The individual's characteristics can result in susceptibility or resistance to TB, and can determine the resistance or susceptibility to specific strains

Generations – Alfredo Sosa Bravo
Acrylic/canvas
(100 x 65.5 cm)

'I am not who engenders you. They are the dead.
They are my father, his father and his ancestors
They are the ones that a long maze of love
traced from Adam and the deserts.'

To the Son (Poem) – Jorge Luis Borges

Alfredo Sosa Bravo
Sagua la Grande, Cuba

Alfredo Sosa Bravo is a painter, designer, engraver, and ceramist. Although he studied at the Academy of Fine Arts in San Alejandro, Cuba, he was mostly self taught. He started exhibiting his works in 1958.

Since 1960, he has received a number of awards in Cuba and abroad, among them the Gold Medal at the XXXIV International Competition of Contemporary Ceramic Art, Faenza, Italy, 1976.

He has also been accorded the highest cultural awards in Cuba like the Orden Félix Varela and the National Prize of Fine Arts; as well as being conferred with an honorary doctorate in Art.

Strain and Antigenic Variation in *Mycobacterium tuberculosis*: Implications for the Development of New Tools for Tuberculosis

Sebastien Gagneux and Michael J Brennan

Strain Variation in *Mycobacterium tuberculosis*

The protective efficacy of BCG vaccines against pulmonary TB ranges from 0 to 80 per cent (1–3). Various factors have been proposed to be responsible for this variation, including differences between BCG vaccine strains (4, 5), the various routes of vaccine administration (5–7), exposure to environmental mycobacteria (3), coinfection by parasites (8), and host genetic factors (9). Although strain variation in the members of the *Mycobacterium tuberculosis* complex (MTBC) that infect humans could also play a role, this possibility is only rarely acknowledged (10), and the potential importance of MTBC diversity for the development of new TB vaccines is, in general, not addressed (11). This is surprising as the importance of strain variation for vaccine efficacy has been well established for several other bacterial pathogens. In *Neisseria meningitides*, for example, ongoing mass immunization in the African meningitis belt has led to serogroup replacement (12, 13). Similarly in *Streptococcus pneumoniae*, changes in serotype-specific nasopharyngeal carriage have been reported following the deployment of novel polysaccharide–protein conjugate vaccines

(14, 15), raising concerns about the long-term effectiveness of these new interventions (16, 17). The re-emergence of pertussis has also been associated with an increase in prevalence of *Bordetella pertussis* strains distinct from the one included in the common vaccine (18, 19).

The reason strain variation in MTBC is not usually considered during vaccine design is because this organism harbours limited DNA sequence diversity compared to other bacterial pathogens. Indeed, early DNA sequencing studies reported little genotypic variation in MTBC (20), and this organism is thus referred to as genetically 'monomorphic' (21). Moreover, MTBC exhibits a clonal population structure with virtually no ongoing horizontal gene exchange (22, 23). These observations support the accepted view among TB researchers that MTBC is essentially a 'clone', and that the 'negligible' amount of diversity has no relevant phenotypic consequences (20, 24). However, recent advances in comparative genomics and DNA sequencing of MTBC are shedding new light on the diversity of this organism, and suggest a role for strain variation in TB.

Nature of Strain Variation in MTBC

With the availability of the first MTB genome sequence (25) and the development of DNA microarrays (4), various groups began to investigate strain diversity in MTBC using comparative genome hybridization. These studies revealed that genomic deletions represent an important source of genome plasticity (26–34). One study identified 68 genomic regions, including 224 genes representing 5 per cent of the H37Rv gene content, that were variably deleted in one to several of 100 clinical MTBC isolates tested (28). Several of the genes deleted in these strains encoded known antigens, suggesting that deletion of antigens might reflect a particular strategy of immune evasion in MTBC (28). Several studies have reported additional genomic deletions (also known as regions of difference or RDs) in clinical strains of MTBC which encode antigens of interest for the development of new TB diagnostics or vaccines. For example, the genomic region RD198a includes genes of a prophage and is absent in a large proportion of strains from South- and South-East Asia (28, 30, 32). Two of these prophage genes, Rv2653c and Rv2654c, are human T cell antigens and have been proposed as new TB diagnostic targets (35). However, considering the broad spread of RD198a-deleted strains, a new diagnostic tool based solely on these two antigens would be of limited use in many parts of the world where TB is highly endemic.

More recently, a study reported on a multilocus sequence analysis of 89 genes in 108 strains representative of the global diversity of MTBC (36). This study also included members of animal-adapted MTBC, such as *M. bovis*, *M. microti*, *M. caprae*, and *M. pinnipedii*. DNA sequencing of ~70kb per strain revealed that the human-adapted MTBC was more genetically diverse than previously

Figure 8.1 Global Phylogeny of MTBC
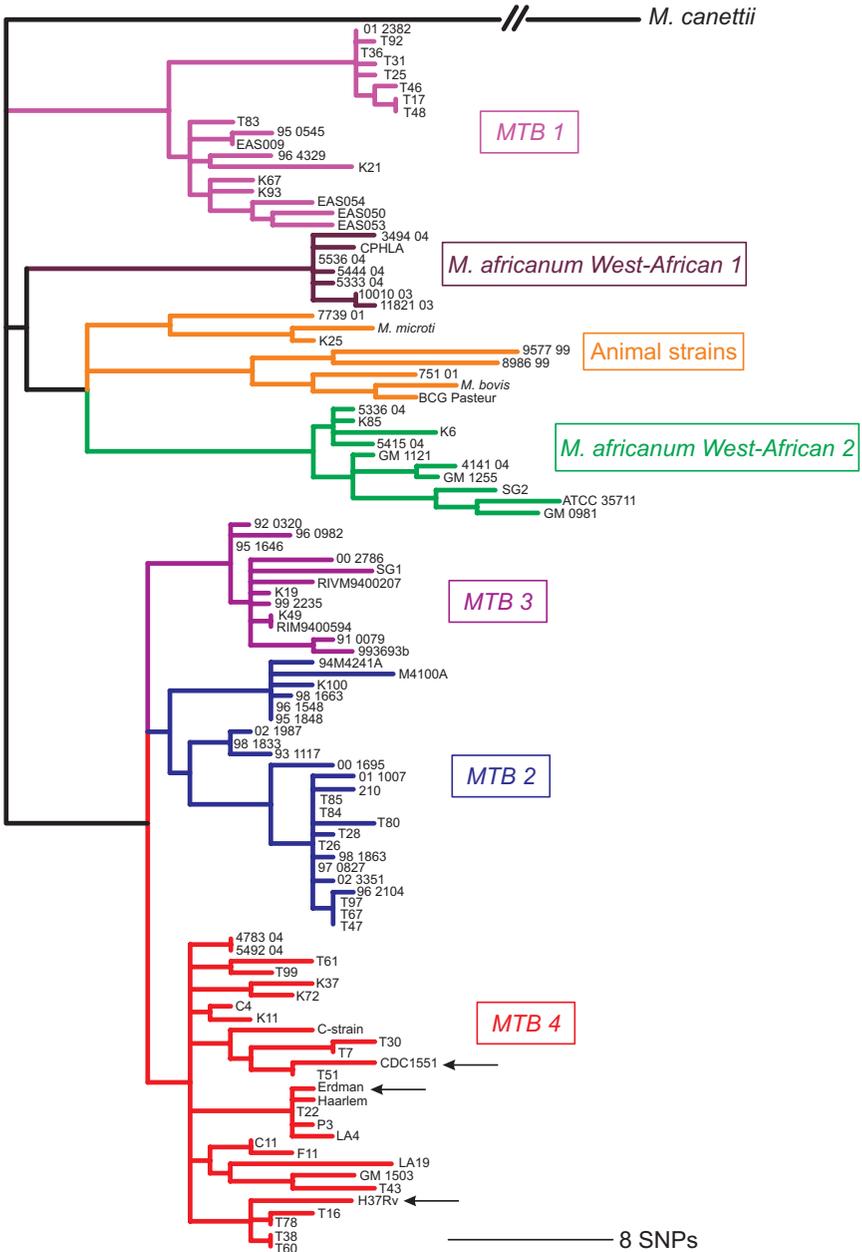
Note: Maximum parsimony phylogenetic tree based on concatenates of 89 gene sequences in each of 108 strains of MTBC, corresponding to ~70kb of DNA sequence data per strain. Six main lineages can be observed among the human-adapted forms of MTBC. All the animal-adapted MTBC cluster together and represent only a subset of all MTBC genetic diversity. Black arrows indicate canonical strains used for basic laboratory research in TB (adapted from (36)).

thought. This was illustrated by the fact that the MTBC strains adapted to animals clustered together and represented only a subset of the overall genetic diversity in the whole MTBC (Figure 8.1). Furthermore, the genetic distance between the various human-adapted strains of MTBC was equivalent to the genetic distance between human- and animal-adapted strains (36). Another important finding in the study by Hershberg et al. was that two-thirds of the single nucleotide polymorphisms (SNPs) in MTBC were non-synonymous; furthermore, that a large proportion of these non-synonymous SNPs occurred at amino acid positions that are conserved in other mycobacterial species and are thus very likely to affect protein function in MTBC (36).

Geographical Distribution of Strain Diversity

Several molecular markers have been used to define the global strain diversity of MTBC. These include genomic deletions or large sequence polymorphisms (LSPs) (32), SNPs (37–39), and those used for traditional genotyping such as spoligotyping and Mycobacterial Interspersed Repetitive Units - Variable Number Tandem Repeats (MIRU-VNTR) (40, 41). Because MTBC is a clonal organism, phylogenies inferred from these various markers are largely congruent, and the emerging consensus shows that human MTBC consists of six main lineages, including two lineages traditionally referred to as *Mycobacterium africanum* (11, 31). This global population structure of MTBC was recently confirmed by DNA sequencing of 89 genes in a globally representative collection of strains (36). This DNA sequence-based study resulted in the most detailed molecular phylogeny of MTBC to date and is reproduced in Figure 8.1 (42). These studies also revealed that the various MTBC lineages are associated with different regions of the world, as shown in Figure 8.2 [reviewed in (11)]. With respect to putative strain effects on TB vaccine efficacy, it is important to note that a large proportion of the variation in effectiveness of BCG vaccination is linked to differences in geography (1, 2). In view of the biogeographical population structure of MTBC (Figure 8.2), differences in strain genetics could also be partly responsible for the variation in BCG efficacy. For example, some of the more prominent examples of BCG vaccine failures were reported from southern India (2), where we know that the MTBC population is dominated by a particular strain lineage (pink in Figure 8.1) (30). Importantly, as can be seen from the phylogeny shown in Figure 8.1, this MTBC lineage is phylogenetically remote from the canonical strains used in basic TB research (i.e. strain H37Rv, Erdman, and CDC1551). In addition to the fact that prolonged in vitro cultivation can reduce the relevance of laboratory strains for pathogenesis research (43, 44), these observations suggest that the global diversity in clinical strains of MTBC should be taken into account during TB product development (11).

Figure 8.2 Global Distribution of the Six Main Lineages of Human-Adapted MTBC
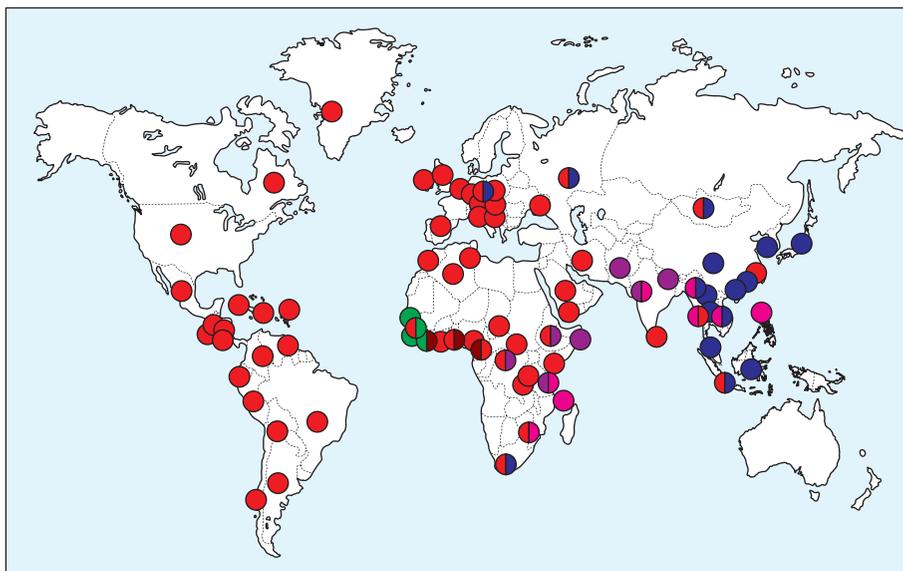
Note: Each dot represents the most frequent lineage(s) circulating in a country. Colours correspond to the lineages defined in Figure 8.1 (adapted from (32)).

Evidence for Phenotypic Effects

There is mounting evidence from in vitro and in vivo models that clinical strains of MTBC differ in immunogenicity and virulence (for comprehensive reviews see (11, 45, 46)). However, whether these differences in laboratory phenotypes have any relevance for human TB remains controversial (47). Several TB outbreaks have been reported during the last few years, where a particular strain of MTBC caused a disproportionate number of secondary cases over a relatively short time period [reviewed in (45)]. However, whether these outbreaks occurred because of the specific biological properties of these outbreak strains, or because of other reasons, has not been established. A few recent studies have shown that in addition to differences between individual strains, variation between MTBC lineages (Figure 8.1) can have a measurable impact on the outcome of TB infection and disease. A study in The Gambia reported MTBC lineage-specific differences in the progression to active disease in recently exposed household contacts of TB patients (48). In Vietnam, studies reported lineage-specific differences in the propensity to cause TB meningitis in HIV-negative adults (49, 50). Associations between MTBC lineage and human genetic polymorphisms have also been reported (49, 51), suggesting possible interactions between bacterial and human genotypes (32).

To date, the evidence for the effect of MTBC strain variation on vaccine efficacy is very limited. Two studies in mice and one study in rabbits have shown that BCG protects less well against disease caused by members of the 'Beijing family' of strains when compared to H37Rv (52–54); this strain family belongs to the blue or East-Asian lineage of MTBC (Figure 8.1). Some have speculated that widespread use of BCG vaccination might select for the spread of Beijing strains (10, 55). Intriguingly, while in some regions of South Africa the proportion of Beijing strains found among TB patient isolates has increased dramatically during the last few years (56, 57), in China the prevalence of this strain family has remained constant over many decades (58). More studies are needed to elucidate the potential impact of BCG vaccination on the spread of the Beijing and other strain families of MTBC.

A study in The Gambia found that TB patients and their household contacts infected with *M. africanum* West-Africa 2 (green lineage in Figure 8.1) exhibited an attenuated T cell response to ESAT-6 as compared to individuals infected by other MTBC lineages (59). DNA sequencing of the RD1 region revealed that strains of the *M. africanum* West-Africa 2 lineage harboured a missense mutation in Rv3879c. However, whether this mutation contributes to the attenuated ESAT-6 response in humans needs to be explored further. Nevertheless, the observation that antigen recognition can vary depending on the particular MTBC lineages circulating in patient populations could have important implications for the implementation of new TB diagnostics and vaccines.

In summary, human-adapted MTBC is more genetically diverse than previously recognized, and there is mounting evidence that this diversity can play a role in the outcome of TB infection and disease in clinical settings. Whether MTBC strain diversity will impact the effectiveness of new TB vaccines is unknown, but should be explored in the future. One important limitation of most of the comparative studies discussed in this section of the chapter is that the PE and PPE multigene families (60) were excluded from most analyses, mainly because of the technical difficulties involved in studying these gene families. In the section below, we will review some of the intriguing characteristics of this gene family, and discuss implications for TB vaccine development.

Variability in Genetic Composition, Expression, and Function of the PE_PGRS Family of MTBC Genes

Molecular analyses of MTBC have provided a number of good examples for the evolution of redundancy in the genome of this human pathogen which is

a reasonable strategy for an organism that must survive in a harsh intracellular environment within the human host. One of the best examples of genetic redundancy is found in the PE and PPE gene families which compose a noteworthy 8 per cent of the coding capacity of the MTBC genome (60). The PE nomenclature is derived from the signature Proline-Glutamic acid amino acid sequence found near the N-terminus and the Proline-Proline-Glutamic Acid sequence found at positions 7-9 in PPE genes. Over 150 highly homologous genes belonging to these two gene families containing unusual glycine rich repeat (the polymorphic Glycine Rich Sequence (PGRS) in the PE family and Major Polymorphic Tandem Repeat (MPTR) in PPEs) are present in MTBC, suggesting that there has been a dedicated selective pressure for the pathogen to maintain these sizable numbers of closely related genes. In this section, we will briefly review the PE/PE_PGRS family but readers are encouraged to examine interesting articles on the PPE genes as well (61). Research from a few laboratories have provided some clues to the importance of PE_PGRS proteins in antigenic and strain variation, immune evasion, infection of host tissues, and in induction of host cell necrosis, but much about the function of these proteins and their importance for MTBC and other mycobacteria remains unknown (62). Nevertheless, this brief narrative will explore what we do know about the PE_PGRS genes/antigens and why it is important that we continue research on this intriguing family of MTBC genes.

The PE family was named by Cole et al. (25) after the commonplace Pro-Glu (PE) residues found in abundant genes encoding for small polypeptides (~110 amino acids on average) found as individual genes or found linked to larger domains containing numerous repeat segments rich in Gly-Ala residues (the polymorphic glycine rich sequences—PGRS) (Figure 8.3A). MTB strain H37Rv contains 38 PE and 63 PE_PGRS genes (a few of these are not expressed as proteins due to frameshift mutations) and the number of PE and PE_PGRS genes can vary among MTBC strains (63). The PGRS domain can contain up to 88 Gly-Gly-Ala repeats (PE_PGRS 54). The PE domain is combined with the PGRS domain via a mostly conserved linker region containing a GRPLIGNG motif which may serve to anchor these proteins to the mycobacterial cell wall (62). The linker region may also serve as a site for specific enzymatic cleavage to release the PE and/or PGRS domains to function independently, perhaps at the surface of the bacteria. The finding that PE genes can be found in a wide variety of mycobacterial species that lack PE_PGRS genes suggests that PE_PGRS evolved through a genetic combination of a PGRS domain with existing PE genes (61). Comparative mycobacterial genomics indicates that the PE_PGRS subfamily arose after *M. avium* and before divergence led to *M. leprae* and *M. marinum*.

The addition of the heterogenous PGRS domain to a very conserved PE domain provides the organism with an extensive variety of genes for expression.

Figure 8.3 (A) Schematic Showing a typical PE_PGRS Protein with a PE and PGRS Domain Connected by a Linker Region. **(B)** Small Colony of MTB *H37Ra* Showing Clusters of Surface PE_PGRS Protein Detected by mAb 7C4 and Anti-Mouse IgG Conjugated with Fluorescein (provided by Marcela Parra, CBER, FDA). X1000

A

| PE | | PGRS |

Linker

B
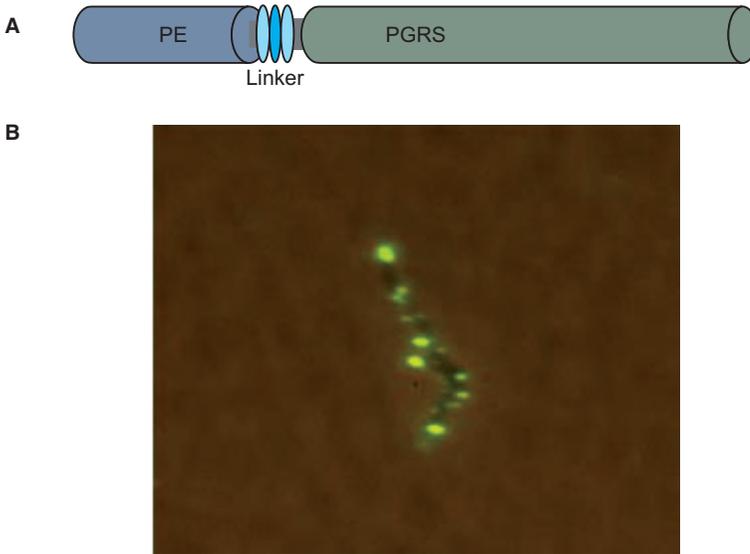
Some of these have been shown to be differentially expressed by MTBC when they are residing inside cells or where they exist under conditions that mimic intracellularity (64). The PE_PGRS genes are randomly distributed in the MTBC genome and there is some evidence to suggest that a variety of environmental signals and regulatory systems could be involved in controlling expression but no common pathway has as yet been deduced (65). It should be recognized that due to their substantial homology it is difficult to develop immunological tools that recognize individual PE_PGRS proteins but the construction of specific molecular probes for Reverse Transcription-Polymerase Chain Reaction (RT-PCR), for example, has been relatively uncomplicated. Real time PCR studies have shown that some PE_PGRS are expressed in vivo as well as in vitro in amounts similar to other genes such as soda and fbpABC which have functionally important roles in the biology of MTBC (66). Interestingly, during infection of mice it has been shown that some PE_PGRS genes (PE_PGRS 9, 16, and 30) are up-regulated over time while others (PE_PGRS26) are reduced in expression, suggesting that they may be useful markers in distinguishing between recent infection and latency (64). Comparative studies using recombinant *M. smegmatis* strains expressing different PE_PGRS proteins have demonstrated that they have different levels of persistence in macrophages and in mice and that they elicit different immune

responses (67). Therefore, the regulated expression of various PE_PGRS genes could influence the virulence of MTBC and the host immune response.

Immunomicroscopy combined with biochemical techniques indicates that PE_PGRS proteins can be found at the surface of the bacterium (68–70) (Figure 8.3B). The presence of these proteins at the surface of the bacterium implicates them in both immunological and cellular interactions with the host. One study shows that the PE domain is critical for targeting PE_PGRS proteins to the cell wall (71), while in a related study Strong et al. (72) have demonstrated that a physical association between one PE (PE25) with a specific PPE protein (PPE41) is required for efficient secretion of the PPE protein (73). These studies have important implications for understanding the functional role of PEs as chaperone-like proteins.

Variation in genetic sequences in PE_PGRS genes is probably best exemplified by the studies of Talarico et al. (74) on human clinical strains which demonstrates that about one-half of 200 clinical strains of MTB examined contain insertions or deletions, mostly within the PGRS domain of the PE_PGRS genes (PE_PGRS16, 26, and 33). Surprisingly few of these mutations interfere with expression of the proteins, suggesting that they have an important role to play in the bacterium's survival. Intolerance to changes within the PE region suggest that it is also critical for maintaining the integrity of the protein or as has been suggested (69, 71) for its localization within the bacterium. This work and other studies with PPE genes (60) indicate that regions within these repetitive genes that are likely to be important for protein function or immunological recognition are continuously changing.

Immune Evasion and Modulation of Immune Responses

A number of studies suggest that expression of PE_PGRS proteins by mycobacteria can alter the immune responses and cellular integrity of infected macrophages. Similarities between the Gly-Ala repeats found in the PGRS domain and the Epstein Barr virus protein EBNA1 which inhibits proteosome-dependent antigen processing (75) suggest that PE_PGRS proteins could play a similar role in altering antigen presentation via the MHC I pathway (62). This could aid MTBC in evading a CD8-mediated cytotoxic T cell response which is thought to be important for an effective immune response to MTBC infection (76, 77). In other experiments, intracellular expression of PE_PGRS33 by a recombinant *M. smegmatis* strain residing in macrophages has been shown to alter the pattern of cytokine expression by the phagocytic cells (67, 78). Concurrently, it also results in accelerated macrophage cell necrosis and, following infection of C57BL/6 mice with the recombinant *M. smegmatis* strain, prolonged mycobacterial persistence in mouse tissues. Certain other PE_PGRS proteins do not give these results, suggesting that PE_PGRS proteins can have selective immunopathogenic functions likely to be dependent upon sequence differences located in the PGRS region. Other studies show that

PE_PGRS33 protein can induce TNF-α production via the TLR2 receptor found on macrophages (79). Recent unpublished results also suggest a role for PE_PGRS33 in regulating host cell necrosis via specific interactions with mitochondria (Cadieux et al., submitted for publication) which may help explain the association found between mitochondrial damage and infection of macrophages with virulent MTBC strains (80).

Also of interest is the fact that ancestral PE genes are found embedded in ESAT-6 gene clusters and are therefore linked to ESX gene families (61). In a related finding, Abdullah et al. have recently shown that PE_PGRS as well as some PPE proteins are specifically secreted by the ESX-5 type VII secretion system found in *M. marinum* (81). This is of particular importance since disruption of ESX-5 results in suppression of cytokine secretion by infected macrophages and therefore implicates PE_PGRS and PPE proteins in modulating cytokine production in macrophages (66, 82).

Limited research on the role of PE_PGRS proteins has been performed in human systems but the presence of specific antibodies directed against PE_PGRS indicates that these proteins are expressed by human subjects infected with MTBC (83). It is of interest that it is very difficult to elicit antibodies towards isolated PE antigens (84). It has been proposed that the presence of a variety of slightly different PE_PGRS proteins at the surface of the bacterium, perhaps through differential regulation of gene expression, could establish a diverse antigenic environment at the surface of MTBC. Although intellectually pleasing, this hypothesis has yet to be verified and in fact antibodies produced against specific PE_PGRS proteins have been shown to cross react with a number of other PE_PGRS antigens (68, 81).

The research presented above suggests that the highly homologous PE/PE_PGRS multigene family is an important adaptation of the MTBC pathogen that is more than likely associated with one or more of the distinguishing traits of MTBC such as intracellular life, tropism for the lung, latency, or other features that establish it as a human pathogen. It is also a viable hypothesis that mycobacteria have evolved this unique set of highly variable genes to modulate host cells and the host immune response as a device to evade detection and destruction.

Conclusion

Researchers developing new tools to prevent, treat, and diagnose MTBC infection should consider the implications of strain variability and the potential impact of the differential expression of members of the multigene families like PE/PE_PGRS when designing clinical trials for new products. Several examples have been provided in this chapter which demonstrates that our lack

of understanding in these areas could result in the development of a 'new tool' that may only be effective against certain 'types' of MTBC or result in products only partially effective against TB. Comparative studies with different MTBC strains have given mixed results in studies that have investigated variation in virulence or response to drugs or vaccines. Other studies have suggested that genetic alterations or changes in expression of the multigene PE/PE_PGRS and PPE families may alter the effectiveness of vaccines. For example, numerous alterations have been found in a PPE gene in MTBC isolates, including deletions in an important immune epitope, that could affect the effectiveness of a novel TB vaccine expressing this PPE antigen (85). Together these investigations indicate that it is important to advance our understanding of strain and antigenic variation in MTBC.

Acknowledgements

❏ References

1 Fine PE. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet*, 1995; 346(8986): 1339–45.

2 Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA*, 1994; 271(9): 698–702.

3 Andersen P and Doherty TM. Opinion: The success and failure of BCG—implications for a novel tuberculosis vaccine. *Nat Rev Microbiol*, 2005.

4 Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*, 1999; 284(5419): 1520–3.

5 Davids V, Hanekom WA, Mansoor N, Gamieldien H, Gelderbloem SJ, Hawkridge A, et al. The effect of bacille Calmette-Guerin vaccine strain and route of administration on induced immune responses in vaccinated infants. *J Infect Dis*, 2006; 193(4): 531–6.

6 Hart PD. Efficacy and applicability of mass BCG vaccination in tuberculosis control. *Br Med J*, 1967; 1(540): 587–92.

7   Al Jarad N, Empey DW, and Duckworth G. Administration of the BCG vaccination using the multipuncture method in schoolchildren: a comparison with the intradermal method. *Thorax*, 1999; 54(9): 762–4.

8   Rook GA, Dheda K, and Zumla A. Opinion: Immune responses to tuberculosis in developing countries: implications for new vaccines. *Nat Rev Immunol*, 2005; 5(8): 661–7.

9   Fortin A, Abel L, Casanova JL, and Gros P. Host genetics of mycobacterial diseases in mice and men: forward genetic studies of BCG-osis and tuberculosis. *Annu Rev Genomics Hum Genet*, 2007; 8: 163–92.

10 Abebe F and Bjune G. The emergence of Beijing family genotypes of Mycobacterium tuberculosis and low-level protection by bacille Calmette-Guerin (BCG) vaccines: is there a link? *Clin Exp Immunol*, 2006; 145(3): 389–97.

11 Gagneux S and Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. *Lancet Infect Dis*, 2007; 7(5): 328–37.

12 Gagneux SP, Hodgson A, Smith TA, Wirth T, Ehrhard I, Morelli G, et al. Prospective study of a serogroup X Neisseria meningitidis outbreak in northern Ghana. *J Infect Dis*, 2002; 185(5): 618–26.

13 Forgor AA, Leimkugel J, Hodgson A, Bugri A, Dangy JP, Gagneux S, et al. Emergence of W135 meningococcal meningitis in Ghana. *Trop Med Int Health*, 2005; 10(12): 1229–34.

14 Mbelle N, Huebner RE, Wasas AD, Kimura A, Chang I, Klugman KP. Immunogenicity and impact on nasopharyngeal carriage of a nonavalent pneumococcal conjugate vaccine. *J Infect Dis*, 1999; 180(4): 1171–6.

15 Bogaert D, van Belkum A, Sluijter M, Luijendijk A, de Groot R, Rumke HC, et al. Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children. *Lancet*, 2004; 363(9424): 1871–2.

16 Bogaert D, De Groot R, and Hermans PW. Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis*, 2004; 4(3): 144–54.

17 Spratt BG and Greenwood BM. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet*, 2000; 356(9237): 1210–11.

18 Elomaa, A, Advani A, Donnelly D, Antila M, Mertsola J, Hallander H, et al. Strain variation among Bordetella pertussis isolates in Finland, where the whole-cell pertussis vaccine has been used for 50 years. *J Clin Microbiol*, 2005; 43(8): 3681–7.

19 Van Loo IH and Mooi FR. Changes in the Dutch Bordetella pertussis population in the first 20 years after the introduction of whole-cell vaccines. *Microbiology*, 2002; 148(Pt 7): 2011–18.

20 Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA*, 1997; 94(18): 9869–74.

21 Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol*, 2008; 62: 53–70.

22 Supply P, Warren RM, Bañuls AL, Lesjean S, van Der Spuy GD, Lewis LA, et al. Linkage disequilibrium between minisatellite loci supports clonal evolution of Mycobacterium tuberculosis in a high tuberculosis incidence area. *Mol Microbiol*, 2003; 47(2): 529–38.

23 Gagneux S and Small PM. Molecular evolution of mycobacteria, in Handbook of tuberculosis, SH Kaufmann and E Rubin, editors. Wiley VCH: Weinheim, 2008.

24 Musser JM, Amin A, and Ramaswamy S. Negligible genetic diversity of mycobacterium tuberculosis host immune system protein targets: evidence of limited selective pressure. *Genetics*, 2000; 155(1): 7–16.

25 Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 1998; 393(6685): 537–44.

26 Mostowy S, Cousins D, Brinkman J, Aranaz A, and Behr MA. Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. *J Infect Dis*, 2002; 186(1): 74–80.

27 Nguyen D, Brassard P, Menzies D, Thibert L, Warren R, Mostowy S, et al. Genomic characterization of an endemic Mycobacterium tuberculosis strain: evolutionary and epidemiologic implications. *J Clin Microbiol*, 2004; 42(6): 2573–80.

28 Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, et al. Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci USA*, 2004;101(14): 4865–70.

29 Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere YO, Kreiswirth BN, Van Soolingen D, et al. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. *J Clin Microbiol*, 2005; 43(7): 3185–91.

30 Narayanan S, Gagneux S, Hari L, Tsolaki AG, Rajasekhar S, Narayanan PR, et al. Genomic interrogation of ancestral Mycobacterium tuberculosis from south India. *Infect Genet Evol*, 2008; 8(4): 474–83.

31 Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, et al. Genomic analysis distinguishes Mycobacterium africanum. *J Clin Microbiol*, 2004; 42(8): 3594–9.

32 Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proc Natl Acad Sci USA*, 2006; 103: 2869–73.

33 Kato-Maeda M, Bifani PJ, Kreiswirth BN, and Small PM. Comparing genomes within the species Mycobacterium tuberculosis. *Genome Res*, 2001; 11(4): 547–54.

34 Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc Natl Acad Sci USA*, 2002; 99(6): 3684–9.

35 Aagaard C, Brock I, Olsen A, Ottenhoff TH, Weldingh K, and Andersen P. Mapping immune reactivity toward Rv2653 and Rv2654: two novel low-molecular-mass antigens found specifically in the Mycobacterium tuberculosis complex. *J Infect Dis*, 2004; 189(5): 812–19.

36 Hershberg R, Mikhail L, Small PM, Sheffer H, Niemann S, Homolka S, et al. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biol*, 2008; 6(12): e311.

37 Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, Bobadilla del Valle M, 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. *J Bacteriol*, 2006; 188(2): 759–72.

38 Gutacker MM, Mathema B, Soini H, Shashkina E,. Kreiswirth BN, Graviss EA, et al. Single-Nucleotide Polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. *J Infect Dis*, 2006; 193(1): 121–8.

39 Baker L, Brown T, Maiden MC, and Drobniewski F. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. *Emerg Infect Dis*, 2004; 10(9): 1568–77.

40 Comas I, Homolka S, Niemann S, and Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in mycobacterium tuberculosis highlights the limitations of current methodologies. *PLoS ONE*, 2009; 4(11): e7815.

41 Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, et al. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol*, 2006; 6(1): 23.

42 Smith NH, Hewinson RG, Kremer K, Brosch R, and Gordon SV. Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol*, 2009.

43 Fux CA, Shirtliff M, Stoodley P, and Costerton JW. Can laboratory reference strains mirror 'real-world' pathogenesis? *Trends Microbiol*, 2005; 13(2): p. 58–63.

44 Domenech P and Reed MB. Rapid and spontaneous loss of Phthiocerol Dimycocerosate (PDIM) from Mycobacterium tuberculosis grown in vitro: implications for virulence studies. *Microbiology*, 2009.

45 Nicol MP and Wilkinson RJ. The clinical consequences of strain diversity in Mycobacterium tuberculosis. *Trans R Soc Trop Med Hyg*, 2008.

46 Malik AN and Godfrey-Faussett P. Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease. *Lancet Infect Dis*, 2005; 5(3): 174–83.

47 Comas I and Gagneux S. The past and future of tuberculosis research. *PLoS Pathog*, 2009; 5(10): e1000600.

48 de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, et al. Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. *J Infect Dis*, 2008.

49 Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, et al. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. *PLoS Pathog*, 2008; 4(3): e1000034.

50 Thwaites G, Caws M, Chau TTH, D'Sa A, Lan NTN, Huyen MNT, et al. The relationship between Mycobacterium tuberculosis genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J Clin Microbiol*, 2008.

51 Intemann CD, Thye T, Niemann S, Browne EN, Amanua Chinbuah M, Enimil A, et al. Autophagy gene variant IRGM -261T contributes to protection from tuberculosis caused by Mycobacterium tuberculosis but not by M. africanum strains. *PLoS Pathog*, 2009; 5(9): e1000577.

52 Grode L, Seiler P, Baumann S, Hess J, Brinkmann V, Nasser Eddine A, et al. Increased vaccine efficacy against tuberculosis of recombinant Mycobacterium bovis bacille Calmette-Guerin mutants that secrete listeriolysin. *J Clin Invest*, 2005; 115(9): 2472–9.

53 Lopez B, Aguilar D, Orozco H, Burger M, Espitia C, Ritacco V, et al. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. *Clin Exp Immunol*, 2003; 133(1): 30–7.

54 Tsenova L, Harbacheuski R, Sung N, Ellison E, Fallows D, Kaplan G. BCG vaccination confers poor protection against M. tuberculosis HN878-induced central nervous system disease. *Vaccine*, 2006.

55 van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, et al. Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. *J Clin Microbiol*, 1995; 33(12): 3234–8.

56 Cowley D, Govender D, February B, Wolfe M, Steyn L, Evans J, et al. Recent and Rapid Emergence of W-Beijing Strains of Mycobacterium tuberculosis in Cape Town, South Africa. *Clin Infect Dis*, 2008.

57 van der Spuy GD, Kremer K, Ndabambi SL, Beyers N, Dunbar R, Marais BJ, et al. Changing Mycobacterium tuberculosis population highlights clade-specific pathogenic characteristics. *Tuberculosis (Edinb)*, 2008.

58 Qian L, Van Embden JD, Van Der Zanden AG, Weltevreden EF, Duanmu H, and Douglas JT. Retrospective analysis of the Beijing family of Mycobacterium tuberculosis in preserved lung tissues. *J Clin Microbiol*, 1999; 37(2): 471–4.

59 de Jong BC, Hill PC, Brookes RH, Gagneux S, Jeffries DJ, Otu JK, et al. Mycobacterium africanum elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. *J Infect Dis*, 2006; 193(9): 1279–86.

60 Brennan MJ, Gey van Pittius NC, and Espitia C. The PE and PPE multigene families of mycobacteria, in tuberculosis and the tubercle bacillus, S Cole, et al., editors, Washington DC: ASM Press, 2005, pp. 513–25.

61 Gey van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD, and Warren R. Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. *BMC Evol Biol*, 2006; 6(1): 95.

62 Brennan MJ and Delogu G. The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends Microbiol*, 2002; 10(5): 246–9.

63 Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-Genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. *J Bacteriol*, 2002; 184(19): 5479–90.

64 Dheenadhayalan V, Delogu G, Sanguinetti M, Fadda G, and Brennan MJ. Variable expression patterns of Mycobacterium tuberculosis PE_PGRS genes: evidence that PE_PGRS16 and PE_PGRS26 are inversely regulated in vivo. *J Bacteriol*, 2006; 188(10): 3721–5.

65 Voskuil MI, Schnappinger D, Rutherford R, Liu Y, and Schoolnik GK. Regulation of the Mycobacterium tuberculosis PE/PPE genes. *Tuberculosis (Edinb)*, 2004; 84(3–4): 256–62.

66 Delogu G, Sanguinetti M, Pusceddu C, Bua A, Brennan MJ, Zanetti S, et al. PE_PGRS proteins are differentially expressed by Mycobacterium tuberculosis in host tissues. *Microbes Infect*, 2006; 8(8): 2061–7.

67 Singh PP, Parra M, Cadieux N, and Brennan MJ. A comparative study of host response to three Mycobacterium tuberculosis PE_PGRS proteins. *Microbiology*, 2008; 154(Pt 11): 3469–79.

68 Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, and Cole ST. Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? *Mol Microbiol*, 2002; 44(1): 9–19.

69 Delogu G, Pusceddu C, Bua A, Fadda G, Brennan MJ, and Zanetti S. Rv1818c-encoded PE_PGRS protein of Mycobacterium tuberculosis is surface exposed and influences bacterial cell structure. *Mol Microbiol*, 2004; 52(3): 725–33.

70 Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, et al. Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun*, 2001; 69(12): 7326–33.

71 Cascioferro A, Delogu G, Colone M, Sali M, Stringaro A, Arancia G, et al. PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. *Mol Microbiol*, 2007; 66(6): 1536–47.

72 Strong M and Goulding CW. Structural proteomics and computational analysis of a deadly pathogen: combating Mycobacterium tuberculosis from multiple fronts. *Methods Biochem Anal*, 2006; 49: 245–69.

73 Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, et al. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol Microbiol*, 2006; 62(3): 667–79.

74 Talarico S, Zhang L, Marrs CF, Foxman B, Cave MD, Brennan MJ, et al. Mycobacterium tuberculosis PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates. *Tuberculosis (Edinb)*, 2008.

75 Levitskaya J, Sharipo A, Leonchiks A, Ciechanover A, and Masucci MG. Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. *Proc Natl Acad Sci USA*, 1997; 94(23): 12616–21.

76 Flynn JL and Chan J. Immunology of tuberculosis. *Annu Rev Immunol*, 2001; 19: 93–129.

77 Kaufmann SH. How can immunology contribute to the control of tuberculosis? *Nat Rev Immunol*, 2001; 1(1): 20–30.

78 Dheenadhayalan V, Delogu G, and Brennan MJ. Expression of the PE_PGRS 33 protein in Mycobacterium smegmatis triggers necrosis in macrophages and enhanced mycobacterial survival. *Microbes Infect*, 2006; 8(1): 262–72.

79 Basu S, Pathak SK, Banerjee A, Pathak S, Bhattacharyya A, Yang Z, et al. Execution of macrophage apoptosis by PE_PGRS33 of Mycobacterium tuberculosis is mediated by toll-like receptor 2-dependent release of tumor necrosis factor-alpha. *J Biol Chem*, 2006.

80 Chen M, Gan H, and Remold HG. A mechanism of virulence: virulent Mycobacterium tuberculosis strain H37Rv, but not attenuated H37Ra, causes significant mitochondrial inner membrane disruption in macrophages leading to necrosis. *J Immunol*, 2006; 176(6): 3707–16.

81 Abdallah AM, Gey van Pittius NC, Champion PAD, Cox J, Luirink J, Vandenbroucke-Grauls CMJE, et al. PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. *Mol Microbiol*, 2009.

82 Bottai D and Brosch R. Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families. *Mol Microbiol*, 2009; 73(3): 325–8.

83 Singh KK, Zhang X, Patibandla AS, Chien P Jr, and Laal S. Antigens of Mycobacterium tuberculosis expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. *Infect Immun*, 2001; 69(6): 4185–91.

84 Parra M, Cadieux N, Pickett T, Dheenadhayalan V, and Brennan MJ. A PE protein expressed by Mycobacterium avium is an effective T-cell immunogen. *Infect Immun*, 2006; 74(1): 786–9.

85 Hebert AM, Talarico S, Yang D, Durmaz R, Marrs CF, Zhang L, et al. DNA polymorphisms in the pepA and PPE18 genes among clinical strains of Mycobacterium tuberculosis: implications for vaccine efficacy. *Infect Immun*, 2007; 75(12): 5798–805.