) were confirmed

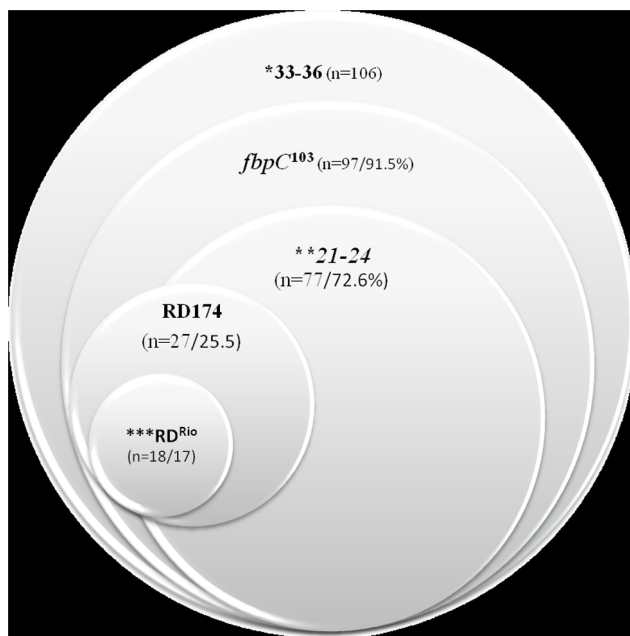


Figure 2. Venn diagram illustrating the different markers in isolated classified as LAM by 24 loci MIRU-VNTR. Notes: The Venn diagram was constructed based on LAM isolates defined by 24 loci MIRU-VNTR. The sizes of the circles is not proportional to the real frequency of these markers. Twenty isolates classified with LAM by MIRU-VNTRs were removed from the final analysis because of showing a mixed genotype or failure in at least one of the three genotype procedures. * absence of spacers 33–36 in the spoligotyping profile, ** absence of spacers 21–24 in the spoligotyping profile, ***Two isolates RD^{Rio} but not RD174 (LAM2 SIT3908).

doi:10.1371/journal.pone.0107747.g002

by 24 MIRU-VNTRs typing, the rest was reclassified as LAM ($n = 31$), S ($n = 9$) and H ($n = 16$). Interestingly, SIT53 was associated with mixed infection in a study conducted in South Africa, a country characterized by a high prevalence of TB [57] and Lazzarini *et al.* [62], using a computational approach, showed that this is the most frequent false spoligotype derived from mixed infections. However, among isolates with this SIT, we did not observe double signals during 24 MIRU typing indicative for mixed infections. We also observed that spoligotypes, indicative for the T family, were sometimes grouped with spoligotypes of the H family by MIRU-VNTR typing. This could be related with the fact that the prototype spoligotype defining T1 SIT53 and H3 SIT50 differ only in the absence of the spacer 31. We also observed that in our population, the absence of spacer 31 is not crucial for classification as being either H or T; what differentiates between the two is the number of copies of ETR-A, ETR-B, ETR-C. The H3 subfamily is characterized by the combination 323 of these alleles while T strains present considerable diversity of these loci (one to three copies of ETR-A, two or three copies of ETR-B and three to five copies of ETR-C).

The RD nominated RD^{Rio} was first reported as new *M. tuberculosis* lineage in Rio de Janeiro in 2007, Lazzarini *et al.* [2] and is a deletion of 26.3 kb restricted to the LAM family and in particular, in subfamilies LAM1, LAM2, LAM4, LAM5, LAM6 and LAM9. This deletion affects 10 genes, including two genes encoding Proline-proline-glutamic Acid Proteins (PPE) [2] and association between RD^{Rio} strains and high prevalence may be related to virulence and/or adaptation specifies the Latin American and European population-based epidemiological and clinical characteristics; however, studies have proven to be inconclusive or contradictory [4], [63], [64]. The lineage RD^{Rio}, was identified in different countries [5], [6], [7], [49] and in Brazil, has been described in different states, including Rio de Janeiro, Rio Grande do Sul, Minas Gerais, Espírito Santo and Rio Grande do Sul [3], [8], [63], [64]. The frequency ranges from 30 to 38% of isolates tested and was associated with MRD-TB [4], [63], [64] and with genotype clustering [5], [63], indicating a higher rate of recent infection and transmissibility. Another deletion, RD174, initially described as a marker for the LAM family and as a co-marker for RD^{Rio} [7], [8] was associated with high transmissibility [10].

The frequency of strains RD^{Rio} in the present study, with an overall frequency of 11% and 20.7% in LAM, is lower than that observed in other studies and even when including the eight isolates that showed mixed RD^{Rio}/WT signals, resulting frequencies of 14.7% and 28.2% still lower than previously observed. This difference could be related with the relative low frequency of LAM1, LAM2, LAM5, LAM6 and LAM9 related subfamilies in our study sample. Earlier studies on classification of RD^{Rio} strains was based on genotypes defined by spoligotyping and/or 12 MIRUs only [2], [5], [6], [7], [8], [63], [64] and again, this is the first study that used 24 MIRU-VNTR typing to conduct a more detailed phylogenetic analysis. We verified the signature of MIRU-VNTR loci for RD^{Rio} and RD174, as compared with that of the hypothetical ancestral of RD^{Rio} as defined by 12 MIRUs [2] and that, besides sharing two copies of MIRU04 and MIRU20, they also share one copy of MIRU24, three copies of MIRU31, two copies of ETR-A, three copies of MTUB21 and one copy of MTUB30, yielding 2213231 as fingerprint for RD^{Rio}. All LAM/RD174 isolates, with or without RD^{Rio}, carried two copies of MIRU20 and ETR-A and one copy of MTUB31. We observed that LAM3 has two copies of MTUB30 (2213232) and we propose

that this subfamily that do not carry RD^{Rio}, has evolved independently. We also observed RD^{Rio} in two isolates (2.6%) with a spoligotype not indicative for being LAM; a small number of such cases have also been observed by other research groups [5], [7], [64]. One of these isolates that had been classified by spoligotyping as H3 SIT53 was reclassified by 24 MIRU as being “LAM-Like” and called attention because it had the MIRU signature of the hypothetical ancestor of RD^{Rio} and carried the SNP *fbpC*¹⁰³ and RD174. Overall, we observed four scenarios: (i) isolates with the RD^{Rio} and RD174, (ii) isolates showing only RD^{Rio} (iii) isolates showing only RD174 and (iv) isolates that did not carry any of the deletions, suggesting that both markers evolved on different time points. This is different from earlier data [7] that claim that RD174 is an absolute co-marker for RD^{Rio} and therefore, studies that use the presence of RD174 to infer the presence of RD^{Rio} [8] may overestimate the frequency of the latter. In 2007, Lazzarini *et al.* [2] suggested that RD^{Rio} arose by homologous recombination between genes and although all neighboring sequences were identical, such event could have happened more than once. This suggests that the RD174 deletion occurred before RD^{Rio} but this needs to be confirmed as we also observed RD^{Rio} strains that had intact RD174. In Figure S3, we propose a possible evolution of members of the Euro-American lineages but with the limitation that the spoligotype defined lineages S, X and other are not represented and this concerns a sampling only from Rio de Janeiro.

In a recent study, Hill *et al.* (2012) [53], mentions the difficulty in studying the evolution of the Euro-American lineage (LAM, Haarlem, T, X and S) using spoligotyping due to the large number of IS6110 copies in such strains that may result in IS6110 mediated deletions in the DR locus. This might be the reason why bacterial evolution exclusively based on spoligotyping is not robust and the wide range of profiles reported as unclassified in SITVITWEB. Our approach, combining spoligotyping, MIRU-VNTRs, SNPs and RDs allowed the reclassification of 13 SITs that did not rank in SITVITWEB, allowing definition of 29 new spoligotypes and refine classification of isolates belonging to Euro-American lineage. Our data are also support the idea that absence of spacers 21–24 is not sufficient for classification as LAM and of spacer 31 to differentiate T and H, the latter indicative for subfamilies H3 SIT50 and T1 SIT53. Possible explanations are that ancestral lineages are currently circulating (plesiomorphic state) or that the isolates are suffering homoplasious evolution and reversion into the plesiomorphic state.

Supporting Information

Figure S1 Dendrogram constructed with BioNumerics software version 6.6 with the results of MIRU-VNTRs 24 loci and spoligotyping by similarity coefficient for categorical data and the neighbor-joining algorithm. 1st column (after spoligotypes): number of isolated label; 2nd column: patients who have more than one isolate in the study ($n = 27$) received numbering 1–27, and the different strains present the same numbering; 3rd column: International Spoligotype Types (SIT); 4th column: classification obtained through SITVITWEB (family and subfamily). (DOC)

Figure S2 Allelic diversity of the 24 loci MIRU-VNTR loci in LAM/RDRio isolates. (DOC)

Figure S3 Possible evolution of *M. tuberculosis* lineage Euro-American (LAM, T and H) according to the markers analyzed in this study. The Euro-American Ancestral evolved into two distinct groups: LAM ancestral and T/H ancestral, characterized by absences of spacers 33–36 and one copy of MIRU24 that is common to all Euro-American lineages. The ancestral LAM strains have *fbpC*¹⁰³ and this was the basis of LAM A (LAM9), with additional absence of spacers 21–24, one copy of MTUB30 and two copies of MIRU04 and ETR-A. LAM A on its turn is the basis for two other groups: LAM B (LAM9-LAM4, LAM1-LAM2-LAM5 and LAM6) and LAM C (LAM3). The LAM B evolved from LAM B1, characterized by a deleted RD174 and on its turn to LAM B2, with both deleted RD174 and RD^{Rio}. The H/T Ancestral lineage is the origin of both groups H and T (difference only in MIRU-VNTR copies), showing absence of spacers 33–36; additional loss of spacer 31 led to subtype H A, observed in high frequency in this study. (DOC)

Table S1 Detailed genotyping results and associated demographic, epidemiologic and nomenclature information on 218 *Mycobacterium tuberculosis* strains

isolated from pulmonary tuberculosis patients in Rio de Janeiro, Brazil. (XLS)

Table S2 Detailed genotyping results of *fbpC*¹⁰³, RD174 and RD^{Rio} on 182 *M. tuberculosis* strains isolated from pulmonary tuberculosis patients in Rio de Janeiro, Brazil. (XLS)

Acknowledgments

We thank the Genomic Platform of Network Technology Platforms (PDTIS) Institute Oswaldo Cruz/FIOCRUZ.RPT01D - Analysis of Fragments Oswaldo Cruz Institute in Rio de Janeiro.

Author Contributions

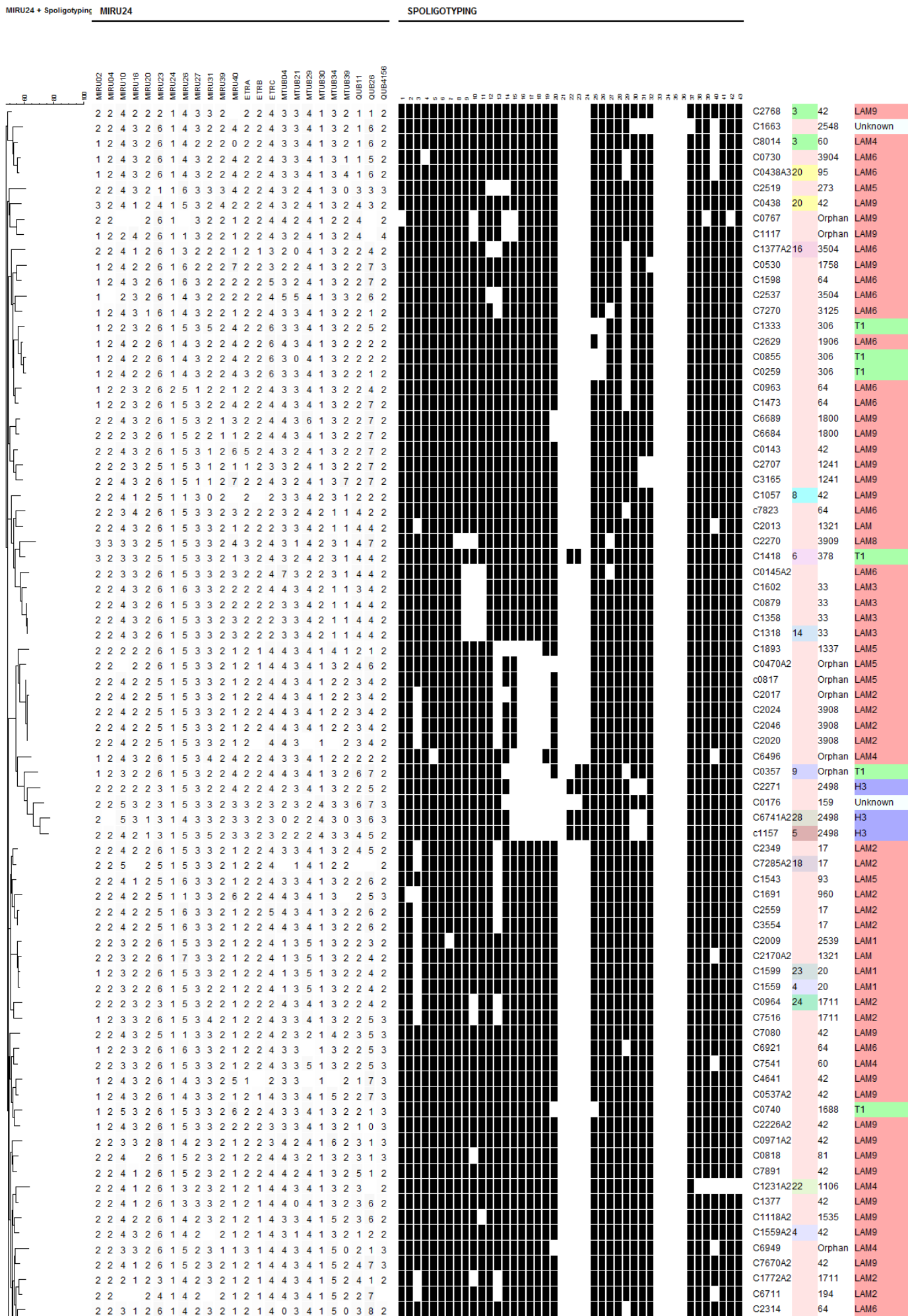
Conceived and designed the experiments: SEGV HMG ARS NB PNS. Performed the experiments: SEGV HMG ARS NB PNS. Analyzed the data: SEGV CCA HMG LLG ECC MIA DC MKG. Contributed reagents/materials/analysis tools: PNS KVL MLL FT PCSC NR CS. Wrote the paper: SEGV CCA PNS NR CS.

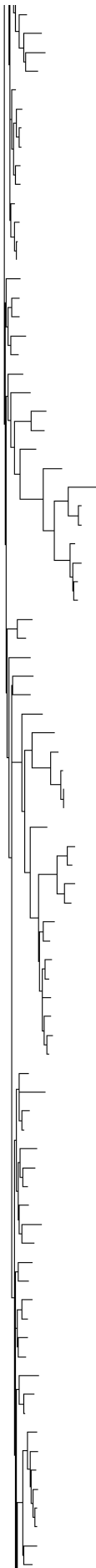
References

- Secretaria de Vigilância em Saúde – Ministério da Saúde, Especial Tuberculose (2012) Boletim Epidemiológico, vol. 43.
- Lazzarini LC, Huard RC, Bocchat NL, Gomes HM, Oelemann MC, et al. (2007) Discovery of a novel *Mycobacterium tuberculosis* lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. *J Clin Microbiol* 45: 3891–3902.
- Lazzarini LC, Spindola SM, Bang H, Gibson AL, Weisenberg S, et al. (2008) RDRio *Mycobacterium tuberculosis* infection is associated with a higher frequency of cavitory pulmonary disease. *J Clin Microbiol* 46: 2175–2183.
- Dalla Costa ER, Lazzarini LC, Perizzolo PF, Diaz CA, Spies FS, et al. (2013) *Mycobacterium tuberculosis* of the RDRio genotype is the predominant cause of tuberculosis and associated with multidrug resistance in Porto Alegre City, South Brazil. *J Clin Microbiol* 51: 1071–1077.
- Weisenberg SA, Gibson AL, Huard RC, Kurepina N, Bang H, et al. (2012) Distinct clinical and epidemiological features of tuberculosis in New York City caused by the RD(Rio) *Mycobacterium tuberculosis* sublineage. *Infect Genet Evol* 12: 664–670.
- David S, Duarte EL, Leite CQ, Ribeiro JN, Maio JN, et al. (2012) Implication of the RD(Rio) *Mycobacterium tuberculosis* sublineage in multidrug resistant tuberculosis in Portugal. *Infect Genet Evol* 12: 1362–1367.
- Gibson AL, Huard RC, Gey van Pittius NC, Lazzarini LC, Driscoll J, et al. (2008) Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio Sublineage of the Latin American-Mediterranean *Mycobacterium tuberculosis* spoligotype family. *J Clin Microbiol* 46: 1259–1267.
- Cardoso Oelemann M, Gomes HM, Willery E, Possuelo L, Batista Lima KV, et al. (2011) The forest behind the tree: phylogenetic exploration of a dominant *Mycobacterium tuberculosis* strain lineage from a high tuberculosis burden country. *PLoS One* 6: e18256.
- Rindi L, Lari N, Garzelli C (2012) 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* 12: 1551–1557.
- de Jong BC, Antonio M, Awine T, Ogungbemi K, de Jong YP, et al. (2009) Use of spoligotyping and large sequence polymorphisms to study the population structure of the *Mycobacterium tuberculosis* complex in a cohort study of consecutive smear-positive tuberculosis cases in The Gambia. *J Clin Microbiol* 47: 994–1001.
- Brudey K, Filliol I, Ferdinand S, Guernier V, Duval P, et al. (2006) Long-term population-based genotyping study of *Mycobacterium tuberculosis* complex isolates in the French departments of the Americas. *J Clin Microbiol* 44: 183–191.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 35: 907–914.
- Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, et al. (2000) Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* 36: 762–771.
- Chacón-Salinas R, Serafin-López J, Ramos-Payán R, Méndez-Aragón P, Hernández-Pando R, et al. (2005) Differential pattern of cytokine expression by macrophages infected in vitro with different *Mycobacterium tuberculosis* genotypes. *Clin Exp Immunol* 140: 443–449.
- Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* 4: e7815.
- Cowan LS, Diem L, Monson T, Wand P, Temporado D, et al. (2005) Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. *J Clin Microbiol* 43: 688–695.
- Dale JW, Nor RM, Ramayah S, Tang TH, Zainuddin ZF (1999) Molecular epidemiology of tuberculosis in Malaysia. *J Clin Microbiol* 37: 1263–1268.
- Kato-Maeda M, Gagneux S, Flores LL, Kim EY, Small PM, et al. (2011) Strain classification of *Mycobacterium tuberculosis*: congruence between large sequence polymorphisms and spoligotypes. *Int J Tuberc Lung Dis* 15: 131–133.
- Abadía E, Sequera M, Ortega D, Méndez MV, Escalona A, et al. (2009) *Mycobacterium tuberculosis* ecology in Venezuela: epidemiologic correlates of common spoligotypes and a large clonal cluster defined by MIRU-VNTR-24. *BMC Infect Dis* 9: 122.
- Cerezo I, Jiménez Y, Hernandez J, Zozio T, Murcia MI, et al. (2012) A first insight on the population structure of *Mycobacterium tuberculosis* complex as studied by spoligotyping and MIRU-VNTRs in Bogotá, Colombia. *Infect Genet Evol* 12: 657–663.
- Gonzalo X, Ambroggi M, Cordova E, Brown T, Poggi S, et al. (2011) Molecular epidemiology of *Mycobacterium tuberculosis*, Buenos Aires, Argentina. *Emerg Infect Dis* 17: 528–531.
- Gomes HM, Elias AR, Oelemann MA, Pereira MA, Montes FF, et al. (2012) Spoligotypes of *Mycobacterium tuberculosis* complex isolates from patients residents of 11 states of Brazil. *Infect Genet Evol* 12: 649–656.
- Demay C, Liens B, Burguière T, Hill V, Couvin D, et al. (2012) SITVITWEB—a publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 12: 755–766.
- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, et al. (2006) *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 6: 23.
- Filliol I, Driscoll JR, Van Soolingen D, Kreiswirth BN, Kremer K, et al. (2002) Global distribution of *Mycobacterium tuberculosis* spoligotypes. *Emerg Infect Dis* 8: 1347–1349.
- Sola C, Filliol I, Gutierrez MC, Mokrousov I, Vincent V, et al. (2001) Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. *Emerg Infect Dis* 7: 390–396.
- Sola C, Filliol I, Legrand E, Lesjean S, Loch C, et al. (2003) Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect Genet Evol* 3: 125–133.
- Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Res* 38: W326–331.
- Allix-Béguec C, Fauville-Dufaux M, Supply P (2008) Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 46: 1398–1406.

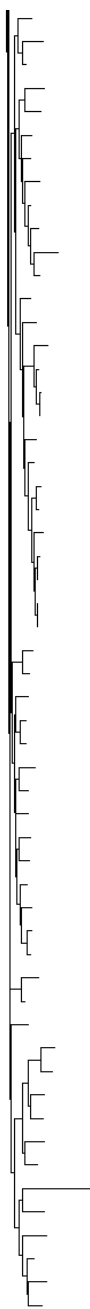
30. García de Viedma D, Mokrousov I, Rastogi N (2011) Innovations in the molecular epidemiology of tuberculosis. *Enferm Infecc Microbiol Clin* 29 Suppl 1: 8–13.
31. Goh KS, Rastogi N, Berchel M, Huard RC, Sola C (2005) Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (*narGHJI*) operon promoter. *J Clin Microbiol* 43: 4010–4014.
32. Gagneux S, Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 7: 328–337.
33. Huard RC, Fabre M, de Haas P, Lazzarini LC, van Soolingen D, et al. (2006) Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. *J Bacteriol* 188: 4271–4287.
34. Vasconcellos SE, Huard RC, Niemann S, Kremer K, Santos AR, et al. (2010) Distinct genotypic profiles of the two major clades of *Mycobacterium africanum*. *BMC Infect Dis* 10: 80.
35. Stucki D, Gagneux S (2013) Single nucleotide polymorphisms in *Mycobacterium tuberculosis* and the need for a curated database. *Tuberculosis (Edinb)* 93: 30–39.
36. Musser JM, Amin A, Ramaswamy S (2000) Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. *Genetics* 155: 7–16.
37. Dos Vultos T, Mestre O, Rauzier J, Golec M, Rastogi N, et al. (2008) Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. *PLoS One* 3: e1538.
38. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, et al. (2012) High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLoS One* 7: e39855.
39. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 31: 406–409.
40. Zhang J, Abadia E, Refregier G, Tafaj S, Boschirolu ML, et al. (2010) *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. *J Med Microbiol* 59: 285–294.
41. Graur D, W-H Li (2000) Dynamics of genes in populations, p. 58. In D. Graur and W.-H. Li (ed.), *Fundamentals of molecular evolution*. Sinauer Associates, Sunderland, Mass.
42. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* 26: 2465–2466.
43. Noguti EN, Leite CQ, Malaspina AC, Santos AC, Hirata RD, et al. (2010) Genotyping of *Mycobacterium tuberculosis* isolates from a low-endemic setting in northwestern state of Paraná in Southern Brazil. *Mem Inst Oswaldo Cruz* 105: 779–785.
44. Mendes NH, Melo FA, Santos AC, Pandolfi JR, Almeida EA, et al. (2011) Characterization of the genetic diversity of *Mycobacterium tuberculosis* in São Paulo city, Brazil. *BMC Res Notes* 4: 269.
45. Miranda SS, Carvalho WaS, Suffys PN, Kritski AL, Oliveira M, et al. (2011) Spoligotyping of clinical *Mycobacterium tuberculosis* isolates from the state of Minas Gerais, Brazil. *Mem Inst Oswaldo Cruz* 106: 267–273.
46. Luiz Ro S, Suffys P, Barroso EC, Kerr LR, Duarte CR, et al. (2013) Genotyping and drug resistance patterns of *Mycobacterium tuberculosis* strains observed in a tuberculosis high-burden municipality in Northeast, Brazil. *Braz J Infect Dis* 17: 338–345.
47. Perizzolo PF, Dalla Costa ER, Ribeiro AW, Spies FS, Ribeiro MO, et al. (2012) Characteristics of multidrug-resistant *Mycobacterium tuberculosis* in southern Brazil. *Tuberculosis (Edinb)* 92: 56–59.
48. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 94: 9869–9874.
49. Mokrousov I, Vyazovaya A, Narvskaya O (2014) *Mycobacterium tuberculosis* Latin American-Mediterranean family and its sublineages in the light of robust evolutionary markers. *J Bacteriol* 196: 1833–1841.
50. Warren RM, Streicher EM, Sampson SL, van der Spuy GD, Richardson M, et al. (2002) Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. *J Clin Microbiol* 40: 4457–4465.
51. Kato-Maeda M, Gagneux S, Flores LL, Kim EY, Small PM, et al. (2011) Strain classification of *Mycobacterium tuberculosis*: congruence between large sequence polymorphisms and spoligotypes. *Int J Tuberc Lung Dis* 15: 131–133.
52. Hill V, Zozio T, Sadikalay S, Viegas S, Streit E, et al. (2012) MLVA based classification of *Mycobacterium tuberculosis* complex lineages for a robust phylogeographic snapshot of its worldwide molecular diversity. *PLoS One* 7: e41991.
53. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, et al. (2006) 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* 188: 759–772.
54. Borile C, Labarre M, Franz S, Sola C, Refrégier G (2011) Using affinity propagation for identifying subspecies among clonal organisms: lessons from *M. tuberculosis*. *BMC Bioinformatics* 12: 224.
55. Borile C, Labarre M, Franz S, Sola C, Refrégier G (2011) Using affinity propagation for identifying subspecies among clonal organisms: lessons from *M. tuberculosis*. *BMC Bioinformatics* 12: 224.
56. Nakanishi N, Wada T, Arikawa K, Millet J, Rastogi N, et al. (2013) Evolutionary robust SNPs reveal the misclassification of *Mycobacterium tuberculosis* Beijing family strains into sublineages. *Infect Genet Evol* 16: 174–177.
57. Stavrum R, Mphahlele M, Ovreås K, Muthivhi T, Fourie PB, et al. (2009) High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates. *J Clin Microbiol* 47: 1848–1856.
58. Lopes JS, Marques I, Soares P, Nebenzahl-Guimaraes H, Costa J, et al. (2013) SNP typing reveals similarity in *Mycobacterium tuberculosis* genetic diversity between Portugal and Northeast Brazil. *Infect Genet Evol* 18: 238–246.
59. Chuang PC, Chen YM, Chen HY, Jou R (2010) Single nucleotide polymorphisms in cell wall biosynthesis-associated genes and phylogeny of *Mycobacterium tuberculosis* lineages. *Infect Genet Evol* 10: 459–466.
60. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, et al. (2013) Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 10: e1001387.
61. Ford C, Yusim K, Ioerger T, Feng S, Chase M, et al. (2012) *Mycobacterium tuberculosis*–heterogeneity revealed through whole genome sequencing. *Tuberculosis (Edinb)* 92: 194–201.
62. Lazzarini LC, Rosenfeld J, Huard RC, Hill V, Lapa e Silva JR, et al. (2012) *Mycobacterium tuberculosis* spoligotypes that may derive from mixed strain infections are revealed by a novel computational approach. *Infect Genet Evol* 12: 798–806.
63. Von Groll A, Martin A, Felix C, Prata PF, Honscha G, et al. (2010) Fitness study of the RDRio lineage and Latin American-Mediterranean family of *Mycobacterium tuberculosis* in the city of Rio Grande, Brazil. *FEMS Immunol Med Microbiol* 58: 119–127.
64. Vinhas SA, Palaci M, Marques HS, Lobo de Aguiar PP, Ribeiro FK, et al. (2013) *Mycobacterium tuberculosis* DNA fingerprint clusters and its relationship with RD(Rio) genotype in Brazil. *Tuberculosis (Edinb)* 93: 207–212.

Figure S1: Dendrogram constructed with BioNumerics software version 6.6 with the results of MIRU-VNTRs 24 loci and spoligotyping by similarity coefficient for categorical data and the neighbor-joining algorithm. 1st column (after spoligotypes): number of isolated label; 2nd column: patients who have more than one isolate in the study (n = 27) received numbering 1–27, and the different strains present the same numbering; 3rd column: International Spoligotype Types (SIT); 4th column: classification obtained through SITVITWEB (family and subfamily).

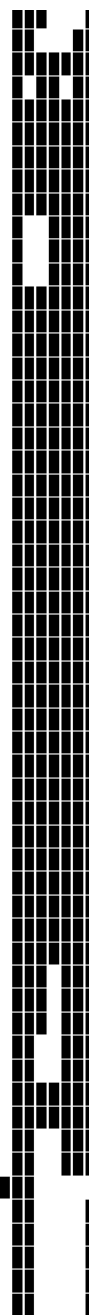
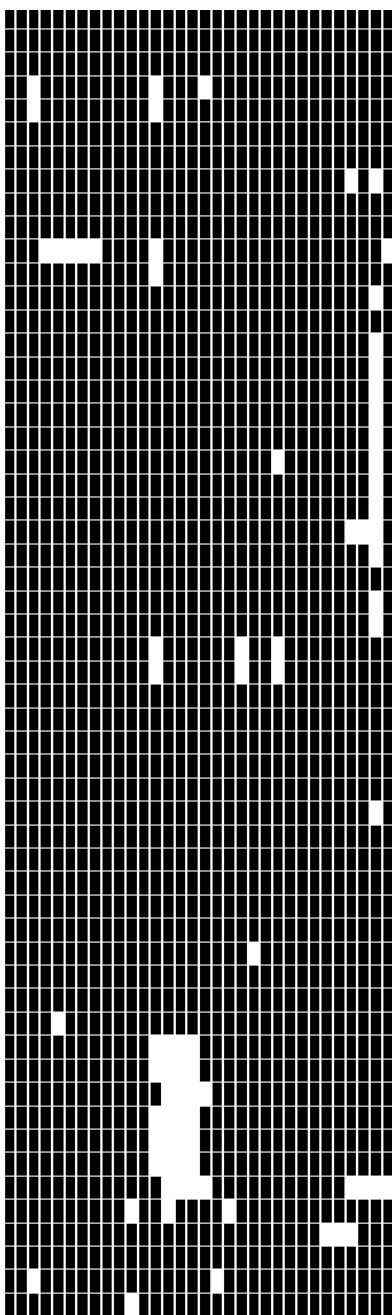




2	2	2	1	2	4	1	4	2	3	2	1	2	1	4	4	3	4	1	5	2	4	7	2
2	2	2	1	2	4	1	4	2	3	2	1	2	1	4	4	3	4	1	5	2	4	4	2
2	2	3	1	2	6	1	4	2	3	2	1	2	1	4	3	3	4	1	4	2	2	8	2
2	2	2	1	2	4	1	4	2	3	2	1	2	1	4	4	3	4	1	5	2	4	7	2
2	2	4	2	2	6	1	5	3	3	2	1	2	1	4	4	3	4	1	4	2	3	1	2
2	2	4	2	2	6	1	5	3	3	2	1	2	1	4	4	3	4	1	2	3	2	7	2
2	2	4	2	2	6	1	5	3	3	2	1	2	4	4	3	3	4	1	5	2	3	7	2
2	4	2	2	6	1	5	3	3	2	1	2	1	4	4	3	4	1	5	2	3	7	2	
2	2	4	2	2	6	1	5	3	3	2	5	2	1	4	5	3	4	1	5	3	1	2	
2	2	5	2	2	6	1	5	3	3	2	3	2	1	4	4	3	4	1	5	2	2	1	2
1	2		2	5	1	4		3	2	1	2	1	4	4	3	4	1		2	3	1	2	
2	2	3	2	2	6	1	4	3	3	2	1	2	1	4	4	3	5	1	5	2	4	5	2
2	2	2	2	2	6	1	4	3	3	2	1	2	1	4	4	3	4	1	6	2	4	1	
2	2	2	2	2	6	1	4	3	3	2	1	2	1	4	4	3	4	1	6	2	4	2	
2	2	4	1	1	5	1	7	3	3	2	1	2	2	3	4	3	4	1	3		2	6	3
2	2	4	2	2	5	1	6	3	3	2	1	2	2	4	4	3	4	1	3	2	2	6	2
2	2	4	2	2	5	1	3	3	3	3	1	2	2	4	4	3	4	1	3	2	2	2	2
2	2	2	2	2	5	1	5	3	3	2	1	2	2	4	4	3	4	1	2	2	3	4	2
2	2	7	2	2	5	1	6	3	3	2	1	2	2	4	4	3	4	1	3	2	2	6	2
2	2	5	3	1	3	1	4	3	3	2	1	2	2	2	4	2	2	1	3	2	1	5	2
2	2		3	2	3	1	1	3	3	2	3	3	2	3	2	2	4	3	3	2	5	2	
2	2	5	3	1	3	1	5	3	3	2	3	3	2	3	1	3	2	4	3	2	2	3	3
2	2	2	3	2	5	1	4	3	3	2	1	3	2	3	2	4	4	3	3	3	1	3	
2	2	3	1	2	5	1	5	3	3	2	1		1	4		2	4	1	4	2	4	5	2
2	2	3	3	2	5	1	5	3	3	2	1	2	4	2	4	3	4	1	2	2	3	4	2
2	2	3	3	2	5	1	8	3	5	3	3		2	0	4	4		3	2	6	1		
2	2	5	3	2	3	1	5	3	3	2	3	3	2	3	2	3	2	4	3	3	3	1	3
2	2	5	3	2	3	1	5	3	3	2	3	2	2	3	1	3	2	4	3	3	3	7	3
3	3	3	3	2	5	1	6	3	2	2	2	4	2	4	3	1	4	1	3	3	3	1	2
3	3	3	2	2	5	1	6	3	3	2	2	4	3	4	3	1	2	2	2		3	5	
3	3	3	3	2	5	1	6	3	2	2	2	4	2	4	3	1	4	2	3	0	3	0	3
3	3	3	2	2	5	1	6	3	2	2	2	4	2	4	3	1	4	2	3	0	3	7	2
2	2	3	3	2	5	1	2	3	3	2	1	2	2	4	2	2	3	2	3	7	2	6	2
2	2		2	2	3	1	5	3		2	1	2	2	4	4	2	4	1	3	2	2	3	3
2	2	5		2	3	2	2	3	1	2	1	2	3	4	2	4	1	1	4	2	2	4	
2	2	4		2	5	1	5	3	3	2	1	3	1	4	1	4	1		2	2	8	2	
2	2	3	1	2	6	1	4	2	3	2	1	2	1	3	5	2	4	1	4	2	2	4	2
2	2	3	2	2	6	1	5	3	3	2	1	1	1	4	3	2	4	1	4	2	3	6	3
2	2	3	3	2	5	1	4	3	3	2	5	2	1	4	2	1	4	2	2	3	3	2	
2	2	4	2	2	6	1	3	3	3	2	4	2	2	4	4	2	4	2	1	4	3	7	0
2	2	4	3	2	6	1	3	3	3	2	3	2	4	4	2	4	2	1	1	3	3	2	
2	2	4	3	2	6	1	3	3	3	2	3	2	4	4	2	4	2	1	1	3	3	2	
2	2	4	3	2	6	1	3	3	3	2	3	2	4	4	2	4	2	1	1	3	3	2	
2	2	4	3	2	6	1	3	3	3	2	3	2	4	4	2	4	2	1	1	3	3	2	
2	2	4	3	2	3	1	5	3	3	2	1	2	2	3	3	3	2	1	3	2	4	5	3
2	2	5	3	1	3	3	5	3	3	2	3	3	2	3	2	2	2	4	3	3	3	5	2
2	2	5	3	1	3	1	5	3	3	2	1	3	2	3	2	2	2	4	3	3	1	5	2
2	2	2	2	2	5	1	3	3	3	2	3	3	2	3	5	2	4	4	3	0	2	3	3
2	2	5	3	2	5	1	3	3	3	2	3	3	2	3	4	2	4	2	3	3	2	4	2
3	2	4	3	2	2	1	6	3	3		3	3	2	3	2	2	2	4	3	7	5	1	2
2	2		3	2	4	1	5	3	3	2	2	3	2	3	2	2	2	4	3	3	2	1	3
2	2	5	3	1	3	1	5	2	3	2	3	3	2	3	2	2	2	4	3	3	3	5	2
2	2	5	3	1	3	1	5	3	3	2	1	3	2	3	2	2	2	4	3	3	1	5	2
2	2	5	3	1	3	1	5	3	3	2	2	3	2	3	2	2	2	4	3	3	1	5	2
2	2	3	3	1	3	1	5	3	3	2	2	3	2	3	2	1	2	4	3	3	4	7	3
2	2	5	3	2	3	2	5	3	3	2	3	3	2	3	2	3	2	4	3	3	6	7	3
2	2	5	3	1	3	1	6	3	3	2	3	3	2	3	2	2	2	4	3	0	5	7	3
2	2	5	3	2	3	1	6	3	3	2	3	3	2	3	2	2	2	4	3	3	5	7	2
2	2		1		1	4	1	2		2	5	2	2	4	5	4	4	1	3	2	2		2
2	2	4	0	1	3	1	5	3	3	2	1	2	2	3	2	2	4	1	3	2	2	4	2
2	2	5	3	2	3	1	5	3	2	2	3	2	2	4	4	2	2	1	3	2	2	5	2
2	2	4		2	3	1	5	3	2	2	1	2	2	4	2	2	4	1	3	2		5	
1	2	4	3	1	6		5	3	2	2	1	2	2	3	2	3	2	1	3	2	3	3	1
1	2	4	3	1	6	1	3	3	2	2	2	1	3	4	3	3	2	1	3	2	2	3	2
2	2	4	3	1	6	1	5	3	2	2		1	2	2	4	3	4	1	3	2	2	1	2
2	2	4	3	1	6	1	5	3	2	2	6	2	2	4	4	3	4	2	3	2	2	4	2
2	2	4	3	1	6	3	5	4	2	2	7	2	2	4	4	5	4	4	3	3	3	2	2
2	2	3	3	1	6	1	5	3	4	2	3	2	2	4	4	3	4	2	1	1	3	4	2
2	2	4	2	2	5	1	5	3	3	6	1	2	2	2	4	1	4	1	3	3	2	7	
1		3	3	2	5	1	5	3	3	2	2	2	2	4	4	3	4	1		2	2	3	2
2	2	3	3	2	6	1	5	3	3	2	1	2	1	4	4	3	4	1		3		2	3
2	2	3	3	2	5	1	5	3	3	2	1	2	1	4	4	2	4	1	5		3	6	3
2	2	4	2	2	6	1	5	3	3	2	1	1	1	4	3	3	4	1	4	2	4	1	3
2	2	4	2	2	6	1	5	3	3	2	1	2	1	4	4	3	4	1	5	2	3	6	2
2	...	5	1	2	5	1	1	3	0	2	2	2	1	3	2	3	4	2	4	2	2	4	3
1	2		2		1	2	3	3		4	2	1	3	4	2	4	2	1	3	3	5	2	
	2	3	3	2	5	1	2	3	3	2	4	2		4	2	2		2					
2	2	4	1	2	5	1	5	2	2	2	1	3	3	4	3	2	4	1	3	2	4	7	3
2	2	4	3	2	5	1	6	3	4	2	1	3	3	4	3	2	4	4	3	2	2	8	3
2	2	4	3	2	5	1	5	3	2	2	1	3	3	3	3	2	4	4	3	2	2	5	3
2	2	4	3	2	5	1	6	3	2	2	1	3	3	4	3		4	4	5	2	1	6	3
2	2	4	3	2	5	1	6	3	2	2	1	1	3	4	3	2	4	4	3	2	3	5	3
2	2	4	3	2	5	1	6	3	2	2	1	3	3	4	3	2	4	4	3	2	3	1	3
2	2		3	4	5	2	6	1	3	2	1	3		4	3	2	4	2	3		3	7	4
2	2	4	3	2																			



2	2	2	3	2	5	1	5	3	3	2	1	2	2	3	2	2	4	1	3	3	2	3	
5	2	5	3	2	2	1	5	3	3	2	4	3	2	3	2	2	4	4	3	3	4	6	3
2	2	5	3	2	5	1	5	3	3	2	4	3	2	3	2	2	4	1	3	2	4	6	2
2	2	4	3	2	5	1	5	3	3	2	3	3	2	4	2	2	6	4	3	3	3	4	
2	2	5	3	2	5	1	3	3	2	2	3	3	2	3	4	2	2	4	3	3	2	4	3
2	2	4	3	2	5	1	1	3	2	2	3	3	3	3	2	3	4	4	3	2	3	7	3
2	2	3	3	2	5	1	5	3	3	2	3	3	2	3	2	3	4	4	4	3	3	7	1
2	1	6	3	2	5	1	5	3	3	2	3	3	2	3	2	4	4	4	3	3	2	3	3
2	2	4	3	2	5	1	4	3	3	2	3	3	2	3	2	4	4	4	3	3	8	3	
2	2	4	3	2	5	1	4	3	3	2	3	3	2	3	2	4	4	4	3	3	5	7	
2	2	4	3	2	5	1	4	3	3	2	3	3	2	3	2	3	4	2	3	3	6	8	3
2	2	4	3	2	5	1	4	3	3	2	2	3	2	3	2	4	4	2	3	3	6	2	2
2	2	5	3	2	3	1	5	3	3	2	1	3	2	3	2	2	2	1	3	1	2	4	2
2	2	5	4	1	3	1	5	3	3	2	3	3	2	3	1	2	2	4	3	2	7	3	
2	2	5	3	2	3	3	5	2	1	1	1	3	3	3	2	3	2	4	3	3	6	7	3
2	2	3	3	1	3	1	5	3	3	1	1	3	2	3	2	3	2	4	3	2	6	7	3
2	2	5	3	1	3	1	5	3	3	1	1	3	2	3	2	3	2	4	3	3	6	7	3
2	2	5	3	1	3	1	5	3	3	1	1	3	2	3	2	3	2	4	3	3	3	7	3
2	2	5	3	1	3	1	5	3	2	2	3	2	2	4	4	2	2	4	3	3	5	7	2
2	2	5	2	2	3	1	5	3	3	2	3	3	2	3	2	2	2	4	3	4	7	2	
3	2	5	3	1	2	1	2	3	3	2	2	3	2	3	2	2	2	4	3	3	4	2	
2	2	5	3	1	3	1	2	3	3	2	2	3	2	3	2	2	2	4	3	3	2	7	2
2	2	5	3	1	3	1	5	3	3	2	3	3	2	3	2	2	2	4	3	3	4	7	2
2	2	5	3	1	3	1	5	3	2	2	3	3	2	3	2	2	2	4	3	3	5	7	2
2	2	5	3	1	3	1	5	3	2	2	3	3	2	3	2	2	2	4	3	3	5	7	2
2	2	2	3	1	3	1	5	3	3	2	3	3	2	3	2	2	2	4	3	3	5	7	2
2	2	4	3	1	3	1	5	3	3	2	3	3	2	3	2	2	2	4	3	3	5	7	2
2	2	2	1	2	6	1	5	3	3	2	4	3	2	5	2	2	4	2	3	3	2	5	2
2	2	3	3	2	6	1	6	3	3	2	3	3	2	4	2	2	4	2	3	2	5	3	
2	2	2	1	2	5	1	5	3	2	2	3	3	2	4	2	1	4	3	4	2	2	2	
2	2	3	3	2	5	1	4	3	3	2	1	3	2	4	2	5	4	2	3	2	2	2	3
2	2	3	3	2	5	1	1	3	3	2	1	3	2	4	2	3	4	2	3	0	2	2	2
2	...	3	3	2	6	1	5	3	3	2	2	3	2	2	3	4	4	2	1	...	3	5	3
2	2	3	3	2	5	1	5	3	3	2	2	3	2	4	3	8	4	2	3	0	4	5	2
2	2	3	3	2	5	1	5	3	3	3	3	2	4	1	2	4	3	4	2	3			
2	2	4	2	2	6	1	1	3	3	2	3	3	2	4	2	2	4	2	1	1	4	2	2
2	2	3	3	2	5	1	5	3	3	2	4	3	2	5	2	2	4	2	1	2	4	2	2
2	2	3	5	2	5	1	5	3	3	2	4	2	2	4	2	2	4	2	3	3	5	5	2
2	2	4	4	2	5	1	6	3	3	2	4	3	2	4	2	2	4	2	3	4	4	5	4
2	2	3	1	2	5	1	2	3	3	2	4	3	2	5	2	2	4	2	3	3	3	5	2
2	2	3	1	2	5	1	2	3	3	2	4	3	2	4	2	2	4	2	3	3	3	5	4
2	2	2	2	1	4	3	5	2	3	2	4	2	2	2	3	4	2	3	0	3	6	3	
2	2	3	3	2	6	1	5	3	3	2	4	2	2	2	3	3	4	2	3	3	6	3	
2	2	3	3	2	5	1	4	3	3	2	3	2	2	4	3	2	4	2	3	2	5	7	2
5	3	4	2	2	1	1	5	3	3	3	4	4	2	4	4	1	4	2	3	4	5	3	2
2	3	4	4	2	5	1	5	3	4	2	5	4	2	4	4	1	4	2	3	5	5	3	3
2	1	4	2	5	1	4	3	3	2	3	4	2	3	4	1	4	2	3	3	2	4	3	
2	3	3	3	2	5	1	4	3	3	2	3	4	2	4	4	1	4	2	3	1	3	8	
2	2	3	3	2	5	1	2	3	3	2	4	2	2	4	2	2	3	2	3	7	2	6	2
2	2	4	3	2	6	1	5	3	2	6	1	2	4	0	2	4	1	3	2	2	7	2	
2	5	4	3	2	6	2	2	3	5	1	3	2	4	2	3	8	3	2	3	2	3	3	1
1	4	2	7	1	5	3	2	2	2	2	2	2	4	3	3	4	1	3	2	2	3	3	
1	2	4	3	2	6	1	5	3	3	2	1	2	2	4	1	4	4	4	3	0	2	1	4
2	2	4	3	2	5	1	5	3	3	2	3	3	2	4	1	4	4	1	3	3	2	3	
2	2	4	3	2	5	1	5	3	3	2	3	3	2	4	2	2	2	4	3	4	3	1	1
2	2	2	3	2	5	1	5	3	2	2	3	3	2	4	2	4	4	4	3	2	3	4	3



C7765	1926	T1	
c0635	521	T1	
C0665	53	T1	
C6697	25	Orphan	T3
C1579	1163	T3	
C7321	8	53	T1
C2104	53	T1	
C1926	746	H3	
C0395	53	T1	
C0837	65	T1	
C1171	27	3906	T3
C0982	157	T3	
C1842	50	H3	
C1120	53	T1	
c0821	50	H3	
C0825	50	H3	
C0859	50	H3	
C0888	50	H3	
c0317	9	50	H3
C2213A2	741	H3	
C2142A2	11	50	H3
C2142	11	50	H3
C0766	777	H4/Ural-1	
C1433	27	50	H3
C1427	5	53	T1
C1999	50	H3	
C2454	50	H3	
C2144	1227	T5-Madrid2	
C2159	1227	T5-Madrid2	
C4509	53	T1	
C7448	53	T1	
C8190	53	T1	
C0874	53	T1	
C0459	53	T1	
c0627A2	17	50	H3
C7829	53	T1	
C8123	53	T1	
C1130	6	53	T1
C0746	53	T1	
C1696	53	T1	
C7157	291	T1	
C0807	52	T2	
C0696	52	T2	
C1210	153	T2	
C2463	12	3907	T1
C2484	12	3907	T1
C1839	189	T1	
C1474	102	T	
C3501	3907	T1	
C1378	10	3907	T1
C7674	Orphan	EAI5	
C0901	Orphan	T4	
C0517	Orphan	T1	
C0627	17	244	T1
C7195	25	18	X2
C1754	Orphan	T1	

Figure S2: Allelic diversity of the 24 loci MIRU-VNTR loci in LAM/RDRio isolates.

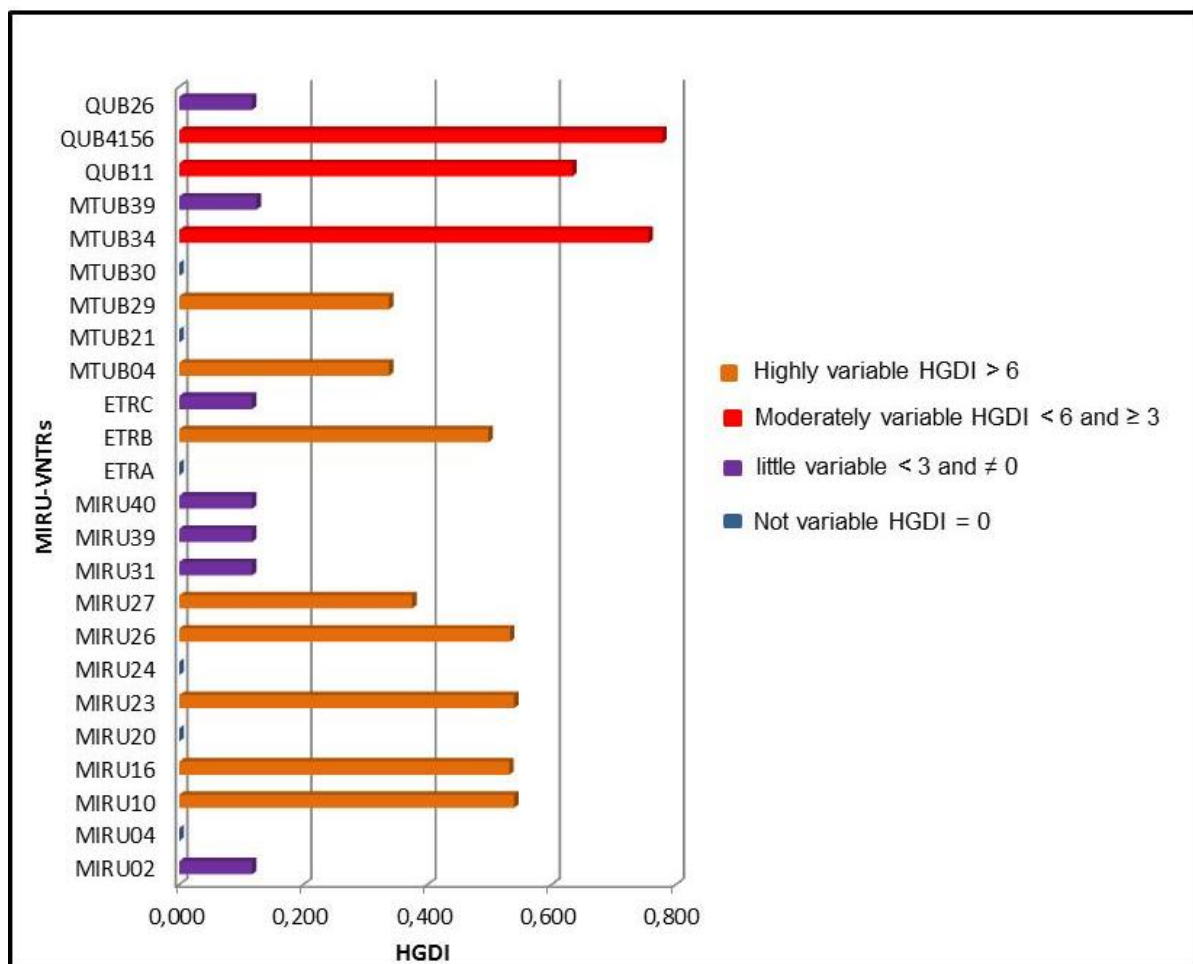
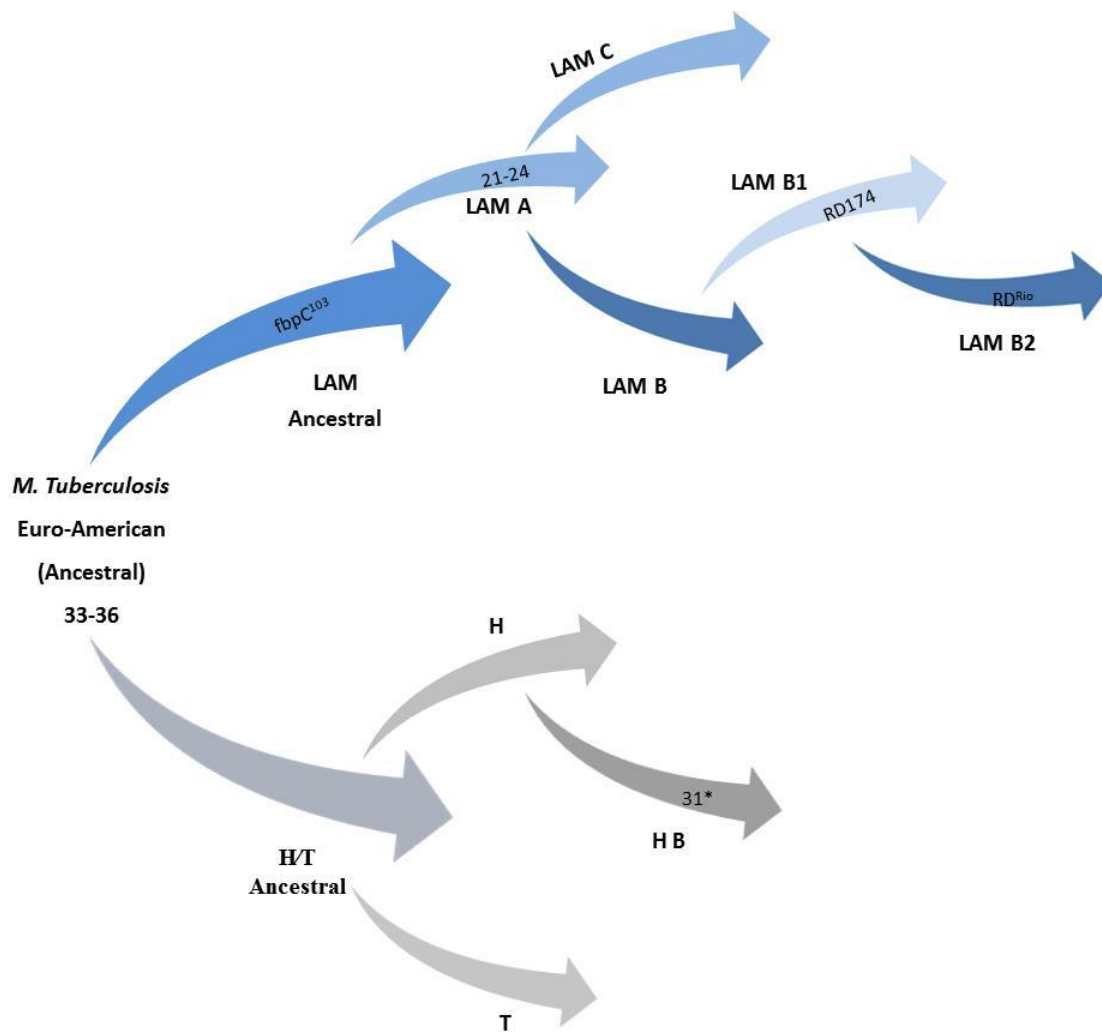


Figure S3: Possible evolution of *M. tuberculosis* lineage Euro-American (LAM, T and H) according to the markers analyzed in this study. The Euro-American Ancestral evolved into two distinct groups: LAM ancestral and T/H ancestral, characterized by absences of spacers 33–36 and one copy of MIRU24 that is common to all Euro-American lineages. The ancestral LAM strains have *fbpC103* and this was the basis of LAM A (LAM9), with additional absence of spacers 21–24, one copy of MTUB30 and two copies of MIRU04 and ETR-A. LAM A on its turn is the basis for two other groups: LAM B (LAM9-LAM4, LAM1-LAM2-LAM5 and LAM6) and LAM C (LAM3). The LAM B evolved from LAM B1, characterized by a deleted RD174 and on its turn to LAM B2, with both deleted RD174 and RDRio. The H/T Ancestral lineage is the origin of both groups H and T (difference only in MIRU-VNTR copies), showing absence of spacers 33–36; additional loss of spacer 31 led to subtype H A, observed in high frequency in this study.



Supplemental Table S1: Detailed genotyping results and associated demographic, epidemiologic and nomenclature information on 218 *M. tuberculosis* strains isolated from pulmonary tuberculosis patients in Rio de Janeiro, Brazil.

IsoNumber	Year	Strain	Spoligotype Description	Octal code	SIT SITVIT2	Clade SITVIT2	SpotClust	MIRU- VNTR	VNTR	5-VIT	MIRU12	12-MIT	MIRU15	15-MIT	MIRU24	24-MIT	Drug-R Code	Sex/A ge	MIT	VIT	MLVA 15- 9
BRA1620090C4772	2009	C4772	□□□□□□□□□□□□	00000000003771	1	BEIJING	ND	BEIJING	32323	102	124326153424	Orphan	2435444242223312	Orphan	132615342422434122	Orphan	0	M/13	84	ND	ND
BRA1620080C7674	2008	C7674	■■■■■■■■■■■■■■■	777741777143600	Orphan	EAI5	EAI5	EAI-5	33422	491	234226153321	Orphan	342531243724312	Orphan	122615332124434152	Orphan	0	M	64	ND	ND
BRA1620090C2754	2009	C2754	■■■■■■■■■■■■■■■	777777770020771	742	H	H	H	32323	102	222226143321	1660	242431244-14312	Orphan	222614332124434162	Orphan	0	M	ND	97	ND
BRA1620080C3275	2009	C3275	■■■■■■■■■■■■■■■	777777770020771	742	H	ND	H	32323	102	224-26153321	Orphan	24-53143214312	Orphan	1-26153321224432132	Orphan	0	M	ND	97	ND
BRA1620080C0462	2008	C0462	■■■■■■■■■■■■■■■	777777774020771	47	H1	ND	H	21423	99	225315153323	42	253533332331342	Orphan	3131533323323132432	Orphan	0	F/28	ND	97	ND
BRA1620082C2139	2008	C2139	■■■■■■■■■■■■■■■	777777774020771	47	H1	ND	H	24255	Orphan	224226113323	Orphan	242133344222221	Orphan	12261133233234224211	Orphan	2	F	ND	97	ND
BRA1620080C6818	2008	C6818	■■■■■■■■■■■■■■■	777777774020771	47	H1	ND	H	32423	21	224226153321	25	242531144133312	Orphan	1226153321114343412	Orphan	0	F	ND	97	ND
BRA1620081C7117	2008	C7117	■■■■■■■■■■■■■■■	777777774020771	47	H1	ND	H	32423	21	224326153321	1286	243531224423321	Orphan	1326153321222334211	Orphan	1	M	ND	97	ND
BRA1620080C7717	2008	C7717	■■■■■■■■■■■■■■■	777777774020771	47	H1	ND	H	32422	491	124-27153222	Orphan	24-522242233312	Orphan	1-2715322224334132	Orphan	0	F	ND	97	ND
BRA1620080C6644	2008	C6644	■■■■■■■■■■■■■■■	717777774020771	431	H1	ND	LAM	32323	102	223326153321	224	23353124-234313	Orphan	326153321244341-3	Orphan	0	M	ND	177	ND
BRA1620080C0991	2008	C0991	■■■■■■■■■■■■■■■	777777774020771	Orphan	H1	H1	H	22323	92	223325123324	1650	233234242622227	Orphan	1325123324224223237	Orphan	0	M/25	174	97	ND
BRA1620080C6918	2008	C6918	□□□□□□□□□□□□	000000004020771	2	H2	ND	H	32423	21	224326153322	1062	243532224423321	Orphan	1326153322222334211	Orphan	0	F/35	43	177	ND
BRA1620080C7285	2008	C7285	□□□□□□□□□□□□	000000004020771	2	H2	ND	H	32523	55	124226162227	Orphan	242627232732212	Orphan	1226162227223224132	Orphan	0	M/43	43	97	ND
BRA1620080C0821	2008	c0821	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	LAM	22223	105	22212442321	1648	221431244724312	584	112414232121434152	232	0	M/39	ND	ND	ND
BRA1620080C1755	2008	C1755	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	T	22423	9	224225163321	130	242631242624312	598	1225163321224434132	234	0	F/43	ND	ND	ND
BRA1620080C0825	2008	C0825	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	22223	105	22-2412-21	Orphan	2-4-12427-4312	Orphan	--24142-21214434152	Orphan	0	F	ND	97	ND
BRA1620082C0859	2008	C0859	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	22223	105	214-25143323	Orphan	14-433432434123	Orphan	+25143324323414233	Orphan	2	M	41	97	ND
BRA1620080C0888	2008	C0888	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	22321	Orphan	223325113321	Orphan	233131342222320	Orphan	1325113321324234230	Orphan	0	M	41	97	ND
BRA1620080C1433	2008	C1433	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	22423	9	223325153323	116	233533333712343	Orphan	1325153323323234433	Orphan	0	F/26	810	197	6904-76
BRA1620080C1842	2008	C1842	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	22423	9	224325163221	949	243621341633-42	Orphan	13251632213343-4452	Orphan	0	M	ND	97	ND
BRA1620080C1999	2008	C1999	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	22423	9	234425153425	Orphan	344545445334125	Orphan	1425153425442414235	Orphan	0	F/52	ND	97	ND
BRA1620090C2142	2009	C2142	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	24423	488	224126132321	579	241331243-24312	Orphan	1126132321214434132	Orphan	0	M	ND	97	ND
BRA1620090C2142A2	2009	C2142A2	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	2-220	Orphan	224126132321	1460	24132123242012	Orphan	1126132221213204132	Orphan	0	M	ND	97	ND
BRA1620090C2454	2009	C2454	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	32322	109	224226143321	1658	24434124362312	Orphan	1261442321214334152	Orphan	0	M	ND	240	97
BRA1620080C0627A2	2008	c0627A2	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	H	21423	99	225413153323	Orphan	254533337-31242	Orphan	3413153323323122432	Orphan	0	F	ND	ND	ND
BRA1620090C2485	2009	C2485	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	LAM	32323	102	124226143224	1659	242424262223012	Orphan	1226143224226304132	Orphan	0	M	ND	ND	ND
BRA1620090C2767	2009	C2767	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	LAM	32323	102	223226143321	10	23241244524312	Orphan	1226143321214351512	Orphan	0	M	ND	229	ND
BRA1620080C0317	2008	c0317	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	LAM	21323	137	222313153323	Orphan	223533335722243	Orphan	1313153323323224233	Orphan	0	F/41	810	191	ND
BRA1620080C1966	2008	C1966	■■■■■■■■■■■■■■■	777777777720771	50	H3	ND	LAM	22423	9	233325143323	1379	3334334438-4121	Orphan	332514332434414231	Orphan	0	M/26	25	198	ND
BRA1620080C7307	2008	C7307	■■■■■■■■■■■■■■■	777776777720771	268	H3	T1	LAM	32523	55	124326163222	328	243622252723212	Orphan	1326163222225324132	Orphan	0	M/28	ND	235	ND
BRA1620090C2213A2	2009	C2213A2	■■■■■■■■■■■■■■■	77777775720771	741	H3	ND	H	2-423	489	224326133323	251	24333234324221	599	13261333232-4424211	235	0	M	607	97	ND
BRA1620082C1926	2008	C1926	■■■■■■■■■■■■■■■	77777777520771	746	H3	ND	H	22423	9	227225163321	Orphan	272631242624312	Orphan	1225163321224434132	Orphan	2	F	ND	184	ND
BRA1620080C1914	2008	C1914	■■■■■■■■■■■■■■■	77777777520771	746	H3	ND	LAM	22423	9	224425163324	Orphan	244634344542224	Orphan	142516332434242434	Orphan	0	F/60	ND	191	ND
BRA1620080C1157	2008	c1157	■■■■■■■■■■■■■■■	777760077620771	2498	H3	ND	H	22422	94	224325143323	30	243433335732443	Orphan	13251433233232344433	Orphan	0	M	ND	ND	ND
BRA1620082C6741A2	2008	C6741A2	■■■■■■■■■■■■■■■	777760077620771	2498	H3	ND	H	32423	21	224226153321	25	242531244523312	Orphan	1226153321224334132	Orphan	2	F	ND	ND	ND
BRA1620080C2271	2008	C2271	■■■■■■■■■■■■■■■	777760077620771	2498	H3	ND	LAM	31423	121	223126142321	1262	231431242823312	Orphan	1126142321214334142	Orphan	0	F	ND	191	ND
BRA1620080C6911	2008	C6911	■■■■■■■■■■■■■■■	777776737720771	Orphan	H3	T1	LAM	32423	21	243516542723212	1663	243516542723212	Orphan	1326153126524324132	Orphan	0	M	ND	230	5320-766
BRA1620090C0766	2009	C0766	■■■■■■■■■■■■■■■	77777777420771	777	H4/Ural-1	ND	H	21423	99	225323153323	43	253533233731343	Orphan	1323153323223132433	Orphan	0	F	42	97	ND
BRA1620090C2170A2	2009	C2170A2	■■■■■■■■■■■■■■■	6777760760731	1321	LAM	ND	LAM	2-423	489	224326133323	251	24333234324221	599	13261333232-4424211	235	0	M	ND	9	ND
BRA1620080C2013	2008	C2013	■■■■■■■■■■■■■■■	6777760760731	1321	LAM	ND	LAM	22424	112	333225163322	Orphan	332632443-5312	Orphan	322516332243431222-	Orphan	0	F	ND	229	1386-259
BRA1620090C145A2	2009	C0145A2	■■■■■■■■■■■■■■■	7761760676071	2310	LAM	ND	LAM	1242-	Orphan	52532323324	Orphan	253534334632243	Orphan	322153324323224433	Orphan	0	F	236	9	MLVA15-9
BRA1620080C1509	2008	C1559	■■■■■■■■■■■■■■■	6777760760771	20	LAM1	ND	LAM	22423	9	224225153321	127	242531243424312	597	142251533212-443-1-23	Orphan	0	F	492	191	ND
BRA1620080C1599	2008	C1599	■■■■■■■■■■■■■■■	6777760760771	20	LAM1	ND	LAM	22423	9	224225153321	1652	24352132532442	Orphan	1325153221333234432	Orphan	0	F	ND	9	ND
BRA1620090C2009	2009	C2009	■■■■■■■■■■■■■■■	6777760760771	2539	LAM1	LAM1/LAM9	LAM	22424	112	333325153324	397	333325153324	Orphan	1325153324324314231	Orphan	0	M	128	9	ND
BRA1620083C2349	2008	C2349	■■■■■■■■■■■■■■■	67773760760771	17	LAM2	ND	LAM	32313	117	223126142321	1262	231431243830301	Orphan	1126142321214034150	Orphan	3	M	25	9	ND
BRA1620090C2559	2009	C2559	■■■■■■■■■■■■■■■	67773760760771	17	LAM2	ND	LAM	32323	102	124326143224	425	243424241523311	Orphan	13261432242424334131	Orphan	0	F/29	130	ND	ND
BRA1620090C3554	2009	C3554	■■■■■■■■■■■■■■■	67773760760771	17	LAM2	ND	LAM	32323	102	123226153321	Orphan	223531242421312	224	1226153321224135132	Orphan	0	Ni	130	9	ND
BRA1620082C785A2	2008	C7285A2	■■■■■■■■■■■■■■■	67773760760771	17	LAM2	ND	LAM	32523	55	122326163321	Orphan	22363124253321	Orphan	32616332122433-132	Orphan	0	M/24	ND	9	ND
BRA1620080C6711	2008	C6711	■■■■■■■■■■■■■■■	67773760760731	194	LAM2	ND	LAM	32422	118	224226153325	Orphan	24253524312531-	Orphan	122615332521453415-	Orphan	0	F/35	ND	ND	ND
BRA1620090C1691	2009	C1691	■■■■■■■■■■■■■■■	47773760760771	960	LAM2	ND	LAM	22423	9	224-25153321	Orphan	24-531342821412	Orphan	4-251533213141441-2	Orphan	0	M	ND	9	ND
BRA1620080C0964	2008	C0964	■■■■■■■■■■■■■■■	67733760760771	1711	LAM2	ND	LAM	22323	92	224325113321	Orphan	243131243532312	Orphan	1325113321224232142	Orphan	0	M/54	ND	225	ND
BRA1620090C1772A2	2009	C1772A2	■■■■■■■■■■■■■■■	67733760760771	1711	LAM2	ND	LAM	22423	9	224325163221	949	243621143533242	Orphan	1325163221134324432	Orphan	0	M	ND	198	ND
BRA1620080C7516	2008	C7516	■■■■■■■■■■■■■■■	67733760760771	1711	LAM2															

BRA1620080C6921	2008	C6921	■■■■■■■■■■	77777760756071	64	LAM6	ND	LAM	32423	21	224362153323	213	24353322443321	603	132615332322234311	236	0	F	ND	9	ND
BRA1620084C0963	2008	C0963	■■■■■■■■■■	77777760756071	64	LAM6	ND	LAM	22322	106	22413622443413	Orphan	24213622443413	Orphan	132615332322234311	Orphan	4	M	ND	191	ND
BRA1620090C1473	2009	C1473	■■■■■■■■■■	77777760756071	64	LAM6	ND	LAM	22423	9	223325153321	401	23353124363421	Orphan	332515332121442415	Orphan	0	M	831	191	6904-76
BRA1620080C1598	2008	C1598	■■■■■■■■■■	77777760756071	64	LAM6	ND	LAM	22423	9	22435332153323	33	24353334213413	Orphan	332515332324144313	Orphan	0	F	ND	ND	6565
BRA1620080C2314	2008	C2314	■■■■■■■■■■	77777760756071	64	LAM6	ND	LAM	322E3	Orphan	223126143231	1262	231431232425212	Orphan	126142321213524142	Orphan	0	M/61	ND	198	ND
BRA1620080C7823	2008	C7823	■■■■■■■■■■	77777760756071	64	LAM6	ND	LAM	3-423	492	22-1-141225	Orphan	2-42542-25412	Orphan	-1-14122524544132	Orphan	0	F	ND	229	ND
BRA1620080C0438A3	2008	C0438A3	■■■■■■■■■■	77777760756071	95	LAM6	ND	LAM	21423	99	224213153523	Orphan	24255334552243	Orphan	121315352332322243	Orphan	0	M/49	425	191	ND
BRA1620080C08131A2	2008	C8131A2	■■■■■■■■■■	77777760756071	95	LAM6	ND	LAM	24234	33	254326225313	64	543253223313822	Orphan	332622351324238322	Orphan	0	F	ND	9	ND
BRA1620080C6582	2008	C6582	■■■■■■■■■■	777760756071	1066	LAM6	ND	LAM	32323	102	223326153323	236	233533244427321	Orphan	332615332322473231	Orphan	0	M	ND	198	ND
BRA1620080C2629	2009	C2629	■■■■■■■■■■	77777760556071	1906	LAM6	ND	LAM	32323	102	22436143325	Orphan	243435121733312	Orphan	43261435213-233-1-21	Orphan	0	M	ND	224	ND
BRA1620080C7270	2008	C7270	■■■■■■■■■■	77773760656071	3125	LAM6	LAM9	LAM	32433	23	22-226153321	Orphan	2-2531244624312	Orphan	226153321214434132	Orphan	0	M/43	ND	191	ND
BRA1620080C1377A2	2008	C1377A2	■■■■■■■■■■	77763760756071	3504	LAM6	LAM9	LAM	24222	94	223125153321	771	231531-4552-212	Orphan	312515332114-24142	Orphan	0	M	ND	214	ND
BRA1620090C2537	2009	C2537	■■■■■■■■■■	77763760756071	3504	LAM6	LAM9	LAM	32323	102	124326143321	Orphan	243431242733312	Orphan	1326143321214334152	Orphan	0	M/68	ND	ND	ND
BRA1620090C0730	2009	C0730	■■■■■■■■■■	73777760756071	3904	LAM6	LAM9	LAM	21423	99	2241352321	Orphan	24-52124-5-2212	Orphan	4-23153221224224132	Orphan	0	F	425	191	ND
BRA1620080C2028	2008	C2028	■■■■■■■■■■	77437760676071	3909	LAM8	ND	LAM	224F2	Orphan	123325153322	1655	233532242324312	Orphan	33251533222244341-2	Orphan	0	M/30	ND	9	4163-328
BRA1620080C2270	2008	C2270	■■■■■■■■■■	77437760676071	3909	LAM8	ND	S	31423	121	124326142230	121	2434202422202434132	Orphan	3326142220224334132	Orphan	0	M	397	23	ND
BRA1620090C0357A2	2009	C0357A2	■■■■■■■■■■	77777760776071	42	LAM9	ND	LAM	21423	99	225313153321	174	2535131531522243	595	331515332132322243	230	0	F/21	ND	198	ND
BRA1620080C1559A2	2008	C1559A2	■■■■■■■■■■	77777760776071	42	LAM9	ND	LAM	24243	9	22425124344312	127	24251234424312	597	122515332124434122	233	0				

Drug-R Code	0 not specified
	1 sensible
	2 INH resistant
	3 MDR
	4 XDR

Supplemental table S2: Detailed genotyping results of *fbpC*¹⁰³, RD174 and RD^{Rio} on 182 *M. tuberculosis* strains isolated from pulmonary tuberculosis patients in Rio de Janeiro, Brazil.

[illegible]

[illegible]

[illegible]

[illegible]