Summary and Conclusions

Tuberculosis remains a scourge to mankind. The causative organism, *M. tuberculosis*, has a special ability to adapt to different conditions and can survive within the host in a dormant state. Drug regimens and vaccines that were affective earlier are now showing poor efficacy against multiple drug-resistant strains that are emerging at a much greater pace. Hence, the need of an hour is to look at other avenues in the search for the "holy grail" (the ultimate drug and/or vaccine).

Identification and characterization of mycobacterial components are critical for diagnosis, vaccination, and in understanding the pathogenesis of the disease. Studies have recently shown that components from other mycobacterial strains (such as the environmental mycobacteria) have the ability to provide protective immune responses against virulent TB. Such environmental mycobacteria have diversified to a large extent due to genome ploidy events and exhibit a repertoire of molecules very different from what is displayed by *M. tuberculosis*. Such strains have potential to be exploited as natural live attenuated vaccine candidates. In this study, we have characterised the molecular components of *Mycobacterium simiae* as well as LppI of *M. tuberculosis*. Also, a database PolysacDB, on microbial polysaccharides, was generated.

The Draft genome of *M. simiae* MTCC 1727 was sequenced using Next-generation DNA sequencing techniques (Ion torrent PGM and Illumina MiSeq). It was assembled from a total of 100 contigs obtained. The size of the genome was found to be 5.7Mb (approximately 1Mb greater than the genome of *M. tuberculosis*). The genome was highly rich in regions of altered GC content indicating lateral gene transfer. Genome annotation and phylogenomic analysis revealed that *M. simiae* somehow occupies the intermediary position between members of *M. avium* complex and *M. ulcerans* (*M. marinum*) complex, possibly indicating an evolutionary link between the two. *M. simiae* also retains functional copies of characteristic *M. avium* proteins such as the proteins of the mycobactin synthesis gene cluster, gramicidin synthase etc. Also, like *M. avium*, the number of PE-PPE genes were extremely less in *M. simiae* in contrast to *M. marinum* that contains one of the largest number of PE-PPE genes ever reported in mycobacterial species.

A comprehensive proteomics analysis of the outer surface of *M. simiae* was carried out. The 2D gels of cell wall and short term culture filtrate (ST-CF) fractions showed abundant expression of common mycobacterial proteins such as GroEL-2, AtpD (F0F1 ATP

synthase subunit beta), elongation factor TU etc. Also, some novel proteins were found to be in abundance. Common among these were proteins related to combating stress such as Rv2005c (Universal stress protein), proteases such as mycobacterial trypsin, different copies of superoxide dismutases, cutinases (especially in the culture filtrate), lipases etc. LC-MS/MS analysis of the digests of cell wall and ST-CF preparations allowed the resolution of higher number of proteins. Extensive searches using high-end algorithms such as Protein Pilot and GenoSuite revealed the expression of 1,503 distinct proteins, with the cell wall accounting for the major portion (1,342 proteins) of the proteins. Four hundred ten proteins were detected from ST-CF. Most of the proteins showed greater than 60% BLAST sequence similarity to *M. tuberculosis* H37Rv proteins. While some proteins showed similarity to other mycobacterial species such as those of the *M. avium* complex, 62 novel proteins were identified that had no match or very low match (below 60% sequence similarity) with any mycobacterial protein as seen by NCBI BLASTP results and was considered specific to *M. simiae*. Most of these novel proteins showed variable similarities to proteins of soil dwelling microbes such as *Nocardia*, *Scionella*, *Actinomadura* etc. These 62 proteins also had the highest altered GC content in their genes, indicating those to have arisen via recent lateral gene transfer. Some of these also were found to be similar to plasmids from rhizobial species.

Analysis of the expressed proteome revealed high abundance of proteins related to energy and lipid metabolism, information storage and processing and cell wall/membrane processes. Approximately 12% (cell wall) and 13.5% (ST-CF) of the total proteins identified had characteristic N-termial sec signal peptide. One novel protease identified as "trypsin" was characteristically identified in the ST-CF fraction with one of the sequenced peptide spanning the mature and processed N-terminal region, thus validating sec-dependent cleavage for this protein. Using mass spectrometry data, the existence of a full length version of mycobacterial cutinase Cut5 was validated for the first time in the secretory fraction. One hundred sixty seven transmembrane proteins were identified in this study including some unusually large proteins such as gramicidin synthetases etc. Twenty one lipoproteins of diverse functionalities were detected. Antigenic proteins were identified by using both in silico analysis and antibodies. Using the VaxiJen server, 15% of the identified proteins were highly antigenic. Antibodies, generated against *M. simiae*, revealed number of immunogenic proteins in 2D gels.

Both polyclonal and monoclonal antibodies, generated against *E. coli* expressed and purified Acy.LppI protein, reacted with the LppI of *M. tuberculosis* H37Rv. Sub-cellular

localisation, using these antibodies, revealed that LppI is predominantly present in the cell wall fraction of *M. tuberculosis* H37Rv. Interestingly, unlike mAb 3F5 which reacted only with the cell wall fraction, polyclonal antibodies reacted with the membrane, cytosolic and culture filtrate fractions. The recognized bands of LppI in different fractions were: 29kDa in cytosol and membrane fractions, 29 kDa and 27 kDa bands in cell wall fraction and a lower 23kDa band in culture filtrate fraction. Partitioning of LppI in detergent phase of the cytosolic fraction indicating that the lipidation of the protein may be occurring in the cytosolic side rather than in the membrane or periplasmic sub-compartment. LppI expression when tested at different growth phases in *M. tuberculosis* H37Rv, i.e. early exponential, late exponential and stationary phase, showed that LppI level was comparable in the pellet fraction of all the phases, while in the cell free extract, the expression increased in the stationary phase.

It was observed that under both aeration and acid stress, the expression of LppI *in vitro* decreases considerably, unlike HspX which was found to be up-regulated as reported earlier, while LpqH, another well characterized lipoprotein of *M. tuberculosis*, level was not affected under these conditions. On probing the *E. coli* and *M. tuberculosis* expressed forms of LppI, ConA specifically bound to *M. tuberculosis* purified LppI and did not bind to *E. coli* expressed LppI, indicating that the protein is also glycosylated in *M. tuberculosis*. Acylated LppI had the ability to stimulate the release of TNF-α from C57BL/6 murine peritoneal macrophages in a dose dependent manner that seemed to be dependent on its lipid structures, since the non-acylated mutant could not elicit TNF-α secretion. Further, like Pam3CysK4, Acy.LppI treatment resulted in less secretion of TNF-α and IL-6 and IL-12 were found to be secreted in a dose dependent manner on exposure to Acy.LppI, Pam3CysK4 and LPS. While, non-acylated mutant failed to show any response indicating that this activity was mediated mainly through the lipid structures on the mature protein.

A database "PolysacDB", of microbial polysaccharide antigens and antibodies, was generated with the help of Dr. G.P.S. Raghava's group. The data was manually curated from more than 400 peer-reviewed publications encompassing 367 different types of microbes. Standard database tools such as "Search", "Advance Search" and "Browse" were incorporated for easy retrieval of information. Provision was provided for online submission and changing existing data. The usefulness of the database was demonstrated by validation experiments. Using PolySacDB, we listed those microbes having glucan or mannan or related polysaccharides on their surface. The results output showed *M. tuberculosis*, *Aspergillus*

fumigatus, *Candida albicans*, *Neurospora crassa*, *Flammulina velutipes* and *Phoma herbarum* as those containing glucan on their surface. Mannan containing microbes were *M. tuberculosis*, *Cryptococcus neoformans*, *Aspergillus fumigatus* and *Aspergillus flavus*. MAb against α(1→4) glucan of *M. tuberculosis* cross-reacted with both *Aspergillus fumigates* and *Aspergillus flavus*, without any significant reactivity with *E. coli* and *Candida albicans* even at higher concentrations. Notably, structures of β-glucans in *Candida* and *Aspergillus* are supposed to be same, yet only the *Aspergillus* species showed reactivity with this mAb. On the other hand, anti-*M.tb* lipoarabinomannan mAb cross-reacted to a appreciable level with *Aspergillus flavus* only, without much reactivity with *A. fumigates*. This indicated a difference between mannan structures of the two species. Also, this antibody did not cross react with any other species tested.

Overall, the important findings of this study are:

1. The genome content and size indicates a predominantly environmental and saprophytic lifestyle for *M. simiae*. It seemed to represent a intermediary state between the *M. avium* complex and the *M. marinum* (*M. ulcerans* complex).
2. Genomics and proteomics analysis indicated large number of genes acquired horizontally from other soil dwelling microbes. Some of these genes may account for some of the unique attributes of *M. simiae* such as antibiotic resistance.
3. Proteomics analysis revealed the expression of a number of virulence factors, antigens and other important mediators. It validated protein processing events such as sec dependent cleavage and existence of full length forms of proteins.
4. Antibodies produced against LppI protein localized the protein in the cell wall. The protein was found to be lipidated and glycosylated. It had the ability to stimulate the release of pro-inflammatory cytokines from macrophages in a acylation dependent and TLR2 dependent manner.
5. The database PolysacDB, generated in this study, was found useful in comparing the antigenic epitopes of polysaccharides on different microbes, particularly in context with mycobacteria.

In nut shell, above findings may provide basic information needed for exploring *M. simiae* as a vaccine candidate. Also, PoslysacDB may be exploited for discovering the specificity and cross reactivity among microbial polysaccharides.