Large Sequence Polymorphisms Classify *Mycobacterium tuberculosis* Strains with Ancestral Spoligotyping Patterns

Laura Flores, Tran Van, Sujatha Narayanan, Kathryn DeRiemer, Midori Kato-Maeda, and Sebastien Gagneux

Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital and the University of California, San Francisco, California; Division of Infectious Diseases and Geographic Medicine, Stanford University Medical Center, Stanford, California; Tuberculosis Research Center, Chennai, India; School of Medicine, University of California, Davis, California; and Institute for Systems Biology, Seattle, Washington

Received 18 April 2007/Returned for modification 24 June 2007/Accepted 29 July 2007

Genomic deletion analysis revealed that strains of *Mycobacterium tuberculosis* exhibiting spoligotyping patterns with almost all spacers present belong either to a strain lineage that includes the W-Beijing strain family or to the ancestral strain lineage of *M. tuberculosis*.

Since its introduction in 1997, spoligotyping has become one of the most widely used molecular typing tools for *Mycobacterium tuberculosis* (12, 13). Spoligotyping adds to the discriminatory power of IS6110 restriction fragment length polymorphism (RFLP) typing, the current gold standard for epidemiological studies of tuberculosis (23), and is frequently used as a complementary genotyping tool. This combinational approach has been especially successful for classifying strains with low numbers of IS6110 copies (5). The global database of spoligotyping patterns SpolDB4 has recently been made freely accessible via the Internet (wwwPasteur-Guadeloupe.fr/tb/bd_myco.html). This database contains the typing data of 39,295 clinical isolates of the *M. tuberculosis* complex recovered from sources worldwide. Many strain families have been identified based on particular spoligotyping patterns (4). Studies using additional markers, such as variable number tandem repeats and single nucleotide polymorphisms, have confirmed some of these strain families (1, 19). However, other strain groupings remain poorly defined by spoligotyping, making an unambiguous strain assignment difficult or even impossible (4).

Spoligotypes evolve through the successive loss of spacer DNA sequences that separate short, tandemly repeated DNA sequences in the direct repeat (DR) locus of *M. tuberculosis* (12). Some of these deletion events are mediated by transposing insertion sequences into the DR region or by slipping strand mispairing. Importantly, deletions of specific spacer sequences are not independent evolutionary events, and identical patterns can emerge in unrelated strain lineages (homoplasmy) as a result of convergent evolution (6, 9, 24). This phenomenon is especially problematic in strains exhibiting spoligotyping patterns in which (almost) all spacers are either present or absent. Given the evolutionary mechanism of the DR region (i.e., sequential loss of spacers without the ability to reacquire lost spacers), spoligotyping patterns exhibiting all spacers can be considered progenitors, or “ancestral.” Such ancestral spoligo-

* Corresponding author. Mailing address: Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103. Phone: (206) 732-1398. Fax: (206) 732-1299. E-mail: sgagneux@systemsbiology.org.

* Published ahead of print on 15 August 2007.
SeqMan program (Lasergen DNAStar Inc.) and compared to reference sequences from strains deleted for the corresponding regions of difference (RD) (7).

Our results showed that four of the eight strains analyzed had a deletion in RD105, including two strains with all spacers present and two strains with only spacer 23 missing (Table 1). We recently showed that the deletion in RD105 is a robust marker for the W-Beijing family of strains (21). DNA sequencing of two prototype strains representing each of the two ancestral spoligotyping patterns revealed that the deletion boundaries in RD105 (Table 1) were identical to the ones reported for W-Beijing strains (21, 22). In addition to the deletion in RD105, classical W-Beijing strains also harbor a deletion in RD207, which leads to the loss of spacers 1 to 34 and thus to the characteristic spoligotype of this family, the current gold standard for defining this strain family (2, 4). The results reported here show that the classical W-Beijing strains (i.e., strains deleted for both RD105 and RD207) are part of a broader strain lineage that is defined by the deletion in RD105 and includes strains with ancestral spoligotypes (i.e., strains without the deletion in RD207 and consequently without the classical W-Beijing spoligotype). Strains with ancestral spoligotypes and a deletion in RD105 can be considered ancestral members of this strain lineage. The fact that all four of the strains analyzed herein originated in East Asia (Table 1) is consistent with the dominance of the RD105 or W-Beijing lineage in that part of the world (7).

There has been an increasing interest in the W-Beijing strain family because some of its members produce a phenolic glycolipid that leads to immune modulation and hypervirulence in mice (17). In a recent study, we demonstrated that the phenolic glycolipid is variably produced across different sublineages of classical W-Beijing strains (18). In contrast, the same study found that compared to other main strain lineages, all members of the RD105 lineage, including all sublineages of the W-Beijing family, constitutively overexpressed the triglyceride synthase encoded by the Rv3033c gene, which was associated with the accumulation of large amounts of triacylglyceride (18). These observations further support the notion that the concept of a broader lineage with a deletion in RD105, which includes the classical W-Beijing family, does indeed represent a biologically meaningful strain classification.

The four additional isolates included in this study exhibited ancestral spoligotyping patterns that differed from the ones discussed above, as they were missing spacer 34 only (Table 1). This spoligotyping pattern was referred to as “Manu1” in a recent study (4). Our genomic deletion analysis revealed that these four strains had a deletion in RD239, which is characteristic of the Indo-Oceanic lineage (7), also referred to as the East African-Indian lineage (4). The deletion boundaries in all four Manu1 strains were identical to those reported for the Indo-Oceanic lineage (22). The four isolates originated in India, which is consistent with the predominance of the Indo-Oceanic lineage in that region (7). In addition to the RD239 deletion, the Indo-Oceanic lineage is further characterized by the presence of the genomic region TbD1 (7), which makes the Indo-Oceanic lineage the most ancestral lineage of M. tuberculosis (3). As expected, additional screening for TbD1 by multiplex real-time PCR (7) showed that all four Manu1 strains had the TbD1 region intact. Strains harboring a deletion in RD239 and an intact TbD1 and exhibiting an ancestral spoligotype are likely some of the most ancestral strains of M. tuberculosis. Further comparative analyses of these strains may result in important new insights into the strain-specific pathogenesis and evolutionary history of tuberculosis.

In summary, this study provides yet another example of the usefulness of LSPs for defining phylogenetically robust mycobacterial groupings and inferring evolutionary relationships (3, 10, 14–16). In particular, our results show that LSPs can be used to classify M. tuberculosis isolates with unusual spoligotypes, including those exhibiting high degrees of homoplasy. In addition, these findings demonstrate that strains with almost identical spoligotypes can belong to very phylogenetically distinct lineages.

This work was supported by the Swiss National Science Foundation and the Novartis Foundation (S.G.), the National Institutes of Health, and the Wellcome Trust.

REFERENCES


