Tuberculosis (2008) **88**, 283–294

Tuberculosis

http://intl.elsevierhealth.com/journals/tube

Mycobacterium tuberculosis PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates

Sarah Talarico[a], Lixin Zhang[a,†], Carl F. Marrs[a,†], Betsy Foxman[a,†], M. Donald Cave[b,c], Michael J. Brennan[d], Zhenhua Yang[a,*]

[a]*Department of Epidemiology, School of Public Health, University of Michigan, 109 S. Observatory Street, 4648 SPH I, Ann Arbor, MI 48109-2029, USA*
[b]*Central Arkansas Veterans Healthcare Center, 4300 West 7th St., Little Rock, AR 72205, USA*
[c]*Department of Neurobiology and Developmental Sciences, College of Medicine, University of Arkansas for Medical Sciences, 4301 W. Markham St., Little Rock, AR 72205, USA*
[d]*Office of Vaccines Research and Review, Center for Biologics Evaluation and Research, Food and Drug Administration, Bldg. 29, Rm. 503, HFM-431, 29 Lincoln Drive, Bethesda, MD 20892, USA*

Received 14 August 2007; received in revised form 1 December 2007; accepted 5 January 2008

KEYWORDS
Mycobacterium tuberculosis;
Genetic diversity;
PE_PGRS gene family

Summary The *Mycobacterium tuberculosis* PE_PGRS multigene family is thought to be involved in antigenic variation, which can be generated by differential regulation of expression and a high frequency of genetic polymorphism. PE_PGRS16 and PE_PGRS26 are inversely regulated during persistent *M. tuberculosis* infection, suggesting that differential regulation of the expression of these two PE_PGRS genes may have a role in latency. To understand how genetic diversity, in addition to differential regulation, contributes to antigenic variability, we investigated the sequence variations in the PE_PGRS16 and PE_PGRS26 genes among 200 clinical *M. tuberculosis* strains, in comparison to the sequenced laboratory strain H37Rv, using PCR and DNA sequencing. Among the 200 strains, 102 (51%) and 100 (50%) had sequence variations within the PE_PGRS16 gene and the PE_PGRS26 gene, respectively. In-frame insertions and deletions, frameshifts, and SNPs were observed in both the PE_PGRS16 gene and the PE_PGRS26 gene. However, the frequency of frameshifts and in-frame deletions differed between the two PE_PGRS genes. Examining the profile of the PE_PGRS16, PE_PGRS26, and the previously investigated PE_PGRS33 amino acid sequences for each of the 200 strains, 72 different profiles were observed with frequencies ranging from 0.5% to 13%. In conclusion, a remarkable level of genetic diversity exists in the PE_PGRS16 and PE_PGRS26 genes

*Corresponding author. Tel.: +1 734 763 4296; fax: +1 734 764 3192.
E-mail address: zhenhua@umich.edu (Z. Yang).
†Authors who have contributed equally to the study.

1472-9792/$ - see front matter
doi:10.1016/j.tube.2008.01.001

of M. tuberculosis clinical strains. The significant sequence variations in the two PE_PGRS genes observed in this study could impact the function of these two PE_PGRS proteins and be associated with differences in the ability of the tubercle bacilli to remain persistent within the host.

Introduction

Mycobacterium tuberculosis, killing approximately 2 million people worldwide each year, has been called the most successful human pathogen. Controlling tuberculosis will require a better understanding of the mechanisms which allow *M. tuberculosis* to evade the immune system and remain persistent in the host. Sequencing of the genomes of *M. tuberculosis* strains has provided important insights into possible mechanisms of persistence, including the discovery of the multigene family of ~60 genes named PE_PGRS that is thought to be involved in antigenic variation. The PE domain of the PE_PGRS protein has a proline–glutamic acid sequence near the amino terminus and the PGRS domain of the protein varies in size and contains many repeats of alanine and glycine.[1] There is evidence that at least some members of this gene family are expressed on the cell surface during *M. tuberculosis* infection and recognized by the host immune system.[2–5] The maintenance of this large multigene family in the *M. tuberculosis* genome suggests that these genes are important to the success of the organism, perhaps because the variability of the PE_PGRS protein surface antigens may contribute to the ability of *M. tuberculosis* to persist in the face of the host immune system.

The regulation of gene expression is one mechanism for generating antigenic diversity. There is evidence that *M. tuberculosis* PE_PGRS genes are variably expressed in different conditions and during different time points of infection.[6–8] In a study of persistent *M. tuberculosis* infection in a mouse model, PE_PGRS16 and PE_PGRS26 were inversely regulated, with expression of PE_PGRS16 being significantly up-regulated and expression of PE_PGRS26 being significantly down-regulated, suggesting that differential regulation of these two PE_PGRS genes may have a role in latency and that the inverse expression of these two genes could potentially serve as a marker of latent infection.[7]

Antigenic variation of an organism can also be generated by a high level of genetic variability of the genes that encode antigens. Thus, to understand the full scope of surface antigen variability generated by the PE_PGRS gene family, it is important to investigate the genetic diversity of these genes among clinical isolates. The sequence variations in one member of this gene family, PE_PGRS33, have been characterized for 123 clinical *M. tuberculosis* strains and included single nucleotide polymorphisms (SNPs), insertions, deletions, and a frameshift mutation. These sequence variations were observed in different combinations resulting in 23 different PE_PGRS33 alleles.[9] Furthermore, in a population-based study of 649 clinical *M. tuberculosis* isolates, patients infected with *M. tuberculosis* isolates having large changes to the PE_PGRS33 protein were 1.9 times more likely to belong to a cluster of tuberculosis cases, defined by *M. tuberculosis* genotyping, and 1.6 times more likely to lack cavitations in the lungs than were patients infected with *M. tuberculosis* isolates having no or minimal change to the PE_PGRS33 protein. This suggests that PE_PGRS33 may have an important role in *M. tuberculosis* persistence.[10]

To extend our knowledge of the genetic diversity of *M. tuberculosis* generated by the PE_PGRS genes and to understand how genetic diversity, in addition to differential regulation, contributes to antigenic variability, we investigated the sequence variations within the PE_PGRS16 and PE_PGRS26 genes among 200 clinical *M. tuberculosis* strains. The frequency of different types of sequence variations was compared between the PE_PGRS16 gene and the PE_PGRS26 gene and the potential antigenic diversity of the 200 strains generated by sequence variations in three PE_PGRS genes, the PE_PGRS16 and PE_PGRS26 genes and the previously investigated PE_PGRS33 gene, was examined.

Materials and methods

M. tuberculosis strains

A study sample of 200 *M. tuberculosis* strains was selected from 705 isolates collected in Arkansas between 1996 and 2000. Strains were selected based on the isolate genotyping data that were available from the Mycobacteriology Research