Published in final edited form as:

Tuberculosis (Edinb). 2017 May ; 104: 79–86. doi:10.1016/j.tube.2017.03.003.

Strains of *Mycobacterium tuberculosis* Transmitting Infection in Brazilian Households and Those Associated with Community Transmission of Tuberculosis

Solange Alves Vinhas¹, Edward C. Jones-López², Rodrigo Ribeiro Rodrigues¹, Mary Gaeddert², Renata Lyrio Peres¹, Patricia Marques-Rodrigues¹, Paola Poloni Lobo de Aguiar¹, Laura Forsberg White², David Alland³, Padmini Salgame³, David Hom², Jerrold J. Ellner², Reynaldo Dietze¹, Lauren F. Collins⁴, Elena Shashkina⁵, Barry Kreiswirth⁵, and Moisés Palaci^{1,*}

¹Núcleo de Doenças Infecciosas (NDI), Universidade Federal do Espírito Santo (UFES), Vitória, Brazil

²Section of Infectious Diseases, Department of Medicine, Boston Medical Center and Boston University School of Medicine, Boston, Massachusetts, USA

³Division of Infectious Diseases, Department of Medicine, New Jersey Medical School, Rutgers, The State University of New Jersey, Newark, New Jersey, USA

⁴Department of Medicine, Duke University Medical Center, Durham, North Carolina, USA

⁵Public Health Research Institute Center, New Jersey Medical School - Rutgers, The State University of New Jersey, Newark, New Jersey, USA

Abstract

Molecular epidemiologic studies have shown that the dynamics of tuberculosis transmission varies geographically. We sought to determine which strains of *Mycobacterium tuberculosis* (MTB) were infecting household contacts (HHC), and which were causing clusters of tuberculosis (TB) disease in Vitoria-ES, Brazil. A total of 741 households contacts (445 TST +) and 139 index cases were characterized according to the proportion of contacts in each household that had a tuberculin skin test positive: low (LT) (40% TST+), high (HT) (70% TST+) and (40–70% TST+) intermediate (IT) transmission. IS6110-RFLP and spoligotyping analysis were performed only 139 MTB isolates from index cases and 841 community isolates. Clustering occurred in 45% of the entire study population. There was no statistically significant association between MTB household transmission category and clustering. Within the household study population, the proportion of clusters in HT and LT groups was similar (31% and 36%, respectively; $p = 0.82$). Among index

*Correspondent author: Moisés Palaci, Núcleo de Doenças Infecciosas, Universidade Federal do Espírito Santo. Av. Marechal Campos, 1468- Maruípe-Vitória-ES, Brazil - mpalaci@ndi.ufes.br.

CONFLICT OF INTEREST

The authors have declared that no conflict of interests exist.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cases isolates associated with households demonstrating TST conversion, the frequency of unique pattern genotypes was higher for index cases of the LT compared to HT households ($p = 0.03$). We concluded that clusters and lineages associated with MTB infection in HT households had no proclivity for increased transmission of TB in the community.

Keywords

Tuberculosis; DNA fingerprinting; *Mycobacterium tuberculosis*; epidemiologic surveillance

INTRODUCTION

Mycobacterium tuberculosis (MTB) is a successful pathogen in humans that infects approximately one-third of the world's population [1]. The epidemiologic and clinical consequences of infection with MTB are dependent on a complex interplay of host, environmental, and bacterial factors (infectivity and virulence) [2]. After exposure to MTB-containing airborne droplets, some contacts will become infected and some go on to develop disease in the next 1–2 years. Population-based genotyping indicates that in urban areas recent transmission represents a variable proportion of new cases [3].

Some patients transmit their infection to large numbers of contacts, whereas others transmit rarely, if at all [4,5]. The risk of a contact becoming infected relates to the infectiousness of the tuberculosis (TB) patient, the duration and proximity of the contact [6,7], and the susceptibility of the contact [8]. Rates of MTB transmission are generally inferred by comparing genotypic clustering between patient isolates within a given epidemiological setting [9]. The proportion of reported TB attributable to recent transmission was based on the concept that epidemiologically related organisms share indistinguishable genotypes, whereas unrelated organisms differ at some genetic loci. TB cases that occur in spatial and temporal clusters and share indistinguishable genotypes are thought to be caused by recently transmitted TB infection; those with nonclustered genotypes are thought to result from progression from an infection acquired >3 years in the past [10]. Another proxy for inferring successful transmission of particular MTB genotypes is measuring increases in the frequency of these genotypes over time in patient populations. Recently, epidemiologic studies have supported the view that overall, strains from “modern” lineages are more transmissible than other MTB strains [11]. It is not known if the bacterial determinants that lead to increased risk of acquiring MTB infection once exposed are the same as those determinants leading to the development of TB disease once infected. Although there are now well-developed tools to study strain ability to successfully cause disease, there are no such tools to evaluate the efficiency with which strains cause infection. Due to this limitation, studies in general explore risk factors among TB patients, household members, and health care workers, and most of them investigate the risk factors relating to the development of TB rather than risk conferred by the source of infection and bacterial strain. The aim of this study was to determine which clinical MTB isolates were infecting efficiently household contacts (HHC), and which were causing clusters of disease in the community, and whether these isolates were the same.

MATERIAL AND METHODS

Study setting

Vitória, Brazil is the capital city of Espírito Santo, a state with a TB incidence rate of 38/100,000 and a very low HIV prevalence in TB cases (<2%) and in the general population (<1%), according to the local surveillance program. Vitória Metropolitan area (Great Vitória) encompasses four municipalities (Vitória, Cariacica, Serra, and Vila Velha), which primarily serve as political divisions with modest variation in socioeconomic characteristics. Whereas migration into this region is limited, there is a high level of inter-municipality mobility amongst inhabitants. Núcleo de Doenças Infecciosas (NDI) is a clinical research laboratory at the State University [Universidade Federal do Espírito Santo (UFES)], which has organized a network of five Mycobacteriology laboratories serving 16 state TB clinics in the metropolitan area of Vitória. Researchers at NDI developed a internet-based software program (*TB notes*, NDI) that manages patient information and MTB strain data, including acid fast bacilli (AFB) smear and culture results, for this network. In 2001, the NDI instituted culture-based diagnosis for all TB patients in greater Vitória and all MTB isolates isolated since then have been cryopreserved at NDI. This network-wide database allows for tracking TB cases in real-time and provides nearly complete access to MTB isolates circulating in this community.

HHC study population and measurements

Between February 2008 and December 2012, a total of 1,577 TB patients were screened; of which, 139 index cases and 741 of their HHC were enrolled according to criteria described by our group previously [12].

Index cases—Patients had 1 sputum specimen with AFB smear grade 2+ with subsequent MTB growth in culture. Patients with 2+ AFB sputum smear result have a higher bacillary load and are supposed to be able to expel bacilli continuously and consequently to infect more efficiently their contacts. Fingerprint analysis was done on one of the positive cultures. The radiologic extent of lung disease was graded on a four-category ordinal scale [12].

Household contacts—HHC were defined by previously described criteria as close contacts [12], summarized here: a close contact had significant exposure to the index TB case for 3 months before enrollment, defined by 1) sleeping under the same roof 5 days/week; or 2) sharing meals 5 days/week; or 3) watching television together on nights or weekends; and; 4) other significant contact. (85% of these visited the household 18 days/month). Contacts were evaluated for MTB infection with a tuberculin skin test (TST) after ascertainment of the index and HHC with a TST <10 mm at baseline were retested after 8–12 weeks to assess for TST conversion. No genotyping method was performed as this group of patient were culture negative at that time.

Household transmission categories—We categorized households into MTB transmission categories according to the percentage of contacts in each household that had a TST 10 mm by 8–12 weeks: low (LT) transmission (<40%) or high (HT) transmission

(70%). A group of households classified as intermediate transmission (IT) households (41–69% contacts with TST 10 mm) were also included in the study for clustering analysis.

Community cases: Historical and concurrent controls

From 2003 to 2007, the period immediately preceding the HHC study, there were a total of 2,868 pulmonary TB cases diagnosed clinically in metropolitan Vitória. In the NDI network database and strain collection, there were an estimated 2,329 culture-confirmed cases, of which, 841 (36.1%) were selected to serve as “historical controls” in order to molecularly characterize the isolates circulating in the community in a five-year period preceding the HHC study. To verify that these isolates were representative of those circulating during the years of the HHC study, 136 isolates from pulmonary TB cases diagnosed from 2008 to 2012 were selected as “concurrent controls” for comparison. These isolates were obtained from (i) index cases excluded from the HHC study (n=86), and (ii) community cases graded AFB smear 1+ (n=50) to decrease sampling bias associated with only characterizing MTB isolates from individuals with high-grade smear results.

Laboratory methods

Genotypic analysis by RFLP and spoligotyping was completed for all historical and concurrent controls. MTB isolates (n=705 from period immediately preceding the HHC study + n= 136 concurrent controls + n=139 HHC study) stored at –70°C were sub-cultured in Ogawa media and all viable isolates were typed by the IS6110RFLP according to a standardized method [13]. The *MTB* 14323 strain was used for comparison of patterns. The band patterns were analyzed with the *BioNumerics* software version 6.5 (Applied Maths, Sint-Martens-Latem, Belgium). A dendrogram was constructed to elucidate the degree of similarity among the isolates. Unweighted Pair Group Method with Arithmetic mean (UPGMA) and the Dice index (1.0% tolerance, 1.5% optimization) were used. Cluster analysis was performed using the strictest definition of a cluster being two or more MTB isolates with identical RFLP hybridization patterns. Spoligotyping was performed by using a commercial kit (Ocimum Biosolutions Inc., Hyderabad, India) according to a standardized protocol [14]. This genotyping method was used as second methods just to confirm cluster. The Hunter-Gaston Discriminatory Index (HGDI) was used to calculate the level of discrimination of each typing method [15]. Results were recorded in a 43-digit binary format and these patterns were compared with those in the updated SITVITWEB database (Pasteur Institute of Guadeloupe) [16].

Statistical analysis

We summarized clustering patterns in index cases from the HHC study, the historical and concurrent controls (separately), as well as the HHC and control populations and both populations combined. Chi square tests were used to analyze differences between clustered and non-clustered populations. Generalized Estimating Equations (GEE) were used to account for potential differences within household clustering when analyzing HHC characteristics. All analyses were run using R 3.2.2 [r-project.org].

Ethics Statement

The study was approved by the Institutional Review Boards of Boston University Medical Center, and New Jersey Medical School – Rutgers University (formerly University of Medicine and Dentistry of New Jersey), and by the Comitê de Ética em Pesquisa do Centro de Ciências da Saúde - Universidade Federal do Espírito Santo (Reference N°050/2007) and the Comissão Nacional de Ética em Pesquisa (CONEP – Reference N°764/2007). We obtained written informed consent and assent in Portuguese in accordance with age-specific ethical guidelines of participating institutions.

RESULTS

Based on household TST results, the 139 index cases and their 741 household contacts were categorized as follows: 25 LT families, 84 HT families; and 30 IT families were selected for enrollment as control households (Figure 1). The median number of household contacts per index patient household was 5.7 for both HT and LT groups. Household contacts were young (median age 20 years); 56.3% were female; 77.0% had a BCG vaccination scar; 99% had no prior diagnosis of MTB and 98% reported no HIV-infection. A lower percentage of contacts (7.5%) were under 5 years old. Contacts were comparable across transmission categories in terms of age, socio-economic condition, family composition, alcohol and tobacco consumption, and history of co-morbidities.

The median age of index cases was 36 years (range 25–46), 64% were male, and most had extensive infectious TB disease as measured by duration of cough (median 12 weeks), sputum AFB smear grade (82% graded 3+), and extent of lung disease on chest radiograph. There were no differences in socioeconomic variables between household transmission categories (Appendix 1).

IS6110RFLP and Spoligotyping were performed on 980 MTB isolates (841 community isolates and 139 HHC group-IC isolates). The RFLP and Spoligotyping analysis revealed 45% of clusters in the entire study population (441/980). Among the 139 index cases comprising the HHC study, clustering was observed in 40 (29%), in comparison to the 841 historical and concurrent controls where clustering was observed in a higher rate (44%) of subjects. Both individual rates of clustering are lower than the overall clustering rate when combined, as is expected when a smaller sample size is analyzed for clustering.

During the study period (2008 – 2015) a total of 22 secondary TB cases occurred in the HHCs and this secondary TB cases were equally genotyped by both methods used in this study. Their RFLP and Spoligotyping patterns were identical to their respective index case RFLP pattern, which suggest recent transmission.

A wide range of clusters and genotypes was identified in both community isolates and HHC isolates (Table 1). The overall mean cluster size was 3.7 isolates (range 2–18 members). The median cluster size was 6 isolates, and almost one half (46.5%) of the clusters had only 2 isolates. To verify the diversity of RFLP profiles, the HGDI was calculated for the community and HHC strain populations individually. Both populations demonstrated adequate genotypic heterogeneity (HGDI = 0.99 for community and 0.94 for HHC isolates).

Table 2 shows the MTB genotypic profiles of the index cases enrolled in the HHC study (HT, LT and IT) and the genotypic profiles characterized when all subjects grouped together (e.g. index cases from the HHC in addition to the historical and concurrent controls). There was no statistically significant association between MTB household transmission category and clustering. Within the household study population, the proportion of clusters in HT and LT groups was similar (31% and 36%, respectively; $p = 0.82$).

Assuming that the sample size and study period potentially influenced clustering, we merged genotypes from the household study and historical and concurrent controls into a single genotype dataset and performed a separate analysis to verify if the proportion of clusters found for each transmission group (high, low and intermediate) differed significantly among them. Although an increase in the proportion of clustering was seen in both HT (31% to 43%) and LT (36 to 60%), there was no statistical difference between groups ($p = 0.20$). In fact, a higher frequency of unique patterns ($p < 0.05$) was observed in all groups, except for the low transmission category when analyzed within the context of historical and concurrent controls. Analyses were performed to determine if certain MTB genotypes or degree of transmissibility were associated with socioeconomic or clinical characteristics of index cases or HHC. The only remarkable finding was that HT households were more likely to have the presence of a BCG scar, compared to LT or IT households ($p = 0.05$). (Appendix 1 and 2).

Genotypes (*IS6110* RFLP patterns and spoligotypes) found in the study transmission groups were grouped according to three approaches (Figure 2): genotype clustering only within HHC study isolates (G1); genotype clustering only with historical and concurrent control isolates combined (G2) and genotype clustering with G1 and G2 grouped together (G3). Most analyzed genotypes belonged to G2 (24/62; 38.7%) and G3 (32/62; 51.6%) categories. Eight cluster genotypes were shared among high and low transmission groups (Appendix 2). To discern the worldwide distribution of the major shared types encountered in the study transmission groups, the isolates were compared with those classified as Spoligotype International Type (SIT) in the database. LAM family and its subfamilies were predominant in our setting, in 34/62 (54.8%) of the isolates. The next most prominent types were T family with 9/62 (14.5%). No associations were found between any family/sub-families and study transmission groups (Appendix 3).

Genotypes (cluster and unique patterns) were compared across transmission categories in terms of household TST status (persistent TST–; baseline TST– → follow-up TST+; and, baseline TST+) (Table 3). There were no differences in the proportion of cluster and unique patterns when comparing the index cases of HT to index cases of LT whose household contacts remained persistently TST negative. Among index cases isolates associated with households demonstrating TST conversion (baseline TST–) → (follow-up TST+), the frequency of unique pattern genotypes was higher for index cases of the LT compared to HT households ($p = 0.03$). Contrarily, index case isolates associated with HHCs who were TST+ at baseline, the frequency of unique pattern genotypes was higher for index cases associated with HT compared to LT households ($p = 0.01$).

DISCUSSION

There is increasing evidence from experiments in animal models and from clinical studies indicating that MTB strains differ in their phenotype [17,18]. Pathogen-related factors that may contribute to MTB genotype selection include variation in transmissibility (the capacity to spread from person to person) and pathogenicity (the ability to cause clinical disease) [19]. In our study, MTB transmissibility to household contacts was not related to whether the isolate from the source case was found in clusters. A similar proportion of clusters were found in both HT and LT groups. Even among the household TST converters (baseline TST – → follow-up TST+) there was no difference between HT and LT groups. Moreover, 8 cluster genotypes were shared between HT and LT groups. These findings contradicted the common assumption that higher rate of clustering would be found in the HT group. Our premise was based on the suggestion by others that MTB isolates causing clusters, independent from patient risk factors, on average would generate more positive contacts [19–21]. Verhagen (2011) et al. in contact investigations around source cases in the Netherlands reported that newly diagnosed index cases in a larger RFLP cluster infected more people than did newly diagnosed cases in smaller clusters. Unlike that study, which concluded that isolates with a similar RFLP pattern showed more similarity in extent of transmission than isolates with different RFLP patterns, we found that MTB transmissibility in household contacts was not associated with cluster genotypes.

Possible explanations for this disparity may lie in the study design and epidemiological situation. Verhagen et al. [20] conducted their study in a low-TB prevalence area and used a different strategy for contact investigation. They enrolled not only household members but also close work colleagues and close friends of all ages. Furthermore, the median number of investigated contacts per index case in their study was considerably higher than in ours and the genetic background of the strains was unknown.

Several studies have demonstrated that the proportion of clustered cases in urban areas indicating recent transmission comprise 20 to more than 50 percent of all local TB cases [22–25]. In our study the clustering proportion (45%) was comparable to those found in the Netherlands, Denmark, and some areas in the United States. However, it is important to note that the percentage of clustering is strongly dependent on the size of the study population, age of the patients, and length of the study period [26]. Furthermore, published meta-analyses and modeling studies using IS6110RFLP data show that the proportion of clustering observed can be affected by (1) study design; (2) features of the typing method; and (3) study setting [6,27,28]. For example, the proportion of clustering increases when the fraction of the total data sampled increases and when study duration increases. The effect of the sampling fraction has been predicted through modeling. Notably, when the average cluster size is small: the ‘missing’ part of the population will lead to more clustered cases being classified as unique pattern, thus underestimating the proportion clustered [27–29]. Based on these premises, we combined genotypes from the household study and the historical and concurrent controls into a single genotype dataset for clustering analyses. The degree of ongoing transmission estimated in our study within a time window of 10 years should be corrected for some under-estimation because of IS6110 instability during this period of time, which is a longer duration than in most similar studies.

Interestingly, the proportion of cluster (42%) and unique pattern (58%) identified from index cases whose HHC were TST >10 mm at baseline (Table 3) was quite similar to that observed in historical and concurrent controls (44% cluster; 56% unique pattern). Thus, we can assume that if contacts from the HT group, which represent most of patients of our study, become secondary cases, the genotypes profile would be similar to that found in the historical and concurrent population.

In a cohort study of TB patients and their household contacts in The Gambia, de Jong et al [2012] enrolled 317 TB patients and followed their 1,808 HIV-negative household contacts for 2 years. Contacts were categorized by the index case's MTB lineage infection. The number of contacts who would go on to become newly infected and develop active TB was then determined. The results showed that, whereas there was no difference in transmission, contacts exposed to 'modern' MTB were more likely to develop active TB compared with individuals exposed to 'ancient' MTB [23]. Gagneux (2012) postulates that these observations are consistent with the notion that 'modern' strains have evolved a strategy of higher virulence and shorter latency, perhaps to better adapt to the increasing availability of susceptible hosts brought by the large human population increases of the last few centuries. Further follow-up of HHCs would provide us with a unique opportunity to determine whether the secondary cases would be more associated with MTB clustering or a particular lineage, and also to assess if secondary cases occur with shorter latency [18].

Finally, we also investigated factors potentially associated with TB transmission within households. In a previous study conducted by our group [12], it was observed that index case factors associated with HT were severity of cough and cavitation on chest radiograph. However, in the present study, no differences in sociodemographic and clinical variables for clustering were observed, with the exception that in the high transmission group, a higher proportion of household contacts was vaccinated with BCG, as demonstrated by the presence of BCG scar ($P=0.05$). Although statistically significant, this difference associated with the presence of BCG scar in high transmission households requires further investigation.

A potential limitation in our study is the possibility of selection bias induced through both the use of RFLP/spoligotyping for cluster definition and for the assumption that MTB strains belonging to cluster RFLP patterns represent recent transmission. In a recent study based on whole-genome sequencing (WGS) conducted by Casali and colleagues [30] it was observed that the relative lack of between-patient diversity compared to within-specimen diversity might suggest that transmission in institutional or household settings may not involve a narrow evolutionary bottleneck, or that the population typically involves one very abundant clone and multiple rare or minor clones. Based on these facts and considering that in the future when decreasing sequencing costs and upcoming bioinformatics developments to facilitate quick and relevant interpretation of the resulting data, it will be possible to a better comprehend MTB transmission dynamics in our setting.

Despite this possible limitation, we found that clusters and lineages associated with MTB infection in HT households had no proclivity for increased transmission of TB in the community. There are several possible explanations that are not mutually exclusive. First,

the household transmission may relate to characteristics of the host (infectiousness of the source case, susceptibility of the household contact). Indeed, TB case production of infectious aerosols has been associated with transmission to household contacts [31]. Second, the MTB isolates that are most transmissible may be inherently less likely to cause disease. Additional characterization of the isolates, including transcriptomics and lipidomics, will be important to address this issue. Last, the MTB isolates in high transmission household may be more likely to induce delayed-type hypersensitivity and manifest infection as a positive TST. This was described previously for the MTB strain CDC 1551 [32].

Additional follow-up of HHCs will be allow determination of whether incident cases of TB are more likely to arise in high or low transmission households and will provide an important perspective to the analysis presented.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the invaluable contribution made by Roberto Rabello de Souza, Aryadne Lyrio de Oliveira and Andressa da Silva Borges (study nurses), Antônio Lourenço Canal (NDI driver), Leduc Mageski, João Batista Pereira da Silva and Mário César Gomes (NDI TB laboratory technicians). We also wish to thank all the personnel affiliated with the TB programs in Vitória, Serra, Cariacica and Vila Velha for their interest and dedication to this study.

FUNDING

This study was supported by NIH-NIAID award U01 AI-065663 (International Collaboration in Infectious Diseases Research) and funds from FAPES (Process No 54694337/2011). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Dye C, Scheele S, Dolin P, Pathania V, Ravigliione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA. 1999; 282:677–686. [PubMed: 10517722]
2. Rhee JT, Piatek AS, Small PM, Harris LM, Chaparro SV, Kramer FR, et al. Molecular epidemiologic evaluation of transmissibility and virulence of *Mycobacterium tuberculosis*. J Clin Microbiol. 1999; 37(6):1764–70. [PubMed: 10325321]
3. Cattamanchi A, Hopewell PC, Gonzalez LC, Osmond DH, Masae Kawamura L, Daley CL, et al. A 13-year molecular epidemiological analysis of tuberculosis in San Francisco. Int J Tuberc Lung Dis. 2006; 10(3):297–304. [PubMed: 16562710]
4. Golub JE, Cronin WA, Obasanjo OO, Coggin W, Moore K, Pope DS, et al. Transmission of *Mycobacterium tuberculosis* through casual contact with an infectious case. Arch Intern Med. 2001; 161:2254. [PubMed: 11575983]
5. MacIntyre CR, Carnie J, Randall M. Risk of transmission of tuberculosis among inmates of an Australian prison. Epidemiol Infect. 1999; 123:445–50. [PubMed: 10694155]
6. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies. Int J Tuberc Lung Dis. 2008; 12:480–492. [PubMed: 18419882]
7. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. Eur Respir J. 2013; 41(1):140–56. [PubMed: 22936710]

8. Yim J, Selvaraj P. Genetic susceptibility in tuberculosis. *Respirology*. 2010; 15:241–256. [PubMed: 20199642]
9. van der Spuy GD, Warren RM, Richardson M, Beyers N, Behr MA, van Helden PD. Use of genetic distance as a measure of ongoing transmission of *Mycobacterium tuberculosis*. *J Clin Microbiol*. 2003; 41:5640–5644. [PubMed: 14662954]
10. Moonan PK, Ghosh S, Oeltmann JE, Kammerer JS, Cowan LS, Navin TR. Using genotyping and geospatial scanning to estimate recent mycobacterium tuberculosis transmission, United States. *Emerg Infect Dis*. 2012; 18(3):458–65. [PubMed: 22377473]
11. Coscolla M, Gagneux S. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov Today Dis Mech*. 2010; 7(1):43–59.
12. Jones-López EC, Kim S, Fregona G, Marques-Rodrigues P, Hadad DJ, Molina LP, et al. Importance of cough and *M. tuberculosis* strain type as risks for increased transmission within households. *PLoS One*. 2014; 9(7):e100984. [PubMed: 24988000]
13. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol*. 1993; 31:406–409. [PubMed: 8381814]
14. Brudey K, Driscoll J, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, 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. [PubMed: 16519816]
15. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol*. 1988; 26:2465–2466. [PubMed: 3069867]
16. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. SITVITWEB--a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol*. 2012; 12:755–766. [PubMed: 22365971]
17. Coscolla M, Gagneux S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin Immunol*. 2014; 26(6):431–44. [PubMed: 25453224]
18. Gagneux S. Host-pathogen coevolution in human tuberculosis. *Philos Trans R Soc Lond B Biol Sci*. 2012; 367(1590):850–859. [PubMed: 22312052]
19. Marais BJ, Hesselink AC, Schaaf HS, Gie RP, van Helden PD, Warren RM. *Mycobacterium tuberculosis* transmission is not related to household genotype in a setting of high endemicity. *J Clin Microbiol*. 2009; 47(5):1338–43. [PubMed: 19261801]
20. Verhagen LM, van den Hof S, van Deutekom H, Hermans PW, Kremer K, Borgdorff MW, et al. Mycobacterial factors relevant for transmission of tuberculosis. *Infect Dis*. 2011; 203(9):1249–55.
21. Nebenzahl-Guimaraes H, Verhagen LM, Borgdorff MW, van Soolingen DJ. Transmission and Progression to Disease of *Mycobacterium tuberculosis* Phylogenetic Lineages in The Netherlands. *Clin Microbiol*. 2015; 53(10):3264–71.
22. Albanna AS, Reed MB, Kotar KV, Fallow A, McIntosh FA, Behr MA, et al. Reduced transmissibility of East African Indian strains of *Mycobacterium tuberculosis*. *PLoS One*. 2011; 6(9)
23. de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, et al. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *J Infect Dis*. 2008; 198(7):1037–43. [PubMed: 18702608]
24. Nava-Aguilera E, Andersson N, Harris E, Mitchell S, Hamel C, Shea B, et al. Risk factors associated with recent transmission of tuberculosis: systematic review and metaanalysis. *Int J Tuberc Lung Dis*. 2009; 13:17–26. [PubMed: 19105874]
25. Houben RM, Glynn JR. Systematic review and analysis of population-based molecular epidemiological studies. *Int J Tuberc Lung Dis*. 2009; 13(2):275. [PubMed: 19146761]
26. van Soolingen D, Borgdorff MW, de Haas PE, Sebek MM, Veen J, Dessens M, et al. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis*. 1999; 180:726–736. [PubMed: 10438361]

27. Houben RM, Glynn JR. A systematic review and meta-analysis of molecular epidemiological studies of tuberculosis: development of a new tool to aid interpretation. *Trop Med Int Health*. 2009; 8:892–909.
28. Borgdorff MW, Van Den Hof S, Kalisvaart N, Kremer K, van Soolingen D. Influence of sampling on clustering and associations with risk factors in the molecular epidemiology of tuberculosis. *Am J Epidemiol*. 2011; 174:243–51. [PubMed: 21606233]
29. Glynn JR, Bauer J, de Boer AS. Interpreting DNA fingerprint clusters of *Mycobacterium tuberculosis*. European concerted action on molecular epidemiology and control of tuberculosis. *Int J Tuberc Lung Dis*. 1999; 3:1055–60. [PubMed: 10599007]
30. Casali N, Broda A, Harris SR, Parkhill J, Brown T, Drobniewski F. Whole Genome Sequence Analysis of a Large Isoniazid-Resistant Tuberculosis Outbreak in London: A Retrospective Observational Study. *PLoS Med*. 2016; 13(10):e1002137. [PubMed: 27701423]
31. Birring SS, Prudon B, Carr AJ, Singh SJ, Morgan MD, Pavord ID. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). *Thorax*. 2003; 58:339–343. [PubMed: 12668799]
32. Manca C, Tsenova L, Barry CE 3rd, Bergtold A, Freeman S, Haslett PA, et al. *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. *J Immunol*. 1999; 162(11):6740–6. [PubMed: 10352293]

Figure 1.
Study Flowchart.

Figure 2. Profile clustering (PCb) of MTB isolates from household transmission study. Genotypes (IS6110 RFLP patterns and spoligotypes) were grouped according to three approaches: genotypes clustering only within household transmission study isolates (G1); genotypes clustering with only historical/concurrent controls isolates (G2) and genotypes clustering with both historical/concurrent controls and household transmission study isolates (G3). Household contacts (HHC)a transmission categories are depicted in the second column as high (H), low (L) and intermediate (I).

Table 1

Numbers and frequencies of *MTB* isolates clustered assorted by RFLP and Spoligotyping in Vitoria Metropolitan Area.

Genotype	Household transmission study ^a (n=139)	Historical and concurrent controls (n=841)	All study population (n=980)
Unique pattern (%)	99 (71.0)	472 (56.0)	541 (55.0)
Cluster (%)	40 (29.0)	369 (44.0)	439 (45.0)
No of clusters	15	101	115
Clustering rate	0.17	0.32	0.34
Total of genotypes	115	573	651
Hunter & Gaston diversity index	0.94	0.99	0.99

^aHigh, low and intermediate categories.

Table 2

Genotypes identified within study transmission groups and similarity to historical and concurrent controls.

Transmission group	Index cases (n)	Household transmission study only			Household transmission study/HCC ^a		
		Unique pattern	Cluster	P	Unique pattern	Cluster	P
High	84	58* (69%)	26 (31%)	0.82	48* (57%)	36 (43%)	0.2
Low	25	16* (64%)	9 (36%)		10 (40%)	15 (60%)	
Intermediate	30	25* (83%)	5 (17%)		19* (63%)	11 (37%)	

* Higher frequency of unique pattern (p<0.05) in all groups, except for Low transmission group when analyzed within the context of historical and concurrent controls.

^a Historical and concurrent controls.

Table 3

MTB genotypes identified from index cases of household contacts with skin test (TST) conversion.

Study group	Index cases	HHC*	Persistent TST –			Follow-up TST +/Baseline TST –			Baseline TST+		
			n (%)	p-value (within group)	p-value (between groups)	n (%)	p-value (within group)	p-value (between groups)	n (%)	p-value (within group)	p-value (between groups)
High	84	376	39			31			306		
Unique pattern	48 (58%)	214 (57%)	22 (56%)			16 (52%)			176 (58%)		
Cluster	36 (42%)	162 (43%)	17 (44%)	0.42		15 (48%)	0.86		130 (42%)	0.01	
Low	25	176	112			27			37		0.5
Unique pattern	10 (40%)	95 (54%)	59 (53%)			19 (70%)			17 (46%)		
Cluster	15 (60%)	81 (46%)	53 (47%)	0.57		8 (30%)	0.03		20 (54%)	0.62	
Total**	139	741	206			80			455		
Unique pattern	77 (55%)	411 (55%)	105 (51%)			48 (60%)			258 (57%)		
Cluster	62 (45%)	330 (45%)	101 (49%)			32 (40%)			197 (43%)		

* Household contacts

**

Total = the sum of index or HHC subjects from high, intermediate and low transmission groups (intermediate not included in the table).

APPENDIX 1

Characteristics of index cases at baseline by *MTB* genotypes and transmission category.

Characteristic	Low			High			Overall		
	Cluster	Non-cluster	p	Cluster	Non-cluster	p	Cluster	Non-cluster	p
N	15	10		35	48		62	77	
Age (years) ^a	39.0 (30.7–42.4) ^c	36.2 (30.2–46.7)	0.78	30.0 (22.1–47.2)	35.2 (27.1–44.9)	0.49	35.9 (23.3–46.7)	36.6 (27.2–45.8)	0.69
Sex			0,79			0,94			0,49
Male	6 (60)	11 (73)		30 (63)	23 (66)		48 (62)	43 (69)	
Female	4 (40)	4 (27)		18 (38)	12 (34)		29 (38)	19 (31)	
Cough Visual Analog Score ^a	5.5 (3.3–8.0)	5.0 (4.3–6.0)	0.91	8.0 (5.3–9.0)	9.0 (6.0–9.0)	0.65	7.0 (4.0–9.0)	8.0 (5.0–9.0)	0.33
Leicester Cough questionnaire ^a	80.5 (68.0–105.5)	88.0 (78.0–100.0)	0.66	79.0 (67.0–91.0)	69.5 (66.0–84.0)	0.21	80.0 (68.0–92.0)	73.0 (67.0–90.)	0.28
Weeks sick prior to enrollment	8.4 (4.2–21.0)	8.4 (8.4–15.8)	0.98	12.6 (8.4–23.1)	12.6 (7.4–22.1)	0.43	10.5 (4.2–16.8)	12.6 (8.4–21.0)	0.97
Housing			1			1			1
Owned	10 (1)	14 (93.3)		38 (79.2)	27 (77.1)		66 (85.7)	53 (85.5)	
Rented	0 (0)	1 (6.7)		10 (20.8)	8 (22.9)		11 (14.3)	9 (14.5)	
Sputum AFB smear			1			0.67			0.84
2+	3 (30)	4 (27)		8 (17)	8 (23)		14 (18)	13 (21)	
3+	7 (70)	11 (73)		40 (83)	27 (77)		63 (82)	49 (79)	
Cavitations			1			1			0.86
Absent	4 (40)	6 (40)		8 (17)	6 (17)		20 (27)	15 (24)	
Present	6 (60)	9 (60)		38 (83)	29 (83)		54 (73)	47 (76)	
Extent of Disease on CXR ^b			0.53			0.25			0.91

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Characteristic	Overall			High			Low		
	Cluster	Non-cluster	p	Cluster	Non-cluster	p	Cluster	Non-cluster	p
Normal	1 (1)	1 (1)		0	0		1 (10)	0 (0)	
Minimal	3 (4)	4 (5)		1 (2)	3 (9)		0 (0)	1 (10)	
Moderate	39 (63)	30 (48)		25 (54)	14 (30)		6 (40)	9 (60)	
Far advanced	31 (50)	27 (44)		20 (57)	18 (39)		3 (20)	5 (33)	

^a median (IQR);
^b Chest X-ray;
^c Confidence interval

APPENDIX 2

Characteristics of household contacts at baseline by *MTB* genotypes and transmission category.

Characteristic	Low			High			Overall		
	Cluster	Non-cluster	p	Cluster	Non-cluster	p	Cluster	Non-cluster	p
N	81	95		162	214		330	411	
Age (years) ^a	27.0 (12.0–47.0)	33.0 (15.0–48.0)	0.42	29.5 (13.0–53.5)	22.0 (10.0–46.0)	0.13	28.5 (12.0–51.8)	26.0 (11.0–47.0)	0.36
Sex			1			0.76			0.29
Male	45 (47)	38 (47)		88 (41)	70 (43)		171 (42)	151 (46)	
Female	50 (53)	43 (53)		126 (59)	92 (57)		240 (58)	179 (54)	
BCG Scar			0.34			0.05			0.58
Absent	18 (19)	9 (11)		29 (14)	31 (19)		57 (14)	55 (17)	
Present	74 (79)	70 (86)		174 (81)	129 (80)		340 (83)	265 (80)	
Uncertain	2 (2)	2 (2)		11 (7)	2 (1)		13 (4)	10 (3)	
Close contact with other TB case			0.21			0.59			0.83
Yes	11 (12)	4 (5)		53 (25)	45 (28)		89 (22)	68 (21)	
No	84 (88)	75 (95)		161 (75)	117 (72)		322 (78)	260 (79)	

^a median (IQR);

^b Confidence interval.

APPENDIX 3

Genotypes profile found only in High and Low transmission categories and shared by them.

High					Low					Shared High & Low				
Index case	RFLP Id	Spoligo family	P.C. ^a	Index case	RFLP Id	Spoligo family	P.C. ^a	Index case	RFLP Id	Spoligo family	P.C. ^a	Index case	RFLP Id	Spoligo family
VT-3020	ES1h	LAM3	G2	SE-3040	ES9c	LAM6	G2	CA-3063	ES1b	LAM3	G3			
CA-3001	ES4	S	G2	SE-3020	ES19f	LAM5	G2	SE-3025	ES1b	LAM3	G3			
CA-3051	ES6	Orphan	G2	SE-3030	ES21	LAM9	G2	CA-3004	ES8	T1	G3			
CA-3040	ES6a	Orphan	G2	VT-3042	ES4	S	G2	SE-3015	ES8	T1	G3			
VV-3005	ES14c	LAM9	G2	CA-3016	ES25	LAM9	G2	SE-3052	ES8	T1	G3			
SE-3027	ES14d	LAM9	G2	CA-3055	ES30c	H1	G2	SE-3066	ES8	T1	G3			
CA-3003	ES14j	Orphan	G2	VT-3021	ES14h	LAM9	G2	SE-3054	ES8	Orphan	G3			
SE-3069	ES14k	LAM9	G2					SE-3064	ES8	Orphan	G3			
SE-3043	ES19b	LAM9	G2					VT-3063	ES11a	T4-CEU1	G3			
SE-3035	ES24a	LAM9	G2					VV-3003	ES11a	T4-CEU1	G3			
VT-3005	ES30a	H3	G2					VT-3050	ES11a	T4-CEU1	G3			
								SE-3070	ES14	LAM9	G3			
CA-3005	ES32	Unknown	G1					CA-3066	ES14	LAM9	G3			
VV-3023	ES32	Unknown	G1					CA-3021	ES19c	LAM1	G3			
SE-3075	ES1	LAM3	G1					CA-3036	ES19c	LAM1	G3			
VT-3066	ES1	LAM3	G1					SE-3077	ES19h	LAM2	G3			
								VT-3048	ES19h	LAM2	G3			
CA-3042	ES6k	Unknown	G3					SE-3072	ES19j	LAM2	G3			
VT-3068	ES6k	Unknown	G3					VT-3032	ES19j	LAM2	G3			
VT-3003	ES13c	Orphan	G3					VT-3049	ES19j	LAM2	G3			
VT-3014	PB-4	X1	G3					VV-3030	ES19l	LAM2	G3			
CA-3034	ES1e	LAM3	G3					SE-3002	ES19l	LAM2	G3			
SE-3031	ES1e	LAM3	G3											
SE-3026	ES30e	H3	G3											

^aProfile clustering: Group 1 (G1) = Clustering only within transmission group strains; Group 2 (G2) = Clustering with only background strains; Group 3 (G3) = Clustering with both background and transmission group strains. Eleven index cases were not included in the table because they belonged to the intermediate group