Identification of unique essential proteins from a *M. tuberculosis* F15/LAM4/KZN phage secretome library

Thamsanqa Emmanuel Chiliza[1], Manormoney Pillay[2], and Balakrishna Pillay[1*]

[1]Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville 3630, South Africa;

[2]Medical Microbiology and Infection Control, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

*Corresponding author:

Balakrishna Pillay

School of Life Sciences (Westville Campus)

Discipline of Microbiology

University of KwaZulu-Natal

P/Bag X54001

South Africa

Durban

4000

Telephone: +27 312 608316

Email: pillayb1@ukzn.ac.za

ABSTRACT

Mycobacterium tuberculosis (*Mtb*) is the causative agent of tuberculosis disease (TB), the leading cause of death from bacterial infection worldwide. Although treatable, the resurgence of multi- and extensively-drug resistance TB is a major setback for the fight against TB globally. Consequently, there is an urgent need for new *Mtb* derived biomarkers for use in the design of new drugs and rapid point-of-care diagnostic or prognostic tools for management of TB transmission. Therefore, the present study aimed to identify unique *Mtb* secreted proteins from the extensively drug resistant *Mtb* F15/LAM4/KZN phage secretome library. A whole genome library was constructed using genomic DNA fragments of *Mtb* F15/LAM4/KZN strain. A phage secretome sub-library of $8x10^3$ clones was prepared and phage DNA was sequenced from 120 randomly selected clones. DNA sequence BLAST analysis identified 86 open reading frames. Using bioinformatics tools and databases, ten proteins essential for *in vivo* growth and survival of *Mtb* (Nrp, PssA, MmpL5, SirA, GatB, EspA, TopA, EccCa1, Rv1634 and Rv3103c). Proteins essential for growth and survival of *Mtb* during infection have potential application in the development of diagnostic tools, new drugs and vaccines. Further studies will be conducted to evaluate their potential application in the fight against TB.

Keywords

M. tuberculosis, biomarkers, phage display, secreted proteins, diagnosis, vaccine

INTRODUCTION

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), is harboured asymptomatically by one-third of the world's population. In 2014, approximately 10.4 million new infections and 1.8 million deaths were recorded globally (WHO, 2016). The emergence and spread of multi- and extensively drug resistant (MDR/XDR) strains of *M. tuberculosis* as well as high human immunodeficiency virus (HIV) co-infection rates (WHO, 2016), have complicated the fight against TB. Improved management of TB transmission can be achieved by early and rapid detection, and timely administration of effective TB treatment. The slow progress in the development of rapid point-of-care diagnostic assays, new drugs, vaccines and immunotherapeutic agents has largely been due to the availability of only a few, novel *M. tuberculosis* derived biomarkers (Wallis *et al.*, 2013). *M. tuberculosis* proteins secreted into its surrounding environment play an important role in host-pathogen interaction and can be ideal candidate biomarkers for the development of new drugs; vaccines; and for use in diagnosis (Andersen *et al.*, 1991, Harth & Horwitz, 1999). Preferably, a biomarker should be easily accessible as targets for effective, less invasive sampling for diagnosis and/or therapeutic interventions.

M. tuberculosis secrete a variety of extracellular virulence factors that play a role in adhesion and invasion of host target cells using various secretion pathways (Harth & Horwitz, 1999, Bendtsen *et al.*, 2005, Abdallah *et al.*, 2006). The unique cell wall and the associated extracellular virulence factors have been credited for the success of this pathogen (Champion & Cox, 2007) and may represent ideal candidates for the development of new drugs, vaccines, and TB diagnostics. The main *M. tuberculosis* protein secretory pathways are the Sec-dependent and Twin-Arginine Translocation (TAT) pathways (Ligon *et al.*, 2012). Two specialized protein secretion systems, the SecA2-dependent and the type VII secretion system or ESX, are also present in *M. tuberculosis* (Bendtsen *et al.*, 2005, Thakur *et*

al., 2016). The ESX system is responsible for secretion of small virulence proteins, such as early secreted antigenic target of 6 kDa (ESAT-6) and culture filtrate protein of 10 kDa (CFP-10). *M. tuberculosis* contains the genetic information for five type VII secretion machineries [ESX-1, ESX-2, ESX-3, ESX-4 and ESX-5] (Gey van Pittius *et al.*, 2006), suggesting the importance of Sec-independent protein secretion for this pathogen.

Generally, mass spectrometry based methods are used to study *M. tuberculosis* secreted proteins in culture filtrate (Ge *et al.*, 2003, de Souza *et al.*, 2011, Zheng *et al.*, 2013). While these methods have been highly effective, their main limitation is inconsistent protein expression owing to varying environmental growth conditions and the inability to detect proteins expressed at low concentrations (Forrellad *et al.*, 2013). Therefore, the phage display method offers a good alternative for bacterial secretome repertoire analysis (Jacobsson *et al.*, 2003, Rosander *et al.*, 2003, Wall *et al.*, 2003, Karlström *et al.*, 2004, Jankovic *et al.*, 2007, Rosander *et al.*, 2011, Gagic *et al.*, 2013). In 2011, *M. tuberculosis* H37Rv phage secretome library was used to identify six immunogenic proteins [MPT64 (Rv1980c), Ag85B (Rv1886c), cpsA (Rv3484), LpqA (Rv3016), PapA2 (Rv3820c) and EsxO (Rv2346c)] with potential as vaccine or diagnostic candidates (Liu *et al.*, 2011, Liu *et al.*, 2013).

Since the availability of the *M. tuberculosis* H37Rv complete genome sequence (Cole *et al.*, 1998), many *in vitro* and *in silico* studies have been conducted to analyse expression, identify open reading frames (ORFs) and predict their potential functions. However, the laboratory strain H37Rv has been reported to accumulate adaptation changes during repeated passage in culture, resulting in phenotypic alterations and partial attenuation (Ioerger *et al.*, 2010, Mehaffy *et al.*, 2010). Thus, the varying degree of pathogenicity among clinical *M. tuberculosis* strains (Palanisamy *et al.*, 2009), the phenotypic changes observed in H37Rv strain (Devasundaram & Raja, 2016), and the threat posed by emergence of highly transmissible MDR/XDR strains (Pillay & Sturm, 2007, Peters *et al.*, 2016) indicate a need to

study clinically relevant strains that may provide more insight into wild-type virulent behavior (Palanisamy *et al.*, 2009, Devasundaram & Raja, 2016).

The current study aim to identify appropriate *M. tuberculosis* derived diagnostic and drug target candidate protein biomarkers. The phage display method was used to construct a whole genome phage library of XDR *M. tuberculosis* F15/LAM4/KZN genomic DNA fragments. A phage secretome sub-library of $8x10^3$ clones was generated and sequence analysis was performed on randomly selected clones. The encoded ORFs were identified from the *M. tuberculosis* comparative database. Different bioinformatic tools, gene enrichment analysis and other databases were used to deduce essential and virulence associated *M. tuberculosis* protein biomarkers. The identified secretory protein biomarkers may have potential as diagnostic biomarkers, vaccine candidates and target for immunotherapeutic agents and drug discovery.

MATERIALS AND METHODS

Bacterial Strains, Phage Display Vector and Helper Phage

The XDR *M. tuberculosis* F15/LAM4/KZN (KZN605) strain was obtained from the archived collection at Medical Microbiology and Infection Control, University of KwaZulu-Natal. The XDR-KZN605 strain (GenBank accession: NC_018078) is resistant to isoniazid, rifampicin, ethambutol, ofloxacin, kanamycin, capreomycin and niacinamide. IS6110-restriction fragment length polymorphism analysis identified KZN605 to belong to the F15/LAM4/KZN strain family, while spoligotyping categorised it as the shared type (ST) 60 based on the absence of spacers 21-24, 33-36, 40 (Naidoo & Pillay, 2014). The strain was cultured aerobically in Middlebrook 7H9 broth (Difco), supplemented with 10 % oleic acid, albumin, dextrose, catalase (OADC), 0.05 % Tween 80 and 0.5 % glycerol at 37^oC with shaking to an OD_{600nm} 1. The pDJ01 phagemid vector, helper phage VCSM13d3 and pJARA plasmid DNA

were a gift from Dr Jasna Rakonjac of Massey University. The *E. coli* TG1 cells were from Lucigen Corporation.

M. tuberculosis Whole Genome Library Construction

Genomic DNA fragments of XDR *M. tuberculosis* F15/LAM4/KZN strain were used to construct a phage secretory protein repertoire library in the pDJ01 phage display system (Jankovic *et al.*, 2007). The minimum size of the library required to represent the whole *M. tuberculosis* genome was calculated using the following formula (Jacobsson *et al.*, 2003):

$$N = \ln (1-P)/\ln (1-f)$$

Where: P is the desired probability

f is the fractional proportion of the genome in a single recombinant

N is the necessary number of recombinants

Since the *M. tuberculosis* genome size is 4.41 $x10^6$ base pairs and the average fragment size is estimated at 500 base pairs,

$$N = \ln (1 - 0.99)/\ln [1-(500bp/4.41x10^6bp)]$$

$$N= 4x10^4 \text{ clones}$$

Genomic DNA was extracted using a modified sodium chloride-cetyl trimethyl-ammonium bromide (CTAB-NaCl) method (vanSoolingen *et al.*, 1994). Genomic DNA (200ng/uL) was fragmented by sonication on ice for 8 min using the Sonic Ruptor 400 (OMNI International). Fragments of 150 to 1500 base pair sizes were purified using a PCR clean-up kit (Macherey-Nagel Kit), followed by repair of blunt ends by End Repair Enzyme Mix (Thermo Fisher Scientific Inc.). The phagemid pDJ01 vector DNA was digested with *SmaI* (Thermo Scientific) and dephosphorylated with FastAP (Thermo Fisher Scientific Inc.) in the same reaction at 37^oC to generate blunt ends. DNA fragments (~8µg) were ligated into the pDJ01 vector (~8µg) in 1:1 ratio using a Rapid Ligation Kit (Thermo Fisher Scientific

Inc.). After desalting, the ligation reaction mixtures were electroporated into electro-competent *E. coli* TG1 cells (Lucigen Corporation), and incubated in 2xTY broth at 37ºC with aeration for 1hr. Serial dilutions were prepared for library size determination and the remaining cells were plated on TYE agar supplemented with 20 µg/mL chloramphenicol (Cm) and incubated overnight at 30ºC. These cultures were used to prepare 1 ml aliquots of 15% glycerol stock culture of the whole genome library and stored at -70ºC.

***M. tuberculosis* Phage Secretome Sub-library Preparation**

The phage display secretome library was prepared as follows: 1 mL of whole genome library stock culture was inoculated into 25 mL of 2xTY-Cm. The exponentially growing culture (OD_{600} 0.2) was infected with helper phage VCSM13d3 [phage to bacterium MOI = 50:1] for 1hr at 37ºC. Cells were centrifuged and the pellet re-suspended in 250 ml of 2xYT-Cm with 50 µg/mL kanamycin and incubated for 4hrs at 37ºC. After centrifugation at 10,000xg for 20 min, phagemid particles in the supernatant were precipitated with 5% (w/v) PEG/0.5 M NaCl overnight at 4ºC. The phagemid particles were centrifuged and pellet re-suspended in TN buffer (10 mmol/L Tris-HCl pH 7.6; 50 mmol/L NaCl). Defective phagemid particles were eliminated by treatment of the phagemid suspension ($1x10^{12}$ CFU/mL) with sarcosyl at a final concentration of 0.1% (w/v), followed by DNase I (100 U) in the presence of 5 mM MgCl2, and then inactivated by EDTA (20 mM). The remaining sarcosyl-resistant phage particles were precipitated with PEG/NaCl solution as above. For preparation of secretome sub-library, the ssDNA was extracted from sarcosyl resistant phagemid particles by first incubating at 70^{o}C for 10 min in the presence of 1.2% (w/v) SDS. The phenol-chloroform DNA isolation method was used for further purification of ssDNA (Su *et al.*, 1998). The surface protein library was amplified by transforming the ssDNA into *E. coli* TG1 that was grown on 2xYT-Cm plates overnight at 37ºC.

DNA Sequencing

The recombinant phage DNA of randomly selected clones was sequenced at Inqaba Biotechnical Industries (Pty) Ltd (South Africa) using the primer set flanking the vector cloning site, pDJ01R02 (5'-CCGGAAACGTCACCAATGAA-3') and pDJF03 (5'-ATGTTGCTGTTGATTCTTCA-3'). The DNA sequences were analysed using the CLCBio Workbench (v. 2.0).

Tuberculist, Signal Peptide and Transmembrane Prediction

BLAST analysis of the nucleotide sequences against the *M. tuberculosis* comparative database was performed to retrieve encoded open reading frames (ORFs). Tuberculist database (http://tuberculist.epfl.ch/) was used to determine functional categories of identified proteins (Lew *et al.*, 2011). Unknown proteins not documented on Tuberculist database were placed in the unknown functional category.

For prediction of Sec-dependent and Tat-dependent amino terminal signal peptide for secretion, SignalP 4.1 and TatP 1.0 were used (Bendtsen *et al.*, 2005a, Petersen *et al.*, 2011). SecretomeP 2.0 was used for prediction of non-classical secreted proteins (Bendtsen *et al.*, 2005b), and transmembrane proteins were predicted using TMHMM 2.0 (Krogh *et al.*, 2001). LipoP 1.0 was used for prediction of lipoprotein signal peptide (Rahman *et al.*, 2008). All software are freely available from the Centre for Biological Sequence Analysis at the Technical University of Denmark (http://www.cbs.dtu.dk/services). Alternative software, PRED-SIGNAL and PRED-TAT were used for supplementary analysis to predict Sec-dependent and Tat-dependent amino terminal signal peptide, respectively (Bagos *et al.*, 2009, Bagos *et al.*, 2010). PRED-LIPO was used for prediction of lipoprotein signal peptide (Bagos *et al.*, 2008). The alternative software were accessed via the Computational Genetics

Research Group at the Department of Computer Science and Biomedical Informatics (http://www.compgen.org/ tools) at the University of Thessaly. The subcellular localization of uncharacterized proteins was predicted with TBPred designed for analysis of mycobacterial proteins (http://www.imtech.res.in/raghava/tbpred/) (Rashid *et al.*, 2007).

Gene Enrichment Analysis and Functional Annotation

The Universal Protein Resource (UniProt) database (http://www.uniprot.org/blast/) accession numbers were retrieved by BLAST analysis of protein sequences against UniProt database (Suzek *et al.*, 2015). Gene enrichment analysis was performed using the database for annotation, visualization and integrated discovery (DAVID) (https://david.ncifcrf.gov/) for Gene Ontology (GO) terms and functional annotation clusters (Huang *et al.*, 2009). The significant enriched ($p < 0.05$) GO biological process (BP), cellular component (CC) and molecular function (MF) were identified. Pathosystems Research Intergrated Centre (PATRIC) database (https://www.patricbrc.org/) was used to identify essential and virulence genes as well as genes required for *M. tuberculosis* growth and survival within host (Wattam *et al.*, 2014).

RESULTS AND DISCUSSION

M. tuberculosis Whole Genome Library

A whole genome XDR *M. tuberculosis* phage library was successfully constructed by cloning of DNA fragment inserts into pDJ01 phage display vector. The library size was ~1.76×10^6 clones, exceeding the desired calculated size. Library diversity was confirmed on randomly selected clones by colony PCR (Fig. 1) and sequence analysis of 15 clones (results not shown). More than 90% of the library contained the *M. tuberculosis* DNA fragments. The library size was approximately ten fold larger than that obtained by Liu et al. (2011). This is

probably due to the total genomic DNA and fragments sizes (150bp to 1500bp) used as compared to 300bp to 1500bp fragments in Liu et al. (2011).

M. tuberculosis Phage Secretome

M. tuberculosis secrete proteins via classical (Sec- and Tat-dependent) and non-classical (Sec-independent) pathways. The classical pathway secreted proteins must have the signal peptide sequence for successful secretion (Ligon *et al.*, 2012). In the present study, the signal peptide sequence and transmembrane helices are responsible for guiding secretome to secretion and anchoring into membrane for display on phage surface, respectively (Gagic *et al.*, 2016). Therefore, successful display of *M. tuberculosis* peptide protein will require the DNA insert to encode for signal or transmembrane helices peptide and be cloned in-frame to the vector resulting in a stable phage particle.

Some *M. tuberculosis* virulence factors are secreted through the non-classical type VII secretion pathway. Usually, the type VII secretion pathway substrates are secreted as complex of two or more proteins as they need each other for successful translocation through membrane channels (Chen *et al.*, 2013). Even though some of the type VII secreted proteins will be in-frame, these might not be displayed on phage surface as their secretion partners required to form complex will be missing. The surrogate *E. coli* used for phage display is not expected to mimic the *M. tuberculosis* type VII secretion pathway. This should limit the discovery of most type VII/ESX secretion substrates. However, the type VII secreted membrane proteins may find their way into the phage secretome via transmembrane helices directing membrane localization.

Generally, it is expected that approximately 1-2% of DNA inserts will be in-frame and result in sarcosyl-resistant phagemid particles (Jacobsson & Frykberg, 2001). Furthermore, some *M. tuberculosis* signal sequences may not function efficiently when

expressed in *E. coli* (Smith, 1985), perhaps due to different codon usage and GC content (Poquet *et al.*, 1998). Therefore, the majority of the *M. tuberculosis* DNA inserts will be packaged into sarcosyl-sensitive phagemid as they lacked in-frame signal sequence or encode some type VII secretion pathways proteins. In this study, *M. tuberculosis* phage secretome sub-library of ~$8x10^3$ clones was obtained from the DNA of sarcosyl resistant phages [Fig. 2(a)]. Colony PCR confirmed the presence of insert DNA in randomly selected clones [Fig. 2(b)].

Phage DNA from 120 randomly selected sub-library clones were sequenced using primers flanking the pDJ01 vector cloning site. Analysis of DNA sequences using CLC Workbench software demonstrated that 18% of sequences comprised of repeat clones and 98 distinct DNA sequences were identified. BLAST analysis of distinct DNA sequences against the *M. tuberculosis* comparative database identified 86 corresponding ORFs and their protein sequences were retrieved for further analysis (Supplementary Information, Table 1).

Functional Categories of Identified Proteins

All 86 proteins were assigned a functional category (Supplementary Information, Table 2) as per Tuberculist database. The distribution of proteins to their functional category is presented in Fig. 3. The majority of proteins (34/86) belong to cell wall and cell processes category. The cell wall and cell processes proteins include two immunogenic proteins Mpt63 (Rv1926c) and Mpt64 (Rv1980c), two lipoproteins, LpqX (Rv1228) and LprO (Rv0179c), metal cation transporter CtpH (Rv0425c), MmpL12 (Rv1522c) involved in fatty acid transport, EmbC (Rv3793), alanine-leucine rich (Rv2693c) and alanine-valine-leucine rich (Rv2729c) suspected to be involved in the active evasion of the host immune response. Two ESX-1 secretion system proteins EspA (Rv3616c) and the conserved component protein EccCa1 (Rv3870) were also associated with cell wall and cell processes.

There were seventeen intermediary metabolism and respiration category proteins that include serine protease PepA (Rv0125), acyltransferase (Rv1254), putative ligase (Rv3712), and 4-hydroxy-2-oxovalerate aldolase (Rv3469c). Sixteen were conserve hypothetical proteins, including ala-, pro-rich protein (Rv1157c) and secreted protein Rv1268c. Eight lipid metabolism category proteins identified include the secreted Ag85B (Rv1886c, Ag85C (Rv0129c), peptide synthetase Nrp (Rv0101) and triacylglycerol synthase Tgs2 (Rv3734c). Other functional categorized proteins were four information pathways (Rv1700, Rv1981c, Rv3009c, and Rv3646c); three PPE family proteins (PPE32, PPE43 and PPE54); two virulence, detoxification and adaptation proteins, Rv1478 and Rv1566c associated with host invasion; and one regulatory protein (Rv1267c). One un-annotated protein, MT3042 was classified as a protein of unknown functional category.

Secretion Signal Peptide and Cellular Localization

Using different bioinformatics tools, the analysis of ORF protein sequences revealed Sec- and Tat- secreted proteins, membrane proteins and the secreted leaderless proteins. Twenty-seven proteins including 12 membrane proteins were predicted to harbor Sec-dependent N-terminal signal peptides by SignalP 4.1 and PRED-SIGNAL. Twenty-one proteins including 3 membrane proteins Rv1230c, Rv3162c and Rv3395A were predicted to contain a Tat signal peptide by TatP 1.0 and PRED-TAT. Twenty-one proteins were predicted trans-membrane proteins with 2 to 14 transmembrane helical structures by TMHMM 2.0 (Table 1).

In this study, 32 identified leaderless proteins were analyzed using SecretomeP 2.0 and TBPred databases to predict their subcellular localization (Table 2). Nine secreted proteins including two PPE proteins were identified by SecretomeP 2.0, whilst 21 of the 23 remaining proteins were confirmed as either secreted, membrane or membrane attached proteins by TBPred. Interestingly, 2 proteins Rv3616c and Rv3703c were predicted to be

cytoplasmic proteins by TBPred database (Table 2). Rv3616c encodes the indirect ESX-1 pathway substrate, EspA, which is associated with *M. tuberculosis* virulence (Garces et al., 2010), whilst the Rv3703c gene is a member of gene cluster or operon (Rv3704c–Rv3700c) responsible for ergothioneine (EGT) biosynthesis in *M. tuberculosis* (Glökler et al., 2010). EGT is a low-molecular weight protein essential for survival of *M. tuberculosis* in macrophages and is associated with antimycobacterial drug resistance (Glökler et al., 2010; Xu et al., 2015). The EGT synthesis gene cluster consists of *EgtA, EgtB, EgtC, EgtD* and *EgtE* respectively. The involuntary prediction of the well-known ESX-1 substrate (Rv3616c) as a cytoplasmic protein suggests that Rv3703c may also be a secreted protein. Similarly, Jankovic et al., (2007) identified two peptides without membrane or signal peptide sequence but were shown to possess undetermined secretion directive. Therefore, further studies should be undertaken in order to establish subcellular localization of all EGT synthesis gene cluster members and determine their possible leakage or secretory pathway.

ESX Proteins

The ESX-1 secretion system, that includes the secreted substrates ESAT6 and CFP10 virulence effectors, plays a critical role in *M. tuberculosis* pathogenicity (Simeone et al., 2009). Two ESX-1 associated proteins, EccCa1 and EspA, were identified in our study, but were not detected by Liu et al. (2011). EccCa1 is a conserved membrane component of the ESX-1 secretion system that is required for substrate exportation (Abdallah et al., 2007). EspA lacks a signal peptide sequence and was determined to be a cytoplasmic protein (Rashid et al., 2007). The gene (*espA*) encoding this protein is located within the *esx-1* gene cluster (Abdallah et al., 2007). This protein is a member of ESX-1 substrates: EsxA (ESAT-6), EsxB (CFP10), and EspA and EspB that require each other for successful secretion (Abdallah et al., 2007; Garces et al., 2010). Upon secretion, EspA forms a disulfide bonded

homodimer that is important for functional integrity of the *M. tuberculosis* cell wall. EspA is the most important determinant of ESX-1 mediated virulence as its disruption resulted in significant attenuation of *M. tuberculosis* virulence *in vivo*. Not only does EspA "guide" known virulence factors (ESAT-6 and CFP10) through ESX-1 apparatus system proteins such as EccCa, but it also appears to regulate activity of proteins interacting with the *M. tuberculosis* cell wall (Chen et al., 2013; Garces et al., 2010). ESX-1 substrates including ESAT-6 are responsible for the translocation of mycobacteria from the phagolysosome to the cytosol (Peng and Sun, 2016). Therefore, ESX-1 system proteins like EccCa1 and EspA are potential biomarkers that can be targeted for development of therapeutic interventions (Bottai et al., 2014) and diagnostic tools.

PE/PPE Proteins

About 10% of the *M. tuberculosis* genome encodes two gene families, the *pe* and *ppe* genes (Cole, 1998), totaling 99 and 69 respectively, in this pathogen (Fishbein et al., 2015; Gey van Pittius et al., 2006). PE and PPE proteins are named after the Proline (Pro) and Glutamic acid (Glu), and Pro–Pro–Glu motifs near the N terminus, respectively (Cole, 1998). These proteins are involved in host-pathogen interactions and may be required for survival *in vivo* (Abdallah et al., 2006; Fishbein et al., 2015). Species within the *M. tuberculosis* complex (MTBC), as well as the non-tuberculosis mycobacteria such as *M. leprae*, *M. marinum*, *M. ulcerans* and *M. avium* harbor the most number of *pe/ppe* genes. Fewer *pe*/*ppe* genes are found in nonpathogenic mycobacteria (Fishbein et al., 2015). Nearly all PPE proteins are secreted through the specialized type VII (ESAT-6 like) secretion system (Abdallah et al., 2006). ESAT-6 like specialized secretion systems [ESX-1, ESX-2, ESX-3 and ESX-5] have been associated with secretion of PE and PPE proteins (Gey van Pittius et al., 2006). However,

some PE and PPE proteins possess N-terminal signal peptide sequences for secretion (Forrellad et al., 2013).

In this study, three PPE proteins PPE32, PPE43 and PPE54 were identified. PE/PPE proteins are categorized into five sublineages (I to V) according to their evolutional relationship (Gey van Pittius et al., 2006). PPE32 and PPE43 belong to sublineage IV (Gey van Pittius et al., 2006) known to be secreted through ESX-5 secretion system (Fishbein et al., 2015). Interestingly, PPE32 have an N-terminal signal peptide for the Sec-dependent pathway. PPE43 and PPE54 lack a signal peptide sequence and were predicted to be extracellular proteins by SecretomeP. PPE54 belongs to the sublineage V and to the PPE_MPTR (major polymorphic tandem repeats) subfamily of PPE proteins. Sublineage V is reportedly highly expressed during *in vivo* infection (Fishbein et al., 2015), whilst PPE54 is expressed in guinea pig lungs (Kruh et al., 2010) and reported to be essential for *in vitro* growth (Tuberculist). Most PPE protein functions remain unknown, however, establishing localization of these proteins could provide some important clues to their function. Therefore, based on our findings PPE32, PPE43 and PPE54 are extracellular proteins that are either surface membrane attached or released proteins, and thus, may be directly involved in host-pathogen interaction.

Gene Enrichment Analysis

Gene Ontology terms and functional annotation clusters with p-value < 0.05 were enriched using DAVID database. The BP term significantly enriched (p = 0.04) for five proteins (Erm(37), EmbC, Ag85B, Ag85C and Rv1634) was response to antibiotic. The CC significantly enriched term (p = 0.00) for 28 genes was extracellular region. The CC enriched genes encoded for hypothetical proteins, immunogenic proteins such as MPT63 and MPT64, antigenic proteins MPT53 and CFP2, and proteins with some enzyme activity. The MF

significantly enriched (p = 0.01) term of four genes (Ag85B, Ag85C, Rv0192 and LdtA) was transferase activity (List of significantly enriched gene, Supplementary Information, Table 3).

DAVID functional annotation clustering of gene list resulted in two significantly enriched annotation clusters. Annotation clusters provides an overview of functions associated related proteins within gene list. Cluster one contain 26 transmembrane/membrane enriched proteins (p -value of 0.02 to 0.04) and cluster two contain 8 proteins (NuoF, DesA1, GltB, EtgB, NarG, NrdF1, SirA and Rv0203) involved in iron binding with p-value of 0.01. Therefore, enrichment of GO terms and functional annotation clustering suggest that our *M. tuberculosis* phage secretome library is rich in transmembrane, surface membrane and extracellular proteins. These membrane and extracellular proteins include essential proteins required for *in vivo* growth and virulence behaviour by assisting pathogen to acquire nutrients and navigate therapeutic interventions.

It is worth noting that only 85 of 86 proteins were analyzed by DAVID database. MT3042 was excluded from analysis and labelled as unknown protein. This could be due to fact that DAVID analysis was based on background *M. tuberculosis* H37Rv.

Essential Mycobacterial Proteins

M. tuberculosis secreted proteins play an important role in host-pathogen interaction and facilitate nutrient acquisition, navigate the host immune response and interfere with therapeutic intervention. Therefore, *M. tuberculosis* secretome consist of proteins essential for successful invasion and *in vivo* growth during host infection. PATRIC is the bacterial bioinformatics resource centre with curated and consistently annotated literature-based data with 7941 *M. tuberculosis* genome sequences (Wattam et al., 2014). In search for unique *M. tuberculosis* secreted protein biomarkers, PATRIC database was used to identify virulence factor and essential proteins.

Of 86 proteins in our list, 19 proteins were identified virulence factors (Table 3). The virulence factors included two cell wall (Ag85B and Ag85C), three intracellular survival (Rv1478, UvrA and AceAb), three proteins involve in modulation of host immune response (Rv1813c, MPT64 and NrdF1), seven virulence associated proteins (Nrp, Cut2, SirA, Rv2693c, Rv3103c, SapM, EspA and EccCa1). The others virulence factor proteins were the PepD a chaperone, NarG for anaerobic respiration, and PPE54 which affect phagosome.

M. tuberculosis essential proteins are ideal targets for the development of diagnostic tools and new drugs because of their key role in *in vivo* bacterial survival and growth. Therefore, identifying essential *M. tuberculosis* proteins required for growth and survival in infected host could lead to discovery of potential useful biomarkers. Using PATRIC database, ten essential proteins were identified, five membrane proteins (PssA, MmpL5, GatB, EccCa1 and Rv1634) and five secreted proteins (Nrp, SirA, EspA, TopA and Rv3103c). Two proteins are responsible for antibiotic resistance (MmpL5 and Rv1634) and one (PssA) is a drug target (Table 4). Generally, bacteria release their membrane proteins into the external environment as a means of membrane surface maintenance (Antelmann et al., 2001). Therefore, membrane proteins PssA, MmpL5, GatB, EccCa1 and Rv1634 will be valuable potential diagnostic biomarkers, since they could be found present in the body fluid of TB infected patients. Hence, these biomarkers have the potential to determine disease stages as it is expected that membrane proteins could be released from the bacterial cell surface at an advanced stage of growth.

There is limited literature on MT3042, the 82 amino acid protein. According to PATRIC database, MT3042 is a hypothetical protein annotated in *M. tuberculosis* CDC1551 genome sequence. Further studies are required to investigation the prevalence and expression of MT3042 among the *M. tuberculosis* clinical strains.

CONCLUSIONS

In this study, we report an elegant and efficient *in vitro* approach, for the selective extraction of secretome genetic information from the extensively drug-resistant *M. tuberculosis*. This genetic information can be used to identify secreted protein ORFs and allow further secretome specific *in silico* characterization. Using this approach, more than 95% of identified ORFs were confirmed as *M. tuberculosis* secretory or surface membrane proteins using different bioinformatics tools. Furthermore, *M. tuberculosis* virulence factor and essential proteins that are a prerequisite for growth and survival during infection were identified. We identified ten essential proteins, Nrp (Rv0101), PssA (Rv0436c), MmpL5 (Rv0676c), SirA (Rv2391), GatB (Rv3009c), EspA (Rv3616c), TopA (Rv3646c), EccCa1 (Rv3870), Rv1634 and Rv3103c. The essential proteins have potential application in the development of diagnostic tools, new drugs and vaccines. Future studies should investigate the suitability of all ten essential proteins as potential *M. tuberculosis* infection specific biomarkers for the development of rapid point-of-care antigen detection test.

Our findings complement the study by Liu et al. (2011) and support the use of phage display in the study of *M. tuberculosis* clinical strains secretome for biomarker discovery. Future studies using *M. tuberculosis* phage secretome libraries should aim to identify useful protein epitopes unique to MDR and XDR strains for use in design of vaccines, drugs and diagnostic tools.

FUNDING

This work was supported by the National Research Foundation (NRF) of South Africa [Grant number: 87989 to TE Chiliza]; and the University of KwaZulu-Natal, College of Agriculture, Engineering and Science, School of Life Science.

ACKNOWLEDGEMENTS

Authors would like to thank Dr. Jasna Rakonjac from the Institute for Molecular Biosciences, Massey University, New Zealand for donation of phage display system.

The authors, Thamsanqa E. Chiliza, Manormoney Pillay and Balakrishna Pillay declare that they have no conflict of interest.

REFERENCES

Abdallah AM, Verboom T, Hannes F*, et al.* (2006) A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Mol Microbiol* **62**: 667-679.

Andersen P, Askgaard D, Ljungqvist L, Bennedsen J & Heron I (1991) Proteins released from *Mycobacterium tuberculosis* during growth. *Infection and Immunity* **59**: 1905-1910.

Abdallah AM, van Pittius NCG, Champion PADG, Cox J, Luirink J, Vandenbroucke-Grauls CMJE, Appelmelk BJ & Bitter W (2007) Type VII secretion—mycobacteria show the way. *Nature Reviews Microbiology* **5**: 883-891.

Antelmann H, Tjalsma H, Voigt B, Ohlmeier S, Bron S, van Dijl JM & Hecker M (2001) A proteomic view on genome-based signal peptide predictions. *Genome Research* **11**: 1484-1502.

Armitige LY, Jagannath C, Wanger AR & Norris SJ (2000) Disruption of the genes encoding antigen 85A and antigen 85B of *Mycobacterium tuberculosis* H37Rv: effect on growth in culture and in macrophages. *Infection and immunity* **68**: 767-778.

Bagos PG, Tsirigos KD, Liakopoulos TD & Hamodrakas SJ (2008) Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *Journal of Proteome Research* **7**: 5082-5093.

Bagos P, Tsirigos K, Plessas S, Liakopoulos T & Hamodrakas S (2009) Prediction of signal peptides in archaea. *Protein Engineering Design and Selection* **22**: 27-35.

Bagos PG, Nikolaou EP, Liakopoulos TD & Tsirigos KD (2010) Combined prediction of Tat and Sec signal peptides with hidden Markov models. *Bioinformatics* **26**: 2811-2817.

Bendtsen JD, Nielsen H, Widdick D, Palmer T & Brunak S (2005a) Prediction of twin-arginine signal peptides. *BMC Bioinformatics* **6**: 167.

Bendtsen J, Kiemer L, Fausboll A & Brunak S (2005b) Non-classical protein secretion in bacteria. *BMC Microbiology* **5**: 58.

Bottai D, Serafini A, Cascioferro A, Brosch R & Manganelli R (2014) Targeting type VII/ESX secretion systems for development of novel antimycobacterial drugs. *Current pharmaceutical design* **20**: 4346-4356.

Bretl DJ, He H, Demetriadou C, White MJ, Penoske RM, Salzman NH & Zahrt TC (2012) MprA and DosR coregulate a *Mycobacterium tuberculosis* virulence operon encoding Rv1813c and Rv1812c. *Infection and immunity* **80**: 3018-3033.

Brodin P, Poquet Y, Levillain F, Peguillet I, Larrouy-Maumus G, Gilleron M, Ewann F, Christophe T, Fenistein D & Jang J (2010) High content phenotypic cell-based visual screen identifies *Mycobacterium tuberculosis* acyltrehalose-containing glycolipids involved in phagosome remodeling. *PLoS Pathog* **6**: e1001100.

Champion PAD, Stanley SA, Champion MM, Brown EJ & Cox JS (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. *Science* **313**: 1632-1636.

Champion PAD & Cox JS (2007) Protein secretion systems in Mycobacteria. *Cellular Microbiology* **9**: 1376-1384.

Chauhan P, Reddy PV, Singh R, Jaisinghani N, Gandotra S & Tyagi AK (2013) Secretory phosphatases deficient mutant of *Mycobacterium tuberculosis* imparts protection at the primary site of infection in guinea pigs. *PloS one* **8**: e77930.

Chen JM, Zhang M, Rybniker J, Basterra L, Dhar N, Tischler AD, Pojer F & Cole ST (2013) Phenotypic Profiling of *Mycobacterium tuberculosis* EspA Point Mutants Reveals that Blockage of ESAT-6 and

CFP-10 Secretion *In Vitro* Does Not Always Correlate with Attenuation of Virulence. *Journal of Bacteriology* **195**: 5421-5430.

Cole ST, Brosch R, Parkhill J*, et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

de Souza GA, Leversen NA, Målen H & Wiker HG (2011) Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *Journal of Proteomics* **75**: 502-510.

Devasundaram S & Raja A (2016) Variable transcriptional adaptation between the laboratory (H37Rv) and clinical strains (S7 and S10) of *Mycobacterium tuberculosis* under hypoxia. *Infection, Genetics and Evolution* **40**: 21-28.

Fishbein S, van Wyk N, Warren RM & Sampson SL (2015) Phylogeny to function: PE/PPE protein evolution and impact on *Mycobacterium tuberculosis* pathogenicity. *Mol Microbiol* **96**: 901-916.

Forrellad MA, Klepp LI, Gioffré A, Sabio y García J, Morbidoni HR, Santangelo MdlP, Cataldi AA & Bigi F (2013) Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence* **4**: 3-66.

Fortune S, Jaeger A, Sarracino D, Chase M, Sassetti C, Sherman D, Bloom B & Rubin E (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 10676-10681.

Gagic D, Wen W, Collett MA & Rakonjac J (2013) Unique secreted–surface protein complex of *Lactobacillus rhamnosus*, identified by phage display. *MicrobiologyOpen* **2**: 1-17.

Gagic D, Ciric M, Wen WX, Ng F & Rakonjac J (2016) Exploring the Secretomes of Microbes and Microbial Communities Using Filamentous Phage Display. *Frontiers in Microbiology* **7**: 429

Gao L-Y, Pak M, Kish R, Kajihara K & Brown EJ (2006) A mycobacterial operon essential for virulence *in vivo* and invasion and intracellular persistence in macrophages. *Infection and immunity* **74**: 1757-1767.

Garces A, Atmakuri K, Chase MR*, et al.* (2010) EspA acts as a critical mediator of ESX1-dependent virulence in *Mycobacterium tuberculosis* by affecting bacterial cell wall integrity. *PLoS Pathog* **6**: e1000957.

Ge Y, ElNaggar M, Sze SK, Oh HB, Begley TP, McLafferty FW, Boshoff H & Barry Iii CE (2003) Top down characterization of secreted proteins from Mycobacterium tuberculosis by electron capture dissociation mass spectrometry. *Journal of the American Society for Mass Spectrometry* **14**: 253-261.

Gey van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD & Warren RM (2006) 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 Evolutionary Biology* **6**.

Glökler J, Schütze T & Konthur Z (2010) Automation in the high-throughput selection of random combinatorial libraries—Different approaches for select applications. *Molecules* **15**: 2478-2490.

Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S & Sherman DR (2004) Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **51**: 359-370.

Harth G & Horwitz MA (1999) An inhibitor of exported *Mycobacterium tuberculosis* glutamine synthetase selectively blocks the growth of pathogenic mycobacteria in axenic culture and in human monocytes: extracellular proteins as potential novel drug targets. *J Exp Med* **189**: 1425-1436.

Houghton J, Townsend C, Williams AR, Rodgers A, Rand L, Walker KB, Böttger EC, Springer B & Davis EO (2012) Important role for *Mycobacterium tuberculosis* UvrD1 in pathogenesis and persistence apart from its function in nucleotide excision repair. *Journal of Bacteriology* **194**: 2916-2923.

Huang DW, Sherman BT & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**: 44-57.

Ioerger TR, Feng Y, Ganesula K, Chen X, Dobos KM, Fortune S, Jacobs WR, Mizrahi V, Parish T & Rubin E (2010) Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. *Journal of Bacteriology* **192**: 3645-3653.

Jacobsson K & Frykberg L (2001) Shotgun phage display cloning. *Comb Chem High Throughput Screen* **4**: 135 - 143.

Jacobsson K, Rosander A, Bjerketorp J & Frykberg L (2003) Shotgun Phage Display - Selection for Bacterial Receptins or other Exported Proteins. *Biol Proced Online* **5**: 123 - 135.

Jankovic D, Collett MA, Lubbers MW & Rakonjac J (2007) Direct selection and phage display of a Gram-positive secretome. *Genome Biology* **8**: R266.

Karlström Å, Jacobsson K, Flock M, Flock J-I & Guss B (2004) Identification of a novel collagen-like protein, SclC, in *Streptococcus equi* using signal sequence phage display. *Veterinary Microbiology* **104**: 179-188.

Kozak RA, Alexander DC, Liao R, Sherman DR & Behr MA (2011) Region of difference 2 contributes to virulence of *Mycobacterium tuberculosis*. *Infection and Immunity* **79**: 59-66.

Krogh A, Larsson B, Von Heijne G & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**: 567-580.

Kruh NA, Troudt J, Izzo A, Prenni J & Dobos KM (2010) Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome *in vivo*. *PLoS One* **5**: e13938.

Lew JM, Kapopoulou A, Jones LM & Cole ST (2011) TubercuList - 10 years after. *Tuberculosis* **91**: 1-7.

Ligon LS, Hayden JD & Braunstein M (2012) The ins and outs of *Mycobacterium tuberculosis* protein export. *Tuberculosis* **92**: 121-132.

Liu S, Han W, Sun C, Lei L & Feng X (2013) Identification of Two New Virulence Factors of *Mycobacterium Tuberculosis* that Induce Multifunctional CD4 T Cell Responses. *J Mycobac Dis S* **6**: 2161-1068.

Liu S, Han W, Sun C, Lei L, Feng X, Yan S, Diao Y, Gao Y, Zhao HL & Liu Q (2011) Subtractive screening with the *Mycobacterium tuberculosis* surface protein phage display library. *Tuberculosis* **91**: 579-586.

MacGurn JA & Cox JS (2007) A genetic screen for *Mycobacterium tuberculosis* mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system. *Infection and immunity* **75**: 2668-2678.

Mehaffy C, Hess A, Prenni JE, Mathema B, Kreiswirth B & Dobos KM (2010) Descriptive proteomic analysis shows protein variability between closely related clinical isolates of *Mycobacterium tuberculosis*. *Proteomics* **10**: 1966-1984.

MohamedMohaideen NN, Palaninathan SK, Morin PM, Williams BJ, Braunstein M, Tichy SE, Locker J, Russell DH, Jacobs Jr WR & Sacchettini JC (2008) Structure and Function of the Virulence-Associated High-Temperature Requirement A of *Mycobacterium tuberculosis*. *Biochemistry* **47**: 6092-6102.

Muñoz-Elías EJ & McKinney JD (2005) *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for *in vivo* growth and virulence. *Nature Medicine* **11**: 638-644.

Naidoo C & Pillay M (2014) Increased *in vitro* fitness of multi-and extensively drug-resistant F15/LAM4/KZN strains of *Mycobacterium tuberculosis*. *Clinical Microbiology and Infection* **20**: O361-O369.

Ocampo M, Rodriguez D, Curtidor H, Vanegas M, Patarroyo M & Patarroyo M (2012) Peptides derived from *Mycobacterium tuberculosis* Rv2301 protein are involved in invasion to human epithelial cells and macrophages. *Amino Acids* **42**: 2067-2077.

Palanisamy GS, DuTeau N, Eisenach KD, Cave DM, Theus SA, Kreiswirth BN, Basaraba RJ & Orme IM (2009) Clinical strains of *Mycobacterium tuberculosis* display a wide range of virulence in guinea pigs. *Tuberculosis* **89**: 203-209.

Peng X & Sun J (2016) Mechanism of ESAT-6 membrane interaction and its roles in pathogenesis of *Mycobacterium tuberculosis*. *Toxicon* **116**: 29-34.

Peters J, Calder B, Gonnelli G, Degroeve S, Rajaonarifara E, Mulder N, Soares NC, Martens L & Blackburn JM (2016) Identification of quantitative proteomic differences between *Mycobacterium tuberculosis* lineages with altered virulence. *Frontiers in Microbiology* **7**.

Petersen TN, Brunak S, von Heijne G & Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785-786.

Pillay M & Sturm AW (2007) Evolution of the extensively drug-resistant F15/LAM4/KZN strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. *Clinical Infectious Diseases* **45**: 1409-1414.

Poquet I, Ehrlich SD & Gruss A (1998) An export-specific reporter designed for gram-positive bacteria: application to *Lactococcus lactis*. *Journal of bacteriology* **180**: 1904-1912.

Rahman O, Cummings SP, Harrington DJ & Sutcliffe IC (2008) Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World Journal of Microbiology and Biotechnology* **24**: 2377-2382.

Rashid M, Saha S & Raghava GP (2007) Support Vector Machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs. *BMC Bioinformatics* **8**: 1.

Rosander A, Frykberg L, Ausmees N & Müller P (2003) Identification of Extracytoplasmic Proteins in *Bradyrhizobium japonicum* Using Phage Display. *Molecular Plant-Microbe Interactions* **16**: 727-737.

Rosander A, Guss B, Frykberg L, Björkman C, Näslund K & Pringle M (2011) Identification of immunogenic proteins in *Treponema phagedenis*-like strain V1 from digital dermatitis lesions by phage display. *Veterinary Microbiology* **153**: 315-322.

Sassetti CM & Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences* **100**: 12989-12994.

Simeone R, Bottai D & Brosch R (2009) ESX/type VII secretion systems and their role in host–pathogen interaction. *Current Opinion in Microbiology* **12**: 4-10.

Smith G (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**: 1315 - 1317.

Su M, Venkatesh T & Bodmer R (1998) Large-and small-scale preparation of bacteriophage lambda lysate and DNA. *BioTechniques* **25**: 44-46.

Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH & Consortium tU (2015) UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**: 926-932.

Thakur P, Gantasala NP, Choudhary E, Singh N, Abdin MZ & Agarwal N (2016) The preprotein translocase YidC controls respiratory metabolism in *Mycobacterium tuberculosis*. *Sci Rep* **6**: 24998.

vanSoolingen D, deHaas PE, Hermans PW & vanEmbden JD (1994) DNA fingerprinting of *Mycobacterium tuberculosis*. *Methods Enzymol* **235**: 196-205.

Wall T, Roos S, Jacobsson K, Rosander A & Jonsson H (2003) Phage display reveals 52 novel extracellular and transmembrane proteins from *Lactobacillus reuteri* DSM 20016T. *Microbiology* **149**: 3493-3505.

Wallis RS, Kim P, Cole S, Hanna D, Andrade BB, Maeurer M, Schito M & Zumla A (2013) Tuberculosis biomarkers discovery: developments, needs and challenges. *Lancet Infect Dis* **13**: 362-372.

Wattam AR, Gabbard JL, Shukla M & Sobral BW (2014) Comparative genomic analysis at the PATRIC, a bioinformatic resource center. *Host-Bacteria Interactions: Methods and Protocols* 287-308.

WHO (2016) Global Tuberculosis Report 2016. *World Health Organization*.

Xu Y, Yang E, Huang Q, Ni W, Kong C, Liu G, Li G, Su H & Wang H (2015) PPE57 induces activation of macrophages and drives Th1-type immune responses through TLR2. *Journal of Molecular Medicine* **93**: 645-662.

Zheng J, Ren X, Wei C, Yang J, Hu Y, Liu L, Xu X, Wang J & Jin Q (2013) Analysis of the Secretome and Identification of Novel Constituents from Culture Filtrate of Bacillus Calmette-Guérin Using High-resolution Mass Spectrometry. *Molecular & Cellular Proteomics* **12**: 2081-2095.

Table 1: List of Sec- and Tat-dependent N-terminal signal harbouring and transmembrane proteins

Sec-dependent		**Tat-dependent**		**TMHMM Predicted**	
Rv no.	**Gene name**	**Rv no.**	**Gene name**	**Rv no.**	**Gene name**
Rv0116c	*ldtA*	-	*MT3042*	[a]Rv0236c	*aftD*
Rv0320	-	Rv0125	*pepA*	[a]Rv0425c	*ctpH*
Rv0559c	-	Rv0129c	*fbpC*	[a]Rv0436c	*pssA*
Rv0603	-	Rv0179c	*lprO*	Rv0676c	*mmpL5*
Rv0675	*echA5*	Rv0203	-	[a]Rv0842	-
Rv1157c	-	Rv0455c	-	[a]Rv1029	*kdpA*
Rv1228	*lpqX*	Rv1268c	-	Rv1200	-
Rv1478	-	Rv1291c	-	[b]Rv1230c	-
Rv1804	-	Rv1435c	-	[a]Rv1254	-
Rv1808	*ppe32*	Rv1566c	-	[a]Rv1522c	*mmpL12*
Rv1926c	*mpt63*	Rv1813c	-	[a]Rv1621c	*cydD*
Rv1980c	*mpt64*	Rv1886c	*fbpB*	Rv1634	-
Rv2376c	*cfp2*	Rv2301	*cut2*	Rv1733c	-
Rv2878c	*mpt53*	Rv2391	*sirA*	[a]Rv2693c	-
Rv3310	*sapM*	Rv3222c	-	[a]Rv2729c	-
		Rv3646c	*topA*	[b]Rv3162c	-
		Rv3712	-	[a]Rv3193c	-
		Rv3835	-	[a]Rv3365c	-
				[b]Rv3395A	-
				Rv3793	*embC*
				Rv3870	*eccCa1*

[a]Sec-dependent N-terminal signal
[b]Tat-dependent N-terminal signal

Table 2: Subcellular localization prediction of 32 proteins using SecretomeP 2.0a and TBPred.

Rv no.	Gene name	Subcellular location	Rv no.	Gene name	Subcellular location
Rv0101	*nrp*	secreted	Rv1981c	*nrdF1*	secreted
[c]Rv0192	-	-	Rv1988	*erm(37)*	transmembrane
Rv0255c	*cobQ1*	membrane attached	Rv2239c	-	secreted
[c]Rv0822c	-	-	[c]Rv2768c	*ppe43*	-
Rv0824c	*desA1*	secreted	[c]Rv2922A	*acyP*	-
[c]Rv0983	*pepD*	-	Rv3009c	*gatB*	membrane attached
[c]Rv1118c	-	-	[c]Rv3103c	-	-
Rv1156	-	membrane attached	Rv3150	*nuoF*	secreted
Rv1161	*narG*	secreted	Rv3280	*accD5*	secreted
Rv1267c	*embR*	membrane attached	[c]Rv3343c	*ppe54*	-
Rv1366	-	secreted	Rv3410c	*guaB3*	membrane attached
Rv1447c	*zwf2*	transmembrane	Rv3469c	*mhpE*	secreted
Rv1613	*trpA*	membrane attached	Rv3616c	*espA*	Cytoplasmic
[c]Rv1638	*uvrA*	-	Rv3703c	*egtB*	Cytoplasmic
Rv1700	-	membrane attached	Rv3734c	*tgs2*	transmembrane
Rv1916	*aceAb*	secreted	Rv3859c	*gltB*	membrane attached

[c]SecretomeP 2.0a predicted proteins

Table 3: List of virulence factor *M. tuberculosis* secretory proteins identified using PATRIC database.

Rv no.	Gene	Classification	References
Rv0101	*nrp*	Virulence	(Sassetti & Rubin, 2003)
Rv0129c	*fbpC*	cell wall	(Puech *et al.*, 2002)
Rv0983	*pepD*	chaperone,protease	(MohamedMohaideen *et al.*, 2008)
Rv1161	*narG*	Cellular metabolism,Anaerobic respiration	PATRIC
Rv1478		invasion,intracellular survival and replication	(Gao *et al.*, 2006)
Rv1638	*uvrA*	intracellular survival and replication	(Houghton *et al.*, 2012)
Rv1813c		modulate host immune response	(Bretl *et al.*, 2012)
Rv1886c	*fbpB*	No evidence of virulence,cell wall	(Armitige *et al.*, 2000, Puech *et al.*, 2002)
Rv1916	*aceAb*	intracellular survival and replication	(Muñoz-Elías & McKinney, 2005)
Rv1980c	*mpt64*	modulate host immune response	(Kozak *et al.*, 2011)
Rv1981c	*nrdF1*	modulate host immune response	(Kozak *et al.*, 2011)
Rv2301	*cut2*	invasion	(Ocampo *et al.*, 2012)
Rv2391	*nirA*	Virulence factor	(Sassetti & Rubin, 2003)
Rv2693c		Virulence factor	(MacGurn & Cox, 2007)
Rv3103c		Virulence factor	(Sassetti & Rubin, 2003)
Rv3310		virulence	(Chauhan *et al.*, 2013)
Rv3343c	*PPE54*	affect phagosome	(Brodin *et al.*, 2010)
Rv3616c	*espA*	Virulence associated secretion systems, Type VII secretion	(Garces *et al.*, 2010, Chen *et al.*, 2013)
Rv3870	*eccCa1*	virulence,Type VII secretion	(Guinn *et al.*, 2004, Champion *et al.*, 2006)

Table 4: List of essential *M. tuberculosis* secretory proteins identified using PATRIC database.

Rv no.	Gene	Product	References
*Rv0101	*nrp*	Peptide synthetase Nrp	(Sassetti & Rubin, 2003)
Rv0436c	*pssA*	Phosphatidylserine synthase	PATRIC
Rv0676c	*mmpL5*	Siderophore exporter MmpL5	PATRIC
Rv1634		Probable multidrug-efflux transporter	PATRIC
*Rv2391	*sirA*	inorganic ion transport and metobolism	(Sassetti & Rubin, 2003)
Rv3009c	*gatB*	Asn/Gln amidotransferase subunit B	PATRIC
*Rv3103c		Hypothetical proline-rich protein	(Sassetti & Rubin, 2003)
**Rv3616c	*espA*	ESX-1 secretion-associated protein EspA	(Sassetti & Rubin, 2003, Fortune *et al.*, 2005)
Rv3646c	*topA*	DNA topoisomerase 1	PATRIC
**Rv3870	*eccCa1*	ESX-1 secretion system protein EccCa1	(Sassetti & Rubin, 2003, Guinn *et al.*, 2004)

*= Sassetti & Rubin, 2003

** = identified by PATRIC and Sassetti & Rubin, 2003

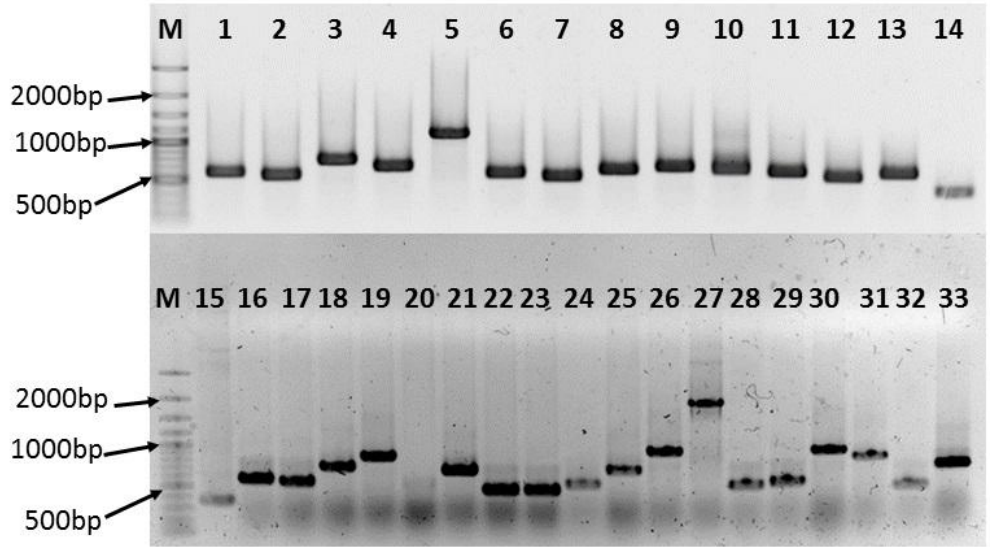

Figure 1: Colony PCR of randomly selected clones of the whole genome *M. tuberculosis* phage library. Lane M is 100bp ladder and 1-33 are PCR amplicons of library clones.

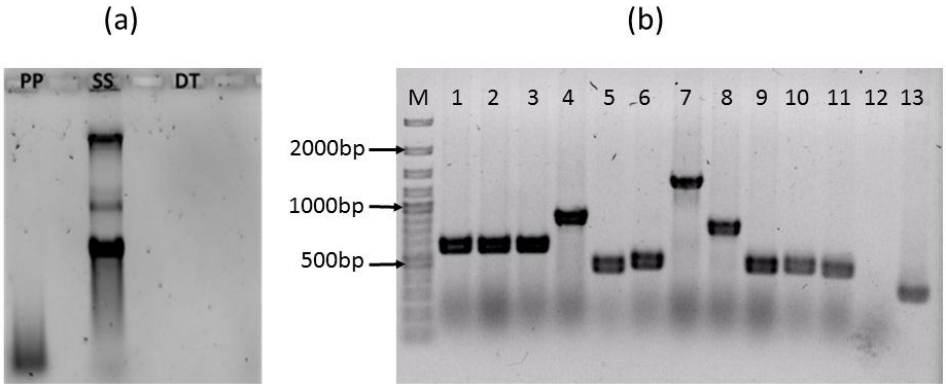

Figure 2: (a) Selective disassembly of phage particles display no or non-secretory protein. PP: phage particles prepared from *M. tuberculosis* phage library; SS: DNA released from sarcosyl sensitive phage particles; DT: DNase I treatment results showing almost complete removal or digestion of phage DNA released during SS step. (b) PCR amplicons of randomly selected clones of the phage secretome sub-library.

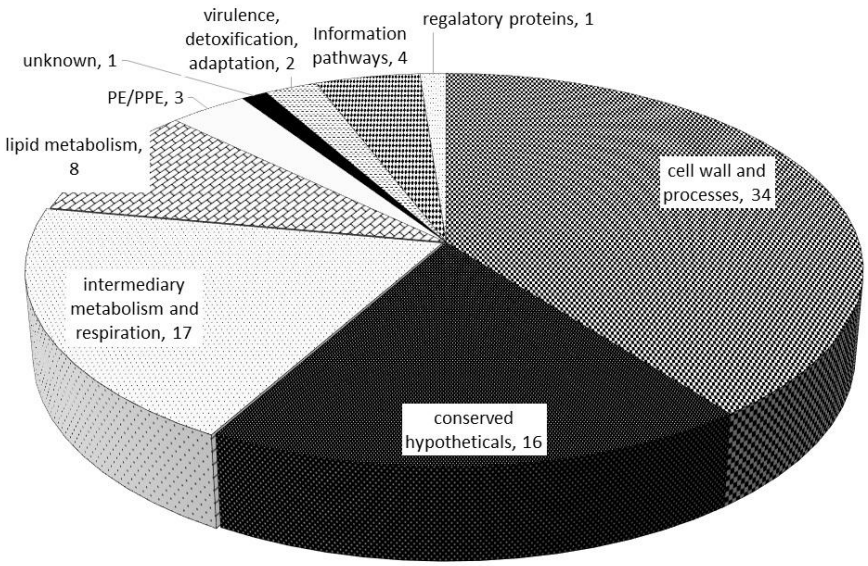

Figure 3: Distribution of the functional categories of 86 identified ORFs from F15/LAM4/KZN phage displayed secretome. The number of proteins in the different functional categories were: cell wall and cell processes (34), conserved hypothetical proteins (16), intermediary metabolism and respiration (17), lipid metabolism (8), information pathways (4), PE/PPE family proteins (3), virulence, detoxification, and adaptation (2), regulatory proteins (1) and unknown (1) function according to UniProt. Functional group codes were obtained from the Tuberculist database web server except for the unknown protein that was confirmed by the UniProt server.