Vol. 42, No. 3

Stability of Polymorphic GC-Rich Repeat Sequence-Containing Regions of *Mycobacterium tuberculosis*

Madalene Richardson,[1]* Gian D. van der Spuy,[1] Samantha L. Sampson,[1]† Nulda Beyers,[2]
Paul D. van Helden,[1] and Robin M. Warren[1]

MRC Centre for Molecular and Cellular Biology, Department of Medical Biochemistry,[1] and Centre for TB Research and Education, Department of Paediatrics and Child Health,[2] Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa

Mycobacterium tuberculosis **cultures were subjected to DNA fingerprinting with IS*6110*- and polymorphic GC-rich sequence (PGRS)-containing probes. The PGRS banding patterns remained highly stable during multiple cultures of specimens from one disease episode (0.5% changed) and during transmission in patients with close contact (1.9% changed). Characteristic PGRS-restriction fragment length polymorphism motifs for different strain groupings may indicate distant evolutionary events leading to the differentiation of *M. tuberculosis* strain lineages.**

The IS*6110* insertion sequence is used internationally to type *Mycobacterium tuberculosis* strains by restriction fragment length polymorphism (RFLP) analysis (16). Additional methods, based on genomic repeat elements, have been developed to enhance strain differentiation. These methods include spoligotyping (5), mycobacterial interspersed repetitive unit-variable-number tandem-repeat analysis (14), and Southern hybridization to the polymorphic GC-rich sequence (PGRS) regions (1, 19).

The combination of data from different probes to increase the accuracy of *M. tuberculosis* strain differentiation is common but often limited to subsets of samples (4, 19), especially strains with fewer than six IS*6110* insertions (7, 9). The differential power of molecular markers varies (6) and is dependent on the evolutionary rate of each marker (15, 21), in that the changes of slowly evolving markers indicate historic genetic events, while changes in rapidly evolving markers may indicate more recent events. Identification of the mechanisms relating to the success of this pathogen (such as lineage-specific pathogenesis) as well as the investigation of recent evolutionary events may be applied in the design of treatment and prevention strategies.

Studies to calculate the evolutionary rate of IS*6110* (3, 8, 21, 23) have suggested different rates, although evolutionary events were mostly observed during the early disease phase and are likely to have occurred during exposure to a new host environment. The IS*6110* probe is therefore likely to reflect recent epidemiological events. Also, the presence and observation of evolved strains after disease transmission represent recent events and will affect the calculation of transmission rates (20). Data on the stability of other markers include the data for the 12 mycobacterial interspersed repetitive unit-variable-number tandem-repeat regions which were shown to be

stable over time (12). Similarly, spoligotype patterns were shown to be stable in patient serial culture isolates (13). The evolutionary rate of the PGRS regions is largely unknown, although there is limited evidence of PGRS pattern stability and a slower evolutionary rate than IS*6110* (22).

This study aims to determine the stability of the PGRS-containing domains in *M. tuberculosis* cultures from patients for whom multiple cultures from one disease episode were available and from patients for whom cultures were obtained during early infection. The study forms part of an ongoing molecular epidemiology study (ethically approved by the Institutional Review Board of Stellenbosch University, Faculty of Health Science, and the local health committee) of tuberculosis in Cape Town, South Africa (18).

M. tuberculosis strains from patients with culture-confirmed tuberculosis and residing in two neighboring urban communities (3.4 km^2) with a total population of 38,656 (Statistics South Africa, census 1996) were analyzed. These communities have a low socioeconomic status and a low human immunodeficiency virus infection rate (human immunodeficiency virus seroprevalence in women attending antenatal clinics in Western Cape Province rose from 1.2 to 5.2% from 1994 to 1998 [Directorate of Health Systems Research and Epidemiology, Western Cape Department of Health, 1999]) but a very high incidence of tuberculosis of 761/100,000 per year (17). The RFLP procedure with probe IS*6110* was used (16), and the patterns were analyzed with the GelCompar II software. PGRS-containing RFLP patterns were generated with probe MTB484(1) in combination with *Hin*fI digestion (19) for at least one culture from all patients in the study. MTB484(1) contains both (GTG)$_n$ and PGRS repeats (GenBank accession numbers AF23598, AF025399, AF025400, AF025401, and AF025402) and hybridizes to a number of related sequences containing these repeats. MTB484(1) generated complex patterns of 30 to 40 bands (of various intensities) of approximately 5 to 0.3 kbp. It was observed, as previously suggested (6), that the distribution of PGRS banding-pattern motifs was characteristic for different IS*6110*-defined strain groupings. The patterns remained stable during multiple cultures generated from a single disease epi-

* Corresponding author. Mailing address: P.O. Box 19063, Tygerberg 7505, South Africa. Phone: 27 21 9389073. Fax: 27 21 9389476. E-mail: mr6@sun.ac.za.
† Present address: Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Mass.
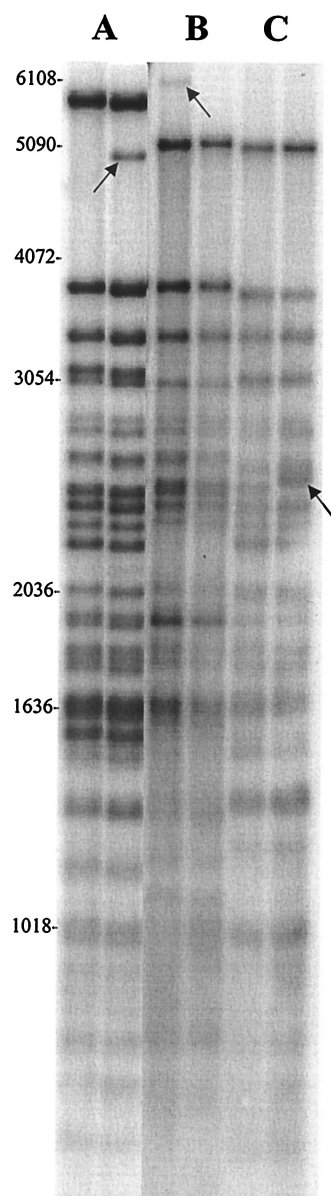
FIG. 1. PGRS banding patterns of strain sets (A to C) with changes. Arrows indicate either an additional band (sets A and B) or a band shift (set B). Characteristic banding-pattern motifs for different strain sets can be observed. Molecular sizes are indicated (in base pairs).

sode and after disease transmission. The limited number of PGRS-RFLP changes observed was characterized by band shifts and additional bands (Fig. 1). The possibility of partial digestion was eliminated because the changes remained after proteinase K repurification and redigestion (HinfI and AluI) of the DNA (19).

From the beginning of the study in July 1992 to the end of 1998, 1,788 cultures representing 938 episodes in 895 patients (70% of all culture-confirmed cases) attending the two community health clinics serving the study communities were analyzed. The IS6110 probe identified 355 strains with six or more copies of IS6110 (high-copy-number strains) and 10

strains with five or fewer copies (low-copy-number strains). The PGRS-containing probe identified 223 banding patterns in the group of 355 high-copy-number strains and 40 in the low-copy-number strains, an additional fourfold resolution for the low-copy-number strains.

The PGRS-RFLP stability of multiple cultures collected during single disease episodes for which both IS6110 and PGRS data were available was investigated. Cases with simultaneous changes in IS6110 and PGRS were excluded to eliminate the possibility of insertion sequence-mediated mutation of the PGRS region. Data for both probes for serial cultures were available for 220 of 384 (57%) patients. An additional band was observed in the PGRS pattern in only 1 of 220 (0.5%) patient cultures, but the pattern reverted in a further follow-up culture. In comparison, the IS6110 patterns changed in 4% of patients with multiple cultures (21). The time range for serial isolates was 0 to 847 days (6 days for the changed strain).

To determine the stability of the PGRS domains after transmission, *M. tuberculosis* patients infected by close-contact transmission were identified. Close-contact transmission was defined as disease caused by strains with identical IS6110 banding patterns in patients residing at the same address or within two closely situated houses (on the same side of the street). It was assumed that transmission took place due to the identified close contact only if the secondary-case strain had not been cultured previously from a patient resident elsewhere in the community during the ongoing molecular epidemiological study (20). Of the 106 events of transmission due to close contact, 79 (74.5%) occurred without any changes in the PGRS domains. Simultaneous changes in IS6110 and PGRS were observed in 8.5% (9 of 106) of close-contact transmission events (as above, these were excluded from the analysis), and in 16 of 106 (15.1%) events, the changes were associated only with IS6110. In 1.9% (2 of 106) of the events, changes (a band shift and a faint additional band) were observed in PGRS regions. The intervals between the isolation of the initial strain and the appearance of the changed isolate were 9 and 12 months.

Many recent studies have measured the differential resolution of a molecular marker by the number of subgroups defined, without considering the evolutionary rate (6, 10). This study demonstrated that *M. tuberculosis* PGRS-containing domains change at a lower rate than does IS6110 and are highly stable. PGRS-RFLP data are therefore unlikely to significantly affect the calculation of recent transmission in molecular epidemiology studies. This finding was predicted by mathematical analysis, where it was shown that secondary probing against the PGRS regions will not significantly influence the calculation of recent transmission unless low-copy-number strains are abundant and highly clustered (10). The inability to distinguish recent events may lead to underestimation of transmission rates, especially when strain patterns occur throughout the study region (11). The greater resolution of PGRS probes for the low-copy-number strains is likely to be a consequence of the inability of IS6110 to undergo change in these strains rather than of a higher rate of PGRS-associated changes. The value of PGRS-containing probes will therefore lie in the confirmation of strain relatedness within strain lineages and their use in combination with other molecular markers in differentiating low-copy-number strains.

Given the characteristic RFLP motifs of the PGRS-containing probes, correlation between strain lineages may allow identification of the evolutionary events leading to the differentiation of strain lineages. Furthermore, strains may appear unrelated with the IS*6110* marker but share a characteristic PGRS pattern, which may indicate that they belong to a common lineage. This was observed for strains shown to belong to the Beijing strain lineage, although this was not indicated by their IS*6110* patterns (our unpublished data). PGRS-containing domains may therefore indicate phylogenetic relatedness but are not an indicator of recent evolutionary events.

The PGRS-containing genomic regions of *M. tuberculosis* have been shown to be associated with a subgroup of the PE (Pro-Glu) gene family (2). The evolution of these genes has been implicated in antigenic variation and evasion of the host defense system (2). However, this study implies that the interface between the bacillus and the host immune system only rarely induces genomic changes which can be observed by PGRS-RFLP analysis, although it does not exclude changes on a level undetected by such analysis.

REFERENCES

1. **Chaves, F., Z. Yang, H. El Hajj, M. Alonso, W. J. Burman, K. D. Eisenach, F. Dronda, J. H. Bates, and M. D. Cave.** 1996. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **34:**1118–1123.
2. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badkock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Olivier, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:**537–544.
3. **De Boer, A. S., M. W. Borgdorff, P. E. de Haas, N. J. Nagelkerke, J. D. van Embden, and D. van Soolingen.** 1999. Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. J. Infect. Dis. **180:**1238–1244.
4. **Ijaz, K., Z. Yang, H. S. Matthews, J. H. Bates, and M. D. Caves.** 2002. *Mycobacterium tuberculosis* transmission between cluster members with similar fingerprint patterns. Emerg. Infect. Dis. **8:**1257–1259.
5. **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.
6. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. M. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. A. van Embden.** 1999. Comparison of methods based on different molecular epidemiologic markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. **37:**2607–2618.
7. **Mathema, B., P. J. Bifani, J. Driscoll, L. Steinlein, N. Kurepina, S. L. Moghazeh, E. Shashkina, S. A. Marras, S. Campbell, B. Mangura, K.**

8. Shilkret, J. T. Crawford, R. Frothingham, and B. N. Kreiswirth. 2002. Identification and evolution of an IS*6110* low-copy-number *Mycobacterium tuberculosis* cluster. J. Infect. Dis. **185:**641–649.
8. **Niemann, S., E. Richter, and S. Rüsch-Gerdes.** 1999. Stability of *Mycobacterium tuberculosis* IS*6110* restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis. J. Clin. Microbiol. **37:**409–412.
9. **Rasolofo-Razanamparany, V., H. Ramarokoto, G. Aurégan, B. Gicquel, and S. Chanteau.** 2001. A combination of two genetic markers is sufficient for restriction fragment length polymorphism typing of *Mycobacterium tuberculosis* complex in areas with a high incidence of tuberculosis. J. Clin. Microbiol. **39:**1530–1535.
10. **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.
11. **Richardson, M., S. W. P. van Lill, G. D. van der Spuy, Z. Munch, C. N. Booysen, N. Beyers, P. D. van Helden, and R. M. Warren.** 2002. Historic and recent events contribute to the disease dynamics of Beijing-like *Mycobacterium tuberculosis* isolates in a high incidence region. Int. J. Tuberc. Lung Dis. **6:**1001–1011.
12. **Savine, E., R. M. Warren, G. D. van der Spuy, N. Beyers, P. D. van Helden, C. Locht, and P. Supply.** 2002. Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **40:**4561–4566.
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. **Supply, P., J. Magdalena, S. Himpens, and C. Locht.** 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol. Microbiol. **26:**991–1003.
15. **Tanaka, M. M., and N. A. Rosenberg.** 2001. Optimal estimation of transposition rates of insertion sequences for molecular epidemiology. Stat. Med. **20:**2409–2420.
16. **Van Embden, J. D. A., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, and P. M. Small.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:**406–409.
17. **Verver, S., R. M. Warren, Z. Munch, E. Vynnycky, P. D. van Helden, M. Richardson, G. D. van der Spuy, D. A. Enarson, M. W. Borgdorff, M. A. Behr, and N. Beyers.** Transmission of tuberculosis in a high incidence urban community in South Africa. Int. J. Epidemiol., in press.
18. **Warren, R., M. Richardson, G. van der Spuy, T. Victor, S. Sampson, N. Beyers, and P van Helden.** 1999. DNA fingerprinting and molecular epidemiology of tuberculosis: use and interpretation in an epidemic setting. Electrophoresis **20:**1807–1812.
19. **Warren, R. M., M. Richardson, S. Sampson, J. H. Hauman, N. Beyers, P. R. Donald, and P. D. van Helden.** 1996. Genotyping of *Mycobacterium tuberculosis* with additional markers enhances accuracy in epidemiological studies. J. Clin. Microbiol. **34:**2219–2224.
20. **Warren, R. M., G. D. van der Spuy, M. Richardson, N. Beyers, C. Booysen, M. A. Behr, and P. D. van Helden.** 2002. Evolution of the IS*6110*-based restriction fragment length polymorphism pattern during the transmission of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **40:**1277–1282.
21. **Warren, R. M., G. D. van der Spuy, M. Richardson, N. Beyers, M. W. Borgdorff, M. A. Behr, and P. D. van Helden.** 2002. Calculation of the stability of the IS*6110* banding pattern in patients with persistent *Mycobacterium tuberculosis* disease. J. Clin. Microbiol. **40:**1705–1708.
22. **Yang, Z. H., F. Chaves, P. F. Barnes, W. Burman, J. Koehler, K. D. Eisenach, J. H. Bates, and M. D. Cave.** 1996. Evaluation of method for secondary DNA typing of *Mycobacterium tuberculosis* with pTBN12 in epidemiologic study of tuberculosis. J. Clin. Microbiol. **34:**3044–3048.
23. **Yeh, R. W., A. Ponce de Leon, C. B. Agasino, J. A. Hahn, C. L. Daley, P. C. Hopewell, and P. M. Small.** 1998. Stability of *Mycobacterium tuberculosis* DNA genotypes. J. Infect. Dis. **177:**1107–1111.