Comparative Genomics and Proteomic Analysis of Four Non-tuberculous Mycobacterium Species and Mycobacterium tuberculosis Complex: Occurrence of Shared Immunogenic Proteins

Nomakorinte Gcebe1,2*, Anita Michel2, Nicolaas C. Gey van Pittius3 and Victor Rutten2,4

1 Tuberculosis Laboratory, Agricultural Research Council - Onderstepoort Veterinary Institute, Onderstepoort, South Africa, 2 Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa, 3 Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Centre of Excellence for Biomedical Tuberculosis Research, Stellenbosch University, Tygerberg, South Africa, 4 Division of Immunology, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands

The Esx and PE/PPE families of proteins are among the most immunodominant mycobacterial antigens and have thus been the focus of research to develop vaccines and immunological tests for diagnosis of bovine and human tuberculosis, mainly caused by Mycobacterium bovis and Mycobacterium tuberculosis, respectively. In non-tuberculous mycobacteria (NTM), multiple copies of genes encoding homologous proteins have mainly been identified in pathogenic Mycobacterium species phylogenetically related to Mycobacterium tuberculosis and Mycobacterium bovis. Only ancestral copies of these genes have been identified in nonpathogenic NTM species like Mycobacterium smegmatis, Mycobacterium sp. KMS, Mycobacterium sp. MCS, and Mycobacterium sp. JLS. In this study we elucidated the genomes of four nonpathogenic NTM species, viz Mycobacterium komanii sp. nov., Mycobacterium malmesburi sp. nov., Mycobacterium nonchromogenicum, and Mycobacterium fortuitum ATCC 6841. These genomes were investigated for genes encoding for the Esx and PE/PPE (situated in the esx cluster) family of proteins as well as adjacent genes situated in the ESX-1 to ESX-5 regions. To identify proteins actually expressed, comparative proteomic analyses of purified protein derivatives from three of the NTM as well as Mycobacterium bovis and Mycobacterium avium was performed. The genomic analysis revealed the occurrence in each of the four NTM, orthologs of the genes encoding for the Esx family, the PE and PPE family proteins in M. bovis and M. tuberculosis. The identification of genes of the ESX-1, ESX-3, and ESX-4 region including esxA, esxB, ppe68, pe5, and pe35 adds to earlier reports of these genes in nonpathogenic NTM like M. smegmatis, Mycobacterium sp. JLS and Mycobacterium KMS. This report is also the first to identify esxN gene situated within the ESX-5 locus in M. nonchromogenicum. Our proteomics analysis
identified a total of 609 proteins in the six PPDs and 22 of these were identified as shared between PPD of M. bovis and one or more of the NTM PPDs. Previously characterized M. tuberculosis/M. bovis homologous immunogenic proteins detected in one or more of the nonpathogenic NTM in this study included CFP-10 (detected in M. maltesburi sp. nov. PPD), GroES (detected in all NTM PPDs but M. maltesburi sp. nov.), DnaK (detected in all NTM PPDs), and GroEL (detected in all NTM PPDs). This study confirms reports that the ESX-1, ESX-3, and ESX-4 regions are ancestral regions and thus found in the genomes of most mycobacteria. Identification of NTM homologs of immunogenic proteins warrants further investigation of their ability to cause cross-reactive immune responses with MTBC antigens.

Keywords: non-tuberculous mycobacteria, Esx family, PE/PPE, M. fortuitum, M. maltesburi sp. nov., M. komanii sp. nov., M. nonchromogenicum

INTRODUCTION

Purified protein derivatives (PPDs) also known as tuberculins have been used for more than 100 years as antigens in the diagnosis of human and bovine tuberculosis (TB), based on the delayed hypersensitivity (skin) reaction they induce. The first preparation of PPD was introduced by Robert Koch in 1890, where a boiled crude extract of Mycobacterium containing a mixture of proteins, both somatic and antigenic, referred to as “old tuberculin,” was prepared as a potential vaccine against tuberculosis in humans. Whilst the “old tuberculin” failed to live up to its initial claim of having curative properties, it formed the foundation for the modern PPD preparations (Burke, 1993). Tuberculin based tests (tuberculin skin test and the interferon gamma assay) have since been used as diagnostic tools for bovine tuberculosis. While these tests generally show satisfactory sensitivities, the major drawback to the tuberculin based assays is the reduced specificity. This is presumably associated with cross-reactive immune responses due to exposure of animals and humans to non-tuberculous mycobacteria (NTM) or environmental mycobacteria and the TB vaccine strain, Bacille Calmette et Guérin (BCG). These cross-reactive immune responses are most likely due to the presence of immunogenic proteins conserved across the Mycobacterium genus as major components of PPD derived from M. bovis (Schiller et al., 2010). PPD prepared from Mycobacterium avium is included as a control antigen representative of the background responses to environmental mycobacterial antigens in the single intradermal comparative cervical tuberculin test (SICCT) as well as the interferon gamma assays used in different countries in bovine tuberculosis (BTB) control strategies (Vordermeier et al., 2006). As an alternative approach to mitigate the frequency of false positive responses to PPD-B induced by environmental mycobacterial antigens a PPD derived from Mycobacterium fortuitum (PPD-F) used in the modified BOVIGAM assay, has shown to increase the test specificity in South Africa (Michel et al., 2011).

Interference of NTM and BCG in immune responsiveness against bacteria of the Mycobacterium tuberculosis complex (MTBC) has triggered investigation of “specific” or defined antigens which are uniquely present in pathogenic mycobacteria, and absent in most NTM as well as M. bovis BCG as markers for the diagnosis of bovine and human tuberculosis (Schiller et al., 2010). Comparative genomics studies have led to the identification and characterization of certain genetic regions in the genomes of mycobacteria that are present in Mycobacterium tuberculosis and Mycobacterium bovis but absent in M. bovis BCG and most NTM. The Region of Difference 1 (RD1), a major deleted region in BCG first described by Mahairas et al. (1996) seems to have been the early event in the attenuation of the present vaccine strain, hence proteins within this region are studied extensively in order to try to understand the virulence of pathogenic MTBC. The two most predominant proteins encoded in the RD1 region, namely the 6 kDa early secreted antigenic target (ESAT-6/EsxA) and the 10 kDa culture filtrate protein (CFP-10/EsxB) have been investigated widely for inclusion into candidate vaccines and used as antigens in the diagnosis of both TB and BTB (Pym et al., 2003; Vordermeier et al., 2006; Ganguly et al., 2008). Functional studies have revealed that RD1 encodes a specialized system for secretion in mycobacteria, involving genes situated both inside the RD1 locus and its flanking regions, known as the ESAT-6 secretion system (Brodin et al., 2004). The ESAT-6 and CFP-10 proteins belong to the Esx family that consists of 23 small secreted proteins encoded by genes esxA-esxW and arranged in 11 gene pairs (Bitter et al., 2009). Five of the 11 genomic pairs are contained in five loci: ESX-1 to ESX-5 encoding components of the type VII secretion system (Uplekar et al., 2011). Four of the five loci, except ESX-4, also harbor 11 proteins of the Proline Glutamate (PE), or Proline Proline Glutamate (PPE) family, several of which have been implicated in immune evasion (Akhter et al., 2012; Mustafa, 2014). The PE and PPE and Esx proteins whose genes are adjacent to each other in the genome of Mycobacterium tuberculosis interact to form pairs which appear to be essential for their secretion (Renshaw et al., 2005). The secretory proteins have been the focus of research aimed at development of TB vaccines and immunodiagnostic assays, because they are thought to have a potential to induce protective immunity and immune responses of diagnostic value (Ize and Palmer, 2006; Ganguly et al., 2008; Marongui et al., 2013; Mustafa, 2014).
Studies have shown however, that some genes encoding for proteins of the Esx family, as well as PE/PPE do occur in pathogenic NTM like Mycobacterium kansasi Mycobacterium marinum, Mycobacterium szulgai, Mycobacterium gordonae, Mycobacterium gastri, Mycobacterium ulcerans, Mycobacterium sp. JDM601, and Mycobacterium riyadhense, as well as in Mycobacterium leprae (Gey van Pittius et al., 2001, 2006; Geluk et al., 2002, 2004; Newton-Foot et al., 2016) while genes encoding for PE/PPE family of proteins were also identified in Mycobacterium abscessus and Mycobacterium avium (Sorensen et al., 1995; Vordermeier et al., 2007; van Ingen et al., 2009; Mustafa, 2014). The immunogenicity and cross-reactivity against MTB antigens of the ESAT-6 and CFP 10 proteins of pathogenic NTM was as well-demonstrated in M. kansasi (Waters et al., 2010). Therefore use of these antigens in BTB diagnosis may be hampered by the exposure of humans and animals to pathogenic NTM leading to false positive BTB immunological test results.

Comparative genomic studies of rapidly growing mycobacteria (RGM) have shown that, contrary to the genomes of pathogenic MTBC and other pathogenic slow-growing NTM like M. kansasi and M. marinum which contain the five ESX regions (ESX-1 to ESX-5), the nonpathogenic RGM such as Mycobacterium flavescens, M. sp. JLS, M. sp. MCS, M. sp. KMS, Mycobacterium phlei, Mycobacterium thermoresistible, Mycobacterium gilvum PYR-GCK, Mycobacterium vaccae, Mycobacterium vanbaalenii, and M. smegmatis genomes only harbor ESX-1, ESX-3, and ESX-4 (Gey van Pittius et al., 2006; Newton-Foot et al., 2016). It was therefore hypothesized that the three regions (ESX-1, ESX-3, and ESX-4) are genetic factors of all mycobacteria and ESX-2 and ESX-5 are probably absent in all RGM (Gey van Pittius et al., 2006).

To assess the presence of homologous proteins in NTM potentially able to elicit cross-reactive immune responses against MTBC antigens we elucidated the genomes of four NTM species, viz Mycobacterium nonchromogenicum, Mycobacterium fortuitum ATCC 6841, Mycobacterium malnbsurii sp. nov., and Mycobacterium komandii sp. nov. and compared these with the genomes of Mycobacterium tuberculosis H37Rv and Mycobacterium bovis AF2122/97 to identify orthologous genes situated in the ESX-1 to ESX-5. In addition, the proteomes of PPDs derived from the NTM viz M. nonchromogenicum (PPD-N), M. fortuitum ATCC 6841 (PPD-F), M. malnbsurii sp. nov (PPD-M), and M. kansasi ATCC 12478 (PPD-K) were defined to identify the proteins actually produced and compared to those of the commercially available PPDs derived from M. avium and M. bovis. This is the first report of the occurrence of selected genes encoding proteins that were previously shown to be immunogenic in M. bovis/M. tuberculosis in the four NTM genomes as well as the expression of proteins they code for.

**MATERIALS AND METHODS**

**Bacterial Cultures**

*M. malnbsurii* sp. nov. and *M. komandii* sp. nov. were isolated from cattle nasal swabs during an NTM survey in South Africa (Gcebe et al., 2013). These two species were previously identified and characterized as novel NTM (Gcebe et al., unpublished). The *M. nonchromogenicum* strain was isolated from soil during an NTM survey (Gcebe et al., 2013). *M. fortuitum* ATCC 6841 and *M. kansasi ATCC 12478 were obtained from the Agricultural Research Council-Onderstepoort Veterinary Institute, South Africa. The identity of all the strains used was confirmed by PCR-sequencing assays targeting the 16S rDNA (Gcebe et al., 2013), *hsps* (Telenti et al., 1993), *rpOB* (Adékambi et al., 2006), and sodA (Adékambi and Drancourt, 2004). For PPD production, liquid cultures were prepared in Middlebrook 7H9 medium (Becton Dickinson, USA) supplemented with 0.1% OADC and glycerol, incubated under continued shaking at 200 g at 37°C (optimum growth temperature) for 4 weeks for the rapid growing mycobacteria (*M. fortuitum* and *M. malnbsurii* sp. nov.) and 6 weeks for the slow growing NTM (*M. kansasi* and *M. nonchromogenicum*) or until turbid growth was observed. The liquid cultures were screened for contaminants before PPDs were prepared by spread plating each culture on two nutrient agar plates. The plates were incubated at 25 and 37°C respectively, and evaluated after 2 and 5 days for fungal or any bacterial growth not typical of mycobacteria.

**Whole Genome Sequencing, Assembly, and Annotation**

DNA was extracted from bacterial cultures (*M. fortuitum*, *M. nonchromogenicum*, *M. komandii* sp. nov., and *M. malnbsurii* sp. nov.) in Löwenstein -Jensen medium (Becton Dickinson, USA) using the Qiagen nucleic acid extraction kit (Whitehead Scientific, South Africa). Genomic DNA paired-end libraries were generated using the Nextera DNA sample preparation kit (Illumina) and indexed using the Nextera index kit (Illumina). Sequencing was performed as paired-end reads (2 x 250 bp) employing the Illumina Miseq system using the Miseq reagent kit v2 at the Agricultural Research Council, as per manufacturer’s instructions. *De novo* assembly of the sequencing reads was performed using SPAdes, which mainly involves construction of DNA sequences of contigs and the mapping of reads to contigs (Bankevich et al., 2012). Each assembly was evaluated using QUAST (Gurevich et al., 2013). The assembled genome sequences were annotated using Prokka annotation pipeline (Seemann, 2014). This involved predicting tRNA, rRNA, and mRNA genes in the sequences and assigning putative gene products to the protein-coding genes (CDSs) based on their similarity to sequences in a database of curated *Mycobacterium* genes. To gauge the similarity of the NTM genomes to those of *M. bovis* AF2122/97 (NC 002945.3) and *M. tuberculosis* H37Rv (NC 000962.2), alignment of each of the processed sets of reads to these reference genomes was performed using Burrow’s Wheeler Aligner (BWA) (Li and Durbin, 2009). BLAST Ring Image Generator (Alikhan et al., 2011) was used to visualize the pairwise BLAST results of CDS features in *M. bovis* and each NTM annotation against that of *M. tuberculosis* and against the respective closest NTM relative.

**Investigating the Closest Sequence Relatives of the NTM Strains**

Basic local Alignment Search tool (BLAST) search of the largest contigs from each NTM assembly was performed against the NCBI Genbank database (http://www.ncbi.nlm.nih.gov).
gov/genbank/) to identify the closest known sequence relative of each NTM species. BLAST searches were also performed for the annotation CDS features for each NTM species against the reference genomes for their respective best hits and visualized using BLAST Ring Image Generator (BRIG) (Alikhan et al., 2011).

### TABLE 1 | List of target genes within ESX-1 to ESX-5 loci of *M. bovis* and *M. tuberculosis*.

| Name/Gene ID | Species | RefSeq/Genbank | Locus Tag | Protein product |
|--------------|---------|----------------|-----------|-----------------|
| esxA         | *M. tb* | WP_000339993.1 | Rv3875    | 6 kDa early secretory antigenic target (ESAT-6) |
|              | *M. bovis* | Mb3905        |           | 6 kDa early secretory antigenic target (ESAT-6) |
| esxB         | *M. tb* | WP_0003399940.1 | Rv3874    | 10 kDa culture filtrate antigen LHP (CFP10) |
|              | *M. bovis* | Mb3904        |           | 10 kDa culture filtrate antigen LHP (CFP10) |
| esxC         | *M. tb* | WP_0003899750  | Rv3890c   | ESAT-6 like protein EsxC (ESAT-6 like protein 11) |
|              | *M. bovis* | Mb3919c       |           | Putative esat-6 like protein 11 |
| esxD         | *M. tb* | WP_003400035.1 | Rv3891C   | Possible esat-6 like protein EsxD |
|              | *M. bovis* | Mb3920C       |           | Conserved hypothetical protein |
| esxG         | *M. tb* | WP_003401503.1 | Rv0287    | ESAT-6 like protein EsxG |
|              | *M. bovis* | Mb0295        |           | Hypothetical protein |
| esxH         | *M. tb* | WP_003401514.1 | Rv0288    | ESAT-6 like protein EsxH |
|              | *M. bovis* | Mb0296        |           | low molecular weight protein antigen 7 CFP7 (TB10.4) |
| esxT         | *M. tb* | WP_003900059.1 | Rv3444C   | ESAT-6 like protein EsxT |
|              | *M. bovis* | Mb3474C       |           | Hypothetical protein |
| esxU         | *M. tb* | WP_003900060.1 | Rv3445C   | ESAT-6 like protein EsxU |
|              | *M. bovis* | Mb3475C       |           | Hypothetical protein |
| esxN         | *M. tb* | WP_003408840  | Rv1793    | ESAT-6 like protein EsxN |
|              | *M. bovis* | Mb1821        |           | ESAT-6 like protein 5 |
| pe35         | *M. tb* | WP_003912361.1 | Rv3872    | PE family protein |
|              | *M. bovis* | Mb3902        |           | PE family protein |
| ppe68        | *M. tb* | WP_003399879.1 | Rv3873    | PPE family protein |
|              | *M. bovis* | Mb3903        |           | PPE family protein |
| pe5          | *M. tb* | WP_003401485.1 | Rv0285    | PE family protein |
|              | *M. bovis* | NP_853967.1   | Mb0293    | PE family protein |
| ppe4         | *M. tb* | WP_003401488.1 | Rv0286    | PPE family protein |
|              | *M. bovis* | Mb0293        |           | PPE family protein |
| ppe25        | *M. tb* | WP_003901254.1 | Rv1787    | PPE family protein |
|              | *M. bovis* | Mb1815        |           | PPE family protein |
| pe18         | *M. tb* | WP_003408976.1 | Rv1788    | PE family protein |
|              | *M. bovis* | Mb1816        |           | PE family protein |
| ppe26        | *M. tb* | WP_003408979.1 | Rv1789    | PPE family protein |
|              | *M. bovis* | Mb1817        |           | PPE family protein |
| ppe27        | *M. tb* | WP_003408805.1 | Rv1790    | PPE family protein |
|              | *M. bovis* | Mb1818        |           | PPE family protein |
| pe19         | *M. tb* | WP_003408807.1 | Rv1791    | PE family protein |
|              | *M. bovis* | Mb1819        |           | PE family protein |
Identification of Predicted Immunogenic Proteins in NTM Genomic Annotations

BLAST searches were conducted to quantify the similarity of the annotated amino acid sequences in the NTM assemblies (M. fortuitum, M. malmesburii sp. nov., M. komani sp. nov., and M. nonchromogenicum) to immunogenic proteins in M. bovis, M. tuberculosis, and M. smegmatis. Esx family proteins, PE and PPE proteins situated within the ESX-1 to ESX-5 clusters were chosen for this investigation (see Table 1 for the complete list of targeted proteins). BLASTP searches against the NCBI Genbank database were also conducted for the respective NTM protein sequences to quantify similarity to other sequences in the database. Multiple amino acid sequence alignments of the individual proteins: (Esx, PE, PPE situated in ESX-1 and in the database. Multiple amino acid sequence alignments of the individual proteins: (Esx, PE, PPE situated in ESX-1 and in the database. Multiple amino acid sequence alignments of the individual proteins: (Esx, PE, PPE situated in ESX-1 and in the database. Multiple amino acid sequence alignments of the individual proteins: (Esx, PE, PPE situated in ESX-1 and in the database. Multiple amino acid sequence alignments of the individual proteins: (Esx, PE, PPE situated in ESX-1 and in the database. Multiple amino acid sequence alignments of the individual proteins: (Esx, PE, PPE situated in ESX-1 and in the database.

TABLE 2 | Data acquisition.

| FULL SCAN | |
|---|---|
| Resolution | 70,000 (@ m/z 200) |
| AGC target value | 3e6 |
| Scan range | 320–2000 m/z |
| Maximal injection time (ms) | 100 |

| DATA-DEPENDENT MS/MS | |
|---|---|
| Resolution | 17,500 (@ m/z 200) |
| AGC target value | 2e5 |
| Maximal injection time (ms) | 50 |
| Isolation window width (Da) | 3 |
| NCE (%) | 27 |

| DATA-DEPENDENT SETTINGS | |
|---|---|
| Underfill ratio (%) | 1 |
| Charge exclusion | Charge states 1,6-8,>8 |
| Peptide match | preferred |
| Exclusion isotopes | on |
| Dynamic exclusion (s) | 60 |

PCR Verification of the Presence of Genes in ESX-1, ESX-3 as well as esxN in NTM

We set out experiments to confirm the presence of ESX-1 in the four NTM genomes by PCR and sequencing of the two most investigated genes in this region: esxA and esxB. The presence of ESX-3 in the four newly sequenced NTM was verified by amplification of esxH. Since it was previously shown by hybridization experiments that ESX-5 might be absent in Mycobacterium nonchromogenicum we verified the presence of esxN ortholog (situated in ESX-5 locus) in M. nonchromogenicum by PCR and sequencing.

The primers used for amplification of esxA were MSMEG_0066-F (5′-AATTTCCGGGCTATCGAGGG-3′) located at position 19–38 bp of the MSMEG_0066 gene (an M. smegmatis MC²155 esxA ortholog), and MSMEG_0066-R (5′-AGGACACATTCGTTGGAC-3′) located at position 278–268 bp of the esxA ortholog of M. smegmatis MC²155 (MSMEG_0066).

The primers used for amplification of esxB were MSMEG_0065-F (5′-GCGAGTTCTCGGCACTTC-3′) situated at position 46–65 bp, and MSMEG_0065-R (5′-GATTGTCTACGGGACGC-3′) situated at position 300–280 bp.

Primers used for esxH were esxH-13 (5′-ARGTACACATCCCGCGATG-3′) situated at position 13–33 bp of MSMEG_0621 gene (an M. smegmatis ortholog of esxH), and esxH-245 (5′-GCGATGTGTGTCTCTGCT-3′) situated at position 245–226 bp of MSMEG_0621 gene.

Finally, primers used for amplification of esxN were esxN-F (5′-GACGATATTACCTACGCT-3′) situated at position 3–20 bp of the gene in M. tuberculosis, and esxN-R (5′-GCTCCTGCTGCTGCGC-3′) situated at position 255–239 bp of the gene.

PCR

Boiled culture suspensions were prepared from colonies as template for DNA amplification. The following PCR conditions were used for amplification of each of the three gene fragments: A 50 µl PCR mixture was prepared, containing 28.5 µl deionized water, 3 µl MgCl₂(25 mM), 1 µl dNTP mix (10 mM), 4.75 µl of 10x PCR buffer (160 mM) (Tris -HCl, MgCl₂, Tween 20, (NH₄)₂SO₄), 0.75 µl Taq DNA Polymerase (5 U/µl) (Supertherm™), 1 µl of each forward and reverse primers (50 pmol), and 10 µl of DNA template. The PCR cycling parameters were as follows: Forty five cycles of denaturation at 94°C for 1 min, annealing at 60°C (55°C for esxN) for 1 min, and elongation at 72°C for 1 min and final extension at 72°C for 10 min. The PCR products were ran on a 1.5% agarose gel and visualized under ultra violet light (UV).

Sequencing and Sequence Analysis

Sequencing of the PCR products was done at Inqaba Biotechnologies (Pretoria, South Africa). Sequencing was performed in both directions using the forward and reverse primer sequences that were initially used for amplification. Sequences from both strands were edited manually and pairwise
alignments undertaken using the BioEdit Sequence alignment editor (version 7.1.9) and Molecular Evolutionary Genetics Analysis (MEGA) platform (http://www.megasoftware.net) (version 6) (Tamura et al., 2013). The resulting consensus sequences were analyzed on the NCBI platform for gene sequence identity/similarity using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

### PPD Production

PPDs were prepared from *M. nonchromogenicum* (PPD-N), *M. malmesburi* sp. nov. (PPD-M), *M. fortuitum* ATCC 6841 (PPD-F), and *M. kansasii* ATCC 12478 (PPD-K). Three preparations of each PPD [one at a different time interval (first preparation) and two done at the same time (second and third preparations)] were done following a modified protocol by Landi (1963). Briefly, liquid cultures were inactivated by steaming at 121°C for 20 min and then filter-sterilized using the Whatman 40 filter paper and a vacuum pump. Each filtrate was then precipitated by adding 40% trichloroacetic acid (TCA) to a final concentration of 4% v/v and left for at least 12 h at 4–8°C. Afterwards the precipitated filtrates were mixed manually by shaking and centrifuged at room temperature for 20 min at 3900 g. The supernatants were discarded and the pellets were washed twice by suspending them in 1% TCA and then centrifuging at room temperature for 20 min at room temperature. The supernatants were discarded and the pellets suspended in 10% NaCl, then centrifuged for 20 min at 3900 g. After discarding the supernatant, the pellet was harvested by turning the tube upside down, on a piece of sterile filter paper and allowed to dry, weighed, diluted with 0.005% PPD buffer (0.005% Tween 80 in PBS: PH = 7.38) and stored at 4–8°C until peptide digestion. The purified protein derivatives from *M. bovis* (PPD-B) and *M. avium* (PPD-A) were obtained from Prionics at the Netherlands.

### In-Solution Digest of PPDs with Trypsin (IAA)

Trypsin digestion of each PPD preparation was done at the Centre for Proteomics and Genomics Research (CPGR, South Africa). All reagents used were analytical grade or equivalent. Twenty microgram of protein was dissolved in 10 µl of 50 mM triethylammonium bicarbonate (TEAB; Sigma). Protein was reduced by adding 1 µl of 100 mM tris (2-carboxyethyl) phosphine (Sigma) prepared in 50 mM TEAB, and incubated at 60°C for 1 h. Samples were cooled to room temperature and the protein was alkylated by adding 1 µl of 200 mM iodoacetamide (Sigma) prepared in 50 mM TEAB. Samples were incubated at room temperature in the dark for 30 min. The sample volume was adjusted to 50 µl with 50 mM TEAB and then 5 µl of 1 µg/µl trypsin (Promega), prepared in MilliPore water, was added and digestion was allowed to take place at 37°C for 18 h, followed by vacuum centrifugation.

### Table 3 | Summary of the de novo assembly results.

| NTM Species     | Strain ID  | number of contigs (n) | Largest contig (bp) | Total Length (bp) | L50 | N50 |
|----------------|------------|-----------------------|---------------------|-------------------|-----|-----|
| *M. fortuitum*  | ATCC 6841  | 179                   | 195 004             | 6 275 573         | 27  | 82 193  |
| *M. nonchromogenicum* | NCK 8460 | 642                   | 425 774             | 6 623 407         | 17  | 131 682 |
| *M. malmesburi* sp. nov. | WCM 7299 | 255                   | 135 683             | 5 426 606         | 42  | 42 651 |
| *M. komani* sp. nov. | GPK 1020  | 63                    | 298 570             | 5 370 343         | 12  | 162 386 |

### Table 4 | Annotation of the assembled genome, showing number of CDS, tRNA, rRNA, and signal peptides.

| NTM species     | Strain ID  | CDS  | tRNA | rRNA | Signal peptides |
|----------------|------------|------|------|------|----------------|
| *M. fortuitum*  | ATCC 6841  | 6097 | 66   | 3    | 499            |
| *M. nonchromogenicum* | NCK 8460 | 6903 | 70   | 3    | 544            |
| *M. malmesburi* sp. nov. | WCM 7299 | 5298 | 54   | 2    | 401            |
| *M. komani* sp. nov. | GPK 1020  | 5141 | 53   | 2    | 408            |

NTM, non-tuberculous mycobacterium; CDS, coding sequences.

### Table 5 | BLAST results using the largest contig of each NTM against Genbank nucleotide database (E < 1 x 10^-6).

| NTM species     | Species hit               | BLAST score | Query coverage (%) | Identity (%) |
|----------------|---------------------------|-------------|-------------------|--------------|
| *M. fortuitum*  | *M. smegmatis* MC^2 155   | 11121       | 65                | 81           |
| *M. nonchromogenicum* | *M. smegmatis* MC^2 155   | 8822        | 69                | 83           |
| *M. malmesburi* sp. nov. | *M. rhodesiae* NBB3       | 8503        | 52                | 81           |
| *M. komani* sp. nov. | Mycobacterium sp. JLS     | 5565        | 45                | 81           |
Mass Spectrometry

LC–MS/MS Analysis

LC-MS/MS analysis was conducted in triplicate for NTM PPDs from the second and the third preparations and once for the first preparation and the commercial PPDs (PPD from *M. bovis* and from *M. avium*) using a Q-Exactive quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, USA) coupled with a Dionex Ultimate 3000 nano-HPLC system at CPGR. The mobile phases consisted of solvent A (0.1% formic acid in water) and solvent B (80% ACN, 10% water, and 0.1% formic acid). The peptides (as an estimate 500 ng for each sample) were re-suspended in sample loading buffer (95% water, 5% Acetonitrile, 0.05% TFA) and loaded on a C18 trap column (100 µm × 20 mm × 5 µm). Chromatographic separation was performed with a C18 column (75 µm × 150 mm × 3 µm). The gradient was delivered at 300 nl/min and consisted of a linear gradient of mobile phase B initiating from solvent B, 6–60% over 156 min. The mass spectrometer was operated in positive ion mode with a capillary temperature of 250°C. The applied electrospray voltage was 1.95 kV. Details of data acquisition are listed in Table 2.

LC-MS/MS Data Analysis

Database Searching

All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.4.1) and X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1). Mascot was set up to search the *Mycobacterium* database (derived from UniProtKB, 843597 entries) assuming the digestion enzyme trypsin. X! Tandem was set up to search a subset of the *Mycobacterium* database including also assuming trypsin. Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.020 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethyl of cysteine were specified in Mascot and X! Tandem as fixed modifications. Gln->pyro-Glu of the n-terminus, deamidation of asparagine and glutamine, oxidation of methionine were specified in Mascot and X! Tandem variable modifications.

Criteria for Protein Identification

Scaffold (version Scaffold_4.3.4, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 9.0% probability to achieve an FDR less than 0.1% by the Scaffold Local FDR algorithm (Keller et al., 2002). Protein identifications were accepted if they could be established at >100.0% probability to achieve an FDR less than 1.0% and contained at least four identified peptides (Keller et al., 2002). Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

![Figure 1](image-url)
RESULTS

Alignment of the Four NTM Sequence Reads with the Genomes of *M. bovis* and *M. tuberculosis*

Alignment of the NTM reads (DNA sequences) and the reference genomes [*M. bovis* (NC 002945.3) and *M. tuberculosis* (NC 000962.3)] showed very little similarity between the NTM genomes and each of the reference strains. *M. komanii* sp. nov. genome seems to be the closest of the NTM to the reference genomes (18.61 and 18.69% of the sequence reads maps to *M. bovis* and *M. tuberculosis* genomes, respectively), followed by *M. nonchromogenicum* (15.71 and 15.8%), and *M. fortuitum* (10.58 and 10.63%). *M. malmesburi sp. nov.* showed the least similarity of its genome to that of the reference strains (8.83 and 8.89%). The GC contents of the NTM genomes were 66.19% for *M. fortuitum*, 66.25% for *M. nonchromogenicum*, 67.4% for *M. malmesburi sp. nov.*, and 67.33% for *M. komanii* sp. nov., which is as high as the GC contents of *M. tuberculosis* and *M. bovis* (both have GC contents of 66.5%).

De novo Assembly of the NTM DNA Sequence Reads

The result for the assembly of the DNA sequence reads is summarized in Table 3. The number of contigs, the length of the largest contig as well as the N50 (the length for which the set of contigs of that length or longer contains at least half the assembly bases) and L50 (the number of contigs of that length for each assembly is presented). The overwhelming majority of the sequenced bases fall within the few largest contigs, but each assembly also produced a large number of very short contigs. The assembly for *M. komanii* sp. nov. produced 63 contigs, with its largest contig 292, 570 bases in size. While the assembly for *M. nonchromogenicum* produced the largest contig of all (425, 774 bases), it also produced the highest number of contigs (642 bases). The *M. fortuitum* and *M. malmesburi sp. nov.* assemblies provided high total numbers of contigs (179 and 255 respectively), their largest contigs were 195, 004 and 135, 683 bases respectively.

Annotation of the Assembled Genome

The summary of the annotation is provided in Table 4. The numbers of protein coding genes, sequences encoding tRNA...
and rRNA and signal peptides identified in each of the four NTM genomes are indicated. The annotated assemblies were deposited into the European Nucleotide Archive (ENA) database (http://www.ebi.ac.uk/ena). The sequence identifiers i.e., locus tag and the accession numbers are BN1849 and CVRX01000001-CVRX01000338, respectively for *M. fortuitum*; BN1848 and CVTC01000001-CVTC01002268, respectively for *M. nonchromogenicum*; BN1837 and CVTB01000001-CVTB01000324, respectively for *M. malmesburi* sp. nov.; and finally BN1838 and CVTA01000001-CVTA0100065, respectively for *M. komani* sp. nov.

**Closest Sequence Relatives of the Four NTM Genomes (M. malmesburi* sp. nov., *M. komani* sp. nov., *M. nonchromogenicum*, and *M. fortuitum* ATCC 6841)**

Table 5 represents the BLAST results for each NTM’s largest contig. *Mycobacterium fortuitum* and to a lesser extent *M. nonchromogenicum* have the closest resemblance to *M. smegmatis* MC^2^ 155 (accession number CP001663.1) with *M. malmesburi* sp. nov. displaying similarity to *M. rhodesiae* (CP003169.1). *M. komani* sp. nov. seems to be the most novel organism scoring its highest BLAST score with *Mycobacterium* sp. JLS (CP000580.1).

**Genome Assembly Visualization by Alignment to Reference Genomes (i.e., *M. rhodesiae*, *M. smegmatis* MC^2^155, *M. tuberculosis* H37Rv, and *Mycobacterium* sp. JLS)**

Visualization of the alignment of the *M. bovis* and the predicted CDS regions of the four NTM species, i.e., *M. nonchromogenicum*, *M. malmesburi* sp. nov., *M. komani* sp. nov., and *M. fortuitum* to those of *M. tuberculosis* H37Rv using BLAST Ring Image Generator (BRIG) is given in Figure 1, while alignment of each NTM predicted CDS to their closest relatives is
shown by the images in Figure 2 (M. fortuitum and M. smegmatis MC^2155), Figure 3 (M. nonchromogenicum and M. smegmatis MC^2155), Figure 4 (M. malmesburi sp. nov. and M. rhodesiae), and Figure 5 (M. komannii sp. nov. and Mycobacterium sp. JLS). The positions of the immunogenic proteins of interest in the reference genomes are also highlighted.

Orthologs of the Esx and PE/PPE Family Proteins Situated within the ESX-1 to ESX-5 Regions
The orthologs of the Esx and PE/PPE family proteins (situated within the ESX-1 to ESX-5 regions) deduced from the newly sequenced NTM genomes are listed in Table 6. The ESX-1, ESX-3 and ESX-4 loci were found in all four of the newly annotated NTM genomes, while the ESX-2 and ESX-5 loci seem to be absent in three of the NTM genomes (M. fortuitum, M. malmesburi sp. nov., and M. komannii sp. nov.). This is consistent with what was observed in the genomes of other rapidly growing NTM like M. smegmatis, Mycobacterium sp. JLS, Mycobacterium vanbaalenii, Mycobacterium vaccae, Mycobacterium sp. MCS, Mycobacterium thermostutes, Mycobacterium sp. KMS, and M. flavescens where only three of the ESAT-6 loci (ESX-1, ESX-3, and ESX-4) were found to be present (Gey van Pittius et al., 2001, 2006; Newton-Foot et al., 2016). In order to confirm the orthology of the NTM esx and pe/ppe genes to those of M. bovis/M. tuberculosis as well as the potential of their protein products to be secreted, we also investigated the newly sequenced NTM genomes for the adjacent genes situated in the ESX loci, as well as the genomic organization of these loci. The organization of these genes in the ESX loci is illustrated in Figure 6. EsxA and esxB as well as esxG and esxH were found next to each other in the genomes of the newly sequenced NTM. The pe35 and ppe68 genes were found to occur next to each other in the genomes of the four newly sequenced NTM. One member of the pe5/ppe4 pair (ppe4), was not detected in any of the newly sequenced NTM. Comparing the ESX loci of M. tuberculosis and M. bovis to those of the newly sequenced NTM, we found that, all the genes of the ESX-1 locus and their organization (as known to occur in M. tuberculosis and M. bovis) except espJ (whose function is unknown), were
found in all four newly sequenced NTM as shown in Figure 6. Likewise all the genes of the ESX-3 (except ppe4) and ESX-4 loci (as known to occur in M. tuberculosis and M. bovis) were also found in all four newly sequenced NTM (Figure 6). The only NTM species that was found to harbor an Esx ortholog of the ESX-5 locus (esxN) as well as 4 of the 12 ESX-2 genes was M. nonchromogenicum.

To quantify amino acid sequence homology of the EsxA and EsxB of Mycobacterium fortuitum in this study to other species in the NCBI database we conducted BLASTP searches. Interestingly, the BLAST searches using the M. fortuitum EsxA and EsxB amino acid sequences revealed the existence of their orthologs in additional NTM species, most notably M. vulneris (98% similarity to M. fortuitum at amino acid level for both proteins), M. mageritense (82 and 74% similarity to M. fortuitum respectively, at amino acid level), and M. farcinogenes (96 and 92% similarity to M. fortuitum respectively, at amino acid level).

Verification of the Presence the esx Genes in the Newly Sequenced NTM

The presence of esxA and esxB were confirmed in M. fortuitum by PCR and sequencing using primers designed from M. smegmatis sequences followed by BLAST searches. We demonstrated the presence of both these gene fragments in Mycobacterium fortuitum ATCC 6841 whose nucleotide sequences were 89 and 85% identical to the M. smegmatis esxA and esxB orthologs. M. tuberculosis/M. bovis derived primers did not amplify any of the NTM (results not shown). Likewise the M. smegmatis derived primers could not amplify these genes in other NTM (M. nonchromogenicum, M. malmesburii sp. nov., and M. komani sp. nov.).

The presence of esxH was also confirmed by PCR and sequencing in M. fortuitum (74% identical to the M. smegmatis esxH ortholog); M. komani sp. nov. (82% identical to the M. smegmatis ortholog); and finally in M. malmesburii sp.
nov., (84% identical to the *M. smegmatis* ortholog). The *M. nonchromogenicum* ortholog did not amplify when these primers were used.

Since DNA-DNA hybridization analysis by Gey van Pittius et al. (2006) previously suggested that ESX-5 may be absent in *M. nonchromogenicum* or the species was evolutionary so far removed from the slow growers that gene homology was insufficient to allow hybridization, we set out to verify the presence of *esxN* in this species by PCR and sequencing using primers designed from the *esxN/rv1793c* gene of *M. tuberculosis*. The highest BLAST hit of the sequenced gene fragment reveals that this sequence was 91% closely related to *Mycobacterium* species JDM601 *esxN* gene.

### Comparison of Amino Acid Sequence of the *M. bovis*/*M. tuberculosis* Immunogenic Epitopes of ESAT-6, CFP-10, TB10.4, PPE68, to the NTM Homologs

**ESAT-6 (EsxA)**

The amino acid sequence alignment of *M. fortuitum*, *M. nonchromogenicum*, *M. smegmatis*, *M. bovis*/*M. tuberculosis* ESAT-6 homologs is demonstrated in Figure 7. Immunogenic epitopes (underlined sequences in Figure 7) of *M. bovis* ESAT-6 as demonstrated by Vordermeier et al. (2000, 2003, 2007) were found in *M. fortuitum*, *M. nonchromogenicum* as well as in *M. smegmatis* orthologs. Comparison of *M. bovis*, *M. nonchromogenicum*, *M. smegmatis*, and *M. fortuitum* immunodominant epitopes revealed the following results: amino acid sequence similarities of 13/16 (81.28%) between the three NTM and *M. bovis* for the epitope at position 1-16; 9/17 (52.9%) and 10/17 (58.8%) amino acid residues were identical between *M. smegmatis* and *M. bovis* and between *M. nonchromogenicum*, *M. fortuitum*, and *M. bovis* respectively, for the epitope at position 16-32, while 12/18 (66%) amino acid sequence identity was observed between the three NTM and *M. bovis*. Finally 11/15 identity between *M. fortuitum*, *M. nonchromogenicum*, and *M. bovis*, and 12/15 between *M. smegmatis* and *M. bovis* for epitope at position 81-95. Amino acid sequence similarities of 90 and 95% for the ESAT-6 were observed between *M. smegmatis* and *M. fortuitum* as well as *M. smegmatis* and *M. nonchromogenicum*, respectively.

**CFP-10 (EsxB)**

Comparison of amino acid sequences of the CFP-10 homologs of *M. fortuitum*, *M. nonchromogenicum*, *M. smegmatis*, *M. malmesburi*, sp. nov, *M. komanii*, sp. nov, *M. bovis*, and *M. tuberculosis* is illustrated by the amino acid alignment in Figure 8 and immunogenic epitopes as demonstrated by Vordermeier et al. (2003, 2007) are underlined. Comparing the CFP-10 amino acid sequence of *M. tuberculosis* to the other NTM species, it is apparent that the epitopes are conserved in all of them. The highest identity was observed between *M. fortuitum* and *M. bovis* with 81.5% identity at position 51-64, followed by 81.28% identity between *M. nonchromogenicum* and *M. bovis* at position 51-64.

---

### TABLE 6 | Orthologs of the Esx and PE/PPE family proteins in the four NTM genomes.

| Locus | esx/pe/ppe gene tag | Percentage homology of Esx and PE/PPE amino acid sequences deducted from the different NTM to *M. bovis* |
|-------|---------------------|-----------------------------------------------------------------------------------------------------------|
|       |                     | *M. komani sp. nov.* | *M. malmesburi sp. nov.* | *M. fortuitum* | *M. nonchromogenicum* |
| ESX-1 | esxA/Rv3875/Mb3905  | 48.75               | 50                          | 75.79            | 71.58            |
|       | esxB/Rv3874/Mb3905  | 57.73               | 57.73                       | 64               | 64               |
|       | pe35/Rv3872/Mb3902  | 51.55               | 51.55                       | 52.22            | 52.22            |
|       | ppe68/Rv3873/Mb3903 | 63.65               | 47.79                       | 59.44            | 56.10            |
| ESX-2 | esxC/Rv3890c/Mb3919c | no                  | no                         | no              | no              |
|       | esxD/Rv3891c/Mb3920c | no                  | no                         | no              | no              |
|       | ppe69/Rv3892c/Mb3921c | no                  | no                         | no              | no              |
|       | pe36/Rv3893c/Mb3922c | no                  | no                         | no              | no              |
| ESX-3 | esxG/Rv0287/Mb0295  | 78.35               | 78.35                       | 81.5             | 81.05            |
|       | esxH/Rv0288/Mb0296  | 76                   | 76                          | 75.79            | 77.94            |
|       | pe5/Rv0285/Mb0293   | 72.28               | 72.28                       | 74.54            | 74.51            |
|       | ppe4/Rv0286/Mb0294  | no                  | no                         | no              | no              |
| ESX-4 | esxT/Rv3444c/Mb3474c | 73.4                | 73.96                       | 73.96            | 73.96            |
|       | esxU/Rv3445c/Mb3475c | 60.75               | 56.01                       | 60.42            | 60.42            |
| ESX-5 | esxV/Rv1792/Mb1820  | no                  | no                         | no              | no              |
|       | esxW/Rv1793/Mb1821  | no                  | no                         | no              | 85.11            |
|       | ppe25/Rv1787/Mb1815 | no                  | no                         | no              | no              |
|       | ppe18/Rv1788/Mb1816 | no                  | no                         | no              | no              |
|       | ppe26/Rv1789/Mb1817 | no                  | no                         | no              | no              |
|       | ppe27/Rv1790/Mb1818 | no                  | no                         | no              | no              |
|       | ppe19/Rv1791/Mb1819 | no                  | no                         | No              | No              |

no, no ortholog detected.
The unannotated genes in NTM encode for hypothetical proteins. They encode for (as well as ESX-2), and ESX-4 of *M. tuberculosis* annotation was also used for the NTM orthologs (except for *M. smegmatis*). Protein coding sequences are represented by blocked arrows reflecting the relative lengths of the genes and the direction of transcription. Genes of *M. tuberculosis* are color coded according to the protein families they encode for (Gey van Pittius et al., 2006). The unannotated genes in NTM encode for hypothetical proteins.
acid sequences of *M. bovis* and the NTM orthologs revealed the following: for the epitope at position 1-19, a total of 14/19 (73.7%) of the amino acid residues were identical between *M. bovis*, *M. smegmatis*, *M. nonchromogenicum*, and *M. fortuitum*, while only 12/19 (63%) amino acid residues were identical between *M. bovis*, *M. komnani* sp. nov., and *M. malmesburii* sp. nov. For the epitope at position 13-28, a total of 11/16 (68.75%) of the amino acids were found to be identical between *M. smegmatis*, *M. nonchromogenicum*, *M. fortuitum*, and *M. bovis*, while 10/16 (62.5%) were identical between *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. komnani* sp. nov., and *M. malmesburii* sp. nov. However, 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, *M. komnani* sp. nov., *M. malmesburii* sp. nov., and *M. bovis*. The epitope at position 28-44 had 9/16 (56.25%) amino acid residues identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 8/16 (50%) were identical between *M. bovis*, "Goebe et al. Genomics and Proteomics of Four NTM"
sp. nov. and *M. malmesburi* sp. nov. The epitope at position 55-72 showed sequence identity of 12/16 (75%) of the amino acid residues between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, *M. smegmatis*, *M. komanii* sp. nov., and *M. malmesb* sp. nov. For the epitope at position 56-76, 12/20 (60%) of the amino acids were identical between all five NTM and *M. bovis*. Finally, for epitope at position 76-93, 9/18 (50%) of the amino acid residues were identical between *M. bovis*, *M. fortuitum*, *M. nonchromogenicum*, and *M. smegmatis*, while 6/18 (33.3%) were identical between *M. bovis*, *M. komanii* sp. nov., and *M. malmesb* sp. nov.

**TB10.4 (EsxH)**

Alignment of the EsxH amino acid sequences of the NTM, *M. bovis*, and *M. tuberculosis* species is illustrated in Figure 9 and immunogenic epitopes described by Skjøt et al. (2002) are underlined. The immunodominant epitope of *M. tuberculosis* described by Mustafa (2014) was also found in *M. fortuitum* and *M. nonchromogenicum* showing 18/25 (72%) of the amino acid residues to be identical between the NTM and *M. tuberculosis*. *M. smegmatis* showed 19/25 (76%) amino acid sequence identity of this region to that of *M. tuberculosis*.

**Identification of Proteins Present in the PPD Preparations**

Using Mass spectrometry, 608 different proteins were identified in all the PPD preparations combined (consensus from the different analysis) (Supplementary data). One hundred and thirty two were identified in PPD-B. Among the NTM PPDs, PPD-A showed the highest number of proteins (*n* = 225) followed by PPD-F (*n* = 193), PPD-N (*n* = 189), and PPD-M (*n* = 136) while 39 proteins were identified in PPD-K. The molecular mass of the majority of the proteins was found to be in the range of 10–50 kDa. Twenty two proteins were identified as shared between the NTM PPD preparations and PPD-B.
| 10 | 20 | 30 | 40 | 50 |
|----|----|----|----|----|
| M. tubercu | M------L WHAMPPEINT ARIMAGAGPA PMLAAAAGWQ TLSALDAQA |
| M. bovis  | M------L WHAMPPEINT ARIMAGAGPA PMLAAAAGWQ TLSALDAQA |
| M. fortuit| MAVYVPVVPPT WLALPPEVT ARIMVGAGPA PMLQASGWG GLAILLQTEQA |
| M. nonchro| MAVYVPVVPPT WLALPPEVT ARIMVGAGPA PMLQASGWG GLAILLQTEQA |
| M. komani | M-------F WHALAPFNT ARIMAGAGAA PMLQAAAGWGE ALAIASLQTA |
| M. smegmat| M-------WIALPPEINT ARIMVGAGPA PMLQAAAGWGE GLAILLQTEQA |
|          | *** *** ** ***** ***** ********* * * ** |
| 60 | 70 | 80 | 90 | 100 |
| M. tubercu | VELTARLNSL EAAWGTSSDD KHALAADPMV VNQLATSTQA KTROMQATQA |
| M. bovis  | VELTARLNSL EAAWGTSSDD KHALAADPMV VNQLATSTQA KTROMQATQA |
| M. fortuit| DELAGALSQL TSVWSTASE RAVATPMPI MWLRTMVQA OQRALAQAQ |
| M. nonchro| DELAGALSQL TSVWSTASE RAVATPMPI MWLRTMVQA OQRALAQAQ |
| M. komani | DELASLTLT TSWSGSGSVE RVQAVTMMPI AWLVTSAQA MKRALAQAQ |
| M. smegmat| AELAALTNL TSWVTGAASE RAVATPMPI TNLQTTAQA OQRALAQAQ |
|          | ** * * >>* ** * * <<* ** * * ** ** ** ** |
Goebe et al. Genomics and Proteomics of Four NTM

Figure 10: Alignment of PPE68 aa sequences of M. fortuitum ATCC 6841 (M. fortuit), M. smegmatis MC^2 155 (M. smegmat), M. nonchromogenicum (M. nonchro), M. komanii sp. nov., M. bovis, and M. tuberculosis (M. tubercu). *Indicates identical aa sequences in all species, and > indicates same aa residue in at least one of the NTM species and M. bovis/M. tuberculosis. Highly immunogenic epitope of M. tuberculosis, as described by Mustafa (2014) are Underlined.

(Table 7). Identities and functions of these proteins are also listed in Supplementary Table 8. These proteins include previously characterized immunodominant MTBC proteins like: the 10 kDa culture filtrate protein/CFP-10 (shared between PPD-B and PPD-M); the 10 kDa chaperonin GroES (shared between the NTM PPDs except PPD-M, and PPD-B); 60 kDa chaperonin GroEL (shared by all the NTM PPDs and PPD-B); Chaperone protein DNAK (shared by all the NTM PPDs and PPD-B); Secreted antigen Ag85C (shared between PPD-A and PPD-B), 6 kDa early secretory antigenic target/ESAT-6 (shared between PPD-K and PPD-B) and EsxN (shared between PPD-K and PPD-B). The Venn diagrams in Figure 11 illustrate the number of shared proteins between PPD-B, PPD-A and different NTM PPDs, as well as unique proteins of each PPD. There was a higher degree of protein overlap between PPD-A and PPD-B (16/22 shared proteins) than between the other NTM PPDs and PPD-B. Eight proteins were shared between PPD-B and each of the following NTM PPDs: PPD-M, PPD- while 11 proteins were shared between PPD-B and PPD-K and 10 between PPD-B and PPD-N. One hundred and ten proteins were identified as unique to PPD-B. Four hundred and seventy seven proteins were identified as unique to NTM. Some of the proteins detected only in NTM PPDs have been previously characterized as immunogenic in M. bovis/M. tuberculosis and were also in the NTM database entries. These include most notably: M. fortuitum MPB63 (identified in PPD-F and PPD-N), M. fortuitum Ag85C (identified in PPD-F, PPD-M, and PPD-N), M. fortuitum Ag85A (identified in PPD-F and PPD-N), M. kansasii Ag85B (identified in PPD-K), M. avium MPB64 (identified in PPD-A), M. avium Ag85A and Ag85B (both identified in PPD-A).

Discussion
In this study we first compared whole genome sequences of four NTM species to those of M. tuberculosis and M. bovis,
| Identified Proteins (16/561) | Accession Number | Molecular Weight | PPD-A | PPD-B | PPD-F | PPD-K | PPD-M | PPD-N | Function | References |
|-------------------------------|------------------|------------------|-------|-------|-------|-------|-------|-------|----------|------------|
| 10 kDa culture filtrate antigen EsxB*, Mycobacterium tuberculosis | A2VMP9_MYCTX | 11 kDa | + | + | + | + | Immunodominant antigen involved in secretion, complex x formation, immunogenicity and virulence | Brodin et al., 2005 |
| 10 kDa chaperonin GroES*, Mycobacterium tuberculosis | A2VPK5_MYCTX | 11 kDa | + | + | + | + | Immunodominant protein, ATP binding | Uniprot, Prasad et al., 2013 |
| 60 kDa chaperonin GroEL*, Mycobacterium tuberculosis | RAUN2_MYCTX | 57 kDa | + | + | + | + | Immunodominant protein | Uniprot, Prasad et al., 2013 |
| 50S ribosomal protein L7/L12, Mycobacterium paraclosatraceum | D5PC72_MYCO | 13 kDa | + | + | + | + | Involved in interaction with translation factors | Gudkov, 1997 |
| Elongation factor Tu, Mycobacterium fortuitum | K0K00_MYCF0 | 44 kDa | + | + | + | + | GTPase activity | Uniprot |
| DivIVA domain containing protein, M. avium subsp. paratuberculosis | F7P246_MYCPC | 28 kDa | + | + | + | + | Cell shape maintenance | He and de Buck, 2010 |
| Chaperone protein DnaK*, Mycobacterium tuberculosis | A2VF50_MYCTX | 67 kDa | + | + | + | + | Highly antigenic and act as co-repressor for heat shock protein transcription repressor (hspR) | Das Gupta et al., 2008 |
| Aconitase hydratase, Mycobacterium sylvaticum | V7KF43_MYCAV | 102 kDa | + | + | + | + | Unknown function | Uniprot |
| Elongation factor Tu, Mycobacterium avium | EFTU_MYCA1 | 44 kDa | + | + | + | + | GTPase activity | Uniprot |
| Secreted antigen 85-c fopC*, Mycobacterium bovis | A1KEV0_MYGBP | 37 kDa | + | + | Proteins of antigen 85 are responsible for high affinity of mycobacteria | Uniprot |
| Malate synthetase, Mycobacterium avium | MASZ_MYCA1 | 80 kDa | + | + | Enzyme of glyoxylate shunt | Kumar and Bhakuni, 2011 |
| Diacylglycerol acyltransferase/mycolyltransferase Ag85B, Mycobacterium smegmatis | A85B_MYCS2 | 35 kDa | + | + | Proteins of antigen 85 are responsible for high affinity of mycobacteria | Uniprot |
| Probable glyceraldehyde 3-phosphate dehydrogenase, Mycobacterium bovis BCG | A1KIM5_MYGBP | 36 kDa | + | + | Phases of glycolysis | Uniprot |
| Iron-dependent repressor IdeR, Mycobacterium avium | A0QIP3_MYCA1 | 25 kDa | + | + | DNA binding transcription factor | Uniprot |
| Putative ESAT-6 like protein 5, esxN*, Mycobacterium bovis (strain BCG/Pasteur 1173P2) | A1KJK3_MYGBP | 10 kDa | + | + | Reported to induce T-cell response | Deng et al., 2014 |
| DNA-directed RNA polymerase subunit beta*, rpoC, Mycobacterium tuberculosis | S5ES53_MYCTX | 147 kDa | + | + | Catalyzes the transcription of DNA into RNA | Uniprot |

(Continued)
TABLE 7 | Continued

| Identified Proteins (16/561) | Accession Number | Molecular Weight | Function | References |
|----------------------------|------------------|-----------------|----------|------------|
| 38S ribosomal protein, 54kDa | KOY712, MYCCO | 38 kDa | Catalyzes the transcription of DNA into RNA | Brodin et al., 2005 |
| ATPase subunit beta | B3_MYCA | 15 kDa | Response to stress | Uniprot |
| DNA-directed RNA polymerase | USW82015 | 38 kDa | | |
| Subunit alpha, 81kDa, Mycobacterium | WSW2_MYCA | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | AOXG_MYCA | 17 kDa | | |
| Mycobacterium avium subunit, 81kDa | NE_MYCA | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | USW82015 | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | MYCCO | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | AOXG_MYCA | 17 kDa | | |
| Mycobacterium avium subunit, 81kDa | NE_MYCA | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | USW82015 | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | MYCCO | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | AOXG_MYCA | 17 kDa | | |
| Mycobacterium avium subunit, 81kDa | NE_MYCA | 15 kDa | | |
| Mycobacterium avium subunit, 81kDa | USW82015 | 15 kDa | | |

mainly focusing on the presence in NTM of genes homologous to those encoding immunogenic proteins of the Esx and the PE/PPE families in M. bovis/M. tuberculosis. As a second approach, comparative proteomic analysis of PPDs derived from four NTM species to those of commercially available PPDs of M. bovis and M. avium—was conducted to identify the actual presence of homologous proteins. Three of the investigated NTM species i.e. M. nonchromogenicum, M. malmsburi sp. nov., and M. komaniai sp. nov. had previously been described as being among the most abundant NTM in cattle, buffalo and the environments of these animals in South Africa. M. fortuitum ATCC 6841 and its PPD was chosen because it is currently used in the modified BOVIGAM™ assay (Michel et al., 2011; Gcebe et al., 2013). M. malmsburi sp. nov and M. komaniai sp. nov. strains used in this study were both isolated from cattle nose mucous membranes and M. nonchromogenicum from soil in South Africa (Gcebe et al., 2013). M. malmsburi sp. nov. and M. komaniai sp. nov. are novel rapidly growing mycobacteria (RGM) (colonies taking <7 days to appear on solid medium). M. nonchromogenicum (a slow growing mycobacterial (SGM) species (colonies take >7 days to appear on solid growth medium) and M. fortuitum (a RGM) have also been isolated from cattle tissue in Great Britain, France, and Northern Ireland (Pollock and Anderson, 1997; Hughes et al., 2005; Vordermeier et al., 2007; Biet and Boschiroli, 2014) These two species were also reported to be among the most frequently isolated NTM in cattle in other parts of Africa (Diguimbaye-Djaib et al., 2006; Berg et al., 2009). Exposure of cattle to these NTM species may constitute a possible cause of cross-reactive immune response to PPD and may lead to misdiagnosis of BTB. This further highlighted the need to investigate these NTM in much more detail in terms of their genetic make-up as well as the proteomes of their PPDs in order to assess the presence of immunogenic proteins in NTM potentially causing cross-reactivity with MTB antigens. Comparative genomics revealed that overall there was very little similarity between the NTM genomes and those of M. bovis and M. tuberculosis. Despite these differences, orthologs of genes encoding for the most investigated antigens of the Esx family of MTBC including EsxA/ESAT-6, EsxB/CFP-10, EsxH/TR10.4, EsxG/TB 9.8 and EsxN; and the PE/PPE family proteins (PE35, PE5 and PPE68) were identified in the four newly sequenced NTM (Skjøt et al., 2000; Vordermeier et al., 2007; Zvi et al., 2008; Hoang et al., 2013; Amoudy et al., 2014; Mustafa, 2014). Only ESX-1, ESX-3, and ESX-4 loci were detected in the newly sequenced RGM, while genes of ESX-2 as well as esxN of ESX-5 were also identified in M. nonchromogenicum by whole genome analysis. The presence of esxA and esxB (situated within the ESX-1 locus) were confirmed in M. fortuitum by PCR and sequencing, while esxN (situated in ESX-5 locus) was confirmed in M. nonchromogenicum also by PCR and sequencing. The esxA expression could not be amplified in M. nonchromogenicum, M. malmsburi sp. nov., and M. komaniai sp. nov. using M. tuberculosis/M. bovis (result not shown) as well as M. smegmatis primers, probably due to huge nucleotide sequence differences between these species. The esxH gene (situated in ESX-3 locus) was also confirmed by PCR and sequencing in three of the NTM except M. nonchromogenicum using M. smegmatis derived
primers. This was possibly due to little homology between the *M. nonchromogenicum* esxH sequence and that of *M. smegmatis*, for these primers to anneal. The esxA, esxB, and esxH genes were however, identified in all the four NTM genomes by whole genome sequence analysis.

Studies have shown that the esx genes like esxA and esxB as well as esxG and esxH interact to form a 1:1 heterodimer which may be essential for their secretion (Uplekar et al., 2011). Likewise the pe and ppe genes are thought to form pairs (Gey van Pittius et al., 2006; Uplekar et al., 2011). In this study the esxA and esxB, esxG and esxH, pe35 and ppe68 genes were found to occur next to each other in the genomes of the four newly sequenced NTM. This confirmed that esxA, esxB, esxG, esxH, pe35, and ppe68 genes in these loci were true orthologs of *M. tuberculosis* and *M. bovis* esx, pe, and ppe genes. A closer inspection of the genes adjacent to the esx, pe, and ppe genes situated within the ESX-loci revealed the occurrence of orthologs of genes encoding for the ATP-dependent chaperones of the AAA family, membrane bound ATPases, transmembrane proteins chaperonin, and mycosin-subtilin like proteins. These genes form part of the type VII secretory system, therefore if expressed in these NTM, the Esx, PE, and PPE proteins may be secreted.

In NTM; ESAT-6, CFP-10, TB10.4, PE35, PE5, and PPE68 antigens of *M. bovis* and *M. tuberculosis* have been mainly shown to cross-react with homologs of other pathogenic Mycobacterium species that are phylogenetically closely related to MTBC including *M. kansasi*, *M. marinum*, *M. ulcerans*, as well as in *M. leprae* (Geluk et al., 2002, 2004; Skjøt et al., 2002; Vordermeier et al., 2007; Arnoudy et al., 2014; Mustafa, 2014). Identification of the genes of the Esx family and the PE/PPE family proteins in nonpathogenic NTM in this study add to earlier reports of their occurrence in nonpathogenic *M. smegmatis*, *Mycobacterium* sp. MGS, *Mycobacterium* sp. JLS, *Mycobacterium phlei*, *Mycobacterium thermoresistible*, *Mycobacterium gilvum* PYR-GCK, *Mycobacterium vaccae*, *Mycobacterium vanbaalenii*, and *Mycobacterium* sp. KMS (Gey van Pittius et al., 2006; Newton-Foot et al., 2016). Expression of the proteins of the Esx family, the PE/PPE family as well as the other antigens in nonpathogenic NTM have not been investigated. For that reason we did comparative investigation for the presence of these proteins in PPDs of three of the nonpathogenic NTM: *M. fortuitum* (PPD-F), *M. malmoeburii* sp. nov. (PPD-M) and *M. nonchromogenicum* (PPD-M) and in addition *M. kansasii* (PPD-K) compared to *M. bovis* (PPD-B) and *M. avium* (PPD-A) by Mass Spectrometric analysis. *Mycobacterium kansasii* derived PPD was included in this study, since its antigens have been investigated for their cross reactivity with *M. bovis* (Vordermeier et al., 2007). This bacterium has been isolated from cattle in Great Britain and the United States of America (Vordermeier et al., 2007; Thacker et al., 2013). Even though proteins are largely degraded during PPD preparation, we identified as many as 609 proteins in the combined six PPDs and we found 22 of these proteins to be shared between PPD-B and one or more of the NTM PPDs, including PPD-A. In the different NTM PPDs we identified homologs of some of previously characterized immunogenic MTBC proteins including CFP-10 (present in PPD-M), GroES (present in all PPDs except PPD-A), GroEL (present in all PPDs), EssN (PPD-K), ESAT-6 (PPD-K), and DnaK (all PPDs except PPD-M). The occurrence of these immunogenic antigens in nonpathogenic RGM PPDs could potentially lead to cross-reactive immune responses that may interfere with the diagnosis of bovine tuberculosis. In fact these proteins may be responsible for cross-reactivity seen between the *M. fortuitum* PPD and *M. bovis* PPD in South Africa (Michel et al., 2011).

In an attempt to predict cross-reactivity of the NTM homologs of ESAT-6, CFP-10, EsxH, and PPE68 more precisely, we compared amino acid sequences in these proteins with those of defined immunogenic epitopes in the proteins of *M. bovis* and *M. tuberculosis* (Skjøt et al., 2000; Vordermeier et al., 2003, 2007; Mustafa, 2014). This was done despite the fact that some of the genes encoding for these proteins were identified at DNA level, but not represented as proteins in the respective NTM PPD preparations. These proteins, with the exception of *M. malmoeburii* sp. nov.’s CFP-10 which was found in PPD-M, could have been degraded during PPD preparation, and possibly differential sensitivity for procedures of PPD production and trypsin treatment before Mass spectrometry or were not abundantly expressed in NTM to allow detection by LC-MS/MS (Borsuk et al., 2009). In all cases as illustrated in the amino acid alignments in Figures 6–10, homologies of >50% of the previously characterized immunogenic epitopes in *M. bovis* were seen in the different NTM homologs. For instance, six epitopes of *M. bovis* ESAT-6 recognized by bovine T-cells in the context of BoLA-DR and BoLA DQ, as described by Vordermeier et al. (2003, 2007) had amino acid sequence homologies of as high as 81.28% and as low as 52.9% to those of the *M. fortuitum* and *M. nonchromogenicum* orthologs. Likewise amino acid sequence homologies of >50% between the four *M. bovis* CFP-10 immunogenic epitopes and those of the NTM species was observed. The same was seen with PPE68 immunogenic epitope (VLATNFFGINTIPIALTEMDYFR).
described by Mustafa (2014) where 72% of the amino acid residues were identical between the NTM and M. tuberculosis. Still with the sequence homology of <100% of the NTM CFP-10, ESAT-6, PPE68, and EsxH epitopes and the M. bovis antigens, we could not unambiguously predict T-cell cross-reactivity. Several studies have reported contrasting results with regards to the relation of sequence identity and cross-reactivity of antigens. For instance, antigen cross-recognition has been observed with M. leprae ESAT-6 and CFP-10 on M. tuberculosis patients despite very low sequence identity (36 and 40% respectively at amino acid level) to the M. tuberculosis homologs (Geluk et al., 2002, 2004). Likewise, Hewinson et al. (2006) also demonstrated that sequence identity between epitopic regions from unrelated otherwise non-homologous mycobacterial antigens of >50% in the 16–20 mer regions indicated cross-reactivity in cattle (Hewinson et al., 2006). Contrary, other peptides that displayed similar degrees of sequence identity were not cross-reactive (Hewinson et al., 2006).

The NTM homologs in this study could potentially induce cross-reactivity against MTBC antigens and should be tested in animal experiments.

In conclusion, in this study we identified genes and proteins present in and expressed by members of the Mycobacterium tuberculosis complex known as markers for BTB diagnosis in selected nonpathogenic NTM genomes and proteomes. The identification of the genes encoding ESAT-6, CFP-10, PE35, PPE68, and PE5 in nonpathogenic NTM in this study adds to earlier findings of these genes in nonpathogenic like M. smegmatis and Mycobacterium sp. KMS. Expression of one of the most studied M. bovis immunogenic proteins, CFP-10 in M. malmesburi sp. nov. and that of other previously characterized MTBC immunogenic proteins such as GroES, DnaK, GroEL was shown by mass spectrometric analysis of the PPDs. We also identified immunogenic epitopes of ESAT-6, CFP-10, EsxH, and PPE68 through genomic analysis, in some or all of the four newly sequenced nonpathogenic NTM, also suggesting the potential of these proteins, to elicit cross reactive immune responses against MTBC antigens. The NTM homologs of the immunogenic proteins need to be further investigated for their cross-reactivity with the M. bovis antigens and consequently their interference in BTB assays.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We would like to thank Mr. Jonathan Featherston of the ARC Biotechnology platform as well as staff of CPGR for the assembly and annotation of the four NTM genomes. We also thank Dr. Zac McDonald for assisting with the proteomic analysis of the PPDs. We are grateful to WOTRO Science for global development; grant number W01.65.321.00, for funding.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00795

REFERENCES

Adékambi, T., Berger, P., Raoult, D., and Drancourt, M. (2006). rpoB gene sequence-based characterisation of emerging non-tuberculous mycobacteria with descriptions of Mycobacterium bolletii sp. nov., Mycobacterium phocaicum sp. nov. and Mycobacterium aubagense sp. nov. Int. J. Syst. Evol. Microbiol. 56, 133–143. doi: 10.1099/ijs.0.63969-0

Adékambi, T., and Drancourt, M. (2004). Dissection of the phylogenetic relationships among 19 rapidly growing Mycobacterium species by 165 rRNA, hsp65, sodA, recA and rpoB gene sequencing. Int. J. Syst. Evol. Microbiol. 54, 2095–2105. doi: 10.1099/ijs.0.63094-0

Akhter, Y., Ehebauer, M. T., Mukhopadhyay, S., and Hasnain, S. (2012). The PE/PPE multigene family codes for virulence factors and is possible source of mycobacterial antigenic variation: perhaps more. Biochimie 94, 110–116. doi: 10.1016/j.biochi.2011.09.026

Alikhan, N. F., Petty, N. K., Zakour, N. L. B., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:12. doi: 10.1186/1471-2164-12-402

Amoudy, H. A., Ebrahimi, B. H., and Mustafa, A. S. (2014). Immune responses against Mycobacterium tuberculosis-specific proteins PE35 and CFP10 in mice immunized with recombinant Mycobacterium vaccae. Saudi Med. J. 35, 350–359.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. doi: 10.1089/cmb.2012.0021
Gcebe, N., Gey van Pittius, N. C., Rutten, V., and Michel, A. (2013). Prevalence and distribution of non-tuberculous mycobacteria (NTM) in cattle, African buffaloes ( Syncerus caffer ) and their environments in South Africa. Trans. Emerg. Dis. 60, 74–84. doi: 10.1111/tbed.12133

Geluk, A., van Meijgaarden, K. E., Franken, K. L. M. C., Subronto, Y. W., Wieles, B., Arend, S. M., et al. (2002). Identification and characterisation ESAT-6 homologue of Mycobacterium leprae and T-cell cross reactivity with Mycobacterium tuberculosis. Infect. Immun. 70, 2544–2548. doi: 10.1128/IAI.70.3.2544-2548.2002

Gey van Pittius, N. C., Sampson, S., Lee, H., Kim, Y., van Helden, P. D., and Warren, R. M. (2006). Evolution and expression of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol. Biol. 6:95. doi: 10.1186/1471-2148-6-95

Gudkov, A. T. (1997). The L7/L12 ribosomal domain of the ribosome: structural and functional studies. FEBS Lett. 407, 253–256. doi: 10.1016/S0014-5793(97)00361-X

Gurevich, A., Saveliev, V., Vyaihi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 8. doi: 10.1093/bioinformatics/btt086

He, Z., and de Buck, J. (2010). Localisation of proteins in the cell wall of Mycobacterium avium subsp paratuberculosis K10 by proteomic analysis. Proteome Sci. 8:21. doi: 10.1186/1477-5956-8-21

Hewinson, R. G., Vordermeier, H. M., Smith, N. H., and Gordon, S. V. (2006). The potential of ESAT-6 antigen secreted by Mycobacterium tuberculosis for diagnostic purposes and its role in the protective immune response. Vet. Microbiol. 70, 254–263. doi: 10.1016/j.vetmic.2005.11.050

Hoang, T., Aagaard, C., Dietrich, J., Cassidy, J. P., Dolganov, G., Schooink, G. K., et al. (2013). Esat-6 (esxA) and TB10.4 (esxII) based vaccines for pre- and post- exposure tuberculosis vaccination. PLoS ONE 8:e80579. doi: 10.1371/journal.pone.0080579

Hughes, M. S., Ball, M. W., Mccarroll, J., Erskine, M., Taylor, J. M., Pollock, J. L., et al. (2005). Molecular analysis of mycobacteria other than the M. tuberculosis complex isolated from Northern Ireland cattle. Vet. Microbiol. 108, 101–112. doi: 10.1016/j.vetmic.2005.03.001

Ize, B., and Palmer, T. (2006). Microbiology. Mycobacteria’s export strategy. Science 313, 1583–1584. doi: 10.1126/science.1132537

Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383–5392. doi: 10.1021/ac025747h

Kumar, R., and Bhakuni, V. (2011). Comparative analysis of malate synthase G from Mycobacterium tuberculosis and E. coli: role of ionic interaction in modulation of structural and functional properties. Int. J. Biol. Macromol. 49, 917–922. doi: 10.1016/j.ijbiomac.2011.08.008

Landi, S. (1963). Preparation, purification and stability of tuberculin. Appl. Microbiol. 11, 408–412.

Li, H., and Dubin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25, 14. doi: 10.1093/bioinformatics/btp324

Maharias, G. G., Sabo, P. J., Hickey, M. J., Sing, D. C., and Stiver, C. H. (1996). Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J. Bacteriol. 178, 1274–1282.

Marongui, L., Donini, M., Toffali, L., Zenaro, E., and Dusi, S. (2013). ESAT-6 and HspX improve effectiveness of BCG to induce human dendritic cells dependant Th1 and NK cells activation. PLoS ONE 8:e75864. doi: 10.1371/journal.pone.0075864

Michel, A. L., Cooper, D., Joooste, J., de Klerk, L. M., and Jolles, A. (2011). Approaches towards optimizing the gamma interferon assay for diagnosing Mycobacterium bovis infection in African buffalo ( Syncerus caffer ). Prev. Vet. Med. 98, 142–151. doi: 10.1016/j.prevetmed.2010.10.016

Mustafa, A. S. (2014). Characterization of a cross-reactive, immunodominant and HLA promiscuous epitope of Mycobacterium tuberculosis-specific major antigenic protein PPE68. PLoS ONE 9:e103679. doi: 10.1371/journal.pone.0103679

Nesvizhskii, A. L., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646–4658. doi: 10.1021/ac0341261

Newton-Foot, M., Warren, R. M., Sampson, S. L., van Heelden, P. D., and Gey van Pittius, N. C. (2016). The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems. BMC Evol. Biol. 16:16. doi: 10.1186/s12862-016-0126-2

Pollock, J. M., and Anderson, P. (1997). The potential of ESAT-6 antigen secreted by virulent mycobacteria for specific diagnosis of tuberculosis. J. Infect. Dis. 175, 1251–1254. doi: 10.1086/593686

Prasad, T. K., Verma, R., Kumar, S., Nirujiogi, R. S., Sathe, G. J., Madugundu, A. K., et al. (2013). Proteomic analysis of purified protein derivative of Mycobacterium tuberculosis. Clin. Proteomics 10:8. doi: 10.1186/1559-0298-7-50

Pym, A. S., Brolin, P., Majlesi, L., Brosch, R., Demangel, C., Williams, A., et al. (2003). Reombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat. Med. 9, 533–539. doi: 10.1038/nm859

Renshaw, P. S., Lightbody, K. L., Veverka, V., Musket, F. W., Kelly, G., Frenkel, T. A., et al. (2005). Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. EMBO J. 24, 2491–2498. doi: 10.1038/sj.emboj.7600732

Schiller, I., Oesch, B., Vordermeier, H. M., Palmer, M. V., Harris, B. N., Orloski, K. A., Bovine tuberculosis, et al. (2010). A review of current and emerging diagnostic techniques in view of their relevance for disease control and eradication. Transbound. Emerg. Dis. 57, 205–220.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 14. doi: 10.1093/bioinformatics/btu153

Skjot, R. L., Oettinger, T., Rosenkrands, I., Ravn, P., Brock, L., Jacobsen, S., et al. (2000). Comparative evaluation of low molecular mass proteins from Mycobacterium tuberculosis identifies members of the ESAT-6 family as immunodominant T-cell antigens. Infect. Immun. 68, 214–220. doi: 10.1128/IAI.68.1.214-220.2000

Skjot, R. L. V., Brock, L., Arend, S. M., Munk, M. E., Theisen, M., Ottenhoff, T. H. M., et al. (2002). Epitope mapping of the immunodominant antigen TB10.4 and the two homologous proteins TB10.3 and TB12.9, which constitute a subfamily of the esat-6 gene family. Infect. Immun. 70, 5446–5453. doi: 10.1128/IAI.70.1.5446-5453.2002

Sorensen, A. L., Nagai, S., Houen, G., Andersen, P., and Andersen, A. B. (1995). Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. Infect. Immun. 63, 1710–1717.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. doi: 10.1093/molbev/mst197

Telfen, A., Marchesi, F., Bala, M., Bally, F., Boggert, E. C., and Bodmer, T. (1993). Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J. Clin. Microbiol. 31, 175–178.
Thacker, T. C., Robbe-Auterman, S., Haris, B., Palmer, M. V., and Water, W. R. (2013). Isolation of mycobacteria from clinical samples collected in the United States from 2004-2011. BMC Vet. Res. 9:100. doi: 10.1186/1746-6148-9-100

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Uplekar, S., Heym, B., Friocourt, V., Rougemont, J., and Cole, S. T. (2011). Comparative genomics of the esx genes from clinical isolates of Mycobacterium tuberculosis provides evidence of gene conversion and epitope variation. Infect. Immun. 79, 4042–4049. doi: 10.1128/IAI.05344-11

van Ingen, J., De Zwaan, R., Dekhuijzen, R., Boeree, M., and van Soolingen, D. (2009). Region of difference 1 in non-tuberculous Mycobacterium species adds a phylogenetic and taxonomic character. J. Bacteriol. 191, 5865–5867. doi: 10.1128/JB.00683-09

Vordermeier, H. M., Brown, J., Cockle, P. I., Franken, W. P. J., Arend, S. M., Ottenhoff, T. H. M., et al. (2007). Assessment of cross-reactivity between Mycobacterium bovis and M. kansasii ESAT-6 and CFP 10 at the T-cell epitope level. Clin. Vac. Immunol. 14, 1203–1209. doi: 10.1128/CVI.00116-07

Vordermeier, H. M., Cockle, P. C., Whelan, A., Rhodes, S., and Hewinson, R. G. (2000). Towards the development of diagnostic assays to discriminate between Mycobacterium bovis infection and BCG vaccination in cattle. Clin. Infect. Dis. 30, S291–S298. doi: 10.1086/313877

Vordermeier, H. M., Huygen, K., Singh, M., Hewinson, R. G., and Xing, Z. (2006). Immune response induced in cattle by vaccination with recombinant adenovirus expressing mycobacterial antigen 85A and Mycobacterium bovis BCG. Infect. Immun. 74, 1416–1418. doi: 10.1128/IAI.74.2.1416-14.2006

Vordermeier, M., Whelan, A. O., and Hewinson, R. G. (2003). Recognition of mycobacterial epitopes by T cells across mammalian species and use of a program that predicts Human HLA-DR binding peptides to predict bovine epitopes. Infect. Immun. 71, 1980–1987. doi: 10.1128/IAI.71.4.1980-1987.2003

Waters, W. R., Whelan, A. O., Lyashchenko, K. P., Greenwald, R., Palmer, M. V., Harris, B. N., et al. (2010). Immune responses in cattle inoculated with Mycobacterium bovis, Mycobacterium tuberculosis, or Mycobacterium kansasii. Clin. Vac. Immunol. 17, 247–252. doi: 10.1128/CVI.00442-09

Zvi, A., Ariel, N., Fulkerson, J., Sadoff, J. C., and Shaffer, A. (2008). Whole genome identification of Mycobacterium tuberculosis vaccine candidates by comprehensive data mining and bioinformatics analyses BMC Med Genomics 1:18. doi: 10.1186/1755-8794-1-18

Copyright © 2016 Gcebe, Michel, Gey van Pittius and Rutten. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.