Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 208371, 6 pages
doi:10.1155/2010/208371

Research Article

Assessment of the Genetic Diversity of *Mycobacterium tuberculosis esxA, esxH,* and *fbpB* Genes among Clinical Isolates and Its Implication for the Future Immunization by New Tuberculosis Subunit Vaccines Ag85B-ESAT-6 and Ag85B-TB10.4

Jose Davila,[1] Lixin Zhang,[1] Carl F. Marrs,[1] Riza Durmaz,[2] and Zhenhua Yang[1]

[1] *Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, USA*
[2] *Department of Clinical Microbiology, Medical Faculty, Inonu University, 44280 Malatya, Turkey*

Correspondence should be addressed to Zhenhua Yang, zhenhua@umich.edu

Received 2 November 2009; Accepted 15 April 2010

Academic Editor: Anne S. De Groot

The effort to develop a tuberculosis (TB) vaccine more effective than the widely used Bacille Calmette-Guérin (BCG) has led to the development of two novel fusion protein subunit vaccines: Ag85B-ESAT-6 and Ag85B-TB10.4. Studies of these vaccines in animal models have revealed their ability to generate protective immune responses. Yet, previous work on TB fusion subunit vaccine candidate, Mtb72f, has suggested that genetic diversity among *M. tuberculosis* strains may compromise vaccine efficacy. In this study, we sequenced the *esxA, esxH,* and *fbpB* genes of *M. tuberculosis* encoding ESAT-6, TB10.4, and Ag85B proteins, respectively, in a sample of 88 clinical isolates representing 57 strains from Ark, USA, and 31 strains from Turkey, to assess the genetic diversity of the two vaccine candidates. We found no DNA polymorphism in *esxA* and *esxH* genes in the study sample and only one synonymous single nucleotide change (C to A) in *fbpB* gene among 39 (44.3%) of the 88 strains sequenced. These data suggest that it is unlikely that the efficacy of Ag85B-ESAT-6 and Ag85B-TB10.4 vaccines will be affected by the genetic diversity of *M. tuberculosis* population. Future studies should include a broader pool of *M. tuberculosis* strains to validate the current conclusion.

1. Introduction

The need for an improved vaccine against tuberculosis (TB) has never been more urgent. One in three people today are infected with *Mycobacterium tuberculosis*, the causative agent of TB, and worldwide approximately three million people die from TB annually. The currently available TB vaccine, Bacille Calmette-Guérin (BCG), has failed to consistently protect against the most contagious form of the disease, adult pulmonary TB, despite its widespread use [1–4]. Developing a new vaccine, which may serve as a booster or a replacement for BCG, is of critical importance in the fight against worldwide TB-related morbidity and mortality [4, 5].

Of the various vaccine candidates proposed, fusion subunit vaccines have received considerable attention in the recent literature, especially those composed of antigenic proteins ESAT-6, Ag85B, and TB10.4 [3–8]. It appears that the multiple epitopes that fusion subunit vaccines offer makes them more effective than single-peptide vaccines in interacting with the complexity of the host immune response against TB and the genetic restriction imposed by major histocompatibility complex molecules [3, 9]. Two fusion subunit vaccines, Ag85B-ESAT-6 and Ag85B-TB10.4, which are the focus of the present study, have been found to induce protective cell-mediated immunity in animal models [3, 6, 7, 9, 10]. Ag85B-ESAT-6 is currently in expanded Phase I studies in which the vaccine is tested in BCG-vaccinated, latently infected, and individuals from TB endemic regions [4]. As these candidates move forward in or toward clinical trials, it will be critically important to evaluate their protective potential as global vaccines via bioinformatic approaches built upon the comparative genomics of the pathogen population and the immunomics of the host population.

Bioinformatics approaches are invaluable to the development of effective vaccine candidates. Comparative genomics of the pathogen population, for instance, allows vaccine candidates that are potentially ineffective due to genetic diversity of the pathogen population to be discredited before they reach the costly stages of clinical trials. In other words, bioinformatic approaches can provide information based on which a rational selection of clinical trial sites can be made. As for subunit vaccines, comparative genomics can help analyze whether antigenic targets are conserved among infectious strains of an organism in order to ensure their protective efficacy across diverse pathogen populations circulating in different geographic region [8].

Although previous studies have suggested that *M. tuberculosis* has a relatively stable genome in comparison with other bacteria [11, 12], recent genomic studies have revealed biologically significant variation among clinical strains [13]. Hebert and colleagues, for instance, revealed considerable genetic variation in the PPE18 gene of *M. tuberculosis*, with important implications for the ability of the Mtb72f vaccine candidate to provoke protective immunity against diverse populations of *M. tuberculosis* [14]. Furthermore, the interaction of the genetic variation of the PPE18 component of Mtb72f with the allelic variation of human MHC-II DRB1 proteins negatively affects vaccine epitope binding to DRB1 proteins [15]. Taken in the context of vaccine development, revelations like these are crucial to the survival of vaccine candidates as potential clinical vaccines. A similar comparative genomics study on Ag85B-ESAT-6 and Ag85B-TB10.4 subunit vaccines may provide useful information for predicting the protective efficacy of these candidates in the pre- or early stages of their clinical evaluation.

Little information has been documented on the genetic variation of the genes encoding for ESAT-6, Ag85B, and TB10.4 proteins. If any of these three genes is highly variable, the protective efficacy of Ag85B-ESAT-6 and Ag85B-TB10.4 subunit vaccines might be compromised on the global stage. To further investigate the ability of these two-vaccine candidates in recognizing naturally occurring *M. tuberculosis* strains, we investigate the genetic diversity of the *esxA*, *esxH*, and *fbpB* genes of *M. tuberculosis* that encode for the components of the two new subunit vaccines in a sample of 88 *M. tuberculosis* strains collected from Turkey and Arkansas, USA.

2. Materials and Methods

2.1. M. tuberculosis Isolates. The clinical strains used in the present study are from Ark, USA, and Turkey. Following the work of Herbert and colleagues, the isolates were selected to represent different geographical regions, including Arkansas and Malatya, Turkey, to assess the impact of regional genetic variability on the two subunit vaccine candidates [14]. Each of the selected isolates represents a different strain of *M. tuberculosis* with a distinct IS*6110* restriction fragment polymorphism (RFLP) pattern with more than five bands or a distinct combination of a common IS*6110* RFLP pattern with five or less bands and a unique spoligo typing pattern (Table 1). The rationale for including isolates from two geographical regions was to discern the potential impact of genetic variation on future vaccination with Ag85B-ESAT-6 and Ag85B-TB10.4 in separate populations.

Our initial intent was to analyze the same set of clinical isolates ($n = 225$) used by Hebert and colleagues in their work on PPE18 and *pepA* [14]. However, after initial sequencing of a randomly selected subset ($n = 41$) of the 225 isolates revealed no genetic variation in the ESAT-6, TB10.4, and Ag85B genes, we selected only 47 of the remaining 84 isolates that had previously shown variation in the PPE18 gene ($n = 47$) for the current study. This decision was made with the consideration of cost-effective lab procedure, reasoning that local genetic variations would be indicative of broader genomic variability. The 88 study isolates represent 57 different strains from Ark, USA, and 31 distinct strains from Turkey. The 88 study isolates shared 49 different spoligotypes that represent 47 different spoligo international types, as determined by using the query tool of the fourth international spoligotyping database (SpolDB4) [16]. The present sample represents 80 of 84 strains (95.2%) that showed PPE18 variation in Hebert's study.

2.2. PCR of esxA, esxH, and fbpB Genes. The three genes under study were amplified using Invitrogen Platinum Taq PCRx Polymerase kit (Invitrogen, Carlsbad, CA) with primers published previously [17]. The primers used for *esxA*, encoding the ESAT-6 protein were esxA-F (5′-GCA-ATCCGGCGGCTCCACCAG-3′) located 533 bp upstream of the *esxA* gene and esxA-R (5′-TCGGCCGCCATGACA-ACCTCTC-3′) located 124 bp downstream from the end of the *esxA* gene. The primers used for *esxH*, encoding the TB10.4 protein were esxH-F (5′-GAGAGGGGGAGG-CGACGGCTTACC-3′) located 411 bp upstream of the esxH gene, and esxH-R (5′-TCCCCGCCCCAATGGTTTCAGC-3′) located 86 bp downstream from the end of the *esxH* gene. Finally, the primers used for *fbpB*, encoding Ag85B protein were fbpB-F2 (5′-ACTCGGCTAACTGGCTGGTGC-3′) located 217 bp upstream of the *fbpB* gene, and fbpB-R (5′-CATACCGCCATACCGTTTGTGAGC-3′) located 164 bp downstream from the end of the *fpbB* gene. The inclusion of the regions flanking the *esxA*, *esxH*, and *fbpB* genes allowed further confirmation that the PCR products were specific. The positive control in all the PCR reactions was *M. tuberculosis* H37Rv, and the negative control was PCR-grade water. The 50 μL PCR reaction mixture used was composed of 5 μL of 10 × PCRx amplification buffer, 1.5 μL of 50 mM $MgSO_4$, 1 μL of 10 mM deoxyribonucleoside triphosphate mixture, 20 pmol of each primer in 1 μL, 0.5 μL of Invitrogen Platinum Taq polymerase mixture, 4 μL of a DNA solution containing 50 ng of DNA template, and 40 μL of PCR-grade water. The thermocycling program used for all genes was one cycle at 94°C for 1 minute; 30 cycles of 94°C for 30 seconds, 62°C for 30 seconds, and 72°C for 2.5 minutes; and a final cycle of 72°C for 10 minutes. The sizes of the PCR products were verified by 1.0% (wt/vol) agarose gel electrophoresis in 1 × Tris-borate-EDTA buffer.

Table 1: Genotyping data of the 88 study isolates of *M. tuberculosis* that were originated from Arkansas, USA, and Turkey.

Isolate no.*	IS*6110* Copy no.	IS*6110* RFLP Pattern ¥	Spoligotype	Isolate no.*	IS*6110* Copy no.	IS*6110* RFLP Pattern ¥	Spoligotype
SA143	14	01046	000000000003771	SA208	03	00370	700036777760471
SA033	18	01063	000000000003771	SA738	07	05660	777777744720771
SA604	18	04495	000000000003771	SA204	10	00636	074377607760700
SA327	21	00237	000000000003771	SA565	09	01321	377737774020731
SA374	22	02623	000000000003771	SA710	17	02661	737377677760771
SA413	22	02660	000000000003771	SA221	01	00129	774377777413771
SA193	14	00638	677737607760771	SA253	14	00619	776377613760771
SA759	12	05774	677777477413771	SA052	03	00403	777776777560601
SA639	15	04804	703777740003771	SA333	04	01658	777776777560601
SA085	11	01336	776177607760771	SA597	07	04433	777776777760761
SA557	11	04222	776177607760771	SA286	08	01330	777777000020771
SA378	09	02620	777777770000000	SA198	03	00370	777777777613771
SA652	08	01646	777777774020771	SA173	19	00401	Not Available
SA053	11	00499	777777774020771	TK035	09	TK035	000000004020771
SA475	05	04430	777777777720771	TK036	08	TK036	000000007760771
SA010	09	00472	777777777720771	ZD073	15	ZD073	703777740003031
SA637	10	04802	777777777720771	TK013	03	TK013	777777404760771
SA213	13	00640	777777777720771	TK091	03	TK091	777777404760771
SA364	13	00807	777777777720771	TK110	08	TK110	777777740460771
SA025	11	00325	777777777760731	ZD039	08	ZD039	777777774020771
SA310	10	00315	776177607760771	ZD056	09	ZD056	777777774020771
SA285	06	01329	777777777760771	ZD061	08	ZD061	777777777720771
SA169	09	00326	777777777760771	ZD063	08	ZD063	777777777720771
SA120	12	00372	777777777760771	ZD034	09	ZD034	777777777720771
SA102	12	00538	777777777760771	TK086	10	TK086	777777777720771
SA471	14	06474	777777777760771	ZD013	10	ZD013	777777777720771
SA131	15	00627	777777777760771	ZD020	11	ZD020	777777777720771
SA157	06	00314	777777557760771	ZD071	11	ZD071	777777777720771
SA030	13	00570	777777607560771	TK065	05	TK065	777777777760771
SA706	10	05458	777777777760471	TK103	08	TK103	777777777760771
SA111	02	00016	777776777760771	TK095	09	TK095	777777777760771
SA352	03	00064	777776777760771	ZD062	09	ZD062	777777777760771
SA019	04	00432	777776777760771	TK030	10	TK030	777777777760771
SA170	06	00327	777776777760771	TK105	11	TK105	777777777760771
SA341	03	00694	777776777760601	ZD093	13	ZD093	757777777760771
SA020	05	00392	777776777760601	TK003	06	TK003	037637777760771
SA596	07	04432	777776777760601	ZD041	06	ZD041	777777764020731
SA526	01	00195	777777776413771	ZD031	09	ZD031	717777777760771
SA162	04	00319	700076777760700	ZD024	09	ZD024	770000770000000
SA778	08	06019	376177607760771	TK007	08	TK007	777767777760671
SA038	11	00496	777777777320771	TK016	07	TK016	037227760160771
SA512	12	00832	776377770760771	TK011	08	TK011	700377777760771
SA189	13	00598	777763777760771	TK112	09	TK112	777207777777661
SA430	01	00129	763777777413771	TK097	05	TK097	Not Available

*Isolate numbers starting with SA refer to strains from Arkansas, USA. Isolate numbers starting with ZD or TK refer to strains from Turkey.

¥ The IS*6110* RFLP patterns of the Arkansas strains were indicated using the national designations assigned by the National Tuberculosis Genotyping and Sentinel Surveillance Network Program of the Centers for Disease Control and Prevention. The IS*6110* RFLP patterns of the Turkish strains were designated using the isolate number in which a given IS*6110* RFLP pattern was observed for the first time.

2.3. Automated DNA Sequencing. PCR products were sequenced to identify any insertions/deletions or single nucleotide polymorphisms (SNPs) in the *esxA*, *esxH*, and *fbpB* genes in the selected isolates. PCR products were purified using Invitrogen PureLink PCR Purification Kit, according to manufacturer instructions (Invitrogen, Carlsbad, CA). DNA sequencing was performed with Applied Biosystems DNA sequencers 3700 and 3730 at the University of Michigan Sequencing Core, using the same primers that were used for the PCR of the three genes. The sequences of *esxA*, *esxH*, and *fbpB* of the study strains were compared to those of *M. tuberculosis* laboratory reference strain H37Rv (GenBank accession number BX842575) using the BLAST and BLAST2 nucleotide sequence alignment program of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST) and the Sequencher 4.9 DNA Sequence Assembly software (Demo version) of Gene Codes Corporation (www.genecodes.com). Primary DNA transcripts were translated to amino acid sequences using the *in silico* simulation of molecular biology experiment of the University of Basque Country (http://insilico.ehu.es/).

3. Results and Discussion

3.1. Genetic Diversity of esxA, esxH, and fbpB and Corresponding Amino Acid Sequences. Among the 88 strains investigated, genetic analysis of *esxA* and *esxH* in this study revealed no nucleotide polymorphisms in the genes encoding for ESAT-6 and TB10.4 proteins. Of the 88 strains, 38 (43.2%) belong to principal genetic group 1, 29 (33.0%) belong to principal genetic group 2, and 21 (23.9%) belong to principal genetic group 3. The principle genetic groups were defined by SNPs in the *katG* and *gyrA* genes as described previously by Sreevatsan and colleagues [12]. Unlike the study by Herbert et al., where genetic group 1 strains were found to have the highest frequency of DNA polymorphisms in the PPE18 protein, a component of the Mtb72f vaccine [14], strains in all of the three principal genetic groups showed no DNA variations in both *esxA* and *esxH*. This observation suggests that these gene regions might be conserved among *M. tuberculosis* strains of different geographic origins and among different genetic groups of the pathogen. In fact, Gey Van Pittius and colleagues have recently posited interspecies conservation of the ESAT-6 gene region as part of a novel Gram-positive secretion system with distant homologues in *Bacillus subtilis, Bacillus anthracis, Staphylococcus aureus,* and *Clostridium acetobutylicum* [18].

The analysis of *fbpB,* the gene encoding for Ag85B revealed only one synonymous C to A SNP, located at position 714 bp of the gene sequence, among 39 (44.3%) of the 88 strains sequenced. Double strand sequencing was conducted on the 39 isolates to confirm the existence of this SNP. Although this SNP had no effect on the amino acid sequence of the peptide when translated, it is indicative of an allelic variation in the *M. tuberculosis* gene pool. Of the 39 strains, 15 (17.0%) belong to principal genetic group 1, 11 (12.5%) belong to principal genetic group 2, and 13 (14.7%) belong to principal genetic group 3. Furthermore, the SNP was not found to be associated with any specific geographic origin of the study strains, suggesting that this particular nucleotide polymorphism is of ancestral origin. These data suggest that *M. tuberculosis* Ag85B antigen is highly conserved in, at least, certain populations of *M. tuberculosis* clinical strains.

3.2. Implications for Immunization. TB remains one of the deadliest infectious diseases of our times. Despite widespread use of the BCG vaccine, the disease continues to claim 2-3 million lives per year. The need for a new vaccine has never been more urgent. In order to gain insight into the efficacy of two new vaccine candidates, two fusion proteins combining Ag85B and ESAT-6, and Ag85B and TB10.4 [6, 9, 10], respectively, the gene regions encoding for ESAT-6, TB10.4, and Ag85B proteins were analyzed for their variability. Experiments involving these two candidates have revealed them to be effective in generating protective immunity in animal models [6, 9, 10].

Yet, while animal models have played an important role in the development of new TB vaccines so far, they are not always representative of the internal human biological environment. As Flynn noted earlier, an important drawback of the murine model is that the pathology of pulmonary TB in mice is quite different from that in humans [19]. Specifically, the heterogeneity of granuloma types observed in the human host is not displayed in the mouse lung [19]. Similar difficulties arise when using other animal models. In the case of bovine infection, the pathology of TB is quite similar to human host response in granulomatous reactions, but differs with respect to cavitation [20]. Non-human primate models also represent the human pathology of TB quite well, but like cattle are limited by economic and infrastructural factors [20].

Furthermore, current preclinical studies of new TB vaccines' protection against *M. tuberculosis* infection in animal models do not take the population diversity of *M. tuberculosis* into consideration. However, as Hebert and colleagues noted previously [14], genetic diversity of *M. tuberculosis* genes can be found among clinical isolates, and such diversity may have important implications for the efficacy of the new vaccines. Thus, comparative genomics of the pathogen population stands as an additional useful tool for pre-clinical evaluation of new vaccines, providing information complementary to those from current *in vivo* and *in vitro* studies. Given the resource-demanding nature of clinical trials, comparative genomics serves as a method for predicting the potential protection of proposed vaccines candidates in the general population. Hebert and colleagues' work revealed that the PPE18 protein, part of the Mtb72f subunit vaccine, was quite variable among isolates collected from Turkey and Arkansas [14]. Analyzing the variability of antigens targeted by potential vaccines in a diverse set of isolates at the genomic level may indeed allow researchers to avoid developing a vaccine that is only variably effective like the current BCG. The findings of such study can also inform the rational selection of the study populations, with a consideration of

covering the diverse pathogen populations in clinical trials of new vaccines.

Our observation that ESAT-6, TB10.4, and Ag85B proteins were highly conserved in our study sample comprising strains from two geographically distant regions and three different principal genetic groups suggest that it is unlikely that the efficacy of Ag85B-ESAT-6 and Ag85B-TB10.4 subunit vaccines will be affected by the genetic diversity of *M. tuberculosis* population. Thus, the protective efficacy of these two novel vaccine candidates may have a wider reach than Mtb72b vaccine, which contains a highly variable antigen of *M. tuberculosis* [14]. However, our findings also indicate the need for further bioinformatics research on the three genes investigated and their specific interaction with the host immune system. While highly conserved genes are indicative of homologous protein antigens, they may also suggest a lack of selective pressure by the host immune system and thus a lack of recognition on behalf of the host's immune response. Previous examples of the inability of animal models to accurately represent the human host environment suggest that the degree to which the human host system interacts with these important peptides remains to be studied.

The goal of pre-clinical evaluation of these two vaccines may be furthered by future studies that include a larger sample of isolates from a greater range of geographic origins, making the data even more representative of the diversity of *M. tuberculosis* worldwide. The finding of this study that *esxA*, *esxH*, and *fbpB* were conserved across 88 clinical strains from Arkansas and Turkey does not confirm that they are conserved globally. Including a larger and more genetically diverse sample of isolates would address the two primary limitations of this study—the number and diversity of the isolates used.

Another factor that must be taken into account is the potential impact that host diversity may have on the global coverage of these two new TB vaccines. This study looked at the diversity of pathogen genes coding for vaccine proteins, but even uniformly conserved proteins may fail to induce protective immunity if host diversity impedes their ability to effectively bind effector immune cells. McNamara and colleagues provide an eloquent, bioinformatic approach to studying the impact that host diversity may have on Mtb27f vaccine coverage by analyzing the allelic variation of human class II MHC DRB1 proteins and its impact on proper vaccine epitope binding [15]. A similar study on Ag85B-ESAT-6 and Ag85B-TB10.4 may provide insightful information on host diversity and its impact on the coverage of these vaccines.

With these future directions in mind, the results of the present study represent an important first step in the pre-clinical bioinformatic assessment of Ag85B-ESAT-6 and Ag85B-TB10.4 vaccine candidates, and the impact that genetic diversity among their respective antigenic protein targets has on their potential success as global vaccines. The finding that *esxA, esxB,* and *fbpB* genes are highly conserved in two distinct populations suggests that Ag85B-ESAT-6 and Ag85B-TB10.4 vaccine candidates may be effective in geographically distinct areas of the world.

Acknowledgments

This study was supported by Grant NIH-R01-AI151975 from the National Institutes of Health and the Research Fund of the Office of the Vice President for Research of the University of Michigan. The Arkansas isolates and the genotyping data of the study isolated used in this study were kindly provided by Dr. Joseph H. Bates at the Arkansas Department of Health and Dr. Donald M. Cave at the University of Arkansas for Medical Sciences.

References

[1] B. R. Bloom and P. E. M. Fine, Eds., *The BCG Experience: Implications for Vaccines against Tuberculosis*, ASM Press, Washington, DC, USA, 1994.

[2] P. Andersen, "Tuberculosis vaccines—an update," *Nature Reviews Microbiology*, vol. 5, no. 7, pp. 484–487, 2007.

[3] L. Brandt, M. Elhay, I. Rosenkrands, E. B. Lindblad, and P. Andersen, "ESAT-6 subunit vaccination against Mycobacterium tuberculosis," *Infection and Immunity*, vol. 68, no. 2, pp. 791–795, 2000.

[4] C. Aagaard, J. Dietrich, M. Doherty, and P. Andersen, "TB vaccines: current status and future perspectives," *Immunology and Cell Biology*, vol. 87, no. 4, pp. 279–286, 2009.

[5] D. F. Hoft, "Tuberculosis vaccine development: goals, immunological design, and evaluation," *The Lancet*, vol. 372, no. 9633, pp. 164–175, 2008.

[6] J. Dietrich, C. Aagaard, R. Leah, et al., "Exchanging ESAT6 with TB10.4 in an Ag85B fusion molecule-based tuberculosis subunit vaccine: efficient protection and ESAT6-based sensitive monitoring of vaccine efficacy," *Journal of Immunology*, vol. 174, no. 10, pp. 6332–6339, 2005.

[7] A. S. Mustafa, Y. A. Skeiky, R. Al-Attiyah, M. R. Alderson, R. G. Hewinson, and H. M. Vordermeier, "Immunogenicity of Mycobacterium tuberculosis antigens in Mycobacterium bovis BCG-vaccinated and M. bovis-infected cattle," *Infection and Immunity*, vol. 74, no. 8, pp. 4566–4572, 2006.

[8] A. L. Sorensen, S. Nagai, G. Houen, P. Andersen, and A. B. Andersen, "Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis," *Infection and Immunity*, vol. 63, no. 5, pp. 1710–1717, 1995.

[9] A. W. Olsen, A. Williams, L. M. Okkels, G. Hatch, and P. Andersen, "Protective effect of a tuberculosis subunit vaccine based on a fusion of antigen 85B and ESAT-6 in the aerosol guinea pig model," *Infection and Immunity*, vol. 72, no. 10, pp. 6148–6150, 2004.

[10] J. A. M. Langermans, T. M. Doherty, R. A. W. Vervenne, et al., "Protection of macaques against Mycobacterium tuberculosis infection by a subunit vaccine based on a fusion protein of antigen 85B and ESAT-6," *Vaccine*, vol. 23, no. 21, pp. 2740–2750, 2005.

[11] M. Marmiesse, P. Brodin, C. Buchrieser, et al., "Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex," *Microbiology*, vol. 150, no. 2, pp. 483–496, 2004.

[12] S. Sreevatsan, X. Pan, K. E. Stockbauer, et al., "Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 18, pp. 9869–9874, 1997.

[13] I. Filliol, A. S. Motiwala, M. Cavatore, et al., "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," *Journal of Bacteriology*, vol. 188, no. 2, pp. 759–772, 2006.

[14] A. M. Hebert, S. Talarico, D. Yang, et al., "DNA polymorphisms in the pepA and PPE18 genes among clinical strains of Mycobacterium tuberculosis: implications for vaccine efficacy," *Infection and Immunity*, vol. 75, no. 12, pp. 5798–5805, 2007.

[15] L. McNamara, Y. He, and Z. Yang, "Using epitope predictions to evaluate efficacy and population coverage of the Mtb72f vaccine for tuberculosis," *BMC Immunology*, vol. 11, no. 1, article 18, 2010.

[16] K. Brudey, J. R. Driscoll, L. Rigouts, et al., "Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology," *BMC Microbiology*, vol. 6, article 23, 2006.

[17] R. Hershberg, M. Lipatov, P. M. Small, et al., "High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography," *PLoS Biology*, vol. 6, no. 12, article e311, pp. 2658–2671, 2008.

[18] N. C. Gey Van Pittius, J. Gamieldien, W. Hide, G. D. Brown, R. J. Siezen, and A. D. Beyers, "The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria," *Genome Biology*, vol. 2, no. 10, p. RESEARCH0044, 2001.

[19] J. L. Flynn, "Lessons from experimental Mycobacterium tuberculosis infections," *Microbes and Infection*, vol. 8, no. 4, pp. 1179–1188, 2006.

[20] A. S. Dharmadhikari and E. A. Nardell, "What animal models teach humans about tuberculosis," *American Journal of Respiratory Cell and Molecular Biology*, vol. 39, no. 5, pp. 503–508, 2008.