NOTES

Detection of a Previously Unamplified Spacer within the DR Locus of *Mycobacterium tuberculosis*: Epidemiological Implications

INGRID FILLIOL, CHRISTOPHE SOLA, AND NALIN RASTOGI*

Unité de la Tuberculose et des Mycobactéries, Institut Pasteur, F-97165 Pointe à Pitre Cedex, Guadeloupe

Spoligotyping, a method based on the variability of distribution of the 43 inter-direct repeat (DR) spacers of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG, is useful to study the molecular epidemiology of bovine and human tuberculosis. Recently, a major family of *M. tuberculosis* clinical isolates named the Haarlem family, which did not contain spacers 31 and 33 to 36, was reported in a multicenter study. Independently, a data bank containing all the published spoligotypes showed that the two most prevalent spoligotypes in the world differed only by the presence or absence of spacer 31. A careful analysis of the DR locus sequence led us to hypothesize that spacer 31 may not have been amplified in some isolates with the primer sets DRa and DRb currently used for spoligotyping. Consequently, a modified spoligotyping method based on different combinations of the 36-bp DR and IS*6110* primers was devised that was able to discriminate between the left and the right parts of the DR locus and demonstrated the presence of the previously unamplified spacer 31 for some of the clinical isolates. By analogy, we suggest that a single-spacer difference in some epidemiologically linked cases of tuberculosis may simply arise due to the insertion of an extra copy of IS*6110* within the DR locus, leading to its asymmetrical disruption and subsequent lack of the DRa or DRb targets. The influence of the IS*6110* preferential insertion sites within the DR locus on spoligotyping results should be further investigated.

Originally described in *Mycobacterium bovis* (12), the direct repeat (DR) locus of the *Mycobacterium tuberculosis* complex genome consists of a variable number of copies of a 36-bp DR interspersed with specific spacer sequences, which has been used to develop a rapid molecular fingerprinting technique called spoligotyping (14), often associated with the IS*6110* restriction fragment-length polymorphism (RFLP) for epidemiological purposes (9, 10). A recent multicenter study involving comparison of various molecular markers for the study of epidemiology of *M. tuberculosis* concluded that, along with other techniques, spoligotyping was a method of choice for reproducible PCR-based genotyping (15). In parallel, this study also underlined that several separate genotype families within the *M. tuberculosis* complex could be recognized on the basis of multiple genetic markers, namely, the Beijing, Africa, and Haarlem families (15). A careful examination of these families showed that the Haarlem family was always characterized by the lack of spacer 31. This observation, along with the fact that epidemiologically related *M. tuberculosis* clinical isolates may sometimes differ by a single spacer (14), and our independent observations that the two most prevalent spoligotypes around the world (named types 53 and 50 in Fig. 2) differ by the presence or absence of spacer 31 (18), prompted us to have a closer look at this spacer by using a modified spoligotyping method.

Bacterial isolates. All of the experimentally studied *M. tuberculosis* isolates were obtained from the collection of the Pasteur Institute of Guadeloupe and were isolated and identified locally by standard mycobacteriological procedures.

Spoligotyping. Standard spoligotyping was performed with the DRa and DRb primers as described previously (14). For the left-right (L-R) spoligotyping, primers IS6 and IS3 (4, 5) were used in combination with biotinylated DRb or biotinylated DRa to amplify either the left or the right part of the IS*6110* boundaries or the central part of the DR locus for strains harboring two IS*6110* copies within the DR locus. The following primers were used: DRa, GGTTTTGGGTCTGACGAC (14); DRb, CCGAGAGGGGACGGAAAC (14); IS3, GCTGCCTACTACGCTCAAC (4, 5); and IS6, CAAGTAGACGGGCGACCTC (4, 5). The primer combinations are shown in Table 1. The PCR protocol and the conditions for membrane preparation, hybridization, and detection were the same as those reported earlier (14), except for the elongation time, which was set to 3 min.

IS*6110* and DRr hybridization. IS*6110* and DRr RFLPs were performed according to published procedures (12, 21) by either direct or indirect labeling. Detection by the enhanced chemoluminescence (ECL) systems was performed according to the supplier's protocols (Amersham, Buckinghamshire, United Kingdom).

Modified LR spoligotyping: basis for development and results. The spoligotyping results for *M. tuberculosis* isolates from our laboratory were pooled with those published previously to construct a database containing 993 individual patterns. The nomenclature of specific spoligotypes numbered 1 to 69 was recently described (18), to which new types 70 to 72 were added during the course of this investigation. The description of these spoligotypes is available on request. Statistical analysis of spacer distribution on these 993 isolates un-

* Corresponding author. Mailing address: Unité de la Tuberculose et des Mycobactéries, Institut Pasteur, Morne Jolivière, BP 484, F-97165 Pointe-à-Pitre Cedex, Guadeloupe. Phone: 590-893-881. Fax: 590-893-880. E-mail: rastogi@ipagua.gp.

TABLE 1. Primer combinations used in this study and expected results according to the number of IS*6110* copies inserted within the DR locus of *M. tuberculosis*

Primer set	Primer[a]		Result with given no. of IS*6110* copies	
	Forward	Reverse	1 copy	2 copies[b]
A[c]	DRb	Biot-DRa	All spacers present	All spacers present
B	Biot-DRb	IS3	No amplification	No amplification
C	IS3	Biot-DRa	Right DR spacers	Right and central DR spacers
D	Biot-DRb	IS6	Left DR spacers	Left and central DR spacers
E	IS6	Biot-DRa	No amplification	No amplification

[a] The directions of the forward and reverse primers are deduced from previous DR sequence data (11). Biot, biotinylated.

[b] If the two copies of IS*6110* are inserted in the same direction.

[c] Positive control.

derlined a region (spacers 21 to 32) that is characterized by successive decreases for some of the spacers, particularly spacer 31 (results not shown). Based on the available sequence of the DR locus (11), a physical map was redrawn for a strain containing two copies of IS*6110* (Fig. 1), with one copy being inserted after the spacer 31. As shown in Fig. 1, a second copy inserted after spacer 31 may lead to an unequal split of the DR, leading to the consequent loss of the DRa primer target, which in turn will not permit the amplification of spacer 31 with primers DRa and DRb, since there was no target for primer DRa at the right side of this spacer. Inversely, the bordering spacers 24 and 25 would be detected (Fig. 1). To experimentally verify this hypothesis, we searched in our collection for isolates lacking spacer 31 after routine spoligotyping (types 3, 50, and 67 [Fig. 2]). Type 50, which represents the second most frequent spoligotype in our database and is overrepresented in our region, differs only by the lack of spacer 31 as compared to type 53, another very common spoligotype (18). The *Pvu*II-DR RFLPs of these types of isolates lacking spacer 31 (spoligotype patterns 3, 50, and 67) mostly showed the presence of three DRr-hybridizing bands (Fig. 2), suggesting the presence of a second IS*6110* copy within the DR locus, very likely inserted asymmetrically after spacer 31 (Fig. 1). This assumption is strengthened by the observation that nearly 82% of the strains of type 53 (universal type; presence of all spacers except 33 to 36) (18) harbored only two DRr-hybridizing bands (Fig. 2). Assessment of the percentage of strains of spoligotype 50 har-

boring a second copy of IS*6110* inserted asymmetrically and located between spacers 31 and 25 of the DR locus is currently in progress.

The modified L-R spoligotyping method, which is theoretically able to selectively amplify the right or the left border of the IS*6110* sequence(s) of the DR locus, was initially tested on reference strains of *M. tuberculosis* (H37Rv) and *M. bovis* BCG (BCG-Pasteur), known to contain only a single IS*6110* copy linked to the DR. As expected, each of the reference strains was split into two distinct left and right spoligotyping patterns by L-R spoligotyping (Fig. 3 [data for BCG not shown]), confirming that the IS*6110* copy was indeed inserted immediately after spacer 24. Accordingly, the primer combinations C and D resulted in a strongly positive spoligotyping result in all cases, each distinctly covering half of the positive control as amplified with primer set A. Primer set B always gave negative results (Fig. 3), whereas primer set E produced no signals or weak and nonreproducible signals (results not shown). These results conclusively showed that primer sets C and D were able to discriminate the right and the left parts of the DR region in the reference *M. tuberculosis* and *M. bovis* BCG strains. However, due to the lesser sensitivity of the L-R spoligotyping technique (explained by the smaller amount of PCR targets [Fig. 1]), spacers located on the extremities of the DR locus were sometimes undetected (results not shown).

This experimental approach was later extended to investigate the DR loci of various clinical isolates lacking spacer 31
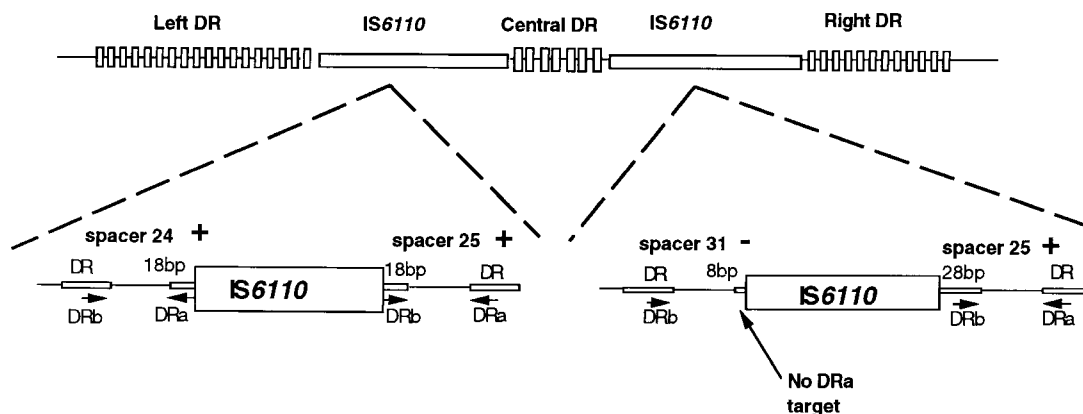

FIG. 1. Schematic representation of the DR locus of *M. tuberculosis* with an enlargement of the two IS*6110* insertion sites. The top of the figure shows the general organization of the DR locus, whereas the bottom of the figure is an enlargement showing the flanking regions of IS*6110* insertions. In the bottom part, the left-hand copy of IS*6110* is inserted exactly in the middle of the DRr sequence between spacers 24 and 25, splitting the DRr into two equal 18-bp fragments which serve as PCR targets for primers DRa and DRb and give a positive hybridization signal for spacers 24 and 25. On the other hand, an asymmetrical disruption of the DRr sequence due to the insertion of a second copy of IS*6110* in the right-hand side of the DR locus between spacers 31 and 25 results in two unequal parts in such a way that only an 8-bp fragment of the DRa target is conserved toward spacer 31, not enough for successful amplification and subsequent detection of this spacer.

Type number	Spoligotype	number of isolates	PvuII-DR-RFLP Results Number of DR bands		
			n = 1	n = 2	n = 3
53		27	1	22	4
50		10	0	0	10
67		2	0	0	2
3		1	0	1	0

FIG. 2. Correlation between spoligotypes and PvuII-DR RFLP results showing a link between the absence of spacer 31, the presence of a third DRr band, and the consequent presence of a second copy of IS6110 inserted within the DR locus.

that were typed in parallel by PvuII-DR RFLP; a total of 12 isolates selected comprised either strains found to harbor a single copy of IS6110 within the DR locus (types 2, 72, 62, and 47) or those found to possess two copies within the DR locus (types 50, 3, and var3; the latter differed from type 3 by absence of a single spacer, spacer 40). Representative experimental results for some of the isolates are shown in Fig. 3; for types 2, 72, 62, and 47, the primer sets C and D showed a split in position 24 resulting in distinct right and left amplifications of the DR locus starting from the point of insertion, whereas for types 50 and var3, an additional overlapping of spacers 25 to 31 by the primer set D corresponded to the insertion of the second IS6110 copy (Table 1 and Fig. 3). The fact that the left part of the DR pattern was absent in the type 2 isolate when

primer set D was used suggests that this portion was effectively missing and that a single copy of IS6110 was inserted immediately after spacer 31 (Fig. 1). Finally, the observation that previously undetected spacer 31 was revealed with primer set D in types 50 and var3 was in agreement with an asymmetric insertion of the second copy of IS6110 within the DR immediately after spacer 31 (Fig. 1 and 3).

DISCUSSION

In M. tuberculosis, various repetitive loci have been described, among which are numerous insertion sequences, such as IS6110 (20), as well as the variable number tandem repeats (8, 13), the DR (12), and the polymorphic GC-rich sequence

Spoligotype	Strain	Primer set	Type
	H37Rv	A	Control
		B	
		C	
		D	
	97024	A	2
		C	
		D	
	94019	A	72
		C	
		D	
	98063	A	62
		C	
		D	
	94041	A	47
		C	
		D	
	97120	A	Var3
		C	
		D	
	96139	A	50
		C	
		D	

FIG. 3. Schematic representation of results obtained by L-R spoligotyping of M. tuberculosis H37Rv and by the routine spoligotyping method of representative clinical isolates that lacked spacer 31. The primer sets A, B, C, and D have been detailed in the text and in Table 1.

(16). All of these markers have recently been used in a multi-center study of *M. tuberculosis* epidemiology (15). Spoligotyping based on the variability of inter-DR sequences has been proposed as a method of choice for reproducible PCR-based genotyping (15). In this context, the existence of preferential IS*6110* insertion sites in the DR locus of *M. tuberculosis* (4, 7) may indirectly contribute to the variability of spoligotyping results, which may have important epidemiological consequences. Moreover, clinical isolates of *M. tuberculosis* differing by a single DR spacer or a single IS*6110* copy that were either epidemiologically related (14) or were serial isolates from the same patient (17) have been reported. Consequently, we decided to investigate the fine structure of the DR locus in relation to the insertion of IS*6110* and its consequences in terms of spoligotyping results.

In this investigation, a modified spoligotyping method based on different combinations of DRr and IS*6110* primers was able to discriminate between the left and the right parts of the DR locus and demonstrated the presence of previously undetected spacer 31 for some of the clinical isolates studied. Although limited to spacer 31 in this study, this interpretation of a missing spacer provides us with a new interpretation for a single negative hybridization spot in spoligotyping. However, this interpretation of the absence of spacer 31 does not mean that all single-spacer absences are caused by an asymmetrical IS*6110* insertion: e.g., the lack of spacers 3, 9, and 16 for *M. bovis* BCG is not linked to any IS*6110* insertion (11).

The study of preferential IS*6110* transposition sites may have important applications in the study of the evolutionary genetics of *M. tuberculosis*, particularly because the DR locus is likely to evolve more slowly than IS*6110*, whose evolution has recently been shown to be too fast for the evolution of multidrug-resistant strains of *M. tuberculosis* to be monitored efficiently (1). A recent study has evaluated the rate of change of IS*6110* RFLP patterns in serial *M. tuberculosis* patient isolates as being 3.2 years (3). Seen from an evolutionary point of view, various mechanisms, such as the transposition of IS*6110* and other mobile genetic elements (2, 5, 19), homologous recombination (11), and slipped-strand mispairing (15), are believed to drive genomic diversification in *M. tuberculosis*. Our study underlines the role of IS*6110* transpositional events in the evolution of the DR locus. Alternatively, a one-spacer difference may also be due to mutations in the spacer sequence itself; however, if the DR locus is neutral (no environmental and/or selective pressure), point mutations in this locus should be extremely rare, as shown by the restricted allelic diversity shown in *M. tuberculosis* (19). More likely, IS*6110*-mediated deletions and insertions and homologous recombination may have driven the evolution of the DR locus (6, 11). Our experimental approach may increase the discriminatory power of spoligotyping to study isolates differing by a single missing spacer and thus result in a better evaluation of strain relatedness in epidemiological studies of tuberculosis.

REFERENCES

1. **Alito, A., N. Morcillo, S. Scipioni, A. Dolmann, M. I. Romano, A. Cataldi, and D. van Soolingen.** 1999. The IS*6110* restriction fragment length polymorphism in particular multidrug-resistant *Mycobacterium tuberculosis* strains may evolve too fast for reliable use in outbreak investigation. J. Clin. Microbiol. **37:**788–791.
2. **Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. M. Musser, and B. N. Kreiswirth.** 1996. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA **275:**452–457.
3. **De Boer, A. S., M. W. Borgdoff, P. E. de Haas, N. J. Nagelkerke, J. D. A. van Embden, and D. van Soolingen.** 1999. Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. J. Infect. Dis. **180:**1238–1244.
4. **Fang, Z., and K. J. Forbes.** 1997. A *Mycobacterium tuberculosis* IS*6110* preferential locus (*ipl*) for insertion into the genome. J. Clin. Microbiol. **35:**479–481.
5. **Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes.** 1998. IS*6110* transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. J. Bacteriol. **180:**2102–2109.
6. **Fang, Z., C. Doig, D. T. Kenna, N. Smittipat, P. Palittapongarnpim, B. Watt, and K. J. Forbes.** 1999. IS*6110*-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. J. Bacteriol. **181:**1014–1020.
7. **Fomukong, N., M. Beggs, H. E. Hajj, G. Templeton, K. Eisenach, and M. D. Cave.** 1998. Differences in the prevalence of IS*6110* insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS*6110*. Tuber. Lung Dis. **78:**109–116.
8. **Frothingham, R., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology **144:**1189–1196.
9. **Goguet de la Salmonière, Y.-O., H. M. Li, G. Torrea, A. Bunschoten, J. van Embden, and B. Gicquel.** 1997. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **35:**2210–2214.
10. **Goyal, M., N. A. Saunders, J. D. A. van Embden, D. B. Young, and R. J. Shaw.** 1997. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS*6110* restriction fragment length polymorphism. J. Clin. Microbiol. **35:**647–651.
11. **Groenen, P. M. A., A. E. Bunschoten, D. van Soolingen, and J. D. A. van Embden.** 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*, application for strain differentiation by a novel typing method. Mol. Microbiol. **10:**1057–1065.
12. **Hermans, P. W. M., D. van Soolingen, E. M. Bik, P. E. W. de Haas, J. W. Dale, and J. D. A. van Embden.** 1991. Insertion element IS*987* 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.
13. **Hermans, P. W. M., D. van Soolingen, and J. D. A. van Embden.** 1992. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. J. Bacteriol. **174:**4157–4165.
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. M. Hermans, C. Martín, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. A. 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. **Poulet, S., and S. T. Cole.** 1995. Characterization of the highly abundant polymorphic GC-rich repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. Arch. Microbiol. **163:**87–95.
17. **Rastogi, N., B. C. Ross, B. Dwyer, K. S. Goh, S. Clavel-Sérès, V. Jeantils, and P. Cruaud.** 1992. Emergence during unsuccessful chemotherapy of multiple drug resistance in a strain of *Mycobacterium tuberculosis*. Eur. J. Clin. Microbiol. Infect. Dis. **11:**901–907.
18. **Sola, C., A. Devallois, L. Horgen, J. Maïsetti, I. Filliol, E. Legrand, and N. Rastogi.** 1999. Tuberculosis in the Carribean: using spacer oligonucleotide typing to understand strain origin and transmission. Emerg. Infect. Dis. **5:**404–414.
19. **Sreevatsan, S., X. Pan, K. Stockbauer, N. Connell, B. Kreiswirth, T. Whittam, and J. Musser.** 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. USA **97:**9869–9874.
20. **Thierry, D., M. D. Cave, K. D. Eisenach, J. T. Crawford, J. H. Bates, B. Gicquel, and J. L. Guesdon.** 1990. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res. **18:**188.
21. **van Embden, J. D. A., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and P. M. Small.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:**406–409.