

RESEARCH ARTICLE

Open Access

Single nucleotide polymorphism in the genes of *mce1* and *mce4* operons of *Mycobacterium tuberculosis*: analysis of clinical isolates and standard reference strains

Rashmi Pasricha¹, Amita Chandolia¹, Prija Ponnann², Neeraj Kumar Saini¹, Sangeeta Sharma³, Madhu Chopra⁴, Mandira Varma Basil¹, Vani Brahmachari⁴, Mridula Bose^{1*}

Abstract

Background: The presence of four mammalian cell entry (*mce*) operons in *Mycobacterium tuberculosis* suggests the essentiality of the functions of the genes in these operons. The differential expression of the four *mce* operons in different phases of *in vitro* growth and in infected animals reported earlier from our laboratory further justifies the apparent redundancy for these genes in the genome.

Here we investigate the extent of polymorphism in eight genes in the *mce1* and *mce4* operons of *M. tuberculosis* from four standard reference strains (H37Rv, H37Ra, LVS (Low Virulent Strain) and BCG) and 112 clinical isolates varying in their drug susceptibility profile, analysed by direct sequencing and Sequenom MassARRAY platform.

Results: We discovered 20 single nucleotide polymorphisms (SNPs) in the two operons. The comparative analysis of the genes of *mce1* and *mce4* operons revealed that *yrbE1A* [*Rv0167*] was most polymorphic in *mce1* operon while *yrbE4A* [*Rv3501c*] and *lprN* [*Rv3495c*] had the highest number of SNPs in the *mce4* operon. Of 20 SNPs, 12 were found to be nonsynonymous and were further analysed for their pathological relevance to *M. tuberculosis* using web servers PolyPhen and PMut, which predicted five deleterious nonsynonymous SNPs. A mutation from proline to serine at position 359 of the native Mce1A protein was most deleterious as predicted by both PolyPhen and PMut servers. Energy minimization of the structure of native Mce1A protein and mutated protein was performed using InsightII. The mutated Mce1A protein showed structural changes that could account for the effects of this mutation.

Conclusions: Our results show that SNPs in the coding sequences of *mce1* and *mce4* operons in clinical isolates can be significantly high. Moreover, *mce4* operon is significantly more polymorphic than *mce1* operon ($p < 0.001$). However, the frequency of nonsynonymous substitutions is higher in *mce1* operon and synonymous substitutions are more in *mce4* operon. *In silico* modeling predict that nonsynonymous SNP at *mce1A* [*Rv0169*], a virulence gene could play a pivotal role in causing functional changes in *M. tuberculosis* that may reflect upon the biology of the bacteria.

Background

The variability in the genome sequence of *M. tuberculosis* between clinical isolates has been analysed earlier and variability in the number and site of integration of transposable element IS6110 is well documented [1]. There are also reports on the analysis of whole genome SNPs in

mycobacteria [2]. Compared to many other bacterial species, *M. tuberculosis* exhibits very little genomic sequence variation [3]. However, there is increasing evidence that even this limited inter-strain genetic variability is biologically significant [4]. *M. tuberculosis* infection in animal models has shown a range of immune responses and variable degrees of virulence depending on the infecting strain [5,6]. In the majority of humans, an effective immune response develops after infection with *M. tuberculosis* and restricts the spread of the pathogen and clinical

* Correspondence: mridulabose@hotmail.com

¹Department of Microbiology, Vallabhbhai Patel Chest Institute, University of Delhi, Delhi-110007, India

Full list of author information is available at the end of the article

manifestation of the disease is seen in less than 10% of those infected. Clinical tuberculosis is influenced by variability in the host's genetic background, immune status, diet, social and environmental factors [7,8]. However, little is known about the bacterial factors, especially, genetic diversity in bacterial virulence factors contributing to variable host responses.

The expression of *mce* genes is of importance for the virulence of mycobacteria [9,10]. The presence of four copies of *mce* genes in four operons each consisting of eight genes [11] and the differential expression of *mce1* and *mce4* operons points towards functional importance of these operons [9,12]. Interestingly, the domain organization in the genes of all the four operons is similar. This conservative arrangement may be of strategic significance to the biology of *M. tuberculosis*. The antigenic and immunogenic effects of *mce* proteins in nature suggest that the variation in amino acid sequence of these proteins may affect host response, apart from their effect on functions of these proteins [13,14]. In the light of these observations, we initiated the present study to understand the possible importance of genetic diversity in the *mce* operon genes which have a role in the pathogenesis of *M. tuberculosis*. Polymorphism in the genes of *mce1* and *mce4* operons in 112 clinical isolates of *M. tuberculosis* was analysed to understand and relate the effect of the genetic variability to structural changes in the proteins by computational methods.

Results

Single nucleotide polymorphism in *mce* operons

We used a discovery platform consisting of four standard reference strains (H37Rv, H37Ra, LVS (Low Virulent Strain) and BCG) and 12 clinical isolates selected at random. Overlapping primers were designed to map eight genes each of *mce1* and *mce4* operons (Figure 1). We identified 7 SNPs in *mce1* operon; 6 of these were nonsynonymous and one was synonymous substitution (Table 1). 100 clinical isolates were then genotyped for these SNPs on Sequenom MassARRAY platform. Among 8 genes of *mce1* operon, *yrbE1A* [*Rv0167*] gene was significantly more polymorphic as compared to the other seven genes, with approximately 26% of clinical isolates showing the polymorphic allele. In comparison 13 SNPs were identified in *mce4* operon (Table 2), of which 6 were nonsynonymous and 7 were synonymous SNPs. In *mce4* operon significant polymorphism was observed in clinical isolates at *yrbE4A* [*Rv3501c*] and *lprN* [*Rv3495c*] genes with 25.50% and 26.50% SNP respectively.

Effect of SNPs on codon usage in *mce* operons

The preferential usage of codons for different amino acids in various organisms including *M. tuberculosis* is well known. The codon bias influences the translational

efficiency in these organisms [15]. Therefore, we analysed the codon usage in *M. tuberculosis* for synonymous changes observed in both *mce1* and *mce4* operons. Analysis revealed that codons of amino acids were changed to the next preferred codon (Table 3). It is possible that such altered preference for certain codons would alter the expression of the respective proteins.

Prediction of functional consequences of nonsynonymous SNPs by PolyPhen and PMut servers

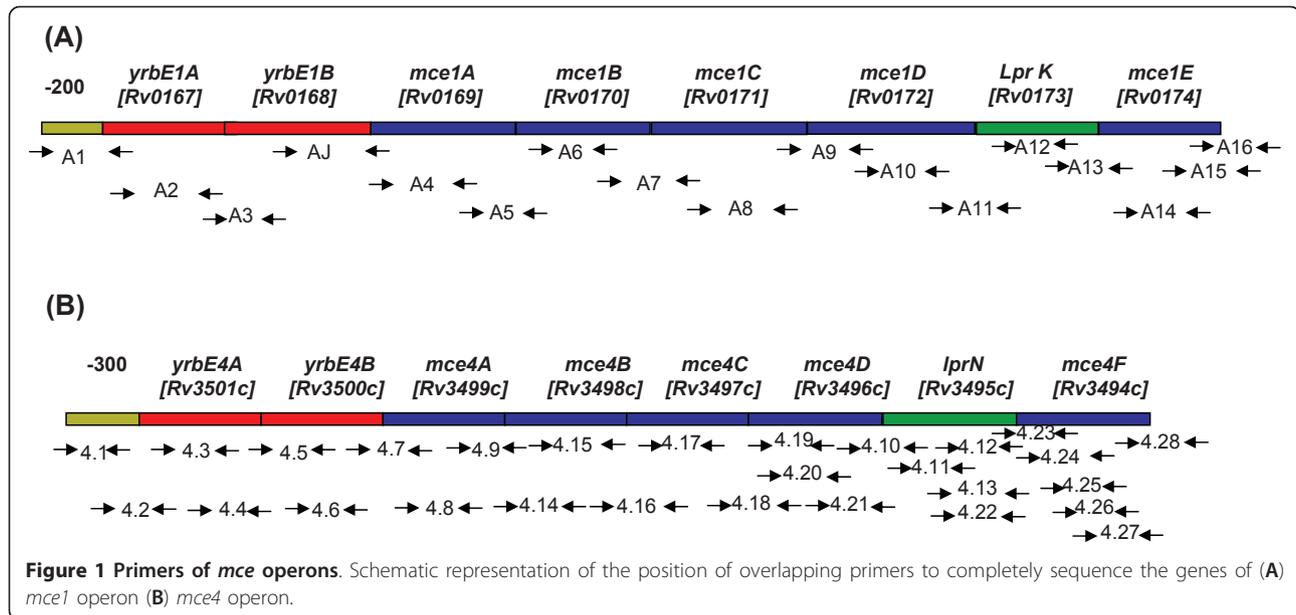
The functional impact of 12 nonsynonymous SNPs in proteins of *mce1* and *mce4* operons was analyzed using PolyPhen <http://genetics.bwh.harvard.edu/pph/> and PMut <http://mmb2.pcb.ub.es:8080/PMut/> servers. Of the 12 nonsynonymous SNPs studied, 5 nonsynonymous SNPs were predicted to be deleterious to the organism by both PolyPhen and PMut programs. These nonsynonymous SNPs were located in the genes *yrbE1B* [*Rv0168*] (NN output; 0.84, PSIC score; 1.6), *mce1A* [*Rv0169*] (NN output; 0.84, PSIC score; 2.04), *mce1B* [*Rv0170*] (NN output; 0.59, PSIC score; 1.6), *lprN* [*Rv3495c*] (NN output; 0.55, PSIC score; 1.73) and *mce4F* [*Rv3494c*] (NN output; 0.52, PSIC score; 2.01). Whereas the other 7 nonsynonymous SNPs had NN output < 0.5 and PSIC score < 1.5. The highest score in this analysis was for *mce1A* gene with C1075T mutation resulting in substitution of proline to serine at 359 amino acid position. Thus, C1075T was considered to be the most deleterious mutation by PolyPhen and PMut programs.

Modeling of mutated protein structure

We selected C1075T (Pro359Ser) polymorphism in *mce1A* gene as shown in Table 1 for further structural analysis. The substitution is positioned at 359 amino acid and we have mapped this in the three dimensional structure [PDB: 1NA9] [16]. Mutation at the specified position was performed by InsightII/Biopolymer and energy minimizations were performed by InsightII/Discover module for both the native structure [PDB: 1NA9] and mutant modeled structure (Pro359Ser). This structural analysis shows that the native (Figure 2A) and the mutant (Figure 2B) protein structure has an RMSD of 3.07 Å. It is interesting to observe that, in the native structure, Proline359 is a part of the helical conformation while the mutated counterpart (Pro359Ser) has a loop structure at this position (Figure 3). Perturbation in the hydrogen bonds as indicated in the HB plots (Figure 4A and 4B) could be attributed to the conformational changes at Ser 359 position and other regions of mutant protein.

Correlation between drug resistance and SNPs

Approximately 59 drug resistant (DR) and 22 drug sensitive (DS) clinical isolates were analysed further. Four first line drugs namely isoniazid, rifampicin, ethambutol



and streptomycin were taken into account to characterize the isolates. The sensitive isolates were sensitive to all the four antitubercular drugs while the resistant isolates were resistant to atleast one drug. The comparison between the two categories revealed that *mce1* and *mce4* operon genes were significantly more polymorphic in

DS clinical isolates than DR isolates (*, $p < 0.05$) (Figure 5A) and (**, $p < 0.01$) (Figure 5B) respectively.

Among 59 DR clinical isolates, 19 were MDR TB (Multi drug resistant, at least to isoniazid and rifampicin). Among 19 MDR TB clinical isolates, polymorphism was observed in *yrbE1A* (15.78%) and *yrbE1B* (5.26%)

Table 1 Polymorphisms in the genes of *mce1* operon

| | | <i>mce1</i> operon | | | | | | | |
|------------------------------------|---|--------------------|------------|--|--------------|--------------|---------------|------------------|--|
| Gene Name (Accession Number) | Nucleotide Change [GenBank Accession Number] | Amino Acid Change | | Frequency Distribution of polymorphism (%) | | | | | |
| | | Non Synonymous | Synonymous | All isolates n = 112 | DS n = 22 | DR n = 59 | SDR n = 15 | MDR TB n = 19 | |
| <i>yrbE1A</i> [<i>Rv0167</i>] | C14T [GenBank:HQ901088] | Thr5Ile | NONE | (25.96) | (29.16) | (29.09) | (41.76) | (15.78) | |
| <i>yrbE1B</i> [<i>Rv0168</i>] | T154G [GenBank:HQ901089] | Tyr52Asp | NONE | (0.9) | NONE | (1.72) | NONE | (5.26) | |
| <i>mce1A</i> [<i>Rv0169</i>] | C1075T C1323T [GenBank:HQ901082] | Pr0359Ser | Tyr441Tyr | (1.87) | (4) | NONE | NONE | NONE | |
| <i>mce1B</i> [<i>Rv0170</i>] | T536C [GenBank:HQ901085] | Ile179Thr | NONE | (0.9) | (3.8) | NONE | NONE | NONE | |
| <i>mce1C</i> [<i>Rv0171</i>] | G636C [GenBank: HQ901086] | Glu212Asp | NONE | (0.9) | (3.8) | NONE | NONE | NONE | |
| <i>mce1D</i> [<i>Rv0172</i>] | NONE | NONE | NONE | NONE | NONE | NONE | NONE | NONE | |
| <i>lprK</i> [<i>Rv0173</i>] | NONE | NONE | NONE | NONE | NONE | NONE | NONE | NONE | |
| <i>mce1F</i> [<i>Rv0174</i>] | G129T [GenBank: HQ901083] | Lys43Asn | NONE | (0.9) | (4) | NONE | NONE | NONE | |

Frequency of single nucleotide polymorphisms detected in the genes of *mce1* operon. The nucleotide changes and the corresponding changes in amino acids are shown here. The frequency of SNPs was calculated from 112 clinical isolates. The data has been subdivided according to the drug susceptibility profile. The single letter nucleotide designations used are as follows: A, adenine; C, cytosine; G, guanine and T, thymidine. The three letter amino acid designations used are as follows: Thr, threonine; Ile, isoleucine; Tyr, tyrosine; Asp, aspartic acid; Pro, proline; Ser, serine; Glu, glutamic acid; Lys, lysine and Asn, asparagine. DS: drug sensitive, DR: drug resistant, SDR: single drug resistant, MDR TB: Multi drug resistant.

Table 2 Polymorphisms in the genes of *mce4* operon

| Gene Name (Accession Number) | Nucleotide Change [GenBank Accession Number] | <i>mce4</i> operon | | Frequency Distribution of polymorphism (%) | | | | |
|---------------------------------|---|-----------------------|----------------------|--|--------------|--------------|---------------|------------------|
| | | Amino Acid Change | | All isolates n = 112 | DS n = 22 | DR n = 59 | SDR n = 15 | MDR TB n = 19 |
| | | Non Synonymous | Synonymous | | | | | |
| <i>yrbE4A</i> [Rv3501c] | G18T C753A [GenBank: HQ901084] | NONE | Ala6Ala Ile251Ile | (25.49) | (20.83) | (29.62) | (41.76) | (21.05) |
| <i>yrbE4B</i> [Rv3500c] | C21T C624T [GenBank: HQ901090] | NONE | Ile7Ile Pro208Pro | (3.7) | (8) | (3.44) | (5.88) | NONE |
| <i>mce4A</i> [Rv3499c] | T32G C873T [GenBank: HQ901091] | Val11Gly | Phe291Phe | (2.25) | (4.55) | NONE | NONE | NONE |
| <i>mce4B</i> [Rv3498c] | NONE | NONE | NONE | NONE | NONE | NONE | NONE | NONE |
| <i>mce4C</i> [Rv3497c] | A136C C571A [GenBank: HQ901092] | Thr46Pro Arg191Ser | NONE | (3.75) | (8.33) | NONE | (5.88) | (5.26) |
| <i>mce4D</i> [Rv3496c] | G35A [GenBank: HQ901093] | Arg12Gln | NONE | (1.25) | (7.69) | NONE | NONE | NONE |
| <i>lprN</i> [Rv3495c] | C798T C1016A [GenBank: HQ901094] | Thr339Lys | Ala266Ala | (26.47) | (29.09) | (30.9) | (31.57) | (31.07) |
| <i>mce4F</i> [Rv3494c] | C117A C1214T [GenBank: HQ901087] | Pro405Lys | Thr39Thr | (8.75) | (9.09) | (7.3) | (10.52) | (5.09) |

Frequency of single nucleotide polymorphisms detected in the genes of *mce4* operon. The nucleotide changes and the corresponding changes in amino acids are shown here. The frequency of SNPs was calculated from 112 clinical isolates. The data has been subdivided according to the drug susceptibility profile. The single letter nucleotide designations used are as follows: A, adenine; C, cytosine; G, guanine and T, thymidine. The three letter amino acid designations used are as follows Ala, alanine; Ile, isoleucine; Pro, proline; Val, valine; Gly, glycine; Phe, phenylalanine; Thr, threonine; Arg, arginine; Ser; serine; Gln, glutamine and Lys, lysine.

DS: drug sensitive, DR: drug resistant, SDR: single drug resistant, MDR TB: Multi drug resistant.

genes of *mce1* operon; and in *yrbE4A* (21.05%), *mce4B* (5.26%), *lprN* (31.57%) and *mce4F* (10.52%) genes of *mce4* operon. Of the 15 single drug resistant (SDR) clinical isolates studied, polymorphism was observed in *yrbE1A* (41.76%) gene of *mce1* operon and in *yrbE4A* (41.76%), *yrbE4B* (5.88%), *mce4C* (5.26%), *lprN* (35.29%) and *mce4F* (5.88%) genes of *mce4* operon. Interestingly, *mce* genes were significantly more polymorphic in SDR strains than MDR TB strains in both *mce1* and *mce4* operons. (**, $p < 0.01$ and ***, $p < 0.001$ respectively).

Table 3 Codon usage in *mce1* and *mce4* operons

| Operon | Gene name (Accession Number) | Wild type codon | Polymorphic codon |
|--------------------|------------------------------|-----------------|-------------------|
| <i>mce1</i> operon | <i>mce1A</i> [Rv0169] | TAC | TAT |
| | <i>yrbE4A</i> [Rv3501c] | GCG ATC | GCT ATA |
| <i>mce4</i> operon | <i>yrbE4B</i> [Rv3500c] | ATC CCC | ATT CCT |
| | <i>mce4A</i> [Rv3499c] | TTC | TTT |
| | <i>lprN</i> [Rv3495c] | GCC | GCT |
| | <i>mce4F</i> [Rv3494c] | ACC | ACA |

The codon usage in the polymorphic regions is shown here. The synonymous changes in the nucleotide sequence, when analysed bioinformatically through Gene Runner software version 3.05 (Hastings Software, Inc.) predicts the usage of less preferred codon which could reflect upon the expression efficiency of the protein encoded by the gene.

Nucleotide highlighted in bold indicates the altered nucleotide.

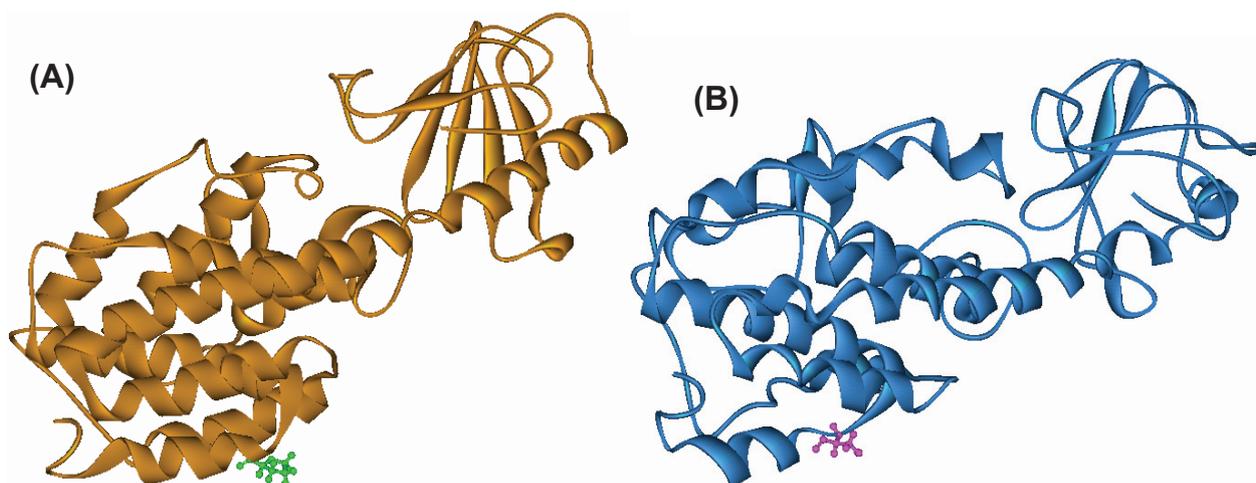


Figure 2 Wild and mutant protein structure of Mce1A. Structure of (A) wild (orange ribbon) and (B) Pro359Ser mutant (blue ribbon) proteins showing Pro359 (green) in wild protein and Ser359 (pink) in the mutant protein represented in ball and stick. The figure was prepared using Discovery studio 2.5 (DS Modeling 2.5, Accelrys Inc.: San Diego, CA).

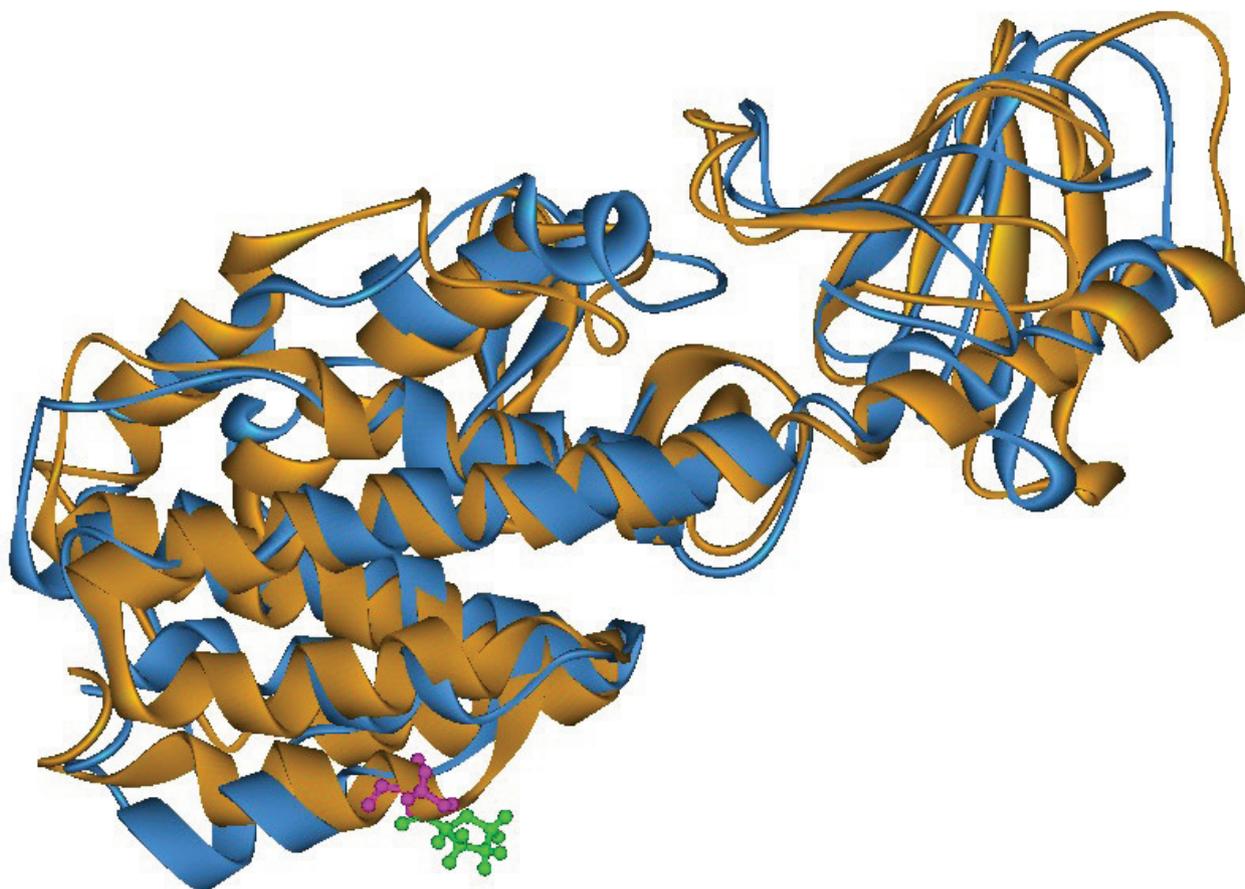
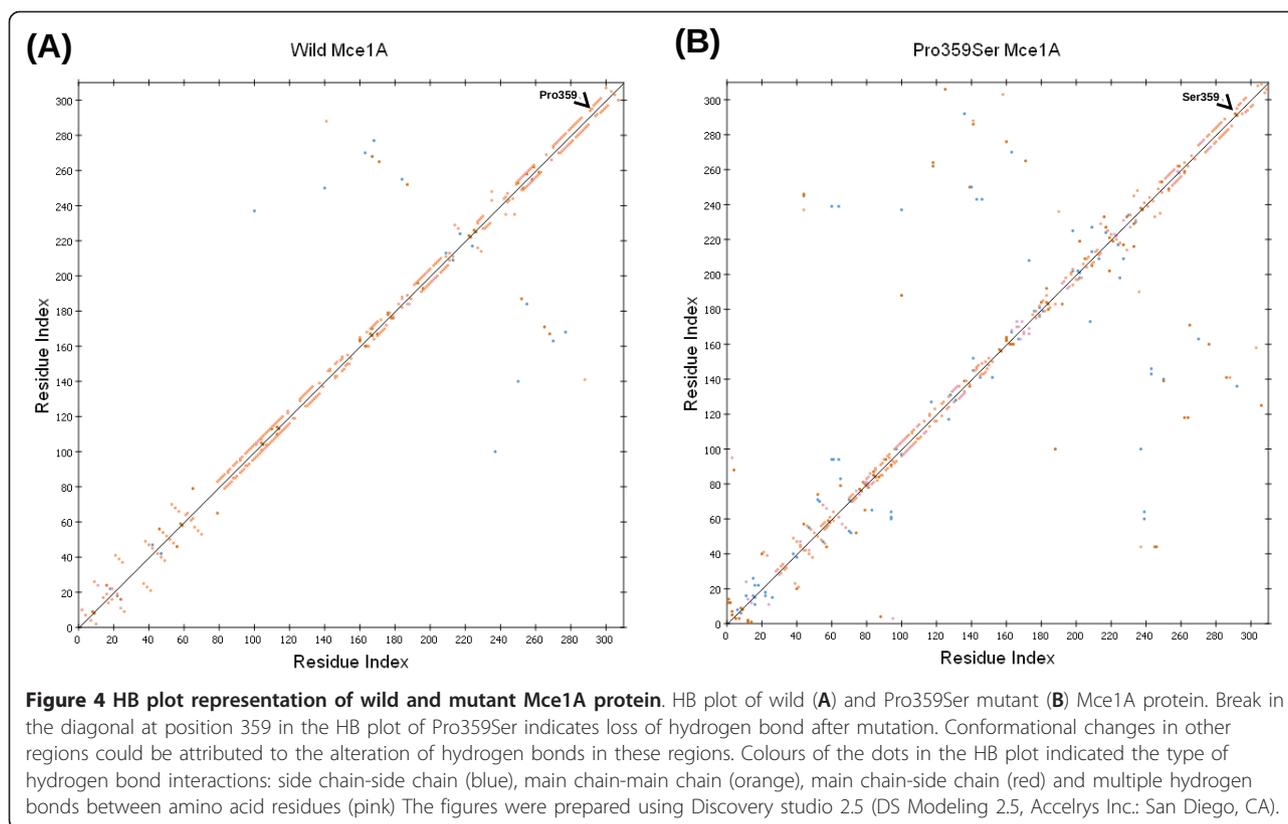


Figure 3 Comparison of Wild and mutant protein structure of Mce1A. Superimposed structure of wild (orange) and Pro359Ser mutant (blue) of Mce1A protein showing a change in helix to loop conformation after energy minimization of protein structures, as described in methods section. The RMSD between native and mutant protein was 3.07Å. Pro359 (green) in wild protein and Ser359 (pink) in the mutant protein are represented in ball and stick.



Discussion

It has been observed that severity of tuberculosis varies in different patients. It is possible that clinical isolates of *M. tuberculosis* encountering the human hosts with individual immune systems need to accordingly modulate their virulence associated biological factors to survive within the host. Therefore, it is important to understand the biology of the pathogen at the genetic level. Genetic polymorphisms in the bacterial hosts have been shown to significantly influence the biology of the organisms [17]. In *M. tuberculosis*, most of the polymorphisms have been studied in the transposable elements and drug resistant genes [1,18]. A study of the genetic mutations in the genes coding for virulence factors interacting with host's immune system would help us in understanding the ways in which various strains of *M. tuberculosis* adapt to different hosts.

The sequencing and Sequenom MassARRAY analysis presented here have revealed that *mce4* operon is significantly more polymorphic than *mce1* operon. Seven out of eight genes of *mce4* operon were found to be polymorphic. The *mce4* operon is expressed in the stationary phase of *in vitro* broth culture and in the late phase of infection [9,12] and has role in the survival of the pathogen inside the host body [19]. This is the time when the bacterium has established itself efficiently in the host

and it is possible that the bacterium then permits itself to undergo genetic substitutions to evade the host immune response. The detailed analysis of codon usage for synonymous changes observed in both *mce1* and *mce4* operons revealed that codons of amino acids were changed to the next preferred codon which would alter the expression of proteins. Our observation of more codon bias in *mce4* operon that may lead to less expression of proteins further supports the possibility that such diversity facilitates better survival of *M. tuberculosis* inside the host's body. Our results further reveal that more than 25% of clinical isolates have SNPs in *yrbE4A* and *lprN* genes of *mce4* operon. The *lprN* gene of *mce4* operon codes for lipoprotein precursor [20]. The lipoproteins of *M. tuberculosis* are known to be effectively antigenic in nature [21]. Thus, high polymorphism in *lprN* gene (both synonymous and nonsynonymous) further supports our hypothesis that such polymorphisms favour intracellular survival of the pathogen. Drug resistance itself makes the organism in a better position to survive within the hostile intracellular environment. But DS isolates being drug susceptible do not have this advantage. Therefore antigenic variation is a tool utilized by DS clinical isolates. For example, the function of PPE proteins is unknown. However several observations and results support that many are cell surface associated and recognized

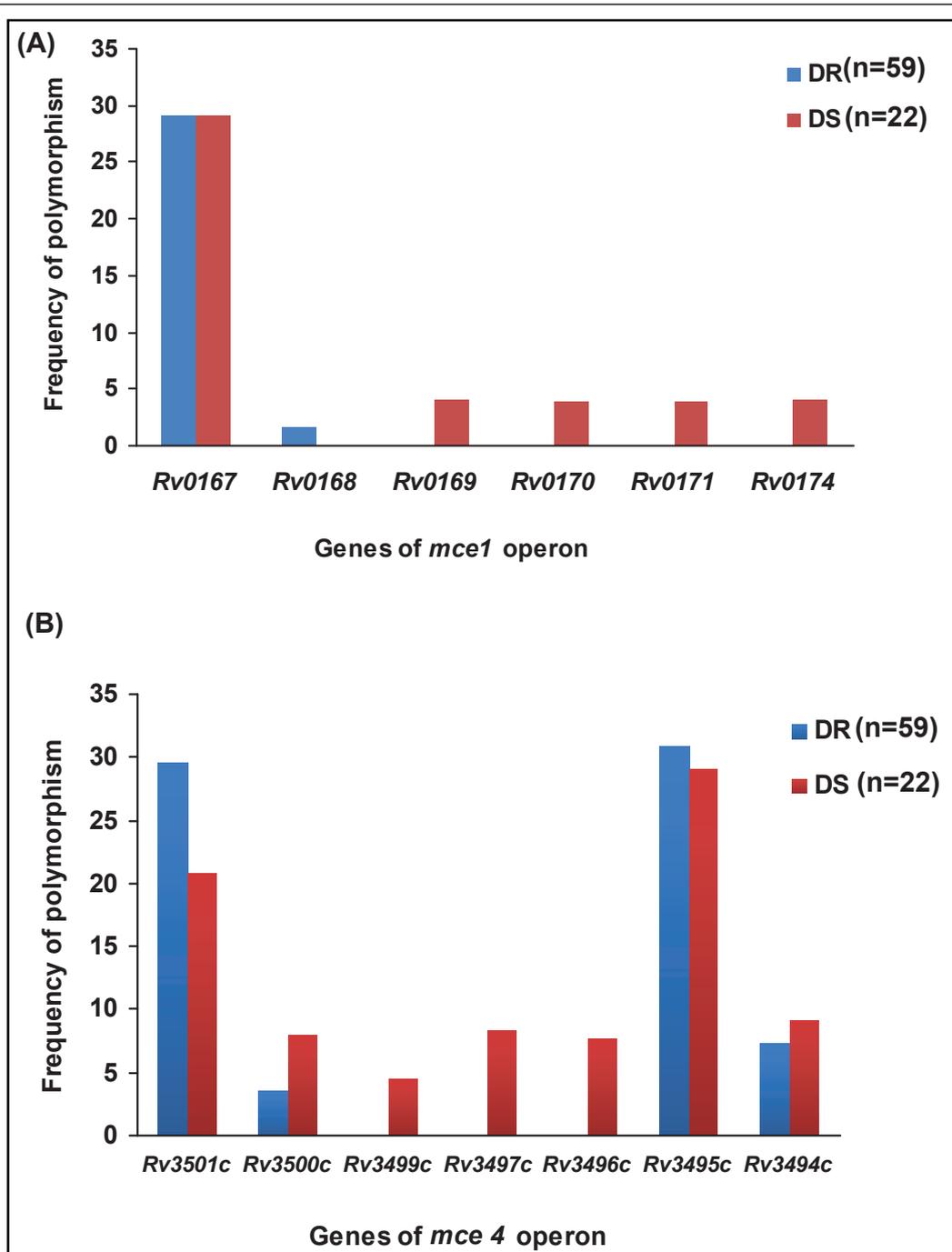


Figure 5 Comparative analysis of the frequency of SNPs in the *mce* operons genes of drug resistant (DR) and drug sensitive (DS) clinical isolates. SNPs were explored using Sequenom MassARRAY platform. DR (n = 59) and DS (n = 22) clinical isolates of *M. tuberculosis* were taken up for this study. The comparison between the two categories revealed that (A) *mce1* and (B) *mce4* operon genes were significantly more polymorphic in DS clinical isolates than DR isolates (*, $p < 0.05$) and (**, $p < 0.01$) respectively.

by the host immune system. The possibility of high antigenic variation associated with these highly antigenic PE and the PPE family proteins have also been reported [22]. The PGRS member *Rv1759* is a fibronectin-binding protein of relative molecular mass 55,000 Da [23] that elicits

a variable antibody response, indicating either that individuals mount different immune responses or that this PGRS protein may vary between strains of *M. tuberculosis*. Bioinformatics analysis have indicated that LprN is also a cell surface associated protein. Therefore it is

possible that SNP observed in this gene could be translated into antigenic variation in the LprN protein to facilitate the intracellular survival of mycobacteria.

In contrast, the *mce1* operon is required for the entry of the pathogen inside the host cell [24] and hence, remains less polymorphic. However, the *yrbE1A* gene is revealed to be highly polymorphic in *mce1* operon. Since, YrbE1A has been predicted to be a transmembrane protein [20], so the observed polymorphism in its gene may influence activity of the protein.

From the computational analysis, we could infer that the results obtained on the basis of structural details (PolyPhen) and sequence details (PMut) were in tune with each other. Both the programs have predicted that the SNP observed in *mce1A* gene is having the highest pathological relevance. Although, programs used here were primarily designed for predicting functional consequences of nonsynonymous SNPs in human proteins. However, concurrent observations on the nonsynonymous SNPs of *mce* operon proteins reported by both PolyPhen and PMut substantiate our hypothesis further. Energy minimization studies on the structure of Mce1A protein show that Pro359Ser mutation resulted in the loss of α -helical structure in the mutated protein. Analysis of wild and mutated Mce1A protein structures by HB plot indicates that change in hydrogen bonding interaction pattern in the mutant protein lead to conformational changes. Mutation of proline to serine residue in proteins are known to cause structural alterations by the reduction of α -helix content of protein and decreases protein stability and increase its susceptibility to proteolysis by trypsin [25]. Yazyu et al. [26] observed that Pro122Ser mutation could bring about the alteration in the pH of the system by changing the cation specificity of melibiose carrier (a membrane bound protein which mediates co transport of α -galactosides with monovalent cations) in *E. coli*. Pro122Ser mutant lost the ability to utilize H⁺ and made the carrier favorable for Li⁺- melibiose co-transport. Serine being a hard Lewis base interacts with hard Lewis acids such as Li⁺ instead of H⁺ [26]. Mce1A protein is a cell surface protein [27] so it may be speculated that the aforementioned changes due to Pro359Ser mutation may have a diminishing effect on the stability of protein and thus on the biological function of it.

In a further analysis, we compared the SNPs in the genes of *mce1* and *mce4* operons in 59 drug resistant (DR) and 22 drug sensitive (DS) clinical isolates. The comparison of SNPs in the *mce* genes in DR and DS clinical isolates revealed that both *mce1* and *mce4* operon genes of DS clinical isolates were more polymorphic than DR clinical isolates. It is possible that while drug resistance provides extra edge to DR isolates,

the DS isolates try to enhance their virulence mechanisms and adaptability to hostile intracellular environment by undergoing mutations in them. This is also supported by a report by Shimono et al. [28] where they have demonstrated that, unlike wild type *M. tuberculosis*, a strain of *M. tuberculosis* with disrupted *mce1* operon become hypervirulent.

Further study of larger number of single and multi drug resistant isolates may give a conclusive answer to the significance of such an observation.

Taken together the SNP analysis and *in silico* modeling reported here predict that the SNPs in the *mce1* and *mce4* operons in the clinical isolates are reasonably frequent. Also, the *in silico* modeling of nonsynonymous SNP in the *mce1A* gene of *mce1* operon indicates that such change may translate into altered function of the gene that may reflect on the virulence and biology of the pathogen.

Conclusions

In the present study, we have investigated the extent of polymorphism in the genes of *mce1* and *mce4* operons of *M. tuberculosis* in a panel of four standard reference strains (H37Rv, H37Ra, LVS (Low virulent Strain) and BCG) and 112 clinical isolates. Our results show that SNPs in the coding sequences of *mce1* and *mce4* operons in clinical isolates can be significantly high. Twenty SNPs were discovered in the two operons out of which 12 were nonsynonymous changes. Further analysis of pathological relevance of these changes revealed that five of the SNPs were deleterious. Overall, *mce4* operon is significantly more polymorphic than *mce1* operon ($p < 0.001$). However, nonsynonymous SNPs detected in *mce1A* gene of *mce1* operon predict effect of such SNPs on the biology of the pathogen.

Methods

Bacterial Strains

A collection of ~112 *M. tuberculosis* clinical isolates and four standard reference strains (H37Rv, H37Ra, LVS (Low virulent Strain) and BCG) were taken for the study. These isolates were from the patients visiting the out patient department (OPD) of Vallabhbhai Patel Chest Institute, Delhi, India. The strains were collected from sputum samples submitted to the Department of Microbiology for laboratory diagnosis of tuberculosis. The study was approved by the institutional ethics committee. Informed consent was also signed by the patients included in the study.

Processing of the sample

The sputum samples were decontaminated by the standard Petroff's method [29] and inoculated on Lowenstein Jensen (LJ) media. DNA was extracted from the cultures by the CTAB method [30]

Drug Susceptibility assays

Drug susceptibility testing was performed by the proportion method. The drug concentrations tested as per WHO recommendations were 0.2 mg/litre for isoniazid, 40 mg/litre for rifampicin, 2 mg/litre for ethambutol and 4 mg/litre for streptomycin. The drug incorporated LJ slants were incubated at 37°C and observed at 28 and 42 days of incubation [31]. The drug susceptibility was carried out on 59 DR and in 22 DS isolates out of the 100 clinical isolates and in 12 random selected isolates.

PCR amplification

Sixteen genes of *mce1* and *mce4* operons were amplified using overlapping primers listed in Additional file 1 for 4 standard reference strains (H37Rv, H37Ra, LVS (Low Virulent Strain) and BCG) and 12 clinical isolates. Thermal cycling was carried out for 40 cycles, with initial denaturation at 95°C for 10 minutes, followed by denaturation at 94°C for 1 minute, annealing at 56°C-64°C for 1 minute depending on primer sequence, elongation at 72°C for 1 minute and a final extension of 72°C for 10 minutes. The amplicons were purified by Qiagen PCR purification kit to remove unincorporated nucleotides and dNTPs.

Sequencing and Data Analysis

The PCR products obtained by using the overlapping primer sets as described above from four standard reference strains (H37Rv, H37Ra, LVS (Low Virulent Strain) and BCG) and 12 clinical isolates were sequenced using a DNA sequencer 3730 (Applied Biosystems). Both strands were sequenced to confirm the sequence data. The sequences were aligned and compared with the reference sequence of H37Rv available in public domain at Tuberculist [http://genolist.pasteur.fr/TubercuList/\[11\]](http://genolist.pasteur.fr/TubercuList/[11]) using Align two sequences (bl2seq) of BLAST <http://blast.ncbi.nlm.nih.gov/Blast.cgi> [32]. The SNPs obtained by the sequence analysis were used to screen other 100 clinical isolates through Sequenom MassARRAY system. All the SNPs were analysed further for the change in amino acids in the corresponding protein sequences through Gene Runner software version 3.05 (Hastings Software, Inc.) available at <http://www.generunner.net>.

Computational methods

Structure homology-based method (PolyPhen) to predict functional and structural changes in proteins

In order to analyze the impact of nonsynonymous SNPs on the structure and function of proteins of *mce* operons, Polyphen server [http://genetics.bwh.harvard.edu/pph/\[33\]](http://genetics.bwh.harvard.edu/pph/[33]) was used. Protein sequences in FASTA format with the position of amino acid variants indicated were submitted as the query. Polyphen server calculates position-specific independent counts (PSIC) scores for each

of the two variants based on the parameters such as sequence-based characterization of the substitution site, profile analysis of homologous sequences, and mapping of the substitution site to a known protein's three dimensional structure and then the difference between the PSIC scores of the two variants are computed. The higher the PSIC score (> 1.5) difference, the higher the functional impact a particular amino acid substitution is likely to have.

Neural network-based sequence information method (PMut) to predict pathological character of nonsynonymous SNPs

PMut server [http://mmb2.pcb.ub.es:8080/PMut/\[34\]](http://mmb2.pcb.ub.es:8080/PMut/[34]) was used to predict pathological relevance of nonsynonymous SNPs in the *mce* operon proteins. The software uses different kinds of sequence information to label mutations from the databases of disease-associated mutations (DAMU), and neural networks (NNs) to process the databases of DAMUs and neutral mutations (NEMUs). The resulting vector of properties is then utilized to decide whether the mutation is pathological or not. Although, PMut is designed to analyze pathological character associated with mutations in the human proteins. A number of workers [35,36] have qualitatively interpreted the functionality of mutated non-human proteins especially that of microbes. We submitted the protein sequences as the query, the location of the mutation and the amino acid residues were also furnished. Small NN (20 nodes, 1 hidden layer) with using 2/3 input parameters (pam40 matrix index, pssm index, variability index) was used to train the database as it is recommended for predictions of non-human proteins [34]. NN output greater than 0.5 is predicted as pathological otherwise neutral.

Molecular Modeling of nonsynonymous SNP located on the protein structure

The homology modeled three dimensional structure of Mce1A protein submitted to protein databank [PDB: 1NA9] [16] was used to study the effect of mutated residue. The authors have modelled Mce1A structure from residues 68 to 376, the N-terminal 67 residues and the C-terminal 78 residues were not modelled due to lack of homology [16]. Biopolymer module implemented in InsightII (Accelrys Inc.: San Diego, CA) was used to modify the mutated residues, from the InsightII fragment library. Using the same module, hydrogen atoms were added to both wild type and mutated protein structures at pH 7.0. The default cvff (Consistent Valence Force Field) force field [37] was applied to both the structures. Further, a series of energy minimization steps were performed on both the protein structures by InsightII/Discover (Accelrys Inc., San Diego, CA) using

the following protocol: (a) In the first step of minimization, all the heavy (all non-hydrogen) atoms were constrained, the hydrogen atoms were allowed to minimize by steepest descent algorithm until the maximum derivative ($|dE/dr|$) of the system was <1 kcal/(mole.Å). (b) This step was followed by another steepest descent minimization with the same parameter as in step (a), but constraining the protein backbone atoms and relaxing all other atoms of the molecule. (c) In the final step, the protein molecule was minimized by conjugate gradient method with the backbone atom fixed and allowing all other atoms relax until the maximum derivative was <0.01 kcal/(mole.Å). The deviation between the two structures is evaluated by their RMSD values which could affect stability and functional activity. Structure analysis of protein after energy minimization of protein structure was analyzed using Discovery Studio 2.5 (DS Modeling 2.5, Accelrys Inc.: San Diego, CA).

Statistical methods

Statistical analysis was done by Fischer's exact t test using Graph Pad Prism software <http://www.graphpad.com/quickcalcs/contingency1.cfm>. A two-tailed p-value < 0.05 was considered statistically significant.

Additional material

Additional file 1: Overlapping primers to sequence entire *mce1* and *mce4* operons.

Acknowledgements

The authors thank Indian Council for Medical Research (ICMR), Govt of India for financial support. RP thank Council for Scientific and Industrial Research (CSIR), Govt of India for Senior Research Fellowship (SRF). The support from Department of Biotechnology, Govt. of India for Bioinformatics Facility (BIF) at Dr. B.R. Ambedkar Center for Biomedical Research is highly acknowledged.

Author details

¹Department of Microbiology, Vallabhbhai Patel Chest Institute, University of Delhi, Delhi-110007, India. ²Department of Biochemistry, Vallabhbhai Patel Chest Institute, University of Delhi, Delhi-110007, India. ³The Centre for Genomic Application, Okhla Industrial Area, Delhi-110020, India. ⁴Dr B.R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi-110007, India.

Authors' contributions

MB and VB conceived the study. MVB provided the clinical isolates of *Mycobacterium tuberculosis*. RP carried out the major experimental work. MC and PP conducted the computational work. AC and NKS helped in experimental design. MB, VB, MVB, RP and PP participated in data interpretation and manuscript preparation. All authors read and approved the manuscript.

Received: 23 August 2010 Accepted: 23 February 2011
Published: 23 February 2011

References

1. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM: Strain identification of

- Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993, **31**:406-409.
2. Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidman J, Khouri H, Gill J, Mikula A, Bishai W, Jacobs WR Jr, Venter JC, Fraser CM: Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* 2002, **184**:5479-5490.
3. Hughes AL, Friedman R, Murray M: Genomewide pattern of synonymous nucleotide substitution in two complete genomes of *Mycobacterium tuberculosis*. *Emerg Infect Dis* 2002, **8**:1342-1346.
4. Gao Q, Kripke KE, Saldanha AJ, Yan W, Holmes S, Small PM: Gene expression diversity among *Mycobacterium tuberculosis* clinical isolates. *Microbiology* 2005, **151**:5-14.
5. Domenech P, Boshoff HI, Reed MB, Manca C, Kaplan G, Barry CE III: In vivo phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. *J Infect Dis* 2005, **192**:600-606.
6. Manca C, Tsenova L, Bergtold A, Freeman S, Tovey M, Musser JM, Barry CE III, Freedman VH, Kaplan G: Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha/beta. *Proc Natl Acad Sci USA* 2001, **98**:5752-5757.
7. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV: Variations in the *NRAMP1* gene and susceptibility to tuberculosis in West Africans. *N Engl J Med* 1998, **338**:640-644.
8. Weiss RA, McMichael AJ: Social and environmental risk factors in the emergence of infectious diseases. *Nat Med* 2004, **10**:S70-S76.
9. Kumar A, Bose M, Brahmachari V: Analysis of expression profile of mammalian cell entry (*mce*) operons of *Mycobacterium tuberculosis*. *Infect Immun* 2003, **71**:6083-6087.
10. Aguilar LD, Infante E, Bianco MV, Cataldi A, Bigi F, Pando RH: Immunogenicity and protection induced by *Mycobacterium tuberculosis mce-2* and *mce-3* mutants in a Balb/c mouse model of progressive pulmonary tuberculosis. *Vaccine* 2006, **24**:2333-2342.
11. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998, **393**:537-544.
12. Sassetti CM, Rubin EJ: Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci USA* 2003, **100**:12989-12994.
13. Ahmad S, Akbar PK, Wiker HG, Harboe M, Mustafa AS: Cloning, expression and immunological reactivity of two mammalian cell entry proteins encoded by the *mce1* operon of *Mycobacterium tuberculosis*. *Scand J Immunol* 1999, **50**:510-518.
14. Ahmad S, Shazly SE, Mustafa AS, Attiyah RA: Mammalian cell-entry proteins encoded by the *mce3* operon of *Mycobacterium tuberculosis* are expressed during natural infection in humans. *Scand J Immunol* 2004, **60**:382-391.
15. Andersson SGE, Sharp PM: Codon usage in the *Mycobacterium tuberculosis* complex. *Microbiology* 1996, **142**:915-925.
16. Das AK, Mitra D, Harboe M, Nandi B, Harkness RE, Das D, Wiker HG: Predicted molecular structure of the mammalian cell entry protein Mce1A of *Mycobacterium tuberculosis*. *Biochem Biophys Res Commun* 2003, **302**:442-447.
17. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM: Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA* 1997, **94**:9869-9874.
18. Ramaswamy S, Musser JM: Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis* 1998, **79**(1):3-29.
19. Saini NK, Sharma M, Chandolia A, Pasricha R, Brahmachari V, Bose M: Characterization of Mce4A protein of *Mycobacterium tuberculosis*: role in invasion and survival. *BMC Microbiol* 2008, **8**:200-208.
20. Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST: Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* 1999, **79**:329-342.

21. Young DB, Garbe TR: **Lipoprotein antigens of *Mycobacterium tuberculosis*.** *Res Microbiol* 1991, **142**:55-65.
22. Karboul Anis, Mazza Albarto, Gey Van Pittious CNicilaas, Ho LJohn, Brausseau Ronald, Mardassi Helmi: **Frequent homologous recombination events in *M tuberculosis* PE/PPE multigene families: Potential role in antigenic variability.** *J Bacteriol* 2008, **190**:7838-7846.
23. Abou-Zeid C, Garbe T, Lathigra R, Wiker HG, Harboe M, Rook GA, Young DB: **Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin binding proteins.** *Infect Immun* 1991, **59**:2712-2718.
24. Arruda S, Bonfim G, Knights R, Huima-Byron T, Riley LW: **Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells.** *Science* 1993, **261**:1454-1457.
25. Gilles AM, Girons IS, Monnot M, Femandjian S, Michelson S, Barzu O: **Substitution of a serine residue for proline-87 reduces catalytic activity and increases susceptibility to proteolysis of *Escherichia coli* adenylate kinase.** *Proc Natl Acad Sci USA* 1986, **83**:5798-5802.
26. Yazyu H, Shiota S, Futai M, Tsuchiya T: **Alteration in cation specificity of the melibiose transport carrier of *Escherichia coli* due to replacement of proline 122 with serine.** *J Bacteriol* 1985, **162**:933-937.
27. Chitale S, Ehart S, Kawamura I, Fujimura T, Shimono N, Anand N, Lu S, Gould LC, Riley L: **Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry.** *Cell Microbiol* 2001, **3**:247-254.
28. Shimono N, Morici L, Casali N, Cantrell S, Sidders B, Ehart S, Riley LW: **Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the *mce1* operon.** *Proc Natl Acad Sci USA* 2003, **100**:15918-15923.
29. Petroff SA: **A new and rapid method for the isolation and cultivation of tubercle bacilli directly from sputum and faeces.** *J Exp Med* 1915, **21**:38-42.
30. Van Soolingen D, Hermans PW, de Haas PE, Soll DR, Van Embden JD: **Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence dependent DNA polymorphism as a tool in the epidemiology of tuberculosis.** *J Clin Microbiol* 1991, **29**:2578-2586.
31. Canetti G, Kreis B, Thibier R, Gay P, Le Lirzin M: **Current data on primary resistance in pulmonary tuberculosis in adults in France. 2nd survey of the Center d'Etudes sur la Resistance Primaire.** *Rev Tuberc Pneumol* 1967, **31**:433-74.
32. Miller W, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
33. Ramensky V, Bork P, Sunyaev S: **Human nonsynonymous SNPs.** *Nucleic Acids Res* 2002, **30**:3894-3900.
34. Costa F, Orozco M, de la Cruz X: **Use of Bioinformatics Tools for the Annotation of Disease-Associated Mutations in Animal Models.** *Proteins* 2005, **61**:878-887.
35. Chasman D, Adams RM: **Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation.** *J Mol Biol* 2001, **307**:683-706.
36. Krishnan VG, Westhead DR: **A comparative study of machine learning methods to predict the effects of single nucleotide polymorphisms on protein function.** *Bioinformatics* 2003, **19**:2199-2209.
37. Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT: **Structure and energetics of ligand binding to proteins: *E. coli* dihydrofolate reductase-trimethoprim, a drug-receptor system.** *Proteins* 1988, **4**:31-47.

doi:10.1186/1471-2180-11-41

Cite this article as: Pasricha et al.: Single nucleotide polymorphism in the genes of *mce1* and *mce4* operons of *Mycobacterium tuberculosis*: analysis of clinical isolates and standard reference strains. *BMC Microbiology* 2011 **11**:41.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

