Discovery and validation of pathogen specific biomarkers and comparative genomics for diagnosis and tracking of bovine tuberculosis

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Sylvia Irene Wanzala

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser, Srinand Sreevatsan

September 2017
© Sylvia Irene Wanzala 2017
Acknowledgements

I would like to thank my adviser, Prof Srinand Sreevatsan for the tireless effort and incredible mentorship throughout my program; my colleagues in the Sri lab- Fernanda, Nicole, Anum and Moti as well as all the friends that I have made during the past few years in Minnesota.

I would like to thank all the different funding agencies that made it possible for this work to be carried out.

These are:

1. The Morris Animal Foundation
2. University of Minnesota Grand Challenges Grant
3. The MNDrive Fellowship
4. The NIH Research Training Grant # R25 TW009345 funded by the Fogarty International Center, the National Institute of Mental Health, the NIH Office of the Director Office of Research on Women's Health and the NIH Office of the Director the Office of AIDS Research
5. The Broad Institute - for sequencing of the isolates
6. The VMED/CMB travel grant
7. The Microbial and Plant Genomics Institute (MPGI) travel grant
8. The Center for Global Health and Social Responsibility Scholarship
Dedication

This thesis is dedicated to my parents, Mr. John Bikala Wanzala and Sarah Khanakwa Wakooli and my six siblings, Thomas, Jacquelyn, Hellen, Gladys, Catherine, and Emmanuel. To my dear nieces, Susan and Precious and my adorable nephew, Marvin. I love you all very much!

Thank you so much for the unwavering support, love and encouragement in this journey. Its surely been Ebenezer! May the Lord bless you always!

Abstract

Current tools for the detection of *Mycobacterium tuberculosis* complex infections notably *M. bovis* are inadequate. New methods are required that are quick, inexpensive and accurate. In addition, the contribution of *Mycobacterium bovis* to the proportion of tuberculosis cases in humans is unknown. In this thesis, these questions were addressed by applying novel minimally invasive tools-pathogen specific peptides and circulating small and miRNA (biomarkers) - to detect infection with *Mycobacterium tuberculosis* complex organisms quickly, efficiently and accurately. Whole genome sequencing was also carried out and used to extract a minimum Single Nucleotide Polymorphism (SNP) set that would be used in a SNP chip as a means to quickly detect *Mycobacterium tuberculosis* complex infection. Genome analysis in this study identified *M. bovis* in humans and great apes suggesting transmission from domesticated ruminants is possible in high TB prevalence regions of the world due to a dynamic and changing interface, which has created opportunity for exposure and transmission. These tools all offer specific approaches for the early identification of *M. bovis* and other *Mycobacterium tuberculosis* complex infections.
TABLE OF CONTENTS

Acknowledgements........................................................................................................... i
Dedication.......................................................................................................................... ii
Abstract.............................................................................................................................. iii
TABLE OF CONTENTS.................................................................................................. iv
List of Tables..................................................................................................................... vi
List of Figures ................................................................................................................... vi
INTRODUCTION.............................................................................................................. 1
CHAPTER ONE ................................................................................................................ 8
Biomarkers in the Diagnosis of Mycobacterium tuberculosis complex Infections . 8
  1.1 Introduction: ............................................................................................................... 8
    1.1.1 Biomarkers of tuberculosis in animals................................................................. 9
  1.2 Biomarkers defined. .................................................................................................. 12
    1.2.1 Characteristics of an Ideal Biomarker. ............................................................ 13
  1.3 Circulating biomarkers. ............................................................................................ 14
    1.3.1 Circulating nucleic acid approach...................................................................... 14
    1.3.2 Proteomics approach .......................................................................................... 18
  1.4 Circulating miRNA. .................................................................................................. 22
  1.5 Serum cytokines. ...................................................................................................... 22
  1.6 Cellular immune responses for potential biomarker application............................... 25
  1.7 Human tuberculosis. ................................................................................................ 26
  1.8 Biomarker challenges. ............................................................................................... 29
  1.9 Conclusions .............................................................................................................. 30

CHAPTER TWO ............................................................................................................. 32
Evaluation of Pathogen-Specific Biomarkers for the Diagnosis of Tuberculosis in
White-Tailed Deer (Odocoileus virginianus) ........................................................................ 32
  2.0 Background .............................................................................................................. 32
  2.1 Materials and Methods ............................................................................................ 35
  2.3 Data analysis ............................................................................................................ 37
  2.4 Results ...................................................................................................................... 38
    2.4.1 Experimentally infected deer ............................................................................. 38
    2.4.2 TB-specific biomarkers in TB-positive hunter-harvested deer............................. 40
    2.4.3 Prevalence of bovine TB in Minnesota deer ....................................................... 40
  2.5 Discussion ................................................................................................................. 41

CHAPTER THREE ......................................................................................................... 47
Develop a Circulating Nucleic-Acid Based Detection System to Detect
Mycobacterium bovis Infection in Animals ....................................................................... 47
  3.0 Background .............................................................................................................. 47
List of Tables

Table 3. 1-Small RNA mapping statistics ........................................................................53
Table 3. 2: sRNA of Mycobacterium bovis discovered at post-infection time points ......55
Table 4. 1: Nucleotide sequence of primers used to determine the presence of
Mycobacterium bovis ........................................................................................................65
Table 4. 2: PCR results for 136 PZA resistant Mycobacterium tuberculosis complex .....69
Table 4. 4: Mycobacterial genomes with Genbank accession numbers .......................70
Table 4. 5: Global Isolates for phylogenetic analysis .....................................................73
Table 4. 6: Drugs, genes and drug-resistance associated mutations .............................76
Table 4. 7: SNPs identified in antibiotic resistance associated regions in Mycobacterium
tuberculosis complex isolates ....................................................................................76

List of Figures

Figure 2. 1—The S/N values for biomarkers pks5 (A), MB2515c (B), and MB1895c (C) in
serum samples obtained from 8 white-tailed deer (Odocoileus virginianus)
experimentally infected with Mycobacterium bovis .................................................39
Figure 2. 2—Box-and-whisker plots of S/N values for 3 biomarkers in sera of 9 M bovis–
infected hunter-harvested white-tailed deer ...............................................................39
Figure 2. 3—Overall TB prevalence in sera of hunter-harvested deer by year for each
biomarker by use of 3 S/N cutoff values (2.5 [squares], 3.0 [circles], and 3.5
[diamonds]). Notice that there is a gradual decrease in TB prevalence from 2007 to
2010. ..........................................................................................................................41
Figure 4. 1: PCR result for RD9 (600bp) for some of the 133 IS6110 positive samples.
.................................................................................................................................68
Figure 4. 2: Progressive Mauve alignment of 17 genomes .............................................72
Figure 4. 3 Molecular Phylogenetic analysis of Mycobacterium tuberculosis complex
organisms by the Maximum Likelihood method. Evolutionary history was inferred
by using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano
model (94). ................................................................................................................74

Supplementary Figures

Figure 1: Studies from the Joint Clinical Research Centre (JCRC) studies from which
pyrazinamide resistant samples were drawn
Figure 2: Distribution by age and gender of case patients with pyrazinamide (PZA)
resistant isolates.
Figure 3: Source of PZA samples tested for Mycobacterium tuberculosis complex.
Figure 4: Resistance patterns for anti-tuberculosis drugs in PZA resistant isolates.
Figure 5: File of SNPs extracted from Mauve
Figure 6: RAST images showing percentage similarity of selected genomes to the reference strain AF2122/97.
Figure 7: Molecular phylogenetic analysis of \textit{Mycobacterium tuberculosis} complex organisms by the Maximum Likelihood Method

List of abbreviations

- BLAST: Basic Local Alignment Search Tool
- BCG: bacille Calmette-Guerin
- BLS-3: Biosafety level three containment
- BSE: bovine spongiform encephalopathy
- bTB: bovine tuberculosis
- CCT: comparative cervical test
- CFP-10: culture filtrate protein-10
- CNAs: Circulating nucleic acids
- CSF: colony stimulating factors
- CFT: Caudal Fold Test
- CHS: College of Health Sciences
- CWRU: Case Western Reserve University
- DIVA: Differentiating Infected from Vaccinated Animals
- DNR: Department of Natural Resources
- DR: drug resistant
- ESAT-6: early secretory antigenic target-6
- FAO: Food and Agricultural Organization
- HYK: Hasegawa-Kishino-Yano
- IFN: interferons
- IGRA: interferon gamma release assay
- IL: interleukin
- IRB: Institutional Review Board
- iTRAQ: isobaric tag for relative and absolute quantitation
- IP: IFN-gamma-induced protein 10
- JCRC: Joint Clinical Research Center
- LAM: lipoarabinomannan
- Mega: Molecular Evolutionary Genetic Analysis
- MTC: Mycobacterium tuberculosis complex
- MALDI-TOF: matrix-assisted laser desorption ionization-time of flight mass-spectrometry
- MDR: multi-drug resistant
- NAATs: nucleic acid amplification tests
- NADC: National Animal Disease Center
- OADC: Oleic Acid-albumin-Dextrose-Catalase enrichment
- OD: optical density
- OIE: World Organisation for Animal Health
- Pks5: polyketide synthetase 5
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPD</td>
<td>purified protein derivative</td>
</tr>
<tr>
<td>PZA</td>
<td>Pyrazinamide</td>
</tr>
<tr>
<td>RAST</td>
<td>Rapid Annotation using Subsystem Technology</td>
</tr>
<tr>
<td>RD</td>
<td>Regions of Difference analysis</td>
</tr>
<tr>
<td>sICAM</td>
<td>soluble intercellular adhesion molecule</td>
</tr>
<tr>
<td>S/N</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>TB</td>
<td>tuberculosis</td>
</tr>
<tr>
<td>TPPs</td>
<td>target product profiles</td>
</tr>
<tr>
<td>TNF</td>
<td>tumor necrosis factor</td>
</tr>
<tr>
<td>UNCST</td>
<td>Uganda National Council for Science and Technology</td>
</tr>
<tr>
<td>WGS</td>
<td>Whole Genome Sequencing</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>XDR</td>
<td>extremely drug resistant (XDR)</td>
</tr>
</tbody>
</table>

**Footnotes**

- a. Sigma-Aldrich, St Louis, MO.
- b. Nunc-Maxisorp, Sigma, St Louis, MO.
- c. Blocker Blotto, Pierce, Rockford, IL.
- d. PBS Tween, ThermoFisher, Carlsbad, CA.
- e. G-BioSciences, St Louis, MO.
- f. Santa Cruz Biotechnology Inc, Dallas, TX.
- g. 1-Step Ultra tetramethylbenzidine, Pierce, Rockford, IL.
- h. SpectraMax M2, Molecular Devices, Sunnyvale, CA.
- i. MS Excel, Microsoft Corp, Redmond, WA.
- j. GraphPad Prism 6, GraphPad Software Inc, La Jolla, CA.
INTRODUCTION

Mycobacterial diseases continue to cause suffering among people and massive economic losses to farmers. Testing for mycobacterial diseases is costly and time-consuming. Zoonotic tuberculosis (TB) is a major concern in low and middle-income countries with significant implications for human and animal health. Pastoral settings of the developing world where people have intimate contact with their animals, pose a significant risk for bi-directional transmission of Mycobacterial tuberculosis complex (MTC) infections i.e. *M. bovis* infection to humans and *M. tuberculosis* infection to animals. Treatment of tuberculosis is lengthy, expensive and the outcome is not always favorable. Co-infection with HIV and the increasing worldwide trends of multi-drug resistant TB further confounds the problem leading to grave animal and public health implications. The degree of interspecies MTC transmission at the human-animal interface presents unique challenges for healthcare and veterinary workers.

In this thesis, mycobacterial zoonotic infections among animals and humans was investigated. Novel non-invasive TB-specific serum biomarkers were applied to provide quick, inexpensive and accurate diagnostics for TB. These biomarkers had previously been established in experimentally and naturally infected deer, cattle and primate sera. Micro and small RNA from experimentally infected deer sera were used to develop a circulating nucleic acid based detection system to identify *M. bovis* infection in animals. Genomic tools were also applied to determine the relatedness of mycobacterial species at the human-animal interface to help trace the source of infection and the most common recent ancestor of the MTC strains.

The overarching goal of this thesis was to test the hypothesis that bi-directional
*Mycobacterium tuberculosis* complex transmission occurs between animals and humans. In resource limited settings this transmission is multi-directional due to the intimate and frequent human animal contact.

This hypothesis was tested under the following specific aims:

1. Evaluation of pathogen-specific biomarkers for the diagnosis of tuberculosis in white-tailed deer (*Odocoileus virginianus*)

2. Biomarkers in the Diagnosis of *Mycobacterium tuberculosis* complex Infections

3. MicroRNA as biomarkers in the diagnosis of *Mycobacterium tuberculosis* complex infections

4. Comparative genomics of archived pyrazinamide resistant *Mycobacterium tuberculosis* complex isolates from Uganda

**Specific Aim 1:** Evaluation of pathogen-specific biomarkers for the diagnosis of tuberculosis in white-tailed deer (*Odocoileus virginianus*). This work has been published in the June 2017 edition of the American Journal of Veterinary Research (AJVR). This tested the hypothesis that *Mycobacterium bovis* specific nucleic acids, peptides, and small RNA are detectable in circulation in sub-clinical infection. The research question sought to be answered was: *Are Mycobacterium tuberculosis complex-specific peptides detectable in circulation to enable detection of subclinical tuberculosis infection?* The overall objective for this specific aim was to develop a biomarker-based detection system specific for *Mycobacterium bovis* infection in animals and humans.

**Specific Aim 2:** MicroRNA as biomarkers in the diagnosis of *Mycobacterium tuberculosis* complex infections. This was addressed in two ways; one was by carrying
out a comprehensive review of all the current data on biomarkers which has been made into a book chapter to the Centre for Agriculture and Biosciences International (CABI) called **Biomarkers in the Diagnosis of Mycobacterium tuberculosis complex Infections**.

The second way this specific aim was addressed was to extract miRNA from experimentally infected deer and use them to design a circulating miRNA detection system. The results of this work have been put together in a short communication to be submitted to the Journal of Clinical Microbiology entitled **MicroRNA as biomarkers in the diagnosis of Mycobacterium tuberculosis complex infections**.

The hypothesis for this specific aim was that pathogen-host interactions would lead to accumulation of pathogen and host specific small molecules in the serum that can be applied to accurately diagnose bovine tuberculosis in animal populations.

**Specific Aim 3:** Comparative genomics of archived pyrazinamide resistant *Mycobacterium tuberculosis* complex isolates from Uganda. The manuscript for this work will be submitted to the Journal of Clinical Microbiology. The following objectives were addressed:

**Objective 1:** Identify the population genetic structure of *M. bovis* through comparative genomics

**Objective 2:** Define a Single Nucleotide Polymorphism (SNP) set to differentiate *Mycobacterium tuberculosis* complex group of organisms

**RESEARCH IMPACT**

New methods of testing for tuberculosis are direly needed for mycobacterial disease control and elimination. Biomarker based diagnostics offer a specific approach to identify
Mycobacterium tuberculosis complex (MTC) infection. This work has demonstrated unique means in which mycobacterial infections can be detected early, quickly and inexpensively by use of a simple diagnostic test. This will help in tracking, preventing and managing mycobacterial infections with the use of biomarker technology. Comparative genomics of MTCs isolated from humans and animals at the complex interface in disease endemic parts of the world will facilitate timely application of health interventions via a better understanding of transmission mechanisms. Mycobacterial infections, particularly those caused by Mycobacterium bovis, require more research and this work will help fill some of the knowledge gaps around this fastidious organism and perhaps provide new ways in which it can be managed both in the low and high disease prevalence regions of the world.

SIGNIFICANCE:

Zoonotic or bovine tuberculosis (bTB) is a major concern in low and middle-income countries with significant implications for human and animal health (World Health, 2015). One third of the world’s population is infected with tuberculosis (TB); in 2013 alone there were 9 million new infections and 1.5 million deaths were attributed to this scourge (World Health, 2015). Outbreaks in low-disease prevalence regions like the US lead to significant economic losses due to herd depopulations and export restrictions ((APHIS), 2017; (MDARD), 2017). The extent of Mycobacterium bovis infection in humans or M. tuberculosis in animals, in pastoral settings of the world is yet to be fully determined.

M. bovis and M. tuberculosis belong to the MTC complex of organisms; grouped due to their high degree of 16s rDNA sequence similarity. Both organisms cause TB with indistinguishable pathology in animals and humans. Members of this complex affect a variety of hosts with M. bovis having the widest host range. Although humans are
primarily infected by *M. tuberculosis*, infection with *M. bovis* goes undetected when routine laboratory tests are applied in most TB testing centers in developing countries ((OIE), 2017; Alemayehu, Girmay, & Gobena, 2008; Ameni et al., 2006; Ameni & Erkihun, 2007; Daborn, Grange, & Kazwala, 1996). Distinguishing between lesions caused by *M. bovis* and *M. tuberculosis* in humans is almost impossible. In animals, the location of the lesion depends on the route of infection and the subsequent dissemination to other organs (Michel, Muller, & van Helden, 2010; Scorpio & Zhang, 1996; Wedlock, Skinner, de Lisle, & Buddle, 2002). Humans are commonly infected by *M. bovis* through the consumption of raw or contaminated milk, and other dairy products as well as undercooked meat (Ameni & Erkihun, 2007; Ameni et al., 2013; Daborn et al., 1996). Infection by aerosols or person-person transmission also occurs but is not the main route of infection (Cosivi et al., 1998b; Michel et al., 2010).

In more industrialized countries, the systematic and sustained test-and-slaughter methods over decades ensured that disease prevalence reduced to such low levels that the disease is no longer considered a public health problem ((OIE), 2017; Michel et al., 2010). However, the cost incurred by a single bovine TB outbreak in such low disease prevalence regions, is expected to be in millions of dollars due to loss of markets and livestock productivity ((APHIS), 2017; Michel et al., 2010). In resource-limited settings, bTB has been termed as a ‘neglected zoonosis’ by WHO, FAO and OIE ((OIE), 2017; World Health, 2006, 2015, 2016b) but it is not reportable in most sub-Saharan countries and this makes its control and epidemiology quite challenging ((OIE), 2017; Cosivi et al., 1998b; Oloya et al., 2008). The application of test-and-slaughter methods to developing countries is simply not feasible yet bTB remains a real threat to human and economic
livelihoods (Cosivi et al., 1998b; Michel et al., 2010; Oloya et al., 2008). In addition, the intricately intimate relationship of livestock owners and their animals—which expose them to the zoonotic spread of tuberculosis - weave a difficult and daunting task to public health and veterinary care professionals. The extent to which bTB contributes to human tuberculosis is largely unknown in these settings and yet the majority of livestock keepers have intimate interactions with their livestock, which exposes them to zoonotic tuberculosis (Biffa et al., 2011; Kankya et al., 2010; Oloya, Kazwala, et al., 2007; Oloya, Muma, et al., 2007). Zoonotic diseases in these vulnerable pastoral communities, in addition to impacting human health, also threaten their livelihoods by compromising sustainable food supply, income and social status (Kankya et al., 2010; Muwonge et al., 2012; Oloya et al., 2008).

Mycobacterial disease diagnostic methods (e.g. Caudal Fold Test (CFT), gamma interferon release assay (IGRA), comparative cervical test (CCT) rely on a single cut-off value to indicate infectious status but diagnostic requirements differ between regions of low and high disease prevalence for bTB. The application of TB-specific biomarkers as diagnostic tests would greatly enhance screening for this traditionally difficult mycobacterial disease. Application of culture independent genomic epidemiology tools will provide unprecedented resolution, specificity and sensitivity to detect and differentiate infecting mycobacterial species as well as identify mixed infection in humans and animals. This work validated Mycobacterium tuberculosis complex biomarkers that will be used in a point-of-care device for the rapid screening of tuberculosis, something that has not been possible until now. The highly specific serum
biomarkers composed of mycobacterial peptides and proteins together with genomics will provide great insights into human and cattle MTC infections.
CHAPTER ONE

Biomarkers in the Diagnosis of Mycobacterium tuberculosis complex Infections

Sylvia I. Wanzala and Srinand Sreevatsan

Centre for Agriculture and Biosciences International (CABI) Book Chapter

1.1 Introduction:

Bovine tuberculosis (bovine TB) is a zoonotic infection in cattle caused by the intracellular bacterium, Mycobacterium bovis (M. bovis) that belongs to the Mycobacterium tuberculosis complex (MTB complex), a group of related mycobacteria that cause TB in mammals. Bovine TB is the most prevalent infectious disease of dairy cattle worldwide (Cosivi et al., 1998a) causing a conservative annual loss of about three billion dollars (M. V. Palmer, Waters, & Thacker, 2007). In the United States, the eradication program of bovine TB uses a test and slaughter strategy that cost about $38 million between 1917 and 1992; the current programs cost approximately $3.5-4 million annually (Charles O. Theon, 2006).

Cattle of all ages are susceptible to infection with M. bovis, however older animals appear to have greater susceptibility (Mackay & Hein, 1989; Munroe, Dohoo, & McNab, 2000; Thoen, 1995). In most cases, M. bovis infection primarily leads to a subclinical disease (95%) with rapid onset in only 5% of the exposed animals. Thus detection of subclinical infected animals with progressing granulomatous infection is critical in the control and eradication of bovine TB. Current USDA surveillance for bovine TB is a laborious multistep procedure involving the caudal fold test (CFT) and the comparative
cervical test (CCT) or γ-interferon release assays. The current diagnostics are problematic: CFT lacks specificity for *M. bovis* and fails to detect all diseased cattle, while the γ-interferon assay is costly and requires blood samples to be processed within 24 hours of collection. Moreover, early detection of subclinical infection by serological tests is hindered, since the humoral immune response in bovine TB occurs at a late stage of disease progression. Early diagnosis of bovine TB is essential to prevent substantial losses of valuable resources, monetary and production losses as well as to minimize the risk of human infection. Thus, there is a need to develop a low cost-effective, early detection assay for bovine TB.

Host macrophages are the main site for *M. bovis* infection in cattle. Infection is mainly transmitted via aerosols that are inhaled into the respiratory tract, and gross lesions involve granuloma formation in lungs and thoracic lymph nodes (Thoen, 1995). The biology of the granuloma involves intense cellular and biological activity at the site of infection leading to “leakage” of RNA, DNA and proteins into circulation that may serve as biomarker(s) for early detection of bovine TB. Recent advances in genomics and proteomics have opened new robust means for biomarker discovery. Discovery of novel biomarkers is essential for developing new diagnostic tests to aid in identification of infected animals in disease surveillance for bovine TB.

1.1.1 Biomarkers of tuberculosis in animals.

Bovine TB presents unique challenges in TB diagnostics and this is further exacerbated by the presence of wildlife reservoirs. Major reservoirs for bovine TB include White-tailed deer (*Odocoileus virginianus*) in the USA, the European badger (*Meles meles*) in the UK
and Ireland, brush tail possum (*Trichosurus vulpecula*) in New Zealand, Cape buffalo (*Syncerus caffer*) and greater kudu (*Tragelaphus strepsiceros*) in southern Africa, elk (*Cervus canadensis*) and American bison (*Bison bison*) in Canada and wild boar (*Sus scrofa*) in Spain (Miller & Sweeney, 2013; M. V. Palmer et al., 2000; M. V. Palmer, Whipple, & Waters, 2001a; Talip, Sleator, Lowery, Dooley, & Snelling, 2013). *Mycobacterium bovis*, with the largest host range among the MTB complex organisms, also causes zoonotic TB in humans. In the USA, TB testing in wildlife is carried out with the *in vivo* tuberculin skin test together with the *in vitro* interferon γ assay (O'Brien et al., 2009; M. V. Palmer, Waters, Whipple, Slaughter, & Jones, 2004; M. V. Palmer et al., 2000; M. V. Palmer et al., 2001a). These strategies identify bacteria in lesions or detect host immune responses but they suffer from low sensitivity, are labor intensive, costly and not always readily available in crucial locations. In cattle, bovine TB is a major welfare and economic challenge. Bovine TB reduces productivity in affected animals, with the identification of infected animals leading to movement controls, testing of herds, culling of affected animals, and trade restrictions (Humblet, Boschiroli, & Saegerman, 2009; Rodriguez-Campos, Smith, Boniotti, & Aranaz, 2014). In countries that practice active bovine TB surveillance, the three main tests used are the caudal fold test (CFT), comparative cervical test (CCT), and a gamma interferon release assay (IGRA). The primary test for screening for bovine TB is the century old tuberculin skin test or CFT, whereby bovine purified protein derivative (PPD), prepared from a culture of *M. bovis*, induces a delayed-type hypersensitivity reaction when injected intra-dermally, resulting in skin swelling, after 72 hours. These tests are all labor-intensive, present logistical problems and have challenges with sensitivity and specificity (Lamont, Janagama, et al., 2014a). The CFT is not very specific for *M. bovis* infection and doesn't detect all
diseased cattle; co-infection with *Mycobacterium avium paratuberculosis* further confounds the results. The IGRA test is based on release of a cytokine IFN-\(\gamma\), when sensitized lymphocytes are re-exposed *in vitro* to *M. bovis* antigens (H. M. Vordermeier, de Val, et al., 2014; H. M. Vordermeier, Jones, Buddle, Hewinson, & Villarreal-Ramos, 2016a). IGRA requires a very quick turnaround for sample processing which is not always possible when working with large herds in a remote location.

Although the tuberculin test is the most common for diagnosis of bovine TB, it has several limitations. The PPD used in the test contains more than 200 antigens that are shared between pathogenic mycobacterial species and other 'atypical' mycobacteria (Chaparas, Maloney, & Hedrick, 1970). In the United States the estimated sensitivity of CFT and CCT are 80.4 % to 88.4 % (CFT) and 75% (CCT) respectively while specificity are 96% (CFT) and 98% (CCT) respectively (Whipple et al., 1995). These tests require accredited veterinarians for testing and the animal needs to be restrained at least twice for each test. Serological tests are not applicable in bovine TB surveillance program as the antibody titers rise very late in the *M. bovis* infection.

Thus, it is quite evident that the diagnosis of bovine TB can be extremely difficult. If missed, the consequences could be disastrous including substantial loss of valuable resources, time (lengthy quarantine period of the animals for the diagnosis of disease in the area of bovine TB outbreak), money, emotional expense and trauma (slaughtering of 100-1000 animals for the identification of a single infected animal in the disease surveillance area) associated with the cattle owners, as well as the risk of human infection. Thus, the development of serological based tests that are more sensitive and specific using novel approaches could be useful in slaughterhouse surveillance programs.
There are several biomarkers related to bovine TB pathology and vaccine efficacy. For example, the *ex vivo* ESAT-6 induced production of IFN-γ from blood is correlated with the degree of pathology following experimental infection of cattle with *M. bovis*. BCG-vaccinated calves had lower or reduced responses as well as reduced gross pathology (H. M. Vordermeier, Jones, Buddle, & Hewinson, 2016). Another marker investigated as a potential predictor of vaccine-induced protection and memory is IL-2. IL-2 production is also a potential biomarker for latency and different disease stages in cattle (O'Brien et al., 2009; M. V. Palmer et al., 2004; M. V. Palmer et al., 2000; M. V. Palmer et al., 2001a; H. M. Vordermeier, Jones, et al., 2016a).

Mycobacteria have evolved an array of sophisticated mechanisms for evading the host immune system and defending itself in the harsh intracellular environment of the macrophage, some of which are well understood while others are yet to be discovered. In the “dance of seduction” with *M. bovis*, the host has also evolved several responses to contain infection, and together with the pathogen mechanisms will make up the gist of our discussion.

1.2 Biomarkers defined.

A biomarker can be defined as a characteristic that can be measured and evaluated as an indicator of a normal biological process, a pathogenic process or pharmacologic responses to a therapeutic intervention (Biomarkers Definitions Working, 2001). Biomarkers can be used as diagnostic tools for the identification of patients with disease (e.g., elevated glucose concentration for diagnosis of *diabetes mellitus*), for disease staging (e.g., prostate specific antigen concentration in blood), as an indicator of disease prognosis (e.g., anatomic measurement of tumor shrinkage of certain cancers) or for prediction and monitoring of clinical response to an intervention (e.g., response to
anti-tuberculosis drug treatment, vaccine efficacy, blood cholesterol concentration for
determination of heart disease etc.). Currently, the technologies used in biomarker
discovery include in vitro analyses of DNA variation (disease susceptibility), circulating
DNA or RNA (disease progression-apoptosis/proliferative pathways), transcriptomes
(disease induced transcriptional alterations), and proteomics (disease progression).

Discovery of novel biomarkers is essential for developing new diagnostic tests to
aid in identification of infected animals in disease surveillance for bovine TB, without
necessarily slaughtering herds of 100s to 1000s of animals for identification of the single
infected animal. Early diagnosis is essential in bovine TB because clinical symptoms
appear very late, when the disease has significantly advanced and the risk of infection
being transmitted is high.

### 1.2.1 Characteristics of an Ideal Biomarker.

For a biomarker to qualify as an ideal diagnostic tool it has to meet a minimum set of
criteria or target product profiles (TPPs) (Gardiner, Karp 2015). The ultimate diagnostic
test for MTB complex infections would be one that is highly sensitive, specific and non-
sputum based with a clear predictive response to therapy, independent of the host
response (Gardiner, Karp 2015). An ideal biomarker-based assay should also have very
high sensitivity and specificity (>98%), be non-invasive or minimally invasive, provide
results rapidly, without the need for the cold chain, and be affordable. Hass and
colleagues (Haas, Roe et al. 2016) argue that having a combination of biomarkers
would enhance the diagnostic value in different settings: for instance having one set of
biomarkers for differentiating between active and latent TB (humans)/ subclinical TB
(animals) and another set to diagnose TB in comparison with other diseases. Such
efforts are possible through collaboration between human and bovine tuberculosis researchers. The pathway taken for a prospective biomarker to reach the market is a tortuous one and potential biomarkers must progress through sequential testing to confirm their efficacy in an independent cohort, then they are further validated in a prospective study. The biomarkers are also required to undergo tests on a platform appropriate for their proposed use; such platforms include TB clinics in endemic countries (Gardiner, Karp 2015).

Existing TB diagnostics are inadequate. Infection with MTB complex organisms presents unique challenges for diagnostic testing. Current diagnostics have several shortcomings; potential new diagnostics are explored below.

1.3 Circulating biomarkers.

The analysis of alterations in circulating protein profiles or circulating DNA or RNA (including microRNA and small RNAs) in plasma are promising diagnostic tools, requiring only a limited blood sample. Recently completed human and animal genomes have brought refinements in technology including nucleotide and protein sequencing, mass spectrometry, and microarrays (nucleic acid and protein arrays), and these have allowed researchers to elucidate fundamental biological processes of chronic diseases such as cancers, neurological disorders, cardiovascular disease and several infectious diseases (Jacobsen, Mattow, Repsilber, & Kaufmann, 2008; Maruvada, Wang, Wagner, & Srivastava, 2005; Scaros & Fisler, 2005).

1.3.1 Circulating nucleic acid approach
Circulating nucleic acids (CNAs) are segments of DNA and RNA that are devoid of cellular material and are detected in biological fluids. They are present in small amounts in the plasma of healthy individuals. However, their increased levels are associated with disease conditions. Mandel and Metais first reported CNA in 1948 in human plasma (Anker, 2000; Mandel P, 1947). Later studies on CNA were mainly focused on autoimmune diseases like *lupus erythematosus* (Tan, Schur, Carr, & Kunkel, 1966) and rheumatoid arthritis (Ayala, Moore, & Hess, 1951). Thirty years later, the diagnostic implications of CNA were recognized by Leon *et al.* in 1977, when he reported high levels of CNAs in patients with pancreatic cancer and demonstrated that levels of plasma CNA decreased after chemotherapy (Leon, Shapiro, Sklaroff, & Yaros, 1977). Since then elevated levels of CNA has been reported in chronic illness (Lui *et al.*, 2002; Schutz *et al.*, 2005), trauma (Lo, Rainer, Chan, Hjelm, & Cocks, 2000), acute stroke (Rainer *et al.*, 2003), myocardial infarction (Chang *et al.*, 2003), prenatal diagnosis (Chim *et al.*, 2005; Lo *et al.*, 1997), and various cancerous diseases (Capone *et al.*, 2000; Hibi *et al.*, 1998; Shao, Wu, Shen, & Nguyen, 2002; Sorenson *et al.*, 1994; Vasioukhin *et al.*, 1994).

In the last decade, CNA has gained more attention because of its potential application as a non-invasive, rapid and sensitive tool for molecular diagnosis and monitoring of acute pathologies. Most CNA based laboratory diagnosis involves amplification of either RNA or DNA with primers designed for single-copy coding regions. These CNA tests primarily detect the functional genes associated with exogenous nucleic acid (Lui *et al.*, 2002) for example those belonging to West Nile virus, parvovirus B-19, human immunodeficiency virus, hepatitis B virus, hepatitis A virus, etc. In addition to detecting single-copy exogenous nucleic acids, CNA diagnostics are being developed that detect endogenous, repetitive sequences (Stroun, Lyautey, Lederrey, Olson-Sand,
& Anker, 2001). Often chronic diseases that lead to cell stress and the release of nucleic acids into the blood show a consistent pattern of endogenous CNAs in serum.

Although most of the diagnostic CNA signatures are associated with human disease, a few researchers have detected CNA in cattle diseases (Brenig, Schutz, & Urvovitz, 2002; Schutz et al., 2005; Shaughnessy, Farrell, Riepema, Bakker, & Gordon, 2015). Researchers have studied repetitive sequences including short Alu repeat sequences (SINE-like sequence of primates) in bovine spongiform encephalopathy (BSE) where they identified the 3' region of Bov-tA fragments in PCR products derived from the serum of confirmed BSE cases or BSE exposed cohorts (Schutz et al., 2005). These repetitive elements identified in BSE and other infectious disease lead us to believe that a similar pattern may exist in a chronic infection such as bovine TB.

1.3.1.1 Mechanism for release of CNA into circulation.

Various hypotheses have been proposed as to the mechanism of release of CNA in biological fluids. However, there are controversies related to these hypotheses and the actual origin of CNA still remains ambiguous. On the one hand, necrosis and apoptosis has been considered as major pathway for CNA release (Lichtenstein, Melkonyan, Tomei, & Umansky, 2001; Lo et al., 2000); in contrast it has been reported that cellular necrosis may not be an important pathway as plasma DNA levels fall rather than rise following radiation therapy (Tan et al., 1966). Researchers have considered cellular apoptosis as a source of plasma CNA based on the fact that the electrophoretic pattern produced by plasma CNA is similar to that found with DNA extracted from apoptotic cells (Kamm & Smith, 1975). Apoptosis-induced increased CNA level has been
demonstrated in plasma of patients with lung cancer (Fournie et al., 1995). Several in vivo experiments in mice have demonstrated increased CNA levels in blood via apoptotic or necrotic pathways (Fournie et al., 1995; Jiang & Pisetsky, 2005; Jiang, Reich, Monestier, & Pisetsky, 2003). Increased CNA level has been documented in plasma as a result of apoptosis induced oxidative stress on placental tissue (Tjoa, Cindrova-Davies, Spasic-Boskovic, Bianchi, & Burton, 2006). Evidence for the active release of CNA from activated lymphocytes or other nucleated cells (Anker, Stroun, & Maurice, 1975; Stroun, Lyautey, Lederrey, Olson-Sand, et al., 2001) and lysis of tumor cells has also been reported (Sorenson, 2000). Apoptosis and necrosis are also associated with TB pathogenesis and are critical for mycobacterial killing, granuloma formation, and chronic inflammatory condition induced by the pathogen (13). Thus an increase in CNA of *M. bovis* infected animals is expected to occur as the disease progresses.

1.3.1.2 Methodologies used for CNA discovery.

Although there has been a controversy about the use of serum or plasma for CNA discoveries, most of the CNA discoveries have been applied to serum (Kopreski et al., 1997; Leon et al., 1977; Lo et al., 1997; Nawroz, Koch, Anker, Stroun, & Sidransky, 1996; Sorenson et al., 1994). It has been reported that CNA recovered from serum is several fold higher than that in plasma. The difference in CNA levels have been considered due to the in vitro lysis of white blood cells during the process of clotting (Chen et al., 1999; Lui et al., 2002). Lui et al.’s study concluded that the serum CNA might not be a true representation of the biological condition of the patient.
Several groups have measured the levels of CNA in different diseases in search of diagnostic or prognostic markers (Ziegler, Zangemeister-Wittke, & Stahel, 2002). Multiple techniques have been used in different studies for quantitative analysis of CNA post discovery using de novo sequencing of total circulating nucleic acids. These include radioimmunoassay (Leon et al., 1977; Shapiro, Chakrabarty, Cohn, & Leon, 1983), competitive PCR (Jahr et al., 2001), quantitative real time PCR (Thijssen, Swinkels, Ruers, & de Kok, 2002), fluorimetric quantification (Thijssen et al., 2002), spectrophotometric determination (Shao, Wu, Shen, & Nguyen, 2001) and visual comparison with known standards (Sozzi et al., 2001). To date, all the studies into CNA demonstrated a significant increase of CNA levels in the diseased condition irrespective of the use of serum or plasma.

1.3.2 Proteomics approach

The term proteome is derived from “protein and genome” and was first coined by Marc Wilkin in 1995. Proteome refers to all the proteins expressed by genome at a given time within a given environment (Solassol et al., 2006). With the completion of the genome sequences for many prokaryotic and eukaryotic organisms, researchers had to assign cellular and molecular functions to thousands of newly predicted gene products and explain how these products cooperate in complex physiological processes. This led to the emergence new field of research termed “proteomics” that aims to characterize biological mechanisms by identifying different proteins involved.

In the last decade, proteomics has provided us an ability to rapidly identify novel protein biomarkers for various cancerous and non-cancerous diseases. Several researchers have reported that not a single biomarker but a battery of biomarkers is
required to show the specificity and sensitivity for the detection or monitoring of most cancerous diseases (E. F. Petricoin, 3rd et al., 2002; E. F. Petricoin et al., 2002; Stone et al., 2005; Tirumalai et al., 2003; Zhang et al., 2004). While substantial research on biomarker discovery exists in fields such as oncology, very few studies have investigated the utility in using the proteomic approach to understanding the pathogenesis of infectious diseases (Agranoff et al., 2006; Gravett et al., 2004; Pang et al., 2006; Poon et al., 2004; Yip et al., 2005).

Nonetheless several studies have looked at the diagnostic potential of proteomic fingerprinting to determine different disease states as well as monitor the treatment response in TB (Haas, Roe, Pollara, Mehta, & Noursadeghi, 2016b). Early proteomic research showed that a combination of four biomarkers (serum amyloid A, transthyretin, neopterin and C reactive protein) could distinguish between active pulmonary TB and non-TB disease and healthy controls (Agranoff et al., 2006; Seth et al., 2009). It was speculated that targeting specific protein variants rather than the total protein would improve the accuracy of the diagnosis. The translation of proteomic biomarkers into diagnostic tests has been faced with some challenges: the protein biomarker candidates reported by independent studies vary considerably and a universal proteomic profile of TB is yet to be agreed upon. In addition, varying results may be due to differences in proteomic techniques and their resolutions, study design, case definitions and statistical analyses (Haas et al., 2016b). There are overlaps though, of serum proteins that are differentially expressed in active TB like CD14, S100A proteins, apolipoproteins, fibrinogen, orosomucoid and serum amyloid A. The challenge with proteomics research is that different investigators use different selection criteria when assessing common protein signatures and the identified proteins are not always evaluated for their
diagnostic potential (i.e. with receiver operator curve analyses or decision trees); signatures may not be cross-validated in independent datasets; or evaluated with external datasets (Scaros & Fisler, 2005).

A novel approach involves detecting circulating mycobacterial peptides and/or lipids or metabolites in the serum or plasma of infected animals. Research in the Sreevatsan laboratory has recently identified sixteen *M. bovis* proteins including, MB2515c (transcriptional regulator (LuxR family), MB1895c (cell wall biosynthesis) and MB1554c or Pks5 (polyketide synthetase 5) in bovine TB positive and exposed cattle and deer (Lamont, Janagama, et al., 2014a; Wanzala et al., 2016). These proteins were first identified by gel-free multi-dimensional isobaric tag for relative and absolute quantitation (iTRAQ) proteomics and subsequently validated using a well-characterized cattle serum repository (Lamont, Janagama, et al., 2014a; Seth et al., 2009). An indirect ELISA using monoclonal antibodies synthesized against these peptides was developed to detect these biomarkers in serum and have been validated in bovine and primate TB (Sreevatsan, Kaushal and Lamont, unpublished data). Given that the current bovine TB diagnostics have a ‘one-size-fits-all’ testing method whereby disease prevalence status for a given region is not considered; these pathogen specific biomarkers (Pks5, MB2515c and MB1895c) are unique in that they take the disease prevalence status into account and also detect TB.

**1.3.2.1 Mechanisms of release of protein biomarkers into circulation.**

Peptide biomarkers are the low molecular, less abundant circulating proteome termed as “peptidome” (Lai, 2015). This peptidome may consist of many types of diagnostic information that may constitute the parent protein, the peptide fragment, the
quantity of peptide or the nature of carrier protein to which it is bound (Petricoin, Belluco, Araujo, & Liotta, 2006). According to the peptidome hypothesis, many proteins and peptides are shed into the local circulation from the disease microenvironment. Apoptosis and necrosis of cells are considered as the main causes for release of proteins and peptides into circulation from the disease microenvironment. Mycobacterial lysis or its release into circulation (mycobacteremia as has been recently proposed using a phage based diagnostic test, (Swift, Convery, & Rees, 2016) can also lead to release of bacterial products into circulation. As a consequence, the blood peptidome could contain ongoing recordings of the molecular cascade of communication that takes place in the tissue microenvironment (Petricoin et al., 2006).

Researchers have explored all different kinds of biological matrices from cell cultures (lysates, supernatants) to clinical samples (serum, plasma, cerebrospinal fluid, bronchoaveolar lavage and urine) for protein biomarker discovery. Of these, serum has many attributes that make it preferable over others for biomarker discovery. Serum is readily available and has dynamic range of proteins. Serum continuously perfuses through the tissues, and thus contains proteins and peptides secreted/released from cells and tissues from the disease microenvironment. However, there are many challenges associated with using serum for biomarker discovery. Candidate biomarkers are expected to exist in a very low concentration and are generally carried with high abundant blood proteins like albumin, which exist in a billion-fold excess. Moreover, serum constitutes 65-97% of high abundance proteins like albumin and immunoglobulins that mask the biologically significant variations among low abundant serum proteins and prevents their detection and identification in proteomic studies (Govorukhina et al., 2003). Thus there is a need for the depletion of these high abundant proteins to enrich
low abundant biomarkers for the proteomic analysis. If not stored properly, serum protein may be depleted due to repeated freeze-thaw cycles.

1.4 Circulating miRNA.

Biomarker research has also pointed to circulating microRNA as a potential prognostic and diagnostic biomarker (Farrell et al., 2015). MicroRNA are short (~22 nt), single-stranded, non-coding RNAs that regulate mRNA expression. Micro-RNAs are important regulators of gene expression and play a key role in regulating both the innate and adaptive immune responses. Recent work (Golby, Villarreal-Ramos, Dean, Jones, & Vordermeier, 2014) has shown that expression of miR-155 was more than 40 times higher in naturally infected cattle with visible pathology compared with infected animals that presented without visible pathology. This suggests that miR-155 could distinguish active from latent infection, and could be used as a diagnostic and prognostic biomarker to identify infected animals as well as be used for DIVA testing although more research is still needed in this area (Abd-El-Fattah, Sadik, Shaker, & Aboulftouh, 2013; Golby et al., 2014). What makes miRNAs particularly appealing as potential biomarkers is that they have tissue specific expression patterns that can serve as a fingerprint for disease and also can be detected by RT-PCR assays and microarray techniques (Williams et al., 2013). In addition, it’s hypothesized that different stages of mycobacterial infection have distinct miRNA signatures (Farrell et al., 2015). Another fascinating fact about them is that serum miRNA is stable to repeated freeze-thaw cycles as well as to heat, acidic and alkaline conditions.

1.5 Serum cytokines.

Alveolar macrophages and pulmonary dendritic cells represent the first line of defence
against invading mycobacterial pathogens (Kaufmann 2004). This process releases chemokines, which attract monocytes and other inflammatory cells to the lungs (Kleinnijenhuis, Oosting et al. 2011). Chemokines are types of cytokines—a group of mainly soluble proteins and glycoproteins that modulate the immune system. Examples include interleukins (ILs), interferons (IFNs), growth factors, colony stimulating factors (CSFs), the tumor necrosis factor (TNF) family, and chemokines (Choi et al., 2016).

Phagocytic cells trigger the adaptive immune response by presentation of mycobacterial antigens to T cells. Once a macrophage is infected with mycobacteria, it releases interleukins 12 and 18 (IL-12 and IL-18). The released cytokines stimulate CD4, CD8 and natural killer cells to produce interferon gamma (IFN-\(\gamma\)) and tumour necrosis factor alpha (Villarreal-Ramos et al., 2003). T cells respond to the released IFN-\(\gamma\) in a positive feedback loop leading to the production of more IFN-\(\gamma\). The IFN-\(\gamma\) activates macrophages to kill the invading mycobacteria by activating nitric oxide synthase, which produces nitric oxide while the TNF-alpha is critical for the initiation of the immune response against infection with mycobacteria (Das, Thomas, Garnica, & Dhandayuthapani, 2016; Kaufmann, 2004). In a recent study (Thacker, Palmer, & Waters, 2007), expression of IFN-\(\gamma\), TNF-\(\alpha\), iNOS and IL-4 by peripheral blood mononuclear cells (PBMC) was increased in response to infection, whereas, IL-10 expression decreased. There was also a positive association between Th1 responses and disease severity but as infection progressed, the differences in gene expression between the low and high pathology groups were indistinguishable implying a possible influence of early Th1 response on pathology (Thacker et al., 2007). Characterization of the bovine immune response has been done by several research teams; examples include use of real time PCR and bovine immune microarrays to characterize bovine
cytokine/ chemokine/ transcription factor etc. responses to bTB (Schiller et al., 2010). Other cellular immunity based tests include the development of bovine cytokine and chemokine multiplex systems detecting several parameters in a single sample as well as the use of monoclonal antibodies that recognize bovine cytokines (Coad et al., 2010; Schiller et al., 2010).

A granuloma is a compact, organized collection of mature macrophages, which arises in response to persistent stimuli (Ramakrishnan, 2012). Necrotic areas called caesium occur within granulomas as a result of dying cells. Cells that make up the granuloma include neutrophils, dendritic cells, B and T cells, natural killer cells, fibroblasts and cells that secrete extracellular matrix components (Harding & Boom, 2010). After formation of granulomas, several scenarios may occur; cessation of infection or dormancy; progression of infection with dissemination to other organs as well as reactivation which may occur months or years after the initial infection due to a compromised immune system (Kaufmann, 2004; Kleinnijenhuis, Oosting, Joosten, Netea, & Van Crevel, 2011; Ramakrishnan, 2012; van Crevel, Ottenhoff, & van der Meer, 2002). The various manifestations of infection with mycobacteria are a reflection of the delicate balance between the bacteria and the host-defense mechanisms (van Crevel et al., 2002). Recent research by (M.V.; Palmer, Thacker, & Waters, 2016) examined cytokine expression of TNF-α, IFN-γ, TGF-β, IL-17A and IL-10 in experimentally infected calves and found a moderate but positive correlation between the level of cytokine expression and cell size or number of nuclei in giant cells of granulomas. Their work demonstrated that these giant cells contribute to the “cytokine milieu necessary to form and maintain granulomas” (M.V.; Palmer et al., 2016).
Another cytokine is IL-17 produced by Th17 cells, has been demonstrated to play a role in tuberculosis immunopathology as well as other chronic illnesses. Antigen specific in vitro expression of IL-17A has been correlated to both increased disease severity and vaccine induced protection in cattle experimentally infected with *M. bovis* (M.V.; Palmer et al., 2016). In tuberculosis IL-17 cytokines play key roles in initiating both protective and harmful inflammatory responses and the use of Th17-associated cytokines have been suggested as possible biomarkers of infection and protection in the immune responses to bovine tuberculosis (Waters et al., 2015).

1.6. Cellular immune responses for potential biomarker application.

Mycobacteria are intracellular pathogens and the host mounts a successful response through a strong cell-mediated response by the adaptive immune system, which, in the case of Bovine TB, also acts as the immunological diagnosis of infection (Goosen et al., 2014; H.M.; Vordermeier, Cockle, Whelan, Rhodes, & Hewinson, 2000). Several cellular immune responses have been the target for biomarker application. BCG vaccination is one example though it’s given variable responses in both humans and cattle and studies to improve the protection conferred by the BCG vaccine are in full gear (H. M. Vordermeier, Jones, et al., 2016a). One such method is the heterologous prime-boost strategy which involves use of supplemental or booster vaccine where the immune system is primed with BCG after which it’s boosted with subunit vaccines containing protective antigens present in BCG (H. M. Vordermeier, Jones, Buddle, & Hewinson, 2016). Other methods include complete replacement of BCG with attenuated *M. bovis* strains leading to over-expression of antigens or the use of genetically modified BCG strains with improved immunogenicity (H. M. Vordermeier, Jones, Buddle, & Hewinson, 2016; Waters et al., 2009).
Identification of genes deleted in BCG when compared to *M. bovis* by comparative genomics identified key targets in cattle and humans like the *M. bovis* proteins early secretory antigenic target (ESAT-6) and the culture filtrate protein (CFP-10) located on the RD1 region of *M. bovis* and in virulent strains of *M. tuberculosis* (Mahairas, Sabo, Hickey, Singh, & Stover, 1996; H.M.; Vordermeier et al., 2000; H. M. Vordermeier, Jones, Buddle, & Hewinson, 2016). In terms of diagnostics, ESAT-6 and CFP-10 can differentiate between *M. tuberculosis* infected from BCG vaccinated humans and ESAT-6 could differentiate between *M. bovis* infected and BCG vaccinated cattle; these peptides through the Bovigam PC-EC assay (BEC) and the Bovigam PC-IHC (BHP) have been harnessed to improve the specificity of IGRAs though the sensitivity of BTB diagnosis has been sub-optimal (Goosen et al., 2014). A study by (Goosen et al., 2014) in African buffalo demonstrated that Monocyte derived chemokine IFN-gamma-induced protein 10 (IP-10) was a useful marker of immune activation by *M. bovis* antigens when using the bovine IP-10 ELISA and they recommended that the diagnostic potential of IP-10 for BTB in cattle be re-evaluated using species-specific reagents.

**1.7 Human tuberculosis.**

In humans, for MTB complex organisms to be detected from sputum, it implies that the airways are close to or have necrotic foci of infection (Gardiner, Karp 2015). Thus for diagnosis with sputum to be possible it means that most likely the patient has had active disease for quite some time, often with severe damage to the lungs (Gardiner, Karp 2015). The most common method used is the direct detection of the pathogen via microscopy, culture or by PCR where the DNA is amplified. The presence of a sustained T cell reactivity with *M. tuberculosis* complex antigens (tuberculin skin tests) or by use of
interferon release assays of peripheral blood is also used to determine infection (Wallis, Pai et al. 2010) The gold standard for TB testing, culture, is lengthy due to the fastidious nature of MTB complex organisms with very slow generation time (20-22 hours for *M. tuberculosis*). Identification of *M. tuberculosis* therefore takes weeks and this delay also pushes back treatment of ill persons who may be actively infecting others. Interferon gamma release assays (IGRA) are also used to detect infection but they only have moderate predictive value marginally higher than that of tuberculin skin test in low/middle income countries (Leung, Lange, & Zhang, 2013). IGRAs work by measuring the IFN-γ released by T-cells in a blood sample after re-stimulation with specific *Mycobacterium tuberculosis* complex (MTC) antigens. A positive outcome for IGRA gives an indication of infection but it cannot distinguish between active and latent TB.

Despite new automated molecular ways for TB detection and drug resistance, a simple, affordable point-of-care test for TB is still not available (Wallis et al., 2010). Low sensitivity is one of the biggest challenges facing microscopy and one may miss diagnosis in more than 30% of samples tested. Research shows that use of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry and nucleic acid amplification tests (NAATs) may soon accelerate this step for the identification of positive cultures (Wallis et al., 2010). The current molecular test (Xpert MTB/RIF) recommended by the World Health Organization (WHO) has a sensitivity of 99.7% and specificity of 98.5% in smear positive samples and is 76.1% sensitive and 98.8% specific in smear-negative samples (Boehme et al., 2010; Centers for Disease & Prevention; Cepheid). The Xpert® MTB/RIF test though is expensive and not widely available in resource limited settings where it is needed the most. In addition, it is not useful for testing extra-pulmonary manifestations of human and bovine TB (Gardiner &
Biomarkers are unique because they provide prognostic information about the future health status of an individual; they can indicate normal or pathological states, as well as responses to anti-tubercular drug therapy. In TB diagnostics biomarkers are required to detect active disease and latency as well as predict non-relapsing treatment success in humans. In addition, they would be useful in determining individuals that are protected from TB by new vaccines. Sputum based biomarkers play a limited role in latent TB.

A simple, non-invasive test using urine, saliva or serum that can serve as both a diagnostic and prognostic test would greatly enhance TB diagnostics. Biomarkers in urine like lipoarabinomannan (a 17.3-kDa immunogenic glycolipid component of the mycobacterial cell wall or LAM) have been tested with varying results as well as detection of volatile organic compounds in patients with pulmonary tuberculosis though much more research needs to done to establish changes in these biomarkers during treatment or clinical outcome (Boehme et al., 2010). There is currently a commercially available urine LAM test but use is limited due to sensitivity issues but is useful if used in combination with other current testing methods (Gardiner & Karp, 2015a; Lamont, Ribeiro-Lima, Waters, Thacker, & Sreevatsan, 2014; Leung et al., 2013).

Analyses to detect antibodies against antigen 85 have been carried out in blood and urine by mass spectrometry with promising results (Young et al., 2014b). But in general antibody analyses have not been very effective mainly due to the heterogeneity of the antibody response to *M. tuberculosis* and thus they haven’t been able to meet the requirements for a diagnostic test (Gardiner & Karp, 2015a). Data from several researchers have concluded that antibody responses are unlikely to provide useful
diagnostics for TB (Gardiner & Karp, 2015a).

There are biomarkers that are increased at baseline in proportion to the degree of the disease and subsequently decline with treatment: these include soluble intercellular adhesion molecule (sICAM), C-reactive protein, soluble urokinase plasminogen activator receptor and procalcitonin (Eckersall & Bell, 2010; Wallis et al., 2010). Assays with these biomarkers are simple, affordable and can be carried out on frozen plasma samples so they can be incorporated into treatment protocols. Studies indicate that they have greatest prognostic value when measured at or near the completion of therapy (Wallis et al., 2010). Use of a panel of biomarkers gives a better response than just one marker. In addition, measuring multiple parameters by proteomics, transcriptomics, and metabolomics would greatly enhance the efficacy of biomarkers.

1.8 Biomarker challenges.

Major challenges in TB diagnostics include the absence of effective tools for appropriate and accurate TB diagnosis as well as well-defined tools to monitor treatment response to enable shorter courses of chemotherapy for human patients and faster turnaround for farmers with a suspected TB case. Although there’s on-going research on biomarkers in humans, new tools for TB diagnosis in animals are not as abundant (H. M. Vordermeier, Jones, Buddle, & Hewinson, 2016). There have been major issues with reproducibility in biomarker research and this has been put down to inattentiveness to methodological issues of study design and performance, which have been repeatedly claimed as a major reason for false-positive findings in biomarker research (Pesch et al., 2014). Another significant challenge with biomarker research is the lack of a collaborative and systematic approach and this is the case for the development of biomarkers in both
human and veterinary medicine (Kondo, 2014). Technical challenges like low sensitivity, reproducibility and throughput have caused biomarker failures but this has increasingly been overcome by use of DNA microarray technology which enables the measurement of the mRNA levels of thousands of genes in a quantitative and reproducible manner at relatively low cost (Kondo, 2014).

Market failure is another important challenge facing the development of new TB diagnostics like biomarkers. In the case of human TB, industry usually avoids developing and marketing products that will be mainly used by patients in resource-limited countries because such products don’t generate profits (Wallis et al., 2010). In addition, even if the products are available, neither their cost nor performance, is adapted for developing countries meaning that their potential benefits are unavailable to the patients and healthcare providers who need them most (Wallis et al., 2010). For bovine TB, major strides have been made and the coming together of the bovine TB vaccine development programme with the international human TB vaccine programme has been beneficial and has led to development of platforms with DIVA (Differentiating Infected from Vaccinated Animals) capability to be applied in countries that plan to use vaccination together with test and slaughter (H. M. Vordermeier, Jones, et al., 2016a). The steps taken for a biomarker to be approved as a diagnostic test are rigorous before they can be approved for animal or human use (Pesch et al., 2014).

1.9 Conclusions

For global bovine TB control to become a reality, there is need for a more accurate, affordable point-of-care diagnostic test. Biomarkers are key in this process but it’s important for researchers to harness the advantages of having multiple biomarkers. The
bovine TB diagnostics pipeline has grown in recent years with many promising candidates. Biomarkers are among such candidates but there is need for improvements in standardization and validation procedures to increase reproducibility and accuracy and promote adoption of these biomarkers. Continuous improvement in our knowledge of these intriguing organisms is the best way of overcoming some of the knowledge gaps surrounding TB biomarkers. Knowledge about the biology of MTC organisms and their host-pathogen interactions is still incomplete thus research into these areas will greatly enhance the development of accurate, and safe biomarkers for bovine TB.
CHAPTER TWO

Evaluation of Pathogen-Specific Biomarkers for the Diagnosis of Tuberculosis in
White-Tailed Deer (Odocoileus virginianus)


Sylvia I. Wanzala BVM, MSc, MPH, Mitchell V. Palmer DVM, PhD, Wade R. Waters
DVM, PhD, Tyler C. Thacker PhD, Michelle Carstensen PhD, Dominic A. Travis DVM,
MS, Srinand Sreevatsan BVSc, MVSc, MPH, PhD.

2.0 Background

Bovine TB is a zoonotic infection of cattle caused by Mycobacterium bovis (Hingley-
Wilson, Sambandamurthy, & Jacobs, 2003). Bovine TB was first diagnosed in free-
ranging, white-tailed deer (Odocoileus virginianus) in Michigan in November 1975
(Schmitt et al., 1997). Since surveillance for bovine TB began in 1994, the extent and
characteristics of outbreaks of the disease in deer and details of local and national
management efforts have been extensively described (O'Brien et al., 2002; O'Brien,
Bovine TB is of concern because of its ability to infect a wide variety of species,(O'Reilly
& Daborn, 1995) including humans,(Wilkins et al., 2003; Wilkins et al., 2008) and
resulting costs of infection for the livestock industry because of herd condemnations and
loss of markets (Morris, Pfeiffer, & Jackson, 1994). After numerous years of surveillance
and research, white-tailed deer remain the only proven reservoir for TB infection of US
cattle (other than other infected cattle) (Corner, 2006). Wildlife reservoirs have also been identified outside the United States, including badgers in United Kingdom,(Corner, 2006; Morris et al., 1994) brush-tailed possums in New Zealand,(Corner, 2006) and elk (Cervus elaphus) in Manitoba, Canada (Corner, 2006). Even with intense efforts to eradicate bovine TB in the United States, this disease is detected in 8 US cattle herds annually ((APHIS), 2017).

Cattle with bovine TB pose serious risks to free-ranging wildlife if poor biosecurity practices allow for inadequate separation at the wildlife-livestock interface. Free-ranging deer (Odocoileus spp) and elk are of most concern because they often are attracted to shared food resources of cattle and disease transmission can occur directly from infected animals or indirectly through fomites.

The complex ecology and continuing reemergence of M. bovis necessitates rapid, thorough national and international surveillance and a better understanding of transmission dynamics among an increasing number of wildlife reservoirs and hosts in a variety of ecosystems. Reliance on the in vivo tuberculin skin test coupled with an assay to measure in vitro release of interferon-\(\gamma\) for the identification of infected free-ranging deer is untenable. Current methods require trapping and handling of animals, which are difficult tasks for farm or ranch deer and are not feasible or cost effective for free-ranging populations. Thus, an option for broad-based surveillance would be an easy, unambiguous diagnostic test that could be performed on serum samples obtained from hunter-harvested deer or as an ante-mortem test for farmed deer. Novel serum- or urine-based molecular markers are needed for identifying and monitoring the progression of M bovis infection in free-ranging animal populations to effectively estimate prevalence of
bovine TB so that cost-effective control measures can be implemented to prevent disease transmission to domesticated animals and humans.

Identification of novel molecular markers involves detection of circulating mycobacterial peptides, lipids, or metabolites in serum or plasma of infected animals. By use of this approach, 16 *M. bovis* proteins, including MB2515c (transcriptional regulator [LuxR family]), MB1895c (cell wall biosynthesis), and MB 1554c or pks5, were detected in bovine TB–infected and –exposed cattle (Lamont, Janagama, et al., 2014b). Biomarkers were first identified by use of gel-free multidimensional isobaric tag for relative and absolute quantitation proteomics and subsequently validated by use of a well-characterized cattle serum repository (Lamont, Janagama, et al., 2014b; Seth et al., 2009). An indirect ELISA that involves use of monoclonal antibodies synthesized against these peptides was developed to detect such biomarkers in serum and has been validated in samples obtained from host cattle (Lamont, Janagama, et al., 2014b) and primates with experimentally induced TB (unpublished data).

Current diagnostic tests have a one-size-fits-all method that does not take into consideration the disease prevalence for a given region. The biggest risk for spread of bovine TB in a region with a low disease prevalence (eg, the United States) is not wildlife or their natural movements; rather the biggest risk is from human movement of infected cattle (Carstensen, O’Brien, & Schmitt, 2011). If disease is detected in deer, efforts to reduce the free-ranging population and limit the risk of disease transmission can be costly as well as logistically and politically challenging (Carstensen et al., 2011; O’Brien et al., 2008).

The purpose of the study reported here was to use novel *Mycobacterium*-specific
peptides to develop a diagnostic test that could be used on deer sera for diagnosis of *M. bovis* infection. Use of tests to detect TB-specific biomarkers would have great potential to aid in the early detection and monitoring of bovine TB in wildlife through the use of hunter-harvested samples.

2.1 Materials and Methods

Sample

Serum samples were obtained from white-tailed deer in a laboratory environment and from MN DNR harvested free-ranging white-tailed deer.

**Experimentally infected deer**—Eight yearling white-tailed deer were maintained in a biosafety level 3 laboratory at the National Animal Disease Center (NADC). The deer were inoculated with $2 \times 10^8$ CFUs of *M. bovis* strain 1315 via intratonsillar instillation (day 0). Sera were obtained on days 0, 19, 48, and 60. A group of 3 age-matched control white-tailed deer were injected with saline (0.9% NaCl) solution on day 0, and serum samples were obtained at the same time points as for the inoculated deer. The age-matched control deer were housed separately from the *M. bovis*–inoculated deer (M. V. Palmer, Whipple, & Waters, 2001b; Seth et al., 2009). Approvals by an USDA IACUC agency were obtained for use of the animals in this study.

**Minnesota DNR free-ranging deer**—Sera were obtained from 393 MN DNR white-tailed deer from 2007 through 2010 during an outbreak of bovine TB in Minnesota (Carstensen & Doncarlos, 2011; Glaser et al., 2016). DNR obtained these samples as a part of TB eradication program in collaboration with the MN Board of Animal Health. Trained staff of the Minnesota Department of Natural Resources performed necropsies
and collected blood from the chest cavity of DNR harvested deer. Nine deer had gross lesions compatible with bovine TB and histologically confirmed granulomas containing acid-fast bacilli, and *M. bovis* was isolated from samples that were submitted to the National Veterinary Diagnostic Laboratory. Disease-negative animals were defined as deer that had negative culture results for *M. bovis*. These 384 disease-negative deer were considered an exposed-uninfected population for the analysis.

**Biomarkers**

Sera from calves experimentally infected with *M. bovis* were analyzed by use of multidimensional proteomics, whereby 32 host and 16 pathogen-specific peptides were identified that specifically increased in serum of *M. bovis*-infected calves (Seth et al., 2009). In addition, 16 *M. bovis*-specific peptides were identified in the same proteomics dataset (Lamont, Janagama, et al., 2014b). The 3 most abundant pathogen-specific peptides (MB1895c [a hypothetical protein with a molybdenum sulfurase domain], MB2515c [a transmembrane family protein as determined by analysis of amino acid sequences], and pks5 [a polyketide synthase potentially involved in lipid metabolism]) were identified by use of a well-characterized cattle serum repository and an antigen capture ELISA (Lamont, Janagama, et al., 2014b) with monoclonal antibodies against the peptides. These 3 pathogen peptide biomarkers were further validated for specificity in deer by use of sera from experimentally infected deer (Carstensen et al., 2011; Lamont, Janagama, et al., 2014b). For each sample tested, contemporaneous control samples were also tested.

**Indirect ELISA**

Serum from each of the 8 experimentally infected white-tailed deer, 3 age-matched
control deer, 9 TB-positive hunter-harvested deer, and 384 exposed-uninfected deer were diluted (1:50 dilution) in 0.05M carbonate-bicarbonate buffer (pH, 9.6). Fifty microliters of each diluted serum sample was transferred to separate wells in polystyrene flat-bottom ELISA plates. Each sample was assayed in duplicate, and positive and negative control samples were included on each plate. Sera were allowed to absorb overnight at 4°C. Plates were washed 3 times with PBS solution (200 µL/well). Plates were blocked by incubation with 5% bovine serum albumin in TBS solution (200 µL/well) for 2 hours at 37°C, washed 3 times (300 µL/well) with 1X phosphate buffered saline solution containing 0.05% Tween 20, and incubated (100 µL/well) with primary monoclonal antibodies against MB2515c (1:5,000 dilution), MB1895c (1:5,000 dilution), or pks5 (1.2 mg/mL) or with a 1% solution of bovine serum albumin in TBS with 0.05% Tween 20 for 2 hours at room temperature (22°C). Plates were then washed as described previously and incubated after addition (100 µL/well) of goat anti-mouse horseradish peroxidase–conjugated IgG diluted 1:10,000 in 1% bovine serum albumin in TBS solution with 0.05% Tween 20 for 2 hours at room temperature. Plates then were washed as described previously and were developed by incubation with tetramethylbenzidine (100 µL/well) in the dark for 30 minutes at room temperature. The tetramethylbenzidine reaction was stopped by the addition of 2M sulfuric acid (50 µL/well), and the OD at 450 nm was recorded by use of a microplate reader.

2.3 Data analysis

The OD data were uploaded into a spreadsheet, and S/N values were calculated for each biomarker; S/N was defined as the ratio of OD for the 9 TB-positive hunter-harvested deer against the mean OD of all the exposed-uninfected deer. Samples for
the 9 TB-positive hunter-harvested deer were used as the positive control samples for comparison against samples from the experimentally infected deer and the remainder of the hunter-harvested deer. These ratios were uploaded to a graphing program, and plots of time-course modulation of biomarkers in experimentally infected or hunter-harvested deer were generated. Box-and-whisker plots were generated for each biomarker by use of the same 9 TB-positive hunter-harvested deer.

Quantities of each biomarker for the 9 TB-positive hunter-harvested deer was simulated over time as a function of the total number of animals with positive test results against the total number of samples tested in a given time period. Prevalence of bovine TB as determined on the basis of biomarker presence was used to establish the extent and degree of disease burden since eradication of TB after the 2007–2010 outbreak was completed.

2.4 Results

2.4.1 Experimentally infected deer

The S/N for Pks5 and MB2515c increased gradually over the infection cycle and reached a peak at 60 days after inoculation (Figure 2.1). In contrast, the S/N for MB1895c increased early in the infection cycle and decreased after day 48 after inoculation.
Figure 2. 1—The S/N values for biomarkers pks5 (A), MB2515c (B), and MB1895c (C) in serum samples obtained from 8 white-tailed deer (*Odocoileus virginianus*) experimentally infected with *Mycobacterium bovis*. Deer were inoculated with 2 X 10^8 CFUs of virulent *M* bovis strain 1315 (day 0), and sera were obtained on days 0, 19, 48, and 60. Note that the scale on the y-axis differs among panels and that there is a gradual increase in concentrations of all 3 biomarkers as the infection progresses.

Figure 2. 2—Box-and-whisker plots of S/N values for 3 biomarkers in sera of 9 *M bovis*—
infected hunter-harvested white-tailed deer. S/N values for pks5 and MB2515c indicate that they would be reliable biomarkers for detection of bovine TB infection, whereas MB1895c would not be a reliable marker. Each box represents the interquartile range (25th to 75th percentiles), the horizontal bar in each box is the median, and the whiskers are the distribution of S/N values around the median.

2.4.2 TB-specific biomarkers in TB-positive hunter-harvested deer

Sera from the 9 TB-positive hunter-harvested deer were assayed. Range of the OD values (measured at 450 nm) was 0.198 to 0.407 for pks5, 0.211 to 0.510 for MB2525c, and 0.153 to 0.78 for MB1895c. Range of the S/N values was 3.394 to 6.977 for pks5, 4.234 to 10.234 for MB2515c, and 1.463 to 4.571 for MB1895c. Thus, S/N values for all 3 biomarkers calculated by use of sera from the 9 TB-positive hunter-harvested deer were between 1.463 and 10.234. Range of the S/N values obtained for each biomarker was used to identify cutoff values to determine infection status. By use of these ranges, an S/N value ≥ 2.5 for any of the 3 biomarkers was considered a positive result (infected), and an S/N value < 2.5 was considered a negative result (uninfected).

Evaluation of box-and-whisker plots of the S/N values for the 3 pathogen-specific biomarkers revealed that pks5 and MB2515c were the most reliable biomarkers for use in determining infection status for the 9 TB-positive hunter-harvested deer (Figure 2.2). Interquartile ranges for pks5 and MB2515c overlapped; thus, these 2 biomarkers were considered to be the most reliable. The S/N values for MB1895c differed slightly from those of the other 2 biomarkers; hence, it was not considered to be a reliable biomarker.

2.4.3 Prevalence of bovine TB in Minnesota deer

Infection status of 384 hunter-harvested white-tailed deer in northwest Minnesota was evaluated by use of serum samples obtained from 2007 through 2010. Prevalence of
bovine TB was estimated by use of 3 S/N cutoff values (2.5, 3.0, and 3.5) to provide information on the prevalence over time (Figure 2.3). Prevalence estimates indicated that bovine TB in wild Minnesota deer peaked in approximately 2009 (all 3 cutoff values) and decreased thereafter to almost undetectable levels (the most conservative cutoff value of 3.5). However, use of a cutoff value of 2.5 or 3.0 indicated that bovine TB appeared to be present, albeit at an extremely low level, in Minnesota deer.

Figure 2.3—Overall TB prevalence in sera of hunter-harvested deer by year for each biomarker by use of 3 S/N cutoff values (2.5 [squares], 3.0 [circles], and 3.5 [diamonds]). Notice that there is a gradual decrease in TB prevalence from 2007 to 2010.

2.5 Discussion

Bovine TB is a zoonotic disease associated with devastating consequences for agriculture and public health. *Mycobacterium bovis* threatens domestic animals, humans, and the economy via a number of pathways. First, wildlife reservoirs of bovine TB have directly caused an increase in the incidence of TB in cattle in recent years,
which resulted in loss of animals as well as trade dollars. Second, TB is a major zoonotic concern in rural areas in developing countries in which the animal-human interface is intensifying as patterns of land use change.

Existing strategies for the detection of TB in imported animals and local wildlife (in vivo tuberculin skin test coupled with assay of in vitro interferon-\(\gamma\) release) are ill-equipped to tackle a problem of this magnitude. Proposed strategies are for detection of host immune responses or identification of intact bacteria in lesions.

Existing strategies are rarely practical for implementation with free-ranging wildlife populations or specimens. Tests have poor performance metrics (eg, low sensitivity) and are labor-intensive, invasive, expensive, and unavailable at critical locations, specifically remote farming areas in the United States and developing countries where bovine TB is widespread. In addition, test results provide no information about the stage of disease.

In the present study, 3 pathogen-specific peptides were evaluated by the use of sera from experimentally infected deer and wild deer. This appeared to be a practical sampling strategy for monitoring disease in free-ranging wildlife and would be sufficient to rapidly identify and allow a response to emerging outbreaks of TB in wildlife. Results indicated that pks5, MB2515c, and MB1895c can be used to detect infection in samples from TB-positive and exposed deer and that they can be used to distinguish between uninfected and infected animals. The most reliable pathogen biomarkers for bovine TB in the present study were pks5 and MB2515c. In another study (Lamont, Janagama, et al.,
investigators found that pks5 was the most reliable biomarker for bovine TB. It was not surprising that the S/N values for experimentally infected deer were lower than those from free-ranging deer because the time frame in which the samples from the experimentally infected deer were obtained was extremely short (60 days), but samples of hunter-harvested deer were obtained from adult deer in which infection potentially had a longer time to develop.

In the present study, the combination of pks5 and MB2515c were optimal indicators of infection (or its absence). Simulation of the bovine TB infection status by use of a range of S/N values indicated that although bovine TB in Minnesota deer decreased after 2009, it may have persisted at extremely low levels as subclinical disease.

These findings confirmed that the calculated prevalence of disease differed depending on the S/N cutoff value. The cutoff value would differ on the basis of the areas in which disease surveillance is performed (eg, regions with endemic infection vs low-prevalence regions). Areas with endemic infection require tests with high sensitivity or low cutoff values to maximize accurate detection and minimize potential spillover to other domestic animals and wildlife. Thus, cutoff values would need to be optimized on the basis of the population prevalence. In areas with a low prevalence of bovine TB (such as Minnesota), a test with high specificity or a high cutoff value would be optimal to ensure that disease-free animals are not misclassified as test-positive animals. A region with a low disease prevalence (such as the United States) could greatly benefit from a diagnostic test that is highly specific, which should prevent misclassification resulting from animal exposure to other mycobacteria such as *Mycobacterium avium* subsp *paratuberculosis*. On the other hand, regions in Africa and Asia with a high disease prevalence need diagnostic tests
with high sensitivity to identify all infected animals and minimize potential transmission to humans and other animals.

Existing diagnostic tests for bovine TB are inadequate for the purpose of disease surveillance and monitoring of free-ranging wildlife. Current methods for diagnosis of bovine TB rely on direct detection of the pathogen (through microscopy or culture) or on DNA amplification (Gardiner & Karp, 2015b). Infection is determined by the presence of sustained T-cell reactivity with antigens of the \( M \) \( \text{tuberculosis} \) complex (tuberculin skin tests or assays of interferon-\( \gamma \) release in blood). These tests, although useful, do not provide an optimal foundation for control of bovine TB or a clear path to the development of improved strategies (Gardiner & Karp, 2015b). Bacterial culture (the criterion-referenced standard for bovine TB testing) is time consuming, which results in delays for obtaining results.

Critical challenges in diagnostic testing to detect disease caused by organisms of the \( M \) \( \text{tuberculosis} \) complex include a lack of effective tools to provide timely and accurate results. Research in TB diagnostic testing has yielded promising results. However, there is still a scarcity of tools for the diagnosis of TB in animals. We speculate that the direct detection of lesions attributable to infection with \( M \) \( \text{tuberculosis} \) or \( M \) \( \text{bovis} \) or their products (eg, peptides and DNA) in blood and urine by use of high-resolution de novo methods has better sensitivity than does bacterial culture of specimens, but the usefulness is still well below that required for a diagnostic test (O'Brien et al., 2008; Young et al., 2014a). Lipoarabinomannan has been extensively studied, and a commercially available test to detect lipoarabinomannan in urine is available, although
its low sensitivity has limited its use (Gardiner & Karp, 2015b). Detection of antibodies in serum and urine against several TB proteins including antigen 85 complex have shown promising results (Gardiner & Karp, 2015b; Young et al., 2014a). Mycobacterium tuberculosis complex–specific antibody responses have been evaluated, but further optimization is needed because of the heterogeneity of the antibody response against the M tuberculosis complex (Gardiner & Karp, 2015b).

A highly sensitive and specific diagnostic test designed to evaluate disease progression and response to treatment, would provide the ultimate diagnostic test for TB. Investigators of a recent study (Gardiner & Karp, 2015b) indicated target product profiles a biomarker should meet to provide the greatest impact as an ideal TB diagnostic test. Although the emphasis of that study (Gardiner & Karp, 2015b) was on TB in humans, the criteria remain the same for bovine TB, a zoonotic disease. An ideal biomarker-based assay should have high (> 98%) sensitivity and specificity, be noninvasive or minimally invasive, provide results rapidly, and be affordable. Investigators of 1 study (Haas, Roe, Pollara, Mehta, & Noursadeghi, 2016a) indicated that a possible combination of biomarkers will confer diagnostic value for different settings (eg, one set of markers for differentiating between active and latent TB infections in humans or subclinical infections in other animals, and another set of markers to differentiate TB from other diseases). However, this would require collaborative efforts on the part of researchers on TB in humans and other animals (especially cattle).

Current methods for detection of bovine TB in animals are inadequate, and control programs cannot rely solely on test-and-slaughter of reactor animals. Methods need to be supplemented with results of comprehensive epidemiological investigations of
outbreaks of bovine TB and surveillance at abattoirs, implementation of movement restrictions of animals, and use of biomarker technology (Carstensen & Doncarlos, 2011; M. V. Palmer et al., 2001b; H. M. Vordermeier, Jones, Buddle, Hewinson, & Villarreal-Ramos, 2016b; H. M. Vordermeier, Perez de Val, et al., 2014). There is a need for a highly sensitive and specific diagnostic test for TB that can rapidly detect latent or subclinical disease in animals of various disease prevalences at a reasonable cost on an easily obtainable sample such as blood, urine, or exhaled breath.

In the present study, a biomarker detection assay was used to detect subclinical as well clinical infection in white-tailed deer and to differentiate between infected and noninfected experimentally infected deer. Information about new biomarkers is commonly published, but refinement, validation, and independent confirmation of such biomarkers often is not accomplished (Gardiner & Karp, 2015b). Three biomarkers (pks5, MB2515c, and MB1895c) were validated for use in detection of TB infection of different host species (deer and cattle), with promising results. Application of TB-specific biomarkers in diagnostic tests would greatly enhance the ability to screen for this mycobacterial disease that traditionally has been difficult to monitor. Biomarkers for the \textit{M tuberculosis} complex appeared to be valid indicators of TB infection, and we propose that that be incorporated into a point-of-care device for the rapid screening of TB. Evaluation of highly specific serum biomarkers composed of mycobacterial peptides and proteins may provide insights into disease progression and interspecies transmission of \textit{M bovis} infections between humans, cattle, and deer.
CHAPTER THREE

Develop a Circulating Nucleic-Acid Based Detection System to Detect

*Mycobacterium bovis* Infection in Animals

(Short Communication-JCM)

Sylvia I. Wanzala, Nicole Herman and Srinand Sreevatsan

3.0 Background

Bovine tuberculosis a zoonotic infection of domestic and wild animals caused by

*Mycobacterium bovis*. Approximately one-third of the world’s population is infected with

*M. tuberculosis* or *M. bovis*, “the world’s most successful pathogen” (Hingley-Wilson et al., 2003), the majority in developing countries. The global spread and increasing severity of tuberculosis are in part due to the increasing intensity of human-animal interactions as land use patterns around the globe rapidly change. Furthermore, the existence of wildlife reservoirs such as deer in the US, complicate the ecology and mitigation of this devastating disease in agricultural animals.

The severe and growing threat of *M. bovis* necessitates rapid, thorough national and international surveillance of strain distribution dynamics in the population. Relying on the existing *in vivo* tuberculin skin test coupled with an *in vitro* interferon - assay to identify infected wild deer is unsustainable. The tests require trapping and handling of animals – a task that is not feasible as an application for surveillance of tuberculosis in wildlife species. We require an easy to perform and unambiguous set of diagnostic tools that identify and can track the progression of *M. bovis* infection in wild free-ranging wild
animal populations to effectively control bovine tuberculosis and prevent its transmission to domesticated animals and humans.

3.1.1 Biomarkers in disease detection

The analysis of alterations in circulating DNA or RNA in plasma are promising diagnostic tools, requiring only a limited blood sample. Recently completed human and animal genomes have brought refinements in technology including nucleotide and protein sequencing, mass spectrometry, and microarrays (nucleic acid and protein arrays), and these have allowed researchers to elucidate fundamental biological processes of chronic diseases such as cancers, neurological disorders, cardiovascular disease and several infectious diseases (Correia et al., 2017). These studies would pioneer the applications of biomarkers in tuberculosis diagnostics, as we will identify *M. bovis*-specific circulating proteins or nucleic acids in the serum of infected deer.

3.1.2 Circulating nucleic acids as a biomarker

A novel approach involves detecting circulating nucleic acids in the serum or plasma of infected animals. Circulating nucleic acids are segments of RNA and DNA in biologic fluids devoid of cellular material, and they are useful laboratory markers in the staging of some chronic illnesses. Currently, most research in plasma or serum DNA focuses on its diagnostic application to several cancers, fetal anomalies, and viral diseases. Chronic bacterial infection is an ideal application of these techniques, though its use has not yet been attempted. Apoptosis and inflammatory necrosis have been proposed as major mechanisms through which free circulating nucleic acids accumulate in plasma and serum (Jiang & Pisetsky, 2005; Lichtenstein et al., 2001; Lui & Dennis, 2002; Stroun,
Lyautey, Lederrey, Mulcahy, & Anker, 2001; Stroun et al., 2000), and these processes are integral to mycobacterial survival in macrophages, granuloma formation, and the chronic inflammatory processes induced in mycobacterial disease (Cassidy, 2006). *M. bovis* constantly recycles its cell wall components and repairs its genome as a result of host induced stress (Cosma, Sherman, & Ramakrishnan, 2003), and these may be released as circulating nucleic acids. Recent studies on mycobacterial pathogenesis have shown an important role for small RNAs in post-transcriptional regulation as well as in apoptosis (Arnvig et al., 2011; Arnvig & Young, 2009; Obregon-Henao et al., 2012). Therefore, we expected to identify any relative increases of specific circulating nucleic acids that occur as the disease progressed.

In this work, we used powerful, novel massively parallel de novo sequencing to detect specific infection in deer by identifying specific nucleotide motifs released from the *M. bovis* genome as circulating nucleic acids in serum. These circulating nucleic acids can serve as surrogates for progression of infection, and will likely be useful to index responses to treatment in related *Mycobacterium tuberculosis complex* infection of animals and humans.

### 3.1.3 Micro and small noncoding RNAs

Apoptosis and inflammatory necrosis have been proposed as major mechanisms through which circulating nucleic acids accumulate in plasma and serum. These processes have been shown to be critical to the survival of mycobacteria within macrophages, granuloma formation and the chronic processes of inflammation. In this specific aim, the characteristics of micro and small RNA extracted from experimentally infected deer and their contemporaneous controls were looked at. MicroRNA are 19-24
nucleotide long single stranded RNAs that regulate the expression of target genes by interacting with complementary sites of the target mRNAs and inhibiting translation. These non-coding RNAs play very key regulatory roles. Previous work in our lab on *M. avium* subsp. *paratuberculosis* demonstrated that there are unique characteristic sequence motifs between infected and uninfected animals. These tools were harnessed to specifically detect infection by *M. bovis* in experimentally infected deer samples by identifying specific nucleotide motifs released into circulation from the *M. bovis* genome as circulating nucleic acids in serum. These nucleic acid motifs can provide a means to determine disease progression, and will possibly provide means of determining treatment responses to MTC infection in animals and humans.

Furthermore, there’s increasing evidence to demonstrate the critical role of microRNA in host-pathogen interactions and how they can serve as unique molecular markers for disease. MicroRNA biomarkers for use in tuberculosis (TB) diagnosis and treatment pose an exciting challenge for TB researchers (Bettencourt, Pires, & Anes, 2016; Ehrlich, 2015; Kim, Kim, Basu, & Jo, 2017; Mehta & Liu, 2014). There is currently limited information about the host-pathogen interactions at the RNA level in mycobacterial infections; TB researchers postulate that there might be a possible regulatory role for miRNA in the interaction between macrophages and bacteria (Kim et al., 2017). Studies have shown that specific miRNAs are present during infection and influence the state of the disease (Ehrlich, 2015; Mehta & Liu, 2014). Better knowledge about the host-pathogen interactions for tuberculosis will ensure that there are enhanced ways to diagnose it at an early stage or find new methods of treatment. These microRNA studies will provide another alternative in the search for better TB diagnostics.
Well-characterized prospective samples were collected from an aerosol infection model of deer tuberculosis in white-tailed deer to identify *M. bovis*-specific nucleotide sequences in the serum as the disease progressed. Distinct sera were analyzed from each time point- baseline, early, mid, and late stage disease (test sera as well as control sera from uninfected animals bled at the same time points) -by standard RNA extraction followed by high throughput ultra-deep sequencing to identify circulating nucleic acids. The studies also examined modulation of biomarkers during the progression of bovine tuberculosis and subsets that can enable the early diagnosis of disease were identified. The identified miRNA nucleic acid motifs predictive of disease progression will be tested on several well-characterized sera from field cases of deer tuberculosis to evaluate the potential for use in routine surveillance of bovine tuberculosis diagnostics. Because many molecules (including proteins and lipids) and nucleic acids (DNA, small RNAs) are unique to mycobacteria or mycobacterial infection, a diagnostic test based on the detection of mycobacterial components will provide exceptional specificity, far beyond what we currently realize with established methods.

In this work, we hypothesized that pathogen-host interactions would lead to accumulation of pathogen and host specific small molecules in the serum that can be applied to accurately diagnose bovine tuberculosis in animal populations. This hypothesis had been earlier tested on a small sample data set from *Mycobacterium paratuberculosis* (MAP) infection studies and results obtained demonstrated that there are characteristic sequence motifs between infected and uninfected animals and we went ahead to determine what would happen in *M. bovis* infected animals. A circulating nucleic acid based detection system to identify *M. bovis* infection in animals was the goal of this study.
3.2 Materials and methods

3.2.1 Samples

Serum samples were obtained from experimentally infected white-tailed deer in a laboratory environment. These were eight yearling white-tailed deer that were maintained in a biosafety level 3 laboratory at the National Animal Disease Center. The deer were inoculated with $2 \times 10^8$ CFUs of *Mycobacterium bovis* strain 1315 via intra-tonsillar instillation at day 0. Sera were collected on days 0, 19, 48, and 60. A group of 3 age-matched control white-tailed deer were injected with saline (0.9% NaCl) solution on day 0, and serum samples were obtained at the same time points as for the inoculated deer. The age-matched control deer were housed separately from the *M. bovis*-inoculated deer. Approvals by a USDA institutional animal care and use committee were obtained for use of the animals in this study.

3.2.2 RNA extraction and processing

Total RNA was extracted using a commercial kit (Qiagen®, miRNeasy for serum/plasma) from eight experimentally infected deer (infected with *Mycobacterium bovis*) and their controls: uninfected controls, 19 days post-infection, 48 days post-infection and 60 days post-infection. After passing quality control checks at the UMGC, small RNA libraries were created from the circulating serum RNA extraction. Size selection of small RNA libraries was performed prior to sequencing using the HiSeq 2500 Rapid Mode, 50bp single read method. The resulting data was extracted as pile up files and analyzed to determine the behavior of micro and small RNAs over different time points of disease progression.
At the UMGC 10 single indexed TruSeq smRNA libraries were created and combined into a single pool. They were sequenced in one lane of a 50 bp SR run on the HiSeq 2500 HO instrument using v4 chemistry and the lane generated about 170 million reads. Libraries were size selected to 145 bp +/- 12% and the mean quality score was above Q30 for all libraries. Adapters were removed using the Trimmomatic software and the resulting sequences were mapped to the references using Bowtie Galaxy version 2.2.6.2 (Langmead & Salzberg, 2012; Langmead, Trapnell, Pop, & Salzberg, 2009). The numbers of mapped reads were subsequently calculated for each miRNA and provided as a number of each of the unique reads mapped to each reference sequence, and as a number of all the reads mapped to each reference sequence. Pile up files for each set of reads were obtained using SAMTools (Li, 2011; Li et al., 2009); the pile up file format displays every base where one or more reads mapped to the reference sequence. Analysis was done in MS Excel of these pile up files to obtain the specific miRNA, their sequences as well as positions relative to the reference genome.

3.3 Results and Discussion

Table 3. 1-Small RNA mapping statistics

<table>
<thead>
<tr>
<th>File</th>
<th>Index</th>
<th># reads (millions)</th>
<th># mapped 1x</th>
<th>% mapped</th>
<th># mapped &gt;1x</th>
<th>Total # mapped</th>
<th>Alignment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Af2122</td>
<td>29222320</td>
<td>216894</td>
<td>0.74%</td>
<td>39</td>
<td>216933</td>
<td>0.74%</td>
</tr>
<tr>
<td>Day 19 test</td>
<td>Af2122</td>
<td>20193684</td>
<td>462155</td>
<td>2.29%</td>
<td>68</td>
<td>462223</td>
<td>2.29%</td>
</tr>
<tr>
<td>Day 19 control</td>
<td>Af2122</td>
<td>22754526</td>
<td>50241</td>
<td>0.22%</td>
<td>39</td>
<td>50280</td>
<td>0.22%</td>
</tr>
<tr>
<td>Day 48 test</td>
<td>Af2122</td>
<td>27167115</td>
<td>843584</td>
<td>3.11%</td>
<td>66</td>
<td>843650</td>
<td>3.11%</td>
</tr>
<tr>
<td>Day 48 control</td>
<td>Af2122</td>
<td>15235975</td>
<td>125843</td>
<td>0.83%</td>
<td>41</td>
<td>125884</td>
<td>0.83%</td>
</tr>
<tr>
<td>Day 60 test</td>
<td>Af2122</td>
<td>22816790</td>
<td>723514</td>
<td>3.17%</td>
<td>45</td>
<td>723559</td>
<td>3.17%</td>
</tr>
<tr>
<td>Day 60 control</td>
<td>Af2122</td>
<td>14507874</td>
<td>194601</td>
<td>1.34%</td>
<td>40</td>
<td>194641</td>
<td>1.34%</td>
</tr>
<tr>
<td>Baseline</td>
<td>Bos taurus</td>
<td>29222320</td>
<td>30633</td>
<td>0.10%</td>
<td>25966</td>
<td>56599</td>
<td>0.19%</td>
</tr>
</tbody>
</table>
Next Generation Sequencing (NGS) data on microRNAs revealed about 170 million reads per lane. NGS data was processed to remove adaptor sequences using Clip and Trimmomatic (Bolger, Lohse, & Usadel, 2014) tools in the Galaxy interface. Size selection of the RNA was done to extract all micro and small RNA, which yielded sufficient concentrations of these species for further analysis. Analysis of this data reveals that circulating microRNA can be used as a nucleic acid based detection system to non-invasively detect TB infection in animals.

Findings from our micro and small RNA analysis indicate that there are different micro and small RNAs present at different time points of disease. Some vary over time while others appear at specific times post infection and can be used to detect and stage the infection stages as disease progresses (See Tables 3.1-3.3) (Table 3.3 in appendices).

Micro RNA are conserved across species, and if a particular miRNA or set of miRNA have been shown to have diagnostic utility in humans or laboratory animals, they are good candidates for veterinary species (Lidbury, 2016). In these results, miRNAs identified from species as diverse as humans beings (*hsa- Homo sapiens*); the dog (*cfa- Canis familiaris*) to the cow (*bta-Bos taurus*) and the soil nematode (*cel-Caenorhabditis elegans*) were seen. See Table 3.3 in appendices.
Table 3. 2: sRNA of *Mycobacterium bovis* discovered at post-infection time points

<table>
<thead>
<tr>
<th>TIME POINTS</th>
<th>depth</th>
<th>No. of bases/length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SEQUENCE) UPREGULATION X COVERAGE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11112</td>
<td>GGGGCTTAGCTCAGTTG GTAGAGCA</td>
<td>16</td>
</tr>
<tr>
<td>787573</td>
<td>CTACCGTCGCA GTTCTACTCC</td>
<td>2</td>
</tr>
<tr>
<td>852879</td>
<td>TGCTCTACCACTG AGCTAAGGAGGC</td>
<td>2</td>
</tr>
<tr>
<td>852879</td>
<td>TGCTCTACCACTGAGCTA AGGAGGCA</td>
<td>4</td>
</tr>
<tr>
<td>925087</td>
<td>AGGTCCGCCGGT GCATCCCG</td>
<td>3</td>
</tr>
<tr>
<td>925087</td>
<td>AGGTCCGCCGGTGCATCCCG</td>
<td>10</td>
</tr>
<tr>
<td>925087</td>
<td>AGGTCCGCCGGTTCGATCCCC</td>
<td>3</td>
</tr>
<tr>
<td>1470163</td>
<td>CGGGTAGGGGGGCTGAGGG</td>
<td>4</td>
</tr>
<tr>
<td>1470295</td>
<td>CGGGCTTCGGGGTTGAAA</td>
<td>3</td>
</tr>
<tr>
<td>1470571</td>
<td>CGGAGATATCA GGAGG</td>
<td>55</td>
</tr>
<tr>
<td>1470754</td>
<td>CGGCTGGGGAGTGTCGGGG</td>
<td>1176</td>
</tr>
<tr>
<td>1470989</td>
<td>GTTCATGTGG CCAG</td>
<td>3</td>
</tr>
<tr>
<td>1471803</td>
<td>GATTCCGAAT GGGAAACC</td>
<td>5</td>
</tr>
<tr>
<td>1471891</td>
<td>AACATCTCAGTTA CCC</td>
<td>6</td>
</tr>
<tr>
<td>1472233</td>
<td>CGGATTAGTAC CGTQAGG</td>
<td>3</td>
</tr>
<tr>
<td>1472262</td>
<td>AGTACCCCQGGG AGGGGAGT</td>
<td>12</td>
</tr>
<tr>
<td>1472278</td>
<td>AGTGGAAAGAGTACCTGAA ACCGTTGG</td>
<td>3</td>
</tr>
<tr>
<td>1472289</td>
<td>ACCTGAAACCG TGTGC</td>
<td>6</td>
</tr>
<tr>
<td>1472428</td>
<td>AAAGGGAGTCT G</td>
<td>5</td>
</tr>
<tr>
<td>1472604</td>
<td>AAGGAGCTAATCAAACCTCC</td>
<td>4</td>
</tr>
<tr>
<td>1472830</td>
<td>CAAGCATGTGTC TAAGTGGGGAA</td>
<td>5</td>
</tr>
<tr>
<td>1473611</td>
<td>CCCGTAACTTCGGGAGA</td>
<td>3</td>
</tr>
</tbody>
</table>

From Tables 3.2 and 3.3 there’s a clear presence of various miRNA in the serum of experimentally infected animals. Since there are no annotated miRNA for mycobacteria, for the purpose of this paper, we focused on the host-miRNA in Table 3.3. The reference genome that we mapped the miRNA against was *Bos taurus* although the serum came from white-tailed deer (*Odocoileus virginianus*). This is because the deer genome has
not been completed yet. The data we have is unique and to the best of our knowledge is the first such study looking at miRNA arising from infection with *Mycobacterium bovis* in deer.

Looking at the data in **table 3.3** there are several miRNA present in serum as early as 19 days post infection. Most miRNA were seen at 19 days post-infection with 40 viable hits in the miRNA database - miRBASE® (Kozomara & Griffiths-Jones, 2011, 2014). There were 8 viable hits at 48 days post-infection and 14 viable hits at 60 days post-infection. Some miRNA like, mmu-miR-8107 and efu-miR-9235b were seen at both D-48 and D-60 post-infection where they were both upregulated.

There were several miRNAs whose sequences had homologs with human miRNA like hsa-miR-6857-5p, hsa-miR-6876-3p, hsa-miR-3907 and hsa-miR-6855-3p. Clearly, distinct miRNA are appearing at specific time points that can be used to develop a circulating nucleic acid based detection system to identify *M. bovis* infection in animals. These shall need to be validated and confirmed with techniques like RT-PCR. Micro RNA have been demonstrated to have high stability in extracellular fluids of mammals like blood plasma, serum, urine, saliva and semen and this makes them attractive biomarker candidates for diagnosis or prognosis of complex diseases since they are still viable after repeated cycles of freeze-thawing and long term storage of frozen samples (Correia et al., 2017).

In our research, we hypothesized that pathogen-host interactions would lead to accumulation of pathogen and host specific small molecules in the serum that can be applied to accurately diagnose bovine tuberculosis in animal populations. With this information, we see that fine-tuning of biomarker technology is critical and its use together
with current BTB detection methods will greatly assist in providing quick, accurate and inexpensive diagnostics for this traditionally difficult disease. These culture-independent pathogenesis based, diagnostic tools will have the ability to detect, and differentiate between mycobacterial species as well as identify mixed infection in humans and animals.
CHAPTER FOUR

Comparative Genomics of Archived Pyrazinamide Resistant *Mycobacterium tuberculosis* Complex Isolates from Uganda

Sylvia I. Wanzala ab, Jesca Nakavumaa, Dominic Travisb, Praiscillia Kiaa, Sam Ogwanga, W. Ray Waters, d Tyler Thacker, d Timothy Johnson, e Srinand Sreevatsanb, #

4.0 Introduction

*Mycobacterium bovis*, a subspecies of the *Mycobacterium tuberculosis* complex (MTC), is a significant cause of bovine tuberculosis (bTB) in the developing world (Banuls, Sanou, Anh, & Godreuil, 2015; Cole et al., 1998). Each MTC member displays varied pathogenicity, geographical distribution, epidemiology, and host range (Dos Vultos et al., 2008). MTC, the causative agents of tuberculosis, are intracellular pathogens that have very low (0.01-0.03) synonymous nucleotide variation with no significant trace of genetic exchange among them (Gutierrez et al., 2005). In the developed world, bovine TB (bTB) has nearly been eradicated from domesticated animals thus the risk of human transmission is low. This was achieved through mandatory testing of cattle, removal of positive reactors and in-contact animals and compulsory pasteurization of milk (Michel et al., 2010; Nelson, 1999; O’Reilly & Daborn, 1995). In contrast, in resource-limited countries where the devastation of MTCs on humanity continues unabated, effective disease control measures like mandatory milk pasteurization and removal of positive reactors is difficult to implement mainly due to economic factors (Berg et al., 2015; Cosivi et al., 1998b; Daborn et al., 1996; El-Sayed, El-Shannat, Kamel, Castaneda-
Vazquez, & Castaneda-Vazquez, 2016; Pesciaroli et al., 2014). For example, bTB is not reportable in many countries in sub-Saharan Africa and there are no compensatory mechanisms in place for farmers whose animals test positive for the disease (Kazwala et al., 2001; Michel et al., 2010; Muller et al., 2013; Wedlock et al., 2002).

The World Health Organization (WHO) estimates that two billion humans are infected with latent tuberculosis with a 5-10% chance of developing active TB ("Mycobacterioses and the acquired immunodeficiency syndrome. Joint Position Paper of the American Thoracic Society and the Centers for Disease Control," 1987; World Health, 2015). The brunt of this burden is largely borne by developing countries with an approximate two million deaths annually; 95% of which are in low or medium-income countries (World Health, 2015). The proportion of tuberculosis infections caused by M. bovis is not well defined especially in resource-limited settings of the world. Prevalence estimates of M. bovis range from 5-16% (Kazwala et al., 2001; Wedlock et al., 2002). In Uganda, routine TB testing does not assign species within MTC organisms hence any human infection with an MTC is assumed to be due to Mycobacterium tuberculosis (sensu stricto). In 2015, Uganda had an estimated population of 39 million people with only about five laboratories that carry out routine TB screening with culture and drug sensitivity testing; these are all located in urban centers (World Health, 2016a, 2016b). Staining for acid fast bacilli (AFB) on sputum smears to screen for TB is carried out by a larger number of network laboratories. Animal TB diagnostics in Uganda are lagging far behind human TB diagnostics and available bTB information is based on passive surveillance data from abattoirs that are active in recording and submitting annual cattle slaughter reports (Oloya, Muma, et al., 2007). Uganda does not have a national bovine TB eradication program and previous studies on bTB in Uganda have shown the presence of M. bovis
among cattle belonging to the pastoral communities (Oloya, Kazwala, et al., 2007; Oloya, Muma, et al., 2007; Oloya et al., 2006). Although thousands of animals were examined in Oloya’s study (Oloya, Kazwala, et al., 2007), only 61 had lesions compatible with bTB reflecting the low sensitivity of post-mortem examinations.

Even though *M. tuberculosis* is the most prevalent of the *M. tuberculosis* complex in Kampala, infections with *M. tuberculosis* or *M. bovis* are clinically and pathologically indistinguishable (Gallivan, Shah, & Flood, 2015; Wedlock et al., 2002). The challenge is physicians treat both diseases similarly although *M. bovis* is naturally resistant to pyrazinamide, one of the first line TB drugs (Huard, Lazzarini, Butler, van Soolingen, & Ho, 2003). More importantly, species assignment of MTCs provides a possible eco-source of the agent that can aid in appropriate treatment and prevention. It is therefore imperative that MTC isolates are completely identified to subspecies level in order to provide sufficient information on their epidemiology and enable appropriate treatment of the patients (Huard et al., 2003).

There is little documented information about the proportion of people infected by *M. bovis* in Uganda. This work was developed to determine and differentiate archived human MTC isolates, from a large reference laboratory, Joint Clinical Research Centre (JCRC), Kampala, Uganda, as *M. bovis* or *M. tuberculosis* (MTB). The JCRC is an HIV/AIDS care and research institution established in 1990 to provide a scientific approach to the national HIV/AIDS problem (Joint Clinical Research, 2015). This retrospective study characterized archived pyrazinamide resistant isolates stored at the JCRC from 1999 through 2012.
4.1 MATERIALS AND METHODS

4.1.1 Institutional Review Board (IRB) and ethical approvals.

Permission to use isolates for molecular characterization was obtained from the research committee of the JCRC; the JCRC Uganda-Case Western Reserve University (CWRU) Research Collaboration and the Molecular Diagnostics Laboratory Department of Medical Microbiology, College of Health Sciences (CHS), Makerere University Kampala. Institutional Review Board (IRB) approval was obtained from the Makerere University, School of Biomedical Sciences Research and Ethics Committee. Scientific and ethical clearance was obtained from the Uganda National Council for Science and Technology (UNCST) reference number HS 1478. Since the work was a retrospective study using archived isolates, this work was exempt from University of Minnesota IRB approval.

Study site. The archived MTC isolates were obtained from the JCRC in Kampala, Uganda (Joint Clinical Research, 2015). Confirmatory tests to identify \textit{M. bovis} were done at the Molecular Biology Lab in the Department of Microbiology, College of Health Sciences, Makerere University.

4.1.2 Study samples.

Pyrazinamide (PZA) resistant isolates were drawn from a collection of 5,676 archived isolates collected between 1999 and 2012 from 19 different research studies previously conducted in partnership with key research collaborators at JCRC; see Figure 1 in Supplementary Figures.
We first determined the proportion of *M. tuberculosis* isolates resistant to first line anti-tuberculosis drugs; (pyrazinamide, rifampin, isoniazid and ethambutol and streptomycin) and then carried out comparative genomics. Furthermore, we sought to determine if there was a nonrandom distribution of genomotypes of MTC organisms infecting animals and humans in Uganda using a defined Single Nucleotide Polymorphism (SNP) set to differentiate the *Mycobacterium tuberculosis* complex group of organisms and enable determination of drug resistance and characterization of these MTC isolates as drug resistant (DR), multi-drug resistant (MDR), or extremely drug resistant (XDR) organisms.

4.1.3 Sampling, culture and isolation of Mycobacteria.

All the work with the MTC isolates was carried out in Biosafety level three (BSL-3) containment facilities. Archived samples from the JCRC were retrieved from -80 °C freezer stocks, thawed and inoculated into Middlebrook 7H9® broth and incubated at 37°C for 42 days (Joseph Sambrook, Fritsch, & Maniatis, 1989; J. Sambrook & Russell, 2006). The samples were sub-cultured on solid media, Middlebrook 7H10® containing sodium pyruvate, and Oleic Acid-albumin-Dextrose-Catalase (OADC®) enrichment. In case of contaminating growth, the cultures were also sub-cultured on selective media, Middlebrook 7H11® containing sodium pyruvate, glucose, sterile calf serum, OADC enrichment, sheep blood, malachite green and glycerin and incubated at 37°C. The cultures were inspected weekly and considered negative when there was no visible growth after 8 weeks. Ziehl Neelsen and auramine fluorescent staining and microscopy were carried out on the isolates to confirm that they were *Mycobacterium sp*.

4.1.4 Molecular testing to confirm MTC.
Isolates were harvested and inoculated in Middlebrook 7H9® broth in cryovials and taken to the Molecular Diagnostic Laboratory, Department of Medical Microbiology, Makerere University for heat killing, DNA extraction and PCR-based Regions of Difference (RD) analysis to speciate *M. bovis* from other members of MTC. DNA was extracted by sonication and quantified using the GeneQuant® UV spectrophotometer (20-30 μg per isolate) (26). Genomic regions of difference (RD analysis) for RD1, RD9, RD4 and RD12 were used to differentiate between the different MTC species as described (Huard et al., 2006; Huard et al., 2003; Pym, Brodin, Brosch, Huerre, & Cole, 2002; Thorne et al., 2011). These RDs are regions within the MTC genome that represent loss of genetic material in *M. bovis* compared to *M. tuberculosis* H37Rv (Brosch et al., 2002; Huard et al., 2003; Kamerbeek et al., 1997).

4.1.5 DNA extraction/ DNA preparation for PCR.

Cells stored in 7H9 at -20°C were harvested and heat killed at 95°C for 1 hour in a hybridization oven and DNA was extracted by sonication. Heat killed samples were centrifuged at 13,000 r.p.m for 2 minutes in a microfuge, supernatant discarded and pellet retained. To the pellet, 500μl of Tris-EDTA buffer, pH 8, was added and centrifuged at 13,000 r.p.m for another 2 minutes and the pellet was retained after pouring off the supernatant. The pellet was re-suspended in 100μl Tris-EDTA buffer and sonicated in an ultra-sonic bath for 45 minutes. The resultant solution was centrifuged at 12,000 r.p.m for 10 minutes to obtain the supernatant that contained DNA. These were temporarily stored at -20°C prior to PCR analysis (Joseph Sambrook et al., 1989).

4.1.6 Genomic deletions (Regions of difference) analysis.
The target gene loci, primer names, primer sequences, amplification product sizes and their genome locations in *M. tuberculosis* H37Rv are listed in **Table 4.1** (Huard et al., 2006; Huard et al., 2003). IS6110 PCR was conducted to confirm that the samples belonged to the *Mycobacterium tuberculosis* complex (Thorne et al., 2011; van Soolingen, Hermans, de Haas, Soll, & van Embden, 1991). IS6110 is insertion element present in the DR region of most *M. tuberculosis* strains (Kamerbeek et al., 1997). Briefly, a master mix consisting of PCR water (Qiagen®), 2xTaq master mix (Qiagen®), forward and reverse primers P₄₃ (5’ –TCA GCC GCG TCC ACG CCG CCA -3’) and P₅₃ (5’ –CCG ACC GCT CCG ACC GAC GGT -3’), was used. The PCR program entailed an initial denaturation at 95°C for 5 minutes, 28 cycles of denaturation at 95°C for 1 minute; annealing at 55°C for 30 seconds; extension at 72°C for 30 seconds, and final extension at 72°C for 10 minutes. The amplicons were electrophoresed on a 1.5% agarose gel stained with ethidium bromide and visualized on UV bio-imager®. *Mycobacterium tuberculosis* strains H37Rv and *M. bovis* JN55 (Asiimwe et al., 2009) were positive controls and nuclease free water was the negative control. Confirmed MTC positive samples were analysed for RD9, RD12, RD4 and RD1 to identify any *M. bovis* amongst the MTC species. Master Mixes consisted of PCR water (Qiagen®), 2xTaq master mix (Qiagen®), forward and reverse primers specific for respective regions of difference (Table 4.1). PCR was performed with an initial denaturation at 95°C for 5 minutes followed by 28 cycles of denaturation at 95°C for 1 minute, annealing at 55°C for 30 seconds, extension at 72°C for 30 seconds, and final extension at 72°C for 10 minutes. Amplicons were resolved on a 1.5% agarose gel and the sizes of the amplicons were estimated by comparison against a100bp DNA ladder. *M. tuberculosis* strains H37Rv and *M. bovis* JN55 were used as the positive controls and nuclease free water as the
negative control. A negative PCR result for RD9, RD12 and RD4 and a positive reaction for RD1 confirmed the presence of *M. bovis* (Huard et al., 2003).

Table 4.1: Nucleotide sequence of primers used to determine the presence of *Mycobacterium bovis*

<table>
<thead>
<tr>
<th>Primer type and target locus</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16SrRNA</td>
<td>5’ ACG GTG GGT ACT AGG TGT GGG TTT C 3’</td>
<td>5’ TCT GCG ATT ACT AGC GAC TCC GAC TTC A3’</td>
<td>543</td>
</tr>
<tr>
<td>IS6110</td>
<td>5’ TCA GCC GCG TCC ACG CCG CCA 3’</td>
<td>5’ CCG ACC GCT CCG ACC GAC GGT 3’</td>
<td>786</td>
</tr>
<tr>
<td>RD9 (Rv2073c)</td>
<td>5’ TCG CCG CTG CCA GAT GAG TC3’</td>
<td>5’ TTT GGG AGC CGC CGG TGG TGA TGA3’</td>
<td>600</td>
</tr>
<tr>
<td>RD4 (Rv1510)</td>
<td>5’ GTG CGC TCC ACC CAA ATA GTT GC3’</td>
<td>5’ TGT CGA CCT GGG GCA CAA ATC AGT C3’</td>
<td>1,033</td>
</tr>
<tr>
<td>RD12 (Rv3120)</td>
<td>5’ GTC GGC GAT AGA CCA TGA GTC CTC CAT3’</td>
<td>5’ GCG AAA AGT GGG CGG ATG CCA GAA TAG T 3’</td>
<td>404</td>
</tr>
<tr>
<td>RD1 (Rv3877/8)</td>
<td>5’ CGA CGG GTC TGA CGG CCA AAC TCA TC3’</td>
<td>5’ CTT GCT CGG TGG CCG GTT TTT CAG C3’</td>
<td>999</td>
</tr>
</tbody>
</table>

4.1.7. Genomic analysis - Sequencing, assembly, and annotation of 8 genomes.

Ten pyrazinamide resistant isolates were selected and shipped to the Broad Institute, USA for whole genome sequencing, WGS (Broad, 2016). Eight of the 10 mycobacterial DNA extracts were subjected to whole-genome shotgun sequencing from a paired-end library at the Broad Institute using the Illumina HiSeq 2000 platform (GA Pipeline version RTA1.17.21.3) (The Olive, 2014). Draft genome sequences and reference genomes *Mycobacterium bovis* Af2122/97 and *M. tuberculosis* CDC 1551 were downloaded from NCBI at ftp://ftp.ncbi.nih.gov/genomes/ in FastA format for genome alignment with Progressive Mauve (A. C. Darling, Mau, Blattner, & Perna, 2004; A. E. Darling, Mau, &
Perna, 2010). Close neighbor references from the Mycobacterium tuberculosis complex were used to identify orthology. The Progressive Mauve software was used to align homologous regions of the genomes in order to identify evolutionary changes in their DNA (A. E. Darling et al., 2010). This software constructs multiple genome alignments in the presence of large-scale evolutionary events like rearrangement and inversions and uses these multiple genome alignments to provide a basis for comparative genomics and the study of genome-wide evolutionary dynamics.

Single nucleotide polymorphisms (SNPs) were extracted from progressive Mauve genome alignments and used to infer phylogeny with phylogenetic tree construction using Mega 7.0 (Molecular Evolutionary Genetic Analysis software) using maximum likelihood method with 1000 bootstrap replicates under the Hasegawa-Kishino-Yano (HYK) model (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). All the genome sequences were re-annotated with RAST (Rapid Annotation using Subsystem Technology), which is a fully automated server for complete or draft archeal and bacterial genomes (Aziz et al., 2008). Comparisons were done using BLAST (Basic Local Alignment Search Tool) at https://blast.ncbi.nlm.nih.gov/Blast.cgi and the results were displayed as graphical maps showing sequence features, gene and protein names, COG category assignments, and sequence composition characteristics (Aziz et al., 2008; National Center for Biotechnology, 2016a). Further, the presence of the following antibiotic resistance genes, pncA, KatG, rpoB, inhA, embABC, gyrA and 16S rRNA (rrs) from the reference strain M. tuberculosis H37Rv were determined using the NCBI tBLASTx tool (translated nucleotide using a translated nucleotide query) where each gene served as a query sequence for alignment against each of the different genomes in this study. This was to determine the exact location of the drug-resistance associated
mutations in the different genomes.

4.2 RESULTS

4.2.1 Culture results.

Out of 5,676 stored MTC isolates, only 136 (2.4%) were resistant to PZA, a first line anti-TB drug. These 136 isolates were retrieved and verified as MTC based on typical growth on media with and without antibiotics, positive results with Ziehl Neelsen and auramine staining and microscopy, and no growth on blood agar. Selective amplification of the *Mycobacterial* IS6110 gene, an MTC specific insertion sequence, confirmed the presence of MTC in 133 of 136 DNA samples. Three of the samples gave indeterminate results hence were removed from further processing steps. The age and gender of individuals from whom isolates were drawn is in Figure 2 and the sample sources are indicated in Figure 3 of the Supplementary Figures. The isolates were all de-identified.

4.2.2 Regions of Difference (RD) analysis

Of the 133 IS6110 positive samples subjected to RD analyses, 126 were RD9 positive while 7 (5.3%) samples were negative. The RD9 positive results suggested that the isolates were either *M. tuberculosis*, *M. canetti*, or *M. africanum* subtype II and this was demonstrated by the presence of a 600bp band on the 1.5% agarose gel. These were not analyzed further. The RD9 negative isolates were further analysed with RD4 to further delineate the MTC species therein. Of the seven RD9 negative samples analysed with RD4, five were RD4 positive and two RD4 negative. RD4 is a region deleted from all *M. bovis* strains (that is classical *M. bovis*, *M. bovis* BCG and MOTTs) but is present in *M. canettii*, *M. tuberculosis*, *M. africanum* type I and II and *M. microti* isolates. To confirm
presence of either Classical *M. bovis* or *M. bovis* BCG, the RD4 negative samples were analysed further using both RD12 and RD1 PCR. The two RD4 negative samples were analyzed for the presence of RD12 and RD1 loci. Both isolates were RD12 and RD1 negative confirming them as *M. bovis*. These samples were further characterized by genome sequencing. (See **Figure 4.1** below for RD9 gel-image).

**Figure 4.1**: PCR result for RD9 (600bp) for some of the 133 IS6110 positive samples. Order on gel: 100 bp Lad, +C: Positive control (M. tuberculosis strain H37Rv), samples 1-12, -C; negative control (nuclease free PCR water).

**Table 4.2** captures a summary of the different RD analyses. The prevalence of *M. bovis* in archived PZA samples at JCRC was determined to be at 1.5 % (2/136). The two *M. bovis* positive samples were **Bz_31150** and **B2_7505**. Sample Bz_31150 was a bronchial wash from a chimpanzee submitted for characterization to the JCRC in May 2007. This sample was sensitive to streptomycin, isoniazid, rifampicin and ethambutol, first line anti-tuberculosis drugs. It was resistant to pyrazinamide, which is expected for *M. bovis* since this organism has inherent resistance to this drug (Huard et al., 2003; Thorne et al., 2011; Wedlock et al., 2002). Another sample taken for sequencing was **B2_7505**, a sputum sample that was a classical *M. bovis* isolate, from a human patient.
collected in 2006. This person was part of a multi-resistant TB drug study. This isolate was resistant to all the anti-tuberculosis antibiotics.

Table 4.2: PCR results for 136 PZA resistant *Mycobacterium tuberculosis* complex

<table>
<thead>
<tr>
<th>Regions of difference (RD)</th>
<th>Suspected MTC</th>
<th>Frequency</th>
<th>Total tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA</td>
<td><em>Mycobacteria</em></td>
<td>133</td>
<td>136</td>
</tr>
<tr>
<td>IS 6110</td>
<td><em>Mycobacterium tuberculosis</em> complex</td>
<td>133</td>
<td>136</td>
</tr>
<tr>
<td>RD9</td>
<td>MTC other than <em>M. africanum</em> and <em>M. bovis</em></td>
<td>126</td>
<td>133</td>
</tr>
<tr>
<td>RD4</td>
<td><em>M. bovis/M. bovis BCG/M. caprae</em></td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>RD12</td>
<td><em>M. bovis/M. bovis BCG</em></td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>RD1</td>
<td><em>M. bovis</em></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

4.2.3 Phenotypic Drug susceptibility profiles.

All isolates were resistant to at least one anti-TB drug ([Figure 4 in Supplementary figures](#)). In [Figure 4 suppl](#) less than 30% of the isolates were sensitive to the first line TB drugs and eighty-three (61%) out of 136 of the isolates were resistant to all the first line TB drugs (e.g. Isoniazid, Rifampicin, Pyrazinamide and Ethambutol) and Streptomycin. These patients were characterized as having multi-drug resistant TB or MDR-TB. Twenty-six or 19% of the isolates were sensitive to all first line TB drugs except for PZA. In addition, 44 or 32% of the isolates were resistant to all the four first line TB drugs and streptomycin. The drug resistance profiles are in Table 4.3 in the appendices.

4.2.4 *Mycobacterium bovis* from human host among sequenced genomes.
Eight genomes sequenced by Illumina are uploaded at NCBI (National Center for Biotechnology, 2016b, 2016c) and the Olive database (The Olive, 2014); their sample names, sources, type and Genbank accession numbers are listed in Table 4.4. Two of these isolates were *M. bovis* and six were *M. tuberculosis*. The draft genomes were assembled at the Broad Institute using the allpaths v. R48559 assembly method. Whole-genome shotgun sequencing from a paired-end library was performed using Illumina HiSeq2000 platform from the Broad Institute (GAPipeline version RTA1.17.21.3) with 1.1 M spots yielding 214.9 Mb for *M. bovis* BZ 31150, and 1.8 M spots yielding 356.7 Mb for *M. bovis* B2 7505. For *M. bovis* Bz 31150, Illumina reads were assembled into 141 contigs and 136 scaffolds, resulting in a genome size of ~4.272 Mb with a GC content of 65.54% at 24x coverage (Wanzala et al., 2015). For *M. bovis* B2_7505 Illumina reads were assembled into 243 contigs and 206 scaffolds, resulting in a genome size of ~4.17 Mb, with a G+C content of 65.5% at 31.0x coverage (Wanzala et al., 2015). The 8 draft genomes together with 9 other genomes from different regions of the world were downloaded from the NCBI ftp database and aligned with Progressive Mauve (A. C. Darling et al., 2004; A. E. Darling et al., 2010). Single nucleotide polymorphisms (SNPs) were extracted to generate a phylogenetic tree using MEGA 7.0 (Tamura et al., 2013). These were aligned against the reference strain (*M. bovis* AF2122/97) to determine phylogenetic relatedness. Figure 4.2 shows the Mauve genome alignment outputs.

**Table 4.3: Mycobacterial genomes with Genbank accession numbers**

<table>
<thead>
<tr>
<th>Bio-Project ID</th>
<th>Bio-Sample ID</th>
<th>Genome/Genbank Accession</th>
<th>Sample Name</th>
<th>Sample Source</th>
<th>Sample Type</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRJNA233395</td>
<td>SAMN02567763</td>
<td>JKAK00000000</td>
<td>Mycobacterium tuberculosis D 4155</td>
<td>Pleural biopsy</td>
<td>Human</td>
<td>Baseline</td>
</tr>
<tr>
<td>PRJNA233396</td>
<td>SAMN02567764</td>
<td>JKAJ00000000</td>
<td>Mycobacterium tuberculosis Kc 32216</td>
<td>Isolate</td>
<td>Human</td>
<td>Baseline</td>
</tr>
<tr>
<td>PRJNA233392</td>
<td>SAMN02567760</td>
<td>JKAN00000000</td>
<td>Mycobacterium</td>
<td>Sputum</td>
<td>Human</td>
<td>Baseline</td>
</tr>
</tbody>
</table>
### 4.2.5. 7200 SNPs obtained by phylogenetic analysis.

In total **7248 SNPs** were extracted across the 17 genomes (Table 4.5) and used to create a phylogenetic tree in MEGA 7.0 using maximum parsimony method with 1000 bootstrap replicates and the Hasegawa-Kishino-Yano (HYK) model (Figure 4.2 shows the Figtree alignment outputs. This is also in **Suppl Figure 7** in appendices).
Figure 4.2: Progressive Mauve alignment of 17 genomes. Each alignment display is organized into one horizontal "panel" per input genome sequence. Each genome's panel contains the name of the genome sequence, a scale showing the sequence coordinates for that genome, and a single black horizontal center line. The 17 genomes are aligned with the reference genome at the very top of figure 8 (M. bovis Af2122/97 here written as Af.fasta/97). Colored block outlines appear above and below the center-line. Each of these block outlines surrounds a region of the genome sequence that aligned to part of another genome, and is presumably homologous and internally free from genomic rearrangement. When a block lies above the center-line the aligned region is in the forward orientation relative to the first genome sequence. Blocks below the center-line indicate regions that align in the reverse complement (inverse) orientation. Regions outside blocks lack detectable homology among the input genomes. Inside each block Mauve draws a similarity profile of the genome sequence. The
height of the similarity profile corresponds to the average level of conservation in that region of the genome sequence. Areas that are completely white were not aligned and probably contain sequence elements specific to a particular genome. The height of the similarity profile is calculated to be inversely proportional to the average alignment column entropy over a region of the alignment (36, 44).

Table 4.4: Global Isolates for phylogenetic analysis

<table>
<thead>
<tr>
<th>ISOLATE</th>
<th>HOST</th>
<th>COUNTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Af2122_97 M. bovis reference strain</td>
<td>Cattle</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>M. bovis B2 7505</td>
<td>Human</td>
<td>Uganda</td>
</tr>
<tr>
<td>M. bovis BZ 31150</td>
<td>Chimpanzee</td>
<td>Uganda</td>
</tr>
<tr>
<td>M. tuberculosis WT 21231*</td>
<td>Human</td>
<td>Uganda</td>
</tr>
<tr>
<td>M. tuberculosis WT 21419*</td>
<td>Human</td>
<td>Uganda</td>
</tr>
<tr>
<td>M. tuberculosis KC_32216</td>
<td>Human</td>
<td>Uganda</td>
</tr>
<tr>
<td>M. tuberculosis KC 9614*</td>
<td>Human</td>
<td>Uganda</td>
</tr>
<tr>
<td>M. tuberculosis D 4155*</td>
<td>Human</td>
<td>Uganda</td>
</tr>
<tr>
<td>M. tuberculosis MR 4387*</td>
<td>Human</td>
<td>Uganda</td>
</tr>
<tr>
<td>MBO_4303- Mycobacterium bovis strain 04-303</td>
<td>Wild boar living on free-range</td>
<td>Argentina</td>
</tr>
<tr>
<td>MBAN5v1= Mycobacterium bovis Strain AN5</td>
<td>Used for Production of Purified Protein Derivative- Brazil</td>
<td>Brazil</td>
</tr>
<tr>
<td>ASM_93432v1=Mycobacterium bovis strain SP38</td>
<td>Isolated from the lungs of a bovine</td>
<td>Sao Paolo, Brazil</td>
</tr>
<tr>
<td>NE_Elk</td>
<td>Elk, Cervus Canadensis</td>
<td>Minnesota, USA</td>
</tr>
<tr>
<td>MBO_Consentino</td>
<td>Elk, Cervus canadensis</td>
<td>Minnesota, USA</td>
</tr>
<tr>
<td>MBO_Elk_94_0704-Mycobacterium bovis</td>
<td>Elk, Cervus Canadensis</td>
<td>(USDA/APHIS/VS/NVSL)</td>
</tr>
<tr>
<td>M. tuberculosis KZN strain (Kwazulu Natal Isolate)</td>
<td>Human</td>
<td>South Africa</td>
</tr>
<tr>
<td>M. tuberculosis CDC 1551</td>
<td>Human</td>
<td>USA</td>
</tr>
</tbody>
</table>
Phylogenetic analysis of SNP distribution on genome sequences was carried out to prove relationships between the isolates and reference genomes as well as likely disease transmission patterns. In our analysis, *M. bovis* from a human being (B2_7505) was distinct in that it formed its own cluster and neither clustered with any *M. bovis* or *M. tuberculosis* strains, as seen in Figure 4.3. See supplementary Figure 5 for SNPs extracted by progressive Mauve.

![Molecular Phylogenetic analysis of Mycobacterium tuberculosis complex organisms by the Maximum Likelihood method.](image)

Figure 4.3 Molecular Phylogenetic analysis of Mycobacterium tuberculosis complex organisms by the Maximum Likelihood method. Evolutionary history was inferred by using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano model (94). The tree with the highest log likelihood (-49783.9307) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to
scale, with branch lengths measured in the number of substitutions per site. The analysis involved 17 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 7248 positions in the final dataset.

In addition, we carried out genome annotation of all the Ugandan genomes and comparative analysis of the two M. bovis genomes (Bz_31150 and B2_7505) with the reference genome M. bovis Af2122/97, M. bovis BCG and M. tuberculosis CDC_1551 with the RAST tool (Aziz et al., 2008) (See Suppl. Figure 6). Portions of the genomes that are 100% similar are indicated in dark purple. There was a distinct white region that run across all the M. bovis genomes but was absent at the same location of the M. tuberculosis CDC_1551 genome. This is an M. bovis specific region or RD9 which is deleted in M. bovis strains and present in M. tuberculosis strains. These results further demonstrate the relatedness of the M. bovis strains and though, M. bovis B2 7505 from the human host formed its own cluster during phylogenetic analysis, it still retains its characteristic M. bovis identity as shown by the RAST map in Suppl. Figure 6.

4.2.6 Genome analysis of multi-drug resistance mutations.

Mycobacteria lack diversity and this implies that when amino acid polymorphisms are detected, it strongly suggests that the variation has functional importance like antibiotic resistance (Sreevatsan, Pan, Stockbauer, et al., 1997). All drug resistance-related mutations of the genomes analyzed in this study are listed in Tables 4.6 and 4.7. All the Ugandan isolates were phenotypically resistant to PZA but on genome analysis, they had mutations in different locations of the 561 bp pncA gene. The animal isolates (M. bovis Af2122/97, M. bovis Bz_31150, M. bovis B2 7505, M. bovis Elk strain, and M. bovis Corsentino strain) and the human isolate M. bovis B2_7505 had the same pncA single C→G point mutation at nucleotide 169 which resulted in an H57D substitution.
(Sreevatsan, Pan, Zhang, Kreiswirth, & Musser, 1997). In addition, both B2_7505, the human TB isolate, and Bz_31150, the chimpanzee isolate had an adenine residue at nucleotide 285 in the oxyR gene (oxyR is a central regulator of the peroxydative and nitrosative stress response dysfunctional in the MTBC species) further confirming them as *M. bovis* isolates (Sreevatsan et al., 1996).

**Table 4. 5: Drugs, genes and drug-resistance associated mutations**

<table>
<thead>
<tr>
<th>DRUG</th>
<th>GENE</th>
<th>GENE PRODUCT</th>
<th>MUTATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>inhA</td>
<td>NADH-dependent enoyl-(ACP) reductase</td>
<td>S94A</td>
</tr>
<tr>
<td>RMP</td>
<td>rpoB</td>
<td>DNA directed RNA polymerase subunit beta</td>
<td>D435Y, D435V, S450L, M390T</td>
</tr>
<tr>
<td>PZA</td>
<td>pncA</td>
<td>Pyrazinamidase/nicotinamidase pncA</td>
<td>H57D, P54Q, A134V, V163A, V131F</td>
</tr>
<tr>
<td>STR</td>
<td>rpsL</td>
<td>rpsL 30S ribosomal protein S12</td>
<td>None seen</td>
</tr>
<tr>
<td></td>
<td>rrs</td>
<td>16S ribosomal RNA</td>
<td>R309C, R338G, R173C</td>
</tr>
<tr>
<td>EMB</td>
<td>embA</td>
<td>Indolylacetylinositol arabinosyltransferase A</td>
<td>G884D, T608N</td>
</tr>
<tr>
<td></td>
<td>embC</td>
<td>Indolylacetylinositol arabinosyltransferase C</td>
<td>T270I, V52X, M800V</td>
</tr>
<tr>
<td>OFX</td>
<td>gyrA</td>
<td>DNA gyrase subunit A</td>
<td>E21Q, S95T, G668D, D639A, R418W, D199G, T80A</td>
</tr>
</tbody>
</table>

**Table 4. 6: SNPs identified in antibiotic resistance associated regions in *Mycobacterium tuberculosis* complex isolates**

<table>
<thead>
<tr>
<th>MUTATIONS IN TARGET GENE (CORRESPONDING DRUGS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>---------</td>
</tr>
<tr>
<td>M. bovis</td>
</tr>
<tr>
<td>M. tuberculosis</td>
</tr>
</tbody>
</table>
Two Ugandan MTb strains (*M. tuberculosis* D_4155 and *M. tuberculosis* KC_32216) were phenotypically PZA resistant but exhibited no mutation at the genome level while *M. tuberculosis* KC_9614 had a mutation 4 amino acids upstream of the *pncA* gene Y(-4)C. This agrees with Sreevatsan *et al.*'s, (Sreevatsan, Pan, Zhang, et al., 1997) work where 28% of the PZA resistant *M. tuberculosis* organisms they tested, lacked nucleotide sequence changes in *pncA*. The rest of the PZA resistant *M. tuberculosis*
organisms displayed unique single point mutations; WT_21419 had a V163A mutation, MR_4387 had an A134V mutation while WT_21231 had a pncA V131F mutation. AhpC is an alkyl hydroperoxide reductase C, an enzyme that reduces organic peroxides, and it’s been shown to be involved in the response to oxidative stress in mycobacterial species. The ahpC gene is located next to oxyR gene and there were mutations seen in the ahpC gene in two isolates; in MR_4387 there were G(-88)A and T(-76)A nucleotide mutations while in D_4155 there was a G(-88)A mutation.

More than 95% of rifampicin (RMP) resistance is associated with mutations in an 81bp region of the rpoB gene (the RIF resistance determining region (RRDR) from codons 507 to 533 of rpoB) but isoniazid (INH) resistance is more complex and has been associated with multiple genes, but most frequently with KatG and inhA (Seifert, Catanzaro, Catanzaro, & Rodwell, 2015). Other common mutation patterns are found at codon 315 (55-90%) of KatG and codon 43 (47-79%) of rpsL for streptomycin resistance (Brammacharry & Muthaiah, 2014). Mutations in the KatG and inhA genes together may be responsible for 80% of all organisms resistant to INH (Sreevatsan, Pan, Zhang, et al., 1997). In our work, M. tuberculosis KC_9614 had the S315T mutation in the KatG gene (catalase-peroxidase) while M. bovis B2_7505 had both the S315T mutation in KatG as well as the S94A mutation in the inhA gene (NADH-dependent enoyl-reductase) defining it as a truly INH resistant organism since inhA is an essential gene and any mutations in this coding region are rare making it sufficient to confer INH and Ethambutol (EMB) resistance in mycobacterial clinical isolates (Vilcheze & Jacobs, 2014).

Phenotypically, M. bovis B2_7505, and M. tuberculosis isolates MR_4387, KC_9614, WT_21419, and WT_21231 were all resistant to rifampicin (RMP) yet the mutations they
exhibited at the genome level fell outside the RRDR (e.g. S450L, D435Y) which is meant to convey high level resistance (Gagneux, 2012). Gagneux et al., suggest that rifampicin resistance-conferring mutations associated with no- or low fitness cost in vitro are the most frequent in clinical strains (Gagneux, 2012). Furthermore, there was an rpoB S531L mutation in WT_21231 and WT_21419 and a D516V mutation for MR_4387, which are classic mutations associated with high level RMP resistance (Dominguez et al., 2016). Several isolates including the reference strain M. bovis Af2122/97 had the R463L mutation in KatG but according to Sekiguchi et al., the R463L mutation is not known to be associated with INH resistance and is more likely to be present in INH-sensitive strains than INH-resistant one (Sekiguchi et al., 2007; Vilcheze & Jacobs, 2014). This means that the phenotypic resistance to INH observed in some of the isolates (MR_4387, WT_21419 and WT_21231) was probably due to mutations in other genes.

Mutations in the gyrA gene related to fluoroquinolone (FQ) resistance were seen in several isolates with the most common being E21Q, G668D and T80A; all these mutations were outside the QRDR yet these isolates demonstrated phenotypic resistance (Avalos et al., 2015). Isolates B2_7505, Bz_31150, D_4155 and MR_4387 had a C206G nucleotide mutation in gyrA. In addition, M. bovis B2_7505 also had a S95T mutation in the QRDR. Fluoroquinolone resistance has been associated with alterations in DNA gyrase, especially mutations in the Quinolone resistance determining region (QRDR) of gyrA (codons 74 to 113) and gyrB (codons 500 to 538) respectively (Avalos et al., 2015; Malik, Willby, Sikes, Tsodikov, & Posey, 2012). Its estimated that 60-90% of Mtb clinical isolates with extremely high FQ resistance have mutations in the QRDR of gyrA, particularly codons 88, 90, 91 and 94 (Avalos et al., 2015).
For ethambutol resistance, a G884D mutation was seen for \textit{embA} in MR\_4387 and D\_4155, Mtb isolates. In addition, there was a T608N mutation in the \textit{embA} gene in KC\_9614. None of these are documented mutations in literature for EMB resistance. The mutations for \textit{embC} were only in the \textit{M. bovis} isolates except for B2\_7505 (see Tables 4.6 and 4.7).

4.3. DISCUSSION

Infections with \textit{Mycobacterium tuberculosis} complex organisms are fraught with a continuum of symptoms in their hosts, ranging from latency to fully-fledged tuberculosis. Although much has been studied on tuberculosis in developed parts of the world and Asia, there are still gaps in the knowledge on tuberculosis-particularly zoonotic tuberculosis-in Africa. In this study, we carried out comparative genomics on archived pyrazinamide resistant MTC isolates from a multicenter study archive in Uganda to determine if there was a non-random distribution of genomotypes of MTC organisms infecting humans and animals in Uganda with the use of a defined Single Nucleotide Polymorphism (SNP) set for differentiation of the MTC organisms. The guiding hypothesis was whether there was multi-directional transmission of tuberculosis occurring in resource-limited settings where animal-human contact is intimate and frequent.

Cattle are the main hosts for bTB in Uganda and the region, though there are studies that indicate that goats can be a source of infection as well but TB in small ruminants is rarely detected at the abattoir as a result of a lower quality of meat inspection than that performed for cattle (Pesciaroli et al., 2014). In Uganda, milk is usually boiled but raw
and fermented milk are also consumed by some pastoral communities (Kankya et al., 2010; Oloya, Kazwala, et al., 2007; Oloya, Muma, et al., 2007; Oloya et al., 2006). With the absence of an active bTB surveillance program in Uganda, there is still little known about the scale of the problem in both the animal and the human population. Pasteurized milk is unaffordable for many people, and this coupled with the consumption of raw milk among the pastoral communities, raises important questions about the actual bovine TB status of the country.

Bovine TB generally results in extra-pulmonary lesions. Thus, a bias toward detection on a very small proportion of *M. bovis* was anticipated as about 80% of the PZA resistant isolates came from sputum. There are limited laboratory facilities for culture and typing of tubercle bacilli in Uganda especially in rural areas (Cosivi et al., 1998b; Cousins, 2001). Sputum smear microscopy with acid-fast stains is the primary approach to pulmonary TB is detection in Ugandan TB laboratories. It is an easy, fast and inexpensive way to screen for TB cases but has low sensitivity especially with persons co-infected with HIV. Furthermore, *M. bovis infections* are generally missed when microscopy on sputum alone is used for diagnostics and as a triage for culturing for mycobacteria.

The World Health indicates that bovine TB is a neglected disease (World Health, 2015) and given its insidious nature, one that policy makers need to address. There is still little published information about the relative proportion of MTC due to *M. bovis*, the epidemiology as well as the public health aspects of this disease (Brudey et al., 2006; Cadmus et al., 2006; Cleaveland et al., 2007; Cosivi et al., 1998b). In fact, Ayele *et al.*, (Ayele, Neill, Zinsstag, Weiss, & Pavlik, 2004) proposed that of all the reports submitted
to the OIE and WHO by different African member countries, none mention bTB as an important human TB infection. Tanzania, one of the countries neighboring Uganda in East Africa, is one of the few countries in Africa that has quantitative data on bTB prevalence in humans with more than 30% of recorded TB cases listed as extrapulmonary in the Arusha region (Cleaveland et al., 2007; Etter et al., 2006); of these *M. bovis* was isolated from 7/65 or 10.8% culture positive cases with cervical lymphadenitis. This also highlights the critical need for physicians to be aware of *M. bovis* and the different routes of transmission and different treatment recommendations.

### 4.3.1. *Mycobacterium bovis* in PZA resistant isolates.

We carried out culture, regions of difference (RD) PCR analysis as well as whole-genome sequencing on archived pyrazinamide resistant isolates from Uganda and found two *M. bovis* isolates; *M. bovis B2_7505* isolated from a human patient and *M. bovis Bz_31150* isolated from a chimpanzee. Out of the over 5,000 stored isolates at the JCRC, only 133 or 2.4% were resistant to pyrazinamide, and only two of these turned out to be *M. bovis*. We expected a higher number of isolates to be positive given that *M. bovis* is naturally resistant to PZA, a first line anti-tuberculosis drug (Wedlock et al., 2002). The rest of the PZA resistant isolates were *M. tuberculosis* pointing to a high degree of drug resistance to anti-TB drugs amongst the humans from whom the isolates were collected (*Supplementary Figure 4*). In addition to resistance to PZA, the case patients were resistant to multiple anti-tuberculosis drugs or had multi-drug resistant or MDR-TB.

Studies in Ethiopia on bTB have indicated that a higher prevalence of *M. tuberculosis* in cattle owned by farmers with active TB than in farmers who did not have active
tuberculosis; this suggests possible bi-directional transmission of mycobacterial species between cattle and their owners (Alemayehu et al., 2008; Ameni et al., 2006; Ameni & Erkahun, 2007; Ameni et al., 2013; Ameni et al., 2011; Aylate, Shah, Aleme, & Gizaw, 2013). In Alemayehu’s study (Alemayehu et al., 2008) *M. bovis* was isolated from the sputum of cattle owners and *M. tuberculosis* from cow’s milk, which suggests a mode of bi-directional transmission of mycobacterial species. Studies in Tanzania also indicate that consumption of raw milk is an important risk factor in bTB transmission (Cleaveland et al., 2007; Daborn et al., 1996; Roug et al., 2014; Wedlock et al., 2002). Humans acquire bTB infection through ingestion of infected contaminated milk and milk products and also by inhalation when there is close physical contact between the owner and his cattle, especially at night since in some cases they share shelters with their animals (Alemayehu et al., 2008; Aylate et al., 2013; Berg et al., 2009; O’Reilly & Daborn, 1995). In this study, the person infected with *M. bovis* (sample B2_7505) was an adult female with MDR TB. Although she was received treatment from the JCRC, in Kampala, the capital city of Uganda, we cannot rule out the possibility that she had close contact with animals and the fact that many Ugandans have contact with animals.

4.3.2 Whole Genome Sequencing (WGS) and analysis.

WGS provides unique diagnostic abilities for drug-resistant TB given that it provides information on entire genes and their promoters, expected sensitivity of strains, genetic background, epidemiological data, an indication of risk of laboratory contamination giving a ‘complete genotypic profile’; features that are critical as incidences of multi-drug resistant and extensively drug resistant TB are increasing (Cirillo et al., 2016; Ramirez-Busby & Valafar, 2015). In this work, WGS of eight Ugandan isolates and genome
analysis with Mauve (A. C. Darling et al., 2004), MEGA (Tamura et al., 2013), RAST (Aziz et al., 2008) and the use of the PATRIC database (Wattam et al., 2014) revealed that the two *M. bovis* isolates from Uganda displayed unique characteristics. The *M. bovis* reference genome Af2122/97 has a genome of \( \sim 4.3 \)Mb while the Ugandan isolates were \( \sim 4.272 \) Mb (*M. bovis* Bz 31150) and 4.17 Mb (*M. bovis* B2 7505) (43). Although these are both draft genomes, there is a clear reduction in genome size when compared to the reference strain and the reduction is greatest in B2 7505, the isolate from a human patient. According to some schools of thought, evolution towards pathogenicity is often accompanied by a reduction in genome size, which is compensated in part by gene duplications and diversification (Le Chevalier, Cascioferro, Majlessi, Herrmann, & Brosch, 2014). It has been proposed that genome downsizing of pathogenic bacteria is as a result of adaptation to a pathogenic lifestyle that includes the exploitation of resources of a host organism which might make certain gene functions redundant and subject to gene loss (Cole et al., 1998; Le Chevalier et al., 2014; Moran, 2002). On the other hand, the reduced genome size could be due to the fact that these were draft genomes with some regions of the genomes not covered. Another possible reason is that it is difficult to obtain complete MTBC genome sequences with second generation sequencing technologies like the Illumina Hiseq platform due to the high GC content and repetitive sequences (Votintseva et al., 2017).

Zoonotic *Mycobacterium bovis* B2_7505 is a multi-drug resistant strain that has adapted itself to the human host and perhaps found a way of successfully circumventing its host defenses, persisting in the host and developing resistance through various mechanisms; many of which are yet to be fully understood. Another possibility is that the patient may have had a mixed infection with both *M. tuberculosis* and *M. bovis*. 
4.3.4 Phylogenetic analysis.

Phylogenetic analysis carried out on the Ugandan isolates together with Af2122/97 reference genome, US *M. bovis* strains isolated from Elk (*M. bovis* Elk and Corsentino strains) and isolates from Argentina and Brazil (Figure 4.3) show that Bz_31150 clustered with the other *M. bovis* strains. The other Ugandan strains clustered with *M. tuberculosis* CDC_1551 and in particular *M. tuberculosis* WT_21231 clustered very closely with *M. tuberculosis* KZN, the Kwazulu Natal strain isolated from a patient in South Africa. *Mycobacterium bovis* B2_7505, did not cluster with any of the two main cluster groups above but formed an intermediate group between the two; demonstrating that genome changes most probably occurred as the organism moved from cattle to the human host. Analysis of selected isolates (*M. bovis* B2_7505, *M. bovis* Bz_31150, *M. bovis* BCG, *M. bovis* AF2122/97 and *M. tuberculosis* CDC_1551) using RAST enabled us to visualize the regions of the genomes that were most similar and others that were not. This is in line with what we obtained from the phylogenetic tree analysis using Single nucleotide polymorphisms (SNPs).

This work demonstrates that although MTC have a high degree of nucleotide relatedness, there are significant genomic differences between them, which can be harnessed to make critical decisions affecting tuberculosis treatment and management. SNP-based phylogenetic analysis helped to differentiate the different MTC genomes; this provides a population genetic framework to aid in identifying factors responsible for the wide host range and disease phenotypes of *M. bovis* and its behavior in different host species (Joshi et al., 2012). Our understanding of Tb epidemiology and genomics is constantly being modified and improved with the rapid advances in whole genome
sequencing (Garcia-Betancur, Menendez, Del Portillo, & Garcia, 2012; Jagielski et al., 2016; Salipante et al., 2015; Takiff & Feo, 2015).

4.3.5. Antibiotic resistance genes of Uganda isolates.

Resistance to drugs develops when mutations or chromosomal errors occur, in genes that encode drug targets or drug metabolic pathways, which impacts the effectiveness of the drugs. This is made worse and amplified when a patient is given less than optimal treatment or there’s low adherence to the drugs. In addition, lack of detection and subsequent unsuitable treatment increase the risk transmission of drug resistant tuberculosis and in fact studies have shown that a large proportion of MDR and XDR cases are due to primary transmission (Liu et al., 2014; Seifert et al., 2015).

The \textit{pncA} gene encodes the pyrazinamidase (PZase) enzyme, which converts PZA, a pro-drug, into the active pyrazinoic acid (Whitfield et al., 2015). PZA kills semi-dormant tubercle bacilli that are not affected by other anti-tuberculous drugs and its inclusion in the treatment of tuberculosis shortens the period of therapy from 12-18 months to 6 months (Maslov et al., 2015; Scorpio & Zhang, 1996). In this study, we sequenced 8 genomes from Uganda that were phenotypically PZA resistant but on genome analysis, three of did not have any mutation in the \textit{pncA} gene (Mtb D_4155, Mtb KC_32216 and Mtb KC_9614). Further, there was no mutation in the \textit{M. bovis} Ravenel strain we used yet this mutation (H57D) should be present in this isolate since \textit{M. bovis} strains are naturally resistant to pyrazinamide (Scorpio & Zhang, 1996; Wedlock et al., 2002). This is a lab strain and may have mutated over the course of several passages.
Table 4.6 shows all drug resistance related mutations and in this research and Table 4.7 shows a more detailed study of SNPs related to different TB drug mutations. DNA sequencing studies have shown that mutations occur across the entire length of the \textit{pncA} gene with some SNPs occurring more frequently than others possibly because these SNPs are rooted in ancestral strains (Sreevatsan, Pan, Zhang, et al., 1997; Whitfield et al., 2015). This means that development of a molecular assay for PZA resistance detection would be more challenging compared to other genes like \textit{rpoB}, \textit{gyrA} and \textit{embB} that have well defined resistance causing mutations (Whitfield et al., 2015). Studies also show that mutations in the \textit{pncA} gene can be absent in a small proportion of isolates that are phenotypically resistant to PZA, this they say, could be indicative of an alternative mechanism of PZA resistance other than mutation in the \textit{pncA} gene (Whitfield et al., 2015).

Isoniazid (INH) targets both active and latent tuberculosis and has relatively low cost and toxicity but INH resistance is quite complex and has been associated with multiple genes, usually \textit{KatG} (S315T) which encodes a catalase peroxidase that transforms INH to its active form, and \textit{inhA} (S94A) which encodes a putative enzyme involved in mycolic acid biosynthesis (Aye et al., 2016; Seifert et al., 2015; Vilcheze et al., 2006).

\textit{Mycobacterium bovis} B2_7505, from a human patient, had both phenotypic resistance to INH as well as mutations in \textit{KatG} (S315T) and \textit{inhA} (S94A). Two isolates had the S315T mutation and according to Vilcheze et al., this variant is more often found in MDR_TB patients than in INH mono-resistant clinical isolates hence it may be related to the higher transmission capabilities of this particular strains (Vilcheze & Jacobs, 2014).
Rifampicin (RMP) is another important first line drug for TB treatment and most resistance to this drug arises from mutations in the *rpoB* gene with the most frequent mutations seen at positions 516, 526 and 531 (Khosravi, Goodarzi, & Alavi, 2012). The most common mutation seen in our study was S450L and D435Y. The isolates screened in this study commonly exhibited the following *gyrA* mutations (E21Q, D668D) which, according to Lui *et al.*, (Liu et al., 2014), were also seen in a fluoroquinolone susceptible strain thus indicating that they may not be the source of fluoroquinolone resistance. Isolates B2_7505, Bz_31150, D_4155 and MR_4387 had a C206G nucleotide mutation in *gyrA*. Although FQs are proven to be very effective second-line anti-mycobacterial drugs for the treatment of drug-resistant TB, the increasing rate of drug resistance among them is a concern.

The presence of these mutations (H57D, S315T, S94A) exclusively in the human isolate suggest that these occurred after transmission from cattle. Genome analysis in this study identified *M. bovis* in humans and great apes suggesting transmission from domesticated ruminants in the area due to a dynamic and changing interface, which has created exposure opportunity. The acquisition, mutation or loss of certain sections of genes could provide pathogens with peculiar advantages during infection and transmission and understanding these mechanisms could perhaps shed more light on the secrets these challenging organisms have that makes them so fastidious and successful. This isolate may also be a mixed infection of both *M. bovis* and *M. tuberculosis* given its unique characteristics. Mutations in key genes responsible for drug resistance have ensured that these organisms circumvent MTC therapy and perpetuate themselves in their hosts and in other susceptible hosts. MDR-TB and XDR-TB are a
real threat to the achievements made in the fight against TB; one that needs a formidable arsenal of tools to overcome.

4.3.6 Application of WGS/SNP Chip in the field.

Genome studies suggest that *M. bovis* transmission occurs between cattle, between cattle and humans and potentially between humans (Srivastava et al., 2008; Tschopp et al., 2010; van Soolingen et al., 1991; H. M. Vordermeier, Perez de Val, et al., 2014) as well as other species like wildlife (Banuls et al., 2015; Dos Vultos et al., 2008; Ehrt, Rhee, & Schnappinger, 2015; Garcia-Betancur et al., 2012). Genotypic assays like the Cepheid Xpert MTB/RIF (Cepheid, Sunnyvale, CA, USA) and Hain line-probe (Hain Lifescience, Nehren Germany) assays provide short turnaround time (less than a day) in identification of mycobacteria; but they are not able to detect all resistance conferring mutations (Pankhurst et al., 2016). In our work we used WGS to identify SNP differences in MTC species and determined those that had multidrug resistance genes (Harris & Okoro, 2014; Kjeldsen, Bek, Rasmussen, Prieme, & Thomsen, 2009; Pankhurst et al., 2016; Sabat et al., 2013; Takiff & Feo, 2015). A total of 7250 SNPs were used to generate the phylogenetic tree that enabled us to determine relatedness of the sequenced genomes to reference strains. Future studies include development of an affordable SNP-chip to achieve sensitive, specific and rapid results for MTC detection and typing to enable real time monitoring of interspecies transmission.

4.4. CONCLUSION

Two of the 136 isolates were *Mycobacterium bovis*. This likely reflects the study population at the JCRC. However, a number of the isolates were MDR-TB. The MTC species are strictly clonal and genetic material is only passed vertically from mother to
daughter cells. This clonality facilitates phylogenetic analysis of the MTC complex where variations occur exclusively by genomic mutations like single nucleotide polymorphisms (SNPs), deletions or insertions of small or large pieces of DNA (Schurch & van Soolingen, 2012; Zakham, Aouane, Ussery, Benjouad, & Ennaji, 2012). In our work, we demonstrated the presence of *M. bovis* in a human and wildlife using archived samples from Uganda. This implies that more work in the pastoral communities would present a clearer picture of the TB disease dynamics and possibly the transmission patterns between humans and animals of MTCs. We plan to use a subset of the SNPs we extracted to make a SNP chip for field applications in order to detect MTCS quickly and accurately. In addition, this work encourages a One Health approach between medical and veterinary departments in order to strengthen the health care and disease surveillance systems, and the need for translating science into policy (Hasegawa, Kishino, & Yano, 1985; Kankya et al., 2010; Kazwala et al., 2001).
In this work we sought to discover and validate pathogen specific biomarkers as well as carry out comparative genomics for the diagnosis and tracking of bovine tuberculosis. We did this through the evaluation of pathogen specific peptides; pks5, MB2515c and MB1895c. Our results showed us that biomarker-based diagnostics may offer a specific approach for the early identification of *M. bovis* infection. We also looked at miRNAs as a tool to determine infection status of experimentally infected animals and we are using this information to design a circulating miRNA detection system for *Mycobacterium bovis* detection in animals and possibly humans. Whole genome sequencing of *Mycobacterium tuberculosis* complex isolates led to the discovery of unique drug resistance profiles in an *M. bovis* isolate from a human patient. In addition, phylogenetic analysis indicated the possibility of multiple genome changes of this isolate as it moved from its animal host to the human host.

**Limitations of the study**

The studies in this thesis had some limitations which could have caused some bias. Bias refers to any tendency which prevents unprejudiced consideration of a question and in research, this occurs when systematic error is introduced into sampling or testing by selecting or encouraging one outcome or answer over others (Pannucci & Wilkins, 2010). One of the limitations of this research is observed in the biomarkers and miRNA studies where there were only three controls and eight test animals. This was because the cost of running this kind of experiment is prohibitively high hence the use of a limited number of animals. This can be overcome by having a larger sample size which is ideal
because it reduces the random error due to sampling variability and there’s increased statistical significance of the outcome (Pannucci & Wilkins, 2010).

For the retrospective study in Uganda, the isolates were drawn from a large repository of frozen samples using PZA resistance as the search criteria. The challenge with this is that the information from these samples was from a single data source, the JCRC, yet cross-referencing the results with other associated records would have been useful to reduce discrepancies due to a single data source. Selection bias is another limitation that may have affected this study since it was a retrospective study where the exposure and outcome had already occurred at the time the individuals were selected for study inclusion. In future studies this can be avoided by ensuring that subjects are selected using rigorous criteria to avoid confounding and also by carrying out a prospective study to avoid selection bias since the outcome is unknown at the time of enrollment.

The way information is solicited, recorded and interpreted can be affected by interviewer bias. This occurs when the disease is known to the interviewer and in this case all the persons whose samples were used in the Uganda retrospective study (comparative genomics of pyrazinamide resistance isolates from Uganda) either had TB or were co-infected with HIV and TB; all drawn from 18 different studies (see supplementary figure 1). This study could have been affected by interviewer bias. This study population was also not truly representative of the country’s or region’s population but was a crucial starting point to address the issue of pyrazinamide resistance and zoonotic TB from archived Mycobacterium tuberculosis complex samples. Such bias can be minimized if the interviewer is blinded to the outcome of interest or if the outcome has not yet occurred like in a prospective study (Pannucci & Wilkins, 2010).
Another limitation is that there may have been misclassification of some of the samples. Our search criterion was resistance to pyrazinamide but on isolate retrieval and re-culture, only 2 of the 133 samples were actually *Mycobacterium bovis*. In addition, after whole genome analysis of the results there were a few discrepancies between the physical drug sensitivity tests and the antibiotic resistance results obtained from whole genome analysis data. This implies that though most of these were multi-drug resistant isolates, the fact that they were not characterized up to species level left a lot to speculation on what they could actually have been in the *Mycobacterium tuberculosis* complex group of organisms.

This work is a clear example of the ongoing, complex interactions between humans, animals and the environment. These interactions are fluid and constantly changing with a myriad of outcomes. We believe that some of the results we have uncovered here are just the tip of the iceberg and hence there is need for more research to be carried out in real time-with a prospective study-where these complex interactions can be fully delved into and perhaps provide a clearer understanding of these interfaces.

All in all, the research here-in has shed more light towards this fascinating area of biomarkers for tuberculosis research. The results we have obtained are a rich resource that will be mined for years to come as a source of invaluable information regarding the *Mycobacterium tuberculosis* complex group of organisms- these formidable organisms that present us with classical host-pathogen interactions.

Important recommendations for this work include having a prospective study carried out in a pastoral community in a resource limited setting where the disease is endemic like
Mbarara district in Uganda. This would provide invaluable information on the actual
disease transmission dynamics at the human animal interface. This would inform
research on this neglected zoonosis and provide a platform to inform both animal and
human health practitioners on the importance of cross-sectoral collaboration in
combating this ancient but formidable enemy—tuberculosis.
REFERENCES


101


Doi 10.1016/S0167-5877(02)00010-7


Mycobacterium bovis through milk in pastoralist households of Tanzania. *Preventive Veterinary Medicine, 115*(3-4), 157-165. doi:10.1016/j.prevetmed.2014.04.014


Scaros, O., & Fisler, R. (2005). Biomarker technology roundup: from discovery to clinical applications, a broad set of tools is required to translate from the lab to the clinic. *Biotecniques, Suppl, 30-32.*


Schutz, E., Urnovitz, H. B., Iakoubov, L., Schulz-Schaeffer, W., Wemheuer, W., & Brenig, B. (2005). Bov-tA short interspersed nucleotide element sequences in circulating nucleic acids from sera of cattle with bovine spongiform encephalopathy (BSE) and sera of cattle exposed to BSE. *Clinical and Diagnostic Laboratory Immunology, 12*(7), 814-820.


Retrieved from


APPENDICES
Supplementary Figure 1: JCRC studies from which samples were drawn

- TBTC 29 Extension
- Adult CD80 Study
- AE-TBC Study
- Evaluation of a rifapentine-...
- GeneXpert TB assay
- Direct MGIT
- PZA Study
- Lipa Assay
- PART Study
- Animal Specimens
- Wellcome Trust Study
- Kawempe Community Health Study
- Treatment Shortening Study
- Household Contact Study
- Multidrug Resistant TB Study
- TB Genetics
- Private Patient (JCRC clinic)
- TB Pleural Fluid Study

% of total samples
Supplementary Figure 2: Gender and age of persons with pyrazinamide resistant MTC

Supplementary Figure 3: Source of samples tested for *Mycobacterium tuberculosis* complex
Supplementary Figure 4: Resistance patterns for anti-tuberculosis drugs in PZA resistant isolates

- Pyrazinamide
- Etambutol
- Rifampicin
- Isoniazid
- Streptomycin

Legend: sensitive (green) and resistant (blue)
Supplementary Figure 6: RAST
Supplemental Figure 7:
Phylogenetic analysis of
Mycobacterium tuberculosis
complex organisms by the
Maximum Likelihood method
### TABLE 3.3: miRNA discovered in sera of the host at post-infection time points

<table>
<thead>
<tr>
<th>START/STOP POSITION</th>
<th>DAY 19</th>
<th>DAY 48</th>
<th>DAY 60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TIME POINTS</td>
<td>(SEQUENCE) UPREGULATION X COVERAGE</td>
<td>depth</td>
</tr>
<tr>
<td>30935</td>
<td>AGAAAAACAGAAGGAACCTCC- (miR-6868-5p)</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>820120</td>
<td>GGAGGTTTCTTCAGTTTCT- (pma-miR-146-3p)</td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>1600699</td>
<td>AAGAAAAAGGATCGAACC- none</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>1647808</td>
<td>TAAGATACAGACCCCTATAC- none</td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>2325878</td>
<td>CACTCCAGGCCAAGATCAT- (mmu-miR-8107)</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>2325878</td>
<td>CACTCCAGGCCAAGTCA- (mmu-miR-8107)</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>2437733</td>
<td>GACTCTGTCTGACCACC- (mmu-miR-1966-5p)</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>3372080</td>
<td>AAATTGTCCCTGCTGCTG- (cte-miR-92b)</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>3921894</td>
<td>AGATTTGTAATGGATGTCAC- none</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>4831799</td>
<td>ACCCGGGCTGTGCACTGTAGAGGGAGCTAGGCAGnone</td>
<td></td>
<td>8013</td>
</tr>
<tr>
<td>4831901</td>
<td>ACAACCCGAAGGCTCAGCTGAGGCTAGGTAGACCnone</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>4831901</td>
<td>ACAACCCGAAGGGCCT- none</td>
<td></td>
<td>35</td>
</tr>
<tr>
<td>4831901</td>
<td>ACAACCCGAAGGCT- none</td>
<td></td>
<td>109</td>
</tr>
<tr>
<td>4831926</td>
<td>CGCGCCATGGCTGAGCTGAGCAG- (aly-miR2443-5p)</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>7698695</td>
<td>CTTGAGGGCTGCTGCCACGACC- (gga-miR-1560-3p)</td>
<td></td>
<td>72</td>
</tr>
<tr>
<td>7739002</td>
<td>GTGGTTTGGTTCAGGTGT- none</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>ID</td>
<td>Sequence</td>
<td>Target</td>
<td>Length</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------</td>
<td>-------------</td>
<td>--------</td>
</tr>
<tr>
<td>9434144</td>
<td>GTGCAGTCTGTTGAATTTTA- (gga-miR-7445-3p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10989401</td>
<td>GAATATAATTTTGTGCCACA- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11890401</td>
<td>GAACAAGTTGCACTTTT- (oha-miR-7-2-3p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12025609</td>
<td>GGAGCTGCTCTCGTACTAG- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12025609</td>
<td>GGAGCTGCTCTCGTACTAGA- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12754092</td>
<td>CTCACAACAGGCCCCAAA- (efu-miR-9262)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13800644</td>
<td>CTAAGGTGTTAATGCGGACA- (pxy-miR-8502)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14660986</td>
<td>ACATTCTATCCTTCTTC- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15642751</td>
<td>GCAAATTCCACCGCAGCGT- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16576559</td>
<td>AAGTAGGAAAACGAGTGAG- (osa-miR1868)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17306629</td>
<td>ACAACCTCTTGTATTTTAT- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17376916</td>
<td>TAGAACCTGAAACCTTGTC- (hco-miR87c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18896825</td>
<td>TGGGAAGCTCCATTTCA- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21188003</td>
<td>CAAATTCTCATAAGGTC- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23125748</td>
<td>CGATAGCATGAAGAAGA- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23782349</td>
<td>TCACTCACAGCCACAGTCA- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24095842</td>
<td>GGTGAACACATCCAACGCT- none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Sequence</td>
<td>Target</td>
<td>ID</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------</td>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>24187343</td>
<td>GTAAGCTCTCATCTCCAGT-</td>
<td>(aca-miR-5463)</td>
<td>24231830</td>
</tr>
<tr>
<td>27084486</td>
<td>GCTTTTCAGTGCTGAATA-</td>
<td>(str-miR-8402-5p)</td>
<td>28791899</td>
</tr>
<tr>
<td>28972001</td>
<td>AAATTCAGCAAGCCTTC-</td>
<td>none</td>
<td>29131068</td>
</tr>
<tr>
<td>29991075</td>
<td>ATGAGTCGCTACGCCCTGA-</td>
<td>(mdo-miR-7269-5p)</td>
<td>31842367</td>
</tr>
<tr>
<td>32151312</td>
<td>CACCCACAGCCACAAGTCA-</td>
<td>(efu-miR-9235b)</td>
<td>32151312</td>
</tr>
<tr>
<td>32369316</td>
<td>CGGCTACCACATCCAAGGA-</td>
<td>none</td>
<td>32382428</td>
</tr>
<tr>
<td>32386575</td>
<td>TTCGATTCCGGAGAG-</td>
<td>none</td>
<td>35158124</td>
</tr>
<tr>
<td>35856977</td>
<td>TTTCCACTTTGT TCTGTG-</td>
<td>none</td>
<td>36631043</td>
</tr>
<tr>
<td>Length</td>
<td>Id Number</td>
<td>Sequence</td>
<td>Match to</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>20</td>
<td>37553202</td>
<td>CCAGTACTGAGAAGGAAAGG- hsa-miR-6876-3p</td>
<td>AGGTCTGGAACCTAAAGGA- (bmo-miR-2772a/ dev-miR-D103p/ cfa-miR-8812)</td>
</tr>
<tr>
<td>21</td>
<td>37970060</td>
<td>TCAATGACTTGGCAGGCTC- (dre-miR-25-5p)</td>
<td>TCAATGACTTGGCAGGCTC- (bmo-miR-2772a/ dev-miR-D103p/ cfa-miR-8812)</td>
</tr>
<tr>
<td>29</td>
<td>38014962</td>
<td>TCAATGACTTGGCAGGCTC- (dre-miR-25-5p)</td>
<td>TCAATGACTTGGCAGGCTC- (bmo-miR-2772a/ dev-miR-D103p/ cfa-miR-8812)</td>
</tr>
<tr>
<td>25</td>
<td>38264438</td>
<td>GTGTAACACTCACACAGCCATGC- (gga-miR-1733)</td>
<td>GTGTAACACTCACACAGCCATGC- (gga-miR-1733)</td>
</tr>
<tr>
<td>26</td>
<td>38901633</td>
<td>AGGTCTGGAACCTAAAGGA- (bmo-miR-2772a/ dev-miR-D103p/ cfa-miR-8812)</td>
<td>AGGTCTGGAACCTAAAGGA- (bmo-miR-2772a/ dev-miR-D103p/ cfa-miR-8812)</td>
</tr>
<tr>
<td>20</td>
<td>39048554</td>
<td>CCCCCCGCCGGGACCCGGGC- none</td>
<td>CCCCCCGCCGGGACCCGGGC- none</td>
</tr>
<tr>
<td>38</td>
<td>39375800</td>
<td>TGAGGCAGCCGGGAGAAGCTC- (hsa-miR-3907)</td>
<td>TGAGGCAGCCGGGAGAAGCTC- (hsa-miR-3907)</td>
</tr>
<tr>
<td>20</td>
<td>39404087</td>
<td>AAAGAAGCAGCTCATCAACT- none</td>
<td>AAAGAAGCAGCTCATCAACT- none</td>
</tr>
<tr>
<td>59</td>
<td>41357265</td>
<td>GTTGTGATTGGAACCTCTCA- none</td>
<td>GTTGTGATTGGAACCTCTCA- none</td>
</tr>
<tr>
<td>18</td>
<td>41640096</td>
<td>GGGTCCTGAGCGTGCGGAGCA- (str-miR-7880i-3p)</td>
<td>GGGTCCTGAGCGTGCGGAGCA- (str-miR-7880i-3p)</td>
</tr>
<tr>
<td>21</td>
<td>42493106</td>
<td>CAGGTGGTAACCTGAGAATACA- (ath-miR398b-3p)</td>
<td>CAGGTGGTAACCTGAGAATACA- (ath-miR398b-3p)</td>
</tr>
<tr>
<td>51</td>
<td>42549589</td>
<td>AGGACGGTGGCCATGGA- none</td>
<td>AGGACGGTGGCCATGGA- none</td>
</tr>
<tr>
<td>29</td>
<td>42809280</td>
<td>AACAGAAAGCAGAAGGGGAAA- (cin-miR-4006c-3p)</td>
<td>AACAGAAAGCAGAAGGGGAAA- (cin-miR-4006c-3p)</td>
</tr>
<tr>
<td>24</td>
<td>44836480</td>
<td>TGCCCTGCCGCTCTGGAAG- none</td>
<td>TGCCCTGCCGCTCTGGAAG- none</td>
</tr>
<tr>
<td>28</td>
<td>44919634</td>
<td>TGCTCTGCAACCCACCAAC- (hsa-miR-6855-3p)</td>
<td>TGCTCTGCAACCCACCAAC- (hsa-miR-6855-3p)</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Sequence</td>
<td>Target Gene</td>
<td>Length</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------</td>
<td>----------------------</td>
<td>--------</td>
</tr>
<tr>
<td>45488028</td>
<td>TAGCGATCCCCCGACCTT-</td>
<td>TGACGTGTGGCTGGGGGTGA- (efu-miR-9215a)</td>
<td>81</td>
</tr>
<tr>
<td>47113519</td>
<td>TAGCGATCCCCCGACCTT-</td>
<td>TAAACCTTGGCTACAAATAT- (hma-miR3012b)</td>
<td>64</td>
</tr>
<tr>
<td>48925149</td>
<td>TGAAGAATTGGTTGGGGC-</td>
<td>TAAACCTTGGCTACAAATAT- (hma-miR3012b)</td>
<td>25</td>
</tr>
<tr>
<td>50085385</td>
<td>GAAAACCTTAAAACACTTTCG-</td>
<td>none</td>
<td>39</td>
</tr>
<tr>
<td>50217131</td>
<td>GAAAACCTTAAAACACTTTCG-</td>
<td>none</td>
<td>23</td>
</tr>
<tr>
<td>51500407</td>
<td>TTAGGACTGAATTTACAAT-</td>
<td>(sme-miR-2204-5p)</td>
<td>45</td>
</tr>
<tr>
<td>52278005</td>
<td>TCTTTAAATTCTGTCCGT-</td>
<td>(ppymiR-1324)</td>
<td>25</td>
</tr>
<tr>
<td>54020302</td>
<td>TGCCAGCTGGGACAGAGGAC-</td>
<td>(hsa-miR-6775-3p)</td>
<td>20</td>
</tr>
<tr>
<td>57858012</td>
<td>GGTAGTGCAAGAGTAGGAAA-</td>
<td>(vvi-miR482/ ghr-miR482a/ pabmiR482c/ pde-miR482a/ sly- miR482e-3p)</td>
<td>26</td>
</tr>
<tr>
<td>58414176</td>
<td>CAGCTTTTTCACTCAATAT-</td>
<td>none</td>
<td>23</td>
</tr>
<tr>
<td>62966793</td>
<td>CAGCGTGCAGGCTGGAAGG-</td>
<td>(gga-miR-1556)</td>
<td>24</td>
</tr>
<tr>
<td>65601473</td>
<td>AAGCAGTCTAATGCTAATG-</td>
<td>none</td>
<td>21</td>
</tr>
<tr>
<td>66270926</td>
<td>AAGCAGTCTAATGCTAATG-</td>
<td>none</td>
<td>20</td>
</tr>
<tr>
<td>69048695</td>
<td>CCGTCATAATCGCCTGT-</td>
<td>none</td>
<td>22</td>
</tr>
<tr>
<td>Index</td>
<td>Sequence</td>
<td>Annotation</td>
<td>Length</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------</td>
<td>------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>69596665</td>
<td>CTTGGGTTTGATGGTACATGCC-</td>
<td>(hsa-miR-1234-3p)</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td><strong>had many</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>72156030</td>
<td>TTTCCACCAGCCCCCAAG-</td>
<td>(eca-miR-1905c/hsa-miR-67565p/hsa-miR-6794-5p/mml-miR6794-5p/bta-miR-2374)</td>
<td>22</td>
</tr>
<tr>
<td>72156030</td>
<td>TTTCCACCAGCCCCCAAG-</td>
<td>(eca-miR-1905c/hsa-miR-6756-5p/hsa-miR-6794-5p/mml-miR6794-5p/bta-miR-2374)</td>
<td>21</td>
</tr>
<tr>
<td>72156030</td>
<td>TTTCCACCAGCCCCCAAG-</td>
<td>(eca-miR-1905c/hsa-miR-6756-5p/hsa-miR-6794-5p/mml-miR6794-5p/bta-miR-2374)</td>
<td>40</td>
</tr>
<tr>
<td>72767441</td>
<td>ATCAGCCTCAAGGACAAAAAA-</td>
<td>(asu-miR-5361-3p)</td>
<td>83</td>
</tr>
<tr>
<td>73103361</td>
<td>CAGGTTCTGTGGTTTTT-</td>
<td>(blamiR-2452)</td>
<td>61</td>
</tr>
<tr>
<td>74854288</td>
<td>TCAACCCAGACTTCTTTCT-</td>
<td>(gsa-miR-9402-5p)</td>
<td>37</td>
</tr>
<tr>
<td>75866549</td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>77049811</td>
<td>AGATGTAATGCGCTACTCTC-</td>
<td>none</td>
<td>28</td>
</tr>
<tr>
<td>77154288</td>
<td>GGGGGCCGCCCATAGGACTTG-</td>
<td>none</td>
<td>97</td>
</tr>
<tr>
<td>77861379</td>
<td>AGATGTAATGCGCTACTCTC-</td>
<td>none</td>
<td>36</td>
</tr>
<tr>
<td>77861379</td>
<td>TCTCCAGCATTTCTCAAT-</td>
<td>none</td>
<td>76</td>
</tr>
<tr>
<td>78408214</td>
<td>ATTAGGTTTTTGGAGTTCTTA-</td>
<td>(atr-miR8601)</td>
<td>428</td>
</tr>
<tr>
<td>ID</td>
<td>Sequence</td>
<td>Corresponding miRNA</td>
<td>Mismatches</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------</td>
<td>-----------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>83837860</td>
<td>CAGCAGTTACAGCAGGCAG-None</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>85224212</td>
<td>GGGCTTTGTTCCAAGCTACTG-None</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>89196344</td>
<td>TTCTTTTCTCTCGCCATGA-(mmu-miR-7211-5p/mdo-miR7398f-3p)</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>89995889</td>
<td>TTCCACTTAGTTTCTGT-None</td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>91880441</td>
<td>GTTGAGCAGTAGTATCAT-None</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>97589551</td>
<td>GGGGATAGTGGTAGCTAT-(cfa-miR-8898)</td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>100767650</td>
<td>GATTAGAAAGCAAAAAGTCCCT-(str-miR-81c-5p)</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>101877964</td>
<td>TCCTTGCCAATGTCACCAGA-(sme-miR-71a-1-3p)</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>102296891</td>
<td>GTTCAGTCAATGGAGGTTTCC-none</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>110133024</td>
<td>TGCCAGTTCCCTTAACCCAAGT-(tca-miR-2765-3p)</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>115327848</td>
<td>ATTTTAAAGTTGTTGGTGCG-(gra-miR8659a)</td>
<td><strong>has several</strong></td>
<td>22</td>
</tr>
<tr>
<td>116326826</td>
<td>AAACCTGCACGTTCATCTGTTTCC-none</td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>117049452</td>
<td>CCATACAAAAACCACCTCCA-(cel-miR-4818-5p)</td>
<td></td>
<td>42</td>
</tr>
<tr>
<td>118444808</td>
<td>TTCTAAAGTCCAGCTGTCCA-none</td>
<td></td>
<td>37</td>
</tr>
<tr>
<td>123638608</td>
<td>GCAGAAACCTTTCAACTGTGA-(cin-miR-4060-5p)</td>
<td></td>
<td>147</td>
</tr>
<tr>
<td>150609293</td>
<td>GACAGAGGCGAAGGCTGAAG-None</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Sample type</td>
<td>Drug resistance patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NNNNN</td>
<td>NNNNR</td>
<td>NSSSR</td>
</tr>
<tr>
<td>Blood</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronchial wash</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastric Aspirates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isolate</td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Pleural fluid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sputum</td>
<td>2</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

S=sensitive, R=resistant N-not done. Order of first line drugs and streptomycin: STM-streptomycin, INH-isoniazid, RMP-rifampicin, EMB-ethambutol, and PZA-pyrazinamide. **NNNN**-not done for all five TB drugs. **NNNNR**-not done for STM, INH, RMP, EMB, resistant to PZA. **NSSSR**-Not done for STM; Sensitive to INH, RMP, EMB, Resistant to PZA. **RRRRR**- resistant to all five TB drugs. **RRRSSR**- resistant to all drugs except EMB. **RRSRR**- Resistant to all TB drugs except RMP. **RRSSR**-resistant to STM, INH, and PZA, sensitive to RMP, EMB. **SRRRR**- Resistant to all drugs except STM. **SRRSR**-resistant to INH, RMP, PZA, sensitive to STM and EMB. **SRSSR**, resistant to RMP, PZA, sensitive to STM, RMP, and EMB. **SSSSR**-sensitive to all TB drugs except PZA.