JOURNAL OF CLINICAL MICROBIOLOGY, Oct. 2010, p. 3614–3623 Vol. 48, No. 10
0095-1137/10/$12.00 doi:10.1128/JCM.00157-10

Genomic Signatures of the Haarlem Lineage of *Mycobacterium tuberculosis*: Implications of Strain Genetic Variation in Drug and Vaccine Development▽

Andrés Cubillos-Ruiz,[1]† Andrea Sandoval,[1]† Viviana Ritacco,[2] Beatriz López,[2] Jaime Robledo,[3,4] Nidia Correa,[3,4] Iván Hernandez-Neuta,[1] Maria Mercedes Zambrano,[1] and Patricia Del Portillo[1,4]*

Corporación Corpogen, Carrera 5 no. 66A-34, Bogotá, Colombia[1]; Instituto Nacional de Enfermedades Infecciosas ANLIS "Carlos G. Malbrán," Velez Sarsfield 563, Buenos Aires, Argentina[2]; Corporación para Investigaciones Biológicas, CIB, Universidad Pontificia Bolivariana, Carrera 72a no. 78B-141, Medellín, Colombia[3]; and Centro Colombiano de Investigación en Tuberculosis, Medellín, Colombia[4]

Received 25 January 2010/Returned for modification 5 April 2010/Accepted 7 July 2010

Tuberculosis is the world's leading cause of death due to a single infectious agent, and efforts aimed at its control require a better understanding of host, environmental, and bacterial factors that govern disease outcome. Growing evidence indicates that certain *Mycobacterium tuberculosis* strains of distinct phylogeographic lineages elicit unique immunopathological events. However, identifying the genetic basis of these phenotypic peculiarities has proven difficult. Here we report the presence of six large sequence polymorphisms which, together with two single-nucleotide changes previously described by our group, consistently differentiate Haarlem strains from the remaining *M. tuberculosis* lineages. The six newly found Haarlem-specific genetic events are four deletions, which altogether involve more than 13 kb, and two intragenic insertions of the element IS*6110*. The absence of the genes involved in these polymorphisms could have an important physiological impact on Haarlem strains, i.e., by affecting key genes, such as *Rv1354c* and *cyp121*, which have been recently proposed as plausible drug targets. These lineage-specific polymorphisms can serve as genetic markers for the rapid PCR identification of Haarlem strains, providing a useful tool for strain surveillance and molecular epidemiology studies. Strain variability such as that described here underscores the need for the definition of a core set of essential genes in *M. tuberculosis* that are ubiquitously present in all circulating lineages, as a requirement in the development of effective antituberculosis drugs and vaccines.

Mycobacterium tuberculosis is the causative agent of tuberculosis, the leading cause of death by a single bacterial agent in the world (36). Infection with *M. tuberculosis* has historically shown to result in a variety of clinical outcomes that are usually associated with host inherited susceptibility and environmental risk factors (2, 31, 32). Moreover, increasing evidence suggests that genetic variation in the tubercle bacilli also plays an important role in the outcome of the disease (4, 19, 33). Due to the absence of exchange of genetic material with a global microbial gene pool, *M. tuberculosis* had long been considered to have a clonal population structure. However, a significant strain-to strain genetic variation within *M. tuberculosis* has recently been unveiled (11, 19).

Changes in neutral regions of the chromosome, such as the direct repeat (DR) locus, and in the mycobacterial interspersed repetitive units (MIRUs) are useful in epidemiological and phylogenetic analyses and in describing the most conspicuous *M. tuberculosis* lineages (3, 21). In addition to the variation in neutral regions, genetic polymorphisms involving coding regions have been described to occur through single-nucleotide changes and through deletion and insertion events, the latter mediated mainly by the IS*6110* element (23, 30). Although these genomic alterations are thought to be among the principal sources of phenotypic variation in *M. tuberculosis*, the specific genomic changes that define each lineage have not yet been fully defined.

There are currently six phylogeographic lineages that make up the *M. tuberculosis* global population (10). One is the Euro-American group, which includes all the spoligotype families predominating in the Western world, such as Haarlem, LAM, and the ill-defined T group (3). In particular, the Haarlem genotype is ubiquitous worldwide (15) and represents about 25% of the isolates in Europe, Central America, and the Caribbean, suggesting a link with the post-Columbus European colonization (8). Haarlem strains are actively transmitted in urban settings in Colombia, causing major public health problems (N. E. Correa, E. Zapata, V. Gómez, G. E. Mejia, A. Restrepo, J. Robledo, and CCITB, presented at the 107th General Meeting of the American Society for Microbiology, Toronto, Canada, 2007) and have also been responsible for a prolonged outbreak of multidrug-resistant tuberculosis in Argentina (26, 29).

An intriguing question is whether *M. tuberculosis* strains differ in terms of pathogenic characteristics as a consequence of long-standing interactions of particular lineages with specific human populations. Animal models that take advantage of an identical genetic background, and therefore a uniform host immune response, have given insight regarding the contribution of strain genetic diversity to the outcome of the infectious process (7, 20). It is currently accepted that genetically different *M. tuberculosis* strains produce markedly different immunopathological events in isogenic mice (4, 18). Thus, under-

* Corresponding author. Mailing address: Corporación Corpogen, Carrera 5 no. 66a-34, Bogotá, D.C., Colombia. Phone: 57-1-8050106. Fax: 57-1-3484607. E-mail: pdelportillo@corpogen.org.
† These authors contributed equally to this work.
▽ Published ahead of print on 14 July 2010.

standing genotypic differences and mechanisms underlying infection variability and identifying specific changes or genes associated with both virulence and immunopathogenicity of the different *M. tuberculosis* lineages have important implications for the future effective control of tuberculosis (7, 33).

In a recent bioinformatic study using multiple genome alignments of six fully sequenced *M. tuberculosis* strains belonging to different lineages, we showed a trend toward accumulation of a limited number of genome-specific polymorphisms preferentially associated with circulating strains and underrepresented in laboratory strains. This suggests that such polymorphisms arise as active mechanisms of adaptation to the human host (5). We speculated that some of these genome-specific polymorphisms might be common to strains of a particular lineage rather than being an exclusive property of the isolate examined. To test this, in the present study we examined whether genome-specific polymorphisms previously identified in fully sequenced strains were present in a broader group of strains and could thus represent a lineage-wide condition. In particular, we explored whether polymorphisms identified as specific to the sequenced *M. tuberculosis* Haarlem strain (5) were prevalent in additional members of the Haarlem lineage and absent from other lineages. In the present paper, we report the presence of eight genomic signatures highly exclusive to the *M. tuberculosis* Haarlem lineage that can prove important for the rapid identification of these strains and also contribute to our understanding of the genetic variations underlying phenotypic differences among the various lineages of the tubercle bacilli.

MATERIALS AND METHODS

***M. tuberculosis* isolates.** A set of 40 *M. tuberculosis* clinical isolates belonging to the Haarlem lineage and 62 non-Haarlem isolates, including LAM, S, T, X, EIA, and Beijing, were selected from the collection of the Instituto Nacional de Enfermedades Infeccciosas ANLIS "Carlos G. Malbrán" in Buenos Aires, Argentina, and from the collection of the Centro Colombiano de Investigación en Tuberculosis (CCITB) held at the Corporación para Investigaciones Biológicas (CIB) in Medellín, Colombia. Isolates were selected based on different IS*6110* restriction fragment length polymorphism (RFLP) patterns to ensure that they represented the most conspicuous patterns of strains circulating in both settings between 1997 and 2005. Laboratory strain H37Rv was also included in the non-Haarlem group (see Fig. 2). DNA was obtained from culture lysates as described previously (25).

IS*6110* RFLP typing and phylogenetic analysis. IS*6110* RFLP and spoligotype patterns (14, 35) were available at genotype databases in Buenos Aires and Medellin laboratories. Computer-assisted analysis of IS*6110* RFLP patterns was performed with the software BioNumerics 5.1 (Applied Maths, Sint-Martens-Latem, Belgium) as described previously (12). Similarity between patterns was calculated using the Dice coefficient with 1% band position tolerance and 1% optimization. Cluster analysis was performed using the unweighted pair group method with arithmetic averages (UPGMA). Phylogenetic lineages and spoligo-international shared types (SITs) were assigned according to SpolDB4, available at www.pasteur-guadeloupe.fr/tb/bd_myco.htlm (3).

Primers and PCR assays. Two sets of primers were designed for each polymorphic region to determine the presence or absence of a specific deletion in each *M. tuberculosis* isolate. IS*6110* insertions were detected using primers that annealed with the IS*6110* flanking regions, generating bands that differed in size (1,362 bp) depending on whether the IS*6110* element was present or absent. Table 1 summarizes the sequences of the primers and the expected amplification products for each region. All PCRs were performed in a iCycler DNA thermal cycler (Bio-Rad) in a final volume of 50 μl containing 2.5 units of TucanTaq DNA polymerase (Corpogen, Bogotá, Colombia), 1× TucanTaq amplification buffer, 1.5 mM $MgCl_2$, 0.5 μM each primer, 0.3 mM deoxynucleoside triphosphates (dNTPs), and 2 μl of DNA from culture lysate extracts. For detection of deletions, PCRs were carried out for 35 cycles consisting of 45 s of denaturation at 94°C, 45 s of annealing at 64°C, and 120 s of extension at 72°C. For detection of insertions, PCRs were carried out for 35 cycles consisting in 45 s of denaturation at 94°C, 45 s of annealing at 66°C for HSI1 and 71°C for HSI2, and 120 s of extension at 72°C. PCR products were verified by 1.5% agarose gel electrophoresis for the presence of a single amplification band. Five randomly chosen products for each region were sequenced using the BigDye terminator cycling conditions (Macrogen, South Korea) in order to confirm that the target region was amplified. For the detection of single-nucleotide polymorphisms (SNPs) in the *ogt* and *ung* genes, the primers and conditions reported previously for allelic discriminatory PCR were used (25).

Statistical analysis. The Fisher exact test was applied to determine significant associations between polymorphisms and *M. tuberculosis* lineages.

RESULTS

Specific polymorphisms in strains of the Haarlem lineage of *M. tuberculosis*. Of 12 deletions and 6 insertions identified in a previous bioinformatic study as unique to the sequenced Haarlem strain (5) (www.broadinstitute.org/), we selected the most conspicuous to investigate if they were lineage-wide mutations. Specifically, the IS*6110* insertions and the largest deletion polymorphisms were chosen for a preliminary analysis using PCR with four Haarlem strains. Polymorphisms spanning repetitive regions, such as Pro-Pro-Glu (PPE) family genes, were excluded from the present analysis in order to avoid possible misinterpretation. Likewise, polymorphisms of ≤200 bp were excluded because they cannot be unequivocally differentiated from intrinsic errors occurred during sequencing and finishing of the Haarlem strain genome. Results from this preliminary analysis indicated that only six polymorphisms were in fact present in the four analyzed Haarlem strains (Table 2). The occurrence of these mutations was therefore further inspected using a larger panel of epidemiologically unrelated isolates from Argentina and Colombia. For this analysis we used these six large-sequence polymorphisms and two additional SNPs located in the *ogt* and *ung* DNA repair genes previously reported by our group as specific to the Haarlem lineage (25).

The analysis of the 102 strains indicated that the presence of all eight studied polymorphisms correlated highly with the Haarlem lineage (Table 3). For this reason, the regions displaying deletions were designated Haarlem-specific deletions (HSD1 to HSD4), and the two IS*6110* element insertions were named Haarlem-specific insertions (HSI1 and HSI2). Likewise, SNPs present in genes *ogt* and *ung* were named Haarlem-specific SNPs (HSSNP1 and HSSNP2, respectively). When analyzed individually, each of these genetic events showed a highly significant association with the Haarlem lineage: HSD1 was found in 37/40 Haarlem versus 1/62 non-Haarlem strains ($P < 0.00001$), HSD2 and HSD3 were found in 38/40 Haarlem versus 1/62 non-Haarlem isolates ($P < 0.00001$), and HSD4 was found in 38/40 Haarlem versus 2/62 non-Haarlem isolates ($P < 0.00001$). The two insertions of the IS*6110* element also correlated highly with the Haarlem lineage: HSI1 was present in 38/40 Haarlem versus 4/62 non-Haarlem isolates ($P < 0.00001$), and HSI2 was present 38/40 Haarlem versus 0/62 non-Haarlem strains ($P < 0.00001$). Lastly, and commensurate with previous reports, HSSNP1 in *ogt* and HSSNP2 in *ung* both were present in 38/40 Haarlem versus 1/62 non-Haarlem isolates ($P < 0.00001$).

Genes involved in the deletions and insertions. The genes involved in large Haarlem-specific polymorphisms are depicted in Fig. 1. HSD1 is a 1,774-bp deletion that removes most of genes *helZ* and *Rv2102*, HSD2 is a 6,480-bp deletion that

TABLE 1. Primers used in this study for the identification of the Haarlem-specific polymorphisms

Region or locus	Lineage	Primer	Sequence	Product Size (bp)	Reference
HSD1	Haarlem	hsd1 A(f)	5′ CGCTCCGTCGACAAGAGAG	2,230	This study
		hsd1 B(r)	5′ TATCCTGGCGAGAATGCTGA		
	Non-Haarlem	hsd1 C(f)	5′ ACGCGGCCCTACATCCT	1,786	This study
		hsd1 B(r)	5′ TATCCTGGCGAGAATGCTGA		
HSD2	Haarlem	hsd2 A(f)	5′ TTGCGCGAATGTGCTTTCTC	1,840	This study
		hsd2 B(r)	5′ CCGGCCGGCTCTTGTC		
	Non-Haarlem	hsd2 A(f)	5′ TTGCGCGAATGTGCTTTCTC	3,507	This study
		hsd2 C(r)	5′ CTTCGGGCCGTCTTCTTGC		
HSD3	Haarlem	hsd3 A(f)	5′ TAAGCCCTCAACGCGCCACC	831	This study
		hsd3 B(r)	5′ GCGCTCGATCCCACGTTGT		
	Non-Haarlem	hsd3 A(f)	5′ TAAGCCCTCAACGCGCCACC	1,737	This study
		hsd3 C(r)	5′ CACACCGTCGGACCTCCTGC		
HSD4	Haarlem	hsd4 A(f)	5′ AACACGCCGATACCTATTTGGTC	849	This study
		hsd4 B(r)	5′ CGTGAGGGCATCGAGGTGGC		
	Non-Haarlem	hsd4 A(f)	5′ AACACGCCGATACCTATTTGGTC	1,287	This study
		hsd4 B(r)	5′ CGTGAGGGCATCGAGGTGGC		
HSI1	Haarlem	hsi1 A(f)	5′ AATGCCGTCGTGGTCAA	2,381	This study
		hsi1 B(r)	5′ CGGTTTCTCGGGTGCTAC		
	Non-Haarlem	hsi1 A(f)	5′ AATGCCGTCGTGGTCAA	1,019	This study
		hsi1 B(r)	5′ CGGTTTCTCGGGTGCTAC		
HSI2	Haarlem	hsi2 A(f)	5′ GGTCAGGCTGCGGGATGTT	2,280	This study
		hsi2 B(r)	5′ AGCGTTGCGGGATACTCTGG		
	Non-Haarlem	hsi2 A(f)	5′ GGTCAGGCTGCGGGATGTT	918	This study
		hsi2 B(r)	5′ AGCGTTGCGGGATACTCTGG		
HSSNP1	Haarlem	ogt F-M(f)	5′ CCCCATCGGGCCATTAAG	545	25
		ogt R(r)	5′ ACTCAGCCGCTCGCGAGC		
	Non-Haarlem	ogt F-W(f)	5′ CCCCATCGGGCCATTAAC	545	25
		ogt R(r)	5′ ACTCAGCCGCTCGCGAGC		
HSSNP2	Haarlem	ung F-M(f)	5′ GCTGGTGGCGATCCTA	287	25
		ung R(r)	5′ GGCAACAAGAAGCGACTC		
	Non-Haarlem	ung F-W(f)	5′ GCTGGTGGCGATCCTG	287	25
		ung R(r)	5′ GGCAACAAGAAGCGACTC		
DR	Haarlem	hsd4 IS7(f)	5′ AACACGCCGATACCTATTTGGTC		This study
		hsd4 INS1(r)	5′ CGTGAGGGCATCGAGGTGGC		
		hsd4 DR30(r)	5′ GAAACTCTTGACGATGCGGTTG		

removes genes *Rv2271* through *Rv2278* and partially truncates *lppN*, HSD3 is a 4,753-bp deletion that affects genes *Rv1353c* through *Rv1356c*, and HSD4 is a 439-bp deletion within the DR locus between genes *Rv2813* and *Rv2814c*. Insertions HSI1 and HSI2 interrupt genes *Rv2336* and *Rv0963c*, respectively.

Detailed examination of the Haarlem isolates in the set showed that every strain classified within the H1 and the H2 subfamilies consistently displayed all polymorphisms, as did 10 out of 12 strains classified within the H3 subfamily (Table 3). Another H3 strain (isolate no. 1511) displayed seven of the eight analyzed polymorphisms. In contrast, two isolates (no. 1633 and 1089, of the H3 and H4 subfamilies, respectively), did not have any of these polymorphisms. Consistently, these two isolates did not fit within the Haarlem branch in the IS*6110* RFLP dendrogram constructed with the whole set of *M. tuberculosis* strains used in this study (Fig. 2). On the other hand, the large majority of isolates belonging to non-Haarlem lineages did not contain these polymorphic regions (Table 3). More importantly, only 6 out of 62 non-Haarlem strains displayed some of these polymorphisms, and none of these six strains harbored all of them. Five of them belonged to the LAM lineage and carried only one Haarlem-specific polymorphism each, suggesting a certain relationship with the Haarlem lineage or, alternatively, an evolutionary process involving similar selection pressures (Table 3 and Fig. 2, DNA numbers UT89, UT272, 1632, 1506, and 1516). The sixth strain, which had an undefined (U) lineage, appeared to be fairly close to the Haarlem family by both IS*6110* pattern and spoligotype, and its relatedness to this lineage was confirmed by its displaying six out of the eight Haarlem-specific polymorphisms (Fig. 2, DNA UT148).

Genomic organization of the DR locus in Haarlem strains. The HSD4 deletion mapped to the DR region and eliminated spacers 26 to 31. PCR was positive for this deletion in all Haarlem 1 and 2 subfamilies and in 11 out of 12 strains belonging to Haarlem 3 subfamily (Fig. 2). This deletion explains

TABLE 2. Identification of indels in four *M. tuberculosis* Haarlem strains

Deletion or insertion[a]	Position	Size (bp)	Gene product	Presence in Haarlem strain:				PCR primers
				1	2	3	4	
HSDs								
1	912152	92	Rv0823c (Possible transcriptional regulatory protein)	ND[b]	ND	ND	ND	
2	965426	481	PE-PGRS15	ND	ND	ND	ND	
3	1451173	115	Intergenic region between ThrB (threonine synthetase) and ThrC (homoserine kinase)	ND	ND	ND	ND	
4	1521604	4,753	Rv1353c (probable transcriptional regulatory protein)-Rv1354c (diguanylate cyclase protein)-MoeY (molybdopterin biosynthesis protein)-Rv1356c (hypothetical protein)	+	+	+	+	5′TAAGCCCTCAACGCGCCACC, 5′GCGCTCGATCCCACGTTGT
5	2060412	643	PE-PGRS33	ND	ND	ND	ND	
6	2365930	1,774	HelZ (Helicase)-Rv2102 (conserved hypothetical protein)	+	+	+	+	5′CGCTCCGTCGACAAGAGAG, 5′TATCCTGGCGAGAATG CTGA
7	2546849	6,480	Rv2271 (hypothetical protein)-Rv2272 (probable conserved transmembrane protein)-Rv2273 (probable conserved transmembrane protein)-Rv2274c (hypothetical protein)-Rv2275 (hypothetical protein)-Cyp121 (cytochrome P450)-Rv2 277c (probable glycerolphosphodiesterase)	+	+	+	+	5′TTGCGCGAATGTGCTTTCTC, 5′CCGGCCGGCTCTTGTC
8	2792261	170	PE-PGRS42	ND	ND	ND	ND	
9	3117314	439	Repeat region (DR) intergenic between Rv2813 and Rv2814	+	+	+	+	5′CGTGAGGGCATCGAGGTGGC, 5′AACACGCCGATACCTATTT GGTC
10	3192481	87	IS*1539* Transposase contained in repeat region	ND	ND	ND	ND	
11	4033587	616	Rv3600c (conserved hypothetical protein)-PanD (aspartate decarboxylase precursor)	(−)	+	(−)	+	5′CGGGTGTTCTTCTTGGAG, 5′CGACGTGCCAGCTTAA ATGA
12	4365135	84	PPE69 and intergenic region with PE36	ND	ND	ND	ND	
HSIs								
1	2606037–2607396	1,359	Insertion of the IS*6110* element on Rv2336 (hypothetical protein)	+	+	+	+	5′AATGCCGTCGTGGTCAA, 5′CGGTTTCTCGGGTGCTAC
2	1071617–1072976	1,359	Insertion of the IS*6110* element on Rv0963c (conserved hypothetical protein)	+	+	+	+	5′GGTCAGGCTGCGGGATGTT, 5′AGCGTTGCGGGATACT CTGG
3	1714841–1716200	1,359	Insertion of the IS*6110* element on Mmpl12 (transmembrane protein)[c]	−	−	−	−	5′CGGCGAGAAAGATGTGC, 5′TTGAGGGCGTTGACCATAG
4	1834511–1834569	58	Rv1637c (hypothetical protein)	ND	ND	ND	ND	
5	1973210–1973436	226	PPE24	ND	ND	ND	ND	
6	4144939–4144997	58	LeuA (2-isopropylmalate synthetase)	ND	ND	ND	ND	

[a] HSD, Haarlem strain deletion; HSI, Haarlem strain insertion.
[b] ND, not done.
[c] No amplification.

TABLE 3. Presence of Haarlem-specific polymorphisms in a set of 102 *M. tuberculosis* strains (40 belonging to the Haarlem lineage and 62 non-Haarlem)[a]

Lineage	No. (%) of:								
	Isolates	Deletions				Insertions		SNPs	
		HSD1	HSD2	HSD3	HSD4	HSI1	HSI2	HSSNP1	HSSNP2
Haarlem									
H1	17	17	17	17	17	17	17	17	17
H2	10	10	10	10	10	10	10	10	10
H3	12	10	11	11	11	11	11	11	11
H4	1	0	0	0	0	0	0	0	0
Total	40 (100)	37 (92.5)	38 (95.0)	38 (95.0)	38 (95.0)	38 (95.0)	38 (95.0)	38 (95.0)	38 (95.0)
Non-Haarlem									
LAM	28	0	0	1	0	3	0	0	0
LAM3/S convergent	1	0	0	0	1	0	0	0	0
S	3	0	0	0	0	0	0	0	0
T	20	0	0	0	0	0	0	0	0
Undetermined	4	1	1	0	1	1	0	1	1
X	3	0	0	0	0	0	0	0	0
EAI	1	0	0	0	0	0	0	0	0
Beijing	1	0	0	0	0	0	0	0	0
H37Rv	1	0	0	0	0	0	0	0	0
Total	62 (100)	1 (1.6)	1 (1.6)	1 (1.6)	2 (3.2)	4 (6.5)	0 (0)	1 (1.6)	1 (1.6)

[a] Abbreviations: H, Haarlem; LAM, Latin America and Mediterranean; EAI, East African and Indian; HSD, Haarlem-specific deletion; HSI, Haarlem-specific insertion; SNP, single-nucleotide polymorphism; HSSNP, Haarlem-specific single-nucleotide polymorphism.

clearly the spoligotype observed in *M. tuberculosis* strains of the Haarlem 1 and 2 subfamilies, which are characterized by the absence of those six spacers. However, most strains of the H3 subfamily lack only spacer 31 in the spoligotyping despite the presence of spacers 26 through 30. In order to understand the DR organization in H3, we designed primers to amplify the DR locus between spacers 30 to 32 and between the IS*6110* and these spacers in strains belonging to Haarlem subfamilies 1, 2, and 3. Figure 3 shows a schematic representation of the DR locus organization in the Haarlem lineage in the H1, H2, and H3 subfamilies. A deletion of spacers 26 to 31 was confirmed by sequencing in all six analyzed strains belonging to H1 and H2. The sequence of H3 showed a different DR locus organization. We identified the presence of spacers 26 to 31 downstream of the ancestral IS*6110* insertion, followed by an extra copy of the insertion element together with spacers 25 and 32. This organization was further confirmed by comparison with the Haarlem strain sequenced by the Broad Institute (www.broadinstitute.org/). The identity was 100%, indicating that the sequenced strain belongs to the Haarlem 3 subfamily.

DISCUSSION

The Haarlem family was described in the Netherlands in 1999 (15). The family is highly diverse and has been amply studied to better understand its evolutionary history. In previous work, we identified two SNPs in DNA repair genes, *ung* and *ogt*, present in all analyzed Haarlem strains (25). In the present study we report additional markers that can constitute a genomic signature of the Haarlem family. These genomic markers include six specific large polymorphisms (four deletions and two IS*6110* insertions) along with the two previously described SNPs in *ung* and *ogt*. Three of the deletions involved in these specific polymorphisms have been previously reported in 5 out of 100 clinical isolates analyzed by microarray technology from a collection of epidemiologically well-characterized isolates from San Francisco (34). Four of these strains belong to the Haarlem family and the other one to the U lineage (M. Kato-Maeda, personal communication), reinforcing the idea that these are Haarlem-specific polymorphisms. The fact that strains from different geographical origins such as Argentina, Colombia, and the United States share the same specific polymorphisms prompts us to propose that these mutations are widely distributed.

The high frequency of these polymorphisms within one of the most widespread and successful genotypes can have key biological significance. In particular, these Haarlem-specific polymorphisms may have important functional consequences for the tubercle bacilli and be relevant in terms of strategies for disease control. This is especially evident with respect to the validity of some recently proposed drug targets. Gene *Rv1354c*, for example, which codes for the only identified putative diguanylate cyclase in the genome, is associated with the inner membrane (22) and thought to be involved in the turnover of cyclic-di-GMP, a multifunctional second-messenger molecule exclusive of the bacterial domain. Based on this, *Rv1354c* has been recently proposed as an ideal target for the design of new drugs (6). However, gene *Rv1354c* is completely deleted in our Haarlem strains as part of HSD3, indicating that strains with this genotype can dispose of this protein and signaling pathway without losing their capacity to infect and cause disease. Thus, gene *Rv1354c* cannot be considered a suitable antituberculosis drug target (6). Similarly, the cytochrome P450 gene *cyp121* was shown to be essential for *M. tuberculosis* H37Rv viability and was proposed as a novel target for azole drugs (24). This gene, however, is also deleted in Haarlem strains as part of HSD2, making necessary a reevaluation of the antimicrobial activity in circulating strains. The results obtained here underscore the importance of strain diversity and the need to iden-

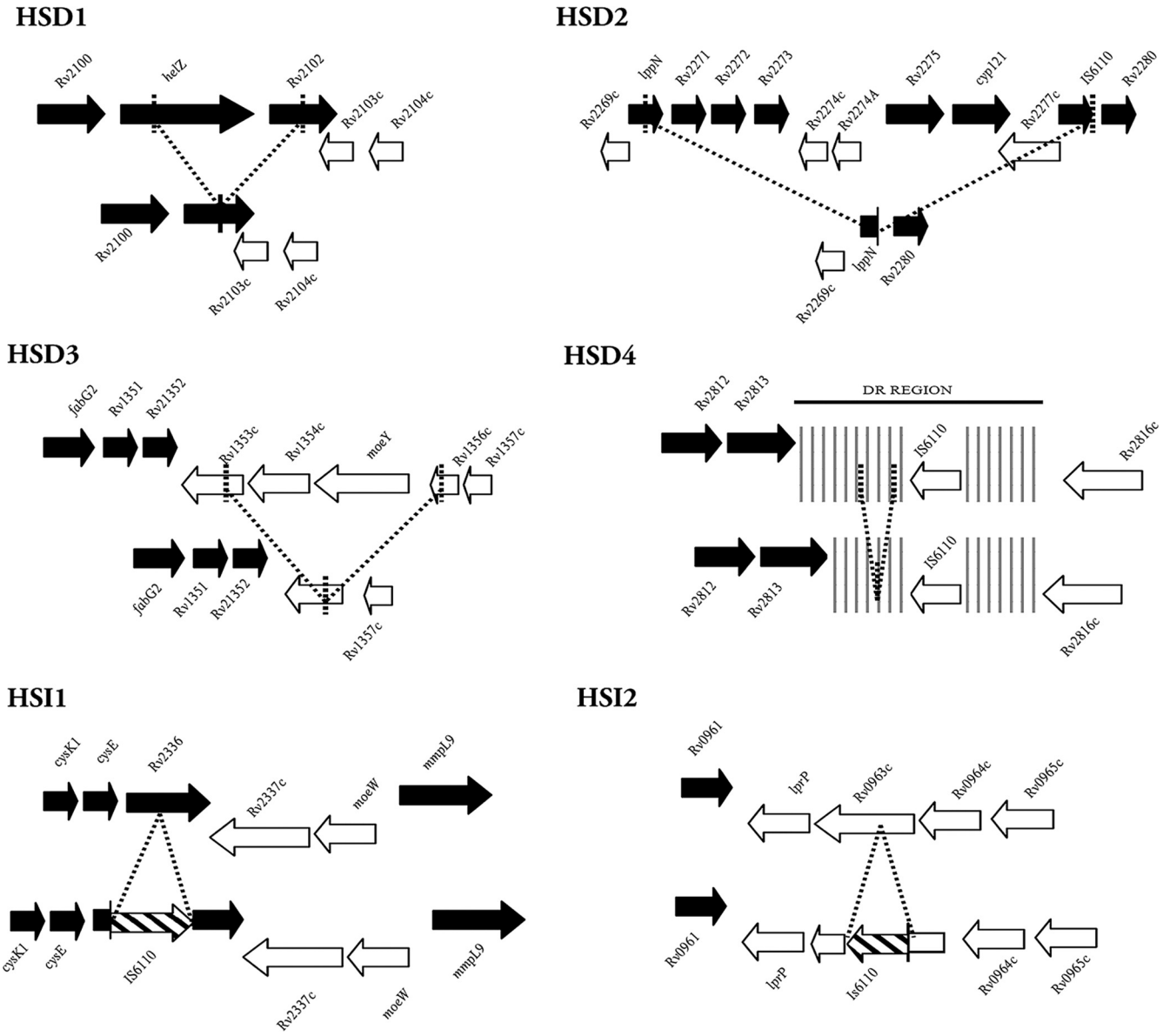


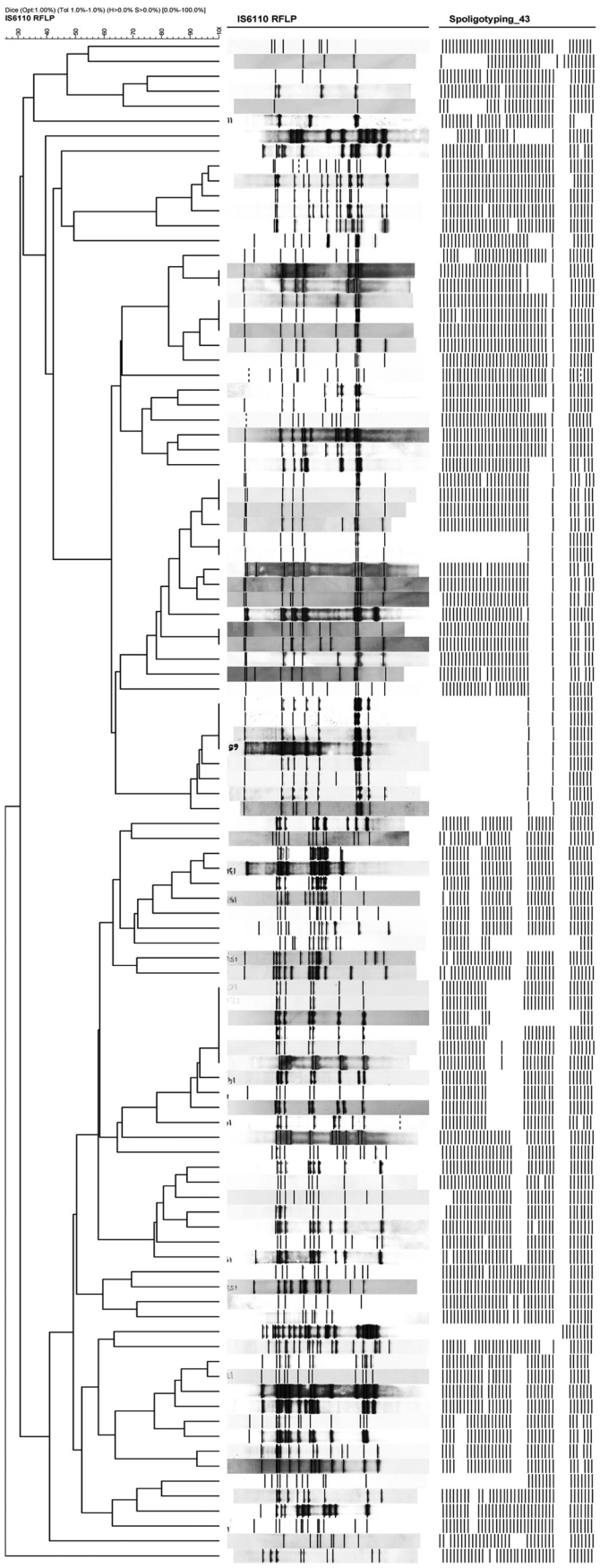
FIG. 1. Haarlem-specific polymorphisms. Genes involved in four Haarlem-specific deletions (HSD1 to HSD4) and two Haarlem-specific insertions (HSI1 and HSI2) are shown. The upper part of each horizontal panel represents the wild-type sequence, and the lower part represents the Haarlem genotype.

tify a core set of genes common to all *M. tuberculosis* lineages as a crucial step in the development of new antituberculosis drugs.

Likewise, genetic variation can reflect differences in antigenic repertoire composition among the different lineages of *M. tuberculosis*, as pointed out previously (34) and exemplified here by deletion of the transmembrane proteins Rv2272 and Rv2273 as part of HSD2. Consequently, vaccine candidates should be effective against challenge not only with laboratory strains but also with strains representative of the major lineages of the global population of *M. tuberculosis.*

In addition to these, other genes affected in Haarlem strains could result in important phenotypic changes. Genes *Rv2274c* and *Rv2274A*, absent due to HSD2, have been annotated among the 38 toxin-antitoxin (TA) operons present in the *M. tuberculosis* genome (1, 27). It has been proposed that these systems can fulfill a variety of roles associated with retardation of cell growth and persistence in stressful environments (1). The IS*6110* element insertion in HSI1 interrupts *Rv2336*, a gene that has been implicated in virulence because it is downregulated in the attenuated strain H37Ra (28). No obvious phenotypic effects can be inferred from the other specific polymorphisms identified here, i.e., HSD1, HSI2, HSSNP1, and HSSNP2, and additional functional analysis of these mutations would be required.

A striking result of our analysis is that the lineage classification given by spoligotyping matches almost perfectly with the one resulting from the presence of these Haarlem-specific



	Country	Strain	SIT	Lineage	ogt/ung genes
	AR	1147*	53	T1 Ghana	wt/wt
	CO	UT393	8	EAI5 or EAI3	wt/wt
	CO	UT147	37	T3	wt/wt
	CO	UT55	119	X1	wt/wt
	CO	UT387	91	X3-variant1	wt/wt
	AR	1153*	439	X2	wt/wt
⊠	AR	1633*	750	H3	wt/wt
	AR	1636*	73	T2-T3	wt/wt
	AR	848*	53	T1 Ghana	wt/wt
	AR	926*	240	U	wt/wt
	AR	839*	53	T1 Ghana	wt/wt
	AR	981*	37	T3	wt/wt
	AR	H37Rv	451	H37Rv	wt/wt
★	CO	UT148	1561	U (H1?)	mut/mut
★	AR	1511*	orphan	(H3?)	mut/mut
➔	CO	UT105	45	H1	mut/mut
➔	CO	UT53	45	H1	mut/mut
➔	CO	UT125	50	H3	mut/mut
➔	CO	UT142	207	H3	mut/mut
➔	CO	UT354	50	H3	mut/mut
➔	CO	UT303	50	H3	mut/mut
➔	CO	UT144	50	H3	mut/mut
➔	AR	1527*	49	H3	mut/mut
➔	AR	1542*	50	H3	mut/mut
➔	AR	951*	47	H1	mut/mut
➔	AR	1148*	50	H3	mut/mut
➔	AR	1518*	50	H3	mut/mut
➔	AR	1629*	50	H3	mut/mut
➔	AR	1248*	62	H1	mut/mut
➔	CO	UT137	727	H1	mut/mut
➔	CO	UT201	62	H1	mut/mut
➔	CO	UT305	62	H1	mut/mut
➔	CO	UT267	62	H1	mut/mut
➔	AR	116*	2	H2	mut/mut
➔	AR	118*	2	H2	mut/mut
➔	CO	UT174	727	H1	mut/mut
➔	CO	UT329	62	H1	mut/mut
➔	CO	UT261	62	H1	mut/mut
➔	AR	1625*	45	H1	mut/mut
➔	CO	UT300	62	H1	mut/mut
➔	CO	UT98	62	H1	mut/mut
➔	CO	UT70	62	H1	mut/mut
➔	CO	UT245	727	H1	mut/mut
➔	AR	986*	151	H1	mut/mut
➔	AR	1003*	2	H2	mut/mut
➔	AR	186*	2	H2	mut/mut
➔	AR	836	2	H2	mut/mut
➔	AR	M-6548	2	H2	mut/mut
➔	AR	1294	2	H2	mut/mut
➔	AR	484*	2	H2	mut/mut
➔	AR	938*	2	H2	mut/mut
➔	AR	P-410*	2	H2	mut/mut
	AR	457	33	LAM3	wt/wt
	CO	UT345	17	LAM2	wt/wt
	AR	M-6006	33	LAM3	wt/wt
	AR	1088*	33	LAM3	wt/wt
	AR	1710	33	LAM3	wt/wt
	AR	1531*	1354	LAM3	wt/wt
	AR	994*	33	LAM3	wt/wt
	AR	1241*	33	LAM3	wt/wt
	AR	1521*	105	U (LAM 3?)	wt/wt
	AR	1541*	20	LAM1	wt/wt
●	CO	UT272	20	LAM1	wt/wt
	AR	1023*	159	LAM (Tuscany)	wt/wt
	AR	1215	159	LAM (Tuscany)	wt/wt
	AR	1528*	105	U (LAM 3?)	wt/wt
	AR	810	159	LAM (Tuscany)	wt/wt
	CO	UT206	379	LAM (Tuscany) like	wt/wt
	CO	UT6	379	LAM (Tuscany) like	wt/wt
	AR	1489*	159	LAM (Tuscany)	wt/wt
	AR	1507*	159	LAM (Tuscany)	wt/wt
	AR	1239*	159	LAM (Tuscany)	wt/wt
	AR	1490*	1228	LAM (Tuscany)	wt/wt
	CO	UT189	42	LAM9	wt/wt
	AR	841*	42	LAM9	wt/wt
	AR	834*	42	LAM9	wt/wt
	CO	UT77	64	LAM6	wt/wt
	AR	1790	469	LAM9	wt/wt
	AR	1238*	177	LAM9	wt/wt
●	AR	1632*	290	LAM8	wt/wt
	AR	1894	93	LAM5	wt/wt
●	AR	1506*	20	LAM1	wt/wt
	AR	853*	37	T3	wt/wt
	AR	1537*	53	T1 Ghana	wt/wt
	AR	1549*	58	T5 MAD2	wt/wt
	AR	1928	58	T5 MAD2	wt/wt
	AR	2152*	1	Beijing	wt/wt
⊠	AR	1089**	262	H4	wt/wt
	AR	1522*	725	LAM5	wt/wt
	AR	1735*	725	LAM5	wt/wt
	AR	1323*	93	LAM5	wt/wt
	AR	1261*	93	LAM5	wt/wt
	AR	1092*	391	LAM4	wt/wt
	AR	1555*	391	LAM4	wt/wt
	AR	1091*	391	LAM4	wt/wt
	AR	1661*	391	LAM4	wt/wt
●	AR	1516*	4	LAM3/S convergent	wt/wt
	AR	846*	71	S	wt/wt
	AR	1699*	34	S	wt/wt
	AR	1499*	34	S	wt/wt
●	CO	UT89	20	LAM1	wt/wt
	AR	930	1547	T3	wt/wt

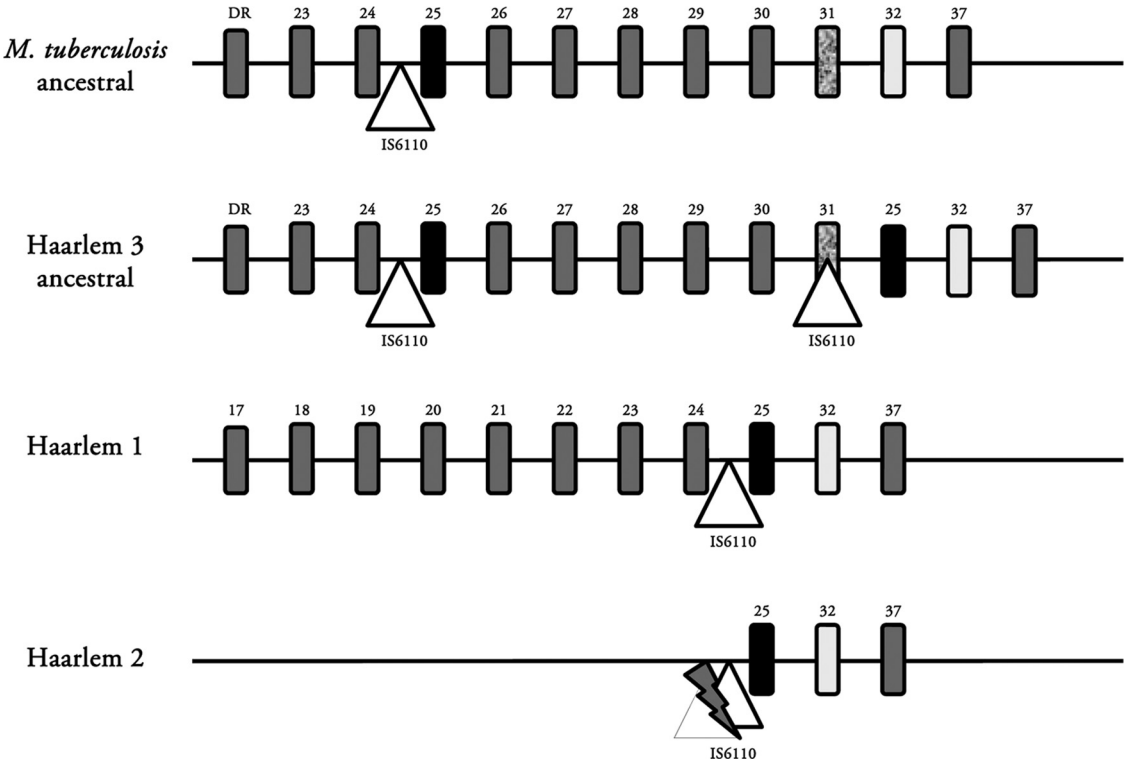


FIG. 3. Proposed evolution of the Haarlem DR locus. A schematic representation of hypothetical changes in the direct repeat (DR) locus of *Mycobacterium tuberculosis* Haarlem strains is shown. H3 probably arose due to a duplication of IS*6110*, together with spacer 25. A deletion, probably mediated by IS*6110* recombination, including 25 to 31 spacers generated H1. H2 could be the result of an ulterior large deletion of all spacers upstream of IS*6110*. Insertion of IS*6110* is shown as a triangle. Partial deletion of IS*6110* is shown as an arrow. Numbered bars represent consecutive spacers between DRs.

polymorphisms and also with the Haarlem branch in the RFLP IS*6110* dendrogram (Fig. 2). HSI1 and HSI2 result from the integration of the *IS*6110 element and are proposed to be the categorical markers of the RFLP *IS*6110 pattern observed in the Haarlem lineage; likewise, HSD4 encompasses the deletion of the DR spacers that give Haarlem strains their unique spoligotyping pattern. This observation demonstrates that the classification given by the polymorphisms reported here is linked with the most commonly used genotyping methods, showing the usefulness of these new markers in phylogenetic studies. The results presented here reinforce the idea that Haarlem is indeed a distinct phylogenetic group.

Another interesting result comes from the analysis of the DR region in Haarlem family strains. Based on our findings and the spoligotyping patterns, we propose the following scenario for the evolution of the Haarlem DR region (Fig. 3). The organization seen in H3 most probably arose due to a duplication of IS*6110*, together with spacer 25. The insertion of an extra copy of IS6110 was previously described (9, 16). A deletion, possibly mediated by IS*6110* recombination, could then generate the observed distribution of the locus in H1, and finally, the H2 DR organization could have arisen as the result of a spacer-mediated recombination that eliminated the 5′ region of the DR encompassing the first spacers up to the IS*6110*. An alternative explanation for the H3 DR locus organization could be that it resulted from a recombination event between different strains, as was previously suggested for *M. tuberculosis* (17). In addition to contributing to understanding the organization of the DR locus, our results also indicate that changes in this region, which is the basis for spoligotyping lineage assignment, correlate with changes in other regions of the genome, some of which may affect the physiology of the tubercle bacilli and contribute to the establishment and worldwide spread of successful lineages.

Increased resolution of the phylogeny of Euro-American lineages is needed to provide more accurate data for evolu-

FIG. 2. Dendrogram of clinical isolates. IS*6110* RFLP and spoligopatterns of 101 clinical isolates from Colombia (CO) and Argentina (AR), obtained between 1997 and 2005, and of laboratory strain H37Rv are shown. The IS*6110* RFLP dendrogram was constructed using arithmetic average linkages and the Dice coefficient with the software BioNumerics v 5.1 (Applied Maths, St-Martens-Latem, Belgium). Spoligo-shared types (SITs) and lineages were assigned according to the SITVIT database (http://www.pasteur-guadeloupe.fr:8081/SITVITDemo/). The 38 isolates sharing all eight polymorphisms here postulated as Haarlem specific are indicated with arrows, two isolates displaying six or seven of these polymorphisms are marked with stars, two isolates classified as Haarlem according to SITVIT and lacking these polymorphisms are indicated with crossed squares, and the five isolates marked with dots displayed only one of the eight Haarlem-specific polymorphisms.

tionary, epidemiological, and public health applications. In this respect, our findings fully support results of a recent SNP study (Christophe Sola, personal communication) indicating that some spoligotypes classified as Haarlem in SpolDB4 (3), especially those defined as H4, are not related to this family and that a more stringent definition is needed for this group. The Haarlem-specific mutations described here may be used to optimize a single-target PCR and/or to include the best-fitted target in multiplex assays aimed to classify strains into the main strain families. Indeed, studies associating distinct lineages with patient clinical and epidemiologic traits will improve our understanding of disease pathogenesis and improve current control measures, thus preventing further spread of epidemic strains.

A recent analysis of *M. tuberculosis* complex strains indicated that much of the observed genetic diversity has phenotypic consequences and that purifying selection is severely reduced in this highly clonal population, which suffers constant bottlenecks, produced when a single cell is enough to establish an infection (13). In this respect, the identification of genomic changes unique to the Haarlem lineage can provide a basis from which to begin to unravel some of the specific phenotypic characteristics that distinguish this particular genotype from the rest of the *M. tuberculosis* lineages.

ACKNOWLEDGMENTS

This work was supported by Colciencias grant 431-2004, the Colombian Center for Excellence in Tuberculosis Research (CCITB), CYTED grant 207RT0311, and grant FP7-HEALTH-2007-A-201690 from the EC.

REFERENCES

1. **Arcus, V. L., P. B. Rainey, and S. J. Turner.** 2005. The PIN-domain toxin-antitoxin array in mycobacteria. Trends Microbiol. **13:**360–365.
2. **Berrington, W. R., and T. R. Hawn.** 2007. *Mycobacterium tuberculosis*, macrophages, and the innate immune response: does common variation matter? Immunol. Rev. **219:**167–186.
3. **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, 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.
4. **Caws, M., G. Thwaites, S. Dunstan, T. R. Hawn, N. T. Lan, N. T. Thuong, K. Stepniewska, M. N. Huyen, N. D. Bang, T. H. Loc, S. Gagneux, D. van Soolingen, K. Kremer, M. van der Sande, P. Small, P. T. Anh, N. T. Chinh, H. T. Quy, N. T. Duyen, D. Q. Tho, N. T. Hieu, E. Torok, T. T. Hien, N. H. Dung, N. T. Nhu, P. M. Duy, N. van Vinh Chau, and J. Farrar.** 2008. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. PLoS Pathog. **4:**e1000034.
5. **Cubillos-Ruiz, A., J. Morales, and M. M. Zambrano.** 2008. Analysis of the genetic variation in *Mycobacterium tuberculosis* strains by multiple genome alignments. BMC Res. Notes **1:**110.
6. **Cui, T., L. Zhang, X. Wang, and Z. G. He.** 2009. Uncovering new signaling proteins and potential drug targets through the interactome analysis of *Mycobacterium tuberculosis*. BMC Genomics **10:**118.
7. **Dormans, J., M. Burger, D. Aguilar, R. Hernandez-Pando, K. Kremer, P. Roholl, S. M. Arend, and D. van Soolingen.** 2004. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different *Mycobacterium tuberculosis* genotypes in a BALB/c mouse model. Clin. Exp. Immunol. **137:**460–468.
8. **Duchene, V., S. Ferdinand, I. Filliol, J. F. Guegan, N. Rastogi, and C. Sola.** 2004. Phylogenetic reconstruction of *Mycobacterium tuberculosis* within four settings of the Caribbean region: tree comparative analyse and first appraisal on their phylogeography. Infect. Genet. Evol. **4:**5–14.
9. **Filliol, I., C. Sola, and N. Rastogi.** 2000. Detection of a previously unamplified spacer within the DR locus of *Mycobacterium tuberculosis*: epidemiological implications. J. Clin. Microbiol. **38:**1231–1234.
10. **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. U. S. A. **103:**2869–2873.
11. **Gagneux, S., and P. M. Small.** 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect. Dis. **7:**328–337.
12. **Heersma, H. F., K. Kremer, and J. D. van Embden.** 1998. Computer analysis of IS6110 RFLP patterns of *Mycobacterium tuberculosis*. Methods Mol. Biol. **101:**395–422.
13. **Hershberg, R., M. Lipatov, P. M. Small, H. Sheffer, S. Niemann, S. Homolka, J. C. Roach, K. Kremer, D. A. Petrov, M. W. Feldman, and S. Gagneux.** 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. **6:**e311.
14. **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.
15. **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.
16. **Legrand, E., I. Filliol, C. Sola, and N. Rastogi.** 2001. Use of spoligotyping to study the evolution of the direct repeat locus by IS6110 transposition in *Mycobacterium tuberculosis*. J. Clin. Microbiol. **39:**1595–1599.
17. **Liu, X., M. M. Gutacker, J. M. Musser, and Y. X. Fu.** 2006. Evidence for recombination in *Mycobacterium tuberculosis*. J. Bacteriol. **188:**8169–8177.
18. **Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen, and D. van Soolingen.** 2003. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. Clin. Exp. Immunol. **133:**30–37.
19. **Malik, A. N., and P. Godfrey-Faussett.** 2005. Effects of genetic variability of *Mycobacterium tuberculosis* strains on the presentation of disease. Lancet Infect. Dis. **5:**174–183.
20. **Marquina-Castillo, B., L. Garcia-Garcia, A. Ponce-de-Leon, M. E. Jimenez-Corona, M. Bobadilla-Del Valle, B. Cano-Arellano, S. Canizales-Quintero, A. Martinez-Gamboa, M. Kato-Maeda, B. Robertson, D. Young, P. Small, G. Schoolnik, J. Sifuentes-Osornio, and R. Hernandez-Pando.** 2009. Virulence, immunopathology and transmissibility of selected strains of *Mycobacterium tuberculosis* in a murine model. Immunology **128:**123–133.
21. **Mathema, B., N. E. Kurepina, P. J. Bifani, and B. N. Kreiswirth.** 2006. Molecular epidemiology of tuberculosis: current insights. Clin. Microbiol. Rev. **19:**658–685.
22. **Mawuenyega, K. G., C. V. Forst, K. M. Dobos, J. T. Belisle, J. Chen, E. M. Bradbury, A. R. Bradbury, and X. Chen.** 2005. *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. Mol. Biol. Cell **16:**396–404.
23. **McEvoy, C. R., A. A. Falmer, N. C. Gey van Pittius, T. C. Victor, P. D. van Helden, and R. M. Warren.** 2007. The role of IS6110 in the evolution of *Mycobacterium tuberculosis*. Tuberculosis (Edinb.) **87:**393–404.
24. **McLean, K. J., P. Carroll, D. G. Lewis, A. J. Dunford, H. E. Seward, R. Neeli, M. R. Cheesman, L. Marsollier, P. Douglas, W. E. Smith, I. Rosenkrands, S. T. Cole, D. Leys, T. Parish, and A. W. Munro.** 2008. Characterization of active site structure in CYP121. A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J. Biol. Chem. **283:**33406–33416.
25. **Olano, J., B. Lopez, A. Reyes, M. P. Lemos, N. Correa, P. Del Portillo, L. Barrera, J. Robledo, V. Ritacco, and M. M. Zambrano.** 2007. Mutations in DNA repair genes are associated with the Haarlem lineage of *Mycobacterium tuberculosis* independently of their antibiotic resistance. Tuberculosis (Edinb.) **87:**502–508.
26. **Palmero, D., V. Ritacco, M. Ambroggi, N. Marcela, L. Barrera, L. Capone, A. Dambrosi, M. di Lonardo, N. Isola, S. Poggi, M. Vescovo, and E. Abbate.** 2003. Multidrug-resistant tuberculosis in HIV-negative patients, Buenos Aires, Argentina. Emerg. Infect. Dis. **9:**965–969.
27. **Pandey, D. P., and K. Gerdes.** 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res. **33:**966–976.
28. **Rindi, L., N. Lari, and C. Garzelli.** 1999. Search for genes potentially involved in *Mycobacterium tuberculosis* virulence by mRNA differential display. Biochem. Biophys. Res. Commun. **258:**94–101.
29. **Ritacco, V., M. Di Lonardo, A. Reniero, M. Ambroggi, L. Barrera, A. Dambrosi, B. Lopez, N. Isola, and I. N. de Kantor.** 1997. Nosocomial spread of

human immunodeficiency virus-related multidrug-resistant tuberculosis in Buenos Aires. J. Infect. Dis. **176:**637–642.

30. **Sampson, S. L., R. M. Warren, M. Richardson, G. D. van der Spuy, and P. D. van Helden.** 1999. Disruption of coding regions by IS*6110* insertion in *Mycobacterium tuberculosis*. Tuber. Lung Dis. **79:**349–359.
31. **Schmidt, C. W.** 2008. Linking TB and the environment: an overlooked mitigation strategy. Environ. Health Perspect. **116:**A478–A485.
32. **Suchindran, S., E. S. Brouwer, and A. Van Rie.** 2009. Is HIV infection a risk factor for multi-drug resistant tuberculosis? A systematic review. PLoS One **4:**e5561.
33. **Thwaites, G., M. Caws, T. T. Chau, A. D'Sa, N. T. Lan, M. N. Huyen, S. Gagneux, P. T. Anh, D. Q. Tho, E. Torok, N. T. Nhu, N. T. Duyen, P. M. Duy, J. Richenberg, C. Simmons, T. T. Hien, and J. Farrar.** 2008. Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. J. Clin. Microbiol. **46:**1363–1368.
34. **Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Y. O. Goguet de la Salmoniere, K. Aman, M. Kato-Maeda, and P. M. Small.** 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc. Natl. Acad. Sci. U. S. A. **101:**4865–4870.
35. **van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, et al.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:** 406–409.
36. **WHO.** 2008. Global tuberculosis control. Surveillance, planning, financing. WHO report 2008. WHO, Geneva, Switzerland.