JOURNAL OF CLINICAL MICROBIOLOGY, Oct. 2007, p. 3393–3395 Vol. 45, No. 10
0095-1137/07/$08.00+0 doi:10.1128/JCM.00828-07

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, California[1]; Division of Infectious Diseases and Geographic Medicine, Stanford University Medical Center, Stanford, California[2]; Tuberculosis Research Center, Chennai, India[3]; School of Medicine, University of California, Davis, California[4]; and Institute for Systems Biology, Seattle, Washington[5]

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 IS*6110* 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 IS*6110* 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 (homoplasy) 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 spoligotypes, 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 IS*6110* 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 IS*6110* 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.

TABLE 1. Genomic deletion analysis of eight strains of *M. tuberculosis* with ancestral spoligotyping patterns

Strain no.	Country of origin	Spoligotyping pattern (spacers 1–43)	Octal value	Frequency in SpolDB4[a]	Deleted RD	Confirmation method(s)
93_2461	Vietnam	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■	77777777777771	19	105	PCR and sequencing
94_M4241A	China	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■	77777777777771		105	PCR
91_2211	Laos	■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■	77777767777771	1	105	PCR and sequencing
96_1548	Laos	■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■	77777767777771		105	PCR
M447	India	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■	77777777773771	20	239	PCR and sequencing
M631	India	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■	77777777773771		239	PCR and sequencing
M1122	India	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■	77777777773771		239	PCR and sequencing
M1140	India	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■	77777777773771		239	PCR and sequencing

[a] Frequency of the three spoligotypes in the international database SpolDB4, which includes the typing data of 39,295 clinical isolates from global sources (4).

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

1. **Baker, L., T. Brown, M. C. Maiden, and F. Drobniewski.** 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg. Infect. Dis. **10:**1568–1577.
2. **Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth.** 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. **10:**45–52.
3. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **99:**3684–3689.
4. **Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Prodinger, A. Gori, S. A. Al-Hajoj, C. Allix, L. Aristimuno, J. Arora, V. Baumanis, L. Binder, P. Cafrune, A. Cataldi, S. Cheong, R. Diel, C. Ellermeier, J. T. Evans, M. Fauville-Dufaux, S. Ferdinand, D. Garcia de Viedma, C. Garzelli, L. Gazzola, H. M. Gomes, M. C. Gutierrez, P. M. Hawkey, P. D. van Helden, G. V. Kadival, B. N. Kreiswirth, K. Kremer, M. Kubin, S. P. Kulkarni, B. Liens, T. Lillebaek, H. M. Ly, C. Martin, I. Mokrousov, O. Narvskaia, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, M. Z. Rahim, V. Rasolofo-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rusch-**

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. Portaels, 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. Drobniewski, 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.