

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

Laura Flores,¹ Tran Van,² Sujatha Narayanan,³ Kathryn DeRiemer,⁴
Midori Kato-Maeda,¹ and Sebastien Gagneux^{2,5*}

Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital and the University of California, San Francisco, California¹; Division of Infectious Diseases and Geographic Medicine, Stanford University Medical Center, Stanford, California²; Tuberculosis Research Center, Chennai, India³; School of Medicine, University of California, Davis, California⁴; and Institute for Systems Biology, Seattle, Washington⁵

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 (www.pasteur-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-

types, however, cannot be classified unambiguously, as they could belong to the same ancestral lineage of *M. tuberculosis* or, alternatively, represent ancestral variants of different strain families. An unequivocal phylogenetic classification of these strains would help link particular clinical phenotypes to specific strain genotypes as well as add to our understanding of the global population genetics and evolutionary history of *M. tuberculosis*.

We recently showed that because in *M. tuberculosis* horizontal DNA transfer is extremely rare (20), large sequence polymorphisms (LSPs) are powerful molecular markers that can be used to construct robust phylogenies of *M. tuberculosis* (10). Our analysis of a global sample of 875 strains from 80 countries reported earlier revealed six main lineages, each of which was defined by a specific LSP (7). These lineages are congruent with groupings defined by single nucleotide polymorphisms (8). In addition, each of these lineages is associated with specific geographic regions (7, 8).

In the present study, we analyzed eight strains exhibiting three different ancestral spoligotypes (Table 1). Strains exhibiting such spoligotypes are generally rare. For example, only 40 (~0.1%) of the 39,295 global isolates included in the SpolDB4 database harbor any of these three spoligotypes (Table 1) (4). Of the eight strains analyzed here, four were isolated during our population-based molecular epidemiological study in San Francisco from East Asian immigrants who developed reactivated tuberculosis, as determined by standard IS6110 RFLP genotyping (i.e., they had “unique” genotyping patterns) (11). The four strains from India were isolated at the Tuberculosis Research Center in Chennai. Three of these strains had IS6110 RFLP data available and exhibited different genotyping patterns, indicating that these strains were also epidemiologically independent. To determine whether the eight strains with ancestral spoligotyping patterns belonged to any of the six main lineages of *M. tuberculosis* as defined by LSPs, we performed genomic deletion analysis using PCR conditions and primers published previously (7). The PCR products were sequenced for at least one representative of each spoligotype (<http://cmgm.stanford.edu/pan/>). The sequences were analyzed with the

* 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.

- Gerdes, A. Sajduda, S. Samper, I. Shemyakin, U. B. Singh, A. Somoskovi, R. Skuce, D. van Soolingen, E. M. Streicher, P. N. Suffys, E. Tortoli, T. Tracevska, V. Vincent, T. C. Victor, R. Warren, S. F. Yap, K. Zaman, F. Portals, N. Rastogi, and C. Sola. 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* **6**:23.
5. Dale, J. W., H. Al-Ghusein, S. Al-Hashmi, P. Butcher, A. L. Dickens, F. Drobniowski, K. J. Forbes, S. H. Gillespie, D. Lamprecht, T. D. McHugh, R. Pitman, N. Rastogi, A. T. Smith, C. Sola, and H. Yesilkaya. 2003. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS6110. *J. Bacteriol.* **185**:2555–2562.
 6. Filliol, I., A. S. Motiwala, M. Cavatore, W. Qi, M. Hernando Hazbon, M. Bobadilla Del Valle, J. Fyfe, L. Garcia-Garcia, N. Rastogi, C. Sola, T. Zozio, M. I. Guerrero, C. I. Leon, J. Crabtree, S. Angiuoli, K. D. Eisenach, R. Durmaz, M. L. Joloba, A. Rendon, J. Sifuentes-Osornio, A. Ponce de Leon, M. D. Cave, R. Fleischmann, T. S. Whittam, and D. Alland. 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* **188**:759–772.
 7. Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell, and P. M. Small. 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **103**:2869–2873.
 8. Gagneux, S., and P. M. Small. 2007. The global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* **7**:328–337.
 9. Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser. 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J. Infect. Dis.* **193**:121–128.
 10. Hirsh, A. E., A. G. Tsolaki, K. DeRiemer, M. W. Feldman, and P. M. Small. 2004. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc. Natl. Acad. Sci. USA* **101**:4871–4876.
 11. Jasmer, R. M., J. A. Hahn, P. M. Small, C. L. Daley, M. A. Behr, A. R. Moss, J. M. Creasman, G. F. Schecter, E. A. Paz, and P. C. Hopewell. 1999. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. *Ann. Intern. Med.* **130**:971–978.
 12. 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.
 13. Mathema, B., N. E. Kurepina, P. J. Bifani, and B. N. Kreiswirth. 2006. Molecular epidemiology of tuberculosis: current insights. *Clin. Microbiol. Rev.* **19**:658–685.
 14. Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr. 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J. Infect. Dis.* **186**:74–80.
 15. Mostowy, S., J. Inwald, S. Gordon, C. Martin, R. Warren, K. Kremer, D. Cousins, and M. A. Behr. 2005. Revisiting the evolution of *Mycobacterium bovis*. *J. Bacteriol.* **187**:6386–6395.
 16. Mostowy, S., A. Onipede, S. Gagneux, S. Niemann, K. Kremer, E. P. Desmond, M. Kato-Maeda, and M. Behr. 2004. Genomic analysis distinguishes *Mycobacterium africanum*. *J. Clin. Microbiol.* **42**:3594–3599.
 17. Reed, M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry III. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* **431**:84–87.
 18. Reed, M. B., S. Gagneux, K. DeRiemer, P. M. Small, and C. E. Barry III. 2007. The W/Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and is constitutively upregulated for the DosR dormancy regulon. *J. Bacteriol.* **189**:2583–2589.
 19. Sola, C., I. Filliol, E. Legrand, S. Lesjean, C. Locht, P. Supply, and N. Rastogi. 2003. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect. Genet. Evol.* **3**:125–133.
 20. Supply, P., R. M. Warren, A. L. Banuls, S. Lesjean, G. D. Van Der Spuy, L. A. Lewis, M. Tibayrenc, P. D. Van Helden, and C. Locht. 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol. Microbiol.* **47**:529–538.
 21. Tsolaki, A. G., S. Gagneux, A. S. Pym, Y. O. Goguet de la Salmoniere, B. N. Kreiswirth, D. Van Soolingen, and P. M. Small. 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **43**:3185–3191.
 22. Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Y. O. Goguet de la Salmoniere, K. Aman, M. Kato-Maeda, and P. M. Small. 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc. Natl. Acad. Sci. USA* **101**:4865–4870.
 23. van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, et al. 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**:406–409.
 24. Warren, R. M., E. M. Streicher, S. L. Sampson, G. D. van der Spuy, M. Richardson, D. Nguyen, M. A. Behr, T. C. Victor, and P. D. van Helden. 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. *J. Clin. Microbiol.* **40**:4457–4465.