Population structure of multidrug-resistant Mycobacterium tuberculosis clinical isolates in Colombia

Juan Germán Rodríguez-Castillo, Claudia Llerena, Lorena Argoty-Chamorro, Julio Guerra, David Couvin, Nalin Rastogi, Martha Isabel Murcia

Population structure of Multidrug-resistant *Mycobacterium tuberculosis* clinical isolates in Colombia

Juan Germán Rodríguez-Castillo [a1] Claudia Llerena[b1], Lorena Argoty-Chamorro[a2], Julio Guerra[a], David Couvin[c], Nalin Rastogi[c*], Martha Isabel Murcia[a*]

[a] Departamento de Microbiología, Facultad de Medicina, Universidad Nacional de Colombia, Bogotá, Colombia. Carrera 30 45 – 03, Facultad de Medicina, Edifcio 471. Bogotá, D.C. Colombia

[b] Grupo de Micobacterias, Red Nacional de Laboratorios, Instituto Nacional de Salud, Bogotá, Colombia, Avenida calle 26 No. 51-20 - Zona 6 CAN. Bogotá, D.C. Colombia.

[c] WHO Supranational TB Reference Laboratory, Institut Pasteur de la Guadeloupe, Abymes, France Institut Pasteur de Guadeloupe, Morne Jolivière, BP484, F97183 Abymes Cedex, Guadeloupe, France.

[1] Contributed equally.

[2] Present adress: Area de Virología. Instituto Nacional de Salud, Avenida calle 26 No. 51-20 - Zona 6 CAN. Bogotá, D.C. Colombia.

*Corresponding authors:

Martha Isabel Murcia Aranguren: Universidad Nacional de Colombia, Facultad de Medicina, Departamento de Microbiología, Bogotá-Colombia, Carrera 30 #.45-03. E-mail: mimurciaa@unal.edu.co

Nalin Rastogi: WHO Supranational TB Reference Laboratory, Institut Pasteur de la Guadeloupe, F97183 Abymes, Guadeloupe, France. E-mail: nrastogi@pasteur-guadeloupe.fr

Abstract

Emergence of multidrug-resistant (MDR) *Mycobacterium tuberculosis* complex (MTBC) isolates is a major public health problem that threatens progress made in tuberculosis (TB) care and control worldwide. In Colombia, the prevalence of MDR tuberculosis (MDR-TB) has increased slowly but steadily since 2001. However, the population structure of the MDR-TB strains circulating in Colombia is sparsely known. In this work, 203 MDR isolates isolated in 2012-2013 were collected, and characterized by spoligotyping, followed by 24-loci MIRU-VNTR (data available for 190 isolates). The most prevalent genotypes corresponded to SIT42/LAM9 (12.81%), SIT62/H1 (10.34%), and SIT190/Beijing (10.34%). A fine analysis showed that although the MDR strains came from 29 of the 33 departments of Colombia, the distribution of these main lineages was not at random and depended on the city of isolation (p-value <0.000001). Both LAM and Beijing lineage strains were significantly associated with MDR-TB (*p-value* <0.0001): LAM lineage was associated with 2 patterns of MDR, namely combined resistance to INH + Rifampin (HR), and to SHRE (Streptomycin + INH + Rifampin + Ethambutol), while the Beijing lineage strains were essentially associated with MDR (SHRE). Interestingly, distribution of genotypic lineages in function of drug resistance information (e.g. pansusceptible vs. MDR) was different in our setting as compared to other countries in Latin America. However, MIRU-VNTR patterns were unique for all strains, an observation that did not support active transmission of circulating MDR clones.

Key words: Multidrug-resistant, molecular epidemiology, tuberculosis, Colombia.

1. Introduction

In 2018, according to the World Health Organization (WHO), there were approximately about half a million (range: 417 000–556 000) new cases of rifampicin-resistant TB (of which 78% had multidrug-resistant TB [1]. These forms of TB are becoming a worldwide public health problem that requires fast and effective action for its control.

Colombia with a population near to 50 million inhabitants represents a developing country where conditions of poverty and inequity are at the same time, an obstacle, and a challenge to the TB control program. In Colombia, the incidence of TB is 26.9 cases per 100,000 inhabitants and for the year 2018, Public Health Surveillance System (Sivigila) reported 14,338 TB cases, of which 428 were drug-resistant, of these 211 were MDR and RR (rifampicin-resistant) [2,3]. At the national level, the prevalence of MDR-TB was 2.38% between the years 2004-2005 in untreated patients. Since then, there has been an increase in the number of cases of MDR–TB: with 601 cases being diagnosed from 2010 to 2013 [2]. However, considering WHO estimates of 580 new cases of MDR-TB per year [1], the cases effectively diagnosed are probably an underrepresentation of the real situation. Furthermore, monitoring of MDR-TB treatment in Colombia for the year 2009 revealed a relatively lower treatment success rate of 43% [2], which is a matter of concern for TB control in the country. During the period between 2009 and 2019, an increase in drug-resistant TB was observed (from 7 cases in 2009 to 427 cases in 2019), thanks to a better access to molecular drug detection and sensitivity tests, and improvement in the notification of these cases [3].

In such a context, improved epidemiologic surveillance today makes use of molecular markers such as IS*6110*-RFLP, genetic polymorphism based on the variability of the Direct Repeat (DR) locus (spoligotyping), and the mycobacterial interspersed repetitive-unit-variable-number tandem-repeats (MIRU-VNTRs) in order to identify the circulating strains [4,5]. As reviewed recently [6–8], spoligotyping in conjunction with other markers such as Large Sequence Polymorphisms (LSPs) and whole genome sequence techniques has allowed to reveal the worldwide population structure of *M. tuberculosis*. Briefly, the global population structure of the *M. tuberculosis* has been composed by seven LSP-based lineages [9–11] or around 12 spoligotyping based lineages [12,13]. The seven LSP-based lineages are known as Indo-Oceanic (L1), East Asian (L2), Indian-East African (L3), Euro-American (L4), West African I & II (L5, L6) and Ethiopian (L7) [6-8]. In SITVIT databases [12,13], these lineages correspond to the following spoligotyping-defined lineages: L1 corresponds to East-African-Indian (EAI); L2 corresponds to Beijing; L3 corresponds to Central-Asian (CAS); L4 is split in at least five lineages (Haarlem or H, X, Latin-American-Mediterranean or LAM, S, and T); L5 and L6 correspond to AFRI or *M. africanum*; while L7 has been recently defined in the database and corresponds to Ethiopian lineage. More recently, a new lineage named L8 – seemingly restricted to the African Great Lakes region – was described; phylogenomic analysis showed that L8 is a sister clade to the MTBC lineages [14]. Furthermore, another new lineage, L9, was recently discovered [15].

Interestingly, study of global phylogeography of tubercle bacilli not only suggested a co-evolution of MTBC lineages with the human host [11, 14] but further suggested that different *M.*

tuberculosis lineages could have different predispositions for the acquisition of multidrug resistance; e.g., L2/Beijing strains acquired drug resistances in vitro more rapidly than L4/ Euro-American strains [16]. However, studies in human population showed that diverse genotypes have similar frequencies of acquisition of drug-resistant TB [17,18].

Hence, a better knowledge of the genetic characteristics of circulating MTBC strains seems a valid prerequisite not only to better comprehend the prevailing epidemiological snapshot, but also to implement a rational and successful therapy. In this regard, previous studies on MTBC population structure in Colombia showed the high proportion of LAM and Haarlem lineages [19–21] a result that was as expected due to the preponderance of the L4/Euro-American strains in the Americas. Nevertheless, studies analyzing MDR-TB strains in Valle del Cauca, the second department in caseloads in Colombia, indicated that not only LAM but also the Beijing genotype were significantly associated with MDR and XDR-TB isolates [22–25]. Consequently, the present study aimed to provide with an up-to-date inventory of the genetic diversity of MDR-TB isolates circulating in Colombia by studying all such isolates obtained between 2012 and 2013. The spoligotyping-based lineages were further compared to circulating MTBC genotypes not only in Colombia but also in neighboring Latin American countries using an international genotyping database.

2. Materials and Methods

2.1 Ethical considerations

This study provides with data on retrospective genotyping of a total of 203 MDR-TB isolates collected between the years 2012-2013 from 29 of the 33 territorial entities of Colombia. Considering the total of 233 cases of MDR TB reported in Colombia for this period [2] (n=97 in 2012 and n=136 in 2013), the 203 isolates analyzed correspond to 77% of all cases of MDR TB in Colombia. The strains in question were processed by the National Network of Laboratories under their routine activity on TB diagnosis and sent to the National Reference Laboratory at the National Institute of Health in Bogotá, for confirmation and evaluation of drug resistance. Note that the results of the second-line drug susceptibility tests were from previously informed providers and territorial health entities, the latter being responsible for treatment under established national TB guidelines. Retrospective genotyping of MTBC strains met with the standards set out in resolution 008430 of 1993 by the Ministry of Social Protection, Colombia, as follows. Briefly, (i) all sampling and testing were made as part of diagnosing the disease and drug resistance for patient benefit, (ii) retrospective genotyping did neither require additional sampling / new intervention, nor any risk for the evaluated cases, and (iii) all genotyping data were anonymized prior to reporting to the TB genotyping database curator for an anonymous comparison focusing on genotypic and phylogenetic analysis of MTBC genetic diversity.

2.2 Bacterial isolates, genotyping, and database comparison

All *Mycobacterium tuberculosis* complex (MTBC) isolates showing resistance either to isoniazid and/or rifampicin by molecular or conventional tests by the National Network of Laboratories, in Colombia by the National Network of Laboratories were sent for confirmation and evaluation of drug

resistance to the National Reference Laboratory at the National Institute of Health. The strains were cultured in Lowenstein-Jensen medium and MGIT liquid medium (BD Diagnostics, Sparks MD, USA), subjected to routine species identification [26], followed by first- and second-line drug susceptibility testing (DST) using Bactec™ MGIT™ 960 methodology according to the manufacturer's recommendations with following drug concentrations: First-line: streptomycin (S), 1µg/ml and 4µg/ml; isoniazid (H), 0,1µg/ml and 0,4µg/ml; rifampicin (R), 1µg/ml; ethambutol (E), 5µg/ml and 7,5µg/ml; and Second-line: kanamycin (Km), 2,5 µg/ml; amikacin (Am), 1µg/ml; capreomycin (Cm), 2,5 µg/ml; ofloxacin (Of), 2 µg/ml. After DST screening, a total of 203 MTBC isolates (1 strain per patient) isolated between the years 2012-2013 from 29 of the 33 territorial entities of Colombia, were retained for this study.

The isolates were spoligotyped as reported [4], the results obtained were converted into binary and octal codes and analyzed using the SITVIT2 proprietary database of the Pasteur Institute of Guadeloupe [13,27] which is an updated version of previously released SpolDB4 and SITVITWEB databases [12]. At the time of this study, SITVIT2 contained a total of 111,635 MTB clinical isolates from 169 countries of patient origin. In this database, Spoligotype International Type (SIT) designates identical patterns shared by two or more isolates, whereas "orphan" designates unique patterns reported for a single isolate. Major phylogenetic lineages were assigned according to spoligotype signatures using SITVIT2 database and for the 7 major MTBC genetic lineages we used the CBN method from TB-Lineage (http://tbinsight.cs.rpi.edu/run_tb_lineage.html) [28] . The lineage distribution in cities or departments of Colombia was compared with those from other cities of Colombia, as well as other countries in Latin America (Mexico, Venezuela, Brazil and Peru), for which data are available in the SITVIT2 database. Lastly, the worldwide distribution of clusters containing 5 or more isolates in this study, and their worldwide distribution was studied both at the country (2 letter country codes according to http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2), as well as macro-geographical level (United Nations subregions at http://unstats.un.org/unsd/methods/m49/m49regin.htm). 24-MIRU-VNTR typing [5] was performed using PCR followed by agarose gel electrophoresis, PCR products were analyzed in a photo documenter (Bio Imagen System, ChemiGenius model), and the molecular weight of each fragment has been automatically determined using GeneSnap® software version 6.07 GeneTools.

2.3 Phylogenetical and statistical analyses

To establish the relationships between different isolates of MDR-TB we calculated the similarity tree using the unweight pair group method with arithmetic averages (UPGMA) using the MIRU-VNTRplus web application (http://www.miru-vntrplus.org/MIRU/index.faces) [29]. A cluster was defined as two or more *M. tuberculosis* isolates with identical patterns (note that the term "cluster" is used here for spoligotyping data solely for convenience and does not imply recent transmission). The evolutionary relationships among all the observed spoligotypes and MIRU-VNTR patterns were obtained by drawing minimum spanning trees (MSTs) using MLVA Compare software from Geno Screen and Ridom Bioinformatics and BioNumerics version 6.6 (https://www.applied-maths.com/bionumerics). MSTs are undirected graphs in which all samples are connected with the

fewest possible connections between nearest neighbors. Finally, statistical analysis was performed using the R software. Chi-squares and P-values were calculated with Pearson's Chi-squared test. A p-value between 0.05 and 0.07 was considered marginally significant. A p-value<0.05 was considered statistically significant. The Hunter-Gaston Discriminatory Index (HGDI) was used to evaluate the discriminatory power of the spoligotyping method.

3. Results

3.1 Characteristics of the population studied

We analyzed two hundred and three MDR isolates that were collected from 29 of 33 territorial entities of Colombia during the years 2012-2013. Most MDR-TB cases occurred in the department of Antioquia, mainly due to the cases brought by the capital city Medellin, which is the city with the highest MDR-TB isolates (29.56%, n=60) (Table S1). The second is Valle del Cauca department (23.16%), n=47), whose cases are mostly provided by Buenaventura (14.29%, n=29), which is a port city boarding the Pacific Ocean, and the capital Cali (8.87%, n=18). The third department that has more cases of MDR-TB is Atlántico, whose cases are concentrated in the capital Barranquilla, a port boarding the Caribbean Sea. The fourth department is Norte de Santander whose cases are in the capital Cucuta (3.45%, n=7) which is a border town with Venezuela. Medellin and Cali are cities with more than two million people, meanwhile Buenaventura has 400.000 people and Cucuta has 650.000 people (http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion). The global Male/Female sex ratio was 129/74, i.e. 1.74 and the mean age was 39 year, with the interquartile range between 26 to 53 years of age (Table S2).

3.2 MDR-TB Population Structure

To determine the population structure of the 203 MDR-TB strains, we carried out the spoligotyping technique. We found 52 SITs, twenty-seven SITs were clustered (51.9%) meanwhile twenty-five SITs were unique (48.1%). A total of 6 SITs were recently created and 4 of them were exclusively found in Colombia (Table 1). Among these newly created SITs, one pattern belonged to an “unknown” signature (SIT4100), another to X lineage (SIT4105), while the remaining 4 patterns (SIT4101 to 4104) belonged to LAM. Lastly, the description of clusters containing 5 or more isolates and their worldwide distribution according to the SITVIT database is shown in Table 2. The most dominant spoligotype family in the MDR cases was Latin American and Mediterranean (SIT42/LAM9, 12.81%, n=26), Haarlem (SIT62/H1, 10.34%, n= 21), Beijing (SIT190/Beijing, 10.34%, n=21), followed by the ill-defined T (SIT53/T1, 4.43%, n=9). Fourteen isolates were orphans. A dendrogram was constructed based on spoligotyping results (Figure S1); according to this tree 163/203 isolates were grouped into 27 clusters, containing from 2 to 26 isolates; the most dominant clusters were represented by the most dominant SITs, as mentioned previously. In this study, the discriminatory power of the spoligotyping method was 0.9464 measured by the Hunter-Gaston Discriminatory Index (HGDI).

24-loci MIRU-VNTR analysis showed an absence of clustering; furthermore, the patterns observed did not match with the patterns previously submitted to the SITVIT database, hence classified as orphans (Table S4). However, despite the absence of clustering, the MST tree (n=190 isolates) obtained with

the combined information on 24-loci MIRU-VNTR and spoligotyping based lineages, showed that the strains tended to group together in function of their spoligotyping pattern (Figure S2). For example, the branches corresponding to the LAM, Haarlem, and Beijing lineages were clearly well defined. The most represented lineage in the tree was LAM with 40% of isolates (n=76/190), followed by Haarlem 22,1% (n=42/190) and Beijing 11% (n=21/190). Lineages S, X and T were less well defined in the tree and corresponded to 22% of the isolates. The other well-defined branch in the tree corresponded to unknown lineage isolates (n=10/19; 0,05%;) though most of them belonged to the SIT881. In the SITVIT2 database, SIT881 was present in Venezuela, United States, Italy, Spain and Colombia; nevertheless, Colombia contributed with 72% (n=18/25) of such isolates. Because of the high variability in the MIRU-VNTR patterns, it was difficult to find an allelic marker for the branches observed; and in this regard, the loci with the highest polymorphism were QUB 26 (0.87), MIRU10 (0.82) and QUB11b (0.81). Further analyses would be needed to better delineate sublineages from genotyping data as performed in another study [30].

3.3 Lineage distribution analysis

We determined the distribution of the eight lineages/sublineages found in this study: Beijing, EAI, Haarlem, LAM, Manu, S, T and X. LAM was the most dominant lineage with 36.45% (n=74), followed by Haarlem with 23.15% (n=47), Beijing and T, each one with 10.34% (n=21), X with 7.39% (n=15), S with 2.46% (n=5) and finally EAI with 0.49% (n=1). Remarkable was the presence of 17 isolates with unknown lineage (8.37%) (Table S3, A). The difference of lineage/sublineages distribution between the two years (2012 and 2013) was “marginally” significant (p-value=0.0504) (Table S3, B). The difference between gender of patients vs. lineage was not statistically significant (p-value=0.88). When comparing distribution of main lineages/sublineages (Beijing, Haarlem, LAM and T) in Medellin (n=68 isolates), Buenaventura (n=29 isolates), Cali (n=18 isolates), and other cities grouped together (n=96 isolates), we noted a statistically significant difference (p-value<0.000001) (Table 3). Beijing lineage was more frequent in Buenaventura, Haarlem lineage was more frequent in Medellin, and LAM was more frequent in Cali as well as the group of other cities. A detailed map of Colombia showing the phylogeographical distribution of MTB lineages/sublineages is illustrated in **figure 1**.

The department with the highest diversity of lineages/sublineages is Valle del Cauca (where Buenaventura and Cali are located; (n=16), followed by Antioquia (with Medellin being the capital; n=11), Atlántico (n=8), and Bogotá and Norte de Santander (n=6 each). Moreover, the distribution of the lineages/sublineages is different in each region, as mentioned above. To sum up, a statistically significant difference was noted when comparing MDR (HR), MDR (SHR), and MDR (SHRE) groups vs. lineages; note however that not enough data were available for MDR (HRE) group. Last but not least, 20/21 of strains belonging to Beijing lineage were associated with MDR (SHRE), while a great proportion of LAM lineage strains was associated with MDR (HR) (p-value<0.0001) (Table 3).

Comparing the distribution of MDR-TB lineages/sublineages obtained in this study (Colombia) with the data available in SITVIT2 for other Latin American countries such as Mexico, Venezuela, Peru and Brazil, we noticed that the most dominant lineage among the MDR-TB isolates varied according to

the countries studied – thus it was LAM in Colombia, Venezuela, Brazil and Peru, whereas the sublineage T was the most dominant lineage in Mexico (45.30%, n=82) **(Figure 2).** The Manu sub lineage was only present in Colombia and Brazil. In a nutshell, each country had its own characteristic distribution, for example, the most prominent lineages/sublineages were LAM, Haarlem, Beijing and T in Colombia, as opposed to T, X and LAM in Mexico.

3.4 Evolutionary relationships

Minimum spanning trees (MSTs) shown in Figure 3 were based on all spoligotypes of this study (n=203 isolates). **Figure 3A** illustrates evolutionary relationships of spoligotypes in function of MTBC lineages. One may notice that the most visible lineages/sublineages were LAM, Haarlem, Beijing, and T. Only 2 isolates belonged to Manu sublineage, and only one strain belonged to EAI (East-African-Indian). The most visible SITs were SIT42/LAM9 (n=26 isolates), SIT62/H1 (n=21 isolates), SIT190/Beijing (n=21 isolates), SIT53/T1 (n=9 isolates), SIT50/H3 (n=8 isolates), SIT217/X1 (n=8 isolates), and SIT881/Unknown (n=8 isolates). Some newly created SITs (such as SITs 4101, 4102, and 4104) as well as some orphan spoligotypes (Or06, Or10, and Or12), and SIT2648 belonging to LAM sublineage appeared at a terminal position in the tree, forming a sub-branch rooted by SIT4101. These spoligotypes may be specific to Colombia, and may constitute a subgroup of LAM sublineage marked by the absence of spacers 5 and 15. **Figure 3B** is the same tree as the precedent, but it was drawn in function of information on first-line MDR-TB drugs. One may notice that SIT190/Beijing, appearing at the bottom of the MST, was highly associated with MDR (SHRE) group. SIT42/LAM9 was mostly associated with both MDR (HR) and MDR (SHRE). Furthermore, a high proportion of MDR (SHRE) was also visible among isolates belonging to SIT62/H1.

The MST analysis based on 24-loci MIRU-VNTR data (n=190 isolates) showed that all patterns were unique (as confirmed by comparison with SITVIT database). However, despite their uniqueness and diversity, we noticed that isolates were rather well organized in function of spoligotyping based lineages, as mentioned before (**Figure 4**). The UPGMA tree drawn in **Figure S2** showed that strains belonging to SIT42/LAM9 and SIT53/T1 were particularly scattered in different branches. This observation corroborated the fact that several sublineages could be discovered among strains sharing a same spoligotyping pattern.

4. Discussion

The present work characterizes the genotypes of the 203 MDR-TB isolates recollected during 2012-2013 from 29 of 33 territorial entities of Colombia. The predominant genotypes within this pool of MDR-TB isolates belonged to the most predominant lineages present in Colombia reported previously [19–21], i.e. LAM and Haarlem sublineages represented by the SIT42/LAM9 (12.81%,) and SIT62/H1 (10.34%) respectively. However, we must highlight the presence of a specific Beijing genotype (SIT190/Beijing) as predominant genotype in MDR-TB isolates, mainly isolated in the seaport of Buenaventura (Valle del Cauca department) (Table 3). Although the current data might not be sufficient to make year to year assessment, prevalence of SIT190/Beijing strains increased from 7

284 isolates in 2012 to 14 in 2013. This trend was already observed previously in a cohort of patients from
285 1999 to 2012 regarding MDR-TB isolates associated with the Beijing lineage [21]. Indeed, the high
286 prevalence of Beijing genotype in the city of Buenaventura has been reported since 1998 [22], with
287 subsequent emergence of a so-called rare "Beijing-like" genotype SIT190 [23], which continues to
288 dominate the MDR-TB cases in Buenaventura [24,25]. Apart from human migration, this may partly be
289 due to deficiencies in the monitoring and treatment of patients by the precarious health system in
290 Colombia. Indeed, despite the recommendations made by the public-private organizations from Valle
291 del Cauca, interested in knowing the MDR-TB situation in this department, much remains to be done
292 to precisely know the proportion of primary vs. previously treated MDR cases, a finding that will have
293 significant impact on TB control and treatment outcome [31].

294 We observed a strong association between Beijing genotype and resistance to isoniazid,
295 rifampicin, ethambutol and streptomycin (20/21 strains, p-value < 0.0001). Several studies have
296 shown that *M. tuberculosis* strains belonging to the L2/East-Asian lineage, to which Beijing genotype
297 belongs, have strong association with drug-resistance and increase virulence [32–34]. A study showed
298 that particular strains of the Beijing genotype acquired drug resistances in vitro more rapidly than L4/
299 Euro-American strains [16], but this observation has not been confirmed in real settings. The
300 association of Beijing strains with MDR-TB can also differ depending on countries in Latin America
301 (Fig. 2) [35] and elsewhere [13] . In this regard, as compared data shown for Latin American countries
302 illustrated in Fig. 2 (Colombia, Peru, Brazil, and Venezuela), a recent study from Peru highlighted the
303 fact that LAM lineage strains were more likely to be drug resistant as compared to Beijing strains [36],
304 while T lineage strains were associated with MDR-TB in Mexico [37]. In conclusion, despite
305 geographic variability observed for association of MDR-TB and Beijing genotype at a global scale [as
306 reviewed recently; [13]], our results do provide with clinical and epidemiological evidence that
307 L2/Beijing genotype strains, and in particular an emerging SIT190/Beijing variant, show an increased
308 capacity to be transmitted and be associated with drug resistance in Colombia [38,39].

309 Considering the prevailing worldwide occurrence of Beijing strains and their dominance in
310 certain countries [40,41], one must be extremely careful with the subtle change in the genetic diversity
311 landscape and drug-resistance characteristics of prevailing *M. tuberculosis* strains in Colombia.
312 Remarkably, a high diversity of lineages is found in Valle del Cauca (n=13) with predominance of the
313 Beijing lineage (see results and Fig. 1) fueled by the seaport of Buenaventura on the Pacific Ocean. In
314 contrast, low diversity and different lineages were observed in the department of Atlántico (n=8; Fig. 1)
315 as compared to Valle del Cauca even though the provider city of Barranquilla is also a port on the
316 Caribbean Sea. Thus, despite both Buenaventura and Barranquilla being port cities, the diversity of
317 lineages and MDR-TB may be explained by the magnitude of the commercial movements in the port
318 of Buenaventura, which moves 33% of exports and 49.6% of Colombian imports
319 (http://colombiatrade.com.co/). The high prevalence of the Beijing lineage in Buenaventura,
320 specifically genotype SIT 190/Beijing variant, suggests that it came from China; in fact, the SIT190 is
321 present in China and is the country that contributes the most to this genotype worldwide (24%)
322 (http://www.pasteur-guadeloupe.fr:8081/SITVIT2/).

According to the 2011 census conducted by the “Departamento Administrativo Nacional de Estadística” (www.dane.gov.co/), Buenaventura with primarily an Afro-Colombian population has a high proportion of people with unmet basic needs (34.52%), while Barranquilla with as little as 13.2% of Afro-Colombians has a significantly lower proportion (17.70%) of people with unmet basic needs. It is therefore tempting to speculate that a high proportion of people with a specific ethnicity living in poverty could be one of the factors favoring the emergence of drug-resistant Beijing genotype variant(s) in Buenaventura. Additionally, MIRU-typing showed that all 24-loci MIRU-VNTR patterns were unique (or orphans as compared to SITVIT database; Table S4) – an observation that did not support active transmission of circulating MDR clones. MDR thus most probably emerged among unrelated cases due to lack of compliance.

In concordance with prevailing MTB lineages present among pulmonary TB patients in Colombia [20], Antioquia with the capital city of Medellin was the second department with a high genetic diversity (n=11) in our study, with Haarlem being the most predominant lineage (p-value<0.0001). Strangely enough, Bogotá – the capital city of Colombia, with nearly eight million population, was characterized by least genetic diversity. This result should be re-explored in future investigations to verify if it could be linked to access to good healthcare and better life conditions or simply an underreporting of TB cases. Finally, this study highlighted the presence of new spoligotypes (defined by the absence of spacers 5 and 15), which may constitute a new subgroup of LAM lineages. However further investigations would be necessary to know if these new patterns are specific to Colombia possibly due to specific host-pathogen interactions, or due to a simple coincidence. To sum up, our study brings a deep overview regarding MDR-TB situation in Colombia as compared to other Latin American countries and argues in favor of better strategies and management for control and prevention of TB in Colombia. A limitation of our study is the fact that Whole Genome Sequencing (WGS) analysis was not performed. Future studies should target WGS data to know if active transmission of MDR-TB clones is ongoing and at what extent.

Acknowledgements

The work related to molecular biology received the support of Departamento de Microbiología, Facultad de Medicina, Universidad Nacional de Colombia.

Author contribution

MIM designed the study. JGRC, CLL, LACH and JG conducted the experiments. CLL collected the samples and prepared patient´s data. JGRC, CLL, LACH, DC, NR and MIM analyzed and interpreted the data. JGRC, DC and NR collectively wrote the text, and NR edited the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare there are no competing interests.

References

[1] World Health Organization. Global tuberculosis report 2019. World Health Organization. https://apps.who.int/iris/handle/10665/329368.

[2] Instituto Nacional de Salud. Informe del Evento Tuberculosis Farmacorresistente, periodo epidemiológico 12 del año 2013. Bogotá D.C. 2013. https://www.ins.gov.co/buscador-eventos/Lineamientos/PRO_Tuberculosis.pdf

[3] Cruz-Martinez OA. Comportamiento de la tuberculosis y avances en la implementación del plan estratégico hacia el fin de la tuberculosis en Colombia 2016-2025. Bogotá D.C. 2020. https://www.risaralda.gov.co/descargar.php?idFile=14217

[4] Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, Van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35: 907-14. https://doi.org/10.1128/JCM.35.4.907-914.1997.

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

[6] Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. Emerg Infect Dis. 2013;19:460-3. https://doi.org/10.3201/eid1903.120256.

[7] Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, et al. Significance of the identification in the Horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. PLoS One. 2012;7(12):e52841. https://doi.org/10.1371/journal.pone.0052841.

[8] Jagielski T, Minias A, van Ingen J, Rastogi N, Brzostek A, Żaczek A, et al. Methodological and clinical aspects of the molecular epidemiology of *Mycobacterium tuberculosis* and other mycobacteria. Clin Microbiol Rev 2016;29:239-90. https://doi.org/10.1128/CMR.00055-15.

[9] Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet 2013;45:1176–82. https://doi.org/10.1038/ng.2744.

[10] Gagneux S. Genetic diversity in *Mycobacterium tuberculosis*. Curr Top Microbiol Immunol 2013;374:1–25. https://doi.org/10.1007/82_2013_329.

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

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

[13] Couvin D, David A, Zozio T, Rastogi N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. Infect Genet Evol 2019;72:31-43. https://doi.org/10.1016/j.meegid.2018.12.030.

[14] Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. Nat Commun 2020;11:2917 https://doi.org/10.1038/s41467-020-16626-6.

[15] Coscolla M, Brites D, Menardo F, Loiseau C, Darko Otchere I, Asante-Poku A, et al. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. BioRxiv 2020;17:19. https://doi.org/10.1101/2020.06.10.141788. Posted June 10, 2020

[16] Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. *Mycobacterium*

tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 2013;45:784–90. https://doi.org/10.1038/ng.2656.

[17] Chisompola NK, Streicher EM, Muchemwa CMK, Warren RM, Sampson SL. Molecular epidemiology of drug resistant *Mycobacterium tuberculosis* in Africa: A systematic review. BMC Infect Dis 2020;20. https://doi.org/10.1186/s12879-020-05031-5.

[18] Oppong YEA, Phelan J, Perdigão J, MacHado D, Miranda A, Portugal I, et al. Genome-wide analysis of *Mycobacterium tuberculosis* polymorphisms reveals lineage-specific associations with drug resistance. BMC Genomics 2019;20. https://doi.org/10.1186/s12864-019-5615-3.

[19] Cerezo I, Jiménez Y, Hernandez J, Zozio T, Murcia MI, Rastogi N. A first insight on the population structure of *Mycobacterium tuberculosis* complex as studied by spoligotyping and MIRU-VNTRs in Bogotà, Colombia. Infect Genet Evol 2012;12:657–63. https://doi.org/10.1016/j.meegid.2011.07.006.

[20] Realpe T, Correa N, Rozo JC, Ferro BE, Gomez V, Zapata E, et al. Population Structure among *Mycobacterium tuberculosis* Isolates from Pulmonary Tuberculosis Patients in Colombia. PLoS One 2014;9:e93848. https://doi.org/10.1371/journal.pone.0093848.

[21] Puerto G, Erazo L, Wintaco M, Castro C, Ribon W, Guerrero MI. *Mycobacterium tuberculosis* Genotypes Determined by Spoligotyping to Be Circulating in Colombia between 1999 and 2012 and Their Possible Associations with Transmission and Susceptibility to First-Line Drugs. PLoS One 2015;10:e0124308. https://doi.org/10.1371/journal.pone.0124308.

[22] Laserson KF, Osorio L, Sheppard JD, Hernández H, Benítez AM, Brim S, et al. Clinical and programmatic mismanagement rather than community outbreak as the cause of chronic, drug-resistant tuberculosis in Buenaventura, Colombia, 1998. Int J Tuberc Lung Dis 2000;4:673–83.

[23] Murcia MI, Manotas M, Jiménez YJ, Hernández J, Cortès MIC, López LE, et al. First case of multidrug-resistant tuberculosis caused by a rare “Beijing-like” genotype of Mycobacterium tuberculosis in Bogot??, Colombia. Infect Genet Evol 2010;10:678–81. https://doi.org/10.1016/j.meegid.2010.03.010.

[24] Ferro BE, Nieto LM, Rozo JC, Forero L, van Soolingen D. Multidrug-resistant Mycobacterium tuberculosis, southwestern Colombia. Emerg Infect Dis 2011;17:1259–62. https://doi.org/10.3201/eid1707.101797.

[25] Nieto LM, Ferro BE, Villegas SL, Mehaffy C, Forero L, Moreira C, et al. Characterization of extensively drug-resistant tuberculosis cases from Valle del Cauca, Colombia. J Clin Microbiol 2012;50:4185–7. https://doi.org/10.1128/JCM.01946-12.

[26] Garzón MC, Restrepo G, Llerena C, Orjuela D, Bueno J, Medina ML. Diagnóstico bacteriológico de tuberculosis y micobacteriosis. Bogotá, Colombia:Instituto Nacional de Salud; 2012. https://www.ins.gov.co

[27] Couvin D, Rastogi N. The establishment of databases on circulating genotypes of *Mycobacterium tuberculosis* complex and web tools for an effective response to better monitor, understand and control the tuberculosis epidemic worldwide. EuroReference, 2014;12:36-48. https://pro.anses.fr/euroreference/Documents/ER12-RESEAUX-MycobacteriumEN.pdf

[28] Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, et al. TB-Lineage: An online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. Infect Genet Evol 2012;12:789–97. https://doi.org/10.1016/j.meegid.2012.02.010.

[29] Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: A web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucleic Acids Res 2010;38(Web Server issue):W326-31. https://doi.org/10.1093/nar/gkq351.

[30] Mokrousov I, Vyazovaya A, Iwamoto T, Skiba Y, Pole I, Zhdanova S, et al. Latin-American-Mediterranean lineage of *Mycobacterium tuberculosis*: Human traces across pathogen’s phylogeography. Mol Phylogenet Evol 2016;99:133-143. https://doi.org/10.1016/j.ympev.2016.03.020.

[31] Villegas SL, Ferro BE, Perez-Velez CM, Moreira CA, Forero L, Martínez E, et al. High initial multidrug-resistant tuberculosis rate in Buenaventura, Colombia: A public-private initiative. Eur Respir J 2012;40:1569-72. https://doi.org/10.1183/09031936.00018212.

[32] Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, et al. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. J Am Med Assoc 2005; 293:2726-31. https://doi.org/10.1001/jama.293.22.2726.

[33] Hanekom M, Gey Van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. *Mycobacterium tuberculosis* Beijing genotype: A template for success. Tuberculosis 2011;91:510–23. https://doi.org/10.1016/j.tube.2011.07.005.

[34] de Steenwinkel JEM, ten Kate MT, de Knegt GJ, Kremer K, Aarnoutse RE, Boeree MJ, et al. Drug susceptibility of *Mycobacterium tuberculosis* Beijing genotype and association with MDR TB. Emerg Infect Dis 2012;18:660–3. https://doi.org/10.3201/eid1804.110912.

[35] Cerezo-Cortés MI, Rodríguez-Castillo JG, Hernández-Pando R, Murcia MI. Circulation of *M. tuberculosis* Beijing genotype in Latin America and the Caribbean. Pathog Glob Health 2019; 113:336-351. https://doi.org/10.1080/20477724.2019.1710066.

[36] Grandjean L, Iwamoto T, Lithgow A, Gilman RH, Arikawa K, Nakanishi N, et al. The Association between *Mycobacterium tuberculosis* genotype and drug resistance in Peru. PLoS One 2015; May 18;10(5):e0126271. https://doi.org/10.1371/journal.pone.0126271.

[37] Martinez-Guarneros A, Rastogi N, Couvin D, Escobar-Gutierrez A, Rossi LMG, Vazquez-Chacon CA, et al. Genetic diversity among multidrug-resistant *Mycobacterium tuberculosis* strains in Mexico. Infect Genet Evol 2013;14:434-43. https://doi.org/10.1016/j.meegid.2012.12.024.

[38] Rodríguez-Castillo JG, Pino C, Niño LF, Rozo JC, Llerena-Polo C, Parra-López CA, et al. Comparative genomic analysis of *Mycobacterium tuberculosis* Beijing-like strains revealed specific genetic variations associated with virulence and drug resistance. Infect Genet Evol 2017;54:314–23. https://doi.org/10.1016/j.meegid.2017.07.022.

[39] Ramirez LMN, Ferro BE, Diaz G, Anthony RM, de Beer J, van Soolingen D. Genetic profiling of *Mycobacterium tuberculosis* revealed “modern” Beijing strains linked to MDR-TB from Southwestern Colombia. PLoS One 2020; Apr 24;15(4):e0224908. https://doi.org/10.1371/journal.pone.0224908.

[40] Glynn JR, Whiteley J, Bifani PJ, Kremer K, Van Soolingen D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: A systematic review. Emerg Infect Dis 2002;8:843–9. https://doi.org/10.3201/eid0808.020002.

[41] Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, et al. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. Jama 2005;293:2726–31. https://doi.org/10.1001/jama.293.22.2726.

Legends to Figures

Fig 1. **Phylogeographic distribution of MTBC sublineages among MDR-TB strains in Colombia**. The map shows the distribution of the MDR-TB sublineages corresponding to different territorial entities of Colombia. In dark blue, the territorial entities that have the largest number of MDR-TB cases, each one with its corresponding pie representing the proportion of sublineages found in that region. These territories provide 65% of the isolates from this study. The big pie represents all the lineages found among the pool of MDR-TB isolates.

506 **Fig 2**. **Map representing distribution of pansusceptible and MDR strains of some Latin**
507 **American countries (Mexico, Venezuela, Brazil, Peru, and Colombia) in the international**
508 **database**. This map was obtained from the following link: http://en.wikipedia.org/wiki/Latin_America
509 according to terms of the Creative Commons 3.0 Attribution License
510 (http://creativecommons.org/licenses/by/3.0/). Note that for a better visualization, distribution of strains
511 belonging to this study (n=203) is shown under the subheading Colombia.

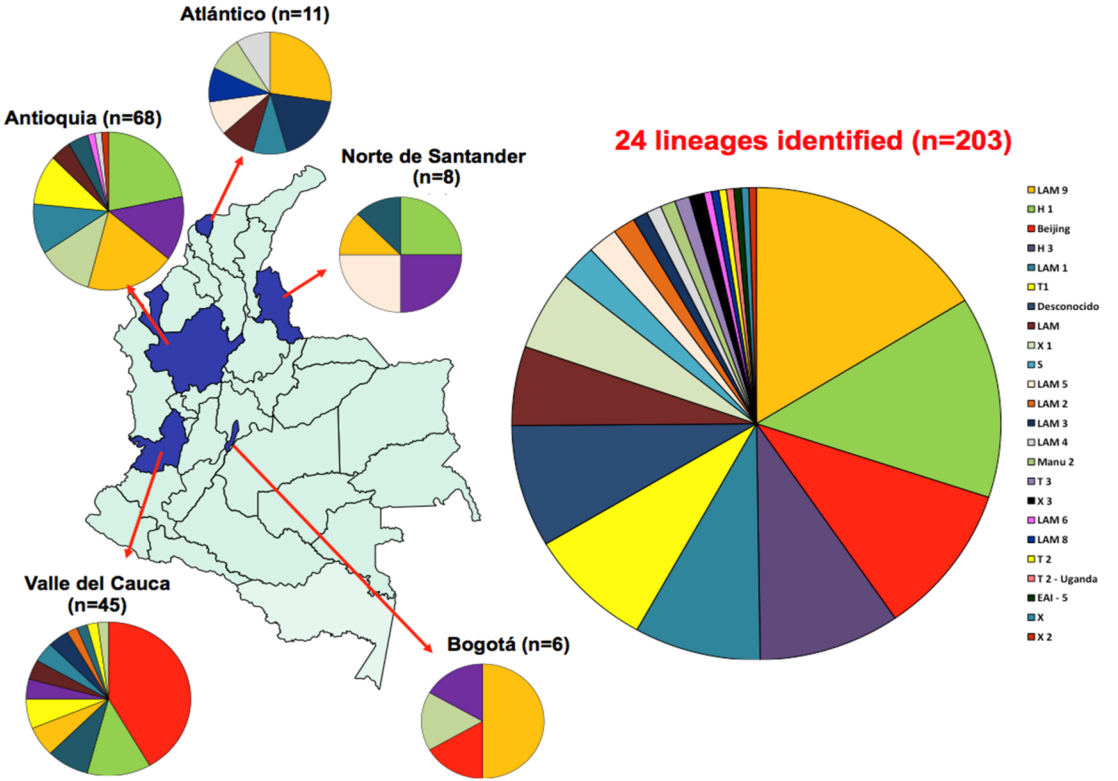
512 **Fig 3**. **A minimum spanning tree (MST) illustrating evolutionary relationships between the *M.***
513 ***tuberculosis* spoligotypes of this study (n=203 isolates)**. The phylogenetic tree connects each
514 genotype based on degree of changes required to go from one allele to another (the distance numbers
515 are visible on each edge). Solid black line denotes one unique change between two patterns, while
516 solid gray line denotes 2 changes, bold dashed line denotes 3 changes, and thin dotted line
517 represents 4 or more changes. The size of the circle is proportional to the total number of isolates. SIT
518 numbers appear inside nodes. Trees were drawn in function of phylogenetic lineages (A), and in
519 function of information on first line MDR drugs (B).

520 **Fig 4. A minimum spanning tree (MST) illustrating evolutionary relationships between all 24-**
521 **loci MIRU patterns (n=190 isolates)**. All circles have the same size because each pattern was
522 unique. Distance between patterns was represented by dashed lines. The tree was drawn in function
523 of the spoligotyping based lineages.

524 **Table 1.** Description of 52 Spoligotype International Types (SITs; n=189 isolates) and corresponding
525 spoligotyping defined lineages/sublineages starting from a total of 203 *M. tuberculosis* strains isolated
526 in Colombia.
527
528 **Table 2.** Description of clusters containing 5 or more isolates in this study, and their worldwide
529 distribution in the SITVIT database

530 **Table 3.** Distribution of MTBC lineages in function of several parameters of the study.
531

532


Atlántico (n=11)
Antioquia (n=68)
Norte de Santander (n=8)
Valle del Cauca (n=45)
Bogotá (n=6)
24 lineages identified (n=203)
LAM 9
H 1
Beijing
H 3
LAM 1
T1
Desconocido
LAM
X 1
S
LAM 5
LAM 2
LAM 3
LAM 4
Manu 2
T 3
X 3
LAM 6
LAM 8
T 2
T 2 - Uganda
EAI - 5
X
X 2

Mexico

Lineage	Pan-susceptible (%)	MDR (%)
BOV	5 (3.16)	2 (1.10)
Beijing	2 (1.27)	3 (1.66)
EAI	6 (3.80)	10 (5.52)
Haarlem	16 (10.13)	13 (7.18)
LAM	27 (17.09)	24 (13.26)
Manu	5 (3.16)	0 (0.00)
S	2 (1.27)	9 (4.97)
T	48 (30.38)	82 (45.30)
X	40 (25.32)	24 (13.26)
Unknown	7 (4.43)	14 (7.73)

*P-value<0.005

Venezuela

Lineage	Pan-susceptible (%)	MDR (%)
Beijing	0 (0.00)	2 (3.64)
Haarlem	3 (3.95)	3 (5.45)
LAM	61 (80.26)	41 (74.55)
Manu	1 (1.32)	0 (0.00)
S	1 (1.32)	1 (1.82)
T	7 (9.21)	8 (14.55)
Unknown	3 (3.95)	0 (0.00)

*P-value>0.3

Brazil

Lineage	Pan-susceptible (%)	MDR (%)
BOV	1 (0.10)	0 (0.00)
Beijing	1 (0.10)	2 (0.68)
CAS	4 (0.40)	0 (0.00)
Cameroon	1 (0.10)	0 (0.00)
EAI	64 (6.44)	17 (5.74)
Haarlem	131 (13.18)	30 (10.14)
LAM	439 (44.16)	116 (39.19)
Manu	12 (1.21)	4 (1.35)
S	13 (1.31)	5 (1.69)
T	191 (19.22)	75 (25.34)
Unknown	91 (9.15)	16 (5.41)
Ural	5 (0.50)	0 (0.00)
X	41 (4.12)	31 (10.47)

*P-value<0.001

Colombia

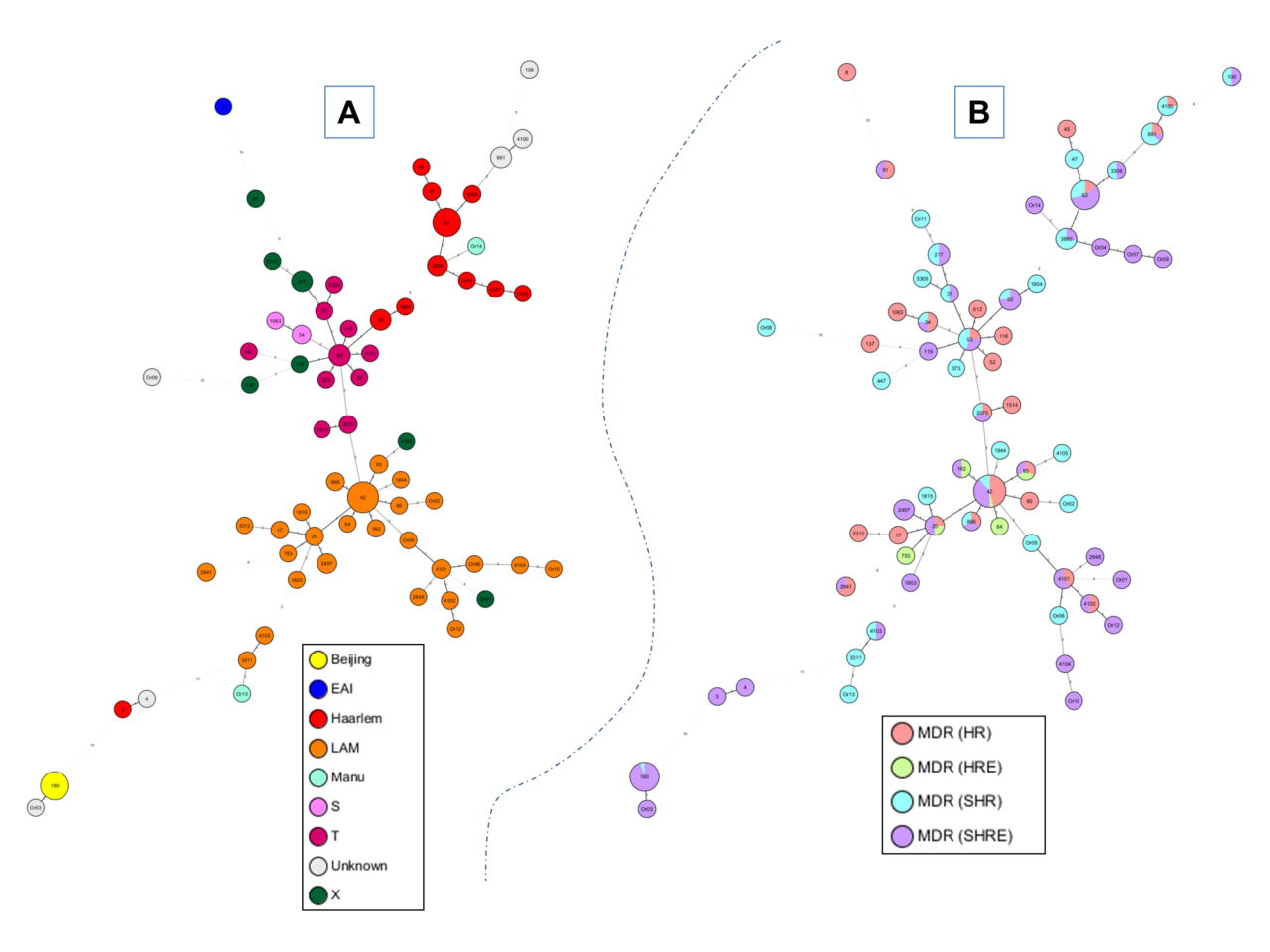
Lineage	Pan-susceptible (%)	MDR (%)	This study (%)
Beijing	1 (0.28)	24 (27.59)	21 (10.34)
Cameroon	1 (0.28)	0 (0.00)	0 (0.00)
EAI	0 (0.00)	0 (0.00)	1 (0.49)
Haarlem	98 (27.22)	19 (21.84)	47 (23.15)
LAM	162 (45.00)	26 (29.89)	74 (36.45)
Manu	0 (0.00)	1 (1.15)	2 (0.99)
S	11 (3.06)	2 (2.30)	5 (2.46)
T	56 (15.56)	2 (2.30)	21 (10.34)
X	13 (3.61)	4 (4.60)	15 (7.39)
Unknown	18 (5.00)	9 (10.34)	17 (8.37)

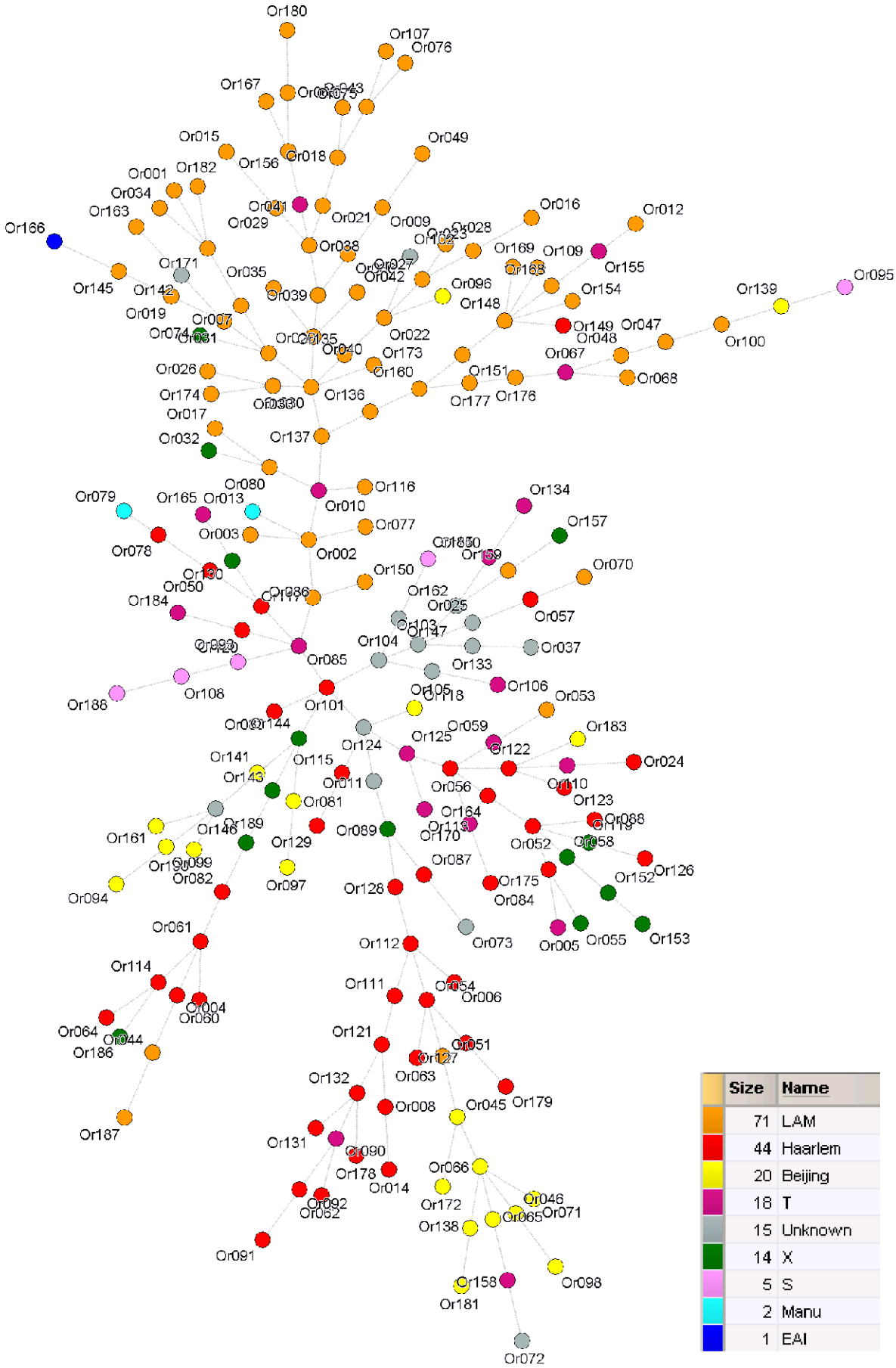
*P-value<0.000001

Peru

Lineage	Pan-susceptible (%)	MDR (%)
AFRI	22 (0.97)	0 (0.00)
BOV	6 (0.27)	0 (0.00)
Beijing	223 (9.86)	32 (9.01)
EAI	3 (0.13)	0 (0.00)
Haarlem	738 (32.63)	54 (15.21)
LAM	511 (22.59)	160 (45.07)
Manu	1 (0.04)	0 (0.00)
S	33 (1.46)	1 (0.28)
T	386 (17.06)	71 (20.00)
Ural	2 (0.09)	2 (0.56)
X	143 (6.32)	13 (3.66)
Unknown	194 (8.58)	22 (6.20)

*P-value<0.000001


A
Beijing
EAI
Haarlem
LAM
Manu
S
T
Unknown
X
B
MDR (HR)
MDR (HRE)
MDR (SHR)
MDR (SHRE)



Size	Name
71	LAM
44	Haarlem
20	Beijing
18	T
15	Unknown
14	X
5	S
2	Manu
1	EAI

Table 1. Description of 52 Spoligotype International Types (SITs; n=189 isolates) and corresponding spoligotyping defined lineages/sublineages starting from a total of 203 *M. tuberculosis* strains isolated in Colombia.

SIT [a]	Spoligotype Description	Octal code	Nb (%) in study	% in study vs. SITVIT	Lineage [b]	Unique or Clustered [c]
3	□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■□■□□□□■■■■■■■	000000007720771	1 (0.49)	0.91	H3	Unique
4	□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■□□□□■■■■■■■	000000007760771	1 (0.49)	0.26	Unknown	Unique
8	■□□□□□□□□□□□□■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	400037777413771	1 (0.49)	0.5	EAI5	Unique
17	■■□■■■■■■■■■□■■■■■■■□□□□■■■■■■■□□□□■■■■■■■	677737607760771	2 (0.99)	0.27	LAM2	Clustered
20	■■□■■■■■■■■■■■■■■■■■□□□□■■■■■■■□□□□■■■■■■■	677777607760771	4 (1.97)	0.44	LAM1	Clustered
34	■■■■■■■■□□■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	776377777760771	4 (1.97)	0.45	S	Clustered
37	■■■■■■■■■■■■□■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	777737777760771	2 (0.99)	0.36	T3	Clustered
42	■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■□□□□■■■■■■■	777777607760771	26 (12.81)	0.71	LAM9	Clustered
45	■■■■■■■■■■■■■■■■■■■■■■□■□□□□□□■□□□□■■■■■■■	777777764020771	1 (0.49)	0.93	H1	Unique
47	■■■■■■■■■■■■■■■■■■■■■■■□□□□□□■□□□□■■■■■■■	777777774020771	3 (1.48)	0.18	H1	Clustered
50	■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■■■■■	777777777720771	8 (3.94)	0.2	H3	Clustered
52	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■□■■■	777777777760731	1 (0.49)	0.1	T2	Unique
53	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	777777777760771	9 (4.43)	0.13	T1	Clustered
60	■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■□□□□■■■□■■■	777777607760731	1 (0.49)	0.22	LAM4	Unique
62	■■■■■■■■■■■■■■■■■■■■■■■□□□□□□■□□□□■■■□■■■	777777774020731	21 (10.34)	3.54	H1	Clustered
64	■■■■■■■■■■■■■■■■■■■■□□□□■■■■□■■■□□□□■■■■■■■	777777607560771	1 (0.49)	0.22	LAM6	Unique
91	■■■□□□□□□□□□■■■■□■■■■■■■■■■■■■■□□□□■■■■■■■	700036777760771	2 (0.99)	0.55	X3	Clustered
93	■■■■■■■■■■■■□■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	777737607760771	3 (1.48)	0.64	LAM5	Clustered
106	■■■■■■■■□□□■■■■■■■■□□□□□□□□□□□□□□□□□□■■■■■	776177400000171	2 (0.99)	1.39	Unknown	Clustered
118	■■■■■■■■■■■■■□■■■■■■■■■■■■■■■■■□□□□■■■■■■■	777767777760771	1 (0.49)	0.6	T1	Unique
119	■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■□□□□■■■■■■■	777776777760771	1 (0.49)	0.09	X1	Unique
137	■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■□□□□■■□□□□■	777776777760601	1 (0.49)	0.09	X2	Unique
162	■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■□■■■■	777777607760671	2 (0.99)	5.26	LAM9	Clustered
190	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■□■■■	000000000003731	21 (10.34)	10.24	Beijing	Clustered
217	■■■■■■■■■■■■□■■■■□■■■■■■■■■■■■■■□□□□■■■■■■■	777736777760771	8 (3.94)	18.18	X1	Clustered
373	■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■□□□□■■■■■■■	777777767760771	1 (0.49)	1.41	T1	Unique
447	■■■■■■■■■■■■■□□□□□■■■■■■■■■■■■■■□□□□■■■■■■■	777760377760771	1 (0.49)	11.11	T1	Unique
612	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■□■■	777777777760751	1 (0.49)	3.23	T1	Unique
753	■□□■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	477777607760771	1 (0.49)	4.35	LAM1	Unique
866	■□■■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	577777607760771	3 (1.48)	5.17	LAM9	Clustered
881	■■■■■■■■□□■■■■■■■■■■■■■□□□□□□□□□□□□■■■□■■■	776377770000731	8 (3.94)	23.53	Unknown	Clustered
1063	■■■■■■■■□□■■■■■■■■■■■■■■■■■■■□■■□□□□■■■■■■■	776377777660771	1 (0.49)	3.85	S	Unique
1514	■■■■□■■■■■■■■■■■■■□□■■■■■■■■■□□□□■■■■■■■	757777637760771	1 (0.49)	14.29	T1	Unique
1803	■■□□□■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	617777607760771	1 (0.49)	7.69	LAM1	Unique
1804	■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■□■□□□□■■■■■■■	777777577720771	1 (0.49)	11.11	H3	Unique
1815	■■□■■■■■■■■■■■■■■■■■□□□□■■□■■■■■□□□□■■■■■■■	677777606760771	1 (0.49)	10.0	LAM8	Unique
1844	■■■■■■■■■■■■■■■□□■■■■□□□□■■■■■■■■□□□□■■■■■■■	777763607760771	1 (0.49)	9.09	LAM9	Unique
2070	■■■■■■■■■■■■■■■■■■■■■□□■■■■■■■■■□□□□■■■■■■■	777777637760771	3 (1.48)	13.64	T1	Clustered
2497	■■□■■■■■■■■■□■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	677677607760771	6 (2.96)	50.0	LAM1	Clustered
2648	■■■■□■■■■■■■□■■■■■□□□□■■■■■■■■□□□□■■□■■■■	757767607760671	1 (0.49)	33.33	LAM9	Unique
2941	□□□■■■■□□□■■■■■■■■■■□□□□■■■■■■■■□□□□■■■□□□□	074377607760700	3 (1.48)	42.86	LAM	Clustered
3306	■■■■■■■■■□■■■■■■■■■■■■■□□□□□□■□□□□■■■□■■■	777377774020731	2 (0.99)	14.29	H1	Clustered
3309	■■■■■■■■■■■□■■■■■■■■■■■■■■■■□□□□■□■■■■■	777737777760571	1 (0.49)	14.29	T2-Uganda	Unique
3311	□□□□□□■■□□□■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	006177607760771	2 (0.99)	40.0	LAM3	Clustered
3312	■■□■■■■□■■■■□■■■■■■■■□□□□■■■■■■■■□□□□□■■■■■■	675737607760371	1 (0.49)	33.33	LAM2	Unique
3890	■■	777777775020731	7 (3.45)	77.78	H3	Clustered
4100*	■■■■■■■■□□■■■■■■■■■■■■■■□□■□□□□□□□□■■■□■■■	776377771000731	4 (1.97)	100.0	Unknown	Clustered
4101*	■■■■□■■■■■■■□■■■■■□□□□■■■■■■■□□□□□■■■■■■■	757767607740771	5 (2.46)	100.0	LAM	Clustered
4102*	■■□■□■■■■■■■■□■■■■■□□□□■■■■■■■□□□□□■■■■■■■	657767607740771	2 (0.99)	100.0	LAM1	Clustered
4103*	■□□□□□■■□□□■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	406177607760771	2 (0.99)	100.0	LAM1	Clustered
4104*	■■■■□■■■■■■■□□□□■■■■□□□□■■■■■■■■□□□□■■■■■■■	757703607760771	1 (0.49)	50.0	LAM5	Unique
4105*	■■■■■■■■■■■■□■■■■□■■■□□□■■■■■■■■□□□□■■■■■■■	777736707760771	1 (0.49)	50.0	X1	Unique

[a] Note that SITs followed by an asterisk indicates "newly created" SITs due to 2 or more strains belonging to an identical new pattern within this study or after a match with an orphan in the database. SITs followed by an asterisk (*) are newly created. Note that SITs 4100 to 4103 were all newly created within this study itself, while SIT4104 and SIT4105 (a single isolate each) matched an orphan strain from Colombia and USA respectively.

[b] Lineage designations according to revised rules; “Unknown” designates patterns with signatures that do not belong to any of the major lineages described in the database.

[c] Clustered strains correspond to a similar spoligotype pattern shared by 2 or more strains “within this study”; as opposed to unique strains harboring a spoligotype pattern that does not match with another strain from this study. Unique strains matching a preexisting pattern in the SITVIT database are classified as SITs, whereas in case of no match, they are designated as “orphan”.

Table 2. Description of clusters containing 5 or more isolates in this study, and their worldwide distribution in the SITVIT database.

SIT/Lineage	Spoligotype Description	Nb (%) in study	% in study vs. database	Distribution in Countries with ≥3% of a given SIT [a]	Distribution in Regions with ≥3% of a given SIT [b]
42/LAM9	■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	26 (12.81)	0.71	BR=14.64, US=10.86, CO=7.69, MA=6.46, IT=5.99, FR=4.65, PE=3.39, ES=3.06, VE=3.04	South America=33.68, North America=10.86, Southern Europe=10.37, Western Europe=8.67, North Africa=7.85, Northern Europe=4.49, West Indies=3.89, East Africa=3.56, Central America=3.26
62/H1	■■■■■■■■■■■■■■■■■■■■■■■□□□□□□■□□□□■■■□■■■	21 (10.34)	3.54	CO=34.85, US=11.45, FR=11.11, IT=6.9, GM=4.04, ES=3.54, FI=3.37	South America=37.37, Western Europe=14.81, Southern Europe=11.45, North America=11.45, Northern Europe=5.05, West Africa=4.21, East Africa=3.54
190/Beijing	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■□■■■	21 (10.34)	10.24	US=28.78, CO=24.39, CN=21.95, JP=5.85	Eastern Asia=31.22, North America=28.78, South America=25.37, Southeastern Asia=4.88
53/T1	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	9 (4.43)	0.13	US=12.32, FR=7.35, BR=5.97, IT=4.97, ZA=4.52, PE=4.42, TR=3.23, AT=3.19	South America=16.13, Western Europe=14.57, North America=12.56, Southern Europe=8.76, Northern Europe=6.97, Western Asia=6.81, East Africa=4.79, Austral Africa=4.63, Eastern Asia=3.97, North Africa=3.28, Eastern Europe=3.04, Central America=3.01
50/H3	■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■■■■■	8 (3.94)	0.2	US=14.75, PE=13.09, BR=8.15, FR=5.77, AT=5.11, ES=4.56, IT=4.56, ZA=3.4, CM=3.12, CZ=3.07	South America=27.14, Western Europe=14.75, North America=14.75, Southern Europe=9.69, West Indies=4.88, Eastern Europe=4.63, Northern Europe=4.58, North Africa=3.57, Austral Africa=3.4, Central Africa=3.17
217/X1	■■■■■■■■■■■□■■■■□■■■■■■■■■■■■□□□□■■■■■■■	8 (3.94)	18.18	US=40.91, CO=25.0, MX=15.91, PA=4.55, BR=4.55	North America=40.91, South America=31.82, Central America=20.45
881/Unknown	■■■■■■■■□□■■■■■■■■■■■■□□□□□□□□□□□■■■□■■■	8 (3.94)	23.53	CO=76.47, US=11.76	South America=82.35, North America=11.76, Southern Europe=5.88
3890/H3	■■■■■■■■■■■■■■■■■■■■■■□■□□□□■□□□□■■■□■■■	7 (3.45)	77.78	CO=77.78, GM=22.22	South America=77.78, West Africa=22.22
2497/LAM1	■■□■■■■■■■□■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	6 (2.96)	50.0	CO=58.33, BR=33.33, PT=8.33	South America=91.67, Southern Europe=8.33
4101/LAM	■■■■□■■■■■■■□■■■■■□□□□■■■■■■■□□□□■■■■■■■	5 (2.46)	100.0	CO=100.0	South America=100.0

[a] The 2 letter country codes are according to http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2; countrywide distribution is only shown for SITs with ≥3% of a given SITs as compared to their total number in the SITVIT database.

[b] Worldwide distribution is reported for UN Sub-regions with more than 3% of a given SITs as compared to their total number in the SITVIT database.

Table 3. Distribution of MTBC lineages in function of several parameters of the study.

	Gender and Age			**City of isolation**				**First line multi-drug resistance**			
Lineage	Female	Male	Mean Age (years)	Buen-aventura	Cali	Medellin	Other cities	MDR (HR)	MDR (HRE)	MDR (SHR)	MDR (SHRE)
Beijing	7	14	37.94	16	4	0	1	0	0	1	20
EAI	0	1	56.00	0	0	0	1	1	0	0	0
Haarlem	16	31	40.20	5	2	22	18	4	0	18	25
LAM	27	47	40.84	4	6	20	44	24	6	12	32
Manu	0	2	38.50	0	0	0	2	0	0	1	1
S	1	4	43.00	1	0	0	4	3	0	1	1
T	9	12	37.75	2	2	6	11	7	0	9	5
X	6	9	35.29	1	1	9	4	2	0	6	7
Unknown	8	9	41.13	0	3	3	11	3	0	10	4