JOURNAL OF CLINICAL MICROBIOLOGY, Feb. 1997, p. 479–481 Vol. 35, No. 2
0095-1137/97/$04.00+0

A *Mycobacterium tuberculosis* IS*6110* Preferential Locus (*ipl*) for Insertion into the Genome

ZHIGANG FANG AND KEN J. FORBES*

Medical Microbiology, Aberdeen University, Foresterhill, Aberdeen AB25 2ZD, United Kingdom

Received 19 June 1996/Returned for modification 30 August 1996/Accepted 5 November 1996

A 267-nucleotide *Mycobacterium tuberculosis* genomic sequence (*ipl*, the IS*6110* preferential locus) which can harbor the insertion sequence IS*6110* at six alternative locations has been identified in some three-quarters of the isolates tested. Only one IS*6110* copy was observed at this locus in the *ipl*::IS*6110*$^+$-containing isolates tested, and all insertions had the same orientation. The implications of this finding for IS*6110* fingerprint typing methods is discussed in this work.

The insertion sequence IS*6110* is found in the *Mycobacterium tuberculosis* complex (3, 9) and is usually present in several copies which are well dispersed in the genome (10, 12). It is generally assumed that there is very little discrimination in the selection of a target sequence by IS*6110* upon transposition (2, 4, 11), and this fact has been utilized in the epidemiological characterization of strains by using IS*6110* fingerprint (restriction fragment length polymorphism [RFLP]) differences between strains (2, 10). Philipp et al. (7), however, have reported an apparently nonrandom distribution of this element on the physical map of *M. tuberculosis* H37Rv. One locus, the direct repeat region, has a high frequency of carriage of IS*6110* and has been proposed as a hot spot for integration of this element, although most of the copies are located at a single site (1). We report here another locus in which IS*6110* was found in six different locations in a 267-nucleotide (nt) genomic region in some three-quarters of the isolates examined.

Our study collection comprised 84 isolates of *M. tuberculosis* collected between August and December 1992, from every culture-positive patient in Scotland (6), and isolates of *M. tuberculosis* H37Ra, *Mycobacterium bovis* (clinical isolate 10772 and strain BCG-Pasteur), and *Mycobacterium africanum* (clinical isolate 110). An anchored polymerase chain reaction (PCR) was used to amplify genomic sequences flanking IS*6110* copies. Anchored PCR mixtures (final volume, 50 μl) comprised 200 ng of genomic DNA, 0.25 μM each primer, 50 mM KCl, 10 mM Tris-HCl (pH 8.0), 1.5 mM $MgCl_2$, 10% glycerol, and 0.5 U of *Taq* polymerase. The PCR mixtures were subjected to 40 cycles of amplification (1 min at 94°C, 2 min at 35°C, and 2 min at 72°C) following an initial denaturation at 94°C for 4 min and were completed by an extension of 8 min at 72°C. With M13 sequencing primer used as an arbitrary primer and the IS*6110*-specific primer IS6 as an anchored primer (Table 1), a 510-nt product was obtained from a clinical *M. tuberculosis* isolate (isolate 229). This 510-nt PCR product was sequenced by cycle sequencing and dye-labelled dideoxynucleotide chemistry on an ABI 373A automatic DNA sequencer and was found to comprise the expected right-hand end of IS*6110* adjacent to a genomic sequence (Fig. 1). The IS6 primer in combination with a new primer (FL3) designed from the genomic sequence allowed the amplification (as described above but at an annealing temperature of 55°C) of a 194-nt product from isolate 229 as anticipated. Subsequent testing of the whole collection with this primer pair identified other isolates with products of this length and also PCR-negative isolates which presumably lacked an IS*6110* copy at this locus. However, many isolates yielded PCR products of other lengths suggestive of the possibility of different IS*6110* insertion sites in this region. Sequencing of the longest of these products (isolate 91) allowed the design of a new primer, FL0, which was located in the genomic DNA just downstream from this particular IS*6110* copy. PCR amplification using primers FL0 and FL3 with DNA from strains without IS*6110* at this locus yielded a continuous sequence for this region (Fig. 1). This locus has been designated the IS*6110* preferential locus (*ipl*).

Systematic screening of the culture collection by PCR using primers FL3 and IS6 to detect the presence of IS*6110* in *ipl* and using product length to localize the positions of IS*6110* copies in *ipl* identified six different IS*6110* insertion sites. The exact location of the IS*6110* copies in isolates selected from each of these six groups was determined by PCR amplification (using IS6 with FL3 and IS3 with FL0) and sequencing of these products to locate the ends of the IS*6110* copies. These six different alleles were designated *ipl*1::IS*6110* to *ipl*6::IS*6110*. The insertion sites were all located in a 267-nt section of *ipl* (Fig. 1). Of the 84 *M. tuberculosis* clinical isolates, 62 (74%) carried an IS*6110* copy in *ipl*. The frequencies of the *ipl*::IS*6110* alleles in these isolates ranged from 1 (1%) to 19 (23%) (Table 2). Sequence analysis indicated that three of the putative *ipl*1::IS*6110* alleles had identical IS*6110* insertion sites; two putative *ipl*3::IS*6110* alleles also had identical insertion points. Not only were the points of insertion identical in each of these two groups of isolates, but so too were the duplicated nucleotides of the direct repeats at each end of the IS*6110* copies which were generated by the transpositional insertion of the IS*6110* element. Additional IS*6110* copies were not observed in the nine *ipl*::IS*6110* isolates in which the *ipl* locus was sequenced (Table 2). No occurrences of IS*6110* in the opposite orientation were detected, since all PCR products exhibited the expected lengths and were not increased in size by the addition of IS*6110* DNA.

The *ipl* sequence without IS*6110* was observed in all of the 22 *ipl*::IS*6110*-negative (*ipl*$^+$) *M. tuberculosis* isolates, since PCR with primers FL0 and FL3 gave a product of 372 nt as predicted. Similarly, *M. bovis* (isolate 10772), *M. bovis* BCG-Pasteur, and *M. africanum* (isolate 110) each also gave a 372-nt product, which comprised the *ipl*$^+$ sequence, with these primers. The *ipl* locus is therefore widespread in the *M. tuberculosis* complex. In all of these isolates, no PCR length polymorphisms in the 372 nt between primers FL0 and FL3 could be detected

* Corresponding author. Phone: 1224 663123, ext. 54953. Fax: 1224 685604.

TABLE 1. PCR primers used in this study

Primer	Sequence	Description
IS3	5′ GCTGCCTACTACGCTCAAC	nt 1272–1290 in IS*6110*[a]
IS6	5′ CAAGTAGACGGGCGACCTC	nt 162–179 in IS*6110*[a]
FL3	5′ GTGGCCTTGACGGTCTTCT	nt 355–373 in *ipl*[b]
FL0	5′ CCCTTTGCGTCTCAGTGTC	nt 2–20 in *ipl*[b]
M13	5′ GTAAAACGACGGCCAGT	M13 forward sequencing primer

[a] EMBL sequence accession number X17348 sequence starting from IS*6110* itself at nt 3 (9).
[b] EMBL sequence accession number X95799.

TABLE 2. Frequencies of *ipl*::IS*6110* alleles in 84 *M. tuberculosis* isolates

Allele	No. of copies (%) as determined by[a]:		Insertion site sequence[b]
	PCR	Sequencing	
*ipl*1::IS*6110*	19 (23)	3[c]	GTGAAGGAGGCAACCACCA
*ipl*2::IS*6110*	12 (14)	1	GTGTGTGCCGCGAGGTGGG
*ipl*3::IS*6110*	10 (12)	2	TGATAGGAGCGTGGCTTTC
*ipl*4::IS*6110*	10 (12)	1	GCCCTTTGCGTCT
*ipl*5::IS*6110*	10 (12)	1	GACCGGCCCAACCTCAACA
*ipl*6::IS*6110*	1 (1)	1	TCACCTGGTGTCTGGCTCA
ipl::IS*6110* negative	22 (26)	2	

[a] Number (frequency) of alleles in the collection as determined by PCR product length and by DNA sequencing across the ends of the IS*6110* copy.
[b] IS*6110* insertions resulted in direct repeats of the genomic sequence indicated (solid underlining), though in the case of *ipl*4::IS*6110* the length of this repeat was not determined.
[c] *ipl*1::IS*6110* was also noted in *M. tuberculosis* H37Ra in addition to being noted in the three isolates sequenced in this collection.

by high-resolution agarose (3% Metaphor Agarose; FMC) electrophoresis, suggesting that these isolates had not undergone any insertion or deletion events (such as the gain or loss of IS*6110* from this site, as the precise excision of IS*6110* is expected to be a rare event). By using appropriate IS*6110* and *ipl* primer combinations (IS6 and FL3; IS3 and FL0), *M. tuberculosis* isolates without a detected IS*6110* at this locus were tested by PCR for IS*6110* copies adjacent to this region, but none were found within the resolution of this technique (1 to 2 kb), suggesting that the *ipl* locus is restricted to the 267-nt region first noted, at least in this collection. While most of the isolates here were from British patients, a number were from patients from around the world, including *M. tuberculosis* H37Ra which carries *ipl*1::IS*6110* and was derived from an isolate collected in 1905 from an American patient (8), implying that these alleles are widespread.

The reason for the strong preference of IS*6110* to insert into this short region, *ipl*, of the *M. tuberculosis* chromosome is not clear. A number of features are of note however. (i) The preferred region is restricted to only 267 nt and does not seem

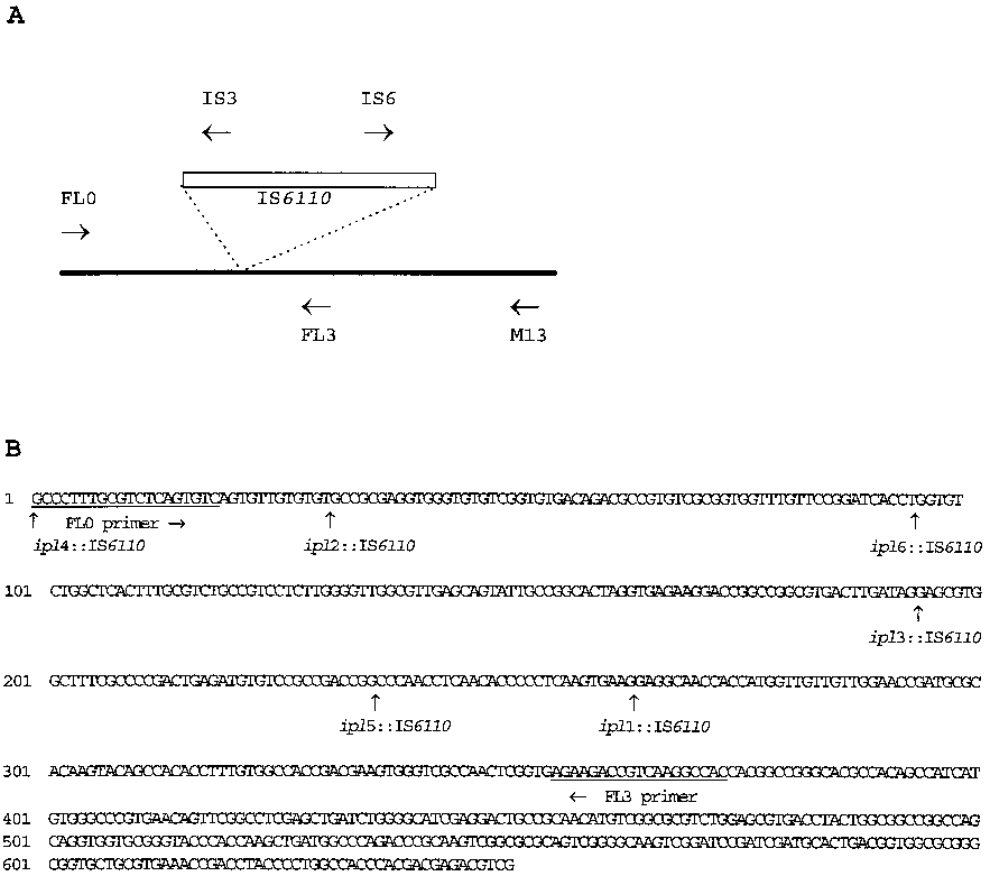

FIG. 1. DNA sequence of the *ipl* locus (EMBL accession number X59799) with the locations of IS*6110* insertions shown. (A) Schematic of locations of IS*6110*-specific primers IS3 and IS6 and genomic primers FL0 and FL3 and presumptive location of primer M13 in isolate 229. (B) IS*6110* insertions are indicated by vertical arrows (EMBL accession numbers X98148 to X98158), and primers FL0 and FL3 are indicated by underlining.

to extend beyond this, at least to a distance of 1 to 2 kb in either direction. Miller et al. (5) have noted a 25- to 50-fold preferential insertion of Tn*9*, a composite transposon which is flanked by IS*1* copies, into a 1-kb region of the *Escherichia coli lac* operon, so such localization is not without precedence. (ii) None of the isolates tested carried more than one copy of IS*6110* in *ipl*. (iii) The six insertions all had the same orientation with respect to the chromosome. It was not possible to determine whether *ipl* was in a coding region of the genome because of the short length of sequence that was available. However, it did contain a number of open reading frames, so it is possible that the influence of externally initiated transcription on the transpositional activity of IS*6110* is a governing feature of IS*6110* orientation. (iv) Although only 10 *ipl*::IS*6110* insertions have been sequenced, the insertions in the other *ipl*::IS*6110* isolates have been located by PCR (using high-resolution agarose electrophoresis), and these mapped to the six sites identified by sequencing (Table 2). The frequency of the six alleles in our strain collection varies greatly.

The most popular method at present for the differentiation of clinical *M. tuberculosis* isolates is *Pvu*II RFLP of IS*6110*-containing fragments (10). This method characterizes isolates with dissimilar RFLP patterns as unrelated and conversely characterizes isolates with similar patterns is related. If the location of IS*6110* copies in the genome is random, then the occurrence of similar RFLP patterns in unrelated isolates will be a rare event, since IS*6110* insertions could occur anywhere in the (comparatively large) genome (7). However, if IS*6110* insertion occurs in preferred regions of the genome, then the variety of RFLP patterns that can be generated will be reduced, especially if these regions are short, as in this study. In this case, the likelihood of unrelated isolates having coincidentally similar patterns would be increased. As the ratio of preferred to randomly selected IS*6110* insertion sites increases, so does the discriminatory power of IS*6110* RFLP typing decrease. It remains to be seen where the balance of these probabilities lies.

Nucleotide sequence accession numbers. The sequences described in this work have been assigned EMBL accession numbers X59799 and X98148 to X98158.

This work was supported by Aberdeen University's Research Committee and ACTR, Aberdeen Royal Hospitals-NHS Trust Endowment Fund, and benefited from the use of the SEQNET facility, CCL, Daresbury Laboratory, Warrington, United Kingdom.

We acknowledge the extensive assistance of B. Watt and the staff of SMRL, City Hospital, Edinburgh, United Kingdom.

REFERENCES

1. **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.
2. **Hermans, P. W. M., D. van Soolingen, J. W. Dale, A. R. J. Schuitema, R. A. McAdam, D. Catty, and J. D. A. van Embden.** 1990. Insertion element IS*986* from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. J. Clin. Microbiol. **28:**2051–2058.
3. **McAdam, R. A., P. W. Hermans, D. van Soolingen, Z. F. Zainuddin, D. Catty, J. D. A. van Embden, and J. W. Dale.** 1990. Characterization of a Mycobacterium tuberculosis insertion sequence belonging to the IS3 family. Mol. Microbiol. **4:**1607–1613.
4. **Mendiola, M. V., C. Martin, I. Otal, and B. Gicquel.** 1992. Analysis of the regions responsible for IS6110 RFLP in a single Mycobacterium tuberculosis strain. Res. Microbiol. **143:**767–772.
5. **Miller, J. H., M. P. Calos, D. Galas, M. Hofer, D. E. Buchel, and B. Muller-Hill.** 1980. Genetic analysis of transpositions in the lac region of Escherichia coli. J. Mol. Biol. **144:**1–18.
6. **Olson, E. S., K. J. Forbes, B. Watt, and T. H. Pennington.** 1995. Population genetics of Mycobacterium tuberculosis complex in Scotland analysed by pulsed-field gel electrophoresis. Epidemiol. Infect. **114:**153–160.
7. **Philipp, W. J., S. Poulet, K. Eiglmeier, L. Pascopella, V. Balasubramanian, B. Heym, S. Bergh, B. R. Bloom, W. R. Jacobs, Jr., and S. T. Cole.** 1996. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. Proc. Natl. Acad. Sci. USA **93:**3132–3137.
8. **Steenken, W., Jr., W. H. Oatway, Jr., and S. A. Petroff.** 1934. Biological studies of the tubercle bacillus. III. Dissociation and pathogenicity of the R and S variants of the human tubercle bacillus (H37). J. Exp. Med. **60:**515–540.
9. **Thierry, D., M. D. Cave, K. D. Eisenach, J. T. Crawford, J. H. Bates, B. Gicquel, and J. L. Guesdon.** 1990. IS6110, an IS-like element of Mycobacterium tuberculosis complex. Nucleic Acids Res. **18:**188.
10. **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.
11. **van Soolingen, D., P. W. Hermans, P. E. de Haas, D. R. Soll, and J. D. A. van Embden.** 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J. Clin. Microbiol. **29:**2578–2586.
12. **Zhang, Y., G. H. Mazurek, M. D. Cave, K. D. Eisenach, Y. Pang, D. T. Murphy, and R. J. Wallace, Jr.** 1992. DNA polymorphisms in strains of *Mycobacterium tuberculosis* analyzed by pulsed-field gel electrophoresis: a tool for epidemiology. J. Clin. Microbiol. **30:**1551–1556.