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

Andrea L. Gibson, Richard C. Huard, Nicolaas C. Pittius, Luiz Claudio Oliveira Lazzarini, Jeffrey Driscoll, Natalia Kurepina, Thierry Zozio, Christophe Sola, Silvana Miranda Spindola, Afranio L. Kritski, et al.

JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 2008, p. 1259–1267 Vol. 46, No. 4
0095-1137/08/$08.00+0 doi:10.1128/JCM.02231-07

Application of Sensitive and Specific Molecular Methods To Uncover Global Dissemination of the Major RD^{Rio} Sublineage of the Latin American-Mediterranean *Mycobacterium tuberculosis* Spoligotype Family▽‡

Andrea L. Gibson,[1]† Richard C. Huard,[1,2]† Nicolaas C. Gey van Pittius,[3] Luiz Claudio Oliveira Lazzarini,[1,4] Jeffrey Driscoll,[5] Natalia Kurepina,[6] Thierry Zozio,[7] Christophe Sola,[7,8] Silvana Miranda Spindola,[9] Afrânio L. Kritski,[4] Daniel Fitzgerald,[1,10] Kristin Kremer,[11] Helmi Mardassi,[12] Poonam Chitale,[1] Jessica Brinkworth,[1,13] Dario Garcia de Viedma,[14,15] Brigitte Gicquel,[8] Jean W. Pape,[10] Dick van Soolingen,[11] Barry N. Kreiswirth,[6] Robin M. Warren,[3] Paul D. van Helden,[3] Nalin Rastogi,[7] Philip N. Suffys,[16] Jose Lapa e Silva,[4] and John L. Ho[1]*

Department of Medicine, Division of International Medicine and Infectious Diseases, Weill Medical College of Cornell University, New York, New York[1]; Clinical Microbiology Service and the Department of Pathology, Columbia University Medical Center, New York-Presbyterian Hospital, New York, New York[2]; DST/NRF Centre of Excellence in Biomedical Tuberculosis Research, US/MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Health Sciences, Stellenbosch University, Cape Town, South Africa[3]; Institute of Thoracic Diseases, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil[4]; Wadsworth Center, New York State Department of Health, Albany, New York[5]; Public Health Research Institute, Newark, New Jersey[6]; Unité de la Tuberculose et des Mycobactéries, Institut Pasteur de Guadeloupe, Guadeloupe[7]; Unité de Génétique Mycobactérienne, Institut Pasteur, Paris, France[8]; Departamento de Clínica Médica/Pneumologia, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil[9]; GHESKIO Centers, Port-au-Prince, Haiti[10]; National Institute of Public Health and the Environment, Bilthoven, The Netherlands[11]; Laboratoire des Mycobactéries, Institut Pasteur de Tunis, Tunisia[12]; Department of Anthropology, Graduate School and University Center, City University of New York/New York Consortium in Evolutionary Primatology, New York, New York[13]; Servicio de Microbiología Clínica y Enfermedades Infecciosas, Hospital Gregorio Marañón, Madrid, Spain[14]; CIBER Enfermedades Respiratorias, Madrid, Spain[15]; and Laboratory of Molecular Biology Applied to Mycobacteria, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil[16]

Received 19 November 2007/Returned for modification 27 December 2007/Accepted 9 January 2008

The Latin American-Mediterranean (LAM) family of *Mycobacterium tuberculosis* is believed to be the cause of ~15% of tuberculosis cases worldwide. Previously, we defined a prevalent sublineage of the LAM family in Brazil by a single characteristic genomic deletion designated RD^{Rio}. Using the Brazilian strains, we pinpoint an Ag85C^{103} single nucleotide polymorphism (SNP) (screened by restriction fragment length polymorphism [RFLP] analysis) that correctly identified all LAM family strains. Importantly, all RD^{Rio} strains concomitantly possessed the RD174 deletion. These genetic signatures, along with a newly developed multiplex PCR for rapid differentiation between "wild-type" and RD^{Rio} strains, were then used to analyze an international collection of *M. tuberculosis* strains. RD^{Rio} *M. tuberculosis* was identified from four continents involving 11 countries. Phylogenetic analysis of the IS*6110*-RFLP patterns from representative RD^{Rio} and LAM strains from Brazil, along with all representative clusters from a South African database, confirmed their genetic relatedness and transcontinental transmission. The Ag85C^{103} SNP RFLP, as compared to results obtained using a PCR method targeting a LAM-restricted IS*6110* element, correctly identified 99.8% of LAM spoligotype strains. Together, these tests were more accurate than spoligotyping at categorizing strains with indefinable spoligotypes and segregated true LAM strains from those with convergent spoligotypes. The fact that RD^{Rio} strains were identified worldwide highlights the importance of this LAM family sublineage and suggests that this strain is a global threat that should be specifically targeted by public health resources. Our provision of simple and robust molecular methods will assist the evaluation of the LAM family and the RD^{Rio} sublineage.

Mycobacterium tuberculosis is a successful obligate pathogen and the etiological agent of tuberculosis (TB). Worldwide, TB remains one of the leading causes of death from an infectious disease (9, 41).

Molecular typing, based on genetic markers, permits the rapid identification and species level identification of mycobacteria within the *M. tuberculosis* complex (MTC), as well as providing useful tools for examining the transmission and evolution of these microorganisms (6, 10, 15, 24, 33). One such technique, based on the amplification of several frequently observed deletion loci, can differentiate between the members of the MTC (20, 21). A study using this MTC PCR typing panel for the analysis of *M. tuberculosis* strains from Rio de Janeiro,

* Corresponding author. Mailing address: Cornell University, Joan and Sanford I. Weill Medical College, Department of Medicine, Division of International Medicine and Infectious Diseases, Rm. A-421, 525 East 68th St., New York, NY 10065. Phone: (212) 746-6320. Fax: (212) 746-8675. E-mail: jlho@med.cornell.edu.

† A.L.G. and R.C.H. contributed equally to this study.

‡ Supplemental material for this article may be found at http://jcm.asm.org/.

▽ Published ahead of print on 30 January 2008.

Brazil, found that 30% of isolates would not amplify from the IS*1561′* locus (25). These strains, designated RD^{Rio}, were found to contain a chromosomal deletion of more than 26 kb. Further analysis of RD^{Rio} strains with the molecular technique spoligotyping found that all RD^{Rio} strains belonged to the Latin American-Mediterranean (LAM) spoligotype family (25).

Spoligotyping is based on the analysis of the direct repeat locus, which is comprised of directly repeated sequences interspersed with nonrepetitive spacer DNA (23). This rapid PCR-based method allows the classification of strains into spoligotype families based on the presence or absence of spacer regions (31). The evolution of the direct repeat locus is believed to be unidirectional, i.e., spacers are lost rather than gained (37). Mechanisms thought to be responsible for this loss are the transposition of the insertion element IS*6110* and the deletion of single or stretches of contiguous direct variant repeats through homologous recombination between neighboring or distant direct repeats (37). The independent loss of similar spacer sets can lead to the convergent evolution of spoligotypes (40), since spoligotype families are defined predominantly by specific spacer patterns. This is therefore problematic since not all strains may truly represent their designated spoligotype family. Recently, other molecular markers have been utilized to define specific spoligotype families, such as the LAM-restricted IS*6110* element (26) and an SNP in *mgtC* which can differentiate between Haarlem and non-Haarlem strains (1).

Genotyping using single nucleotide polymorphisms (SNPs) has been applied to phylogenetically map the evolution of *M. tuberculosis* (2, 13, 17, 18, 34). SNPs frequently observed in the genes *katG* and *gyrA* permit the classification of MTC members into one of three principal genetic groups (PGG). The identification of SNPs at multiple loci within the *M. tuberculosis* genome provided enough variation to subdivide *M. tuberculosis* into several phylogenetic groups (2, 13, 17, 18). These phylogenetic groupings proved to be concordant with the three PGG, as well as IS*6110* restriction fragment length polymorphism (RFLP) data and spoligotyping families (3, 12, 38). The LAM family, to which RD^{Rio} strains belong, is found within PGG2 and in phylogenetic cluster VI according to Gutacker et al. (17, 18).

In the present study, we validated a multiplex PCR method for the classification of RD^{Rio} and "wild-type" non-RD^{Rio} (WT) strains using a collection of previously characterized clinical *M. tuberculosis* isolates from Rio de Janeiro, Brazil (25). This collection was then analyzed for the presence of an SNP previously identified at codon 103 of the gene encoding Ag85C (Rv0129c) (29). The $Ag85C^{103}$ SNP was found to be LAM lineage specific. A global collection of *M. tuberculosis* isolates was also examined to determine the worldwide distribution of RD^{Rio} and the $Ag85C^{103}$ SNP. Supplemental to this, these isolates were examined for the presence or absence of the LAM-restricted IS*6110* element as well as that of the region-of-difference (RD) loci within the *pks15/1* gene (27) and RD174, which are markers for, respectively, the purported major Euro-American lineage of *M. tuberculosis* defined by Gagneux et al. and its West African sublineage (15). Each of these phylogenetically informative markers was identified in *M. tuberculosis* strains from multiple worldwide locales linking them all to a common progenitor in the MTC evolutionary tree.

TABLE 1. Proportion of RD^{Rio} *M. tuberculosis* isolates observed in sample sets from each country[a]

Country[b]	No. of patient isolates	No. of strains (%)		WT/RD^{Rio} ratio (%)
		WT	RD^{Rio}	
Brazil (RJ)	314	218 (69)	93 (30)	3 (1)
South Africa	150	110 (73)	37 (25)	3 (2)
Brazil (MG)	105	64 (61)	39 (37)	2 (2)
The Netherlands	83	61 (73)	22 (27)	0
United States*	67	34 (51)	33 (49)	0
Tunisia	49	49 (100)	0	0
Guadeloupe	26	2 (8)	24 (92)	0
French Guiana	23	10 (43)	13 (57)	0
Haiti	17	13 (76)	4 (24)	0
Spain*	4	1 (25)	3 (75)	0
Chile*	2	2 (100)	0	0
Martinique	2	1 (50)	1 (50)	0
Canada*	1	1 (100)	0	0
Burundi*	1	1 (100)	0	0
Mongolia*	1	1 (100)	0	0
Tahiti*	1	1 (100)	0	0
Honduras*	1	1 (100)	0	0
Czech Republic	1	1 (100)	0	0
Russia	1	1 (100)	0	0
Ecuador*	1	0	1 (100)	0
Republic of Djibouti	1	0	1 (100)	0
Total	851	572 (67)	271 (32)	8 (1)

[a] The number of *M. tuberculosis* isolates is not a representative sample of each country or region. With the exception of Brazil and Haiti, the isolates were enriched for LAM spoligotypes.

[b] RJ, Rio de Janerio; MG, Minas Gerais. The numbers for the Rio de Janeiro *M. tuberculosis* isolates were as published previously (25). The data for three isolates from the United States, three isolates from The Netherlands, and one isolate from the Republic of Djibouti were also taken from previously published sources (21, 25). *, one isolate from each of these countries, except Spain and Chile, from which two isolates were obtained for analysis, was from Kremer et al. (24).

MATERIALS AND METHODS

Bacterial strains. A previously characterized cohort of *M. tuberculosis* isolates (25) was analyzed in the present study. Briefly, these isolates were randomly selected from a repository of mycobacterial cultures at the Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, between January 2002 and August 2003. Cell lysates of each isolate had previously been prepared as previously described (25). A patient previously identified as having a mixed infection in the Rio de Janeiro study was also included. From this patient isolate, single colonies were previously extracted, and two isolates from each genotype were subsequently analyzed. In addition, *M. tuberculosis* isolates (as DNA/cell lysates) were sent from 20 countries covering five continents (as detailed in Table 1) to the Cornell Weill Medical Center for analysis. These strains were provided as a nonrepresentative sample, with a bias toward LAM strains, since it was previously ascertained that RD^{Rio} *M. tuberculosis* strains were members of the LAM family (25) and the $Ag85C^{103}$ SNP appeared to be linked to LAM strains. It was therefore logical to select more strains belonging to this lineage to discern its sensitivity as a marker for the LAM family. All isolates and their corresponding molecular genotyping results are listed in Table S1 in the supplemental material. The following *M. tuberculosis* strains and MTC members were also analyzed to determine whether they contained the $Ag85C^{103}$ SNP (described below): *M. tuberculosis* H37Rv, *M. tuberculosis* CDC1551, *M. bovis* (ATCC 29210), *M. bovis* BCG Japan (ATCC 35737), *M. africanum* I (strain AF0271), and *M. microti* (strain 15496) (20, 21).

Multiplex PCR analysis. A rapid multiplex PCR was developed in the present study to distinguish between WT and RD^{Rio} strains. Previously described primers, used in single PCRs, were applied to develop the multiplex (20, 21, 25), and are detailed in Table S1 in the supplemental material. These primers either

target IS*1561′*, which is internal to the RD^{Rio} deleted region and so a marker for WT *M. tuberculosis* (530 bp), or flank the RD^{Rio} locus and bridge the deletion in RD^{Rio} strains (1,175 bp). Each PCR mix contained 25 μl of High Fidelity PCR Master Mix (Roche Diagnostics, Indianapolis, IN), 2.5 μl of dimethyl sulfoxide, 1 μl of each primer (20 μM), and 2.5 μl of bacterial lysate and was adjusted to a final volume of 50 μl with sterile distilled water. Amplification was performed in a Perkin-Elmer 9700 Thermocycler (PE Applied Biosystems, Foster City, CA) for 5 min at 95°C, followed by 45 cycles of 1 min at 95°C, 1 min at 60°C, and 4 min at 72°C, with a final extension cycle of 10 min at 72°C. When purified DNA samples were used, the number of cycles was reduced to 30. The amplified products were electrophoretically fractionated in a 1.5% agarose gel together with a 100-bp ladder (Invitrogen, Carlsbad, CA) and visualized by transillumination and ethidium bromide staining. The image was captured by using a Nighthawk Imaging System and Quality One Software (PDI, Inc., New York, NY).

Ag85C PCR analysis and restriction enzyme digest of Ag85C amplicons. A region of the Ag85C gene previously found to contain a silent SNP (GAG to GAA) at codon 103 (29) was targeted. Primers were designed by using the Lasergene program (DNASTAR, Inc., Madison, WI), and are detailed in Table S1 in the supplemental material. The PCR mix and cycle conditions used for the multiplex PCR were also used to amplify the Ag85C region. The amplified products (519 bp) were also visualized as described above. All Ag85C amplicons were digested by using the restriction enzyme MnlI (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions. The digested products were then separated on a 4% agarose gel containing ethidium bromide and visualized over UV light. Each gel also contained a 100-bp ladder to measure the size of the digested products. The presence of the SNP results in the loss of one of three restriction enzyme digestion sites and is visualized by the presence of a 461-bp band and a second band of 58 bp (which was not always detected using the gel method). An absence of the SNP is seen as three bands of 365, 96, and 58 bp (which also was not always visible on the gel). The protocol was therefore internally controlled for product digestion of the 519-bp amplicon.

Sequence analysis. DNA sequencing was performed by the Cornell University BioResource Center (Ithaca, NY [http://www.brc.cornell.edu]), using a BigDye terminator kit (PE Applied Biosystems) and analyzed by using an ABI 3700 DNA sequencer. The primers used for PCR amplification were also used for the sequencing reaction, with the exception of the reverse RD^{Rio} sequencing primer, where an internal reverse primer closer to the site of deletion was used (detailed in Table S1 in the supplemental material).

Analysis of *pks15/1* and RD174. A 7-bp deletion in the *pks15/1* gene (27) and RD174 are markers that have been previously reported (15). Detailed in Table S1 in the supplemental material are the primers used to amplify *pks15/1* region and the RD174 (forward flanking, internal forward [internal to RD174], and reverse flanking primers). The PCR conditions used to amplify both loci were the same as given above. Once the specific *pks15/1* site was amplified, sequence analysis (detailed above) was performed to determine whether the 7-bp deletion common to PGG2 and PGG3 strains was present. For RD174 analysis, the amplified products were analyzed on a 1.5% agarose gel. A 300-bp band indicated an intact locus, whereas a 500-bp band indicated the locus was deleted.

Analysis of the LAM-restricted IS*6110* element. PCR primers that identify the presence or absence of an IS*6110* insertion element unique to all LAM strains (26) were used to amplify all discrepant isolates (Table S1 in the supplemental material). Briefly, a 205-bp band indicates a LAM strain by the presence of an IS*6110* element, whereas a 141-bp band indicates a non-LAM strain lacking the IS*6110* element.

Statistical analysis. Statistical analysis was performed by using Prism 4 (GraphPad Software, Inc., San Diego, CA). Categorical variables were compared by a two-tailed Fisher exact test. A P value of ≤ 0.05 was considered significant.

RESULTS

RD^{Rio} multiplex PCR development. In our previous study, patient isolates were all characterized by using single PCRs containing primer pairs specific to either RD^{Rio} or WT strains (25). Since this is time- and resource-consuming, we developed a more rapid multiplex PCR protocol to differentiate RD^{Rio} and WT *M. tuberculosis*. A subset of isolates from our previous Rio de Janeiro study (20 WT strains, 20 RD^{Rio} strains and one mixed sample) was reevaluated to validate the protocol. In all cases, the expected amplicon size(s) was observed, i.e., 530 bp for WT strains, 1,175 bp for RD^{Rio} strains, and a double band for the mixed infection sample (Fig. 1A).

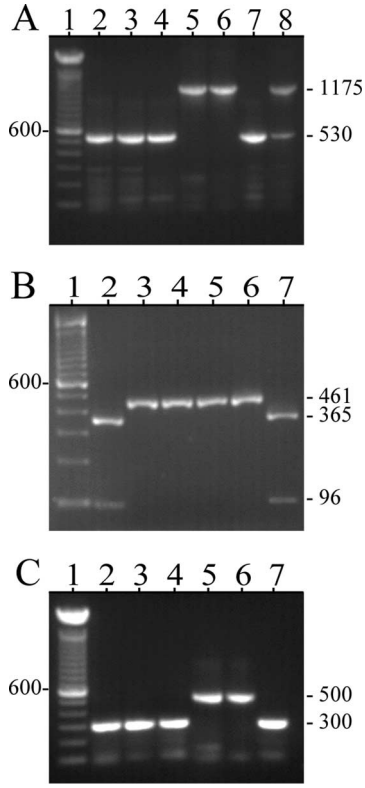

FIG. 1. (A) RD^{Rio} multiplex PCR. The presence of a 530-bp band indicates a WT strain, while the presence of an 1,175-bp band indicates an RD^{Rio} strain. Lanes: 1, 100-bp ladder; 2, *M. tuberculosis* H37Rv; 3 and 4, WT LAM strains; 5 and 6, RD^{Rio} LAM strains; 7, WT non-LAM strain (Haarlem 1); 8, a mixed (WT/RD^{Rio}) isolate. (B) Ag85C^{103} SNP digest analysis. The presence of a 461-bp band indicates that the SNP is present, while 365- and 96-bp bands indicate the SNP is absent. Lanes 1 to 7 are described as in panel A. (C) RD174 PCR analysis. A 300-bp band is indicative of an intact region, while a 500-bp band indicates that RD174 has been deleted. Lanes 1 to 7 are as described in panel B.

Identification and evaluation of the Ag85C^{103} SNP. The Rio de Janeiro collection was screened for the Ag85C^{103} SNP (29), using a restriction enzyme digest approach. A total of 306/310 patient isolates (90 RD^{Rio} *M. tuberculosis* and 216 WT patient isolates) and 1 isolate from a patient with a mixed infection, from which WT and RD^{Rio} colonies were previously extracted (four isolates in total), were available for analysis. Overall, 180 of 306 (58.8%) patient isolates contained the Ag85C^{103} SNP. A total of 20 isolates were selected, 10 for each digestion pattern, and sequenced to confirm for mutant and WT isolates. The sequencing results were concordant with the Ag85C^{103} digest results (data not shown). Figure 1B illustrates examples of the Ag85C^{103} restriction digest patterns. Overall, all RD^{Rio} isolates analyzed contained the SNP (n = 90), whereas less than half of the WT (43.5%) strains did. From the patient with the mixed infection, the RD^{Rio} clones both harbored the polymorphism, while the two WT isolates did not. When the spoligotyping profiles of the Rio de Janeiro strains were compared (n = 293), all LAM strains (n = 161) were found to contain

TABLE 2. Division of LAM and non-LAM isolates (designated by spoligotyping) by the presence or absence of Ag85C[103] according to country of origin

Country of origin (no. of isolates)[a]	No. of LAM isolates (%)		No. of non-LAM isolates (%)	
	SNP present	SNP absent	SNP present	SNP absent
Brazil (RJ) (293)	161 (55)	0	10 (3)	122 (42)
South Africa (145)	101 (70)	0	17 (12)	27 (18)
Brazil (MG) (96)	62 (65)	0	1 (1)	33 (34)
The Netherlands (73)	68 (93)	4 (6)	0	1 (1)
United States (63)	50 (79)	2 (3)	0	11 (18)
Tunisia (47)	44 (94)	0	0	3 (6)
Guadeloupe (21)	19 (90)	0	0	2 (10)
French Guiana (20)	20 (100)	0	0	0
Haiti (15)	3 (20)	1 (7)	0	11 (73)
Spain (3)	3 (100)	0	0	0
Chile (2)	2 (100)	0	0	0
Martinique (2)	2 (100)	0	0	0
Ecuador (1)	1 (100)	0	0	0
Mongolia (1)	1 (100)	0	0	0
Tahiti (1)	1 (100)	0	0	0
Honduras (1)	1 (100)	0	0	0
Republic of Djibouti	1 (100)	0	0	0
Czech Republic (1)	0	0	0	1 (100)
Russia (1)	0	0	0	1 (100)
Total (787)	540 (68)	7 (1)	28 (4)	212 (27)

[a] The numbers in parentheses differ from those in Table 1 because the spoligotype pattern was not available, the spoligotype pattern was not definable (see Table 3), or the isolate appeared to be a mixed sample according to at least one of the molecular tests and thus the spoligotype was not reliable. The mixed isolate from Rio de Janeiro that was separated into single colonies was included in the analysis.

Ag85C[103], whereas the majority of WT non-LAM strains (92.4%) did not ($P < 0.0001$; two-tailed Fisher exact test, Table 2). A total of 10 non-LAM isolates, as defined by spoligotyping, contained the Ag85C[103] SNP. These isolates were sequenced, and the results were found to be concordant with the Ag85C[103] digest results (data not shown). These discrepant strains were later analyzed by using the LAM IS*6110*-specific PCR assay, the results of which are detailed below.

RD174 and *pks15/1* 7-bp deletion PCR Analysis. Gagneux et al. (15) constructed a phylogenetic tree based on the presence or absence of common long sequence polymorphisms (LSPs) found in *M. tuberculosis* strains and used these RD loci to construct an evolutionary tree. To determine whether RD174 was a marker for LAM and/or RD[Rio] strains, we evaluated the Rio de Janeiro strain collection for this LSP. Remarkably, RD174 was deleted in all RD[Rio] strains but remained intact in all WT strains, including WT LAM strains (Fig. 1C). This suggests that RD174 is in fact a parallel marker for the RD[Rio] genotype. The RD174 polymorphism is believed to be at the node of a sub-branch of the Euro-American lineage, defined by the *pks15/1* 7-bp deletion. Therefore, we tested a selection of strains from the Rio de Janeiro collection to confirm that isolates harboring the RD174 polymorphism did belong to this lineage and also to ascertain whether the Euro-American lineage contained both LAM and non-LAM strains. A total of 10 RD[Rio] isolates, 15 WT LAM isolates, and 13 WT non-LAM isolates, representing ca. 10% of each group from Rio de Janeiro, were analyzed. All LAM strains including RD[Rio] isolates (PGG2) were found to possess the 7-bp *pks15/1* deletion, as did the non-LAM strains belonging to PGG2 and PGG3 (data stated herein). The only strains not containing the *pks15/1* deletion were two W/Beijing strains and an EAI strain, both of which are PGG1 isolates, confirming that the *pks15/1* 7-bp deletion is PGG2 and PGG3 restricted (21).

RD[Rio] genotype evaluation of global LAM strains. An international collection of *M. tuberculosis* strains (total of 851 isolates [see Materials and Methods]) was analyzed by multiplex PCR to determine whether RD[Rio] *M. tuberculosis* has disseminated outside of Rio de Janeiro. Overall, the RD[Rio] genotype was identified in isolates from 11 of the 20 countries represented, showing that RD[Rio] strains are circulating internationally and are not restricted to Brazil. Table 1 shows the classification of isolates by genotype for each geographical area. As well as identifying RD[Rio] *M. tuberculosis* in a second state of Brazil (Minas Gerais, where 36% of the provided strains were RD[Rio]), RD[Rio] was also identified in Ecuador and French Guiana, both South American countries, and in North America (i.e., the United States). Moreover, RD[Rio] *M. tuberculosis* was observed in all three Caribbean islands studied (Guadeloupe, Haiti, and Martinique). Of great importance was the fact that RD[Rio] strains were also identified outside of the Americas, in Europe (The Netherlands and Spain) and in Africa (South Africa and the Republic of Djibouti). These results clearly demonstrate that RD[Rio] *M. tuberculosis* has indeed disseminated globally. To provide additional support that the RD[Rio] and LAM strains from Brazil were related to those found elsewhere, we performed IS*6110* RFLP phylogenetic analysis of isolates from the Rio de Janeiro collection previously studied by IS*6110* RFLP together with one isolate from each cluster (strain/RFLP signature) from a South African database (data not shown). The genetic relatedness of LAM strains and specifically RD[Rio] strains from Rio de Janeiro with their counterparts from South Africa were confirmed by IS*6110* RFLP phylogenetic analysis (see Fig. S1 in the supplemental material).

Ag85C[103] SNP correlates with LAM-restricted IS*6110* insertion PCR in analysis of global LAM strains. The *M. tuberculosis* isolates from each country were examined for the Ag85C[103] SNP, and the results were compared to their spoligotypes (Table 2; the spoligopatterns and spoligotypes designation assigned by SpolDB4 are presented in Table S2 in the supplemental material). In almost all cases, LAM strains contained the Ag85C[103] SNP (540 of 547 [98.7%]), while the majority of non-LAM strains did not contain this SNP (212 of 240 [88.3%]) ($P \leq 0.0001$; two-tailed Fisher exact test). As previously observed in the Rio de Janeiro cohort, all RD[Rio] strains contained the SNP. At first, this suggested that the Ag85C[103] SNP was not as sensitive or specific as previously seen in the Rio de Janeiro cohort. Overall, 7 LAM strains without the Ag85C[103] SNP and 28 non-LAM strains with the SNP were observed. Possible explanations for these two scenarios may be the convergent evolution of spoligotypes and/or the presence of SNP-containing ancestral progenitor strains. To further explore these discrepant strains, we analyzed them for a site-specific IS*6110*-element that is unique to LAM strains (26). Interestingly, 27 of 28 non-LAM strains (according to spoligotyping) with the Ag85C[103] SNP were found to be LAM when analyzed with the LAM-specific IS*6110* PCR, in-

TABLE 3. Indistinguishable spoligotype patterns[a]

Spoligotype pattern	No. of isolates	SpolDB4 designation[b]	No. of isolates with the Ag85C[103] SNP:		No. of LAM or non-LAM isolates[c]
			Present	Absent	
■■■■■■■■■■■■■■■■■■■■■■□□□□□□□□□□□□□□□□□□□□□□□□□	2	U	2	0	2 LAM
■■■■■■■■■■□■■■■■■■■■■□□□□□□□□□□□□□□□□□□□□□□□□□	1	NA[d]	1	0	1 LAM
■■■■■■■■■□□■■■■■■■■■■□□□□□□□□□□□□□□□□□□□□□□□□□	1	NA	1	0	1 LAM
■■■■■■■■□□□■■■■■■■■■□□□□□□□□□□□□□□□□□□□□■■■■■	4	U (LAM3)?	3	1[e]	3 LAM, 1 non-LAM[e]
■■■■■■■■□□□■■■□□□■□□□□□□□□□□□□□□□□□□□□□■■■■■	1	U	1	0	1 LAM
■■■■□□□□□□□□□□□□□□□□□□□□□■■■■■■■■□□□□■■■□■■■	7	T2 (LAM type?)	7	0	7 LAM
□□□□■■■■■■■□■■■■■■□□□□□□□□□□□□□□□□□□□□■■■■■■	1	U	0	1	1 Non-LAM
□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■□□□□■■■■■■■	15	LAM3/S convergent	5	10[f]	5 LAM, 10 non-LAM[f]
□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■□□□□■■■□■■■	4	T2	0	4	3 non-LAM
■□■□□□□■■■■■■■□□■■■■■■□□□□□□□□□□□□□□□□□□■■■■■■	1	NA	0	1	1 Non-LAM
■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■[g]					

[a] That is, spoligotype patterns that could not be segregated into LAM or non-LAM strains.
[b] That is, the definition given by the international spoligotyping database (SpolDB4; Brudey et al. [7]), which tentatively classifies spoligotypes into clades.
[c] Defined by LAM-specific IS*6110* PCR.
[d] NA, not available.
[e] The same isolate.
[f] The same isolates.
[g] The final spoligotype is the LAM9 reference spoligotype.

dicating that these strains are of the LAM lineage. Of the seven LAM strains (according to spoligotyping) without the Ag85C[103] SNP, six strains were PCR amplified as non-LAM, indicating that the spoligotypes of these strains were the result of convergence. Therefore, this left only two discrepant strains, a LAM strain without the Ag85C[103] SNP and a non-LAM strain with the SNP, resulting in a sensitivity and specificity rate for the Ag85C[103] SNP of 99.8 and 99.6%, respectively.

From the international collection, an additional 37 *M. tuberculosis* isolates had spoligotypes that could not be segregated into LAM or non-LAM groups (Table 3). Although these strains had spacers 21 to 24 and spacers 33 to 36 deleted, some of the surrounding spacers were also missing, thus preventing accurate clade designation. To confirm whether these isolates belonged to the LAM lineage, all were analyzed for the LAM-specific IS*6110* PCR and segregated into LAM/non-LAM groups according to the presence or absence of this element (Table 3). The Ag85C[103] SNP results were 100% concordant to the IS*6110* PCR results. Importantly, this shows that the Ag85C[103] SNP is an accurate marker for the differentiation of LAM/non-LAM strains and particularly useful when a spoligotype pattern is indistinguishable.

In addition to the international *M. tuberculosis* strain collection, several other MTC isolates were examined for the Ag85C[103] SNP. It was found to be absent from *M. tuberculosis* H37Rv and *M. tuberculosis* CDC1551, as well as representative strains from other MTC members, namely, *M. africanum* I, *M. microti*, *M. bovis*, and *M. bovis* BCG (Japan).

RD^{Rio} deletion correlates with RD174. The international strain collection was also analyzed for RD174 by PCR to confirm that this polymorphism was indeed a marker for RD^{Rio} strains. All RD^{Rio} strains had RD174 deleted, and all WT strains had this region intact (n = 848). Of note, the data indicated that four patients had mixed strain *M. tuberculosis* infections that included an RD^{Rio} *M. tuberculosis* strain, as has been noted previously (25). One isolate from Minas Gerais and one from South Africa that were found to be mixed by the multiplex PCR and also had a double band corresponding to both an intact and a deleted RD174 region. The remaining two mixed isolates (according to the multiplex PCR), one from South Africa and one from Minas Gerais, were not mixed by RD174, suggesting that the multiplex RD^{Rio} PCR may be more sensitive at identifying mixed infections that include these strains than the use of RD174. Since only one sample from each of the above mixed isolates was evaluated, the possibility of laboratory cross-contamination could not be excluded.

DISCUSSION

This study aimed to identify specific markers able to define the LAM family lineage and one of its major sublineages, RD^{Rio}, and to develop a multiplex PCR to rapidly identify RD^{Rio} strains. The newly developed multiplex PCR was then utilized to reveal that the RD^{Rio} sublineage was not just a problem localized to Rio de Janeiro in Brazil, where it was first identified, but that it is found globally.

We identified a synonymous SNP in the gene for Ag85C that was linked to LAM strains. Initially, the Ag85C[103] SNP did not appear to be a particularly accurate marker for LAM strains compared to spoligotyping. Initially, the sensitivity of this marker by comparison to spoligotyping was 98.7%, while the specificity was 88.3% in our global collection. However, further analysis of the discrepant isolates based on the presence or absence of an IS*6110* element unique to LAM strains proved that the sole use of spoligotype patterns are not altogether reliable. As previously observed, the loss of these spacers can lead to the convergent evolution of strains (40), since spoligotype families are mainly defined by specific spacer patterns. We observed this phenomenon in seven strains that had "LAM" spoligotypes but did not contain Ag85C[103]. All but one strain lacked the LAM-specific IS*6110* element, proving that these strains were not true LAM strains. Conversely, the 28 "non-LAM" discrepant strains containing Ag85C[103] could not be explained by convergent evolution, considering that spacers are lost rather than gained. IS*6110* PCR analysis found that

these apparent "non-LAM" strains were in fact true LAM lineage strains since 27 of 28 possessed the LAM-specific IS*6110* element. Overall, the sensitivity and specificity of the Ag85C^{103} SNP were 99.8 and 99.6%, respectively. Although an isolate with a LAM spoligotype but without Ag85C^{103} could be considered as an example of convergent evolution, i.e., not a true LAM strain, the single discrepant isolate was confirmed as a LAM strain by the presence of the LAM-specific IS*6110* element. Conceptually, the insertion of this IS*6110* element and the Ag85C^{103} mutation are unlikely to have occurred simultaneously, and so it is reasonable to conclude that one polymorphism predates the other. In this case one could argue that the IS*6110* insertion occurred first; however, one Ag85C^{103}-positive non-LAM strain (confirmed by IS*6110* PCR) was observed, which would argue the converse. We cannot at present be certain in what order these polymorphisms arose, but hopefully further genetic analyses of these strains will settle the issue. Interestingly, to date there has been scant evidence of horizontal genetic exchange in *M. tuberculosis*, although data support that it has occurred in the MTC progenitor species "*Mycobacterium canettii*" (prototuberculosis) (5, 36). Given that mixed strain infections appear to be more common than previously assumed, one or the other of the above discordant stains may well represent a rare example of just such a genomic homogenizing event in *M. tuberculosis*.

In some instances the spoligotype pattern was indefinable and could not be designated as a LAM or a non-LAM strain. Both the Ag85C^{103} SNP and the LAM-specific IS*6110* PCR were able to differentiate these isolates into LAM and non-LAM strains, proving that spoligotyping alone is not always the best method for defining LAM strains. As we have demonstrated, Ag85C^{103} SNP analysis can also define LAM family strains that are misclassified by spoligotyping (i.e., have a non-LAM spoligotype pattern). Ag85C^{103} SNP analysis is not a replacement for spoligotyping, since the latter can both define and differentiate between LAM family strains. However, it is relatively cheap compared to spoligotyping and does not require any specialized equipment. The Ag85C^{103} SNP PCR assay could therefore serve as a handy postspoligotyping confirmatory tool in epidemiological studies and/or in resource poor settings as a screening method for LAM strains to limit the number of isolates requiring full characterization with spoligotyping.

Specific SNPs have been linked to each of the MTC species and various tubercle bacilli, as well as many of the *M. tuberculosis* lineages (1, 2, 13, 16, 21, 34, 39), thereby providing consistent phylogenetic markers and allowing inter- and intraspecies segregation with relative ease. For example, the "*M. canettii*" *hsp65*^{631} SNP differentiates it from the rest of the MTC (21). Within *M. tuberculosis* lineages, a single SNP in *mgtC* was found to differentiate between Haarlem and non-Haarlem strains of *M. tuberculosis* (1), while an SNP in *rrs*^{491} was specifically associated with the F11 (LAM3) clade, a dominant *M. tuberculosis* lineage in the Western Cape of South Africa (39). Genome SNP analysis has previously been used to classify *M. tuberculosis* into phylogenetic lineages (2, 13, 17). One such tree divided *M. tuberculosis* into nine distinct lineages; LAM strains and a subset of T strains segregated into SNP cluster VI (17). The Ag85C^{103} SNP was not specific to all strains in cluster VI since several "T" strains from our USA collection that were known to belong to cluster VI did not contain the Ag85C^{103} SNP (data not shown). Nonetheless, the identification of a single SNP that is specifically associated with the LAM family lineage (LAM strains comprise ~80% of cluster VI isolates) could potentially be utilized as an indicator to identify cluster VI strains. As such, this test would reduce the amount of SNP analysis required to characterize this major phylogenetic lineage. In addition, since we provide a restriction digest approach for the Ag85C^{103} SNP, the need for more costly DNA sequence analysis would be eliminated.

LSPs have also been used to define both MTC members and *M. tuberculosis* lineages (6, 15, 19, 21). Gagneux et al. (15) developed a global phylogenetic tree based on several LSPs, identified through whole-genome screening, and found that geographical regions are associated with certain lineages. Using markers from this tree we linked RD^{Rio} *M. tuberculosis* to the West African sublineage of the major Euro-American lineage (15). Our study demonstrated that RD174 was specific for RD^{Rio} *M. tuberculosis* and not observed in any other LAM strains. In addition, the major branch of the Euro-American lineage is defined by a 7-bp deletion in a polyketide synthase gene (*pks15/1*). Interestingly, the *pks15/1* deletion was previously found to be synonymous with the SNP at *katG*^{463} CGG; therefore, the Euro-American lineage consists of PGG2 and PGG3 *M. tuberculosis* strains (21, 27) and not PGG1/ancestral strains. Our data concurred with this observation.

Surprisingly, there are a limited number of studies focusing solely on LAM strains. Many studies have investigated W/Beijing strains predominantly since these strains are prevalent across Asia and also contributed greatly to the TB resurgence in the United States in the early 1990s (4, 42). In comparison to LAM strains, which are responsible for ca. 15% of the world's TB cases (SpolDB4 [7]), W/Beijing strains account for just ~10% of *M. tuberculosis* strains worldwide. Therefore, there is a need for more research efforts geared toward the LAM family. The development of a multiplex PCR enabled the rapid segregation of *M. tuberculosis* strains into WT and RD^{Rio} genotypes. We found that TB caused by RD^{Rio} *M. tuberculosis* is not restricted to Brazil or even the Americas but is found in four of the five continents examined. Although RD^{Rio} was not identified in Asia, only two isolates in the present study were examined from this continent, and so further studies are needed to assess whether RD^{Rio} *M. tuberculosis* is circulating in Asia as well. Studies have shown that W/Beijing and LAM strains are the major cause of TB in Russian prisons (11, 22, 32). Of the LAM strains, both LAM9 and LAM1 strains were major contributors. Our previous study found that these two LAM types were associated with the RD^{Rio} genotype (25) and that LAM1 strains were exclusively found to be RD^{Rio} *M. tuberculosis*. Although we did identify one WT LAM1 strain and two mixed infections with LAM1 spoligotypes, 48 of 51 (94.1%) of LAM1 strains in this study were nonetheless RD^{Rio} *M. tuberculosis*. These data thus indirectly suggest that RD^{Rio} may well be a significant cause of TB in Northern Asia, as well by virtue of the LAM1 prevalence in this region. This hypothesis, however, remains to be substantiated by evaluating Russian LAM strains.

Interestingly, two of the RD^{Rio} *M. tuberculosis* strains isolated in Madrid, Spain, were from the most prevalent cluster of *M. tuberculosis* reported in that city, namely, strain 5 (28).

Strain 5 (ST20 or LAM1) has an identical mycobacterial interspersed repetitive unit (MIRU) pattern to a hypothetical progenitor MIRU-type for RD^{Rio} strains and a very similar RFLP pattern (25). These Spanish RD^{Rio} strains were isolated in 2002 and 2003, but strain 5 has remained the predominant strain in Madrid for the last 13 years, whereas other others strain types have declined. This suggests that RD^{Rio} *M. tuberculosis* was circulating in Madrid in 1993 and possibly earlier. As previously hypothesized (25), the persistence of the RD^{Rio} sublineage may be attributed to either intrinsic genetic factors affording biological advantage for the tubercle bacilli or risk factors of the host population that sustain dissemination.

Taking into account that RD^{Rio} has disseminated outside of the Americas and the fact that it is part of the LAM family, RD^{Rio} strains could indeed be a significant global problem. Complementary to the presented genetic data, IS*6110* RFLP phylogenetic analysis of isolates from the Rio de Janeiro collection (25), along with one isolate from each cluster (strain/RFLP signature) from a South African database (data not shown), confirmed genetic relatedness; their all of the LAM strains from Rio de Janeiro fall within one of the RFLP families previously described in South Africa (30) (see Fig. S1 in the supplemental material). Moreover, the RD^{Rio} strains from Rio de Janeiro grouped with the South African F9 and F13 families (LAM11 and LAM1, respectively) (35), which in the present study were shown to be of the RD^{Rio} genotype. Recently, Chihota et al. (8) identified a group of strains (named the Southern Africa 1 or SAF1 family) that made up a predominant proportion of strains isolated in Zimbabwe (47.2%) and Zambia (65%) and which were shown to fall into the South African F9 family (and thus the RD^{Rio} lineage). These findings further support the present genetic data and highlight the fact that RD^{Rio} indeed appears to be a significant cause of TB worldwide. Our rapid PCR technique may be useful in the epidemiological tracking of the RD^{Rio} genotype and in ascertaining the level of RD^{Rio} *M. tuberculosis* circulating in a population. It is of great importance, when developing public health control strategies for TB, to understand the strains circulating not only on a regional scale but globally as well.

Recently, Flores et al. (14) assigned strains to the W/Beijing lineage based upon possession of the prototypic W/Beijing RD105 deletion, despite the absence of typical W/Beijing family spoligotype patterns. Likewise, in the present study, we identified several genetic signatures of the LAM family and its major RD^{Rio} sublineage that allow us to propose to extend the genetic definition for a LAM lineage strain (as distinguished from the defining characteristics of a LAM spoligotype family strain). We classify the LAM lineage by (i) being PGG2, as well as possessing (ii) the 7-bp *pks15/1* deletion, (iii) the SNP at $Ag85C^{103}$, and (iv) the restricted site-specific IS*6110* insertion element. On the other hand, LAM spoligotype family strains (within the LAM lineage, as defined by the above markers) exhibit prototypic spoligotype patterns with an absence of spacers 21 to 24 and spacers 33 to 36 (7). LAM RD^{Rio} *M. tuberculosis* can be further characterized by the RD^{Rio} and RD174 deletions. Now that we have highlighted the global problem of the LAM family and its prominent RD^{Rio} sublineage and developed a series of simple genetic signatures for their characterization, we hope that research interests will focus on this highly prevalent *M. tuberculosis* lineage and provide much-needed information regarding the bases of their success.

In conclusion, our finding that RD^{Rio} strains were identified worldwide highlights the importance of this LAM family sublineage. Limited IS*6110* RFLP analysis confirms the genetic relatedness of LAM and RD^{Rio} strains and cross-transmission between continents. The findings from this report furthermore demonstrate that the RD^{Rio} strain is a global threat and should be specifically targeted by public health resources. Our provision in this report of simple and robust molecular methods should assist in the evaluation of the LAM family and the RD^{Rio} sublineage.

ACKNOWLEDGMENTS

We thank Maryse Timothee for the microbiological processing of the *M. tuberculosis* strains from Haiti and Scott Weisenberg, Hongxia Zhu, and Warren D. Johnson for their suggestions, support, and encouragement.

Funding for this project was provided by National Institutes of Health (NIH) grants R21 AI063147 and R21 AI063147 (to J.L.H.), NIH Fogarty International Center training grant (FICTG) (D43 TW00018), Innovative Approaches for TB Control in Brazil (Rio-ICOHRTA)-U2R TW006885 NIH, FIC (J.R.L.E.S., R. Chaisson, and J.L.H.), Haiti AIDS Research Training: Models to Implementation (Haiti-ICOHRTA)-U2R TW006901 (W. D. Johnson), a grant from the Coordenação de Aperfeicoamento de Pessoal de Nivel Superior (Ministry of Education of Brazil) from the Brazilian Ministry of Health (024/94 DST/AIDS), Fundação Universitária José Bonifácio/FUJB, the Brazilian Research Council/CNPq, the Brazilian Research Council/World Bank Millennium Institute of Science, the Programa de Apoio a Núcleos de Excelência, and PAPES IV/FIOCRUZ (P.N.S. and J.R.L.E.S.), and a grant from the Laura Cook Hull Trust Fund (LCHTF) (W. D. Johnson). A.L.G. was supported in part by the LCHTF, and L.C.O.L was a FICTG and is now a Rio-ICOHRTA and a CNPq trainee. The study performed in N.R.'s laboratory was partially funded by the European Regional Development Fund, European Commission (ERDF/FEDER, A34-05). T.Z. received a Ph.D. fellowship awarded by the European Union and the Regional Council of Guadeloupe and the International Network of Pasteur Institutes. Funding for the Spanish isolates was from the Fondo de Investigaciones Sanitarias (FIS030654; FIS060882) and the Instituto de Salud Carlos III (CIBER Enfermedades Respiratorias CIBERES; CB06/06/0058).

A.L.G. performed the majority of the laboratory work, codeveloped the restriction digest of Ag85C SNP analysis (with R.C.H.), posed and tested the hypotheses (with J.L.H. and R.C.H.), and wrote the manuscript. R.C.H. initiated the multiplex assays and Ag85C SNP evaluation and, together with J.L.H., gathered the international *M. tuberculosis* collection, as well as guided the presentation, provided input during preparation, and edited the manuscript. N.C.G.V.P. made the suggestion to cross-compare LAM-specific IS*6110* PCR with the Ag85C SNP, provided critical input in the evaluation of strains with certain spoligopatterns, and performed the phylogenetic analysis of the IS*6110* RFLP pattern for the present study. N.C.G.V.P., together with R.M.W. and P.D.V.H., provided the samples with spoligopattern and RFLP data from South Africa. L.C.O.L. assisted in some of the earlier genetic analysis, assisted in generating more DNA lysates from Rio de Janeiro needed for the present study, coordinated shipment of these samples to Cornell, provided input in some of the data interpretation, and edited the manuscript for reference order and format. J.D. performed spoligotyping on samples without data and repeated the spoligotyping on select strains. N.K. contributed to the phylogenetic analysis of the IS*6110* RFLP pattern for this study and, together with B.N.K., provided strains with spoligotyped data from North America. T.Z., C.S., and N.R. provided samples with spoligotyped data from French Caribbean countries, and C.S. provided suggestions to improve the manuscript. S.M.S. provided the samples with spoligotyped data from Belo Horizonte, Brazil, which was performed by S.M.S. in France in the laboratory of B.G. A.L.K. assisted S.M.S. in organizing the data file of the Belo Horizonte collection. D.F. and J.W.P. provided samples

from Haiti. K.K. and D.V.S. provided samples with spoligotyped data from The Netherlands. H.M. provided samples with spoligotyped data from Tunisia. P.C. and J.B. assisted in some of the genotyping. D.G.D.V. provided samples from Spain and perspective on the prevalent strains in the epidemiology of TB in Spain. P.N.S. supervised the spoligotyping of the Rio de Janeiro samples. J.R.L.S., together with J.L.H., supervised the laboratory and clinical studies in Rio de Janeiro and shipment of the DNA lysates to United States. J.L.H. posed the hypotheses, coordinated the international collaborations, provided the overall management of the project, guided data analyses and interpretation, provided input for presentation and discussions during manuscript preparation. and edited the manuscript.

REFERENCES

1. **Alix, E., S. Godreuil, and A. B. Blanc-Potard.** 2006. Identification of a Haarlem genotype-specific single nucleotide polymorphism in the *mgtC* virulence gene of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **44:**2093–2098.
2. **Baker, L., T. Brown, M. C. Maiden, and F. Drobniewski.** 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg. Infect. Dis. **10:**1568–1577.
3. **Bhanu, N. V., D. van Soolingen, J. D. van Embden, L. Dar, R. M. Pandey, and P. Seth.** 2002. Predominace of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. Tuberculosis **82:**105–112.
4. **Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth.** 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. **10:**45–52.
5. **Brisse, S., P. Supply, R. Brosch, V. Vincent, and M. C. Gutierrez.** 2006. "A re-evaluation of M. prototuberculosis": continuing the debate. PLoS Pathog. **2:**e95.
6. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **99:**3684–3689.
7. **Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Prodinger, A. Gori, S. A. Al-Hajoj, C. Allix, L. Aristimuno, J. Arora, V. Baumanis, L. Binder, P. Cafrune, A. Cataldi, S. Cheong, R. Diel, C. Ellermeier, J. T. Evans, M. Fauville-Dufaux, S. Ferdinand, D. Garcia de Viedma, C. Garzelli, L. Gazzola, H. M. Gomes, M. C. Guttierez, P. M. Hawkey, P. D. van Helden, G. V. Kadival, B. N. Kreiswirth, K. Kremer, M. Kubin, S. P. Kulkarni, B. Liens, T. Lillebaek, M. L. Ho, C. Martin, C. Martin, I. Mokrousov, O. Narvskaia, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, Z. Rahim, V. Rasolofo-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rusch-Gerdes, A. Sajduda, S. Samper, I. G. Shemyakin, U. B. Singh, A. Somoskovi, R. A. Skuce, D. van Soolingen, E. M. Streicher, P. N. Suffys, E. Tortoli, T. Tracevska, V. Vincent, T. C. Victor, R. M. Warren, S. F. Yap, K. Zaman, F. Portaels, N. Rastogi, and C. Sola.** 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. **6:**23.
8. **Chihota, V., L. Apers, S. Mungofa, W. Kasongo, I. M. Nyoni, R. Tembwe, G. Mbulo, M. Tembo, E. M. Streicher, G. D. van der Spuy, T. C. Victor, P. van Helden, and R. M. Warren.** 2007. Predominance of a single genotype of *Mycobacterium tuberculosis* in regions of Southern Africa. Int. J. Tuberc. Lung Dis. **11:**311–318.
9. **Corbett, E. L., C. J. Watt, N. Walker, D. Maher, B. G. Williams, M. C. Raviglione, and C. Dye.** 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch. Intern. Med. **163:**1009–1021.
10. **Dale, J. W., H. Al-Ghusein, S. Al-Hashmi, P. Butcher, A. L. Dickens, F. Drobniewski, K. J. Forbes, S. H. Gillespie, D. Lamprecht, T. D. McHugh, R. Pitman, N. Rastogi, A. T. Smith, C. Sola, and H. Yesilkaya.** 2003. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS*6110*. J. Bacteriol. **185:**2555–2562.
11. **Drobniewski, F., Y. Balabanova, M. Ruddy, L. Weldon, K. Jeltkova, T. Brown, N. Malomanova, E. Elizarova, A. Melentyev, E. Mutovkin, S. Zhakharova, and I. Fedorin.** 2002. Rifampin- and multidrug-resistant tuberculosis in Russian civilians and prison inmates: dominance of the Beijing strain family. Emerg. Infect. Dis. **8:**1320–1326.
12. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, M. L. Ho, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi.** 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. J. Clin. Microbiol. **41:**1963–1970.
13. **Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. H. Hazbon, M. Bobadilla del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. Leon, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Joloba, A. Rendon, J. Sifuentes-Osornio, A. Ponce de Leon, M. D. Cave, R. Fleischmann, T. S. Whittam, and D. Alland.** 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J. Bacteriol. **188:**759–772.
14. **Flores, L., T. Van, S. Narayanan, K. Deriemer, M. Kato-Maeda, and S. Gagneux.** 2007. Large sequence polymorphisms classify *Mycobacterium tuberculosis* strains with ancestral spoligotyping patterns. J. Clin. Microbiol. **45:**3393–3395.
15. **Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell, and P. M. Small.** 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA **103:**2869–2873.
16. **Goh, K. S., N. Rastogi, M. Berchel, R. C. Huard, and C. Sola.** 2005. Molecular evolutionary history of tubercle bacilli assessed by study of the polymorphic nucleotide within the nitrate reductase (narGHJI) operon promoter. J. Clin. Microbiol. **43:**4010–4014.
17. **Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser.** 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J. Infect. Dis. **193:**121–128.
18. **Gutacker, M. M., J. C. Smoot, C. A. Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser.** 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics **162:**1533–1543.
19. **Hirsh, A. E., A. G. Tsolaki, K. DeRiemer, M. W. Feldman, and P. M. Small.** 2004. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc. Natl. Acad. Sci. USA **101:**4871–4876.
20. **Huard, R. C., L. C. de Oliveira Lazzarini, W. R. Butler, D. van Soolingen, and J. L. Ho.** 2003. PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. J. Clin. Microbiol. **41:**1637–1650.
21. **Huard, R. C., M. Fabre, P. de Haas, L. C. Lazzarini, D. van Soolingen, D. Cousins, and J. L. Ho.** 2006. Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. J. Bacteriol. **188:**4271–4287.
22. **Ignatova, A., S. Dubiley, V. Stepanshina, and I. Shemyakin.** 2006. Predominance of multidrug-resistant LAM and Beijing family strains among *Mycobacterium tuberculosis* isolates recovered from prison inmates in Tula Region, Russia. J. Med. Microbiol. **55:**1413–1418.
23. **Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. **35:**907–914.
24. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden.** 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. **37:**2607–2618.
25. **Lazzarini, L. C., R. C. Huard, N. L. Boechat, H. M. Gomes, M. C. Oelemann, N. E. Kurepina, E. Shashkina, F. C. Q. Mello, A. Gibson, M. J. Virginio, A. G. Marsicob, W. R. Butler, B. N. Kreiswirth, P. N. Suffys, J. R. Lapa e Silva, and J. L. Ho.** 2007. Discovery of a novel *Mycobacterium tuberculosis* lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. J. Clin. Microbiol. **45:**3891–3902.
26. **Marais, B. J., T. C. Victor, A. C. Hesseling, M. Barnard, A. Jordaan, W. Brittle, H. Reuter, N. Beyers, P. D. van Helden, R. M. Warren, and H. S. Schaaf.** 2006. Beijing and Haarlem genotypes are overrepresented among children with drug-resistant tuberculosis in the Western Cape Province of South Africa. J. Clin. Microbiol. **44:**3539–3543.
27. **Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch.** 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. Microbiology **150:**483–496.
28. **Martin, A., F. Chaves, J. Inigo, M. Alonso, C. Sola, N. Rastogi, M. J. Ruiz Serrano, E. Palenque, E. Bouza, and D. Garcia de Viedma.** 2007. Molecular, epidemiological, and infectivity characterisation of a *Mycobacterium tuberculosis* strain prevalent in Madrid. Clin. Microbiol. Infect. doi:10.1111/j.1469-0691.2007.01820.
29. **Musser, J. M., A. Amin, and S. Ramaswamy.** 2000. Negligible genetic diver-

sity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. Genetics **155:**7–16.

30. **Richardson, M., S. W. van Lill, G. D. van der Spuy, Z. Munch, C. N. Booysen, N. Beyers, P. D. van Helden, and R. M. Warren.** 2002. Historic and recent events contribute to the disease dynamics of Beijing-like *Mycobacterium tuberculosis* isolates in a high incidence region. Int. J. Tuberc. Lung Dis. **6:**1001–1011.
31. **Sebban, M., I. Mokrousov, N. Rastogi, and C. Sola.** 2002. A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis*. Bioinformatics **18:**235–243.
32. **Shemyakin, I. G., V. N. Stepanshina, I. Y. Ivanov, M. Y. Lipin, V. A. Anisimova, A. G. Onasenko, O. V. Korobova, and T. M. Shinnick.** 2004. Characterization of drug-resistant isolates of *Mycobacterium tuberculosis* derived from Russian inmates. Int. J. Tuberc. Lung Dis. **8:**1194–1203.
33. **Sola, C., I. Filliol, E. Legrand, I. Mokrousov, and N. Rastogi.** 2001. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS1081, IS6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J. Mol. Evol. **53:** 680–689.
34. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **94:**9869–9874.
35. **Streicher, E. M., T. C. Victor, G. van der Spuy, C. Sola, N. Rastogi, P. D. van Helden, and R. M. Warren.** 2007. Spoligotype signatures in the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **45:**237–240.
36. **Supply, P., R. M. Warren, A. L. Banuls, S. Lesjean, G. D. Van Der Spuy, L. A. Lewis, M. Tibayrenc, P. D. Van Helden, and C. Locht.** 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. Mol. Microbiol. **47:**529–538.
37. **van Embden, J. D., T. van Gorkom, K. Kremer, R. Jansen, B. A. van Der Zeijst, and L. M. Schouls.** 2000. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J. Bacteriol. **182:**2393–2401.
38. **van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden.** 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. **33:**3234–3238.
39. **Victor, T. C., P. E. de Haas, A. M. Jordaan, G. D. van der Spuy, M. Richardson, D. van Soolingen, P. D. van Helden, and R. Warren.** 2004. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a western cape F11 genotype. J. Clin. Microbiol. **42:**769–772.
40. **Warren, R. M., E. M. Streicher, S. L. Sampson, G. D. van der Spuy, M. Richardson, D. Nguyen, M. A. Behr, T. C. Victor, and P. D. van Helden.** 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. J. Clin. Microbiol. **40:**4457–4465.
41. **WHO.** 2006. Global tuberculosis control: surveillance, planning, financing, p.77. World Health Organization, Geneva, Switzerland.
42. **Yang, Z., P. F. Barnes, F. Chaves, K. D. Eisenach, S. E. Weis, J. H. Bates, and M. D. Cave.** 1998. Diversity of DNA fingerprints of *Mycobacterium tuberculosis* isolates in the United States. J. Clin. Microbiol. **36:**1003–1007.