JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 2006, p. 1558–1560 Vol. 44, No. 4
0095-1137/06/$08.00+0 doi:10.1128/JCM.44.4.1558–1560.2006
Copyright © 2006, American Society for Microbiology. All Rights Reserved.

# Microevolution of the Direct Repeat Locus of *Mycobacterium tuberculosis* in a Strain Prevalent in San Francisco

Roxanne S. Aga,[1]† Elizabeth Fair,[1]‡ Neil F. Abernethy,[1]‡ Kathryn DeRiemer,[1]† E. Antonio Paz,[2] L. Masae Kawamura,[2] Peter M. Small,[1]§ and Midori Kato-Maeda[1]*

*Division of Infectious Diseases and Geographic Medicine, Stanford University Medical Center, 300 Pasteur Drive, Grant Building, Room S-169, Stanford, California 94305-5107,[1] and San Francisco TB Control Section, Department of Public Health Ward 94, 1001 Potrero Avenue, San Francisco, California 94110[2]*

Received 10 February 2006/Accepted 16 February 2006

**We describe a microevolutionary event of a prevalent strain of *Mycobacterium tuberculosis* that caused two outbreaks in San Francisco. During the second outbreak, a direct variable repeat was lost. We discuss the mechanisms of this change and the implications of analyzing multiple genetic loci in this context.**

The polymorphism in the direct repeat (DR) locus of *Mycobacterium tuberculosis* is used to study the evolution and the transmission dynamics of specific strains (10). The DR locus is composed of 36-base-pair DR copies, interspersed by nonrepetitive short sequences or direct variable repeats (DVRs), and it is studied using spoligotyping (10). Together with other genetic markers such as the insertion element IS*6110* and polymorphic GC-rich repeat sequence, the DR locus determines the DNA fingerprint of *M. tuberculosis* strains. Strains with an identical fingerprint are considered part of a recent chain of transmission, while different DNA fingerprints are considered reactivation from a previous tuberculosis infection (9).

Clinical strains (8) and laboratory strains (1) that are considered clonal have mutational events in the DR region. Here, we report the microevolution of the DR region in a strain of *M. tuberculosis* within the context of a community outbreak.

The *M. tuberculosis* strains included in this study were collected as part of an ongoing population-based, molecular epidemiology study of tuberculosis in San Francisco, California (9). From 1991 to 2002, 2,953 tuberculosis cases were diagnosed in San Francisco, and a DNA fingerprint was available for 2,501 (84.7%) strains. For this project, we included patients with the strain 0080000p009, which has two IS*6110* bands according to restriction fragment length polymorphism (12) and a characteristic polymorphic GC-rich repeat sequence (11). This is the most frequent strain in San Francisco, causing drug-susceptible tuberculosis in 51 patients between 1995 and 2002. We determined the spoligotype (10) in 48 strains (DNA was unavailable for 3 strains). Spoligotypes were compared with the CDC database and were assigned types accordingly (J. Crawford, personal communication). We sequenced the flanking area of the DVR 23 (5′ GAGTTCCCGTCAGCGTCGTAAATC and 5′ TCCGCGCAG CCAACACCAAGTAGA) at the Stanford University Protein and Nucleic Acid Facility and compared these sequences with the H37Rv strain (http://genolist.pasteur.fr/TubercuList/) using MegAlign 5.01 (DNASTAR Inc.; Windows 32). Epidemiological and clinical data were available from the CDC's Report of a Verified Case of Tuberculosis form and contact investigation documents and were analyzed using SAS software. Chi-square tests and Wilcoxon rank sum tests were performed and *P* values were calculated for variables with sufficient sample size (>5 patients). The study was approved by the institutional review boards of the University of California, San Francisco, and Stanford University.

The strain 0080000p009 caused two outbreaks; the first in 1995 to 1996 and the second in 2001 to 2002. Spoligotyping divided the strain 0080000p009 into two substrains. The first type, lacking DVRs 18, 33 to 36, and 39 to 42, was named S09 by CDC (2, 5, 13). The second substrain was similar to S09 except that DVR 23 was also missing. This substrain, named SU1, was not described in any database during the outbreak (5) nor was it described in the database of spoligotypes maintained at CDC (J. Crawford, personal communication) until 2004. The sequence of the DVR 23 region of SU1 showed a 74-bp-long deletion (genomic address in reference strain H37Rv, 3121974 to 3122048), with no signature elements of an insertion sequence-mediated mutation, compared with S09 and H37Rv. We concluded that both substrains were clonally related, and the substrain SU1 lost a single DVR due to homologous recombination of adjacent direct repeats or due to a deletion of a single DVR (4, 14).

The substrain S09 was isolated during both outbreaks, while SU1 was only detected during the second outbreak. SU1 accounted for 42% of the total 0080000p009 cases in San Francisco during 2001 and 77% of the total 0080000p009 cases in 2002. We analyzed the clinical and epidemiological information to explore whether the emergence and spread of SU1 were by chance (drift), as has been modeled for other organisms (6),

* Corresponding author. Present address: Division of Pulmonary and Critical Care Medicine, Department of Medicine-SFGH, UCSF, 1001 Potrero Avenue, San Francisco General Hospital, Mail Box 0841, San Francisco, CA 94110. Phone: (415) 206-8121. Fax: (415) 695-1551. E-mail: midori.kato-maeda@ucsf.edu.

‡ Present address: Division of Pulmonary and Critical Care Medicine, Department of Medicine-SFGH, UCSF, 1001 Potrero Avenue, San Francisco General Hospital, Mail Box 0841, San Francisco, CA 94110.

† Present address: School of Medicine, University of California, Davis, One Shields Avenue, Davis, CA 95616.

§ Present address: Institute for Systems Biology, 1441 N 34th St., Seattle, WA 98103-8904.

TABLE 1. Demographic and clinical characteristics of tuberculosis patients in San Francisco infected with strain 0080000p009 and with substrain S09 versus substrain SU1

| Characteristic[a] | Total no. (%) infected with 0080000p009 | No. (%) infected with substrain | | *P* value |
|---|---|---|---|---|
| | | S09 | SU1 | |
| *n* | 48 | 33 | 15 | |
| Median age (yr) | 41 | 40 | 46 | 0.018 |
| Age range (yr) | 27–80 | 27–63 | 36–80 | |
| Male gender | 36 (75) | 27 (81.8) | 9 (60.0) | 0.152 |
| Homeless[b] | 32 (67) | 18 (54.6) | 14 (93.3) | 0.023 |
| HIV+ | 28 (58) | 22 (66.7) | 6 (40.0) | 0.147 |
| IVDU[b] | 19 (40) | 14 (42.4) | 5 (33.3) | 0.571 |
| Race | | | | |
| Black | 32 (67) | 23 (69.7) | 9 (60.0) | 0.661 |
| White | 15 (31) | 9 (27.3) | 6 (40.0) | |
| Asian | 1 (2) | 1 (3.0) | | |
| Place of birth | | | | |
| United States | 45 (94) | 30 (90.9) | 15 (100) | |
| Other | 3 (6) | 3 (9.0) | | |
| Pulmonary disease | 47 (98) | 32 (97.0) | 15 (100) | |
| Prior TB | | | | |
| Yes | 2 (4) | 1 (3.0) | 1 (6.7) | |
| No | 46 (96) | 32 (97.0) | 14 (93.3) | |
| Treatment outcome | | | | |
| Completed treatment | 33 (69) | 23 (69.7) | 10 (66.7) | |
| Died | 5 (10) | 5 (15.2) | 0 | |
| Moved | 4 (8) | 3 (9.1) | 1 (6.7) | |
| Unknown | 6 (13) | 2 (6.0) | 4 (26.6) | |

[a] HIV, human immunodeficiency virus; HIV+, HIV seropositive; TB, tuberculosis; IVDU, intravenous drug use.
[b] Reported within the 12 months prior to diagnosis with tuberculosis.

or were due to a selection process whereby the lack of DVR 23, or any mutation in linkage disequilibrium with it, conferred an advantage to the bacteria. We found that homelessness and older age were associated with the SU1 substrain (Table 1). Analysis of the first cases of the 2001 outbreak showed that nine of the 12 cases occurred among homeless persons. Among this group, the first six patients (three with substrain SU1 and three with substrain S09) stayed in one of two specific residences (a shelter or a residential hotel) during their infectious period. It is possible that the high rate of replication during the expansion of the outbreak in these two places resulted in increased chances for mutation, such as the deletion of DVR 23 in the isolates of substrain SU1.

An alternative explanation is that some patients may have had disease due to both S09 and SU1, clonally heterogenic substrains (7). The median age of the patients with SU1 was significantly greater than patients with S09 by nonparametric tests, suggesting that some homeless patients may have been infected with both strains during the first outbreak and developed active disease years later with the SU1 substrain. Future studies using fitness models will be required to evaluate whether SU1 gained an evolutionary advantage allowing it to spread more widely in the second outbreak.

Regardless of the mechanisms of its emergence and successful dissemination, our finding confirms that mutations in the DR region can erase the molecular trail in epidemiological studies. Our analysis of an international 812 strain database, SpolDB3 (5), indicates that 90% of strains would be reclassified by a change in a single DR. This lends further credence to the trend toward using multiple genetic loci in molecular epidemiology studies (3).

This study is limited by the relatively small numbers of patients studied for each substrain, making it difficult to identify the risk factors associated with the mutation. Sputum samples were unavailable for single-colony analysis to rule out mixed infections. However, the genetic study of these outbreak strains provided a unique opportunity (i) to document a mutation in a prevalent strain of *M. tuberculosis*, (ii) to explore the various factors contributing to the microevolution of the DR region, and (iii) to exemplify the limitation of spoligotyping for the study of transmission dynamics during this outbreak.

R.S.A., E.F., N.A., and M.K.-M. were supported by National Institute of Allergy and Infectious Diseases grant AI 34238 (to P.M.S.).

## REFERENCES


1. **Aranaz, A., B. Romero, N. Montero, J. Alvarez, J. Bezos, L. de Juan, A. Mateos, and L. Dominguez.** 2004. Spoligotyping profile change caused by deletion of a direct variable repeat in a *Mycobacterium tuberculosis* isogenic laboratory strain. J. Clin. Microbiol. **42:**5388–5391.
2. **Bauer, J., A. B. Andersen, K. Kremer, and H. Miorner.** 1999. Usefulness of spoligotyping to discriminate IS*6110* low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. J. Clin. Microbiol. **37:**2602–2606.
3. **Cowan, L. S., L. Diem, T. Monson, P. Wand, D. Temporado, T. V. Oemig, and J. T. Crawford.** 2005. Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. J. Clin. Microbiol. **43:**688–695.
4. **Fang, Z., N. Morrison, B. Watt, C. Doig, and K. J. Forbes.** 1998. IS*6110* transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. J. Bacteriol. **180:**2102–2109.
5. **Filliol, I., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. A. Dang, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, M. L. Ho, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolofo-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Suffys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi.** 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. J. Clin. Microbiol. **41:**1963–1970.
6. **Fraser, C., W. P. Hanage, and B. G. Spratt.** 2005. Neutral microepidemic evolution of bacterial pathogens. Proc. Natl. Acad. Sci. USA **102:**1968–1973.
7. **Garcia de Viedma, D., M. Marin, M. J. Ruiz, and E. Bouza.** 2004. Analysis of clonal composition of *Mycobacterium tuberculosis* isolates in primary infections in children. J. Clin. Microbiol. **42:**3415–3418.
8. **Garcia de Viedma, D., M. Marin, M. J. Ruiz Serrano, L. Alcala, and E. Bouza.** 2003. Polyclonal and compartmentalized infection by Mycobacterium tuberculosis in patients with both respiratory and extrarespiratory involvement. J. Infect. Dis. **187:**695–699.
9. **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.
10. **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.

11. **Rhee, J. T., M. M. Tanaka, M. A. Behr, C. B. Agasino, E. A. Paz, P. C. Hopewell, and P. M. Small.** 2000. Use of multiple markers in population-based molecular epidemiologic studies of tuberculosis. Int. J. Tuberc. Lung Dis. **4:**1111–1119.
12. **Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston, G. F. Schecter, C. L. Daley, and G. K. Schoolnik.** 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N. Engl. J. Med. **330:**1703–1709.
13. **Soini, H., X. Pan, A. Amin, E. A. Graviss, A. Siddiqui, and J. M. Musser.** 2000. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J. Clin. Microbiol. **38:**669–676.
14. **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.