IS6110, a *Mycobacterium tuberculosis* Complex-Specific Insertion Sequence, Is Also Present in the Genome of *Mycobacterium smegmatis*, Suggestive of Lateral Gene Transfer among Mycobacterial Species

Abbie Coros, Erin DeConno, and Keith M. Derbyshire

Division of Infectious Disease, Wadsworth Center, New York State Department of Health, School of Public Health, University at Albany, and Department of Biomedical Sciences, Albany, New York 12201-2002

Received 3 January 2008/Accepted 26 February 2008

IS6110 is an insertion element found exclusively within the members of the *Mycobacterium tuberculosis* complex (MTBC), and because of this exclusivity, it has become an important diagnostic tool in the identification of MTBC species. The restriction of IS6110 to the MTBC is hypothesized to arise from the inability of these bacteria to exchange DNA. We have identified an IS6110-related element in a strain of *Mycobacterium smegmatis*. The presence of IS6110 indicates that lateral gene transfer has occurred among mycobacterial species, suggesting that the mycobacterial gene pool is larger than previously suspected.

Genetic exchange is thought to be a driving force behind the ability of bacterial species to evolve and adjust to environmental challenges. Lateral gene transfer (LGT) in bacteria is mediated by one of three processes, namely, conjugation, transformation, or transduction; examples of these processes have been described for almost all bacterial species (9, 19). By contrast, the *Mycobacterium tuberculosis* complex (MTBC) species, comprising *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. canneti*, *M. caprae*, *M. microti*, and *M. pinnipedi* (10, 15), are clonal populations evolved from a single progenitor species that has diversified by the acquisition of spontaneous mutations rather than by LGT. Genome comparisons between these seven species show that they have almost identical 16S rRNA sequences and highly similar genome sequences, and there is no strong evidence for genetic exchange (>99% identity [2, 3, 16]). The lack of clearly documented LGT among members of the MTBC is thought to be a consequence of the organisms’ solitary lifestyles within their hosts, preventing their contact with other mycobacterial species, or perhaps even other bacteria. Thus, it has become generally accepted within the scientific community that the MTBC species do not undergo genetic exchange (10, 15, 16).

IS6110 is an insertion element that is found exclusively within the MTBC; the assumption has been that this restriction is a result of the lack of genetic exchange with other mycobacterial species. A benefit of this exclusivity is that IS6110 has become an important diagnostic tool in the differentiation of MTBC species from other mycobacteria. Moreover, the element’s presence in multiple copies, and at differing locations in the genome, has provided an excellent method by which strains can be genotyped; because of these characteristics, IS6110 has been used extensively for epidemiological studies (12, 18, 20).

Our studies have focused on DNA transfer between strains of *M. smegmatis*. This work has shown that DNA transfer occurs by a process most similar to conjugation: distinct donor and recipient strains exist and transconjugants are detected only after prolonged cell-cell contact (14, 21, 22). The transfer process is chromosomally encoded and can occur only from a donor to a recipient. The donor and recipient strains are independent isolates of *M. smegmatis* with distinct colony morphologies (13). The genetic basis for donor and recipient ability is unknown. In order to identify genes required for DNA transfer, we have used transposon mutagenesis to isolate transfer-defective mutants in both donor and recipient strains (8; A. Coros, B. Callahan, and K. M. Derbyshire, unpublished data). The transposon mutagenesis screen performed in the recipient strain of *M. smegmatis* (MKD8) identified several insertions into regions of DNA unique to the recipient genome (i.e., the regions are absent from the sequenced mc2155 donor genome; The Institute for Genomic Research), suggesting that they encode recipient-specific functions necessary for transfer (Coros et al., unpublished). We have used a PCR-based chromosome-walking technique to sequence outward from these unique regions into sequences that align with the known *M. smegmatis* genome sequence. Analysis of one of these recipient-specific regions identified an insertion sequence that is very closely related to IS6110 (3e−105, 67% amino acid [aa] identity; Fig. 1). This is a remarkable finding, because IS6110-like elements were thought to be exclusive to the MTBC (12, 17). The presence of this element within the *M. smegmatis* genome is most consistent with the occurrence of a horizontal transfer event, from an MTBC species to *M. smegmatis*; it provides strong supporting evidence for a natural genetic exchange occurring between these two species. It also suggests that the potential shared gene pool of the mycobacteria is much larger than previously thought.

The element (we use “Ms” to differentiate the element from its *M. tuberculosis* counterpart) is not completely intact, because it lacks one inverted repeat (IR); however, the entire transposase gene (encoding 267 aa) and the second IR are highly homologous to ISMt6110 elements found in the *M.
tuberculosis and M. bovis genomes. Immediately adjacent to ISMs6110 is a second insertion sequence (IS) with weak homologies to IS3-like elements (Fig. 1). It is possible that this IS3-like element inserted into ISMs6110 and in the process either deleted or displaced the 5′/H11032 end of the element and the now-missing IR. Not only is the ISMs6110 transposase protein highly conserved throughout its length, but also the nucleotide sequence itself contains extensive regions of identity (e.g., 189/248 nucleotide matches, 76% identity) with the M. tuberculosis sequences. Most importantly, the transposase-binding site, located immediately 3′ of the transposase gene, is highly conserved (25/26 positions; Fig. 1). This conservation is significant since the IR is a critical cis-acting sequence, and its immediate proximity to the gene underscores the functional correlation with the IS6110 transposase; IS transposases work preferentially in cis, and IRs are generally found within a few base pairs of the 3′ end of their cognate transposase gene (4). The one mismatch occurs at the fifth position of the IR; the variations in this region of the IR are not uncommon, because it is a junction between two functional domains of the IR (6, 7, 23).

A Southern analysis was used to determine how many copies of ISMs6110 were present in the recipient M. smegmatis genome. Only one cross-hybridizing band was detected, indicating that only one element was present and that subsequent transposition events had not occurred (see Fig. S1 in the supplemental material). ISMs6110 was not detected in any of these isolates, which included two recipient strains (strains Jucho and Takeuchi) and three donor strains (strains mc2155, Rabinowitchi, and Nishi) (13). Thus, ISMs6110 is unique to the recipient strain MKD8; its absence in other M. smegmatis isolates is a strong indication that MKD8 acquired the element relatively recently. In addition, we included M. tuberculosis genomic DNA as a control in the Southern analysis. No cross-hybridization was detected under the experimental conditions used. It has been shown that an ISMt6110 probe does not cross-hybridize with ISMs6110 under standard diagnostic conditions (M. McGarry, personal communication). The lack of cross-hybridization with ISMt6110 is not surprising given the lack of extended nucleotide identity between the two elements.

IS6110 has never before been detected outside the MTBC complex. After completion of this work, a bioinformatic survey identified a second IS6110-like element in an environmental species of mycobacterium (Mycobacterium sp. strain JLS; Mjls_2222, YP_001070499.1) that is being sequenced. This JLS IS6110 element carries a conserved IR, and its transposase is even more closely related to that of ISMt6110 (86% aa identity). So, how did the M. smegmatis recipient strain acquire a copy of IS6110? The most parsimonious explanation is that the segment of DNA was inherited directly from a member of the
MTBC, when the two species co-occupied a single environmental niche, and the sequence subsequently diverged. A second possibility is that the exchange place through an intermediate species, such as *Mycobacterium* sp. strain JLS. Regardless, the presence of IS6110 in a non-MTBC bacterium indicates that DNA exchange has occurred between mycobacterial species.

Although plasmid transformation has been described for *M. smegmatis*, exchange of chromosomal DNA by this mechanism has not been demonstrated (1). By contrast, genetic exchange of chromosomal DNA by both conjugation and transduction is known to occur (5, 11). Currently, it is impossible to determine which of the latter processes (if not both) mediated the genetic exchange, but it has not escaped our attention that the MKD8 strain of *M. smegmatis* is a recipient for conjugal DNA transfer. Moreover, many of the genes that we have identified as being required for transfer in *M. smegmatis* are conserved in *M. tuberculosis*. An argument against conjugation is that it requires both parents to have coexisted in a single environmental niche, which might seem unlikely given what is known of the species’ normal habitats. By contrast, phage-mediated transfer does not require that the donor and recipient coexist.

Although an IS element has been acquired by the recipient, it is not clear whether the element integrated into the genome via transposition or recombination. As mentioned above, the element lacks one IR, and downstream of the IS are three genes that have only weak homologies with other mycobacterial genes (Coros et al., unpublished), indicating that they too have been acquired from another mycobacterium. It is therefore possible that the entire segment of DNA was inherited by homologous recombination between the upstream IS3 element—of which there are multiple copies in the recipient genome—and genes more distal to the ISms6110 element. Regardless of the mechanism of inheritance, the presence of an IS6110-like element in *M. smegmatis* implicates the existence of horizontal gene transfer among the mycobacteria and suggests that the transfer event described here occurred in the recent past. Although the present example suggests that the transfer event was from the MTBC to *M. smegmatis*, it is not unreasonable to assume that transfer could occur in the opposite direction, i.e., into or between other members of the MTBC. The only requirement would be that recipient strains of the MTBC exist.

**Nucleotide sequence accession number.** The ISms6110 sequence has been deposited at GenBank (EU366287).

We thank Mike McGarry of the Wadsworth mycobacterial genotyping laboratory for confirming that IS6110 from *M. tuberculosis* does not cross-hybridize with IS6110 from *M. smegmatis*. We thank J. Driscoll, V. Escuyer, A Verschoor, and members of the Derbyshire laboratory for comments on the manuscript. The Wadsworth Center Molecular Genetics Core performed DNA sequence analysis.

This work was supported by funding from the NIH (AI 42308) to K.M.D.

**REFERENCES**