Journal of Clinical Microbiology

Global Study of IS*6110* in a Successful *Mycobacterium tuberculosis* Strain: Clues for Deciphering Its Behavior and for Its Rapid Detection

Maria Isabel Millán-Lou, Ana Isabel López-Calleja, Cristina Colmenarejo, Maria Antonia Lezcano, Maria Asunción Vitoria, Patricia del Portillo, Isabel Otal, Carlos Martín and Sofía Samper

J. Clin. Microbiol. **2013, 51(11):3631. DOI: 10.1128/JCM.00970-13.**
Published Ahead of Print 28 August 2013.

JCM
Journals.ASM.org

Global Study of IS*6110* in a Successful *Mycobacterium tuberculosis* Strain: Clues for Deciphering Its Behavior and for Its Rapid Detection

Maria Isabel Millán-Lou,[a,b] Ana Isabel López-Calleja,[a] Cristina Colmenarejo,[a] Maria Antonia Lezcano,[a,b] Maria Asunción Vitoria,[b,c] Patricia del Portillo,[e] Isabel Otal,[b,d] Carlos Martín,[b,d] Sofía Samper[a,b]

IIS Aragón, Hospital Universitario Miguel Servet, Zaragoza, Spain[a]; CIBER de Enfermedades Respiratorias, Zaragoza, Spain[b]; IIS Aragón, Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain[c]; Universidad de Zaragoza, Zaragoza, Spain[d]; CorpoGen, Bogotá, Colombia[e]

The *Mycobacterium tuberculosis* insertion sequence IS*6110*, besides being a very useful tool in molecular epidemiology, seems to have an impact on the biology of bacilli. In the present work, we mapped the 12 points of insertion of IS*6110* in the genome of a successful strain named *M. tuberculosis* Zaragoza (which has been referred to as the MTZ strain). This strain, belonging to principal genetic group 3, caused a large unsuspected tuberculosis outbreak involving 85 patients in Zaragoza, Spain, in 2001 to 2004. The mapping of the points of insertion of IS*6110* in the genome of the Zaragoza strain offers clues for a better understanding of the adaptability and virulence of *M. tuberculosis*. Surprisingly, the presence of one copy of IS*6110* was found in *Rv2286c*, as was recently described for a successful Beijing sublineage. As a result of this analysis, a rapid method for detecting this particular *M. tuberculosis* strain has been designed.

The airborne spread and long incubation period of tuberculosis (TB) make it difficult to prevent the transmission of the disease, which affects about 8 million people each year. Efficient molecular methods allow the genetic typing of *Mycobacterium tuberculosis*. In a population-based molecular study conducted in Zaragoza, Spain, from 2001 to 2004, the occurrence of a large TB outbreak involving 85 patients was detected. A change in the epidemiological pattern of TB was observed, with a single epidemic strain, named *M. tuberculosis* Zaragoza, causing 18.7% of all TB cases (1). The spoligotyping pattern SIT-773 presented by this strain was described in 5 cases geographically distributed in the United States and the Netherlands and was reported as an unclassified pattern in the updated database SITVITWEB (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE) (2). Some strains either are prevalent in different geographic areas or are more successfully transmitted in the population, as has been described for the Beijing or Haarlem family (3, 4). Relative to isolates of other genotypes, an *M. tuberculosis* Zaragoza isolate achieved the highest fitness ranking in the lungs of low-dose-aerosol-infected mice (5). However, the reasons for the dominance and spread of the Zaragoza strain in our population remain unclear.

The *M. tuberculosis* insertion sequence IS*6110* was first described in 1990 (6, 7). The IS*6110* sequence in the genome of *M. tuberculosis* has shown the stability required for use in molecular epidemiology. In some cases, when a specific location is detected, it has been used as a target to design a multiplex PCR method that allows rapid strain identification. For example, the IS*6110* located within the designated region NTF-1 (*Rv3128c* to *Rv3129*), *Rv2180c*, and *Rv2664* to *Rv2665* allowed identification of the W-multidrug-resistant Beijing strain, the GC1237 Beijing strain, and clone B0/W148 Beijing, respectively, to better control the transmission of specific epidemic strains (8–10).

Moreover, IS*6110* induces loss of gene activity either by mediating deletion events (11–16) or by disrupting coding sequences and regulatory domains (17, 18). The IS*6110* element itself modulates the expression of neighboring genes by acting as a promoter sequence, driving or enhancing their expression (19–22). All of these changes seem to have effects on the molecular structure of the mycobacterial genome, modifying the biology and fitness of the bacilli.

In the present work, we have characterized the successful Zaragoza strain and localized all IS*6110* copies in its genome. The study offers clues for better understanding the adaptability and virulence of *M. tuberculosis* and, in addition, has allowed us to design an accurate multiplex PCR assay that may be used for rapid detection and surveillance of this variant of *M. tuberculosis*.

MATERIALS AND METHODS

Characterization of the Zaragoza strain. The presence or absence of the TbD1 region was assessed by PCR amplification using the primers TBD1fla1-F and TBD1fla1-R (Table 1). TbD1 deletion results in an amplicon size of 400 bp, whereas the presence of the TbD1 region results in an amplicon size of 2,153 bp. Three single nucleotide polymorphisms (SNPs) described as markers of the *M. tuberculosis* Haarlem genotype strains, i.e., $mgtC^{182}$(Arg-His) (23), ogt^{44}(Thr-Ser), and ung^{501}(Leu-Leu) (24), were studied by PCR and DNA sequencing using the primers described in Table 1.

Automated DNA sequencing was performed by the DNA sequencing support service of the Spanish National Cancer Research Center (Madrid, Spain), by capillary electrophoresis with an Applied Biosystems 3730 DNA analyzer using the ABI Prism BigDye Terminator system (Applied Biosystems). The sequence generated was aligned with the sequence of *M. tuberculosis* H37Rv (http://genolist.pasteur.fr/TubercuList).

Mapping of IS*6110*. We used two different methodologies to identify the points of insertion of IS*6110* in the genome of the *M. tuberculosis* Zaragoza strain. Ligation-mediated PCR, as described previously by Prod'hom et al. (25), was used to amplify the flanking sequences of the IS*6110* of the DNA obtained from one isolate in 2004, considered the reference for the Zaragoza strain, and *Mycobacterium bovis* BCG, as a

Received 10 April 2013 Returned for modification 7 May 2013
Accepted 1 August 2013
Published ahead of print 28 August 2013
Address correspondence to Sofía Samper, ssamper.iacs@aragon.es.

doi:10.1128/JCM.00970-13

TABLE 1 Primers used in this study

Gene(s)	Primer	Sequence (5′ to 3′)	Reference or source
TbD1	TBD1fla1-F	CTA CCT CAT CTT CCG GTC CA	Brosch et al. (16)
	TBD1fla1-R	CAT AGA TCC CGG ACA TGG TG	Brosch et al. (16)
mgtC	mgtC-F	CGC CTA GGC TCA AAC TGC TG	Alix et al. (23)
	mgtC-R	CAA TAC CCG GCG GAT CTA CC	Alix et al. (23)
ogt	ogt-F	CCC AGC ACC TGT GGA CCA	Olano et al. (24)
	ogt-R	ACT CAG CCG CTC GCG AGC	Olano et al. (24)
ung	ung-F	GCT GGC AAT CGT TTG G	Olano et al. (24)
	ung-R	GGC AAC AAG AAG CGA CTC	Olano et al. (24)
	Salgd	TAG CTT ATT CCT CAA GGC ACG AGC	Prod'hom et al. (25)
	Salpt	TCG AGC TCG TGC	Prod'hom et al. (25)
	ISA1	CCT GAC ATG ACC CCA TCC TTT CC	Mendiola et al. (40)
	ISA3	GAG GCT GCC TAC TAC GCT CAA CG	Mendiola et al. (40)
Rv0755A, *thrV*	755-F	ATC TTG ATC GCA TCG GAA GC	This work
	755-R	AAC CCG CTG AGC AGC ATC GC	This work
Rv0794c, *Rv0797*	795-F	CCT ATC GCC CAC CAG ACG C	This work
	795-R	ACA CCG TGC GCC TCT ACC TGC	This work
Rv1668c, *Rv1869*	1668-F2	ACC CGC AAT CCG CTA CGC	This work
	1668-R2	AGC ACG AAT CAT GGA CTC GGC	This work
Rv2286c, *yjcE*	2286-F	AGC TCA TGA CCG ATC GCT GC	This work
	2286-R	ACC GCC GCG TTG ACC TGT TTC G	This work
Rv2349c (*plcC*)	2349-F	AAT GCC GAC GTC ATA TCG CC	Vera-Cabrera et al. (34)
	2349-R	ACG TCA TCA ACA ACA CGC TGC	Vera-Cabrera et al. (34)
Rv2813, *Rv2814* (DR region)	DR43-F	ACC CGG TGC GAT TCT GCG	This work
	DR22-R	AGA CGG CAC GAT TGA GAC	This work
Rv2823c	2823-F	AAG AGC TGT GCG GTC AGG	This work
	2823-R	AAG GTG ATC GAG GAG AAG TAC CGG C	This work
Rv3229c (*desA3*)	3229-F	AGC ATC TGC CGT AGG TAC CAC	This work
	3229-R	AAC ACC ATC CTT GCG ATC G	This work
MT1802, *MT1814* (RvD2, RvD3)	Cut1-F	ACG TAT CCG AGA GTG TGA C	This work
	RvD3-R	CAA TCA TGC TGT TTG GC	This work
PPE38	PPE38-F	AAG TGC GGA TTT TCG GTG TGG	This work
	PPE38-R	ATC AAA CGA GGA CGC CGA GG	This work
MT2423 (RvD4)	RvD4-F2	GTT GCC GTT GTT TCC GTT ACC	This work
	RvD4-R2	GGT GGT GGT GGC GGC T	This work
MT3429, *MT3430* (RvD5)	MT3429-F	GCA ATC AGA ACG TCG GTG T	This work
	RvD5-R	GTA CCC GCA CCA CCT GCT	This work
gyrB	MTUB-F	TCG GAC GCG TAT GCG ATA TC	Herrera-León et al. (27)
	MTUB-R	ACA TAC AGT TCG GAC TTG CG	Herrera-León et al. (27)
Rv2823c	MTZ-RF	CGG GGC GGT TCA TTG GTG	This work
	2823c-R2	AGG TGA TCG AGG AGA AGT ACC G	This work

control. Briefly, the total genomic DNA was restricted with SalI (Boehringer Mannheim) and ligated to a linker containing a SalI restriction site. The oligonucleotides used as linkers and primers are indicated in Table 1. SalI was then used to digest the resulting template. This strategy was also applied to digested fragments of DNA separated by electrophoresis, in order to gain effective resolution with the technique. DNA was recovered in seven different size ranges through overnight electrophoresis in low-melting-point agarose gels (Bio-Rad), following the protocol described by the manufacturer (Fermentas Life Science). For PCR amplification, two sets of primers were used, (i) Salgd and ISA1, which amplified the DNA sequence next to the 5′ end of the IS*6110*, and (ii) Salgd and ISA3, which amplified the DNA sequence next to the 3′ end of the IS*6110* (Table 1). The PCR products obtained were purified using the GFX PCR DNA gel band purification kit (GE Healthcare), to be sequenced as described above. The sequence generated was aligned with the sequences of *M. tuberculosis* H37Rv (http://genolist.pasteur.fr/TubercuList) and CDC1551 (http://blast.ncbi.nlm.nih.gov).

In a second approach, four isolates of the Zaragoza strain, including the 2004 reference isolate, were analyzed in the context of a high-throughput survey of *in vivo* IS*6110* transposition in multiple *M. tuberculosis* genomes for the simultaneous identification of points of insertion of IS*6110* with Illumina sequencing technology. This new methodology was described by Reyes et al. (26).

Based on the results obtained with the two methods, new primers were designed to verify the points of insertion of IS*6110* (Table 1). Most of the expected PCR products were in the range of 1,500 to 2,000 bp if the IS*6110* was present and in the range of 300 to 600 bp if the IS*6110* was absent from the site of amplification. The PCR products were sequenced and aligned with the *M. tuberculosis* H37Rv genome with BLAST.

Design of a rapid method for detection of the Zaragoza strain. The points of insertion of IS*6110* detected in the Zaragoza strain that had not been previously reported as preferential loci for IS*6110* transposition were analyzed in 100 samples from a collection of isolates in the University of Zaragoza that had been isolated between 2005 and 2008. This collection included different strains selected on the basis of their diverse IS*6110* restriction fragment length polymorphism (RFLP) patterns. Separate PCRs were performed and, in cases in which the IS*6110* appeared in non-Zaragoza isolates, another point of insertion was tested (Table 1). Four isolates of the Zaragoza strain were included as positive controls. The copy inserted at point 3129520/25 of the genome of strain H37Rv (GenBank

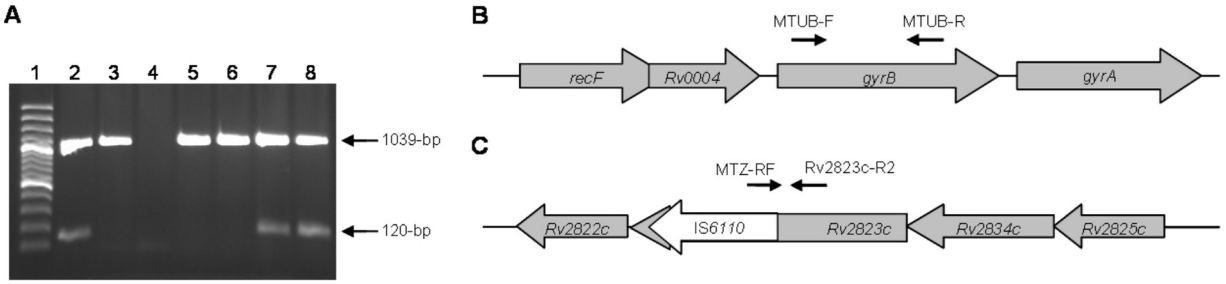

FIG 1 (A) Results of multiplex PCR visualized in agarose gel electrophoresis. Lane 1, molecular size marker (100-bp DNA ladder); lane 2, the *M. tuberculosis* Zaragoza strain (1,039 bp and 120 bp); lane 3, H37Rv strain (1,039 bp); lane 4, negative PCR control; lanes 5, 6, 7, and 8, different clinical isolates of *M. tuberculosis*. (B and C) Schematic representations of the targets used for the diagnostic test, i.e., the *gyrB* region in the *M. tuberculosis* complex (B) and the position and orientation of the IS*6110* insertion in the *Rv2823c* region in the Zaragoza strain (C). White arrow, IS*6110* element; arrowhead, direction of transcription of the putative transposase; small black arrows, primers.

accession no. NC_000962.2), in the *Rv2823c* gene, was selected as the specific polymorphism for detection of the Zaragoza strain.

A multiplex PCR detecting two different targets was designed for rapid identification of the *M. tuberculosis* Zaragoza genotype. The first target was the specific IS*6110* in *Rv2823c* in the Zaragoza strain. For this purpose, the primer MTZ-RF, annealing to the end of IS*6110* and the following sequence of the genome, and Rv2823c-R2 were designed. The second target was the *gyrB* gene, specific for *M. tuberculosis* complex strains and amplified with primers MTUB-F and MTUB-R, as described previously (27, 28) (Table 1). Multiplex PCR was performed in a total volume of 25 μl, containing 200 μM each deoxynucleoside triphosphate in 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 2.5 U of DNA polymerase (puReTaq Ready-to-go; GE Healthcare) in reaction buffer, 25 μM each primer, and 2 μl of template DNA. Temperature cycling conditions included 95°C for 5 min, 35 cycles of 94°C for 30 s, 65°C for 30 s, and 72°C for 90 s, and a final extension at 72°C for 10 min. The PCR products were analyzed by electrophoresis on 1.2% agarose gels. The molecular sizes of the multiplex PCR products were estimated by comparison with the migration distances of the DNA bands of a 100-bp DNA ladder. The Zaragoza strain was expected to produce two amplicons, i.e., 1,039-bp (*gyrB*) and 120-bp (IS*6110* in *Rv2823c* specific for the Zaragoza strain) products. A non-Zaragoza *M. tuberculosis* complex isolate was expected to produce only the 1,039-bp product. A non-*M. tuberculosis* complex strain was expected to produce no amplicon (Fig. 1).

Verification of multiplex PCR for detection of the *M. tuberculosis* Zaragoza strain. In order to probe the specificity of the multiplex PCR assay, we tested 122 positive stored cultures of *M. tuberculosis* isolated in the Hospital Universitario Miguel Servet in 2000. The strains were defrosted and recovered. The extraction of genomic DNA was done in two different ways. For 68 of the 122 isolates, DNA was extracted from solid cultures by the chemical lysis method (29). For the remaining 54 isolates, rapid DNA extraction was performed directly from the glycerol-stored samples. An aliquot of 200 μl was spin dried at 1,000 rpm for 15 min, and the supernatant was removed; 50 μl of Tris-EDTA (TE) buffer was added, and the mixture was heated at 100°C for 20 min. The sample was then spin dried at 1,000 rpm for 5 min, and the supernatant was used to amplify DNA by the multiplex PCR. In addition, these strains were analyzed by spoligotyping with a commercially available kit, according to the manufacturer's instructions (Ocimum Biosolutions Ltd., Hyderabad, India), as described by Kamerbeek et al. (30).

RESULTS

Classification of the Zaragoza strain. The TbD1 region was absent from the Zaragoza strain, and the $mgtC^{182}$(Arg-His), ogt^{44}(Thr-Ser), and ung^{501}(Leu-Leu) SNPs characteristic of the Haarlem genotype were not present, which allowed us to classify Zaragoza as a non-Haarlem but "modern" strain of *M. tuberculosis*.

Localization of the points of insertion of IS*6110*. The application of two different methodologies allowed us to identify the 12 points of insertion of IS*6110* in the *M. tuberculosis* Zaragoza genome (Table 2). Eight IS*6110* sites were localized by both methodologies, and four additional IS*6110* sites were found by the Illumina sequencing technology. The 12 insertion points were analyzed by PCR to confirm the sites of insertion, the presence or absence of direct repeat (DR) nucleotides at these points, and the IS*6110* orientation. Eleven of the 12 copies generated direct repeat sequences of 3, 4, 6, and 8 bp next to the insertion site.

There were six intragenic insertions, localized in *Rv2286c*, *plcC* (*Rv2349c*), *PPE38* (*Rv2352c*), *Rv2823c*, *desA3* (*Rv3229c*), and *MT2423* (RvD4 region). Five IS*6110* insertions were localized in different intergenic regions, at positions 393, 308, 13, 18, and 594 bp upstream of the *RV0794c*, *Rv1668c*, *Rv0755A*, *MT3429* (RvD5 region), and *Rv2813* genes, respectively (Table 2). The 12th IS*6110* site was located between *MT1802* and *MT1814* (RvD2 and RvD3 regions).

A deletion was found next to the point of insertion of a copy of IS*6110* in two cases; the first was in the highly conserved 36-bp DR region within a DR (direct copy) sequence of 36 bp. There was a duplication of an 8-bp direct repeat flanking this insertion point. The analysis of the DR region indicated an 863-bp deleted region, which included the 25th to 40th interspaces (see Fig. 2 for a detailed representation of this region). The second deletion was from the RvD2 to RvD3 regions of H37Rv, which resulted in the absence of the *cut1*, *wag22*, *Rv1760*, *Rv1761c*, *Rv1762c*, *Rv1763*, *Rv1764*, and *Rv1765c* genes. The 5′ end of the IS*6110* was inserted into the *cut1* gene, at the same insertion point as in the CDC1551 strain but in a different orientation. This region is included in the RD152 region characteristic of the Beijing lineage.

Selection of the *Rv2823c* gene as the specific target for rapid detection. We studied the 12 insertion points in order to select a specific target to develop a method for rapid detection of *M. tuberculosis* Zaragoza isolates. Three locations were ruled out as they were previously described as preferential loci for IS*6110* transposition, i.e., *Rv0794c* to *Rv0797*, *plcC* (*Rv2349c*), and the site in the DR region (31–34). Thus, we investigated the presence of IS*6110* in the other locations (*Rv0755A*, *Rv2286c*, *Rv2823c*, and *desA3* [*Rv3229c*]) in 30 isolates. Six of the isolates tested presented the IS*6110* in the same nucleotide position in the *Rv0755A-thrV*(*tRNA-Thr*) intergenic region as in the Zaragoza strain. Further, 100 isolates were tested for the *Rv2823c* point of insertion

TABLE 2 Distribution and characteristics of the 12 points of insertion of the copies of IS*6110* found in the *M. tuberculosis* Zaragoza strain[a]

Gene(s)	Insertion point position in the Zaragoza strain	No. of positive isolates/no. tested[b]	Testing	Orientation	DR	Intergenic	Distance to gene[c]
Rv0755A–thrV	850536–850539	6/30	Hot spot for strains analyzed in our control study	Direct	TGTC	Yes	13 bp to *Rv0755A* 103 bp to *thrV*
Rv0794c–Rv0797	889029–890376		Hot spot (17, 31); insertion similar to H37Rv	Direct	GAGG	Yes	393 bp to *Rv0794c* 12 bp to *Rv0797*
Rv1668c–Rv1669	1895651–1895654			Inverse	TAGG	Yes	308 bp to *Rv1668c* 75 bp to *Rv1669*
MT1802–MT1814	1985524–1995905			Direct	—[d]	No	471 bp to *MT1802* 93 bp to *MT1814.2*
Rv2286c	2559568–2559570	0/30		Inverse	ATC	No	758 bp to *Rv2285* 135 bp to *yjcE*
Rv2349c (*plcC*)	2627492–2627494	0/30	Hot spot (48)	Direct	TCA	No	512 bp to *Rv2348c* 1,289 bp to *plcB*
PPE38	2633977–2633979			Inverse	GCG	No	1,902 bp to *plcA* 551 bp to *PPE39*
MT2423	2633461–2633464			Inverse	TTTC	No	1,382 bp to *MT2422* 970 bp to *MT2423.1*
Rv2813–Rv2816c	3119660–3121887		Hot spot (17, 37)	Direct	GTCTGACG	Yes	624 bp to *Rv2813* 1,738 bp to *Rv2816c*
Rv2823c	3129520–3129525	0/100		Inverse	TTGTGA	No	173 bp to *Rv2822c* 2,250 bp to *Rv2824c*
Rv3229c (*desA3*)	3606308–3606310	0/30		Inverse	ACG	No	590 bp to *Rv3228* 804 bp to *Rv3230c*
MT3429–MT3430	3709089–3709091			Direct	GCC	Yes	18 bp to *MT3429* 197 bp to *MT3430*

[a] Results of the analysis to detect the specificity of the IS*6110* localization in 100 different clinical isolates are shown. The orientation of each IS*6110* copy with regard to the genome sequence was direct or inverse. The direct repeat nucleotide duplications generated at the points of insertion of IS*6110* and the distances to neighboring genes are detailed. The numbering was made with respect to the H37Rv sequence; when an insertion locus could not be mapped to H37Rv, it was referred to CDC1551.
[b] Different clinical isolates analyzed for detection of the specificity of the localization of IS*6110*.
[c] Distance from IS*6110* to neighboring genes.
[d] —, absence of direct repeat.

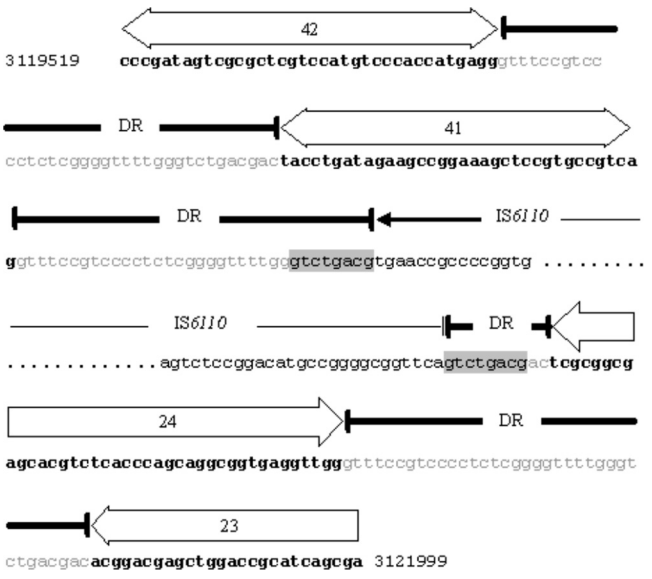

FIG 2 Map of the DR region of the *M. tuberculosis* Zaragoza strain at the point of insertion of IS*6110*. The numbering refers to the H37Rv genome. Black arrow, IS*6110* sequence; black bars, DRs; white arrows, spacer sequences (numbered with reference to the spoligotyping results); gray shading, eight direct repeat nucleotides at the insertion point.

(position 3129525 in H37Rv). All isolates showed the same size of amplified product as the control H37Rv, indicating that IS*6110* was not present (Fig. 1). Subsequently, this insertion was selected to design a rapid diagnostic test. The results are shown in Table 2. In addition, the 12 IS*6110* sites were present in four *M. tuberculosis* Zaragoza isolates used as positive controls, as expected.

Validation of multiplex PCR as a rapid diagnostic test for the Zaragoza strain. We analyzed 122 clinical *M. tuberculosis* complex isolates from 2000 by multiplex PCR and spoligotyping. The results for the detection of Zaragoza isolates by the two techniques were concordant for a total of 26 samples (Table 3). In all Zaragoza isolates, the 120-bp Zaragoza-specific amplicon was produced. In only nine Zaragoza isolates, for which DNA was extracted directly from glycerol-stored samples and might have been insufficient or degraded, the 1,039-bp *M. tuberculosis* complex-specific amplicon of the *gyrB* gene was not obtained (Table 3). This rapid multiplex PCR test achieved sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of 100% in identifying Zaragoza isolates.

DISCUSSION

A large unsuspected TB outbreak involving 85 patients in Zaragoza, Spain, in 2001 to 2004 was attributed to the *M. tuberculosis* Zaragoza strain (35). The lack of the TbD1 region described here confirms that Zaragoza is a modern strain, in agreement with our previous work showing that the strain belongs to principal genetic group 3 (PGG-3) (35). Other authors found that isolates belong-

TABLE 3 Results of the analysis of 122 isolates from 2000 by multiplex PCR rapid testing for detection of the *M. tuberculosis* Zaragoza strain and by spoligotyping

Extraction and testing[a]	No. of positive isolates of:	
	M. tuberculosis	Zaragoza strain
Extraction of genomic DNA		
Multiplex PCR of the Zaragoza strain	68	18
Spoligotyping	68	18
Rapid DNA extraction		
Multiplex PCR of the Zaragoza strain	45[b]	8[c]
Spoligotyping	54	8

[a] The two tests achieved sensitivities, specificities, PPVs, and NPVs of 100% in identifying *M. tuberculosis* Zaragoza isolates.
[b] The absence of the *gyrB* amplicon in the rapid extraction samples could be due to degradation or insufficient amounts of DNA.
[c] The 1,039-bp amplicon specific for the *M. tuberculosis* complex was not produced in one Zaragoza isolate.

ing to PGG-1 and PGG-2 are more frequently associated with clustered cases of TB (36), whereas the Zaragoza strain (fitted in PGG-3) was responsible for a widespread outbreak. In an attempt to better understand the Zaragoza strain, we proposed to determine the localization of all of the IS*6110* integration sites in its genome. Depending on the position of integration, different alterations in the bacteria can be produced, ranging from lethality to improved fitness (20–22).

Transposition is the more common mechanism used by IS*6110* to move along genomes, which leads to the generation of 3- to 4-bp direct repeats immediately flanking the IS*6110* sequences (7). The direct repeat nucleotides flanking the insertion site were present in 11 of 12 copies in the strain. As far as we know, this is the first time that 6-bp and 8-bp direct repeat sequences have been described.

Similar to most *M. tuberculosis* complex strains, the Zaragoza strain contains a copy of IS*6110* in the DR region, which is the clustered regularly interspaced short palindromic repeat (CRISPR) locus of *M. tuberculosis* (33, 37). Unlike H37Rv, *M. bovis* BCG, and *M. bovis* AF2122/97, the Zaragoza strain carries IS*6110* in this locus inserted at a different point, 3 nucleotides from the end of a direct repeat sequence. The analysis of the DR region indicated an 863-bp deletion that is reflected in its discriminative spoligotype, which lacks amplified spacers 24 to 40 (35). Since 8-nucleotide repeats flanking the insertion were detected, it seems that this copy was the consequence of transposition; therefore, we cannot determine whether IS*6110* is implicated in the process of deletion of the adjacent region (Fig. 2).

Genomic regions absent in the *M. tuberculosis* H37Rv genome named RvD regions were described, and some of them were apparently created by homologous recombination of two adjacent copies of IS*6110* (38). Only one copy of IS*6110* was not flanked by direct repeats in the *M. tuberculosis* Zaragoza strain. In this region, a rearrangement between two genomic regions absent in the *M. tuberculosis* H37Rv genome, named RvD2 and RvD3 (*MT1802* to *MT1814*), was detected. This could be interpreted as a recombination process involving neighboring copies of IS*6110* elements, which resulted in the loss of *wag22*, *Rv1760*, *Rv1761c*, *Rv1762c*, *Rv1763*, and *Rv1764* and interruption of the *cut1* gene, strongly suggesting that these genes are not essential for *M. tuberculosis*. This IS*6110*-mediated deletion was included in the RD152 region characteristic of the Beijing lineage (39). This result is in agreement with other studies indicating that strains with greater numbers of IS*6110* copies have lost genomic regions (16, 22). In contrast, there is no description of these events in strains with only few copies, as all IS*6110* copies are flanked by direct repeats (33, 40).

According to the distance and the orientation of IS*6110* (21), IS*6110* could have a transcriptional influence on neighboring genes. The possibility of such an influence on *Rv0797*, *MT1814.2*, *Rv1668c*, *Rv2823c*, and *MT3430* should be analyzed.

Five of the 12 insertion points were located in intergenic regions representing 41.6% of the known sites, i.e., the DR region, *Rv0755A* to *thrV*, *Rv0794c* to *Rv0797*, *Rv1668c* to *Rv1669*, and *MT3429* to *MT3430* (RvD5 region). According to other authors, these results suggest that transposition is relatively more frequent in noncoding sequences of the genome (17, 18, 33).

According to the intragenic points of insertion of the IS*6110*, we could assume that *Rv2286c*, *plcC* (*Rv2349c*), *PPE38* (*Rv2352c*), *Rv2823c*, *desA3* (*Rv3229c*), and *MT2423*, coding for a PPE family protein (RvD4 region), are naturally inactivated in the Zaragoza strain. Among them, at least three copies of IS*6110* interrupt different described operons, i.e., *Rv2349c* (*plcB* and *plcC*), *Rv2823c* (*Rv2824c* and *Rv2819c*), and *Rv3229c* (*Rv3230c* and *Rv3229c*) (41–43), and the potential transcriptional influence on these operons is an interesting point to be studied.

The insertion in the *desA3* gene caught our attention, as it is thought to be involved in lipid metabolism. Its product is a stearoyl-coenzyme A desaturase (DesA3) implicated in the synthesis of oleic acid and considered to be essential for intracellular growth in macrophages (44, 45). In addition, DesA3 is a target of the second-line antituberculosis drug isoxyl, which was used together with isoniazid in multiple-drug therapy in the 1960s (43, 44). Isoxyl inhibits the synthesis of oleic and mycolic acids (43) and was demonstrated to be effective against various clinical isolates of multidrug-resistant strains of *M. tuberculosis* (46). The finding of the insertion of IS*6110* in the *desA3* (*Rv3229c*) gene suggests that isoxyl would not be effective against the Zaragoza strain. However, our preliminary studies showed that the Zaragoza strain is susceptible (data not shown), suggesting that multiple targets are used by isoxyl. More studies should be carried out to elucidate this process.

The *plcABC* locus, described as a probable virulence factor (47), is known as a preferential site for IS*6110* transposition (48). Confirming this finding, the Zaragoza strain has three IS*6110* copies, in *plcC* (*Rv2349c*), *PPE38* (*Rv2352c*), and *MT2423* (RvD4 region), within a region of around 10 kb. Rv2286c and Rv2823c are conserved hypothetical proteins with unknown functions (http://genolist.pasteur.fr/TubercuList). Interestingly, one IS*6110* copy was found to be interrupting the *Rv2286c* gene in a recent study of a highly transmitted Beijing strain (22). IS*6110* had the same orientation in the Zaragoza strain and the Beijing strain just mentioned, but the insertion points differed by 64 nucleotides (position 2559504 in the Beijing strain versus position 2559568 in the Zaragoza strain). The possibility exists that this copy of IS*6110* could be affecting regulation of the transcription of the adjacent gene *yjcE* (*Rv2287*). The IS*6110* location in the *Rv2823c* gene was specific for the *M. tuberculosis* Zaragoza strain, and the region was chosen to design the multiplex PCR for identification of the strain.

The use of whole-genome sequencing with last-generation systems shows difficulties in localizing the repeated sequences in the genome, such as IS*6110* sequences in the *M. tuberculosis* complex

genome. Nevertheless, knowledge of the specific locations of these insertion sequences has allowed the design of rapid techniques to identify specific genotypes of the W Beijing strain (8–10). Another objective of this study was to develop a rapid diagnostic method based on a specific point of insertion of IS*6110* to identify *M. tuberculosis* Zaragoza isolates. The specific target chosen, IS*6110* inserted in *Rv2328c*, was not found in any other isolates tested. We could verify its specificity in a study that analyzed the insertion points of IS*6110* in 570 strains (data not shown), although another point of insertion in the *Rv2823c* gene was described for 9 strains belonging to the Latin American-Mediterranean 5 family (26). During the selection of the point of insertion of IS*6110* to be used as the target, the *Rv0755A-thrV*(*tRNA-Thr*) intergenic region was found in the same nucleotide position in 6 other isolates. Surprisingly, these 6 different isolates were classified as family T by their spoligotyping patterns, which raises the possibility that these strains are epidemiologically or evolutionally related.

Application of the diagnostic test should make possible the rapid identification of patients who have been infected with the Zaragoza strain within a brief period. The multiplex PCR developed as part of this study could identify 26 Zaragoza isolates (21.3%) among the 122 isolates from 2000 that were analyzed, showing that the outbreak was already ongoing at that time. All of those isolates demonstrated amplification of the specific target of the *M. tuberculosis* Zaragoza genotype. We confirmed these results by comparison with spoligotyping results, supporting the specificity of the test. Currently, spoligotyping and IS*6110* RFLP assays are performed routinely for all *M. tuberculosis* complex isolates in the region; therefore, the incidence of this strain is being monitored, which allows us to verify that spoligotype SIT-773 has remained specific for the Zaragoza strain in our population to 2012 (data not shown). This multiplex PCR is a quick easy technique with high specificity, which can be used without delay for a single sample in the laboratory.

We conclude that IS*6110* transposition in the Zaragoza strain may be a driving force in adaptation to the human host, especially for sites that do not correspond to hot spots for other *M. tuberculosis* strains, including interruption of the *desA3* gene (*Rv3229c*), which is considered a target for new drugs against tuberculosis. The similarity of the Zaragoza strain to Beijing lineage strains with respect to the deletion between the RvD2 and RvD3 regions and the insertion of IS*6110* in *Rv2286c*, which was described for the highly transmitted Beijing strain although at a different insertion point, could provide a clue for the success of this strain in our population. Finally, the specific site found in *Rv2823c*, which was not present in any other strain tested, has allowed us to use it for identification of the Zaragoza strain.

ACKNOWLEDGMENTS

We thank Dessislava Marinova for her kind help in English revision of and her valuable comments on the manuscript.

This work was supported by the Instituto de Salud Carlos III (grants FIS09/051 and FIS12/1970). María Isabel Millán-Lou was financially supported by the Instituto de Salud Carlos III (grant CM09/000123).

REFERENCES

1. **López-Calleja AI, Lezcano MA, Vitoria MA, Iglesias MJ, Cebollada A, Lafoz C, Gavin P, Aristimuño L, Revillo MJ, Martin C, Samper S.** 2007. Genotyping of *Mycobacterium tuberculosis* over two periods: a changing scenario for tuberculosis transmission. Int. J. Tuberc. Lung Dis. **11:**1080–1086.
2. **Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N.** 2012. SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect. Genet. Evol. **12:**755–766.
3. **Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, Allix C, Aristimuño L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes HM, Guttierez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Mokrousov I, Narvskaïa O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofo-Razanamparany V, Rasolonavalona T, Rossetti ML, Rüsch-Gerdes S, Sajduda A, Samper S, Shemyakin IG, Singh UB, Somoskovi A, Skuce RA, van Soolingen D, Streicher EM, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren RM, Yap SF, Zaman K, Portaels F, Rastogi N, Sola C.** 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. **6:**23. doi:10.1186/1471-2180-6-23.
4. **Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN.** 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. **10:**45–52.
5. **Caceres N, Llopis I, Marzo E, Prats C, Vilaplana C, de Viedma DG, Samper S, Lopez D, Cardona PJ.** 2012. Low dose aerosol fitness at the innate phase of murine infection better predicts virulence amongst clinical strains of *Mycobacterium tuberculosis*. PLoS One **7:**e29010. doi:10.1371/journal.pone.0029010.
6. **McAdam RA, Hermans PW, van Soolingen D, Zainuddin ZF, Catty D, van Embden JD, Dale JW.** 1990. Characterization of a *Mycobacterium tuberculosis* insertion sequence belonging to the IS3 family. Mol. Microbiol. **4:**1607–1613.
7. **Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, Guesdon JL.** 1990. IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res. **18:**188.
8. **Plikaytis BB, Marden JL, Crawford JT, Woodley CL, Butler WR, Shinnick TM.** 1994. Multiplex PCR assay specific for the multidrug-resistant strain W of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **32:**1542–1546.
9. **Millán-Lou MI, Alonso H, Gavín P, Hernández-Febles M, Campos-Herrero MI, Copado R, Cañas F, Kremer K, Caminero JA, Martín C, Samper S.** 2011. Rapid test for identification of a highly transmissible *Mycobacterium tuberculosis* Beijing strain of sub-Saharan origin. J. Clin. Microbiol. **50:**516–518.
10. **Mokrousov I, Narvskaya O, Vyazovaya A, Otten T, Jiao WW, Gomes LL, Suffys PN, Shen AD, Vishnevsky B.** 2012. Russian "successful" clone B0/W148 of *Mycobacterium tuberculosis* Beijing genotype: a multiplex PCR assay for rapid detection and global screening. J. Clin. Microbiol. **50:**3757–3759.
11. **Fang Z, Doig C, Morrison N, Watt B, Forbes KJ.** 1999. Characterization of IS1547, a new member of the IS900 family in the *Mycobacterium tuberculosis* complex, and its association with IS6110. J. Bacteriol. **181:**1021–1024.
12. **Sampson SL, Warren RM, Richardson M, Victor TC, Jordaan AM, van der Spuy GD, van Helden PD.** 2003. IS6110-mediated deletion polymorphism in the direct repeat region of clinical isolates of *Mycobacterium tuberculosis*. J. Bacteriol. **185:**2856–2866.
13. **Sampson SL, Richardson M, Van Helden PD, Warren RM.** 2004. IS6110-mediated deletion polymorphism in isogenic strains of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **42:**895–898.
14. **Yesilkaya H, Dale JW, Strachan NJ, Forbes KJ.** 2005. Natural transposon mutagenesis of clinical isolates of *Mycobacterium tuberculosis*: how many genes does a pathogen need? J. Bacteriol. **187:**6726–6732.
15. **Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST.** 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol. Microbiol. **32:**643–655.
16. **Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. U. S. A. **99:**3684–3689.
17. **Sampson S, Warren R, Richardson M, van der Spuy G, van Helden P.**

2001. IS6110 insertions in *Mycobacterium tuberculosis*: predominantly into coding regions. J. Clin. Microbiol. **39:**3423–3424.
18. **Sampson SL, Warren RM, Richardson M, van der Spuy GD, van Helden PD.** 1999. Disruption of coding regions by IS6110 insertion in *Mycobacterium tuberculosis*. Tuber. Lung Dis. **79:**349–359.
19. **Beggs ML, Eisenach KD, Cave MD.** 2000. Mapping of IS6110 insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **38:**2923–2928.
20. **Soto CY, Menéndez MC, Pérez E, Samper S, Gómez AB, García MJ, Martín C.** 2004. IS6110 mediates increased transcription of the *phoP* virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J. Clin. Microbiol. **42:**212–219.
21. **Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, Howard ST.** 2004. IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. Mol. Microbiol. **52:**999–1012.
22. **Alonso H, Aguilo JI, Samper S, Caminero JA, Campos-Herrero MI, Gicquel B, Brosch R, Martín C, Otal I.** 2011. Deciphering the role of IS6110 in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. Tuberculosis (Edinb.) **91:**117–126.
23. **Alix E, Godreuil S, Blanc-Potard AB.** 2006. Identification of a Haarlem genotype-specific single nucleotide polymorphism in the *mgtC* virulence gene of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **44:**2093–2098.
24. **Olano J, López B, Reyes A, Lemos MP, Correa N, Del Portillo P, Barrera L, Robledo J, Ritacco V, Zambrano MM.** 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.
25. **Prod'hom G, Guilhot C, Gutierrez MC, Varnerot A, Gicquel B, Vincent V.** 1997. Rapid discrimination of *Mycobacterium tuberculosis* complex strains by ligation-mediated PCR fingerprint analysis. J. Clin. Microbiol. **35:**3331–3334.
26. **Reyes A, Sandoval A, Cubillos-Ruiz A, Varley KE, Hernández-Neuta I, Samper S, Martín C, Garcia MJ, Ritacco V, López L, Robledo J, Zambrano MM, Mitra RD, Del Portillo P.** 2012. IS-seq: a novel high throughput survey of in vivo IS6110 transposition in multiple *Mycobacterium tuberculosis* genomes. BMC Genomics **13:**249. doi:10.1186/1471-2164-13-249.
27. **Herrera-León L, Molina T, Saíz P, Sáez-Nieto JA, Jiménez MS.** 2005. New multiplex PCR for rapid detection of isoniazid-resistant *Mycobacterium tuberculosis* clinical isolates. Antimicrob. Agents Chemother. **49:**144–147.
28. **Kasai H, Ezaki T, Harayama S.** 2000. Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences. J. Clin. Microbiol. **38:**301–308.
29. **Somerville W, Thibert L, Schwartzman K, Behr MA.** 2005. Extraction of *Mycobacterium tuberculosis* DNA: a question of containment. J. Clin. Microbiol. **43:**2996–2997.
30. **Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. **35:**907–914.
31. **Fang Z, Forbes KJ.** 1997. A *Mycobacterium tuberculosis* IS6110 preferential locus (*ipl*) for insertion into the genome. J. Clin. Microbiol. **35:**479–481.
32. **McEvoy CR, Falmer AA, Gey van Pittius NC, Victor TC, van Helden PD, Warren RM.** 2007. The role of IS6110 in the evolution of *Mycobacterium tuberculosis*. Tuberculosis (Edinb.) **87:**393–404.
33. **Otal I, Isabel O, Gomez AB, Kremer K, de Haas P, García MJ, Martín C, van Soolingen D.** 2008. Mapping of IS6110 insertion sites in *Mycobacterium bovis* isolates in relation to adaptation from the animal to human host. Vet. Microbiol. **129:**333–341.
34. **Vera-Cabrera L, Hernández-Vera MA, Welsh O, Johnson WM, Castro-Garza J.** 2001. Phospholipase region of *Mycobacterium tuberculosis* is a preferential locus for IS6110 transposition. J. Clin. Microbiol. **39:**3499–3504.
35. **López-Calleja AI, Gavín P, Lezcano MA, Vitoria MA, Iglesias MJ, Guimbao J, Lázaro MA, Rastogi N, Revillo MJ, Martín C, Samper S.** 2009. Unsuspected and extensive transmission of a drug-susceptible *Mycobacterium tuberculosis* strain. BMC Pulm. Med. **9:**3. doi:10.1186/1471-2466-9-3.
36. **Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. U. S. A. **94:**9869–9874.
37. **Hermans PW, van Soolingen D, Bik EM, de Haas PE, Dale JW, van Embden JD.** 1991. Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. Infect. Immun. **59:**2695–2705.
38. **Brosch R, Philipp WJ, Stavropoulos E, Colston MJ, Cole ST, Gordon SV.** 1999. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. Infect. Immun. **67:**5768–5774.
39. **Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere YO, Kreiswirth BN, Van Soolingen D, Small PM.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **43:**3185–3191.
40. **Mendiola MV, Martín C, Otal I, Gicquel B.** 1992. Analysis of the regions responsible for IS6110 RFLP in a single *Mycobacterium tuberculosis* strain. Res. Microbiol. **143:**767–772.
41. **Viana-Niero C, de Haas PE, van Soolingen D, Leão SC.** 2004. Analysis of genetic polymorphisms affecting the four phospholipase C (*plc*) genes in *Mycobacterium tuberculosis* complex clinical isolates. Microbiology **150:**967–978.
42. **Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P.** 2007. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. Mol. Biol. Evol. **24:**1861–1871.
43. **Chang Y, Fox BG.** 2006. Identification of Rv3230c as the NADPH oxidoreductase of a two-protein DesA3 acyl-CoA desaturase in *Mycobacterium tuberculosis* H37Rv. Biochemistry **45:**13476–13486.
44. **Chang Y, Wesenberg GE, Bingman CA, Fox BG.** 2008. In vivo inactivation of the mycobacterial integral membrane stearoyl coenzyme A desaturase DesA3 by a C-terminus-specific degradation process. J. Bacteriol. **190:**6686–6696.
45. **Sassetti CM, Boyd DH, Rubin EJ.** 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. **48:**77–84.
46. **Phetsuksiri B, Jackson M, Scherman H, McNeil M, Besra GS, Baulard AR, Slayden RA, DeBarber AE, Barry CE, Baird MS, Crick DC, Brennan PJ.** 2003. Unique mechanism of action of the thiourea drug isoxyl on *Mycobacterium tuberculosis*. J. Biol. Chem. **278:**53123–53130.
47. **Kumar M, Khan FG, Sharma S, Kumar R, Faujdar J, Sharma R, Chauhan DS, Singh R, Magotra SK, Khan IA.** 2011. Identification of *Mycobacterium tuberculosis* genes preferentially expressed during human infection. Microb. Pathog. **50:**31–38.
48. **Viana-Niero C, Rodriguez CA, Bigi F, Zanini MS, Ferreira-Neto JS, Cataldi A, Leão SC.** 2006. Identification of an IS6110 insertion site in *plcD*, the unique phospholipase C gene of *Mycobacterium bovis*. J. Med. Microbiol. **55:**451–457.