High Throughput Discovery of Novel Diagnostic Antigens for *Mycobacterium bovis* Using a Whole Genome Approach

Nadia Assal

Thesis submitted to the University of Ottawa

In partial fulfillment of the requirements for the

Doctorate of Philosophy degree in Microbiology and Immunology

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

University of Ottawa

Abstract

Bovine tuberculosis (BTB) is a chronic infectious disease caused by the bacterium *Mycobacterium bovis.* It infects animals and can be a source of zoonosis. In the last five years, Canada has faced two outbreaks of bovine tuberculosis in the years 2018 and 2016. BTB is mainly diagnosed using the Tuberculin skin test, a test that detects the cellular immune response to administered purified protein derivative (PPD). A drawback of this test is the high level of false-positive test results caused by the immune response to PPD proteins that are conserved in non-tuberculous mycobacteria. Current serodiagnostic tests can detect the disease especially in the advanced state with, low sensitivity ranging between 9-45%. It is hypothesized that the profiling of the humoral immune responses to selected proteins will lead to the discovery of novel immunogenic protein antigens for improved BTB diagnostic tests. Bioinformatic tools used in this research for the prediction of extracellular or outermembrane proteins from *M. bovis* genome sequences identified 96 protein candidates. Also, a proteomic study conducted to identify proteins secreted by *M. bovis*, together with the review of previous proteomic studies of the PPD, led to the identification of an additional 92 protein candidates. A high throughput system was developed to efficiently express and analyze these antigens involving the PCR amplification, i*n vivo* cloning, and *in vitro* expression of open reading frames (ORFs) coding for selected protein candidates. This was followed by a high throughput recombinant protein purification and a microarray analysis of these purified recombinant proteins with sera from *M. bovis*-infected animals. The system successfully amplified, cloned, and expressed 159 recombinant proteins. From the microarray screening, 13 antigens exhibited immunological reactions and were able to differentiate the sera of the infected animals from the controls. Out of those 13 antigens, 4 novel antigens Mb2740c, Mb0598c, Mb3469c, and Mb3453c, in addition to two well-known antigens MPB70 and MPB83, gave the highest reactivity; they were selected for further evaluation for diagnostic applications using ELISA and dot blot assays. Two antigens Mb3469c and Mb3453c had a significant ability to differentiate between infected and control cattle as tested in ELISA and dot blot assays, respectively and demonstrated by ROC curve analysis. These two novel antigens could be added to the panel of serodiagnostic antigens for improving BTB serodiagnosis and could be beneficial in the detection of outbreaks caused by certain *M. bovis* strains.

Acknowledgment

I would like to express my gratitude and appreciation to all who helped and supported me with this work.

Thank you to my supervisor, Dr. Min Lin for his guidance throughout my Ph.D. journey. He accepted me when I was a newcomer to Canada and continued to teach, support, and approved my new ideas. I would also like to thank my co-supervisor Dr. Jim Sun for his valuable advice and support to finish the thesis.

I express my sincere gratitude to the members of my thesis advisory committee, Dr. Alain Stintzi, Dr. Alex Wong, and Dr. Mehdi Arbabi Ghahroudi for their continuous support, guidance, and constructive advice. Thank you for encouraging me and make me feel I am doing a good job.

Thank you to all current and former members of Lin's lab, with special thanks to Hanhong Dan for all the help, advice, support, and encouragement. Thank you to current and former members of the Canadian Food Inspection Agency Dr. Liz Rohonczy, Dr. Olga Andrievskaia, Dr. Marc-Olivier Duceppe, Dr. Mingsong Kang, Bryan Rennie, Dr. Hongsheng Huang, Beverly Todd, and Dr. Dele Ogunremi for attending my rehearsals and providing me with advice and invaluable feedback.

Thanks to the collaborators at Heath Canada and NRC, to Lisa Walrond for the help and time, and to Dr. Qing Liu and her lab members for the advice and the technical support.

Thank you to my parents in Egypt, Fouad Assal and khairyaSoffar, who continue to support me through the different chapters of my life. I can hear your prayers and I know I am blessed to have you in my life. Thanks to my husband, Mohamed, for believing in me and continuously encouraging me. Thank you to my sister, Hala, and my brother, Walid, for their love and support. Thanks to my children, Malak and Oday, who shared the journey with me. Your hugs and laughs brighten my days.

Table of Contents

List of Tables

Supplementary Tables

List of Figures

Supplementary Figures

List of Abbreviations

All abbreviations have been defined at first mention in the text. For quick reference, a non-comprehensive list of abbreviations is presented here:

ADC	Albumin-dextrose-catalase
BCG	Bacille Calmette-Guérin
BSA	Bovine serum albumin
BTB	Bovine Tuberculosis
CFIA	Canadian Food Inspection Agency
CMI	Cell-mediated immunity
DCs	Dendritic cells
DTH	Delayed type hypersensitivity
EC	Extracellular
ECMV	Extracellular support vector machine
ELISA	Enzyme-linked immunosorbent assay
FFPE	Formalin-fixed, paraffin-embedded tissues
GFP	Green fluorescent protein
IFN-γ	Interferon-gamma
IVTT	*In vitro* transcription and translation
kDa	kiloDalton
LJ	Lowenstein–Jensen
MFI	Mean fluorescent intensity
MTB	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
NVSL	National Veterinary Services Laboratories
NEB	New England Biolabs
OD	Optical density
OIE	Office International des Epizooties
OLF	Ottawa laboratory Fallowfield

OM	Outer membrane
ORF	Open reading frame
PCR	Polymerase chain reaction
PGRS	Polymorphic GC-rich sequence
PMT	Photomultiplier tube
PPD	Purified protein derivative
PPDB	Purified protein derivative of *M. bovis*
ROC	Receiver operating characteristic
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
SICCT	Single intradermal comparative cervical tuberculin
SIT	Skin intradermal test
TST	Tuberculin skin test
SNP	Single nucleotide polymorphism
USDA	US Department of Agriculture
WB	Western blot
WGS	Whole-genome sequencing

Chapter 1. Introduction

1.1 Research Rationale

Tuberculosis (TB) remains a major cause of ill health and among the top 10 causes of death worldwide (1). Bovine tuberculosis (BTB) is a chronic infection that mainly infects cattle but can be transmitted to humans (zoonotic transmission). Zoonosis mainly occurs through the respiratory pathways after close or direct contact with infected cattle or through the consumption of contaminated animal products such as undercooked meat or unpasteurized milk (2, 3). Currently, after the introduction of regular milk pasteurization, BTB is well controlled and cases of zoonotic TB are rarely seen especially in developed countries. However, in Africa BTB is still considered prevalent in most countries due to the lack of effective disease control policies. It is estimated that there are about 70 000 zoonotic tuberculosis cases occurring annually in Africa (4). Moreover, substantial evidence suggests that the health and economic burden of *M. bovis* infection might be undervalued in humans (5, 6).

The diagnosis of BTB is challenging due to the limitations of the current diagnostic tests in sensitivity and specificity. The accurate diagnosis seems to require the results of multiple tests where a single test fails to reliably identify all infected animals. Conventional diagnosis of TB depends on the use of the “Tuberculin skin test” (TST). The test is based on the intradermal skin injection of the purified protein derivative (PPD) of *M. bovis* (PPDB) followed by the examination of the size of induration 72 hours later (7). The relative sensitivity of the TST is usually reported around 60%; however, reports also show a between-herd heterogeneity with lower sensitivities between 40-52% in some herds (8). The

second diagnostic test is "*In vitro* interferon-gamma (IFN-γ) assay". The test measures IFN-γ released from memory T-cells after stimulation with PPDB (9). Both tests depend on the evaluation of the cell-mediated immune response after exposure to the PPDB, also known as "tuberculin". This reagent has raised a lot of controversies about its specificity, standardization, and quality in diagnostics (7, 10). The main concern with using PPDB in the current BTB diagnostics is its association with high levels of false-positive results. These false-positive results may be attributed to vaccination with *M. bovis* Bacille Calmett-Guérin (BCG) or exposure to non-tuberculous mycobacteria. The exposures provoke an immune response that may be similar to BTB infection, given that 60-80% of the PPDB proteomic antigen content is shared by environmental non-tuberculous mycobacteria (11–16).

Serodiagnostic tests offer the advantages of being simple, with short turnaround time and requiring only one-time handling. These tests detect advanced stages of the disease in which the cell-mediated immune response is lacking. At these stages the animals tend to be infectious, reflecting a more serious condition (17). The most widely used serological test is enzyme-linked immunosorbent assay (ELISA) developed by IDEXX laboratories (Westbrook, ME), which detects antibodies to two *M. bovis* proteins, MPB70 and MPB83. The assay has a high specificity ~98%, nevertheless, its sensitivity varies by geographic region, with sensitivities of 45%, and 9% in bovine serum samples from the United States ($n = 146$), and Mexico ($n = 128$), respectively (18).

The control of BTB primarily depends on accurate diagnosis. In this study, the improvement of the serological diagnosis of BTB was targeted by aiming for the discovery of new specific antigens. These antigens when applied appropriately will form the basis for the development of improved diagnostic tests to advance disease control.

1.2 Hypothesis

The availability of complete genome sequences of *M. bovis* isolates makes it possible to predict surface or extracellular proteins encoded by *M. bovis* genomes using bioinformatics tools. Also, the proteomic analysis of the mycobacterial culture filtrate allowed the identification of additional, secreted, extracellular protein candidates produced by *M. bovis.* Profiling the humoral immune responses to these selected proteins will lead to the discovery of novel immunogenic protein antigens for the development of improved diagnostic tests.

1.3 Research Objective

The ultimate goal of this study is to discover novel antigens of *M. bovis* and use them to develop laboratory tools that enhance the diagnosis of *M. bovis* infection in animals.

Specific aims

1. Identifying a repertoire of the open reading frames (ORFs) from the genome sequence of *M. bovis* coding for the surface or secreted proteins using bioinformatics analysis as well as additional ORFs revealed by proteomic studies.

2. Profiling the humoral immune response to *M. bovis* infection by protein microarray technology with recombinant proteins expressed from the identified ORFs to discover novel diagnostic antigens.

3. Evaluating the identified antigens in an appropriate immunological assay format for the detection and identification of the *M. bovis* infected animals.

Chapter 2. Literature Review

2.1 Tuberculosis and Bovine Tuberculosis

2.1.1 Epidemiology

Tuberculosis (TB) constitutes a major global health problem that causes millions to fall ill each year. Since 2007, the World Health Organization (WHO) has stated that TB is the leading cause of death from a single infectious disease outplacing HIV/AIDS. Globally in 2018, there were 1.2 million TB deaths and an additional 0.25 million deaths among HIV-positive people. New TB cases were reported to be 10.0 million, 57% of which were adult-men, 32% were adult women, and children accounted for 11% (1)

In 2018, there were an estimated 143 000 new human cases of zoonotic TB according to the WHO, and mortality estimates were reported to be 12 300 globally. Of these figures, the African region carried the heaviest burden with 69 800 new cases (48.8%) and 9100 deaths (73.9%) in 2018. It is worth mentioning that zoonotic BTB is prevalent in most countries in Africa due to a lack of effective disease control policies including regular milk pasteurization and slaughterhouse meat inspection (1). This situation is exacerbated by the presence of multiple additional risk factors such as human behavior and the high prevalence of HIV infections (5, 19, 20). Nevertheless, data from high-income countries with a low burden of BTB might have led to insufficient awareness and underestimation of the health impacts of zoonotic TB (1, 5, 21).

2.1.2 Transmission and Manifestations of BTB

Bovine tuberculosis (BTB) is a chronic infectious disease affecting animals. Cattle are the main reservoir of *M. bovis* that can infect all mammals including wildlife (4, 22). Transmission of the infection to humans (zoonotic transmission) is well described and occurs primarily through close or direct contact with infected cattle. The transmission of infection occurs through the inhalation of infectious aerosols or the ingestion of contaminated animal products such as undercooked meat, unpasteurized milk, or milk products (19, 20).

In countries where eradication plans are implemented, BTB-infected cattle are identified early before the appearance of any clinical symptoms. The infection may remain dormant for months or years (latent TB) and reactivate under stress or in old age. This makes it difficult to diagnose the disease clinically. In progressive and late stages, symptoms of TB disease caused by *M. bovis* are similar to those of TB caused by *M. tuberculosis*; the symptoms can include low-grade fever, night sweats, weakness, loss of appetite, and weight loss. If the disease affects primarily the lungs, especially with the aerosol transmission, the disease can be associated with a productive cough and dyspnea. In cases of food-borne disease or due to the ingestion of sputum, the disease affects the gastrointestinal system causing abdominal pain, diarrhea, or constipation. The disease is often associated with the enlargement of lymph nodes in the retropharyngeal, pulmonary, or abdominal regions. If untreated, the disease can lead to death. *M. bovis* was reported to cause more extrapulmonary than pulmonary manifestations. This may be related to the important role of unpasteurized milk in transmitting the disease especially in the African countries where control measures are lacking (3, 23–26).

2.1.3 BTB Impact and Control

BTB is a disease listed in the World Organization for Animal Health (OIE) Terrestrial Animal Health Code and is an officially reportable disease under *Canada's Health of Animals Act*. In the last ten years, Canada has suffered three outbreaks of BTB affecting beef cattle (Table 2-1) (27).

Table 2-1. Outbreaks of BTB in Canada (2011-2020).

Year	Date confirmed	Location	Infected animal
2018	November 9	British Columbia	Beef Cattle
2016	September 16	Alberta	Beef Cattle
2011	May 16	British Columbia	Beef Cattle

BTB causes significant global economic losses estimated annually at a total of about US$ 3 billion. Those are due to the reduced productivity of affected animals and their slaughter causing trading restrictions in addition to expenses related to regular surveillance, regular testing, movement control, and procedures related to TB control (28). In Argentina, US$63 million were lost annually loss due to bovine TB (20). In the USA, the annual federal appropriation for the bovine tuberculosis program has settled at approximately US$15 million since 2005. (5).

For many years, the control of BTB has focused on the control of BTB in domestic cattle herds using the standard control measure, which is "test and slaughter". Effective disease detection is mandatory to implement this strategy. Two elements are essential for proper detection: passive surveillance and active surveillance. Passive surveillance depends

on voluntary, nontargeted sample submissions for detecting the disease. Examples of passive surveillance include field cases that are sent to a provincial veterinary diagnostic laboratory, submission of animals presenting disease signs or found dead by hunters for post-mortem analysis, and submission of carcasses, organs or samples of all wild mammals presenting lesions suggestive of tuberculosis (29).

Active surveillance involves conducting surveys in order to quantify the level of disease and follow its progression over time. For example, regular TB testing of livestock herds at high risk are carried out because of their geographic proximity to diseased wildlife (27, 29). Meat inspection represents a long-standing form of active surveillance that serves both food safety and animal health. Slaughterhouse inspection is an effective surveillance tool for BTB which produces slowly progressive but evident lesions. Detecting infected animals prevents unsafe meat from entering the food chain and allows tracing of infection. Pasteurization of milk of infected animals has prevented the spread of the disease in humans. Treatment of infected animals is rarely attempted due to being costly and time-consuming (2, 29, 30). Notably, BTB control strategies have proven to be effective in the Western countries where multiple European countries managed to eliminate BTB maintaining an Officially bovine tuberculosis-free (OTF) status regarding freedom from bovine tuberculosis, in cattle (31).

Even though the “test and slaughter” approach may be convenient for domestic animals, the complete control of the disease using this strategy has been hampered by the existence of wildlife reservoirs of *M. bovis* especially in the UK and North America (32). In the province of Alberta, Canada, *M. bovis* is endemic where the prevalence of infection in 1925–2013 was 42.0 – 49.0% mainly in Wood bison in the Buffalo National Park (BNP)

(32). Similarly, a herd of elk in Manitoba was considered responsible for BTB outbreaks in 11 cattle herds surrounding Riding Mountain National Park (RMNP) between the years 1991 and 2003 (29, 33). The Canadian Food Inspection Agency (CFIA) set up the Riding Mountain Eradication Area (RMEA) as a special surveillance zone. In the RMEA, the CFIA has worked closely with livestock producers, the livestock industry, and federal and provincial partners to conduct periodic TB testing of livestock herds. This activity allows for early detection and guards against the potential further spread of the disease and thus protects the health status of Canada's domestic livestock herds (34).

Vaccination in TB is based on the *M. bovis* BCG vaccine and is practiced in human medicine; however, it is not popular in animals. Vaccinating cattle is not currently implemented within any international control program. It is even illegal within the European Union (E.U). This is due to the variability in the efficacy of existing vaccines and the interference with the tuberculin skin test, thus creating more barriers against the elimination of the disease (2, 30, 35).

2.2 *Mycobacterium bovis*

2.2.1 Taxonomy

M. bovis is the agent responsible for BTB. It is a facultative intracellular bacterial pathogen. *M. bovis* is a member of the *Mycobacterium tuberculosis* complex (MTBC). The complex constitutes a notably genetically homogeneous group, characterized by very high similarity at the nucleotide level, but the members differ widely in terms of their host tropisms, phenotypes, and pathogenicity. The MTBC includes *M. tuberculosis, M.*

africanum, and *M. bovis* which are well-known human pathogens (36). *M. bovis* is a microaerophilic bacterium, different from the aerobic *M. tuberculosis* (37).

The *Mycobacterium* genus belongs to the family Mycobacteriaceae within the order Actinomycetales. *Mycobacterium* is closely related to other genera (i.e., *Nocardia*, and *Corynebacterium*). They are, however, distinguished by their morphology, motility, and staining characters (36). *M. bovis* displays the broadest spectrum of host infection, affecting humans, domestic or wild bovines, and goats (36).

2.2.2 Characteristics and Cell Envelope

M. bovis bacilli appear under the light microscope as a straight or slightly curved rod of approximately 3–5 μm in length and 0.2–0.6 μm in width. It is well known for being an acid-fast bacterium that has unique characteristics in its cell wall, where it exhibits the combined properties of both Gram-positive and Gram-negative bacteria. The cell wall consists of multilayered peptidoglycan and covalently linked arabinogalactan. It can be stained weakly by the Gram stain leading to the classification of mycobacteria as Gram-positive bacteria. On the other hand, the most important and distinctive feature of the mycobacterial cell wall is the substantial amount of lipid component (up to 60% of the total mass). Recently, by the use of cryo-electron tomography and electron microscopy of ultrathin cryosections, the native three-dimensional organization of the cell envelope *of M. smegmatis* and *M. bovis* BCG was revealed as the bilayer structure of the outer membrane, supporting the model that the cell wall lipids are organized in an outer membrane, similar to the Gram-negative bacteria (38, 39). The organization and structural components of a mycobacterial cell envelope are illustrated in **Figure 2-1**.

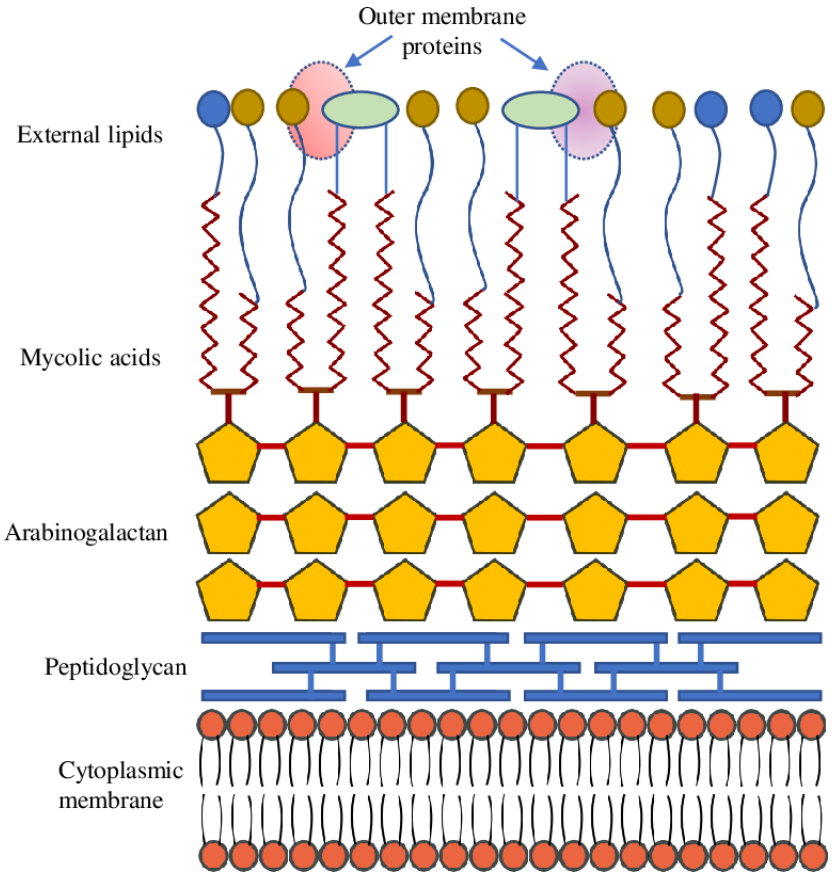


Figure 2-1. Schematic diagram of the *Mycobacterium* cell envelope.
The proposed arrangement is the cytoplasmic membrane and cell wall components (peptidoglycan, arabinogalactan, mycolic acids, and external lipids). The peptidoglycan is multilayered similar to Gram-positive bacteria. Also, mycolic acids and external lipids are arranged in an outer layer similar to the outer membrane of the Gram-negative bacteria.

Mycobacterium, being a member of the Actinomycetes, are known to be high G+C Gram-positive bacteria because their DNA has more than 55 to 60% G+C content. Based on the data in Genbank (at www.ncbi.nlm.nih.gov), *M. tuberculosis* and *M. bovis* have 65.6% G+C content, similar to that of *Pseudomonas aeruginosa* (66.6%), suggesting that *M.*

tuberculosis and *M. bovis* could be more closely related to some Gram-negative bacteria with respect to the G+C content (39).

2.2.3 Pathogenesis and Immune Response to infection with *M. bovis*

Aerosol exposure is the most common mode of transmission of *M. bovis* infection allowing the bacteria to be carried through the small air passages to the pulmonary alveoli. The bacteria are uptaken by alveolar macrophages and other phagocytes, including dendritic cells (DCs), where some bacteria are killed; however, others may survive and replicate intracellularly (40, 41). The phagocytes transport the bacteria through the circulation to the draining lymph nodes and even protect them from bactericidal components in serum. Inside the macrophage phagosome, mycobacteria can cause phagosome maturation arrest. They can block phagolysosome fusion and detoxify reactive oxygen and nitrogen radicals (40, 41). Studies on *M. tuberculosis* revealed that the bacteria block the acidification of .phagosome; phagosome acidification is an essential event that promotes the degradation of invading particles into constituents for antigen presentation with the consequence of stimulating the adaptive immune responses (42, 43). PtpA, a secreted protein tyrosine phosphatase, is found to be essential for Mtb pathogenicity. It inhibits phagosome acidification and plays a role in the arrest of phagosome maturation causing inhibition of phagosome-lysosome fusion (43, 44).

In the draining lymph nodes, the adaptive immune response is stimulated, in the form of a cell-mediated immune response (CMI) which usually requires a period of 10-14 days. T-lymphocytes play an important role in CMI where they recognize "processed" mycobacterial antigens in association with major histocompatibility complex (MHC) products on antigen-

presenting cells such as the dendritic cells. This interaction leads to the production of T cell populations including effector memory T cells (T_{EM}) and central memory T cells (T_{CM}).

The effector memory T cells (T_{EM}) appear in the blood for a brief period and are usually not detected after the bacteria are cleared. Consequently, the presence of live, non-replicating bacteria is usually accompanied by persistent effector memory T cells. T_{EM} cells home to peripheral tissue and can differentiate directly into effector cells. The measurement of interferon-γ (IFN-γ) production in specific tests after the stimulation of peripheral blood mononuclear cells with select mycobacterial antigens is the way used to detect T_{EM} cells. On the other hand, the central memory T cells (T_{CM}). remain for life after TB infection even in the absence of persistent antigen. These cells home to secondary lymphoid organs, nevertheless may not provide protection in infected subjects altogether (45, 46).

The immune interaction between the antigen-presenting cells and T-cells results in the secretion of cytokines and chemokines. This keeps the macrophages in an activated state and ensuring the recruitment of other immune cells to the site of infection where virulent mycobacteria or their products exist. The cellular immune response may lead to cellular hypersensitivity, which may contribute to cell death and tissue destruction (caseous necrosis) leading to the sequelae of liquefaction and cavity formation. The cavities may rupture contributing to the aerosol spread of the bacilli. Lymph nodes are the most pathologically affected tissues especially bronchial and mediastinal lymph nodes in respiratory infection. In food-borne infection, the lymph tissues associated with the intestinal tract will be commonly affected (45).

The hallmark of mycobacterial infection is the formation of the granuloma, which develops due to the cellular response where a large number of phagocytic mononuclear cells accumulate in an attempt to control the infection. Granulomas are generally regarded as being structured clusters containing mycobacteria-infected macrophages in the center surrounded by different types of immune cells, particularly macrophages and T_{CM} lymphocytes, B- lymphocytes. They may remain localized by fibrous tissue capsules in an attempt to localize the disease and destroy the bacteria (47). Granulomas affect tissues including the lungs and draining lymph nodes and can finally develop into macroscopic lesions known as tubercles (19, 48, 49). Analysis of asymptomatic cattle naturally infected with *M. bovis* has revealed these animals have numerous granulomatous regions containing epithelioid macrophages, giant cells, lymphocytes, and neutrophils (40, 41, 46). Thus, CMI may benefit the host by initiating processes that may destroy or inhibit mycobacteria (50).

Similarly, the antibody-mediated immune responses are induced following mycobacterial infection. The role of the humoral immune response in the protection against mycobacterial infection in mouse models has been controversial (51). In murine TB, it has been demonstrated that B cells can regulate the level of granulomatous reaction, cytokine production, and the T cell response. Antibodies can also contribute to the host response to pathogens through various means including antigen-specific neutralization and complement activation. Additionally, specific monoclonal antibodies against mycobacterial components were shown to be protective in maintaining a lower bacterial load in infected mice (51). On the other hand, studies showed that antibody production may also contribute to pathogenesis through the formation of antigen-antibody complexes which may subsequently damage tissues by activating complement or may inhibit phagocytosis by macrophages (50, 51). IgG1

in advanced tuberculosis was shown to increase proinflammatory cytokines such as TNFα produced by monocytes and contribute to additional inflammation (52). A representation of the immune response to BTB infection is illustrated in **Figure 2-2**.

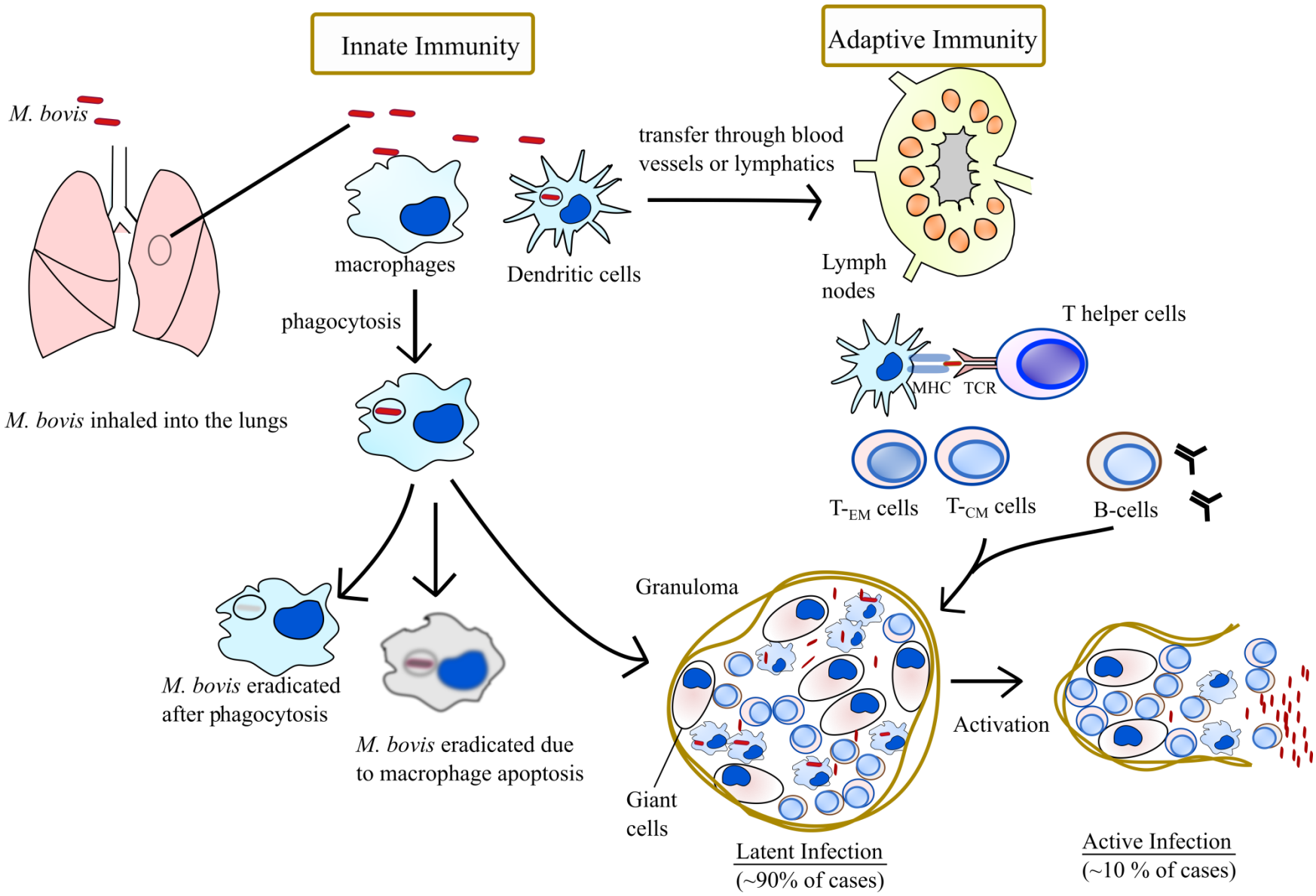

Innate Immunity
Adaptive Immunity
M. bovis
macrophages
Dendritic cells
transfer through blood vessels or lymphatics
Lymph nodes
phagocytosis
T helper cells
MHC
TCR
M. bovis inhaled into the lungs
T-EM cells
T-CM cells
B-cells
Granuloma
M. bovis eradicated after phagocytosis
M. bovis eradicated due to macrophage apoptosis
Activation
Giant cells
Latent Infection
(~90% of cases)
Active Infection
(~10 % of cases)

Figure 2-2. Immune response to *M. bovis* bacilli following inhalation.

The figure shows the activation of the innate immune response after the inhalation of the bacteria. Macrophages phagocytose *M. bovis* bacteria in an attempt to eradicate it either by phagolysosomal activation or due to macrophage apoptosis. Dendritic cells also phagocytose the bacteria and transfer it to the regional lymph nodes initiating the adaptive immune response. In the lymph nodes, the dendritic cells present the processed bacterial particles on MHC molecules to the TCR on the T-cells. This leads to the activation of lymphocytes populations, including T_{EM}, T_{CM} cells, and B-cells. The latter two cell types are involved in the formation of the granuloma which also includes giant cells and infected macrophages that failed to eradicate the bacteria. The granuloma is surrounded by fibrous tissue and denotes latent TB status. If activated, the fibrous tissue is disrupted releasing the bacteria to the surrounding tissues and bloodstream (Active TB).

There is a great variation in the reported minimum infectious dose for *M. bovis*. According to the Public Health Agency of Canada's Pathogen Safety Datasheet, the infectious dose of MTBC is unknown (53). A study by Dean *et al.* (54) showed that the intra-tracheal inoculation of 1 CFU of *M. bovis* in cattle was enough to develop pulmonary pathology typical of bovine tuberculosis. Other studies reported that 4 x 10^5 bacilli are required for infection via the respiratory route (55, 56). Infection through the oral route seems to be elicited by higher doses (55). Only a minority (~10%) of infected subjects develop tuberculosis because in most healthy subjects the immune defense system is sufficient to keep the organism in check to prevent the development of the disease (57).

2.2.4 *M. bovis* Proteins with Diagnostic Implication

MPB70. Some of the known proteins are given the prefix "MPB" which stands for protein from *M. bovis*. These proteins have homologues named 'MPT' which refers to

proteins from *M. tuberculosis*. Of these, MPB70 is one of the well-known *M. bovis* proteins. It was found in large quantities in the culture filtrate of the virulent strains of *M. bovis* and some of *M. bovis* BCG strains as in BCG Tokyo, Moreau, Russia, Sweden, Birkhaug, and Romania. MPB70 is a stable and active component of the bovine tuberculin, the reagent used in the BTB diagnosis (58, 59). MPB70 is recognized as an immunodominant antigen of *M. bovis*. The protein elicits a strong delayed-type hypersensitivity skin (DTH) reaction in guinea pigs sensitized with the BCG-Tokyo vaccine (60). It also elicits a strong DTH response and stimulates T-lymphocyte proliferation in *M. bovis*-infected animals (58). The protein was additionally found to boost the immunoglobulin G1 antibody responses upon intradermal skin testing in *M. bovis*-infected cattle with macroscopic lesions (58).

MPB70 has a molecular weight of 22.0 kDa and is actively secreted in the culture media by the bacterium. Also, it was present in the form of fragments, due to the cleavage by signal peptidase, or in the form of larger molecules due to the formation of stable aggregates or dimers. These forms were revealed by SDS-PAGE and Western blot using monoclonal mouse anti-MPB70 antibodies (59, 61). MPB70 is a highly species-specific protein for *M. bovis* with minimal cross-reaction from other species of mycobacteria. Interestingly, it was also observed that *M. tuberculosis* expressed low levels of the MPB70 homologue, MPT70, in contrast to high levels of MPB70 expression by virulent strains of *M. bovis* (62). On the other hand, the MPB70 protein cross-reacted with a corresponding component in *Nocardia asteroids* (63). MPB70 was shown to have homology with osteoblast-specific factor 2 (OSF-2). This association was related to the occurrence of osteitis in neonates after the vaccination by *M. bovis BCG* substrains that produced large amounts of anti-MPB70. These observations

provide evidence that MPB70 induces an immune response in the host following vaccination or infection with mycobacteria (64).

MPB83. MPB83 is another well-studied mycobacterial antigen. It is a glycosylated lipoprotein located at the bacterial surface. It is attached by a lipid tail coupled to the outermembrane and is processed by signal peptidase II. It is highly homologous to MPB70 proteins. Both proteins are highly expressed by *M. bovis*, and their homologues are markedly less expressed by *M. tuberculosis*. Both proteins share common features as the tendency to self-assemble, form dimers and the association of BCG osteitis after vaccination (59, 64). The expression of both *mpb70* and *mpb83* genes is controlled by the transcriptional regulator SigK. Some *M. bovis* Bacille Calmette-Guérin (BCG) strains acquire a point mutation in SigK, leading to the expression of minimal amounts of MPB70 and MPB83. (65).

MPB83 is an antigen recognized by the immune system of BTB-infected animals. This antigen induces early antibodies, as early as at 5 days of experimental infection in animals. The typical antibody response, however, usually develop around 4 weeks post-infection (66). On the other hand, antibodies to MPB70 are reported to develop 18–22 months after experimental infection (67). Despite this, serological sensitivity responses to MPB83 alone have not been good enough for use in the BTB-diagnostics (68).

ESAT-6 (Early secreted antigenic target of 6 kDa) **and CFP-10 (**10 kDa culture filtrate antigen, ESAT-6-like protein EsxB**)**. These proteins are members of the ESAT-6 family of small proteins. There is some degree of similarity between these proteins. Both proteins are secreted in the early stages of infection by metabolically active bacteria and were identified in the culture filtrate of *M. tuberculosis* as well as that of *M. bovis* (11, 69). They

possessed strong stimulatory antigenic properties. This is demonstrated as the induction of a strong T-cell response with considerable release of IFN-γ. The antigens are considered specific and well-conserved for the *M. tuberculosis* complex. The genes coding for these antigens are present in all isolates of *M. tuberculosis* and virulent *M. bovis* but are absent from all *M. bovis* BCG strains (70). The response to these antigens has been found in tuberculosis-infected subjects but not in BCG-vaccinated or unvaccinated subjects (70–72). Additionally, both antigens are recognized strongly by a high percentage of TB patients (73), as well as cattle infected with BTB (74). Therefore, ESAT-6 and CFP-10 are potentially useful antigens for the early detection of both human and BTB infection when applied in the tuberculin skin test, IFN-γ tests, or multiantigen serodiagnostic tests (75–78).

Antigen 85 proteins. The antigen 85 proteins have three known forms, Ag85A, Ag85B, and Ag85C. These proteins play a role in the bacterial virulence demonstrated by the high affinity of these proteins for the host fibronectin, a large adhesive glycoprotein. The latter host protein is responsible for the attachment of *M. tuberculosis* to the alveolar macrophages. Antigen 85 proteins also critically affect the bacterial cell survival through their role in the synthesis of covalently linked mycolic acid and the transfer of these mycolic acids to the cell wall, thus maintaining the cell wall integrity (79, 80). Antigen 85 proteins are extracellular proteins that share 70.8 to 77.5% sequence homology. They interact with the host immune system as demonstrated by the detection of immunoglobulin G antibodies against the 85A and 85B components of the *M. bovis* BCG antigen 85 complex. In addition, Antigen 85C from *M. bovis*, BCG and *M. tuberculosis* promotes monocyte-mediated uptake of microbeads coated with mycobacterial products via the complement receptor CR3 and are strong T-cell stimulators (81–83).

TB10.4 and TB7.3. TB10.4 is a secreted protein that belongs to the ESAT-6 family of low molecular mass proteins whereas TB7.3 is a low-molecular-mass protein outside this family. Both antigens were deployed in the IFN-γ based assays in human TB patients in attempts to differentiate TB-infected subjects from vaccinated subjects. Like CFP-10 and ESAT-6, TB10.4 demonstrated a strong immunogenic induction of IFN-γ release. The TB7.3, nevertheless, demonstrated a weaker induction of IFN-γ release (71).

MPB53. The gene coding for this protein is conserved among the members of the *M. tuberculosis* complex as well as nontuberculous mycobacteria. MPT53 of *M. tuberculosis*, a homologue of MPB53, induces strong, tuberculosis-specific antibody responses evidenced by the presence of anti-MPT53 IgG antibodies in high levels in tuberculous guinea pigs after aerosol infection but induces no delayed-type hypersensitivity. The protein exhibits minimal immune responses during human tuberculosis (84). The protein has a possible antioxidant activity (85).

EspC protein homologue. EspC of *M. tuberculosis* was identified in the membrane fraction of the bacteria and not the culture filtrate (86). Its homologue of 10.8 kDa size, stimulates IFN-γ responses in *M. bovis*-infected but not in naïve or BCG-vaccinated cattle. The protein was found to stimulate IFN-γ responses in a significant proportion of infected cattle that did not respond to the well-characterized mycobacterial antigens ESAT-6 and CFP-10 (87).

Other antigens. Thiol peroxidase (TPX) is a 16.9 kDa protein that is upregulated in oxidative stress associated with the infection. The protein has been detected in the extracellular space and is a strong T-cell antigen (69, 88, 89). It also has an antioxidant

activity where it catalyzes the reduction of hydrogen peroxide and thus protects the bacterial cell against oxidative stress (90). It may also have a role in virulence (91). **pepA** is a protein identified from the proteomic analysis of the PPDB, a probable serine protease involved in the processing of intracellular proteins. Its homologue in *M. tuberculosis* was found to induce the *in vitro* IFN-γ responses in TB patients and was a candidate for vaccine development (92, 93)**. TB8.4** is a secreted low molecular weight T-cell antigen that has a role in both resistance of *M. bovis* to intracellular stress and its adaptation to hypoxia (94)**.** Other antigens that induce a T-cell response include ESAT-6-like protein EsxG and PPE14 that were characterized based on their ability to induce *in vitro* IFN-γ responses in infected or BCG-vaccinated calves (95, 96).

2.3 BTB Diagnostics

The diagnosis of bovine tuberculosis, like human tuberculosis, remains extremely challenging. Currently, a single test is not enough to fulfill the criteria necessary to identify all infected animals. A combination of different tests is likely needed to achieve an adequate level of diagnostic accuracy. Sensitivity and specificity estimates for diagnostic tests can be dependent upon the animal population studied and the disease characteristics within this population. Ante-mortem diagnosis of BTB depends on “Indirect Diagnostics” which are based upon measuring various parameters of the immune response to the *M. bovis* organism. Broadly they fall into two categories: those based upon the cellular immune response and those based upon antibody responses. Post-mortem tests are mainly confirmatory tests that directly detect the causative agent.

2.3.1 Ante-mortem BTB Diagnostics - Assessment of Cellular Immunity

The two main tests under this category are the Tuberculin skin test and the IFN-γ assay. The Tuberculin skin test (TST), also known as the skin intradermal test (SIT), is the international standard for ante-mortem diagnosis of BTB in cattle herds and other animals. The test is based on the intradermal injection of tuberculin purified protein derivative of *M. bovis* (PPDB) usually performed on the mid-neck or in the caudal fold of the tail. This is followed by the detection of the elicited delayed-type hypersensitivity (DTH) response within 72 hours in response to the injection (PPD) (7). It was found that delayed hypersensitivity may not develop for a period of 3 to 6 weeks following the infection (2, 97). The interpretation of the test is based on an observed increase in skinfold thickness in addition to the evaluation of any clinical signs, if present (3).

Estimates of the sensitivity of the TST range from 50-60% (17, 26). The sensitivity of the test is affected by the interval post-infection, desensitization as well as subjective to the observer's variation. Moreover, the potency and dose of tuberculin administered may also be confounding factors. The performance of the test demonstrated a high incidence of cattle with false positive (non-specific) reactions to the tuberculin in the UK and Ireland (17, 97, 98). This is influenced by sensitization as a result of exposure to *M. avium*, *M. paratuberculosis,* and environmental mycobacteria (7). Also, the subjectivity of the test result interpretation and reporting is an additional compromise.

To decrease incidents of false-positive reactions, the single intradermal comparative cervical tuberculin (SICCT) test was developed which entails the simultaneous injection of both bovine and avian tuberculin side-by-side into the skin of the neck. The interpretation of the SICCT test is based on the assessment of the higher response of the BTB-infected

animals to bovine tuberculin compared to avian tuberculin. SICCT, therefore, provides better discrimination, compared to the TST, between animals infected with *M. bovis* and those sensitized to tuberculin after exposure to *M. avium* or environmental non-pathogenic mycobacteria. However, it is tedious and time-consuming to perform the SICCT. The (non-comparative) caudal fold TST is the primary ante-mortem test for routine TB screening of individual cattle and herds in the USA, Canada, and New Zealand (17). The U.S. and Canada use the caudal fold TST for preliminary screening of cattle and re-test reactors with the SICCT (99).

In addition to the classical intradermal tuberculin test, laboratory-based assays are used as ancillary tests either to maximize the detection of infected animals (parallel testing) or to confirm or negate the results of an intradermal skin test (serial testing). The second cellular immunity-based assay for bovine tuberculosis, is the *in vitro* IFN-γ assay. This laboratory test is based on the detection of IFN-γ released in response to a specific antigen (9). The principle of the test is the measurement of the release of IFN-γ from sensitized lymphocytes in a whole-blood culture system during a 16 to 24-hour period of incubation usually with PPD. The test compares the IFN-γ production following the stimulation with each of the avian and bovine PPD. The detection of bovine IFN-γ is carried out with a sandwich ELISA that uses two monoclonal antibodies to bovine IFN-γ (9, 100). The IFN-γ test can be used in parallel testing to the skin test to allow the detection of a greater number of infected animals thus reducing the risk of the transmission of infection to other animals or the potential contamination of the environment with the pathogen (9, 100).

An improvement of the specificity of the IFN-γ test is provided by using distinct mycobacterial antigens such as ESAT-6 and CFP-10. These antigens also offer the ability to

differentiate BCG-vaccinated from unvaccinated animals. Some countries such as the United Kingdom and New Zealand employ these antigens for serial testing (101). However, the use of these antigens has had a detrimental effect on the sensitivity of the test. The sensitivity of the test has decreased from 60% in the conventional IFN-γ assay to about 48% when these antigens alone were employed (102).

One of the controversies about the IFN-γ assay is the effect of storage on the IFN-γ release from the lymphocytes. In practice, laboratories tend to postpone setting up the blood culture on the day of sampling due to logistic and financial obstacles. Schiller *et al.* (103) compared the stimulation of lymphocytes by PPD within 8 hours and 24 hours of blood collection and found a modest reduction in the assay sensitivity from 96% to 90% which was statistically insignificant. In New Zealand, however, the delayed initiation of culture is not recommended, and their policy requires the initiation of the assay within 24 h of sampling. Ryan *et al.* (104) found that the optical density (OD) values obtained after overnight storage of the blood samples were substantially lower than those obtained when the blood culture was started on the day of sampling (104). Another study reported a change in the diagnostic outcome of using ESAT-6 but not PPDs when blood was tested 24 h after sample collection (105). Lack of consensus on the time of initiation of lymphocyte culture is evident, and the time to start the lymphocyte culture is thus decided by the authorized laboratories.

2.3.2 The Purified Protein Derivative of Tuberculin (PPD)

Generally, both the TST and the IFN-γ assay depend on the use of the PPD where there is still a controversy about the PDD's specificity and standardization and whether it is the best reagent for TB diagnosis.

The origin of the PPD dates to 1890 after Robert Koch isolated the tubercle bacilli. He produced a culture filtrate of the bacterium in the hope that it could be used to treat tuberculosis. Although this "Old Tuberculin" did not achieve its goal, it was the basis for the TST. Standardized tuberculin was produced in 1934 and was then termed PPD (7, 10). The bovine tuberculin currently in use in the UK and worldwide is a PPD preparation from *M. bovis* AN5, a field strain isolated in England in 1948 and selected for its optimal growth *in vitro* (106).

The PPD is still used in the current tests despite being discovered more than a hundred years ago. Its use has been associated with a high level of false-positive results in the previously described tests. Given the fact that 60 to 70% of the PPD proteomic content is made up of homologous antigens encoded by other mycobacteria, this explains why the PPD-based tests are unable to distinguish infection with *M. bovis* from vaccination with *M. bovis* Bacille Calmett-Guérin (BCG), or even exposure to non-tuberculosis mycobacteria. Cases of false-positive responses are generally attributed to an immune response triggered by homologous antigens from either vaccination with BCG or from environmental nontuberculous mycobacteria (11–15).

2.3.3 Ante-mortem BTB Diagnostics - Assessment of the Humoral Immunity

Serodiagnostic tests for detecting an antibody response have proved to be clinically useful for the detection of advanced disease and thus the identification of animals that are more likely to be infectious (52). A certain proportion of 'anergic' cattle can be detected by serological assays (17). These animals are in the late stages of disease development in which cell-mediated immunity is usually lost.

In practice, however, these tests generally complement tests based on cellular immunity. The tests require handling only once, are simple, and deliver the test results in a short turnaround time, making them very suitable for screening. Nevertheless, the low sensitivity of the antibody-detection tests remains a concern. Multiple antigens have been attempted in the tests with limitations in sensitivity and specificity as well as variability in the test results in different animal species (3, 107, 108).

ELISA is the most widely used serodiagnostic test. ELISA has been reported to be useful for detecting *M. bovis* infections in some domestic animals as well as wildlife. It is approved in some countries like New Zealand as an ancillary parallel test for farmed deer. According to the World Organisation for Animal Health, also known as, Office International des Epizooties (OIE), MPB70 and MPB83 are the antigens used in ELISA for the serodiagnosis of BTB. The use of MPB70 in ELISA achieved high specificity (96.4%) but had poor sensitivity (10 - 18.1%) in some studies (97, 98, 109). The late and irregular development of antibodies to MPB70 in the sera, 18 to 22 weeks after experimental infection of cattle, might be responsible for the observed low sensitivity (63). IDEXX TB ELISA, one of the most common commercially available tests and approved by the US Department of Agriculture (USDA), has also shown low sensitivity in testing. The assay has a sensitivity that varies by geographic region, with sensitivities of 45% and 9% in serum samples from cattle in the United States ($n = 146$), and Mexico ($n = 128$), respectively (18). This variability also extends to the species of the tested animals where IDEXX TB ELISA shows a sensitivity of 37.5% for the African buffalo (18). In contrast, the sensitivity of ELISA has been reported to be as high as 85% in deer due to earlier and more predictive antibody response in this species (3).

The Fluorescent polarization assay (FPA) was developed by Lin *et al.* (110) using fluorescein-labeled MPB70 to detect anti-MPB70 antibodies for the diagnosis of BTB. The test reported higher sensitivity (60-81%) (111, 112). Despite that, one report shows FPA did not perform better when compared to SICCT concerning the sensitivity (113). Another format for the serological assessment of BTB is a lateral flow-based rapid test that has proved to be useful in some domestic and wildlife animals. The test is efficient where IFN-γ tests are inaccessible or scarce and the skin test is unreliable. This test is licensed by the USDA for species such as elephants and is approved for use in the United Kingdom for badgers. However, the sensitivity of the latter test in cattle is relatively low, at 42% of tuberculin positive cattle (3, 114).

In general, serological tests have only been validated for a small number of hosts. Nevertheless, they are being used to help diagnose tuberculosis in some domestic animals and wildlife infected with *M. bovis* (99). Altogether, the discovery of new antigens to be employed in diagnostic tests is urgently required.

2.3.4 Post-mortem Testing of BTB

Post-mortem testing and confirmation of *M. bovis* infection involve culturing samples for about 12 weeks in addition to the detection of *M. bovis* DNA in clinical samples using Polymerase Chain Reaction (PCR) (3, 27). PCR is typically done on formalin-fixed, paraffin-embedded tissues (FFPE) which have demonstrated evidence of *M. bovis* infection in the form of the classical lesions and the colonizing mycobacteria. PCR is recognized as the most widely used direct method for the detection of the disease (3, 27).

In the Canadian Food Inspection Agency's Ottawa Laboratory (Fallowfield) (CFIA-OLF) as well as other laboratories, PCR on FEPE follows the histologic confirmation of the presence of a granuloma and acid-fast bacteria. Testing with PCR directly on fresh tissues is disfavored because it poses the risk of false positives due to laboratory cross-contamination. The most commonly used PCR targets for several acid-fast organisms are the insertion sequences IS6110 and IS1081 in addition to genes coding for MTBC-specific proteins, such as MPB70 (115). Results of PCR can be confounded by false-positive in cases of sample contamination. False-negative results are also reported especially in specimens containing low numbers of bacilli or due to inefficient DNA extraction of mycobacteria, difficulty in amplification of the bacterial DNA, or the presence of PCR inhibitors in the samples (17).

The culture method is the gold standard for TB diagnosis. For isolation of the bacteria, most laboratories use a variety of solid and liquid media in culture. The lesions are often paucibacillary; the presence of acid-fast organisms in histological sections may not be detected. This may lead to false-negative culture results. To avoid that, pooling of lymph node samples from the head and thorax for culture is recommended in animal cases with positive TST or IFN-γ test results and negative post-mortem histopathologic examination.

For primary isolation, solid egg-based media, such as Lowenstein–Jensen (LJ) or agar-based media such as Middlebrook 7H10 or 7H11 may be utilized. Liquid culture systems are deployed routinely in some hospital and veterinary laboratories; in these systems growth is measured by fluorometric means. Cultures are incubated for a minimum of 8 weeks (and preferably for 10–12 weeks) at 37°C with or without CO_2 (2, 97, 115). Confirmation after the culture can be done by biochemical tests or PCR. Currently, biochemical tests became less commonly applied because they are time-consuming, labour-

intensive, and are difficult to interpret. Thus, molecular methods are generally preferred if available (27, 99). The main laboratory procedures for in-herd testing for BTB as well as quarantine related procedures are outlined in **Figure 2-3**.

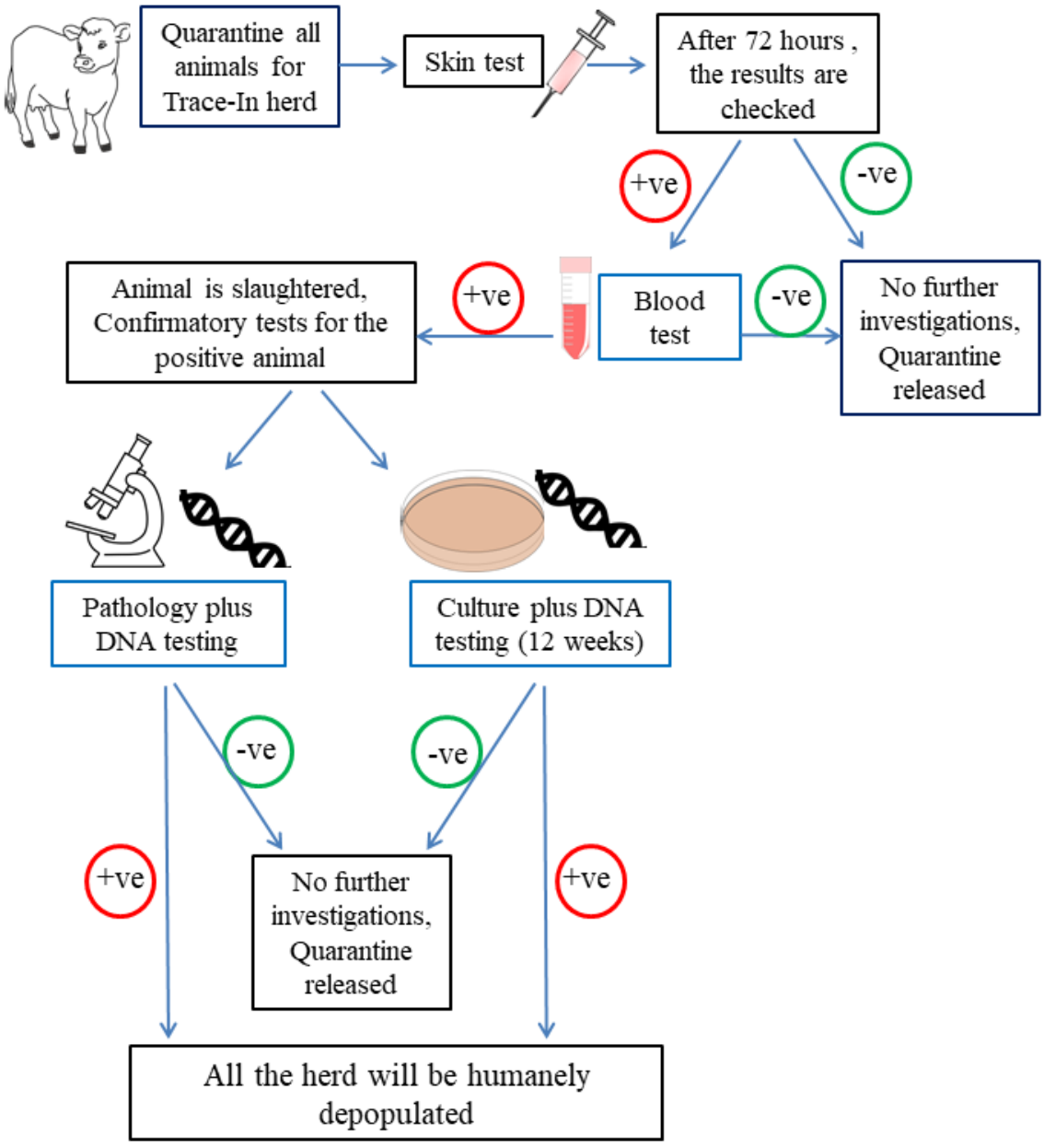


Figure 2-3. Outline of the main quarantine-related measures and laboratory investigations associated with trace in-herd bovine tuberculosis infection as adapted by the CFIA, Canada.

Trace-In herd means testing all herds that provided animals to the index herd. The skin test refers to the Tuberculin skin test. The blood test refers to the IFN-γ assay or the serological assays. Confirmatory tests refer to tests done post-mortem to confirm or negate BTB. Detected bacteria on histopathological examination of

suspected tissue as well as any growth after culturing the samples are confirmed by testing for *M. bovis* DNA. The testing procedures take about 14 weeks or more.

Molecular techniques are applied primarily for epidemiological purposes. The purpose of these tests is to identify the *M. tuberculosis* complex isolates and strains and to describe the patterns of transmission and spread of *M. bovis*. The techniques included DNA fingerprinting techniques, restriction fragment length polymorphism analysis, spoligotyping, and multiple loci variable number tandem repeat analysis (VNTR) (3). Currently, the molecular methods are rapidly advancing; the whole-genome sequencing (WGS) for BTB has largely replaced the traditional molecular techniques including spoligotyping and VNTR. Proving to be more efficient, WGS is used to detect single nucleotide polymorphisms (SNP) in the target strains to identify and trace the transmission pathways as parts of the official tuberculosis eradication program. WGS was capable of tracing and identifying sources of infections either from wildlife or from trade movement for better control measures (116).

WGS is used by countries such as the USA, New Zealand, and Canada (116, 117). The method helps resolve indistinguishable genotypes identified by other molecular methods and possibly identify new strains. The results of WGS can be obtained within the same time frame as traditional genotyping techniques (typically within 4–6 weeks after tissue submission) for the field investigation (116, 117). Traditional molecular techniques and WGS results show that Canadian strains of *M. bovis* exhibit diversity in genotypes and are linked to some strains of the geographical origins including the USA and Mexico (118, 119).

Chapter 3. Materials and Methods

3.1 Bacterial Strains, Growth Conditions and Preparation of Genomic DNA and Plasmid DNA samples

M. bovis isolate 2011/0690 was used in the study. This is a field strain isolated from the BTB outbreak in cattle in British Columbia, Canada in 2011. The cells were grown in Middlebrook 7H9 with Albumin-dextrose-catalase (ADC) enrichment for 7 weeks at 37°C, heat-killed at 95°C for 20 min. Sterility was proven by inoculating heat-killed samples into Mycobacterium Growth Indicator Tube (MGIT) and incubated in the Bactec MGIT 960 system for 5 weeks. Genomic DNA was extracted using DNeasy ® Blood & Tissue Kit (Qiagen, Germany) following pretreatment using lysozyme 50mg/ml at 37°C for 30 min as per the manufacturer instructions for Gram-positive bacterial cells.

Listeria monocytogenes HCC23 serotype 4a cells were cultured in BHI broth, resuspended in TE buffer containing 50mg/ml lysozyme at 37°C for 30 min, and then mixed with DNAzol (Invitrogen, USA) and ethanol. The cell lysate was centrifuged at 13000 rpm for 1 min, the supernatant was discarded, and the DNA pellet was dissolved in NaOH.

Escherichia coli strains (DH5α and Rosetta DE3/pLysS) were used for cloning and *in vivo* recombinant protein expression, respectively. *E. coli* cells harboring recombinant plasmids were cultured in Luria-Bertani (LB) broth (BD Biosciences, Sparks, MD, USA) supplemented with 100 µg/ml carbenicillin to maintain the derivatives of pIVEX2.3d (Biotechrabbit, Germany) and pT7 CFE (ThermoScientific, USA) or 50 µg/ml kanamycin to maintain the derivative of pLIC plasmid constructed in house (120). All bacterial strains were obtained from the CFIA, OLF.

3.2 Using Bioinformatics to Identify Protein Candidates

PSORTb, an online tool, was used for computational prediction of the subcellular localization of proteins encoded by the bacterial genomes (121). The genome sequences of the reference strain, *M. bovis* AF2122/97, as well as *M. bovis* strain 2011/0690, also known as MBWGS010, sequenced by the CFIA (119, 122), were submitted for the analysis with PSORTb. Proteins that had the prediction of their final localization as extracellular or outer membrane proteins were selected as candidates for the study. In addition, other proteins with a propensity to be extracellular were selected based on the output of the Extracellular Support Vector Machine (ECSVM) algorithm, provided by PSORTb. ECSVM is a machine learning-based classifier trained using the sequences of proteins with known extracellular localization. Protein-BLAST was used to identify proteins that share 50% or higher sequence identity within multiple *M. bovis* genomes. The proteins were also blasted against a list of serologically cross-reacting mycobacteria obtained from the CFIA. The list included *M. non-chromogenicum*, *M. intermedium*, and *M. avium complex*. Protein candidates with less than 40 % sequence homology with cross-reacting strains were selected.

3.3 Bacterial Growth of *M. bovis* for the Proteomic Analysis

One ml of concentrated, previously frozen *M. bovis* isolate 2011/0690 was inoculated into 20 ml Middlebrook 7H9 with ADC enrichment and incubated at 37 ºC. After 10 days, 1 ml of the culture was inoculated into 10 ml vials of 7H9 liquid medium with ADC enrichment. The vials were incubated at 37ºC. At least three vials were used from each of three time points: at 10 days, 21 days, and 6 weeks of culture. Cultures were sterile filtered twice: primarily, using Vivaspin centrifugal concentrators (Sartorius) that were centrifuged at 1000 g for 5 min, and secondly, the supernatant from each vial was filtered through a sterile,

0.22 µm syringe filter to ensure sterility and absence of any viable bacteria. The OD at 600 nm was measured for the bacterial culture and sterility was further assessed by culturing 500 µl of each supernatant sample into 7H9 medium- ADC enriched then LJ at 37ºC. Sterility was confirmed when no bacterial growth was observed with the supernatant samples, indicating that the materials to be used for proteomic analysis are free of viable mycobacteria.

3.4 Sample Preparation for the Proteomic Analysis of *M. bovis* Proteins Secreted into the Culture medium

3.4.1 Bovine albumin depletion from the culture supernatant. The bovine albumin from the culture supernatant samples was depleted prior to proteomic analysis using a Pierce Albumin Depletion Kit (Thermo Scientific, USA) following the manufacturer's instructions.

3.4.2 Protein precipitation. The culture supernatant proteins were precipitated using 10% TCA (v/v), allowed to precipitate overnight in a -20°C freezer, then spun for 30 min at 10, 000 g at 4°C. The supernatant was removed, and the pellet was washed twice using 50% ether in 50% acetone (ice cold). The sample was centrifuged again for 5 min at 10, 000 g to remove the supernatant, and the residual supernatant was left in the air to evaporate.

3.4.3 Reduction and Alkylation. The dry protein pellet was solubilized and denatured using 7M urea in 50 mM ammonium bicarbonate and ProteaseMAX™ Surfactant (Promega, USA) to help protein solubilization (~0.05% w/v). Dithiothreitol (DTT) (Millipore sigma, Canada) was added to reduce disulfide bridges at a final concentration of 10 mM. Iodoacetamide (IAA) (Millipore sigma, Canada) was added to alkylate free cysteines at a final concentration of 25 mM. Protein is quantitated using a Pierce™ BCA Protein Assay

Kit (Thermo Scientific, USA). The protein was diluted 5 times to bring the urea concentration down to 1M.

3.4.4 Trypsin Digestion and Peptide Purification. Trypsin Gold, Mass Spectrometry Grade, (Promega, USA) was used at a 1:20 ratio of trypsin mass to protein mass, for 100 µg of protein. ProteaseMAX™ Surfactant (Promega, USA) is added to a final concentration of 0.05 %. Digestion was done overnight at 37°C and quenched by adding TCA to a final concentration of 10%. This was followed by centrifugation at 16,000 g for 10 min. Peptides were purified from the soluble fraction using Pierce C18 columns (Thermo Scientific, USA) as per the manufacturer's instruction. Peptides were eluted from C18 columns using 70% (v/v) acetonitrile which is later removed with a Spin Vac (ThermoFisher Scientific, USA).

3.5 MS Analysis and Data Processing

3.5.1 LC-MS/MS analysis of tryptic digests of protein sample. The peptide samples were diluted in 0.1% (v/v) formic acid. For each sample, approximately 300 ng was analyzed by loading onto a Thermo Acclaim PepMap® 100 C18 trap column (75 µm x 2 cm with 3 µm particles) and desalting with 0.1% formic acid in water (solvent A) at 5.0 µl/min before separating on a Thermo PepMap® RSLC 100 C18 reverse-phase analytical column (75 µm x 15 cm with 2 µm particles). Chromatographic separation was achieved at a flow rate of 0.300 µl/min as follows: (solvent B was 0.1% (v/v) formic acid in acetonitrile) initial, 2% B; increasing to 20% B in 40 min: then to 40 % B by 60 min, then to 60% B by 70 min, increasing to 95% B in 5 min where it is held at 95% B for 10 min before being returned to 2% B for column equilibration. The eluting peptides were analyzed with a Thermo

OrbitrapTM FusionTM Tribrid mass spectrometer (Milford, MA, USA) operating in data-dependent acquisition (DDA) 'Top-Speed' mode. All samples were acquired in triplicate.

3.5.2 Data Processing and Proteome Discoverer Database Search. MS/MS spectra were processed using Proteome Discoverer 2.3 (Thermo Scientific). Searches were performed using Sequest, Mascot, and MS Amanda search algorithms against a SwissProt *M. bovis* database and a protein contaminants database. Searches were performed specifying trypsin as the proteolytic enzyme, carbamidomethylation of cysteine as a fixed modification, deamidation of asparagine and glutamine, and protein N-terminal acetylation as variable modifications. Peptide and fragment ion mass tolerances were set at 5 ppm and 0.6 Da, respectively. Protein identifications have been reported at a 1% false discovery rate (FDR). Protein quantitation was done using label-free quantitation. Protein localization prediction was obtained using UniProt (https://www.uniprot.org/) and using PSORTb (https://www.psort.org/psortb/), and locus tags were obtained using PROTEIN BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and using mycobrowser (https://mycobrowser.epfl.ch/).

3.6 PCR Amplification of Candidate Genes

3.6.1. Primer design for PCR amplification of candidate genes

For cloning a target open reading frame (ORF) into pIVEX2.3d (C-His) by PCR, the forward primer was designed to contain 5'GTTTAACTTTAAGAAGGAGATATACC3' followed by a gene-specific sequence (beginning with ATG codon) coding for at least 6 N-terminal residues of the target protein. The reverse primer is made up of 5'GATGATGATGATGATGAGAACCCCC3' followed by a sequence reverse

complementary to the sequence coding for at least 6 C-terminal residues with the omission of the stop codon. For cloning a target DNA into pT7CFE1(C-His), the forward primer is made up of 5'CGATGATAATATGGCCACCACCCAT3' followed by a gene-specific sequence (beginning with ATG codon) coding for at least 6 N-terminal residues of the target protein. The reverse primer is made up of 5'GTGGTGGTGGTGGTGGTGCTCGAG3' followed by a sequence reverse complementary to the sequence coding for at least 6 C-terminal residues with the omission of the stop codon. For an ORF of larger than 3 kb, the sequence was divided into two or three smaller segments, maintaining an overlap of at least 150 bp between the sequences prior to being cloned. Sequences of the primers used in this study are described in Appendix 2. *LMHCC_RS00060,* an ORF from *L. monocytogenes*, was used as a control gene for comparison. Primers for this gene are as follows:
FP:5'GTTTAACTTTAAGAAGGAGATATACCATGGGGTGGTATGATCAACCGAA3',
RP: 5'GATGATGATGATGATGAGAACCCCCTACCGGAGCTACTGGGTTTT3'

3.6.2. DNA polymerase enzymes for the optimization of PCR procedures

Amplification of some selected ORFs by PCR was done using high fidelity polymerase enzymes including Phusion® High-Fidelity DNA Polymerase (New England Biolabs (NEB)), Q5® High-Fidelity DNA Polymerases (NEB), Platinum™ SuperFi™ DNA Polymerase (Invitrogen), PrimeSTAR GXL DNA Polymerase (Takara Bio) and Taq polymerase.

3.6.3. Master mixes for the optimization of PCR procedures.

Enhancers included in a polymerase kit were used if supplied; otherwise, 2-5% DMSO (Sigma, Cat# D9170) was used instead. Master mixes with an enhancer and without an enhancer were prepared following the manufacturer's instruction for each ORF candidate.

Those with the enhancer had either the maximum kit-supplied enhancer (10 µl per 50 µl reaction) or 5% DMSO (0.7M) for both *mbp83* and *Mb0129* from *M. bovis*. The *LMHCC_RS00060* reaction mixes with the enhancer contained either 50% of the kit-supplied enhancer or 2.5% DMSO (0.35 M). The primers were added at a final concentration of 0.5 µM, and the dNTPs were at a final concentration of 0.2 mM. The genomic DNA was used at 5 ng/reaction. All PCR reactions were done in a 50 µl final reaction volume.

3.6.4. Annealing temperature calculation for the optimization of PCR procedures.

Annealing temperatures for the 3-step PCR reactions were used as specified by the manufacturer's recommendations, 60°C for PrimeSTAR GXL DNA Polymerase, or using the *Tm* calculator on the manufacturers' websites for Phusion and Q5 polymerases (http://tmcalculator.neb.com/#!/main) and Platinum™ SuperFi polymerase (www.thermofisher.com/tmcalculator).

3.6.5. Thermal cycle conditions for the optimization of PCR procedures.

The Biometra TAdvanced thermal cycler (Analytik Jena, Germany) was used for the study. Three protocols were used to assess the productivity of each enzyme on its own. Amplification using 3-step (3St), 2-step (2St), and a 3Step with touchdown (TD) protocols were attempted following the manufacturer's recommendations for the different enzymes. The protocols for the 3St and 2St are described in Table 3-1, and the TD protocol is presented in Table 3-2.

Table 3-1. 3St and 2St protocols for the different enzymes used.

Protocol		**3St**						**2St**					
Enzyme		Phusion, Q5, SuperFi[b]		GXL[b]		Taq		Phusion, Q5, SuperFi		GXL		Taq	
Step		Temp[a]	Time	Temp[a]	Time	Temp	Time	Temp[a]	Time	Temp[a]	Time	Temp[a]	Time
Initial denaturation		98	30s	-	-	95	2m30s	98	30s	-	-	95	2m30s
35 cycles	denature	98	10s	98	15s	95	20s	98	10s	98	15s	95	20s
	anneal	63	20s	60	15s	60	20s	-	-	-	-	-	-
	extend	72	1m30s	68	2m	72	2m30s	72	1m30s	68	2m	72	2m30s
Final extension		72	5m	-	-	72	5m	72	5m	-	-	72	5m
Hold		4	∞	4	∞	4	∞	4	∞	4	∞	4	∞

[a] Temp refers to temperatures in °C

[b]"SuperFi" stands for Platinum™ SuperFi™ DNA Polymerase and "GXL" for PrimeSTAR GXL DNA Polymerase

Table 3-2. TD protocol for the different enzymes used.

Enzyme		Phusion, Q5, SuperFi[b]		GXL[b]		Taq	
Step		Temp[a]	Time	Temp[a]	Time	Temp[a]	Time
Initial denaturation		98	30s	-	-	95	2m30s
10 cycles	denature	98	10s	98	15s	95	20s
	anneal[c]	72 (-1.6/ cycle)	20s	72 (-1.2/ cycle)	15s	72 (-1.6/ cycle)	20s
	extend	72	1m30s	68	2m	72	2m30s
30 cycles	denature	98	10s	98	15s	95	20s
	anneal	55	20s	60	15s	55	20s
	extend	72	1m30s	68	2m	72	2m30s
Final extension		72	5m	-	-	72	5m
Hold		4	∞	4	∞	4	∞

[a,b] as described in

Table 3-1.

[c] For the TD PCR the annealing temperature decreases by 1.2 or 1.6 °C per cycle for the first ten cycles until it reaches the desired annealing temperature.

3.6.6. Final protocol for the PCR amplification of all *M. bovis* ORFs.

The amplification of *M. bovis* ORFs by PCR was performed using the high fidelity polymerase, PrimeSTAR GXL DNA Polymerase (Takara Bio, Cat# R050Q). Master mixes were prepared following the manufacturer's instruction with the addition of 5% DMSO (0.7M) (Sigma, Cat# D9170). The primers were added at 0.5 μM final concentration, and the dNTPs were at 0.2mM final concentration. The genomic DNA was used at 5 ng/reaction. All experiments were done in a 50 μl final reaction volume. The thermal cycling conditions consisted of 35 cycles of denaturation at 98°C for 10 seconds, annealing and extension at 68 °C for 1 min/kb, and the reaction mix was held at 4°C after the completion of all the cycles. The thermal cycler ramp speed was adjusted at 2°C/sec.

3.7 Creation of Fusion by Overlap Extension PCR

Twelve genes were selected to be synthesized in the form of six fusion-genes using overlap extension PCR as described before (123) with modifications. Briefly, each fusion is composed of two ORFs. The forward primer of the first ORF and the reverse primer of the second ORF are designed as described in Section 3.6 for cloning the ORFs into pIVEX2.3d. The reverse primer of the first ORF and the forward primer of the second ORF are constructed such that each has a 10 -15 bp 5' overhang complementary to the end of the other ORF and 20 bp specific to the ORF. The first round of PCR is run to amplify each ORF using the *M. bovis* DNA as the template. The PCR products were analyzed by agarose gel electrophoresis and the correct-sized PCR product was extracted using a QIAquick® Gel

Extraction Kit (Qiagen). The second round of PCR is run using the forward primer of the first ORF, the reverse primer of the second, and the amplicons of both ORFs from the first PCR round as the DNA template. The primers for the six gene fusions are listed in Table 3-3.

Table 3-3. Primers used for the creation of fusion genes.

Fus#	locus tag	Primer#	D	Sequence (5' to 3')
1	Mb1387	2332	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCCGCACGCTTGCGTT
		2619	R	CGGGATATCGCTCATTGAGATCCGCATCCAAAGCT
	Mb1857	2620	F	TGGATGCGGATCTCAATGAGCGATATCCCGTCCGA
		2335	R	GATGATGATGATGATGAGAACCCCCCTCGGTCAGTGTGCCGCGGT
2	Mb2659c	2362	F	GTTTAACTTTAAGAAGGAGATATACCATGACCACCGCACGCGACAT
		2363	R	TGCGGTCCATGCTGGCGAGGGCCATGGGCG
	Mb2801c	2364	F	CCTCGCCAGCATGGACCGCAGAATCCTGAG
		2365	R	GATGATGATGATGATGAGAACCCCCGATGGGCGCCGCCCCGCGCA
3	Mb3219c	2384	F	GTTTAACTTTAAGAAGGAGATATACCATGTCAGCAAGATCAACGCA
		2385	R	GTAGGTTCATACCCAGTCGTCCCTGCTCCG
	Mb3389	2386	F	ACGACTGGGTATGAACCTACGGCGCCATCA
		2387	R	GATGATGATGATGATGAGAACCCCCGGACCCGGGCAGCCCCGGCA
4	Mb0340	2408	F	GTTTAACTTTAAGAAGGAGATATACCATGGCAAAAATGGCTGGATC
		2409	R	GCGCCGACATATCGACCAAATCGTCCTCGG
	Mb3743c	2410	F	TTTGGTCGATATGTCGGCGCTGCTCGCTCA
		2411	R	GATGATGATGATGATGAGAACCCCCGATCCCTGGTACAGGTGGTG
5	Mb0824	1998	F	GTTTAACTTTAAGAAGGAGATATACCATGGCACTCAAGGTAGAGAT
		2413	R	GGCACCGGCCATGGCACTCAAGGTAGAGAT
	Mb0723	2412	F	TGAGTGCCATGGCCGGTGCCCCGAACAGGT
		1989	R	GATGATGATGATGATGAGAACCCCCGGCCGGTGCCCCGAACAGGT
6	Mb3905	1534	F	GTTTAACTTTAAGAAGGAGATATACCATGACAGAGCAGCAGTGGAA
		2415	R	AATGGGCTTCATGACAGAGCAGCAGTGGAA
	Mb3904	2414	F	GCTCTGTCATGAAGCCCATTTGCGAGGACA
		1533	R	GATGATGATGATGATGAGAACCCCCGAAGCCCATTTGCGAGGACA

Fus# = Fusion gene number, D = Primer Direction

3.8 *In vivo* Cloning of ORF or Gene Fragment into pIVEX2.3d

The plasmid vector pIVEX2.3d was minipreped from five ml of *E. coli* culture using a QIAprep Spin Miniprep Kit (Qiagen, Germany). One µg of plasmid DNA was linearized using NotI HF (NEB, United States) following the manufacturer's instructions. The linearized vector was then PCR-amplified, using primers, 5'GGGGGTTCTCATCATCATCAT3' and 5'GGTATATCTCCTTCTTAAAGTTAAAC3' using the Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Canada) following the manufacturer's instructions. This was followed by gel extraction of the amplified vector using a QIAquick® Gel Extraction Kit (Qiagen).

The PCR products of target ORFs were purified using a QIAquick PCR Purification Kit (Qiagen) or using a QIAquick® Gel Extraction Kit (Qiagen). A mixture was prepared by combining purified vector and each purified target DNA in a 1:2 vector to target DNA ratio together with chemically competent DH5α (124). The mixture was incubated on ice for 30 min, heat-shocked at 42°C for 35 secs and chilled on ice for 2 min. Two hundred and fifty microliters of S.O.C medium were added to the mixture, which was then grown on a plate of LB-agar containing carbenicillin (100 µg/ml) at 37°C overnight. The following day, a few colonies were selected from each plate and grown in 5 ml LB-broth containing carbenicillin (100 µg/ml) at 37°C overnight. Plasmid DNA was minipreped from the overnight culture. The presence of insert in the recombinant plasmid was verified by colony PCR using T7 promoter and T7 terminator primers or by restriction digestion of the recombinant plasmid using Hind IIIHF enzyme (NEB, Canada). The presence of an insert in all recombinant plasmids was confirmed by DNA sequencing (McGill University and Génome Québec Innovation Centre).

3.9 *In vitro* Expression of Recombinant Proteins using *E. coli* lysate

The recombinant vector DNA derived from pIVEX2.3d was diluted to 0.05 µg/ µl in water, and 10 µl of the diluted vector is used to express the recombinant protein. The RTS 100 *E. coli* kit (Biotechrabbit, Germany) was used for the *in vitro* transcription and translation of recombinant proteins (IVTT) following the manufacturer's recommendation. The *E. coli* lysate containing the recombinant DNA was incubated for 6 hours at 30°C. A positive control plasmid pIVEX2.3-GFP and a negative (no DNA) control were also prepared. The expressed target protein was analyzed by SDS-PAGE and Western blotting.

3.10 *In vivo* Cloning of ORF or Gene Fragment into pT7CFE1 Plasmid Vector

The procedures were similar to section 3.8 except for using the pT7CFE1-CHis vector (ThermoScientific, USA). The vector was linearized using the NotI-HF enzyme (NEB). The vector was PCR-amplified, using primers 5'CTCGAGCACCACCACCACCA3' and 5'ATGGGTGGTGGCCATATTAT3' using PrimeSTAR GXL DNA Polymerase (Takara Bio) following the manufacturer's instructions. Recombinant plasmids were analyzed for the presence of an insert by colony PCR using T7 promoter and T7 terminator primers or by double restriction digestion of the recombinant plasmid using PvuII-HF and XhoI enzymes (NEB, Canada) as per the manufacturer's instruction. DNA sequencing was performed to confirm the presence of an insert in recombinant plasmids (McGill University and Génome Québec Innovation Centre).

3. 11 *In vitro* Expression of Recombinant Proteins using Mammalian Cell Lysate

For the cell-free expression of recombinant proteins in a mammalian system, 1-Step Human Coupled IVT Kit (ThermoScientific, USA) was used following the manufacturer's instructions. One µg of recombinant plasmid derived from pT7CFE1 was used in a final 25 µl volume. A positive control plasmid pCFE-GFP and a negative (no DNA) control were also prepared. The lysate containing a recombinant plasmid was incubated at 30°C for 4 hours to express the corresponding protein.

3.12 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western Blotting

Proteins were separated by size using electrophoresis with a 4% stacking gel and 12% separating gel containing SDS. Separated denatured proteins in the gel were visualized by staining with Coomassie blue. For recombinant proteins expressed by the IVTT as well as those expressed *in vivo* in *E. coli*, Western blotting is performed after transferring proteins onto a nitrocellulose membrane using a semi-dry transfer apparatus (Bio-Rad, Mississauga, ON, Canada) according to manufacturer's instructions. Probing the immobilized proteins using primary or secondary antibodies were done using the iBind Western System (Life technologies, USA) following the manufacturer's recommendations. For visualization of the protein bands, the nitrocellulose membrane was then stained using horseradish peroxidase (HRP) substrate kit (Bio-Rad) for colour detection or Clarity ECL Western blotting substrate (Bio-Rad) for chemiluminescent signal detection. The manufacturer's instructions for each kit were followed. Monoclonal mouse Anti·Penta-His Antibody (Qiagen) at a dilution of

1:1000 in PBS containing 3% (w/v) bovine serum albumin (BSA) or at 1:200 dilution if prepared in ibind solution (Life technologies) was used to probe membrane-bound proteins. Primary antibodies in the form of serum from *M. bovis*-infected cattle were used at 1:100 dilution in either of the two above mentioned solutions. Secondary antibodies of Peroxidase-AffiniPure goat anti-mouse IgG, H+L (Jackson ImmunoResearch, USA) or Peroxidase-AffiniPure goat anti-bovine IgG, H+L (Jackson ImmunoResearch) were used at a 1:1000 dilution in PBS containing 3% (w/v) BSA or 1:200 dilution if prepared in ibind solution (Life technologies).

3.13 High Throughput Purification of Recombinant Proteins

The 96-wells filter plate method was developed and optimized for use in the purification of the recombinant soluble proteins using metal-affinity chromatography. AcroPrep™ Advance 96-well filter plates for aqueous filtration (Pall laboratory, USA) were used. Ni-NTA agarose (Qiagen, Germany) resin was washed three times by the wash/equilibrium buffer pH 8.0 (50 mM NaH_2PO4, 300 mM NaCl, 20 mM imidazole). The slurry is transferred in 100 µl aliquots to filter plate wells (equivalent to 50 µl of sorbent per well). Sample proteins are loaded into the wells, washed with the wash/equilibrium buffer by centrifugation, and eluted using the elution buffer pH 8.0 (50 mM NaH_2PO_4, 250 mM imidazole). Insoluble proteins were purified similarly, with the modification of the wash/ equilibrium buffer pH 8.0 (8M urea, 50 mM NaH_2PO4, 300 mM NaCl, 20 mM imidazole) and the elution buffer pH 8.0 (8M urea, 50 mM NaH_2PO_4, 250 mM imidazole)

3.14 Protein Microarray Printing and Processing

Recombinant proteins expressed from the *in vitro* system were purified and used for microarray printing. For each protein, a working solution composed of 25% (v/v) of the

purified sample protein, 25% (v/v) glycerol, in PBST (0.2%). Proteins were printed in duplicates on the Oncyte-Supernova nitrocellulose glass slides (Grace-Biolabs, USA) using the TopSpot D microarray workstation printer (Biofluidix, Germany). The slides were left overnight to dry at 4° C in the dark. On the next day, the slides were assembled in a proplate separator (Grace Biolabs, USA). This separator divides the slide into 16 wells, each well has a nitrocellulose pad on which proteins are printed. Each well was incubated with different primary or secondary antibodies. The slides were blocked using a 50% Super G protein array blocking buffer in PBST (Grace Biolabs, USA). The blocking was left one hour at room temperature with shaking 50 rpm/min and then overnight at 4° C. On the third day, the slides were washed using PBST. The slides were then incubated with sera from infected cattle (1:100) or with monoclonal mouse Anti·Penta-His Antibodies (1:1000) (primary antibodies), washed with PBST, and incubated with the secondary antibody (1:200). Antibodies were diluted in 5% Super G protein array blocking buffer in PBST (Grace Biolabs, USA). Arrays are washed between each step thrice by PBST and dried after all incubations and prior to detection. The slides were kept in the dark at 4° C until reading by the scanner.

For the primary antibodies, sera were obtained from 10 adult naturally infected animals, 5 juvenile naturally infected cattle (less than 2 years of age), 5 experimentally infected cattle, and 8 sera from non-infected cattle. Sera from naturally infected cattle were obtained from the USDA, National Veterinary Services Laboratories (NVSL). Sera from experimentally-infected animals were obtained four months after the intra-tracheal pathogen inoculation as previously described (125). All infected animals, natural and experimental, were confirmed for BTB infection by post-mortem culture and PCR analysis. In addition, monoclonal mouse Anti·Penta-His Antibody (Qiagen) was used as a primary antibody. Cy-3-

conjugated AffiniPure Goat Anti-Bovine IgG (H+L) (Jackson ImmunoResearch) and Cy-3-conjugated Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch) antibodies were used as secondary antibodies.

For the controls, a negative protein control (a sample containing all the constituents of the *in vitro* expression kit except the vector DNA and purified as described in section 3.8) and another negative control, His-tagged recombinant E2 envelope glycoprotein from the classical swine fever virus (CSFV) (126), were printed on each pad and probed with the primary and secondary antibodies.

3.15 Image Scanning and Analysis

The fluorescent signal from microarray slides was read using the GenePix Pro 4200A laser scanner (CA, USA) at the National Research Council, Ottawa. The resolution was set at high resolution (5 μm pixel size). Laser power was adjusted to 95%, PMT was at 290, and wave-length was at 532 nm. Images were saved as (.tif) files. The fluorescent intensity of each spot was calculated using ImageJ (Version 1.51). Using ImageJ, the pixel intensities of each spot were measured in duplicates and the average was reported as numerical fluorescence intensities (FI). The fluorescence of the background was subtracted from each spot. For each array, the average signal of controls ("No DNA" controls and negative His-tagged recombinant protein control) was subtracted from spots' signal intensities. Recombinant proteins were verified to be present on the spot when the signal intensity from the reactions with monoclonal mouse Anti·Penta-His Antibody was above the mean of "No DNA" control signals plus 2 standard deviations.

3.16 Cloning, Large Scale Expression and Purification of the Selected Recombinant Proteins

Four antigens were selected based on the results of microarray screening for the evaluation of their diagnostic potentials. The coding sequences of these antigens were amplified by PCR using specific primers (Table 3-4) for the purpose of cloning into the pLIC-CHis vector using a ligation independent cloning technique. The genomic DNA from *M. bovis* isolate 2011/0690 was used as the template in PCR with PrimeSTAR GXL DNA Polymerase (Takara Bio), suitable for high GC content genes. All PCR products were purified and then treated with T4 DNA polymerase in the presence of dCTP.

Table 3-4. Oligonucleotide primers for cloning protein candidates identified by Aim#2.

PCR Product Description	Primer Name *	primer sequence
Mb0598c	P2911 For	5'TTTAAGAAGGAGATATAAGTC ATGAAGCACTTCACGGCGGC3'
	P2912 Rev	5'AGTGGTGGTGGTGGTGGTGAGTCCC GGGCGTGATGGTCGTCTGCT3'
Mb3469c	P2913 For	5'TTTAAGAAGGAGATATAAGTC ATGGCTGACCGGTTGAACGT3'
	P2914 Rev	5'AGTGGTGGTGGTGGTGGTGAGTCCC CGGACCCGCTTGCGGAAGCT3'
Mb2740c	P2915 For	5'TTTAAGAAGGAGATATAAGTC ATGAACGGGCAGAGAGGTCA3'
	P2916 Rev	5'AGTGGTGGTGGTGGTGGTGAGTCCC ATCCGCCCTAATGAAGCCGT3'
Mb3453c	P2917 For	5'TTTAAGAAGGAGATATAAGTC ATGACGACAGTCTTGGGCAT3'
	P2918 Rev	5'AGTGGTGGTGGTGGTGGTGAGTCCC CCGCACCTGCCCCTGCATCA3'

*For= Forward, Rev= Reverse
Underlined sequences are added sequences of homology with the pLIC-CHis.

The pLIC-CHis vector was developed in house (120, 127–129). The *E. coli* cells harboring the vector were grown in 15 - 20 ml of LB broth containing kanamycin (50µg/ µl). The pLIC-CHis vector was minipreped and linearized using ScaI (NEB). The linearized vector was purified using a gel extraction kit (Qiagen) and then treated with a T4 DNA polymerase in the presence of dGTP. The treated insert and vector were mixed in a 2:1 molar

ratio and used to transform *E. coli* DH5α strain competent cells, which were then plated on kanamycin-containing LB agar at 37°C overnight. Recombinant plasmids were, minipreped from selected clones and verified for the correct insert by restriction digestion and DNA sequencing. *E. coli* Rosetta pLysS strain transformed with the recombinant plasmid was used for the large scale production of the identified antigens as described (127). The induction of protein synthesis in *E. coli* was done using 1 mM IPTG. The expression of recombinant proteins was confirmed by SDS-PAGE and Western blot analysis using monoclonal mouse Anti·Penta-His Antibody. For purification of the recombinant protein expressed in *E. coli*, the cell pellets were resuspended in phosphate-buffered saline (PBS) (pH 7.2) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) and lysed by passage through a French press at 1500 lb/in^2. The homogenates were spun at 27,000 g at 4°C. Three recombinant proteins, Mb2740c, Mb0598c, and Mb3469c, were purified using Ni-agarose affinity chromatography under native conditions and the fourth protein, Mb3453c, was purified under denaturing conditions from the insoluble fraction as previously described (126, 127).

3.17 Evaluation of the Recombinant Proteins in ELISA Format

The purified recombinant proteins were evaluated in ELISA for the serological detection of animals infected with *M. bovis*. Purified antigens were diluted to a final concentration of 2 µg/ml using 0.06 M carbonate buffer, pH 9.6. The antigen coating was done overnight at room temperature using Nunc 96-well plates (Invitrogen), and plates were frozen at -20 °C until use. When used, the plates were washed using PBST. A blocking step was performed using 2.5% horse serum for 1 hour at 37 °C followed by a wash. The primary antibody panels included positive sera from 50 naturally infected cattle and negative sera from 110 healthy cattle. Multiple serum-dilutions were used ranging from 1/100 to 1/1600.

Serum-dilutions were added as the primary antibody and incubated at room temperature for 1 hour. This was followed by washing with PBST thrice and adding to the wells Peroxidase-conjugated AffiniPure goat anti-bovine IgG, H+L (Jackson ImmunoResearch). The plates were again washed thrice, and the antibody-antigen reactions were detected by adding ABTS substrate. The OD of each sample, in duplicates, was read at 414 nm using an ELISA microplate spectrophotometer, Biotek Epoch (Thermo Fisher Scientific). All blank wells on a plate were similarly treated compared to other wells except that they were not coated with the antigens. The mean OD of blank wells was subtracted from the mean OD value of each sample (from 2 readings).

Fifty positive sera samples from cattle, confirmed to be infected by *M. bovis* infection by postmortem investigations, were obtained from the USDA, NVSL. Fourteen sera from experimentally-infected animals were obtained four months after the intra-tracheal pathogen inoculation as previously described (125). One hundred and ten negative sera samples were obtained from the CFIA. These positive and negative sera were used for the evaluation study. The serological sensitivity and specificity of the assays were determined using receiver operating characteristic (ROC) curves analysis.

3.18 Evaluation of the Recombinant Proteins in Dot Blot Format

Zoom Blot 96-well pates (Vitrozm, USA) were used, according to the manufacturer's recommendations with few modifications. Briefly, 5µl of the antigen (2µg/ml) was immobilized on the nitrocellulose base of each well. After 20 min, the wells are blocked using 50 µl iBind blocking buffer (Life technologies) for 25 min and then probed using 2 µl/ well of the sera from *M. bovis*-infected animals at a1:1000 dilution. The sera were used as outlined in Section 3.17. For the secondary antibody, 2 µl of Peroxidase-AffiniPure goat

anti-bovine IgG (H+L) (Jackson ImmunoResearch) was used at a 1:1000 dilution. The plate was rinsed as recommended, followed by the addition of 10 µl of the Clarity ECL substrate (Bio-Rad). The chemiluminescent signal was detected directly by putting the plate in the ChemiDoc Imager (BioRad). Images were saved as a ".tif" file and analyzed using the ImageJ software. The mean signal from two antibody-antigen reactions was background corrected by subtracting the mean signal of two "no antigen" reactions. The corrected signals were used to construct the ROC curve analysis.

3.19 Statistical Methods

Statistics were performed using IBM Statistic Product and Service Solutions (SPSS) Statistics version 21 (IBM, IL, USA). The program was also used to perform ROC curve analysis as well as to draw the bar charts and the box plots. A comparison between two groups was done using the independent samples t-test for the microarray data. The results are expressed as means ± standard deviation (S.D.). The independent samples t-test for parametric statistics and Mann-Whitney U test for non-parametric statistics were used for comparison between two groups for the ELISA results. A comparison between 3 or more groups was done by the Kruskal-Wallis H test for non-parametric statistics. p-value is significant at $p<0.05$. The heat map was drawn using Graph-Pad version 8 (CA, USA).

Chapter 4. Identification of Candidate Proteins for M. bovis Diagnostics Using Bioinformatics and Proteomic Detection Methods

4.1 Introduction

BTB is a chronic zoonotic infectious disease that mainly affects cattle. In addition, it can affect all mammals including wildlife and contributes to a significant economic loss (56, 57, 130). Attention has been given to the identification of immunogenic proteins of *M. bovis* origin that can be used for diagnostic reagent development. The choice of antigens critically influences the performance of the serological assays used for the diagnosis of *M. bovis* infection. For example, the use of the specific antigen, MPB83, in the serological test was presumed to be useful in the early diagnosis of the disease (66). Also, MPB70 and MPB83 antigens used in ELISA are specific but have low sensitivity (18). ESAT-6 and CFP-10 are antigens that beneficially improve the specificity of the IFN-γ assay (101).

Consequently, *M. bovis* protein antigens were explored for the improvement of BTB diagnostics. Special consideration was given to the unique characters of the *M. bovis* bacterial cell wall. The bacterial cell wall consists of multilayered peptidoglycan, similar to the Gram-positive bacteria. Additionally, the cell wall lipids are distinctively organized in an outer membrane, similar to the Gram-negative bacteria (38, 39). For those reasons, *M. bovis* is regarded as Gram-positive bacteria with an outer membrane.

To identify potentially immunogenic proteins, one approach is to identify the extracellular or outer membrane proteins encoded by the *M. bovis* genome. These proteins

are more likely to stimulate the immune response based on their localization. The mentioned microbial protein-antigens can activate B cells directly in the absence of T-cell help through direct interaction of the antigens with the B-cell receptor. The ability of B cells to respond directly to these antigens provides a rapid response to many important bacterial pathogens (131). Currently, being in the era of whole-genome sequencing, the complete genome sequences of *M. bovis* isolates are available in the public database. These genome sequences can be analyzed by specific bioinformatic analysis for the detection of surface and extracellular proteins encoded by *M. bovis* genomes.

Another approach to identifying potentially immunogenic protein candidates is the proteomic analysis of the culture supernatant of *M. bovis*. Researchers have worked on the analysis of the Purified Protein Derivative (PPDB), a standardized preparation of the culture filtrate of the *M. bovis* AN5 strain (106). The PPDB is used to assess the cell-mediated immune response to infection with *M. bovis*, being the reagent deployed in both the Tuberculin skin test and the IFN-γ test, the major diagnostic tests of BTB (7, 9). To understand the mechanism by which PPDB induces the cell-mediated immune response, researchers analyzed the proteomic content of the PPDB. This was done to identify the protein components of the PPDB that contribute to the development of the delayed hypersensitivity reaction in the Tuberculin skin test or produce a reaction in the IFN-γ assay (11, 12, 132). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of PPDB followed by immunoblotting using sera from infected animals were also attempted to identify immunogenic *M. bovis* proteins (125).

Analysis of the PPDB has allowed the identification of some of the proteins that are secreted in early phases of bacterial growth including ESAT-6 and CFP-10. These two

proteins were found to be highly immunogenic proteins (58, 59, 101). It is worth mentioning that the different proteomic studies analyzing PPDB report a majority of cytoplasmic proteins, which may be related to the heat killing procedure applied. For instance, Borsuk *et al.* (13) report a majority of the proteins that are cytoplasmic (77.9%) and to a less extent involved in intermediary metabolism and respiration (24.25%).

This study was undertaken to select the protein candidates with the aid of the bioinformatics and proteomics tools. First is the use of Bioinformatics to identify potential extracellular or outer membrane proteins encoded by the *M. bovis* genome and the second is through the identification of the protein antigens by proteomic studies. For the second approach, literature review of the previous proteomic studies was attempted to identify potential candidate proteins. Studying the literature revealed that data is lacking regarding the proteomic analysis of PPD from North American *M. bovis* strains (11, 13, 132). Also, most of the contents in PPDB are cytoplasmic proteins (11, 13). This triggered the need to conduct a proteomic study using one of the Canadian field strains, *M. bovis* isolate 2011/0690. The culture filtrate was collected at three different time points and heat-killing was circumvented. This culture supernatant was analyzed to identify possible proteins that may be secreted by the bacteria in different stages of its growth.

4.2 Results and Discussion

4.2.1 Identification of Candidate Proteins for *M. bovis* Diagnostics Using Bioinformatics

The genome sequence of a reference strain of *M. bovis*, *M. bovis* AF2122/97, retrieved from the public database, was submitted for analysis with the bioinformatics tool PSORTb (122). This online tool was used to execute the computational prediction of the

subcellular localization of different proteins encoded by the genome (121). PSORTb version 3.0, used in this study, allowed the analysis of the mycobacterial species in an advanced search option under the category "Gram-positive with an outer membrane". The tool analyzed the genome and out-turned the final localization prediction with high prediction score values based on the combined results of multiple incorporated algorithms. The algorithms include the combination of BLAST-P, detection of specific motifs, Support vector machine, and Signal P.

The predicted proteins from the analysis were submitted to PROTEIN-BLAST to identify the homologous proteins in *M. bovis* isolate 2011/0690. The results obtained from the analysis of the genome of *M. bovis* AF2122/97 are shown in Table 4-1 and Figure 4-1. Seventy-seven proteins were predicted to have their final localization as extracellular (EC) or outer membrane (OM) proteins and were selected as candidates in Table 4-2. These proteins were also checked for homologues in the *M. tuberculosis* (MTB) database (Table 4-2).

Table 4-1. The results of the final localization prediction of *M. bovis* AF2122/97 using PSORTb.

Predicted localization	Protein count	Percentage
Cytoplasmic	1652	42.16%
Cytoplasmic membrane	770	19.65%
Extracellular	70	1.79%
Outer membrane	7	0.18%
Periplasmic	72	1.84%
Unknown	1347	34.38%
Total	3918	100.00%

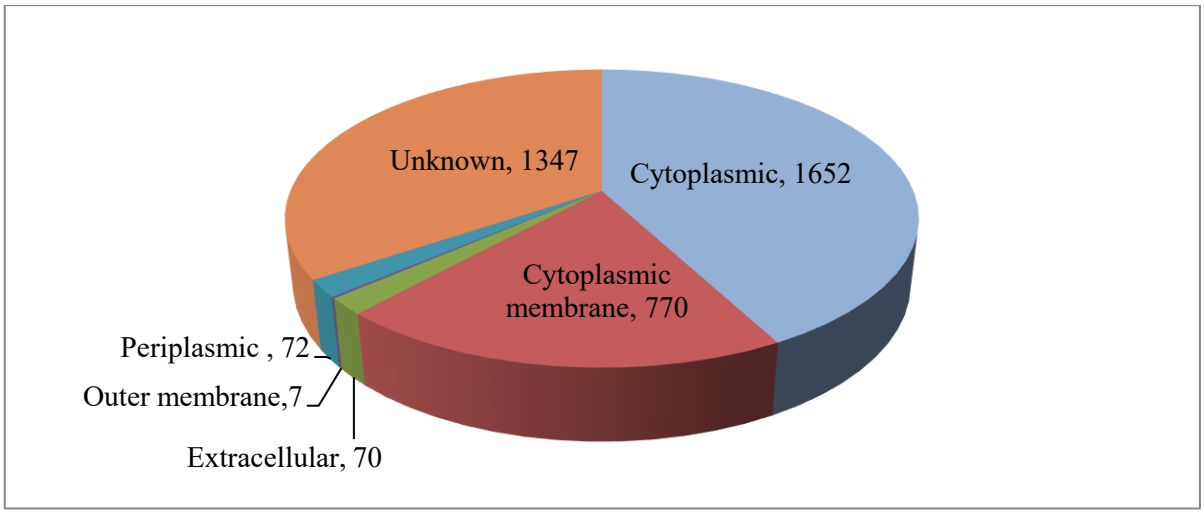


Figure 4-1. Pie chart showing the breakdown of the final localization prediction after analysis of *M. bovis* AF2122/97 genome using PSORTb version 3.

Table 4-2. Locus tag of the 77 proteins predicted to have a final localization as extracellular (EC) or outer membrane (OM), the homologues in *M. tuberculosis*, and supporting literature.

*Strain ID	Mb #	Gene size (bp)	final localization	Rv#	MTB reference
GNBNAFKM_01274	Mb0041c	993	EC	Rv0040c	(69)
GNBNAFKM_01297	Mb0064	1440	EC	Rv0063	(69)
GNBNAFKM_00514	Mb0129	1617	EC	Rv0124	-
GNBNAFKM_00456	Mb0184	735	EC	Rv0178	(133)
GNBNAFKM_00352	Mb0285c	2607	EC	Rv0278c	(69)
GNBNAFKM_00352	Mb0286c	2634	EC	Rv0278c	(69)
GNBNAFKM_00351	Mb0287c	2496	EC	Rv0279c	-
GNBNAFKM_00332	Mb0305	1821	EC	Rv0297	-
GNBNAFKM_01718	Mb0368	828	EC	Rv0361	(133)
GNBNAFKM_01658	Mb0426	1503	EC	Rv0418	(134)
GNBNAFKM_00827	Mb0545	1746	EC	Rv0532	-

GNBNAFKM_00779	Mb0593c	3921	EC	Rv0578c	-
GNBNAFKM_00676	Mb0690	843	EC	Rv0671	-
GNBNAFKM_00599	Mb0768	2730	EC	Rv0747	-
GNBNAFKM_00590	Mb0777c	1938	EC	Rv0755c	-
GNBNAFKM_02091	Mb0845c	2055	EC	Rv0822c	-
GNBNAFKM_02076	Mb0856	2322	EC	Rv0833	-
GNBNAFKM_02074	Mb0858	645	EC	Rv0835	-
GNBNAFKM_02035	Mb0896c	1827	EC	Rv0872c	-
GNBNAFKM_01923	Mb1002	2772	EC	Rv0977	(135)
GNBNAFKM_04091	Mb1006c	1374	EC	Rv0980c	-
GNBNAFKM_03519	Mb1018c	333	EC	Rv0991c	-
GNBNAFKM_03545	Mb1044c	759	EC	Rv1016c	(136)
GNBNAFKM_03552	Mb1050	732	EC	Rv1022	(133)
GNBNAFKM_03993	Mb1096c	2016	EC	Rv1067c	-
GNBNAFKM_03992	Mb1097c	2292	EC	Rv1068c	-
GNBNAFKM_03487	Mb1116	2325	EC	Rv1087	-
GNBNAFKM_03485	Mb1118	306	EC	Rv1089	(137)
GNBNAFKM_03483	Mb1121	2553	EC	Rv1091	-
GNBNAFKM_02907	Mb1262c	1263	EC	Rv1230c	(134)
GNBNAFKM_02894	Mb1275c	1689	EC	Rv1243c	-
GNBNAFKM_03672	Mb1360c	1812	EC	Rv1325c	
GNBNAFKM_03315	Mb1476c	1476	EC	Rv1441c	(135)
GNBNAFKM_03326	Mb1485c	4227	EC	Rv1450c	(136)
GNBNAFKM_04081	Mb1487c	2388	EC	Rv1452c	-
GNBNAFKM_02782	Mb1575c	2037	EC	Rv1548c	-
GNBNAFKM_01856	Mb1679c	3057	EC	Rv1651c	(138)

GNBNAFKM_04023	Mb1783c	1692	EC	Rv1754c	(139)
GNBNAFKM_04024	Mb1784c	1545	EC	Rv1755c	-
GNBNAFKM_04028	Mb1788	657	EC	Rv1758	-
GNBNAFKM_02276	Mb1831c	1623	EC	Rv1803c	-
GNBNAFKM_03873	Mb2006c	654	EC	Rv1984c	(69)
GNBNAFKM_02192	Mb2108	2166	EC	Rv2082	-
GNBNAFKM_02178	Mb2125c	1476	EC	Rv2099c	-
GNBNAFKM_00971	Mb2323	693	EC	Rv2301	(69)
GNBNAFKM_01062	Mb2410c	465	EC	Rv2389c	(140)
GNBNAFKM_01469	Mb2506c	324	EC	Rv2481c	-
GNBNAFKM_03111	Mb2517c	3453	EC	Rv2490c	
GNBNAFKM_03007	Mb2606c	465	EC	Rv2576c	(69)
GNBNAFKM_03951	Mb2667c	2337	EC	Rv2634c	-
GNBNAFKM_02542	Mb2734	1026	EC	Rv2715	(141)
GNBNAFKM_02536	Mb2740c	2100	EC	Rv2721c	(142)
GNBNAFKM_03396	Mb2878	1848	EC	Rv2853	-
GNBNAFKM_01517	Mb3183c	1770	EC	Rv3159c	-
GNBNAFKM_03611	Mb3402	1881	EC	Rv3367	-
GNBNAFKM_03778	Mb3453c	1035	EC	Rv3419c	-
GNBNAFKM_03760	Mb3469c	1404	EC	Rv3439c	-
GNBNAFKM_03748	Mb3481	789	EC	Rv3451	-
GNBNAFKM_03834	Mb3537	4083,	EC	Rv3507,	-
GNBNAFKM_03835	Mb3538	4383	EC	Rv3508	-
GNBNAFKM_04065	Mb3541	5817	EC	Rv3511	-
GNBNAFKM_03835	Mb3543	2979	EC	Rv3514	-
GNBNAFKM_00272	Mb3562	1221	EC	Rv3532	-

GNBNAFKM_00080	Mb3751	702	EC	Rv3724A	-
GNBNAFKM_00069	Mb3761c	1365	EC	Rv3734c	(141)
GNBNAFKM_00036	Mb3792	690	EC	Rv3766	-
GNBNAFKM_01101	Mb3833c	900	EC	Rv3803c	(69)
GNBNAFKM_01109	Mb3841	1620	EC	Rv3811	-
GNBNAFKM_01183	Mb3909c	2235	EC	Rv3879c	(139)
GNBNAFKM_01195	Mb3921c	1200	EC	Rv3892c	-
GNBNAFKM_01666	Mb0419c	987	OM	Rv0411c	(69)
GNBNAFKM_01951	Mb0975c	999	OM	Rv0950c	-
GNBNAFKM_03236	Mb1397c	663	OM	Rv1362c	-
GNBNAFKM_02737	Mb1530	2253	OM	Rv1493	(139)
GNBNAFKM_01807	Mb1724	945	OM	Rv1698	(143)
GNBNAFKM_01028	Mb2377c	1845	OM	Rv2356c	-
GNBNAFKM_02953	Mb2554c	723	OM	Rv2525c	(144)

Strain ID=locus tag from *M. bovis* 2011/0690, Mb # = locus tag from *M. bovis* AF2122/97, Rv# = locus tag of the homologue in *M. tuberculosis* H37Rv, MTB reference =evidence from the *M. tuberculosis* database

It was recognized that the number of 7 predicted outer membrane proteins is considered low when compared to the number of outer membrane proteins predicted in other bacteria. For example, *Pseudomonas aeruginosa or E. coli* are predicted to have 174 and 113 outer membrane proteins, respectively. One explanation is that PSORTb uses stringent criteria resulting in the identification of a lower number of candidates with a high prediction score (7.5/10 or more). Also, the unique composition of the mycobacterial cell wall might make the prediction of outer membrane proteins more difficult.

Analysis of the 77 proteins showed that 27 out of 77 proteins belonged to the PE-PGRS (polymorphic GC-rich sequence), and 6 out of the 77 belonged to the PPE family **(**Table S5, Appendix 3). The PE-PGRS belongs to the PE family of proteins. The PE and PPE proteins were named as such due to the highly conserved Proline-Glutamate and Proline-Proline-Glutamate residues present near the N-termini of the proteins, respectively. These proteins are considered unique mycobacterial proteins and are particularly abundant in pathogenic mycobacteria such as *M. tuberculosis*. The PE/PPE proteins have been shown to elicit B cell responses due to several factors. First, high-throughput proteomics studies suggest that cell wall/surface localization is a characteristic of several PE/PPE proteins making these proteins directly targeted by the humoral immune response. Another factor is the nature of these proteins; evidence suggests that PE/PPE family members contain a potentially rich source of B cell epitopes, especially the PE PGRS families that possess highly repetitive domains (145).

The rest of the 77 proteins include 6 lipoproteins, 22 hypothetical proteins, 1 protein from the ESAT-6 family of small proteins, EspK, and some enzymes. The results of the analysis were compared with the proteins encoded by the *M. tuberculosis* genome and only 28 protein candidates (<50%) had *M. tuberculosis* homologues reported in the literature to have an extracellular localization (Table 4-2). This illustrates that more research needs to be done to identify mycobacterial extracellular proteins as described previously (38).

Surprisingly, the final localization prediction output by PSORTb did not include the known *M. bovis* proteins, MPB70 and MPB83. After careful observation, these proteins were predicted as extracellular by the ECMV algorithm (provided by PSORTb as described in section 3.2 in Materials and Methods). Consequently, the ECSVM algorithm was used to

expand the list of protein candidates. Initially, 45 additional protein candidates were identified by that algorithm. These were further analyzed using Protein-BLAST to identify proteins that have 50% or more sequence identity conserved within multiple *M. bovis* genomes. Also, they were blasted against the proteins encoded by the genomes of a list of serologically cross-reacting mycobacteria obtained from the CFIA. The list included *M. non-chromogenicum*, *M. intermedium*, and *M. avium complex*. The decision was made to select protein candidates with less than 40 % sequence homology with cross-reacting strains. This resulted in the identification of 19 additional protein candidates by the ECSVM algorithm (Table 4-3). These 19 proteins were cytoplasmic membrane (n = 12), periplasmic (n = 2), and unknown proteins (n = 5). The localization of 4 out of the 19 proteins was reported in the literature (Table 4-3). Eleven of the 19 proteins belong to the PE-PGRS family and six of the 19 belonged to the PPE family (Appendix 3, Table S6).

Table 4-3. Gene locus tags and predicted final localization of 19 proteins identified by the ECSVM.

Strain ID	Mb #	Gene size (bp)	Final localization *	Rv #	MTB reference
GNBNAFKM_00208	Mb3626C	1479	CM	Rv3595c	-
GNBNAFKM_00213	Mb3621C	1755	CM	Rv3590c	-
GNBNAFKM_00531	Mb0113	1491	CM	Rv0109	-
GNBNAFKM_00600	Mb0767	2394	CM	Rv0746	-
GNBNAFKM_01019	Mb2369C	1242	U	Rv2340c	-
GNBNAFKM_01070	Mb2418	1083	CM	Rv2396	-
GNBNAFKM_01176	Mb3903	1107	U	Rv3873	(146)
GNBNAFKM_01633	Mb0450C	1413	U	Rv0442c	-
GNBNAFKM_02111	Mb2186C	1473	P	Rv2162c	-
GNBNAFKM_02273	Mb1828	1968	CM	Rv1800	-
GNBNAFKM_02518	Mb2761	1449	CM	Rv2741	(147)
GNBNAFKM_03022	Mb2622	1641	U	Rv2591	-
GNBNAFKM_03269	Mb1431C	1359	P	Rv1396c	-
GNBNAFKM_03590	Mb3420	2244	CM	Rv3388	-

GNBNAFKM_03872	Mb2005	1677	CM	Rv1983	
GNBNAFKM_03989	Mb1797	1875	CM	Rv1768	-
GNBNAFKM_03804	Mb1228	1182	CM	Rv1196	(148)
GNBNAFKM_04029	Mb1789C	2463	CM	Rv1759c	(69)
GNBNAFKM_01516	Mb3184C	2043	U	Rv3160c	-

Strain ID=locus tag from *M. bovis* isolate 2011/0690, Mb # = locus tag from *M. bovis* AF2122/97, Rv# = locus tag from *M. tuberculosis* H37Rv, MTB reference =evidence from the *M. tuberculosis* database

*CM= Cytoplasmic Membrane, P= Periplasmic, U= Unknown

The predicted proteins were analyzed for the presence or absence of signal peptide. Analysis of the 77 proteins identified by the PSORTb showed that 30 of 77 had a signal peptide and 47 contained no signal peptide (Appendix 3, Table S5). Out of the 19 proteins identified by the ECMV, sixteen had a signal peptide and 3 had no signal peptide (Appendix 3, Table S6). Signal peptides are signals at the N-terminus that help direct proteins to other locations including cellular and extracellular. Proteins with signal peptide can be secreted proteins. However, the protein localization patterns characterized by the signal peptide can be complex (149–151). Thus, the use of the signal peptide algorithm solely was not adopted for the detection of extracellular proteins.

4.2.2. Proteomic Identification of Candidate Proteins for *M. bovis* Diagnostics

M. bovis culture filtrate was collected at three stages, early exponential phase (growth for 10 days), late exponential phase (growth for 21 days), and stationary phase (growth for 6 weeks) and processed as described in Materials and Methods sections 3.3-3.5. The data obtained from the proteomic analysis of culture filtrate samples have been deposited to the ProteomeXchange Consortium via the PRIDE (152) partner repository with the dataset identifier PXD017817. The proteins identified are shown in Table 4-4.

Table 4-4. A summary of proteins detected in the different stages of growth of *M. bovis* strain 2011/0690.

Mb#	locus tag in *M. bovis* isolate 2011/0565	Detection stage			[a]predicted localization	protein name	[b]Added
		Early	Mid	Late			
Mb3904	GNBNAFKM_01177	x	x	x	EC	CFP-10	Y
Mb3905	GNBNAFKM_01178	x	x	x	EC	Early secretory antigenic target(ESAT-6)	Y
Mb2900	GNBNAFKM_03372	x	x	x	EC	MPB70	Y
Mb0448	GNBNAFKM_01635	x	x	x	C/ EC	60 kDa chaperonin 2- GroL2	N
Mb2268	GNBNAFKM_00919	x	x	x	C	Meromycolate extension acyl carrier protein	N
Mb2898	GNBNAFKM_03375	x	x	x	EC/ CW	Cell surface glycolipoprotein MPB83	Y
Mb1918C	GNBNAFKM_02365	x	x	x	EC	Antigen 85 complex B (Ag85B)	Y
Mb2397C	GNBNAFKM_01048	x	x	x	EC	Low molecular weight antigen MTB12	Y
Mb2002C	GNBNAFKM_03867	x	x	-	EC	Antigen MPB64	N
Mb0192A	GNBNAFKM_00447	x	x	x	U	Metallothionein OS	N
Mb1858	GNBNAFKM_02302	x	x	x	CW/ EC	Uncharacterized protein Mb1858	N
Mb1891	GNBNAFKM_02338	x	x	x	EC	Fibronectin attachment protein (FAP-B)	N
Mb2970C	GNBNAFKM_02669	x	x	x	CM/CW/EC	Lipoprotein LppX	Y
Mb3834C	GNBNAFKM_01102	x	x	-	EC	Antigen 85 complex A	Y
Mb0358	GNBNAFKM_01728	-	x	x	C/ECV	Heat shock protein 70 - DnaK	N
Mb1967	GNBNAFKM_02417	-	x	x	C	Thiol peroxidase (TPX)	Y

Mb1662	GNBNAFKM_01872	-	x	-	CW/PM	Iron-regulated universal stress protein family protein TB15.3	Y
Mb0055	GNBNAFKM_01288	-	x	-	EC/ PM	Single-stranded DNA-binding protein	N
Mb3672C	GNBNAFKM_00165	-	x	-	C/ CW	Probable cold shock protein A cspA	N
Mb2057C	GNBNAFKM_02243	-	x	x	CW/EC	14 kDa antigen- HspX-HSP 16.3	Y
Mb3945	GNBNAFKM_01220	-	x	x	C	Thioredoxin/ MPT46	N

[a] C- Cytoplasmic; CW- Cell wall; EC- Extracellular; ECV- Extracellular vesicle; PM- Plasma membrane; S- Secreted; U- Unknown;

[b] Added=Eleven protein candidate selected to be added to the study. Y=Yes; N=No

In total, 21 proteins were detected from different stages of bacterial growth. Proteins reported here were detected in at least two biological repeats. The experiments were done primarily on albumin-undepleted samples and then repeated using an albumin depletion kit. The removal of bovine albumin from the media was to improve the resolution of mass spectrometry (153). Protein depletion caused an increase in the number of detected proteins from (7 to 14 proteins) at 10 days of bacterial growth, (15 to 21 proteins) at 21 days of growth, and (15 to 16 proteins) after 6 weeks of growth.

Out of the 21 proteins detected, 17 (80.9%) were reported or predicted to be secreted or cell wall or extracellular proteins. Eight out of the 17 proteins had signal peptides (Appendix 3, Table S8). The number of proteins detected in the late exponential or the stationary phase is lower than the number reported in other studies (11, 13, 14, 69, 132). This might be because this study used a field strain, whereas the previous studies used laboratory strains, for example, *M. bovis* AN5, and *M. tuberculosis* H37Rv. Another factor may be related to the different media used to culture bacteria. The previous studies use Sauton or Reid media (albumin or protein poor) whereas this study used albumin-enriched media. The albumin-enriched media may have an advantage; it could mimic the composition of serum, in which albumin is a major constituent. Thus, the proteins secreted by *M. bovis* grown in this media may closely resemble what free serum mycobacteria would secrete in infected hosts.

Twelve proteins were consistently detected in the three stages of bacterial growth. Of these proteins, CFP-10 was the most abundant then followed by MPB70, early secretory antigenic target (ESAT-6), and MPB83. Currently, the two proteins MPB70 and MPB83, being abundant proteins identified in the PPDB analysis, are used as antigens in the bovine TB serological test (3, 109). CFP-10 and ESAT-6 have been employed in the IFN-γ assay to

increase specificity and to differentiate BCG-vaccinated from unvaccinated animals. Some countries such as the United Kingdom and New Zealand use these antigens for serial TB testing (3, 101).

In addition to the four above-mentioned proteins, the proteins identified in this proteomic study were similar to those reported by Roperto *et al.* (132) and by Rennie *et al.* (125). The proteins Ag85A and Ag85B but not Ag85C were identified in this study. An explanation could be that a different strain used in this study may secrete a lower level of Ag85C. This was reported in the literature as variation in abundance of Antigen 85 among different *M. tuberculosis* strains (154). Ag85A, Ag85B, and Ag85C play a role in the affinity of mycobacteria for fibronectin and were considered candidates for vaccine development (155).

Furthermore, three heat shock proteins from *M. bovis*, GroL2 (Mb0448), HspX (Mb2057C), and DnaK (Mb0358), were identified in this study. The protein GroL2 was continuously detected from the early exponential phase of growth until the stationary phase. HspX and DnaK were detected in the latter two stages. The homologous counterparts of these proteins (GroEL2, HspX, and DnaK) were reported by Cho *et al.* (14) after the analysis of the PPD from *M. tuberculosis* origin. The latter study showed that these proteins dominated the composition of PPD. These proteins are cytoplasmic and possible T cell antigens, but they were also reported to be extracellular or surface proteins (69, 156, 157). The rest of the proteins from the mid or late phase which may have extracellular localization are, Mb2970C (LppX) that functions in lipid transport and Apa (Mb1891), a protein with a role in extracellular matrix binding (158, 159). Two proteins that may play a role in the development of immune responses, MTB12 (Mb2397C) and MPB64 (Mb2002C), were also

identified (160). MPB64 is a possible plasminogen-binding protein that may have a role in the bacterial virulence (161). This is in addition to the detection of Mb1858; a presumed regulator of glutamate metabolism (157, 162). Two proteins with possible extracellular localization were detected in the late exponential phase but were not detected in the stationary phase, probably due to protein degradation. These two proteins are Mb1662, a stress protein and Mb0055, a single-strand DNA binding protein; both have been reported to play a role in antibiotic response (163).

Moreover, this study identified two proteins that were not detected from the PPDB preparations in the recent studies of Borsuk *et al.*, Cho *et al.* and Roperto *et al.* (11, 13, 132). These first is Mb0192A, a metallothionein of unknown localization detected in three phases of growth. The second is Mb3672C, a probable cold shock protein A, of possible cell wall localization that was detected only in the late exponetial phase of growth. Mb3672C was reported to be of high structural fragility (139, 164).

Eleven proteins were selected from this proteomic study as shown in Table 4-4. Nine of them were reported in the literature to have the diagnostic potential. These proteins are CFP-10, ESAT-6, MPB70, MPB83, Antigen 85 complex B (Ag85B), Thiol peroxidase (TPX), Iron-regulated universal stress protein (TB15.3), HSP 16.3, and Antigen 85 complex A (58, 59, 79, 83, 88, 165). Two proteins, MTB12 and Lipoprotein LppX, were selected based on being uniquely identified in *M. bovis* but not in the cross-reacting *M. avium* species.

4.2.3. Identification of Candidate Proteins for *M. bovis* Diagnostics through Literature Review of Published Proteomic Studies

A review of previous studies on the proteomic analysis of the PPDB was conducted thoroughly to identify potentially immunogenic protein candidates. Proteins commonly expressed by *M. avium* were excluded from the candidate list to increase the antigen specificity for *M. bovis* diagnostics. Also, some proteins were already discovered by the bioinformatics approach or by the present proteomic study. Recent proteomic studies on the PPDB preparations by Borsuk *et al.*, Cho *et al.,* and Roperto *et al.* (11, 13, 132) were reviewed. The identified proteins were blasted against the reference genome of *M. bovis* AF2122.97 and against the genome of *M. bovis* strain 2011/0690 to identify the homologous locus tags in both latter strains. Sixty-eight proteins were identified and added to the list of protein candidates. As reported in the MTB database, 11 candidates out of the 68 proteins had an extracellular localization, 5 had probable extracellular or cell wall localization and 29 had cell wall but not extracellular localization (Table 4-5).

Table 4-5. Sixty-eight proteins detected from the review of previous proteomic literature.

RV#	Mb #	strain ID	protein identity	Localization*
Rv0242c	Mb0248c	GNBNAFKM_00389	3-oxoacyl-[acyl-carrier protein] reductase	U
Rv0243	Mb0249	GNBNAFKM_00388	acetyl-CoA acyltransferase	CW/ PM/ C
Rv0248c	Mb0254c	GNBNAFKM_00382	succinate dehydrogenase	CW/ PM
Rv0270	Mb0276	GNBNAFKM_00360	fatty-acid-CoA ligase	CW/ PM
Rv0333	Mb0340	GNBNAFKM_00295	hypothetical protein	U
Rv0431	Mb0439	GNBNAFKM_01644	unknown	CW
Rv0462	Mb0471	GNBNAFKM_01612	dihydrolipoamide dehydrogenase	C/ PM/ EC
Rv0467	Mb0476	GNBNAFKM_01607	isocitrase: glyoxylate bypass	C/ PM/ EC
Rv0469	Mb0478	GNBNAFKM_01605	mycolic acid synthase	C/PM
Rv0475	Mb0485	GNBNAFKM_01598	heparin-binding hemagglutinin HBHA	CW/ PM/ C
Rv0503c	Mb0515c	GNBNAFKM_01567	cyclopropane-fatty-acyl-phospholipid synthase 2	C/PM
Rv0569	Mb0584	GNBNAFKM_00789	Conserved hypothetical protein	CW/C
Rv0577	Mb0592	GNBNAFKM_00780	hypothetical	EC/PM
Rv0583c	Mb0598c	GNBNAFKM_00774	lipoprotein LpqN	CW/ EC/PM
Rv0632c	Mb0649c	GNBNAFKM_00721	enoyl hydrase	U
Rv0639	Mb0658	GNBNAFKM_00709	transcription antitermination protein	CW/ PM/ C
Rv0703	Mb0723	GNBNAFKM_00643	50S ribosomal protein L23	C/PM
Rv0707	Mb0727	GNBNAFKM_00639	30S ribosomal protein	CW/PM
Rv0719	Mb0740	GNBNAFKM_00627	50S ribosomal protein L6	CW/PM
Rv0733	Mb0754	GNBNAFKM_00613	adenylate kinase	C/PM/CW
Rv0761c	Mb0784c	GNBNAFKM_00582	Zinc-containing alcohol dehydrogenase	C/PM
Rv0801	Mb0824	GNBNAFKM_04049	hypothetical	-
Rv0831c	Mb0854c	GNBNAFKM_02082	hypothetical protein	C/PM/CW
Rv0884c	Mb0908c	GNBNAFKM_02023	phosphoserine aminotransferase	C/EC/PM
Rv0896	Mb0920	GNBNAFKM_02009	type II citrate synthase	C/PM
Rv0905	Mb0929	GNBNAFKM_02000	enoyl-CoA hydratase	CW/PM
Rv0932c	Mb0956c	GNBNAFKM_01968	phosphate-binding lipoprotein	PM/EC
Rv0984	Mb1010	GNBNAFKM_03511	pterin-4-alpha-carbinolamine dehydratase	PM
Rv1070c	Mb1099c	GNBNAFKM_03504	enoyl-CoA hydratase	CW/PM
Rv1080c	Mb1109c	GNBNAFKM_03494	transcription elongation factor	CW/ PM/ cytosol
Rv1093	Mb1123	GNBNAFKM_03481	serine hydroxymethyltransferase 1	CW/ PM/ cytosol

Rv1133c	Mb1164c	GNBNAFKM_03438	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	CW/ PM/ cytosol
Rv1181	Mb1213	GNBNAFKM_03721	polyketide beta-ketoacyl synthase	PM
Rv1270c	Mb1301c	GNBNAFKM_02865	lipoprotein LPRA	CW/EC/PM
Rv1352	Mb1387	GNBNAFKM_03644	hypothetical protein	EC
Rv1826	Mb1857	GNBNAFKM_02301	glycine cleavage system protein H	U
Rv1392	Mb1427	GNBNAFKM_03265	S-adenosylmethionine synthetase	CW/ PM/ cytosol
Rv1630	Mb1656	GNBNAFKM_01879	ribosomal protein S1	CW/ PM/ cytosol
Rv1915	Mb1951c	GNBNAFKM_02401	PPE family protein PPE34	U
Rv1916	Mb1950	GNBNAFKM_02399	isocitrate lyase	U
Rv2140c	Mb2164c	GNBNAFKM_02136	hypothetical	EC/PM
Rv2215	Mb2238	GNBNAFKM_00886	dihydrolipoamide acetyltransferase	CW/ PM/ C
Rv2220	Mb2244	GNBNAFKM_00892	glutamine synthetase	CW/ PM/ C/EC
Rv2241	Mb2265	GNBNAFKM_00916		CW/EC/PM
Rv2246	Mb2270	GNBNAFKM_00921	3-oxoacyl-[acyl-carrier protein] synthase 2	CW/ PM/ C
Rv2557	Mb2587	GNBNAFKM_02987	hypothetical protein	U
Rv2593c	Mb2624c	GNBNAFKM_03024	Holliday junction ATP-dependent DNA helicase RuvA	PM
Rv2623	Mb2656	GNBNAFKM_03963	Universal stress protein	CW/ PM/ C
Rv2626c	Mb2659c	GNBNAFKM_03960	Hypoxic response protein 1	CW/ PM/ C/EC
Rv2779c	Mb2801c	GNBNAFKM_02481	DNA-binding transcriptional activator DecR	C
Rv2787	Mb2810		ESX-1 secretion-associated protein EspI	CW
Rv2889c	Mb2913c	GNBNAFKM_03359	Elongation factor Ts	CW/ PM/ C
Rv2940c	Mb2965c	GNBNAFKM_02675	Mycocerosic acid synthase	U
Rv3001c	Mb3026c	GNBNAFKM_02608	Ketol-acid reductoisomerase (NADP(+))	PM/C
Rv3045	Mb3071	GNBNAFKM_03213	NADP-dependent alcohol dehydrogenase C	CW/ PM/ C
Rv3048c	Mb3074c	GNBNAFKM_03210	Ribonucleoside-diphosphate reductase subunit beta nrdF2	C
Rv3196A	Mb3219c	GNBNAFKM_01478	hypothetical protein	U
Rv3354	Mb3389	GNBNAFKM_03624	Putative lipoprotein LprJ	EC
Rv3392c	Mb3424c	GNBNAFKM_03586	Cyclopropane mycolic acid synthase 1	C/PM
Rv3417c	Mb3451c	GNBNAFKM_03780	60 kDa chaperonin 1	C/PM/CW

Rv3443c	Mb3473c	GNBNAFKM_03756	50S ribosomal protein L13	C/PM/CW
Rv3457c	Mb3486c	GNBNAFKM_03743	DNA-directed RNA polymerase subunit alpha	C/PM/CW
Rv3458c	Mb3487c	GNBNAFKM_03742	30S ribosomal protein S4	C/PM/CW
Rv3663c	Mb3687c	GNBNAFKM_00149	Glutathione import ATP-binding protein GsiA	U
Rv3722c	Mb3749c	GNBNAFKM_00084	Putative aminotransferase	EC
Rv3797	Mb3826	GNBNAFKM_00001	Putative acyl-CoA dehydrogenase AidB	U
Rv3841	Mb3871	GNBNAFKM_01140	Ferritin BfrB	CW/ PM/ C/EC
Rv3846	Mb3876	GNBNAFKM_01149	Superoxide dismutase	C/P/PM/EC

Strain ID=locus tag from *M. bovis* isolate 2011/0690, Mb# = locus tag from *M. bovis* AF2122/97, Rv# = locus tag from *M. tuberculosis* H37Rv

*C= Cytoplasmic, CW=Cell wall, EC= Extracellular, PM= Plasma membrane, U= Unknown

4.2.4 Known *M. bovis* Antigens Added to the Candidate List

Thirteen of the known *M. bovis* antigens reported in the literature were added to the list of candidates (Table 4-6). These proteins served as control candidate proteins.

Table 4-6. Thirteen *M. bovis* known antigenic candidates.

Protein name	Strain ID	Mb #	Protein size (kDa)
Ag85C	GNBNAFKM_00509	Mb0134c	36.7
PE5	GNBNAFKM_00344	Mb0293	9.5
Esxl	GNBNAFKM_04088	Mb1230	10
Pe13	GNBNAFKM_03735	Mb1227	9.8
TB10.4(esxH)	GNBNAFKM_00341	Mb0296	10.4
Espc	GNBNAFKM_00192	Mb3645c	10.7
TB7.3	GNBNAFKM_01446	Mb3247c	7.3
MPB53	GNBNAFKM_03369	Mb2903c	18.3
Mb0130	GNBNAFKM_00513	Mb0130	53.4
PPE14	GNBNAFKM_01988	Mb0939c	41.4
Esxg	GNBNAFKM_00342	Mb0295	9.7
Rv3740c	GNBNAFKM_00064	Mb3766c	48
Mb1207c	GNBNAFKM_03715	Mb1207c	10.8

Strain ID=locus tag from *M. bovis* isolate 2011/0690, Mb# = locus tag from *M. bovis* AF2122/97

In summary, the final complete list of 188 protein candidates comprised 96 candidates identified from the bioinformatics, 79 candidates from the proteomic studies and 13 known *M. bovis* antigens.

Chapter 5. PCR Procedures to Amplify GC-Rich DNA Sequences of Mycobacterium bovis

As a disclaimer, the work presented in this chapter has been published online (https://doi.org/10.1016/j.mimet.2020.106121) (166).

5.1. Introduction

The development of the polymerase chain reaction (PCR) amplification of DNA has been considered a revolution in molecular biology since its discovery (167). It is an indispensable tool in both medical and biological research fields. Applications of PCR include the diagnosis of infectious diseases, molecular genetic analysis, and the creation of recombinant DNA constructs for protein expression purposes (168, 169). Despite being a straightforward technique, amplification of target DNA sequences of high GC content (>60%) can be an obstacle to a successful PCR. GC-rich sequences are present in important regulatory elements of human DNA including promoters and enhancers (170), in addition to some bacterial genomes characterized by high GC content such as *Mycobacteria* and *Pseudomonas* (171, 172). Cloning genes of high GC content by PCR using long primers can pose an extra challenge to the amplification process because of the high melting temperature (*Tm*), added to the possibility of forming secondary structures such as hairpin, self and cross dimer formation (173).

Numerous studies aiming to overcome problems associated with PCR amplification of high GC sequences are documented in the literature. Special consideration is given to the primers, enhancer solutions, and cyclic conditions. It is recommended to keep the length of the primers between 15 to 30 nucleotide residues (bases) and to have the *Tm* between 52-58

°C. Moreover, di-nucleotide repeats, formed by “G” and “C”, should be avoided to prevent hairpin structure formation (173). Designing such primers can be achieved using a commonly used program, Primer3 (http://primer3.ut.ee/). The addition of enhancers can modify the melting characters of the double-stranded DNA helices allowing easy separation of the strands. The most commonly described enhancers in the literature are betaine, dimethyl sulfoxide (DMSO), and formamide. Betaine is an amino acid analogue that decreases the energy required for DNA strands denaturation. DMSO interferes with the hydrogen bond formation preventing inter- and intra-strand reannealing. However, DMSO has the drawback of reducing the activity of Taq polymerase making it important to consider maintaining a balance between template accessibility and the polymerase activity (174–177). Formamide increase PCR specificity when working with GC-rich targets (178). Modification of cyclic conditions is also considered; this includes optimizing the annealing temperature, applying Hot start PCR, and using nested PCR and touchdown PCR (179). Touchdown PCR (TD-PCR) uses an annealing temperature higher than the expected annealing temperature. This higher temperature is decreased by 1-2°C every cycle usually for ten cycles until the desired annealing temperature is reached. TD-PCR makes benefit from the exponential nature of PCR reactions giving an advantage of 2-fold per cycle for the production of the desired product (179, 180). The use of TD-PCR has been described to overcome multiple issues including difficulties in the amplification of high GC content targets (181).

It is noted that even with adopting the previously mentioned approaches, amplification challenges still exist especially when using GC-rich templates, lengthy primers, and aiming to amplify longer products. Moreover, most of the published literature focuses on the amplification of genes less than 1 kilobase and optimization of the conditions relating to

a single PCR product (171, 174, 175, 182, 183); however, they do not address working with multiple lengthy GC-rich targets (>1kb). In this project, for the amplification and cloning of more than 100 GC-rich ORFs from the *M. bovis* genome, a standard procedure that enables amplification of multiple targets simultaneously was required without the need to optimize individual conditions for each target. This provoked us to design this study that tested and compared the PCR amplification of one "difficult" target as a model, *Mb0129* from *M. bovis*, a large-sized gene 1794 bp and a GC content 77.5%, and two relatively "easier" targets, *mpb83 (Mb2898)* from *M. bovis*, a 663 bp gene and 63% GC content, and *LMHCC_RS00060* from *L. monocytogenes*, a 1617 bp gene and 41.5 % GC.

5.2 Results and Discussion

Difficulties were experienced in our attempts to amplify various ORF targets in the *M. bovis* genome by PCR for the purpose of cloning. Most of the ORFs targets are more than 1 kb in size and have a GC content of more than 65% which presumably hindered PCR amplification of the *M. bovis* ORF targets with DNA polymerases under a conventional PCR cycling. Unsuccessful attempts were experienced with Taq polymerase, One*Taq* DNA Polymerase (NEB, M0480S), Platinum™ Pfx DNA Polymerase (Invitrogen, 11708039), Expand Long Template PCR System (an enzyme mix that contains thermostable Taq DNA Polymerase and a thermostable DNA polymerase with proofreading activity, Roche, 11681834001). This observation prompted the establishment of a reliable PCR procedure that not only can be used for cloning the *M. bovis* sequences but also can be applied to the amplification of other targets of high GC content. Here, several PCR protocols were evaluated and compared for the amplification of three selected ORFs of high, moderate, and low GC content. These ORFs are *Mb0129* (a gene from *M. bovis* composed of 1794 bp with

a GC content of 77.5%), *mpb83* (a smaller gene of 663 bp with a GC content of 63.0% from *M. bovis*), and *LMHCC_RS00060* (a large gene of 1617 bp with a GC content 41.5% from *L. monocytogenes*). Four high fidelity enzymes were chosen for use in the PCR protocols, based on their ability to amplify high GC-content genes as indicated by the suppliers, in addition to the widely used Taq polymerase. PCR amplification of the three selected targets was attempted with each of the 5 enzymes described in the Materials and Methods, section 3.6.2. Three PCR protocols, 3St, 2St, and TD protocols, were designed as per the instructions of the enzyme manufacturers or as described by Don *et al.* (181) and tested with each of the 5 selected DNA polymerases. Given that enhancers such as DMSO have been recognized for having the ability to improve the PCR performance (174, 175), we also investigated the effect of the enhancers on PCR amplification of the GC-rich gene targets. A summary of PCR amplification results of the three targets is illustrated in Table 5-1.

Table 5-1. Summary of the PCR results from the amplification of three target genes by using 5 different DNA polymerases in the three protocols.

Enzyme	Protocol	*mpb83*		*Mb0129*		*LMHCC_RS00060*	
		Enhancer added [a]		Enhancer added [a]		Enhancer added [a]	
		N	Y	N	Y	N	Y
Taq	3 St	-	+	-	-	-	+
	2 St	-	+	-	-	-	+
	TD	-	+	-	-	-	+
Q5	3 St	+	+	-	+	+	-
	2 St	+	+	-	+	+	-
	TD	+	+	-	-	+	+
Platinum SuperFi	3 St	+	+	-	+	+	+
	2 St	+	+	-	+	+	-
	TD	+	+	-	-	+	+
Prime STAR GXL	3 St	+	+	-	+	+	+
	2 St	+	+	-	+	+	+
	TD	+	+	-	-	+	+
Phusion	3 St	+	+	-	-	+	weak
	2 St	+	+	-	weak	+	-
	TD	+	+	-	-	+	weak

[a] N refers to reactions done without the added enhancer, Y refers to those with the enhancer added

+ refers to a successful amplicon production, - refers to no amplicon production

5.2.1 PCR Amplification using Taq DNA Polymerase

Taq is the most commonly used DNA polymerase in PCR for amplification of a target DNA in laboratories. With Taq DNA polymerase, three PCR protocols (Materials and Methods, section 3.6.5) were designed and evaluated for amplifying selected ORFs of low, moderate, and high GC contents. Both *mpb83* (63% GC) and *LMHCC_RS00060* (41.5% GC) were successfully generated by employing the three protocols in the presence of DMSO (an enhancer) at 5% and 2.5%, respectively. For *Mb0129* (77.5% GC), all of the three protocols failed to produce the amplicons in the presence of 5% DMSO (Figure 5-1A). Without DMSO, Taq failed to amplify target products for all three genes, *mpb83, LMHCC_RS00060*

and *Mb0129* (Figure 5-1B) suggesting that an enhancer is required for the PCR reactions. One benefit of using an enhancer in PCR is to prevent the formation of secondary structures in long primers. These results revealed the difficulty of amplifying a larger gene by Taq, especially a target of more than 1 kb in size and higher GC content (> 65%). The touchdown protocol drastically improved the yield of both *mpb83* and *LMHCC_RS00060* amplification; however, it still failed to produce the *Mb0129* amplicon (Figure 5-1A).

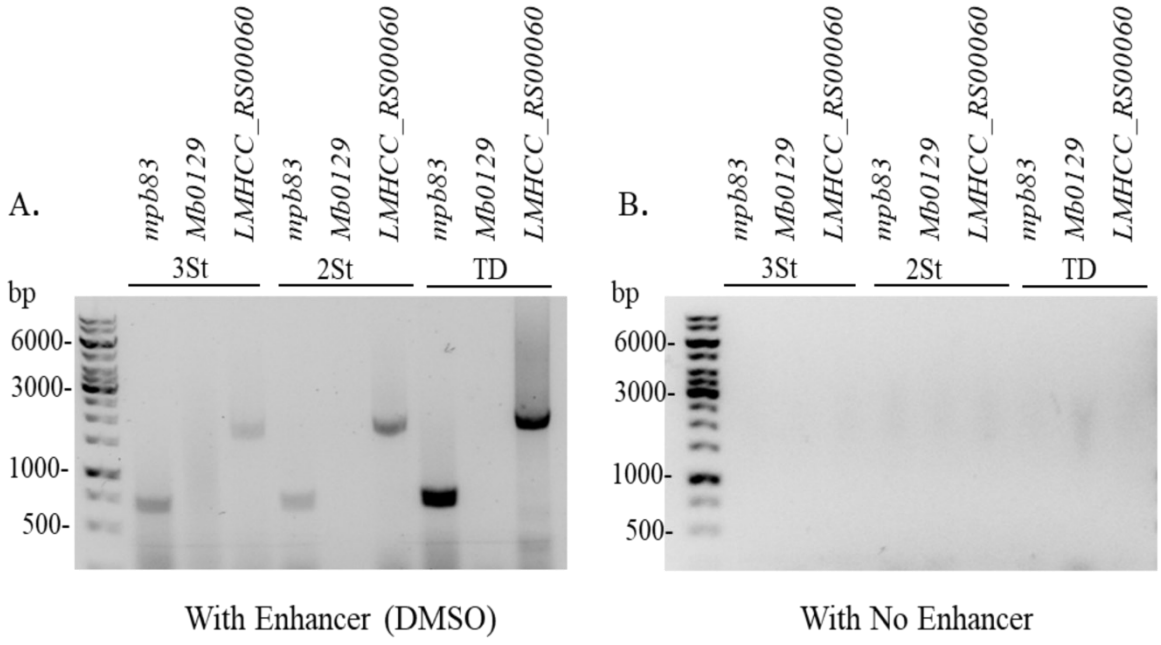


Figure 5-1. Amplification results of three target genes using Taq polymerase.
(A) Attempted amplification of the three DNA targets (*mpb83*, *Mb012,* and *LMHCC_RS00060*) sequences by Taq in the presence of DMSO using 3St, 2St, and TD protocols; *mpb83* and *LMHCC_RS00060* successfully showed amplicons using the three protocols. *Mb012* failed to show an amplicon. DMSO was used at 5% for *mpb83* and *Mb0129* and 2.5% for *LMHCC_RS00060.* (B) Attempts of amplification of the three DNA target sequences in the absence of DMSO using 3St, 2St, and TD protocols showing no amplicon.

5.2.2 The Superiority of Q5 DNA Polymerase over Phusion in the Synthesis of *Mb0129*

Phusion and Q5 DNA polymerases were selected for use in the PCR protocols based on their high fidelity and suitability for amplification of high GC-content DNA templates and were used in the three protocols. Both enzymes were able to generate the *mpb83* amplicon in all the reactions of three PCR protocols with and without the presence of DMSO (Figure 5-2A). Despite the anticipated capacity of Phusion to amplify higher GC-content DNA sequences, two of the three protocols (3St and TD) failed to generate a detectable PCR product from the *Mb0129* template with this enzyme, while a 2St PCR reaction weakly generated the *Mb0129* amplicon in the presence of DMSO. In contrast to Phusion, Q5 allowed the synthesis of *Mb0129* in good yield from the 2St reaction with the addition of the enhancer and produced a much lower yield of the amplicon in the 3St PCR reaction (Figure 5-2B). However, non-specific bands were produced by both 3St and 2St PCR using Q5 in *Mb0129* amplification. These bands may be due to mispriming in PCR. The low GC-content gene, *LMHCC_RS00060*, was successfully amplified by using the three protocols with either Phusion or Q5. PCR amplification of this gene appeared to be better without adding the enhancer DMSO in the reaction (Figure 5-2C). This may be attributed to the relatively low GC content of *LMHCC_RS00060*. These results demonstrated that Phusion worked well for the amplification of a shorter target sequence of moderate GC content in the presence of DMSO but not for a longer GC-rich sequence. Q5 DNA polymerase, when used in the 2St PCR protocol, was superior over Phusion in amplifying a longer DNA sequence of high GC-content.

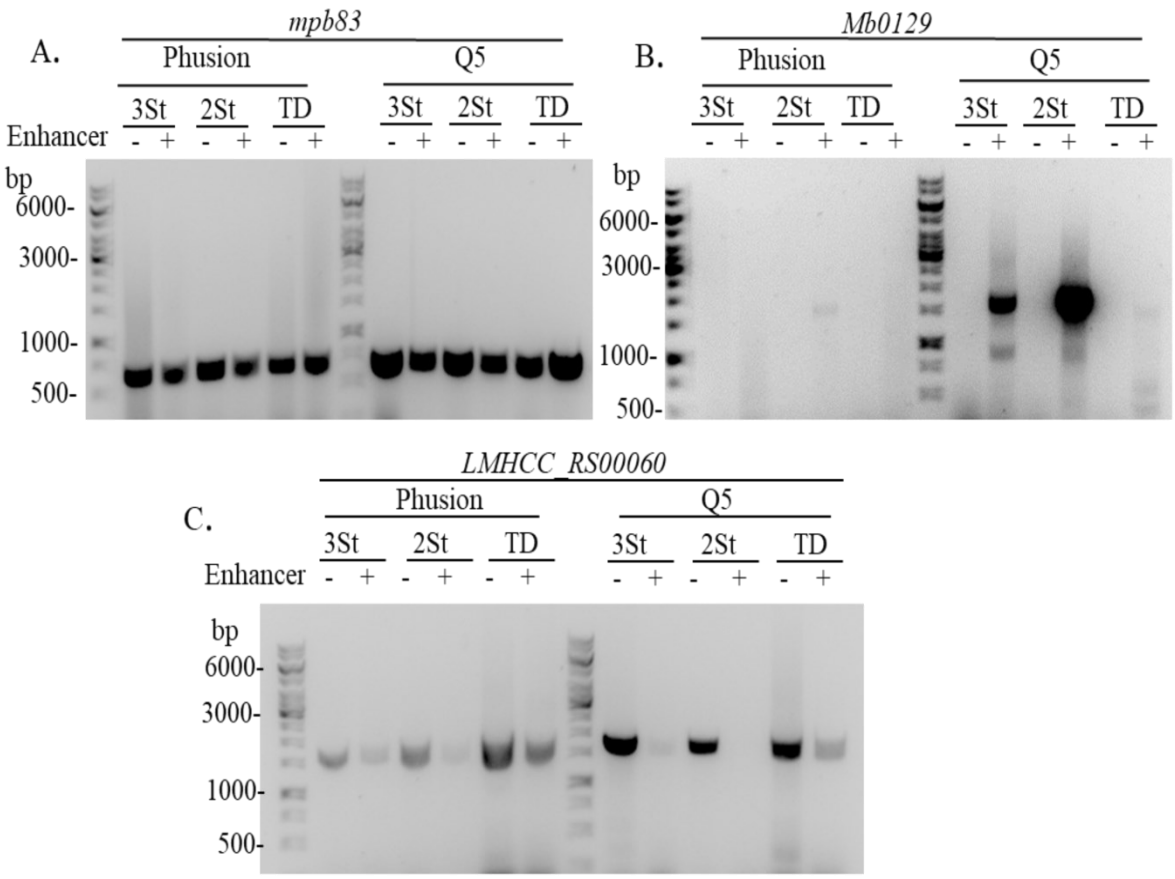


Figure 5-2. Results for attempted amplification of the three target genes using Phusion and Q5 polymerase.

(A) *mpb83* is readily amplified using both Phusion and Q5 polymerase using the three different protocols. (B) *Mb0129* is only amplified using the Q5 enzyme using the 2St and 3St protocols in the presence of the enhancer. (C) *LMHCC_RS00060* shows amplicons with both enzymes in the three protocols with no need for the enhancer. Products are shown in the presence (+) and absence (-) of the enhancer. *mpb83* with added enhancer in 3St PCR is used as a positive control. *Mb0129* with no enhancer in 3St PCR is used as a negative control.

5.2.3 The Successful Synthesis of *Mb0129* by using PrimeSTAR GXL and Platinum SuperFi Polymerases

All the three PCR protocols with PrimeSTAR GXL DNA polymerase were able to amplify *mpb83* in all attempts. With this enzyme, both the 3St reaction and the 2St reaction produced a good yield of the *Mb0129* amplicon in the presence of 5% DMSO (Figure 5-3A). Similarly, Platinum SuperFi polymerase was capable of amplifying *mpb83* by using all three PCR protocols. *Mb0129* was amplified in the presence of the enhancer via the 2St reaction which gave a higher yield than the 3St PCR reaction (Figure 5-3B). With no requirement for the enhancer, *LMHCC_RS00060* amplicons were generated using both enzymes. The PrimeSTAR GXL DNA polymerase, however, demonstrated a better performance than the Platinum SuperFi polymerase (Figure 5-3C, and Figure 5-3D).

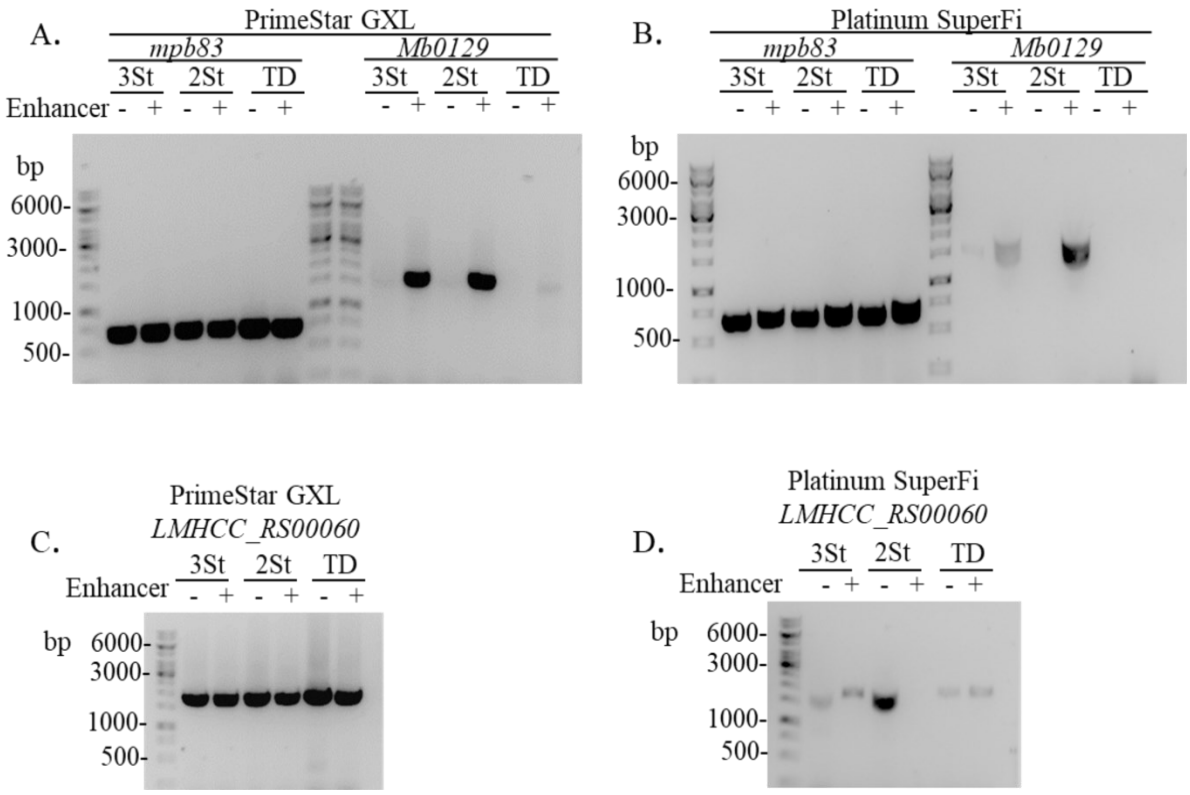


Figure 5-3. Results for attempted amplification of the three target genes using PrimeSTAR GXL and Platinum SuperFi polymerase in the three protocols.

(A) With PrimeSTAR GXL polymerase, *mpb83* is readily amplified by using the three different protocols whereas *Mb0129* is only amplified with the 2St and 3St protocols in the presence of the enhancer. (B) Amplification of *mpb83* and *Mb0129* with Platinum SuperFi polymerase show similar results as in panel A with a lower yield of *Mb0129* amplicon in the 3St reaction. *LMHCC_RS00060 are successfully amplified using* PrimeSTAR GXL polymerase (C) and Platinum SuperFi polymerase (D) without the need for the enhancer. The presence (+) and absence (-) of the enhancer in PCR reactions are indicated. *mpb83* with added enhancer in 3St PCR is used as a positive control. *Mb0129* with no enhancer in 3St PCR is used as a negative control.

5.2.4 Comparison of the Three PCR Protocols

Examination of the results obtained with the 3St, 2St, and TD PCR protocols revealed that *mpb83*, the short gene (<1kb) with 63% GC content, and *LMHCC_RS00060*, the long gene (>1.5 kb) with 41.5 % GC were easily amplified in most of the reactions and protocols (Table 5-1). On the other hand, *Mb0129,* the long DNA target (> 1.5 kb) with a high GC-content (77.5%) behaved differently. *Mb0129* amplification was achieved by both the 2St and 3St protocols, but not the TD protocol. The success of implementing the 2St or 3St PCR protocols is dependent on the use of a selected high fidelity DNA polymerase such as Q5, Platinum SuperFi, and PrimeSTAR GXL and the enhancer. Apparently, the PCR protocol involved in the 2St reaction performed better than other protocols. The 3St PCR protocol performed similarly compared to the 2St protocol in some cases but generated a low yield for the amplification of *Mb0129* using Q5 and Platinum SuperFi polymerases.

Touchdown PCR has been widely recommended for the amplification of high GC-content genes. This method was also suggested for use in cases where the determination of primers' melting temperatures is difficult (179, 181). The advantage of the TD PCR procedure is in exploring the exponential nature of PCR. Given that primer annealing and extension are very critical steps for PCR, starting by annealing at higher temperatures then decreasing the annealing temperature gradually gives a two-fold advantage per cycle towards the correct product synthesis (180). As opposed to the expectation, the TD protocol failed to synthesize the larger GC-rich gene *Mb0129* while using several selected enzymes, although it was able to amplify the small gene *mpb83* of moderate GC-content and the large gene *LMHCC_RS00060* of low GC content.

The success in amplifying the *Mb0129* gene by the 2St PCR reaction but not the TD protocol indicated that the PCR reaction, including specific primer annealing and extension, works better at higher temperatures for a long GC-rich DNA template. In the TD protocol, the final annealing temperature at 55°C, lower than that used in the 2St or 3St protocol, might be a factor responsible for the observed difference in performance between the protocols. It should not be surprising that the temperature played a critical role in the PCR protocols for the amplification of the *M. bovis* ORFs especially with the use of lengthy primers. Primers were designed to clone the target ORFs into a plasmid vector, in which a stretch of nucleotides derived from the vector sequence was added at the 5' end of gene-specific primers, (Materials and Methods, Section 3.6.1). These long primers, with extra non-gene specific sequences, exhibit high melting temperatures (≥72°C) thus giving an advantage to the 2St PCR. In addition, it is highly possible that the high GC content of the *M. bovis* genome forms a hairpin or other secondary structures at the PCR annealing stage and thus interferes with the function of the polymerase by hindering product synthesis; another factor that favours a higher temperature in the reaction.

5.2.5 The Ramp Speed of the Thermal Cycler

The ramp speed of the thermal cycler may play a role in amplifying difficult targets by decreasing the ramp speed. Attempts were made to amplify a total of 7 genes (having lengths between 1617-2166 bp and GC content between 60-77.6% (Appendix 4, Table S9) by using the PrimeSTAR GXL enzyme in a 2St PCR. The thermal cycler was explored at a full ramp speed (6 °C /s) and lower speed (2 °C/s). Results show the successful amplification of the 7 targets with the lower ramp speed, in contrast to only a single target amplified using a full ramp speed (Figure 5-4). These results agree with the findings of Frey *et al.* (183)

using a slow ramp speed for amplification of GC-rich sequences. Frey *et al.* (183), however, used an additional step in which they added 7'-deaza-2'-deoxyguanosine, a dGTP analogue that disrupts hydrogen bonds and decreases the annealing temperatures. This step would disrupt the gene structure for the downstream cloning and protein expression applications and thus was not implemented in our experiments. Frey *et al.* (183) also demonstrated the amplification of short DNA targets (<1kb) and did not discuss the obstacles with the amplification of longer targets (>1kb). In this experiment, the Takara PrimeSTAR GXL enzyme was selected because it showed a better performance in the amplification of *Mb0129.* This was evidenced by the absence of non-specific bands seen in the case of the Q5 polymerase reactions and by a better yield than that of the Platinum SuperFi polymerase reactions (Figure 5-2B and Figure 5-3AB).

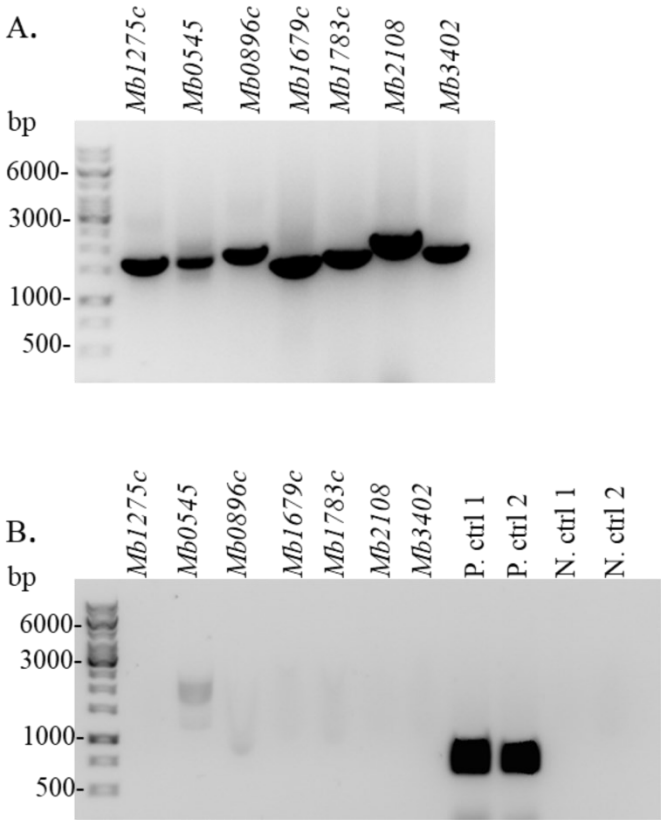


Figure 5-4. Comparison of the effects of modification of the thermal cycler ramp speed on amplification of seven target genes using PrimeSTAR GXL enzyme in a 2St PCR.

A. Seven gene targets are readily amplified by adjusting the thermal cycler to a slow ramp speed (2 °C /s). B. Only one target out of the seven (*Mb0545*) showed an amplicon in a full-speed thermal cycler (6 °C /s). P. ctrl1 and P. ctrl2 and represent positive controls 1 and 2, *mpb83* amplified using 2 St PCR with an added enhancer at low and full ramp speed, respectively. N. ctrl1 and N. ctrl2 represent negative controls 1 and 2, *Mb0129* amplified using 2St PCR without an enhancer at low and full ramp speed respectively.

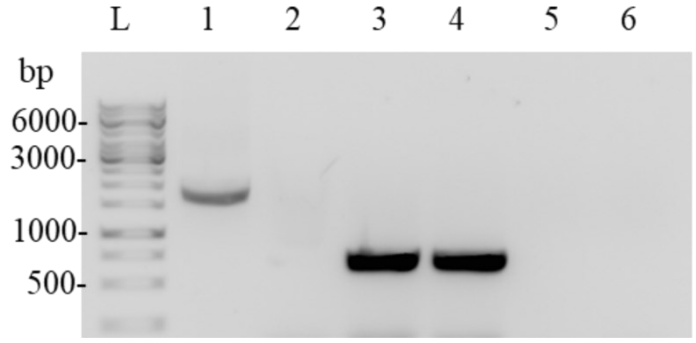
5.2.6 The Role of the Lengths of the Primers

The results obtained with the combination of three factors showed a better performance with the 2St PCR. These factors are the GC-rich DNA template, the long DNA target (>1kb), and the use of long primers. We aimed to investigate the role of the length of the primers in these reactions. Takara PrimeSTAR GXL enzyme was utilized to amplify a 1737 bp (fragment of *Mb0287c*), of 78% GC content from the genome of *M. bovis*. Primers were designed using the "Primer 3" software to have a short length of 20 nucleotides. The primers had a calculated melting temperature of 63°C. Results show the successful amplification of the target using 2St PCR with annealing and extension at 68 °C. Conversely, the 3St PCR with annealing at 63 °C did not produce an amplicon (S Figure 1). Clearly, the primers with shorter lengths did not result in the success of the 3St PCR to amplify the long (1.7kb), GC-rich gene. This indicates that the two factors, a long target, and a GC-rich template, maybe the key players that require a higher temperature and thus favour the 2St PCR for successful amplification.

A.

FP	CAGCGTCATAGCCGTAAACA
RP	GATCGCGTTGAGGATTTGTT

B.



S Figure 1. Primer design and amplification results for *Mb0287c* amplification.

A. The sequence of the primers (5' to 3') designed for the amplification of a 1737 bp sequence (fragment of *Mb0287c*). B. An agarose gel electrophoresis picture of the amplification results showing successful amplification of *Mb0287c* with the 2St PCR, lane1, whereas no product of *Mb0287c* with 3St PCR, lane 2. Lanes 3 and 4 show the positive control (*mpb83*) amplified while using an enhancer in 2St PCR and 3St PCR respectively. Lanes 5 and 6 show the negative control (*Mb0129*), with no enhancer added, and amplified using 2St PCR and 3St PCR respectively.

5.2.7 Formulation of a Final PCR Protocol

A PCR protocol was finalized based on the experimental results for the amplification of lengthy GC-rich DNA sequences from the *M. bovis* genome. The protocol involved the use of PrimeSTAR GXL enzyme, DMSO at 5%, and a master mix as described in Materials and Methods, Section 3.6.6. The cyclic conditions were set as 2St PCR to allow primer annealing and extension at high temperatures while slowing down the ramp speed to 2°C/s. This manipulation created a favourable environment for amplifying lengthy GC-rich targets (>1kb). The protocol proved to be effective and successful despite the relatively large difference in the T_m of primer pairs (in a range of 10 to 15°C) associated with the amplification of a sequence from the *M. bovis* genome. The final protocol enabled the successful amplification of GC-rich ORFs including *Mb0129,* (Figure 5-3A) and 7 other lengthy targets (Figure 5-4A).

Chapter 6. Profiling the Humoral Immune Response to M. bovis Infection by Protein Microarray Technology with Recombinant Proteins

6.1 Introduction

BTB is a chronic debilitating disease in cattle. The symptoms of bovine tuberculosis usually last for months, can remain dormant for years, and reactivate during periods of stress or in old age. In countries where eradication programs are implemented, there is a needed to identify early asymptomatic cases. In such situations, the clinical diagnosis is rarely adopted and most cases are diagnosed by routine testing or identified at the slaughterhouse (26).

The tuberculin skin test (TST), one of the major tests used in BTB diagnosis, is estimated to be able to detect around 50 - 60% of infected animals. Other tests such as IFN-γ assay can be used as an ancillary tool. An IFN-γ assay is based on the detection of IFN-γ released from sensitized lymphocytes after incubation with a specific antigen for 16 - 24 hours. Both tests have drawbacks including long turnaround time, high costs, in addition to difficulties in the standardization of antigens used in both tests (3, 17, 97).

Another test developed for the BTB diagnosis is the serological test. As an ancillary test, the serological test is being incorporated into BTB eradication programs in many countries. It is applied either in a serial testing regime after the TST to enhance specificity or in a parallel testing regime to enhance the sensitivity of cellular immunity-based assays(184). Serological tests may be useful at a later stage of the disease when the cell-mediated immunity wanes and cellular-immunity based tests give false-negative results. At this stage,

the humoral immune response, however, is steadily induced and the serological tests can be extremely advantageous especially in the developing countries where BTB is endemic (184, 185).

Serological tests are simple, economic with short turnaround time. These tests detect the binding of antibodies in the sera of infected hosts to specific *M. bovis* proteins. Indirect ELISA is a commonly used assay format in serodiagnosis. Serological tests, however; have been faced with some challenges. One of those is the variability in the antibody responses observed between different animal populations tested. Cattle, for example, are characterized by late and irregular development of the humoral immune response during the course of BTB. Another major issue that needs to be addressed is the choice of the antigen and the differential antibody response to various antigens. To date, MPB70 and MPB83 from *M. bovis* are the most widely used in commercial serological tests for BTB diagnosis. Other antigens such as the PPDB and antigens of the Ag85 complex were tested in research but were not found promising (79, 108, 186). To date, the diagnosis of BTB in the early stages of infection using indirect ELISA is hindered by the low sensitivity of the test compared to T cell-based assays. One possible solution is the identification of novel humoral immune response inducing antigens that may improve the sensitivity and specificity of the ELISA assay (59, 107, 184)

One of the approaches to identifying potentially immunogenic *M. bovis* proteins is to investigate the bacterial proteins specifically induced *in vivo* during infection. The identification of all *in vivo* induced bacterial proteins under different conditions of infection can be difficult. Bacterial proteins induced *in vivo* can be affected by several factors, for example, type of tissues infected, stage of infection, and exposure to stress (187, 188).

As a result, the choice of using bioinformatics and proteomic studies can help identify protein candidates that have the potential to induce the immune response. As described in Chapter 4, we have selected candidate proteins that may have extracellular localization and thus can potentially directly interact with the B-cell receptor. Here, a high throughput system was developed to clone and express the ORFs coding for the candidate proteins *in vitro* in a cell-free expression system. Protein microarray screening of the recombinant proteins for reactivity to sera from *M. bovis*-infected cattle allowed the identification of the antigens with the diagnostic potential.

6.2 Results

The development of a high throughput system has allowed generating recombinant proteins *in vitro* from a large number of selected ORFs for the identification of *M. bovis* diagnostic antigens. The overall approach in the system involves PCR amplification, *in vivo* cloning, *in vitro* transcription and translation, and protein microarray screening.

6.2.1 PCR Amplification of Target ORFs.

One-hundred and eighty-eight ORFs were targeted for amplification by PCR. Out of these selected ORFs, 16 candidates had their DNA larger than 2.5 kb in size and were divided into two or three fragments making the total number of DNA sequences for PCR amplification 208. Out of the 208 sequences, 186 were successfully amplified. Figure 6-1 and Figure 6-2 show some examples of ORFs derived by PCR from the *M. bovis* genome.

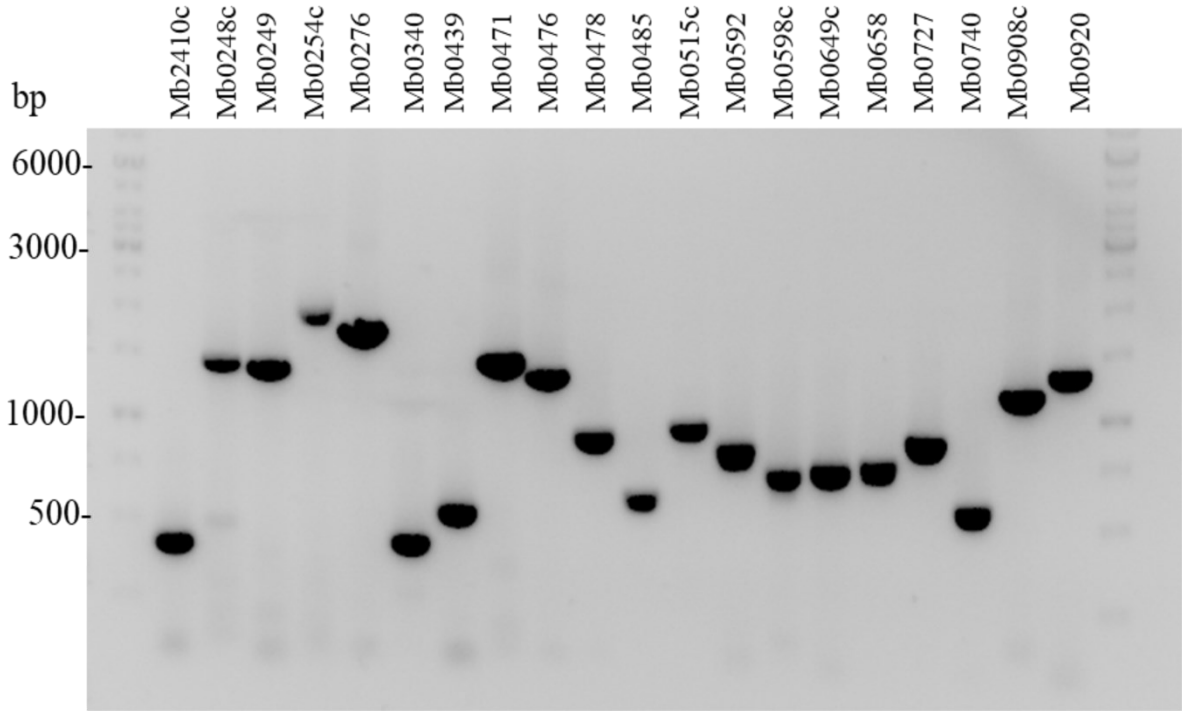


Figure 6-1. Agarose gel electrophoresis picture showing successful amplification of target DNA gene-sequence of *M. bovis*. Lanes represent the amplicons of 20 GC-rich ORFs candidates that were selected for this study.

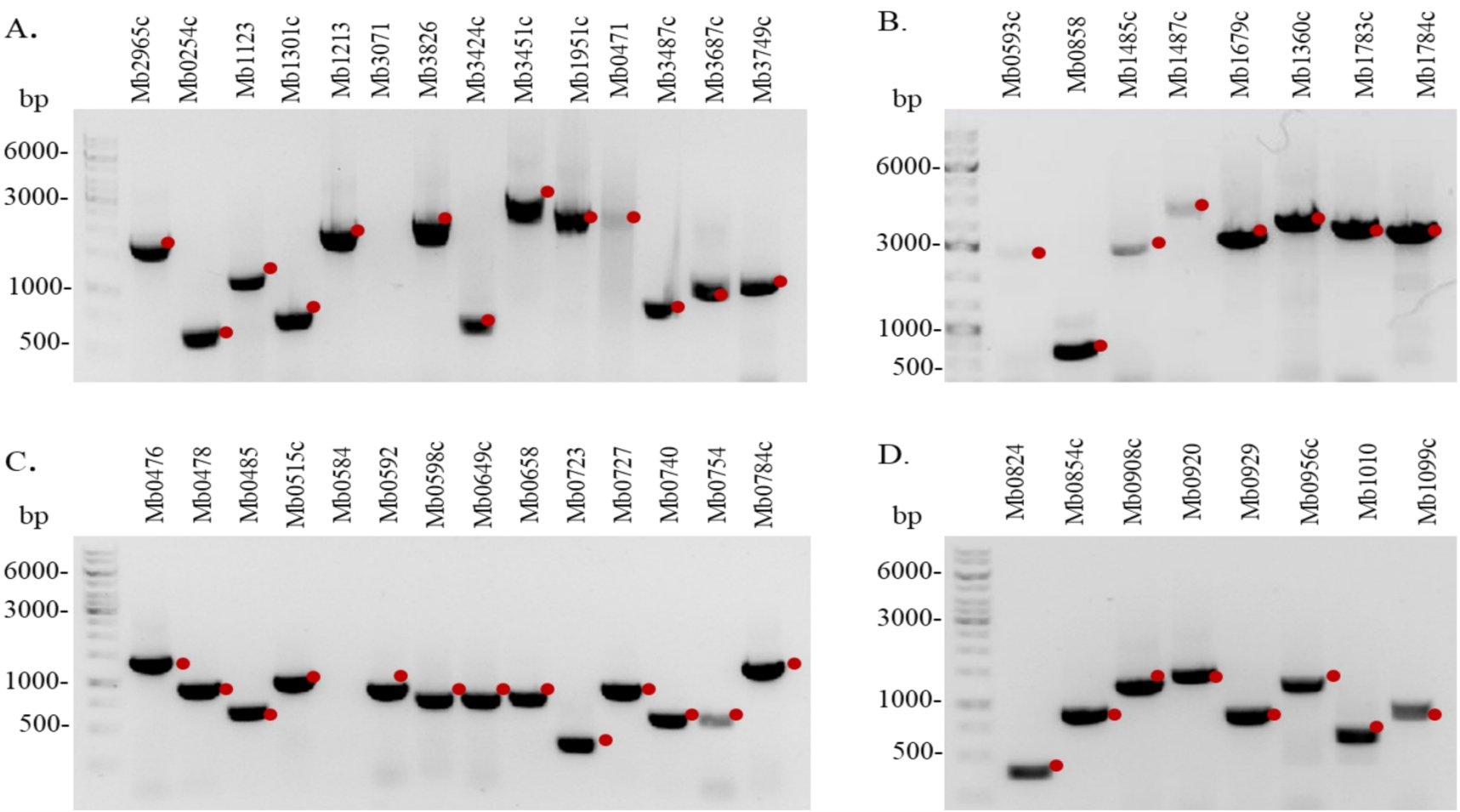


Figure 6-2. Panels A-D. Forty-two *M. bovis* gene sequences were successfully amplified using the designed protocol. Forty-four *M. bovis* genes were attempted, 42 of which were successfully amplified. Only 2 genes, *Mb3071*and *Mb0584,* were not amplified.

In addition, 12 ORF candidates were amplified and fused together using overlap extension PCR to create 6 chimeric genes, each of which consists of two ORFs (Figure 6-3).

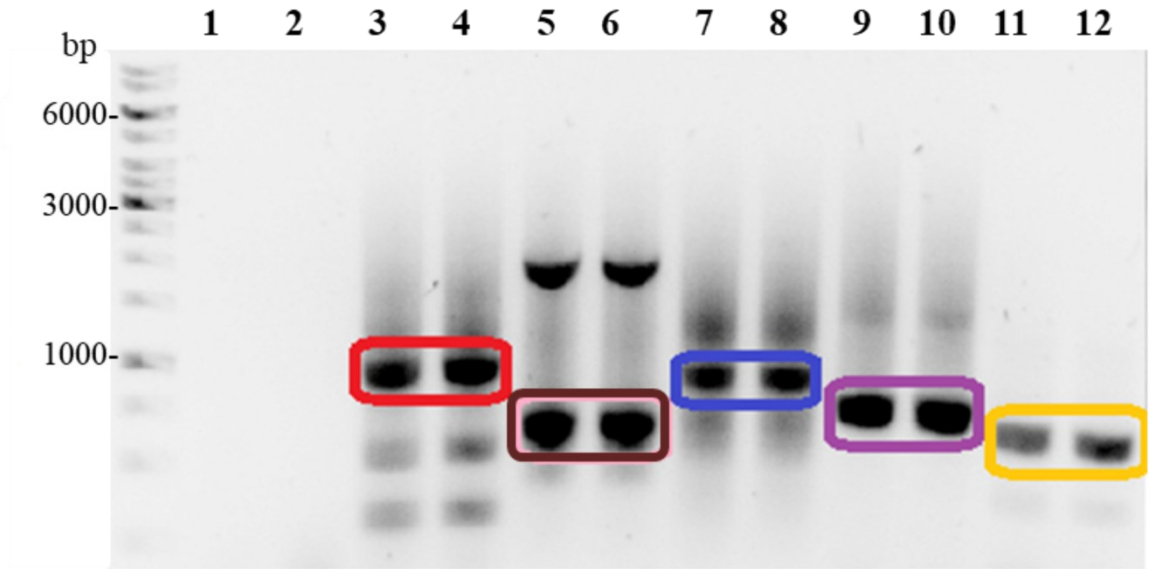


Figure 6-3. The use of overlap extension PCR to create a fusion between 2 genes; each less than 500 bp to create hybrid proteins.

Lanes1 &2= a fusion between *Mb1387*and *Mb1857*, lanes 3 &4= a fusion between *Mb2659*c and *Mb2801c*, lanes 5&6= a fusion between *Mb3219c* and *Mb3389*, lanes7&8= a fusion between *Mb0340* and *Mb3743*, lanes 9&10= a fusion between *Mb0723* and *Mb0824*, lanes 11& 12= a fusion between *esat6* "*Mb3905*" and *cfp10* "*Mb3904*". Each band represents a fusion gene amplicon, in duplicates. All lanes show amplicons except lanes 1 &2.

6.2.2 *In vivo* Cloning of ORFs in *E. coli*

Figure 6-4 shows the PCR amplification of the vector pIVEX2.3d (Appendix 1. Vector Map) that was pre-linearized by restriction digestion using NotI HF. Cloning of an ORF into

pIVEX2.3d was achieved by homologous recombination in *E. coli* that had been transformed with a mixture of purified PCR-amplified ORF and purified PCR-amplified linearized vector.

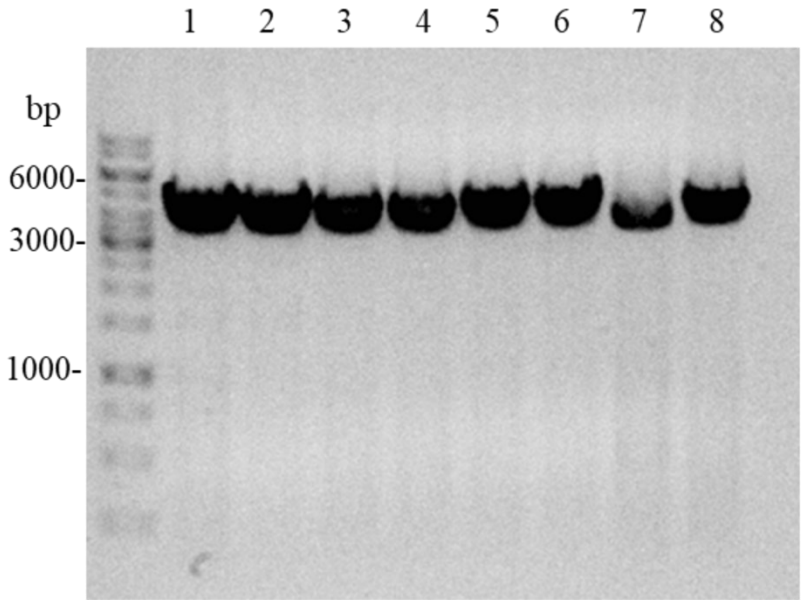


Figure 6-4. Agarose gel electrophoresis showing eight lanes of the successfully amplified pIVEX2.3d vector, 3.5kb.

Multiple attempts at using colony PCR failed to identify recombinant plasmids containing the inserts. Restriction digestion of recombinant plasmids derived from pIVEX2.3d with HindIII HF was found to be effective in detecting the presence of insert. The treatment of pIVEX 2.3d with HindIII HF yields two fragments, one fragment of ~2700 bp and a second fragment of ~800 bp. Successful cloning of an ORF into pIVEX2.3d, as exemplified by the cloned *esat6* gene (lane 2, Figure 6-5). increased the size of the second 800 bp fragment to 1100 bp due to the presence of cloned ORF.

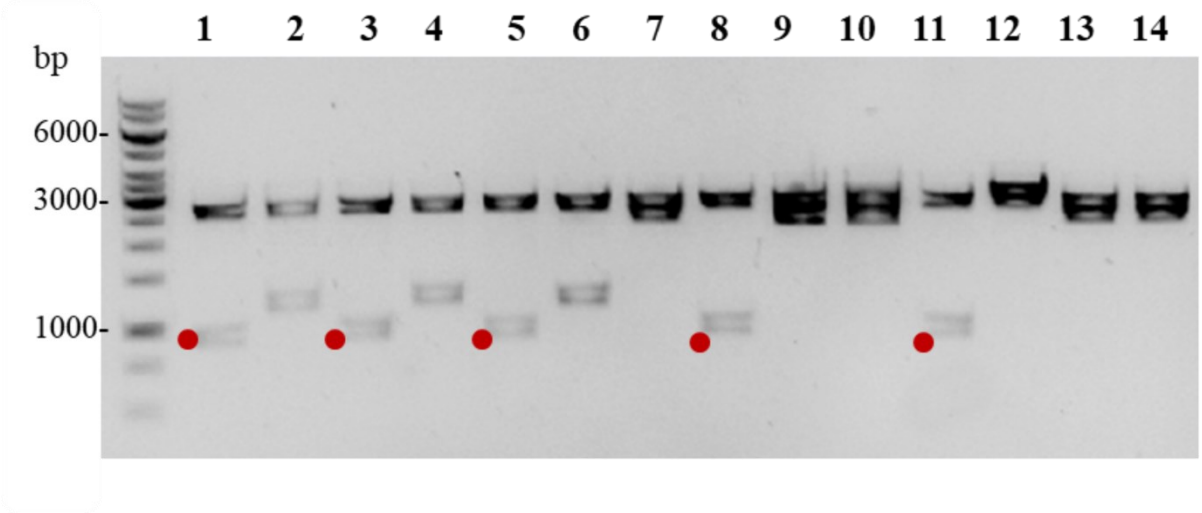


Figure 6-5. Restriction digestion of the recombinant plasmids and observing the size of the cut fragments.

Lanes 1,3,5,8 and 11 show a band of about 800 bp (red dots), similar to fragments produced by digestion of an empty vector meaning there are no inserts in these vectors. Lanes 2,4,6,7,9,10,12 and 13 are recombinant vectors that contain inserts. These lanes show an increase in the size of the 800 bp fragment by the size of the cloned ORF.

6.2.3 *In vitro* Expression of Recombinant Proteins

Proteins are shown to be expressed from the ORFs cloned into pIVEX2.3d in an *E. coli* based *in vitro* transcription-translation (IVTT) system by using the monoclonal mouse Anti·penta-His Antibody to detect the presence of a 6 ×His tag at the C-terminus on Western blot (Figure 6-6). A total of 145 *M. bovis* proteins were successfully expressed using the *E. coli*-based cell-free system.

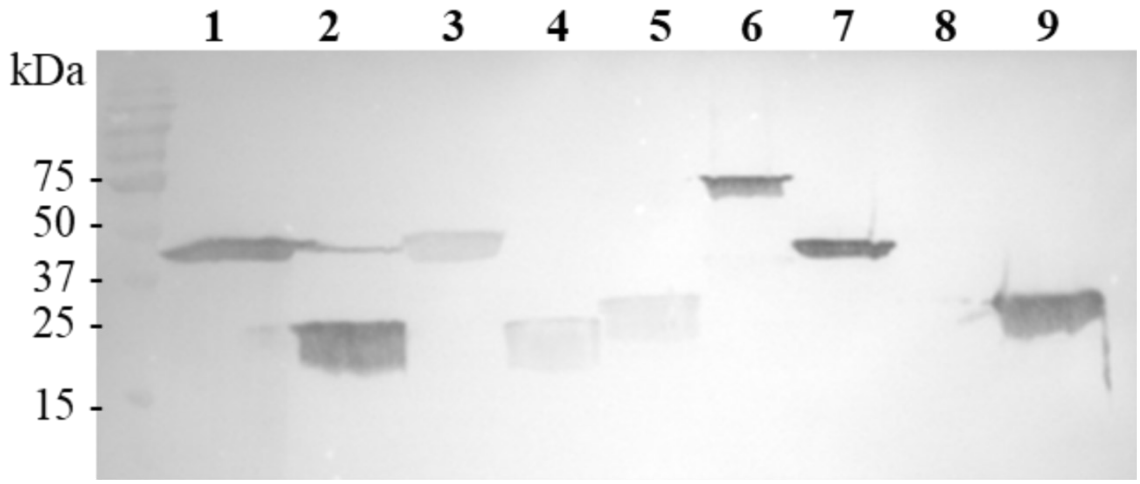


Figure 6-6. Western blot picture of expressed ORFs using the *E. coli* cell-free expression system.

Lanes 1 to 7 represent recombinant protein product of Mb3562 (41 kDa), Mb3751 (23 kDa), Mb3761c (49 kDa), Mb3792 (25 kDa), Mb3833c (31 kDa), Mb3841(57 kDa) and Mb3921c (39 kDa). Lane 8= Negative control, no plasmid DNA sample, Lane 9= positive control, GFP 28 kDa.

For the proteins that had failed to express in the *E. coli*-based cell-free expression system, another cell-free expression system based on the HeLa cell lysate was assessed and shown to express some *M. bovis* proteins from their ORFs cloned into the pT7CFE1-CHis vector (Figure 6-7). Fourteen proteins were successfully expressed with the mammalian cell-free expression system.

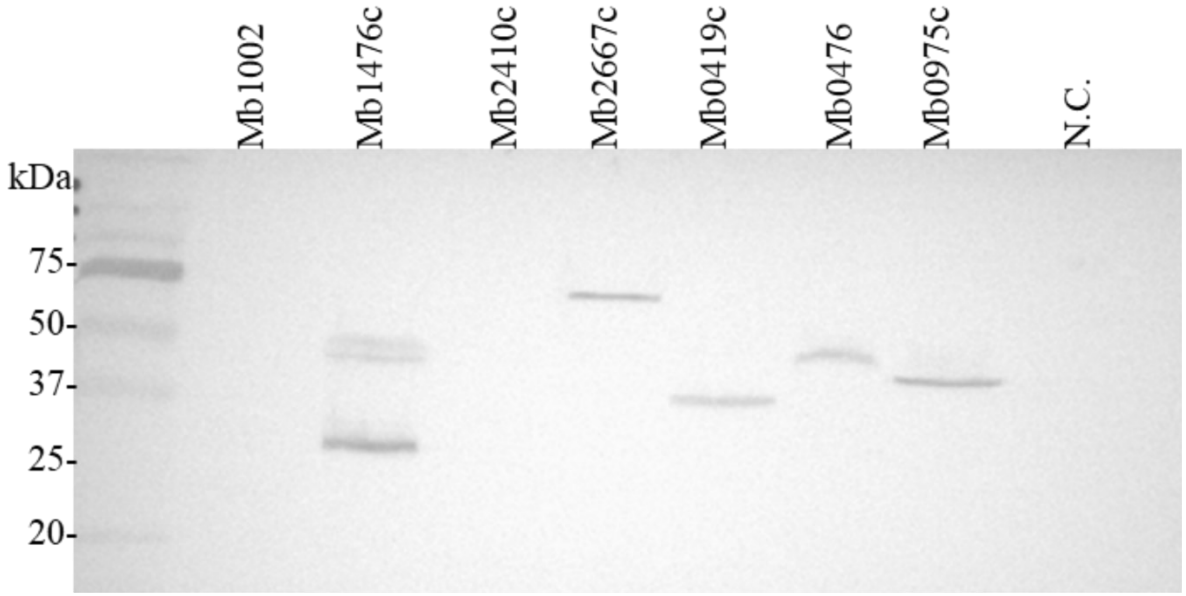


Figure 6-7. Western blot analysis showing five of *M. bovis* recombinant proteins that were expressed using the mammalian cell-free expression system. Proteins were probed using monoclonal mouse Anti·Penta-His Antibody; Mb1476c (41kDa), Mb2667c (63kDa), Mb0419c (34kDa), Mb0476 (47 kDa) and Mb0975c (35 kDa). Mb1002 (31 kDa) and Mb2410c (15 kDa) were not expressed. N.C. =negative control, no plasmid DNA sample.

Table 6-1 shows the number of successfully expressed proteins in full-length, fragments and fusions, along with the number of successfully amplified and cloned ORFs.

Table 6-1. Summary of the total number of ORF candidates from bioinformatics and proteomic sources and their application success.

Candidate source	whole Protein	Fragment	Amplified	Cloned	Expressed
Bioinformatics	80	33	97	91	82
Proteomic	80	3	77	77	67
Fusions*	12		12	12	10
Total	172	36	186	180	159
Grand total	208		186	180	159

* Fusions candidates were originally detected by both bioinformatics and proteomic studies, included in this table separately to demonstrate the success rate.

6.2.4 Optimization of Conditions for Probing Membrane-Bound Recombinant Proteins with Bovine Sera

To optimize the conditions for protein microarray screening two recombinant proteins, MPB70 and MPB83 were spotted on nitrocellulose and probed with the serum from an experimentally infected animal at three dilutions (S Figure 3, Appendix 5). As a negative control, a protein sample from the *E. coli* lysate of the IVTT system without the added plasmid DNA was also spotted for dot blot analysis. Results show a strong reaction of the serum to the negative control. To reduce the background reaction, four positive sera were treated with *E.coli* lysate (20% v/v), prepared in house to adsorb anti-*E. coli* antibodies, and used at 1:100 dilution for dot blot analysis as above (S Figure 4, Appendix 5). Results showed significant dampening of reactivity of the sera to *M. bovis* proteins. Subsequently, the experiment was designed and executed to compare the reactivity of one confirmed ELISA-positive serum, with and without pre-adsorption, to purified or unpurified *M. bovis* proteins (Figure 6-8).

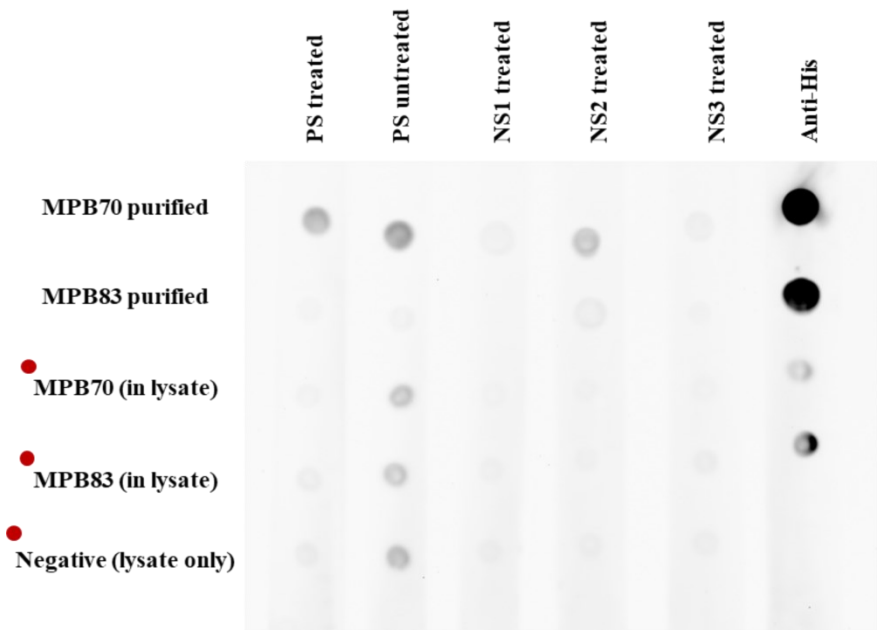


Figure 6-8. Dot blot analysis of the immune reactivity of blotted proteins against bovine sera and monoclonal mouse Anti·Penta-His Antibody.

Dot Blot analysis of purified MPB70, purified MPB83 (both from large-scale expression), unpurified MPB70, unpurified MPB83 (from the *in-vitro* system, red mark) and a negative protein (lysate only, from the *in vitro* system, red mark) against a panel of primary antibodies including serum from a TB ELISA positive control serum pre-adsorbed (PS treated), TB ELISA positive control serum without pre-adsorption (PS untreated), 3 negative control sera (NS1, NS2, NS3) with pre-adsorption (treated) and against monoclonal mouse Anti·Penta-His Antibody. The signal was detected using chemiluminescence.

The dot blot image was analyzed using ImageJ to obtain a numerical value of the signal intensity, based on which a heat map was constructed (Table 6-2). Results show lower reactivity of the treated positive serum to all *M. bovis* protein samples. By calculation, the reactivity of the positive serum to purified MPB70 decreased by almost 30% due to the

treatment. The treated positive serum reacted similarly to both the unpurified recombinant proteins and to the negative control (lysate only).

Table 6-2. Analysis and heat-map of the signal intensity of the dot blot analysis of Figure 6-8.

	PS treated	PS untreated	NS1 treated	NS2 treated	NS3 treated	Anti-His
MPB70 purified	49211	74398	2053	41419	2956	370207
MPB83 purified	1105	2091	600	5021	1399	409051
MPB70 (in lysate)	1697	23264	666	897	839	11445
MPB83(in lysate)	3119	15641	527	310	652	10576
Negative (lysate only)	3355	10492	916	792	708	542

PS = positive serum, NS = negative serum, 3 negative sera were used; NS1, NS2, NS3

Lowest Highest

To further optimize microarray screening conditions, preparations of purified or unpurified MPB70 and negative control proteins, printed on microarray chips, were probed using positive and negative bovine sera. Without protein purification, the positive serum shows the difficulty in differentiating the reaction to MPB70 from that to the negative control (N.C.). Similar results were observed with both *E. coli* pre-adsorbed and unabsorbed positive serum. With protein purification, the positive serum shows a strong reaction to purified MPB70 but not N.C. in comparison to the negative serum (S Figure 5, Appendix 5).

6.2.5 Purification of Recombinant Proteins

Recombinant proteins expressed using the IVTT system and purified using the method described in Chapter 3, section 3.13, were analyzed by SDS-PAGE and found to be acceptably pure proteins, as exemplified by recombinant MPB70 (Figure 6-9). Out of the 159

in vitro expressed recombinant proteins, 129 proteins were successfully recovered from the purification process. Purified proteins were used for printing on nitrocellulose-based microarray chips.

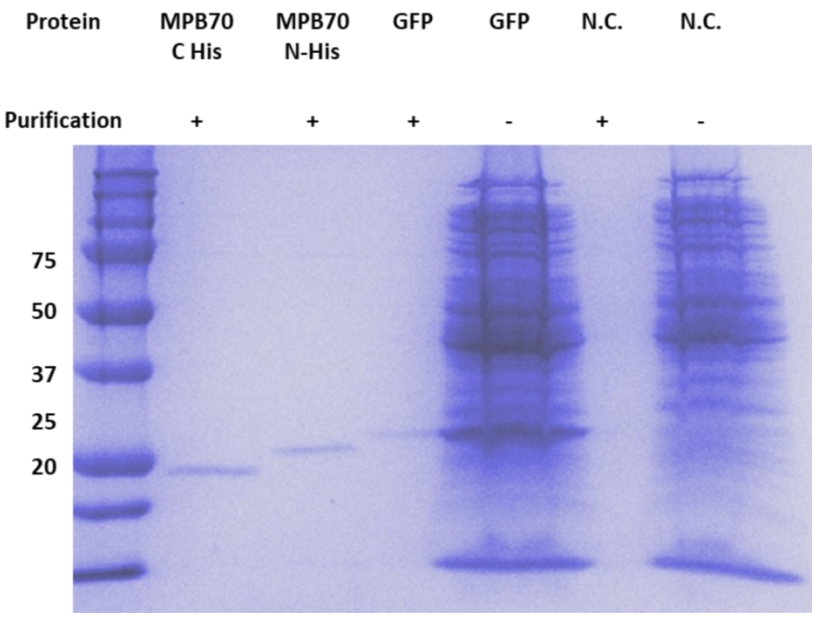


Figure 6-9. SDS-PAGE and Coomassie blue of recombinant proteins from the IVTT before and after purification.

The image shows two purified proteins, MPB70-C-His, MPB70-N-His, and the before and after purification both GFP and the negative control.

N.C.= negative control, reaction containing all the constituents of the IVTT except the plasmid DNA

6.2.6 Identification of Immunoreactive Antigens by Protein Microarray Screening

Printing proteins on microarray slides was monitored and confirmed by the microarray camera. The presence of the printed proteins on the slides was confirmed by probing with monoclonal mouse Anti·Penta-His Antibodies. The proteins were also probed by sera from 10 naturally infected adult cattle, 5 naturally infected juvenile cattle (less than 2

years of age), 5 experimentally infected cattle (at 4 months after inoculation), and 8 negative animals.

Figure 6-11 shows the reactivity of 129 purified proteins to the above panel of selected sera, which is depicted as a microarray heatmap. The protein antigens are used in the heatmap rows. The mean and the standard deviation of the reactivity of all infected and control animal sera to each printed recombinant protein are presented in Appendix 6, Table S10.

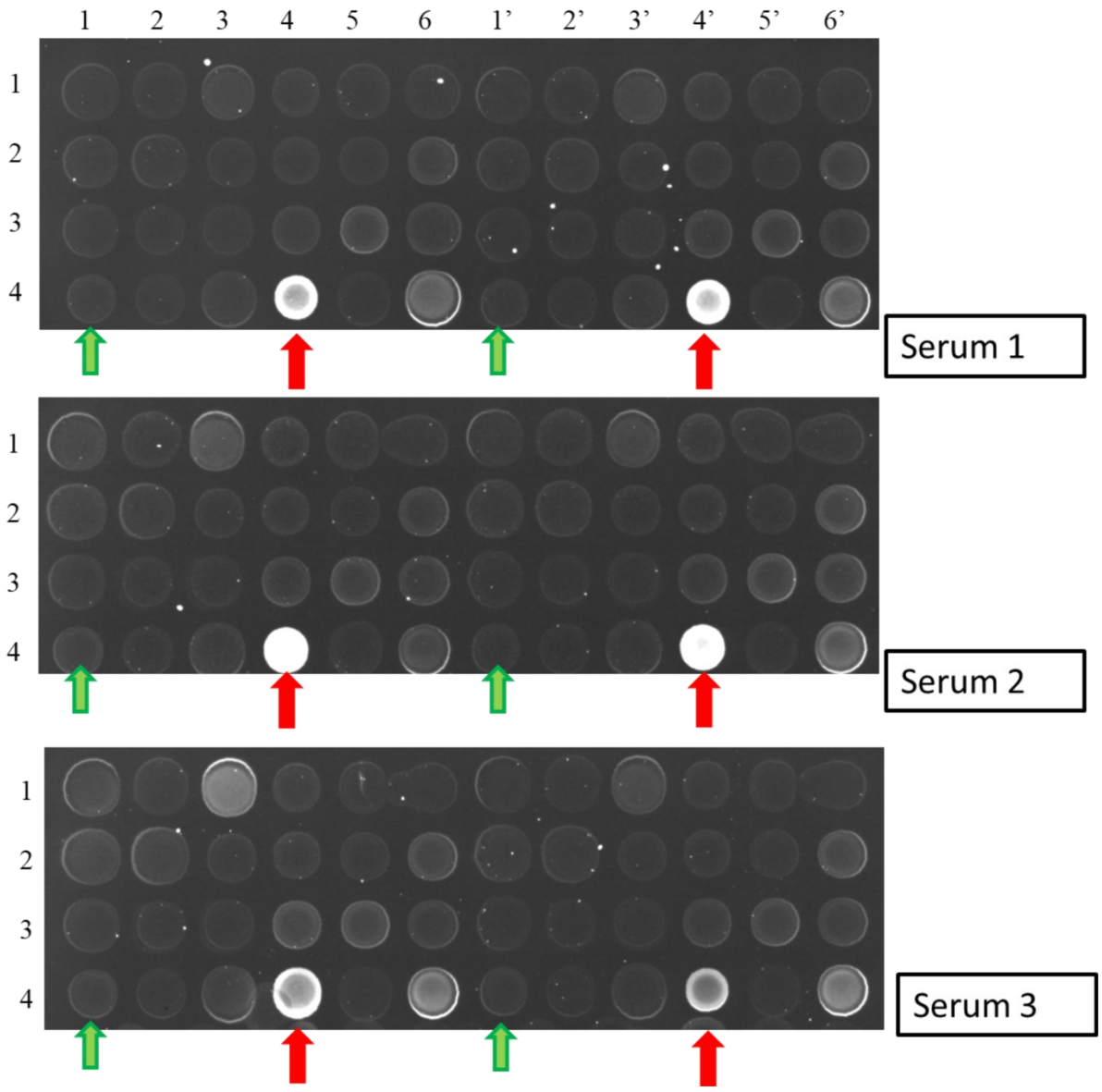


Figure 6-10. Three arrays each with 24 spots of printed recombinant proteins in duplicates, probed using sera from 3 different *M. bovis*-infected cattle.

Twenty-four proteins in columns 1 to 6 and reprinted again as 1' to 6'. Green arrows point to negative control proteins and red arrows point to a consistent positive reaction in the three arrays.

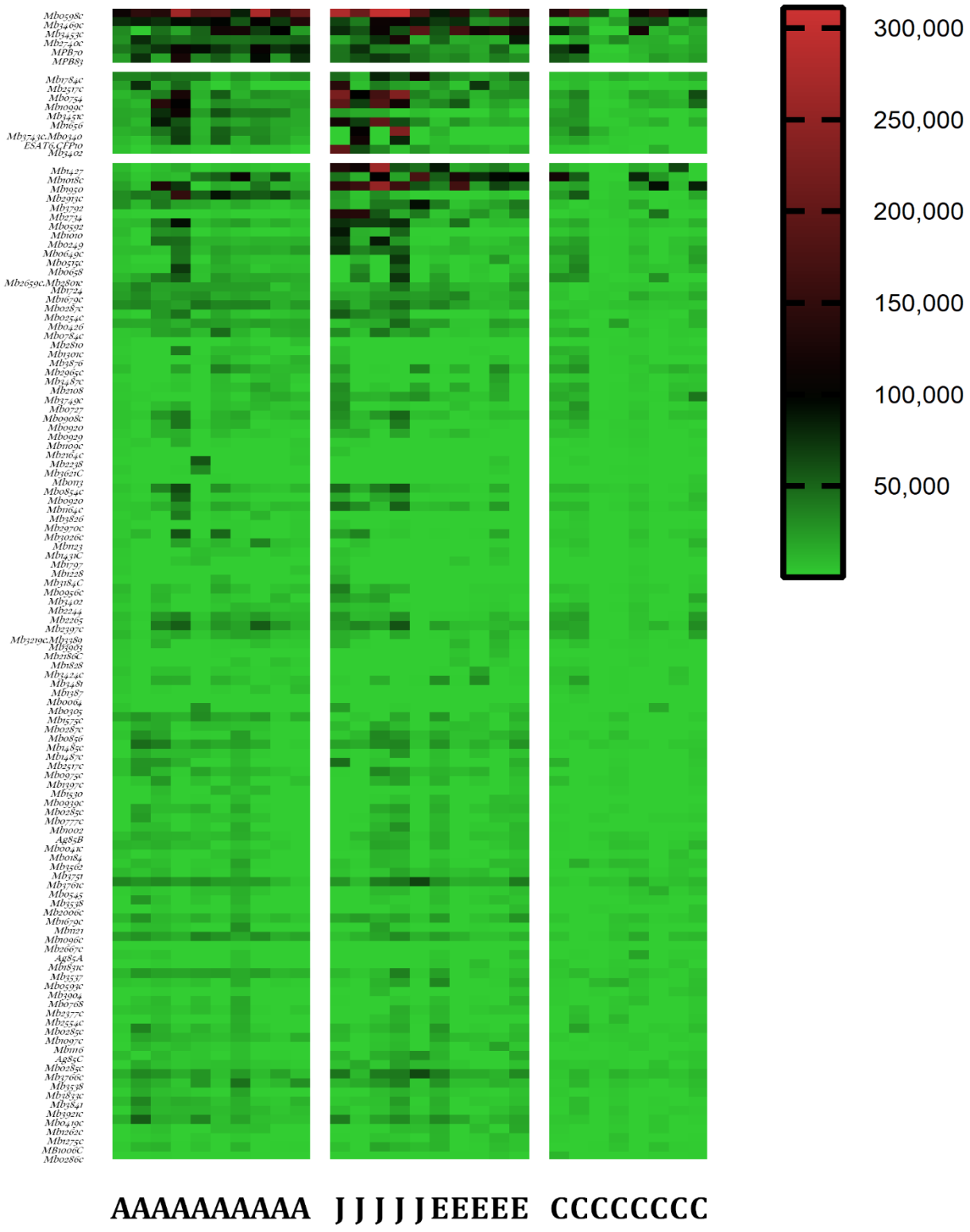


Figure 6-11. Microarray Heat map.

One-hundred and twenty-nine recombinant proteins (rows) probed with a panel of sera (columns). The sera were obtained from 15 naturally infected animals (10 sera from adult animals (A), 5 sera from juvenile animals (J)), 5 sera from experimentally infected animals (E), and 8 negative control sera (C).The colour of each cell represent FI with the red strongest and the bright green weakest.

The microarray heatmap revealed that most of the antigens were weakly reactive to the sera from infected animals. After a careful examination of the microarray results, two criteria were established to select the immunoreactive antigens: (i) the mean fluorescent intensity (MFI) from the reactivity of the infected animal sera to an antigen is greater than 40,000 (arbitrary unit) and/or (ii) the antigen has at least a single reaction with the FI of more than 100,000. Twenty-one antigens were found to meet the criteria and are placed on the top of the heatmap.

The independent sample t-test conducted to compare the MFI values between the reactivity of the sera from all infected cattle (adult naturally infected + juvenile naturally infected + experimentally infected) versus the control sera to the 129 antigens, revealed that 13 of the top 21 antigens reacted differently to the positive sera and the control sera with statistical significance ($p^* < 0.05$) (Table 6-3).

Table 6-3. Thirteen recombinant proteins that gave a statistically significant ($p<0.05$) difference when comparing the reactivity between positive sera versus control (negative) sera.

	Recombinant protein name	MFI controls	S.D. controls	MFI positive*	S.D. positive*	t	Sig. (2-tailed)
Potential serodiagnostic antigens	Mb0598c*	105,380	55,597	180,374	74,377	2.58	0.016
	Mb3453c*	37,354	43,361	84,062	50,683	2.287	0.031
	Mb3469c*	28,026	22,053	79,039	26,281	4.837	.000
	MPB70*	28,255	33,720	57,008	28,842	2.273	.031
	MPB83*	23,278	20,084	50,063	27,917	2.459	.021
	Mb2740c*	15,509	9,819	48,303	27,716	3.235	.003
Immune-reactive antigens	Mb1784c*	10,044	5,354	40,257	25,579	5.015	.000
	Mb0754*	21,584	20,103	61,832	70,061	2.340	.028
	Mb1099c*	16,499	15,770	55,410	59,220	2.708	.012
	Mb1656*	11,503	11,133	42,416	47,107	2.749	.011
	Mb0592*	7,937	6,117	30,396	30,359	3.152	.005
	Mb2517c*	2,739	2,952	28,509	34,566	3.304	.004
	ESAT-6. CFP-10*	4,775	7,363	23,128	32,832	2.356	.027

* positive sera= adult naturally infected+ juvenile naturally infected + experimentally infected

Further, the reactivity of sera to the above 13 antigens was compared between each of the three different infected animal groups and the control group as well as between the naturally infected adult and naturally infected juvenile using the independent sample t-test. The antigens showing a significant difference in reactivity between each two compared groups were tabulated in Tables 6-4 to 6-7.

Table 6-4. Antigens showing a significant difference in the sera reactivity between the adult naturally infected and the control groups.

	Adult (n=10)		Control (n=8)			
	Mean Adults	S.D. Adults	Mean Control	S.D. controls	t	Sig. (2-tailed)
Mb0598c	173,963	61,602	105,380	55,597	2.449	0.026
Mb3469c	88,642	20,536	28,026	22,053	6.024	0.000
MPB70	77,411	22,070	28,255	33,720	3.731	0.002
MPB83	56,504	35,363	23,278	20,084	2.361	0.031
Mb2740c	49,963	12,390	15,509	9,819	6.407	0.000
Mb1784c	31,433	11,771	10,044	5,354	4.741	0.000
ESAT6.CFP10	25,658	21,908	4,775	7,363	2.569	0.021

Table 6-5. Antigens showing a significant difference in the sera reactivity between the juvenile naturally infected and the control groups.

	Juvenile (n=5)		Control (n=8)			
	Mean Juvenile	S.D. Juvenile	Mean Control	S.D. controls	t	Sig. (2-tailed)
Mb0598c	256,000	57,840	105,380	55,597	4.683	0.001
Mb3453c	100,013	54,305	37,354	43,361	2.307	0.041
Mb3469c	83,114	30,113	28,026	22,053	3.822	0.003
MPB83	56,773	13,839	23,278	20,084	3.252	0.008
Mb1784c	64,735	39,844	10,044	5,354	3.052	0.037
Mb0754	152,015	83,510	21,584	20,103	3.431	0.024
Mb1099c	117,109	73,071	16,499	15,770	3.035	0.036
Mb1656	96,158	63,837	11,503	11,133	2.937	0.041
Mb0592	61,169	28,549	7,937	6,117	4.111	0.013

Table 6-6. Antigens showing a significant difference in the sera reactivity between the experimentally infected and the control groups.

	Experimental (n=5)		Control (n=8)			
	Mean Experimental	S.D. Experimental	Mean Control	S.D. controls	t	Sig. (2-tailed)
Mb3453c	120,018	40,629	37,354	43,361	3.42	0.01
Mb1784c	33,427	14,374	10,044	5,354	3.49	0.02

Table 6-7. Antigens showing a significant difference in the sera reactivity between the naturally infected adult and juvenile groups.

	Adult (n=10)		Juvenile (n=5)			
	Mean Adult	S.D. Adult	Mean Juvenile	S.D. Juvenile	t	Sig. (2-tailed)
Mb0598c	173,963	61,602	256,000	57,840	-2.477	.028
MPB70	77,411	22,070	46,431	17,452	2.725	.017
Mb0754	35,144	34,110	152,015	83,510	-3.006	.032
Mb1099c	39,304	44,401	117,109	73,071	-2.590	.022
Mb0592	23,046	26,916	61,169	28,549	-2.538	.025

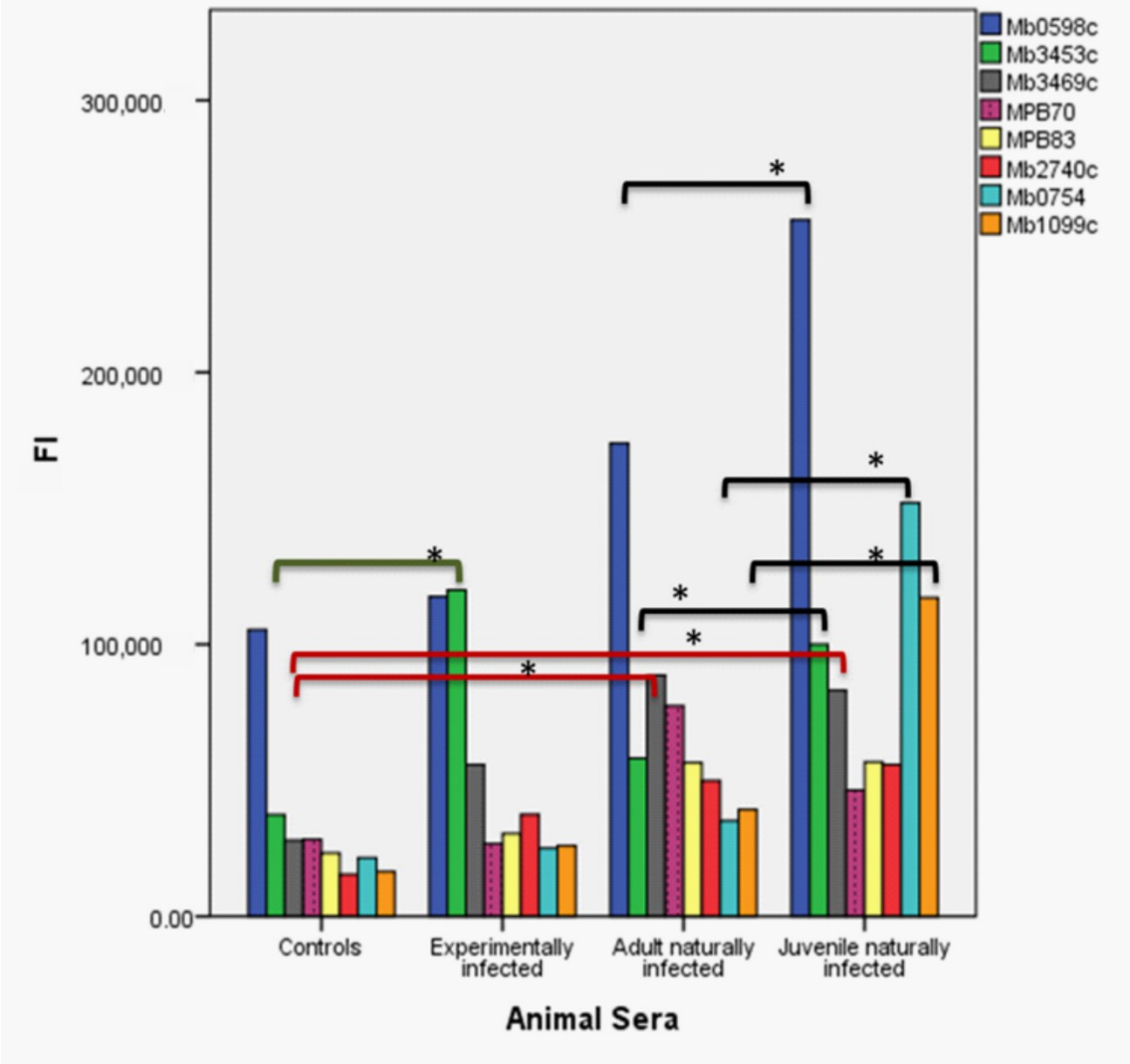


Figure 6-12. Clustered bar chart showing the mean reactivity of sera to 8 selected antigens on the comparison of the four animal-groups.

The mean reactivity of four antigens, Mb0598c, Mb3453c, Mb0754, and Mb1099c are significantly higher in the "Juvenile" group than the "Adult" group. The mean reactivity of Mb3469c in the control group is significantly lower than the mean of both the "Adult" and "Juvenile" group. The mean reactivity of Mb3453s is higher in the "Experimental" group than the control group. FI = Fluorescence intensity.

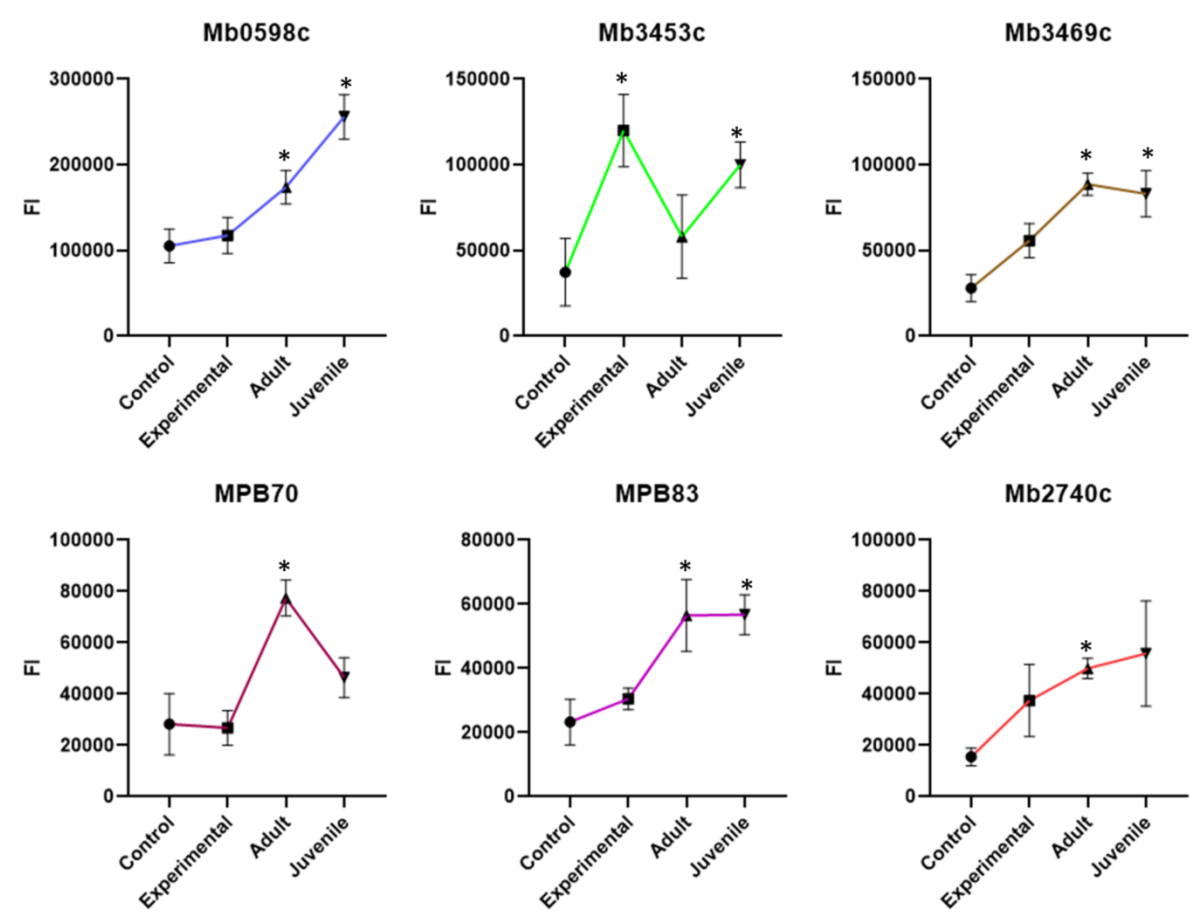


Figure 6-13. Patterns of fluorescence intensities and standard errors of six selected antigens in the 4 animal groups.

The asterisk (*) represents mean intensities significantly higher than the control group. Mb0598c, Mb3469c, and MPB83 reactivity are significantly higher in both the adults and the juvenile group. Mb3453c reactivity is significantly higher in both the adults and the experimental group. MPB70 and Mb2740c reactivity are significantly higher in the adult group only. FI= fluorescence intensity.

6.3 Discussion

In this part of the study, a high throughput system was developed to allow for the production of the candidate proteins in a recombinant form. This was followed by the

screening of these recombinant proteins against sera from BTB-infected animals to determine their serodiagnostic potential. These candidate proteins were selected based on the possible extracellular localization as predicted by bioinformatics or identified by proteomic studies as outlined in Chapter 4.

PCR amplification of the target ORFs was the first step in the system for the high throughput expression of the selected proteins. Given that the *M. bovis* genome has a high GC content, an effective protocol was developed as outlined in Section 5.2.7. Additionally, other solutions to address the challenging problems in amplifications were adopted. First, ORFs larger than 2.5 kb in length were covered by two or three shorter sequences keeping overlap sequences of at least 150 bp to avoid loss of possible epitopes. These shorter sequences had a higher success rate in the amplification and cloning. Also, nested PCR was used for the amplification of some sequences (189). Finally, the wobble technique described by Kumar and Kaur (171) was also applied. The latter technique modifies the primers by substituting "G" or "C" bases with "A" or "T" bases without changing the encoded native amino acid sequence based on the degeneracy of codon. The application of these procedures combined resulted in the successful amplification of 186 out of 208 (~89.4%) DNA sequences attempted.

An *in vivo* cloning strategy was employed to clone the DNA targets. This strategy is more robust and more compatible with cloning numerous ORF candidates compared to the conventional restriction digestion/ligation cloning method. The primary plasmid vector used for the cloning, pIVEX2.3d, was selected because of its suitability in the downstream protein expression by IVTT. The amplification of the linearized vector by PCR followed by purification allowed the elimination of the residual circular vector, which may interfere with

the cloning process. In addition, amplification of the vector allowed a high yield of ready-to-use vector for multiple *in vivo* cloning experiments. Using specific primers, each candidate ORF amplicon was generated containing additional sequences at both ends. These sequences were homologous with the end sequences of the linearized amplified vector. A length of 25 to 26 bps in homology proved to be effective for the *in vivo* homologous recombination after the transformation of competent *E. coli* cells with a mixture of the PCR amplified targets and the linearized vector. This is shorter than the homology length of 33 to 49 bps described by others (190, 191). The homologous recombination *in vivo* resulted in the cloning of an ORF into the plasmid vector. Interestingly, the success rate of cloning was found higher when transformation plates were left in the room temperature for 48 hours than when they were left at 37 °C overnight. The procedure with a longer incubation time was only considered for a second trial, after the failure of the first trial. In this study, the colony PCR using the T7 promoter and T7 terminator primers to confirm the presence of an insert was sometimes unsuccessful due to the amplification difficulty caused by the high GC content of a target ORF. For this reason, the restriction digestion of plasmids prepared from the *E. coli* clones was conducted instead to verify the success of cloning in comparison to the digestion of the empty vector (Figure 6-6). The restriction digestion strategy proved to be successful in identifying the clones correctly. The cloned sequences were determined at Genome Quebec and found to be correct and in frame. The *in vivo* cloning system was able to clone 97% of the attempted ORF targets.

The expression of the recombinant proteins was carried out *in vitro* using the *E. coli* cell-free expression system. The system enabled a high-throughput expression of 145 recombinant proteins. Proteins were expressed in a small but sufficient amount for the

microarray screening. All expressed recombinant proteins had a C-terminal 6× His-tag that was confirmed by Western blot analysis using the monoclonal mouse Anti·Penta-His Antibody (Figure 6-6). The choice for the expression of C-terminal 6× His-tagged proteins was to ensure that the correct translation for a full-length recombinant protein was achieved. Detection of smaller protein fragments would indicate that the protein undergoes proteolysis. This phenomenon was observed on a few occasions where protein fragments of different sizes were detected on Western blots. To improve the expression of certain proteins *in vitro*, the HeLa "mammalian" cell-free expression system was another system deployed in this study (Figure 6-7). One reason for utilizing this system is that the mammalian cells were found to deliver a higher level of protein expression in response to GC-rich codons compared to GC-poor codons, offering this system an advantage for expressing *M. bovis* genes (192). The system resulted in the expression of fourteen recombinant proteins that failed to express using the *E. coli* cell-free system.

In this work, it was observed that small proteins (less than 20 kDa) failed to express using both the *E. coli* cell-free expression system and the mammalian cell-free expression system. An explanation for this observation may be that small proteins are relatively easily degraded by proteases in the lysate. To resolve this problem, these smaller proteins were attempted to be expressed in chimeras through the use of overlap extension PCR to create a fusion of two target ORFs (Figure 6-3). The strategy proved to be successful in achieving the expression of 6 hybrid proteins representing 12 candidate proteins.

Screening of recombinant proteins against sera from experimentally infected animals was carried out using dot blot analysis for the optimization purpose prior to the microarray analysis. Surprisingly, dot blots showed a strong reaction of the positive sera to the negative

IVTT control, a protein sample containing all the constituents of the *in vitro* expression kit except the vector DNA (S Figure 3, Appendix 5). This reaction was almost similar to the reactions of the positive sera to specific IVTT *M. bovis* recombinant proteins. This observation may be explained by the pre-existence of anti-*E. coli* antibodies in the cattle sera and the fact that IVTT contains a majority of *E. coli* lysate proteins. To eliminate the background reaction, treatment of the cattle sera by pre-adsorption using *E. coli* lysate was attempted as a solution. This solution was adopted by Davies *et al.* (191) who found that the pretreatment of sera by adding *E. coli* lysate before probing the microarray resulted in a reduced *E. coli* background, without interfering with the specific reactivity. Unexpectedly, the reactivity of positive sera to *M. bovis* proteins was markedly decreased after the pre-adsorption (S Figure 4, Appendix 5). Consequently, another dot blot analysis was carried out to compare the reaction of a positive serum previously confirmed to react to MPB70 by ELISA. The positive serum was tested in both pre-adsorbed and unadsorbed forms against *M. bovis* proteins in both purified and unpurified forms. The reactivity of pre-adsorbed positive serum to purified MPB70 protein decreased significantly compared to the untreated serum. Clearly, a considerable amount of specific anti-MPB70 antibodies were lost in the experiment due to the treatment of the serum (Figure 6-8). As a result, the pre-adsorption was not implemented in this study.

Another solution sought to reduce the high background from the reaction of cattle sera to protein antigens was the development of a high-throughput procedure for the purification of recombinant proteins. The procedure, developed in-house as described in Chapter 3.13, was successful in the preparation of well purified proteins. One hundred and twenty-nine recombinant proteins were recovered following the application of the procedure

to the purification of 159 recombinant proteins. The purified proteins were directly used in the downstream screening process.

Protein microarray is a high-throughput system used here to screen for protein antigens that induced the humoral immune response in BTB-infected cattle. All immunological reactions from the screened proteins were plotted on the heat map (Figure 6-11). The heat map showed that most of the antigens were weakly reactive. Twenty-one antigens fulfilled the criteria established for identifying highly reactive antigens. Thirteen out of these 21 antigens exhibited a statistically significant difference in the reactivity to sera between the infected and control animals (Table 6-3). These 13 antigens were further tested for their ability to differentiate different groups of infected animals from the control group.

Juvenile naturally infected animals remarkably reacted with a larger number of antigens. This was revealed from the microarray heat map and also confirmed by the statistical analysis. Nine recombinant proteins showed significantly higher immunoreactivity in the juvenile naturally infected group than in the control group (Table 6-5). Additionally, 4 antigens demonstrated significantly high immunoreactivity in the juvenile naturally infected group compared to that of the adult naturally infected group (Table 6-7). This may be explained by the pre-existence of maternal cross-reacting antibodies or due to a more active immune system in the younger age animals.

Another observation was that the adult naturally infected group showed higher immunoreactivity to the antigens than both the experimentally infected and the control group. The reactivity of the sera to seven antigens could differentiate the adult naturally infected group from the controls (Table 6-4). Clearly, the experimentally infected group showed the lowest immunoreactivity to the antigens among the three infected groups. Only

two antigens demonstrated immunoreactivity that significantly differentiated the experimentally infected group from the control group (Table 6-6). One possible explanation for this observation is that the serum samples were taken early after experimental infection, only four months after pathogen inoculation.

To the best of our knowledge, this study represents the first study to screen a significant number of *M. bovis* proteins against infected cattle sera in a microarray format. Other studies used protein microarray to screen for immunoreactive antigens, in the context of diseases such as brucellosis and leptospirosis (191, 193). The study is also the first to include a comparison of antigens reactivity in two different age groups, adult and juvenile naturally infected, in addition to experimentally infected animals. One limitation of this study is its ability to analyze only protein antigens. Lipid or polysaccharide antigens were not assessed. However, for the objective of improving BTB serodiagnostics, this may not be a serious limitation.

In this study six antigens were highlighted. They exhibited the highest individual immunoreactivity in the different groups of infected animals as well as a significant difference in immunoreactivity between the positive and control sera (Figure 6-13). Four novel antigens, Mb0598c, Mb3469c, Mb3453c, and Mb2740c, were identified as potentially immunodiagnostic antigens in addition to the well-known MPB70 and MPB83 (Figure 6-13). One antigen, Mb0598c demonstrated the highest reactivity with the positive sera among all the antigens tested. The immunoreactivity of antigen Mb3469c demonstrated the highest ability to differentiate the positive from the control sera (p=.000). The six antigens also demonstrated the ability to differentiate positive groups from the controls, with superior reactivity in specific groups. For example, Mb0598c, Mb3469c together with MPB83

showed significantly higher immunoreactivity in both the adult naturally infected group and the juvenile naturally infected group compared to the control group. The reactivity of Mb3453c with sera was significantly higher in both the adult naturally infected group and the experimentally infected group compared to the control group. Mb2740c was similar in immunoreactivity to MPB70; both showed significantly higher reactivity with sera in the adult naturally infected group than the control group. The four novel antigens, Mb0598c, Mb3469c, Mb3453c, and Mb2740c, were selected as potential immunodiagnostic antigens for evaluation in the next part of the study.

Chapter 7. Evaluating the Identified Antigens in an Appropriate Immunological Assay Format for the Detection and Identification of the M. bovis Infected Animals

7.1 Introduction

In the work reported in Chapters 4 to 6, a stepwise approach was adopted to identify potential immunogenic candidates. The approach comprised the selection of potential candidates by exploring the *M. bovis* genome sequences using bioinformatics tools to identify potentially extracellular proteins. In addition, we identified *M. bovis* proteins secreted into the bacterial culture media by the proteomic approach. For that purpose, a proteomic study was conducted in this study to identify proteins secreted by *M. bovis* in different stages of growth. Proteomic studies previously published by others were also explored to identify additional *M. bovis* protein candidates.

The next steps involved the PCR amplification of candidate genes and the *in vivo* cloning of these genes into an expression vector through homologous recombination in *E. coli*. This was followed by the expression of the target genes using the cell-free expression systems. All expressed proteins were purified and screened against a panel of sera from naturally and experimentally infected animals using protein microarray. This study resulted in a high throughput system for gene cloning, protein expression, protein purification, and screening of recombinant proteins for the identification of *M. bovis* antigens with diagnostic potentials. As described in Chapter 6, the protein microarray screening showed that the majority of selected protein candidates were not or were weakly immunoreactive to the sera from infected cattle. Four novel antigens, Mb0598c, Mb3469c, Mb3453c, and Mb2740c,

were identified as potentially immunodiagnostic antigens. Promising results were obtained with these four proteins as their reactivity with sera showed a significant difference between infected and non-infected animals. These antigens are novel antigens because they were not previously reported for use in BTB diagnosis.

This part of the study reports the evaluation of the four novel proteins for BTB serodiagnosis in an appropriate immunological assay format. The ELISA format was selected as the primary diagnostic application. It is widely used, easy to perform, and cost-effective. The ORFs coding for the four proteins were amplified and cloned as described in Chapter 3, Section 3.16. The proteins were then expressed on a large scale, purified and evaluated in ELISA for the detection of specific antibodies in sera from *M. bovis*-infected cattle. One of the four antigens, Mb3453c, was also evaluated in dot blot assay.

7.2 Results

7.2.1 Cloning of the Four Selected ORFs *Mb2740c, Mb0598c*, *Mb3453c*, and *Mb3469c*

The four proteins encoded respectively by *Mb2740c, Mb0598c*, *Mb3453c*, and *Mb3469c* have been identified as novel immunoreactive antigens. Expression of these proteins on a large scale in *E. coli* is needed for evaluating their diagnostic potential. The pLIC-CHis vector was linearized with ScaI into a band of ~5.2 kb (Figure 7-1, Panel A). The four ORFs of *Mb2740c, Mb0598c*, *Mb3453c*, and *Mb3469c* were amplified by PCR from the *M. bovis* genomic DNA, together with *mpb70* and *mpb83* (Figure 7-1, Panels B-G). These ORFs were successfully cloned into pLIC-Chis using the ligation independent cloning technique described earlier in Chapter 3, section 3.16. DNA sequencing of recombinant plasmids verified that the target ORFs were correctly inserted into the vector.

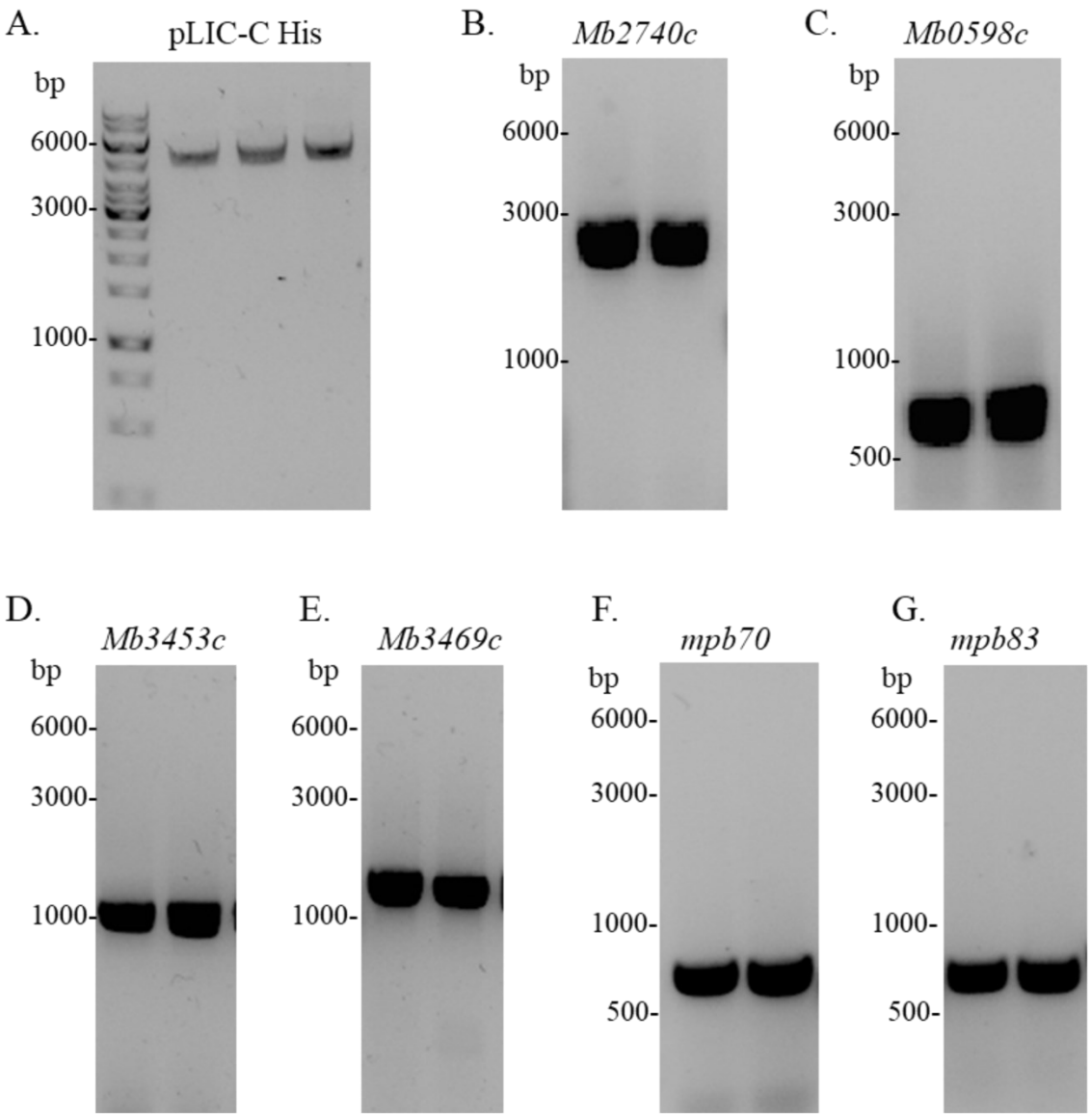


Figure 7-1. Agarose gel electrophoresis picture of the pLIC-CHis vector, digested, and the six selected genes, PCR-amplified.

Panel A. The pLIC-CHis vector was linearized using ScaI. **Panels B-G.** Successfully amplified DNA sequences of the antigens, *Mb2740c* (2100 bp), *Mb0598c* (687bp), *Mb3453c* (1035bp), *Mb3469c* (1404bp), *mpb70* (582 bp) and *mpb83* (663 bp) respectively.

7.2.2 Large Scale Expression and Purification of the Four Novel Immunoreactive Proteins

The large scale production of the novel antigens encoded by *Mb2740c*, *Mb0598c*, *Mb3453c*, and *Mb3469c* was carried out by transformation of *E. coli* Rosetta pLysS, with the recombinant vectors obtained from the previous experiment. Figure 7-2 shows the IPTG-induced expression of the four recombinant proteins, as depicted by SDS-PAGE. The other two important protein antigens of *M. bovis*, MPB70 and MPB83, were similarly expressed (Figure 7-3). All the selected proteins except Mb3453c were successfully purified from the *E. coli* expression host using Ni-agarose affinity chromatography under non-denaturing conditions. The Mb3453c protein was purified under denaturing conditions from the insoluble fraction prepared from the *E. coli* expression host (S Figure 6, Appendix 7).

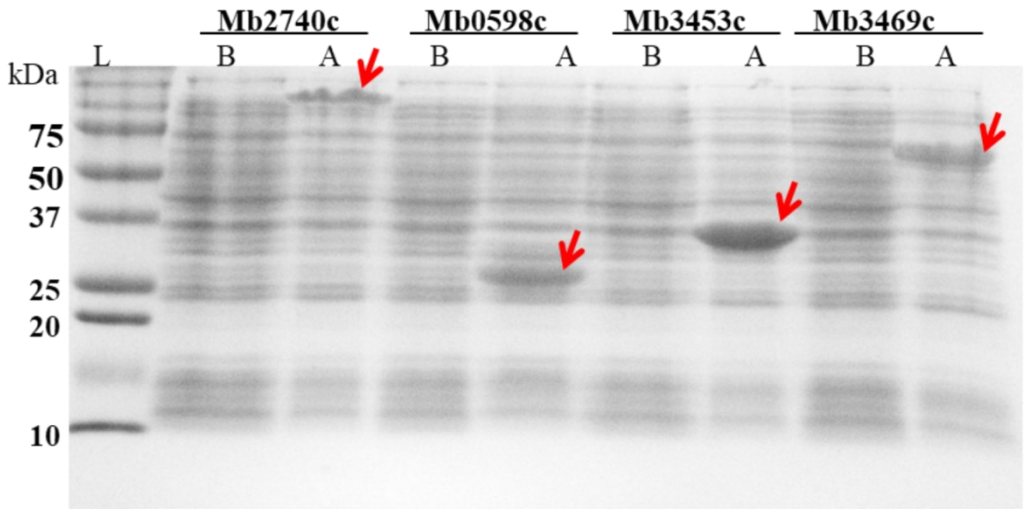


Figure 7-2. SDS-PAGE and Coomassie blue staining large scale production of the four selected proteins.

Each antigen is represented by two lanes, (B) before and (A) after IPTG induction. Expressed proteins are highlighted by the red arrows. Proteins are Mb2740c (72 kDa), Mb0598c (24 kDa) Mb3453c (38kDa) and Mb3469c (48 kDa). L=ladder

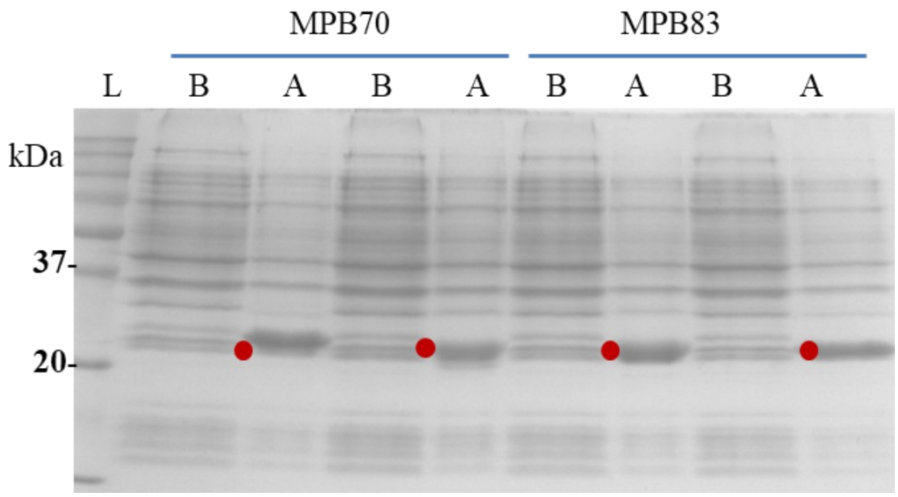


Figure 7-3. SDS-PAGE and Coomassie blue staining after the large scale production of MPB70 (19kDa) and MPB83 (22 kDa) proteins.

Each antigen is represented by two lanes, (B) before and (A) after IPTG induction in duplicates. Expressed proteins are highlighted by red dots. L=ladder

7.2.3 Confirmation of the Immunoreactivity of the Purified Proteins, Mb2740c, Mb0598c, Mb3453c, and Mb3469c

The reactivity of the four purified proteins with a positive serum at a dilution of 1/100 and a negative serum at a dilution of 1/100 was analyzed by dot blots (Figure 7-4). Mb2740c, Mb0598c, Mb3453c, and Mb3469c at three different blotted masses (0.5, 1, 2 µg) showed much stronger reactivity with the positive serum than the negative serum. Interestingly, Mb3469c stands out in immunoreactivity with remarkable visual discrimination of the positive sera from the negative sera.

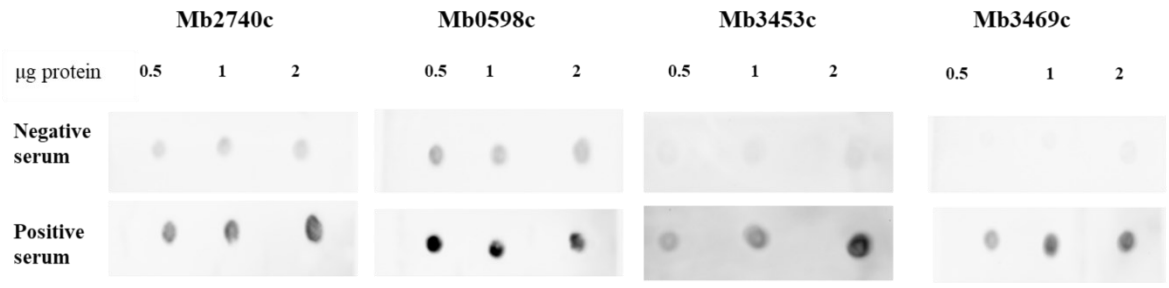


Figure 7-4. Dot blot analysis of the four selected antigens, against a negative and a positive serum.

Each antigen was blotted in three spots, each represents a different total antigenic mass (0.5, 1, 2 µg). The positive serum showed higher reactivity than the negative serum especially against the 2 µg blotted antigen.

7.2.4 Evaluation of Mb2740c, Mb0598c, Mb3453c and Mb3469c Antigens by ELISA for BTB Diagnostics

Before the ELISA work, multiple optimization procedures for the coating of the four antigens on the ELISA plates were carried out using four positive sera that demonstrated high reactivity in the microarray experiments compared to the negative sera. Several plates were tested with different binding capacities. The plates included Nunc MaxiSorp™, hydrophilic plates with enhanced binding. Medisorp™ plates, slightly hydrophilic plates with moderate binding, and Nunc cat#269620, non-treated polystyrene plates. The final decision was to use the untreated plates after the other types demonstrated strong background reactivity with negative sera. All antigens were tried and three proteins, Mb2740c, Mb0598c, and Mb3469c successfully demonstrated differences in reactivity against positive and negative serum. The protein, Mb3453c did not demonstrate a difference in the reactivity with the positive and negative serum in the ELISA optimization. As a result, Mb2740c, Mb0598c, and Mb3469c were tested in ELISA-based assays for the detection of antibodies in 50 sera

from *M. bovis*-naturally infected cattle and 110 sera from control non-infected cattle. Other optimization procedures included testing multiple antigen dilutions, between 0.5µg/ml to 10µg/ml in addition to the evaluation of the coating temperature at room temperature versus at 4°C. Different blocking reagents were experimented including horse serum (2.5-5%), gelatin, and fractionated casein. The result was choosing 2.5% horse serum due to its better performance. Sera dilutions ranging from 1/100 to 1/1600 were assessed. The selected sera dilutions for each antigen were as follows: 1:800 for Mb2740c, 1:1600 for Mb0598c, and 1:400 for Mb3469c. Overall, the final conditions selected after ELISA optimization are described in Chapter 3, Section 3.17.

The measured OD values of ELISA resulting from the reactivity of sera to each of the three antigens Mb2740c, Mb0598c, and Mb3469c were used to calculate the serological sensitivity and specificity by the ROC curve analysis (Figure 7-5 and Table 7-1).

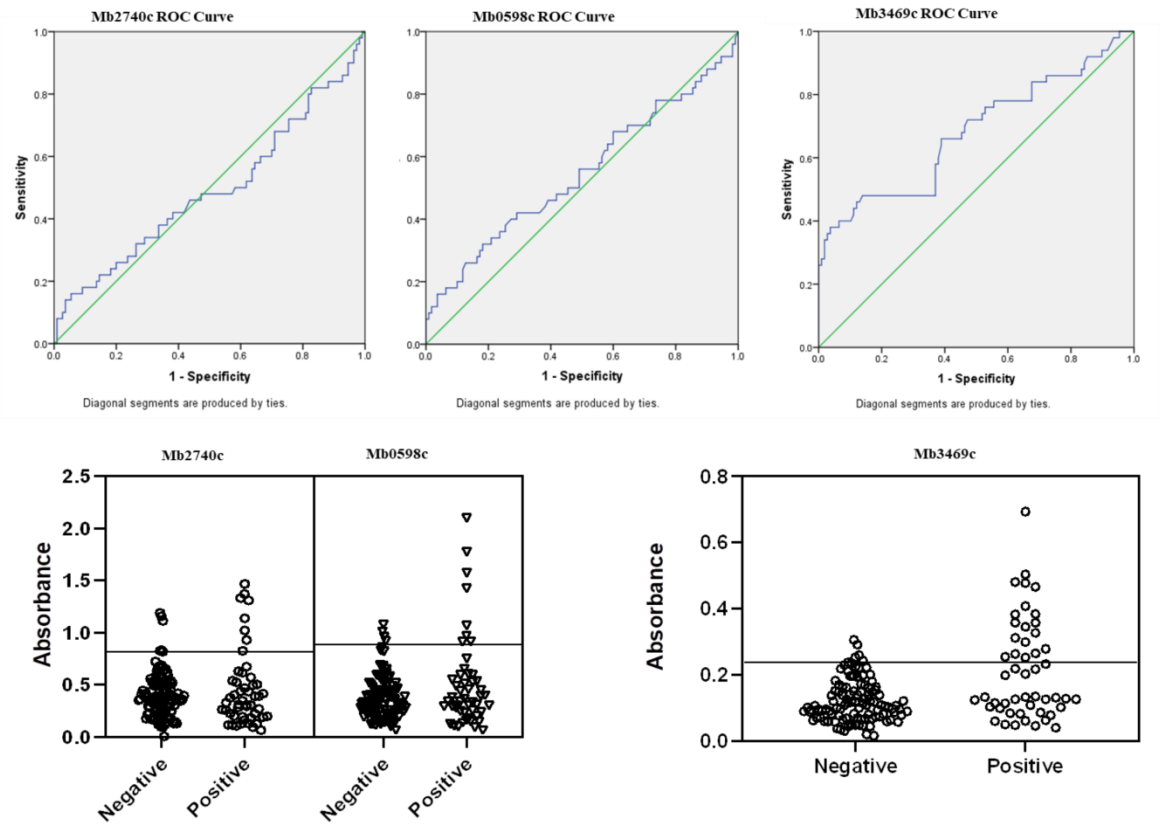


Figure 7-5. ROC curve analysis (upper panel) and plot distribution of ELISA absorbances (lower panel) for antigens Mb2740c, Mb0598c, and Mb3469c.

Fifty positive sera and 110 negative sera were used as primary antibodies. The plates were coated with the antigens at a concentration of 2µg/ml.

Table 7-1. The sensitivity and specificity of ELISA assays for Mb2740c, Mb0598c, and Mb3469c by ROC curve analysis.

Ag	cutoff	Sensitivity (%)	Specificity (%)	AUC#	Sig (p)
Mb2740c	0.822	16	94.5	0.53	0.832
Mb0598c	0.887	16	96.5	0.547	0.345
Mb3469c	0.231	40	93.5	0.68	0.000*

#AUC = Area under curve (for the ROC curve analysis)

The ROC curve analysis showed a statistically significant difference between the reactivity of Mb3469c with positive and negative sera. Mb2740c and Mb0598c; however, did not show a significant difference in reactivity with positive and negative sera.

7.2.5 ELISA Analysis of the Reactivity of Positive and Negative Sera to MPB70 and MPB83

MPB70 and MPB83 being important antigens in *M. bovis* serodiagnostic were tested for the comparison with the diagnostic performance of the novel antigens identified in this study. Both antigens were tested by ELISA for antibody detection using the same panels of positive and negative sera at 1:100 dilution (see details in Figure 7-5). The ELISA results showed that each of the two antigens was able to differentiate the positive group from the control group with statistical significance (Figure 7-6 and Table 7-2).

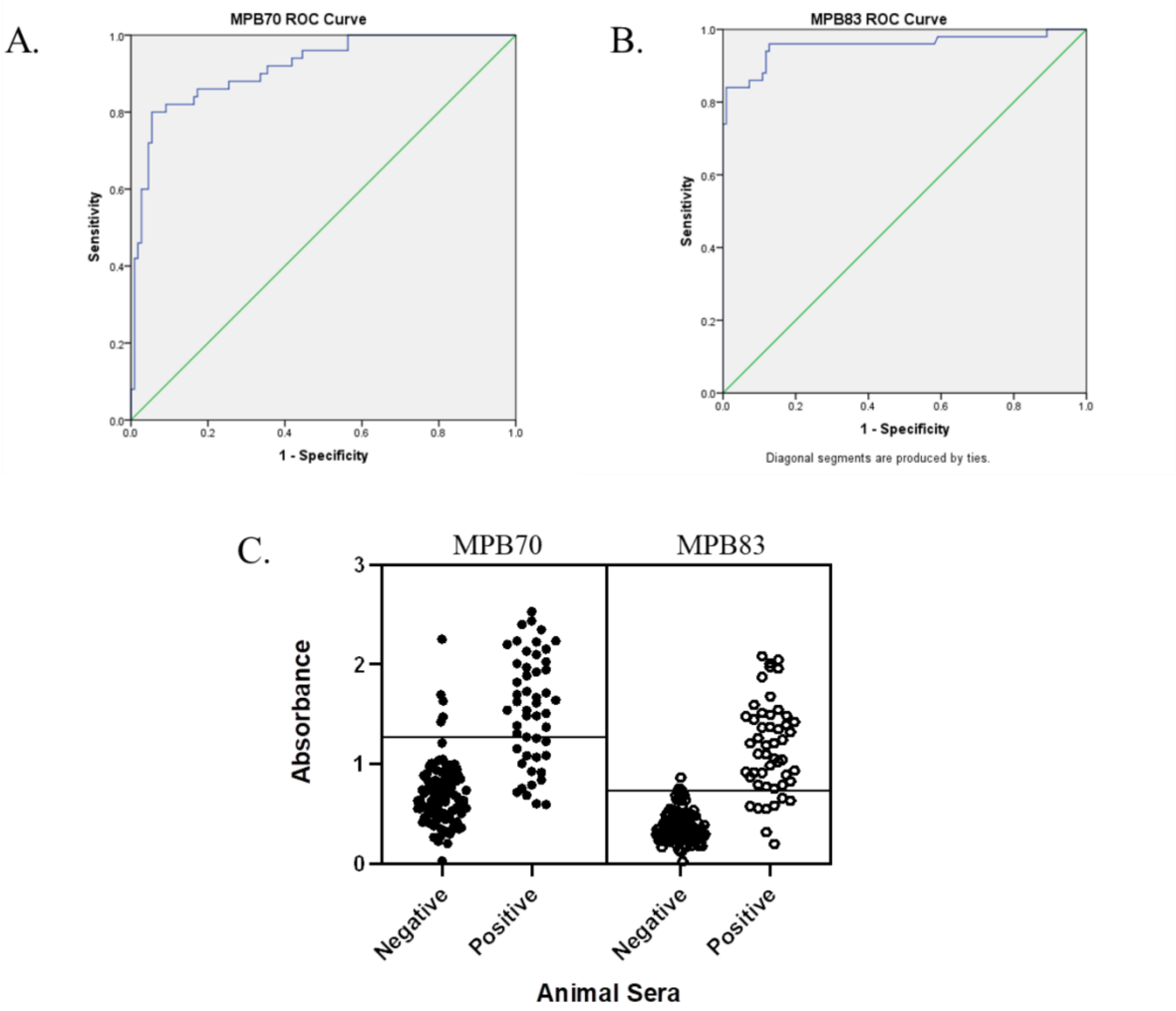


Figure 7-6. ROC curve analysis for MPB70 (A) and MPB83 (B) as well as dot plot distribution of the ELISA absorbances for both antigens (C).

Fifty positive sera and 110 negative sera were used as primary antibodies. The plates were coated with the antigens at a concentration of 2µg/ml.

Table 7-2. The sensitivity and specificity of ELISA assays for MPB70 and MPB83 by ROC curve analysis.

Ag	cutoff	Sensitivity (%)	Specificity (%)	AUC	Sig
MPB70	1.63	46	98	0.915	0.000*
MPB83	0.737	84	97.3	0.956	0.000*

7.2.6 Analysis of the Immune Response to Mb3469c in Bovine Hosts Infected with Different *M. bovis* Strains

The ELISA data obtained with the antigen Mb3469c (Figure 7-5) were further analyzed to compare the reactivity of the antigen with the sera from different infected cattle hosts. The 50 positive sera used in the ELISA were obtained from four outbreaks in the USA. Each outbreak represents infection with a different *M. bovis* strain. In this analysis, the positive sera were divided into 4 subgroups and named after the state at which the outbreak was identified: 1CO for Colorado outbreak 2010, 2CA for California outbreak 2011, 3IA for Iowa outbreak 2009, and 4MI for Michigan outbreak 2013. The reactivity of Mb3469c with each of the 4 positive groups was compared to the control group's reactivity (Figure 7-7 and Table S12, Appendix 7).

The results showed a significant difference in the reaction against Mb3469c between different sera groups as illustrated by the Kruskal-Wallis H test statistic, p*= 0.0001. When comparing each of the four positive serum groups individually to the control serum group, only the two sera groups, 1CO and 2CA, showed significantly higher reactivity to the antigen than the control group. Sera groups 3IA and 4MI showed no significant difference in immunoreactivity when compared to controls.

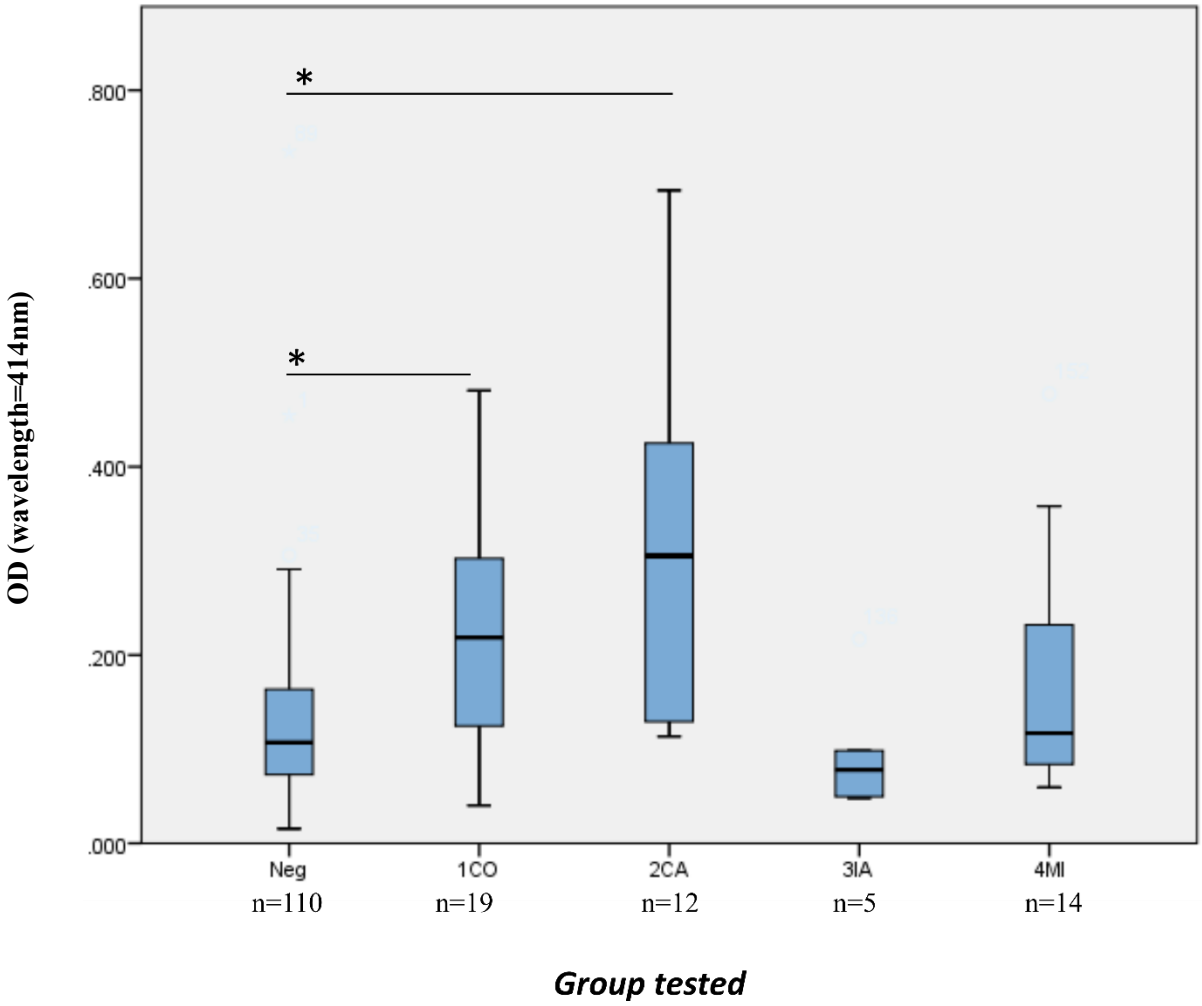


Figure 7-7. Box plot analysis showing the Mb3469c ELISA reactivity against sera from different populations of cattle.

The reactivity is displayed against negative control sera (Neg, n=110) and four positive sera groups: (1CO, n=19), (2CA, n=12), (3IA, n=5), (4MI, n=14). Two sera groups 1CO and 2CA showed significantly higher reactivity to Mb3469c than the negative sera. ($p^* < 0.05$).

7.2.7 Evaluation of Mb3453c by Dot Blots for BTB Diagnostics

The Mb3453c antigen, immobilized onto nitrocellulose-based membranes at 0.01 µg/spot, was tested for reactivity to the same positive and negative sera panels used earlier in the ELISA. The data obtained from the quantitative analysis of the dot blot images with ImageJ followed by the ROC curve analysis showed that there is a statistically significant difference between the positive and negative sera in their reactivity to Mb3453c (Figure 7-8 and Table 7-3).

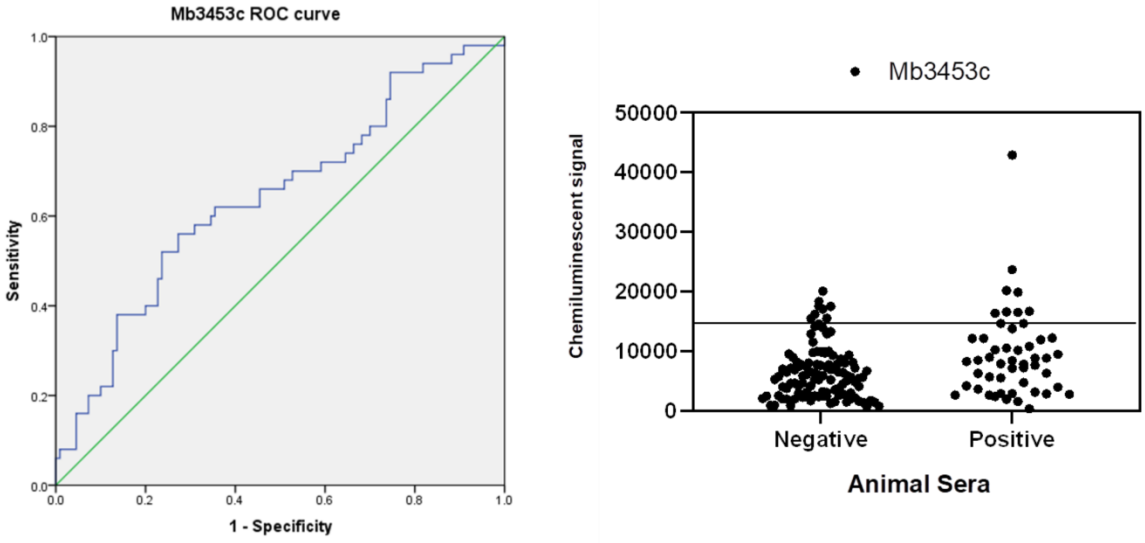


Figure 7-8. ROC curve analysis (left) and dot plot distribution of the chemiluminescent signals (right) for the Mb3453c antigen.

Fifty positive sera and 110 negative sera were used as primary antibodies. The antigens were spotted at 0.01 µg/spot.

Table 7-3. The sensitivity and specificity of Mb3453c by ROC curve analysis.

Ag	cutoff	Sensitivity (%)	Specificity (%)	AUC	Sig (p)
Mb3453c	14605	20	92.7	0.642	0.004*

The reactivity of sera to Mb3453c was compared between the control group and each positive serum subgroup (ICO, 2CA, 3IA, and 4MI) (Figure 7-9 and Table S11, Appendix 7). This comparison reveals a significant difference in the reactivity to the antigen Mb3453c between the control group and the 3IA subgroup (p=0.0086*) as well as between the control group and the 4MI subgroup (p=0.02*). No statistical difference was found between either the ICO or 2CA subgroups and the control group concerning the reactivity of sera to the antigen.

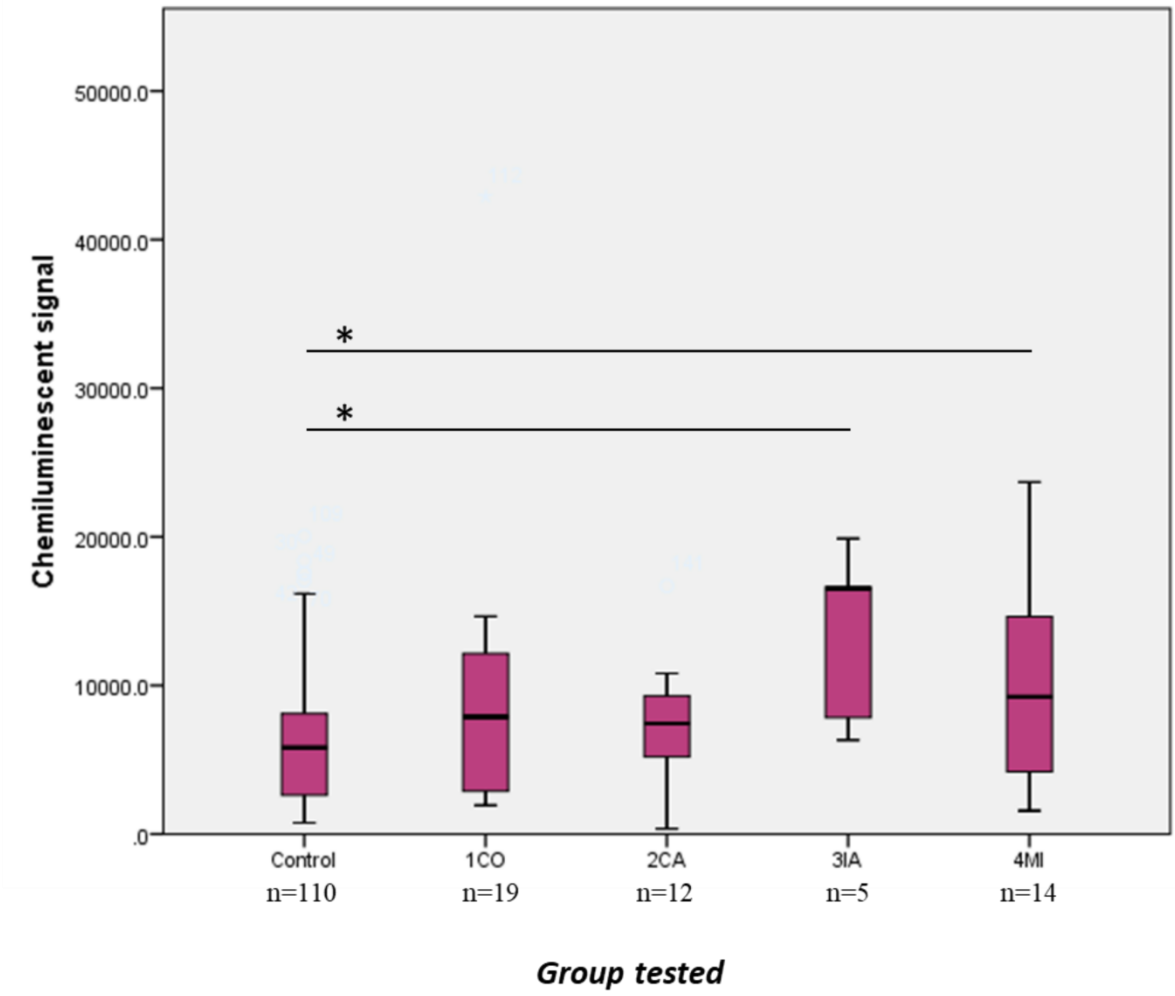


Figure 7-9. Box plot analysis showing the Mb3453c dot blot reactivity against sera from different populations of cattle.

The reactivity is displayed against negative control sera (Neg, n=110) and four positive sera groups: (1CO, n=19), (2CA, n=12), (3IA, n=5), (4MI, n=14). Two sera groups 3IA and 4MI showed significantly higher reactivity to Mb3453c than the negative sera (p significant <0.05).

7.3 Discussion

In this part of the study, four novel antigens were selected for the evaluation of their potential serodiagnostic capabilities based on the results of the microarray studies (Chapter 6). Three of those antigens, Mb2740c, Mb3453c, and Mb3469c were identified from the present bioinformatic analysis as possible extracellular antigens, and the antigen, Mb0598c, was identified from the review of previous proteomic analysis of PPDB.

To understand more about these proteins, *the M. tuberculosis* H37Rv genome was examined for homologues of the four selected antigens. Rv2721c, Rv0583c, Rv3419c, and Rv3439c in *M. tuberculosis* H37Rv were found 100% identical in sequence to the antigens Mb2740c, Mb0598c, Mb3453c, and Mb3469c, respectively. Information on these protein homologues was limited. Rv2721c, Rv0583c, the homologues of Mb2740c, Mb0598c respectively, are reported as the proteins that function in the cell wall and cell processes (142). Rv2712c, the homologue of Mb2740c, is a possible conserved transmembrane alanine and glycine-rich protein whereas Rv0583c, the homologue of Mb0598c, is a conserved lipoprotein. Both have been identified in the membrane fraction of *M. tuberculosis* H37Rv using nanoLC-MS/MS (142). Both have also been identified in culture filtrates of *M. tuberculosis* H37Rv and contain a signal peptide (69). The third protein, Rv3439c, the homologue of Mb3469c, is a conserved hypothetical alanine-proline-rich protein with unknown function. The last protein, Rv3419c, homologue of Mb3453c, is a probable

glycoprotease that functions in intermediary metabolism and respiration. The latter two proteins were not reported to be present in the membrane or *in vitro* culture filtrate of *M. tuberculosis*. The transposon mutagenesis study shows that the four proteins are non-essential genes for *in vitro* growth of *M. tuberculosis* H37Rv (194). The two proteins Mb3469c and Mb3453c have not been detected in the *in vitro* studies to date; nevertheless, the detection of a humoral immune response to these proteins indicates that they are expressed *in vivo* under specific conditions. The four proteins are well conserved in both *M. bovis* and MTB and thus are potentially promising antigens for BTB and TB serodiagnostics.

In this part of the study, ligation independent cloning was carried out using the pLIC-CHis vector derived from the pET30a vector (Novagen). This cloning system was previously used by our group and proved efficient independent of the target gene sequence (120, 129, 195). The system was used to generate constructs for large scale protein expression in *E. coli.* This was a necessity to provide enough protein to allow for the evaluation of these proteins in ELISA-based assays.

During the optimization of ELISA, the protein Mb3453c showed a high background and failed to demonstrate a difference between the reactivity of the positive and negative sera. One explanation may be that the protein, being expressed as an insoluble protein in *E. coli*, failed to coat the plate surface effectively. A limitation related to the binding capacity of the plates may also be suspected. Other explanations are the possible disruption or the concealment of important epitopes when the antigen binds to the plate surface or the incorrect folding of the protein during purification. All these factors may have interfered with antigen−antibody reaction. An effort was made to resolve the solubility problem by solubilizing the protein using SDS and then diluting the SDS concentration in the protein

before proceeding to the coating process as described by Lechtzier *et al.* (196). Despite that, the trial did not prove successful, thus hindering the use of Mb3453c in the ELISA application.

The results of ELISA with Mb3469c confirmed the microarray screening results and showed a significant difference in the reactivity of sera to the antigen between naturally infected animals and the controls. This significant difference was shown by the ROC curve analysis with a sensitivity of 40% and a specificity of 93.5% (p=0.000*). Also, similar to the microarray screening results, Mb3469c showed no statistically significant difference in reactivity with the sera when comparing experimentally infected cattle to the controls (Table S12). The difference between the reactivity of each positive serum subgroup and the negative controls to Mb3469c was highlighted (Table S12 and Figure 7-7). Sera subgroups from two different geographical locations, 1CO and 2CA, showed a significant difference in the reactivity compared to the control sera group. This indicates that Mb3469c is capable of detecting cattle infected with different strains of *M. bovis*. This reaction pattern is similar to the observation that the immune response of cattle to MPB70 varies in different geographic locations where infections are caused by different *M. bovis* strains (18, 197). Unpublished observations in the CFIA also indicated that the cattle infected by different *M. bovis* strains displayed different reactivity of sera to MPB70.

The ELISA results obtained with Mb2740c and Mb0598c demonstrated a lack of a significant demarcation between the positive and control serum samples unlike the microarray screening results (Figure 7-5). This discrepancy could be attributed to several factors. One factor is the different surfaces used to immobilize an antigen may have resulted in a different degree of concealing important epitopes. Another possible explanation is that

the possible incorrect folding of the protein antigen may interfere with antigen−antibody recognition. Additionally, ELISA blocking buffer containing 2.5% horse serum may have created a confounding factor. The horse serum may have cross-reactive IgG that leads to a high background reaction. On the other hand, a specific commercial blocking reagent, which lacks mammalian proteins, was used in the microarray screening. Furthermore, the microarray screening method relies on the detection of the fluorescent signal that offers the advantage of higher sensitivity than the colorimetric ELISA method.

The Mb3469c ELISA results gave a sensitivity of 40%, which is comparable to the sensitivity of 46 % obtained with the MPB70-based ELISA in this study. The sensitivity of MPB70-based ELISA was similarly found by the study of Waters *et al.* (197) that showed the sensitivity of the MPB70 and MPB83-based ELISA on NVSL sera at 48.4%. It should be noted that MBP83 resulted in a sensitivity of 84% in ELISA, much higher than that obtained with Mb3469c. This is in contrast to the microarray screening experiments showing that Mb3469c was comparable to both MPB70 and MPB83 in the reactivity with positive sera but had a higher statistical significance in differentiating positive sera from the controls (Table 6-3). This suggests that certain immunoassay formats may be adopted to realize the full potentials of diagnostic antigens. The contrasting results may also be explained by the potential exposure of the tested cattle to the TST antigens prior to serum withdrawal, which could boost the antibody production and thus enhance the sensitivity of MPB83-based ELISA. This is consistent with a previous study of Casal *et al.* (198) showing an anamnestic increase from ~24% to ~70% in the sensitivity of commercial ELISA tests following the TST. Given that MPB83 is a major component in the PPDB used in the TST and is also an

antigen used in the commercial ELISA tests, the immune boosting by the PPDB is likely the factor responsible for the elevated sensitivity of the MPB83-based ELISA observed here.

In contrast to ELISA, the dot blot assay with Mb3453c demonstrated a significant demarcation between sera from naturally infected cattle and the control sera ($p=0.004$). Comparing the reactivity of sera to Mb3453c between the control sera and different sera subgroups revealed that reactivity to this antigen was significant in two subgroups, 3IA and 4MI (Figure 7-9). This shows that Mb3453c is capable of detecting infections with specific *M. bovis* strains, in contrast to Mb3469c that detects animals infected with different *M. bovis* strains. Mb3453c unexpectedly did not demonstrate a significant reactivity to the sera of experimentally infected animals compared to the controls, despite that the p was near statistical significant ($p=0.09$) (Table S11). This observation is different from the results obtained with the microarray experiments (Chapter 6). One possible explanation of this difference is that the fluorescent-based microarray screening method offered a higher sensitivity of detection and is more optimized.

Interestingly, the two novel protein antigens identified here, Mb3469c and Mb3453c had no homologues detected as membrane proteins or as extracellular proteins in *M. tuberculosis* studies (Table 4-2). They were, however, identified by the bioinformatics analysis as potential extracellular antigens. This demonstrates the capability of bioinformatics to identify potentially immunogenic proteins based on their extracellular localization prediction. Both proteins have no detectable signal peptides, illustrating that the prediction of extracellular proteins based solely on the presence of a signal peptide has some limitations. This is consistent with the study of Antelmann *et al.* (149) showing that ~50% of the extracellular proteome lacked a signal peptide. Thus, it is postulated that these proteins

may be exposed to the immune system due to direct interaction with the B-cells or due to macrophage antigen processing and presentation.

The results of the present investigation show that the two novel antigens, Mb3469c and Mb3453c, in addition to two known antigens, MPB70 and MPB83, are capable of differentiating positive sera from the negative controls with statistical significance. These results are consistent with the findings of Amadori *et al.* (199), which demonstrates that a panel of antigens is needed rather than a single antigen for the serological diagnosis of *M. bovis*. Similar findings were seen in the present study with the evaluation of different antigens. Notably, the sera from 1CO and 2CA subgroups showed significantly higher reactivity to Mb3469c than the other two sera subgroups (3IA and 4MI) whereas the reactivity to Mb3453c was significantly elevated in the 3IA and 4MI subgroups. This indicates that each antigen can be used to detect infections with specific strains of *M. bovis*. The use of both novel antigens in BTB serodiagnostics can increase the sensitivity and detect more infected cattle herds.

Chapter 8. General Discussion and Conclusion

Diagnosis of *M. bovis* infection in cattle has long been a concern to veterinarians and researchers. Serological tests that assess the humoral immune response to specific antigens are used as ancillary methods for BTB diagnostics. The role of humoral immunity in tuberculosis infection has been highlighted over the course of the disease despite the causative agent being an intracellular pathogen. Studies on *M. tuberculosis* have confirmed the presence of specific antibodies in association with the mycobacterial infection (46). Antigens secreted by the mycobacterial pathogen during infection may interact directly with the B-cells, followed by the secretion of specific immunoglobulins. Also, the bacterium may exist at some point in the disease in the extracellular space exposing it to the action of immunoglobulins (51). The serological tests mainly detect specific antibodies to the mycobacterial infection. Antibody-production is presumed to occur in the later stage of the disease with a high level of antibodies existing in the serious or advanced stages of the disease (200). These tests have been hampered by their lower sensitivity that is as low as 9% in some infected animal populations such as Mexican cattle (18).

In this study, three aims were established with the final objective of identifying and evaluating novel antigens that elicit an antibody response in BTB infection for the improvement of BTB diagnostics in cattle. First of all, the study aimed for choosing candidate antigens that might elicit specific immunoglobulins during infection. This work demonstrated that PSORTb, a bioinformatic tool for the prediction of potentially extracellular antigens, can be successfully applied to the identification of potentially immunogenic proteins in *M. bovis*. The candidate list was further expanded by conducting a proteomic study to identify proteins secreted into the culture supernatant by live *M. bovis* and

by reviewing previously published proteomic studies on the PPD. This effort led to the identification of a total of 188 protein candidates, including several known *M. bovis* antigens, for the subsequent investigation. Previous studies aiming to identify immunogenic proteins that induce an antibody response have mostly focused on the analysis of the PPDB. Studies, including those by Casal *et al.* (102), Souza *et al.* (107), and Waters *et al.* (197), deployed a limited number of proteins that were originally discovered in the PPDB for evaluation in TB diagnostic tests.

Secondly, the study aimed to clone and express the ORFs coding for the candidate proteins followed by using these proteins to profile the humoral immune response to *M. bovis* infection. The initial attempt to PCR-amplify some ORFs from the *M. bovis* genome proved to be unsuccessful mainly due to the template sequence with high GC content. The development of a robust system that allows the amplification of such high GC content ORFs without the need to optimize the conditions for each ORF became critically important to the success of the study. The long length of some target ORFs sequences was another obstacle to the successful PCR amplification. To date, most studies reported the amplification of GC-rich gene targets of not larger than 1kb (171, 174, 175, 182, 183). Also, Taq DNA polymerase used in those studies was not favoured because it is a low fidelity enzyme. This study has devised the solutions to the amplification problems, including fragmentation of larger sequences to keep the template sequence length below 2.5 kb and manipulation of the cycling conditions to perform the annealing and extension steps at higher temperatures using a 2-step PCR. Additionally, the slow thermal cycler ramp speed for amplification of GC-rich sequences was used in accordance with the procedures described by Frey at al. (183); however, the dGTP analogue used in the latter study was deliberately omitted to avoid the

disruption of the gene structure. The use of a high fidelity enzyme such as the PrimeStar GXL enzyme was necessary for the expression of a cloned ORF to generate the corresponding recombinant protein. This enzyme was demonstrated to have superior performance compared to other attempted enzymes. All the efforts led to the development of a final protocol that allowed the successful amplification of 186 out of 208 (~89.4%) selected ORFs.

To clone ORFs coding for the candidate proteins in a high throughput manner for the microarray screening, the *in vivo* cloning strategy was applied in line with the study of Davies *et al.* (191). This strategy is robust and more reliable in contrast to the time-consuming conventional cloning technique that relies on the restriction digestion and the ligation. The system exhibited a high success rate by cloning 180 out of 186 (~97%) of the DNA sequences. This was followed by the protein expression using the cell-free expression system, also known as *in vitro* transcription and translation (IVTT). This methodology was adopted by Davies *et al.* (191), Liang *et al.* (201), and Li *et al.* (202). The system was selected because of its efficiency in the expression of potentially toxic proteins and membrane proteins as well as its applicability in the high throughput protein expression (203). In this study, the majority of the recombinant proteins were expressed using the *E. coli* based IVTT system. Several proteins that failed to be expressed in *E. coli* based IVTT system were expressed using the mammalian cell-free expression system. Similarly, Labaer *et al.* (172). employed the mammalian cell-free expression system effectively to express the *Pseudomonas* genes of high GC-content, a feature commonly encountered in *M. bovis* genes. Overall, this study demonstrated an 88% success rate in protein expression by IVTT. Unlike the studies of Davies *et al.* (191), Liang *et al.* (201), Li *et al.* (202), all the recombinant

proteins generated in this study were purified using a high throughput system that was developed here to facilitate the screening of immunogenic proteins. Purification of recombinant proteins eliminated the need for the pre-adsorption of serum thus reduced the high background signal caused by the non-specific reaction of serum antibodies. As a result, the interpretation of the results of the microarray screening became easier.

This study represents the first in the literature to use protein microarray screening for the identification of novel antigens for the diagnosis of *M. bovis*-infection in cattle. Microarray studies have addressed other mycobacterial infections. The screening of MTB infection in humans using protein microarray was carried out by Zhou *et al.* (204). The latter study constructed a protein microarray to screen 409 TB proteins. They found fourteen strong seropositive reactions including the reactions to six novel proteins, 38 kDa antigen, LprG, Rv1566c, Rv1623c, MPT64, and HspX. The HspX protein, which had been previously investigated in BTB diagnostics (165), was included as one of the candidates selected for this study. The other 5 proteins had homologues in *M. avium* species with a degree of similarity between 55% and 80%. This high degree of homology is the reason for the exclusion of these proteins from the present candidate list. The exposure of cattle to the environmental *M. avium* could lead to the induction of antibodies specific to these proteins homologues; these antibodies are considered cross-reacting antibodies and can cause false-positive results in BTB serodiagnostics.

In the present investigation, all the protein candidates were expressed with the constructs containing the corresponding ORFs derived from the *M. bovis* genome. This experimental design is superior to the study of Li *et al.* (202). The latter study was carried out to screen the antigens for *M. avium* paratuberculosis (MAP), a pathogen in cattle, by

using ready to express clones derived from *M. tuberculosis*. Li *et al.* (202) acknowledged that the MTB proteins when compared to MAP proteins show an average of 62% identity at the amino acid level (range from 19% to 100%). Contrastingly, the present study constructed the expression clones using the ORFs from the *M. bovis* genome and thus provided a more reliable identification of antigens for BTB diagnostics.

The present study, in both the microarray screening and the ELISA application, has the advantage over other previous studies concerning the sera used for evaluation of the BTB serological diagnostic tests. Souza *et al.* (107) defined the positive cattle serum samples as SICCT positive and determined the sensitivity and specificity of ELISA based on the results of SICTT, while Marassi *et al.* (109) defined the positive samples as both PCR and SICCT positive or SICCT positive alone when calculating the sensitivity and specificity in ELISA. On the other hand, this study obtained positive sera of cattle from the NVSL that had been confirmed to have BTB infection by extensive postmortem testing including culture, histopathologic examination, and molecular tests. Similarly, Waters *et al.* (197) also chose to use positive sera of cattle confirmed by postmortem testing in order to obtain an accurate and reliable estimation of the sensitivity and specificity in ELISA.

Comparison of the reactivity to selected antigens on protein microarray among the sera from naturally infected adult and juvenile cattle, experimentally infected, and control cattle reveals an interesting phenomenon. The juvenile group exhibited a superior sera immunoreactivity compared to other groups, evidenced by a higher number of significantly reacting antigens that differentiated juvenile infected group from the control group. In addition, four of these antigens demonstrated significantly higher reactivity to the sera from the juvenile group compared to the adult group.

Microarray screening of protein candidates showed that 13 antigens were capable of eliciting an immune response leading to the successful differentiation between the infected cattle and the controls. In addition to the two known antigens MPB70 and MPB83, four of the 13 antigens, Mb2740c, Mb0598c, Mb3453c, and Mb3469c showed the highest reactivity to positive sera and had the potential for use in BTB serodiagnostic tests. Like MPB70 and MPB83, the four novel antigens were able to further differentiate the different positive groups, including adult naturally infected, juvenile naturally infected, and experimentally infected, from the control group.

To date, studies seeking novel antigens for the serodiagnosis of *M. bovis* infection have been limited. In a recent study, Lyashchenko *et al.* (205) tested a number of MTB recombinant proteins in search of novel *M. bovis* antigens. By using the multi-antigen print immunoassay (MAPIA) for screening, they identified 12 reactive antigens. On top of those 12 antigens were MPB70 and MPB83. Similar results were obtained from the microarray screening experiment conducted here, showing that MPB70 and MPB83 were highly reactive to positive sera. Statistical analysis of the microarray screening results was performed to compare the reactivity to the antigens between the positive and negative sera. This was in contrast to the study of Lyashchenko *et al.* (205) in which positive sera were compared to a single negative serum concerning their reactivity to the antigens. Also, the latter study did not further evaluate the discovered antigens in ELISA or other immunoassays for BTB diagnostic applications.

In line with the third aim, the four novel antigens identified from the microarray screening experiment were evaluated further for the serological detection of BTB infection using ELISA and dot blot assay. The results obtained demonstrated that the two novel

antigens, Mb3469c and Mb3453c, had the immunodiagnostic potential. The two antigens can be added to the current list of BTB diagnostic antigens including MPB70 and MPB83. Both Mb3469c and MPB70 exhibited comparable sensitivities and similarity in the detection of infected cattle. However, the evaluation of Mb3453c by the dot blot assay demonstrated an added value in BTB serodiagnosis where 7 out of 50 positive cattle (14%) were discovered only by using Mb3453c but not MPB70 or Mb3469c. Therefore, the use of Mb3453c as an additional serodiagnostic antigen could help the discovery of more infected cattle.

Interestingly, Mb3469c and Mb3453c displayed a difference in the reactivity to test sera obtained from four outbreaks caused by different *M. bovis* strains. The gene sequences coding for these two antigens are found well conserved among multiple *M. bovis* strains by searching the public databases for sequence similarity. Each of the two antigens was able to detect two different outbreaks. This observation is consistent with the heterogeneous humoral immune response of cattle against *M. bovis* proteins reported previously (186, 205). Also, different sensitivities of the IDEXX TB ELISA were reported when cattle sera from different geographic locations of infections caused by different strains were tested (197). One study looked for an explanation for that, by checking the genes of the diagnostic proteins used in the IDEXX TB ELISA, MPB70 and MPB83, as well as regulatory genes. They found more than 80% conserved sequence for each gene and reported that the variation does not explain geographic differences in ELISA sensitivity (18).

The two antigens, Mb3469c and Mb3453c, are expected to be beneficial if they are included in the panel of antigens used in the BTB diagnostics. It should be noted that the two novel antigens exhibited a slightly lower sensitivity and specificity than MPB70 and MPB83 in ELISA. Despite that, it may be possible that the two antigens perform well in situations

where MPB70 and MPB83 are not able to give good test results. For example, Mb3469c and Mb3453c could be experimented with Mexican cattle serum samples where the sensitivity of the IDEXX TB ELISA with the current antigens is 9% (18).

In conclusion, this study demonstrated the success of applying bioinformatics to the identification of potentially immunogenic proteins encoded by the *M. bovis* genome. The study also showed the usefulness of proteomics in the identification of immunogenic proteins from *M. bovis*. A high throughput system was established and optimized to express recombinant proteins from the ORFs of *M. bovis*. This system is composed of an optimized PCR amplification protocol, the *in vivo* cloning strategy, and the *in vitro* recombinant protein expression. Further, a method for high throughput purification of recombinant proteins was developed here to allow the successful optimization and implementation of protein microarray for the screening of immunogenic proteins. The microarray screening of 129 recombinant proteins identified 13 antigens in total that were capable of detecting a significant difference in the reactivity of sera between the infected cattle and the controls. Six of these antigens, MPB70, MPB83, Mb3469c, Mb3453c, Mb0598c, and Mb2740c displayed high reactivity to positive sera and were further evaluated in ELISA and chemiluminescent-based dot-blot immunoassay for serodiagnostic applications. The two novel antigens Mb3469c and Mb3453c, together with the two known antigens MPB70 and MPB83, demonstrated significant immunodiagnostic potentials in differentiating between infected cattle and the controls. In addition, the two novel antigens had the ability to detect BTB outbreaks in some specific geographical locations.

Future research can be directed to the evaluation of Mb3469c and Mb3453c, together with MPB70 and MPB83, for their ability to detect antibodies in sera groups from different

geographic locations other than the USA. This may help to further expose their diagnostic potentials. The evaluation of these antigens for TB diagnosis in humans or different animal species other than cattle such as farmed cervids and badgers could also be undertaken. Another recommendation for the use of these antigens in BTB diagnostics is the development of an immunoassay in more sensitive platforms that combine the four antigens with fluorescence or chemiluminescence-based detection methods. Other immunoreactive antigens discovered in the microarray screening experiment may also be considered for further evaluation of their diagnostic potentials.

Chapter 9. References

1. World Health Organization. 2019. Global Tuberculosis Report. World Health Organization, Geneva, Switzerland. Accessed 2019-12-20. www.who.int/tb/publications/global_report/en/
2. OIE. 2016. OIE - World Organisation for Animal Health. Terrestrial Animal Health. Accessed 2019-12-20. http://www.oie.int/
3. OIE. 2009. Bovine tuberculosis, p. 1–16. *In* OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animal. Accessed 2019-12-20. https://www.oie.int/standard-setting/terrestrial-manual/
4. WHO. 2016. Global Tuberculosis report. Geneva, Switzerland. Accessed 2017-08-10. https://apps.who.int/iris/handle/10665/250441.
5. Olea-Popelka F, Muwonge A, Perera A, Dean AS, Mumford E, Erlacher-Vindel E, Forcella S, Silk BJ, Ditiu L, El Idrissi A, Raviglione M, Cosivi O, LoBue P, Fujiwara PI. 2017. Zoonotic tuberculosis in human beings caused by *Mycobacterium bovis*-a call for action. Lancet Infect Dis 17:e21–e25.
6. Müller B, Dürr S, Alonso S, Hattendorf J, Laisse CJM, Parsons SDC, van Helden PD, Zinsstag J. 2013. Zoonotic *Mycobacterium bovis*-induced tuberculosis in humans. Emerg Infect Dis 19:899–908.
7. Monaghan ML, Doherty ML, Collins JD, Kazda JF, Quinn PJ. 1994. The tuberculin test. Vet Microbiol 40:111–124.
8. Karolemeas K, de la Rua-Domenech R, Cooper R, Goodchild A V, Clifton-Hadley RS, Conlan AJK, Mitchell AP, Hewinson RG, Donnelly CA, Wood JLN, McKinley TJ. 2012. Estimation of the relative sensitivity of the comparative tuberculin skin test

in tuberculous cattle herds subjected to depopulation. PLoS One 7:e43217.

9. Wood PR, Corner LA, Plackett P. 1990. Development of a simple, rapid in vitro cellular assay for bovine tuberculosis based on the production of gamma interferon. Res Vet Sci 49:46–9.

10. Shingadia D, Novelli V. 2008. The tuberculin skin test: a hundred, not out? Arch Dis Child 93:189–90.

11. Cho YS, Jang YB, Lee SE, Cho JY, Ahn JM, Hwang I, Heo E, Nam HM, Cho D, Her M, Jean YH, Jung SC, Kim JM, Lee HS, Lee K, Belisle JT. 2015. Short communication: Proteomic characterization of tuberculin purified protein derivative from *Mycobacterium bovis*. ResVetSci 101:117–119.

12. Cho YS, Dobos KM, Prenni J, Yang H, Hess A, Rosenkrands I, Andersen P, Ryoo SW, Bai GH, Brennan MJ, Izzo A, Bielefeldt-Ohmann H, Belisle JT. 2012. Deciphering the proteome of the in vivo diagnostic reagent "purified protein derivative" from *Mycobacterium tuberculosis*. Proteomics 12:979–991.

13. Borsuk S, Newcombe J, Mendum TA, Dellagostin OA, McFadden J. 2009. Identification of proteins from tuberculin purified protein derivative (PPD) by LC-MS/MS. Tuberculosis 89:423–430.

14. Cho YS, Dobos KM, Prenni J, Yang H, Hess A, Andersen P, Ryoo SW, Bai G, Brennan MJ, Izzo A, Bielefeldt-ohmann H, Belisle JT. 2013. Deciphering the proteome of the in vivo diagnostic reagent "purified protein derivative" from *Mycobacterium tuberculosis*. Proteomics 12:979–991.

15. Yang H, Kruh-Garcia NA, Dobos KM. 2012. Purified protein derivatives of tuberculin--past, present, and future. FEMS Immunol Med Microbiol 66:273–80.

16. Jenkins AO, Gormley E, Gcebe N, Fosgate GT, Conan A, Aagaard C, Michel AL,

Rutten VPMG. 2018. Cross reactive immune responses in cattle arising from exposure to *Mycobacterium bovis* and non-tuberculous mycobacteria. Prev Vet Med 152:16–22.

17. de la Rua-Domenech R, Goodchild AT, Vordermeier HM, Hewinson RG, Christiansen KH, Clifton-Hadley RS. 2006. Ante mortem diagnosis of tuberculosis in cattle: A review of the tuberculin tests, γ-interferon assay and other ancillary diagnostic techniques. Res Vet Sci 81:190–210.
18. Trost B, Stuber T, Surujballi O, Nelson J, Robbe-Austerman S, Smith NH, Desautels L, Tikoo SK, Griebel P. 2016. Investigation of the cause of geographic disparities in IDEXX ELISA sensitivity in serum samples from *Mycobacterium bovis*-infected cattle. Nat Publ Gr 6:22763.
19. Ayele WY, Neill SD, Zinsstag J, Weiss MG, Pavlik I. 2004. Bovine tuberculosis: an old disease but a new threat to Africa. Int J Tuberc Lung Dis 8:924–937.
20. Cosivi O, Grange JM, Daborn CJ, Raviglione MC, Fujikura T, Cousins D, Robinson RA, Huchzermeyer HFAK, De Kantor I, Meslin FX. 1998. Zoonotic tuberculosis due to *Mycobacterium bovis* in developing countries. Emerg Infect Dis 4:59–70.
21. Thoen CO, Lobue PA, Isabel De K. 2010. Why has zoonotic tuberculosis not received much attention? Int J Tuberc Lung Dis 14:1073–1074.
22. Cosivi O, Meslin F-X, Daborn CJ, Grange JM. 1995. Epidemiology of *Mycobacterium bovis* infection in animals and humans, with particular reference to Africa. Rev sci tech Off int Epiz 14:733–746.
23. Centers for Disease Control and Prevention. 2011. *Mycobacterium bovis* (Bovine Tuberculosis) in Humans. Accessed 2020-01-22. http://www.cdc.gov/tb/topic/basics/signsandsymptoms.htm
24. OIE. Bovine Tuberculosis General Disease Information Sheet. Accessed 2020-02-17.

http://www.oie.int/fileadmin/Home/eng/Media_Center/docs/pdf/Disease_cards/BOVINE-TB-EN.pdf

25. Debi U, Ravisankar V, Prasad KK, Sinha SK, Sharma AK. 2014. Abdominal tuberculosis of the gastrointestinal tract: revisited. World J Gastroenterol 20:14831–40.

26. Ramos DF, Silva PEA, Dellagostin OA. 2015. Diagnosis of bovine tuberculosis: review of main techniques. Braz J Biol 75:830–837.

27. Bovine Tuberculosis - Animals - Canadian Food Inspection Agency. Accessed 2017-08-10. https://www.inspection.gc.ca/

28. WHO. 2015. The End TB Strategy. Accessed 2017-10-23. http://www.who.int/tb/End_TB_brochure.pdf?ua=1

29. Lees VW. 2004. Learning from outbreaks of bovine tuberculosis near Riding Mountain National Park: Applications to a foreign animal disease outbreak. Can Vet J 45:28–34.

30. Fitzgerald SD, Kaneene JB. 2012. Wildlife Reservoirs of Bovine Tuberculosis Worldwide: Hosts, Pathology, Surveillance, and Control. Vet Pathol 50:488–499.

31. European Food Safety Authority (EFSA) and European Centre for Disease Prevention and Control (ECDC). 2016. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2015. EFSA J 14:20449.

32. Miller RS, Sweeney SJ. 2013. *Mycobacterium bovis* (bovine tuberculosis) infection in North American wildlife: current status and opportunities for mitigation of risks of further infection in wildlife populations. Epidemiol Infect 141:1357–70.

33. Lees VW, Copeland S, Rousseau P. 2003. Bovine tuberculosis in elk (Cervus elaphus

manitobensis) near Riding Mountain National Park, Manitoba, from 1992 to 2002. Can Vet J 44(10): 830–831.

34. 2012. Backgrounder - Canadian Food Inspection Agency Bovine Tuberculosis Surveillance in the Riding Mountain Eradication Area - Animals - Canadian Food Inspection Agency. Accessed 2017-11-02. http://www.inspection.gc.ca/animals/terrestrial-animals/diseases/reportable/tuberculosis/backgrounder/eng/1354387734000/13543

35. Conlan AJK, Brooks Pollock E, McKinley TJ, Mitchell AP, Jones GJ, Vordermeier M, Wood JLN. 2015. Potential Benefits of Cattle Vaccination as a Supplementary Control for Bovine Tuberculosis. PLoS Comput Biol 11.

36. Velayati AA, Farnia P. 2017. The Species Concept, p. 1–16. *In* Atlas of *Mycobacterium tuberculosis*. Elsevier Academic Press, London, UK.

37. Grange JM, Yates MD, De Kantor IN. 1996. WHO/EMC/ZOO/96.4 Guidelines for speciation within the *Mycobacterium tuberculosis* complex. World Health Organization 2, 1–23.

38. Niederweis M, Danilchanka O, Huff J, Hoffmann C, Engelhardt H. 2010. Mycobacterial outer membranes: in search of proteins. Trends Microbiol 18:109–116.

39. Fu LM, Fu-Liu CS. 2002. Is *Mycobacterium tuberculosis* a closer relative to Gram-positive or Gram–negative bacterial pathogens? Tuberculosis (Edinb) 82:85–90.

40. Thoen CO, Barletta RG. 2014. Pathogenesis of tuberculosis caused by *Mycobacterium bovis*, p. 51–62. *In* Charles O. Thoen, James H. Steele, JBK (ed.), Zoonotic Tuberculosis: *Mycobacterium bovis* and other pathogenic mycobacteria, 3rd ed. John Wiley & Sons, Inc, Chichester, UK.

41. Yuk J-M, Jo E-K. 2014. Host immune responses to mycobacterial antigens and their

implications for the development of a vaccine to control tuberculosis. Clin Exp Vaccine Res 3:155–67.

42. Hackam DJ, Rotstein OD, Zhang WJ, Demaurex N, Woodside M, Tsai O, Grinstein S. 1997. Regulation of phagosomal acidification. Differential targeting of Na +/H + exchangers, Na +/K +-ATPases, and vacuolar-type H +-ATPases. J Biol Chem 272:29810–29820.

43. Bach H, Papavinasasundaram KG, Wong D, Hmama Z, Av-Gay Y. 2008. *Mycobacterium tuberculosis* Virulence Is Mediated by PtpA Dephosphorylation of Human Vacuolar Protein Sorting 33B. Cell Host Microbe 3:316–322.

44. Wong D, Bach H, Sun J, Hmama Z, Av-Gay Y. 2011. *Mycobacterium tuberculosis* protein tyrosine phosphatase (PtpA) excludes host vacuolar-H +-ATPase to inhibit phagosome acidification. Proc Natl Acad Sci U S A 108:19371–19376.

45. Dannenberg AM. 1989. Immune Mechanisms in the Pathogenesis of Pulmonary Tuberculosis. Clin Infect Dis 11:S369–S378.

46. Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A. 2011. Immunological biomarkers of tuberculosis. Nat Rev Immunol 11:343–354.

47. Gordon S, Keshav S, Stein M. 1994. BCG-Induced Granuloma Formation in Murine Tissues. Immunobiology 191:369–377.

48. Subbian S, Tsenova L, Kim M-J, Wainwright HC, Visser A, Bandyopadhyay N, Bader JS, Karakousis PC, Murrmann GB, Bekker L-G, Russell DG, Kaplan G. 2015. Lesion-Specific Immune Response in Granulomas of Patients with Pulmonary Tuberculosis: A Pilot Study. PLoS One 10:e0132249.

49. Hegde S, Rithesh KB, Baroudi K, Umar D. 2014. Tuberculous lymphadenitis: early diagnosis and intervention. J Int oral Heal JIOH 6:96–8.

50. Neill SD, Pollock JM, Bryson DB, Hanna J. 1994. Pathogenesis of *Mycobacterium bovis* infection in cattle. Vet Microbiol 40:41–52.

51. Kozakiewicz L, Phuah J, Flynn J, Chan J. 2013. The Role of B Cells and Humoral Immunity in *Mycobacterium tuberculosis* Infection, p. 225–250. *In* M. Divangahi (ed.) (ed.), The new paradigm of immunity to tuberculosis. Springer Science+Business Media, New York, NY.

52. Hussain R, Shiratsuchi H, Phillips M, Ellner J, Wallis RS. 2001. Opsonizing antibodies (IgG1) up-regulate monocyte proinflammatory cytokines tumour necrosis factor-alpha (TNF-a) and IL-6 but not anti-inflammatory cytokine IL-10 in mycobacterial antigen-stimulated monocytes–implications for pathogenesis. Clin Exp Immunol 123:210–218.

53. Public Health Agency of Canada. Pathogen Safety Data Sheets: Infectious Substances – Mycobacterium spp. Accessed 2017-11-13. www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/mycobacterium.html

54. Dean GS, Rhodes SG, Coad M, Whelan AO, Cockle PJ, Clifford DJ, Hewinson RG, Vordermeier HM. 2005. Minimum Infective Dose of *Mycobacterium bovis* in Cattle. Infect Immun 73:6467–6471.

55. Palmer M V, Waters WR. 2006. Advances in bovine tuberculosis diagnosis and pathogenesis: What policy makers need to know. Vet Microbiol 112:181–190.

56. O'Reilly LM, Daborn CJ. 1995. The epidemiology of *Mycobacterium bovis* infections in animals and man: A review. Tuber Lung Dis 76:1–46.

57. World Health Organization. 2006. Accessed 2020-02-12. https://www.who.int/whr/2006/en/

58. Koo HC, Park YH, Ahn J, Waters WR, Palmer M V., Hamilton MJ, Barrington G, Mosaad AA, Park KT, Jung WK, Hwang IY, Cho S-NN, Shin SJ, Davis WC. 2005. Use of rMPB70 protein and ESAT-6 peptide as antigens for comparison of the enzyme-linked immunosorbent, immunochromatographic, and latex bead agglutination assays for serodiagnosis of bovine tuberculosis. J Clin Microbiol 43:4498–4506.

59. Wiker HG. 2009. MPB70 and MPB83 - Major antigens of *Mycobacterium bovis*. Scand J Immunol 69:492–499.

60. Haga S, Nakagawa M, Nagai S, Miura K, Honda M. 1992. Purification of MPB70 and Production of Specific Monoclonal Antibodies. Hybridoma 11:483–492.

61. Hasløv K, Andersen B, Bentzon MW. 1987. Biological Activity in Sensitized Guinea Pigs of MPB 70, a Protein Specific for Some Strains of *Mycobacterium bovis* BCG. Scand J Immunol 26:445–454.

62. Harboe M, Nagai S. 1984. MPB70, a unique antigen of *Mycobacterium bovis* BCG. Am Rev Respir Dis 129:444–452.

63. Harboe M, Wiker HG, Duncan JR, Garcia MM, Dukes TW, Brooks BW, Turcotte C, Nagai S. 1990. Protein G-based enzyme-linked immunosorbent assay for anti-MPB70 antibodies in bovine tuberculosis. J Clin Microbiol 28:913–921.

64. Ulstrup JC, Jeansson S, Wiker HG, Harboe M. 1995. Relationship of secretion pattern and MPB70 homology with osteoblast- specific factor 2 to osteitis following *Mycobacterium bovis* BCG vaccination. Infect Immun 63:672–675.

65. Saïd-Salim B, Mostowy S, Kristof AS, Behr MA. 2006. Mutations in *Mycobacterium tuberculosis* Rv0444c, the gene encoding anti-SigK, explain high level expression of MPB70 and MPB83 in *Mycobacterium bovis*. Mol Microbiol 62:1251–1263.

66. Mcnair J, Corbett DM, Girvin RM, Mackie DP, Pollock JM, McNair J. 2001. Characterization of the Early Antibody Response in Bovine Tuberculosis: MPB83 is an Early Target with Diagnostic Potential. Scand J Immunol 53:365–371.

67. Fifis T, Corner LA, Rothel JS, Wood PR. 1994. Cellular and Humoral Immune Responses of Cattle to Purified *Mycobacterium bovis* Antigens. Scand J Immunol 39:267–274.

68. Green LR, Jones CC, Sherwood AL, Garkavi I V., Cangelosi GA, Thacker TC, Palmer M V., Waters WR, Rathe C V. 2009. Single-antigen serological testing for bovine tuberculosis. Clin Vaccine Immunol 16:1309–1313.

69. Målen H, Berven FS, Fladmark KE, Wiker HG. 2007. Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. Proteomics 7:1702–1718.

70. Dillon DC, Alderson MR, Day CH, Bement T, Campos-Neto A, Skeiky YA, Vedvick T, Badaro R, Reed SG, Houghton R. 2000. Molecular and immunological characterization of *Mycobacterium tuberculosis* CFP-10, an immunodiagnostic antigen missing in *Mycobacterium bovis* BCG. J Clin Microbiol 38:3285–90.

71. Skjøt RL, Oettinger T, Rosenkrands I, Ravn P, Brock I, Jacobsen S, Andersen P. 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 68:214–20.

72. Ulrichs T, Anding P, Porcelli S, Kaufmann SH, Munk ME. 2000. Increased numbers of ESAT-6- and purified protein derivative-specific gamma interferon-producing cells in subclinical and active tuberculosis infection. Infect Immun 68:6073–6.

73. Ulrichs T, Munk ME, Mollenkopf H, Behr-Perst S, Colangeli R, Gennaro ML, Kaufmann SHE. 1998. Differential T cell responses to *Mycobacterium tuberculosis*

ESAT6 in tuberculosis patients and healthy donors. Eur J Immunol 28:3949–3958.

74. Pollock JM, Andersen P. 1997. The Potential of the ESAT-6 Antigen Secreted by Virulent Mycobacteria for Specific Diagnosis of Tuberculosis. J Infect Dis 175:1251–1254.

75. Mori T, Sakatani M, Yamagishi F, Takashima T, Kawabe Y, Nagao K, Shigeto E, Harada N, Mitarai S, Okada M, Suzuki K, Inoue Y, Tsuyuguchi K, Sasaki Y, Mazurek GH, Tsuyuguchi I. 2004. Specific Detection of Tuberculosis Infection. Am J Respir Crit Care Med 170:59–64.

76. Andersen, P; Munk, M E; Pollock, J M; Doherty TM. 2000. Specific immune-based diagnosis of tuberculosis. Lancet 356:1099–1104.

77. Slogotskaya L, Bogorodskaya E, Ivanova D, Sevostyanova T. 2018. Comparative sensitivity of the test with tuberculosis recombinant allergen, containing ESAT6-CFP10 protein, and Mantoux test with 2 TU PPD-L in newly diagnosed tuberculosis children and adolescents in Moscow. PLoS One 13:e0208705.

78. van Pinxteren LA, Ravn P, Agger EM, Pollock J, Andersen P. 2000. Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. Clin Diagn Lab Immunol 7.

79. Wiker HG, Harboe M. 1992. The antigen 85 complex: a major secretion product of *Mycobacterium tuberculosis*. Microbiol Rev 56:648–61.

80. Warrier T, Tropis M, Werngren J, Diehl A, Gengenbacher M, Schlegel B, Schade M, Oschkinat H, Daffe M, Hoffner S, Eddine AN, Kaufmann SHE. 2012. Antigen 85C inhibition restricts *Mycobacterium tuberculosis* growth through disruption of cord factor biosynthesis. Antimicrob Agents Chemother 56:1735–43.

81. Hetland G, Wiker HG. 1994. Antigen 85C on *Mycobacterium bovis*, BCG and *M.*

tuberculosis promotes monocyte-CR3-mediated uptake of microbeads coated with mycobacterial products. Immunology 82:445–9.

82. Jackson M, Raynaud C, Laneelle M-A, Guilhot C, Laurent-Winter C, Ensergueix D, Gicquel B, Daffe M. 1999. Inactivation of the antigen 85C gene profoundly affects the mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope. Mol Microbiol 31:1573–1587.

83. Content J, de la Cuvellerie A, De Wit L, Vincent-Levy-Frébault V, Ooms J, De Bruyn J. 1991. The genes coding for the antigen 85 complexes of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG are members of a gene family: cloning, sequence determination, and genomic organization of the gene coding for antigen 85-C of *M. tuberculosis*. Infect Immun 59:3205–12.

84. Johnson S, Brusasca P, Lyashchenko K, Spencer JS, Wiker HG, Bifani P, Shashkina E, Kreiswirth B, Harboe M, Schluger N, Gomez M, Gennaro ML. 2001. Characterization of the secreted MPT53 antigen of *Mycobacterium tuberculosis*. Infect Immun 69:5936–9.

85. Wiker HG, Michell SL, Hewinson RG, Spierings E, Nagai S, Harboe M. 1999. Cloning, expression and significance of MPT53 for identification of secreted proteins of *Mycobacterium tuberculosis*. Microb Pathog 26:207–219.

86. de Souza GA, Leversen NA, Målen H, Wiker HG. 2011. Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. J Proteomics 75:502–510.

87. Sidders B, Pirson C, Hogarth PJ, Hewinson RG, Stoker NG, Vordermeier HM, Ewer K. 2008. Screening of highly expressed mycobacterial genes identifies Rv3615c as a useful differential diagnostic antigen for the *Mycobacterium tuberculosis* complex.

Infect Immun 76:3932–3939.

88. Alito A, McNair J, Girvin RM, Zumarraga M, Bigi F, Pollock JM, Cataldi A. 2003. Identification of *Mycobacterium bovis* antigens by analysis of bovine T-cell responses after infection with a virulent strain. Brazilian J Med Biol Res 36:1523–1531.

89. Dosanjh NS, Rawat M, Chung J-H, Av-Gay Y. 2005. Thiol specific oxidative stress response in *Mycobacteria*. FEMS Microbiol Lett 249:87–94.

90. Jaeger T, Budde H, Flohé L, Menge U, Singh M, Trujillo M, Radi R. 2004. Multiple thioredoxin-mediated routes to detoxify hydroperoxides in *Mycobacterium tuberculosis*. Arch Biochem Biophys 423:182–191.

91. Hu Y, Coates ARM. 2009. Acute and persistent *Mycobacterium tuberculosis* infections depend on the thiol peroxidase TPX. PLoS One 4.

92. Brosch R, Gordon S V., Garnier T, Eiglmeier K, Frigui W, Valenti P, Dos Santos S, phanie Duthoy S, line Lacroix C, Garcia-Pelayo C, Inwald JK, Golby P, Nuñ ez Garcia J, Glyn Hewinson R, Behr MA, Quail MA, Churcher C, Barrell BG, Parkhill J, Cole ST, Duthoy S, Lacroix C, Garcia-Pelayo C, Inwald JK, Golby P, Garcia JN, Hewinson RG, Behr MA, Quail MA, Churcher C, Barrell BG, Parkhill J, Cole ST. 2007. Genome plasticity of BCG and impact on vaccine efficacy. Proc Natl Acad Sci 104:5596–5601.

93. Hebert AM, Talarico S, Yang D, Durmaz R, Marrs CF, Zhang L, Foxman B, Yang Z. 2007. DNA Polymorphisms in the pepA and PPE18 Genes among Clinical Strains of *Mycobacterium tuberculosis*: Implications for Vaccine Efficacy. Infect Immun 75:5798–5805.

94. Bottai D, Batoni G, Esin S, Florio W, Brancatisano FL, Favilli F, Maisetta G, Campa M. 2006. The secretion antigen SA5K has a role in the adaptation of *Mycobacterium*

bovis bacillus Calmette–Guérin to intracellular stress and hypoxia. Microbes Infect 8:2254–2261.

95. Mustafa AS, Skeiky YA, Al-Attiyah R, Alderson MR, Hewinson RG, Vordermeier HM. 2006. Immunogenicity of *Mycobacterium tuberculosis* Antigens in *Mycobacterium bovis* BCG-Vaccinated and *M. bovis*-Infected Cattle. Infect Immun 74:4566–4572.

96. Skeiky YAW, Ovendale PJ, Jen S, Alderson MR, Dillon DC, Smith S, Wilson CB, Orme IM, Reed SG, Campos-Neto A. 2000. T Cell Expression Cloning of a *Mycobacterium tuberculosis* Gene Encoding a Protective Antigen Associated with the Early Control of Infection. J Immunol 165:7140–7149.

97. Strain SAJ, Mcnair J, Mcdowell SWJ. 2011. Bovine tuberculosis: A review of diagnostic tests for *M. bovis* infection in cattle. N. Ireland. Accessed 2017-10-16. https://www.daera-ni.gov.uk/sites/default/files/publications/dard/afbi-literature-review-tb-review-diagnostic-tests-cattle.pdf

98. Wadhwa A, Hickling GJ, Eda S. 2012. Opportunities for improved serodiagnosis of human tuberculosis, bovine tuberculosis, and paratuberculosis. Vet Med Int 2012.

99. Spickler AR. 2019. Zoonotic Tuberculosis in Mammals, including Bovine and Caprine Tuberculosis. Iowa, USA. Accessed 2020-02-17. https://www.cfsph.iastate.edu/Factsheets/pdfs/bovine_tuberculosis.pdf

100. Wood PR, Jones SL. 2001. BOVIGAM (TM): an in vitro cellular diagnostic test for bovine tuberculosis. Tuberculosis 81:147–155.

101. Vordermeier HM, Whelan A, Cockle PJ, Farrant L, Hewinson RG, Palmer N, Hewinson RG. 2001. Use of Synthetic Peptides Derived from the Antigens ESAT-6 and CFP-10 for Differential Diagnosis of Bovine Tuberculosis in Cattle. Clin Diagn

Lab Immunol 8:571–578.

102. Casal C, Bezos J, Díez-Guerrier A, Álvarez J, Romero B, de Juan L, Rodriguez-Campos S, Vordermeier M, Whelan A, Hewinson RG, Mateos A, Domínguez L, Aranaz A. 2012. Evaluation of two cocktails containing ESAT-6, CFP-10 and Rv-3615c in the intradermal test and the interferon-γ assay for diagnosis of bovine tuberculosis. Prev Vet Med 105:149–154.

103. Schiller I, Waters WR, Vordermeier HM, Nonnecke B, Welsh M, Keck N, Whelan A, Sigafoose T, Stamm C, Palmer M, Thacker T, Hardegger R, Marg-Haufe B, Raeber A, Oesch B. 2009. Optimization of a whole-blood gamma interferon assay for detection of *Mycobacterium bovis*-infected cattle. Clin Vaccine Immunol 16:1196–202.

104. Ryan TJ, Buddle BM, De Lisle GW. 2000. An evaluation of the gamma interferon test for detecting bovine tuberculosis in cattle 8 to 28 days after tuberculin skin testing. Res Vet Sci 69:57–61.

105. Coad M, Hewinson RG, Clifford D, Vordermeier HM, Whelan AO. 2007. Influence of skin testing and blood storage on interferon-γ production in cattle affected naturally with *Mycobacterium bovis*. Vet Rec 160:660–662.

106. Inwald J, Hinds J, Palmer S, Dale J, Butcher PD, Hewinson RG, Gordon S V. 2003. Genomic Analysis of *Mycobacterium tuberculosis* Complex Strains Used for Production of Purified Protein Derivative. J Clin Microbiol 41:3929–3932.

107. Souza II, Melo ES, Ramos CA, Farias TA, Osório ALA, Jorge KS, Vidal CE, Silva AS, Silva MR, Pellegrin AO, Araújo FR. 2012. Screening of recombinant proteins as antigens in indirect ELISA for diagnosis of bovine tuberculosis. Springerplus 1:77.

108. Aagaard C, Govaerts M, Meikle † V, Vallecillo AJ, Gutierrez-Pabello JA, Suarez-Güemes F, Mcnair J, Cataldi A, Espitia C, Andersen P, Pollock JM. 2006. Optimizing

Antigen Cocktails for Detection of *Mycobacterium bovis* in Herds with Different Prevalences of Bovine Tuberculosis: ESAT6-CFP10 Mixture Shows Optimal Sensitivity and Specificity. J Clin Microbiol 44:4326–4335.

109. Marassi CD, Medeiros L, McNair J, Lilenbaum W. 2011. Use of recombinant proteins MPB70 or MPB83 as capture antigens in ELISAs to confirm bovine tuberculosis infections in Brazil. Acta Trop 118:101–104.

110. Lin M, Sugden EA, Jolley ME, Stilwell K. 1996. Modification of the *Mycobacterium bovis* extracellular protein MPB70 with fluorescein for rapid detection of specific serum antibodies by fluorescence polarization. Clin Diagn Lab Immunol 3:438–43.

111. Surujballi O, Lutze-Wallace C, Turcotte C, Savic M, Stevenson D, Romanowska A, Monagle W, Berlie-Surujballi G, Tangorra E. 2009. Sensitive diagnosis of bovine tuberculosis in a farmed cervid herd with use of an MPB70 protein fluorescence polarization assay. Can J Vet Res 73:161–166.

112. Jolley ME, Nasir MS, Surujballi OP, Romanowska A, Renteria TB, De la Mora A, Lim A, Bolin SR, Michel AL, Kostovic M, Corrigan EC. 2007. Fluorescence polarization assay for the detection of antibodies to *Mycobacterium bovis* in bovine sera. Vet Microbiol 120:113–121.

113. Ngandolo BNR, Müller B, Diguimbaye-Djaïbe C, Schiller I, Marg-Haufe B, Cagiola M, Jolley M, Surujballi O, Akakpo AJ, Oesch B, Zinsstag J. 2009. Comparative assessment of fluorescence polarization and tuberculin skin testing for the diagnosis of bovine tuberculosis in Chadian cattle. Prev Vet Med 89:81–89.

114. Nasr E, Marwah M, Melika L, Tammam A, Gorge S. 2016. Comparison of tuberculin skin test and lateral flow rapid test for detection of bovine tuberculosis in dairy cattle. J Appl Vet Sci 1:21–27.

115. Robbe-Austerman S, Turcotte C. 2014. New and current approaches for isolation, identification, and genotyping of *Mycobacterium bovis*, p. 89–98. *In* Zoonotic Tuberculosis. John Wiley & Sons, Inc, Chichester, UK.

116. Orloski K, Robbe-Austerman S, Stuber T, Hench B, Schoenbaum M. 2018. Whole Genome Sequencing of *Mycobacterium bovis* Isolated From Livestock in the United States, 1989-2018. Front Vet Sci 5:253.

117. Price-Carter M, Brauning R, de Lisle GW, Livingstone P, Neill M, Sinclair J, Paterson B, Atkinson G, Knowles G, Crews K, Crispell J, Kao R, Robbe-Austerman S, Stuber T, Parkhill J, Wood J, Harris S, Collins DM. 2018. Whole Genome Sequencing for Determining the Source of *Mycobacterium bovis* Infections in Livestock Herds and Wildlife in New Zealand. Front Vet Sci 5:272.

118. Andrievskaia O, Duceppe M-O LD. 2018. Genome Sequences of Five *Mycobacterium bovis* Strains Isolated from Farmed Animals and Wildlife in Canada. Genome Announc 6:e00258-18.

119. Andrievskaia O, Turcotte C, Berlie-Surujballi G, Battaion H, Lloyd D. 2018. Genotypes of *Mycobacterium bovis* strains isolated from domestic animals and wildlife in Canada in 1985–2015. Vet Microbiol 214:44–50.

120. Dan H, Balachandran A, Lin M. 2009. A pair of ligation-independent Escherichia coli expression expression vectors for rapid addition of a polyhistidine affinity tag to the N- or C-termini of recombinant proteins. J Biomol Tech 20:241–248.

121. Peabody MA, Laird MR, Vlasschaert C, Lo R, Brinkman FSL. 2016. PSORTdb: Expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. Nucleic Acids Res 44:D663–D668.

122. Malone KM, Farrell D, Stuber TP, Schubert OT, Aebersold R, Robbe-Austerman S,

Gordon S V. 2017. Updated Reference Genome Sequence and Annotation of *Mycobacterium bovis* AF2122/97. Genome Announc 5.

123. Bryksin A V, Matsumura I. 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. Biotechniques 48:463–5.

124. Sambrook J, Russell DW. 2001. Molecular Cloning: A Laboratory Manual.Third Edit. Cold Spring Harbor Laboratory Press., Cold Spring Harbor, NY.

125. Rennie B, Filion LG, Smart N. 2010. Antibody response to a sterile filtered PPD tuberculin in *M. bovis* infected and *M. bovis* sensitized cattle. BMC Vet Res 6.

126. Lin M, Trottier E, Pasick J. 2005. Antibody responses of pigs to defined Erns fragments after infection with classical swine fever virus. Clin Diagn Lab Immunol 12:180–6.

127. Wang L, Lin M. 2007. Identification of IspC, an 86-kilodalton protein target of humoral immune response to infection with *Listeria monocytogenes* serotype 4b, as a novel surface autolysin. J Bacteriol 189:2046–54.

128. Ronholm J, Wang L, Hayashi I, Sugai M, Zhang Z, Cao X, Lin M. 2012. The *Listeria monocytogenes* serotype 4b autolysin IspC has N-acetylglucosaminidase activity. Glycobiology 22:1311–1320.

129. Zhang CXY, Brooks BW, Huang H, Pagotto F, Lin M. 2016. Identification of Surface Protein Biomarkers of *Listeria monocytogenes* via Bioinformatics and Antibody-Based Protein Detection Tools. Appl Environ Microbiol 82:5465–76.

130. 2019. WHO | Global tuberculosis report. World Health Organization. Accessed 2019-12-20. https://www.who.int/tb/publications/global_report/en/

131. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. 2001. B-cell activation by armed helper T cells, p. . *In* Janeway CA Jr, Travers P, Walport M, et al. (ed.),

Immunobiology: The Immune System in Health and Disease. 5th edition. Garland Science, New York.

132. Roperto S, Varano M, Russo V, Lucà R, Cagiola M, Gaspari M, Ceccarelli DM, Cuda G, Roperto F. 2017. Proteomic analysis of protein purified derivative of *Mycobacterium bovis*. J Transl Med 15:1–8.

133. Vizcaíno C, Restrepo-Montoya D, Rodríguez D, Niño LF, Ocampo M, Vanegas M, Reguero MT, Martínez NL, Patarroyo ME, Patarroyo MA. 2010. Computational Prediction and Experimental Assessment of Secreted/Surface Proteins from *Mycobacterium tuberculosis* H37Rv. PLoS Comput Biol 6:e1000824.

134. McCann JR, McDonough JA, Sullivan JT, Feltcher ME, Braunstein M. 2011. Genome-Wide Identification of *Mycobacterium tuberculosis* Exported Proteins with Roles in Intracellular Growth. J Bacteriol 193:854–861.

135. Singh PP, Parra M, Cadieux N, Brennan MJ. 2008. A comparative study of host response to three *Mycobacterium tuberculosis* PE_PGRS proteins. Microbiology 154:3469–3479.

136. Wolfe LM, Mahaffey SB, Kruh NA, Dobos KM. 2010. Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. J Proteome Res 9:5816–26.

137. Tiwari B, Ramakrishnan UM, Raghunand TR. 2015. The *Mycobacterium tuberculosis* protein pair PE9 (Rv1088)-PE10 (Rv1089) forms heterodimers and induces macrophage apoptosis through Toll-like receptor 4. Cell Microbiol 17:1653–1669.

138. Chatrath S, Gupta VK, Garg LC. 2014. The PGRS domain is responsible for translocation of PE_PGRS30 to cell poles while the PE and the C-terminal domains localize it to the cell wall. FEBS Lett 588:990–994.

139. Mawuenyega KG, Forst C V., Dobos KM, Belisle JT, Chen J, Bradbury EM,

Bradbury ARMM, Chen X. 2005. *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. Mol Biol Cell 16:396–404.

140. Mukamolova G V., Turapov OA, Young DI, Kaprelyants AS, Kell DB, Young M. 2002. A family of autocrine growth factors in *Mycobacterium tuberculosis*. Mol Microbiol 46:623–635.

141. Gu S, Chen J, Dobos KM, Bradbury EM, Belisle JT, Chen X. 2003. Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain. Mol Cell Proteomics 2:1284–1296.

142. Xiong Y, Chalmers MJ, Gao FP, Cross TA, Marshall AG. 2005. Identification of *Mycobacterium tuberculosis* H37Rv integral membrane proteins by one-dimensional gel electrophoresis and liquid chromatography electrospray ionization tandem mass spectrometry. J Proteome Res 4:855–861.

143. Siroy A, Mailaender C, Harder D, Koerber S, Wolschendorf F, Danilchanka O, Wang Y, Heinz C, Niederweis M. 2008. Rv1698 of *Mycobacterium tuberculosis* represents a new class of channel-forming outer membrane proteins. J Biol Chem 283:17827–37.

144. Saint-Joanis B, Demangel C, Jackson M, Brodin P, Marsollier L, Boshoff H, Cole ST. 2006. Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of *Mycobacterium tuberculosis*, increases β-lactam susceptibility and virulence. J Bacteriol 188:6669–6679.

145. Sampson SL. 2011. Mycobacterial PE/PPE proteins at the host-pathogen interface. Clin Dev Immunol 2011:497203.

146. Demangel C, Brodin P, Cockle PJ, Brosch R, Majlessi L, Leclerc C, Cole ST. 2004. Cell envelope protein PPE68 contributes to *Mycobacterium tuberculosis* RD1 immunogenicity independently of a 10-kilodalton culture filtrate protein and ESAT-6.

Infect Immun 72:2170–6.

147. Saini NK, Baena A, Ng TW, Venkataswamy MM, Kennedy SC, Kunnath-Velayudhan S, Carreño LJ, Xu J, Chan J, Larsen MH, Jacobs WR, Porcelli SA. 2016. Suppression of autophagy and antigen presentation by *Mycobacterium tuberculosis* PE_PGRS47. Nat Microbiol 1:16133.

148. Nair S, Ramaswamy PA, Ghosh S, Joshi DC, Pathak N, Siddiqui I, Sharma P, Hasnain SE, Mande SC, Mukhopadhyay S. 2009. The PPE18 of *Mycobacterium tuberculosis* Interacts with TLR2 and Activates IL-10 Induction in Macrophage. J Immunol 183:6269–6281.

149. 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 Res 11:1484–1502.

150. Nielsen H, Tsirigos KD, Brunak S, von Heijne G, Konstantinos, Tsirigos D, Søren Brunak, Gunnar Von Heijne. 2019. A Brief History of Protein Sorting Prediction. Protein J 38:200–216.

151. Roshanak S, Tabatabaei Yazdi F, Shahidi F, Javadmanesh A, Movaffagh J. 2020. Comparison of Different Signal Sequences to Use for Periplasmic Over-Expression of Buforin I in Escherichia coli: An In Silico Study. Int J Pept Res Ther 1–10.

152. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz Ş, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaíno JA. 2019. The PRIDE database and related tools and resources in 2019: Improving support for quantification data1. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss

J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeu. Nucleic Acids Res 47:D442–D450.

153. Chromy BA, Gonzales AD, Perkins J, Choi MW, Corzett MH, Chang BC, Corzett CH, McCutchen-Maloney SL. 2004. Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. J Proteome Res 3:1120–1127.

154. Kruh-Garcia NA, Murray M, Prucha JG, Dobos KM. 2014. Antigen 85 variation across lineages of *Mycobacterium tuberculosis*—Implications for vaccine and biomarker success. J Proteomics 97:141–150.

155. Backus KM, Dolan MA, Barry CS, Joe M, McPhie P, Boshoff HIM, Lowary TL, Davis BG, Barry CE, III. 2014. The three *Mycobacterium tuberculosis* antigen 85 isoforms have unique substrates and activities determined by non-active site regions. J Biol Chem 289:25041–53.

156. Hickey TBM, Thorson LM, Speert DP, Daffé M, Stokes RW. 2009. *Mycobacterium tuberculosis* Cpn60.2 and DnaK are located on the bacterial surface, where Cpn60.2 facilitates efficient bacterial association with macrophages. Infect Immun https://doi.org/10.1128/IAI.00143-09.

157. Lee BY, Hefta SA, Brennan PJ. 1992. Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis*. Infect Immun 60:2066–2074.

158. Sulzenbacher G, Canaan S, Bordat Y, Neyrolles O, Stadthagen G, Roig-Zamboni V, Rauzier J, Maurin D, Laval F, Daffé M, Cambillau C, Gicquel B, Bourne Y, Jackson M. 2006. LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of *Mycobacterium tuberculosis*. EMBO J 25:1436–1444.

159. Kuo C-J, Gao J, Huang J-W, Ko T-P, Zhai C, Ma L, Liu W, Dai L, Chang Y-F, Chen T-H, Hu Y, Yu X, Guo R-T, Chen C-C. 2019. Functional and structural investigations of fibronectin-binding protein Apa from *Mycobacterium tuberculosis*. Biochim Biophys Acta - Gen Subj 1863:1351–1359.

160. Webb JR, Vedvick TS, Alderson MR, Guderian JA, Jen SS, Ovendale PJ, Johnson SM, Reed SG, Skeiky YA. 1998. Molecular cloning, expression, and immunogenicity of MTB12, a novel low-molecular-weight antigen secreted by *Mycobacterium tuberculosis*. Infect Immun 66:4208–14.

161. Stamm CE, Pasko BL, Chaisavaneeyakorn S, Franco LH, Nair VR, Weigele BA, Alto NM, Shiloh MU. 2019. Screening *Mycobacterium tuberculosis* Secreted Proteins Identifies Mpt64 as a Eukaryotic Membrane-Binding Bacterial Effector . mSphere 4.

162. Gaudet P, Livstone MS, Lewis SE, Thomas PD. 2011. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief Bioinform 12:449–62.

163. Sharma P, Kumar B, Singhal N, Katoch VM, Venkatesan K, Chauhan DS, Bisht D. 2010. Streptomycin induced protein expression analysis in *Mycobacterium tuberculosis* by two-dimensional gel electrophoresis & mass spectrometry. Indian J Med Res 132:400–8.

164. D'Auria G, Esposito C, Falcigno L, Calvanese L, Iaccarino E, Ruggiero A, Pedone C, Pedone E, Berisio R. 2010. Dynamical properties of cold shock protein A from *Mycobacterium tuberculosis*. Biochem Biophys Res Commun 402:693–698.

165. Mon ML, Moyano RD, Viale MN, Colombatti Olivieri MA, Gamietea IJ, Montenegro VN, Alonso B, Santangelo M de la P, Singh M, Duran R, Romano MI. 2014. Evaluation of Cocktails with Recombinant Proteins of *Mycobacterium bovis* for a

Specific Diagnosis of Bovine Tuberculosis. Biomed Res Int 2014:1–12.

166. Assal N, Lin M. 2020. PCR procedures to amplify GC-rich DNA sequences of *Mycobacterium bovis*. J Microbiol Methods 2020.02.18.953695.

167. Kary B. Mullis. 1990. The Unusual Origin of the Polymerase Chain Reaction. Sci Am 262:56–61, 64–5.

168. Agne M, Valones A, Guimarães RL, André L, Brandão C, Eleutério De Souza PR, De Albuquerque A, Carvalho T, Crovela S. 2009. Principles and Applications of Polymerase Chain Reaction in Medical Diagnostic Fields: A Review. Brazilian J Microbiol 40:1–11.

169. Garibyan L, Avashia N. 2013. Research Techniques Made Simple: Polymerase Chain Reaction (PCR). J Invest Dermatol 133:e6. doi:10.1038/jid.2013.1.

170. Akan P, Deloukas P. 2008. DNA sequence and structural properties as predictors of human and mouse promoters. Gene 410:165–76.

171. Kumar A, Kaur J. 2014. Primer Based Approach for PCR Amplification of High GC Content Gene: Mycobacterium Gene as a Model. Mol Biol Int 2014:1–7.

172. Labaer J, Qiu Q, Anumanthan A, Mar W, Zuo D, Murthy TVS, Taycher H, Halleck A, Hainsworth E, Lory S, Brizuela L. 2004. The *Pseudomonas aeruginosa* PA01 gene collection. Genome Res 14:2190–2200.

173. Borah P. 2011. Primer designing for PCR. Sci Vis 11:134–136.

174. Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S. 2006. An efficient and economic enhancer mix for PCR. Biochem Biophys Res Commun 347:747–751.

175. Jensen MA, Fukushima M, Davis RW. 2010. DMSO and Betaine Greatly Improve Amplification of GC- Rich Constructs in De Novo Synthesis. PLoS One 5:e11024.

176. Geiduschek EPP, Herskovits TT. 1961. Nonaqueous solutions of DNA. Reversible and irreversible denaturation in methanol. Arch Biochem Biophys 95:114–129.

177. Rees WA, Yager TD, Korte J, von Hippel PH. 1993. Betaine can eliminate the base pair composition dependence of DNA melting. Biochemistry 32:137–44.

178. Sarkar G, Kapelner S, Sommer SS. 1990. Formamide can dramatically improve the specificity of PCR. Nucleic Acids Res 18:7465.

179. Lorenz TC. 2012. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. J Vis Exp 63:1–15.

180. Korbie DJ, Mattick JS. 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc 3:1452–1456.

181. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008.

182. Mamedov TG, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R, Subramanian A, Viljoen HJ, Mammedov T, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R, A.Subramanian, HJ. V. 2008. A Fundamental Study of the PCR Amplification of GC-Rich DNA Templates. Comput biol Chem 32:452–457.

183. Frey UH, Bachmann HS, Peters J, Siffert W. 2008. PCR-amplification of GC-rich regions: "slowdown PCR." Nat Protoc 3:1312–1317.

184. Schiller I, Oesch B, Vordermeier HM, Palmer M V., Harris BN, Orloski KA, Buddle BM, Thacker TC, Lyashchenko KP, Waters WR. 2010. Bovine Tuberculosis: A Review of Current and Emerging Diagnostic Techniques in View of their Relevance for Disease Control and Eradication. Transbound Emerg Dis 57:205–220.

185. de la Rua-Domenech R. 2006. Human *Mycobacterium bovis* infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of

bovine tuberculosis. Tuberculosis 86:77–109.

186. Lyashchenko KP, Pollock JM, Colangeli R, Gennaro ML. 1998. Diversity of Antigen Recognition by Serum Antibodies in Experimental Bovine Tuberculosis. Infect Immun 66:5344–5349.

187. Green ER, Mecsas J. 2016. Bacterial Secretion Systems: An Overview, p. 213–239. *In* Indira T. Kudva, Nancy A. Cornick, Paul J. Plummer, Qijing Zhang, Tracy L. Nicholson, John P. Bannantine, Bryan H. Bellaire (eds.), Virulence Mechanisms of Bacterial Pathogens, 5th ed. ASM Press, Washington, DC, USA.

188. 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.

189. Severini GM, Mestroni L, Falaschi A, Camerini F, Giaccal M. 1993. Nested Polymerase Chain Reaction for High-Sensitivity Detection of Enteroviral RNA in Biological Samples. J Clin Microbiol 31:1345–1349.

190. Oiiner JD, Kinzler KW, Vogelstein B. 1993. In vivo Cloning of PCR products in E.coli. Nucleic Acids Res 21:5192–5197.

191. Davies DH, Liang X, Hernandez JE, Randall A, Hirst S, Mu Y, Romero KM, Nguyen TT, Kalantari-Dehaghi M, Crotty S, Baldi P, Villarreal LP, Felgner PL. 2005. Profiling the humoral immune response to infection by using proteome microarrays: High-throughput vaccine and diagnostic antigen discovery. Proc Natl Acad Sci U S A 102:547–552.

192. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. PLoS Biol 4:0933–0942.

193. Lessa-Aquino C, Wunder EA, Lindow JC, Rodrigues CB, Pablo J, Nakajima R, Jasinskas A, Liang L, Reis MG, Ko AI, Medeiros MA, Felgner PL. 2015. Proteomic

features predict seroreactivity against leptospiral antigens in leptospirosis patients. J Proteome Res 14:549–556.

194. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM. 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7.

195. Boivin T, Elmgren C, Brooks BW, Huang H, Pagotto F, Lin M. 2016. Expression of Surface Protein LapB by a Wide Spectrum of *Listeria monocytogenes* Serotypes as Demonstrated with Anti-LapB Monoclonal Antibodies. Appl Environ Microbiol 82:6768–6778.

196. Lechtzier V, Hutoran M, Levy T, Kotler M, Brenner T, Steinitz M. 2002. Sodium dodecyl sulphate-treated proteins as ligands in ELISA. J Immunol Methods 270:19–26.

197. Waters WR, Buddle BM, Vordermeier HM, Gormley E, Palmer M V, Thacker TC, Bannantine JP, Stabel JR, Linscott R, Martel E, Milian F, Foshaug W, Lawrence JC. 2011. Development and Evaluation of an Enzyme-Linked Immunosorbent Assay for Use in the Detection of Bovine Tuberculosis in Cattle. Clin Vaccine Immunol 18:1882–1888.

198. Casal C, Díez-Guerrier A, Álvarez J, Rodriguez-Campos S, Mateos A, Linscott R, Martel E, Lawrence JC, Whelan C, Clarke J, O'Brien A, Domínguez L, Aranaz A. 2014. Strategic use of serology for the diagnosis of bovine tuberculosis after intradermal skin testing. Vet Microbiol 170:342–351.

199. Amadori M, Lyashchenko KP, Gennaro ML, Pollock JM, Zerbini I. 2002. Use of recombinant proteins in antibody tests for bovine tuberculosis. Vet Microbiol 85:379–389.

200. de Martino M, Lodi L, Galli L, Chiappini E. 2019. Immune Response to *Mycobacterium tuberculosis*: A Narrative Review. Front Pediatr 7:350.

201. Liang L, Leng D, Burk C, Nakajima-Sasaki R, Kayala MA, Atluri VL, Pablo J, Unal B, Ficht TA, Gotuzzo E, Saito M, Morrow WJW, Liang X, Baldi P, Gilman RH, Vinetz JM, Tsolis RM, Felgner PL. 2010. Large scale immune profiling of infected humans and goats reveals differential recognition of Brucella melitensis antigens. PLoS Negl Trop Dis 4.

202. Li L, Bannantine JP, Campo JJ, Randall A, Grohn YT, Katani R, Schilling M, Radzio-Basu J, Kapur V. 2017. Identification of sero-reactive antigens for the early diagnosis of Johne's disease in cattle. PLoS One 12:e0184373.

203. Khambhati K, Bhattacharjee G, Gohil N, Braddick D, Kulkarni V, Singh V. 2019. Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. Front Bioeng Biotechnol 7:248.

204. Zhou F, Xu X, Wu S, Cui X, Fan L, Pan W. 2015. Protein array identification of protein markers for serodiagnosis of *Mycobacterium tuberculosis* infection. Sci Rep 5.

205. Lyashchenko KP, Grandison A, Keskinen K, Sikar-Gang A, Lambotte P, Esfandiari J, Ireton GC, Vallur A, Reed SG, Jones G, Vordermeier HM, Stabel JR, Thacker TC, Palmer M V., Waters WR. 2017. Identification of novel antigens recognized by serum antibodies in bovine tuberculosis. Clin Vaccine Immunol 24:1–9.

206. 2015. RTS^{TM} pIVEX E. coli His-tag 2 nd Generation Vector Set Manual. Accessed 2020-01-31. https://www.biotechrabbit.com

Contribution of Collaborators

Dr. Olga Andrievskaia, CFIA, supplied me with the *M. bovis* genomic DNA as well as heat-killed *M. bovis* cells for extracting the genomic DNA.

Dr. Marc-Olivier Duceppe, CFIA, helped me with the Bioinformatic analysis; he supplied me with the genome sequence of *M. bovis* isolate 2011/069 (unpublished) and performed a high throughput Protein-BLAST analysis.

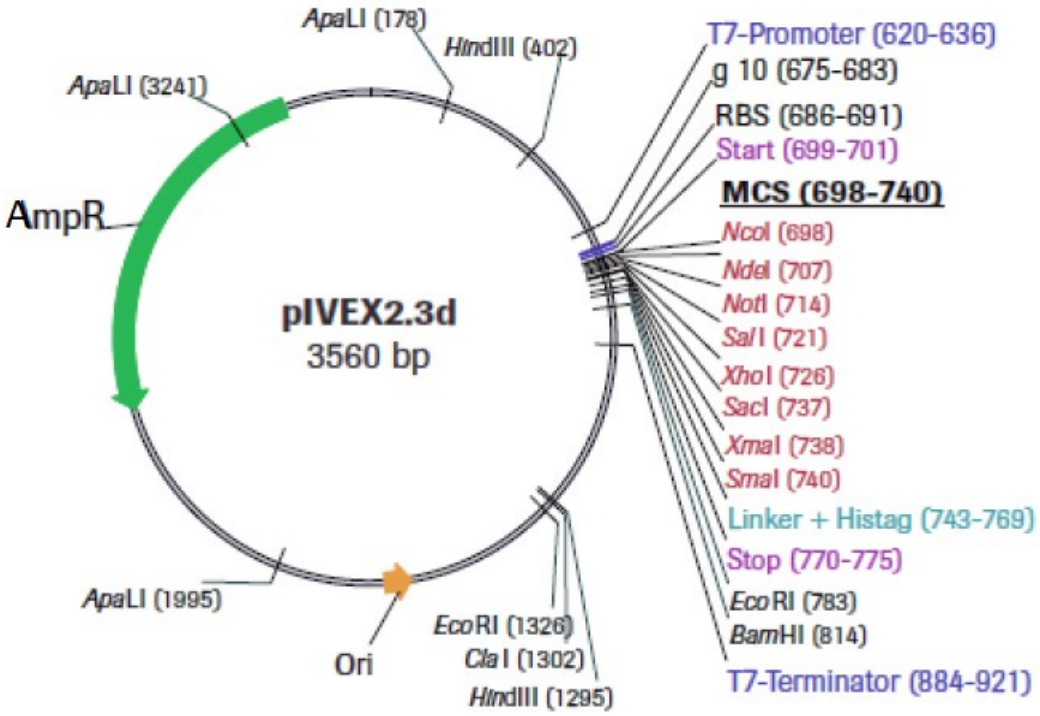
Dara Lloyd, CFIA, grew *M. bovis* cultures in Biological Safety Level 3 (BSL-3) laboratory for the proteomic study and use in DNA extraction

Bryan Rennie performed the sterile filtration of the culture supernatant, OD measurement, and the subsequent confirmation of sterility before the sample release for the proteomic study.

The proteomic study was carried in collaboration with Lisa Walrond and Dr. Terry Cyr, Health Canada. Lisa performed the mass spectrometry analysis of the purified peptides that I prepared. She also documented the mass spectrometry procedures.

Dr. Qing Yan Liu at the National Research Council, Canada allowed me to work on the scanner in her lab to take pictures of my microarray chips.

Appendix 1. Vector Map



S Figure 2. pIVEX 2.3d plasmid vector map.

The plasmid vector pIVEX 2.3d is 3560 bp long and contains a T7 promoter, prokaryotic ribosome-binding sequence, ATG start codon, multiple cloning site and T7 terminator. It enables the attachment of a 6xHis-tag at the C-terminus of the protein. The plasmid contains an ampicillin/carbenicillin resistance gene (206).

Appendix 2. List of Primers used in this Study

Table S1. Primers used for cloning some known *M. bovis* antigenic targets in pIVEX2.3d.

Gene name	locus tag	Primer#	D	Sequence (5' to 3')
MPB70	Mb2900	1440	F	TTAACTTTAAGAAGGAGATATACCATGAAGGTAAAGAACACAATTGCG
		1441	R	GATGATGATGATGATGAGAACCCCCCGCCGGAGGCATTAGCACGC
MPB83	Mb2898	1442	F	GTTTAACTTTAAGAAGGAGATATACCATGATCAACGTTCAGGCCAA
		1443	R	GATGATGATGATGATGAGAACCCCCCTGTGCCGGGGGCATCAGCA
Ag85A	Mb3834c	1444	F	GTTTAACTTTAAGAAGGAGATATACCATGCAGCTTGTTGACAGGG
		1445	R	GATGATGATGATGAGAACCCCCGTTGGGCGTGGCACCCAGT
Ag85B	Mb1918c	1446	F	GTTTAACTTTAAGAAGGAGATATACCATGACAGACGTGAGCCGA
		1447	R	GATGATGATGATGATGAGAACCCCCGCCGGCGCCTAACGAACTCT
Ag85C	Mb0134c	1448	F	GTTTAACTTTAAGAAGGAGATATACCATGACGTTCTTCGAACAGGT
		1449	R	GATGATGATGATGAGAACCCCCTGTCGCGCCGTTGAGCACA
hspX	Mb2057c	1450	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCACCACCCTTCCCGT
		1451	R	GATGATGATGATGATGAGAACCCCCGTTGGTGGACCGGATCTGAA
PE5	Mb0293	1452	F	GTTTAACTTTAAGAAGGAGATATACCATGACGTTGCGAGTGGTTCC
		1453	R	GATGATGATGATGATGAGAACCCCC GCCGCCCACGACCCCGTACG
Esxl	Mb1230	1454	F	GTTTAACTTTAAGAAGGAGATATACC ATGACCATCAACTATCAATT
		1455	R	GATGATGATGATGATGAGAACCCCCGGCCCAGCTGGAGCCGACGG
PE13	Mb1227	1456	F	GTTTAACTTTAAGAAGGAGATATACCATGTCTTTCGTGATGGCATA
		1457	R	GATGATGATGATGATGAGAACCCCCGCTGGCCGCCGCCGCATTGG
TB10.4	Mb0296	1458	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGCAAATCATGTACAA
		1459	R	GATGATGATGATGATGAGAACCCCCGCCGCCCCATTTGGCGGCTT
TB15.3	Mb1662	1460	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCGCCTATAAGACCGT
		1461	R	GATGATGATGATGATGAGAACCCCCGGTGGTGTGCACGATCAGCA
Espc	Mb3645c	1462	F	GTTTAACTTTAAGAAGGAGATATACCATGACGGAAAACTTGACCGT
		1463	R	GATGATGATGATGATGAGAACCCCCGGTAAACAACCCGTCGATAG
TB7.3	Mb3247c	1464	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCGAGGATGTTCGCGC
		1465	R	GATGATGATGATGATGAGAACCCCCGCTGATCACCGCGATAAGGT
MPB53	Mb2903c	1466	F	GTTTAACTTTAAGAAGGAGATATACCATGAGTCTTCGCCTGGTGTC
		1467	R	GATGATGATGATGATGAGAACCCCCGGACGTCAGCGCAGCCACCC
Mb0130	Mb0130	1468	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCAATTCGCGCCGCCG
		1469	R	GATGATGATGATGATGAGAACCCCCGGCCGGGGGTCCCTCGGCCA
PPE14	Mb0939c	1470	F	GTTTAACTTTAAGAAGGAGATATACCATGGATTTCGGGCTTTTACC
		1471	R	GATGATGATGATGATGAGAACCCCCCCGCGGGGGGTTTCCGGGCG
Mb0295	Mb0295	1472	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCCTTTTGGATGCTCA
		1473	R	GATGATGATGATGATGAGAACCCCC GAACCCGGTATAGGTCGACG
Rv3740c	Mb3766c	1474	F	GTTTAACTTTAAGAAGGAGATATACCATGTCACCGATCGATGCGCT
		1475	R	GATGATGATGATGATGAGAACCCCCTAGCCCGACCGCGCGCTCCA

Mb1207c	Mb1207c	1476	F	GTTTAACTTTAAGAAGGAGATATACCATGAGGCTGTCGTTGACCGC
		1477	R	GATGATGATGATGATGAGAACCCCCATAGTTGTTGCAGGAGCCGG
TPX	Mb1967	1478	F	GTTTAACTTTAAGAAGGAGATATACCATGGCACAGATAACCCTGCG
		1479	R	GATGATGATGATGATGAGAACCCCCGGCGCCCAGCGCGGCCAGCG
CFP10	Mb3904	1532	F	GTTTAACTTTAAGAAGGAGATATACCATGGCAGAGATGAAGACCGA
		1533	R	GATGATGATGATGATGAGAACCCCCGAAGCCCATTTGCGAGGACA
ESAT6	Mb3905	1534	F	GTTTAACTTTAAGAAGGAGATATACCATGACAGAGCAGCAGTGGAA
		1535	R	GATGATGATGATGATGAGAACCCCCTGCGAACATCCCAGTGACGT

Table S2. Primers used for cloning Bioinformatic targets in pIVEX2.3d

locus tag	Primer#	D	Sequence (5' to 3')
Mb0041c	1536	F	GTTTAACTTTAAGAAGGAGATATACCATGATCCAGATCGCGCGCAC
	1537	R	GATGATGATGATGATGAGAACCCCCGCGCGGCGGGACTGGTGTCA
Mb0064	1538	F	GTTTAACTTTAAGAAGGAGATATACCATGGCGCGTGAGATCTCACG
	1539	R	GATGATGATGATGATGAGAACCCCCGAAGTTCAGCCCGGAGAACA
Mb0129	1540	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTCAGCGTAGC
	1541	R	GATGATGATGATGATGAGAACCCCCGCCGGGTTGGCCGGTGTCGC
Mb0184	1542	F	GTTTAACTTTAAGAAGGAGATATACCATGGAAGATCAGCAGTCTGC
	1543	R	GATGATGATGATGATGAGAACCCCCCACCTGCGGCGTCAAATCCA
Mb0285c*	1544	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTGATTGCGGC
	1544.1	R	GATGATGATGATGATGAGAACCCCCTGAGGTGCTGACTGGAGCATT
Mb0285c*	1545	F	GATGATGATGATGATGAGAACCCCCTGGCAAGCCGTTGAGTCCGT
	1545.1	R	GTTTAACTTTAAGAAGGAGATATACCATGGGTGGCATACCCGGTTTGAT
Mb0286c*	1546	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTGATTGCGGC
	46.1	R	GATGATGATGATGATGAGAACCCCCGGGAGCATTGAATCCGTCTA
Mb0286c*	1547	F	GATGATGATGATGATGAGAACCCCCTGGCAGGCCGTTGAGCCCGT
	47.1	R	GTTTAACTTTAAGAAGGAGATATACCATGGGCGTGTTGATCGGCAATGG
Mb0287c*	1548	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTGATTGCGGC
	48.1	R	GATGATGATGATGATGAGAACCCCCACTACCACCGTTACCGCCATT
Mb0287c*	1549	F	GATGATGATGATGATGAGAACCCCCCAGGCCGTTGAGCCCGTTCT
	49.1	R	GTTTAACTTTAAGAAGGAGATATACCATGGGAAAGGCGGGTCTGCTC
Mb0305	1550	F	GTTTAACTTTAAGAAGGAGATATACCATGTCCTTTGTGATCGCACA
	1551	R	GATGATGATGATGATGAGAACCCCCGCCGGGGTTGCCGTTGTTGC
Mb0368	1552	F	GTTTAACTTTAAGAAGGAGATATACCATGTCCAACGCACCCGAGCC
	1553	R	GATGATGATGATGATGAGAACCCCCGTTGCTGGAGGACTGGCAGA
Mb0426	1554	F	GTTTAACTTTAAGAAGGAGATATACCATGGTGAACAAATCCAGGAT
	1555	R	GATGATGATGATGATGAGAACCCCCCGGTTTGGCAATCAGGTGGC
Mb0545	1556	F	GTTTAACTTTAAGAAGGAGATATACCATGTCTAATTTGCTGGTAAC
	1557	R	GATGATGATGATGATGAGAACCCCCGCCGTGCGGGCCGGTCGTGC
Mb0593c*	1558	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTCGTAATCGCGAC
	1690	R	GATGATGATGATGATGAGAACCCCCGGACGCATTGAAGCCGATGCT

Mb0593c*	1688	F	GTTTAACTTTAAGAAGGAGATATACCATGGCTATCAGCAGCGGACAGAC
	1559	R	GATGATGATGATGATGAGAACCCCCGAAGCTGCCGCCGGCGCCGC
Mb0690	1560	F	GTTTAACTTTAAGAAGGAGATATACCATGCTCCGCAGGGTCGCCAT
	1561	R	GATGATGATGATGATGAGAACCCCCTCTGAAGTGCGCTGCAAAGA
Mb0768	1562	F	GTTTAACTTTAAGAAGGAGATATACCATGTCATGGGTGATGGTTTC
	62.1	R	GATGATGATGATGATGAGAACCCCCGTCCAGGCCCAACAGCAC
Mb0777c	1564	F	GTTTAACTTTAAGAAGGAGATATACCATGGTCGGTTTCGCGTGGTT
	64.1	R	GATGATGATGATGATGAGAACCCCCGAAGTTGTAGCTGCCGCTGTT
Mb0845c	1566	F	GTTTAACTTTAAGAAGGAGATATACCATGAGTGACGGCGAGAGCGC
	66.1	R	GATGATGATGATGATGAGAACCCCCGAAACCGACGAAGTCAATCG
Mb0856*	1568	F	GTTTAACTTTAAGAAGGAGATATACCATGATCGGCAACGGCGGGGCC
	68.1	R	GATGATGATGATGATGAGAACCCCCGTCCAAACCCAACAACAGC
Mb0856*	69.1	F	GTTTAACTTTAAGAAGGAGATATACCATGTTCGCCAACGGCTCTACC
	1691	R	GATGATGATGATGATGAGAACCCCCCTCCGCGCCGATCAGCGAA
Mb0858	1570	F	GTTTAACTTTAAGAAGGAGATATACCATGTGTTGCTCCACTGCTGC
	1571	R	GATGATGATGATGATGAGAACCCCCCTGGATGCGATCGCTGATCG
Mb0896c	1572	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTATGTATTGGCGAC
	1573	R	GATGATGATGATGATGAGAACCCCCGCCGGGTTGGCCGGGTTGGC
Mb1002*	1638	F	GTTTAACTTTAAGAAGGAGATATACCATGTCCTTCGTGGTCACAGC
	1639	R	GATGATGATGATGATGAGAACCCCCGTTGCCGCCATTGCCGCCAT
Mb1002*	1640	F	GTTTAACTTTAAGAAGGAGATATACCATGCACGGCGGCGCCGCCCAAT
	1641	R	GATGATGATGATGATGAGAACCCCCGTTAAAGGTGGTCGTCCCGA
Mb1006c	1642	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTCAACGTGGC
	1643	R	GATGATGATGATGATGAGAACCCCCTATGGCCGCCGAACACACCG
Mb1018c	1644	F	GTTTAACTTTAAGAAGGAGATATACCATGCCAACCTACAGCTACGA
	1645	R	GATGATGATGATGATGAGAACCCCCTGTGGAGCTGGTTGACTT
Mb1044c	2526	F	GTTTAACTTTAAGAAGGAGATATACC ATGAGACCCTTGGCGGTCGC
	1647	R	GATGATGATGATGATGAGAACCCCCGCAGTAGCCATGGGTTAGCC
Mb1050	1648	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGCCGAGACGCTGGTT
	1649	R	GATGATGATGATGATGAGAACCCCCCAGCGGATGACCCGCCGCAT
Mb1096c	1650	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTGTTGGTGTC
	1651	R	GATGATGATGATGATGAGAACCCCCCTGCCCCGGCGTGCCGGCGT
Mb1097c	1652	F	GTTTAACTTTAAGAAGGAGATATACCATGTCCTACATGATTGCGGT
	1653	R	GATGATGATGATGATGAGAACCCCCTTGCCCGGGCGTGCCGGGGT
Mb1116	1654	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTAGTCGTGGC
	1655	R	GATGATGATGATGATGAGAACCCCCGTTCGCCCCCGGCTCGCCGA
Mb1118	1656	F	GTTTAACTTTAAGAAGGAGATATACCATGCGGGGCGCCGTCAACGC
	1657	R	GATGATGATGATGATGAGAACCCCCGGCAAAGTGAATTGCAACCG
Mb1121	1658	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTGATTGCTGC
	1659	R	GATGATGATGATGATGAGAACCCCCGTTGGCGTTACCGCTGAGGC
	1660	F	GTTTAACTTTAAGAAGGAGATATACCATGGGTAATGGCGTCAGCAGCAG
	1661	R	GATGATGATGATGATGAGAACCCCCCGGCGACACGCCAGCCGTGC

Mb1262c	1662	F	GTTTAACTTTAAGAAGGAGATATACCATGCACATTGGGGGACGCTG
	1663	R	GATGATGATGATGATGAGAACCCCCGCCGGCAGGAGCCGGCGCCG
Mb1275c	1664	F	GTTTAACTTTAAGAAGGAGATATACCATGGAGTATCTGATTGCAGC
	1665	R	GATGATGATGATGATGAGAACCCCCGCCGTTCAGGCCGTGCTGAC
Mb1360c	1666	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTAATCGCCGC
	1667	R	GATGATGATGATGATGAGAACCCCCCGACAACCCGGGTGATCCGT
Mb1476c	1668	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGAACGTGATGGTAGT
	1669	R	GATGATGATGATGATGAGAACCCCCCCCGTGCTTTCCTTGCGCGC
Mb1485c*	1670	F	GTTTAACTTTAAGAAGGAGATATACCATGAACGGAATCCAGGGTGTCG
	1671	R	GATGATGATGATGATGAGAACCCCCCACGAATTCGCCGCCGCCGA
Mb1485c*	1672	F	GTTTAACTTTAAGAAGGAGATATACCATGGGCAACCAGGCCGTCAAC
	1673	R	GATGATGATGATGATGAGAACCCCCCTTGGCGCCTGGGGTAGAG
Mb1485c*	1674	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTGGTGATCGTGGC
	1675	R	GATGATGATGATGATGAGAACCCCCGTGGGCACCGATCGTGGAG
Mb1487c	1676	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTGGTGATCGTGGC
	1677	R	GATGATGATGATGATGAGAACCCCCGATGGTTGGATCGCCACCGG
Mb1575c	1678	F	GTTTAACTTTAAGAAGGAGATATACCATGAATTTTTCGGTGTTGCC
	1679	R	GATGATGATGATGATGAGAACCCCCATGGATCCAGCCCGACGCCC
Mb1679c*	1680	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTCTTACTCGTGGA
	1681	R	GATGATGATGATGATGAGAACCCCCGCCATTGCCGATGAAGCTACT
Mb1679c*	1682	F	GTTTAACTTTAAGAAGGAGATATACCATGGGGACGAGCAGCGATCTG
	1683	R	GATGATGATGATGATGAGAACCCCCAGGGGCAATTGCCTGCGCTA
Mb1783c	1684	F	GTTTAACTTTAAGAAGGAGATATACCATGTACCGGTACCAAGTCCG
	1685	R	GATGATGATGATGATGAGAACCCCCTGGCAGCACGGTGTGCATCA
Mb1784c	1686	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCCAAAGCCACATCGG
	1687	R	GATGATGATGATGATGAGAACCCCCGCACGGACCGCTCGGAATAC
Mb1788	1786	F	GTTTAACTTTAAGAAGGAGATATACCATGGTCACGACGTGGGCACT
	1787	R	GATGATGATGATGATGAGAACCCCCACCGAGCCGACCGGCGGCAA
Mb1831c	1788	F	GTTTAACTTTAAGAAGGAGATATACCATGCAAATCGTGGGCCAAAC
	1789	R	GATGATGATGATGATGAGAACCCCCCGGCTGCCCGTTCTGCCCAG
Mb2006c	1790	F	GTTTAACTTTAAGAAGGAGATATACCATGACTCCACGCAGCCTTGT
	1791	R	GATGATGATGATGATGAGAACCCCCTCCGGCGTGATCGAGCCTGT
Mb2108	1792	F	GTTTAACTTTAAGAAGGAGATATACCATGGCGGGTGATCTGCCAC
	2527	R	GATGATGATGATGATGAGAACCCCCTCTGGATCTCCTCGTTGATA
Mb2125c	1794	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTGATTGCGAG
	1795	R	GATGATGATGATGATGAGAACCCCCCCCGTCCGGTCCGGATGG
Mb2323	1796	F	GTTTAACTTTAAGAAGGAGATATACCATGAATGATTTACTGACCCG
	1797	R	GATGATGATGATGATGAGAACCCCCTTGCAGCTTTCCGGCAACGA
Mb2410c	1798	F	GTTTAACTTTAAGAAGGAGATATACCATGACACCGGGTTTGCTTAC
	1799	R	GATGATGATGATGATGAGAACCCCCATCGTCCCTGCTCCCCGAAC
Mb2506c	1800	F	GTTTAACTTTAAGAAGGAGATATACCATGGCGTTGCGTAGACGCCA
	1801	R	GATGATGATGATGATGAGAACCCCCGTGAGCTATAGACGACTTCG

Mb2517c*	1802	F	GTTTAACTTTAAGAAGGAGATATACCATGACGGCGATCTTCCTGGG
	1803	R	GATGATGATGATGATGAGAACCCCCGGTCAGTGGCCCGTGTTTGCT
Mb2517c	1804	F	GTTTAACTTTAAGAAGGAGATATACCATGAACGGCGGCATTGGTATCAC
	1805	R	GATGATGATGATGATGAGAACCCCCCGTGACGTTAACGGAGATGG
Mb2606c	1806	F	GTTTAACTTTAAGAAGGAGATATACCATGCCGGCCGGCGTCGGTAA
	1807	R	GATGATGATGATGATGAGAACCCCCCCCAGGGACGAGTCGAGCACT
Mb2667c	1808	F	GTTTAACTTTAAGAAGGAGATATACCATGTCATTTGTGATTGCGGT
	1809	R	GATGATGATGATGATGAGAACCCCCTCCATCAGCGCCCTGGGTGC
Mb2734	1810	F	GTTTAACTTTAAGAAGGAGATATACCATGACCGAGCGGAAGCGAAA
	1811	R	GATGATGATGATGATGAGAACCCCCGGTAGCGCTGCGTTCGTTGG
Mb2740c	1812	F	GTTTAACTTTAAGAAGGAGATATACCATGAACGGGCAGAGAGGTCA
	1813	R	GATGATGATGATGATGAGAACCCCCATCCGCCCTAATGAAGCCGT
Mb2878	1814	F	GTTTAACTTTAAGAAGGAGATATACCATGTTGTATGTAGTTGCGTC
	1815	R	GATGATGATGATGATGAGAACCCCCTCCTGGCTGTCCAGGGTTGC
Mb3183c	1816	F	GTTTAACTTTAAGAAGGAGATATACCATGAATTATTCGGTGTTGCC
	1817	R	GATGATGATGATGATGAGAACCCCCGGGCTGACCGAAGAAGCCCG
Mb3402	1818	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTCGTAGCAGT
	1819	R	GATGATGATGATGATGAGAACCCCCTCCTGGGTTGCCGTCGGTGC
Mb3453c	1820	F	GTTTAACTTTAAGAAGGAGATATACCATGACGACAGTCTTGGGCAT
	1821	R	GATGATGATGATGATGAGAACCCCCCCCGCACCTGCCCCTGCATCA
Mb3469c	1822	F	GTTTAACTTTAAGAAGGAGATATACCATGGCTGACCGGTTGAACGT
	1823	R	GATGATGATGATGATGAGAACCCCCCGGACCCGCTTGCGGAAGCT
Mb3481	2528	F	GTTTAACTTTAAGAAGGAGATATACCATGAATTACGCCGCCAGCCG
	2529	R	GATGATGATGATGATGAGAACCCCCGCGACCGTGCAGCGATTCGG
Mb3537*	2530	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTTGTGTTGGTTTC
	2531	R	GATGATGATGATGATGAGAACCCCCGCCGGGAATCACGGCGTTG
Mb3537*	2532	F	GTTTAACTTTAAGAAGGAGATATACCATGATACCGGGCACTGGCACAGA
	2529	R	GATGATGATGATGATGAGAACCCCCGCCAGCGTTGCCGTTGTC
Mb3538*	1832	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTCGTGTTGATCGC
	1833	R	GATGATGATGATGATGAGAACCCCCAGCACCGGTGGGGTTGTC
Mb3538*	1834	F	GTTTAACTTTAAGAAGGAGATATACCATGGGTACGGGTGGTGTCGGTGG
	2530	R	GATGATGATGATGATGAGAACCCCCGCCATCACCGCCGTTACC
Mb3541*	1836	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTCGTGTTGATCTC
	1837	R	GATGATGATGATGATGAGAACCCCCGGTGGTGGTGTTGTCTGCGC
Mb3541*	1838	F	GTTTAACTTTAAGAAGGAGATATACCATGCAACAAGGCAACGGCGGCAA
	1839	R	GATGATGATGATGATGAGAACCCCCGAGGAAGGAGCCGGAACTGC
Mb3541*	1840	F	GTTTAACTTTAAGAAGGAGATATACCATGGCAGCCTCCTCAGCTACCA
	1841	R	GATGATGATGATGATGAGAACCCCCGACACCACCGCTACCGCC
Mb3543*	1842	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTCGTGTTGATCGC
	1843	R	GATGATGATGATGATGAGAACCCCCGTTGCCTGTGGTGCCGATCA
Mb3543*	1844	F	GTTTAACTTTAAGAAGGAGATATACCATGGGCGGACAAGGCGGGATGG
	1845	R	GATGATGATGATGATGAGAACCCCCGCTGCCGTTGTTGCCGGGTT

Mb3562	1846	F	GTTTAACTTTAAGAAGGAGATATACCATGTTCATGGATTTCGCGAT
	1847	R	GATGATGATGATGATGAGAACCCCCCCCGCCCGAGAGTGGACGT
Mb3751	1848	F	GTTTAACTTTAAGAAGGAGATATACCATGGATGTCATCAGATGGGC
	1849	R	GATGATGATGATGATGAGAACCCCCCAGGCGGCTGGCGGCGAAAT
Mb3761c	1850	F	GTTTAACTTTAAGAAGGAGATATACCATGGATCTGATGATGCCCAA
	1851	R	GATGATGATGATGATGAGAACCCCCGATCCCGACGGCCTGTTCCA
Mb3792	1852	F	GTTTAACTTTAAGAAGGAGATATACCATGCGCTCGGCATTCGACAG
	1853	R	GATGATGATGATGATGAGAACCCCCTTCCGCGAGCATGGCCGCCA
Mb3833c	1854	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGGGTCGGTCGGCGCT
	1855	R	GATGATGATGATGATGAGAACCCCCGCGGATCGCACCGACGATAT
Mb3841	1856	F	GTTTAACTTTAAGAAGGAGATATACCATGGCAGCGACCGTCGTCAT
	1857	R	GATGATGATGATGATGAGAACCCCCGGTGATCGGATGCGTTG
Mb3909c	1858	F	GTTTAACTTTAAGAAGGAGATATACCATGAGTATTACCAGGCCGAC
	1859	R	GATGATGATGATGATGAGAACCCCCGGATGCGGCGGCCAGGGC
Mb3921c	1860	F	GTTTAACTTTAAGAAGGAGATATACCATGCCCGATCCAGGATGGGC
	1861	R	GATGATGATGATGATGAGAACCCCCCCGCCGGTCATCGACAACAC
Mb0419c	1862	F	GTTTAACTTTAAGAAGGAGATATACCATGACGCGCCGGGCCCTCCT
	1863	R	GATGATGATGATGATGAGAACCCCCGTCCACATACCTCGGCGTGG
Mb0975c	1864	F	GTTTAACTTTAAGAAGGAGATATACCATGGCAGCGATTCGCACACC
	1865	R	GATGATGATGATGATGAGAACCCCCACCGGTGTAATTGCCGACGC
Mb1397c	1866	F	GTTTAACTTTAAGAAGGAGATATACCATGACCGATGACGTACGCGA
	1867	R	GATGATGATGATGATGAGAACCCCCAACCGGGGTGAACTTGGTGA
Mb1530	1868	F	GTTTAACTTTAAGAAGGAGATATACCATGACAACCAAGACACCCGT
	1869	R	GATGATGATGATGATGAGAACCCCCATCCAGCGTGTACCCCAGCC
Mb1724	1870	F	GTTTAACTTTAAGAAGGAGATATACCATGATCTCGTTGCGTCAACA
	1871	R	GATGATGATGATGATGAGAACCCCCCTGGGAAACCGTGACTGACA
Mb2377c	1872	F	GTTTAACTTTAAGAAGGAGATATACCATGGTGAATTTTTCGGTGTT
	1873	R	GATGATGATGATGATGAGAACCCCCCCTGAAGAATCCCGAAAGTC
Mb2554c	1874	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGGTCTCTCGGCGTGA
	1875	R	GATGATGATGATGATGAGAACCCCCAGCCCACTGCCCGAATTGGG
Mb3626C	2569	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTTGTGATTGCCGT
	2570	R	GATGATGATGATGATGAGAACCCCC GCCGAGCTCGGCAATAAGGT
Mb3621C	2571	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTTGTGATCGTGGC
	2572	R	GATGATGATGATGATGAGAACCCCC GCCGGGTAGGCCGTCCGCAC
Mb0113	2573	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGCTTTTGATCAC
	2574	R	GATGATGATGATGATGAGAACCCCC GGTTCCGGCGAAGCCGAATA
Mb0767	2575	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTCGTGCTGGCGAT
	2576	R	GATGATGATGATGATGAGAACCCCC CGGCAACCCGTTGAGCCCGT
Mb2369C	2577	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGCACGTTACCGCGGC
	2578	R	GATGATGATGATGATGAGAACCCCC TTCGTGCCCGGGCGGTCGAA
Mb2418	2579	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTCCTGATTGCTTC
	2580	R	GATGATGATGATGATGAGAACCCCC CGGCAGCCCGTTCGTGCCGT
Mb3903	2581	F	GTTTAACTTTAAGAAGGAGATATACC ATGCTGTGGCACGCAATGCC
	2582	R	GATGATGATGATGATGAGAACCCCC CCAGTCGTCCTCTTCGTCCC
Mb0450C	2583	F	GTTTAACTTTAAGAAGGAGATATACC ATGTTTGCCGGTCCCGGGTC

	2584	R	GATGATGATGATGATGAGAACCCCC TTCGTGCCCGGGCGGTCGAA
Mb2186C	2585	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTTGTGATTGCGGC
	2586	R	GATGATGATGATGATGAGAACCCCC GCCGTCGGCGCCGTTTGCT
Mb1828	2587	F	GTTTAACTTTAAGAAGGAGATATACC ATGCTGCCGAATTTCGCGGT
	2588	R	GATGATGATGATGATGAGAACCCCC GGGGACAATGGCCGAGATGT
Mb2761	2589	F	GTTTAACTTTAAGAAGGAGATATACC ATGGACTTGGCGAGCATCGG
	2590	R	GATGATGATGATGATGAGAACCCCC GCTAGGCAGCAATCCGTTCT
Mb2622	2591	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTTGTCACGGCAGC
	2592	R	GATGATGATGATGATGAGAACCCCC CGACGGTCCGTTCGCGCCGT
Mb1431C	2593	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTTTTGTTTGCACA
	2594	R	GATGATGATGATGATGAGAACCCCC CTGGCCGAACAGGGTTCCCC
Mb3420	2595	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCGTTTGTGATCGCGAA
	2596	R	GATGATGATGATGATGAGAACCCCC GAACAGCGTGCCGCCGGTGC
Mb2005	2597	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCATTTCTGGTCGTGGT
	2598	R	GATGATGATGATGATGAGAACCCCC CGCCGGATGATCAAAGACTG
Mb1797	2599	F	GTTTAACTTTAAGAAGGAGATATACC ATGTCCTATCTCGTCGTGGT
	2600	R	GATGATGATGATGATGAGAACCCCC CGGCCCGGGCATCCCATTTTG
Mb1228	2601	F	GTTTAACTTTAAGAAGGAGATATACC ATGGTGGATTTCGGGGCGTT
	2602	R	GATGATGATGATGATGAGAACCCCC CGGTGTGCGGGGTATCGCG
Mb1789C	2603	F	GTTTAACTTTAAGAAGGAGATATACC ATGACGCCGCTGCTCAACTC
	2604	R	GATGATGATGATGATGAGAACCCCC CGTCAAGCCGTGCTGCCCATC
Mb3184C	2605	F	GTTTAACTTTAAGAAGGAGATATACC ATGCTTCTGACGGGAACCAG
	2606	R	GATGATGATGATGATGAGAACCCCC CGGCTGGCCGAAGAACCCGG

***Targets are represented by 2 or three fragments**

Table S3. Primers used for cloning Proteomic targets in pIVEX2.3d

locus tag	Primer#	D	Sequence (5' to 3')
Mb0248c	1956	F	GTTTAACTTTAAGAAGGAGATATACCATGGCTCCCAAGCGTTCGTC
	1957	R	GATGATGATGATGATGAGAACCCCCCGCGCCGATCATGGCCTGGC
Mb0249	1958	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCCCTGCTGCTAAGAA
	1959	R	GATGATGATGATGATGAGAACCCCCGGCCTCCAAAATCGCGGCCA
Mb0254c	1960	F	GTTTAACTTTAAGAAGGAGATATACCATGGTTGAGGTCGAGCGGCA
	1961	R	GATGATGATGATGATGAGAACCCCCGCCTCTCCGTCCTGGATGCT
Mb0276	1962	F	GTTTAACTTTAAGAAGGAGATATACCATGCCTAACCTCACTGATCT
	1963	R	GATGATGATGATGATGAGAACCCCCCAGCTTGCGTAGCTCACGTT
Mb0340	1964	F	GTTTAACTTTAAGAAGGAGATATACCATGACCACATCGGAGATCGC
	1965	R	GATGATGATGATGATGAGAACCCCCATCGACCAAATCGTCCTCGG
Mb0439	1966	F	GTTTAACTTTAAGAAGGAGATATACCATGCTGGTTACAGTGGGCTC
	1967	R	GATGATGATGATGATGAGAACCCCCGCCGGTCACCACGACGATGA
Mb0471	1968	F	GTTTAACTTTAAGAAGGAGATATACCATGACCCACTATGACGTCGT
	1969	R	GATGATGATGATGATGAGAACCCCCGAAATTGATCATGTGGCCAA
Mb0476	1970	F	GTTTAACTTTAAGAAGGAGATATACCATGTCTGTCGTCGGCACCCC
	1971	R	GATGATGATGATGATGAGAACCCCCGTGGAACTGGCCCTCTTCGG
Mb0478	1972	F	GTTTAACTTTAAGAAGGAGATATACCATGACTGAGTTAAGGCCCTT

	1973	R	GATGATGATGATGATGAGAACCCCCCTTGGTCAGTGTGAACTGTC
Mb0485	1974	F	GTTTAACTTTAAGAAGGAGATATACCATGGCTGAAAACTCGAACAT
	1975	R	GATGATGATGATGATGAGAACCCCCCTTCTGGGTGACCTTCTTGG
Mb0515c	1976	F	GTTTAACTTTAAGAAGGAGATATACCATGACGTCACAGGGCGACAC
	1977	R	GATGATGATGATGATGAGAACCCCCTTTGACCAGAGTGAACTGGC
Mb0584	1978	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGGCAAAGGTCGGGGA
	1979	R	GATGATGATGATGATGAGAACCCCCCGTTCCCCTGGCATGGAGGA
Mb0592	1980	F	GTTTAACTTTAAGAAGGAGATATACCATGCCCAAGAGAAGCGAATA
	1981	R	GATGATGATGATGATGAGAACCCCCTTGCTGCGGTGCGGGCTTCA
Mb0598c	1982	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGCACTTCACGGCGGC
	1983	R	GATGATGATGATGATGAGAACCCCCGGGCGTGATGGTCGTCTGCT
Mb0649c	1984	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCGACCCGGTCAGCTA
	1985	R	GATGATGATGATGATGAGAACCCCCCAGCCCGAACTCGGCTGCTA
Mb0658	1986	F	GTTTAACTTTAAGAAGGAGATATACCATGACTACCTTCGACGGTGA
	1987	R	GATGATGATGATGATGAGAACCCCCGATCTTGGAGACTTGGCCAA
Mb0723	1988	F	GTTTAACTTTAAGAAGGAGATATACCATGGCGACGCTCGCTGACCCC
	1989	R	GATGATGATGATGATGAGAACCCCCGGCCGGTGCCCCGAACAGGT
Mb0727	1990	F	GTTTAACTTTAAGAAGGAGATATACCATGGGCCAAAAGATCAATCC
	1991	R	GATGATGATGATGATGAGAACCCCCGCTCTCCGTGCTCTGCGCTT
Mb0740	1992	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGCGTATTGGTAAGCA
	1993	R	GATGATGATGATGATGAGAACCCCCCTTACCTGTCTTTCCGACCT
Mb0754	1994	F	GTTTAACTTTAAGAAGGAGATATACCATGAGAGTTTTGTTGCTGGG
	1995	R	GATGATGATGATGATGAGAACCCCCCTTTCCCAGAGCCCGCAACG
Mb0784c	1996	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGACCAAGGGCGCACT
	1997	R	GATGATGATGATGATGAGAACCCCCGCGGTCGTCGTCCGTGTACC
Mb0824	1998	F	GTTTAACTTTAAGAAGGAGATATACCATGGCACTCAAGGTAGAGAT
	1999	R	GATGATGATGATGATGAGAACCCCCTTGACCCGCCACGCAAAAAG
Mb0854c	2000	F	GTTTAACTTTAAGAAGGAGATATACCATGCTCCCCGAGACAAATCA
	2001	R	GATGATGATGATGATGAGAACCCCCCTGGCGAAGCAGCTCATCTT
Mb0908c	2002	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCGACCAGCTCACCCC
	2003	R	GATGATGATGATGATGAGAACCCCCAAGCCGCTCGACCACCCAGT
Mb0920	2004	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCGACACCGACGACAC
	2005	R	GATGATGATGATGATGAGAACCCCCCCGCGCGTCTATGGTGACGT
Mb0929	2006	F	GTTTAACTTTAAGAAGGAGATATACCATGATCGGTATCACCCAGGC
	2007	R	GATGATGATGATGATGAGAACCCCCAGCCCCTTGGAACTTCGGCG
Mb0956c	2008	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGTTCGCCCGATCCGG
	2009	R	GATGATGATGATGATGAGAACCCCCAGAAATAGCATTCACCGCGG
Mb1010	2010	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGGTCGCCGCGCAGTG
	2011	R	GATGATGATGATGATGAGAACCCCCGATCTCCAAGCTCGACAACT
Mb1099c	2012	F	GTTTAACTTTAAGAAGGAGATATACCATGACGTACGAAACCATCCT
	2013	R	GATGATGATGATGATGAGAACCCCCTCGGTGGGTGAACTGGGGAG
Mb1109c	2322	F	GTTTAACTTTAAGAAGGAGATATACCATGACGGATACTCAAGTCAC

	2323	R	GATGATGATGATGATGAGAACCCCCGGAGTGGTACGGCTCGGCGC
Mb1123	2324	F	GTTTAACTTTAAGAAGGAGATATACCATGTCTGCCCCGCTCGCTGA
	2325	R	GATGATGATGATGATGAGAACCCCCGCGGCCGACCAGACTCCACT
Mb1164c	2326	F	GTTTAACTTTAAGAAGGAGATATACCATGACCCAGCCTGTACGTCG
	2327	R	GATGATGATGATGATGAGAACCCCCGACCATGTTGTGCAGCGACG
Mb1213	2328	F	GTTTAACTTTAAGAAGGAGATATACCATGCGCACCGCAACCGCCAC
	2329	R	GATGATGATGATGATGAGAACCCCCTCGGTGTGCACGGCGGCAGG
Mb1301c	2330	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGCATCCACCTTGTTC
	2331	R	GATGATGATGATGATGAGAACCCCCGACCGGTTTGGTGGCGGTGA
Mb1427	2336	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCGAAAAGGGTCGGCT
	2337	R	GATGATGATGATGATGAGAACCCCCGATGGCGCGCTTGAGGTCGT
Mb1656	2338	F	GTTTAACTTTAAGAAGGAGATATACCATGCCGAGTCCCACCGTCAC
	2339	R	GATGATGATGATGATGAGAACCCCCAGCGCTGCCGGCGAGTTTTT
Mb1951c	2340	F	GTTTAACTTTAAGAAGGAGATATACCATGCCGGCTATTCCGTTGGGT
	2341	R	GATGATGATGATGATGAGAACCCCCACCACCCAGCGCGTTCAGGC
Mb1950	2342	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCATCGCCGAAACGGA
	2343	R	GATGATGATGATGATGAGAACCCCCCGCCTCCTTCGTGATCAACT
Mb2164c	2344	F	GTTTAACTTTAAGAAGGAGATATACCATGACAACTTCACCCGACCC
	2345	R	GATGATGATGATGATGAGAACCCCCACGCTGCTCGTAGGTGCCGA
Mb2238	2346	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCTTCTCCGTCCAGAT
	2347	R	GATGATGATGATGATGAGAACCCCCCAGTCCTAAATCGGCCTCGA
Mb2244	2348	F	GTTTAACTTTAAGAAGGAGATATACCATGACGGAAAAGACGCCCGA
	2349	R	GATGATGATGATGATGAGAACCCCCAACGTCGTAGTACAGCGCGA
Mb2265	2350	F	GTTTAACTTTAAGAAGGAGATATACCATGGCGTCGTATTTGCCCGA
	2351	R	GATGATGATGATGATGAGAACCCCCGGCCCCGGGACCGGGATCCG
Mb2270	2352	F	GTTTAACTTTAAGAAGGAGATATACCATGGGGGTCCCCCCGCTTGC
	2353	R	GATGATGATGATGATGAGAACCCCCGTACCGTCCGAAGGCGATTG
Mb2397c	2354	F	GTTTAACTTTAAGAAGGAGATATACCATGAAGATGGTGAAATCGAT
	2355	R	GATGATGATGATGATGAGAACCCCCGTTCCCTGCGGCCTGCAGCA
Mb2587	2356	F	GTTTAACTTTAAGAAGGAGATATACCATGACCGGTGGCGCCACCGG
	2357	R	GATGATGATGATGATGAGAACCCCCTTGACCGGTGGCGCCACCGG
Mb2624c	2358	F	GTTTAACTTTAAGAAGGAGATATACCATGATCGCCTCGGTCCGCGG
	2359	R	GATGATGATGATGATGAGAACCCCCTCGGGCCTTCCCCAGCAACG
Mb2656	2360	F	GTTTAACTTTAAGAAGGAGATATACCATGTCATCGGGCAATTCATC
	2361	R	GATGATGATGATGATGAGAACCCCCAGTCAGCGACTCGCGTGCC
Mb2810	2366	F	GTTTAACTTTAAGAAGGAGATATACCATGAGTACGTTTAGAGAATG
	2367	R	GATGATGATGATGATGAGAACCCCCTCCGACCTGACGCATCGTTT
Mb2913c	2368	F	GTTTAACTTTAAGAAGGAGATATACCATGGCGAACTTCACTGCCGC
	2369	R	GATGATGATGATGATGAGAACCCCCAGCCTGGCCCACCTCGAAGC
Mb2965c*	2370	F	GTTTAACTTTAAGAAGGAGATATACCATGGAATCACGTGTCACTCC
	2371	R	GATGATGATGATGATGAGAACCCCCGAAGACCCGGTAGCCGTCCT
Mb2965c*	2372	F	GTTTAACTTTAAGAAGGAGATATACCATGACGCCCAAGGTGCCGTACTA

	2373	R	GATGATGATGATGATGAGAACCCCCGACCTGCACCCGCATACCGT
Mb2965c*	2374	F	GTTTAACTTTAAGAAGGAGATATACCATGGACGAACACACGGACGT
	2375	R	GATGATGATGATGATGAGAACCCCCTGATGCCGCCGGTGCCGCCG
Mb2970c	2376	F	GTTTAACTTTAAGAAGGAGATATACCATGAATGATGGAAAACGGGC
	2377	R	GATGATGATGATGATGAGAACCCCCGTCGACGTTGACGGGTTCGT
Mb3026c	2378	F	GTTTAACTTTAAGAAGGAGATATACCATGGCACTAGAGATGTTCTA
	2379	R	GATGATGATGATGATGAGAACCCCCGGCCGTCTCGGTGATCGGGC
Mb3071	2380	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCACTGTTGCCGCCTA
	2381	R	GATGATGATGATGATGAGAACCCCCCAGGGCTGAGATGTCGATGA
Mb3074c	2382	F	GTTTAACTTTAAGAAGGAGATATACCATGACTGGAAACGCAAAGCT
	2383	R	GATGATGATGATGATGAGAACCCCCGAAGTCCCAGTCATCGTCCT
Mb3424c	2388	F	GTTTAACTTTAAGAAGGAGATATACCATGCCCGACGAGCTGAAGCC
	2389	R	GATGATGATGATGATGAGAACCCCCCTTCTGGCAGGTGAACTGGT
Mb3451c	2390	F	GTTTAACTTTAAGAAGGAGATATACCATGAGCAAGCTGATCGAATA
	2391	R	GATGATGATGATGATGAGAACCCCCGTGCGCGTGCCCGTGGTGAT
Mb3473c	2392	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGGTCGCCGAGACCAA
	2393	R	GATGATGATGATGATGAGAACCCCCTTGCGCCACCTGCTTGAGCT
Mb3486c	2394	F	GTTTAACTTTAAGAAGGAGATATACCATGCTGATCTCACAGCGCCC
	2395	R	GATGATGATGATGATGAGAACCCCCAAGCTGTTCGGTTTCGGCGT
Mb3487c	2396	F	GTTTAACTTTAAGAAGGAGATATACCATGGCTCGTTACACCGGACC
	2397	R	GATGATGATGATGATGAGAACCCCCCTTTGAGTAGTACTCGACGA
Mb3687c	2398	F	GTTTAACTTTAAGAAGGAGATATACCATGGCGGGCATGAGCGTCCC
	2399	R	GATGATGATGATGATGAGAACCCCCCAGGTTGCCAACCTTTCGGG
Mb3749c	2400	F	GTTTAACTTTAAGAAGGAGATATACCATGTCGTTCGACTCTCTTAG
	2401	R	GATGATGATGATGATGAGAACCCCCGCGCACGTTCGGCGCCGAGG
Mb3826	2402	F	GTTTAACTTTAAGAAGGAGATATACCATGCCCGAGTACGACCTAGA
	2403	R	GATGATGATGATGATGAGAACCCCCCGTCTGCTCGGCAGTGAACG
Mb3871	2404	F	GTTTAACTTTAAGAAGGAGATATACCATGACAGAATACGAAGGGCC
	2405	R	GATGATGATGATGATGAGAACCCCCGAGGCGGCCCCCGGCAGCGT
Mb3876	2406	F	GTTTAACTTTAAGAAGGAGATATACCATGGCCGAATACACCTTGCC
	2407	R	GATGATGATGATGATGAGAACCCCCGCCGAATGTCAACCCCTTGG

***Targets are represented by 2 or three fragments**

Table S4. Primers used for cloning targets in pT7CFE1

locus tag	Primer#	D	Sequence (5' to 3')
Mb3645c	1632	F	CGATGATAATATGGCCACCACCCATATGACGGAAAACTTGACCGT
	1633	R	GTGGTGGTGGTGGTGGTGCTCGAGGGTAAACAACCCGTCGATAG
Mb3905	1634	F	CGATGATAATATGGCCACCACCCATATGACAGAGCAGCAGTGGAA
	1635	R	GTGGTGGTGGTGGTGGTGCTCGAGTGCGAACATCCCAGTGACGT
Mb0368	1636	F	CGATGATAATATGGCCACCACCCATATGTCCAACGCACCCGAGCC

	1637	R	GTGGTGGTGGTGGTGGTGCTCGAGGTTGCTGGAGGACTGGCAGA
Mb0768	2533	F	CGATGATAATATGGCCACCACCCATATGGCCGGCGGCATTAGTAGAA
	2534	R	GTGGTGGTGGTGGTGGTGCTCGAGCATCAAGCCGTTCAGCCCGT
MB0856	2535	F	CGATGATAATATGGCCACCACCCATATGGGGTTGATCGGTGTCGGAG
	2536	R	GTGGTGGTGGTGGTGGTGCTCGAGGTTACCGAGCCCGAAGGC
Mb1002	2537	F	CGATGATAATATGGCCACCACCCATATGCACGGCGGCGCCGCCCAAT
	2538	R	GTGGTGGTGGTGGTGGTGCTCGAGGTTAAAGGTGGTCGTCCCGA
Mb1121	2539	F	CGATGATAATATGGCCACCACCCATATGGGTAATGGCGTCAGCAGCAG
	2540	R	GTGGTGGTGGTGGTGGTGCTCGAGCGGCGACACGCCAGCCGTGC
Mb1360c	2541	F	CGATGATAATATGGCCACCACCCATATGTCGTTTGTAATCGCCGC
	2542	R	GTGGTGGTGGTGGTGGTGCTCGAGCGACAACCCGGGTGATCCGT
Mb1476c	2543	F	CGATGATAATATGGCCACCACCCATATGTCGAACGTGATGGTAGT
	2544	R	GTGGTGGTGGTGGTGGTGCTCGAGCCCGTGCTTTCCTTGCGCGC
Mb2410c	2545	F	CGATGATAATATGGCCACCACCCATATGACACCGGGTTTGCTTAC
	2546	R	GTGGTGGTGGTGGTGGTGCTCGAGATCGTCCCTGCTCCCCGAAC
Mb2667c	2547	F	CGATGATAATATGGCCACCACCCATATGGCCAACATTGGGTCGACG
	2548	R	GTGGTGGTGGTGGTGGTGCTCGAGGCCCATATTGCCCTGGATG
Mb2878	2549	F	CGATGATAATATGGCCACCACCCATATGATCAGCACGGCAAATGGTG
	2550	R	GTGGTGGTGGTGGTGGTGCTCGAGCGAACCGTTCTTGCCGGAG
Mb0419c	2551	F	CGATGATAATATGGCCACCACCCATATGACGCGCCGGGCCCTCCT
	2552	R	GTGGTGGTGGTGGTGGTGCTCGAGGTCCACATACCTCGGCGTGG
Mb0975c	2553	F	CGATGATAATATGGCCACCACCCATATGGCAGCGATTCGCACACC
	2554	R	GTGGTGGTGGTGGTGGTGCTCGAGACCGGTGTAATTGCCGACGC
Mb0476	2555	F	CGATGATAATATGGCCACCACCCATATGTCTGTCGTCGGCACCCC
	2556	R	GTGGTGGTGGTGGTGGTGCTCGAGGTGGAACTGGCCCTCTTCGG
Mb0485	2557	F	CGATGATAATATGGCCACCACCCATATGGCTGAAAACTCGAACAT
	2558	R	GTGGTGGTGGTGGTGGTGCTCGAGCTTCTGGGTGACCTTCTTGG
Mb2270	2559	F	CGATGATAATATGGCCACCACCCATATGGGGGTCCCCCCGCTTGC
	2560	R	GTGGTGGTGGTGGTGGTGCTCGAGGTACCGTCCGAAGGCGATTG
Mb2624c	2561	F	CGATGATAATATGGCCACCACCCATATGATCGCCTCGGTCCGCGG
	2562	R	GTGGTGGTGGTGGTGGTGCTCGAGTCGGGCCTTCCCCAGCAACG
Mb2965c	2563	F	CGATGATAATATGGCCACCACCCATATGACGCCCAAGGTGCCGTACTA
	2564	R	GTGGTGGTGGTGGTGGTGCTCGAGGACCTGCACCCGCATACCGT
Mb3473c	2565	F	CGATGATAATATGGCCACCACCCATATGTCGGTCGCCGAGACCAA
	2566	R	GTGGTGGTGGTGGTGGTGCTCGAGTTGCGCCACCTGCTTGAGCT
Mb3871	2567	F	CGATGATAATATGGCCACCACCCATATGACAGAATACGAAGGGCC
	2568	R	GTGGTGGTGGTGGTGGTGCTCGAGGAGGCGGCCCCCGGCAGCGT

Appendix 3. Information on the Selected Candidates

Table S5. Protein identity and signal peptide of proteins detected by Bioinformatics.

Gene ID	Mb#	Protein name	Signal peptide detected
GNBNAFKM_01274	Mb0041c	secreted proline rich protein MTC28 (proline rich 28 kDa antigen)	No
GNBNAFKM_01297	Mb0064	oxidoreductase	Yes
GNBNAFKM_00514	Mb0129	PE-PGRS family protein	Yes
GNBNAFKM_00456	Mb0184	mce associated membrane protein	No
GNBNAFKM_00352	Mb0285c	hypothetical protein	Yes
GNBNAFKM_00352	Mb0286c	PE-PGRS family protein	Yes
GNBNAFKM_00351	Mb0287c	PE-PGRS family protein	Yes
GNBNAFKM_00332	Mb0305	PE-PGRS family protein	Yes
GNBNAFKM_01718	Mb0368	hypothetical protein	No
GNBNAFKM_01658	Mb0426	glutamine-binding lipoprotein	Yes
GNBNAFKM_00827	Mb0545	lipoprotein aminopeptidase LpqL	No
GNBNAFKM_00779	Mb0593c	hypothetical protein	Yes
GNBNAFKM_00676	Mb0690	PE-PGRS family protein	Yes
GNBNAFKM_00599	Mb0768	lipoprotein LpqP	Yes
GNBNAFKM_00590	Mb0777c	PE-PGRS family protein	Yes
GNBNAFKM_02091	Mb0845c	PPE family protein	Yes
GNBNAFKM_02076	Mb0856	hypothetical protein	No
GNBNAFKM_02074	Mb0858	PE-PGRS family protein	Yes
GNBNAFKM_02035	Mb0896c	lipoprotein LpqQ	Yes
GNBNAFKM_01923	Mb1002	PE-PGRS family protein	Yes
GNBNAFKM_04091	Mb1006c	hypothetical protein	No
GNBNAFKM_03519	Mb1018c	PE-PGRS family protein	Yes
GNBNAFKM_03545	Mb1044c	PE-PGRS family protein	Yes
GNBNAFKM_03552	Mb1050	serine-rich protein	No
GNBNAFKM_03993	Mb1096c	lipoprotein LpqT	Yes

GNBNAFKM_03992	Mb1097c	lipoprotein LpqU	Yes
GNBNAFKM_03487	Mb1116	PE-PGRS family protein	Yes
GNBNAFKM_03485	Mb1118	PE-PGRS family protein	Yes
GNBNAFKM_03483	Mb1121	PE-PGRS family protein	Yes
GNBNAFKM_02907	Mb1262c	PE family protein	No
GNBNAFKM_02894	Mb1275c	PE-PGRS family protein	Yes
GNBNAFKM_03672	Mb1360c	hypothetical protein	Yes
GNBNAFKM_03315	Mb1476c	PE-PGRS family protein	Yes
GNBNAFKM_03326	Mb1485c	PE-PGRS family protein	Yes
GNBNAFKM_04081	Mb1487c	hypothetical protein	No
GNBNAFKM_02782	Mb1575c	PE-PGRS family protein	Yes
GNBNAFKM_01856	Mb1679c	PE-PGRS family protein	Yes
GNBNAFKM_04023	Mb1783c	PE-PGRS family protein	Yes
GNBNAFKM_04024	Mb1784c	methylmalonyl-CoA mutase	No
GNBNAFKM_04028	Mb1788	PPE family protein	Yes
GNBNAFKM_02276	Mb1831c	PE-PGRS family protein	Yes
GNBNAFKM_03873	Mb2006c	hypothetical protein	Yes
GNBNAFKM_02192	Mb2108	hypothetical protein	No
GNBNAFKM_02178	Mb2125c	phospholipase C 4 PLCD	No
GNBNAFKM_00971	Mb2323	cutinase	No
GNBNAFKM_01062	Mb2410c	hypothetical protein	No
GNBNAFKM_01469	Mb2506c	cutinase CFP21	No
GNBNAFKM_03111	Mb2517c	hypothetical protein	No
GNBNAFKM_03007	Mb2606c	hypothetical protein	Yes
GNBNAFKM_03951	Mb2667c	cutinase	Yes
GNBNAFKM_02542	Mb2734	PPE family protein	Yes
GNBNAFKM_02536	Mb2740c	resuscitation-promoting factor RpfD	No
GNBNAFKM_03396	Mb2878	hypothetical protein	No
GNBNAFKM_01517	Mb3183c	hypothetical protein	No
GNBNAFKM_03611	Mb3402	hypothetical protein	Yes
GNBNAFKM_03778	Mb3453c	hypothetical protein	No
GNBNAFKM_03760	Mb3469c	PE-PGRS family protein	Yes

GNBNAFKM_03748	Mb3481	hydrolase	No
GNBNAFKM_03834	Mb3537	hypothetical protein	Yes
GNBNAFKM_03835	Mb3538	PE-PGRS family protein	Yes
GNBNAFKM_04065	Mb3541	PPE family protein	Yes
GNBNAFKM_03835	Mb3543	PE-PGRS family protein	Yes
GNBNAFKM_00272	Mb3562	DNA-binding/iron metalloprotein/AP endonuclease	No
GNBNAFKM_00080	Mb3751	hypothetical protein	No
GNBNAFKM_00069	Mb3761c	cutinase	Yes
GNBNAFKM_00036	Mb3792	PE-PGRS family protein	Yes
GNBNAFKM_01101	Mb3833c	PE-PGRS family protein	Yes
GNBNAFKM_01109	Mb3841	PE-PGRS family protein	Yes
GNBNAFKM_01183	Mb3909c	PE-PGRS family protein	Yes
GNBNAFKM_01195	Mb3921c	PPE family protein	No
GNBNAFKM_01666	Mb0419c	cutinase	No
GNBNAFKM_01951	Mb0975c	hypothetical protein	No
GNBNAFKM_03236	Mb1397c	hypothetical protein	No
GNBNAFKM_02737	Mb1530	secreted MPT51/MPB51 antigen protein FbpD	Yes
GNBNAFKM_01807	Mb1724	hypothetical protein	No
GNBNAFKM_01028	Mb2377c	hypothetical protein	No
GNBNAFKM_02953	Mb2554c	PPE family protein	No

Table S6. Protein identity and signal peptide of proteins detected by ECSVM.

Strain ID	Mb #	Protein name	Signal Peptide detected
GNBNAFKM_00208	MB3626C	PE-PGRS family protein	Yes
GNBNAFKM_00213	MB3621C	PE-PGRS family protein	Yes
GNBNAFKM_00531	MB0113	PPE family protein	No
GNBNAFKM_00600	MB0767	PE-PGRS family protein	Yes
GNBNAFKM_01019	MB2369C	PE-PGRS family protein	Yes
GNBNAFKM_01070	MB2418	PE-PGRS family protein	Yes

GNBNAFKM_01176	MB3903	PPE family protein	Yes
GNBNAFKM_01633	MB0450C	PPE family protein	Yes
GNBNAFKM_02111	MB2186C	hypothetical protein	Yes
GNBNAFKM_02273	MB1828	PPE family protein	Yes
GNBNAFKM_02518	MB2761	PE-PGRS family protein	Yes
GNBNAFKM_03022	MB2622	PE-PGRS family protein	Yes
GNBNAFKM_03269	MB1431C	PE-PGRS family protein	Yes
GNBNAFKM_03590	MB3420	PE-PGRS family protein	Yes
GNBNAFKM_03872	MB2005	PE-PGRS family protein	Yes
GNBNAFKM_03989	MB1797	PE-PGRS family protein	Yes
GNBNAFKM_03804	MB1228	PPE family protein	No
GNBNAFKM_04029	MB1789C	hypothetical protein	No
GNBNAFKM_01516	MB3184C	PPE family protein	Yes

Table S7. The gene size and localization of proteins identified from the review of previous proteomic studies.

RV#	Mb #	strain ID	Size (bp)	localisation
Rv0242c	Mb0248c	GNBNAFKM_00389	1365	U
Rv0243	Mb0249	GNBNAFKM_00388	1323	CW/ PM/ C
Rv0248c	Mb0254c	GNBNAFKM_00382	1941	CW/ PM
Rv0270	Mb0276	GNBNAFKM_00360	1683	CW/ PM
Rv0333	Mb0340	GNBNAFKM_00295	375	U
Rv0431	Mb0439	GNBNAFKM_01644	495	CW
Rv0462	Mb0471	GNBNAFKM_01612	1935	C/ PM/ EC
Rv0467	Mb0476	GNBNAFKM_01607	1287	C/ PM/ EC
Rv0469	Mb0478	GNBNAFKM_01605	861	C/PM
Rv0475	Mb0485	GNBNAFKM_01598	600	CW/ PM/ C
Rv0503c	Mb0515c	GNBNAFKM_01567	909	C/PM
Rv0569	Mb0584	GNBNAFKM_00789	267	CW/C
Rv0577	Mb0592	GNBNAFKM_00780	786	EC/PM
Rv0583c	Mb0598c	GNBNAFKM_00774	687	CW/ EC/PM
Rv0632c	Mb0649c	GNBNAFKM_00721	696	U
Rv0639	Mb0658	GNBNAFKM_00709	717	CW/ PM/ C
Rv0703	Mb0723	GNBNAFKM_00643	303	C/PM
Rv0707	Mb0727	GNBNAFKM_00639	825	CW/PM
Rv0719	Mb0740	GNBNAFKM_00627	540	CW/PM

Rv0733	Mb0754	GNBNAFKM_00613	546	C/PM/CW
Rv0761c	Mb0784c	GNBNAFKM_00582	1128	C/PM
Rv0801	Mb0824	GNBNAFKM_04049	348	-
Rv0831c	Mb0854c	GNBNAFKM_02082	816	C/PM/CW
Rv0884c	Mb0908c	GNBNAFKM_02023	1131	C/EC/PM
Rv0896	Mb0920	GNBNAFKM_02009	1296	C/PM
Rv0905	Mb0929	GNBNAFKM_02000	732	CW/PM
Rv0932c	Mb0956c	GNBNAFKM_01968	1113	CM/EC
Rv0984	Mb1010	GNBNAFKM_03511	546	PM
Rv1070c	Mb1099c	GNBNAFKM_03504	774	CW/PM
Rv1080c	Mb1109c	GNBNAFKM_03494	495	CW/ PM/ cytosol
Rv1093	Mb1123	GNBNAFKM_03481	1281	CW/ PM/ cytosol
Rv1133c	Mb1164c	GNBNAFKM_03438	2280	CW/ PM/ cytosol
Rv1181	Mb1213	GNBNAFKM_03721	1494	PM
Rv1270c	Mb1301c	GNBNAFKM_02865	735	CW/EC/PM
Rv1352	Mb1387	GNBNAFKM_03644	372	EC
Rv1826	Mb1857	GNBNAFKM_02301	405	U
Rv1392	Mb1427	GNBNAFKM_03265	1212	CW/ PM/ cytosol
Rv1630	Mb1656	GNBNAFKM_01879	1446	CW/ PM/ cytosol
Rv1915	Mb1951c	GNBNAFKM_02401	1929	U
Rv1916	Mb1950	GNBNAFKM_02399	2301	U
Rv2140c	Mb2164c	GNBNAFKM_02136	531	EC/PM
Rv2215	Mb2238	GNBNAFKM_00886	1662	CW/ PM/ C
Rv2220	Mb2244	GNBNAFKM_00892	1437	CW/ PM/ C/EC
Rv2241	Mb2265	GNBNAFKM_00916	2706	CW/EC/PM
Rv2246	Mb2270	GNBNAFKM_00921	1317	CW/ PM/ C
Rv2376c	Mb2397c	GNBNAFKM_01048	507	CW/EC
Rv2557	Mb2587	GNBNAFKM_02987	675	U
Rv2593c	Mb2624c	GNBNAFKM_03024	591	PM
Rv2623	Mb2656	GNBNAFKM_03963	894	CW/ PM/ C
Rv2626c	Mb2659c	GNBNAFKM_03960	432	CW/ PM/ C/EC
Rv2779c	Mb2801c	GNBNAFKM_02481	438	C
Rv2787	Mb2810		1764	CW
Rv2889c	Mb2913c	GNBNAFKM_03359	816	CW/ PM/ C
Rv2940c	Mb2965c	GNBNAFKM_02675	6336:	U
Rv2945c	Mb2970c	GNBNAFKM_02669	702	EC/CW/PM
Rv3001c	Mb3026c	GNBNAFKM_02608	1014	PM/C
Rv3045	Mb3071	GNBNAFKM_03213	1041	CW/ PM/ C
Rv3048c	Mb3074c	GNBNAFKM_03210	975	C
Rv3196A	Mb3219c	GNBNAFKM_01478	174	U

Rv3354	Mb3389	GNBNAFKM_03624	390	EC
Rv3392c	Mb3424c	GNBNAFKM_03586	864	C/PM
Rv3417c	Mb3451c	GNBNAFKM_03780	1620	C/PM/CW
Rv3443c	Mb3473c	GNBNAFKM_03756	534	C/PM/CW
Rv3457c	Mb3486c	GNBNAFKM_03743	1044	C/PM/CW
Rv3458c	Mb3487c	GNBNAFKM_03742	606	C/PM/CW
Rv3663c	Mb3687c	GNBNAFKM_00149	1656	U
Rv3722c	Mb3749c	GNBNAFKM_00084	1308	EC
Rv3797	Mb3826	GNBNAFKM_00001	1782	U
Rv3841	Mb3871	GNBNAFKM_01140	546	CW/ PM/ C/EC
Rv3846	Mb3876	GNBNAFKM_01149	624	C/P/PM/EC

*C= Cytoplasmic, CW=Cell wall, EC= Extracellular, P= Periplasmic, PM= Plasma membrane U= Unknown

Table S8. Predicted signal peptide and size of the proteins identified after conducting the proteomic study.

MB#	locus tag in *M. bovis* MBWGS010	[a]Signal peptide	Protein name	Size (kDa)
MB3904	GNBNAFKM_01177	N	CFP-10	10
MB3905	GNBNAFKM_01178	N	Early secretory antigenic target(ESAT-6)	6
MB2900	GNBNAFKM_03372	Y	MPB70	19
MB0448	GNBNAFKM_01635	N	60 kDa chaperonin 2- GroL2	57
MB2268	GNBNAFKM_00919	N	Meromycolate extension acyl carrier protein	13
MB2898	GNBNAFKM_03375	Y	Cell surface glycolipoprotein MPB83	22
MB1918C	GNBNAFKM_02365	Y	Antigen 85 complex B (Ag85B)	35
MB2397C	GNBNAFKM_01048	Y	Low molecular weight antigen MTB12	17
MB2002C	GNBNAFKM_03867	Y	Antigen MPB64	25
MB0192A	GNBNAFKM_00447	N	Metallothionein OS	5.7
MB1858	GNBNAFKM_02302	N	Uncharacterized protein Mb1858	17
MB1891	GNBNAFKM_02338	Y	Fibronectin attachment protein (FAP-B)	33
MB2970C	GNBNAFKM_02669	Y	Lipoprotein LppX	24
MB3834C	GNBNAFKM_01102	Y	Antigen 85 complex A	35
MB0358	GNBNAFKM_01728	N	Heat shock protein 70 - DnaK	67
MB1967	GNBNAFKM_02417	N	Thiol peroxidase (TPX)	17
MB1662	GNBNAFKM_01872	N	Iron-regulated universal stress protein family protein TB15.3	15
MB0055	GNBNAFKM_01288	N	Single-stranded DNA-binding	17

			protein	
MB3672C	GNBNAFKM_00165	N	Probable cold shock protein A cspA	7
MB2057C	GNBNAFKM_02243	N	14 kDa antigen- HspX-HSP 16.3	16
MB3945	GNBNAFKM_01220	N	Thioredoxin/ MPT46	13

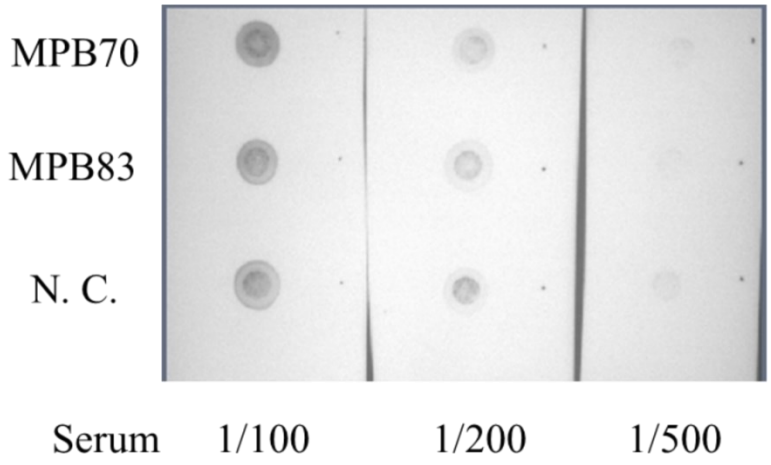
[a] Y- Signal peptide present; N- Signal peptide absent

Appendix 4. Genes used in PCR Optimization

Table S9. The seven genes used in PCR in Figure 5-4, the gene-lengths and GC content.

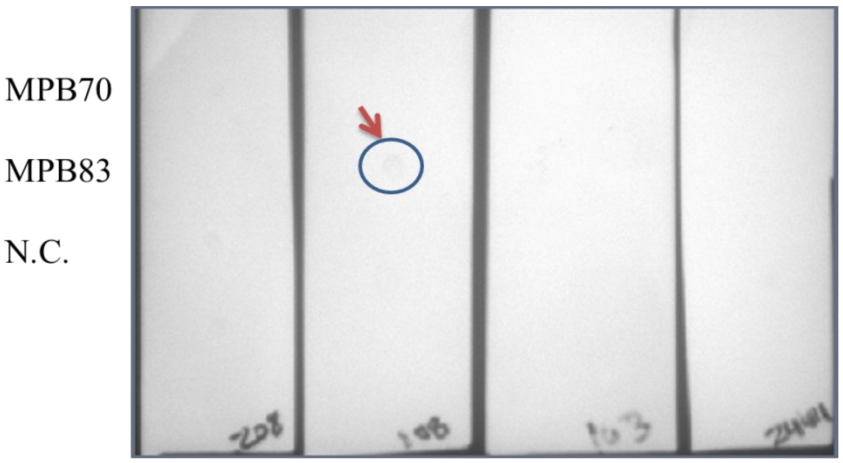
Gene	length (bp)	GC content (%)
Mb1275c	1689	74.8
Mb0545	1746	77.6
Mb0896c	1827	74.9
Mb1679c	1794	70
Mb1783c	1692	60
Mb2108	2166	67
Mb3402	1881	74.6

Appendix 5. Microarray Optimization



S Figure 3. Dot blot analysis showing two proteins, MPB70 and MPB83, in addition to negative control protein, probed against a positive serum.

The blotted proteins were probed against positive serum from experimentally infected animals at three dilutions. Strong reactions are observed against the N.C. N.C.= negative control, reaction containing all the constituents of the IVTT except the plasmid DNA.

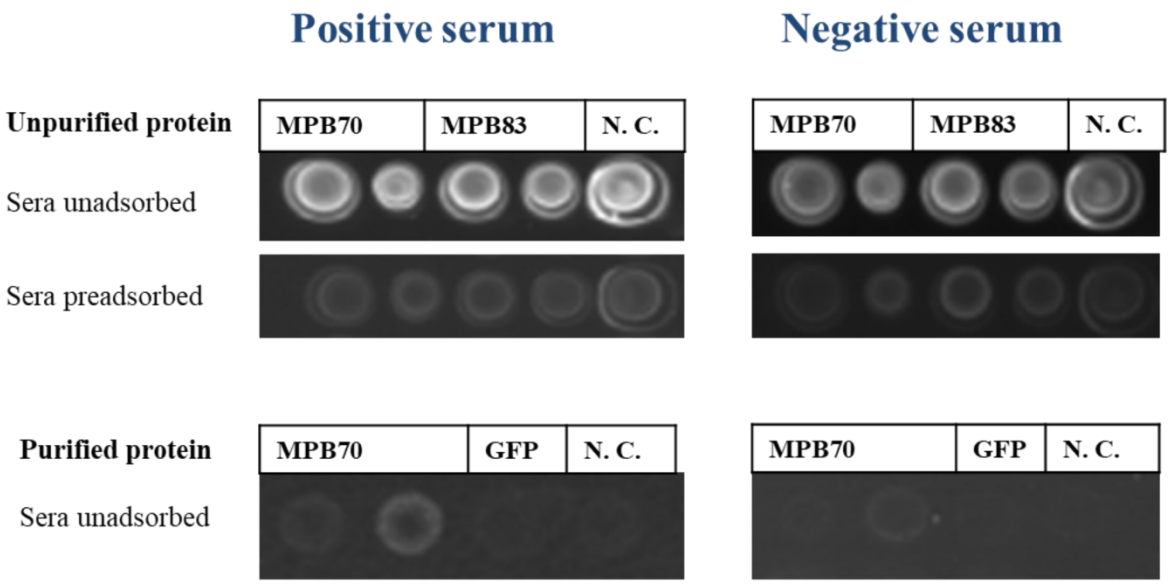


Serum#	S#208	S#108	S#103	S#2441
Dilution	1/100	1/100	1/100	1/100

S Figure 4. Dot blot analysis showing two proteins, MPB70 and MPB83, and negative control probed against 4 pre-adsorbed sera from experimentally infected cattle at 1:100 dilution.

The pre-adsorbed sera fail to show a positive reaction against the proteins, MPB70 and MPB83, except for a weak reaction against MPB83 by S#108 (red arrow).

N.C.= negative control, reaction containing all the constituents of the IVTT except the plasmid DNA.



S Figure 5. Microarray optimization using purified and unpurified proteins.
MPB70, MPB83, and control proteins were printed on microarray chips before and after purification. All proteins were probed using a confirmed positive serum and a negative serum. Before protein purification, the positive serum picture shows the difficulty to differentiate between the expected reaction to MPB70 and negative control (N.C.). The results are similar with both the positive unadsorbed and the pre-adsorbed serum. After purification, the positive serum shows a strong reaction to MPB70 but not N.C. A negative serum is shown for comparison.
N.C. is a "lysate only, no DNA control"

Appendix 6. Recombinant Proteins used in the Microarray Heat Map, Figure 6-11

Table S10. Results of recombinant proteins MFI used in the microarray in the order illustrated in the microarray heat map.

	Recombinant protein name	MFI controls	S.D. controls	MFI positive	S.D. positive
Potential serodiagnostic antigens	Mb0598c*	105,380	55,597	180,374	74,377
	Mb3453c*	37,354	43,361	84,062	50,683
	Mb3469c*	28,026	22,053	79,039	26,281
	MPB70*	28,255	33,720	57,008	28,842
	MPB83*	23,278	20,084	50,063	27,917
	Mb2740c*	15,509	9,819	48,303	27,716
Immune-reactive antigens	Mb1784c*	10,044	5,354	40,257	25,579
	Mb0754*	21,584	20,103	61,832	70,061
	Mb1099c*	16,499	15,770	55,410	59,220
	Mb1656*	11,503	11,133	42,416	47,107
	Mb0592*	7,937	6,117	30,396	30,359
	Mb2517c*	2,739	2,952	28,509	34,566
	ESAT6.CFP10*	4,775	7,363	23,128	32,832
Weak/ cross-reacting antigens	Mb3451c	8,279	11,555	26,006	29,080
	Mb3743c.Mb0340	9,510	14,415	32,738	54,103
	Mb1427	14,511	22,185	42,704	64,659
	Mb1018c	35,978	44,970	57,852	47,552
	Mb1950	31,873	36,207	72,408	70,858
	Mb2913c	13,224	19,897	35,178	38,223
	Mb3402	2,851	5,035	21,735	43,506
	Mb2734	6,346	15,206	24,081	41,463
	Mb3792	4,259	5,488	24,710	21,048
	Mb1010	2,210	2,963	12,358	22,708
	Mb0249	7,282	7,829	20,315	24,122
	Mb0649c	10,834	10,661	21,598	20,067
	Mb0515c	9,449	11,909	10,732	18,400
	Mb0658	9,898	11,812	12,327	19,303
	Mb2659c.Mb2801c	11,382	15,024	19,009	20,227
	Mb1724	2,844	2,476	19,675	9,988
	Mb1679c	7,284	6,524	20,976	7,730

Mb0287c	4,153	1,478	28,349	11,206
Mb0254c	4,237	5,241	13,611	17,015
Mb0426	5,679	7,945	18,479	9,576
Mb0784c	4,731	6,093	15,453	13,526
Mb2810	427	625	767	2,024
Mb1301c	4,217	5,201	3,854	9,459
Mb3876	3,980	6,803	2,890	4,675
Mb2965c	10,082	10,022	11,797	13,019
Mb3487c	2,502	2,169	3,556	6,260
Mb2108	6,088	7,020	7,304	9,288
Mb3749c	10,734	14,637	9,509	9,443
Mb0727	5,689	9,353	5,432	8,149
Mb0908c	7,687	7,981	13,993	14,146
Mb0920	2,860	4,534	8,393	12,998
Mb0929	2,550	2,896	5,381	6,482
Mb1109c	588	846	1,475	3,653
Mb2164c	341	549	513	1,456
Mb2238	316	768	3,430	13,130
Mb3621C	87	226	1,977	7,364
Mb0113	528	1,374	67	208
Mb0854c	4,565	4,134	18,076	20,986
Mb0920	2,238	3,343	6,932	13,649
Mb1164c	4,753	4,157	15,351	15,233
Mb3826	977	1,751	3,218	9,298
Mb2970c	1,084	1,501	806	1,487
Mb3026c	3,033	2,721	10,401	17,932
Mb1123	3,233	4,413	5,094	9,059
Mb1431C	1,044	1,651	2,138	2,937
Mb1797	125	351	1,591	2,798
Mb1228	288	812	921	2,510
Mb3184C	508	972	1,540	4,720
Mb0956c	2,639	2,416	7,217	9,618
Mb1476c	3,314	6,285	5,346	8,282
Mb2244	4,382	5,220	3,826	3,919
Mb2265	9,080	11,188	13,020	11,392
Mb2397c	11,728	11,693	22,948	18,165
Mb3219c.Mb3389	7,235	9,223	8,779	7,319
Mb3903	1,315	2,135	3,637	5,326
Mb2186C	99	278	894	2,331
Mb1828	543	868	1,401	2,501
Mb3424c	2,293	3,102	2,515	4,783

Mb3481	4,147	7,693	7,335	10,048
Mb1387.Mb1857	872	2,058	1,095	2,526
Mb0064	230	647	71	312
Mb0305	2,290	6,258	3,510	8,252
Mb1575c	1,282	2,403	12,466	10,346
Mb0287c	238	670	6,733	5,945
Mb0856	1,783	2,824	11,102	11,031
Mb1485c	1,960	1,631	18,643	13,691
Mb1487c	717	1,668	5,576	7,054
Mb2517c	2,400	4,808	7,999	12,706
Mb0975c	663	1,231	10,689	9,660
Mb1397c	1,890	3,005	5,810	7,943
Mb1530	374	716	2,179	4,075
Mb0939c	1	1	2,675	4,219
Mb0285c	563	826	7,585	8,030
Mb0777c	1,141	1,328	4,931	6,197
Mb1002	133	374	5,298	7,784
Ag85B	2,033	2,133	8,539	6,040
Mb0041c	3,600	4,333	6,890	6,749
Mb0184	1,407	2,516	3,276	4,175
Mb3562	5,995	5,206	5,250	5,876
Mb3751	1,664	2,114	5,868	5,598
Mb3761c	6,724	5,136	27,190	14,404
Mb0545	3,529	5,041	4,145	4,449
Mb3538	774	1,468	4,274	6,736
Mb2006c	1,022	2,158	1,621	3,122
Mb1679c	3,466	3,182	13,209	8,742
Mb1121	948	1,851	4,074	8,461
Mb1096c	6,271	6,260	18,964	8,960
Mb2667c	154	412	71	311
Ag85A	3,151	6,849	1,717	2,986
Mb1831c	1,684	1,966	1,314	2,691
Mb3537	4,386	3,112	14,018	9,974
Mb0593c	4,886	5,611	6,560	9,129
Mb3904	2,391	2,777	1,105	2,691
Mb0768	1,986	2,863	3,425	5,488
Mb2377c	1,857	2,869	5,114	5,287
Mb2554c	3,125	4,716	2,706	5,208
Mb0285c	4,967	7,300	8,666	10,047
Mb1097c	1,620	1,195	8,882	6,052
Mb1116	825	1,224	4,675	5,611

Ag85C	1,033	1,741	7,101	8,697
Mb0285c	2,677	2,342	6,613	6,022
Mb3766c	7,573	4,114	25,286	15,196
Mb3538	4,194	3,963	16,265	12,116
Mb3833c	1,460	2,213	5,048	6,232
Mb3841	1,878	3,061	9,550	8,751
Mb3921c	2,571	3,137	8,505	6,609
Mb0419c	1,951	1,985	18,852	13,727
Mb1262c	1,025	1,582	4,490	5,296
Mb1275c	965	2,474	2,566	3,520
MB1006C	1,739	2,705	5,857	5,882
Mb0286c	1,560	3,774	505	1,040

Results are expressed as MFI and S.D of controls and positive serum samples.

*= statistically significant results of the independent t-test to compare the means between the control and infected groups.

Appendix 7. Additional Information related to Chapter 7

Table S11. Descriptive and comparative statistics for the zoom blot analysis of Mb3453c.

	Control (n = 110)	1CO (n = 19)	2CA (n =12)	3IA (n = 5)	4MI (n =14)	Experimentally infected (n=14)
Number of values	110	19	12	5	14	14
Minimum	756.0	1936	366.0	6322	1566	1841
25% Percentile	2605	2870	4962	7077	3938	4182
Median	5815	7881	7438	16529	9243	7955
75% Percentile	8142	12159	9721	18248	15064	12275
Maximum	20049	42902	16701	19877	23692	17594
Range	19293	40966	16335	13555	22126	15753
Mean	6443	9038	7421	13436	10260	8435
Std. Deviation	4458	9289	4155	5983	6612	4849
Std. Error of Mean	425.0	2131	1199	2676	1767	1296
p value for the statistical test of difference between control and the group [$]	N/A	0.2647	0.4691	0.001**	0.0053*	0.0908

[$]Comparative study done by Mann-Whitney U for the difference between the control and 1CO, Independent t-test for the difference between the control and each of 2CA, 3IA, 4MI and sera from experimentally infected animals.

[#]p. p-value 0.0191[*] for comparing between the six groups using the Kruskal Wallis test.

p= Two-tailed p test significant at $p<0.05$

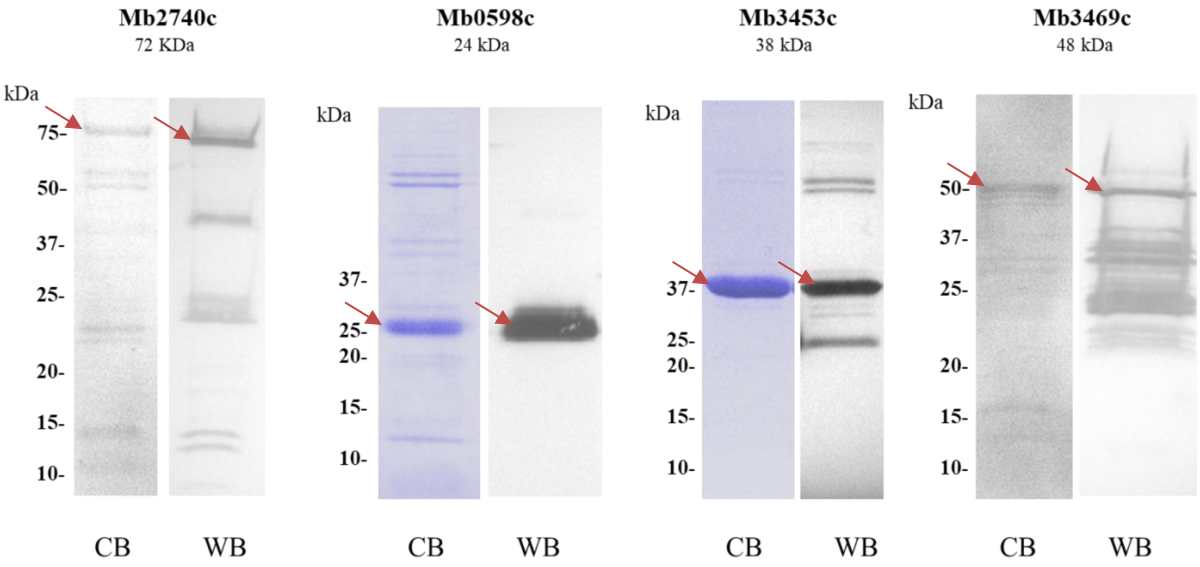
Table S12. Descriptive and comparative statistics for Mb3469c ELISA results.

	Control (n = 110)	**1CO (n = 19)**	**2CA (n =12)**	**3IA (n = 5)**	**4MI (n =14)**	**Experimentally infected (n=14)**
Number of values	110	19	12	5	14	14
Minimum	0.01600	0.04000	0.1140	0.04800	0.06000	0.05400
25% Percentile	0.07225	0.1240	0.1278	0.04900	0.08350	0.06713
Median	0.1075	0.2190	0.3055	0.07800	0.1175	0.1015
75% Percentile	0.1640	0.3270	0.4463	0.1580	0.2373	0.2645
Maximum	0.7350	0.4810	0.6940	0.2170	0.4780	0.6120
Range	0.7190	0.4410	0.5800	0.1690	0.4180	0.5580
Mean	0.1300	0.2178	0.3077	0.09840	0.1686	0.1984
Std. Deviation	0.09109	0.1314	0.1835	0.06958	0.1241	0.1855
Std. Error of Mean	0.008685	0.03014	0.05298	0.03112	0.03316	0.04957
the p-value for the statistical test of the difference between control and the group [$]		0.0033**	<0.0001**	0.4469	0.3148	0.4584

[$] Comparative study done by Mann-Whitney U test for the difference between the control and each of 1CO, 4MI and experimental sera, Independent t-test for the difference between the control and each of 2CA and 3IA

[#]p. p-value 0.0001** for comparing between the six groups using the Kruskal Wallis test.

p= Two-tailed p test significant at p<0.05



S Figure 6. Mb2740c (72 kDa), Mb0598c (24 kDa), Mb3453c(38 kDA), and Mb3469c (48 kDa) purified proteins.

Proteins are represented by Coomassie blue staining (CB) and Western blot (WB) using monoclonal mouse Anti·Penta-His Antibodies. The protein band appears as the right sized band, smaller bands may denote fragments of the proteins that still carry the 6xHis-tag and were detected by WB. Larger fragments may be due to the protein clumping. Red arrows show the proteins of expected sizes.