Published in final edited form as:
*Infect Genet Evol*. 2012 June ; 12(4): 664–670. doi:10.1016/j.meegid.2011.07.018.

# Distinct Clinical and Epidemiological Features of Tuberculosis in New York City Caused by the RD$^{Rio}$ *Mycobacterium tuberculosis* Sublineage

**Scott A. Weisenberg**[1,9,‡], **Andrea L. Gibson**[1,‡], **Richard C. Huard**[1,4], **Natalia Kurepina**[3], **Heejung Bang**[2], **Luiz C O. Lazzarini**[1,5], **Yalin Chiu**[2], **Jiehui Li**[6], **Shama Ahuja**[6], **Jeff Driscoll**[7,8], **Barry N. Kreiswirth**[3], and **John L. Ho**[1,*]

[1]Division of Infectious Diseases, Department of Medicine, Weill Medical College of Cornell University, New York, USA.

[2]Division of Biostatistics and Epidemiology, Department of Public Health, Weill Medical College of Cornell University, New York, USA.

[3]Public Health Research Institute Tuberculosis Center, Newark, New Jersey, USA.

[4]Clinical Microbiology Service and the Department of Pathology, New York-Presbyterian Hospital, Columbia University Medical Center, New York, NY.

[5]Tuberculosis Research Unit, Medical School of Federal University of Rio de Janeiro. Brazil.

[6]Bureau for Tuberculosis Control, New York City Department of Health and Mental Hygiene, New York, NY.

[7]Wadsworth Center, New York State Department of Health, Albany, NY.

## Abstract

**Background—**Genetic tracking of *Mycobacterium tuberculosis* is a cornerstone of tuberculosis (TB) control programs. The RD$^{Rio}$ *M. tuberculosis* sublineage was previously associated with TB in Brazil. We investigated 3847 *M. tuberculosis* isolates and registry data from New York City (NYC) (2001–2005) to: 1) affirm the position of RD$^{Rio}$ strains within the *M. tuberculosis* phylogenetic structure, 2) determine its prevalence, and 3) define transmission, demographic, and clinical characteristics associated with RD$^{Rio}$ TB.

**Methods—**Isolates classified as RD$^{Rio}$ or non-RD$^{Rio}$ *M. tuberculosis* by multiplex PCR were further classified as clustered (≥2 isolates) or unique based primarily upon IS*6110*-RFLP patterns and lineage-specific cluster proportions were calculated. The secondary case rate of RD$^{Rio}$ was compared with other prevalent *M. tuberculosis* lineages. Genotype data were merged with the data from the NYC TB Registry to assess demographic and clinical characteristics.



*Correspondence should be addressed to: John L. Ho; M.D, current address: Southern Maine Medical Center, One Medical Drive, Biddeford, ME; Tel: 207-283-7919; millennium.john@gmail.com.
[8]Currently at Laboratory Branch, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, Atlanta, GA.
[9]Currently at Alta Bates Summit Medical Center, Oakland, CA.
‡Authors contributed equally to the work, (the rest of the author credits go in a different section, depending on the journal)



Disclosures

This study was presented in part as an abstract at the 2008 46th Annual IDSA/48th Annual ICAAC Joint Meeting (Abstract F1-2015)

**Results—**RD$^{Rio}$ strains were found to: 1) be restricted to the Latin American-Mediterranean family, 2) cause approximately 8% of TB cases in NYC, and 3) be associated with heightened transmission as shown by: *i*) a higher cluster proportion compared to other prevalent lineages, *ii*) a higher secondary case rate, and *iii*) cases in children. Furthermore, RD$^{Rio}$ strains were significantly associated with US-born Black or Hispanic race, birth in Latin American and Caribbean countries, and isoniazid resistance.

**Conclusions—**The RD$^{Rio}$ genotype is a single *M. tuberculosis* strain population that is emerging in NYC. The findings suggest that expanded RD$^{Rio}$ case and exposure identification could be of benefit due to its association with heightened transmission.

### Keywords

tuberculosis; lineage; epidemiology; transmission; RD$^{Rio}$

## Introduction

*Mycobacterium tuberculosis* is a major human pathogen that has undergone clonal evolution into divergent lineages that are associated with specific geographic regions and possibly with distinct human ethnic populations (Hirsh et al., 2004; Gagneux et al., 2006). The major *M. tuberculosis* lineages and select sublineages can be defined by specific chromosomal large-sequence polymorphisms (Tsolaki et al., 2004; Reed et al., 2009) and/or groups of single nucleotide polymorphisms (SNPs) (Gutacker et al., 2002). Lineages may also be estimated through molecular epidemiologic tools, including IS*6110*-restriction fragment length polymorphism (RFLP) analysis (van Embden et al., 1993), spoligotyping (Brudey et al., 2006), and mycobacterial interspersed repetitive units (MIRU)-typing (Supply et al., 2000).

Recent reports have suggested that genetic variation between *M. tuberculosis* lineages may translate into measurable biological differences. Lineage-specific differences in rate of growth, ability to induce or evade immune responses, pathogenicity, and expression of virulence factors have been observed previously *in vitro* (Manca et al., 2004; Pheiffer et al., 2005; Riley, 2006) and/or in animal studies (Lpez et al., 2003; Reed et al., 2004; Tsenova et al., 2005). Lineage-specific differences in transmissibility and the resulting clinical manifestations of human tuberculosis (TB) have been previously suggested (van Crevel et al., 2001; Sun et al., 2006; Hanekom et al., 2007a; Kong et al., 2007; de Jong et al., 2008; Feng et al., 2008; Thwaites et al., 2008; van der Spuy et al., 2009). More data are needed to clarify the potential differential capacity of lineages to transmit and cause disease which, in turn, could influence local TB disease burden (Gagneux et al., 2006) and impact public health strategies for TB control, such as the prioritization of certain strains for investigation. Even though, *M. tuberculosis* lineage-specific differences in host ethnic preference are beginning to emerge (Gagneux et al., 2006; Caws et al., 2008), confirmation by large and longitudinal data of combined heterogenous human and *M. tuberculosis* populations is needed.

The Latin American-Mediterranean (LAM) family of *M. tuberculosis* is responsible for ~15% of TB cases reported in the spoligotype-defined SpolDB4 database (over 39,000 international strains) and, as such, is a major *M. tuberculosis* lineage (Brudey et al., 2006; Lazzarini et al., 2007). We recently described the RD$^{Rio}$ genotype as a clonally-derived sublineage within the LAM family (Lazzarini et al., 2007). It is defined by a ~26.3 kb deletion resulting in the loss or modification of 10 genes, including two PPE mycobacterial family genes known to be recognized by host immunity (Singh et al., 2005; Gibson et al., 2008). RD$^{Rio}$ *M. tuberculosis* strains (hereafter called RD$^{Rio}$) possess a coincident deletion,

RD174, at a separate genomic locus (Gibson et al., 2008), and that classifies it as part of the Euro-American/African superlineage (Reed et al., 2009). $RD^{Rio}$ is the most prevalent cause of TB in Rio de Janeiro, Brazil (Lazzarini et al., 2007), and is present in Europe, Africa, and the Americas (Gagneux et al., 2006; de Jong et al., 2008; Gibson et al., 2008). It has been associated with a more "clinically extensive" form of TB with increased sputum bacillary load, lung cavities, hemoptysis, and weight loss, suggesting potential for increased transmissibility (Lazzarini et al., 2007; Lazzarini et al., 2008). It was therefore important to ascertain if the potentially heightened transmissible nature of $RD^{Rio}$ is maintained outside of endemic areas, particularly in locales with a high diversity of human ethnic and racial groups and multiple circulating *M. tuberculosis* lineages; New York City (NYC) is just such a context.

TB in NYC (Li et al., 2003), like the rest of the United States (Talbot et al., 2000), is increasingly a disease of foreign-born individuals. Foreign-born TB cases are often infected by lineages prevalent in their native region (Hirsh et al., 2004). Knowledge of the lineages circulating in NYC and their relative capacity to spread within the multiethnic environment of NYC is therefore critical to understanding current and future *M. tuberculosis* transmission trends. This study aimed to: 1) affirm the phylogenetic position of the $RD^{Rio}$ sublineage within the global *M. tuberculosis* population structure, 2) determine the prevalence of $RD^{Rio}$, and 3) define transmission, demographic, and clinical characteristics associated with $RD^{Rio}$ TB in comparison with other lineages of *M. tuberculosis*.

## Materials and Methods

### Study Population

From 2001–2005, 5508 cases of TB were diagnosed in NYC; 4202 (76.3%) were culture-positive for *M. tuberculosis*. IS*6110*-RFLP and spoligotyping data, as well as DNA for further analysis, was available for 3911 (93%) isolates. Isolates from cases with two different strains (n = 22) and those with an indeterminate $RD^{Rio}$/non-$RD^{Rio}$ genotype (n = 42) were excluded from the analysis, leaving a study total of 3847 TB cases. The incidence of TB in NYC in 2005 was 12.3 per $10^5$ (among non-US born 24 per $10^5$(TB Annual Summary, 2005). IRB approval for the study was granted by the Weill Medical College of Cornell University and the NYC Department of Health and Mental Hygiene (NYCDOHMH).

### Molecular Characterization of *M. tuberculosis* isolates

*M. tuberculosis* isolates from NYC TB cases routinely undergo IS*6110*-RFLP, spoligotyping, and MIRU-typing (Clark et al., 2006). In addition, SNP cluster analysis was performed on all isolates from 2001 to 2004 (Gibson et al., 2008), while isolates from 2005 were assigned a SNP cluster based on IS*6110*-RFLP and spoligotyping data by investigators at the Public Health Research Institute. Spoligotype family was assigned using SpolClust [http://cgi2.cs.rpi.edu/~bennek/SPOTCLUST.html] with consideration of IS*6110*-RFLP pattern and major SNP-based cluster data (Vitol et al., 2006). A multiplex PCR assay that differentiates between $RD^{Rio}$ and non-$RD^{Rio}$ strains (Gibson et al., 2008; Lazzarini et al., 2008) was performed on all isolates from 2005, Cluster VI isolates from 2001–2004, and a representative but randomized selection (10%) of isolates from non-Cluster VI groups from 2001–2004 (LAM strains are known to group in Cluster VI) (Gibson et al., 2008).

### Demographic and Clinical Characteristics of NYC TB Cases

Demographic and clinical case data was obtained from the NYC TB Registry (Table 4). Drug use was consolidated as one variable (yes if any injection drug use, non-injection cocaine use, or non-injection drug use), and non-US country of birth was categorized into

region of birth: Latin American-Caribbean (included Puerto Rico for purposes of this study only), Asia (East Asia and South East Asia), and Other (including Canada, Europe, and world regions not already stated).

### Epidemiological Cluster Analysis

Isolates with identical IS*6110*-RFLP fingerprints (with identical spoligotyping if ≤ 5 IS*6110* bands) were considered clustered (Alland et al., 1994; Maguire et al., 2002). The proportion clustered was calculated, and recent transmission was estimated by the "n-1" method (Small et al., 1994; Murray and Alland, 2002). The secondary case rate [number of secondary cases / (index + unique cases)] (Borgdorff et al., 1998) was calculated for $RD^{Rio}$ and other major lineages. The first chronological case cultured was identified as the "index case."

### Statistical Analysis

Categorical variables were compared for different lineages and sublineages using Pearson's chi-square. Fisher's exact test was used for small cell counts. The Jonckheere-Terpstra test (Jonckheere, 1954) was used to test for age trend. The proportion of $RD^{Rio}$ among TB cases in children (age 0–10) was compared to the proportion of $RD^{Rio}$ in older TB cases (age 11–100). All *Mycobacterium bovis* cases were excluded from the subanalysis, as the *M. bovis*-infected children were part of a food borne outbreak (CDC, 2005). Confidence intervals for the secondary case rate with a given family were calculated using the bootstrap method with 5000 resamples (Efron and B, 1982), noting that this rate formulation is not based on standard binomial distribution. To characterize cases harboring $RD^{Rio}$ compared to non-$RD^{Rio}$ *M. tuberculosis*, we examined the demographic, clinical and social history variables using univariate and multivariate logistic regression models, from which unadjusted and adjusted odds ratios along with the 95% confidence intervals were estimated. Statistical analysis was performed using SAS (version 9.1, SAS Systems, Cary, NC).

## Results

### $RD^{Rio}$ Prevalence in NYC and Phylogenetic Position

The analysis included 3847 unique-case *M. tuberculosis* strains from 2001 to 2005. There was a widespread distribution of lineages, including LAM ($n$ = 666, 17.3%), W/Beijing ($n$ = 581, 15.1%), Haarlem ($n$ = 626, 16.3%), T ($n$ = 568 14.8%), Central Asian (CAS, ($n$ = 182, 4.7%), and East African-Indian (EAI, $n$ = 304, 8%) (Table 1). We found 7.9% ($n$ = 302) of the NYC *M. tuberculosis* isolates were $RD^{Rio}$ strains. Of these $RD^{Rio}$ isolates, 289 (95.7%) had been previously categorized as LAM while 13 (4.3%) had been previously categorized as "non-LAM". Upon further genetic analysis, the latter 13 isolates each had a LAM-defining $Ag85C^{103}$ SNP (Gibson et al., 2008); thus confirming that all $RD^{Rio}$ isolates are members of the LAM family (Gibson et al., 2008). Furthermore, this study also affirmed that all $RD^{Rio}$ strains belong to the SNP Cluster VI grouping (Gutacker et al., 2002).

### Transmission Characteristics of $RD^{Rio}$

Fifty-one percent of the $RD^{Rio}$ strains were part of a cluster, a proportion higher than the phylogenetically related non-$RD^{Rio}$ LAM (42.6%, p=0.035) (Table 2). $RD^{Rio}$ strains also had a greater degree of clustering than other prominent lineages, including W/Beijing, but not the Haarlem family (Table 2). Because the index case could represent reactivation, we next used the "n-1" method to estimate recent transmission (Small et al., 1994; Murray and Alland, 2002). By this method, $RD^{Rio}$ strains showed a greater proportion of clustering than the non-$RD^{Rio}$ LAM and each of the unrelated families, excluding Haarlem (Table 2). A higher secondary case rate (Borgdorff et al., 2000) was also found for the $RD^{Rio}$ sublineage than non-$RD^{Rio}$ LAM strains and other prominent lineages (except Haarlem) (Table 3).

Finally, $RD^{Rio}$ caused 19% (7/37; p=0.02) of all TB cases in children 10 years and younger, a proxy marker for recently transmitted TB. A trend was observed in the prevalence of $RD^{Rio}$ across age groups using the Jonckheere-Terpstra test, (p=0.03).

### Demographic and Clinical Characteristics of $RD^{Rio}$ TB Cases

Multiple logistic regression analysis found a statistically significant association between TB cases with $RD^{Rio}$ and race/ethnicity, year of isolation (increased over study period), region of birth, and drug resistance (Table 4). TB amongst Blacks and Hispanics showed, respectively, nearly 2.3 (95% CI: 1.17–4.63) and 3.3 (95% CI: 1.60–6.99) times greater odds of having an $RD^{Rio}$ strain, even after adjusting for birth in Latin American-Caribbean countries along with other factors. INH resistance was associated with $RD^{Rio}$ (OR 1.65 [95% CI: 1.09–2.51]), as was cavitary disease (though not statistically significant on multivariate analysis (OR 1.35 [95% CI: 0.98–1.87])).

Further descriptive exploratory analyses of transmission patterns revealed that TB caused by $RD^{Rio}$ was associated with case race/ethnicity and birthplace, with differences noted in the percentage of Hispanics and Blacks harboring clustered or non-clustered (unique) $RD^{Rio}$ strains. Of the 154 cases with a clustered $RD^{Rio}$ strain, 145 (94.1%) were Hispanic or non-Hispanic Black. Of the 93 clustered Hispanic cases, 78 (83.9%) were foreign-born from Latin-American or Caribbean regions. In contrast, 33 (64.7%) of the 51 clustered non-Hispanic Black cases were US-born. Among 148 unique “non-clustered” $RD^{Rio}$ case-strains, 67 (85.9%) Hispanics and 24 (43.6%) Blacks were born in Latin American or Caribbean countries. Differences in age at presentation were also noted between US-born Hispanic and Black $RD^{Rio}$ TB cases – the median age was 24 years for Hispanics (from $n = 25$ with the range of 0–52) and 50 years for Blacks ($n = 52$, range of 7–97).

We also looked at the demographic characteristics of each individual $RD^{Rio}$ cluster. Of 43 $RD^{Rio}$ clusters, 5 (11.6%) consisted of US-born cases only, 18 (41.9%) consisted of foreign-born cases only and 20 (46.5%) consisted of mixed (both US- and foreign-born) cases. Thirty of the 43 (69.8%) clusters involved only Hispanic (n=21) or Black (n=9) TB cases; while the remaining 13 (30.2%) clusters contained at least one Hispanic and/or Black TB case. Four of the US-born-only clusters were exclusively comprised of Black or Hispanic cases (n=3 or n=1, respectively), with a single cluster involving both Black and Hispanic cases. Three Hispanic-only, 2 Black-only, and 2 multi-race clusters contained a TB case age 10 years or younger (both children in the multi-race clusters were Hispanic). Overall, these data suggest distinct transmission patterns for different clusters of $RD^{Rio}$ strains that may be host ethnicity related. Among the foreign-born only clusters, 12 (66.7%) were constituted entirely of cases originating from the same country (Dominican Republic [$n$ clusters=5], Ecuador [$n = 3$], Mexico [$n = 2$], Barbados [$n = 1$], and Puerto Rico [$n = 1$]). These data add to the number of countries from which $RD^{Rio}$ has previously been associated (Gibson et al., 2008). With respect to the mixed origin $RD^{Rio}$ clusters, the index case was foreign-born in 15 (75%) and US-born in 5 (25%).

## Discussion

The $RD^{Rio}$ deletion is thought to have arisen from a homologous recombination event between non-mobile genes that could have theoretically occurred more than once as homoplastic events in strains from different lineages (Lazzarini et al., 2007). In this study, we confirmed that $RD^{Rio}$ is strictly a sublineage of the LAM family and further showed that it is enfolded within the broader SNP Cluster VI grouping. The localization of $RD^{Rio}$ strains to a single family suggests that the global distribution of $RD^{Rio}$ strains (Gibson et al., 2008) is the result of the dissemination of a single successful clonal sublineage.

A diverse *M. tuberculosis* clade distribution was observed in this study ($n$=3847, 2001 to 2005) which is consistent with the *M. tuberculosis* families found in other cities with significant and diverse immigrant communities (Hirsh et al., 2004; Gagneux et al., 2006; Reed et al., 2009) (Supplemental Table 1 and 2). We report that the LAM spoligotype family was responsible for the highest proportion by lineage (17.3%) of the NYC TB cases from 2001–2005, consistent with its numerical importance amongst *M. tuberculosis* families in the SpolDB4 international collection database (Brudey et al., 2006). However, less attention has been dedicated to studying this lineage in part because, unlike W/Beijing in the 1990s, it has not been associated with large outbreaks of multi-drug resistant TB or with vulnerable populations (Frieden et al., 1995; Geng et al., 2002). Moreover, the $RD^{Rio}$ sublineage of LAM caused ~8% of TB cases in NYC and 45.3% of all LAM-related TB - a substantial proportion for a single *M. tuberculosis* sublineage competing for hosts against multiple *M. tuberculosis* genotypes in a multiethnic environment such as NYC.

In this study, we used complimentary methods to estimate $RD^{Rio}$ transmission. Prior to this study, few investigations have calculated secondary case rates for TB (Borgdorff et al., 1998) or have reported *M. tuberculosis* sublineage-specific rates of clustering (Hanekom et al., 2007b; Iwamoto et al., 2009). The data indicate that $RD^{Rio}$ strains may be more transmissible as compared to most other major lineages in NYC, including the phylogenetically related non-$RD^{Rio}$ LAM. Heretofore, such an analysis has not been possible as there are few recognized sublineages that make up a sufficient proportion of the other major *M. tuberculosis* families. These data therefore suggest that $RD^{Rio}$ will persist in NYC and increasingly become a source of domestic TB transmission amongst and between foreign- and native-born populations in the US. Of note, strains possessing an $RD^{Rio}$-coincident deletion RD174 had an elevated secondary case rate ratio in San Francisco, CA, (note in ref. de Jong et al., 2008), and higher active TB case rate in RD174 strain-exposed persons in The Gambia compared to those exposed to other lineages (de Jong et al., 2008). Since the RD174 deletion is restricted to $RD^{Rio}$ strains (Gibson et al., 2008), the prior reports were likely describing $RD^{Rio}$ strains, thereby, supporting our observations.

*M. tuberculosis* is an exquisitely well-adapted human pathogen as a result of a long co-evolutionary history with its human host (Hershberg et al., 2008; Wirth et al., 2008). Recently, Gagneux et al. (Gagneux et al., 2006) proposed that certain *M. tuberculosis* lineages have differentially adapted to specific human populations, such that transmission within these ethnic groups may be more efficient. Our data indicates that $RD^{Rio}$ exhibits human host associations consistent with its phylogeography. The dual host association (Hispanics and Blacks) exhibited by $RD^{Rio}$ in NYC is a key difference from previous studies (Gagneux et al., 2006; Reed et al., 2009) and raises the question of whether genetically distinct $RD^{Rio}$ stains, with different host associations, may be circulating. The high proportion of racially segregated Black or Hispanic clusters supports this hypothesis and suggests that separate mini-epidemics of $RD^{Rio}$ may be occurring in NYC with distinct transmission dynamics. Intriguingly, in a post-hoc analysis of secondary cases, Hispanics with $RD^{Rio}$ strains had a higher proportion of cavities on chest radiographs than those with non-$RD^{Rio}$ strains (31.5% vs. 21.8%), whereas non-Hispanics had similar cavitation rates for both (23.3% vs. 21.4%). This suggests possible host-pathogen specific disease manifestations, but is a hypothesis requiring further study (Lazzarini et al., 2007; Lazzarini et al., 2008).

An important limitation of the current study is that we are comparing a genomic deletion-defined sublineage to the less precise IS*6110*-RFLP and spoligotyping methods to define other lineages. Nevertheless, these other lineages provide a background against which to measure the characteristics of the $RD^{Rio}$ sublineage. There may well be sublineages within each of these lineages with higher cluster rates and unique demographic characteristics

(Hanekom et al., 2007b). Another limitation is that, although self-reported race/ethnicity is a reasonably good predictor of racial heritage (Rosenberg et al., 2002), some Latin American countries have greatly admixed populations that blur racial lines. We also could not completely exclude all potential confounding factors, including unmeasured social-economic factors associated with both $RD^{Rio}$ and with delayed access to health care. However, by comparing the transmissibility of the $RD^{Rio}$ sublineage to that of the non-$RD^{Rio}$ LAM lineage (from which $RD^{Rio}$ was derived) we blunt the effect of some potential confounding factors, such as differing cultural norms and a lack of social admixing among ethnic groups. This has not been possible in prior analyses of other lineages with minimally overlapping phylogeographies (Gagneux et al., 2006; Reed et al., 2009), thus strengthening the validity of our observations. Even though this study lacked comparative pathogenicity data in an animal model, the comparison of multiple *M. tuberculosis* lineages in an ethnically and racially diverse human population with $RD^{Rio}$, and in particular the closely related non-$RD^{Rio}$ LAM strains with $RD^{Rio}$ strains, overcomes this perceived limitation. Lastly, the association of clustered IS*6110*-RFLP strains and recent transmission is presumed in molecular epidemiology studies, but may include cases of coincident reactivation while it may miss clusters where the index or secondary case is not in the database.

In conclusion, the present study is one of the few to link data on molecular epidemiology, case demographics, clinical disease characteristics, and strain genotypes to clarify the transmission dynamics of a *M. tuberculosis* sublineage within a major metropolitan area (Hanekom et al., 2007b; Iwamoto et al., 2009). The $RD^{Rio}$ clade is clearly emerging as an important cause of TB, both internationally and in the US, with an increased propensity to cause cavitary disease and with increased transmission efficiency among US-born Blacks and Hispanics and persons of Latin American-Caribbean heritage. Using $RD^{Rio}$ as a model, future studies should aim to enhance our knowledge of the contributions the host and the pathogen on TB transmission and will aid the future targeting and evaluation of public health efforts in TB control.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Funding for this study was provided by National Institutes of Health (NIH) R21 AI063147, and R21 AI063147 (JLH.), NIH Fogarty International Center Training Grant (FICTG) D43 TW00018 under the AIDS International Training and Research Program (Warren D. Johnson and JLH), ICOHRTA AIDS/TB 5 U2R TW006883 (Jose.R. Lapa e Silva and JLH) and grant KL2RR024997 (SAW) of the Clinical and Translation Science Center at Weill Cornell Medical College (CTSC funded by Grant (UL1 RR024996). The authors thank Dr. Barun Mathema at the PHRI TB Center for his critical review of the manuscript. The authors also thank Tracy Agerton, Cynthia Driver, and Sonal Munsiff of the Bureau for Tuberculosis Control, New York City Department of Health and Mental Hygiene for their assistance in facilitating the conduct of study without which this project would not have been accomplished.

## REFERENCES

Alland D, Kalkut G, Moss A, McAdam R, Hahn J, Bosworth W, Drucker E, Bloom B. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med. 1994; 330:1710–1716. [PubMed: 7993412]

Borgdorff M, Behr M, Nagelkerke N, Hopewell P, Small P. Transmission of tuberculosis in San Francisco and its association with immigration and ethnicity. Int J Tuberc Lung Dis. 2000; 4:287–294. [PubMed: 10777075]

Borgdorff M, Nagelkerke N, van Soolingen D, de Haas P, Veen J, van Embden J. Analysis of tuberculosis transmission between nationalities in the Netherlands in the period 1993–1995 using DNA fingerprinting. Am J Epidemiol. 1998; 147:187–195. [PubMed: 9457010]

Brudey K, 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]

Caws M, et al. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog. 2008; 4 e1000034.

CDC. MMWR Morb Mortal Wkly Human Tuberculosis caused by Mycobacterium bovis--New York City, 2001–2004. 2005 Jun 24; 54(24):605–608.

Clark C, Driver C, Munsiff S, Driscoll J, Kreiswirth B, Zhao B, Ebrahimzadeh A, Salfinger M, Piatek A, Abdelwahab J. Universal genotyping in tuberculosis control program, New York City, 2001–2003. Emerging infectious diseases. 2006; 12:719–724. [PubMed: 16704826]

de Jong B, Hill P, Aiken A, Awine T, Antonio M, Adetifa I, Jackson-Sillah D, Fox A, Deriemer K, Gagneux S, Borgdorff M, McAdam K, Corrah T, Small P, Adegbola R. Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J Infect Dis. 2008; 198:1037–1043. [PubMed: 18702608]

Efron, B. The jackknife, the bootstrap and other resampling plans. Philadelphia: CBNS-NSF; 1982.

Feng J, Su W, Tsai C, Chang S. Clinical impact of Mycobacterium tuberculosis W-Beijing genotype strain infection on aged patients in Taiwan. J Clin Microbiol. 2008; 46:3127–3129. [PubMed: 18596137]

Frieden TR, Fujiwara PI, Washko RM, Hamburg MA. Tuberculosis in New York City--turning the tide. New England Journal of Medicine, The. 1995; 333:229–233.

Gagneux S, Small P. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. The Lancet infectious diseases. 2007; 7:328–337. [PubMed: 17448936]

Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong B, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell P, Small P. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:2869–2873. [PubMed: 16477032]

Geng E, Kreiswirth B, Driver C, Li J, Burzynski J, DellaLatta P, LaPaz A, Schluger N. Changes in the transmission of tuberculosis in New York City from 1990 to 1999. New England Journal of Medicine, The. 2002; 346:1453–1458.

Gibson A, et al. Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio Sublineage of the Latin American-Mediterranean Mycobacterium tuberculosis spoligotype family. J Clin Microbiol. 2008; 46:1259–1267. [PubMed: 18234868]

Gutacker M, Smoot J, Migliaccio CAL, Ricklefs S, Hua S, Cousins D, Graviss E, Shashkina E, Kreiswirth B, Musser J. Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics. 2002; 162:1533–1543. [PubMed: 12524330]

Hanekom M, van der Spuy G, Gey van Pittius N, McEvoy C, Ndabambi S, Victor T, Hoal E, van Helden P, Warren R. Evidence that the spread of Mycobacterium tuberculosis strains with the Beijing genotype is human population dependent. J Clin Microbiol. 2007a; 45:2263–2266. [PubMed: 17475755]

Hanekom M, van der Spuy G, Streicher E, Ndabambi S, McEvoy C, Kidd M, Beyers N, Victor T, van Helden P, Warren R. A recently evolved sublineage of the Mycobacterium tuberculosis Beijing strain family is associated with an increased ability to spread and cause disease. J Clin Microbiol. 2007b; 45:1483–1490. [PubMed: 17360841]

Hershberg R, Lipatov M, Small P, Sheffer H, Niemann S, Homolka S, Roach J, Kremer K, Petrov D, Feldman M, Gagneux S. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol. 2008; 6:e311. [PubMed: 19090620]

Hirsh A, Tsolaki A, DeRiemer K, Feldman M, Small P. Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proceedings of the National

Academy of Sciences of the United States of America. 2004; 101:4871–4876. [PubMed: 15041743]

Iwamoto T, Fujiyama R, Yoshida S, Wada T, Shirai C, Kawakami Y. Population structure dynamics of Mycobacterium tuberculosis Beijing strains during past decades in Japan. J Clin Microbiol. 2009; 47:3340–3343. [PubMed: 19710282]

Jonckheere A. A Distribution Free Kappa Distribution Test Against Ordered Alternatives. Biometrika. 1954; 41:133–145.

Kong Y, Cave MD, Zhang L, Foxman B, Marrs CF, Bates JH, Yang ZH. Association between Mycobacterium tuberculosis Beijing/W lineage strain infection and extrathoracic tuberculosis: Insights from epidemiologic and clinical characterization of the three principal genetic groups of M. tuberculosis clinical isolates. Journal of clinical microbiology. 2007; 45:409–414. [PubMed: 17166963]

Lazzarini L, Spindola S, Bang H, Gibson A, Weisenberg S, da Silva Carvalho W, Augusto C, Huard R, Kritski A, Ho J. RDRio Mycobacterium tuberculosis infection is associated with a higher frequency of cavitary pulmonary disease. J Clin Microbiol. 2008; 46:2175–2183. [PubMed: 18463217]

Lazzarini L, Huard R, Boechat N, Gomes H, Oelemann M, Kurepina N, Shashkina E, Mello F, Gibson A, Virginio M, Marsico A, Butler W, Kreiswirth B, Suffys P, Lapa E, Silva J, Ho J. Discovery of a novel Mycobacterium tuberculosis lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. J Clin Microbiol. 2007; 45:3891–3902. [PubMed: 17898156]

Li J, Driver C, Munsiff S, Yip R, Fujiwara P. Differential decline in tuberculosis incidence among US- and non-US-born persons in New York City. Int J Tuberc Lung Dis. 2003; 7:451–457. [PubMed: 12757046]

Lopez B, Aguilar D, Orozco H, Burger M, Espitia C, Ritacco V, Barrera L, Kremer K, Hernandez-Pando R, Huygen K, van Soolingen D. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. Clinical and experimental immunology. 2003; 133:30–37. [PubMed: 12823275]

Maguire H, Dale J, McHugh T, Butcher P, Gillespie S, Costetsos A, Al-Ghusein H, Holland R, Dickens A, Marston L, Wilson P, Pitman R, Strachan D, Drobniewski F, Banerjee D. Molecular epidemiology of tuberculosis in London 1995–7 showing low rate of active transmission. Thorax. 2002; 57:617–622. [PubMed: 12096206]

Manca C, Reed M, Freeman S, Mathema B, Kreiswirth B, Barry C, Kaplan G. Differential monocyte activation underlies strain-specific Mycobacterium tuberculosis pathogenesis. Infection and immunity. 2004; 72:5511–5514. [PubMed: 15322056]

Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. Am J Epidemiol. 2002; 155:565–571. [PubMed: 11882530]

Pheiffer C, Betts J, Flynn H, Lukey P, van Helden P. Protein expression by a Beijing strain differs from that of another clinical isolate and Mycobacterium tuberculosis H37Rv. Microbiology. 2005; 151:1139–1150. [PubMed: 15817781]

Reed M, Domenech P, Manca C, Su H, Barczak A, Kreiswirth B, Kaplan G, Barry C. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature. 2004; 431:84–87. [PubMed: 15343336]

Reed M, Pichler V, McIntosh F, Mattia A, Fallow A, Masala S, Domenech P, Zwerling A, Thibert L, Menzies D, Schwartzman K, Behr M. Major Mycobacterium tuberculosis lineages associate with patient country of origin. Journal of clinical microbiology. 2009; 47:1119–1128. [PubMed: 19213699]

Riley L. Of mice, men, and elephants: Mycobacterium tuberculosis cell envelope lipids and pathogenesis. Journal of Clinical Investigation. 2006; 116:1475–1478. [PubMed: 16741572]

Rosenberg N, Pritchard J, Weber J, Cann H, Kidd K, Zhivotovsky L, Feldman M. Genetic structure of human populations. Science. 2002; 298:2381–2385. [PubMed: 12493913]

Singh K, Dong Y, Patibandla S, McMurray D, Arora V, Laal S. Immunogenicity of the Mycobacterium tuberculosis PPE55 (Rv3347c) protein during incipient and clinical tuberculosis. Infect Immun. 2005; 73:5004–5014. [PubMed: 16041015]

Small P, Hopewell P, Singh S, Paz A, Parsonnet J, Ruston D, Schecter G, Daley C, Schoolnik G. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med. 1994; 330:1703–1709. [PubMed: 7910661]

Sun Y, Lim T, Ong A, Ho B, Seah G, Paton N. Tuberculosis associated with Mycobacterium tuberculosis Beijing and non-Beijing genotypes: a clinical and immunological comparison. BMC Infect Dis. 2006; 6:105. [PubMed: 16820066]

Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C. Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. Molecular microbiology. 2000; 36:762–771. [PubMed: 10844663]

TB Annual Summary. New York: New York City Department of Health and Mental Hygiene; 2005. 2006.

Talbot E, Moore M, McCray E, Binkin N. Tuberculosis among foreign-born persons in the United States, 1993–1998. JAMA. 2000; 284:2894–2900. [PubMed: 11147986]

Thwaites G, Caws M, Chau TTH, D'Sa A, Lan NTN, Huyen MNT, Gagneux S, Anh PTH, Tho D, Torok E, Nhu NTQ, Duyen NTH, Duy P, Richenberg J, Simmons C, Hien T, Farrar J. Relationship between Mycobacterium tuberculosis genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. Journal of clinical microbiology. 2008; 46:1363–1368. [PubMed: 18287322]

Tsenova L, Ellison E, Harbacheuski R, Moreira A, Kurepina N, Reed M, Mathema B, Barry C, Kaplan G. Virulence of selected Mycobacterium tuberculosis clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. The journal of infectious diseases. 2005; 192:98–106. [PubMed: 15942899]

Tsolaki A, Hirsh A, DeRiemer K, Enciso J, Wong M, Hannan M, Goguet de la Salmoniere Y-OL, Aman K, Kato-Maeda M, Small P. Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:4865–4870. [PubMed: 15024109]

van Crevel R, Nelwan RH, de Lenne W, Veeraragu Y, van der Zanden AG, Amin Z, van der Meer JW, van Soolingen D. Mycobacterium tuberculosis Beijing genotype strains associated with febrile response to treatment. Emerging infectious diseases. 2001; 7:880–883. [PubMed: 11747703]

van der Spuy GD, Kremer K, Ndabambi SL, Beyers N, Dunbar R, Marais BJ, van Helden PD, Warren RM. Changing Mycobacterium tuberculosis population highlights clade-specific pathogenic characteristics. Tuberculosis (Edinb). 2009; 89:120–125. [PubMed: 19054717]

van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. Journal of clinical microbiology. 1993; 31:406–409. [PubMed: 8381814]

Vitol I, Driscoll J, Kreiswirth B, Kurepina N, Bennett KP. Identifying Mycobacterium tuberculosis complex strain families using spoligotypes. Infect Genet Evol. 2006; 6:491–504. [PubMed: 16632413]

Wirth T, Hildebrand F, Allix-Bguec C, Wlbeling F, Kubica T, Kremer K, van Soolingen D, Rsch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S. Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathogens. 2008; 4 e1000160.

## Table 1

### *M. tuberculosis* Lineages in New York City, 2001–2005

Number of TB cases by spoligotype family in the study database. Spoligotype families were defined by SpotClust, and all $RD^{Rio}$ isolates were counted as LAM. Clusters were defined by identical IS*6110*-RFLP pattern (with identical spoligotype if IS*6110*-RFLP ≤5 bands). W/Beijing (also know as East-Asian (Gagneux and Small, 2007)), Central Asian (CAS also know as East African-Indian (Gagneux and Small, 2007)) and East African-Indian (EAI, also known as Indo-Oceanic (Gagneux and Small, 2007)). Others category included X (*n* = 434), *Mycobacterium africanum* (*n* = 56), *M. bovis* (*n* = 43). Data are presented only for the major lineages as many of the "other" were low IS*6110*-RFLP copy number strains.

| Spoligofamily | Number of isolates (%) | Clusters (n) | Cluster sizes |
|---|---|---|---|
| LAM* | 666 (17.3) | 90 | 2–16 |
| Haarlem | 626 (16.3) | 71 | 2–31 |
| T | 568 (14.8) | 65 | 2–8 |
| Beijing | 581 (15.1) | 63 | 2–18 |
| EAI | 304 (7.9) | 27 | 2–20 |
| CAS | 182 (4.7) | 11 | 2–3 |
| Others | 920 (23.9) | | |
| Total | 3847 (100) | | |

*LAM *M. tuberculosis* spoligofamily that includes 302 $RD^{Rio}$ case-isolates represented by 43 clusters (range: 2–16 case-isolates).

## Table 2
### Proportions of Clustered and Recently Transmitted Isolates by Prevalent Families

Cluster proportions of $RD^{Rio}$, related groups, and major families are shown. Clustered percentage was calculated by the total number of clustered isolates in each group divided by the total number of isolates for the group. The "n-1" cluster percentage was calculated by the total number of secondary cases in each group divided by the total number of isolates for the group. Pearson's chi-square was used to test the difference between clustered versus non-clustered cases and secondary verses all others cases. P-values were not adjusted for multiple testing. Ref denotes reference group.

| Groups | Clustered (%) | *P* | "n-1" Clustered (%) | *P* |
|---|---|---|---|---|
| $RD^{Rio}$ (n=302) | 154 (51.0) | Ref | 113 (37.4) | Ref |
| Non-$RD^{Rio}$ LAM (n=364) | 155 (42.6) | 0.035 | 107 (29.4) | 0.03 |
| Beijing (n=581) | 202 (34.8) | <0.0001 | 139 (23.9) | <0.0001 |
| Haarlem (n=626) | 292 (46.7) | 0.23 | 221 (35.3) | 0.56 |
| T (n=568) | 185 (32.6) | <0.0001 | 120 (21.1) | <0.0001 |
| EAI (n=304) | 96 (31.6) | <0.0001 | 69 (22.7) | <0.0001 |
| CAS (n=182) | 26 (14.3) | <0.0001 | 15 (8.2) | <0.0001 |

## Table 3

### Secondary Case Rates of RD$^{Rio}$ and Prevalent *M. tuberculosis* Families

**Secondary Case Rates** for RD$^{Rio}$ and major *M. tuberculosis* families are shown. 'Other LAM' refers to non-RD$^{Rio}$ LAM *M. tuberculosis*. Secondary case rates for each group is calculated as [secondary cases/(index cases + unique cases)].

| Family | Secondary Case / (Unique+Index Cases) | Secondary Case Rate | 95% CI* |
|---|---|---|---|
| RD$^{Rio}$ | 113/189 | 0.60 | (0.49, 0.70) |
| Other LAM | 107/257 | 0.42 | (0.35, 0.49) |
| Beijing | 139/442 | 0.31 | (0.27, 0.36) |
| Haarlem | 221/405 | 0.55 | (0.48, 0.62) |
| T | 120/448 | 0.27 | (0.23, 0.31) |
| EAI | 69/235 | 0.29 | (0.24, 0.36) |
| CAS | 15/167 | 0.09 | (0.05, 0.13) |

*CI = confidence interval, computed by Bootstrap method.

**Table 4**

Demographic and Clinical Characteristics of NYC TB Cases Harboring RD$^{Rio}$ and non-RD$^{Rio}$ *M. tuberculosis*: Associations by Univariate and Multivariate Analysis

| Variable | Group | RD$^{Rio}$ (%)* | Non-RD$^{Rio}$ | Odds Ratio (95% CI) | Adjusted Odds Ratio (95% CI) |
|---|---|---|---|---|---|
| Age | | | | 0.99 (0.985–1.0) * | 1 (0.99–1.01) |
| Gender | Female | 111 (36.8) | 1290 (36.4) | 1 (reference) | 1 |
| | Male | 191 (63.3) | 2255 (63.6) | 0.98 (0.77–1.26) | 1.00 (0.77–1.30) |
| Ethnicity | Hispanic | 171 (56.6) | 984 (27.8) | 5.42 (2.83–10.4) ** | 3.34 (1.60–6.99) ** |
| | Black | 107 (35.4) | 1138 (32.1) | 2.93 (1.52–5.68) ** | 2.33 (1.17–4.63) * |
| | Asian | 14 (4.6) | 1111 (31.3) | 0.39 (0.17–0.89) * | 1.69 (0.57–5.02) |
| | Other | 10 (3.3) | 312 (8.8) | 1 | 1 |
| Birthplace | LAC | 189 (62.6) | 1151 (32.5) | 2.91 (1.83–4.63) ** | 1.95 (1.11–3.43) * |
| | US | 82 (27.2) | 944 (26.6) | 1.54 (0.94–2.52) | 1.42 (0.83–2.41) |
| | Asia | 9 (3.0) | 1061 (30.0) | 0.15 (0.07–0.33) ** | 0.19 (0.06–0.60) ** |
| | Other | 21 (7.0) | 372 (10.5) | 1 | 1 |
| Residence | Manhattan | 65 (21.5) | 702 (19.8) | 1.52 (1.08–2.14) * | 1.49 (1.03–2.17) * |
| | Bronx | 72 (23.8) | 489 (13.8) | 2.42 (1.73–3.38) ** | 1.89 (1.31–2.71) ** |
| | Brooklyn | 87 (28.8) | 1074 (30.3) | 1.33 (0.97–1.82) | 1.27 (0.90–1.78) |
| | Other | 78 (25.8) | 1280 (36.1) | 1 | 1 |
| HIV Status | Positive | 50 (16.6) | 590 (16.6) | 0.89 (0.64–1.23) | 0.76 (0.52–1.11) |
| | Negative | 184 (60.9) | 1927 (54.4) | 1 | 1 |
| | Missing | 68 (22.5) | 1028 (29.0) | 0.69 (0.52–0.92) * | 1.01 (0.73–1.39) |
| Alcoholism | Yes | 55 (18.2) | 563 (15.9) | 1.19 (0.87–1.61) | 0.91 (0.64–1.29) |
| | No | 238 (78.8) | 2886 (81.4) | 1 | 1 |
| | Missing | 9 (3.0) | 96 (2.7) | 1.14 (0.57–2.28) | 1.79 (0.31–10.15) |
| Smoking | Yes | 21 (7.0) | 324 (9.1) | 0.79 (0.48–1.28) | 0.79 (0.46–1.35) |
| | No | 87 (28.8) | 1053 (29.7) | 1 | 1 |
| | Missing | 194 (64.2) | 2168 (61.2) | 1.08 (0.83–1.41) | 1.91 (0.87–4.22) |
| Homeless | Yes | 10 (3.3) | 103 (2.9) | 1.22 (0.62–2.41) | 1.17 (0.56–2.47) |

| Variable | Group | RD$^{Rio}$ (%)* | Non-RD$^{Rio}$ | Odds Ratio (95% CI) | Adjusted Odds Ratio (95% CI) |
|---|---|---|---|---|---|
| | No | 104 (34.4) | 1310 (37.0) | 1 | 1 |
| | Missing | 188 (62.3) | 2132 (60.1) | 1.11 (0.87–1.43) | 0.87 (0.39–1.95) |
| Drug Use | Yes | 30 (9.9) | 315 (8.9) | 1.13 (0.76–1.68) | 0.97 (0.61–1.54) |
| | No | 264 (87.4) | 3136 (88.5) | 1 | 1 |
| | Missing | 8 (2.7) | 94 (2.7) | 1.01 (0.49–2.10) | 0.54 (0.09–3.35) |
| Year | 2001 | 59 (19.5) | 804 (22.7) | 0.86 (0.59–1.26) | |
| | 2002 | 61 (20.2) | 692 (19.5) | 1.04 (0.71–1.51) | 1.20 (1.03–1.41)* |
| | 2003 | 76 (25.2) | 728 (20.5) | 1.23 (0.85–1.76) | |
| | 2004 | 51 (16.9) | 675 (19.0) | 0.89 (0.60–1.32) | |
| | 2005 | 55 (18.2) | 646 (18.2) | 1 | |
| **TB History** | Latent TB History | 23 (7.6) | 148 (4.2) | 1.89 (1.20–3.00)** | 1.41 (0.86–2.31) |
| | TB Hx | 5 (1.7) | 73 (2.1) | 0.83 (0.33–2.07) | 0.81 (0.31–2.07) |
| | Other | 274 (90.7) | 3324 (93.8) | 1 | 1 |
| **Tuberculin Skin Test result** | Positive | 187 (61.9) | 2167 (61.1) | 0.89 (0.64–1.23) | 0.96 (0.69–1.36) |
| | Negative | 51 (16.9) | 525 (14.8) | 1 | 1 |
| | Unknown | 64 (21.2) | 853 (24.1) | 0.77 (0.53–1.13) | 0.94 (0.63–1.42) |
| **Sputum Smear** | Positive | 153 (50.7) | 1725 (48.7) | 1.13 (0.86–1.47) | 0.94 (0.69–1.27) |
| | Negative | 97 (32.1) | 1231 (34.7) | 1 | 1 |
| | Unknown | 52 (17.2) | 589 (16.6) | 1.12 (0.79–1.59) | 1.03 (0.69–1.53) |
| **Disease Site** | Pulmonary | 208 (68.9) | 2452 (69.2) | 1.09 (0.73–1.62) | 1.13 (0.74–1.75) |
| | Extrapulm. | 60 (19.9) | 663 (18.7) | 1.16 (0.73–1.83) | 1.07 (0.62–1.84) |
| | Both | 30 (9.9) | 384 (10.8) | 1 | 1 |
| | Unknown | 4 (1.3) | 46 (1.3) | 1.11 (0.38–3.30) | 1.02 (0.33–3.11) |
| **Cavitary Dis** | Yes | 69 (22.9) | 637 (18.0) | 1.42 (1.06–1.91)* | 1.35 (0.98–1.87) |
| | No | 177 (58.6) | 2325 (65.6) | 1 | 1 |
| | Unknown | 56 (18.5) | 583 (16.5) | 1.26 (0.92–1.73) | 1.21 (0.81–1.82) |
| **Drug Resistance** | INH | 31 (10.3) | 261 (7.4) | 1.41 (0.95–2.09) | 1.65 (1.09–2.51)* |
| | MDR | 8 (2.7) | 107 (3.0) | 0.89 (0.43–1.84) | 0.99 (0.46–2.12) |
| | Other | 249 (82.5) | 2958 (83.4) | 1 | 1 |

| Variable | Group | $RD^{Rio}$ (%)* | Non-$RD^{Rio}$ | Odds Ratio (95% CI) | Adjusted Odds Ratio (95% CI) |
|---|---|---|---|---|---|
| | Unknown | 14 (4.6) | 219 (6.2) | 0.76 (0.44–1.32) | 0.71 (0.40–1.27) |

Of note, the robustness of the MVA model in determining associations with $RD^{Rio}$ *M. tuberculosis* is shown by the high area under the curve (AUC) of 0.74.

Percentages (%) refer to the column percentage of cases harboring $RD^{Rio}$ (or non-$RD^{Rio}$) *M. tuberculosis* with each variable.

CI denotes confidence interval.

* denotes $0.01 \leq p < 0.05$;

** denotes $p < 0.01$.

The results from the column of "odds ratio" were from the simple logistic regression that associated the individual factor and the outcome ($RD^{Rio}$ vs. non- $RD^{Rio}$), while the results from the column of "Adjusted odds ratio" were from the multiple logistic regression that associated all of the listed factors jointed and the outcome ($RD^{Rio}$ vs. non- $RD^{Rio}$).

Selected characteristics of TB associated with $RD^{Rio}$ and non- $RD^{Rio}$ *M. tuberculosis* defined by univariate (OR) and multivariate analysis (MVA, Adjusted OR) are shown. Year was included as a continuous variable in the adjusted model. Adjusting for all factors, TB caused by $RD^{Rio}$ strains also appeared to become more prevalent over the five year period (approximately 20% increase in odds ratio per 1 year increase with p=0.02).